Exploring Transfer Learning via Convolutional Neural Networks for Image Classification and Super-Resolution.

by

Eduardo Ribeiro

Cumulative dissertation submitted to the
Faculty of Natural Sciences, University of Salzburg
in partial fulfillment of the requirements
for the Doctoral Degree.

Thesis Supervisor
Uhl Andreas Univ.-Prof. Mag.rer.nat. Dr.rer.nat.

Department of Computer Sciences
University of Salzburg
Jakob Haringer Str. 2
5020 Salzburg, AUSTRIA

Salzburg, February 2018

Abstract

This work presents my research about the use of Convolutional Neural Network (CNNs) for transfer learning through its application for colonic polyp classification and iris super-resolution.

Traditionally, machine learning methods use the same feature space and the same distribution for training and testing the tools. Several problems in this approach can emerge as, for example, when the number of samples for training (especially in a supervised training) is limited. In the medical field, this problem is recurrent mainly because obtaining a database large enough with appropriate annotations for training is highly costly and may become impractical. Another problem relates to the distribution of textural features in a image database which may be too large such as the texture patterns of the human iris. In this case a single and specific training database might not get enough generalization to be applied to the entire domain. In this work we explore the use of texture transfer learning to surpass these problems for two applications: colonic polyp classification and iris super-resolution.

The leading cause of deaths related to intestinal tract is the development of cancer cells (polyps) in its many parts. An early detection (when the cancer is still at an early stage) can reduce the risk of mortality among these patients. More specifically, colonic polyps (benign tumors or growths which arise on the inner colon surface) have a high occurrence and are known to be precursors of colon cancer development. Several studies have shown that automatic detection and classification of image regions which may contain polyps within the colon can be used to assist specialists in order to decrease the polyp miss rate.

However, the classification can be a difficult task due to several factors such as the lack or excess of illumination, the blurring due to movement or water injection and the different appearances of polyps. Also, to find a robust and a global feature extractor that summarizes and represents all these pit-patterns structures in a single vector is very difficult and Deep Learning can be a good alternative to surpass these problems.

One of the goals of this work is show the effectiveness of CNNs trained from scratch for colonic polyp classification besides the capability of knowledge transfer between natural images and medical images using off-the-shelf pretrained CNNs for colonic polyp classification. In this case, the CNN will project the target database samples into a vector space where the classes are more likely to be separable.

The second part of this work dedicates to the transfer learning for iris super-resolution. The main goal of Super-Resolution (SR) is to produce, from one or more images, an image with a higher resolution (with more pixels) at the same time that produces a more detailed and realistic image being faithful to the low resolution image(s). Currently, most iris recognition systems require the user to present their iris for the sensor at a close distance. However, at present, there is a constant pressure to make that relaxed conditions of acquisitions in such systems could be allowed. In this work we show that the use of deep learning and transfer learning for single image super resolution applied to iris recognition can be an alternative for Iris Recognition of low resolution images. For this purpose, we explore if the nature of the images as well as if the pattern from the iris can influence the CNN transfer learning and, consequently, the results in the recognition process.

Abstract (German)

Diese Arbeit präsentiert meine Forschung hinsichtlich der Verwendung von "Transfer-Learning" (TL) in Kombination mit Convolutional Neural Networks (CNNs), um dadurch die Klassifikation von Dickdarmpolypen und die Qualität von Iris Bildern ("Iris-Super-Resolution") zu verbessern.

Herkömmlicherweise verwenden Verfahren des maschinellen Lernens den gleichen Merkmalsraum und die gleiche Verteilung zum Trainieren und Testen der abgewendeten Methoden. Mehrere Probleme können bei diesem Ansatz jedoch auftreten. Zum Beispiel ist es möglich, dass die Anzahl der zu trainierenden Daten (insbesondere in einem "supervised training" Szenario) begrenzt ist. Im Speziellen im medizinischen Anwendungsfall ist man regelmäßig mit dem angesprochenen Problem konfrontiert, da die Zusammenstellung einer Datenbank, welche über eine geeignete Anzahl an verwendbaren Daten verfügt, entweder sehr kostspielig ist und/oder sich als über die Maßen zeitaufwändig herausstellt. Ein anderes Problem betrifft die Verteilung von Strukturmerkmalen in einer Bilddatenbank, die zu groß sein kann, wie es im Fall der Verwendung von Texturmustern der menschlichen Iris auftritt. Dies kann zu dem Umstand führen, dass eine einzelne und sehr spezifische Trainingsdatenbank möglicherweise nicht ausreichend verallgemeinert wird, um sie auf die gesamte betrachtete Domäne anzuwenden. In dieser Arbeit wird die Verwendung von TL auf diverse Texturen untersucht, um die zuvor angesprochenen Probleme für zwei Anwendungen zu überwinden: in der Klassifikation von Dickdarmpolypen und in Iris Super-Resolution.

Die Hauptursache für Todesfälle im Zusammenhang mit dem Darmtrakt ist die Entwicklung von Krebszellen (Polypen) in vielen unterschiedlichen Ausprägungen. Eine Früherkennung kann das Mortalitätsrisiko bei Patienten verringern, wenn sich der Krebs noch in einem frühen Stadium befindet. Genauer gesagt, Dickdarmpolypen (gutartige Tumore oder Wucherungen, die an der inneren Dickdarmoberfläche entstehen) haben ein hohes Vorkommen und sind bekanntermaßen Vorläufer von Darmkrebsentwicklung. Mehrere Studien haben gezeigt, dass die automatische Erkennung und Klassifizierung von Bildregionen, die Polypen innerhalb des Dickdarms möglicherweise enthalten, verwendet werden können, um Spezialisten zu helfen, die Fehlerrate bei Polypen zu verringern.

Die Klassifizierung kann sich jedoch aufgrund mehrerer Faktoren als eine schwierige Aufgabe herausstellen. Zum Beispiel kann das Fehlen oder ein Übermaß an Beleuchtung zu starken Problemen hinsichtlich der Kontrastinformation der Bilder führen, wohingegen Unschärfe aufgrund von Bewegung/Wassereinspritzung die Qualität des Bildmaterials ebenfalls verschlechtert. Daten, welche ein unterschiedlich starkes Auftreten von Polypen repräsentieren, bieten auch die Möglichkeit zu einer Reduktion der Klassifizierungsgenauigkeit. Weiters ist es sehr schwierig, einen robusten und vor allem globalen Feature-Extraktor zu finden, der all die notwendigen Pit-Pattern-Strukturen in einem einzigen Vektor zusammenfasst und darstellt. Um mit diesen Problemen adäquat umzugehen, kann die Anwendung von CNNs eine gute Alternative bieten.

Eines der Ziele dieser Arbeit ist es, die Wirksamkeit von CNNs, die von Grund auf für die Klassifikation von Dickdarmpolypen konstruiert wurden, zu zeigen. Des Weiteren soll die Anwendung von TL unter der Verwendung vorgefertigter CNNs für die Klassifikation von Dickdarmpolypen untersucht werden. Hierbei wird zusätzliche Information von nicht-medizinischen Bildern hinzugezogen und mit den verwendeten medizinischen Daten verbunden: Information wird also transferiert - TL entsteht. Auch in diesem Fall projiziert das CNN die Zieldatenbank (die Polypenbilder) in einen vorher trainierten Vektorraum, in dem die zu separierenden Klassen dann eher trennbar sind, da Wissen aus den nicht-medizinischen Bildern einfließt.

Der zweite Teil dieser Arbeit widmet sich dem TL hinsichtlich der Verbesserung der Bildqualität von Iris Bilder - "Iris- Super-Resolution". Das Hauptziel von Super-Resolution (SR) ist es, aus einem oder mehreren Bildern gleichzeitig ein Bild mit einer höheren Auflösung (mit mehr Pixeln) zu erzeugen, welches dadurch zu einem detaillierteren und somit realistischeren Bild wird, wobei der visuelle Bildinhalt unverändert bleibt. Gegenwärtig fordern die meisten Iris-Erkennungssysteme, dass der Benutzer seine Iris für den Sensor in geringer Entfernung präsentiert. Jedoch ist es ein Anliegen der Industrie die bisher notwendigen Bedingungen - kurzer Abstand zwischen Sensor und Iris, sowie Verwendung von sehr teuren hochqualitativen Sensoren - zu verändern. Diese Veränderung betrifft einerseits die Verwendung von billigeren Sensoren und andererseits die Vergrößerung des Abstandes zwischen Iris und Sensor. Beide Anpassungen führen zu Reduktion der Bildqualität, was sich direkt auf die Erkennungsgenauigkeit der aktuell verwendeten Iris- erkennungssysteme auswirkt. In dieser Arbeit zeigen wir, dass die Verwendung von CNNs und TL für die "Single Image Super-Resolution", die bei der Iriserkennung angewendet wird, eine Alternative für die Iriserkennung von Bildern mit niedriger Auflösung sein kann. Zu diesem Zweck untersuchen wir, ob die Art der Bilder sowie das Muster der Iris das CNN-TL beeinflusst und folglich die Ergebnisse im Erkennungsprozess verändern kann.

Acknowledgments

In the last three years I have learned so much more than I could have ever imagined. And one of the things that I have been practicing a lot is ”Gratitude”.

Firstly, I would like to thank my advisor Univ.-Prof. Dr. Andreas Uhl not only for accepting me in his group giving me a chance to study abroad and expanding my limits, but also for his guidance, patience, advice and suggestions that made me grow as a researcher and as a person.

I would especially like to thank my family and friends that I love so much and missed every day of my staying in Austria and to my role model Maurilio Hoffmann.

I would also like to thank my colleagues from the Wavelab Research Group for all the support and help that I received. I could not ask for better people to work with.

Finally, I would like to thank the ”*Universidade Federal do Tocantins*” (UFT) and the ”*Conselho Nacional de Desenvolvimento Científico e Tecnológico*” (CNPq) for the financial support. This thesis was partially supported by CNPq-Brazil under grant No. 00736/2014-0.

Now I ask you, reader, permission to say some words in portuguese: *Quero apenas dizer que a vida é fantástica. Através dela podemos realizar tantas coisas! Esta tese exigiu muito de mim. Tive que saber ouvir, refazer, repensar. Tive que aprender a ser forte, a conhecer meus limites e a pedir ajuda. Aprendi também a ser ajudado com dignidade e a ser grato, sabendo que a ajuda é parte da vida. Cultive a gratidao, ela é uma credencial que habilita o ser humano a evoluir. Obrigado à todos que me ajudaram de alguma forma a chegar até aqui e a relizar um dos meus grandes sonhos: o título de Doutor.*

Salzburg, March 2018 *Eduardo Ribeiro*

Contents

1. Introduction

This cumulative dissertation covers my research performed at the University of Salzburg in the *Wavelab* research group. The main topic of my research is the transfer learning using Convolutional Neural Networks (CNNs) for Colonic Polyp Classification and Iris Super-Resolution. Traditional applications in machine learning assume that training data and testing/real data must have the same distribution and the same feature space. But in many real-world applications specially based on Deep Learning, finding a large enough database to train a CNN with its millions of parameters can turn into an impossible task. In such cases, a strategy called knowledge transfer or transfer learning can be used, avoiding expensive data-labeling efforts as in the case of medical images.

The focus of my research is laid on the application of Convolution Neural Networks using Texture Transfer Learning for two different problems: Colonic Polyp Classification and Iris Super Resolution. These two areas have the problem of the lack of a general, complete, annotated database that generalizes the most complex issue in Computer Vision field: affine invariant and general texture description. Therefore, the employment of CNNs and transfer learning to surpass these problems is a highly intuitive idea for both applications.

This thesis is organized as follows. Section 1.1 gives a introduction of Transfer Learning and its model. Section 1.2 gives an brief overview of Convolutional Neural Networks and how the transfer learning can be applied in this context. Section 1.3 introduces the problem of Colonic Polyp classification and related work in the computer computer-assisted diagnosis of colonic polyps. Section 1.4 explains about the problem of Iris Super Resolution and related work in the use of deep learning for this problem.

Moreover, to improve the readability, the presented contributions are divided into 2 categories:

1. Convolutional Neural Networks and Transfer Learning applied to Colonic Polyp Classification (Section 2)

2. Convolutional Neural Networks and Transfer Learning applied to Iris Super Resolution (Section 3)

Besides that, the publications are presented in Section 4 and the Section 5 gives a general conclusion of the research presented in this thesis. The breakdown of the authors contributions of the publications is listed in the appendix.

1.1. Transfer Learning

The transfer learning occurs when a machine learning approach is trained by a certain domain and is applied in another task being similar or not. The motivation for using this technique comes from observing examples in the real world where nature can intelligently uses previously acquired knowledge and adapt it to a quicker solution of a new problem. One simple example is when a person who already knows how to play a guitar can easily learn to play a bass. The analysis of this kind of natural behavior was the main motivation for the NIPS-95 workshop "Learning to learn" which focused on the need to adapt machine learning methods reusing pre-acquired knowledge giving an impulse to this promising field of research [22].

Nowadays, several machine learning applications use transfer learning techniques to improve their methods including text sentiment classification [37], software defect classification [18] image classification [44], and multi-language text classification [43]. In computer vision, examples of transfer learning include object detection [4] and image classification [26] [21].

According to Pang and Yang [23], transfer learning can be defined by the following model. Given a domain D having two components: A feature space $X = \{x_1, x_2, ...x_n\}$ and a probabilistic distribution $P(X)$ i.e. $D = \{X, P(X)\}$. Also, given a task T with two components: a ground truth $Y = \{y_1, y_2, ...y_n\}$ and an objective function $T = \{Y, f(.)\}$ assuming that this function can be learned through a training database. Function $f(.)$ can be used to predict the correspondent class $f(x)$ of a new instance x. From a probabilistic point of view, $f(x)$ can be written as $P(y \mid x)$. A given training database X associated to the ground truth Y consisting of the pairs $\{x_i, y_i\}$ is used to train and "learn" the function $f(.)$ or $P(y \mid x)$ until it reaches a defined and acceptable error rate between the result of the function $f(x)$ and the ground truth Y. In case of transfer learning, given a source domain $D_S = \{(x_{S_1}, y_{S_1}) \quad , (x_{S_2}, y_{S_2}) \quad , ... \quad , (x_{S_n}, y_{S_n})\}$ and the learning task T_S, the target domain $D_T = \{(x_{T_1}, y_{T_1}) \quad , (x_{T_2}, y_{T_2}) \quad , ... \quad , (x_{T_m}, y_{T_m})\}$ and the learning task T_T, transfer learning aims to help improving the learning of the target predictive function $f_T(.)$ using the knowledge in D_S and T_S where $D_T \neq D_S$ and $T_T \neq T_S$. Among the various categories of transfer learning, one, called inductive transfer learning, has been used with success in the pattern recognition area. In the inductive transfer learning approach an annotated database is necessary for the source domain as well as for the target domain.

In recent years there has been an increased interest in machine learning techniques that is based not on hand-engineered feature extractors but using raw data to learn the representations [25]. Among the development of efficient parallel solvers together with GPUS, the use of Deep Learning has been extensively explored in the last years in different fields of application. Deep learning is intimately related to the use of raw data to do high level representations of this knowledge through a large volume of annotated data. The most used Deep Learning technique for Computer Vision is the Convolutional Neural Network (CNN) approach that is a class of a feed-forward artificial neural network inspired by the biological behavior. This approach will be used in this work for transfer learning and will be detailed in the next section.

1.2. Convolutional Neural Networks and Transfer Learning

In this section we briefly describe the components of a CNN and how it can be used to perform the CNN from scratch, using fine tunning and performing the transfer learning approach.

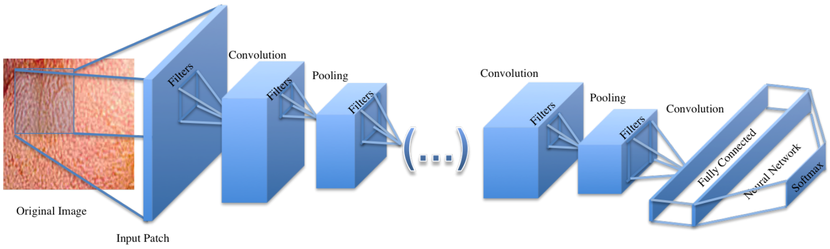


Figure 1.1.: An illustration of the CNN architecture for colonic polyp classification.

A CNN is very similar to traditional Neural Networks in the sense of being constructed by neurons with their respective weights, biases and activation functions. The structure is basically formed by a sequence of convolution and pooling layers ending in a fully connected Neural

Network as shown in Figure 1.1. Generally, the input of a CNN is a $(m \times m \times d)$ image (or patch) where $(m \times m)$ is the dimension of the image and d the number of channels (depth) of the image. The convolutional layer consists of k learnable filters (also called kernels) with size $(n \times n \times d)$ where $(n \leq m)$ that are convolved in the input resulting in the so-called activation maps or feature maps. As classic Neural Networks, the convolution layer outputs are submitted to an activation function, e.g. the ReLU rectifier function $f(x) = \max(0, x)$ where x is the neuron input. After the convolution, a pooling layer is included to subsample the image by average functions (mean) or max-pooling over regions of size $(p \times p)$. These functions are used to reduce the dimensionality of the data in the following layers (upper layers) and to provide a form of invariance to translation thus making over-fitting control. In the convolution and pooling layers the stride has to be specified: the larger the stride, the smaller the overlapping, decreasing the output volume dimensions.

At the end of the CNN there is a fully connected layer as a regular Multilayer Neural Network with the Softmax function that generates a well-formed probability distribution on the outputs. After a supervised training, the CNN is ready to be used as a classifier or as a feature extractor in the case of transfer learning.

Many strategies exploiting CNNs can be used for medical image classification. These strategies can be employed according to the intrinsic characteristics of each database [12] and two of them, mostly used when it comes to CNN training, are described in the following.

1. **CNN Trained From Scratch** When the available training database is large enough, diverse and very different from the database used in all the available pre-trained CNNs (in a case of transfer learning), the most appropriate approach would be to initialize the CNN weights randomly, and train it according to the image database for the kernels domain adaptation, that is, to find the best way to extract the features of the data in order to classify the images properly.

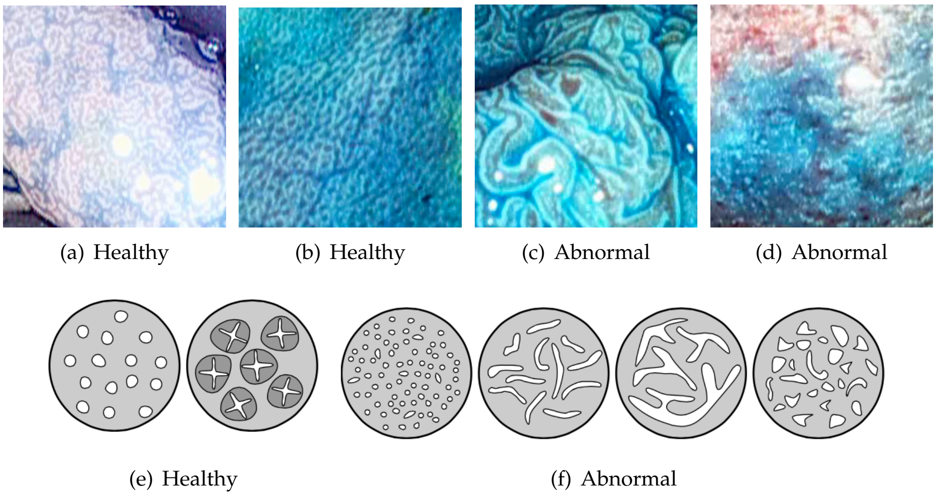
2. **CNN Fine-Tuning** In fine-tuning the pre-trained network training is continued with new entries (with a new database) for the weights to adjust properly to the new scenario reinforcing the more generic features with a lower probability of overfitting.

3. **CNN Transfer Learning** The Transfer Learning occurs when a CNN is trained with a source database different from the domain of the future database. Normally, in image classification, using a pre-trained CNN, the last or next-to-last linear fully connected layer is removed and the remaining pre-trained CNN is used as a feature extractor to generate a feature vector for each input image from a different database.

In this work we explore the three mentioned above techniques to verify which one is more suitable for the colonic polyp classification and iris super-resolution which will be introduced in the next sections.

1.3. Colonic Polyp Classification

The leading cause of deaths related to the intestinal tract is the development of cancer cells (polyps) in its many parts. An early detection (when the cancer is still at an early stage) and a regular exam to everyone over an age of 50 years can reduce the risk of mortality among these patients. More specifically, colonic polyps (benign tumors or growths which arise on the inner colon surface) have a high occurrence and are known to be precursors of colon cancer development.



(a) Healthy (b) Healthy (c) Abnormal (d) Abnormal

(e) Healthy (f) Abnormal

Figure 1.2.: Example images of the two classes (a-d) and the pit-pattern types of these two classes (e-f).

Endoscopy is the most common method for identifying colon polyps and several studies have shown that automatic detection of image regions which may contain polyps within the colon can be used to assist specialists in order to decrease the polyp miss rate [6] [39].

The use of an integrated endoscopic apparatus with high-resolution acquisition devices has been an important object of research in clinical decision support system area. With high-magnification colonoscopies it is possible to acquire images up to 150-fold magnified, revealing the fine surface structure of the mucosa as well as small lesions. Recent work related to classification of colonic polyps used highly-detailed endoscopic images in combination with different technologies divided into three categories: high-definition endoscope (with or without staining the mucosa) combined with the i-Scan technology (1, 2, 3) [40], high-magnification chromoendoscopy [9] and high-magnification endoscopy combined with narrow band imaging [8].

Specifically, the i-Scan technology (Pentax) used in this work is an image processing technology consisting of the combination of surface enhancement and contrast enhancement aiming to help detect dysplastic areas and to accentuate mucosal surfaces and applying post-processing to the reflected light being called virtual chromoendoscopy (CVC) [11].

There are three i-Scan modes available: i-Scan1, which includes surface enhancement and contrast enhancement, i-Scan2, that includes surface enhancement, contrast enhancement and tone enhancement and i-Scan3 that, besides including surface, contrast and tone enhancement, also increases lighting emphasizing the features of vascular visualization [40]. In this work we use an endoscopic image database (CC-i-Scan Database) with 8 different imaging modalities acquired by an HD endoscope (Pentax HiLINE HD+ 90i Colonoscope) with images of size 256×256 extracted from video frames either using the i-Scan technology or without any computer virtual chromoendoscopy ($\neg$CVC).

The automatic detection of polyps in a computer-aided diagnosis (CAD) system is usually performed through a statistical analysis based on color, shape, texture or spatial features applied to the videos frames [3] [24] [38] [35]. The main problems for the detection are the different aspects of color, shape and textures of polyps, being influenced, for example, by the viewing angle, the distance from the capturing camera or even by the colon insufflation as well as the degree of colon muscular contraction [38]. After detection, the colonic polyps can be classified into three different categories: hyperplasic, adenomatous and malignant. Kudo et al. [16] proposed the so-called "pit-pattern" scheme to help in diagnosing tumorous lesions once suspicious areas have been detected. In this scheme, the mucosal surface of the colon can be classified into 5 different types designating the size, shape and distribution of the pit structure [9] [10].

As can be seen in the figure 1.3 (a-d), these five patterns also allow the division of the lesions into two main classes: (1) normal mucosa or hyperplastic polyps (healthy class) and (2) neoplastic, adenomatous or carcinomatous structures (abnormal class). This approach is quite relevant in clinical practice as shown in a study by Kato et al. [15].

1.4. Iris Super-Resolution

Iris recognition technology is considered one of the most accurate and reliable biometric modalities for authentication today mainly due its stability and high degree of freedom in texture [19] [7]. Currently, most systems require the user to present their iris for the sensor at a close distance. However, currently there is a constant pressure to make that relaxed conditions of acquisitions in such systems could be allowed [2] [14]. One of the major problems in these conditions (for example at distance or on the move) is related to the quality of the images which are degraded as well as their resolutions which become low, i.e. the number of pixels in the iris region to allow a good recognition rate is constantly reduced when the resolution decreases as shown in [19].

One of the most relevant areas related to this problem is the Single-Image Super Resolution, which aim to recover a high-resolution image from a low resolution one. Examples are the use of internal patch recurrence [13], regression functions [17] [36] and sparse dictionary methods [41]. However, the use of SR techniques for biometric systems especially for iris recognition is still limited including methods based on PCA eigen-patch transformation [2] and non-parametric Bayesian dictionary learning [1].

Despite the vast literature in SR area and the great interest in the use of Deep-Learning in Biometrics, the application of Deep Learning Super Resolution in iris recognition is still an unexplored field, mainly because approaches generally focus on general and/or natural scenes to produce overall visual enhancement and produce better quality images regarding to photorealism, while iris recognition focuses on the best recognition performance itself [20] [5]. In [34], three multilayer perceptrons (MLPs) are used to perform single image super-resolution for Iris Recognition. The method is based on merging the bilinear interpolation approach with the output pixels values from the trained multiple MLPs considering the edge direction of the iris patterns. Recently, Zhang et.al [42] uses the classic Super-resolution Convolutional Neural Networks (SRCNN) and Super-resolution Forest (SRF) to perform super-resolution in Mobile Iris Recognition systems. The algorithms are applied in the segmented and normalized iris images and the results show a limited effectiveness of the super-resolution method for the iris recognition accuracy. Different from the methods presented in the DLSR literature, in this work we explore if the architectures, and the the database training can have influence in the quality results, and consequently in the recognition performance.

Typically, in a Deep Learning system, the main question is to find a good training database that can provide relevant information to the desired application. In the case of Super Resolution, it is necessary to achieve, during the proposed method training (also called the off-line phase), a mapping between a high-resolution (HR) image with high frequency information and a low-resolution (LR) image with low-frequency information. Figure 1.3 shows this phase, which a training database is chosen and the images are prepared for deep learning SR method training.

In the training phase, the only pre-processing required is, given an image in high resolution X, that image needs to be downscaled to one or more factors followed by a upscaling using

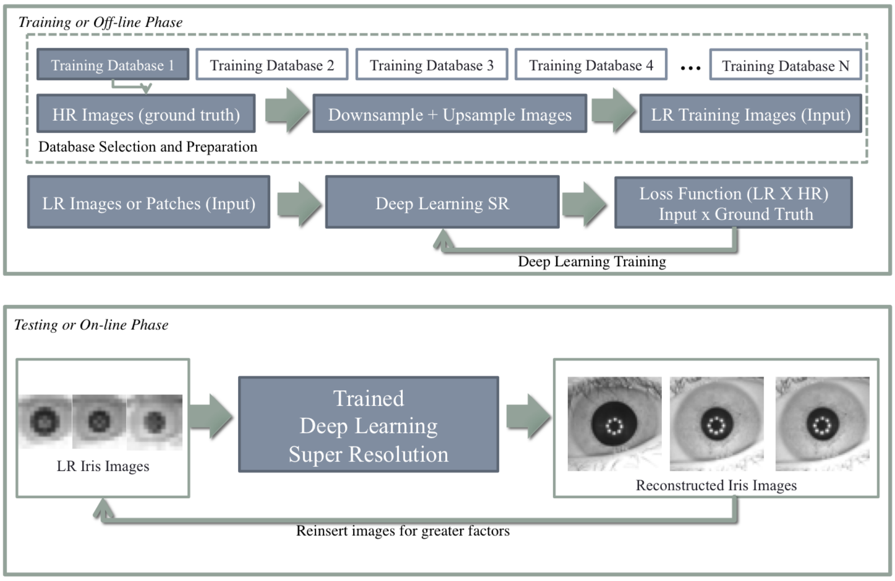


Figure 1.3.: General overview of the training and reconstruction method for the Iris Super Resolution using CNNs.

bicubic interpolation to the same size as the original image X. This image, although it has the same size as X is called ”low resolution” image and is denoted as the LR image Y. The purpose of Deep Learning SR training is, after feeding the network with a LR image or patch Y as input, try to obtain a result F(Y) (the reconstructed image) as much as similar to the HR image or patch X, in this case, the ground truth.

After training, the deep learning method is applied in a low resolution database for the proposed application which is, in the case of this thesis, an iris database also called target database. If so, the deep learning process is a pre-processing step before the iris recognition, in which the low resolution image is introduced as input to the network that will produce the reconstructed image in HR to be used in the process recognition as is shown in Figure 1.3 (on-line phase) that will be reconstructed based on the factor training.

2. Contributions: Convolutional Neural Networks and Transfer Learning applied to Colonic Polyp Classification

Automatic polyp classification based on the so-called pit pattern scheme can help in diagnosing tumorous lesions once suspicious areas have been detected. Deep Learning and Convolutional Neural Networks can help in this task by exploiting directly the input image pixels being successful in handling distortions such as different light conditions, presence of partial occlusions, etc.

Our contribution in this section is showing that CNNs can be used for the automated classification of colonic mucosa, and, to surpass the problem of lack of data to training, "off-the-shelf" CNNs features and texture transfer learning can be an useful alternative to generate rich features for the texture characterization.

Publications (sorted chronologically)

[33] Ribeiro, E., Uhl, A., and Häfner, M. Colonic polyp classification with convolutional neural networks. In *Proceedings of the 29th IEEE International Symposium on Computer-Based Medical Systems (CBMS'16)* (June 2016), pp. 253–258

[28] Ribeiro, E., A. Uhl, G. W., and Häfner, M. Transfer learning for colonic polyp classification using off-the-shelf cnn features (best paper award, 3rd place). In *Proceedings of the 3rd International Workshop on Computer-Assisted and Robotic Endoscopy (CARE'16)* (2016), vol. 10170 of *Springer LNCS*, pp. 1–13

[27] Ribeiro, E., A. Uhl, G. W., and Häfner, M. Exploring deep learning and transfer learning for colonic polyp classification. *Computational and Mathematical Methods in Medicine 2016* (2016), Article ID 6584725

[29] Ribeiro, E., Häfner, M., Wimmer, G., Tamaki, T., Tischendorf, J., S. Yoshida, S. T., and Uhl, A. Exploring texture transfer learning for colonic polyp classification via convolutional neural networks. In *14th International IEEE Symposium on Biomedical Imaging (ISBI'17)* (April 2017)

2.1. Colonic Polyp Classification with Convolutional Neural Networks [33]

The initial work [33] applies CNNs trained from scratch for the automated classification of colonic mucosa for colon polyp staging in the context of colon cancer screening. We show experimentally that this model is more efficient than some of the commonly used features for colonic polyp classification despite the fact that the leave-one-patient-out strategy is used for the training stage because of the lack of sufficient data to proper train the CNN.

2.2. Transfer Learning for Colonic Polyp Classification using Off-the-Shelf CNN Features [28]

In this work [28] we evaluate and analyze the use of CNNs as a general feature descriptor doing transfer learning to generate "off-the-shelf" CNNs features for the colonic polyp classification task. The good results obtained by off-the-shelf CNNs features in many different databases suggest that features learned from CNN with natural images can be highly relevant for colonic polyp classification.

2.3. Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification [27]

This work [27] is an extension of the two first works ([28] and [33]). We compare our results with some commonly used features for colonic polyp classification and the good results suggest that features learned by CNNs trained from scratch and the "off-the-shelf" CNNs features can be highly relevant for automated classification of colonic polyps. Moreover, we also show that the combination of classical features and "off-the-shelf" CNNs features can be a good approach to further improve the results. For the training of CNNs from scratch, we explore data augmentation with image patches to increase the size of the training database and consequently the information to perform the Deep Learning. Different architectures are tested to evaluate the impact of the size and number of filters in the classification as well as the number of output units in the fully connected layer. We also explore and evaluate several different pre-trained CNNs architectures to extract features from colonoscopy images by knowledge transfer between natural and medical images providing what it is called "off-the-shelf" CNNs features. We show that the off-the shelf features may be well suited for the automatic classification of colon polyps even with a limited amount of data. Also, the combination of classical features with off-the-shelf features yields the best prediction results complementing each other.

2.4. Exploring Texture Transfer Learning for Colonic Polyp Classification via Convolutional Neural Networks

In this paper [29] we explore even more the texture transfer learning among different texture databases, using different labels and different distributions via Convolutional Neural Networks (CNNs) for the automated classification of colonic polyps. We show that in texture classification problems with limited amounted of data, as the case of medical area and specifically, the colonic polyp classification task, the transfer learning can be a successfully alternative to extract relevant features by leveraging knowledge learned on other bigger datasets even in very different tasks. We also prove that the bigger the database and the higher the classes in the texture transfer learning, the better the results.

3. Contributions: Convolutional Neural Networks and Transfer Learning applied to Iris Super Resolution

The use of low-resolution images adopting more relaxed acquisition conditions such as mobile phones and surveillance videos is becoming increasingly common in Iris Recognition nowadays. Concurrently a great variety of single image Super-Resolution (SR) techniques are emerging, specially with the use of Convolutional Neural Networks (CNNs). The main objective of these methods is try to recover finer texture details generating more photo-realistic images based on the optimization of a objective function depending basically on the CNN architecture and the training approach. Our contribution for this field is the discussion if the well known Deep-Learning Super-Resolution method for natural images is also valuable for Iris Super-Resolution and, consequently, for Iris Recognition.

Publications (sorted chronologically)

[32] Ribeiro, E., Uhl, A., Alonso-Fernandez, F., and Farrugia, R. A. Exploring deep learning image super-resolution for iris recognition. In *Proc. of the 25th European Signal Processing Conference (EUSIPCO 2017), Kos Island, Greece, August 28 - September 2, 2017* (2017)

[30] Ribeiro, E., and Uhl, A. Exploring texture transfer learning via convolutional neural networks for iris super resolution. In *Proceedings of the 2017 International Conference of the Biometrics Special Interest Group (BIOSIG'17), Darmstadt, Germany 2017* (2017), LNI, GI / IEEE

[31] Ribeiro, E., Uhl, A., and Alonso-Fernandez, F. Iris super-resolution using cnns: is photo-realism important to iris recognition? *Submitted to: IET Biometrics –*, – (2017), –

3.1. Exploring Deep Learning Image Super-Resolution for Iris Recognition [32]

In this work [32] we test the ability of deep learning methods to provide an end-to-end mapping between low and high resolution images applying it to the iris recognition problem. We propose the use of two deep learning single-image super-resolution approaches: Stacked Auto-Encoders (SAE) and Convolutional Neural Networks (CNN) trained from scratch with the most possible lightweight structure to achieve fast speed, preserve local information and reduce artifacts at the same time. When we evaluate the recognition rate by iris comparison experiments, the CNNs in general present better results, but there is no particular CNN approach being the best in all scenarios.

3.2. Exploring Texture Transfer Learning via Convolutional Neural Networks for Iris Super Resolution [30]

In this paper [30] we explore the use of texture transfer learning for super resolution applied to low resolution images. For this, we test if the nature of the images as well as the pattern from the iris can influence the CNN transfer learning and, consequently, the results in the recognition process. The good results obtained by the texture transfer learning using a deep architecture suggest that features learned by Convolutional Neural Networks used for image super-resolution can be highly relevant to increase iris recognition rate. We also show how the features from completely different nature can be transferred in the feature domain, improving the recognition performance if applied to bigger reduction factors comparing to the classical interpolation approaches.

3.3. Iris Super-Resolution using CNNs: is Photo-Realism Important to Iris Recognition? [30]

This work [30] is an extension of the two previous work ([32] and [32]). Here, we discuss if the well known Deep-Learning Super-Resolution method for natural images is also valuable for Iris Super-Resolution and, consequently, for Iris Recognition. We demonstrate by the experiments that there is a dichotomy between the quality assessment and the recognition results showing that, a good photo-realism does not necessarily lead to a good recognition performance specially for very low-resolution images. Differently from the previous work, we focus in the relation between the quality and the performance of the iris recognition. Besides that, the super-resolution is performed in the original image without any segmentation. We also use a new iris database as target database that simulates a real world situation where the images are acquired using mobile phones. Additionally, we test a new application that is the use o Generative Adversarial Networks (SRGANs) to verify if the good performance of this method for natural images.

4. Publications

This chapter presents the publications as originally published. The copyright of the original publications is held by the respective copyright holders, see the following copyright notices. In order to fit the paper dimension, reprinted publications may be scaled in size and/or cropped.

2016 IEEE 29th International Symposium on Computer-Based Medical Systems

Colonic Polyp Classification with Convolutional Neural Networks

Eduardo Ribeiro[1,2] and Andreas Uhl[1]
[1]*University of Salzburg - Department of Computer Sciences - Salzburg, Austria*
[2] *Federal University of Tocantins - Department of Computer Sciences - Tocantins, Brasil*

Michael Häfner
St. Elisabeth Hospital
Vienna, Austria

***Abstract*—Texture patch classification is an important task in many different computer-aided medical systems. Convolutional Neural Networks (CNN's) have become state-of-the-art for many computer vision tasks in recent years. In this paper, we propose the use of CNN's for the automated classification of colonic mucosa for colon polyp staging in the context of colon cancer screening. This deep learning approach has the property of extracting features and classifying images in the same architecture by exploiting directly the input image pixels being successful in handling distortions such as different light conditions, presence of partial occlusions, etc. For this type of deep learning approach it is common to require that the database contains large amounts of data, which is quite rare in the medical field. The method proposed allows the use of small patches (subimages) to increase the size of the database as well to classify different regions in the same image. We show experimentally that this model is more efficient than some of the commonly used features for colonic polyp classification.**

***Index Terms*—Deep Learning, Colonic Polyp Classification, Convolutional Neural Networks**

I. Introduction

Due to the size and complexity of the gastrointestinal tract, many diseases are associated with it, for example: adenomas, polyps, Crohn's disease, celiac disease, Helicobacter pylori infection, among others. However, the leading cause of death related to intestinal tract is caused by the growth of cancerous cells (polyps) in its various parts. Especially in the final segment of the large intestine (colon) and rectum, the colonic polyps have a rather high prevalence and are known to either develop into cancer or to be precursors of colon cancer.

The diagnosis of cancer in an advanced stage increases the mortality risk among patients with color-rectal cancer and can be detected by a physician through an endoscopy procedure. The use of this endoscopic apparatus integrated with high resolution acquisition devices further expanded the research in clinical decision support system area. Intelligent systems can assist in many aspects of colon polyp diagnosis such as accentuating parts of the colon that can possibly have lesions or polyps while the physician performs the colonoscopy procedure, or generating automatic reports about parts of colonoscopy videos that require more attention when they are being analyzed by the physician. Such systems are used to support medical diagnosis, detecting abnormal lesions and/or classifying them, improving the readability of the information, segmenting areas of interest or even predicting possible diagnosis automatically [1], [2].

In the literature, apart from being based on traditional low-resolution white-light colonoscopy, some studies focus mainly on the use of computer-aided diagnosis (CAD) systems related to more advanced colonoscopic images and videos. For computer assisted staging of colon polyps, high-magnification colonoscopes have been used, providing images which are up to 150-fold magnified, thus uncovering the fine surface structure of the mucosa as well as small lesions. Depending on the light source used, colon cancer-oriented CAD systems are divided into two categories: High-magnification chromoendoscopy [3], [1] and high-magnification endoscopy combined with narrow band imaging [4], [5]. However, these expensive devices are only used in larger center and require intensive training of the endoscopist to deliver high quality imagery. Recently, High-Definition (HD) colonoscopes represent a significant advance and are on the way to become clinical standard due to the significantly better image quality (and reasonable costs). Example images of colonic polyps, acquired with such an endoscope, are given in Fig. 1 (a).

In this work we used highly detailed images acquired by a HD endoscope without chromoendoscopy (staining the mucosa). Instead, we employ Pentax virtual chromo-endoscopy (i-Scan technology) which is a method consisting of the combination of surface enhancement and contrast enhancement aiming to help detect dysplastic areas and to accentuate mucosal surfaces [6]. In Fig. 1 (b), an adenomatus polyp acquired using the i-Scan 1 image enhancement technology can be seen [7].

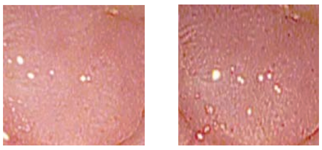

(a) Original
(b) i-Scan 1

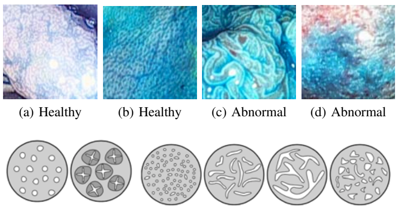
Fig. 1: Images of a polyp without image enhancement (a) and using digital i-Scan 1 technology (b).

For classic white-light endoscopies, several studies have shown that automatic image analysis can be successfully employed to *detect* colorectal polyps in order to assist physicians to decrease the polyp miss rate by detecting image regions that may contain polyps within the colon [8], [9]. Such detection can be performed by analyzing the polyp appearance generally based on color, shape, texture or spatial features applied to the video frames [10], [11], [12]. Colonic polyps may present different aspects of color, shape and texture depending on the

2372-9198/16 $31.00 © 2016 IEEE
DOI 10.1109/CBMS.2016.39

way they are captured by the camera, being influenced, for example, by the viewing angle, the distance from the capturing camera or even by the colon insufflation as well as the degree of colon muscular contraction [11].

Besides that, automatic polyp *classification*, e.g. based on the so-called pit pattern scheme [13], can help in diagnosing tumorous lesions once suspicious areas have been detected [2], [14], [3]. In this paper we also focus on classification and aim to differentiate polyps into two classes: normal mucosa or hyperplastic polyps (class healthy) and neoplastic, adenomatous or carcinomatous structures (class abnormal) as can be seen in Fig 2 (a-d). The different types of pit patterns [13] of these two classes can be observed in Fig. 2 (e-f) [7]. However, the classification can be a difficult task due to several factors such as the lack or excess of illumination, the blurring due to movement or water injection and the appearance of polyps [14], [11].



(e) Healthy (f) Abnormal

Fig. 2: Example images of the two classes (a-d) and the pit-pattern types of these two classes (e-f).

In the literature, existing computer-aided diagnosis techniques generally make use of feature extraction methods of color, shape and texture in combination with machine learning classifiers to perform the classification of colon polyps [15], [16]. Convolution Neural Networks are a promising methodology to help to improve these tasks.

Convolution Neural Networks (CNN's) have been demonstrated to be effective for discriminative pattern recognition in big data and in real-world problems mainly to learn both the global and local structures of images [17]. More recently, CNN were also tested for Computer-aided diagnosis systems such as the analysis, segmentation and prediction of knee cartilage as well as feature extraction from lung CT images [18]. The main advantage of this approach is that the same method can be used for the extraction of strong features that are invariant to distortion and position at the same time of the image classification. The intrinsic feature extractor is formed during the CNN training adapting to the context of the database. Finally, the neural network classifier can make use of these inputs to delineate more accurate hyperplanes helping the generalization of the network. However, one of the problems in the application of this approach is that the deep layers of the CNN work best with structures based on edges, lines and curves, originating from object detection, however most medical databases have more texture-like images having no distinct structures of exactly these types. Another concern is the limitation of the availability of annotated images from medical image databases, since to avoid overfitting a large number of images is necessary to be available during the network training. In this work, we use smaller subimages and some strategies such as Dropout and ReLU activation functions to minimize this problem.

II. METHODOLOGY

We use an architecture of Convolutional Neural Network based on [17] to show that is possible to use this approach to also classify colonic polyp images. The network will need some modification to allow texture pattern recognition. Fig. 3 shows an illustration of the Convolutional Neural Network used in one of the experiments of this work.

A CNN is very similar to traditional Neural Networks in the sense of being constructed by neurons with their respective weights, biases and activation functions. As in Neural Networks, each neuron receives a series of inputs (representing dendrites) which are weighted and summed by the output neurons (representing a neuron's axon). In the case of CNN's, convolutional layers form the first levels (usually with a subsampling step) followed by one or more fully-connected neural networks similar to the multilayer neural networks [19].

In this work, the CNN input is a $(m \times m \times d)$ image (or patch) where $(m \times m)$ is the dimension of the patch and d the number of channels (depth) of the image, in the case of this work: the 3 RGB channel, $d = 3$. The convolutional layer consists of k learnable filters (also called kernels) with size $(n \times n \times d)$ where $(n \leq m)$. Such filters are convolved throughout the image by the product between the inputs and the filter resulting in a new output matrix. Convolving all the k filters and stacking these matrices will form the output volume also called activation maps or feature maps.

In addition, in the convolution step a padding in the input volume is used with zeros (zero padding) to control the spatial volume of output maps as it is appropriate to preserve the exact size of the original inputs. Besides, the stride of the convolution along the spatial dimension has to be specified: the larger the stride, the smaller the overlapping, decreasing the output volume dimensions.

After the convolution, a pooling layer is included to subsample the image by average functions (mean) or max-pooling over regions of size $(p \times p)$. These functions are used to reduce the dimensionality of the data in the following layers (upper layers) and to provide a form of invariance to translation thus making over-fitting control.

One of the most used activation functions in the CNN's and also used in this work is the ReLU rectifier function $f(x) = \max(0, x)$ where x is the neuron input that is demonstrably more efficient than other activation functions [20]. This function accelerates the convergence of the stochastic

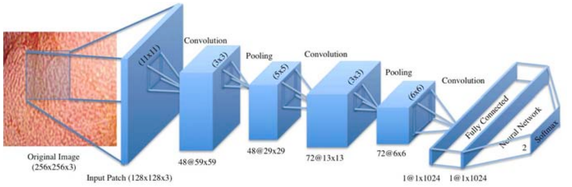


Fig. 3: An illustration of the CNN architecture for colonic polyp classification (CNN-05).

gradient descent algorithm mainly because of its non-linear and unsaturated characteristics.

An alternative to prevent overfitting in large neural networks also used in this work is the Dropout approach [21]. The Dropout disables (drops) feature detector nodes that are weak in the hidden layers of the network during the training forward pass. This is done to reduce interdependence between nodes simulating the training of many large networks with different connections in each iteration [21].

At the end of CNN there is a fully connected layer as a regular Multilayer Neural Network with the activation functions and its offset bias. The activation function used in this part is the Softmax function that generates a well-formed probability distribution on the outputs.

III. Experimental Setup and Results

Due to the limitation of colonic polyp images to train a good CAD system, the main elements of the proposed method are: (1) extracting and preprocessing images in order to have a database with a suitable size (2) the use of CNN's for feature learning and good generalization, (3) the use of methods to avoid overfitting in the training phase.

For the evaluation tests we use a colonic polyp image database consisting of 100 images of size 256×256 from 62 patients using a high-definition (HD) endoscope (Pentax HiLINE HD+ 90i Colonoscope) with i-Scan mode 1 without chromoscopy (staining the mucosa) [6], [7], [22]. These images were extracted from HD video frame regions having histological findings, thus polyp detection is covered in this stage of data preparation. Despite the fact the frames being high-definition, the image size was chosen (i) to be large enough to describe a polyp and (ii) small enough to cover just one class of mucosa type (only healthy or only abnormal area). The database consists of two classes containing 25 healthy images from 18 patients and 75 abnormal images from 56 patients. Some patients may appear in both classes considering that different types of lesions or healthy tissues may be established inside the colon of a single patient. The videos were acquired during colonoscopy sessions between the years 2011 and 2013 at the Department for Internal Medicine (St. Elisabeth Hospital, Vienna).

Usually, some simple preprocessing techniques are necessary for the image feature generation. In this work we apply the normalization by subtracting the mean and dividing by the standard deviation of its elements as in [23] corresponding to local brightness and normalization contrast. We also perform data augmentation by flipping each original image horizontally and vertically, and rotating the original image 90° for the right and left. Besides that, we flipped horizontally the rotated images, then we flipped vertically the horizontally flipped image, totalizing 7 new samples for each original image. After the data augmentation (resulting in 800 images), we randomly extract 75 subimages from each healthy image and 25 subimages from each abnormal image for the training set.

In this work we propose to extract subimages of size 128×128 form the original images. We explored the hypothesis that the colonic polyp classification with the CNN can be done only with a part of the image, and then we trained the network with smaller subimages instead of the entire image. This helps to reduce the size of the network, reducing its complexity and can allow different polyp classifications in the same image using different subimages in different parts of the image. Additionally, choosing smaller regions in a textured image can diminish the degree of intra-image variances in the dataset as the neighborhood is limited.

The CNN proposed by this work to satisfy the requirements cited in the beginning of this section is presented in Fig. 3 and consists of the following layers, parameters and configuration.

- Input Layer: subimages from the original image, of size $128 \times 128 \times 3$.
- Two combinations of convolutional and pooling layers: first convolutional layer consisting of 48 filters of size 11×11 and second convolutional layer consisting of 72 filters of size 5×5. Both layers have padding 0 and stride set to 2 being followed by a ReLU rectifier function. After each convolutional layer there is a max-pooling layer consisting of windows with size 3×3 and stride set to 2;
- One convolutional layer to map the feature maps to the fully-connected output layer consisting of 1024 filters of size 6×6.
- Fully-connected output layer: consists of a neural network with a hidden layer (with 1024 neurons) and a Softmax output layer depending on the number of the classes (in this case, two classes). Also, the Dropout method was used to regularize the two last fully-connected layers.

These hyperparameters were selected based on the works

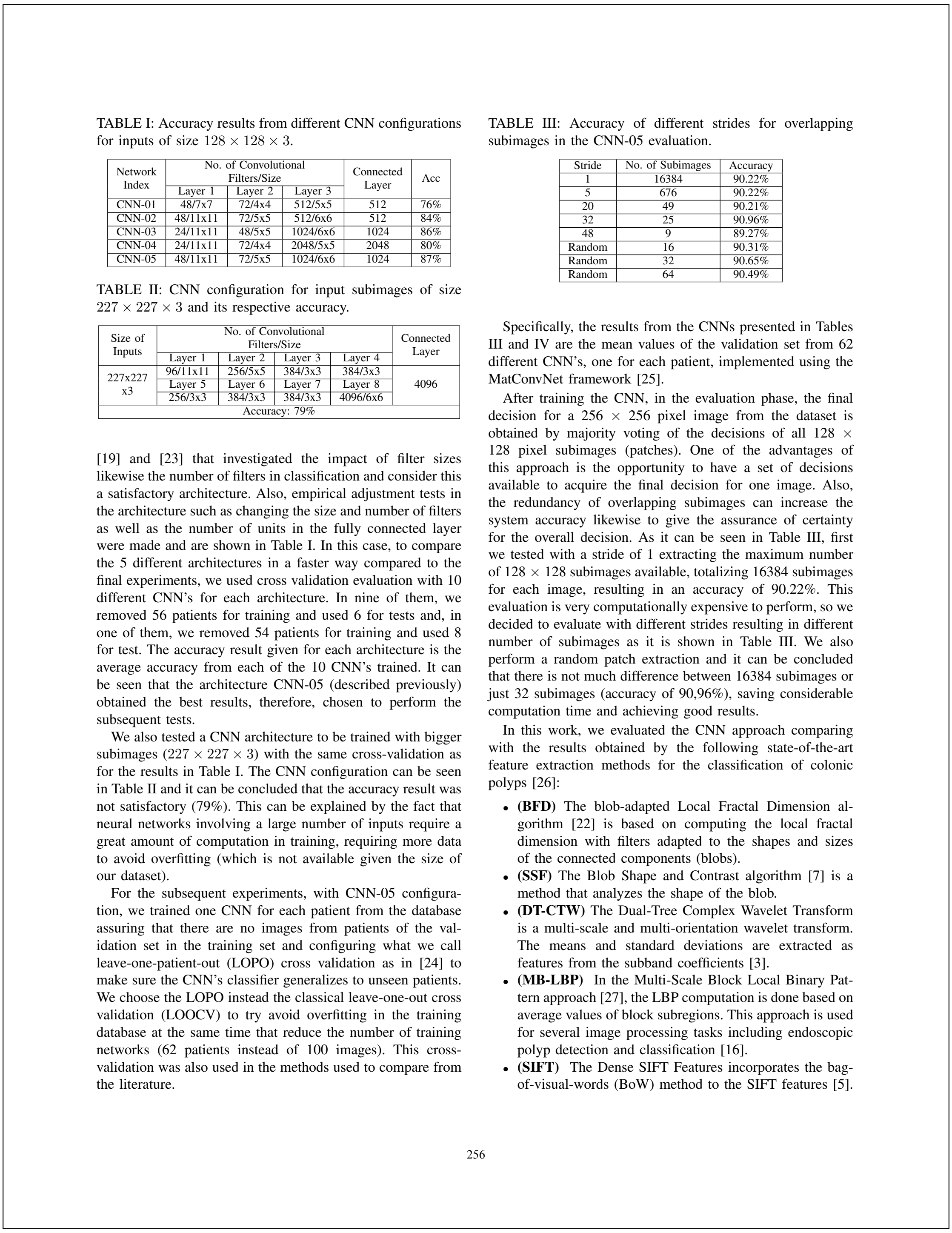


TABLE I: Accuracy results from different CNN configurations for inputs of size $128 \times 128 \times 3$.

Network Index	No. of Convolutional Filters/Size			Connected Layer	Acc
	Layer 1	Layer 2	Layer 3		
CNN-01	48/7x7	72/4x4	512/5x5	512	76%
CNN-02	48/11x11	72/5x5	512/6x6	512	84%
CNN-03	24/11x11	48/5x5	1024/6x6	1024	86%
CNN-04	24/11x11	72/4x4	2048/5x5	2048	80%
CNN-05	48/11x11	72/5x5	1024/6x6	1024	87%

TABLE II: CNN configuration for input subimages of size $227 \times 227 \times 3$ and its respective accuracy.

Size of Inputs	No. of Convolutional Filters/Size				Connected Layer
	Layer 1	Layer 2	Layer 3	Layer 4	
227x227 x3	96/11x11	256/5x5	384/3x3	384/3x3	4096
	Layer 5	Layer 6	Layer 7	Layer 8	
	256/3x3	384/3x3	384/3x3	4096/6x6	
Accuracy: 79%					

[19] and [23] that investigated the impact of filter sizes likewise the number of filters in classification and consider this a satisfactory architecture. Also, empirical adjustment tests in the architecture such as changing the size and number of filters as well as the number of units in the fully connected layer were made and are shown in Table I. In this case, to compare the 5 different architectures in a faster way compared to the final experiments, we used cross validation evaluation with 10 different CNN's for each architecture. In nine of them, we removed 56 patients for training and used 6 for tests and, in one of them, we removed 54 patients for training and used 8 for test. The accuracy result given for each architecture is the average accuracy from each of the 10 CNN's trained. It can be seen that the architecture CNN-05 (described previously) obtained the best results, therefore, chosen to perform the subsequent tests.

We also tested a CNN architecture to be trained with bigger subimages ($227 \times 227 \times 3$) with the same cross-validation as for the results in Table I. The CNN configuration can be seen in Table II and it can be concluded that the accuracy result was not satisfactory (79%). This can be explained by the fact that neural networks involving a large number of inputs require a great amount of computation in training, requiring more data to avoid overfitting (which is not available given the size of our dataset).

For the subsequent experiments, with CNN-05 configuration, we trained one CNN for each patient from the database assuring that there are no images from patients of the validation set in the training set and configuring what we call leave-one-patient-out (LOPO) cross validation as in [24] to make sure the CNN's classifier generalizes to unseen patients. We choose the LOPO instead the classical leave-one-out cross validation (LOOCV) to try avoid overfitting in the training database at the same time that reduce the number of training networks (62 patients instead of 100 images). This cross-validation was also used in the methods used to compare from the literature.

TABLE III: Accuracy of different strides for overlapping subimages in the CNN-05 evaluation.

Stride	No. of Subimages	Accuracy
1	16384	90.22%
5	676	90.22%
20	49	90.21%
32	25	90.96%
48	9	89.27%
Random	16	90.31%
Random	32	90.65%
Random	64	90.49%

Specifically, the results from the CNNs presented in Tables III and IV are the mean values of the validation set from 62 different CNN's, one for each patient, implemented using the MatConvNet framework [25].

After training the CNN, in the evaluation phase, the final decision for a 256×256 pixel image from the dataset is obtained by majority voting of the decisions of all 128×128 pixel subimages (patches). One of the advantages of this approach is the opportunity to have a set of decisions available to acquire the final decision for one image. Also, the redundancy of overlapping subimages can increase the system accuracy likewise to give the assurance of certainty for the overall decision. As it can be seen in Table III, first we tested with a stride of 1 extracting the maximum number of 128×128 subimages available, totalizing 16384 subimages for each image, resulting in an accuracy of 90.22%. This evaluation is very computationally expensive to perform, so we decided to evaluate with different strides resulting in different number of subimages as it is shown in Table III. We also perform a random patch extraction and it can be concluded that there is not much difference between 16384 subimages or just 32 subimages (accuracy of 90,96%), saving considerable computation time and achieving good results.

In this work, we evaluated the CNN approach comparing with the results obtained by the following state-of-the-art feature extraction methods for the classification of colonic polyps [26]:

- **(BFD)** The blob-adapted Local Fractal Dimension algorithm [22] is based on computing the local fractal dimension with filters adapted to the shapes and sizes of the connected components (blobs).
- **(SSF)** The Blob Shape and Contrast algorithm [7] is a method that analyzes the shape of the blob.
- **(DT-CTW)** The Dual-Tree Complex Wavelet Transform is a multi-scale and multi-orientation wavelet transform. The means and standard deviations are extracted as features from the subband coefficients [3].
- **(MB-LBP)** In the Multi-Scale Block Local Binary Pattern approach [27], the LBP computation is done based on average values of block subregions. This approach is used for several image processing tasks including endoscopic polyp detection and classification [16].
- **(SIFT)** The Dense SIFT Features incorporates the bag-of-visual-words (BoW) method to the SIFT features [5].

The visual words are the cluster centers from the k-means clustering applied to the means of the SIFT descriptors.

- **(VASC-F)** The Vascularization Features represent the shape, contrast, size and underlying color of connected components (blood vessels) [15]. These vessel structures on polyps are segmented by means of the phase symmetry filter.

As the focus of several of the original publications was the feature extraction, all the previously cited feature extraction algorithms were evaluated using a k-NN classifier to allow comparison wrt. discriminativeness of the features [22], [7]. In order to stay consistent to the results published, the results of the feature extraction methods presented in Table IV are the mean values of the 10 results of the k-NN classifier (k-values $k = 1-10$) also using the leave-one-patient-out cross (LOPO) validation.

Experiment 1 from Table IV shows our best result using overlapped subimages with stride of 32 resulting in 25 subimages for each image in the evaluation tests compared to the feature extraction methods applied to the original images of size 256 × 256. The results demonstrated that our proposed method has a superior performance (90.96%) to the feature extraction methods generally used for colonic polyp image classification. In Experiment 2 from Table IV we also applied the feature extraction methods to overlapped 128 × 128 pixel subimages with stride of 32 (25 subimages) using majority voting in the final classification as in the CNN evaluation. It can be seen that the results do not exhibit a significant change and our method still outperforms all other feature extraction methods. Some of the reasons for this surpassing result may be the use of three RGB bands from the original image by the CNN instead gray-scale images used by the presented feature extraction methods and the use of k-NN classifier instead of the SVM classifier. Table IV also shows the statistical significance of our results using the McNemar test [28] for the Experiment 1. In this case, number 1 indicates that the CNN is significantly different from the method (with significance level $\alpha = 0.05$). As we can see, the DT-CWT and the SIFT approach are classifying images significantly different to the CNN. However, the McNemar test is highly dependent of the database size [26], which may explain the "no significant differences" between the CNN and the other approaches.

The detailed classification results for the CNN evaluation result with stride of 32 (25 subimages) can be consulted in the confusion matrix displayed in Table V. It is also presented its respective Sensitivity (SE) and Specificity (SP) to delineate the CNN's ability to correctly identify the polyps. The confusion matrix represents the mean of the normalized 62 confusion matrices obtained by the LOPO evaluation with 62 patients.

TABLE IV: The classification results comparing our proposed method with feature extraction algorithms used for colonic polyp classification.

Methods	Acc. Exp. 1	Acc. Exp. 2	Sig.
BFD [22]	87.80%	87.00%	0
SSF [7]	84.70%	85.00%	0
DT-CWT [3]	83.90%	81.00%	1
MB-LBP [16]	82.90%	86.00%	0
SIFT [5]	82.00%	89.00%	0
VASC-F [15]	73.00%	62.00%	1
CNN	**90.96%**	**90.96%**	0

TABLE V: Confusion Matrix associated with CNN Colonic Polyp Classification.

		Prediction Outcome		
		p	**n**	**total**
Actual Value	**p′**	True Positive 47.2	False Negative 2.4	P′ = 49.6
	n′	False Positive 3.2	True Negative 9.2	N′ = 12.4
	total	P = 50.35	N = 11.64	
		SE = 95.16%	**SP = 74.19%**	

From the confusion matrix presented in Table V it can be concluded that, the classification accuracy was 90.96% while the sensitivity was 95.16% which represents a quite positive result since it meant that most of the abnormal polyp images were genuinely classified as such. Besides that, there is a reduced score for false negatives which is relevant for this type of application concerning to be cautious with non-detected disorders. In contrast, the specificity score (SP) was lower than the sensibility with 74.19% meaning that the false positive rate was high. It can be explained by the fact that the number of negative samples was quite low comparing to the positive images for the CNN training. In future work, we intend to decrease this false positive percentage by increasing the training database. Even so, in general, the results were very effective.

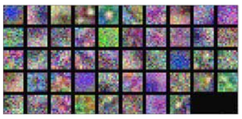


Fig. 4: Filters from the first convolutional layer visualized as small image patches.

The weight matrices in the convolutional layer represent sets of features learned by the network (filters). These features from the first convolution layer of our trained network are presented in Fig. 4. It can be seen that the network has learned a collection of frequency and orientation-selective kernels, as well as many colored blobs intrinsic to the colonic polyp patterns. Some of them are like Laplacian/Gaussian filters, some are like edge detectors at different directions and others like texture extractors. Based on this observation, it can be inferred that the shape, color ant texture information has been learned by the network as good discriminative features to

distinguish the mucosal texture of the colonic polyp image patches. Significant visual features should be captured by these filters for being directly connected to the input image source. Too small filters or too few filters may not capture all the features and generate poor feature maps for the subsequent layers, however, too big or too much filters require a large number of data to improve the accuracy of classification.

IV. CONCLUSION

In this paper, we propose the use of Convolutional Neural Networks (CNN's) to improve the accuracy of colonic polyp classification. This method has the advantage of combining image patches to enlarge the training database, increasing the data volume and consequently the information to perform the deep learning, by the fact that databases containing large amounts of annotated data are often limited for this type of research. The CNN's also use all the intrinsic features of the images such as color, shape and texture, by sharing the filter weights generating strong and representative features that are invariant to local distortions and translations. Different architectures were tested to evaluate the impact of the size and number of filters in the classification as well as the number of output units in the fully connected layer. Our method achieves superior performance compared to the state-of-the-art feature extraction techniques for colonic polyp classification. In future work, to enable even fairer comparison, we will use the outputs of the one-but last CNN layer as inputs into an SVM classifier, and apply an SVM classifier to the classically generated features as well.

ACKNOWLEDGMENT

This work has been partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.

REFERENCES

[1] M. Liedlgruber and A. Uhl, "Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review," *Biomedical Engineering, IEEE Reviews in*, vol. 4, pp. 73–88, 2011.

[2] M. Häfner, L. Brunauer, H. Payer, R. Resch, A. Gangl, A. Uhl, F. Wrba, and A. Vécsei, "Computer-aided classification of zoom-endoscopical images using fourier filters," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 4, pp. 958–970, July 2010.

[3] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, and A. Vécsei, "Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images," *Pattern Analysis and Applications*, vol. 12, no. 4, pp. 407–413, 2009.

[4] M. Ganz, Xiaoyun Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic narrow-band imaging data," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 8, pp. 2144–2151, Aug 2012.

[5] T. Tamaki, J. Yoshimuta, M. Kawakami, B. Raytchev, K. Kaneda, S. Yoshida, Y. Takemura, K. Onji, R. Miyaki, and S. Tanaka, "Computer-aided colorectal tumor classification in {NBI} endoscopy using local features," *Medical Image Analysis*, vol. 17, no. 1, pp. 78 – 100, 2013.

[6] M.l Häfner, A. Uhl, and G. Wimmer, "A novel shape feature descriptor for the classification of polyps in hd colonoscopy," in *Medical Computer Vision. Large Data in Medical Imaging*, vol. 8331 of *Lecture Notes in Computer Science*, pp. 205–213. Springer International Publishing, 2014.

[7] M. Häfner, A. Uhl, and G. Wimmer, "A novel shape feature descriptor for the classification of polyps in hd colonoscopy," in *MICCAI*, vol. 8331, pp. 205–213. Springer International Publishing, 2014.

[8] J. Bernal, J. Snchez, and F. Vilario, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, vol. 45, no. 9, pp. 3166 – 3182, 2012, Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).

[9] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, and P. C. de Groen, "Polyp-alert: Near real-time feedback during colonoscopy," *Computer Methods and Programs in Biomedicine*, vol. 120, no. 3, pp. 164 – 179, 2015.

[10] Sun Young Park, D. Sargent, I. Spofford, K.G. Vosburgh, and Y. A-Rahim, "A colon video analysis framework for polyp detection," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 5, pp. 1408–1418, May 2012.

[11] Yi W., W. Tavanapong, J. Wong, J. Oh, and P.C. de Groen, "Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy," *Biomedical and Health Informatics, IEEE Journal of*, vol. 18, no. 4, pp. 1379–1389, July 2014.

[12] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung fr die Medizin 2009*, Informatik aktuell, pp. 346–350. Springer Berlin Heidelberg, 2009.

[13] Kudo S, Hirota S, and Nakajima T, "Colorectal tumours and pit pattern," *Journal of Clinical Pathology*, vol. 10, pp. 880–885, Oct 1994.

[14] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy.," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 565–581, 2012.

[15] S. Gross, S. Palm, J. Tischendorf, A. Behrens, C. Trautwein, and T. Aach, "Automated classification of colon polyps in endoscopic image data," *SPIE*, vol. 8315, pp. 83150W–83150W–8, 2012.

[16] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Color treatment in endoscopic image classification using multi-scale local color vector patterns," *Medical Image Analysis*, vol. 16, no. 1, pp. 75 – 86, 2012.

[17] Alex K., I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. Curran Associates, Inc., 2012.

[18] Q. Li, W. Cai, X. Wang, Y. Zhou, D.D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *ICARCV, 2014*, Dec 2014, pp. 844–848.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Intelligent Signal Processing*. 2001, pp. 306–351, IEEE Press.

[20] T. N. Sainath, B. Kingsbury, A. Mohamed, G. E. Dahl, G.e Saon, H. Soltau, T. Beran, A. Y. Aravkin, and B. Ramabhadran, "Improvements to deep convolutional neural networks for lvcsr," in *ASRU, 2013 IEEE Workshop on*, 2013, pp. 315–320.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[22] A. Uhl, G. Wimmer, and M. Häfner, "Shape and size adapted local fractal dimension for the classification of polyps in hd colonoscopy," in *ICIP, 2014*, Oct 2014, pp. 2299–2303.

[23] A. Coates, H. Lee, and A.Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," in *AISTATS*. 2011, vol. 15 of *JMLR*, pp. 215–223, JMLR W&CP.

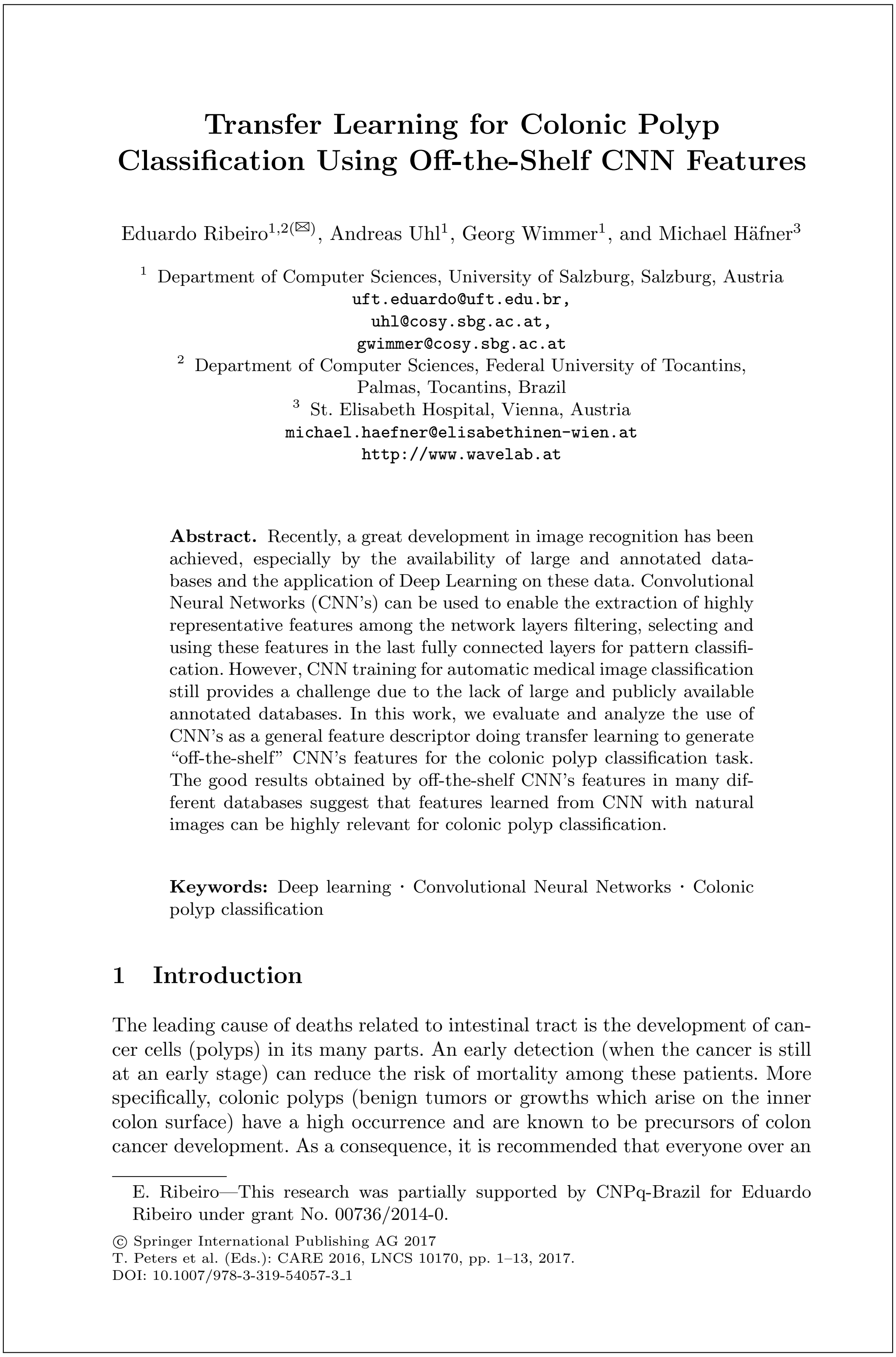
[24] M. Häfner, M. Liedlgruber, S. Maimone, A. Uhl, A. Vécsei, and F. Wrba, "Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, June 2012, pp. 1–6.

[25] A. Vedaldi and K. Lenc, "Matconvnet - convolutional neural networks for MATLAB," *CoRR*, vol. abs/1412.4564, 2014.

[26] G. Wimmer, T. Tamaki, J.J.W. Tischendorf, M. Häfner, S. Yoshida, S. Tanaka, and A. Uhl, "Directional wavelet based features for colonic polyp classification," *Medical Image Analysis*, vol. 31, pp. 16 – 36, 2016.

[27] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S.. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*, vol. 4642 of *Lecture Notes in Computer Science*, pp. 828–837. Springer Berlin Heidelberg, 2007.

[28] Quinn McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157.



Transfer Learning for Colonic Polyp Classification Using Off-the-Shelf CNN Features

Eduardo Ribeiro[1,2(✉)], Andreas Uhl[1], Georg Wimmer[1], and Michael Häfner[3]

[1] Department of Computer Sciences, University of Salzburg, Salzburg, Austria
uft.eduardo@uft.edu.br,
uhl@cosy.sbg.ac.at,
gwimmer@cosy.sbg.ac.at
[2] Department of Computer Sciences, Federal University of Tocantins,
Palmas, Tocantins, Brazil
[3] St. Elisabeth Hospital, Vienna, Austria
michael.haefner@elisabethinen-wien.at
http://www.wavelab.at

Abstract. Recently, a great development in image recognition has been achieved, especially by the availability of large and annotated databases and the application of Deep Learning on these data. Convolutional Neural Networks (CNN's) can be used to enable the extraction of highly representative features among the network layers filtering, selecting and using these features in the last fully connected layers for pattern classification. However, CNN training for automatic medical image classification still provides a challenge due to the lack of large and publicly available annotated databases. In this work, we evaluate and analyze the use of CNN's as a general feature descriptor doing transfer learning to generate "off-the-shelf" CNN's features for the colonic polyp classification task. The good results obtained by off-the-shelf CNN's features in many different databases suggest that features learned from CNN with natural images can be highly relevant for colonic polyp classification.

Keywords: Deep learning · Convolutional Neural Networks · Colonic polyp classification

1 Introduction

The leading cause of deaths related to intestinal tract is the development of cancer cells (polyps) in its many parts. An early detection (when the cancer is still at an early stage) can reduce the risk of mortality among these patients. More specifically, colonic polyps (benign tumors or growths which arise on the inner colon surface) have a high occurrence and are known to be precursors of colon cancer development. As a consequence, it is recommended that everyone over an

E. Ribeiro—This research was partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.

T. Peters et al. (Eds.): CARE 2016, LNCS 10170, pp. 1–13, 2017.
DOI: 10.1007/978-3-319-54057-3_1

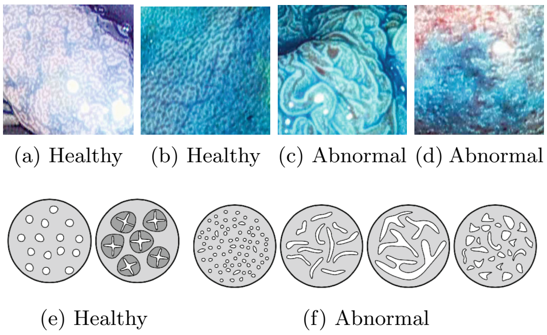
age of 50 years be examined regularly [32]. This exam can be done through an endoscopy procedure that is a minimally invasive and relatively painless diagnostic medical procedure that enables specialists to obtain images of internal human body cavities.

Several studies have shown that automatic detection of image regions which may contain polyps within the colon can be used to assist specialists in order to decrease the polyp miss rate [3,28,31]. Such detection can be performed by analyzing the polyp appearance that is generally based on color, shape, texture or spatial features applied to the video frames denoted as polyp detection [1,21,30].

Subsequently, the polyps can be automatically classified using different aspects of shape, color or texture into hyperplastic, adenomatous and malignant. The so-called "pit-pattern" scheme proposed by Kudo et al. [18] can help in diagnosing tumorous lesions once suspicious areas have been detected. In this scheme, the mucosal surface of the colon can be classified into 5 different types designating the size, shape and distribution of the pit structure [6,9,12]. These five pit-pattern types can allow to group the lesions into two main classes: normal mucosa or hyperplastic polyps (healthy class) and neoplastic, adenomatous or carcinomatous structures (abnormal class) as can be seen in Fig. 1(a–d). This approach is quite relevant in clinical practice as shown in a study by Kato et al. [17].

In this work we focus on the polyp classification into these two classes. The different types of pit patterns [18] of these two classes can be observed in Fig. 1(e–f) [14]. However, the classification can be a difficult task due to several factors such as the lack or excess of illumination, the blurring due to movement or water injection and the different appearances of polyps [32]. Also, to find a robust and a global feature extractor that summarizes and represents all these pit-patterns structures in a single vector is very difficult and Deep Learning can be a good alternative to surpass these problems.

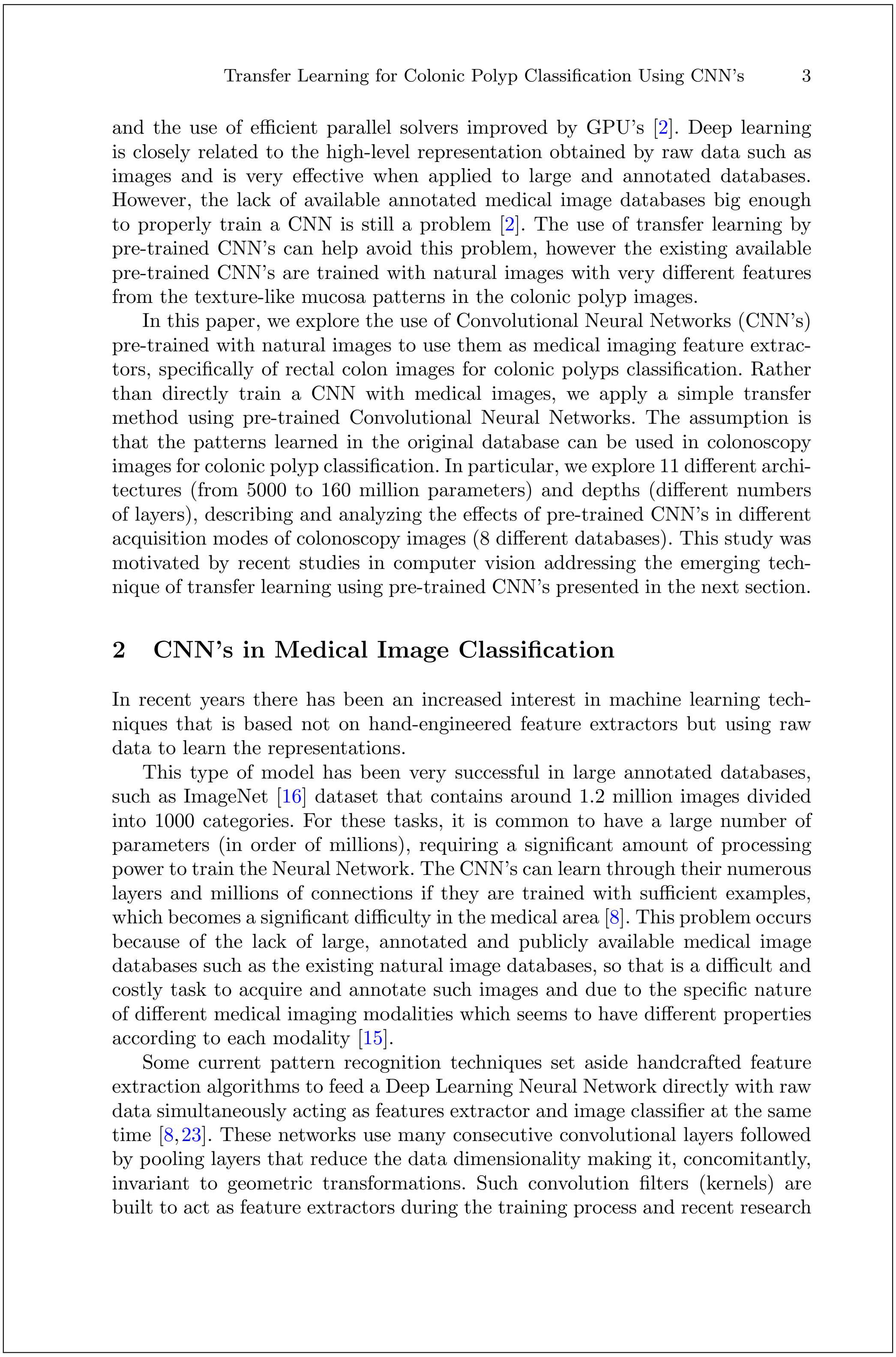
Deep learning Neural Networks have been of great interest in recent years, mainly due to the new variations of so-called Convolutional Neural Networks



(a) Healthy (b) Healthy (c) Abnormal (d) Abnormal

(e) Healthy (f) Abnormal

Fig. 1. Example images of the two classes (a–d) and the pit-pattern types of these two classes (e–f).



and the use of efficient parallel solvers improved by GPU's [2]. Deep learning is closely related to the high-level representation obtained by raw data such as images and is very effective when applied to large and annotated databases. However, the lack of available annotated medical image databases big enough to properly train a CNN is still a problem [2]. The use of transfer learning by pre-trained CNN's can help avoid this problem, however the existing available pre-trained CNN's are trained with natural images with very different features from the texture-like mucosa patterns in the colonic polyp images.

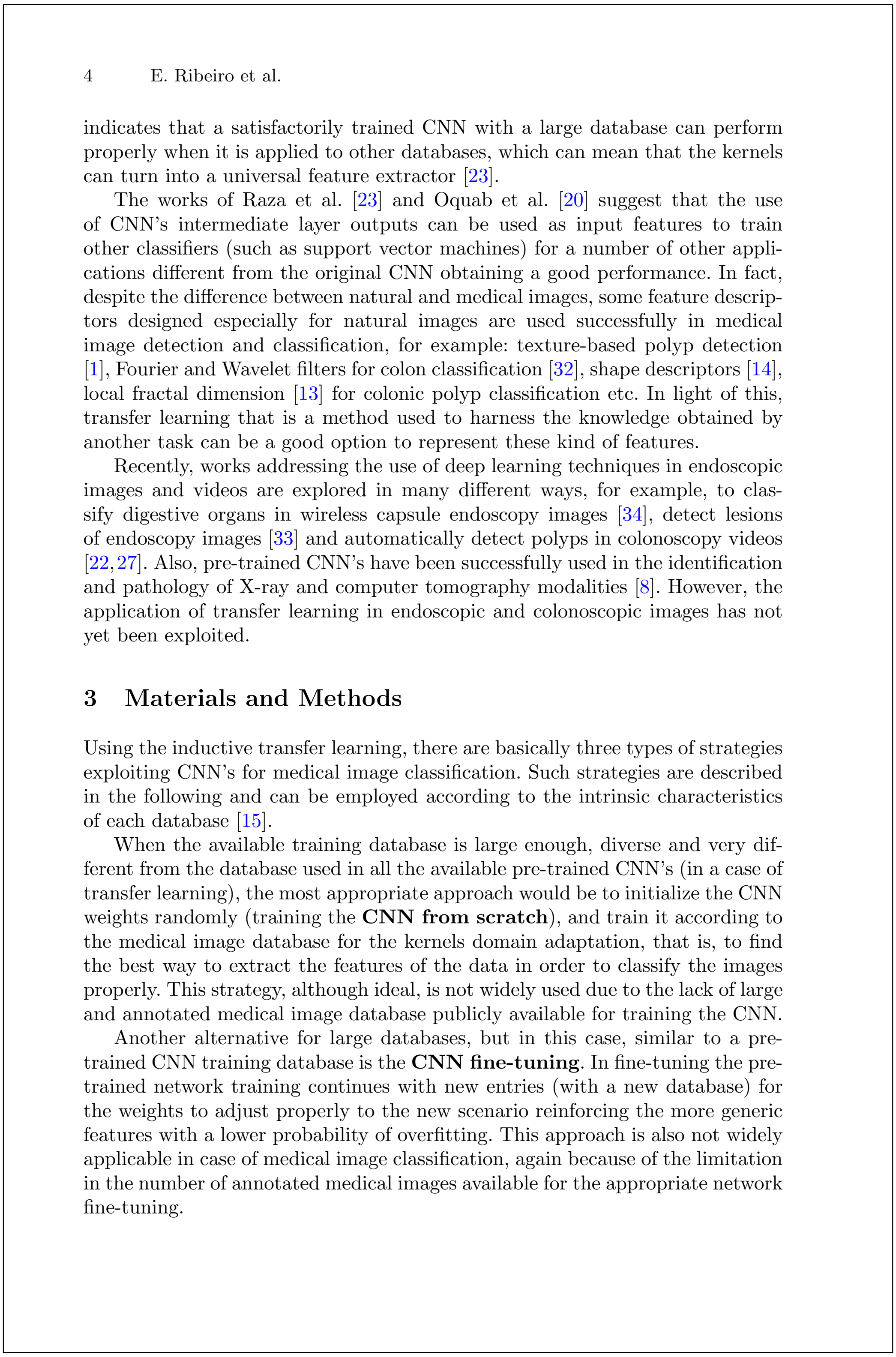
In this paper, we explore the use of Convolutional Neural Networks (CNN's) pre-trained with natural images to use them as medical imaging feature extractors, specifically of rectal colon images for colonic polyps classification. Rather than directly train a CNN with medical images, we apply a simple transfer method using pre-trained Convolutional Neural Networks. The assumption is that the patterns learned in the original database can be used in colonoscopy images for colonic polyp classification. In particular, we explore 11 different architectures (from 5000 to 160 million parameters) and depths (different numbers of layers), describing and analyzing the effects of pre-trained CNN's in different acquisition modes of colonoscopy images (8 different databases). This study was motivated by recent studies in computer vision addressing the emerging technique of transfer learning using pre-trained CNN's presented in the next section.

2 CNN's in Medical Image Classification

In recent years there has been an increased interest in machine learning techniques that is based not on hand-engineered feature extractors but using raw data to learn the representations.

This type of model has been very successful in large annotated databases, such as ImageNet [16] dataset that contains around 1.2 million images divided into 1000 categories. For these tasks, it is common to have a large number of parameters (in order of millions), requiring a significant amount of processing power to train the Neural Network. The CNN's can learn through their numerous layers and millions of connections if they are trained with sufficient examples, which becomes a significant difficulty in the medical area [8]. This problem occurs because of the lack of large, annotated and publicly available medical image databases such as the existing natural image databases, so that is a difficult and costly task to acquire and annotate such images and due to the specific nature of different medical imaging modalities which seems to have different properties according to each modality [15].

Some current pattern recognition techniques set aside handcrafted feature extraction algorithms to feed a Deep Learning Neural Network directly with raw data simultaneously acting as features extractor and image classifier at the same time [8,23]. These networks use many consecutive convolutional layers followed by pooling layers that reduce the data dimensionality making it, concomitantly, invariant to geometric transformations. Such convolution filters (kernels) are built to act as feature extractors during the training process and recent research



indicates that a satisfactorily trained CNN with a large database can perform properly when it is applied to other databases, which can mean that the kernels can turn into a universal feature extractor [23].

The works of Raza et al. [23] and Oquab et al. [20] suggest that the use of CNN's intermediate layer outputs can be used as input features to train other classifiers (such as support vector machines) for a number of other applications different from the original CNN obtaining a good performance. In fact, despite the difference between natural and medical images, some feature descriptors designed especially for natural images are used successfully in medical image detection and classification, for example: texture-based polyp detection [1], Fourier and Wavelet filters for colon classification [32], shape descriptors [14], local fractal dimension [13] for colonic polyp classification etc. In light of this, transfer learning that is a method used to harness the knowledge obtained by another task can be a good option to represent these kind of features.

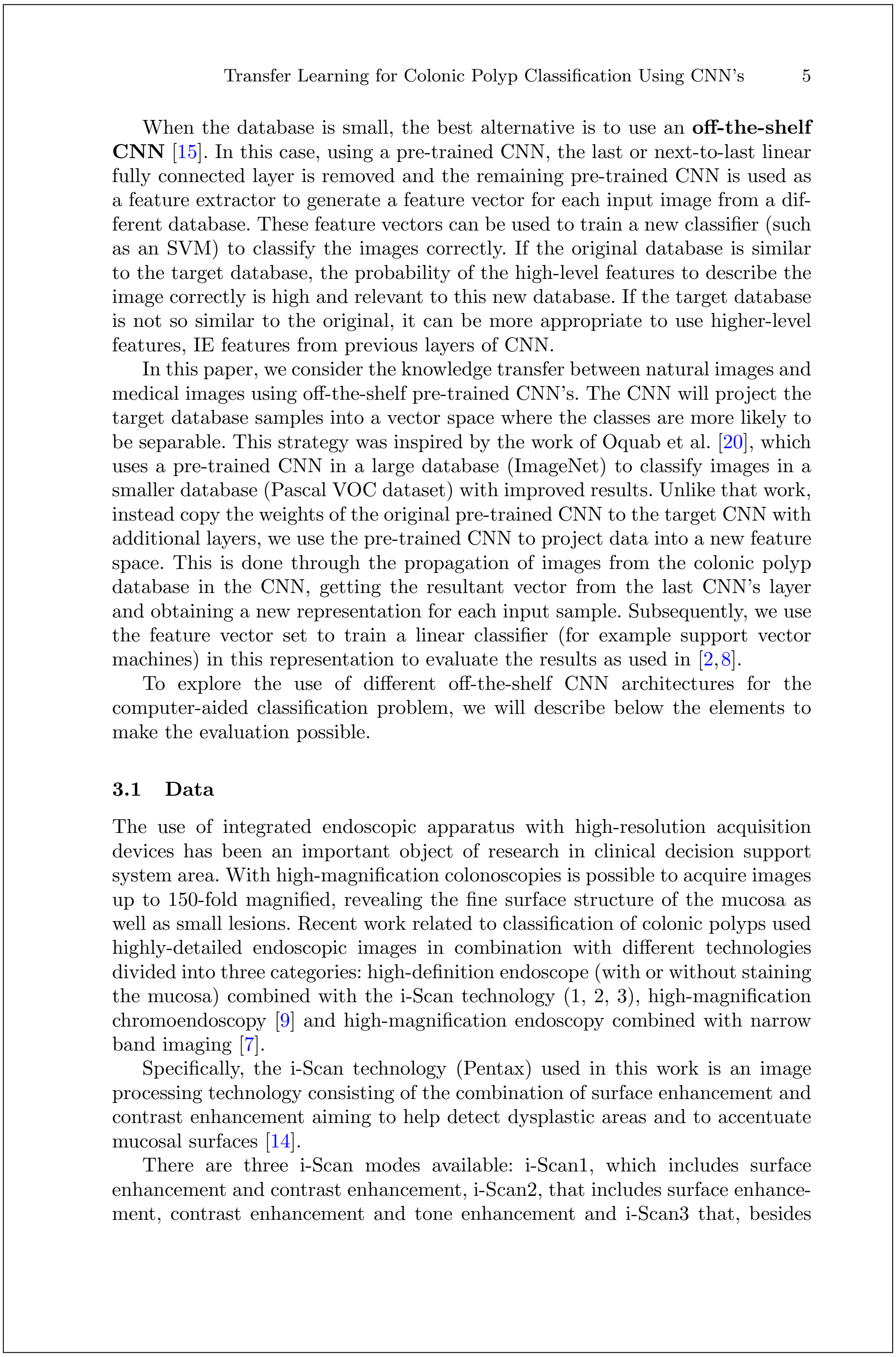
Recently, works addressing the use of deep learning techniques in endoscopic images and videos are explored in many different ways, for example, to classify digestive organs in wireless capsule endoscopy images [34], detect lesions of endoscopy images [33] and automatically detect polyps in colonoscopy videos [22,27]. Also, pre-trained CNN's have been successfully used in the identification and pathology of X-ray and computer tomography modalities [8]. However, the application of transfer learning in endoscopic and colonoscopic images has not yet been exploited.

3 Materials and Methods

Using the inductive transfer learning, there are basically three types of strategies exploiting CNN's for medical image classification. Such strategies are described in the following and can be employed according to the intrinsic characteristics of each database [15].

When the available training database is large enough, diverse and very different from the database used in all the available pre-trained CNN's (in a case of transfer learning), the most appropriate approach would be to initialize the CNN weights randomly (training the **CNN from scratch**), and train it according to the medical image database for the kernels domain adaptation, that is, to find the best way to extract the features of the data in order to classify the images properly. This strategy, although ideal, is not widely used due to the lack of large and annotated medical image database publicly available for training the CNN.

Another alternative for large databases, but in this case, similar to a pre-trained CNN training database is the **CNN fine-tuning**. In fine-tuning the pre-trained network training continues with new entries (with a new database) for the weights to adjust properly to the new scenario reinforcing the more generic features with a lower probability of overfitting. This approach is also not widely applicable in case of medical image classification, again because of the limitation in the number of annotated medical images available for the appropriate network fine-tuning.



When the database is small, the best alternative is to use an **off-the-shelf CNN** [15]. In this case, using a pre-trained CNN, the last or next-to-last linear fully connected layer is removed and the remaining pre-trained CNN is used as a feature extractor to generate a feature vector for each input image from a different database. These feature vectors can be used to train a new classifier (such as an SVM) to classify the images correctly. If the original database is similar to the target database, the probability of the high-level features to describe the image correctly is high and relevant to this new database. If the target database is not so similar to the original, it can be more appropriate to use higher-level features, IE features from previous layers of CNN.

In this paper, we consider the knowledge transfer between natural images and medical images using off-the-shelf pre-trained CNN's. The CNN will project the target database samples into a vector space where the classes are more likely to be separable. This strategy was inspired by the work of Oquab et al. [20], which uses a pre-trained CNN in a large database (ImageNet) to classify images in a smaller database (Pascal VOC dataset) with improved results. Unlike that work, instead copy the weights of the original pre-trained CNN to the target CNN with additional layers, we use the pre-trained CNN to project data into a new feature space. This is done through the propagation of images from the colonic polyp database in the CNN, getting the resultant vector from the last CNN's layer and obtaining a new representation for each input sample. Subsequently, we use the feature vector set to train a linear classifier (for example support vector machines) in this representation to evaluate the results as used in [2,8].

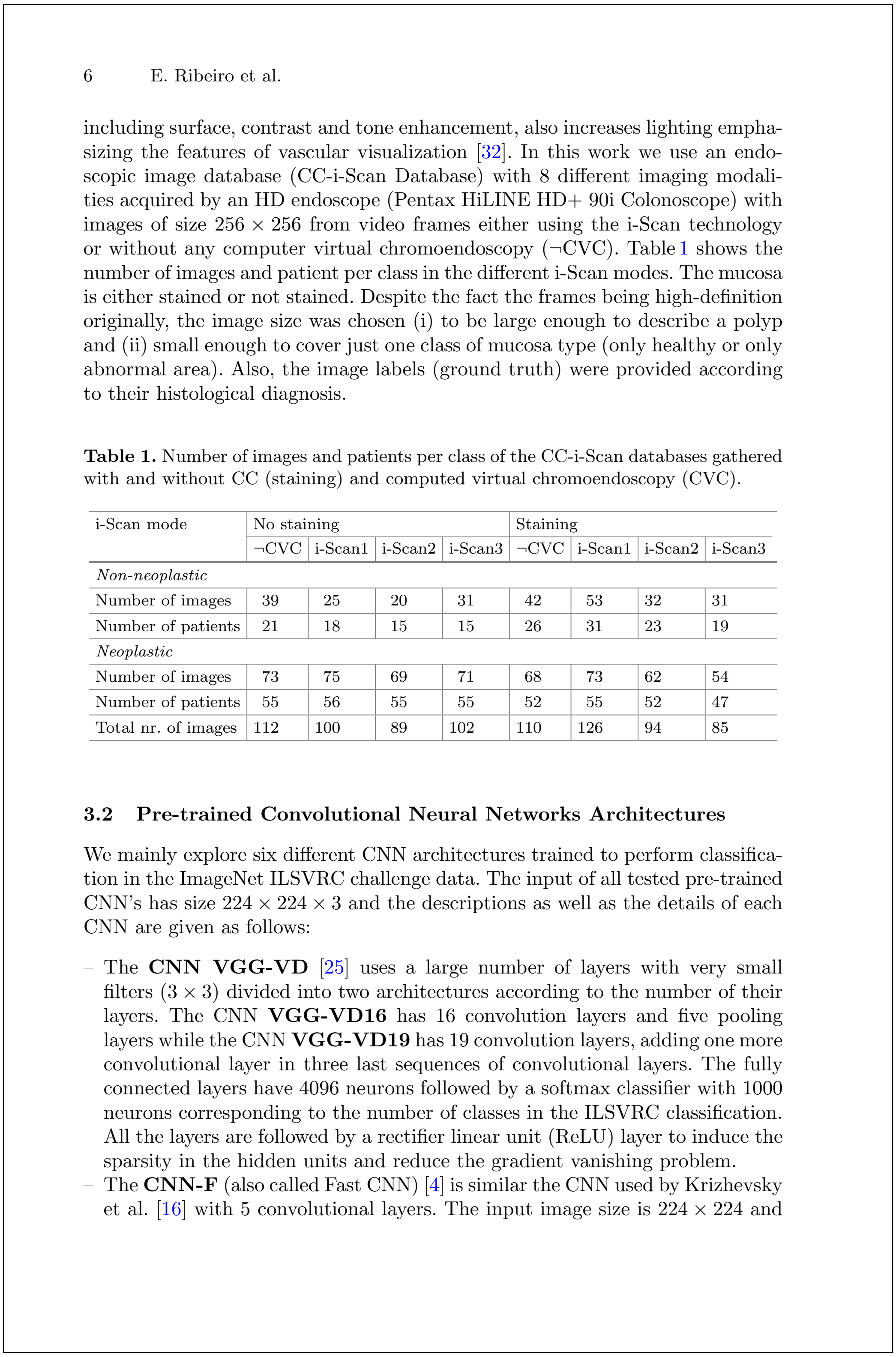
To explore the use of different off-the-shelf CNN architectures for the computer-aided classification problem, we will describe below the elements to make the evaluation possible.

3.1 Data

The use of integrated endoscopic apparatus with high-resolution acquisition devices has been an important object of research in clinical decision support system area. With high-magnification colonoscopies is possible to acquire images up to 150-fold magnified, revealing the fine surface structure of the mucosa as well as small lesions. Recent work related to classification of colonic polyps used highly-detailed endoscopic images in combination with different technologies divided into three categories: high-definition endoscope (with or without staining the mucosa) combined with the i-Scan technology (1, 2, 3), high-magnification chromoendoscopy [9] and high-magnification endoscopy combined with narrow band imaging [7].

Specifically, the i-Scan technology (Pentax) used in this work is an image processing technology consisting of the combination of surface enhancement and contrast enhancement aiming to help detect dysplastic areas and to accentuate mucosal surfaces [14].

There are three i-Scan modes available: i-Scan1, which includes surface enhancement and contrast enhancement, i-Scan2, that includes surface enhancement, contrast enhancement and tone enhancement and i-Scan3 that, besides



including surface, contrast and tone enhancement, also increases lighting emphasizing the features of vascular visualization [32]. In this work we use an endoscopic image database (CC-i-Scan Database) with 8 different imaging modalities acquired by an HD endoscope (Pentax HiLINE HD+ 90i Colonoscope) with images of size 256 × 256 from video frames either using the i-Scan technology or without any computer virtual chromoendoscopy (¬CVC). Table 1 shows the number of images and patient per class in the different i-Scan modes. The mucosa is either stained or not stained. Despite the fact the frames being high-definition originally, the image size was chosen (i) to be large enough to describe a polyp and (ii) small enough to cover just one class of mucosa type (only healthy or only abnormal area). Also, the image labels (ground truth) were provided according to their histological diagnosis.

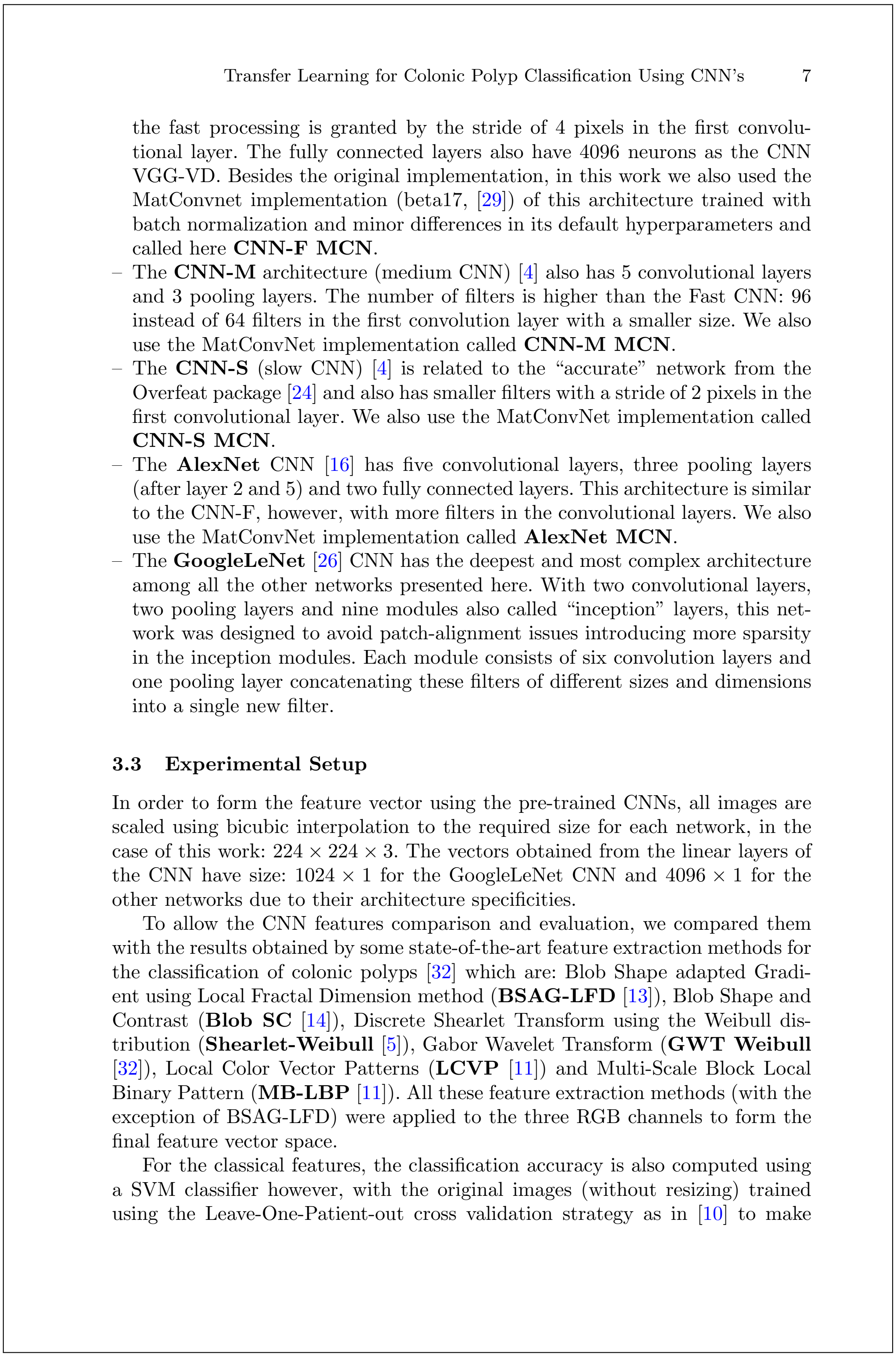
Table 1. Number of images and patients per class of the CC-i-Scan databases gathered with and without CC (staining) and computed virtual chromoendoscopy (CVC).

i-Scan mode	No staining				Staining			
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3
Non-neoplastic								
Number of images	39	25	20	31	42	53	32	31
Number of patients	21	18	15	15	26	31	23	19
Neoplastic								
Number of images	73	75	69	71	68	73	62	54
Number of patients	55	56	55	55	52	55	52	47
Total nr. of images	112	100	89	102	110	126	94	85

3.2 Pre-trained Convolutional Neural Networks Architectures

We mainly explore six different CNN architectures trained to perform classification in the ImageNet ILSVRC challenge data. The input of all tested pre-trained CNN's has size 224 × 224 × 3 and the descriptions as well as the details of each CNN are given as follows:

- The **CNN VGG-VD** [25] uses a large number of layers with very small filters (3 × 3) divided into two architectures according to the number of their layers. The CNN **VGG-VD16** has 16 convolution layers and five pooling layers while the CNN **VGG-VD19** has 19 convolution layers, adding one more convolutional layer in three last sequences of convolutional layers. The fully connected layers have 4096 neurons followed by a softmax classifier with 1000 neurons corresponding to the number of classes in the ILSVRC classification. All the layers are followed by a rectifier linear unit (ReLU) layer to induce the sparsity in the hidden units and reduce the gradient vanishing problem.
- The **CNN-F** (also called Fast CNN) [4] is similar the CNN used by Krizhevsky et al. [16] with 5 convolutional layers. The input image size is 224 × 224 and



the fast processing is granted by the stride of 4 pixels in the first convolutional layer. The fully connected layers also have 4096 neurons as the CNN VGG-VD. Besides the original implementation, in this work we also used the MatConvnet implementation (beta17, [29]) of this architecture trained with batch normalization and minor differences in its default hyperparameters and called here **CNN-F MCN**.

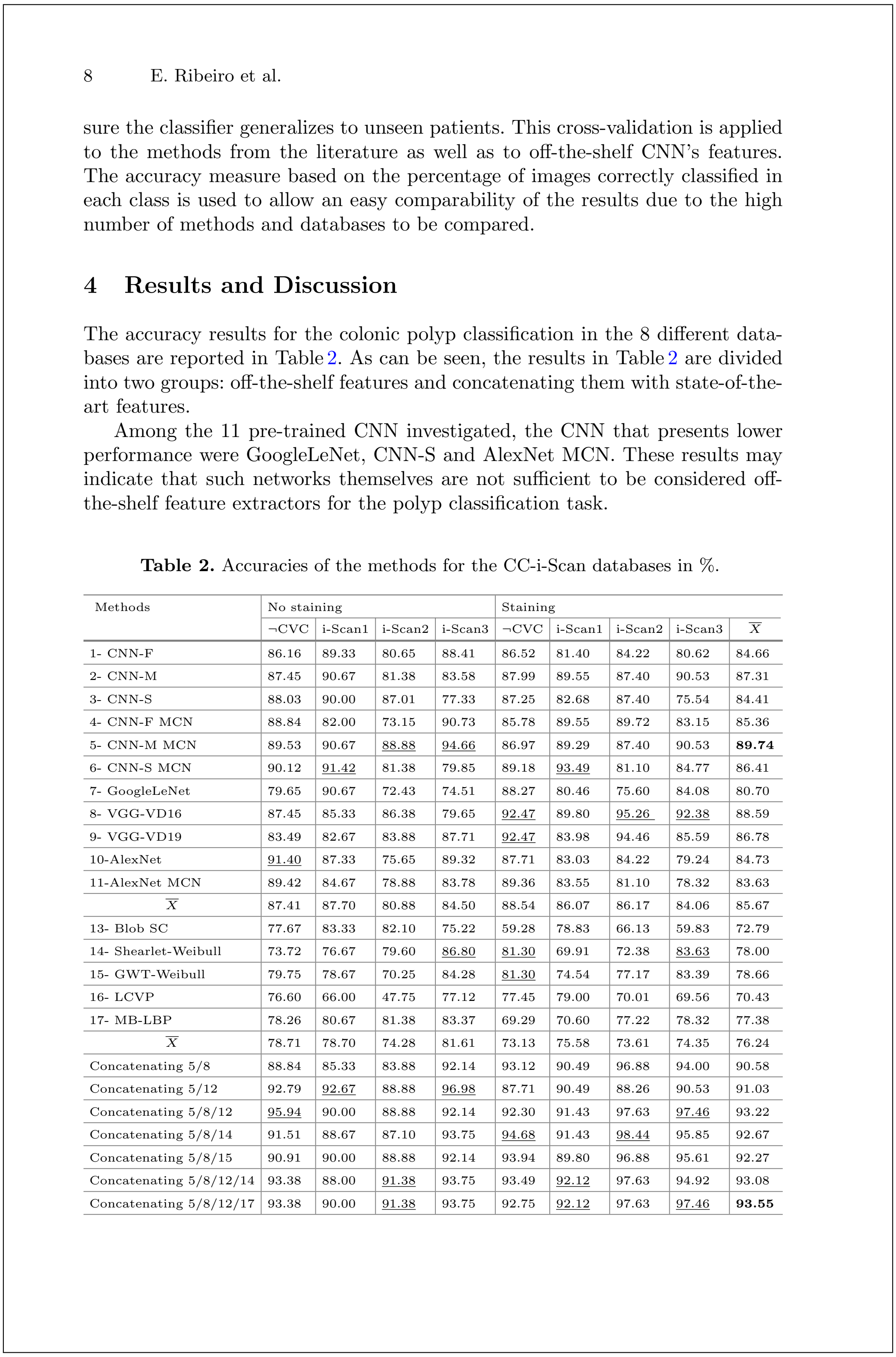
– The **CNN-M** architecture (medium CNN) [4] also has 5 convolutional layers and 3 pooling layers. The number of filters is higher than the Fast CNN: 96 instead of 64 filters in the first convolution layer with a smaller size. We also use the MatConvNet implementation called **CNN-M MCN**.
– The **CNN-S** (slow CNN) [4] is related to the "accurate" network from the Overfeat package [24] and also has smaller filters with a stride of 2 pixels in the first convolutional layer. We also use the MatConvNet implementation called **CNN-S MCN**.
– The **AlexNet** CNN [16] has five convolutional layers, three pooling layers (after layer 2 and 5) and two fully connected layers. This architecture is similar to the CNN-F, however, with more filters in the convolutional layers. We also use the MatConvNet implementation called **AlexNet MCN**.
– The **GoogleLeNet** [26] CNN has the deepest and most complex architecture among all the other networks presented here. With two convolutional layers, two pooling layers and nine modules also called "inception" layers, this network was designed to avoid patch-alignment issues introducing more sparsity in the inception modules. Each module consists of six convolution layers and one pooling layer concatenating these filters of different sizes and dimensions into a single new filter.

3.3 Experimental Setup

In order to form the feature vector using the pre-trained CNNs, all images are scaled using bicubic interpolation to the required size for each network, in the case of this work: $224 \times 224 \times 3$. The vectors obtained from the linear layers of the CNN have size: 1024×1 for the GoogleLeNet CNN and 4096×1 for the other networks due to their architecture specificities.

To allow the CNN features comparison and evaluation, we compared them with the results obtained by some state-of-the-art feature extraction methods for the classification of colonic polyps [32] which are: Blob Shape adapted Gradient using Local Fractal Dimension method (**BSAG-LFD** [13]), Blob Shape and Contrast (**Blob SC** [14]), Discrete Shearlet Transform using the Weibull distribution (**Shearlet-Weibull** [5]), Gabor Wavelet Transform (**GWT Weibull** [32]), Local Color Vector Patterns (**LCVP** [11]) and Multi-Scale Block Local Binary Pattern (**MB-LBP** [11]). All these feature extraction methods (with the exception of BSAG-LFD) were applied to the three RGB channels to form the final feature vector space.

For the classical features, the classification accuracy is also computed using a SVM classifier however, with the original images (without resizing) trained using the Leave-One-Patient-out cross validation strategy as in [10] to make



sure the classifier generalizes to unseen patients. This cross-validation is applied to the methods from the literature as well as to off-the-shelf CNN's features. The accuracy measure based on the percentage of images correctly classified in each class is used to allow an easy comparability of the results due to the high number of methods and databases to be compared.

4 Results and Discussion

The accuracy results for the colonic polyp classification in the 8 different databases are reported in Table 2. As can be seen, the results in Table 2 are divided into two groups: off-the-shelf features and concatenating them with state-of-the-art features.

Among the 11 pre-trained CNN investigated, the CNN that presents lower performance were GoogleLeNet, CNN-S and AlexNet MCN. These results may indicate that such networks themselves are not sufficient to be considered off-the-shelf feature extractors for the polyp classification task.

Table 2. Accuracies of the methods for the CC-i-Scan databases in %.

Methods	No staining				Staining				
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3	$\overline{X}$
1- CNN-F	86.16	89.33	80.65	88.41	86.52	81.40	84.22	80.62	84.66
2- CNN-M	87.45	90.67	81.38	83.58	87.99	89.55	87.40	90.53	87.31
3- CNN-S	88.03	90.00	87.01	77.33	87.25	82.68	87.40	75.54	84.41
4- CNN-F MCN	88.84	82.00	73.15	90.73	85.78	89.55	89.72	83.15	85.36
5- CNN-M MCN	89.53	90.67	88.88	94.66	86.97	89.29	87.40	90.53	**89.74**
6- CNN-S MCN	90.12	91.42	81.38	79.85	89.18	93.49	81.10	84.77	86.41
7- GoogleLeNet	79.65	90.67	72.43	74.51	88.27	80.46	75.60	84.08	80.70
8- VGG-VD16	87.45	85.33	86.38	79.65	92.47	89.80	95.26	92.38	88.59
9- VGG-VD19	83.49	82.67	83.88	87.71	92.47	83.98	94.46	85.59	86.78
10-AlexNet	91.40	87.33	75.65	89.32	87.71	83.03	84.22	79.24	84.73
11-AlexNet MCN	89.42	84.67	78.88	83.78	89.36	83.55	81.10	78.32	83.63
$\overline{X}$	87.41	87.70	80.88	84.50	88.54	86.07	86.17	84.06	85.67
13- Blob SC	77.67	83.33	82.10	75.22	59.28	78.83	66.13	59.83	72.79
14- Shearlet-Weibull	73.72	76.67	79.60	86.80	81.30	69.91	72.38	83.63	78.00
15- GWT-Weibull	79.75	78.67	70.25	84.28	81.30	74.54	77.17	83.39	78.66
16- LCVP	76.60	66.00	47.75	77.12	77.45	79.00	70.01	69.56	70.43
17- MB-LBP	78.26	80.67	81.38	83.37	69.29	70.60	77.22	78.32	77.38
$\overline{X}$	78.71	78.70	74.28	81.61	73.13	75.58	73.61	74.35	76.24
Concatenating 5/8	88.84	85.33	83.88	92.14	93.12	90.49	96.88	94.00	90.58
Concatenating 5/12	92.79	92.67	88.88	96.98	87.71	90.49	88.26	90.53	91.03
Concatenating 5/8/12	95.94	90.00	88.88	92.14	92.30	91.43	97.63	97.46	93.22
Concatenating 5/8/14	91.51	88.67	87.10	93.75	94.68	91.43	98.44	95.85	92.67
Concatenating 5/8/15	90.91	90.00	88.88	92.14	93.94	89.80	96.88	95.61	92.27
Concatenating 5/8/12/14	93.38	88.00	91.38	93.75	93.49	92.12	97.63	94.92	93.08
Concatenating 5/8/12/17	93.38	90.00	91.38	93.75	92.75	92.12	97.63	97.46	**93.55**

As it can be seen, the pre-trained CNN that presents the best result on average for the different imaging modalities ($\overline{X}$) is the CNN-M network trained with the MatConvNet parameters (89.74%) followed by the CNN VGG-VD16 (88.59%). These deep models with smaller filters generalize well with other datasets as it shown in [25], including texture recognition, which can explain the better results in the colonic polyp database. However, there is a high variability in the results and thus it is difficult to draw general conclusions.

Many results obtained by the pre-trained CNN's surpassed the classic feature extractors for colonic polyp classification in the literature. The database that presents the best results using off-the-shelf features is the database staining the mucosa without any i-Scan technology (88.54% on average). In the case of classical features, the database with the best result in the average is the database using the i-Scan3 technology without staining the mucosa (81.61%).

To investigate this difference in the results we asses the significance of them using the McNemar test [19]. By means of this test, we analyze if the images from a database are classified differently or similarly by the other methods. With a high accuracy it is suppose of that the methods will have a very similar response, so the significance level α must be small enough to differentiate between classifying an image as correct or incorrect.

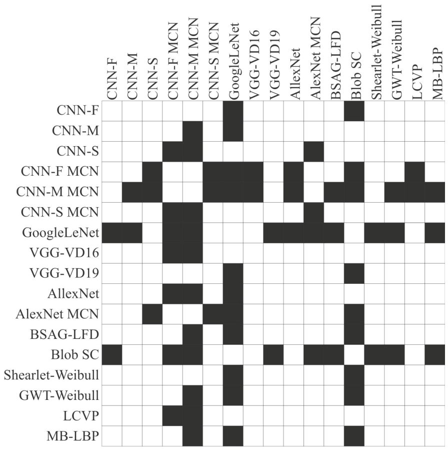
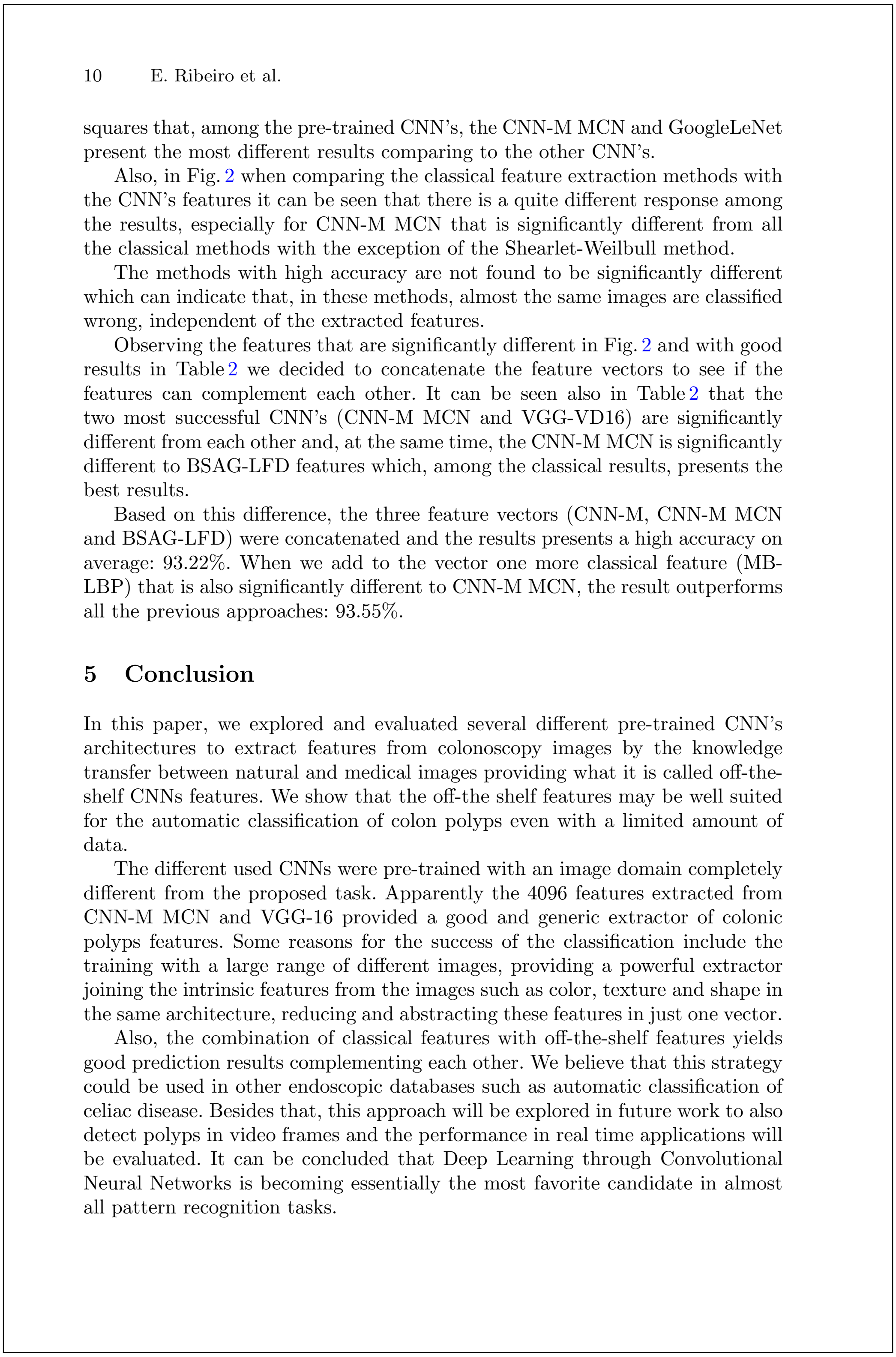


Fig. 2. Results of the McNemar test for the i-Scan3 database without staining. A black square in the matrix means that the methods are significantly different with significance level $\alpha = 0.01$. If the square is white then there is no significant difference between the methods.

The test is carried out on the database that presents the best results with the classic features (i-Scan3 without staining the mucosa) using significance level $\alpha = 0.01$. The results are presented in Fig. 2. It can be observed by the black



squares that, among the pre-trained CNN's, the CNN-M MCN and GoogleLeNet present the most different results comparing to the other CNN's.

Also, in Fig. 2 when comparing the classical feature extraction methods with the CNN's features it can be seen that there is a quite different response among the results, especially for CNN-M MCN that is significantly different from all the classical methods with the exception of the Shearlet-Weilbull method.

The methods with high accuracy are not found to be significantly different which can indicate that, in these methods, almost the same images are classified wrong, independent of the extracted features.

Observing the features that are significantly different in Fig. 2 and with good results in Table 2 we decided to concatenate the feature vectors to see if the features can complement each other. It can be seen also in Table 2 that the two most successful CNN's (CNN-M MCN and VGG-VD16) are significantly different from each other and, at the same time, the CNN-M MCN is significantly different to BSAG-LFD features which, among the classical results, presents the best results.

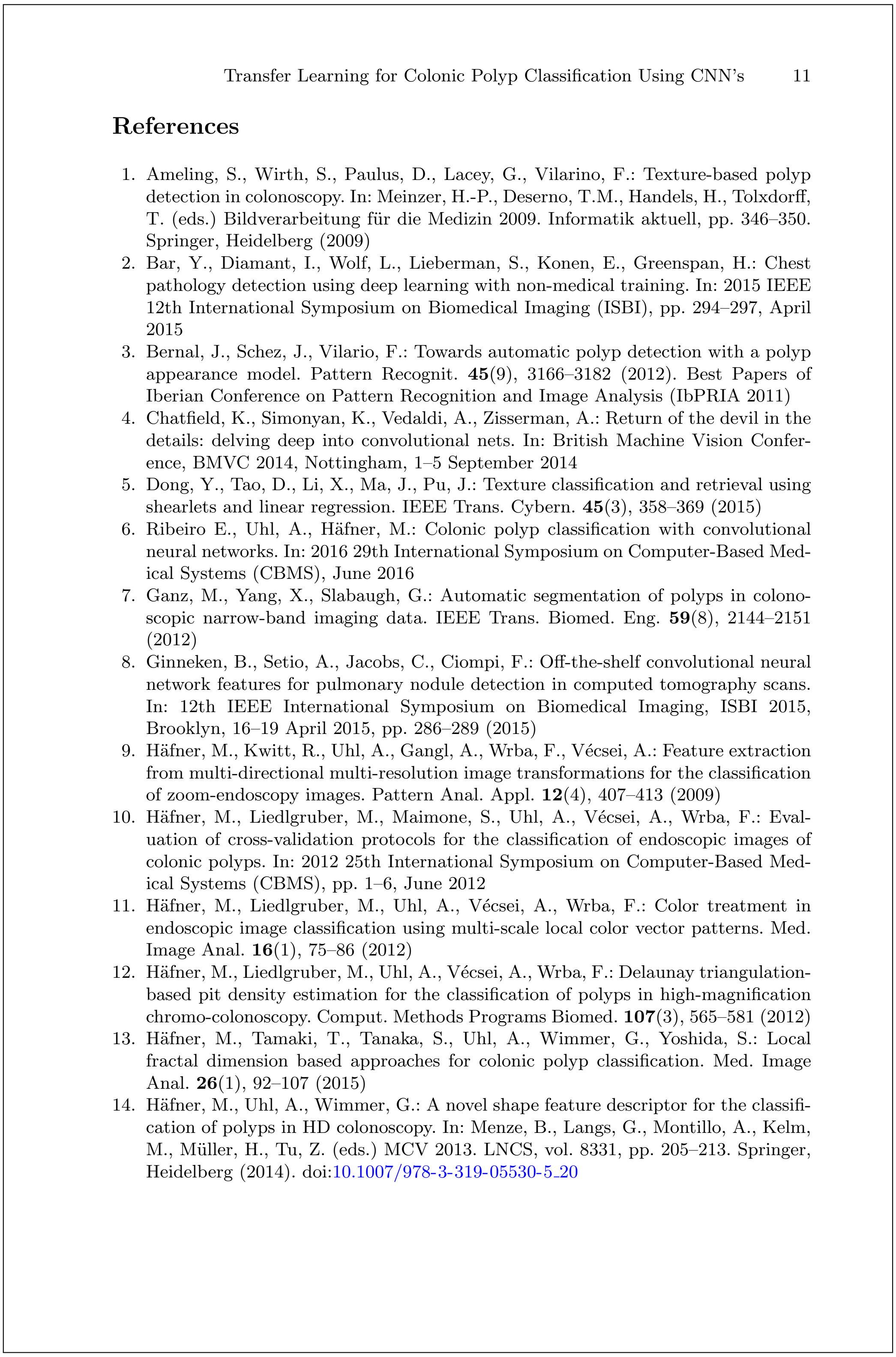
Based on this difference, the three feature vectors (CNN-M, CNN-M MCN and BSAG-LFD) were concatenated and the results presents a high accuracy on average: 93.22%. When we add to the vector one more classical feature (MB-LBP) that is also significantly different to CNN-M MCN, the result outperforms all the previous approaches: 93.55%.

5 Conclusion

In this paper, we explored and evaluated several different pre-trained CNN's architectures to extract features from colonoscopy images by the knowledge transfer between natural and medical images providing what it is called off-the-shelf CNNs features. We show that the off-the shelf features may be well suited for the automatic classification of colon polyps even with a limited amount of data.

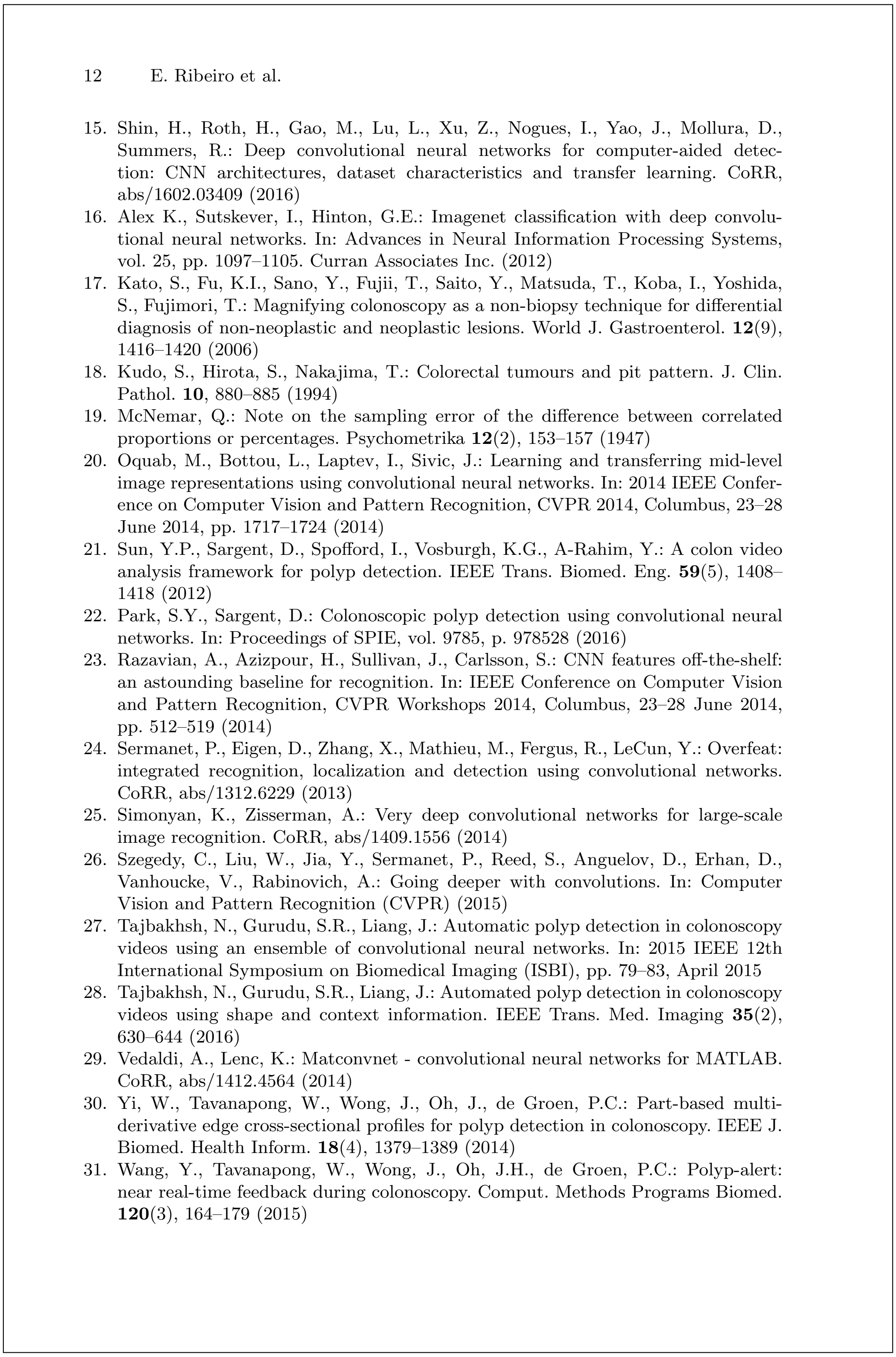
The different used CNNs were pre-trained with an image domain completely different from the proposed task. Apparently the 4096 features extracted from CNN-M MCN and VGG-16 provided a good and generic extractor of colonic polyps features. Some reasons for the success of the classification include the training with a large range of different images, providing a powerful extractor joining the intrinsic features from the images such as color, texture and shape in the same architecture, reducing and abstracting these features in just one vector.

Also, the combination of classical features with off-the-shelf features yields good prediction results complementing each other. We believe that this strategy could be used in other endoscopic databases such as automatic classification of celiac disease. Besides that, this approach will be explored in future work to also detect polyps in video frames and the performance in real time applications will be evaluated. It can be concluded that Deep Learning through Convolutional Neural Networks is becoming essentially the most favorite candidate in almost all pattern recognition tasks.

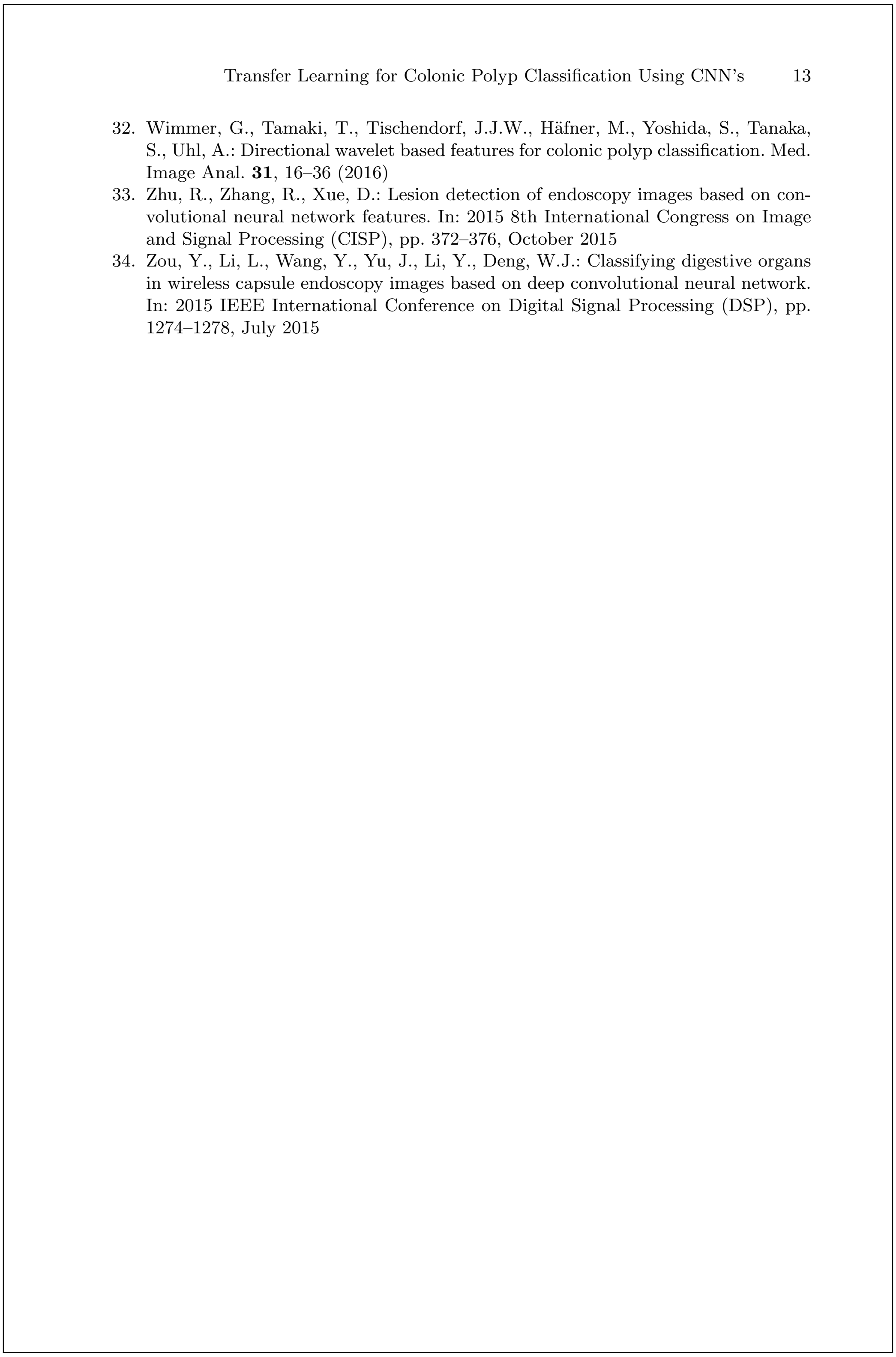


References

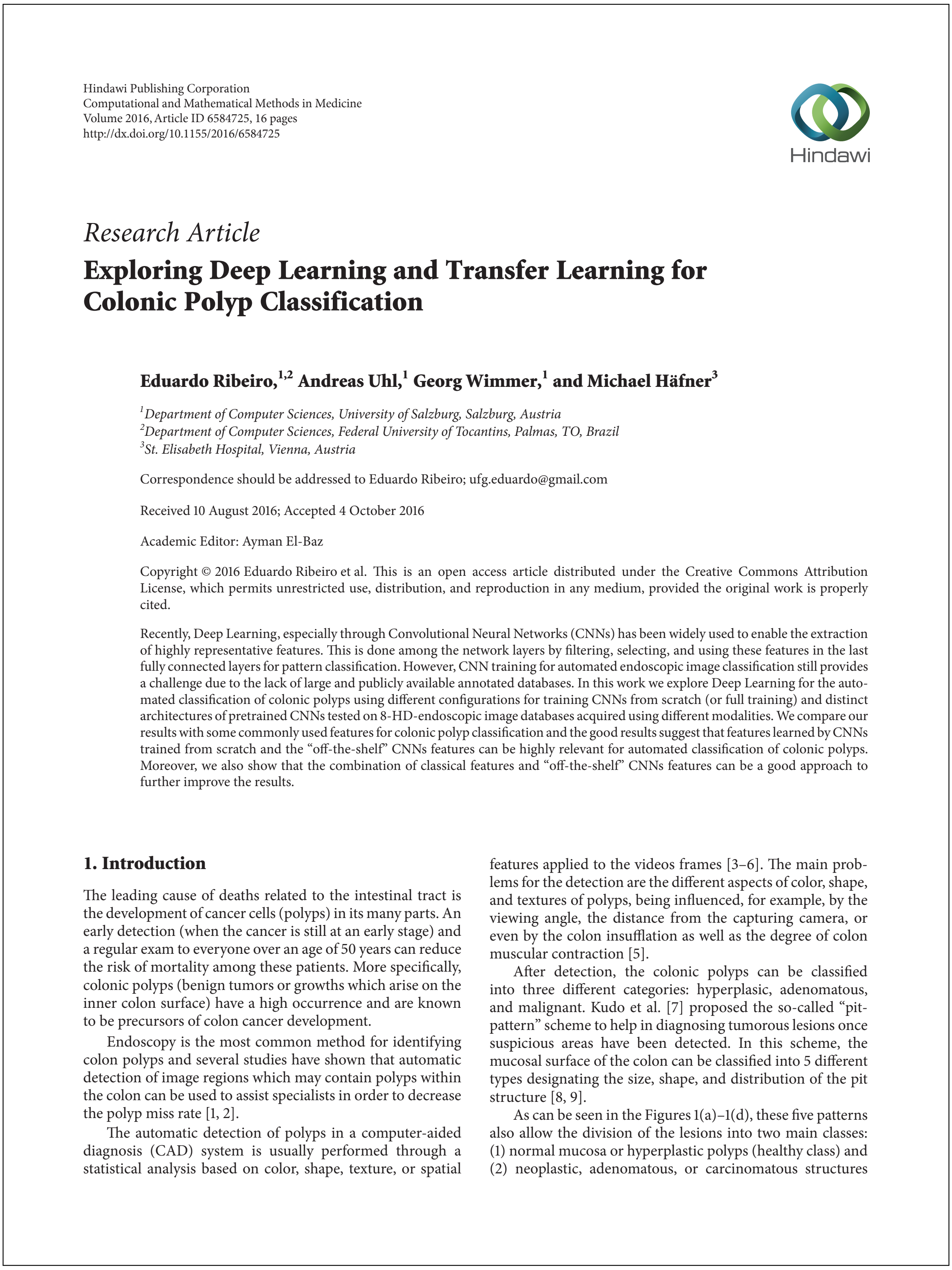
1. Ameling, S., Wirth, S., Paulus, D., Lacey, G., Vilarino, F.: Texture-based polyp detection in colonoscopy. In: Meinzer, H.-P., Deserno, T.M., Handels, H., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2009. Informatik aktuell, pp. 346–350. Springer, Heidelberg (2009)
2. Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., Greenspan, H.: Chest pathology detection using deep learning with non-medical training. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 294–297, April 2015
3. Bernal, J., Schez, J., Vilario, F.: Towards automatic polyp detection with a polyp appearance model. Pattern Recognit. **45**(9), 3166–3182 (2012). Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2011)
4. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. In: British Machine Vision Conference, BMVC 2014, Nottingham, 1–5 September 2014
5. Dong, Y., Tao, D., Li, X., Ma, J., Pu, J.: Texture classification and retrieval using shearlets and linear regression. IEEE Trans. Cybern. **45**(3), 358–369 (2015)
6. Ribeiro E., Uhl, A., Häfner, M.: Colonic polyp classification with convolutional neural networks. In: 2016 29th International Symposium on Computer-Based Medical Systems (CBMS), June 2016
7. Ganz, M., Yang, X., Slabaugh, G.: Automatic segmentation of polyps in colonoscopic narrow-band imaging data. IEEE Trans. Biomed. Eng. **59**(8), 2144–2151 (2012)
8. Ginneken, B., Setio, A., Jacobs, C., Ciompi, F.: Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In: 12th IEEE International Symposium on Biomedical Imaging, ISBI 2015, Brooklyn, 16–19 April 2015, pp. 286–289 (2015)
9. Häfner, M., Kwitt, R., Uhl, A., Gangl, A., Wrba, F., Vécsei, A.: Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images. Pattern Anal. Appl. **12**(4), 407–413 (2009)
10. Häfner, M., Liedlgruber, M., Maimone, S., Uhl, A., Vécsei, A., Wrba, F.: Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps. In: 2012 25th International Symposium on Computer-Based Medical Systems (CBMS), pp. 1–6, June 2012
11. Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F.: Color treatment in endoscopic image classification using multi-scale local color vector patterns. Med. Image Anal. **16**(1), 75–86 (2012)
12. Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., Wrba, F.: Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy. Comput. Methods Programs Biomed. **107**(3), 565–581 (2012)
13. Häfner, M., Tamaki, T., Tanaka, S., Uhl, A., Wimmer, G., Yoshida, S.: Local fractal dimension based approaches for colonic polyp classification. Med. Image Anal. **26**(1), 92–107 (2015)
14. Häfner, M., Uhl, A., Wimmer, G.: A novel shape feature descriptor for the classification of polyps in HD colonoscopy. In: Menze, B., Langs, G., Montillo, A., Kelm, M., Müller, H., Tu, Z. (eds.) MCV 2013. LNCS, vol. 8331, pp. 205–213. Springer, Heidelberg (2014). doi:10.1007/978-3-319-05530-5_20



15. Shin, H., Roth, H., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. CoRR, abs/1602.03409 (2016)
16. Alex K., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25, pp. 1097–1105. Curran Associates Inc. (2012)
17. Kato, S., Fu, K.I., Sano, Y., Fujii, T., Saito, Y., Matsuda, T., Koba, I., Yoshida, S., Fujimori, T.: Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions. World J. Gastroenterol. **12**(9), 1416–1420 (2006)
18. Kudo, S., Hirota, S., Nakajima, T.: Colorectal tumours and pit pattern. J. Clin. Pathol. **10**, 880–885 (1994)
19. McNemar, Q.: Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika **12**(2), 153–157 (1947)
20. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, 23–28 June 2014, pp. 1717–1724 (2014)
21. Sun, Y.P., Sargent, D., Spofford, I., Vosburgh, K.G., A-Rahim, Y.: A colon video analysis framework for polyp detection. IEEE Trans. Biomed. Eng. **59**(5), 1408–1418 (2012)
22. Park, S.Y., Sargent, D.: Colonoscopic polyp detection using convolutional neural networks. In: Proceedings of SPIE, vol. 9785, p. 978528 (2016)
23. Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2014, Columbus, 23–28 June 2014, pp. 512–519 (2014)
24. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: integrated recognition, localization and detection using convolutional networks. CoRR, abs/1312.6229 (2013)
25. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014)
26. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Computer Vision and Pattern Recognition (CVPR) (2015)
27. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 79–83, April 2015
28. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE Trans. Med. Imaging **35**(2), 630–644 (2016)
29. Vedaldi, A., Lenc, K.: Matconvnet - convolutional neural networks for MATLAB. CoRR, abs/1412.4564 (2014)
30. Yi, W., Tavanapong, W., Wong, J., Oh, J., de Groen, P.C.: Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy. IEEE J. Biomed. Health Inform. **18**(4), 1379–1389 (2014)
31. Wang, Y., Tavanapong, W., Wong, J., Oh, J.H., de Groen, P.C.: Polyp-alert: near real-time feedback during colonoscopy. Comput. Methods Programs Biomed. **120**(3), 164–179 (2015)



32. Wimmer, G., Tamaki, T., Tischendorf, J.J.W., Häfner, M., Yoshida, S., Tanaka, S., Uhl, A.: Directional wavelet based features for colonic polyp classification. Med. Image Anal. **31**, 16–36 (2016)
33. Zhu, R., Zhang, R., Xue, D.: Lesion detection of endoscopy images based on convolutional neural network features. In: 2015 8th International Congress on Image and Signal Processing (CISP), pp. 372–376, October 2015
34. Zou, Y., Li, L., Wang, Y., Yu, J., Li, Y., Deng, W.J.: Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network. In: 2015 IEEE International Conference on Digital Signal Processing (DSP), pp. 1274–1278, July 2015



Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2016, Article ID 6584725, 16 pages
http://dx.doi.org/10.1155/2016/6584725

Research Article

Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification

Eduardo Ribeiro,[1,2] Andreas Uhl,[1] Georg Wimmer,[1] and Michael Häfner[3]

[1] *Department of Computer Sciences, University of Salzburg, Salzburg, Austria*
[2] *Department of Computer Sciences, Federal University of Tocantins, Palmas, TO, Brazil*
[3] *St. Elisabeth Hospital, Vienna, Austria*

Correspondence should be addressed to Eduardo Ribeiro; ufg.eduardo@gmail.com

Received 10 August 2016; Accepted 4 October 2016

Academic Editor: Ayman El-Baz

Recently, Deep Learning, especially through Convolutional Neural Networks (CNNs) has been widely used to enable the extraction of highly representative features. This is done among the network layers by filtering, selecting, and using these features in the last fully connected layers for pattern classification. However, CNN training for automated endoscopic image classification still provides a challenge due to the lack of large and publicly available annotated databases. In this work we explore Deep Learning for the automated classification of colonic polyps using different configurations for training CNNs from scratch (or full training) and distinct architectures of pretrained CNNs tested on 8-HD-endoscopic image databases acquired using different modalities. We compare our results with some commonly used features for colonic polyp classification and the good results suggest that features learned by CNNs trained from scratch and the "off-the-shelf" CNNs features can be highly relevant for automated classification of colonic polyps. Moreover, we also show that the combination of classical features and "off-the-shelf" CNNs features can be a good approach to further improve the results.

1. Introduction

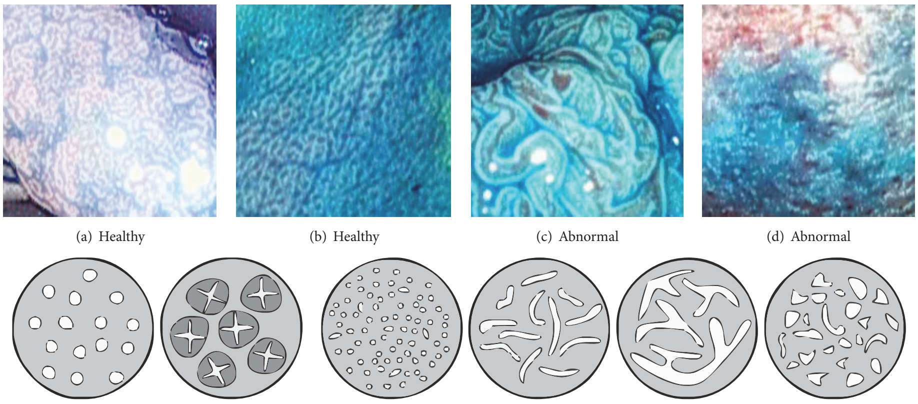
The leading cause of deaths related to the intestinal tract is the development of cancer cells (polyps) in its many parts. An early detection (when the cancer is still at an early stage) and a regular exam to everyone over an age of 50 years can reduce the risk of mortality among these patients. More specifically, colonic polyps (benign tumors or growths which arise on the inner colon surface) have a high occurrence and are known to be precursors of colon cancer development.

Endoscopy is the most common method for identifying colon polyps and several studies have shown that automatic detection of image regions which may contain polyps within the colon can be used to assist specialists in order to decrease the polyp miss rate [1, 2].

The automatic detection of polyps in a computer-aided diagnosis (CAD) system is usually performed through a statistical analysis based on color, shape, texture, or spatial features applied to the videos frames [3–6]. The main problems for the detection are the different aspects of color, shape, and textures of polyps, being influenced, for example, by the viewing angle, the distance from the capturing camera, or even by the colon insufflation as well as the degree of colon muscular contraction [5].

After detection, the colonic polyps can be classified into three different categories: hyperplasic, adenomatous, and malignant. Kudo et al. [7] proposed the so-called "pit-pattern" scheme to help in diagnosing tumorous lesions once suspicious areas have been detected. In this scheme, the mucosal surface of the colon can be classified into 5 different types designating the size, shape, and distribution of the pit structure [8, 9].

As can be seen in the Figures 1(a)–1(d), these five patterns also allow the division of the lesions into two main classes: (1) normal mucosa or hyperplastic polyps (healthy class) and (2) neoplastic, adenomatous, or carcinomatous structures



(a) Healthy (b) Healthy (c) Abnormal (d) Abnormal

(e) Healthy (f) Abnormal

Figure 1: Example images of the two classes (a–d) and the pit-pattern types of these two classes (e–f).

(abnormal class). This approach is quite relevant in clinical practice as shown in a study by Kato et al. [10].

In the literature, existing computer-aided diagnosis techniques generally make use of feature extraction methods of color, shape, and texture in combination with machine learning classifiers to perform the classification of colon polyps [9, 11, 12]. For example, the dual-tree complex wavelet transform DT-CWT features proved to be quite suitable for the distinction of different types of polyps as can be seen in many works like, for example, [13–15]. Other features were also proved to be quite suitable for colonic polyp classification as the Gabor wavelets [16], vascularization features [17], and directional wavelet transform features [18]. Particularly, in the work of Wimmer et al. [18], using the same 8 colonic polyp databases of this work, an average accuracy of 80.3% was achieved in the best scenario. In this work, we achieve an average accuracy of 93.55% in our best scenario.

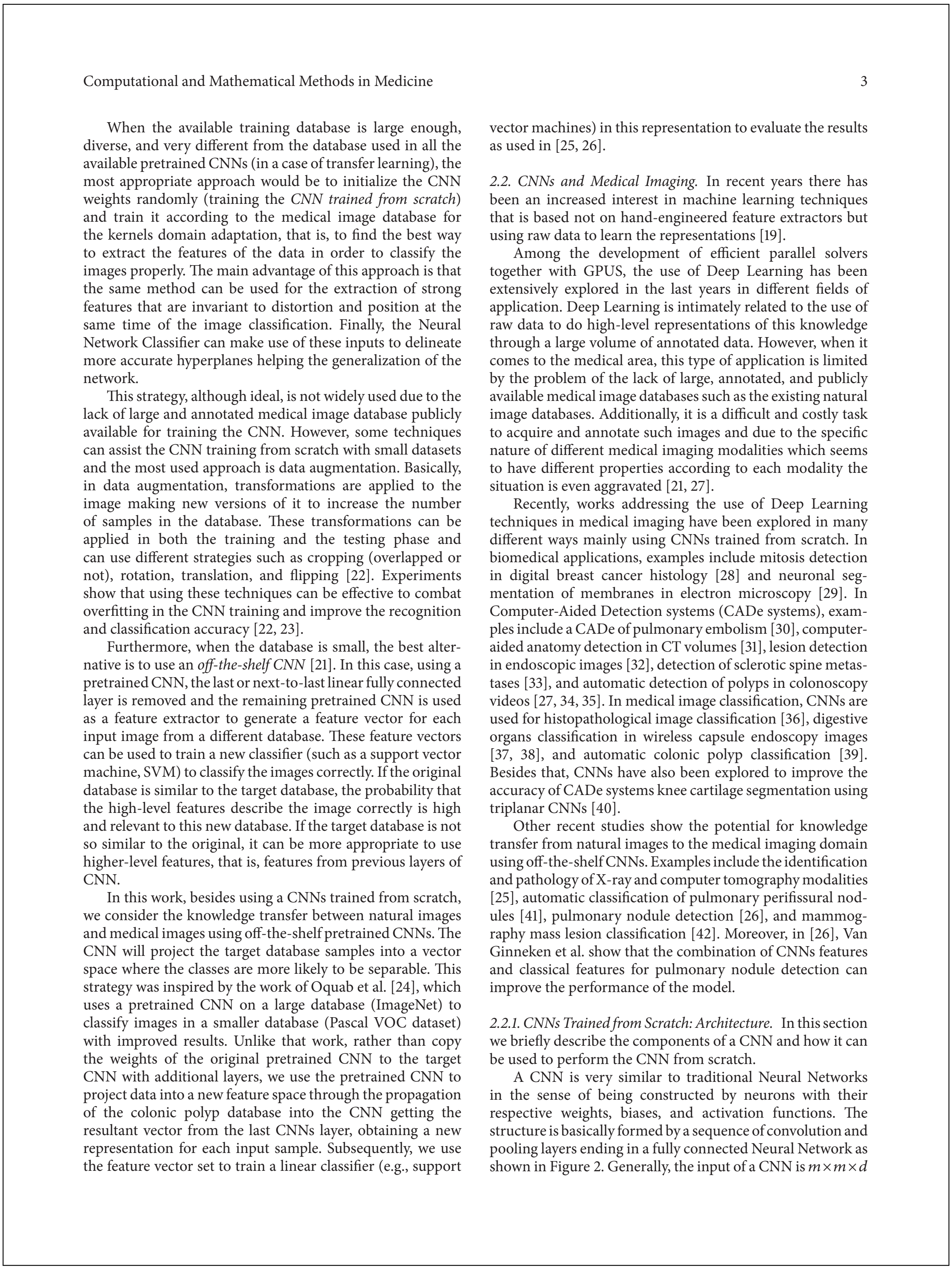
The main difficulty of the feature extraction methods is the proper characterization of these patterns due to several factors as the lack or excess of illumination, the blurring due to movement or water injection, and the appearance of polyps [5, 9]. Also, to find a robust and a global feature extractor that summarizes and represents all these pit-pattern structures in a single vector is very difficult and Deep Learning can be a good alternative to surpass these problems. In this work we explore the use of Deep Learning through Convolutional Neural Networks (CNNs) to develop a model for robust feature extraction and efficient colonic polyp classification.

To achieve this, we test the use of CNNs trained from scratch (or full training) and off-the-shelf CNNs (or pre-trained) using them as medical imaging feature extractors. In the case of the CNN full training we assume that a feature extractor is formed during the CNN training, adapting to the context of the database and particularly in the case of off-the-shelf CNNs we consider that the patterns learned in the original database can be used in colonoscopy images for colonic polyp classification. In particular, we explore two different architectures for the training from scratch and six different off-the-shelf architectures, describing and analyzing the effects of CNNs in different acquisition modes of colonoscopy images (8 different databases). This study was motivated by recent studies in computer vision addressing the emerging technique of Deep Learning presented in the next section.

2. Materials and Methods

2.1. Using CNNs on Small Datasets. Some researchers propose replacing handcrafted feature extraction algorithms with Deep Learning approaches that act as features extractor and image classifier at the same time [19]. For example, the Deep Learning approach using CNNs takes advantage of many consecutive convolutional layers followed by pooling layers to reduce the data dimensionality making it, concomitantly, invariant to geometric transformations. Such convolution filters (kernels) are built to act as feature extractors during the training process and recent research indicates that a satisfactorily trained CNN with a large database can perform properly when it is applied to other databases, which can mean that the kernels can turn into a universal feature extractor [19]. Also, Convolutional Neural Networks (CNNs) have been demonstrated to be effective for discriminative pattern recognition in big data and in real-world problems, mainly to learn both the global and local structures of images [20].

Many strategies exploiting CNNs can be used for medical image classification. These strategies can be employed according to the intrinsic characteristics of each database [21] and two of them, mostly used when it comes to CNN training, are described in the following part.



When the available training database is large enough, diverse, and very different from the database used in all the available pretrained CNNs (in a case of transfer learning), the most appropriate approach would be to initialize the CNN weights randomly (training the *CNN trained from scratch*) and train it according to the medical image database for the kernels domain adaptation, that is, to find the best way to extract the features of the data in order to classify the images properly. The main advantage of this approach is that the same method can be used for the extraction of strong features that are invariant to distortion and position at the same time of the image classification. Finally, the Neural Network Classifier can make use of these inputs to delineate more accurate hyperplanes helping the generalization of the network.

This strategy, although ideal, is not widely used due to the lack of large and annotated medical image database publicly available for training the CNN. However, some techniques can assist the CNN training from scratch with small datasets and the most used approach is data augmentation. Basically, in data augmentation, transformations are applied to the image making new versions of it to increase the number of samples in the database. These transformations can be applied in both the training and the testing phase and can use different strategies such as cropping (overlapped or not), rotation, translation, and flipping [22]. Experiments show that using these techniques can be effective to combat overfitting in the CNN training and improve the recognition and classification accuracy [22, 23].

Furthermore, when the database is small, the best alternative is to use an *off-the-shelf CNN* [21]. In this case, using a pretrained CNN, the last or next-to-last linear fully connected layer is removed and the remaining pretrained CNN is used as a feature extractor to generate a feature vector for each input image from a different database. These feature vectors can be used to train a new classifier (such as a support vector machine, SVM) to classify the images correctly. If the original database is similar to the target database, the probability that the high-level features describe the image correctly is high and relevant to this new database. If the target database is not so similar to the original, it can be more appropriate to use higher-level features, that is, features from previous layers of CNN.

In this work, besides using a CNNs trained from scratch, we consider the knowledge transfer between natural images and medical images using off-the-shelf pretrained CNNs. The CNN will project the target database samples into a vector space where the classes are more likely to be separable. This strategy was inspired by the work of Oquab et al. [24], which uses a pretrained CNN on a large database (ImageNet) to classify images in a smaller database (Pascal VOC dataset) with improved results. Unlike that work, rather than copy the weights of the original pretrained CNN to the target CNN with additional layers, we use the pretrained CNN to project data into a new feature space through the propagation of the colonic polyp database into the CNN getting the resultant vector from the last CNNs layer, obtaining a new representation for each input sample. Subsequently, we use the feature vector set to train a linear classifier (e.g., support vector machines) in this representation to evaluate the results as used in [25, 26].

2.2. CNNs and Medical Imaging. In recent years there has been an increased interest in machine learning techniques that is based not on hand-engineered feature extractors but using raw data to learn the representations [19].

Among the development of efficient parallel solvers together with GPUS, the use of Deep Learning has been extensively explored in the last years in different fields of application. Deep Learning is intimately related to the use of raw data to do high-level representations of this knowledge through a large volume of annotated data. However, when it comes to the medical area, this type of application is limited by the problem of the lack of large, annotated, and publicly available medical image databases such as the existing natural image databases. Additionally, it is a difficult and costly task to acquire and annotate such images and due to the specific nature of different medical imaging modalities which seems to have different properties according to each modality the situation is even aggravated [21, 27].

Recently, works addressing the use of Deep Learning techniques in medical imaging have been explored in many different ways mainly using CNNs trained from scratch. In biomedical applications, examples include mitosis detection in digital breast cancer histology [28] and neuronal segmentation of membranes in electron microscopy [29]. In Computer-Aided Detection systems (CADe systems), examples include a CADe of pulmonary embolism [30], computer-aided anatomy detection in CT volumes [31], lesion detection in endoscopic images [32], detection of sclerotic spine metastases [33], and automatic detection of polyps in colonoscopy videos [27, 34, 35]. In medical image classification, CNNs are used for histopathological image classification [36], digestive organs classification in wireless capsule endoscopy images [37, 38], and automatic colonic polyp classification [39]. Besides that, CNNs have also been explored to improve the accuracy of CADe systems knee cartilage segmentation using triplanar CNNs [40].

Other recent studies show the potential for knowledge transfer from natural images to the medical imaging domain using off-the-shelf CNNs. Examples include the identification and pathology of X-ray and computer tomography modalities [25], automatic classification of pulmonary perifissural nodules [41], pulmonary nodule detection [26], and mammography mass lesion classification [42]. Moreover, in [26], Van Ginneken et al. show that the combination of CNNs features and classical features for pulmonary nodule detection can improve the performance of the model.

2.2.1. CNNs Trained from Scratch: Architecture. In this section we briefly describe the components of a CNN and how it can be used to perform the CNN from scratch.

A CNN is very similar to traditional Neural Networks in the sense of being constructed by neurons with their respective weights, biases, and activation functions. The structure is basically formed by a sequence of convolution and pooling layers ending in a fully connected Neural Network as shown in Figure 2. Generally, the input of a CNN is $m \times m \times d$

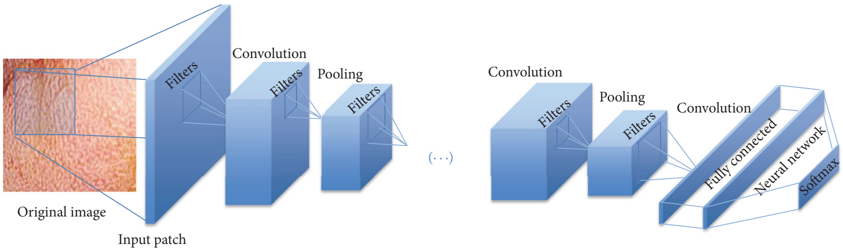


Figure 2: An illustration of the CNN architecture for colonic polyp classification.

image (or patch), where $m \times m$ is the dimension of the image and d is the number of channels (depth) of the image. The convolutional layer consists of k learnable filters (also called kernels) with size $n \times n \times d$ where $n \leq m$ which are convolved with the input image resulting in the so-called activation maps or feature maps. As classic Neural Networks, the convolution layer outputs are submitted to an activation function, for example, the ReLU rectifier function $f(x) = \max(0, x)$, where x is the neuron input. After the convolution, a pooling layer is included to subsample the image by average functions (mean) or max-pooling over regions of size $p \times p$. These functions are used to reduce the dimensionality of the data in the following layers (upper layers) and to provide a form of invariance to translation thus making overfitting control. In the convolution and pooling layers the stride has to be specified; the larger the stride, the smaller the overlapping, decreasing the output volume dimensions.

At the end of the CNN there is a fully connected layer as a regular Multilayer Neural Network with the Softmax function that generates a well-formed probability distribution on the outputs. After a supervised training, the CNN is ready to be used as a classifier or as a feature extractor in the case of transfer learning.

2.2.2. CNNs and Transfer Learning. Transfer learning is a technique used to improve the performance of machine learning by harnessing the knowledge obtained by another task. According to Pan and Yang [43], transfer learning can be defined by the following model. We give a domain D having two components: a feature space $X = \{x_1, x_2, \ldots, x_n\}$ and a probabilistic distribution $P(X)$; that is, $D = \{X, P(X)\}$. Also, we give a task T with two components: a ground truth $Y = \{y_1, y_2, \ldots, y_n\}$ and an objective function $T = \{Y, f(\cdot)\}$ assuming that this function can be learned through a training database. Function $f(\cdot)$ can be used to predict the corresponding class $f(x)$ of a new instance x. From a probabilistic point of view, $f(x)$ can be written as $P(y \mid x)$. In colonic polyp classification, usually, a feature extractor is used to generate the feature space. A given training database X associated to the ground truth Y consisting of the pairs $\{x_i, y_i\}$ is used to train and "learn" the function $f(\cdot)$ or $P(y \mid x)$ until it reaches a defined and acceptable error rate between the result of the function $f(x)$ and the ground truth Y.

In case of transfer learning, given a source domain $D_S = \{(x_{S_1}, y_{S_1}), (x_{S_2}, y_{S_2}), \ldots, (x_{S_n}, y_{S_n})\}$ and the learning task T_S and the target domain $D_T = \{(x_{T_1}, y_{T_1}), (x_{T_2}, y_{T_2}), \ldots, (x_{T_m}, y_{T_m})\}$ and the learning task T_T, transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ using the knowledge in D_S and T_S, where $D_T \neq D_S$ and $T_T \neq T_S$.

Among the various categories of transfer learning, one, called inductive transfer learning, has been used with success in the pattern recognition area. In the inductive transfer learning approach an annotated database is necessary for the source domain as well as for the target domain. In this work, we apply transfer learning between two very different tasks using different labels ($Y_T \neq Y_S$) and different distributions ($P(Y_T \mid X_T) \neq P(Y_S \mid X_S)$). To bypass the difference between the probability distribution of the images $P(X_S)$, the last layer from the original function $f_S(\cdot)$ directly connected to the classification is removed being replaced by other linear function (as SVM) to adapt it to the new task T_T turning into the function $f_T(\cdot)$. In the following sections the functions $f_S(\cdot)$ used in this work are presented. Also, the use of transfer learning using pretrained CNNs can help to avoid the problem of lack of data in the medical field. The works of Razavian et al. [19] and Oquab et al. [24] suggest that the use of CNNs intermediate layer outputs can be used as input features to train other classifiers (such as support vector machines) for a number of other applications different from the original CNN obtaining a good performance.

Despite the difference between natural and medical images, some feature descriptors designed especially for natural images are used successfully in medical image detection and classification, for example, texture-based polyp detection [3], Fourier and Wavelet filters for colon classification [18], shape descriptors [44], and local fractal dimension [45] for colonic polyp classification. Additionally, recent studies show the potential of the knowledge transfer between natural and medical images using pretrained (off-the-shelf) CNNs [34, 46].

2.3. Experimental Setup

2.3.1. Data. The use of an integrated endoscopic apparatus with high-resolution acquisition devices has been an

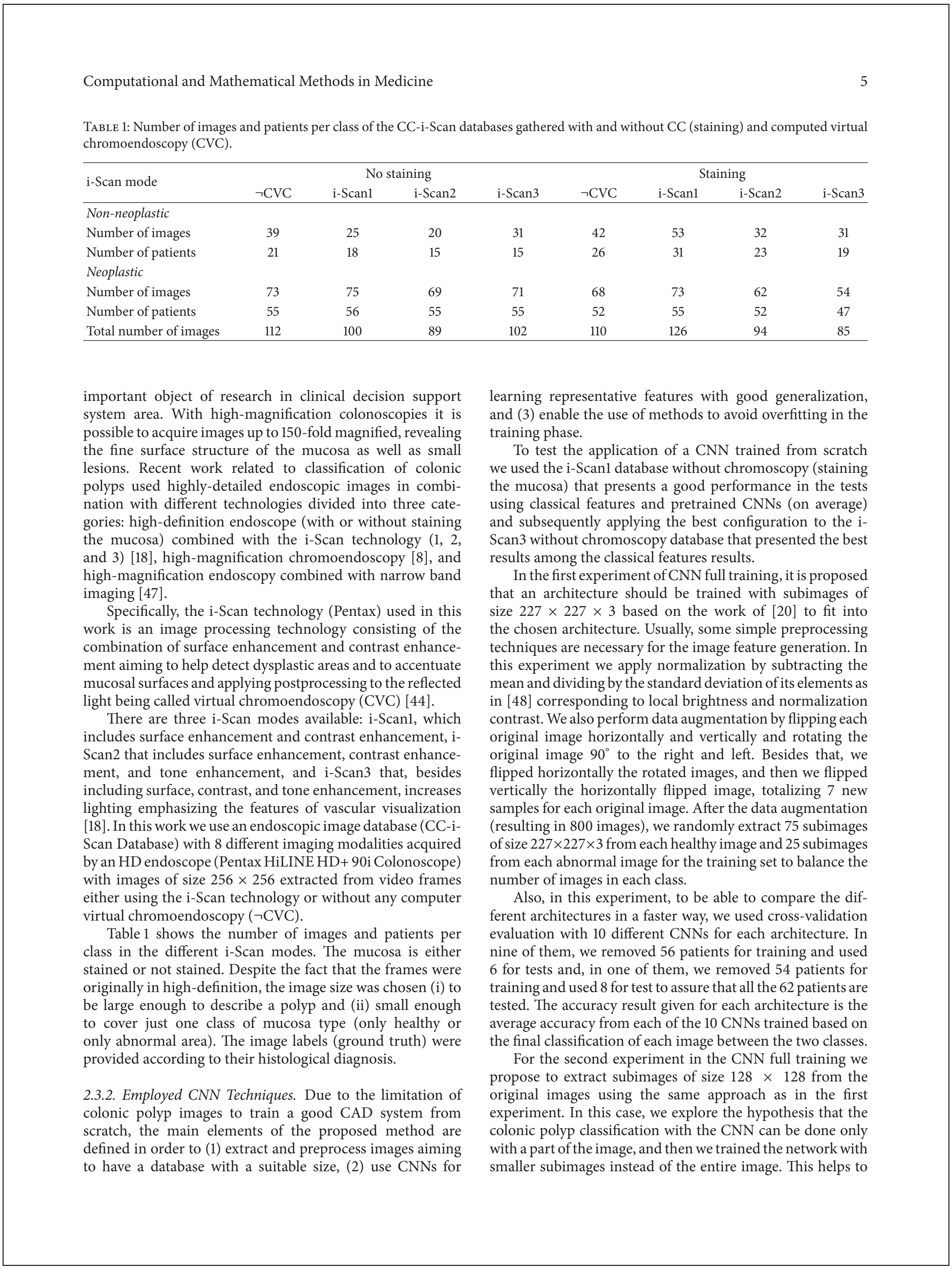


Table 1: Number of images and patients per class of the CC-i-Scan databases gathered with and without CC (staining) and computed virtual chromoendoscopy (CVC).

i-Scan mode	No staining				Staining			
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3
Non-neoplastic								
Number of images	39	25	20	31	42	53	32	31
Number of patients	21	18	15	15	26	31	23	19
Neoplastic								
Number of images	73	75	69	71	68	73	62	54
Number of patients	55	56	55	55	52	55	52	47
Total number of images	112	100	89	102	110	126	94	85

important object of research in clinical decision support system area. With high-magnification colonoscopies it is possible to acquire images up to 150-fold magnified, revealing the fine surface structure of the mucosa as well as small lesions. Recent work related to classification of colonic polyps used highly-detailed endoscopic images in combination with different technologies divided into three categories: high-definition endoscope (with or without staining the mucosa) combined with the i-Scan technology (1, 2, and 3) [18], high-magnification chromoendoscopy [8], and high-magnification endoscopy combined with narrow band imaging [47].

Specifically, the i-Scan technology (Pentax) used in this work is an image processing technology consisting of the combination of surface enhancement and contrast enhancement aiming to help detect dysplastic areas and to accentuate mucosal surfaces and applying postprocessing to the reflected light being called virtual chromoendoscopy (CVC) [44].

There are three i-Scan modes available: i-Scan1, which includes surface enhancement and contrast enhancement, i-Scan2 that includes surface enhancement, contrast enhancement, and tone enhancement, and i-Scan3 that, besides including surface, contrast, and tone enhancement, increases lighting emphasizing the features of vascular visualization [18]. In this work we use an endoscopic image database (CC-i-Scan Database) with 8 different imaging modalities acquired by an HD endoscope (Pentax HiLINE HD+ 90i Colonoscope) with images of size 256 × 256 extracted from video frames either using the i-Scan technology or without any computer virtual chromoendoscopy (¬CVC).

Table 1 shows the number of images and patients per class in the different i-Scan modes. The mucosa is either stained or not stained. Despite the fact that the frames were originally in high-definition, the image size was chosen (i) to be large enough to describe a polyp and (ii) small enough to cover just one class of mucosa type (only healthy or only abnormal area). The image labels (ground truth) were provided according to their histological diagnosis.

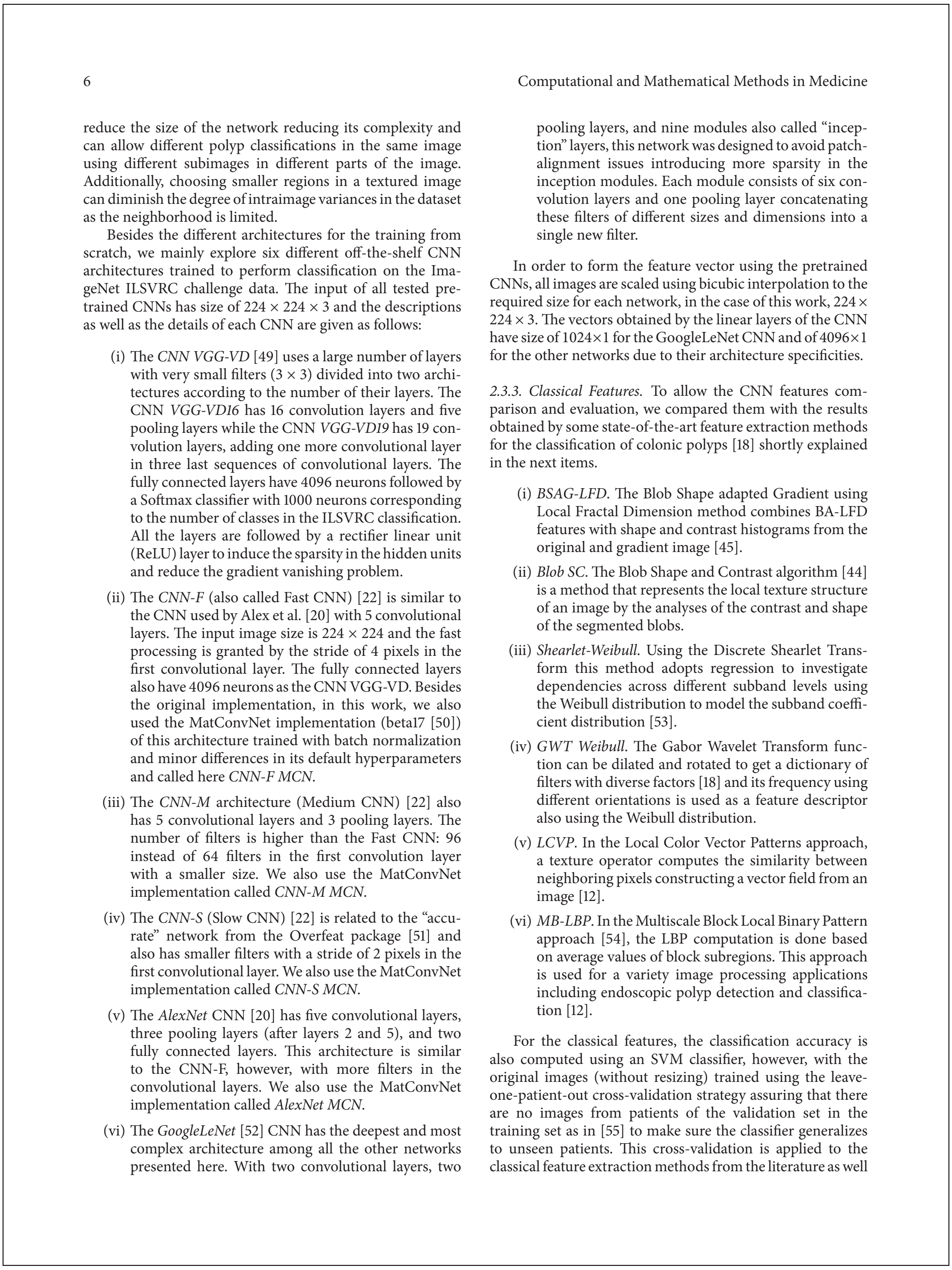
2.3.2. Employed CNN Techniques. Due to the limitation of colonic polyp images to train a good CAD system from scratch, the main elements of the proposed method are defined in order to (1) extract and preprocess images aiming to have a database with a suitable size, (2) use CNNs for learning representative features with good generalization, and (3) enable the use of methods to avoid overfitting in the training phase.

To test the application of a CNN trained from scratch we used the i-Scan1 database without chromoscopy (staining the mucosa) that presents a good performance in the tests using classical features and pretrained CNNs (on average) and subsequently applying the best configuration to the i-Scan3 without chromoscopy database that presented the best results among the classical features results.

In the first experiment of CNN full training, it is proposed that an architecture should be trained with subimages of size 227 × 227 × 3 based on the work of [20] to fit into the chosen architecture. Usually, some simple preprocessing techniques are necessary for the image feature generation. In this experiment we apply normalization by subtracting the mean and dividing by the standard deviation of its elements as in [48] corresponding to local brightness and normalization contrast. We also perform data augmentation by flipping each original image horizontally and vertically and rotating the original image 90° to the right and left. Besides that, we flipped horizontally the rotated images, and then we flipped vertically the horizontally flipped image, totalizing 7 new samples for each original image. After the data augmentation (resulting in 800 images), we randomly extract 75 subimages of size 227×227×3 from each healthy image and 25 subimages from each abnormal image for the training set to balance the number of images in each class.

Also, in this experiment, to be able to compare the different architectures in a faster way, we used cross-validation evaluation with 10 different CNNs for each architecture. In nine of them, we removed 56 patients for training and used 6 for tests and, in one of them, we removed 54 patients for training and used 8 for test to assure that all the 62 patients are tested. The accuracy result given for each architecture is the average accuracy from each of the 10 CNNs trained based on the final classification of each image between the two classes.

For the second experiment in the CNN full training we propose to extract subimages of size 128 × 128 from the original images using the same approach as in the first experiment. In this case, we explore the hypothesis that the colonic polyp classification with the CNN can be done only with a part of the image, and then we trained the network with smaller subimages instead of the entire image. This helps to



reduce the size of the network reducing its complexity and can allow different polyp classifications in the same image using different subimages in different parts of the image. Additionally, choosing smaller regions in a textured image can diminish the degree of intraimage variances in the dataset as the neighborhood is limited.

Besides the different architectures for the training from scratch, we mainly explore six different off-the-shelf CNN architectures trained to perform classification on the ImageNet ILSVRC challenge data. The input of all tested pretrained CNNs has size of 224 × 224 × 3 and the descriptions as well as the details of each CNN are given as follows:

(i) The *CNN VGG-VD* [49] uses a large number of layers with very small filters (3 × 3) divided into two architectures according to the number of their layers. The CNN *VGG-VD16* has 16 convolution layers and five pooling layers while the CNN *VGG-VD19* has 19 convolution layers, adding one more convolutional layer in three last sequences of convolutional layers. The fully connected layers have 4096 neurons followed by a Softmax classifier with 1000 neurons corresponding to the number of classes in the ILSVRC classification. All the layers are followed by a rectifier linear unit (ReLU) layer to induce the sparsity in the hidden units and reduce the gradient vanishing problem.

(ii) The *CNN-F* (also called Fast CNN) [22] is similar to the CNN used by Alex et al. [20] with 5 convolutional layers. The input image size is 224 × 224 and the fast processing is granted by the stride of 4 pixels in the first convolutional layer. The fully connected layers also have 4096 neurons as the CNN VGG-VD. Besides the original implementation, in this work, we also used the MatConvNet implementation (beta17 [50]) of this architecture trained with batch normalization and minor differences in its default hyperparameters and called here *CNN-F MCN*.

(iii) The *CNN-M* architecture (Medium CNN) [22] also has 5 convolutional layers and 3 pooling layers. The number of filters is higher than the Fast CNN: 96 instead of 64 filters in the first convolution layer with a smaller size. We also use the MatConvNet implementation called *CNN-M MCN*.

(iv) The *CNN-S* (Slow CNN) [22] is related to the "accurate" network from the Overfeat package [51] and also has smaller filters with a stride of 2 pixels in the first convolutional layer. We also use the MatConvNet implementation called *CNN-S MCN*.

(v) The *AlexNet* CNN [20] has five convolutional layers, three pooling layers (after layers 2 and 5), and two fully connected layers. This architecture is similar to the CNN-F, however, with more filters in the convolutional layers. We also use the MatConvNet implementation called *AlexNet MCN*.

(vi) The *GoogleLeNet* [52] CNN has the deepest and most complex architecture among all the other networks presented here. With two convolutional layers, two pooling layers, and nine modules also called "inception" layers, this network was designed to avoid patch-alignment issues introducing more sparsity in the inception modules. Each module consists of six convolution layers and one pooling layer concatenating these filters of different sizes and dimensions into a single new filter.

In order to form the feature vector using the pretrained CNNs, all images are scaled using bicubic interpolation to the required size for each network, in the case of this work, 224 × 224 × 3. The vectors obtained by the linear layers of the CNN have size of 1024×1 for the GoogleLeNet CNN and of 4096×1 for the other networks due to their architecture specificities.

2.3.3. Classical Features. To allow the CNN features comparison and evaluation, we compared them with the results obtained by some state-of-the-art feature extraction methods for the classification of colonic polyps [18] shortly explained in the next items.

(i) *BSAG-LFD*. The Blob Shape adapted Gradient using Local Fractal Dimension method combines BA-LFD features with shape and contrast histograms from the original and gradient image [45].

(ii) *Blob SC*. The Blob Shape and Contrast algorithm [44] is a method that represents the local texture structure of an image by the analyses of the contrast and shape of the segmented blobs.

(iii) *Shearlet-Weibull*. Using the Discrete Shearlet Transform this method adopts regression to investigate dependencies across different subband levels using the Weibull distribution to model the subband coefficient distribution [53].

(iv) *GWT Weibull*. The Gabor Wavelet Transform function can be dilated and rotated to get a dictionary of filters with diverse factors [18] and its frequency using different orientations is used as a feature descriptor also using the Weibull distribution.

(v) *LCVP*. In the Local Color Vector Patterns approach, a texture operator computes the similarity between neighboring pixels constructing a vector field from an image [12].

(vi) *MB-LBP*. In the Multiscale Block Local Binary Pattern approach [54], the LBP computation is done based on average values of block subregions. This approach is used for a variety image processing applications including endoscopic polyp detection and classification [12].

For the classical features, the classification accuracy is also computed using an SVM classifier, however, with the original images (without resizing) trained using the leave-one-patient-out cross-validation strategy assuring that there are no images from patients of the validation set in the training set as in [55] to make sure the classifier generalizes to unseen patients. This cross-validation is applied to the classical feature extraction methods from the literature as well

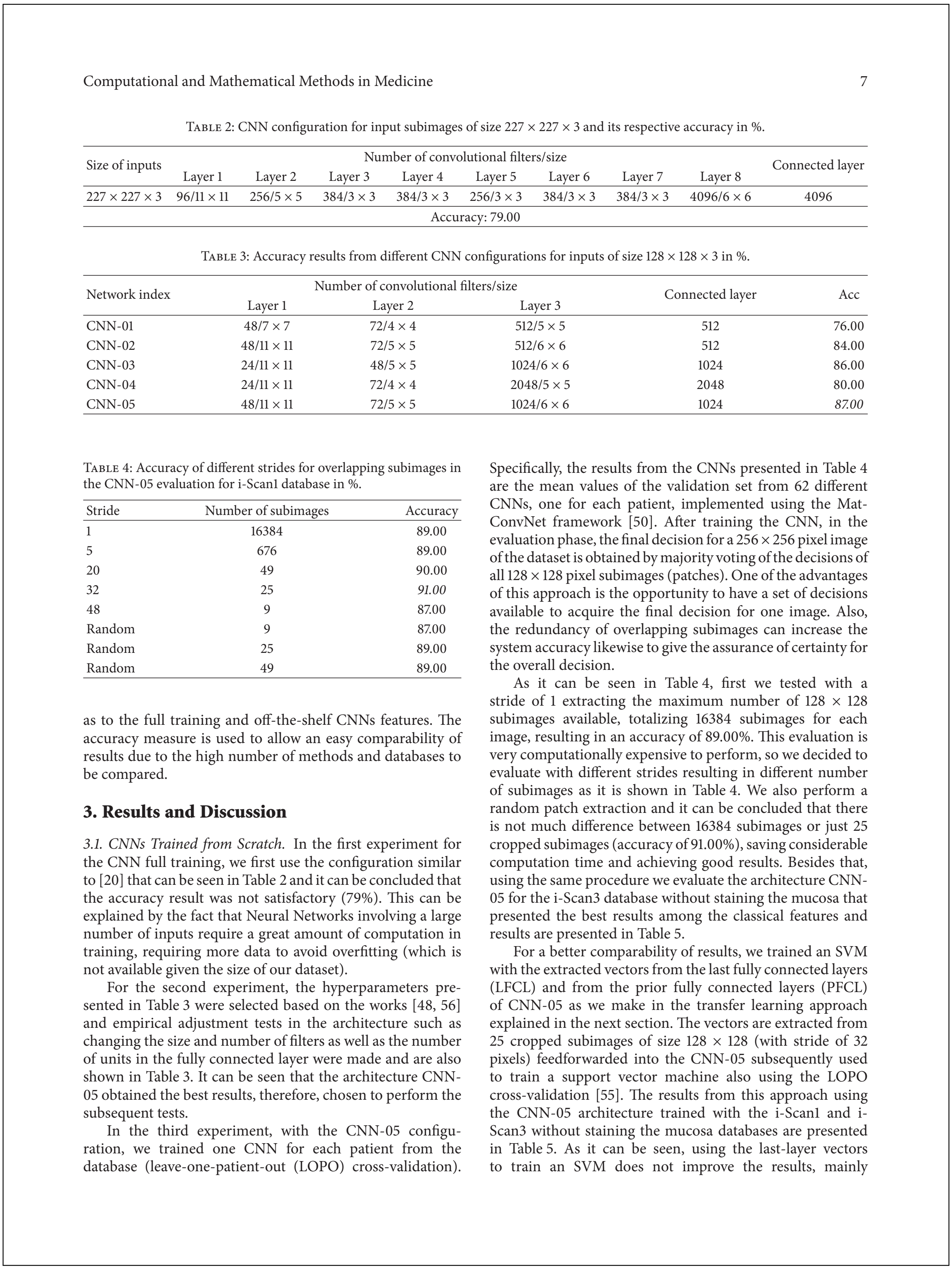


Table 2: CNN configuration for input subimages of size 227 × 227 × 3 and its respective accuracy in %.

Size of inputs	Number of convolutional filters/size								Connected layer
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7	Layer 8	
227 × 227 × 3	96/11 × 11	256/5 × 5	384/3 × 3	384/3 × 3	256/3 × 3	384/3 × 3	384/3 × 3	4096/6 × 6	4096
Accuracy: 79.00									

Table 3: Accuracy results from different CNN configurations for inputs of size 128 × 128 × 3 in %.

Network index	Number of convolutional filters/size			Connected layer	Acc
	Layer 1	Layer 2	Layer 3		
CNN-01	48/7 × 7	72/4 × 4	512/5 × 5	512	76.00
CNN-02	48/11 × 11	72/5 × 5	512/6 × 6	512	84.00
CNN-03	24/11 × 11	48/5 × 5	1024/6 × 6	1024	86.00
CNN-04	24/11 × 11	72/4 × 4	2048/5 × 5	2048	80.00
CNN-05	48/11 × 11	72/5 × 5	1024/6 × 6	1024	*87.00*

Table 4: Accuracy of different strides for overlapping subimages in the CNN-05 evaluation for i-Scan1 database in %.

Stride	Number of subimages	Accuracy
1	16384	89.00
5	676	89.00
20	49	90.00
32	25	*91.00*
48	9	87.00
Random	9	87.00
Random	25	89.00
Random	49	89.00

as to the full training and off-the-shelf CNNs features. The accuracy measure is used to allow an easy comparability of results due to the high number of methods and databases to be compared.

3. Results and Discussion

3.1. CNNs Trained from Scratch. In the first experiment for the CNN full training, we first use the configuration similar to [20] that can be seen in Table 2 and it can be concluded that the accuracy result was not satisfactory (79%). This can be explained by the fact that Neural Networks involving a large number of inputs require a great amount of computation in training, requiring more data to avoid overfitting (which is not available given the size of our dataset).

For the second experiment, the hyperparameters presented in Table 3 were selected based on the works [48, 56] and empirical adjustment tests in the architecture such as changing the size and number of filters as well as the number of units in the fully connected layer were made and are also shown in Table 3. It can be seen that the architecture CNN-05 obtained the best results, therefore, chosen to perform the subsequent tests.

In the third experiment, with the CNN-05 configuration, we trained one CNN for each patient from the database (leave-one-patient-out (LOPO) cross-validation). Specifically, the results from the CNNs presented in Table 4 are the mean values of the validation set from 62 different CNNs, one for each patient, implemented using the MatConvNet framework [50]. After training the CNN, in the evaluation phase, the final decision for a 256 × 256 pixel image of the dataset is obtained by majority voting of the decisions of all 128 × 128 pixel subimages (patches). One of the advantages of this approach is the opportunity to have a set of decisions available to acquire the final decision for one image. Also, the redundancy of overlapping subimages can increase the system accuracy likewise to give the assurance of certainty for the overall decision.

As it can be seen in Table 4, first we tested with a stride of 1 extracting the maximum number of 128 × 128 subimages available, totalizing 16384 subimages for each image, resulting in an accuracy of 89.00%. This evaluation is very computationally expensive to perform, so we decided to evaluate with different strides resulting in different number of subimages as it is shown in Table 4. We also perform a random patch extraction and it can be concluded that there is not much difference between 16384 subimages or just 25 cropped subimages (accuracy of 91.00%), saving considerable computation time and achieving good results. Besides that, using the same procedure we evaluate the architecture CNN-05 for the i-Scan3 database without staining the mucosa that presented the best results among the classical features and results are presented in Table 5.

For a better comparability of results, we trained an SVM with the extracted vectors from the last fully connected layers (LFCL) and from the prior fully connected layers (PFCL) of CNN-05 as we make in the transfer learning approach explained in the next section. The vectors are extracted from 25 cropped subimages of size 128 × 128 (with stride of 32 pixels) feedforwarded into the CNN-05 subsequently used to train a support vector machine also using the LOPO cross-validation [55]. The results from this approach using the CNN-05 architecture trained with the i-Scan1 and i-Scan3 without staining the mucosa databases are presented in Table 5. As it can be seen, using the last-layer vectors to train an SVM does not improve the results, mainly

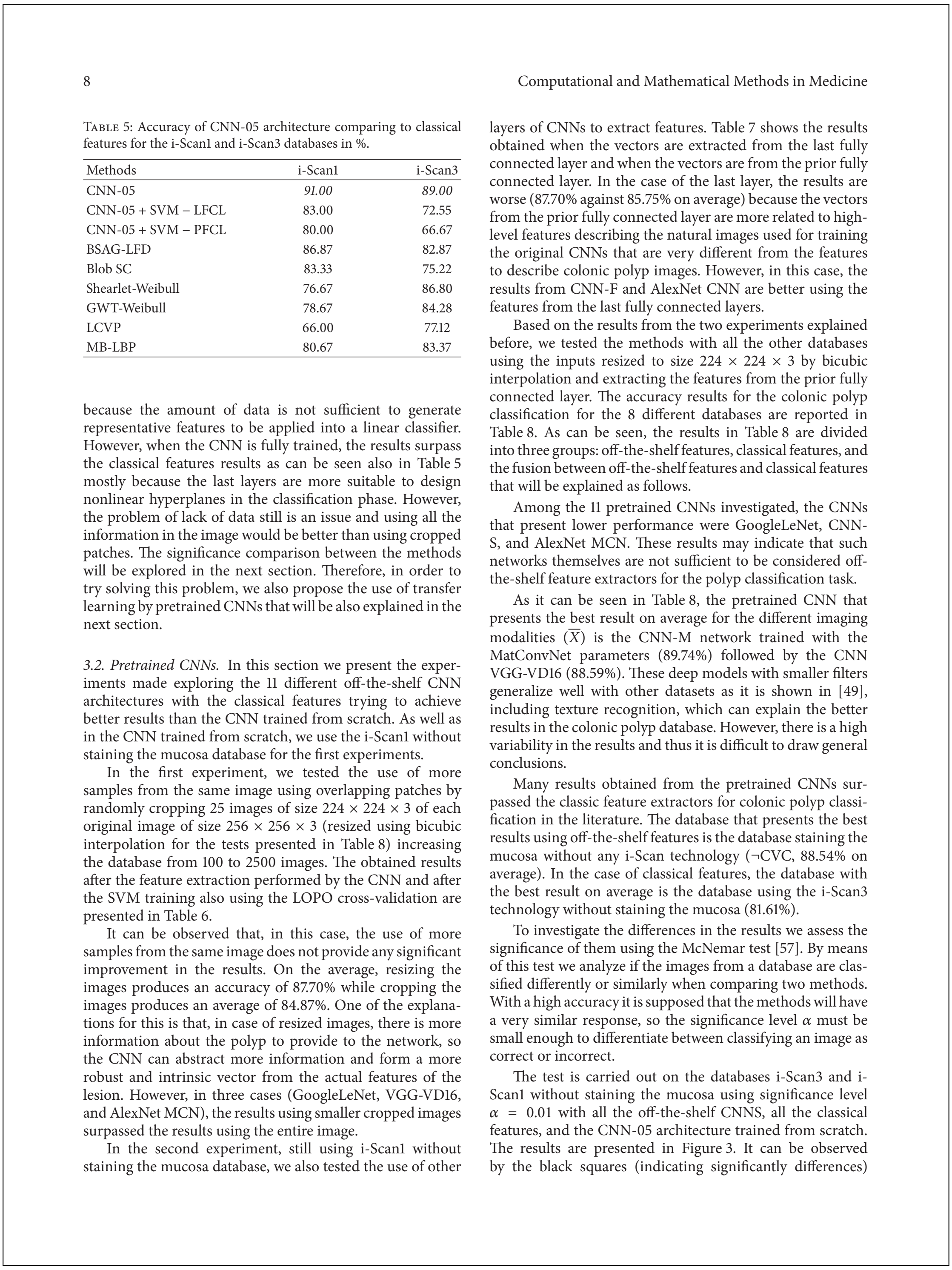


Table 5: Accuracy of CNN-05 architecture comparing to classical features for the i-Scan1 and i-Scan3 databases in %.

Methods	i-Scan1	i-Scan3
CNN-05	*91.00*	*89.00*
CNN-05 + SVM − LFCL	83.00	72.55
CNN-05 + SVM − PFCL	80.00	66.67
BSAG-LFD	86.87	82.87
Blob SC	83.33	75.22
Shearlet-Weibull	76.67	86.80
GWT-Weibull	78.67	84.28
LCVP	66.00	77.12
MB-LBP	80.67	83.37

because the amount of data is not sufficient to generate representative features to be applied into a linear classifier. However, when the CNN is fully trained, the results surpass the classical features results as can be seen also in Table 5 mostly because the last layers are more suitable to design nonlinear hyperplanes in the classification phase. However, the problem of lack of data still is an issue and using all the information in the image would be better than using cropped patches. The significance comparison between the methods will be explored in the next section. Therefore, in order to try solving this problem, we also propose the use of transfer learning by pretrained CNNs that will be also explained in the next section.

3.2. Pretrained CNNs. In this section we present the experiments made exploring the 11 different off-the-shelf CNN architectures with the classical features trying to achieve better results than the CNN trained from scratch. As well as in the CNN trained from scratch, we use the i-Scan1 without staining the mucosa database for the first experiments.

In the first experiment, we tested the use of more samples from the same image using overlapping patches by randomly cropping 25 images of size 224 × 224 × 3 of each original image of size 256 × 256 × 3 (resized using bicubic interpolation for the tests presented in Table 8) increasing the database from 100 to 2500 images. The obtained results after the feature extraction performed by the CNN and after the SVM training also using the LOPO cross-validation are presented in Table 6.

It can be observed that, in this case, the use of more samples from the same image does not provide any significant improvement in the results. On the average, resizing the images produces an accuracy of 87.70% while cropping the images produces an average of 84.87%. One of the explanations for this is that, in case of resized images, there is more information about the polyp to provide to the network, so the CNN can abstract more information and form a more robust and intrinsic vector from the actual features of the lesion. However, in three cases (GoogleLeNet, VGG-VD16, and AlexNet MCN), the results using smaller cropped images surpassed the results using the entire image.

In the second experiment, still using i-Scan1 without staining the mucosa database, we also tested the use of other layers of CNNs to extract features. Table 7 shows the results obtained when the vectors are extracted from the last fully connected layer and when the vectors are from the prior fully connected layer. In the case of the last layer, the results are worse (87.70% against 85.75% on average) because the vectors from the prior fully connected layer are more related to high-level features describing the natural images used for training the original CNNs that are very different from the features to describe colonic polyp images. However, in this case, the results from CNN-F and AlexNet CNN are better using the features from the last fully connected layers.

Based on the results from the two experiments explained before, we tested the methods with all the other databases using the inputs resized to size 224 × 224 × 3 by bicubic interpolation and extracting the features from the prior fully connected layer. The accuracy results for the colonic polyp classification for the 8 different databases are reported in Table 8. As can be seen, the results in Table 8 are divided into three groups: off-the-shelf features, classical features, and the fusion between off-the-shelf features and classical features that will be explained as follows.

Among the 11 pretrained CNNs investigated, the CNNs that present lower performance were GoogleLeNet, CNN-S, and AlexNet MCN. These results may indicate that such networks themselves are not sufficient to be considered off-the-shelf feature extractors for the polyp classification task.

As it can be seen in Table 8, the pretrained CNN that presents the best result on average for the different imaging modalities ($\overline{X}$) is the CNN-M network trained with the MatConvNet parameters (89.74%) followed by the CNN VGG-VD16 (88.59%). These deep models with smaller filters generalize well with other datasets as it is shown in [49], including texture recognition, which can explain the better results in the colonic polyp database. However, there is a high variability in the results and thus it is difficult to draw general conclusions.

Many results obtained from the pretrained CNNs surpassed the classic feature extractors for colonic polyp classification in the literature. The database that presents the best results using off-the-shelf features is the database staining the mucosa without any i-Scan technology (¬CVC, 88.54% on average). In the case of classical features, the database with the best result on average is the database using the i-Scan3 technology without staining the mucosa (81.61%).

To investigate the differences in the results we assess the significance of them using the McNemar test [57]. By means of this test we analyze if the images from a database are classified differently or similarly when comparing two methods. With a high accuracy it is supposed that the methods will have a very similar response, so the significance level α must be small enough to differentiate between classifying an image as correct or incorrect.

The test is carried out on the databases i-Scan3 and i-Scan1 without staining the mucosa using significance level $\alpha = 0.01$ with all the off-the-shelf CNNS, all the classical features, and the CNN-05 architecture trained from scratch. The results are presented in Figure 3. It can be observed by the black squares (indicating significantly differences)

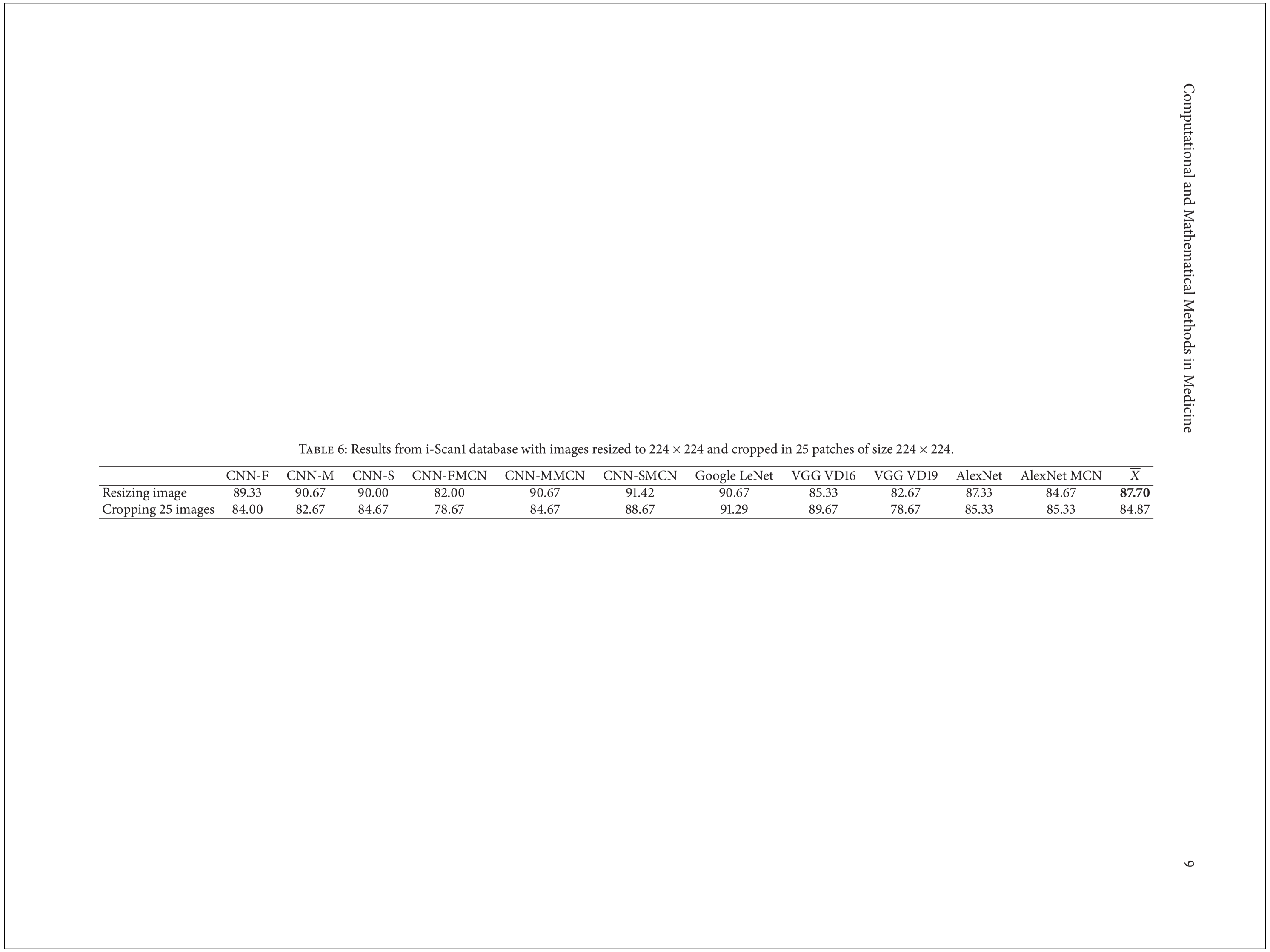


Table 6: Results from i-Scan1 database with images resized to 224 × 224 and cropped in 25 patches of size 224 × 224.

	CNN-F	CNN-M	CNN-S	CNN-FMCN	CNN-MMCN	CNN-SMCN	Google LeNet	VGG VD16	VGG VD19	AlexNet	AlexNet MCN	$\overline{X}$
Resizing image	89.33	90.67	90.00	82.00	90.67	91.42	90.67	85.33	82.67	87.33	84.67	**87.70**
Cropping 25 images	84.00	82.67	84.67	78.67	84.67	88.67	91.29	89.67	78.67	85.33	85.33	84.87

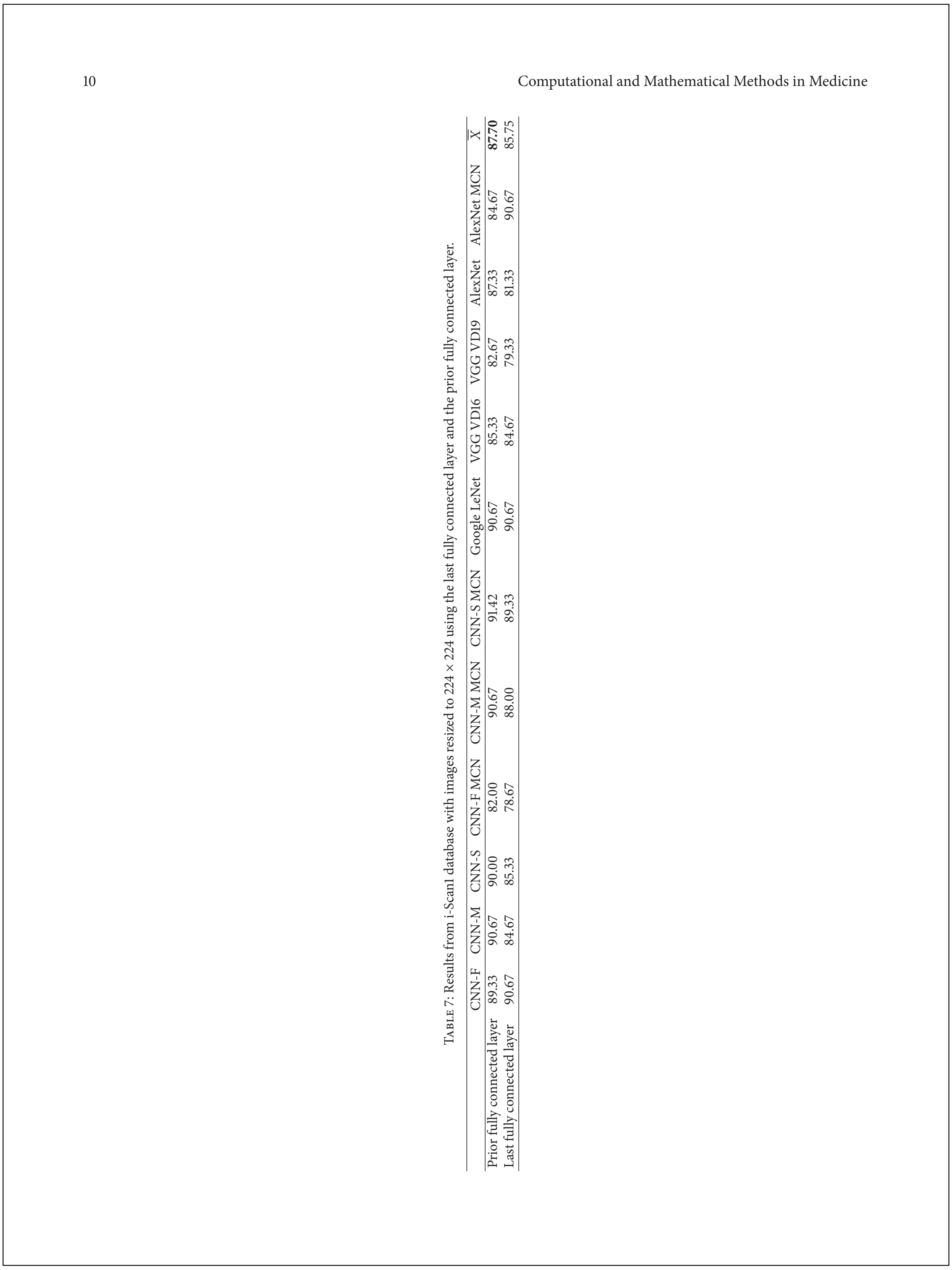


Table 7: Results from i-Scan1 database with images resized to 224 × 224 using the last fully connected layer and the prior fully connected layer.

	CNN-F	CNN-M	CNN-S	CNN-F MCN	CNN-M MCN	CNN-S MCN	Google LeNet	VGG VD16	VGG VD19	AlexNet	AlexNet MCN	$\overline{X}$
Prior fully connected layer	89.33	90.67	90.00	82.00	90.67	91.42	90.67	85.33	82.67	87.33	84.67	**87.70**
Last fully connected layer	90.67	84.67	85.33	78.67	88.00	89.33	90.67	84.67	79.33	81.33	90.67	85.75

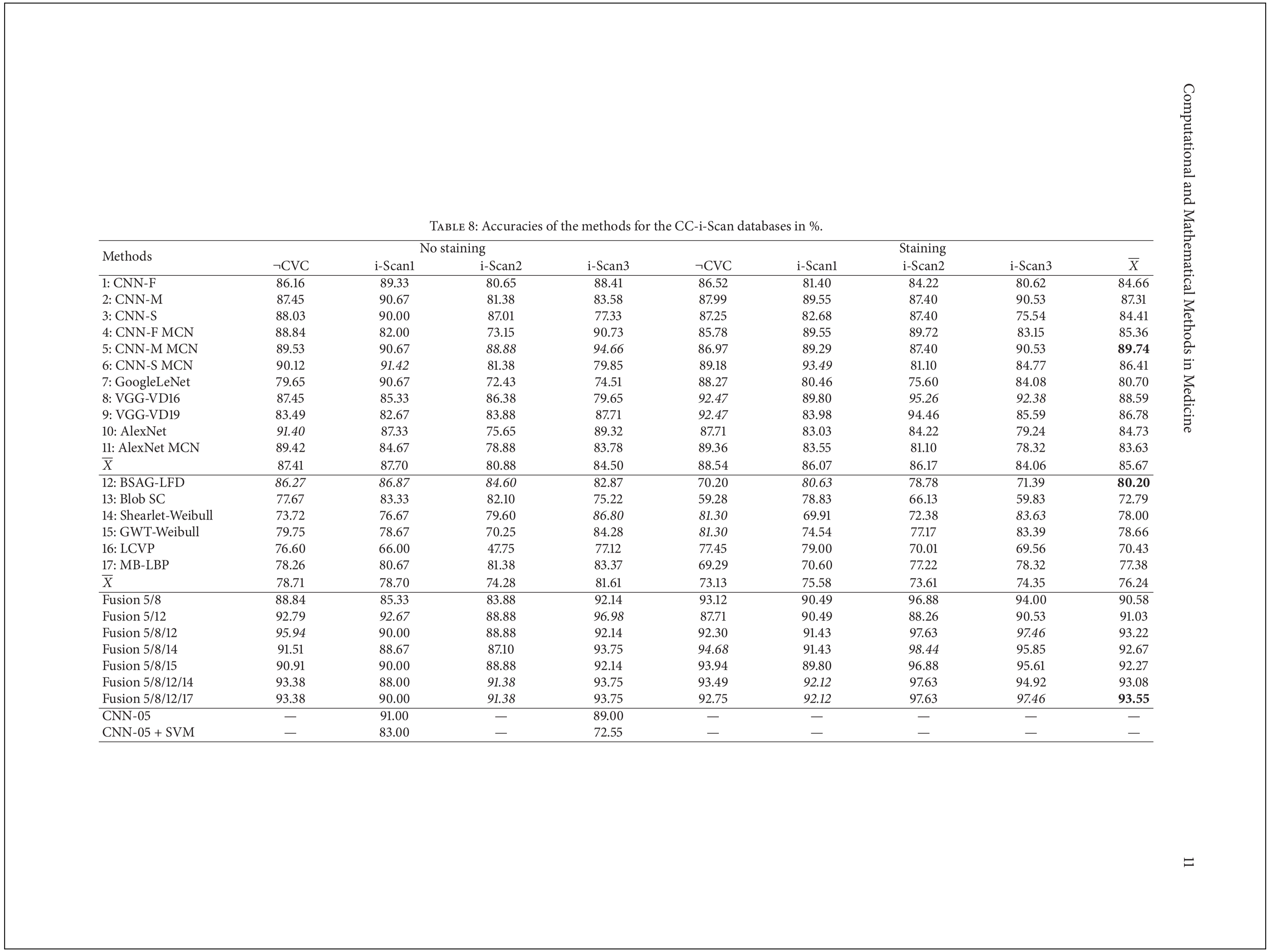
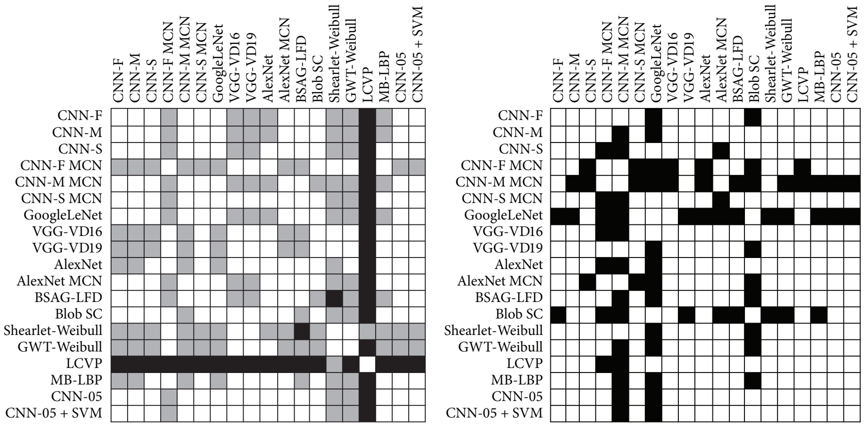


Table 8: Accuracies of the methods for the CC-i-Scan databases in %.

Methods	No staining				Staining				
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3	$\overline{X}$
1: CNN-F	86.16	89.33	80.65	88.41	86.52	81.40	84.22	80.62	84.66
2: CNN-M	87.45	90.67	81.38	83.58	87.99	89.55	87.40	90.53	87.31
3: CNN-S	88.03	90.00	87.01	77.33	87.25	82.68	87.40	75.54	84.41
4: CNN-F MCN	88.84	82.00	73.15	90.73	85.78	89.55	89.72	83.15	85.36
5: CNN-M MCN	89.53	90.67	*88.88*	*94.66*	86.97	89.29	87.40	90.53	**89.74**
6: CNN-S MCN	90.12	*91.42*	81.38	79.85	89.18	*93.49*	81.10	84.77	86.41
7: GoogleLeNet	79.65	90.67	72.43	74.51	88.27	80.46	75.60	84.08	80.70
8: VGG-VD16	87.45	85.33	86.38	79.65	*92.47*	89.80	*95.26*	*92.38*	88.59
9: VGG-VD19	83.49	82.67	83.88	87.71	*92.47*	83.98	94.46	85.59	86.78
10: AlexNet	*91.40*	87.33	75.65	89.32	87.71	83.03	84.22	79.24	84.73
11: AlexNet MCN	89.42	84.67	78.88	83.78	89.36	83.55	81.10	78.32	83.63
$\overline{X}$	87.41	87.70	80.88	84.50	88.54	86.07	86.17	84.06	85.67
12: BSAG-LFD	*86.27*	*86.87*	*84.60*	82.87	70.20	*80.63*	78.78	71.39	**80.20**
13: Blob SC	77.67	83.33	82.10	75.22	59.28	78.83	66.13	59.83	72.79
14: Shearlet-Weibull	73.72	76.67	79.60	*86.80*	*81.30*	69.91	72.38	*83.63*	78.00
15: GWT-Weibull	79.75	78.67	70.25	84.28	*81.30*	74.54	77.17	83.39	78.66
16: LCVP	76.60	66.00	47.75	77.12	77.45	79.00	70.01	69.56	70.43
17: MB-LBP	78.26	80.67	81.38	83.37	69.29	70.60	77.22	78.32	77.38
$\overline{X}$	78.71	78.70	74.28	81.61	73.13	75.58	73.61	74.35	76.24
Fusion 5/8	88.84	85.33	83.88	92.14	93.12	90.49	96.88	94.00	90.58
Fusion 5/12	92.79	*92.67*	88.88	*96.98*	87.71	90.49	88.26	90.53	91.03
Fusion 5/8/12	*95.94*	90.00	88.88	92.14	92.30	91.43	97.63	*97.46*	93.22
Fusion 5/8/14	91.51	88.67	87.10	93.75	*94.68*	91.43	*98.44*	95.85	92.67
Fusion 5/8/15	90.91	90.00	88.88	92.14	93.94	89.80	96.88	95.61	92.27
Fusion 5/8/12/14	93.38	88.00	*91.38*	93.75	93.49	*92.12*	97.63	94.92	93.08
Fusion 5/8/12/17	93.38	90.00	*91.38*	93.75	92.75	*92.12*	97.63	*97.46*	**93.55**
CNN-05	—	91.00	—	89.00	—	—	—	—	—
CNN-05 + SVM	—	83.00	—	72.55	—	—	—	—	—



(a) i-Scan1

(b) i-Scan3

Figure 3: Results of the McNemar test for the i-Scan1 (a) and i-Scan3 (b) databases without staining. A black square in the matrix means that the methods are significantly different with significance level $\alpha = 0.01$ and a grey square in (a) means that the methods are significantly different with significance level $\alpha = 0.05$. If the square is white then there is no significant difference between the methods.

that, among the pretrained CNNs, in the i-Scan1 database the results are not significantly different and in the i-Scan3 database the CNN-M MCN and GoogleLeNet present the most significantly different results comparing to the other CNNs. It also can be seen that the CNN-05 does not have significantly different results comparing to the other CNNs in the i-Scan1 database and has significantly different results with CNN-M MCN and GoogleLeNet in the i-Scan3 database.

Also, in Figure 3, when comparing the classical feature extraction methods with the CNNs features it can be seen that there is a quite different response among the results in i-Scan3 database, especially for CNN-M MCN that is significantly different from all the classical methods with the exception of the Shearlet-Weibull method. The CNN-05 and CNN-05 + SVM did not present significantly different results with the classical features (except with LCVP in i-Scan1 database) and with the pretrained CNNs (except with CNN-M and GoogleLeNet in i-Scan3 database). Likewise, the methods with high accuracy in the i-Scan3 database (BSAG-LFD, VGG-VD16, and VGG-VD19) are not found to be significantly different.

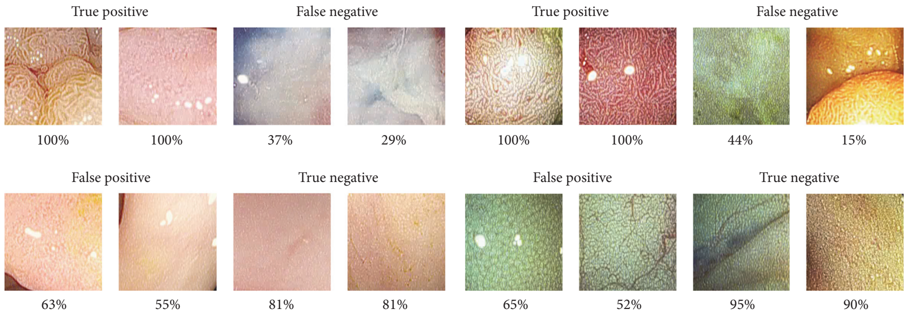
In the i-Scan1 database, with the significance level $\alpha = 0.05$, the results are not significantly different in general (except for LCVP features). However, with the significance level $\alpha = 0.01$, the significance results represented by the grey squares in Figure 3(a) show that the two databases presented different correlation between methods which means that it is difficult to predict a good feature extractor that can satisfy both databases at the same time.

Observing the methods that presented significantly different results in Figure 3 and with good results in Table 8 we decided to produce a feature level fusion in the feature vectors concatenating them to see if the features can complement each other. It can be seen in Figure 3 that the two most successful CNNs CNN-M MCN and VGG-VD16 are significantly different from each other in both databases and the feature level fusion of these two vectors improve the results from 89.74% and 88.59%, respectively, to an accuracy of 90.58% in average as can be seen in Table 8 (Fusion 5/8).

In Figure 3(b) it can also be observed that the results from CNN-M MCN are significantly different to the classical features BSAG-LFD in the i-Scan3 database. With the feature level fusion of these two features the accuracy increases to 91.03% on average. Concatenating the three feature vectors (CNN-M MCN, VGG-VD16, and BSAG-LFD) leads to an even better accuracy: 93.22%. It is interesting to note that in both databases the results from CNN-M MCN and VGG-VD16 are significantly different. Besides that, BSAG-LFD results are significantly different to VGG-VD16 in database i-Scan1. Furthermore, BSAG-LFD results are significantly different to CNN-M MCN in database i-Scan3 which can explain the improvement in the feature level fusion between these three methods.

Making the fusion with these two off-the-shelf CNNs (CNN-M MCN and VGG-VD16) to other classical feature vectors also increases the accuracy as it can be seen in Table 8 (Fusion 5/8/14 and Fusion 5/8/15).

When we add to the vector Fusion 5/8/12 one more classical feature (MB-LBP) that is also significantly different to CNN-M MCN in database i-Scan3 and at the same time



(a) i-Scan1

(b) i-Scan3

FIGURE 4: Example results of the classification in agreement from the methods tested in the McNemar test for each prediction outcome.

significantly different to BSAG-LFD in database i-Scan1, the result outperforms all the previous approaches: 93.55% as it can be seen in Table 8.

In Figure 4 we present some example images from the classification results of all the methods used in the McNemar test with the higher agreement for each prediction outcome. The percentage above each image shows the average classification rate of the prediction. For example, in the i-Scan1 database and i-Scan3 database (Figures 4(a) and 4(b)), the two images presented in the true positive box were classified as such in all classifiers. However, from i-Scan3 database, in the case of the false negative box, one image had 44% of misclassification and another 15% of misclassification in average.

Comparing the results from all off-the-shelf CNNs and classical features with the CNN-05 trained from scratch using the databases i-Scan1 and i-Scan3 in Table 8 it can be observed that the full training CNN outperformed the results obtained by the classical features and some of the pretrained CNNs. This approach can be considered an option for automatic colonic polyp classification, although the training time and processing complexity are not worthwhile if comparing to the off-the-shelf features.

4. Conclusion

In this work, we propose to explore Deep Learning and Transfer Learning approach using Convolutional Neural Networks (CNNs) to improve the accuracy of colonic polyp classification based on the fact that databases containing large amounts of annotated data are often limited for this type of research. For the training of CNNs from scratch, we explore data augmentation with image patches to increase the size of the training database and consequently the information to perform the Deep Learning. Different architectures were tested to evaluate the impact of the size and number of filters in the classification as well as the number of output units in the fully connected layer.

We also explored and evaluated several different pretrained CNNs architectures to extract features from colonoscopy images by knowledge transfer between natural and medical images providing what is called off-the-shelf CNNs features. We show that the off-the shelf features may be well suited for the automatic classification of colon polyps even with a limited amount of data.

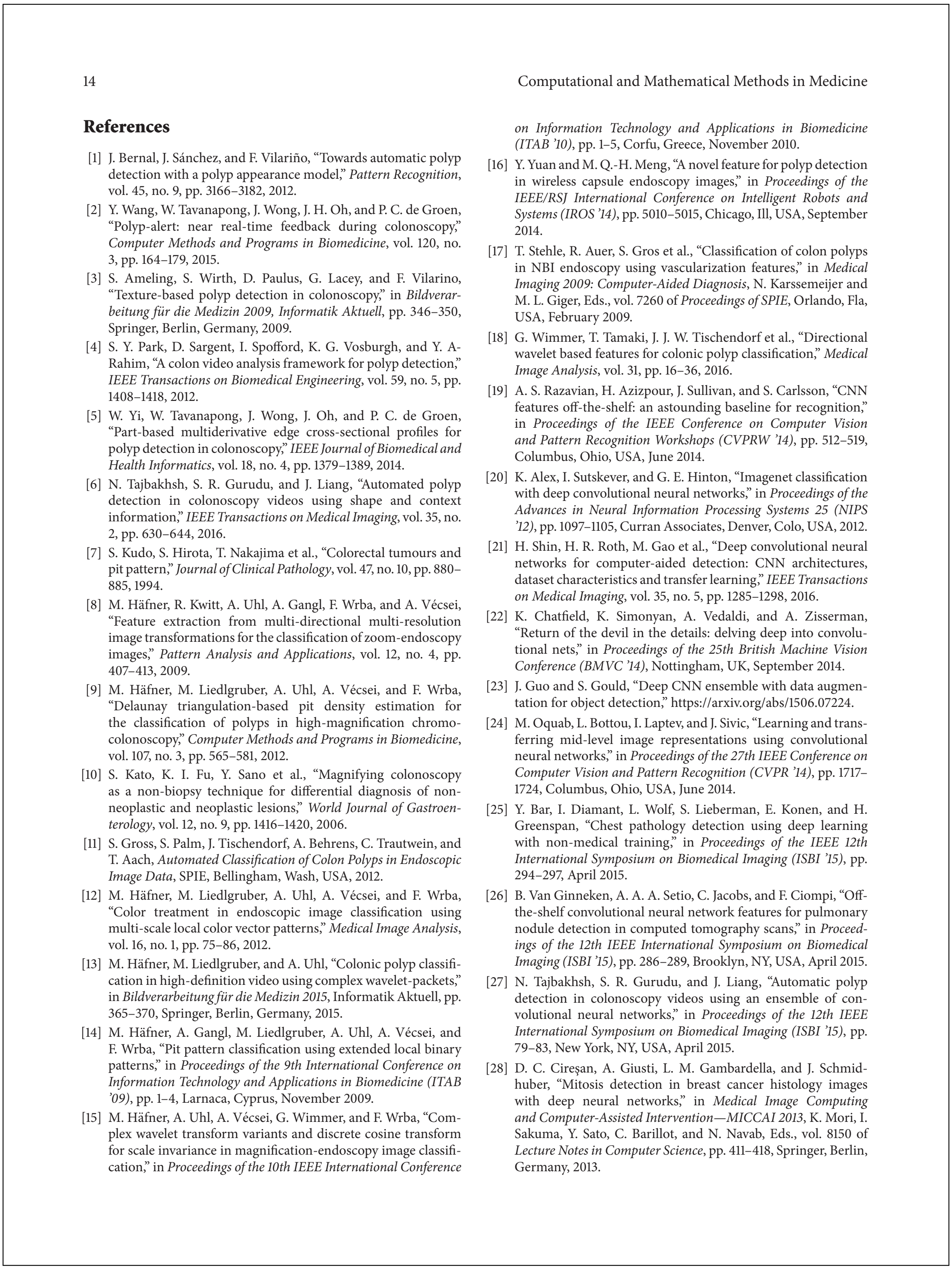
Besides the fact that the pretrained CNNs were trained with natural images, the 4096 features extracted from CNN-M MCN and VGG-16 provided a good feature descriptor of colonic polyps. Some reasons for the success of the classification include the training with a large range of different images providing a powerful extractor joining the intrinsic features from the images such as color, texture, and shape in the same architecture reducing and abstracting these features in just one vector. Also, the combination of classical features with off-the-shelf features yields the best prediction results complementing each other. It can be concluded that Deep Learning using Convolutional Neural Networks is a good option for colonic polyp classification and the use of pretraining CNNs is the best choice to achieve the best results being improved by feature level fusion with classical features. In future work we plan to use this strategy to also test the detection of colonic polyps directly into video frames and evaluate the performance in real time applications as well as to use this strategy in other endoscopic databases such as automatic classification of celiac disease.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research was partially supported by CNPq, Brazil, for Eduardo Ribeiro under Grant no. 00736/2014-0.



References

[1] J. Bernal, J. Sánchez, and F. Vilariño, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, vol. 45, no. 9, pp. 3166–3182, 2012.

[2] Y. Wang, W. Tavanapong, J. Wong, J. H. Oh, and P. C. de Groen, "Polyp-alert: near real-time feedback during colonoscopy," *Computer Methods and Programs in Biomedicine*, vol. 120, no. 3, pp. 164–179, 2015.

[3] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung für die Medizin 2009, Informatik Aktuell*, pp. 346–350, Springer, Berlin, Germany, 2009.

[4] S. Y. Park, D. Sargent, I. Spofford, K. G. Vosburgh, and Y. A-Rahim, "A colon video analysis framework for polyp detection," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1408–1418, 2012.

[5] W. Yi, W. Tavanapong, J. Wong, J. Oh, and P. C. de Groen, "Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1379–1389, 2014.

[6] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, 2016.

[7] S. Kudo, S. Hirota, T. Nakajima et al., "Colorectal tumours and pit pattern," *Journal of Clinical Pathology*, vol. 47, no. 10, pp. 880–885, 1994.

[8] M. Häfner, R. Kwitt, A. Uhl, A. Gangl, F. Wrba, and A. Vécsei, "Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images," *Pattern Analysis and Applications*, vol. 12, no. 4, pp. 407–413, 2009.

[9] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy," *Computer Methods and Programs in Biomedicine*, vol. 107, no. 3, pp. 565–581, 2012.

[10] S. Kato, K. I. Fu, Y. Sano et al., "Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions," *World Journal of Gastroenterology*, vol. 12, no. 9, pp. 1416–1420, 2006.

[11] S. Gross, S. Palm, J. Tischendorf, A. Behrens, C. Trautwein, and T. Aach, *Automated Classification of Colon Polyps in Endoscopic Image Data*, SPIE, Bellingham, Wash, USA, 2012.

[12] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Color treatment in endoscopic image classification using multi-scale local color vector patterns," *Medical Image Analysis*, vol. 16, no. 1, pp. 75–86, 2012.

[13] M. Häfner, M. Liedlgruber, and A. Uhl, "Colonic polyp classification in high-definition video using complex wavelet-packets," in *Bildverarbeitung für die Medizin 2015*, Informatik Aktuell, pp. 365–370, Springer, Berlin, Germany, 2015.

[14] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Pit pattern classification using extended local binary patterns," in *Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine (ITAB '09)*, pp. 1–4, Larnaca, Cyprus, November 2009.

[15] M. Häfner, A. Uhl, A. Vécsei, G. Wimmer, and F. Wrba, "Complex wavelet transform variants and discrete cosine transform for scale invariance in magnification-endoscopy image classification," in *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB '10)*, pp. 1–5, Corfu, Greece, November 2010.

[16] Y. Yuan and M. Q.-H. Meng, "A novel feature for polyp detection in wireless capsule endoscopy images," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '14)*, pp. 5010–5015, Chicago, Ill, USA, September 2014.

[17] T. Stehle, R. Auer, S. Gros et al., "Classification of colon polyps in NBI endoscopy using vascularization features," in *Medical Imaging 2009: Computer-Aided Diagnosis*, N. Karssemeijer and M. L. Giger, Eds., vol. 7260 of *Proceedings of SPIE*, Orlando, Fla, USA, February 2009.

[18] G. Wimmer, T. Tamaki, J. J. W. Tischendorf et al., "Directional wavelet based features for colonic polyp classification," *Medical Image Analysis*, vol. 31, pp. 16–36, 2016.

[19] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '14)*, pp. 512–519, Columbus, Ohio, USA, June 2014.

[20] K. Alex, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS '12)*, pp. 1097–1105, Curran Associates, Denver, Colo, USA, 2012.

[21] H. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[22] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: delving deep into convolutional nets," in *Proceedings of the 25th British Machine Vision Conference (BMVC '14)*, Nottingham, UK, September 2014.

[23] J. Guo and S. Gould, "Deep CNN ensemble with data augmentation for object detection," https://arxiv.org/abs/1506.07224.

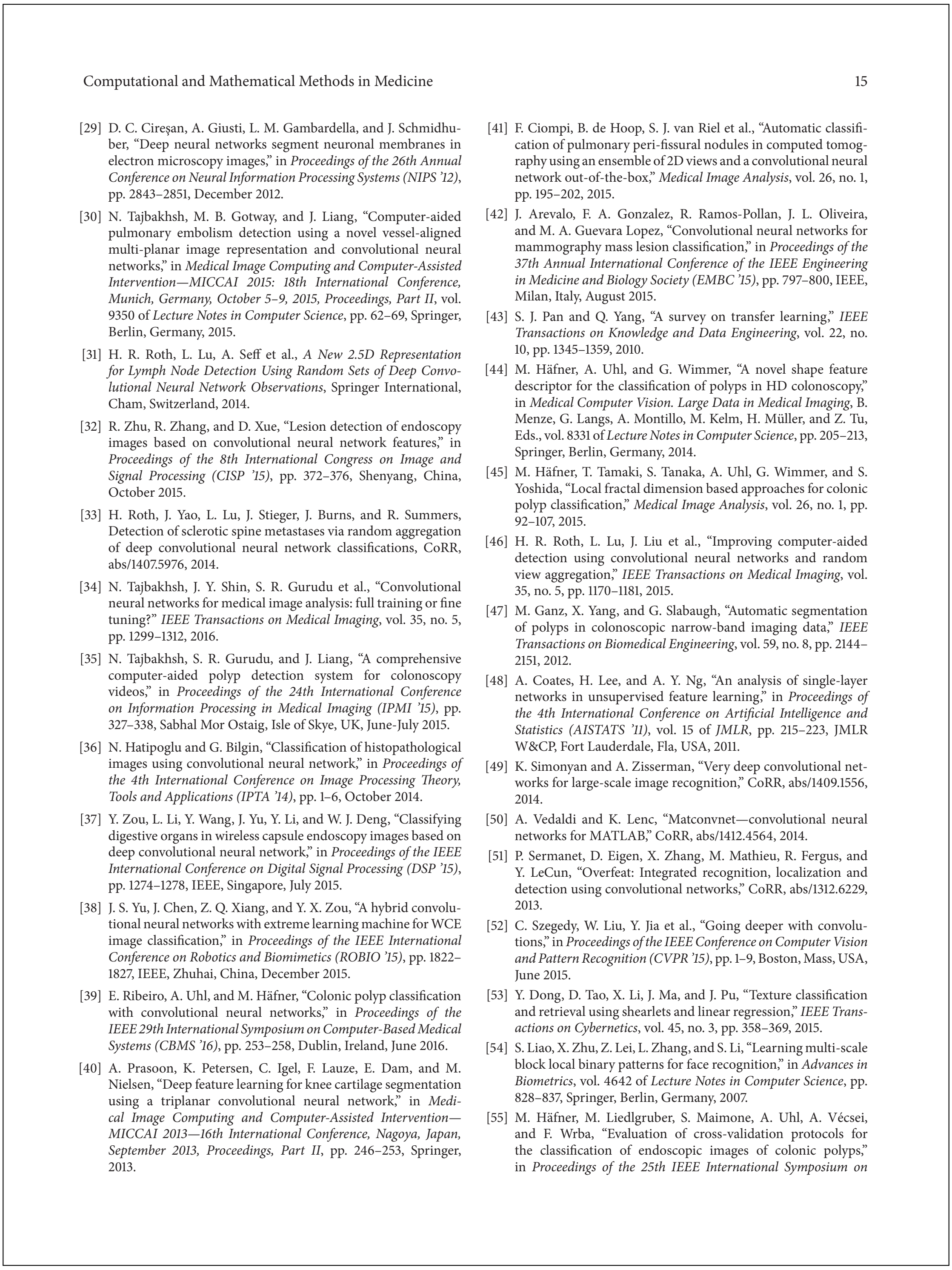
[24] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 1717–1724, Columbus, Ohio, USA, June 2014.

[25] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *Proceedings of the IEEE 12th International Symposium on Biomedical Imaging (ISBI '15)*, pp. 294–297, April 2015.

[26] B. Van Ginneken, A. A. A. Setio, C. Jacobs, and F. Ciompi, "Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans," in *Proceedings of the 12th IEEE International Symposium on Biomedical Imaging (ISBI '15)*, pp. 286–289, Brooklyn, NY, USA, April 2015.

[27] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks," in *Proceedings of the 12th IEEE International Symposium on Biomedical Imaging (ISBI '15)*, pp. 79–83, New York, NY, USA, April 2015.

[28] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., vol. 8150 of *Lecture Notes in Computer Science*, pp. 411–418, Springer, Berlin, Germany, 2013.



[29] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS '12)*, pp. 2843–2851, December 2012.

[30] N. Tajbakhsh, M. B. Gotway, and J. Liang, "Computer-aided pulmonary embolism detection using a novel vessel-aligned multi-planar image representation and convolutional neural networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part II*, vol. 9350 of *Lecture Notes in Computer Science*, pp. 62–69, Springer, Berlin, Germany, 2015.

[31] H. R. Roth, L. Lu, A. Seff et al., *A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations*, Springer International, Cham, Switzerland, 2014.

[32] R. Zhu, R. Zhang, and D. Xue, "Lesion detection of endoscopy images based on convolutional neural network features," in *Proceedings of the 8th International Congress on Image and Signal Processing (CISP '15)*, pp. 372–376, Shenyang, China, October 2015.

[33] H. Roth, J. Yao, L. Lu, J. Stieger, J. Burns, and R. Summers, Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications, CoRR, abs/1407.5976, 2014.

[34] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., "Convolutional neural networks for medical image analysis: full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[35] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "A comprehensive computer-aided polyp detection system for colonoscopy videos," in *Proceedings of the 24th International Conference on Information Processing in Medical Imaging (IPMI '15)*, pp. 327–338, Sabhal Mor Ostaig, Isle of Skye, UK, June-July 2015.

[36] N. Hatipoglu and G. Bilgin, "Classification of histopathological images using convolutional neural network," in *Proceedings of the 4th International Conference on Image Processing Theory, Tools and Applications (IPTA '14)*, pp. 1–6, October 2014.

[37] Y. Zou, L. Li, Y. Wang, J. Yu, Y. Li, and W. J. Deng, "Classifying digestive organs in wireless capsule endoscopy images based on deep convolutional neural network," in *Proceedings of the IEEE International Conference on Digital Signal Processing (DSP '15)*, pp. 1274–1278, IEEE, Singapore, July 2015.

[38] J. S. Yu, J. Chen, Z. Q. Xiang, and Y. X. Zou, "A hybrid convolutional neural networks with extreme learning machine for WCE image classification," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '15)*, pp. 1822–1827, IEEE, Zhuhai, China, December 2015.

[39] E. Ribeiro, A. Uhl, and M. Häfner, "Colonic polyp classification with convolutional neural networks," in *Proceedings of the IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS '16)*, pp. 253–258, Dublin, Ireland, June 2016.

[40] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013—16th International Conference, Nagoya, Japan, September 2013, Proceedings, Part II*, pp. 246–253, Springer, 2013.

[41] F. Ciompi, B. de Hoop, S. J. van Riel et al., "Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box," *Medical Image Analysis*, vol. 26, no. 1, pp. 195–202, 2015.

[42] J. Arevalo, F. A. Gonzalez, R. Ramos-Pollan, J. L. Oliveira, and M. A. Guevara Lopez, "Convolutional neural networks for mammography mass lesion classification," in *Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '15)*, pp. 797–800, IEEE, Milan, Italy, August 2015.

[43] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[44] M. Häfner, A. Uhl, and G. Wimmer, "A novel shape feature descriptor for the classification of polyps in HD colonoscopy," in *Medical Computer Vision. Large Data in Medical Imaging*, B. Menze, G. Langs, A. Montillo, M. Kelm, H. Müller, and Z. Tu, Eds., vol. 8331 of *Lecture Notes in Computer Science*, pp. 205–213, Springer, Berlin, Germany, 2014.

[45] M. Häfner, T. Tamaki, S. Tanaka, A. Uhl, G. Wimmer, and S. Yoshida, "Local fractal dimension based approaches for colonic polyp classification," *Medical Image Analysis*, vol. 26, no. 1, pp. 92–107, 2015.

[46] H. R. Roth, L. Lu, J. Liu et al., "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, 2015.

[47] M. Ganz, X. Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic narrow-band imaging data," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2144–2151, 2012.

[48] A. Coates, H. Lee, and A. Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the 4th International Conference on Artificial Intelligence and Statistics (AISTATS '11)*, vol. 15 of *JMLR*, pp. 215–223, JMLR W&CP, Fort Lauderdale, Fla, USA, 2011.

[49] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, abs/1409.1556, 2014.

[50] A. Vedaldi and K. Lenc, "Matconvnet—convolutional neural networks for MATLAB," CoRR, abs/1412.4564, 2014.

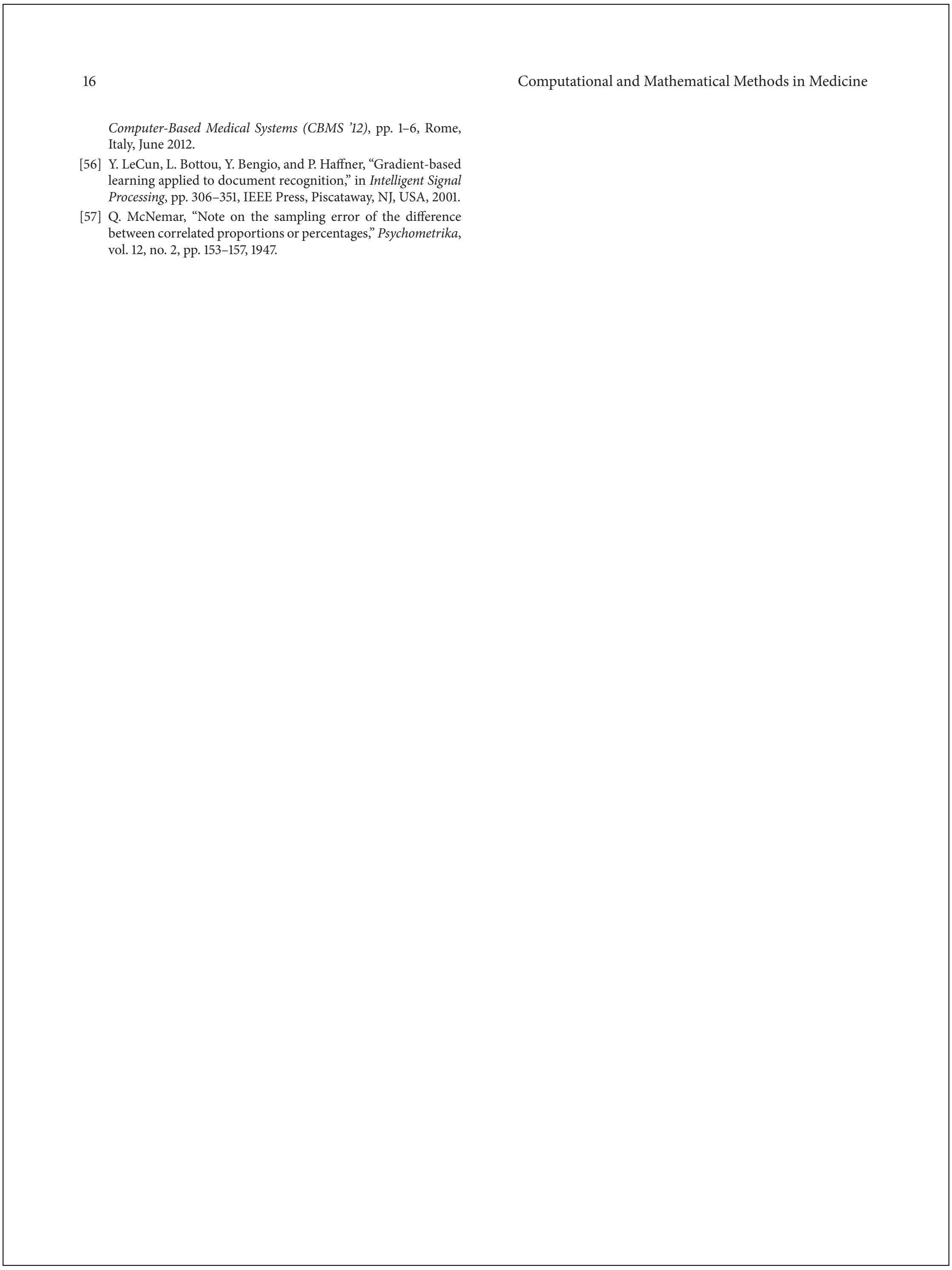
[51] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," CoRR, abs/1312.6229, 2013.

[52] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, pp. 1–9, Boston, Mass, USA, June 2015.

[53] Y. Dong, D. Tao, X. Li, J. Ma, and J. Pu, "Texture classification and retrieval using shearlets and linear regression," *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 358–369, 2015.

[54] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Li, "Learning multi-scale block local binary patterns for face recognition," in *Advances in Biometrics*, vol. 4642 of *Lecture Notes in Computer Science*, pp. 828–837, Springer, Berlin, Germany, 2007.

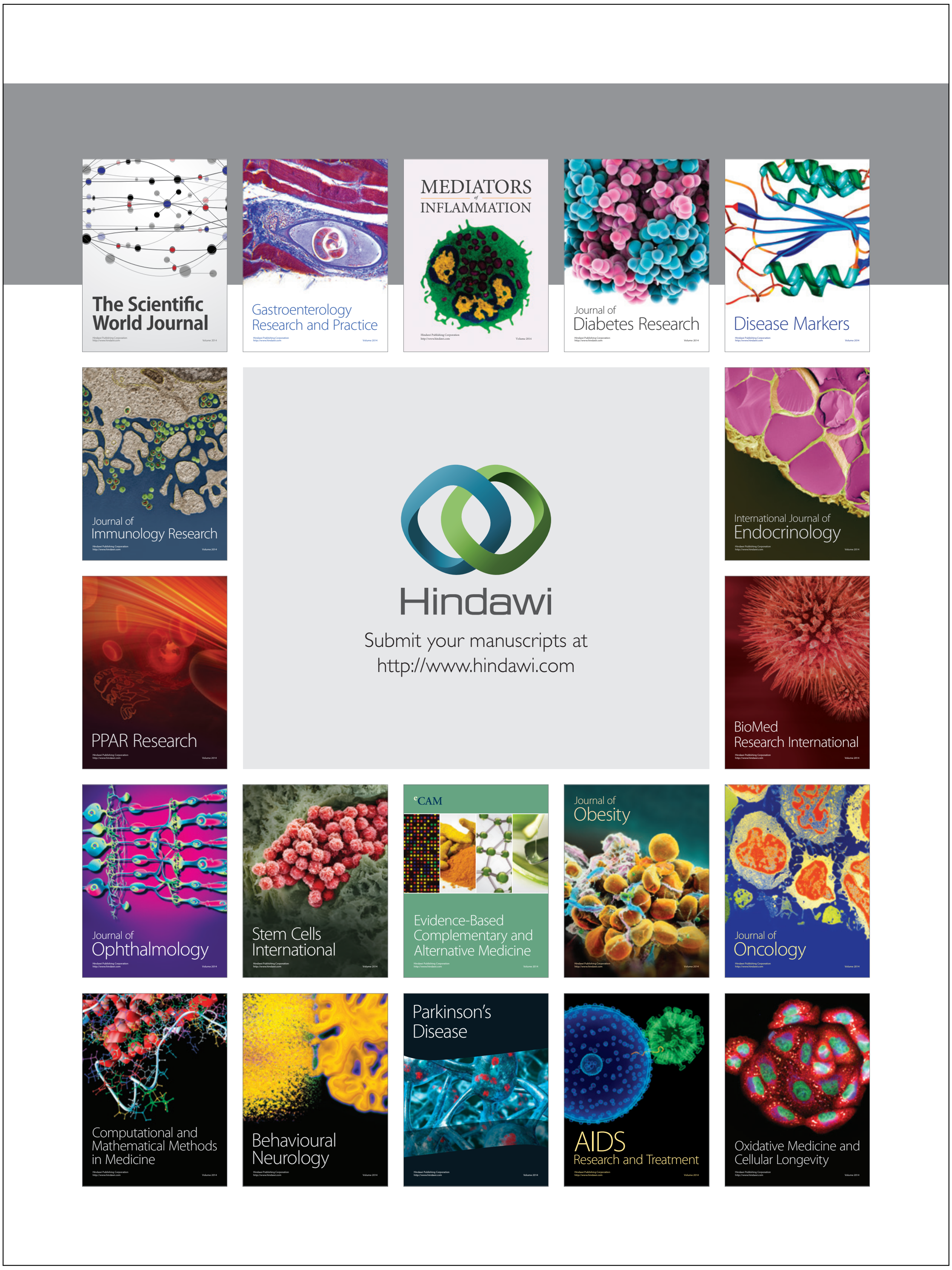
[55] M. Häfner, M. Liedlgruber, S. Maimone, A. Uhl, A. Vécsei, and F. Wrba, "Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps," in *Proceedings of the 25th IEEE International Symposium on*

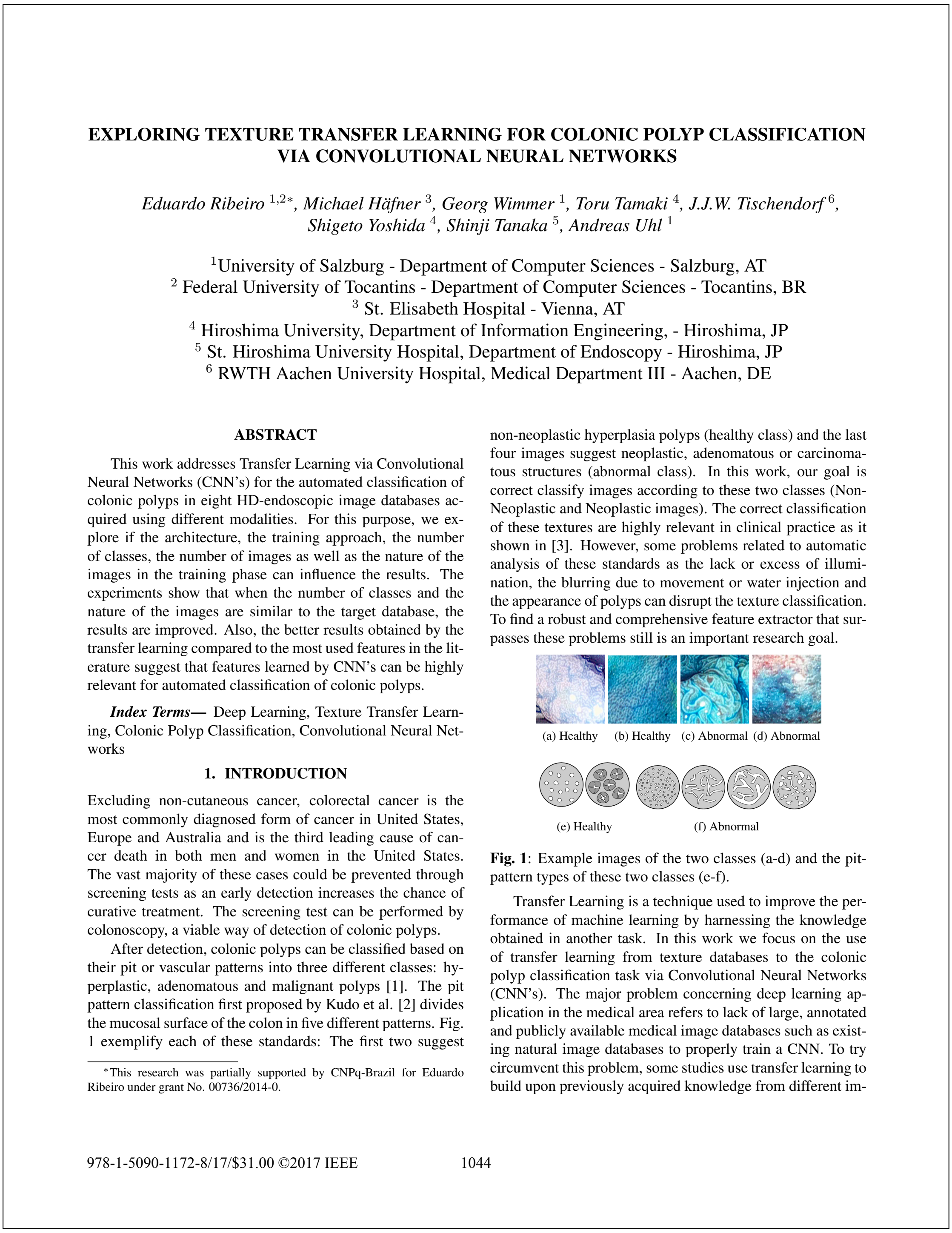


Computer-Based Medical Systems (CBMS '12), pp. 1–6, Rome, Italy, June 2012.

[56] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Intelligent Signal Processing*, pp. 306–351, IEEE Press, Piscataway, NJ, USA, 2001.

[57] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.





EXPLORING TEXTURE TRANSFER LEARNING FOR COLONIC POLYP CLASSIFICATION VIA CONVOLUTIONAL NEURAL NETWORKS

Eduardo Ribeiro [1,2]*, *Michael Häfner* [3], *Georg Wimmer* [1], *Toru Tamaki* [4], *J.J.W. Tischendorf* [6], *Shigeto Yoshida* [4], *Shinji Tanaka* [5], *Andreas Uhl* [1]

[1]University of Salzburg - Department of Computer Sciences - Salzburg, AT
[2] Federal University of Tocantins - Department of Computer Sciences - Tocantins, BR
[3] St. Elisabeth Hospital - Vienna, AT
[4] Hiroshima University, Department of Information Engineering, - Hiroshima, JP
[5] St. Hiroshima University Hospital, Department of Endoscopy - Hiroshima, JP
[6] RWTH Aachen University Hospital, Medical Department III - Aachen, DE

ABSTRACT

This work addresses Transfer Learning via Convolutional Neural Networks (CNN's) for the automated classification of colonic polyps in eight HD-endoscopic image databases acquired using different modalities. For this purpose, we explore if the architecture, the training approach, the number of classes, the number of images as well as the nature of the images in the training phase can influence the results. The experiments show that when the number of classes and the nature of the images are similar to the target database, the results are improved. Also, the better results obtained by the transfer learning compared to the most used features in the literature suggest that features learned by CNN's can be highly relevant for automated classification of colonic polyps.

Index Terms— Deep Learning, Texture Transfer Learning, Colonic Polyp Classification, Convolutional Neural Networks

1. INTRODUCTION

Excluding non-cutaneous cancer, colorectal cancer is the most commonly diagnosed form of cancer in United States, Europe and Australia and is the third leading cause of cancer death in both men and women in the United States. The vast majority of these cases could be prevented through screening tests as an early detection increases the chance of curative treatment. The screening test can be performed by colonoscopy, a viable way of detection of colonic polyps.

After detection, colonic polyps can be classified based on their pit or vascular patterns into three different classes: hyperplastic, adenomatous and malignant polyps [1]. The pit pattern classification first proposed by Kudo et al. [2] divides the mucosal surface of the colon in five different patterns. Fig. 1 exemplify each of these standards: The first two suggest non-neoplastic hyperplasia polyps (healthy class) and the last four images suggest neoplastic, adenomatous or carcinomatous structures (abnormal class). In this work, our goal is correct classify images according to these two classes (Non-Neoplastic and Neoplastic images). The correct classification of these textures are highly relevant in clinical practice as it shown in [3]. However, some problems related to automatic analysis of these standards as the lack or excess of illumination, the blurring due to movement or water injection and the appearance of polyps can disrupt the texture classification. To find a robust and comprehensive feature extractor that surpasses these problems still is an important research goal.

(a) Healthy (b) Healthy (c) Abnormal (d) Abnormal

(e) Healthy (f) Abnormal

Fig. 1: Example images of the two classes (a-d) and the pit-pattern types of these two classes (e-f).

Transfer Learning is a technique used to improve the performance of machine learning by harnessing the knowledge obtained in another task. In this work we focus on the use of transfer learning from texture databases to the colonic polyp classification task via Convolutional Neural Networks (CNN's). The major problem concerning deep learning application in the medical area refers to lack of large, annotated and publicly available medical image databases such as existing natural image databases to properly train a CNN. To try circumvent this problem, some studies use transfer learning to build upon previously acquired knowledge from different im-

*This research was partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.

978-1-5090-1172-8/17/$31.00 ©2017 IEEE 1044

age databases applying it to the medical imaging domain. For example, transfer learning has been used for mammography mass lesion classification [4], pulmonary nodule detection [5] as well as identification, pathology of X-ray and computer tomography modalities [6] and Colonic Polyp Classification [7]. Additionally, Ginneken et al. [5] show that the combination of CNN's features and classical features for pulmonary nodule detection can improve the performance of the model. Furthermore, texture classification using CNN's is not yet a well-explored mainly because most textured databases available are small and or have few classes in order to properly train a CNN.

In this work we aim to answer the following questions: Is the similarity of the dataset used to train/fine-tune a CNN to the data material finally classified important for the obtained classification result of transfer learning? In particular, do we get better result in classifying colonic polyp mucosa when training CNN's on other endoscopic datasets, texture datasets, or collections of natural images? Is it better to train with more similar images or is it better to just use as many images as possible? Another question tackled is about the number of classes: For optimal results of transfer learning, should we have an equal number of classes in the training data and the data subject to classification (recall that we employ the CNNs for feature extraction only)?

Of course, the CNN transfer learning approach [8] assumes that a feature extractor is formed during the training and patterns learned from the training dataset can be used to correctly classify colonic polyps. The CNN's used in this work operate as feature extractors only but not as classifiers: CNNs are either trained from scratch (full training) using one of the training datasets or are employed by fine-tuning using one of the training datasets to a pre-trained CNN. In either case, the CNNs are used to extract features from our colonoscopic datasets finally subjected to classification. The images are classified among different acquisition modes of colonoscopy images (eight different sub-databases in the CC-i-Scan Database) as explained in the next section.

2. METHODOLOGY

2.1. CC-i-Scan Database

In this work colonic polyp classification is explored using an endoscopic database containing 8 sub-databases with 8 different categories. The image frames are from videos acquired by an HD endoscope (Pentax HILINE HD + 90i Colonoscope) either using the i-Scan technology or computer without any virtual chromoendoscopy (¬CVC in Table 1).

The mucosa can be either stained or not stained. Despite the fact frames being originally in high-definition, the image size (255x255x3) was chosen (i) to be large enough to describe a polyp and (ii) small enough to cover just one class of mucosa type (only healthy or only abnormal area). The image labels (ground truth) were provided according to their histological diagnosis.

Table 1: Number of images and patients per class of the CC-i-Scan databases.

	No staining				Staining			
i-Scan mode	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3
Non-neoplastic								
Nr. of images	39	25	20	31	42	53	32	31
Nr. of patients	21	18	15	15	26	31	23	19
Neoplastic								
Nr. of images	73	75	69	71	68	73	62	54
Nr. of patients	55	56	55	55	52	55	52	47
Total	112	100	89	102	110	126	94	85

2.2. Training Databases

For the CNN training, we use nine different databases including three endoscopic databases, three texture databases and three natural image databases described as follows ordered according to their similarity with the target database.

Colonic Polyp Image Databases: The NBI high magnification database Hiroshima (**NBI1**) is a database containing 563 images of colonic polyps divided into 3 classes [1]. The NBI high magnification database Aachen (**NBI2**) is a database containing 387 endoscopic color images from 211 patients divided into two classes [1].

Endoscopic Image Database: The Celiac Disease Database (**CELIAC**) containing 612 idealistic patches of size 128x128 divided into two classes (March-0 and Marsh-03) [9].

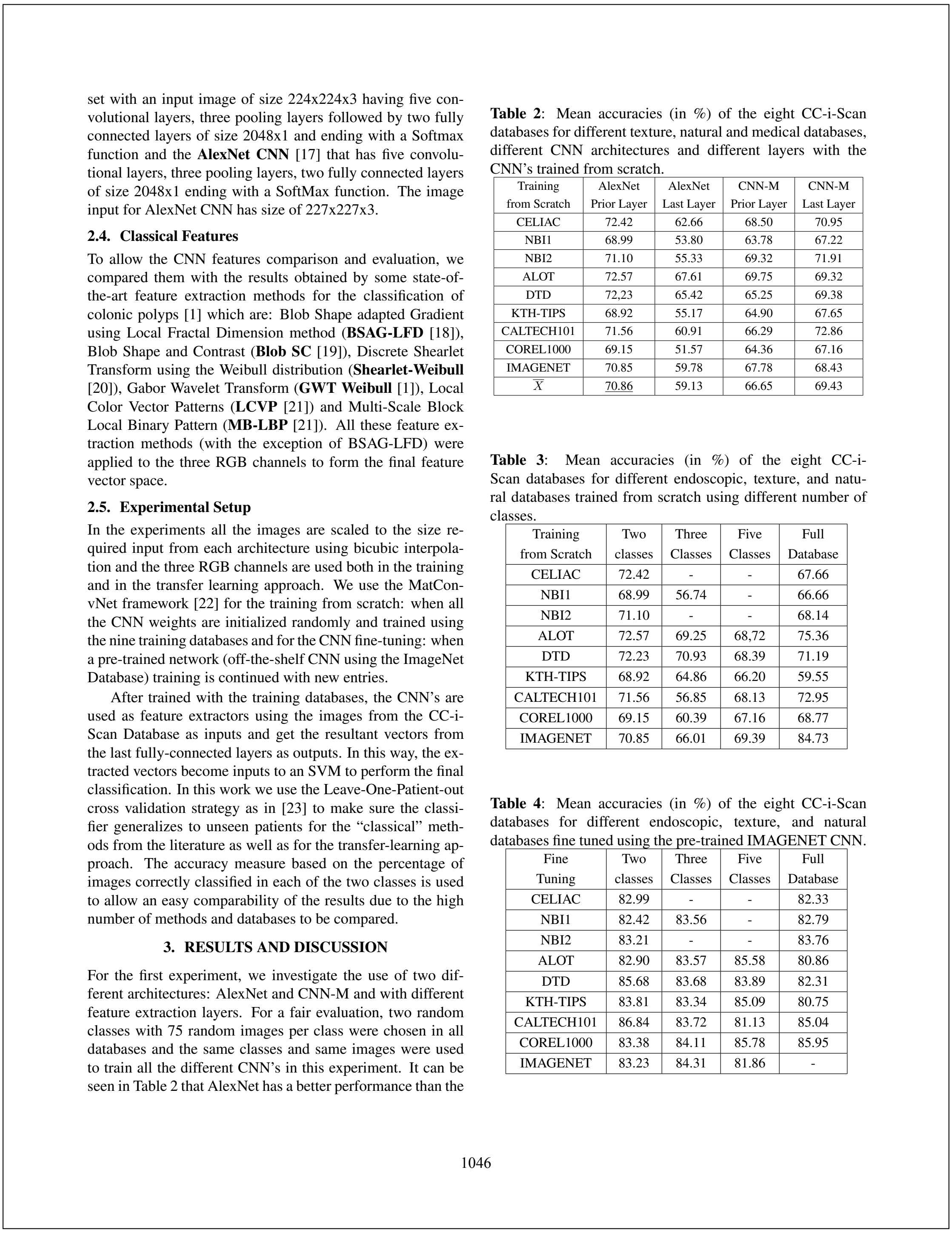
Texture Image Databases: The Amsterdam Library of Textures (**ALOT**) with 27500 rough texture images of size 384x256 divided into 250 classes [10]. The Describable Texture Dataset (**DTD**) with 5640 images of sizes range betwenn 300x300 and 640x640 categorized in 47 classes [11]. The Textures under varying Illumination, Pose and Scale (**KTH-TIPS**) database with 10 different materials containing 81 cropped images of size 200x200 in each class [12].

Natural Image Databases: The **IMAGENET** database [13] with 1.2 million images of size 256x256 categorized in 1000 classes. The **CALTECH101** Database is a natural image dataset with a list of objects belonging to 101 categories [14]. The **COREL1000** database is a natural image database containing 1000 color photographs showing natural scenes of ten different categories [15].

2.3. CNN Architectures

A Convolutional Neural Network is similar to traditional Neural Networks in the sense of being constructed by neuron layers with their respective weights, biases and activation functions. The architecture of a CNN is formed by a stack of distinctive convolutional, activation and pooling layers transforming the input volumes into an output volume through a differentiable function. After a series of convolutional and pooling layers, the CNN ended up with a fully connected layer for the high-level reasoning using a loss layer to train the weights in the back-propagation training.

Two CNN architectures widely used in the literature and that have obtained good results using off-the-shelf features for colonic polyp classification in [7] were chosen for the experiments: The **CNN-M** architecture (medium CNN) [16] that is



set with an input image of size 224x224x3 having five convolutional layers, three pooling layers followed by two fully connected layers of size 2048x1 and ending with a Softmax function and the **AlexNet CNN** [17] that has five convolutional layers, three pooling layers, two fully connected layers of size 2048x1 ending with a SoftMax function. The image input for AlexNet CNN has size of 227x227x3.

2.4. Classical Features

To allow the CNN features comparison and evaluation, we compared them with the results obtained by some state-of-the-art feature extraction methods for the classification of colonic polyps [1] which are: Blob Shape adapted Gradient using Local Fractal Dimension method (**BSAG-LFD** [18]), Blob Shape and Contrast (**Blob SC** [19]), Discrete Shearlet Transform using the Weibull distribution (**Shearlet-Weibull** [20]), Gabor Wavelet Transform (**GWT Weibull** [1]), Local Color Vector Patterns (**LCVP** [21]) and Multi-Scale Block Local Binary Pattern (**MB-LBP** [21]). All these feature extraction methods (with the exception of BSAG-LFD) were applied to the three RGB channels to form the final feature vector space.

2.5. Experimental Setup

In the experiments all the images are scaled to the size required input from each architecture using bicubic interpolation and the three RGB channels are used both in the training and in the transfer learning approach. We use the MatConvNet framework [22] for the training from scratch: when all the CNN weights are initialized randomly and trained using the nine training databases and for the CNN fine-tuning: when a pre-trained network (off-the-shelf CNN using the ImageNet Database) training is continued with new entries.

After trained with the training databases, the CNN's are used as feature extractors using the images from the CC-i-Scan Database as inputs and get the resultant vectors from the last fully-connected layers as outputs. In this way, the extracted vectors become inputs to an SVM to perform the final classification. In this work we use the Leave-One-Patient-out cross validation strategy as in [23] to make sure the classifier generalizes to unseen patients for the "classical" methods from the literature as well as for the transfer-learning approach. The accuracy measure based on the percentage of images correctly classified in each of the two classes is used to allow an easy comparability of the results due to the high number of methods and databases to be compared.

3. RESULTS AND DISCUSSION

For the first experiment, we investigate the use of two different architectures: AlexNet and CNN-M and with different feature extraction layers. For a fair evaluation, two random classes with 75 random images per class were chosen in all databases and the same classes and same images were used to train all the different CNN's in this experiment. It can be seen in Table 2 that AlexNet has a better performance than the

Table 2: Mean accuracies (in %) of the eight CC-i-Scan databases for different texture, natural and medical databases, different CNN architectures and different layers with the CNN's trained from scratch.

Training from Scratch	AlexNet Prior Layer	AlexNet Last Layer	CNN-M Prior Layer	CNN-M Last Layer
CELIAC	72.42	62.66	68.50	70.95
NBI1	68.99	53.80	63.78	67.22
NBI2	71.10	55.33	69.32	71.91
ALOT	72.57	67.61	69.75	69.32
DTD	72,23	65.42	65.25	69.38
KTH-TIPS	68.92	55.17	64.90	67.65
CALTECH101	71.56	60.91	66.29	72.86
COREL1000	69.15	51.57	64.36	67.16
IMAGENET	70.85	59.78	67.78	68.43
$\overline{X}$	70.86	59.13	66.65	69.43

Table 3: Mean accuracies (in %) of the eight CC-i-Scan databases for different endoscopic, texture, and natural databases trained from scratch using different number of classes.

Training from Scratch	Two classes	Three Classes	Five Classes	Full Database
CELIAC	72.42	-	-	67.66
NBI1	68.99	56.74	-	66.66
NBI2	71.10	-	-	68.14
ALOT	72.57	69.25	68,72	75.36
DTD	72.23	70.93	68.39	71.19
KTH-TIPS	68.92	64.86	66.20	59.55
CALTECH101	71.56	56.85	68.13	72.95
COREL1000	69.15	60.39	67.16	68.77
IMAGENET	70.85	66.01	69.39	84.73

Table 4: Mean accuracies (in %) of the eight CC-i-Scan databases for different endoscopic, texture, and natural databases fine tuned using the pre-trained IMAGENET CNN.

Fine Tuning	Two classes	Three Classes	Five Classes	Full Database
CELIAC	82.99	-	-	82.33
NBI1	82.42	83.56	-	82.79
NBI2	83.21	-	-	83.76
ALOT	82.90	83.57	85.58	80.86
DTD	85.68	83.68	83.89	82.31
KTH-TIPS	83.81	83.34	85.09	80.75
CALTECH101	86.84	83.72	81.13	85.04
COREL1000	83.38	84.11	85.78	85.95
IMAGENET	83.23	84.31	81.86	-

Table 5: Accuracies of the methods for the CC-i-Scan databases in %.

Methods	No staining				Staining				
	¬CVC	i-Scan1	i-Scan2	i-Scan3	¬CVC	i-Scan1	i-Scan2	i-Scan3	$\overline{X}$
1: CALTECH101 AlexNet FT (Two Classes)	94,66	85.33	83.15	87.51	89.18	85.18	85.03	84.68	**86.84**
2: DTD AlexNet FT (Two Classes)	92.09	84.00	88.88	84.98	90.83	79.78	84.27	80.62	85.68
3: BSAG-LFD	86.27	86.87	84.60	82.87	70.20	80.63	78.78	71.39	80.20
4: Blob SC	77.67	83.33	82.10	75.22	59.28	78.83	66.13	59.83	72.79
5: Shearlet-Weibull	73.72	76.67	79.60	86.80	81.30	69.91	72.38	83.63	78.00
6: GWT-Weibull	79.75	78.67	70.25	84.28	81.30	74.54	77.17	83.39	78.66
7: LCVP	76.60	66.00	47.75	77.12	77.45	79.00	70.01	69.56	70.43
8: MB-LBP	78.26	80.67	81.38	83.37	69.29	70.60	77.22	78.32	77.38
Concatenating 1/2/3/6	*96.63*	*89.33*	*88.88*	*85.89*	*89.64*	*85.51*	*88.96*	*88.23*	***89.13***

CNN-M architecture specially using the prior fully connected layer.

Using the best configuration obtained in the first experiment (AlexNet trained from scratch using the prior fully connect layer as feature extractor), in the second experiment we decided to examine different number of classes maintaining the number of images: two classes of 75 images each, three classes of 50 images and 5 classes of 30 images each class besides testing the use of the full database to train the CNN's. It can be seen in Table 3 that with the same number of images and classes, texture databases perform better than natural image databases specially in the ALOT, CELIAC and DTD databases. Despite the fact that the CELIAC database presents good results, the databases containing colonic polyp images (NBI2 and NBI2) do not present better results. This can be explained by the different nature of NBI imaging where the pits are indirectly observable due to the spectral transmittance. It also can be noted that, in a fair comparison (with the same number of images in all database) when the number of classes is the same of the target database (two classes), the results are better than using more classes. It is also interesting to note that, when the number of images and classes are increased (in case of the use of the full database) some results are worse than using a lower number of classes and images classes, e.g. as in the case of DTD, KTH-TIPS, CELIAC, NBI1, NBI2 and COREL1000 databases.

In the third experiment we used the trained IMAGENET CNN to perform fine tuning using the other databases and Table 4 present the obtained results. It can be noticed that, in the case of fine tuning when the number of classes becomes closer to the number of classes from the original IMAGENET CNN, the results are improved. It can also be seen that using databases more related to the original database the results can be better, even surpassing the results from the original IMAGENET CNN in the case of CALTECH101 using two classes (86.84 %)) and the full database (85.04%)) and COREL1000 using the full database (85.95%) against the IMAGENET trained from scratch (84.73%).

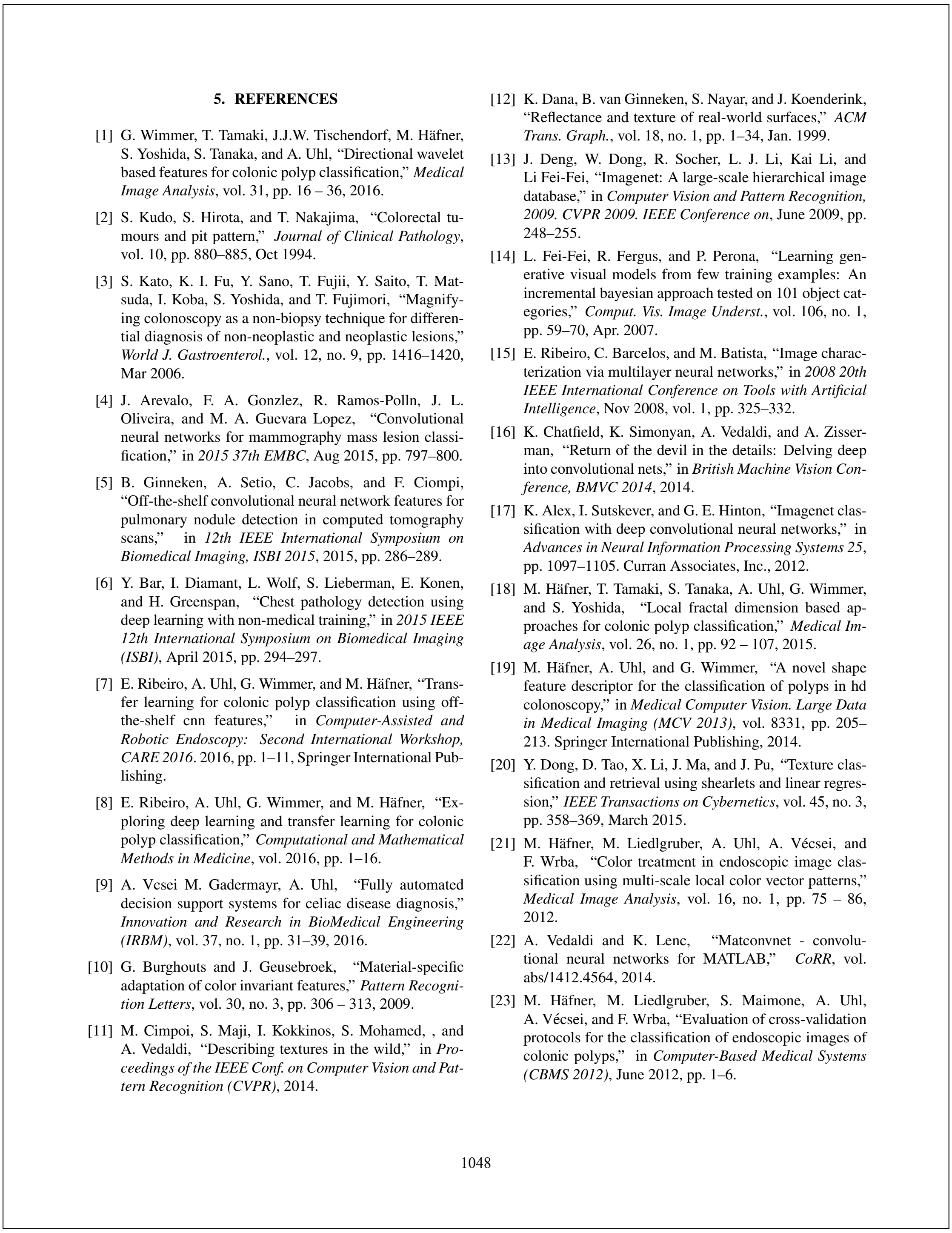
In Table 5 we present the results in a more detailed way separating the accuracies from each of the eight CC-i-Scan databases. We choose the best results obtained from the previous experiments comparing them with the classical features used for colonic polyp classification. It can be seen that the CNN's perform better than all the classic features, especially when trained with more images which is the case of the AlexNet CNN fine tuned (FT) with the CALTECH101 database with two classes (86.84% of accuracy). Applying feature fusion in the classification process with these two bests CNN's with the two classic features that presented the best results in average (BSAG-LFD and GWT-Weibull) presented the best result of all: 89.13% in average showing that different features from completely different nature can complement each other.

4. CONCLUSION AND FUTURE WORKS

In this work, we explored transfer learning across different classification problems via CNN's to surpass the lack of training data in the Colonic Polyp Classification task. We showed that transfer learning can be a successfully alternative to extract relevant features by leveraging knowledge learned on other datasets even in very different tasks.

We also proved that when the number of classes and the nature of the images are similar to the target database, the results are better as well as with the number of the images in the training database. On the basis of the good results obtained compared to the classical features we can conclude that the CNN's have a good generalization capability for the transfer learning specially using texture databases and with the fine tunning approach. We also showed that when the texture database for the CNN trained is also limited, the fine tuning with a bigger database can be a good alternative to surpass this problem even with a completely different original database since the number of images is very high.

As we have chosen fixed classes (randomly) in the training datasets for this work, in future work we plan to randomize the procedure by repeatedly applying this strategy and explore the average accuracy of the results to look deeper into the transfer learning final classification. We also plan to build a massive texture database to improve the results and use this strategy to also test the detection of colonic polyps directly into video frames and evaluate the performance in real time applications as well as to use this strategy in other endoscopic databases such as automatic classification of celiac disease.



5. REFERENCES

[1] G. Wimmer, T. Tamaki, J.J.W. Tischendorf, M. Häfner, S. Yoshida, S. Tanaka, and A. Uhl, “Directional wavelet based features for colonic polyp classification,” *Medical Image Analysis*, vol. 31, pp. 16 – 36, 2016.

[2] S. Kudo, S. Hirota, and T. Nakajima, “Colorectal tumours and pit pattern,” *Journal of Clinical Pathology*, vol. 10, pp. 880–885, Oct 1994.

[3] S. Kato, K. I. Fu, Y. Sano, T. Fujii, Y. Saito, T. Matsuda, I. Koba, S. Yoshida, and T. Fujimori, “Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions,” *World J. Gastroenterol.*, vol. 12, no. 9, pp. 1416–1420, Mar 2006.

[4] J. Arevalo, F. A. Gonzlez, R. Ramos-Polln, J. L. Oliveira, and M. A. Guevara Lopez, “Convolutional neural networks for mammography mass lesion classification,” in *2015 37th EMBC*, Aug 2015, pp. 797–800.

[5] B. Ginneken, A. Setio, C. Jacobs, and F. Ciompi, “Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans,” in *12th IEEE International Symposium on Biomedical Imaging, ISBI 2015*, 2015, pp. 286–289.

[6] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, “Chest pathology detection using deep learning with non-medical training,” in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, April 2015, pp. 294–297.

[7] E. Ribeiro, A. Uhl, G. Wimmer, and M. Häfner, “Transfer learning for colonic polyp classification using off-the-shelf cnn features,” in *Computer-Assisted and Robotic Endoscopy: Second International Workshop, CARE 2016*. 2016, pp. 1–11, Springer International Publishing.

[8] E. Ribeiro, A. Uhl, G. Wimmer, and M. Häfner, “Exploring deep learning and transfer learning for colonic polyp classification,” *Computational and Mathematical Methods in Medicine*, vol. 2016, pp. 1–16.

[9] A. Vcsei M. Gadermayr, A. Uhl, “Fully automated decision support systems for celiac disease diagnosis,” *Innovation and Research in BioMedical Engineering (IRBM)*, vol. 37, no. 1, pp. 31–39, 2016.

[10] G. Burghouts and J. Geusebroek, “Material-specific adaptation of color invariant features,” *Pattern Recognition Letters*, vol. 30, no. 3, pp. 306 – 313, 2009.

[11] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi, “Describing textures in the wild,” in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[12] K. Dana, B. van Ginneken, S. Nayar, and J. Koenderink, “Reflectance and texture of real-world surfaces,” *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, Jan. 1999.

[13] J. Deng, W. Dong, R. Socher, L. J. Li, Kai Li, and Li Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 248–255.

[14] L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories,” *Comput. Vis. Image Underst.*, vol. 106, no. 1, pp. 59–70, Apr. 2007.

[15] E. Ribeiro, C. Barcelos, and M. Batista, “Image characterization via multilayer neural networks,” in *2008 20th IEEE International Conference on Tools with Artificial Intelligence*, Nov 2008, vol. 1, pp. 325–332.

[16] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, “Return of the devil in the details: Delving deep into convolutional nets,” in *British Machine Vision Conference, BMVC 2014*, 2014.

[17] K. Alex, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. Curran Associates, Inc., 2012.

[18] M. Häfner, T. Tamaki, S. Tanaka, A. Uhl, G. Wimmer, and S. Yoshida, “Local fractal dimension based approaches for colonic polyp classification,” *Medical Image Analysis*, vol. 26, no. 1, pp. 92 – 107, 2015.

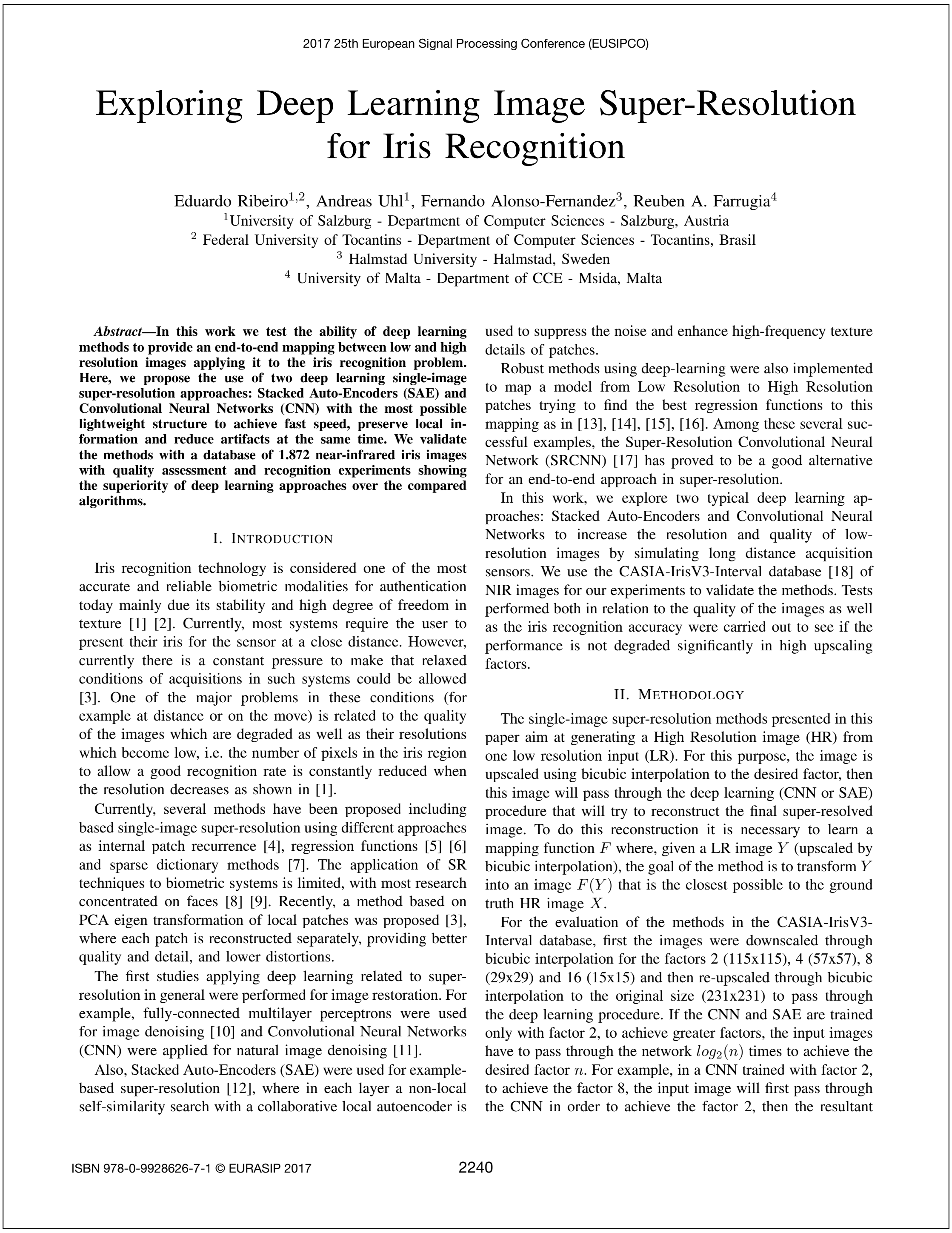
[19] M. Häfner, A. Uhl, and G. Wimmer, “A novel shape feature descriptor for the classification of polyps in hd colonoscopy,” in *Medical Computer Vision. Large Data in Medical Imaging (MCV 2013)*, vol. 8331, pp. 205–213. Springer International Publishing, 2014.

[20] Y. Dong, D. Tao, X. Li, J. Ma, and J. Pu, “Texture classification and retrieval using shearlets and linear regression,” *IEEE Transactions on Cybernetics*, vol. 45, no. 3, pp. 358–369, March 2015.

[21] M. Häfner, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, “Color treatment in endoscopic image classification using multi-scale local color vector patterns,” *Medical Image Analysis*, vol. 16, no. 1, pp. 75 – 86, 2012.

[22] A. Vedaldi and K. Lenc, “Matconvnet - convolutional neural networks for MATLAB,” *CoRR*, vol. abs/1412.4564, 2014.

[23] M. Häfner, M. Liedlgruber, S. Maimone, A. Uhl, A. Vécsei, and F. Wrba, “Evaluation of cross-validation protocols for the classification of endoscopic images of colonic polyps,” in *Computer-Based Medical Systems (CBMS 2012)*, June 2012, pp. 1–6.



2017 25th European Signal Processing Conference (EUSIPCO)

Exploring Deep Learning Image Super-Resolution for Iris Recognition

Eduardo Ribeiro[1,2], Andreas Uhl[1], Fernando Alonso-Fernandez[3], Reuben A. Farrugia[4]
[1]University of Salzburg - Department of Computer Sciences - Salzburg, Austria
[2] Federal University of Tocantins - Department of Computer Sciences - Tocantins, Brasil
[3] Halmstad University - Halmstad, Sweden
[4] University of Malta - Department of CCE - Msida, Malta

***Abstract*—In this work we test the ability of deep learning methods to provide an end-to-end mapping between low and high resolution images applying it to the iris recognition problem. Here, we propose the use of two deep learning single-image super-resolution approaches: Stacked Auto-Encoders (SAE) and Convolutional Neural Networks (CNN) with the most possible lightweight structure to achieve fast speed, preserve local information and reduce artifacts at the same time. We validate the methods with a database of 1.872 near-infrared iris images with quality assessment and recognition experiments showing the superiority of deep learning approaches over the compared algorithms.**

I. Introduction

Iris recognition technology is considered one of the most accurate and reliable biometric modalities for authentication today mainly due its stability and high degree of freedom in texture [1] [2]. Currently, most systems require the user to present their iris for the sensor at a close distance. However, currently there is a constant pressure to make that relaxed conditions of acquisitions in such systems could be allowed [3]. One of the major problems in these conditions (for example at distance or on the move) is related to the quality of the images which are degraded as well as their resolutions which become low, i.e. the number of pixels in the iris region to allow a good recognition rate is constantly reduced when the resolution decreases as shown in [1].

Currently, several methods have been proposed including based single-image super-resolution using different approaches as internal patch recurrence [4], regression functions [5] [6] and sparse dictionary methods [7]. The application of SR techniques to biometric systems is limited, with most research concentrated on faces [8] [9]. Recently, a method based on PCA eigen transformation of local patches was proposed [3], where each patch is reconstructed separately, providing better quality and detail, and lower distortions.

The first studies applying deep learning related to super-resolution in general were performed for image restoration. For example, fully-connected multilayer perceptrons were used for image denoising [10] and Convolutional Neural Networks (CNN) were applied for natural image denoising [11].

Also, Stacked Auto-Encoders (SAE) were used for example-based super-resolution [12], where in each layer a non-local self-similarity search with a collaborative local autoencoder is used to suppress the noise and enhance high-frequency texture details of patches.

Robust methods using deep-learning were also implemented to map a model from Low Resolution to High Resolution patches trying to find the best regression functions to this mapping as in [13], [14], [15], [16]. Among these several successful examples, the Super-Resolution Convolutional Neural Network (SRCNN) [17] has proved to be a good alternative for an end-to-end approach in super-resolution.

In this work, we explore two typical deep learning approaches: Stacked Auto-Encoders and Convolutional Neural Networks to increase the resolution and quality of low-resolution images by simulating long distance acquisition sensors. We use the CASIA-IrisV3-Interval database [18] of NIR images for our experiments to validate the methods. Tests performed both in relation to the quality of the images as well as the iris recognition accuracy were carried out to see if the performance is not degraded significantly in high upscaling factors.

II. Methodology

The single-image super-resolution methods presented in this paper aim at generating a High Resolution image (HR) from one low resolution input (LR). For this purpose, the image is upscaled using bicubic interpolation to the desired factor, then this image will pass through the deep learning (CNN or SAE) procedure that will try to reconstruct the final super-resolved image. To do this reconstruction it is necessary to learn a mapping function F where, given a LR image Y (upscaled by bicubic interpolation), the goal of the method is to transform Y into an image $F(Y)$ that is the closest possible to the ground truth HR image X.

For the evaluation of the methods in the CASIA-IrisV3-Interval database, first the images were downscaled through bicubic interpolation for the factors 2 (115x115), 4 (57x57), 8 (29x29) and 16 (15x15) and then re-upscaled through bicubic interpolation to the original size (231x231) to pass through the deep learning procedure. If the CNN and SAE are trained only with factor 2, to achieve greater factors, the input images have to pass through the network $log_2(n)$ times to achieve the desired factor n. For example, in a CNN trained with factor 2, to achieve the factor 8, the input image will first pass through the CNN in order to achieve the factor 2, then the resultant

ISBN 978-0-9928626-7-1 © EURASIP 2017

image will pass again to the CNN to achieve the factor 4 and so on.

In this work we take advantage of a common strategy used in image restoration, which is the extraction of patches and their representations as a series of pre-trained bases (such as PCA, DCT, Haar among other). Such filters are convolved with the image and in the case of this work will be optimized so that the mapping is the best possible. This can be done in one, two, or more layers and in the case of this work are followed by a reconstruction step which the predicted overlapping high-resolution patches are averaged to produce the final image. This strategy is used both in the SAEs and CNNs that will be explained in the next subsections.

A. Convolutional Neural Networks

Generally, the input of a CNN is a $(m \times m \times d)$ image where $(m \times m)$ is the dimension of the patch and d the number of channels (depth) of the image [19]. In this work, for the CNN training, patches are extracted from the HR images where $m = 33$ and $d = 1$, then the patches are downscaled (depending on the factor chosen for the method) and re-upscaled to the original size both using bicubic interpolation as it can be seen in the Figure 1.

In this work, the implemented CNN has three convolutional layers, where: the first layer consists of 64 filters of size 9x9x1 with stride 1 and padding 0, the second layer with 32 filters of size 1x1x64 with stride 1 and padding 0, and the last layer with 1 filter of size 5x5x32 with stride 1 and padding 0. With all paddings set to zero, the feature maps will decrease in size resulting in a patch of size 21x21. When the training is done, overlapping patches will be extracted from the LR images (upscaled using bicubic interpolation) with stride 1 and only the central pixel of the resulting feature map will be used which means that the smaller size of the resulted feature map will not influence the final image result.

After each convolutional layer a non-linearity (or activation) function is applied to the feature maps mainly to accelerate the convergence of the stochastic gradient algorithm called ReLU rectifier function: $f(x) = \max(0, x)$, where x is the neuron input.

For the training with the high-resolution patches with their correspondent low-resolution patches we use the Mean Squared Error (MSE) as the loss function trying to achieved the best PSNR as possible when the CNN is completely trained and the loss minimization is done using stochastic gradient descent with the standard backpropagation method.

In this work we tested three different approaches for the CNN training:

- From scratch (CNN FS): When the CNN weights are initialized randomly and trained according to the target image database (in the case of this work: the CASIA Interval V3 Iris Database) for the kernels domain adaptation, that is, to find the best way to map the data in order to perform the super-resolution.
- Transfer Learning (CNN TL): When an **off-the-shelf CNN** is chosen, which means that the CNN is pre-trained with a different database (in the case of this work: the ImageNet Database [20]) then it is used to perform the super-resolution in the target image database.
- Fine Tuning (CNN FT): The pre-trained network (off-the-shelf CNN) training is continued with new entries (with the target image database) for the weights to adjust properly to the new scenario reinforcing the more generic features with a lower probability of overfitting.

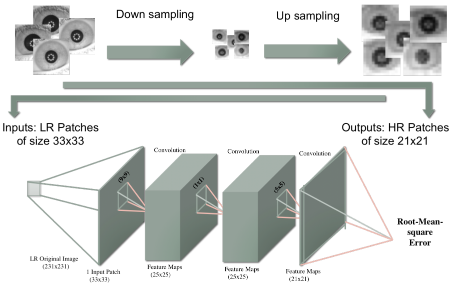


Fig. 1: An illustration of the Convolutional Neural Network architecture for Iris Super-Resolution.

B. Stacked Auto-Enconders

For the Layer-wise pre-training of Stacked Auto-Encoders we use the HR patches downscaled and upscaled again using bicubic interpolation in the same way as for the CNN. However, in this case, the matrix is turned to a vector in order to fit in the auto-encoder architecture. These vectors are used for the first auto-encoder as can be seen in Figure 2 that are trained until a threshold is reached. In the second auto-encoder, we use the vector that we got from the hidden layer of the previous trained auto-encoder as input, and proceed in the first auto-encoder. The same process is applied to the third layer and so on [21]. Then, we use the original images (HR patches) as the targets in the last layer of the output auto-encoder. These targets are used to update the parameter of the deep multi-layered neural network (Stacked Auto-Encoders) by means of a supervised error backpropagation algorithm. This process tries to reconstruct the image patch by generalizing the missing pixels with the auto-encoder weights learned from the all images of the training database.

When the training is completed, the auto-encoder is used to propagate all the LR patches upscaled using bicubic intepolation resulting in the reconstructed super-resolution patches in a magnification of 2 (when the training is done with this magnification). To achieve a magnification factor of 4, it is necessary to reinsert the reconstructed super-resolution images to the network in the same way as explained for the CNN approach.

For the experiments we trained four auto-encoders with the empirically chosen configuration: 1089-1000-1089 (where

1089 means the 33x33 input patches), 1000-2000-1000, 2000-2600-2000, 2600-2000-2600. Consequently, in the fine-tuning phase, the NN configuration for the Stacked Auto-Encoder experiment is: 1089-1000-2000-2600-2000-441. The size of the output (21x21 pixels) is because, in this case, a triangular architecture with more inputs than outputs can help the convergence in the fine-tuning phase.

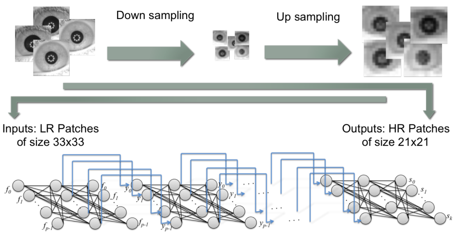

Down sampling
Up sampling
Inputs: LR Patches of size 33x33
Outputs: HR Patches of size 21x21

Fig. 2: An illustration of the Stacked Auto-Encoder architecture for Iris Super-Resolution.

III. Experimental Setup

For the experiments we use the CASIA Interval v3 iris database that contains a total of 2.655 NIR images of size 280x320 pixels, from 249 subjects captured with a self-developed close-up camera, resulting in 396 different eyes. Manual segmentation annotation of the database is available [3], which is used as input for our experiments. In the pre-processing step all images are resized via bicubic interpolation in order to have the same sclera radius and are aligned by extracting a square region of 231x231 around the pupil center. All images that do not fit in this requirement (for example when the eye is close to the image border) are discarded. After this, the 1.872 remaining images are used in the experiments. For the deep learning training and tests, the pre-processed dataset is divided into two separated sets: 925 images from the first 116 users for the training and 947 images from the remaining 133 users for the tests (we consider each eye as a different user). This set division by users is important to make sure that the same pattern (in the patches) will not be used both in training and testing steps.

To evaluate the performance of the methods by quality assessment algorithms we use the Peak Signal to Noise Ratio **(PSNR)**, that is the ratio between the peak signal and the power of corrupting noise that affects the fidelity of its representation, the Structural Similarity Index Measure **(SSIM)** that extracts three separate scores (visual influence, contrast and structural score) combining them to the final score, and the Visual Information Fidelity **(VIF)** that calculates the mutual information between input and the output of the HVS channel when no distortion is present and the mutual information between the input of the distortion channel and the output of the HVS channel for the test signal [22]. In these metrics, a high metric score reflects a high quality. For the quality tests, all images from the database were used in high resolution as reference images. We compare our method with bilinear and bicubic interpolation as well as to PCA hallucination of local patches used in [3].

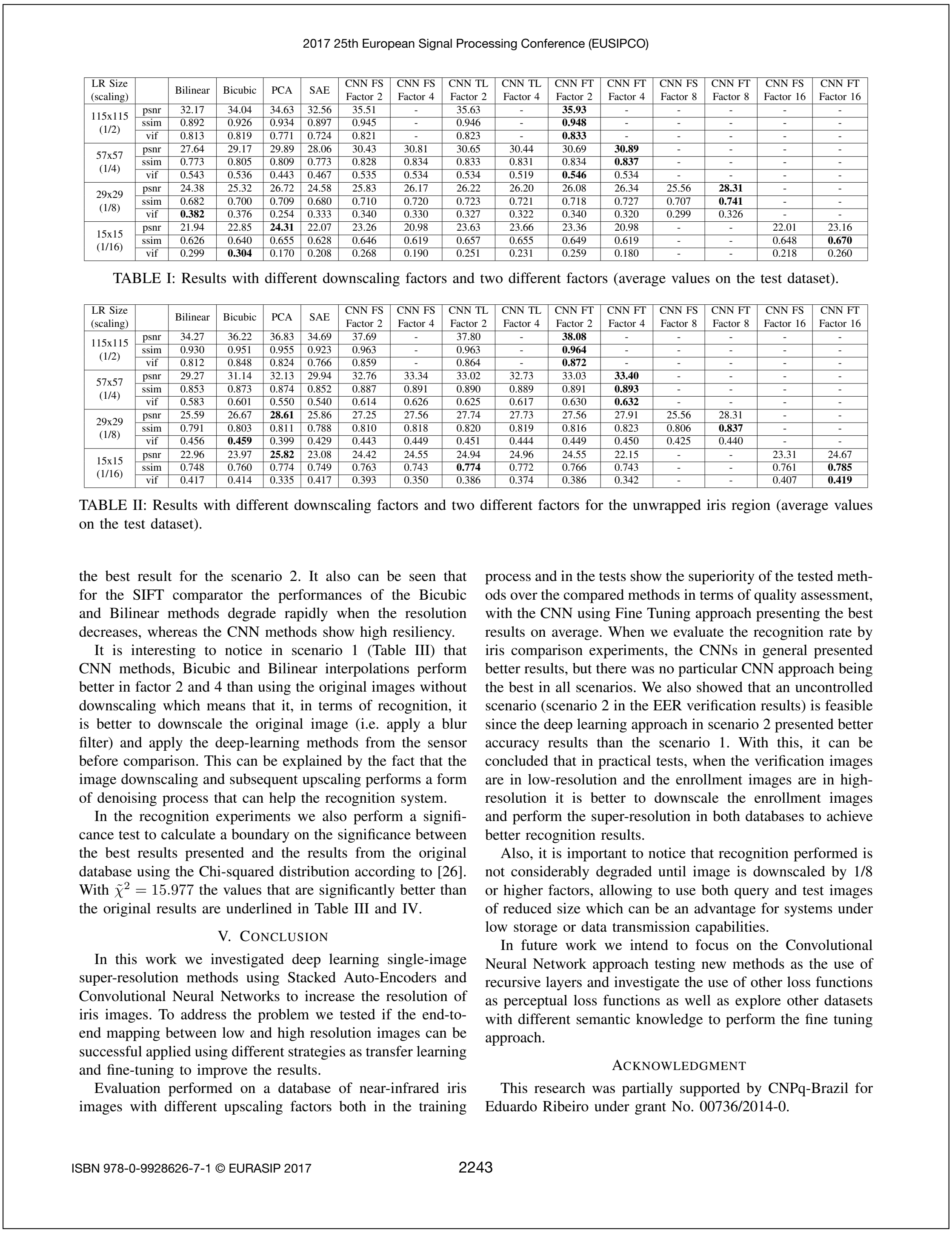
We also conduct recognition experiments using reconstructed images to evaluate the iris recognition performance. In this procedure, first the iris is unwrapped to a normalized rectangle of 20x240 pixels using the Daugman's rubber sheet model [23], then a 1D Log-Gabor (LG) wavelet is applied with a phase binary quantization to 4 levels [24]. The comparison between the binary vectors is done by the normalized Hamming Distance [23] where the rotation is accounted for by shifting the grid of the query image in counter- and clock-wise directions, and selecting the lowest distance that corresponds to the best match. We also implemented a SIFT comparator in which SIFT feature points in scale space are extracted from the iris region (without unwrapping) and the comparison is performed based on the texture information around the feature points using the SIFT operator [25].

IV. Results

The results of the quality assessment for the test images and for the normalized iris region (20x240) are shown in Table I and Table II. It can be seen in Table I that the use of the Convolutional Neural Networks outperforms the traditional methods of interpolation (bicubic and bilinear) as well as the eigen-patch hallucination (PCA) method, mainly for small downscaling factors. It also can be noticed that the use of the Fine Tuning strategy improves the results by merging the use of natural and iris images during the CNN training. Also, when the CNN is trained with the same downscaling factor as the tests, the results are also becoming more resilient for lower resolutions. It can also be noticed that, for low resolutions, the quality assessment algorithms present different best results which can make the results interpretation difficult.

In iris recognition verification we consider two scenarios: 1) enrollment samples taken from original HR input images, and query samples taken from reconstructed super-resolution results (Table III) simulating a controlled enrollment scenario (for example, when the user is registered using a HR sensor and make use of the system using a cellphone camera with certain distance); and 2) both enrollment and query samples taken from the reconstructed super-resolution results (Table IV) simulating a totally uncontrolled scenario (for example, when the user is registered using a cellphone and make use of the system also using a cellphone camera with certain distance).

It can be observed that the performance of CNNs are the best for small downscaling factors in both scenarios in general, despite of the diversity of good results among the training approaches. Using the Log-gabor comparator the CNN using Fine Tuning and Transfer Learning approach beats the other methods except for the lowest resolution that PCA does best. For the SIFT comparator the CNNs are better. However, there is no particular winning training approach, in this case, using the downscaling factor of 2 the SAE method present



LR Size (scaling)		Bilinear	Bicubic	PCA	SAE	CNN FS Factor 2	CNN FS Factor 4	CNN TL Factor 2	CNN TL Factor 4	CNN FT Factor 2	CNN FT Factor 4	CNN FS Factor 8	CNN FT Factor 8	CNN FS Factor 16	CNN FT Factor 16
115x115 (1/2)	psnr	32.17	34.04	34.63	32.56	35.51	-	35.63	-	**35.93**	-	-	-	-	-
	ssim	0.892	0.926	0.934	0.897	0.945	-	0.946	-	**0.948**	-	-	-	-	-
	vif	0.813	0.819	0.771	0.724	0.821	-	0.823	-	**0.833**	-	-	-	-	-
57x57 (1/4)	psnr	27.64	29.17	29.89	28.06	30.43	30.81	30.65	30.44	30.69	**30.89**	-	-	-	-
	ssim	0.773	0.805	0.809	0.773	0.828	0.834	0.833	0.831	0.834	**0.837**	-	-	-	-
	vif	0.543	0.536	0.443	0.467	0.535	0.534	0.534	0.519	**0.546**	0.534	-	-	-	-
29x29 (1/8)	psnr	24.38	25.32	26.72	24.58	25.83	26.17	26.22	26.20	26.08	26.34	25.56	**28.31**	-	-
	ssim	0.682	0.700	0.709	0.680	0.710	0.720	0.723	0.721	0.718	0.727	0.707	**0.741**	-	-
	vif	**0.382**	0.376	0.254	0.333	0.340	0.330	0.327	0.322	0.340	0.320	0.299	0.326	-	-
15x15 (1/16)	psnr	21.94	22.85	**24.31**	22.07	23.26	20.98	23.63	23.66	23.36	20.98	-	-	22.01	23.16
	ssim	0.626	0.640	0.655	0.628	0.646	0.619	0.657	0.655	0.649	0.619	-	-	0.648	**0.670**
	vif	0.299	**0.304**	0.170	0.208	0.268	0.190	0.251	0.231	0.259	0.180	-	-	0.218	0.260

TABLE I: Results with different downscaling factors and two different factors (average values on the test dataset).

LR Size (scaling)		Bilinear	Bicubic	PCA	SAE	CNN FS Factor 2	CNN FS Factor 4	CNN TL Factor 2	CNN TL Factor 4	CNN FT Factor 2	CNN FT Factor 4	CNN FS Factor 8	CNN FT Factor 8	CNN FS Factor 16	CNN FT Factor 16
115x115 (1/2)	psnr	34.27	36.22	36.83	34.69	37.69	-	37.80	-	**38.08**	-	-	-	-	-
	ssim	0.930	0.951	0.955	0.923	0.963	-	0.963	-	**0.964**	-	-	-	-	-
	vif	0.812	0.848	0.824	0.766	0.859	-	0.864	-	**0.872**	-	-	-	-	-
57x57 (1/4)	psnr	29.27	31.14	32.13	29.94	32.76	33.34	33.02	32.73	33.03	**33.40**	-	-	-	-
	ssim	0.853	0.873	0.874	0.852	0.887	0.891	0.890	0.889	0.891	**0.893**	-	-	-	-
	vif	0.583	0.601	0.550	0.540	0.614	0.626	0.625	0.617	0.630	**0.632**	-	-	-	-
29x29 (1/8)	psnr	25.59	26.67	**28.61**	25.86	27.25	27.56	27.74	27.73	27.56	27.91	25.56	28.31	-	-
	ssim	0.791	0.803	0.811	0.788	0.810	0.818	0.820	0.819	0.816	0.823	0.806	**0.837**	-	-
	vif	0.456	**0.459**	0.399	0.429	0.443	0.449	0.451	0.444	0.449	0.450	0.425	0.440	-	-
15x15 (1/16)	psnr	22.96	23.97	**25.82**	23.08	24.42	24.55	24.94	24.96	24.55	22.15	-	-	23.31	24.67
	ssim	0.748	0.760	0.774	0.749	0.763	0.743	**0.774**	0.772	0.766	0.743	-	-	0.761	**0.785**
	vif	0.417	0.414	0.335	0.417	0.393	0.350	0.386	0.374	0.386	0.342	-	-	0.407	**0.419**

TABLE II: Results with different downscaling factors and two different factors for the unwrapped iris region (average values on the test dataset).

the best result for the scenario 2. It also can be seen that for the SIFT comparator the performances of the Bicubic and Bilinear methods degrade rapidly when the resolution decreases, whereas the CNN methods show high resiliency.

It is interesting to notice in scenario 1 (Table III) that CNN methods, Bicubic and Bilinear interpolations perform better in factor 2 and 4 than using the original images without downscaling which means that it, in terms of recognition, it is better to downscale the original image (i.e. apply a blur filter) and apply the deep-learning methods from the sensor before comparison. This can be explained by the fact that the image downscaling and subsequent upscaling performs a form of denoising process that can help the recognition system.

In the recognition experiments we also perform a significance test to calculate a boundary on the significance between the best results presented and the results from the original database using the Chi-squared distribution according to [26]. With $\tilde{\chi}^2 = 15.977$ the values that are significantly better than the original results are underlined in Table III and IV.

V. Conclusion

In this work we investigated deep learning single-image super-resolution methods using Stacked Auto-Encoders and Convolutional Neural Networks to increase the resolution of iris images. To address the problem we tested if the end-to-end mapping between low and high resolution images can be successful applied using different strategies as transfer learning and fine-tuning to improve the results.

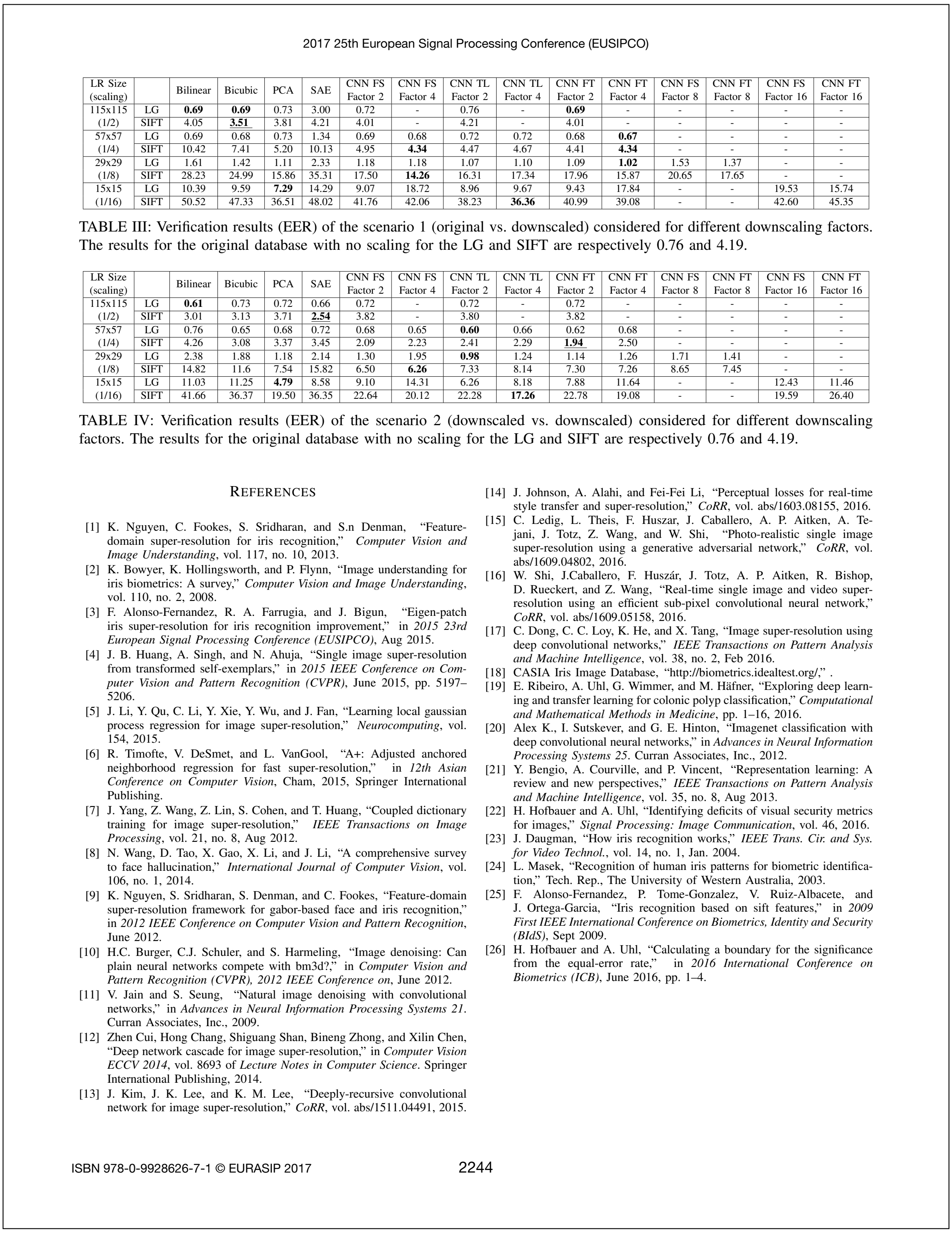
Evaluation performed on a database of near-infrared iris images with different upscaling factors both in the training process and in the tests show the superiority of the tested methods over the compared methods in terms of quality assessment, with the CNN using Fine Tuning approach presenting the best results on average. When we evaluate the recognition rate by iris comparison experiments, the CNNs in general presented better results, but there was no particular CNN approach being the best in all scenarios. We also showed that an uncontrolled scenario (scenario 2 in the EER verification results) is feasible since the deep learning approach in scenario 2 presented better accuracy results than the scenario 1. With this, it can be concluded that in practical tests, when the verification images are in low-resolution and the enrollment images are in high-resolution it is better to downscale the enrollment images and perform the super-resolution in both databases to achieve better recognition results.

Also, it is important to notice that recognition performed is not considerably degraded until image is downscaled by 1/8 or higher factors, allowing to use both query and test images of reduced size which can be an advantage for systems under low storage or data transmission capabilities.

In future work we intend to focus on the Convolutional Neural Network approach testing new methods as the use of recursive layers and investigate the use of other loss functions as perceptual loss functions as well as explore other datasets with different semantic knowledge to perform the fine tuning approach.

Acknowledgment

This research was partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.



LR Size (scaling)		Bilinear	Bicubic	PCA	SAE	CNN FS Factor 2	CNN FS Factor 4	CNN TL Factor 2	CNN TL Factor 4	CNN FT Factor 2	CNN FT Factor 4	CNN FS Factor 8	CNN FT Factor 8	CNN FS Factor 16	CNN FT Factor 16
115x115	LG	**0.69**	**0.69**	0.73	3.00	0.72	-	0.76	-	**0.69**	-	-	-	-	-
(1/2)	SIFT	4.05	**3.51**	3.81	4.21	4.01	-	4.21	-	4.01	-	-	-	-	-
57x57	LG	0.69	0.68	0.73	1.34	0.69	0.68	0.72	0.72	0.68	**0.67**	-	-	-	-
(1/4)	SIFT	10.42	7.41	5.20	10.13	4.95	**4.34**	4.47	4.67	4.41	**4.34**	-	-	-	-
29x29	LG	1.61	1.42	1.11	2.33	1.18	1.18	1.07	1.10	1.09	**1.02**	1.53	1.37	-	-
(1/8)	SIFT	28.23	24.99	15.86	35.31	17.50	**14.26**	16.31	17.34	17.96	15.87	20.65	17.65	-	-
15x15	LG	10.39	9.59	**7.29**	14.29	9.07	18.72	8.96	9.67	9.43	17.84	-	-	19.53	15.74
(1/16)	SIFT	50.52	47.33	36.51	48.02	41.76	42.06	38.23	**36.36**	40.99	39.08	-	-	42.60	45.35

TABLE III: Verification results (EER) of the scenario 1 (original vs. downscaled) considered for different downscaling factors. The results for the original database with no scaling for the LG and SIFT are respectively 0.76 and 4.19.

LR Size (scaling)		Bilinear	Bicubic	PCA	SAE	CNN FS Factor 2	CNN FS Factor 4	CNN TL Factor 2	CNN TL Factor 4	CNN FT Factor 2	CNN FT Factor 4	CNN FS Factor 8	CNN FT Factor 8	CNN FS Factor 16	CNN FT Factor 16
115x115	LG	**0.61**	0.73	0.72	0.66	0.72	-	0.72	-	0.72	-	-	-	-	-
(1/2)	SIFT	3.01	3.13	3.71	**2.54**	3.82	-	3.80	-	3.82	-	-	-	-	-
57x57	LG	0.76	0.65	0.68	0.72	0.68	0.65	**0.60**	0.66	0.62	0.68	-	-	-	-
(1/4)	SIFT	4.26	3.08	3.37	3.45	2.09	2.23	2.41	2.29	**1.94**	2.50	-	-	-	-
29x29	LG	2.38	1.88	1.18	2.14	1.30	1.95	**0.98**	1.24	1.14	1.26	1.71	1.41	-	-
(1/8)	SIFT	14.82	11.6	7.54	15.82	6.50	**6.26**	7.33	8.14	7.30	7.26	8.65	7.45	-	-
15x15	LG	11.03	11.25	**4.79**	8.58	9.10	14.31	6.26	8.18	7.88	11.64	-	-	12.43	11.46
(1/16)	SIFT	41.66	36.37	19.50	36.35	22.64	20.12	22.28	**17.26**	22.78	19.08	-	-	19.59	26.40

TABLE IV: Verification results (EER) of the scenario 2 (downscaled vs. downscaled) considered for different downscaling factors. The results for the original database with no scaling for the LG and SIFT are respectively 0.76 and 4.19.

References

[1] K. Nguyen, C. Fookes, S. Sridharan, and S.n Denman, "Feature-domain super-resolution for iris recognition," *Computer Vision and Image Understanding*, vol. 117, no. 10, 2013.

[2] K. Bowyer, K. Hollingsworth, and P. Flynn, "Image understanding for iris biometrics: A survey," *Computer Vision and Image Understanding*, vol. 110, no. 2, 2008.

[3] F. Alonso-Fernandez, R. A. Farrugia, and J. Bigun, "Eigen-patch iris super-resolution for iris recognition improvement," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015.

[4] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 5197–5206.

[5] J. Li, Y. Qu, C. Li, Y. Xie, Y. Wu, and J. Fan, "Learning local gaussian process regression for image super-resolution," *Neurocomputing*, vol. 154, 2015.

[6] R. Timofte, V. DeSmet, and L. VanGool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *12th Asian Conference on Computer Vision*, Cham, 2015, Springer International Publishing.

[7] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, Aug 2012.

[8] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *International Journal of Computer Vision*, vol. 106, no. 1, 2014.

[9] K. Nguyen, S. Sridharan, S. Denman, and C. Fookes, "Feature-domain super-resolution framework for gabor-based face and iris recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012.

[10] H.C. Burger, C.J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012.

[11] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Advances in Neural Information Processing Systems 21*. Curran Associates, Inc., 2009.

[12] Zhen Cui, Hong Chang, Shiguang Shan, Bineng Zhong, and Xilin Chen, "Deep network cascade for image super-resolution," in *Computer Vision ECCV 2014*, vol. 8693 of *Lecture Notes in Computer Science*. Springer International Publishing, 2014.

[13] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," *CoRR*, vol. abs/1511.04491, 2015.

[14] J. Johnson, A. Alahi, and Fei-Fei Li, "Perceptual losses for real-time style transfer and super-resolution," *CoRR*, vol. abs/1603.08155, 2016.

[15] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016.

[16] W. Shi, J.Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," *CoRR*, vol. abs/1609.05158, 2016.

[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, Feb 2016.

[18] CASIA Iris Image Database, "http://biometrics.idealtest.org/," .

[19] E. Ribeiro, A. Uhl, G. Wimmer, and M. Häfner, "Exploring deep learning and transfer learning for colonic polyp classification," *Computational and Mathematical Methods in Medicine*, pp. 1–16, 2016.

[20] Alex K., I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., 2012.

[21] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, Aug 2013.

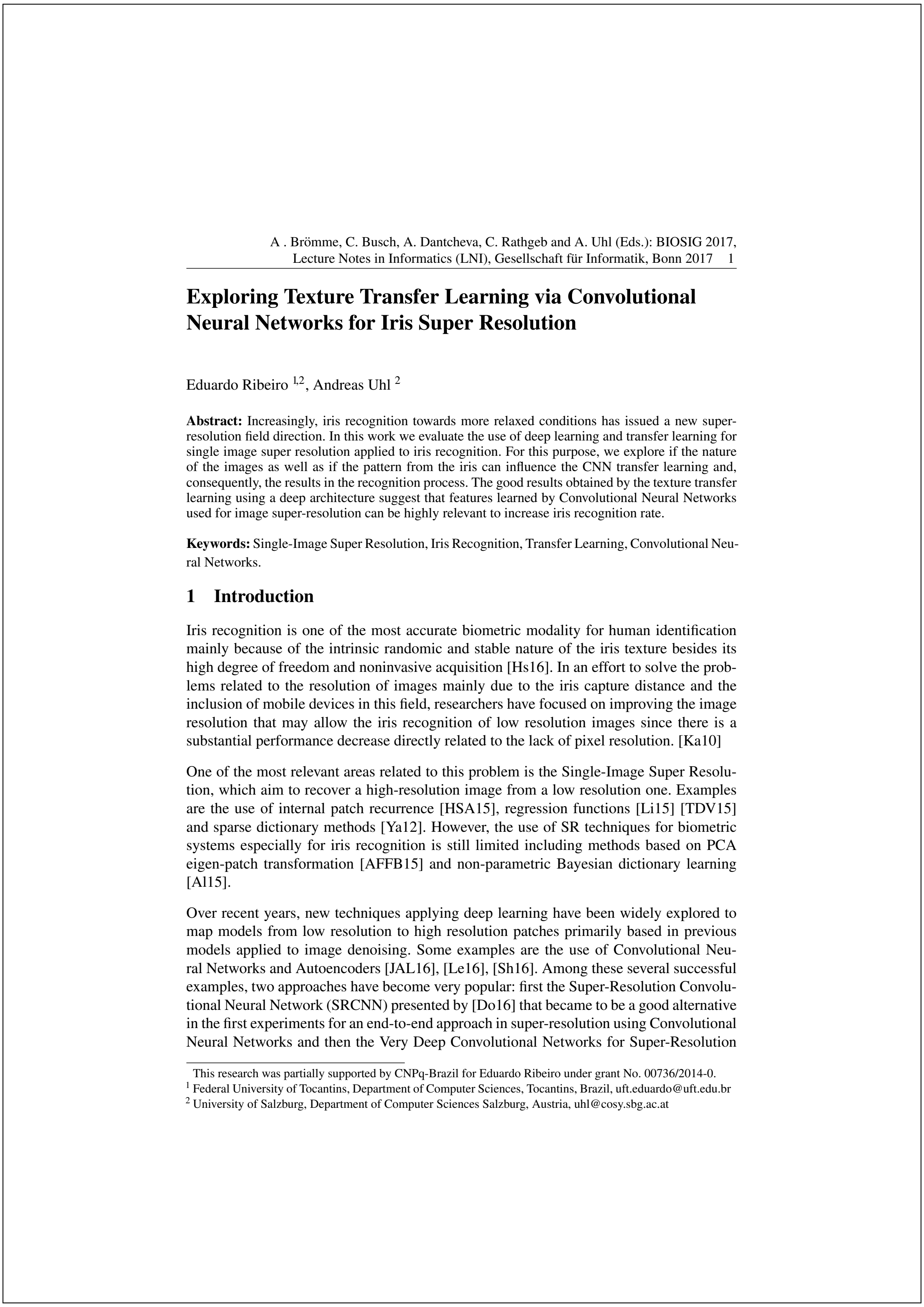
[22] H. Hofbauer and A. Uhl, "Identifying deficits of visual security metrics for images," *Signal Processing: Image Communication*, vol. 46, 2016.

[23] J. Daugman, "How iris recognition works," *IEEE Trans. Cir. and Sys. for Video Technol.*, vol. 14, no. 1, Jan. 2004.

[24] L. Masek, "Recognition of human iris patterns for biometric identification," Tech. Rep., The University of Western Australia, 2003.

[25] F. Alonso-Fernandez, P. Tome-Gonzalez, V. Ruiz-Albacete, and J. Ortega-Garcia, "Iris recognition based on sift features," in *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)*, Sept 2009.

[26] H. Hofbauer and A. Uhl, "Calculating a boundary for the significance from the equal-error rate," in *2016 International Conference on Biometrics (ICB)*, June 2016, pp. 1–4.



A . Brömme, C. Busch, A. Dantcheva, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2017,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2017 1

Exploring Texture Transfer Learning via Convolutional Neural Networks for Iris Super Resolution

Eduardo Ribeiro [1,2], Andreas Uhl [2]

Abstract: Increasingly, iris recognition towards more relaxed conditions has issued a new super-resolution field direction. In this work we evaluate the use of deep learning and transfer learning for single image super resolution applied to iris recognition. For this purpose, we explore if the nature of the images as well as if the pattern from the iris can influence the CNN transfer learning and, consequently, the results in the recognition process. The good results obtained by the texture transfer learning using a deep architecture suggest that features learned by Convolutional Neural Networks used for image super-resolution can be highly relevant to increase iris recognition rate.

Keywords: Single-Image Super Resolution, Iris Recognition, Transfer Learning, Convolutional Neural Networks.

1 Introduction

Iris recognition is one of the most accurate biometric modality for human identification mainly because of the intrinsic randomic and stable nature of the iris texture besides its high degree of freedom and noninvasive acquisition [Hs16]. In an effort to solve the problems related to the resolution of images mainly due to the iris capture distance and the inclusion of mobile devices in this field, researchers have focused on improving the image resolution that may allow the iris recognition of low resolution images since there is a substantial performance decrease directly related to the lack of pixel resolution. [Ka10]

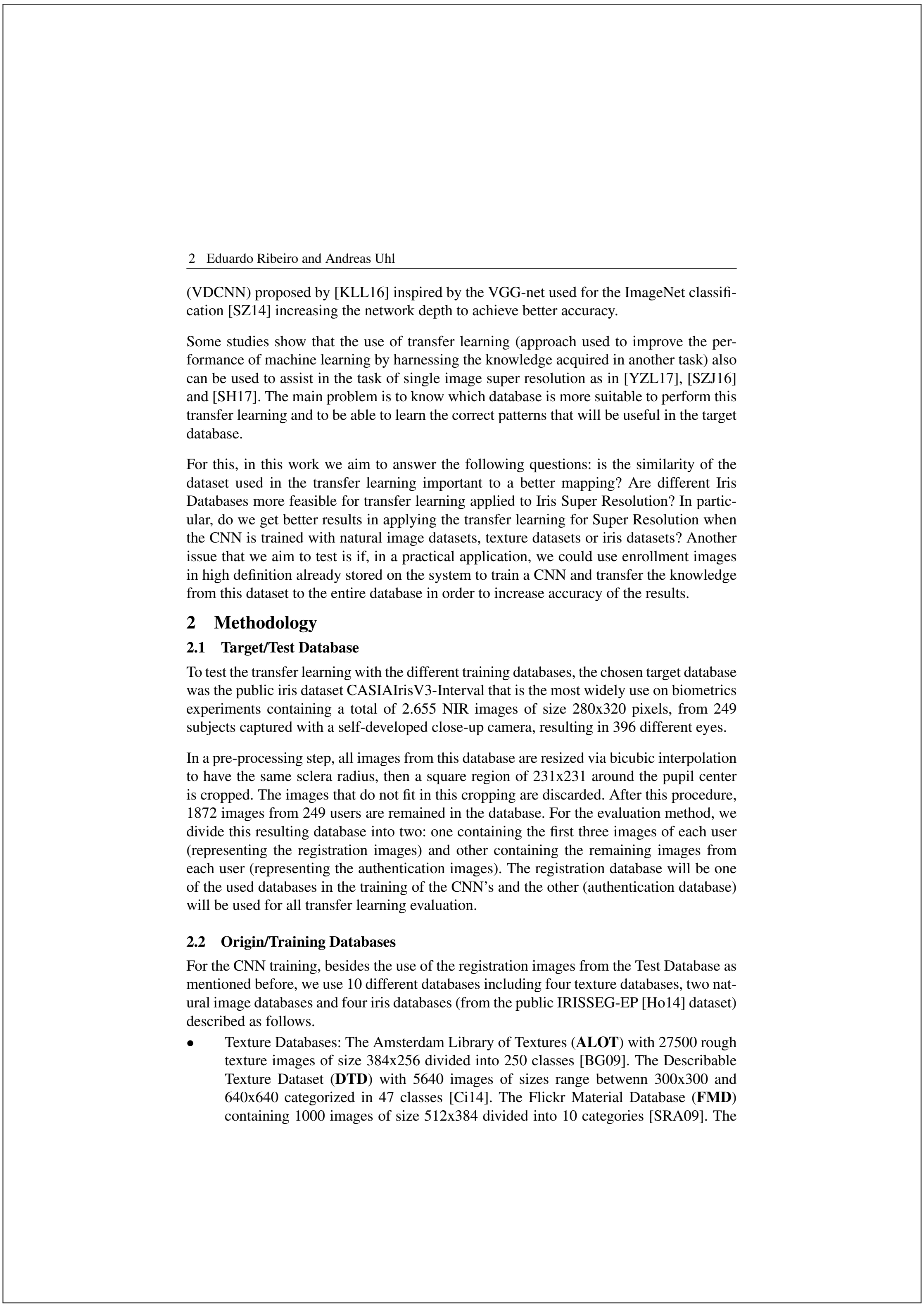
One of the most relevant areas related to this problem is the Single-Image Super Resolution, which aim to recover a high-resolution image from a low resolution one. Examples are the use of internal patch recurrence [HSA15], regression functions [Li15] [TDV15] and sparse dictionary methods [Ya12]. However, the use of SR techniques for biometric systems especially for iris recognition is still limited including methods based on PCA eigen-patch transformation [AFFB15] and non-parametric Bayesian dictionary learning [Al15].

Over recent years, new techniques applying deep learning have been widely explored to map models from low resolution to high resolution patches primarily based in previous models applied to image denoising. Some examples are the use of Convolutional Neural Networks and Autoencoders [JAL16], [Le16], [Sh16]. Among these several successful examples, two approaches have become very popular: first the Super-Resolution Convolutional Neural Network (SRCNN) presented by [Do16] that became to be a good alternative in the first experiments for an end-to-end approach in super-resolution using Convolutional Neural Networks and then the Very Deep Convolutional Networks for Super-Resolution

This research was partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.

[1] Federal University of Tocantins, Department of Computer Sciences, Tocantins, Brazil, uft.eduardo@uft.edu.br

[2] University of Salzburg, Department of Computer Sciences Salzburg, Austria, uhl@cosy.sbg.ac.at



(VDCNN) proposed by [KLL16] inspired by the VGG-net used for the ImageNet classification [SZ14] increasing the network depth to achieve better accuracy.

Some studies show that the use of transfer learning (approach used to improve the performance of machine learning by harnessing the knowledge acquired in another task) also can be used to assist in the task of single image super resolution as in [YZL17], [SZJ16] and [SH17]. The main problem is to know which database is more suitable to perform this transfer learning and to be able to learn the correct patterns that will be useful in the target database.

For this, in this work we aim to answer the following questions: is the similarity of the dataset used in the transfer learning important to a better mapping? Are different Iris Databases more feasible for transfer learning applied to Iris Super Resolution? In particular, do we get better results in applying the transfer learning for Super Resolution when the CNN is trained with natural image datasets, texture datasets or iris datasets? Another issue that we aim to test is if, in a practical application, we could use enrollment images in high definition already stored on the system to train a CNN and transfer the knowledge from this dataset to the entire database in order to increase accuracy of the results.

2 Methodology

2.1 Target/Test Database

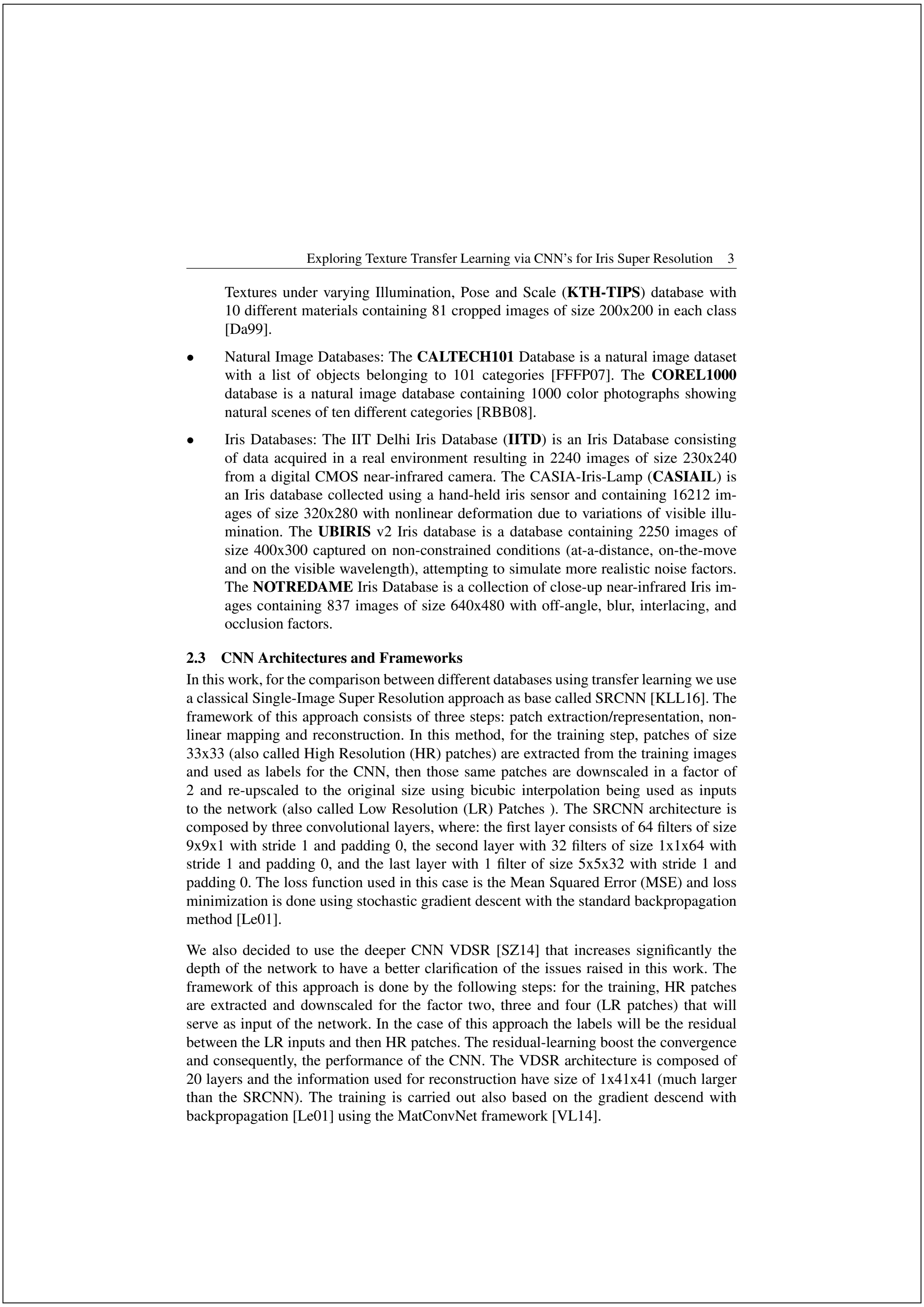
To test the transfer learning with the different training databases, the chosen target database was the public iris dataset CASIAIrisV3-Interval that is the most widely use on biometrics experiments containing a total of 2.655 NIR images of size 280x320 pixels, from 249 subjects captured with a self-developed close-up camera, resulting in 396 different eyes.

In a pre-processing step, all images from this database are resized via bicubic interpolation to have the same sclera radius, then a square region of 231x231 around the pupil center is cropped. The images that do not fit in this cropping are discarded. After this procedure, 1872 images from 249 users are remained in the database. For the evaluation method, we divide this resulting database into two: one containing the first three images of each user (representing the registration images) and other containing the remaining images from each user (representing the authentication images). The registration database will be one of the used databases in the training of the CNN's and the other (authentication database) will be used for all transfer learning evaluation.

2.2 Origin/Training Databases

For the CNN training, besides the use of the registration images from the Test Database as mentioned before, we use 10 different databases including four texture databases, two natural image databases and four iris databases (from the public IRISSEG-EP [Ho14] dataset) described as follows.

- Texture Databases: The Amsterdam Library of Textures (**ALOT**) with 27500 rough texture images of size 384x256 divided into 250 classes [BG09]. The Describable Texture Dataset (**DTD**) with 5640 images of sizes range betwenn 300x300 and 640x640 categorized in 47 classes [Ci14]. The Flickr Material Database (**FMD**) containing 1000 images of size 512x384 divided into 10 categories [SRA09]. The



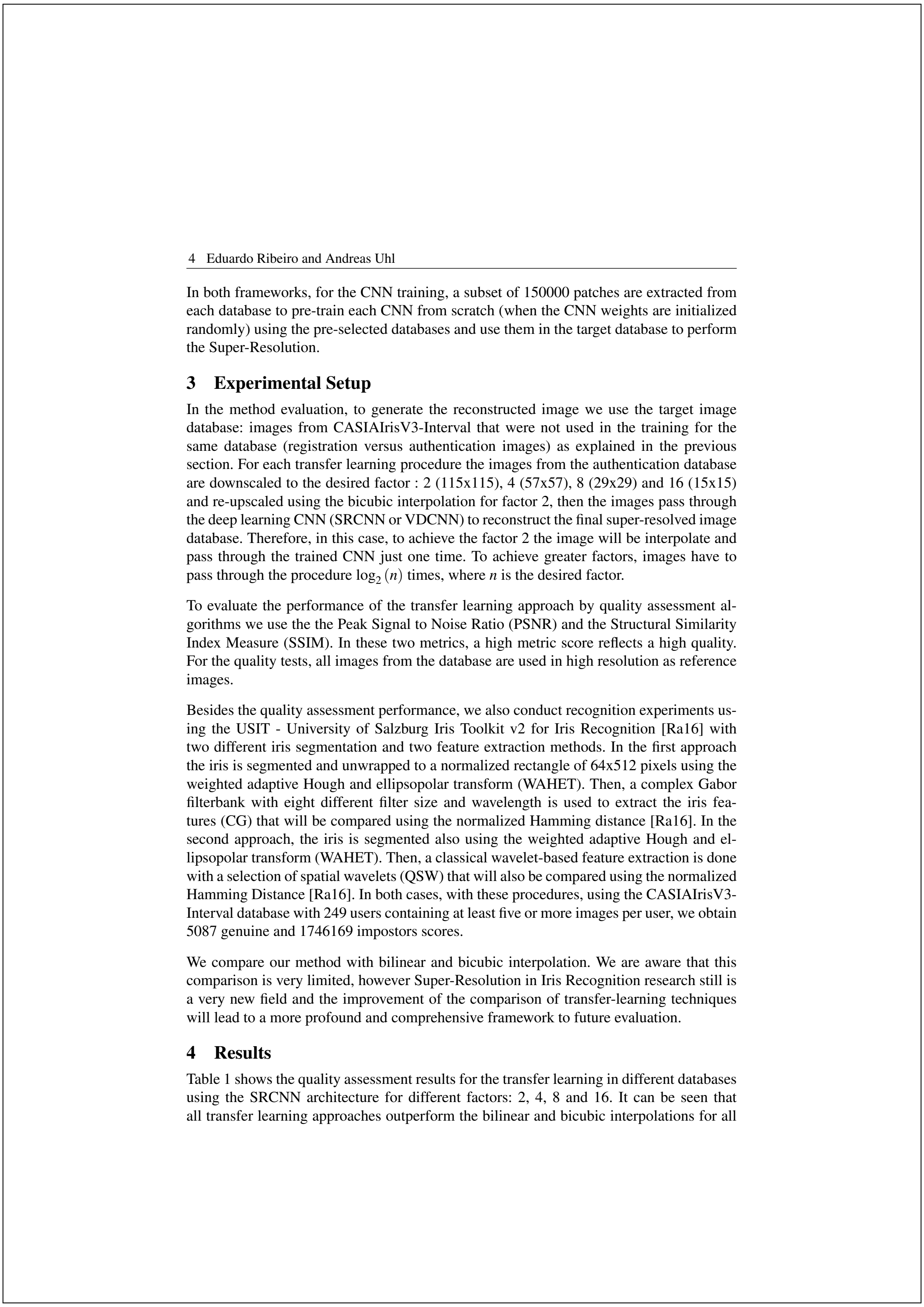
Textures under varying Illumination, Pose and Scale (**KTH-TIPS**) database with 10 different materials containing 81 cropped images of size 200x200 in each class [Da99].

- Natural Image Databases: The **CALTECH101** Database is a natural image dataset with a list of objects belonging to 101 categories [FFFP07]. The **COREL1000** database is a natural image database containing 1000 color photographs showing natural scenes of ten different categories [RBB08].
- Iris Databases: The IIT Delhi Iris Database (**IITD**) is an Iris Database consisting of data acquired in a real environment resulting in 2240 images of size 230x240 from a digital CMOS near-infrared camera. The CASIA-Iris-Lamp (**CASIAIL**) is an Iris database collected using a hand-held iris sensor and containing 16212 images of size 320x280 with nonlinear deformation due to variations of visible illumination. The **UBIRIS** v2 Iris database is a database containing 2250 images of size 400x300 captured on non-constrained conditions (at-a-distance, on-the-move and on the visible wavelength), attempting to simulate more realistic noise factors. The **NOTREDAME** Iris Database is a collection of close-up near-infrared Iris images containing 837 images of size 640x480 with off-angle, blur, interlacing, and occlusion factors.

2.3 CNN Architectures and Frameworks

In this work, for the comparison between different databases using transfer learning we use a classical Single-Image Super Resolution approach as base called SRCNN [KLL16]. The framework of this approach consists of three steps: patch extraction/representation, nonlinear mapping and reconstruction. In this method, for the training step, patches of size 33x33 (also called High Resolution (HR) patches) are extracted from the training images and used as labels for the CNN, then those same patches are downscaled in a factor of 2 and re-upscaled to the original size using bicubic interpolation being used as inputs to the network (also called Low Resolution (LR) Patches). The SRCNN architecture is composed by three convolutional layers, where: the first layer consists of 64 filters of size 9x9x1 with stride 1 and padding 0, the second layer with 32 filters of size 1x1x64 with stride 1 and padding 0, and the last layer with 1 filter of size 5x5x32 with stride 1 and padding 0. The loss function used in this case is the Mean Squared Error (MSE) and loss minimization is done using stochastic gradient descent with the standard backpropagation method [Le01].

We also decided to use the deeper CNN VDSR [SZ14] that increases significantly the depth of the network to have a better clarification of the issues raised in this work. The framework of this approach is done by the following steps: for the training, HR patches are extracted and downscaled for the factor two, three and four (LR patches) that will serve as input of the network. In the case of this approach the labels will be the residual between the LR inputs and then HR patches. The residual-learning boost the convergence and consequently, the performance of the CNN. The VDSR architecture is composed of 20 layers and the information used for reconstruction have size of 1x41x41 (much larger than the SRCNN). The training is carried out also based on the gradient descend with backpropagation [Le01] using the MatConvNet framework [VL14].



In both frameworks, for the CNN training, a subset of 150000 patches are extracted from each database to pre-train each CNN from scratch (when the CNN weights are initialized randomly) using the pre-selected databases and use them in the target database to perform the Super-Resolution.

3 Experimental Setup

In the method evaluation, to generate the reconstructed image we use the target image database: images from CASIAIrisV3-Interval that were not used in the training for the same database (registration versus authentication images) as explained in the previous section. For each transfer learning procedure the images from the authentication database are downscaled to the desired factor : 2 (115x115), 4 (57x57), 8 (29x29) and 16 (15x15) and re-upscaled using the bicubic interpolation for factor 2, then the images pass through the deep learning CNN (SRCNN or VDCNN) to reconstruct the final super-resolved image database. Therefore, in this case, to achieve the factor 2 the image will be interpolate and pass through the trained CNN just one time. To achieve greater factors, images have to pass through the procedure $\log_2(n)$ times, where n is the desired factor.

To evaluate the performance of the transfer learning approach by quality assessment algorithms we use the the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM). In these two metrics, a high metric score reflects a high quality. For the quality tests, all images from the database are used in high resolution as reference images.

Besides the quality assessment performance, we also conduct recognition experiments using the USIT - University of Salzburg Iris Toolkit v2 for Iris Recognition [Ra16] with two different iris segmentation and two feature extraction methods. In the first approach the iris is segmented and unwrapped to a normalized rectangle of 64x512 pixels using the weighted adaptive Hough and ellipsopolar transform (WAHET). Then, a complex Gabor filterbank with eight different filter size and wavelength is used to extract the iris features (CG) that will be compared using the normalized Hamming distance [Ra16]. In the second approach, the iris is segmented also using the weighted adaptive Hough and ellipsopolar transform (WAHET). Then, a classical wavelet-based feature extraction is done with a selection of spatial wavelets (QSW) that will also be compared using the normalized Hamming Distance [Ra16]. In both cases, with these procedures, using the CASIAIrisV3-Interval database with 249 users containing at least five or more images per user, we obtain 5087 genuine and 1746169 impostors scores.

We compare our method with bilinear and bicubic interpolation. We are aware that this comparison is very limited, however Super-Resolution in Iris Recognition research still is a very new field and the improvement of the comparison of transfer-learning techniques will lead to a more profound and comprehensive framework to future evaluation.

4 Results

Table 1 shows the quality assessment results for the transfer learning in different databases using the SRCNN architecture for different factors: 2, 4, 8 and 16. It can be seen that all transfer learning approaches outperform the bilinear and bicubic interpolations for all factors including bigger factors showing the resilience of the deep-learning method when image resolution decreases.

It also can be noticed that the transfer learning using texture databases perform better in terms of similarity to the original HR database than the transfer learning using iris databases. However, the results from the Casia Interval transfer learning present good results compared to the other iris databases. The best result in this case is when the CNN is trained with the DTD database especially for higher factors and the Caltech101 database for smaller factors.

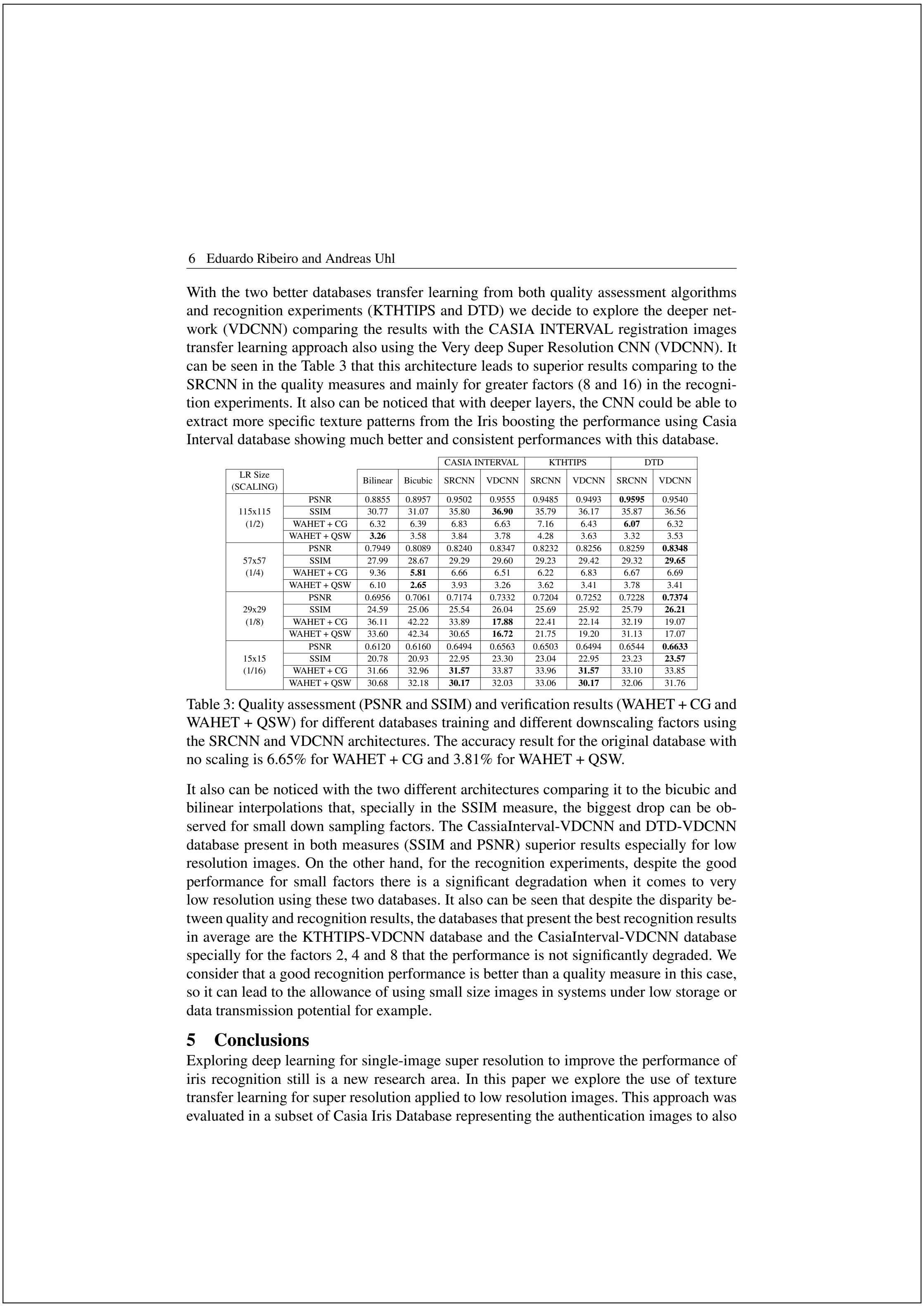
				Texture Databases				Natural Image Databases			Iris Databases			
LR Size (SCALING)		Bilinear	Bicubic	ALOT	DTD	FMD	KTH TIPS	CALTECH 101	COREL 1000	IITD	CASIAIL	UBIRIS	NOTRE DAME	CASIA INTERVAL
115X115	PSNR	0.8855	0.8957	0.9481	**0.9595**	0.9509	0.9485	0.9492	0.9491	0.9483	0.9422	0.9414	0.9495	0.9502
(1/2)	SSIM	30.77	31.07	35.17	**35.87**	35.82	35.79	35.85	35.34	35.43	35.12	34.67	35.70	35.80
57X57	PSNR	0.7949	0.8089	0.8243	**0.8259**	0.8245	0.8232	0.8250	0.8255	0.8214	0.8129	0.8131	0.8216	0.8240
(1/4)	SSIM	27.99	28.67	29.20	**29.32**	29.29	29.23	29.24	28.97	29.18	29.01	28.86	29.24	29.29
29X29	PSNR	0.6956	0.7061	0.7198	0.7228	0.7157	0.7204	**0.7251**	0.7236	0.7127	0.7064	0.7085	0.7128	0.7174
(1/8)	SSIM	24.59	25.06	25.61	25.79	25.57	25.69	**25.80**	25.50	25.44	25.15	25.12	25.44	25.54
15X15	PSNR	0.6120	0.6160	0.6510	0.6544	0.6471	0.6503	**0.6557**	0.6553	0.6439	0.6406	0.6430	0.6447	0.6494
(1/16)	SSIM	20.78	20.93	23.09	**23.23**	23.07	23.04	23.21	23.05	23.01	22.67	22.69	22.97	22.95

Table 1: Results of quality assessment algorithms for different databases training with different downscaling factors (average values on the test dataset) using the SRCNN architecture comparing to the Bilinear and Bicubic approach.

In the iris recognition verification, it can be seen from Table 2 that the results present different best results among the databases as well as presents mismatch results between the quality experimental results from table 2 and the verification results. In the case of EER the best result for the factor 2 (115X115) is when the DTD database is used (accuracy of 6.07%) in accordance with the quality assessment results (PSNR and SSIM) presenting even better results than the original database (6.657% of accuracy). Nonetheless, for the factor 4 (57x57), the best result is from the bicubic interpolation even better than all the results from the factor 2 and from the original HR database results. Among the training databases, for the recognition experiments, the more consistently beneficial for the transfer learning is the KTHTIPS database especially for the factors 4 and 8. Using the enrollment images from the same target database (Casia Interval) does not lead to good recognition performances, which means that the CNN poorly memorize the patterns from the users focusing more in general patterns, mainly because the depth of the network that does not allow a high feature discrimination.

				Texture Databases				Natural Image Database			Iris Databases			
LR Size (SCALING)		Bilinear	Bicubic	ALOT	DTD	FMD	KTH TIPS	CALTECH 101	COREL 1000	IITD	CASIAIL	UBIRIS	NOTRE DAME	CASIA INTERVAL
115X115	WAHET + CG	6.32	6.39	6.50	**6.07**	6.66	7.16	6.74	6.39	6.68	6.61	6.37	6.64	6.83
(1/2)	WAHET+QSW	**3.26**	3.58	3.58	3.32	3.81	4.28	4.02	3.53	3.89	3.92	3.42	4.02	3.84
57X57	WAHET + CG	9.36	**5.81**	7.19	6.67	6.88	6.22	6.83	6.51	7.90	7.84	8.41	7.59	6.66
(1/4)	WAHET+QSW	6.10	**2.65**	4.58	3.78	4.09	3.62	3.95	3.74	5.11	5.22	5.75	4.66	3.93
29X29	WAHET + CG	36.11	42.22	32.97	32.19	36.86	**22.41**	32.88	33.81	38.19	39.88	39.75	39.15	33.89
(1/8)	WAHET+QSW	33.60	42.34	30.62	31.13	34.89	**21.75**	32.10	33.26	36.50	38.53	37.33	37.04	30.65
15X15	WAHET + CG	31.66	32.96	33.95	33.10	33.03	33.96	33.02	34.68	32.73	**28.52**	29.62	31.50	31.57
(1/16)	WAHET+QSW	30.68	32.18	32.57	32.06	31.60	33.06	31.66	33.18	31.84	**27.60**	28.02	31.25	30.17

Table 2: Verification results (EER) for different databases training for different downscaling factors using the SRCNN architecture comparing to the Bilinear and Bicubic approach. The accuracy result for the original database with no scaling is 6.65% for WAHET + CG and and 3.81% for WAHET + QSW.



With the two better databases transfer learning from both quality assessment algorithms and recognition experiments (KTHTIPS and DTD) we decide to explore the deeper network (VDCNN) comparing the results with the CASIA INTERVAL registration images transfer learning approach also using the Very deep Super Resolution CNN (VDCNN). It can be seen in the Table 3 that this architecture leads to superior results comparing to the SRCNN in the quality measures and mainly for greater factors (8 and 16) in the recognition experiments. It also can be noticed that with deeper layers, the CNN could be able to extract more specific texture patterns from the Iris boosting the performance using Casia Interval database showing much better and consistent performances with this database.

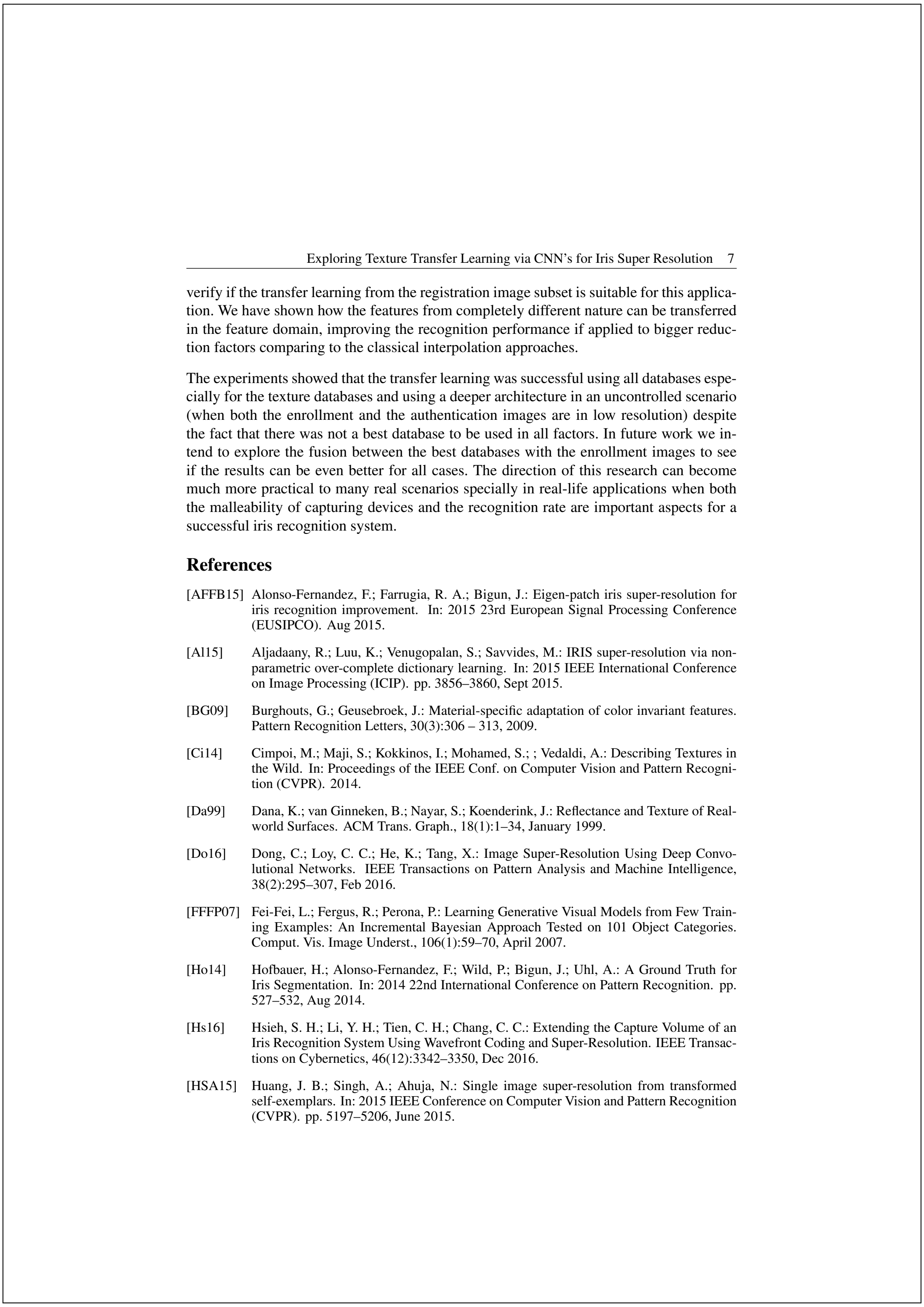
				CASIA INTERVAL		KTHTIPS		DTD	
LR Size (SCALING)		Bilinear	Bicubic	SRCNN	VDCNN	SRCNN	VDCNN	SRCNN	VDCNN
115x115 (1/2)	PSNR	0.8855	0.8957	0.9502	0.9555	0.9485	0.9493	**0.9595**	0.9540
	SSIM	30.77	31.07	35.80	**36.90**	35.79	36.17	35.87	36.56
	WAHET + CG	6.32	6.39	6.83	6.63	7.16	6.43	**6.07**	6.32
	WAHET + QSW	**3.26**	3.58	3.84	3.78	4.28	3.63	3.32	3.53
57x57 (1/4)	PSNR	0.7949	0.8089	0.8240	0.8347	0.8232	0.8256	0.8259	**0.8348**
	SSIM	27.99	28.67	29.29	29.60	29.23	29.42	29.32	**29.65**
	WAHET + CG	9.36	**5.81**	6.66	6.51	6.22	6.83	6.67	6.69
	WAHET + QSW	6.10	**2.65**	3.93	3.26	3.62	3.41	3.78	3.41
29x29 (1/8)	PSNR	0.6956	0.7061	0.7174	0.7332	0.7204	0.7252	0.7228	**0.7374**
	SSIM	24.59	25.06	25.54	26.04	25.69	25.92	25.79	**26.21**
	WAHET + CG	36.11	42.22	33.89	**17.88**	22.41	22.14	32.19	19.07
	WAHET + QSW	33.60	42.34	30.65	**16.72**	21.75	19.20	31.13	17.07
15x15 (1/16)	PSNR	0.6120	0.6160	0.6494	0.6563	0.6503	0.6494	0.6544	**0.6633**
	SSIM	20.78	20.93	22.95	23.30	23.04	22.95	23.23	**23.57**
	WAHET + CG	31.66	32.96	**31.57**	33.87	33.96	**31.57**	33.10	33.85
	WAHET + QSW	30.68	32.18	**30.17**	32.03	33.06	**30.17**	32.06	31.76

Table 3: Quality assessment (PSNR and SSIM) and verification results (WAHET + CG and WAHET + QSW) for different databases training and different downscaling factors using the SRCNN and VDCNN architectures. The accuracy result for the original database with no scaling is 6.65% for WAHET + CG and 3.81% for WAHET + QSW.

It also can be noticed with the two different architectures comparing it to the bicubic and bilinear interpolations that, specially in the SSIM measure, the biggest drop can be observed for small down sampling factors. The CassiaInterval-VDCNN and DTD-VDCNN database present in both measures (SSIM and PSNR) superior results especially for low resolution images. On the other hand, for the recognition experiments, despite the good performance for small factors there is a significant degradation when it comes to very low resolution using these two databases. It also can be seen that despite the disparity between quality and recognition results, the databases that present the best recognition results in average are the KTHTIPS-VDCNN database and the CasiaInterval-VDCNN database specially for the factors 2, 4 and 8 that the performance is not significantly degraded. We consider that a good recognition performance is better than a quality measure in this case, so it can lead to the allowance of using small size images in systems under low storage or data transmission potential for example.

5 Conclusions

Exploring deep learning for single-image super resolution to improve the performance of iris recognition still is a new research area. In this paper we explore the use of texture transfer learning for super resolution applied to low resolution images. This approach was evaluated in a subset of Casia Iris Database representing the authentication images to also



verify if the transfer learning from the registration image subset is suitable for this application. We have shown how the features from completely different nature can be transferred in the feature domain, improving the recognition performance if applied to bigger reduction factors comparing to the classical interpolation approaches.

The experiments showed that the transfer learning was successful using all databases especially for the texture databases and using a deeper architecture in an uncontrolled scenario (when both the enrollment and the authentication images are in low resolution) despite the fact that there was not a best database to be used in all factors. In future work we intend to explore the fusion between the best databases with the enrollment images to see if the results can be even better for all cases. The direction of this research can become much more practical to many real scenarios specially in real-life applications when both the malleability of capturing devices and the recognition rate are important aspects for a successful iris recognition system.

References

[AFFB15] Alonso-Fernandez, F.; Farrugia, R. A.; Bigun, J.: Eigen-patch iris super-resolution for iris recognition improvement. In: 2015 23rd European Signal Processing Conference (EUSIPCO). Aug 2015.

[Al15] Aljadaany, R.; Luu, K.; Venugopalan, S.; Savvides, M.: IRIS super-resolution via nonparametric over-complete dictionary learning. In: 2015 IEEE International Conference on Image Processing (ICIP). pp. 3856–3860, Sept 2015.

[BG09] Burghouts, G.; Geusebroek, J.: Material-specific adaptation of color invariant features. Pattern Recognition Letters, 30(3):306 – 313, 2009.

[Ci14] Cimpoi, M.; Maji, S.; Kokkinos, I.; Mohamed, S.; ; Vedaldi, A.: Describing Textures in the Wild. In: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2014.

[Da99] Dana, K.; van Ginneken, B.; Nayar, S.; Koenderink, J.: Reflectance and Texture of Real-world Surfaces. ACM Trans. Graph., 18(1):1–34, January 1999.

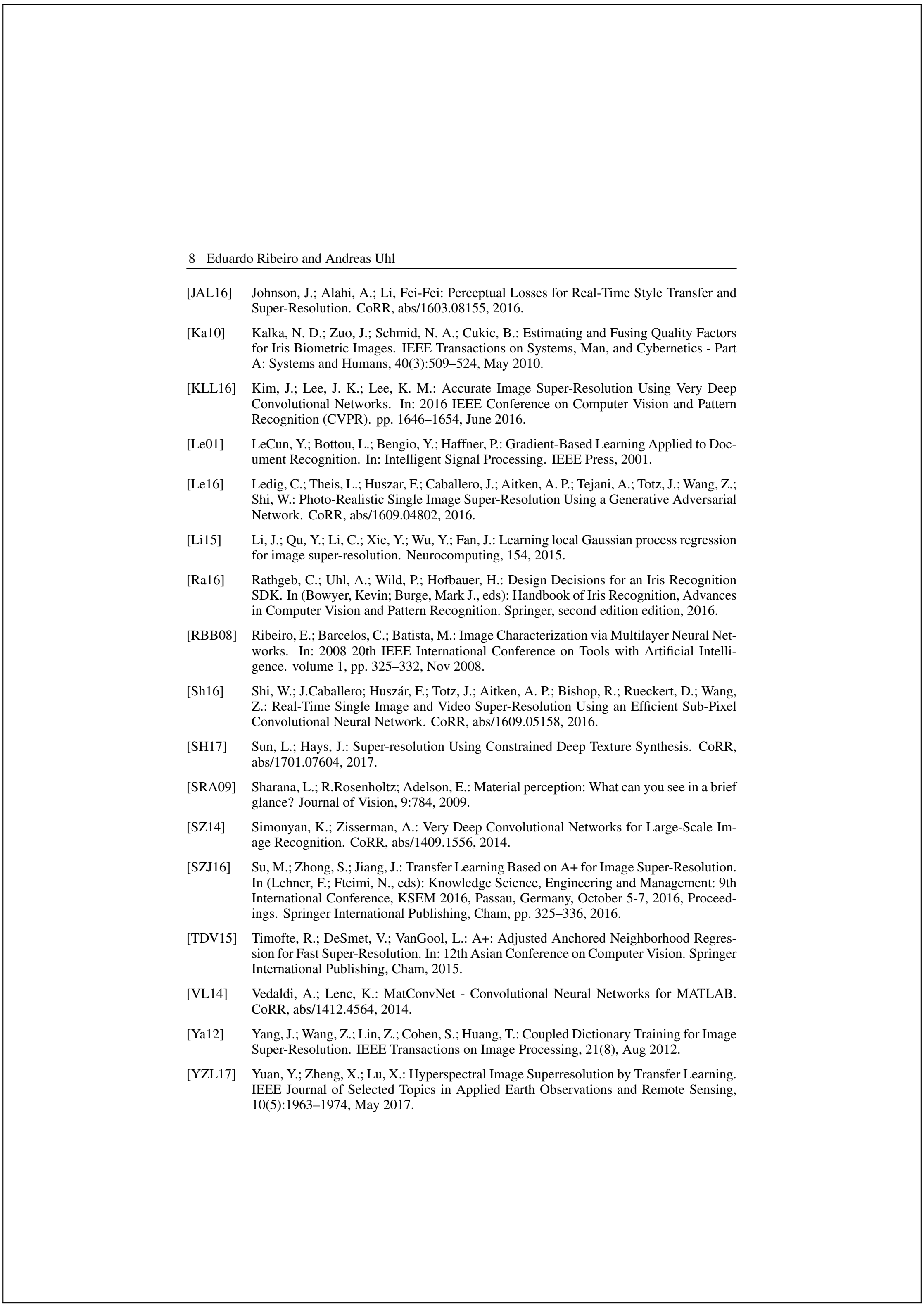
[Do16] Dong, C.; Loy, C. C.; He, K.; Tang, X.: Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):295–307, Feb 2016.

[FFFP07] Fei-Fei, L.; Fergus, R.; Perona, P.: Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. Comput. Vis. Image Underst., 106(1):59–70, April 2007.

[Ho14] Hofbauer, H.; Alonso-Fernandez, F.; Wild, P.; Bigun, J.; Uhl, A.: A Ground Truth for Iris Segmentation. In: 2014 22nd International Conference on Pattern Recognition. pp. 527–532, Aug 2014.

[Hs16] Hsieh, S. H.; Li, Y. H.; Tien, C. H.; Chang, C. C.: Extending the Capture Volume of an Iris Recognition System Using Wavefront Coding and Super-Resolution. IEEE Transactions on Cybernetics, 46(12):3342–3350, Dec 2016.

[HSA15] Huang, J. B.; Singh, A.; Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5197–5206, June 2015.



[JAL16] Johnson, J.; Alahi, A.; Li, Fei-Fei: Perceptual Losses for Real-Time Style Transfer and Super-Resolution. CoRR, abs/1603.08155, 2016.

[Ka10] Kalka, N. D.; Zuo, J.; Schmid, N. A.; Cukic, B.: Estimating and Fusing Quality Factors for Iris Biometric Images. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 40(3):509–524, May 2010.

[KLL16] Kim, J.; Lee, J. K.; Lee, K. M.: Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1646–1654, June 2016.

[Le01] LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P.: Gradient-Based Learning Applied to Document Recognition. In: Intelligent Signal Processing. IEEE Press, 2001.

[Le16] Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Aitken, A. P.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. CoRR, abs/1609.04802, 2016.

[Li15] Li, J.; Qu, Y.; Li, C.; Xie, Y.; Wu, Y.; Fan, J.: Learning local Gaussian process regression for image super-resolution. Neurocomputing, 154, 2015.

[Ra16] Rathgeb, C.; Uhl, A.; Wild, P.; Hofbauer, H.: Design Decisions for an Iris Recognition SDK. In (Bowyer, Kevin; Burge, Mark J., eds): Handbook of Iris Recognition, Advances in Computer Vision and Pattern Recognition. Springer, second edition edition, 2016.

[RBB08] Ribeiro, E.; Barcelos, C.; Batista, M.: Image Characterization via Multilayer Neural Networks. In: 2008 20th IEEE International Conference on Tools with Artificial Intelligence. volume 1, pp. 325–332, Nov 2008.

[Sh16] Shi, W.; J.Caballero; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; Wang, Z.: Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. CoRR, abs/1609.05158, 2016.

[SH17] Sun, L.; Hays, J.: Super-resolution Using Constrained Deep Texture Synthesis. CoRR, abs/1701.07604, 2017.

[SRA09] Sharana, L.; R.Rosenholtz; Adelson, E.: Material perception: What can you see in a brief glance? Journal of Vision, 9:784, 2009.

[SZ14] Simonyan, K.; Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556, 2014.

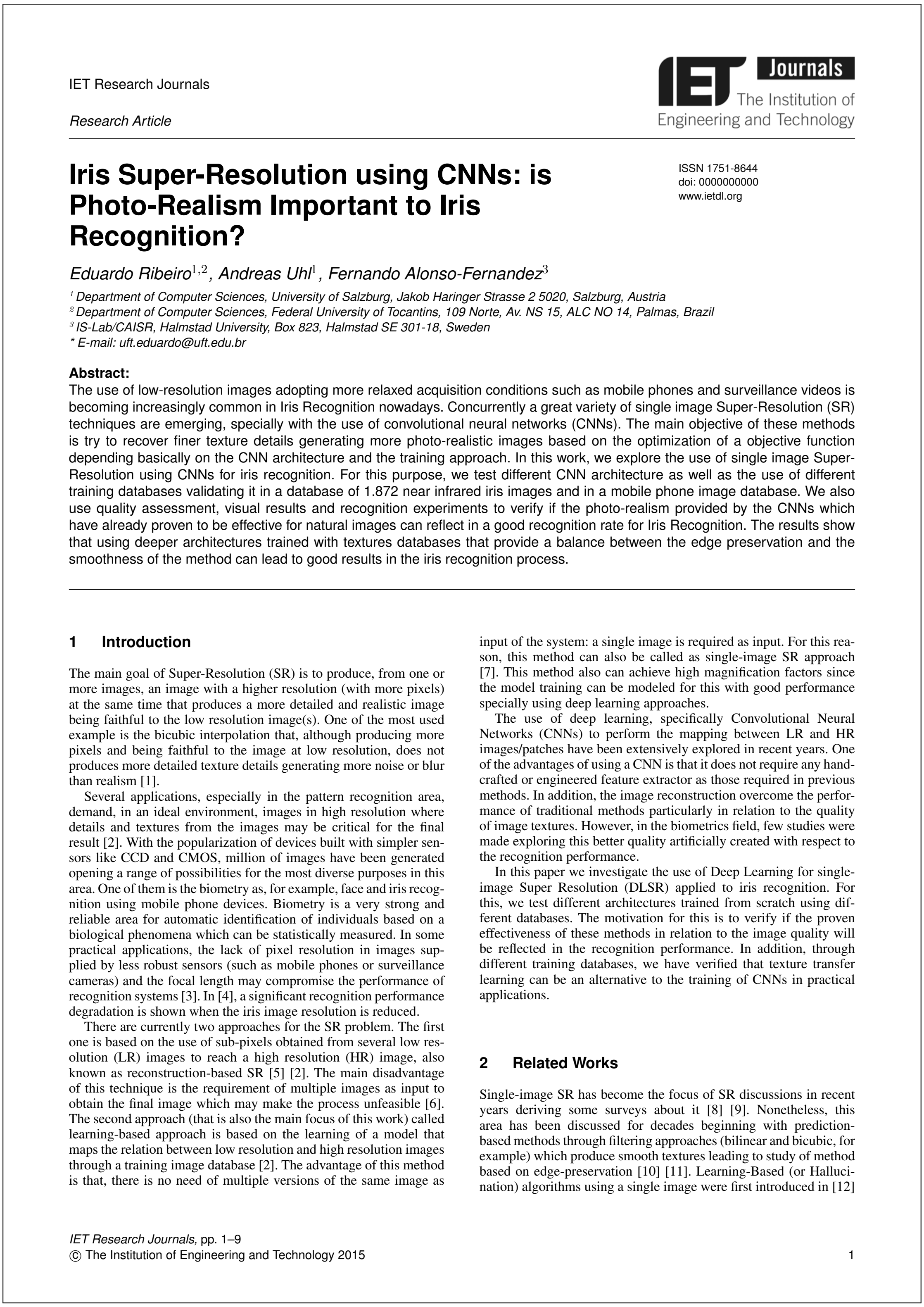
[SZJ16] Su, M.; Zhong, S.; Jiang, J.: Transfer Learning Based on A+ for Image Super-Resolution. In (Lehner, F.; Fteimi, N., eds): Knowledge Science, Engineering and Management: 9th International Conference, KSEM 2016, Passau, Germany, October 5-7, 2016, Proceedings. Springer International Publishing, Cham, pp. 325–336, 2016.

[TDV15] Timofte, R.; DeSmet, V.; VanGool, L.: A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution. In: 12th Asian Conference on Computer Vision. Springer International Publishing, Cham, 2015.

[VL14] Vedaldi, A.; Lenc, K.: MatConvNet - Convolutional Neural Networks for MATLAB. CoRR, abs/1412.4564, 2014.

[Ya12] Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; Huang, T.: Coupled Dictionary Training for Image Super-Resolution. IEEE Transactions on Image Processing, 21(8), Aug 2012.

[YZL17] Yuan, Y.; Zheng, X.; Lu, X.: Hyperspectral Image Superresolution by Transfer Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(5):1963–1974, May 2017.



IET Research Journals

Research Article

Journals
The Institution of
Engineering and Technology

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

Iris Super-Resolution using CNNs: is Photo-Realism Important to Iris Recognition?

Eduardo Ribeiro^{1,2}*, Andreas Uhl*^{1}*, Fernando Alonso-Fernandez*^{3}

[1] *Department of Computer Sciences, University of Salzburg, Jakob Haringer Strasse 2 5020, Salzburg, Austria*
[2] *Department of Computer Sciences, Federal University of Tocantins, 109 Norte, Av. NS 15, ALC NO 14, Palmas, Brazil*
[3] *IS-Lab/CAISR, Halmstad University, Box 823, Halmstad SE 301-18, Sweden*
** E-mail: uft.eduardo@uft.edu.br*

Abstract:
The use of low-resolution images adopting more relaxed acquisition conditions such as mobile phones and surveillance videos is becoming increasingly common in Iris Recognition nowadays. Concurrently a great variety of single image Super-Resolution (SR) techniques are emerging, specially with the use of convolutional neural networks (CNNs). The main objective of these methods is try to recover finer texture details generating more photo-realistic images based on the optimization of a objective function depending basically on the CNN architecture and the training approach. In this work, we explore the use of single image Super-Resolution using CNNs for iris recognition. For this purpose, we test different CNN architecture as well as the use of different training databases validating it in a database of 1.872 near infrared iris images and in a mobile phone image database. We also use quality assessment, visual results and recognition experiments to verify if the photo-realism provided by the CNNs which have already proven to be effective for natural images can reflect in a good recognition rate for Iris Recognition. The results show that using deeper architectures trained with textures databases that provide a balance between the edge preservation and the smoothness of the method can lead to good results in the iris recognition process.

1 Introduction

The main goal of Super-Resolution (SR) is to produce, from one or more images, an image with a higher resolution (with more pixels) at the same time that produces a more detailed and realistic image being faithful to the low resolution image(s). One of the most used example is the bicubic interpolation that, although producing more pixels and being faithful to the image at low resolution, does not produces more detailed texture details generating more noise or blur than realism [1].

Several applications, especially in the pattern recognition area, demand, in an ideal environment, images in high resolution where details and textures from the images may be critical for the final result [2]. With the popularization of devices built with simpler sensors like CCD and CMOS, million of images have been generated opening a range of possibilities for the most diverse purposes in this area. One of them is the biometry as, for example, face and iris recognition using mobile phone devices. Biometry is a very strong and reliable area for automatic identification of individuals based on a biological phenomena which can be statistically measured. In some practical applications, the lack of pixel resolution in images supplied by less robust sensors (such as mobile phones or surveillance cameras) and the focal length may compromise the performance of recognition systems [3]. In [4], a significant recognition performance degradation is shown when the iris image resolution is reduced.

There are currently two approaches for the SR problem. The first one is based on the use of sub-pixels obtained from several low resolution (LR) images to reach a high resolution (HR) image, also known as reconstruction-based SR [5] [2]. The main disadvantage of this technique is the requirement of multiple images as input to obtain the final image which may make the process unfeasible [6]. The second approach (that is also the main focus of this work) called learning-based approach is based on the learning of a model that maps the relation between low resolution and high resolution images through a training image database [2]. The advantage of this method is that, there is no need of multiple versions of the same image as input of the system: a single image is required as input. For this reason, this method can also be called as single-image SR approach [7]. This method also can achieve high magnification factors since the model training can be modeled for this with good performance specially using deep learning approaches.

The use of deep learning, specifically Convolutional Neural Networks (CNNs) to perform the mapping between LR and HR images/patches have been extensively explored in recent years. One of the advantages of using a CNN is that it does not require any hand-crafted or engineered feature extractor as those required in previous methods. In addition, the image reconstruction overcome the performance of traditional methods particularly in relation to the quality of image textures. However, in the biometrics field, few studies were made exploring this better quality artificially created with respect to the recognition performance.

In this paper we investigate the use of Deep Learning for single-image Super Resolution (DLSR) applied to iris recognition. For this, we test different architectures trained from scratch using different databases. The motivation for this is to verify if the proven effectiveness of these methods in relation to the image quality will be reflected in the recognition performance. In addition, through different training databases, we have verified that texture transfer learning can be an alternative to the training of CNNs in practical applications.

2 Related Works

Single-image SR has become the focus of SR discussions in recent years deriving some surveys about it [8] [9]. Nonetheless, this area has been discussed for decades beginning with prediction-based methods through filtering approaches (bilinear and bicubic, for example) which produce smooth textures leading to study of method based on edge-preservation [10] [11]. Learning-Based (or Hallucination) algorithms using a single image were first introduced in [12] where the mapping between the LR and HR image was learned by a Neural Network applied to fingerprint images.

With the popularization of Convolutional Neural Networks, several methods were proposed obtaining excellent results. Wang et. al. [13] showed that encoding a sparse representation particularly designed for SR can make the end-to-end mapping between the LR and HR image through a reduced model size. However, the most famous architecture of this end-to-end mapping is the SRCNN proposed by Dong et. al. [14] that used a bicubic interpolation to up-sample the input LR image using a trained three-layer deep fully CNN to reconstruct the HR image acting as a denoising tool. The most common concern of the work that followed was to find an architecture that minimizes the mean squared error (MSE) between the reconstructed HR image and the ground truth. Besides that, also reflecting the maximization of the peak signal-to -noise ratio (PSNR), one of the most used metrics to evaluate the quality of the result in the comparison of the proposed methods [15].

In [16] a deeper CNN architecture is presented inspired by VGG-net used for ImageNet Classisifcation [17] also called VDCNN. That work demonstrates that the use of the cascading of small filters many times in a deep network structure and the use of residual-learning can affect the accuracy of the SR method.

In [15] a SR generative adversarial network is proposed to try to recover finer texture details from LR images inferring photo-realistic natural images through a novel perceptual loss function using high-level maps from VGG network. The SRCNN, VDCNN and SRGAN architectures will be used in this work and will be detailed in the next sections.

Research about Super-Resolution in Biometrics (specially for Iris Recognition) has been increasing in the last years specially using reconstruction-based methods. For example, Kien et.al. [3] use the feature domain to super-resolve low resolution images relying only in the features incorporating domain specific information for iris models to constrain the estimation. In [18]. Nguyen et.al. introduces a signal-level fusion to integrate quality scores to reconstruction-based super-resolution process performing a quality weighted super-resolution for a low resolution video sequence of a less constrained iris at distance or on the move obtaining good results. However, in this case, as in [19] that performs the best frame selection, it is necessary many LR images to reconstruct the HR image which is one of the disadvantages of this kind of reconstruction-based methods.

In [20] an iris recognition algorithm based on PCA is presented by constructing coarse iris images with PCA coefficients and enhancing them using super-resolution. In [21] a reconstruction based SR is proposed for iris SR from LR video frames using an auto-regressive signature model between consecutive LR images to fill the sub pixels in the constructed image. In [22], two SR approaches are tested for iris recognition, one based on PCA Eigen transformation and other based on Locality-Constrained Iterative Neighbor Embedding (LINE) of local image patches. Both methods use coupled dictionaries to learn the mapping between LR and HR images in a very low resolution simulation using infrared iris images obtaining good results for very small images.

Despite the vast literature in SR area and the great interest in the use of Deep-Learning in Biometrics, the application of Deep Learning Super Resolution in iris recognition is still an unexplored field, mainly because approaches generally focus on general and/or natural scenes to produce overall visual enhancement and produce better quality images regarding to photo-realism, while iris recognition focuses on the best recognition performance itself [23] [24]. In [25], three multilayer perceptrons (MLPs) are used to perform single image super-resolution for Iris Recognition. The method is based on merging the bilinear interpolation approach with the output pixels values from the trained multiple MLPs considering the edge direction of the iris patterns. Recently, Zhang et.al [26] uses the classic Super-resolution Convolutional Neural Networks (SRCNN) and Super-resolution Forest (SRF) to perform super-resolution in Mobile Iris Recognition systems. The algorithms are applied in the segmented and normalized iris images and the results show a limited effectiveness of the super-resolution method for the iris recognition accuracy. Different from the methods presented in the DLSR literature, in this work we explore if the architectures, and the the database training can have influence in the quality results, and consequently in the recognition performance.

In our previous works [27] [28], we demonstrated that basic deep learning techniques for super-resolution such as stacked auto-encoders and the classic SRCNN can be successfully applied to Iris Super Resolution. In that case, we used the CASIA Interval database as target database focusing more in the recognition process. In this work we focus in the relation between the quality and the performance of the recognition and the super-resolution is performed in the original image without any segmentation. We also use a new iris database as target database that simulates a real world situation where the images are acquired using mobile phones. Additionally, we test a new application that is the use o Generative Adversarial Networks (SRGANs) to verify if the good performance of this method for natural images in terms of photo-realism is also valid for iris images in iris recognition context.

3 Reconstruction of Low Resolution Iris Images via CNNs

Typically, in a Deep Learning system, the main question is to find a good training database that can provide relevant information to the desired application. In the case of Super Resolution, it is necessary to achieve, during the proposed method training (also called the off-line phase), a mapping between a high-resolution (HR) image with high frequency information and a low-resolution (LR) image with low-frequency information. Figure 1 shows this phase, which a training database is chosen and the images are prepared for deep learning SR method training.

In the training phase, the only pre-processing required is, given an image in high resolution X, that image needs to be downscaled to one or more factors followed by a upscaling using bicubic interpolation to the same size as the original image X. This image, although it has the same size as X is called "low resolution" image and is denoted as the LR image Y. The purpose of Deep Learning SR training is, after feeding the network with a LR image or patch Y as input, try to obtain a result F(Y) (the reconstructed image) as much as similar to the HR image or patch X, in this case, the ground truth. The weight adjustment of the method will depend on both the chosen architecture and the loss function that will be better explained in the following sessions.

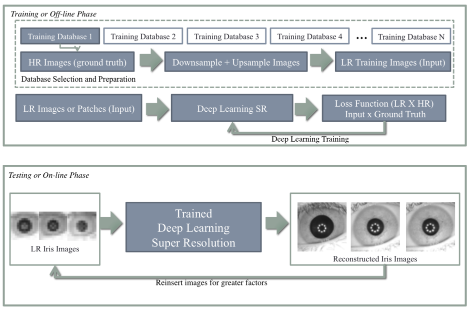


Fig. 1: General overview of the training and reconstruction method for the Iris Super Resolution using CNNs proposed for this work.

After training, the deep learning method is applied in a low resolution database for the proposed application which is, in the case of this work, an iris database also called target database. If so, the deep learning process is a pre-processing step before the iris recognition, in which the low resolution image is introduced as input to the network that will produce the reconstructed image in HR to be used in the process recognition as is shown in Figure 1 (on-line phase) that will be reconstructed based on the factor training.

In Deep Learning, the preparation of individual machines for all possible scenarios to deal with different scales, poses, illumination and textures is still a challenge. In this work, we test the main SR architectures, using different databases for the training, to evaluate some questions such as: if the similarity of the training database with the target database can aid in the process of super resolution or if the use of the target database itself (obtained during the enrollment of the individuals) can be used and if this knowledge can be transferred in a practical application.

4 CNN Architectures

Convolutional Neural Networks are considered the evolution of traditional Neural Networks, however, they share the same essence: a map of neurons with learnable weights, biases, activation functions and loss functions. The main impulse that contributed to the CNN popularity was the capability of treating 3D volumes of neurons (width, height and depth). Generally, the input of a CNN is formed by this 3d volume of size $m \times m \times d$, for example an image, where where $(m \times m)$ is the dimension of this image and d is the number of channels. The architecture of a CNN is defined depending on the application and generally is constructed stacking those layers using three main types: convolutional layers, pooling layers and the fully connected layer (exactly as seen in the traditional Neural Networks). A convolutional layer is formed by a series of k learnable filters with size $(n \times n \times d)$ where $(n \leq m)$. These filters (also known as kernels) are convolved in the input volume resulting in the so-called activation maps or feature maps. As classic Neural Networks, the convolution layer outputs are submitted to an activation function, e.g. the ReLU rectifier function $f(x) = \max(0, x)$ where x is the neuron input.

In this work we use three different deep learning approaches for super resolution: Super-resolution CNN (**SRCNN**), Very Deep Super-Resolution CNN (**VDCNN**) and Super Resolution Adversarial CNN (**SRGAN**). Each one of these approaches, architectures and methodologies used to make the image reconstruction are explained in the next subsections.

*4.1 Super Resolution Convolutional Neural Network (**SRCNN**)*

One of the first CNN architectures in Super-Resolution presented was the SRCNN [14]. This classical approach consists of three layers representing: the patch extraction, a non-linear mapping and the reconstruction step. As a pre-processing step, patches of size 33x33 (also called High Resolution (HR) patches) are extracted from the training images, then, as mentioned in the previous section, the patches are downscaled for the factor two and upscaled for the original size using bicubic interpolation. These also called Low-Resolution (LR) patches are used as the input for the CNN in the training phase.

The SRCNN architecture is specified as follows: the first layer (patch extraction) consists of 64 filters of size $9 \times 9 \times 1$ with stride one and padding zero, the second layer (non-linear mapping) has 32 filters of size $1 \times 1 \times 64$ with stride one and padding zero, and the last layer (reconstuction) has one filter of size $5 \times 5 \times 32$ with stride one and padding zero. The loss-function used in the CNN training is the Mean Squared Error (MSE) between the output (reconstructed patch) and the ground truth (HR patch) as well as the loss minimization is done using stochastic gradient descent and using the MatConvNet framework [29].

*4.2 Super-Resolution Very Deep Convolutional Neural Network (**VDCNN**)*

This architecture proposed in [16] relies on the use of a deeper CNN inspired by the VGG-net used for ImageNet classification. In the training phase, a pre-processing step is done by extracting HR patches and downscaling them for the two, three and four, re-upscaling them to the same size as the HR patches serving as the input for the CNN (LR patches).

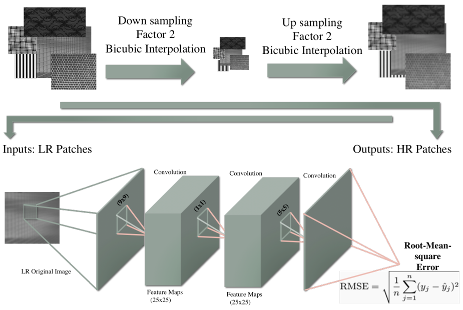


Fig. 2: An illustration of the Convolutional Neural Network architecture for Iris Super-Resolution (SRCNN).

The VDCNN architecture is composed of 20 layers with the same parameterization (except for the first and the last layers): 64 filters of size $3 \times 3 \times 64$. The loss function used in the training is the Mean Squared Error between the residual input error (difference between the reconstructed patch and the HR patch) and the residual ground truth that, in this case is the difference between LR and HR patch. This residual-learning boost the convergence of the CNN training and, consequently, its performance. The loss minimization is done also based on the gradient descent with backpropagation [30] using the MatConvNet framework [29].

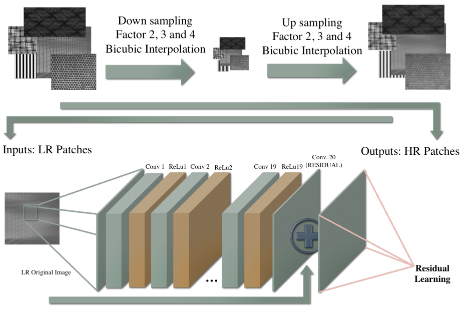


Fig. 3: An illustration of the Very Deep Convolutional Neural Network architecture for Iris Super-Resolution (VDCNN).

*4.3 Super-Resolution Generative Adversarial Network (**SRGAN**)*

This architecture proposed in [15] relies on two different CNNs with a new scheme of objective functions in an attempt to recover finer texture details from very low resolution images. While the generator architecture is responsible to generate the HR reconstructed image from the LR one, the discriminator architecture is trained to differentiate the reconstructed image from the original photo-realistic one.

As it can be seen in 4, the generator network is basically a series of residual blocks with identical layout: two convolutional layers of size with $3 \times 3 \times 64$ followed by a batch-normalization and ParametricRELU layers as activation function. The discriminator network is also based on VGG network and contains 8 convolutional layers with filters of size $3 \times 3 \times T$, where T is increased by a factor 2 through the layers from 64 to 512 filters as in the VGG network. The loss function used to training the method called perceptual loss function uses the output of both CNNâĂŹs (content loss and adversarial loss) trying to assess a solution with respect to perceptually relevant characteristics. For the training, the images were cropped in a size of 96x96 pixels and down-sampled for the factor 4 (LR input) as a pre-processing step.

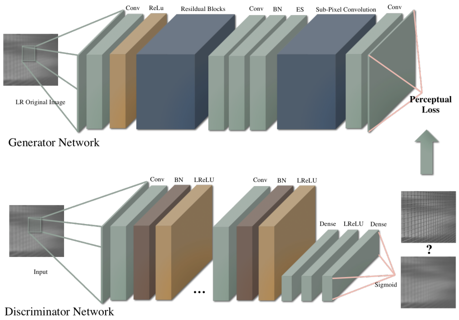


Fig. 4: An illustration of the Generative Adversarial Network architecture for Iris Super-Resolution (SRGAN).

5 Databases

5.1 Target/Test Database

In this work we use as the target database one of the most widely used databases on biometrics experiments: the CASIA Interval V3 database. This database contains 2655 NIR images of size 280×320 from 249 subjects captured with a self-developed close-up camera, resulting in 396 different eyes which will be considered as different subjects for this work. For the experiments, all the images from this database are interpolated using bicubic interpolation in order to have the same sclera radius followed by a cropping around the pupil in a square region of size 231x231. When the images do not fit in this cropping (e.g. if the iris is close to a margin), they are discarded. With this pre-processing step, 1872 images from 249 users are remained in the database.

In the experiments we explore the texture transfer learning between different databases, which means that the CNN is pre-trained with a different database (texture, natural or iris database), then it is used to perform the Super-Resolution in the target image database. For this part, we divide the target database into two: one with the first three images of each user (representing the registration images in a real world situation) and other with the remaining images form each user (representing the authentication database). The registration database is one of the iris training database among the others texture and natural databases that are explained in the next section.

5.2 Training Databases

As mentioned in the previous section, for CNN training we use 10 different databases from different nature to test the transfer learning and its impact in the recognition process. The databases include four texture datasets, two natural image datasets and four iris datasets (from the public IRISSEG-EP [31] dataset) detailed as follows.

5.2.1 Texture Databases: The Amsterdam Library of Textures (**ALOT**) with 27500 rough texture images of size 384×256 divided into 250 classes [32]. The Describable Texture Dataset (**DTD**) with 5640 images of sizes range betwenn 300×300 and 640×640 categorized in 47 classes [33]. The Flickr Material Database (**FMD**) containing 1000 images of size 512×384 divided into 10 categories [34]. The Textures under varying Illumination, Pose and Scale (**KTH-TIPS**) database with 10 different materials containing 81 cropped images of size 200×200 in each class [35].

5.2.2 Natural Image Databases: The **CALTECH101** Database is a natural image dataset with a list of objects belonging to 101 categories [36]. The **COREL1000** database is a natural image database containing 1000 color photographs showing natural scenes of ten different categories [37].

5.2.3 Iris Databases: The IIT Delhi Iris Database (**IITD**) is an Iris Database consisting of data acquired in a real environment resulting in 2240 images of size 230×240 from a digital CMOS near-infrared camera. The CASIA-Iris-Lamp (**CASIAIL**) is an Iris database collected using a hand-held iris sensor and containing 16212 images of size 320×280 with nonlinear deformation due to variations of visible illumination. The **UBIRIS** v2 Iris database is a database containing 2250 images of size 400×300 captured on non-constrained conditions (at-a-distance, on-the-move and on the visible wavelength), attempting to simulate more realistic noise factors. The **NOTREDAME** Iris Database is a collection of close-up near-infrared Iris images containing 837 images of size 640×480 with off-angle, blur, interlacing, and occlusion factors.

6 Experimental Setup

For the experiments, we test different down-sampling factors for the target database. For example, if the original image has size of 231×231 and is down-sampled for factor 4, this will correspond to 16x reduction in image pixels in a new image of size 57×57. Regardless of the chosen factor, for comparison criteria, all images are reconstructed by the CNNs until they reach the original size.

All methods evaluation and comparison in all stages of this work are based on the quality evaluation of the images as well as on the accuracy of the iris recognition. The qualitative assessment data will be given by two measures: the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) where a high score reflects a high quality using the HR image as the reference image.

For the recognition experiments we use one iris segmentation algorithm and two different feature extraction methods from the USIT - University of Salzburg Iris Toolkit v2 for Iris Recognition [38] . In the segmentation process, the iris is segmented and wrapped to a normalized rectangle of size 64×512 via the weighted adaptive Hough and ellipsopolar transform (WAHET). The first feature extraction is based on a complex Gabor filterbank with eight different filter size and wavelength (CG) while the second method is a classical wavelet-based feature extraction with a selection of spatial wavelets (QSW). In both cases, the bit-code vectors are compared using the normalized Hamming Distance [38]. Using the target database (CASIAIrisV3- Interval) with 249 users containing at least five or more images per user, we obtain 5087 genuine and 1746169 impostors scores.

7 Experimental Results

7.1 Texture Transfer Learning Comparison

In this section we explore the use of the texture transfer learning as an alternative to the training of CNNs in practical applications. For this, we chose to use the most basic architecture (SRCNN) trained with 10 different databases, including texture databases (ALOT, DTD, FMD and KTHTIPS), natural image databases (CALTECH 101, COREL1000 and IITD) and Iris Databases (CASIAIL, UBIRIS and NOTREDAME) applying it to the target database (CasiaInterval). In all frameworks, for a fair comparison between the databases,

Table 1 Results of quality assessment algorithms for texture transfer learning comparison with different downscaling factors (average values on the test dataset) using the SRCNN architecture comparing to the Bilinear and Bicubic approach.

				Texture Databases				Natural Image Databases			Iris Databases			
LR Size (SCALING)		Bilinear	Bicubic	ALOT	DTD	FMD	KTH TIPS	CALTECH 101	COREL 1000	IITD	CASIAIL	UBIRIS	NOTRE DAME	CASIA INTERVAL
115X115	PSNR	0.8855	0.8957	0.9481	**0.9595**	0.9509	0.9485	0.9492	0.9491	0.9483	0.9422	0.9414	0.9495	0.9502
(1/2)	SSIM	30.77	31.07	35.17	**35.87**	35.82	35.79	35.85	35.34	35.43	35.12	34.67	35.70	35.80
57X57	PSNR	0.7949	0.8089	0.8243	**0.8259**	0.8245	0.8232	0.8250	0.8255	0.8214	0.8129	0.8131	0.8216	0.8240
(1/4)	SSIM	27.99	28.67	29.20	**29.32**	29.29	29.23	29.24	28.97	29.18	29.01	28.86	29.24	29.29
29X29	PSNR	0.6956	0.7061	0.7198	0.7228	0.7157	0.7204	**0.7251**	0.7236	0.7127	0.7064	0.7085	0.7128	0.7174
(1/8)	SSIM	24.59	25.06	25.61	25.79	25.57	25.69	**25.80**	25.50	25.44	25.15	25.12	25.44	25.54
15X15	PSNR	0.6120	0.6160	0.6510	0.6544	0.6471	0.6503	**0.6557**	0.6553	0.6439	0.6406	0.6430	0.6447	0.6494
(1/16)	SSIM	20.78	20.93	23.09	**23.23**	23.07	23.04	23.21	23.05	23.01	22.67	22.69	22.97	22.95

Table 2 Verification results (EER) for texture transfer learning comparison for different downscaling factors using the SRCNN architecture comparing to the Bilinear and Bicubic approach. The accuracy result for the original database with no scaling is 6.65% for WAHET + CG and and 3.81% for WAHET + QSW.

				Texture Databases				Natural Image Database			Iris Databases			
LR Size (SCALING)		Bilinear	Bicubic	ALOT	DTD	FMD	KTH TIPS	CALTECH 101	COREL 1000	IITD	CASIAIL	UBIRIS	NOTRE DAME	CASIA INTERVAL
115X115	WAHET + CG	6.32	6.39	6.50	**6.07**	6.66	7.16	6.74	6.39	6.68	6.61	6.37	6.64	6.83
(1/2)	WAHET+QSW	**3.26**	3.58	3.58	3.32	3.81	4.28	4.02	3.53	3.89	3.92	3.42	4.02	3.84
57X57	WAHET + CG	9.36	**5.81**	7.19	6.67	6.88	6.22	6.83	6.51	7.90	7.84	8.41	7.59	6.66
(1/4)	WAHET+QSW	6.10	**2.65**	4.58	3.78	4.09	3.62	3.95	3.74	5.11	5.22	5.75	4.66	3.93
29X29	WAHET + CG	36.11	42.22	32.97	32.19	36.86	**22.41**	32.88	33.81	38.19	39.88	39.75	39.15	33.89
(1/8)	WAHET+QSW	33.60	42.34	30.62	31.13	34.89	**21.75**	32.10	33.26	36.50	38.53	37.33	37.04	30.65
15X15	WAHET + CG	31.66	32.96	33.95	33.10	33.03	33.96	33.02	34.68	32.73	**28.52**	29.62	31.50	31.57
(1/16)	WAHET+QSW	30.68	32.18	32.57	32.06	31.60	33.06	31.66	33.18	31.84	**27.60**	28.02	31.25	30.17

a subset of 150000 patches are extracted from each database to pre-train each CNN from scratch, when the CNN weights are initialized randomly.

We also compare the results with the use of two basic interpolation methods: Bilinear and Bicubic interpolation. Besides that, we train the CNN with the remaining images from the target database after the splitting (as explained in the Target Database section) to compare if images from the same individual can be beneficial to the CNN training.

In the Table 1 is presented the quality assessment results for the transfer learning in these databases for different factors: 2, 4, 8 and 16. For all the cases, the images were downscaled for these factors and reconstructed using the CNNs trained with the chosen databases showed in each column in the table. It can be noticed that the quality of the reconstructed images are more similar to the HR images than the interpolated by the traditional methods in all factors, including bigger down-sample factors as factor 8 and 16, demonstrating the flexibility of deep-learning when image resolution decreases. It also can be seen that the group with the best performance the texture database group showing that the texture patterns can provide a better generalization for the iris texture reconstruction. On the other hand, the results from the CASIA Interval (with different images from the same individual for training and testing) also present a good quality compared to the other databases.

In Table 2 we present the results for the iris recognition in order to be able to compare the photo-realism of the reconstructed images with the iris recognition performance. It can be seen that the best results were diversified among the methods and training databases also showing a divergence from the best results presented by the quality performance. The only database with the best result both for the quality assessment results and the recognition accuracy is the DTD database for the factor 2 (115x115) with 6.657% of EER. Another interesting point to notice is that, for the factor 2 and 4, almost all reconstruction methods surpasses the results using original images (the results for the original images are in the Table 2 caption) including the Bicubic interpolation which is, for small factors better than all the CNNs results.

Using the enrollment images from the same target database (CASIA Interval) does not lead to good recognition performances, which means that the CNN poorly memorize the patterns from the users focusing more in general patterns, mainly because the depth of the network that does not allow a high feature discrimination.

7.2 Architectures Comparison

To compare the three different CNN approaches, we take into consideration two databases from the transfer learning experiments: DTD database that presented, in general, a good performance both for quality and recognition measures and CASIA Interval database that uses the same database divided into training (simulating registration/enrollment images already stored in the system in a real situation) and testing (simulating the verification images) to see how the other CNN behaves with the same patterns presented in the training.

In this experiment we test all the architectures explained in section VI analyzing the performance using quality assessment algorithms as well as the recognition performance which are presented in table 3 and the visual results presented in the example in Figure 5. It can be seen that there is not the best approach for all the factors showing that there is not a "universal approach and general training database" to be used that can lead to the best results for quality and recognition process in all factors.

It is interesting to notice that, down-sampling the images for the factor 4 and reconstructing it, the results are better than using the

Table 3 Quality assessment (PSNR and SSIM) and verification results (WAHET + CG and WAHET + QSW) for different databases training and different downscaling factors using different architectures. The accuracy result for the original database with no scaling is 6.65% for WAHET + CG and 3.81% for WAHET + QSW.

		CASIA Interval					DTD		
LR Size(Scaling)		Bilinear	Bicubic	SRCNN	VDCNN	SRGAN	SRCNN	VDCNN	SRGAN
115x115 (1/2)	PSNR	0.8855	0.8957	0.9502	0.9555	0.9075	**0.9595**	0.9540	0.8937
	SSIM	30.77	31.07	35.80	**36.90**	23.46	35.87	36.56	21.68
	WAHET+CG	6.32	6.39	6.83	6.63	6.70	**6.07**	6.32	7.71
	WAHET+QSW	**3.26**	3.58	3.84	3.78	4.27	3.32	3.53	3.94
57x57(1/4)	PSNR	0.7949	0.8089	0.8240	0.8347	0.7914	0.8259	**0.8348**	77.47
	SSIM	27.99	28.67	29.29	29.60	24.10	29.32	**29.65**	22.15
	WAHET+CG	9.36	**5.81**	6.66	6.51	6.98	6.67	6.69	8.57
	WAHET+QSW	6.10	**2.65**	3.93	3.26	4.06	3.78	3.41	4.00
29x29 (1/8)	PSNR	0.6956	0.7061	0.7174	0.7332	0.6333	0.7228	**0.7374**	0.6488
	SSIM	24.59	25.06	25.54	26.04	22.08	25.79	**26.21**	21.00
	WAHET+CG	36.11	42.22	33.89	17.88	**13.58**	32.19	19.07	21.09
	WAHET+QSW	33.60	42.34	30.65	16.72	**13.38**	31.13	17.07	19.50
15x15 (1/16)	PSNR	0.6120	0.6160	0.6494	0.6563	0.5568	0.6544	**0.6633**	60.79
	SSIM	20.78	20.93	22.95	23.30	20.66	23.23	**23.57**	20.83
	WAHET+CG	31.66	32.96	**31.57**	33.87	38.32	33.10	33.85	34.46
	WAHET+QSW	30.68	32.18	**30.17**	32.03	38.41	32.06	31.76	35.92

Table 4 Quality assessment (PSNR, SSIM) results for different methods employed in the VSSIRIS database.

		FULL IMAGE					IRIS REGION				
LR Size(Scaling)		Bilinear	Bicubic	PCA	VDCNN	SRGAN	Bilinear	Bicubic	PCA	VDCNN	SRGAN
13x13 (1/22)	PSNR	24.44	24.97	**26.00**	25.26	24.59	24.35	24.89	**25.45**	24.89	18.08
	SSIM	0.7200	0.7200	**0.7300**	0.7256	0.5862	0.6200	0.6400	**0.6700**	0.6476	0.5395

original images (the results for the original images are in the Table 3 caption). This means that, in terms of recognition, it is better to downscale the original image (i.e. apply a blur filter) and apply the deep-learning methods from the sensor before comparison to perform a kind of denoising process in order to achieve better results for the recognition algorithms.

It also can be noticed that, for greater factors, the best approach using the quality assessment algorithms as a comparison measure, the best approach is the VDCNN using the DTD database as training. However for the recognition algorithms the results were divided between the approaches. For the factor 8, the SRGAN architecture presents a great result comparing to the other approaches for the CASIA INTERVAL database showing that deeper layers are allowed to extract more specific texture patterns from the users showing much better and consistent performance with this CNN. It is also worth to notice that for very small images (specially in the factor 1/16) the difference between the methods in this case is not debatable since the accuracy above 30% is unacceptable for a recognition system.

The example image from Figure 5 allows to compare the photo-realism presented by the methods in each factor. It can be noticed that the SRGAN approach tries to maximize the edge preservation generating a more consistent photo-realism as long as the factor decreases. However, this leads to too many artifacts that can lead to poorly results in the recognition process. As it can see by the red-squared images, the recognition performance is better when there is a balance between the texture, edge preservation and photo-realism of the iris.

7.3 Methods comparison using mobile phone images databases.

In this section we explore the use of CNN's in a real world situation where the images are captured from mobile devices comparing our results with another method found in the literature for iris SR (PCA-SR). For a complete comparison, in this case we use three quality assessment algorithms and two different recognition approaches (also used in the iris SR literature) that will be explained next.

For this experiment we chose the Visible Spectrum Smart-phone Iris (VSSIRIS [39]) database with images captured using two different mobile phone devices: Apple Iphone 5S (3264x2448 pixels) and Nokia Lumia 1020 (3072x1728 pixels). For each device, five images of the two eyes from 28 subjects were captured totaling 280 images per device or 560 in total. Figure 6 shows some example images from each device. As a pre-processing step, all the images are resized to have the same sclera radius followed by a cropping around the pupil in a square region of size 319x319 pixels. The down-sampling factor used for this experiment is the factor 1/22 following the previous studies in [40] and [41] to generate a very small iris region (13x13 pixels) for the real-world low resolution simulation.

We compare the CNN (trained with the DTD database) results with a method used in [41] and [22] for iris-super resolution called PCA hallucination of local patches based on the algorithm for face images of [42] where a PCA eigen-transformation is conducted in the set of LR basis patches to use the weights provided by the projection of the eingen-patches to reconstruct the images.

Besides the two quality assessment measures (PSNR and SSIM) we also compute the Feature Similarity Index for Image Quality Assessment (FSIM) from [43] that extracts low-level features including the significance of local structures and the image gradient magnitude. Table 4 presents the results using these quality assessment metrics comparing the CNNs trained with the DTD database with the bicubic, bilinear and PCA approaches for the full image and for the iris region.

It can be seen that, using the traditional measures (PSNR and SSIM) the aproach that presented the best results was the PCA reconstruction. However analyzing the images from Figure 7 it can be noticed that, in fact, the images that present more photo-realism are the images from the SRGAN CNN reconstruction in both devices. The PCA approach, although present the best results for PSNR and SSIM, visually looks the most artificial result between the methods, generating big squares of pixels due to the Eigen-Patch reconstruction nature.

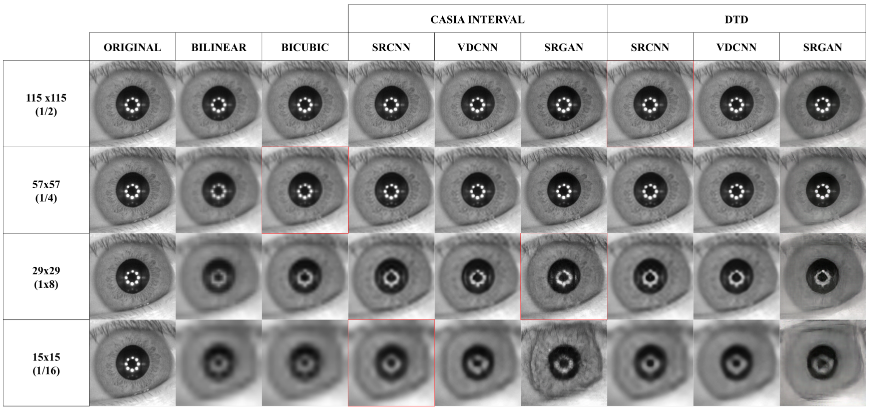


Fig. 5: Resulting images for different sampling factors in different approaches. The original image is replicates in all rows in the column. The red-squared images represent the approach that have the best recognition performance for the factor.

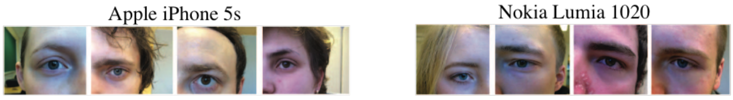


Fig. 6: Sample images from VSIRIS database [39].

However using the FSIM that is based on the human visual system, the best results are from the SRGAN CNN. Using this metric and analyzing the images from Figure 7 it can be noticed that, in fact, the images that present more photo-realism are the images from the SRGAN CNN reconstruction in both devices. The PCA approach, although present the best results for PSNR and SSIM, visually looks the most artificial result between the methods, generating big squares of pixels due to the Eigen-Patch reconstruction nature.

For the verification experiments we use the same recognition algorithms from the last experiments to evaluate the performance of the reconstruction methods. Table 5 presents the EER for these methods with the experiments done separately for each smart-phone with 560 genuine and 38500 impostors scores per device. We also use a new comparator for this experiment called SIFT comparator [44] in which SIFT feature points are extracted from the iris and compared based on the texture information around the points [45]. This is motivated by the factor that this feature extractor does not need any segmentation stage in the process which would be good for images with low quality. Nonetheless, this feature extraction algorithm was not used in the experiments 7.1 and 7.2 because it does not provide any additional information for discussion, unlike this experiment where the results are different comparing to the other feature extraction methods.

It can be noticed that for the WAHET+QSW and the WAHET+CG features, the best results for recognition are different from the best quality assessment results. The VDCNN reconstruction method presents the best result for both IPHONE and NOKIA images showing the robustness of this method for different databases since it was the best approach for the CASIA Interval database as well. Some of the results (specially the VDCNN and PCA methods) surpass the recognition results from the original database which means that blurring the texture can be beneficial to the recognition. The PCA approach that presents good quality results do not present the best result in the recognition experiments as well as the SRGAN approach showing that a good photo-realism does not reflect directly in a good recognition approach. Using the SIFT comparator that is based most in the edges and shapes of the images, the PCA approach presents good results followed by the SRGAN approach that presents, as mentioned before, in the visual results (Figure 7) a good photo-realism.

8 Conclusion

Deep-Learning Super Resolution via CNNs has been extensively explored to provide photo-realistic images from low-resolution images (mainly from natural scenes) obtaining impressive results. Meanwhile, more relaxed acquisition circumstances such as iris recognition via mobile phones or iris recognition via surveillance videos are boosting the need of SR methods to improve the iris recognition process. In this paper we explore different CNN architectures that are proven to be effective in reconstruct natural images for iris SR.

In the first part of experiments we choose a target database (CASIA Interval) and train the most basic CNN (SRCNN) to test the texture transfer learning between different databases from different natures: natural scenes, texture images or iris images. We also perform the training using the same database separated into training (registration images) and testing (validation images) to see if training the CNN with images from the same user can help in the SR. From these experiments we conclude that texture databases are more suitable to train the CNN for iris recognition as well as the use of the same database can also contribute to better results.

In the second part of the experiments we chose two different databases (CASIA Interval and DTD databases) to explore the use of three different sr CNNs: SRCNN, VDCNN and SRGAN. It can be seen that CNNs can produce more photo-realistic images with better quality than the traditional approaches. In specific, the VDCNN presents the best results in terms of quality, however, this does not reflect in terms of recognition rate. The visual analysis helps to understand this disparity on the results, in which where there is much photo-realism (for example, in the case of SRGAN) there is also too

Table 5 Verification results (EER) for different methods employed. The accuracy result for the original database with no scaling is 38.36% (WAHET+QSW) and 39.89% (WAHET+CG) for IPHONE images, 31.60% (WAHET+QSW) and 36.75% (WAHET+CG) for NOKIA images, 0.33% (SIFT) for IPHONE images and and 0.68% (SIFT) for NOKIA images.

		IPHONE					NOKIA				
LR Size(Scaling)		Bilinear	Bicubic	PCA	VDCNN	SRGAN	Bilinear	Bicubic	PCA	VDCNN	SRGAN
13x13 (1/22)	WAHET+QSW	32.64	33.16	33.80	**31.17**	39.29	30.78	30.81	32.40	**28.11**	39.09
	WAHET+CG	35.99	35.93	35.55	**32.23**	42.74	31.13	31.18	36.08	**27.54**	41.93
	SIFT	23.54	22.80	**9.30**	24.28	12.00	26.50	29.80	**11.13**	26.61	14.09

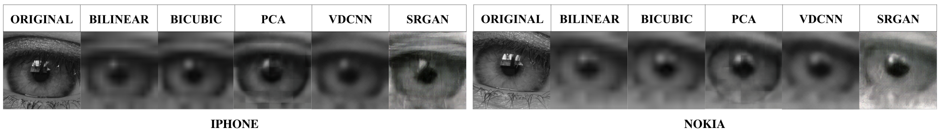


Fig. 7: Resulting images for different sampling factors in different approaches using the VSIRIS database.

many artifacts that can lead to poor results in the feature extraction. The balance between both photo-realism and smoothing images (as the case of SRGAN for the factor 8) is the perfect match for a good result.

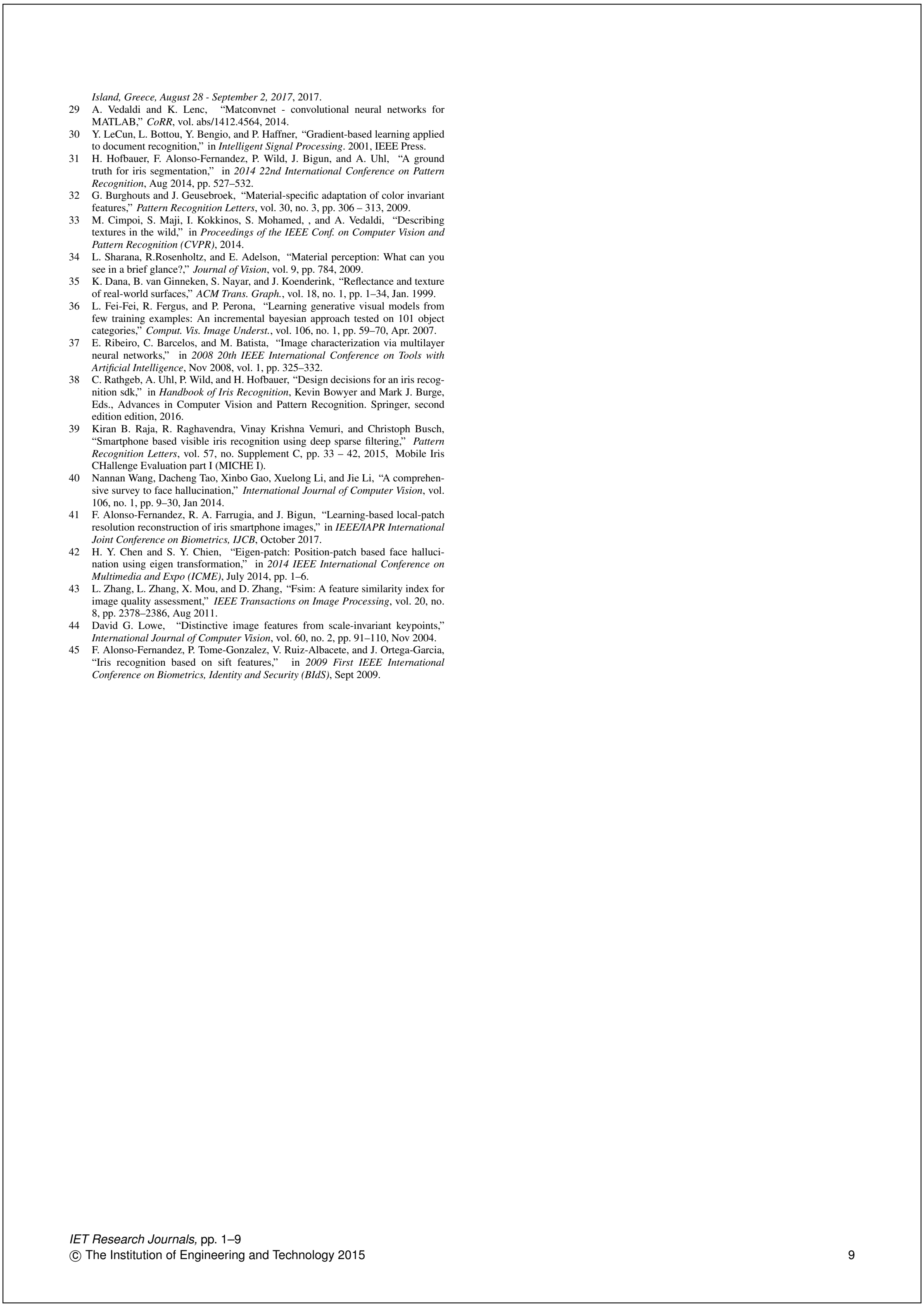
In the last part of the experiments we test the use of deep-learning super resolution for very low resolution images from mobile devices trying to simulate a real-world situation. Also in this experiment there is a dichotomy between the quality assessment and the recognition results showing that, a good photo-realism does not lead to a good recognition performance specially for very low-resolution images. In the future, based on the results of this work, we intend to create and test new CNN architectures specially designed for iris super-resolution that can provide a good balance between edge-preservation and smoothing to serve as a good pre-processing step mainly for images taken from distance and for mobile device iris recognition systems.

Acknowledgment

This research was partially supported by CNPq-Brazil for Eduardo Ribeiro under grant No. 00736/2014-0.

9 References

1 S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, Sep 2002.
2 Sung Cheol Park, Min Kyu Park, and Moon Gi Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, May 2003.
3 Kien Nguyen, Clinton Fookes, Sridha Sridharan, and Simon Denman, "Feature-domain super-resolution for iris recognition," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1526 – 1535, 2013.
4 N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Estimating and fusing quality factors for iris biometric images," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 3, pp. 509–524, May 2010.
5 P. Jonathon Phillips, Patrick J. Flynn, J. Ross Beveridge, W. Todd Scruggs, Alice J. O'Toole, David Bolme, Kevin W. Bowyer, Bruce A. Draper, Geof H. Givens, Yui Man Lui, Hassan Sahibzada, Joseph A. Scallan, and Samuel Weimer, *Overview of the Multiple Biometrics Grand Challenge*, pp. 705–714, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
6 K. Nguyen, S. Sridharan, S. Denman, and C. Fookes, "Feature-domain super-resolution framework for gabor-based face and iris recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 2642–2649.
7 F. Alonso-Fernandez, R. A. Farrugia, and J. Bigun, "Iris super-resolution using iterative neighbor embedding," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, July 2017, pp. 655–663.
8 Kamal Nasrollahi and Thomas B. Moeslund, "Super-resolution: a comprehensive survey," *Machine Vision and Applications*, vol. 25, no. 6, pp. 1423–1468, Aug 2014.
9 Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang, *Single-Image Super-Resolution: A Benchmark*, pp. 372–386, Springer International Publishing, Cham, 2014.
10 J. Allebach and Ping Wah Wong, "Edge-directed interpolation," in *Proceedings of 3rd IEEE International Conference on Image Processing*, Sep 1996, vol. 3, pp. 707–710 vol.3.
11 Xin Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521–1527, Oct 2001.
12 Eric Mjolsness, *Neural networks, pattern recognition, and fingerprint hallucination*, Thesis (dissertation (phd)), California Institute of Technology, 1986.
13 Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas S. Huang, "Deeply improved sparse coding for image super-resolution," *CoRR*, vol. abs/1507.08905, 2015.
14 Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *CoRR*, vol. abs/1501.00092, 2015.
15 Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016.
16 Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," *CoRR*, vol. abs/1511.04587, 2015.
17 Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
18 K. Nguyen, C. Fookes, S. Sridharan, and S. Denman, "Quality-driven super-resolution for less constrained iris recognition at a distance and on the move," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1248–1258, Dec 2011.
19 A. Deshpande and P. P. Patavardhan, "Super resolution and recognition of long range captured multi-frame iris images," *IET Biometrics*, vol. 6, no. 5, pp. 360–368, 2017.
20 Jiali Cui, Yunhong Wang, JunZhou Huang, Tieniu Tan, and Zhenan Sun, "An iris image synthesis method based on pca and super-resolution," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Aug 2004, vol. 4, pp. 471–474 Vol.4.
21 G. Fahmy, "Super-resolution construction of iris images from a visual low resolution face video," in *2007 9th International Symposium on Signal Processing and Its Applications*, Feb 2007, pp. 1–4.
22 F. Alonso-Fernandez, R. A. Farrugia, and J. Bigun, "Improving very low-resolution iris identification via super-resolution reconstruction of local patches," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sept 2017, pp. 1–6.
23 Kien Nguyen, Sridha Sridharan, Simon Denman, and Clinton Fookes, "Feature-domain super-resolution framework for gabor-based face and iris recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, 2012, pp. 2642–2649.
24 S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, Sep 2002.
25 K. Y. Shin, K. R. Park, B. J. Kang, and S. J. Park, "Super-resolution method based on multiple multi-layer perceptrons for iris recognition," in *Proceedings of the 4th International Conference on Ubiquitous Information Technologies Applications*, Dec 2009, pp. 1–5.
26 Qi Zhang, Haiqing Li, Zhaofeng He, and Zhenan Sun, *Image Super-Resolution for Mobile Iris Recognition*, pp. 399–406, Springer International Publishing, 2016.
27 Eduardo Ribeiro and Andreas Uhl, "Exploring texture transfer learning via convolutional neural networks for iris super resolution," in *Proceedings of the 2017 International Conference of the Biometrics Special Interest Group (BIOSIG'17), Darmstadt, Germany 2017*. 2017, LNI, GI / IEEE.
28 Eduardo Ribeiro, Andreas Uhl, Fernando Alonso-Fernandez, and Reuben A. Farrugia, "Exploring deep learning image super-resolution for iris recognition," in *Proc. of the 25th European Signal Processing Conference (EUSIPCO 2017), Kos*



Island, Greece, August 28 - September 2, 2017, 2017.

29 A. Vedaldi and K. Lenc, “Matconvnet - convolutional neural networks for MATLAB,” *CoRR*, vol. abs/1412.4564, 2014.

30 Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” in *Intelligent Signal Processing*. 2001, IEEE Press.

31 H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, and A. Uhl, “A ground truth for iris segmentation,” in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 527–532.

32 G. Burghouts and J. Geusebroek, “Material-specific adaptation of color invariant features,” *Pattern Recognition Letters*, vol. 30, no. 3, pp. 306 – 313, 2009.

33 M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi, “Describing textures in the wild,” in *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

34 L. Sharana, R.Rosenholtz, and E. Adelson, “Material perception: What can you see in a brief glance?,” *Journal of Vision*, vol. 9, pp. 784, 2009.

35 K. Dana, B. van Ginneken, S. Nayar, and J. Koenderink, “Reflectance and texture of real-world surfaces,” *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, Jan. 1999.

36 L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories,” *Comput. Vis. Image Underst.*, vol. 106, no. 1, pp. 59–70, Apr. 2007.

37 E. Ribeiro, C. Barcelos, and M. Batista, “Image characterization via multilayer neural networks,” in *2008 20th IEEE International Conference on Tools with Artificial Intelligence*, Nov 2008, vol. 1, pp. 325–332.

38 C. Rathgeb, A. Uhl, P. Wild, and H. Hofbauer, “Design decisions for an iris recognition sdk,” in *Handbook of Iris Recognition*, Kevin Bowyer and Mark J. Burge, Eds., Advances in Computer Vision and Pattern Recognition. Springer, second edition edition, 2016.

39 Kiran B. Raja, R. Raghavendra, Vinay Krishna Vemuri, and Christoph Busch, “Smartphone based visible iris recognition using deep sparse filtering,” *Pattern Recognition Letters*, vol. 57, no. Supplement C, pp. 33 – 42, 2015, Mobile Iris CHallenge Evaluation part I (MICHE I).

40 Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li, “A comprehensive survey to face hallucination,” *International Journal of Computer Vision*, vol. 106, no. 1, pp. 9–30, Jan 2014.

41 F. Alonso-Fernandez, R. A. Farrugia, and J. Bigun, “Learning-based local-patch resolution reconstruction of iris smartphone images,” in *IEEE/IAPR International Joint Conference on Biometrics, IJCB*, October 2017.

42 H. Y. Chen and S. Y. Chien, “Eigen-patch: Position-patch based face hallucination using eigen transformation,” in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, July 2014, pp. 1–6.

43 L. Zhang, L. Zhang, X. Mou, and D. Zhang, “Fsim: A feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.

44 David G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004.

45 F. Alonso-Fernandez, P. Tome-Gonzalez, V. Ruiz-Albacete, and J. Ortega-Garcia, “Iris recognition based on sift features,” in *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)*, Sept 2009.

5. Conclusion

In this thesis, we explored the use of Deep Learning and Transfer Learning as an accessory for two different applications showing that, these approaches can be successfully applied to different operations with good results. Our focus was to manage the Deep Learning training in order to avoid using alternative methods as the ”leave-one-out” cross validation that takes too much computation for this big data approach that is the Deep Learning method. Besides avoiding this problem, we found out, that with a big data training, the CNNs can be adapted and generalize very well to a new domain providing better results than using the same data distribution.

In the first contribution: the application of CNNs for Colonic Polyp Classification we explored and evaluated several different pre-trained CNNs architectures to extract features from colonoscopy images by the knowledge transfer between natural and medical images providing what it is called ”off-the-shelf” CNNs features.

We showed that the ”off-the shelf” features may be well suited for the automatic classification of colon polyps even with a limited amount of data. The different used CNNs were pre-trained with an image domain completely different from the proposed task, however, they provided a good and generic extractor of colonic polyps features. Some reasons for the success of the classification include the training with a large range of different images, providing a powerful extractor joining the intrinsic features from the images such as color, texture and shape in the same architecture, reducing and abstracting these features in just one vector.

Also, the combination of classical features with off-the-shelf features yielded good prediction results complementing each other. We believe that this strategy could be used in other endoscopic databases such as automatic classification of celiac disease. Besides that, this approach will be explored in future work to also detect polyps in video frames and the performance in real time applications will be evaluated. It can be concluded that Deep Learning through Convolutional Neural Networks is becoming essentially the most favorite candidate in almost all pattern recognition tasks.

In the second contribution: the application of CNNs for Iris Super Resolution we explored the use of texture transfer learning for super resolution applied to low resolution images. We have shown how the features from completely different nature can be transferred in the feature domain, improving the recognition performance if applied to bigger reduction factors comparing to the classical interpolation approaches.

The experiments showed that the transfer learning was successful using all databases especially for the texture databases and using a deeper architecture in an uncontrolled scenario (when both the enrollment and the authentication images are in low resolution) despite the fact that there was not a best database to be used in all factors. We also verified that there is a dichotomy between the quality assessment and the recognition results showing that, a good photo-realism in the Super-Resolution context does not lead to a good recognition performance specially for very low-resolution images in the case of Iris Recognition. In the future, based on the results of this thesis, we intend to create and test new CNN architectures specially designed for iris super-resolution that can provide a good balance between edge-preservation and smoothing to serve as a good pre-processing step mainly for images taken from distance and for mobile device iris recognition systems.

Bibliography

[1] ALJADAANY, R., LUU, K., VENUGOPALAN, S., AND SAVVIDES, M. Iris super-resolution via nonparametric over-complete dictionary learning. In *2015 IEEE International Conference on Image Processing (ICIP)* (Sept 2015), pp. 3856–3860.

[2] ALONSO-FERNANDEZ, F., FARRUGIA, R. A., AND BIGUN, J. Eigen-patch iris super-resolution for iris recognition improvement. In *2015 23rd European Signal Processing Conference (EUSIPCO)* (Aug 2015).

[3] AMELING, S., WIRTH, S., PAULUS, D., LACEY, G., AND VILARINO, F. Texture-based polyp detection in colonoscopy. In *Bildverarbeitung für die Medizin 2009*, Informatik aktuell. Springer Berlin Heidelberg, 2009, pp. 346–350.

[4] AYTAR, Y., AND ZISSERMAN, A. Tabula rasa: Model transfer for object category detection. In *2011 International Conference on Computer Vision* (Nov 2011), pp. 2252–2259.

[5] BAKER, S., AND KANADE, T. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 9 (Sep 2002), 1167–1183.

[6] BERNAL, J., SANCHEZ, J., AND VILARINO, F. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition 45*, 9 (2012), 3166 – 3182. Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).

[7] BOWYER, K., HOLLINGSWORTH, K., AND FLYNN, P. Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding 110*, 2 (2008).

[8] GANZ, M., YANG, X., AND SLABAUGH, G. Automatic segmentation of polyps in colonoscopic narrow-band imaging data. *Biomedical Engineering, IEEE Transactions on 59*, 8 (Aug 2012), 2144–2151.

[9] HÄFNER, M., KWITT, R., UHL, A., GANGL, A., WRBA, F., AND VÉCSEI, A. Feature extraction from multi-directional multi-resolution image transformations for the classification of zoom-endoscopy images. *Pattern Analysis and Applications 12*, 4 (2009), 407–413.

[10] HÄFNER, M., LIEDLGRUBER, M., UHL, A., VÉCSEI, A., AND WRBA, F. Delaunay triangulation-based pit density estimation for the classification of polyps in high-magnification chromo-colonoscopy. *Computer Methods and Programs in Biomedicine 107*, 3 (2012), 565–581.

[11] HÄFNER, M., UHL, A., AND WIMMER, G. A novel shape feature descriptor for the classification of polyps in hd colonoscopy. In *Medical Computer Vision. Large Data in Medical Imaging (Proceedings of the 3rd International MICCAI - MCV Workshop 2013)*, vol. 8331. Springer International Publishing, 2014, pp. 205–213.

[12] H.SHIN, ROTH, H., GAO, M., LU, L., XU, Z., NOGUES, I., YAO, J., MOLLURA, D., AND SUMMERS, R. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *CoRR abs/1602.03409* (2016).

[13] Huang, J. B., Singh, A., and Ahuja, N. Single image super-resolution from transformed self-exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 5197–5206.

[14] Kalka, N. D., Zuo, J., Schmid, N. A., and Cukic, B. Estimating and fusing quality factors for iris biometric images. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 40*, 3 (May 2010), 509–524.

[15] Kato, S., Fu, K. I., Sano, Y., Fujii, T., Saito, Y., Matsuda, T., Koba, I., Yoshida, S., and Fujimori, T. Magnifying colonoscopy as a non-biopsy technique for differential diagnosis of non-neoplastic and neoplastic lesions. *World J. Gastroenterol. 12*, 9 (Mar 2006), 1416–1420.

[16] Kudo, S., Hirota, S., and Nakajima, T. Colorectal tumours and pit pattern. *Journal of Clinical Pathology 10* (Oct 1994), 880–885.

[17] Li, J., Qu, Y., Li, C., Xie, Y., Wu, Y., and Fan, J. Learning local gaussian process regression for image super-resolution. *Neurocomputing 154* (2015).

[18] Nam, J., and Kim, S. Heterogeneous defect prediction. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering* (New York, NY, USA, 2015), ESEC/FSE 2015, ACM, pp. 508–519.

[19] Nguyen, K., Fookes, C., Sridharan, S., and Denman, S. Feature-domain super-resolution for iris recognition. *Computer Vision and Image Understanding 117*, 10 (2013).

[20] Nguyen, K., Sridharan, S., Denman, S., and Fookes, C. Feature-domain super-resolution framework for gabor-based face and iris recognition. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012* (2012), pp. 2642–2649.

[21] Oquab, M., Bottou, L., Laptev, I., and Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (June 2014), pp. 1717–1724.

[22] Pan, S. J., and Yang, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering 22*, 10 (Oct 2010), 1345–1359.

[23] Pan, S. J., and Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng. 22*, 10 (2010), 1345–1359.

[24] Park, S. Y., Sargent, D., Spofford, I., Vosburgh, K., and A-Rahim, Y. A colon video analysis framework for polyp detection. *Biomedical Engineering, IEEE Transactions on 59*, 5 (May 2012), 1408–1418.

[25] Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2014, Columbus, OH, USA, June 23-28, 2014* (2014), pp. 512–519.

[26] Razavian, A. S., Azizpour, H., Sullivan, J., and Carlsson, S. CNN features off-the-shelf: an astounding baseline for recognition. *CoRR abs/1403.6382* (2014).

[27] Ribeiro, E., A. Uhl, G. W., and Häfner, M. Exploring deep learning and transfer learning for colonic polyp classification. *Computational and Mathematical Methods in Medicine 2016* (2016), Article ID 6584725.

[28] Ribeiro, E., A. Uhl, G. W., and Häfner, M. Transfer learning for colonic polyp classification using off-the-shelf cnn features (best paper award, 3rd place). In *Proceedings of the 3rd International Workshop on Computer-Assisted and Robotic Endoscopy (CARE'16)* (2016), vol. 10170 of *Springer LNCS*, pp. 1–13.

[29] Ribeiro, E., Häfner, M., Wimmer, G., Tamaki, T., Tischendorf, J., S. Yoshida, S. T., and Uhl, A. Exploring texture transfer learning for colonic polyp classification via convolutional neural networks. In *14th International IEEE Symposium on Biomedical Imaging (ISBI'17)* (April 2017).

[30] Ribeiro, E., and Uhl, A. Exploring texture transfer learning via convolutional neural networks for iris super resolution. In *Proceedings of the 2017 International Conference of the Biometrics Special Interest Group (BIOSIG'17), Darmstadt, Germany 2017* (2017), LNI, GI / IEEE.

[31] Ribeiro, E., Uhl, A., and Alonso-Fernandez, F. Iris super-resolution using cnns: is photo-realism important to iris recognition? *Submitted to: IET Biometrics –*, – (2017), –.

[32] Ribeiro, E., Uhl, A., Alonso-Fernandez, F., and Farrugia, R. A. Exploring deep learning image super-resolution for iris recognition. In *Proc. of the 25th European Signal Processing Conference (EUSIPCO 2017), Kos Island, Greece, August 28 - September 2, 2017* (2017).

[33] Ribeiro, E., Uhl, A., and Häfner, M. Colonic polyp classification with convolutional neural networks. In *Proceedings of the 29th IEEE International Symposium on Computer-Based Medical Systems (CBMS'16)* (June 2016), pp. 253–258.

[34] Shin, K. Y., Park, K. R., Kang, B. J., and Park, S. J. Super-resolution method based on multiple multi-layer perceptrons for iris recognition. In *Proceedings of the 4th International Conference on Ubiquitous Information Technologies Applications* (Dec 2009), pp. 1–5.

[35] Tajbakhsh, N., Gurudu, S. R., and Liang, J. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medical Imaging 35*, 2 (Feb 2016), 630–644.

[36] Timofte, R., DeÂ Smet, V., and VanÂ Gool, L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *12th Asian Conference on Computer Vision* (Cham, 2015), Springer International Publishing.

[37] Vinokourov, A., Shawe-Taylor, J., and Cristianini, N. Inferring a semantic representation of text via cross-language correlation analysis. In *Proceedings of the 15th International Conference on Neural Information Processing Systems* (Cambridge, MA, USA, 2002), NIPS'02, MIT Press, pp. 1497–1504.

[38] W., Y., Tavanapong, W., Wong, J., Oh, J., and de Groen, P. Part-based multiderivative edge cross-sectional profiles for polyp detection in colonoscopy. *Biomedical and Health Informatics, IEEE Journal of 18*, 4 (July 2014), 1379–1389.

[39] Wang, Y., Tavanapong, W., Wong, J., Oh, J. H., and de Groen, P. C. Polyp-alert: Near real-time feedback during colonoscopy. *Computer Methods and Programs in Biomedicine 120*, 3 (2015), 164 – 179.

[40] Wimmer, G., Tamaki, T., Tischendorf, J., Häfner, M., Yoshida, S., Tanaka, S., and Uhl, A. Directional wavelet based features for colonic polyp classification. *Medical Image Analysis 31* (2016), 16 – 36.

[41] Yang, J., Wang, Z., Lin, Z., Cohen, S., and Huang, T. Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing 21*, 8 (Aug 2012).

[42] Zhang, Q., Li, H., He, Z., and Sun, Z. *Image Super-Resolution for Mobile Iris Recognition*. Springer International Publishing, 2016, pp. 399–406.

[43] Zhou, J. T., Pan, S. J., Tsang, I. W., and Yan, Y. Hybrid heterogeneous transfer learning through deep learning. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence* (2014), AAAI'14, AAAI Press, pp. 2213–2219.

[44] Zhu, Y., Chen, Y., Lu, Z., Pan, S. J., Xue, G.-R., Yu, Y., and Yang, Q. Heterogeneous transfer learning for image classification. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence* (2011), AAAI'11, AAAI Press, pp. 1304–1309.

A. Appendix

A.1. Breakdown of Authors' Contribution

This section detail a breakdown of itemized authors contributions of the papers included in this thesis. The author names are listed alphabetical order except for the name of the author of this thesis (Eduardo Ribeiro) that comes first.

Univ.-Prof. Dr. Andreas Uhl is the thesis advisor of Eduardo Ribeiro. Since the explicit contribution of an advisor cannot be stated for a single paper, it is omitted in the following breakdown. The medical experts (Michael Häfner, Toru Tamaki, Shinji Tanaka, J.J.W. Tischendorf, Shigeto Yoshida) provided the endoscopic images but were not involved in the production of the papers and hence are not listed in the following breakdown.

	Contribution (in %)			
Publication	Eduardo Ribeiro	Fernando Alonso-Fernandez	Georg Wimmer	Reuben A. Farrugia
Convolutional Neural Networks and Transfer Learning applied to Colonic Polyp Classification				
RIBEIRO, E., UHL, A., AND HÄFNER, M. Colonic polyp classification with convolutional neural networks. In *Proceedings of the 29th IEEE International Symposium on Computer-Based Medical Systems (CBMS'16)* (June 2016), pp. 253–258	100			
RIBEIRO, E., A. UHL, G. W., AND HÄFNER, M. Transfer learning for colonic polyp classification using off-the-shelf cnn features (best paper award, 3rd place). In *Proceedings of the 3rd International Workshop on Computer-Assisted and Robotic Endoscopy (CARE'16)* (2016), vol. 10170 of *Springer LNCS*, pp. 1–13	90		10	
RIBEIRO, E., A. UHL, G. W., AND HÄFNER, M. Exploring deep learning and transfer learning for colonic polyp classification. *Computational and Mathematical Methods in Medicine 2016* (2016), Article ID 6584725	90		10	
RIBEIRO, E., HÄFNER, M., WIMMER, G., TAMAKI, T., TISCHENDORF, J., S. YOSHIDA, S. T., AND UHL, A. Exploring texture transfer learning for colonic polyp classification via convolutional neural networks. In *14th International IEEE Symposium on Biomedical Imaging (ISBI'17)* (April 2017)	90		10	

	Contribution (in %)			
Publication	Eduardo Ribeiro	Fernando Alonso-Fernandez	Georg Wimmer	Reuben A. Farrugia
Convolutional Neural Networks and Transfer Learning applied to Iris Super Resolution				
Ribeiro, E., Uhl, A., Alonso-Fernandez, F., and Farrugia, R. A. Exploring deep learning image super-resolution for iris recognition. In *Proc. of the 25th European Signal Processing Conference (EUSIPCO 2017), Kos Island, Greece, August 28 - September 2, 2017* (2017)	85	10		5
Ribeiro, E., and Uhl, A. Exploring texture transfer learning via convolutional neural networks for iris super resolution. In *Proceedings of the 2017 International Conference of the Biometrics Special Interest Group (BIOSIG'17), Darmstadt, Germany 2017* (2017), LNI, GI / IEEE	100			
Ribeiro, E., Uhl, A., and Alonso-Fernandez, F. Iris super-resolution using cnns: is photo-realism important to iris recognition? *Submitted to: IET Biometrics –*, – (2017), –	90	10		