Triplet network for classification of benign and pre-malignant polyps

Triplet Network for classification of benign and pre-malignant polyps

Roger Fonollà[a], Maciej Smyl[a], Fons van der Sommen[a], Ramon M. Schreuder[b], Erik J. Schoon[b], and Peter H.N. de With[a]

[a]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[b]Department of Gastroenterology and Hepatology, Catharina Hospital, Eindhoven, The Netherlands

ABSTRACT

Colorectal polyps are critical indicators of colorectal cancer (CRC). Classification of polyps during colonoscopy is still a challenge for which many medical experts have come up with visual models, albeit with limited success. An early detection of CRC prevents further complications in the colon, which makes identification of abnormal tissue a crucial step during routinary colonoscopy. In this paper, a classification approach is proposed to differentiate between benign and pre-malignant polyps using features learned from a Triplet Network architecture. The study includes a total of 154 patients, with 203 different polyps. For each polyp an image is acquired with White Light (WL), and additionally with two recent endoscopic modalities:Blue Laser Imaging (BLI) and Linked Color Imaging (LCI). The network is trained with the associated triplet loss, allowing the learning of non-linear features, which prove to be a highly discriminative embedding, leading to excellent results with simple linear classifiers. Additionally, the acquisition of multiple polyps with WL, BLI and LCI, enables the combination of the posterior probabilities, yielding a more robust classification result. Threefold cross-validation is employed as validation method and accuracy, sensitivity, specificity and area under the curve (AUC) are computed as evaluation metrics. While our approach achieves a similar classification performance compared to state-of-the-art methods, it has a much lower inference time (from hours to seconds, on a single GPU). The increased robustness and much faster execution facilitates future advances towards patient safety and may avoid time-consuming and costly histhological assessment.

Keywords: Triplet Network, Polyp classification, Linear SVM, Blue Laser Imaging, Linked Color Imaging

1. INTRODUCTION

Colorectal cancer (CRC) is the second most diagnosed cancer in women and the third in men worldwide, with more than half of its incidence rates rising in developed countries.[1] An early diagnosis of CRC can prevent spreading throughout the colon and avoid further complications. In the initial stages of the cancer, abnormal colorectal tissue or polyps can be classified in three different groups, hyperplastic polyps (HPs), adenomas (ADs) and sessile serrated adenomas (SSAs). HPs are considered benign polyps, whilst ADs and SSAs are identified as pre-maligannt polyps, capable of developing to CRC when kept untreatead.[2] Visual differentiation of benign and pre-malignant polyps is an on-going challenge in the clinical endoscopy routine, often dictated by the expertise of the clinician. White light endoscopy (WL) is the most common technique to assess lesions in the intestinal tract, but it falls behind when enhancing the visualization of the vessels and the mucosa. Compared to WL, chromoendoscopy techniques[3] are capable of achieving high-contrast results, but they require the injection of chemical dyes into the body. Similar visual effects can be achieved with the use of in-vivo optical filters, like Narrow-Band Imaging (NBI)[4,5] Alternatively, LED-based techniques like Blue Laser Imaging (BLI) and Linked Color Imaging (LCI)[6] exploit the absorption rate of the hemoglobin in the range of the blue-violet light to achieve comparable enhanced results. The advance towards less invasive modalities could potentially avoid the

Further author information:
E-mail: r.fonolla.navarro@tue.nl

use of chemical stains, while still providing the same amount of visual information. The increasing need for early diagnosis of CRC imposes a growing requirement of in-vivo discrimination between benign and pre-malignant polyps. Computer-aided diagnosis (CAD) systems facilitate with an extra opinion for more detail in clinical decisions, while potentially preventing costly resections and histopathological analysis. In the past years, CAD systems have taken advantage of machine learning methods and visual descriptors or features to automatically classify several pathologies in the gastrointestinal tract.[7] Previous studies on polyp classification extracted local features from blood vessels using NBI images[8,9],[10] while others achieved similar results by combining chromoendoscopy, WL and NBI.[11] In Scheeve *et al.*[12] handcrafted features were used to predict the histology of polyps using Support Vector Machines (SVMs) and clinical classification models. In recent years, Convolutional Neural Networks (CNNs) have proven to be strong feature extractors for a wide variety of areas and applications, including medical image analysis.[13] In the study of Zhang *et al.*[14] several features of a CNN from non-medical databases were used to detect and classify colorectal polyps in three different classes (This approach is further referred to as CNN features). Similar work was achieved by Murata *et al.*,[15] where CNN features were combined with multiple SVMs and a voting system to classify polyps in between benign and pre-malignant using dye imaging, WL and NBI.

In our previous work[16] we extracted CNN features from already trained residual networks and combined them with multiple non-linear SVMs in conjunction with data augmentation at test-time. The proposed workflow increased the performance of polyp classification compared to the above studies, at the cost of high computational time and meticulous parameter tuning for each modality.

In this work, we explore the use of the Triplet Network architecture and its associated triplet loss in order to learn non-linear representations between polyps in three different modalities. We show that the learned features serve as a highly discriminative basis for subsequent machine learning models. This is demonstrated by our experiments, in which we use threefold cross-validation for validation (from now on referred as 3FCV). We compare our results with previous studies and we show that the features learned from a Triple Network are capable of describing the non-linearity of a relatively small and imbalanced dataset and hence suitable to be used in a linear classifier. We achieve results comparable to using non-linear classifiers, without the need of an increased computational cost due excessive steps during test-time, thus decreasing the number of trained classifiers by tenfold.

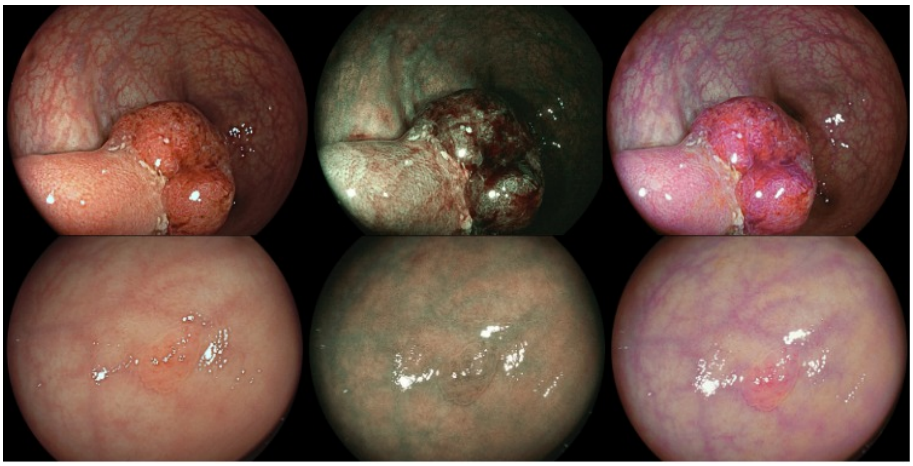

Figure 1. (Top) Example of a pre-malignant polyp from the CZE dataset. (Bottom) Example of a benign polyp. From left to right, a visualization of the three image acquisition modalities used in the study; White Light Endoscopy, Blue Light Imaging and Linked Color Imaging, respectively.

2. METHODS

2.1 Data description

The data collection was carried out in a prospective fashion, according to a pre-defined image acquisition protocol, in the Catharina Hospital Eindhoven (CZE). A total of 154 patients, with a total of 203 polyps were included in this study. After histopathology examination, 172 polyps were found to be pre-malignant and 31 were definitively benign polyps. For each patient, a single image of the polyp was acquired at different time steps with three different modalities, WL, BLI and LCI. All collected data was fully anonymized prior to the study.

2.2 Triplet Network

The Triplet Network architecture (inspired by the Siamese network[17]) consists of three identical sub-networks with shared parameters. Each sub-network is trained to learn embedded features of three different samples, called anchor, positive and negative sample. The combination of an anchor, positive and negative sample is called a *triplet*. The network output consist of the L2 distance between the anchor and the positive sample, and the anchor and the negative sample. Given this learned information, the cost function is calculated through the triplet loss in Eq. (1), where f_i^a denotes the anchor embedding, f_i^p is the positive embedding and f_i^n is the negative embedding. The triplet loss function is associated with the triplets and their individual distances, leading to the following definition:

$$\mathcal{L} = \max\left(0, ||f_i^a - f_i^p||_2^2 - ||f_i^a - f_i^n||_2^2 + \alpha\right). \tag{1}$$

The parameter α is called the margin and determines how successful the network is in identifying negatives samples that are not the easiest (far away from the anchor) or the hardest to learn (too close to the anchor in the embedding space).

2.3 Preprocessing

In order to obtain optimal classification, a manual selection of the ROI was traced for each polyp. The cropped region ensures a coverage of the polyp area, as well as its surrounding texture. Successively, the dataset was normalized by subtracting the mean and dividing the standard deviation of the pre-trained ImageNet weights. As last step, each input image was resized to 224×224 pixels in the RGB color space. To increase the generalization of the network data augmentation is used to enhance the model capabilities for our classification task. In this study, the training images are augmented by a combination of *flipping, shifting* and $\pm$ 90°*rotation*.

2.4 Training

In order to preserve a balanced training for both classes, the same amount of triplets were selected for both the positive and the negative class. To achieve the balanced training, data augmentation (rotations, horizontal flip and vertical flip) was applied during training to successfully obtain an equal amount of positive and negative sample, only leaving the anchor sample unaffected. We assigned the anchor and positive samples as the benign polyps, hence the negative class is referred to as the pre-malignant polyps. In total three Triplet networks were trained with the exact same hyperparameters, where SGD was chosen as preferred optimizer with learning rate 0.001, momentum 0.9 and decay 0.001. Each of the networks was initiliazed with ImageNet weights and trained from scratch. To analyze the extracted features, one Linear SVM for modality and fold was trained and evaluated.

3. EVALUATION AND RESULTS

Threefold cross-validation (3FCV) was employed as validation method for our approach. For each fold, the features were extracted from the Triplet Network. The training features of each fold were partitioned by modality (WL, BLI and LCI) and a Linear SVM was trained for each fold and modality using the best parameters obtained with a grid search on the training set. For each modality, we computed the accuracy and the area under the curve (AUC). Moreover, sensitivity, defined as the rate of correct pre-malignant polyps classified as such, and specificity, defined as the correct rate of benign polyps classified as benign, were calculated as well. In the work of Murata *et al.*,[15] the authors proposed a voting system from the predictions of each combination. Although

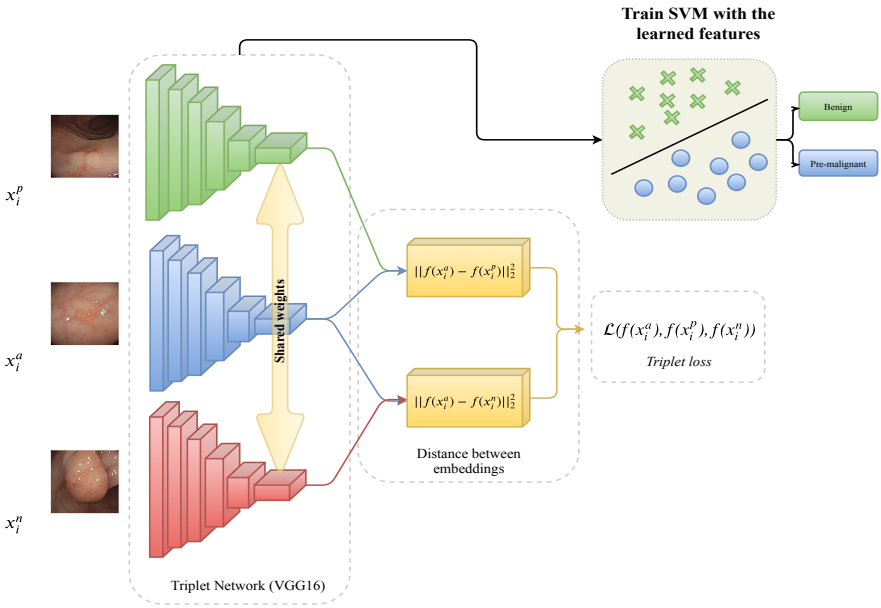

Figure 2. Workflow of the proposed algorithm, where a Triplet Network is trained using the triplet loss. From the learned features several SVMs are trained for each modality (WL, BLI, LCI). The combination of all SVMs are used to increase the classification of the model.

the voting system would be suitable, we prefer to adopt the same methodology as our previous study[16] based on residual networks, where the posterior probabilities acquired for each modality are averaged as a method to compute one unique probability per polyp. The results of the last approach are shown in Table 1, where the performance of each combination of features is presented. To further extend the evaluation of our method, we compared our algorithm with the implementations of Zhang *et al.*,[14] Murata *et al.*[15] as well as our previous proposed approach, since as far our knowledge extends no further studies have been conducted on a similar topic. The first two studies used the Alexnet[18] architecture to extract CNN features. Of those two, the first study showed its best results using features obtained from Places205,[19] the second study was based on the weights from ILSVR15.[20] The third (our previous work) and most recent study obtained features from Resnet50 pre-trained with ImageNet weights and further increased the results by applying test-time data augmentation. We implemented the methods proposed in the original studies and tested and compared the results with our own dataset. The results are presented in Table 2, which shows that the proposed method has a similar performance at default operating point but with a much lower inference time when compared to previous work.

	Accuracy	**Sensitivity**	**Specificity**	**AUC**
Residual Network[16]				
WL	77.3%	77.3%	77.4%	**0.88**
BLI	82.8%	82.6%	83.9%	**0.91**
LCI	82.3%	81.9%	83.9%	**0.90**
WL+BLI+LCI	83.3%	83.1%	83.9%	**0.90**
Triplet Network				
WL	81.8%	82.6%	77.4%	**0.87**
BLI	86.7%	87.2%	83.9%	**0.90**
LCI	84.7%	83.1%	93.5%	**0.89**
WL+BLI+LCI	89.2%	89.0%	90.3%	**0.91**

Table 1. Threefold cross-validation (3FCV) results for polyp malignancy classification using WL, BLI and LCI features extracted from the Triplet Network.

Comparison results	Accuracy	Sensitivity	Specificity	AUC	Kernel	Inference time
Zhang *et al.*[14]	70.4%	70.4%	70.9%	0.90	RBF	**53 s**
Murata *et al.*[15]	86.2%	88.4%	74.2%	N.A.	RBF*	**>10 min**
Residual Network features[16†]	90.2%	90.1%	90.3%	0.97	RBF	**> 1h**
Triplet Network features	89.2%	89.0%	90.3%	0.91	Linear	**56 s**

†.The authors do not specify the kernel used. Our best results were obtained with RBF.

⋆.Test-time augmentation was used to achieve the results.

Table 2. Comparison of three polyp classification studies against our proposed method using the CZE dataset.

4. DISCUSSION

In endoscopy, removal of all polyps is the established procedure, either if the polyp is found benign or is pre-malignant. Table 1 demonstrates the ability to classify a polyp from visual images obtained during colonoscopy. We improve on our previous study[16] on two aspects. First, we show that with our proposed method we achieve higher sensitivity and specificity for each single modality. Second, similar to state-of-the-art studies, we have employed information of different modalities (WL, BLI and LCI), achieving a sensitivity of 89.0% and specificity of 90.3%. In Table 2, we compare our work with previous studies, all which used non-linear SVMs to train on previously extracted features. By using simple linear SVM classifiers, we have shown that the features from the Triplet Network are capable of capturing the non-linear relations in the data. This property allows us to achieve similar results compared to state-of-the-art studies, while decreasing the evaluation time from 4 hours to 56 seconds (using a Titan Xp GPU), without compromising the combined classifications results. This renders our newly presented method more suitable towards real-time colonoscopy. In addition, we have highlighted that combining the predictions of multiple modalities (WL, BLI and LCI), increases the final results for our dataset.

Lastly, the resulting embedding from the trained Triplet Network is capable of separating the anchor from the negative class, while maintaining the proximity between the anchor and the positive class. We can observe the results of the training in Fig. 3, where positive samples stay together, whereas negative samples are pushed away from its anchor.

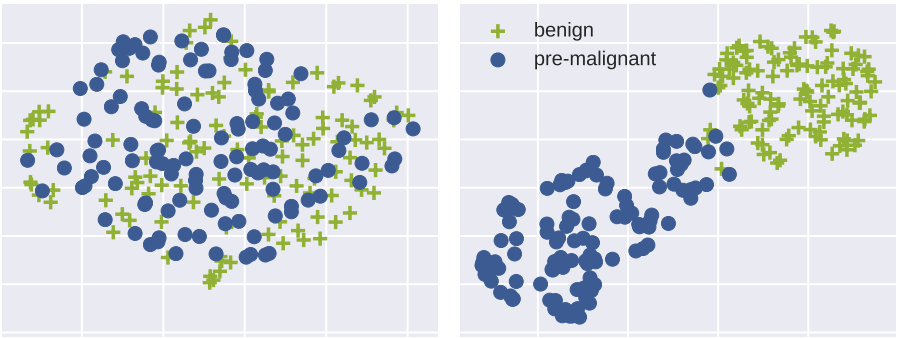

Figure 3. Uniform Manifold Approximation & Projection (UMAP) is used to reduce the dimensionality of the feature embedding into a 2D representation. Left: Initial features after the 1-st training epoch, showing the mixing of classes in the embedding space without any differentiation. Right: representation of the final embedding after model convergence to an optimal solution. The negative class (dots) is separated from the anchor, the positive class (crosses) is kept together.

5. CONCLUSIONS

In this work, we have presented a novel approach for classification of benign and pre-malignant polyps. The proposed method exploits the discriminative visual information in three different endoscopic modalities (White Light, Blue Laser Imaging and Linked Color Imaging). We have explored the learning capabilities of the triplet loss with a Triplet Network architecture, showing that the employed techniques are capable of learning non-linear class relations from a highly unbalanced dataset. Our experiments demonstrate that our approach achieves an

improved performance compared to similar studies. Despite the promising results of our approach, we consider that and end-to-end CNN could potentially even further improve the classification results. However, a considerably larger dataset would be necessary to adapt the proposed changes. Overall, the proposed CAD system can provide support to endoscopists with in-vivo polyp classification, possibly avoiding unnecessary resections of benign polyps, thereby improving patient safety and avoiding expensive histopathological assessment.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 721766. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

REFERENCES

[1] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D., Forman, D., and Bray, F., "Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012," *International Journal of Cancer* **136**(5), E359–E386 (2015).

[2] Jass, J. R., "Classification of colorectal cancer based on correlation of clinical, morphological and molecular features," *Histopathology* **50**(1), 113–130 (2007).

[3] Har-Noy, O., Katz, L., Avni, T., Battat, R., Bessissow, T., Yung, D. E., Engel, T., Koulaouzidis, A., Eliakim, R., Ben-Horin, S., and Kopylov, U., "Chromoendoscopy, narrow-band imaging or white light endoscopy for neoplasia detection in inflammatory bowel diseases," *Digestive Diseases and Sciences* **62**, 2982–2990 (Nov 2017).

[4] East, J. E., Guenther, T., Kennedy, R. H., and Saunders, B. P., "Narrow band imaging avoids potential chromoendoscopy risks," *Gut* **56**(8), 1168–1169 (2007).

[5] Visovan, I., Tantau, M., Pascu, O., Ciobanu, L., and Tantau, A., "The role of narrow band imaging in colorectal polyp detection," *Bosnian Journal of Basic Medical Sciences* **17**(2), 152–158 (2017).

[6] Nishida, H., Ozawa, S., Shimomura, K., Abe, S., Yoshida, K., Ohashi, E., and Mishima, T., "Development of a New Endoscope platform," Tech. Rep. 62-2017, Fujifilm Research & Development (2016).

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

[8] Gross, S., Stehle, T., Behrens, A., Auer, R., Aach, T., Winograd, R., Trautwein, C., and Tischendorf, J., "A comparison of blood vessel features and local binary patterns for colorectal polyp classification," *SPIE Medical Imaging* **7260**, 72602Q–72602Q–8 (2009).

[9] Tischendorf, J., Gross, S., Winograd, R., Hecker, H., Auer, R., Behrens, A., Trautwein, C., Aach, T., and Stehle, T., "Computer-aided classification of colorectal polyps based on vascular patterns: A pilot study," *Endoscopy* **42**(3), 203–207 (2010).

[10] Gross, S., Trautwein, C., Behrens, A., Winograd, R., Palm, S., Lutz, H., Schirin-Sokhan, R., Hecker, H., Aach, T., and Tischendorf, J., "Computer-based classification of small colorectal polyps by using narrow-band imaging with optical magnification," *Gastrointestinal Endoscopy* **74**(6), 1354–1359 (2011).

[11] Tamaki, T., Yoshimuta, J., Kawakami, M., Raytchev, B., Kaneda, K., Yoshida, S., Takemura, Y., Onji, K., Miyaki, R., and Tanaka, S., "Computer-aided colorectal tumor classification in nbi endoscopy: Using local features," *Medical Image Analysis* **17**(1), 78–100 (2013).

[12] Scheeve, T., Schreuder, R.-M., van der Sommen, F., IJspeert, J. E. G., Dekker, E., Schoon, E. J., and With, P. H. N. D., "Computer-aided classification of colorectal polyps using blue-light and linked-color imaging," in [*Medical Imaging 2019: Computer-Aided Diagnosis*], Mori, K. and Hahn, H. K., eds., **10950**, 268 – 275, International Society for Optics and Photonics, SPIE (2019).

[13] Tamaki, T., Yoshimuta, J., Kawakami, M., Raytchev, B., Kaneda, K., Yoshida, S., Takemura, Y., Onji, K., Miyaki, R., and Tanaka, S., "Computer-aided colorectal tumor classification in nbi endoscopy: Using cnn features," (2016).

[14] Zhang, R. I., Zheng, Y., Mak, W. C. T., Yu, R., Wong, S. H., Lau, J., and Poon, C., "Automatic detection and classification of colorectal polyps by transferring low-level cnn features from non-medical domain," *IEEE Journal of Biomedical and Health Informatics* **21**(1), 1–1 (2016).

[15] Murata, M., Usami, H., Iwahori, Y., and Aili, W., "Polyp classification using multiple cnn-svm classifiers from endoscope images," *The Ninth International Conferences on Pervasive Patterns and Applications* (c), 109–112 (2017).

[16] Fonolla, R., van der Sommen, F., Schreuder, R., Schoon, E. J., and de With, P. H. N., "Multi-modal classification of polyp malignancy using CNN features with balanced class augmentation," in [*16th IEEE International Symposium on Biomedical Imaging, ISBI 2019, Venice, Italy, April 8-11, 2019*], 74–78 (2019).

[17] Koch, G., Zemel, R., and Salakhutdinov, R., "Siamese neural networks for one-shot image recognition," (2015).

[18] Krizhevsky, A., Sutskever, I., and E. Hinton, G., "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems* , 1–9 (2012).

[19] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A., "Learning deep features for scene recognition using places database," 487–495 (2014).

[20] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A.and Bernstein, M., C. Berg, A., and Fei-Fei, L., "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision (IJCV)* **115**(3), 211–252 (2015).