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.

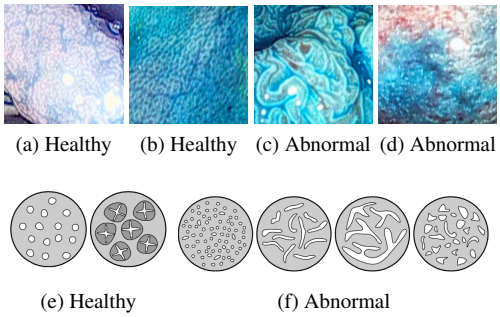

Fig. 1: Example images of the two classes (a-d) and the pit-pattern types of these two classes (e-f).

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.

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
$\overline{X}$	<u>70.86</u>	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.