Colorectal Polyp Classification from White-Light Colonoscopy Images via Domain Alignment

Qin Wang[1,2], Hui Che[1,2], Weizhen Ding[1,2], Li Xiang[3], Guanbin Li[2,4], Zhen Li[1,2(✉)], and Shuguang Cui[1,2]

[1] The Chinese University of Hong Kong, Shenzhen, China
qinwang1@link.cuhk.edu.cn, lizhen@cuhk.edu.cn
[2] Shenzhen Research Institute of Big Data, Shenzhen, China
[3] Longgang District People's Hospital of Shenzhen, Shenzhen, China
[4] Sun Yat-sen University, Guangzhou, China

Abstract. Differentiation of colorectal polyps is an important clinical examination. A computer-aided diagnosis system is required to assist accurate diagnosis from colonoscopy images. Most previous studies attempt to develop models for polyp differentiation using Narrow-Band Imaging (NBI) or other enhanced images. However, the wide range of these models' applications for clinical work has been limited by the lagging of imaging techniques. Thus, we propose a novel framework based on a teacher-student architecture for the accurate colorectal polyp classification (CPC) through directly using white-light (WL) colonoscopy images in the examination. In practice, during training, the auxiliary NBI images are utilized to train a teacher network and guide the student network to acquire richer feature representation from WL images. The feature transfer is realized by domain alignment and contrastive learning. Eventually the final student network has the ability to extract aligned features from only WL images to facilitate the CPC task. Besides, we release the first public-available paired CPC dataset containing WL-NBI pairs for the alignment training. Quantitative and qualitative evaluation indicates that the proposed method outperforms the previous methods in CPC, improving the accuracy by **5.6%** with very fast speed.

1 Introduction

Colorectal cancer (CRC) is one of the most common malignancies with a high mortality rate around the world [1]. Colorectal polyps are recognized as indicators of CRC, and they are roughly classified into two categories: hyperplastic and adenomatous [2]. Hyperplastic polyps are benign while adenomatous polyps have a high possibility of malignant transformation. Considering only the latter ones are required for surgical resection, precise differentiation is important to decrease unnecessary resection and unsuitable treatment. Colonoscopy is

Q. Wang and H. Che—Equal first authorship.

M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12907, pp. 24–32, 2021.
https://doi.org/10.1007/978-3-030-87234-2_3

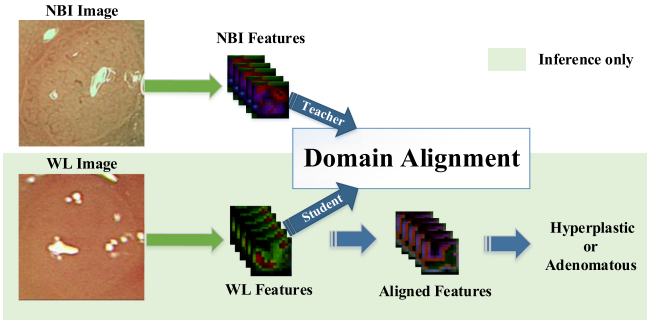

Fig. 1. The proposed teacher-student approach for polyp classification only utilizing WL images during inference. To improve the WL image based polyp differentiation accuracy, we adopt domain alignment to shift the distribution from WL features to NBI features during training, with the assistance of corresponding paired NBI images.

the preferred detection and diagnostic tool for colorectal polyps. However, due to varying illumination conditions, similar tissue representation, and occlusion, it is usually difficult to discriminate between benign and pre-cancerous polyps by conventional white-light (WL) observation, even for well-experienced endoscopists [3]. Therefore, an accurate and objective computer-aided classification system is demanded to assist clinical work.

Recent studies have achieved promising performance in colorectal polyp classification (CPC) by employing deep learning-based methods. Most works prefer to use datasets containing Narrow-Band Imaging (NBI) or Blue Light Imaging (BLI) images, owing to the enhanced visibility and superior performance [7]. For example, Usami *et al.* [11] proposed to distinguish benign/malignant polyps using WL, dye, and NBI images. In [2], authors achieved the highest accuracy of 95% by combining WL, BLI, and Linked Color Imaging (LCI) modalities.

Nevertheless, the widely used colonoscopy devices only have WL and NBI modes. Moreover, the acquisition of those advanced images is required to switch manually when polyps have been detected, while it usually suffers from missing detection in real clinic scenarios. Thus, the colorectal polyp detection using only WL endoscopy images is important but has not drawn sufficient attentions. Recently, Yang *et al.* [12] reported the classification results using WL images with the accuracy 79.5%. As shown, there is a large gap for the classification accuracy between using WL endoscopy images and enhanced images.

In this paper, we propose a novel framework as illustrated in Fig. 1 to facilitate the CPC task from WL colonoscopy images. To enhance low representative WL features, we adopt domain alignment to minimize the distance between WL and NBI feature distributions. Better feature representation in NBI images is transferred to the student network through domain alignment using adversarial learning and contrastive learning. Our main contributions are summarized in three-fold: (1) Through experiments, we prove that the CPC accuracy using

WL images is nearly 10% lower than that of using NBI images (88.9%) as input. Based on this observation, we define a new scheme that exploits NBI features to improve the WL-based classification results. (2) We propose a teacher-student model with GAN-based domain alignment and contrastive learning strategies to improve CPC. (3) We further release the first public-available polyp classification dataset named CPC-Paired, including WL-NBI image pairs. Our method achieves state-of-the-art performance (*i.e.*, $\sim 6\%$ improvement).

2 Related Work

Domain Alignment (DA). DA methods aim to align feature distributions between the source and target domains. Deep CORAL [9] defines a loss function to constrain the distance between the source and target domains in deep layer activations. In [6], correlation alignment is connected with entropy minimization to provide a solid performance. The above methods are applicable when target labels cannot be accessed. Other methods turn to minimize the difference between the source and target distributions in a shared feature space. The joint maximum mean discrepancy (JMMD) [4] is introduced to learn a transfer network by aligning the joint distributions of the network activations in domain-specific layers. Adversarial learning is adopted in domain adaptation to learn representations that the discriminator cannot distinguish between domains [10,13]. In this paper, we adopt this concept to align the features in different domains.

3 Method

3.1 Adversarial Learning for Domain Alignment

Inspired by [8], we adopt generative adversarial networks (GAN) to align the WL features with NBI features. As shown in Fig. 2, a teacher-student scheme is designed for the feature alignment. More specifically, we first pretrain a teacher feature extractor by only utilizing NBI images for CPC, where rich features can be extracted from NBI images to classify polyps. Then, we fix the teacher extractor to output NBI features X_p for aligning features X_a from student extractor. Particularly, the student extractor aims to extract features from WL images for polyp classification. However, the WL features X_a extracted from WL image are unsatisfactory for accurate polyp classification, rather than the features X_p from NBI images. Hence, to improve the classification accuracy of WL images, a discriminator D is introduced to align the WL features X_a with the rich NBI features X_p. The discriminator is optimized to distinguish between aligned WL features X_a and NBI features X_p (*i.e.*, NBI features are real and WL features are fake). As same with the GAN training manner, the discriminator D and student extractor are optimized alternatively. Therefore, the adversarial loss $\mathcal{L}_a$ supervises the student extractor to align its output with the teacher's (*i.e.*, NBI features X_p), which is shown in Eq. 1 where CE is the cross-entropy loss, Y_{nbi} indicates real label and takes 1 in practice.

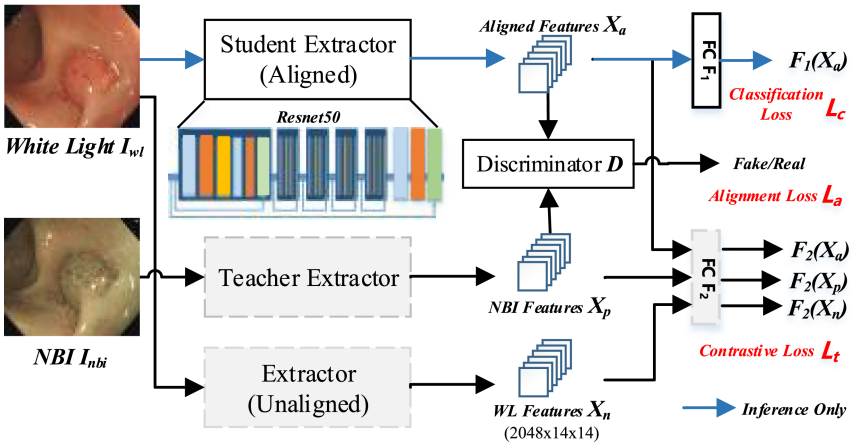

Fig. 2. The overview pipeline of the proposed method. First, we pretrain teacher and unaligned extractors by utilizing NBI images and WL images separately, which is shown in grey parts. To align the WL features X_a with more representative NBI features, an alignment loss $\mathcal{L}_a$ is designed to optimize the student extractor by introducing a discriminator for adversarial learning. Particularly, the discriminator aims to distinguish WL and NBI features. Section 3.1 illustrates details about alignment loss. Finally, the aligned accurate features X_a are fed in fully connected layer F_1 for polyp classification. Moreover, we exploit the contrastive learning loss $\mathcal{L}_t$ to shift the aligned WL features X_a much closer to NBI features X_p and far from the unaligned WL features X_n, which is introduced in details in Sect. 3.2. The blue arrow path indicates the inference phase. (Color figure online)

$$\mathcal{L}_a = CE(D(X_a), Y_{nbi}) = -log(D(X_a))Y_{nbi} \tag{1}$$

3.2 Contrastive Learning on CPC

As shown in Fig. 2, to further facilitate the model convergence and boost the performance, we design a novel contrastive loss $\mathcal{L}_t$ to take advantage of contrastive learning. More specifically, a naive unaligned feature extractor is pretrained to extract WL features X_n for CPC by only utilizing WL images as input. Then, the contrastive loss $\mathcal{L}_t$ can be formulated to supervise the student extractor to generate more representative features which are more similar to NBI features X_p and dissimilar to WL features X_n. Particularly, we take NBI features X_p as positive samples and unaligned WL features X_n as negative samples. To optimize aligned features X_a, the Kullback-Leibler (KL) divergence is adopted to constrain the distribution distance from X_a to WL features X_n (*i.e.*, negative samples) and NBI features X_p (*i.e.*, positive samples) in high-level semantic space, which is shown in Eq. 2. And F_2 is the fully connected (FC) layer to take feature maps for probability vectors generation, which is pretrained with the teacher extractor for classifying NBI images.

Algorithm 1: WL Image CPC via Domain Alignment

Input: NBI Images I_{nbi}; Paired WL Images I_{wl}; CPC Label Y; Test WL Images I_s

Pretrain $Extractor_{teacher}$ and FC F_2 by I_{nbi} only;
Pretrain $Extractor_{unaligned}$ by I_{wl} only;
//Training Phase
For I^i_{nbi}, I^i_{wl}, Y^i in $\{I_{nbi}, I_{wl}, Y\}$
  //Extract Features
  $X_p \longleftarrow Extractor_{teacher}(I^i_{nbi})$
  $X_a \longleftarrow Extractor_{student}(I^i_{wl})$
  $X_n \longleftarrow Extractor_{unaligned}(I^i_{wl})$
  //Train Discriminator D
  Minimize Loss $CE(D(X_p), 1) + CE(D(X_a), 0)$
  //Train Student Extractor
  Minimize CPC Loss $\mathcal{L}_c = CE(F_1(X_a), Y^i)$
  Fix D and Minimize Alignment Loss $\mathcal{L}_a = CE(D(X_a), 1)$
  Minimize Contrastive Loss $\mathcal{L}_t \leftarrow Triplet(X_p, X_a, X_n)$
End
//Inference Phase
$\hat{Y} \longleftarrow F_1(Extractor_{student}(I_s))$
Output: CPC Prediction $\hat{Y}$

$$KL_{margin}(F_2(X_a), F_2(X_p)) \leq KL_{margin}(F_2(X_a), F_2(X_n)) \tag{2}$$

Finally, the triplet loss $\mathcal{L}_t$ is defined for contrastive learning in Eq. 3, where μ is a hyper-parameter we set as 0.85 in practice.

$$\mathcal{L}_t = \max(KL\left(F_2(X_a), F_2(X_p)\right) - KL\left(F_2(X_a), F_2(X_n)\right) + \mu, 0) \tag{3}$$

3.3 Loss Function

The overall training loss $\mathcal{L} = \mathcal{L}_c + \mathcal{L}_a + \mathcal{L}_t$ contains three parts. First, a conventional cross-entropy loss $\mathcal{L}_c = CE(F_1(X_a), Y)$ is applied to supervise student extractor and FC layer F_1 for binary classification (*i.e.*, hyperplastic or adenomatous), where Y is the ground truth. Then, the alignment loss $\mathcal{L}_a$ and contrastive loss $\mathcal{L}_t$ are utilized to align the WL features with NBI features, which make use of GAN and contrastive learning separately. Three loss functions are optimized jointly with equal weights. The Algorithm 1 illustrates the whole training and inference procedures.

4 Experiments and Results

4.1 Implementation Details

We implement our work by PyTorch. All models were trained for 500 epochs by Adam optimizer with learning rate 10^{-3} and weight decay 10^{-8} on single Nvidia V100 GPU. We randomly split the dataset into training and validation set by ratio 8:2 for training and evaluation. The training batch size is 16. We adopt random flipping and rotation for data augmentation. Additionally, we apply 5-fold cross-validation for all experiments, which randomly generates 5-fold train-valid settings.

4.2 Dataset and Preprocessing

We conduct the experiments on our CPC-Paired dataset[1]. The paired data means each WL image has a corresponding NBI image with the same polyp label. For each modal, a total of 307 adenomatous and 116 hyperplastic images are included. Our dataset consists of two parts: collated data from ISIT-UMR Colonoscopy Dataset [5] and clinical data collected from the hospital. ISIT-UMR Colonoscopy Dataset contains 76 short video sequences with category information. For our CPC task, we choose 21 hyperplastic lesions and 40 adenomas sequences. Each lesion in the video is recorded using both NBI and WL. We extract paired frames from videos to build an available dataset. The eventual collated data covers 102 adenomatous and 63 hyperplastic images in each modal. In addition, we collected 258 WL-NBI image pairs from 123 patients consisting of 205 adenoma images and 53 hyperplastic polyp images. We further annotate the bounding box of polyps to crop the corresponding area and scale it to 448×448 as input for the CPC task.

4.3 Network Architecture

In our framework, three extractors share the same backbone design. The backbone can be popular network architectures (*e.g.*, VGG, ResNet50, Inception-V3). More specifically, each extractor is utilized to mapping the original NBI I_{nbi} or WL images I_{wl} to a high-level feature space with the shape $2048 \times 14 \times 14$ (*e.g.*, ResNet50 backbone). Finally, each extractor is followed with a single FC layer to predict the final polyp class. In the pretrain stage, extractors and FC layers are optimized jointly(*e.g.*, teacher extractor and FC layer F_2) which will be fixed during the alignment training phase. The discriminator D consists of two convolution layers and two fully connected layers which aims to distinguish aligned WL features X_a and NBI features X_p.

[1] https://drive.google.com/drive/folders/1e2t5HhQf08sTAE_CPRNVgpi6YUKgQSHn?usp=sharing.

Table 1. The comparison between our approach and the previous best method in [12]. From the comparison, we can clearly notice that our approach surpasses the previous approach with a large margin(*e.g.*, $\sim 6\%$) among all backbones. 'FOLD X' indicates the different cross-validation split settings. 'Speed' indicates the inference time per image in millisecond. The 'Mean' averages the accuracy among all split settings, which gains 5.6% improvement by our approach and exactly proves the superior performance of the proposed alignment method.

	Backbone	Speed	FOLD1	FOLD2	FOLD3	FOLD4	FOLD5	Mean
Yang [12]	VGG	20.62 ms	78.2%	79.5%	77.3%	78.0%	77.9%	78.2%
Our			79.4%	81.1%	78.5%	79.1%	80.1%	79.7%
Yang [12]	InceptionV3	30.72 ms	81.1%	82.1%	80.5%	82.0%	81.3%	81.5%
Our			84.6%	85.7%	83.1%	85.3%	84.4%	84.6%
Yang [12]	ResNet50	17.07 ms	79.5%	80.5%	78.0%	80.3%	78.8%	79.7%
Our			**85.9%**	**86.1%**	**84.2%**	**85.8%**	**85.2%**	**85.3%**
Our w/o DA	ResNet50	17.07 ms	82.6%	83.3%	81.1%	83.7%	83.3%	82.9%
Our w/o CL			83.0%	84.1%	83.5%	84.4%	84.0%	84.0%

4.4 Results

Extensive experiments are conducted to demonstrate the superior performance of the proposed approach. Particularly, by 5-fold cross comparison in Table 1, we can observe that our method outperforms previous state-of-the-art approach [12] among all folds. Specifically, we improve the accuracy on all backbones including VGG, Inception-V3 and ResNet50, which proves the generality of the proposed model. We obtain the best performance of our approach with ResNet50 backbone (*i.e.*, the highest classification accuracy 85.3%, and 5.6% improvement on the mean score compared to the previous method). The ablation study further examines the gains of each component within the proposed approach as shown in Table 1. 'Our w/o DA' indicates removal of alignment loss and 'Our w/o CL' indicates ablation of contrastive loss. The CPC accuracy degradation of the ablation study exactly proves the effectiveness of each proposed component.

The qualitative analysis is shown in Fig. 3. Particularly, we extract aligned WL features X_a, unaligned WL features X_n and NBI features X_p for comparison. Obviously, the aligned WL features are more similar to NBI features than unaligned ones, which further demonstrates the superiority of aligned features and improvement of the proposed alignment approach.

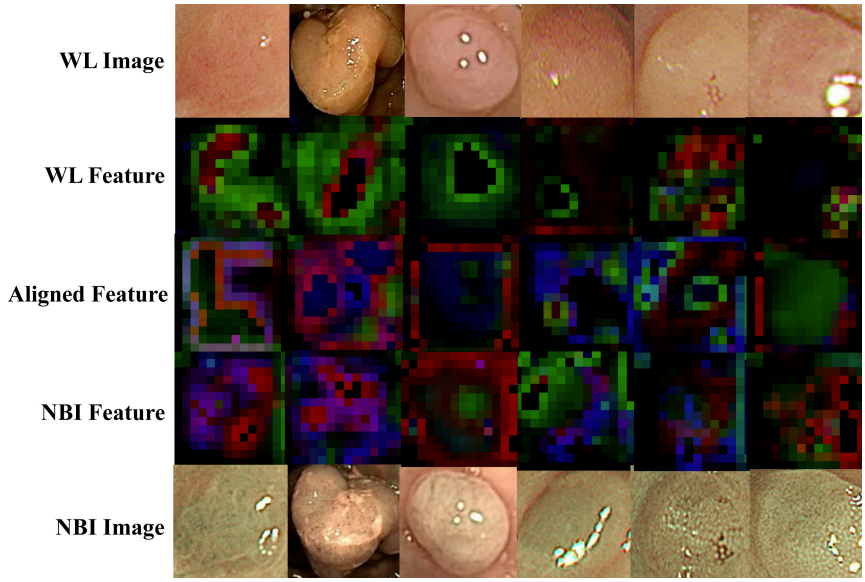

Fig. 3. The visualization comparison between WL feature, aligned feature, and NBI feature. From the comparison, we can obviously notice that the aligned feature (third row) is more similar to the NBI feature (fourth row) and less similar to the WL feature (second row), which exactly prove the effectiveness of the proposed domain alignment and contrastive learning approaches for domain shifting from WL to NBI. The aligned WL feature not only contains the original WL information but also provides more essential NBI domain representation for polyp classification.

5 Conclusion

For the purpose of investigating CPC, we release a polyp classification dataset CPC-Paired. To the best of our knowledge, this is the first public-available dataset including WL-NBI image pairs for this task. To improve the CPC accuracy of white-light (WL) images, we propose a teacher-student model for shifting the feature domain of WL images to NBI images which will be more representative for the CPC. Particularly, the novel alignment loss and contrastive loss are constructed to supervise the student model to generate more satisfactory features for the CPC. Extensive experiments consist of comparison, ablation study, and qualitative visualization, which sufficiently illustrate the effectiveness and superiority of our approach (*i.e.*, 5.6% accuracy improvement beyond the previous state-of-the-art approach on average).

Acknowledgement. The work was supported in part by Key Area R&D Program of Guangdong Province with grant No.2018B030338001, by the National Key R&D Program of China with grant No. 2018YFB1800800, by Shenzhen Outstanding Talents Training Fund, by Guangdong Research Project No. 2017ZT07X152, by NSFC-Youth 61902335, by Guangdong Regional Joint Fund-Key Projects 2019B1515120039, by The National Natural Science Foundation Fund of China (61931024), by helixon biotechnology company Fund and CCF-Tencent Open Fund.

References

1. Chen, P.J., Lin, M.C., Lai, M.J., Lin, J.C., Lu, H.H.S., Tseng, V.S.: Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology **154**(3), 568–575 (2018)
2. Fonollà, R., van der Zander, Q.E., Schreuder, R.M., Masclee, A.A., Schoon, E.J., van der Sommen, F., et al.: A CNN CADx system for multimodal classification of colorectal polyps combining WL, BLI, and LCI modalities. Appl. Sci. **10**(15), 5040 (2020)
3. Komeda, Y., et al.: Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience. Oncology **93**(Suppl. 1), 30–34 (2017)
4. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Deep transfer learning with joint adaptation networks. In: International Conference on Machine Learning, pp. 2208–2217. PMLR (2017)
5. Mesejo, P., et al.: Computer-aided classification of gastrointestinal lesions in regular colonoscopy. IEEE Trans. Med. Imaging **35**(9), 2051–2063 (2016)
6. Morerio, P., Cavazza, J., Murino, V.: Minimal-entropy correlation alignment for unsupervised deep domain adaptation. arXiv preprint arXiv:1711.10288 (2017)
7. Rondonotti, E., et al.: Blue-light imaging compared with high-definition white light for real-time histology prediction of colorectal polyps less than 1 centimeter: a prospective randomized study. Gastrointest. Endosc. **89**(3), 554–564 (2019)
8. Sankaranarayanan, S., Balaji, Y., Castillo, C.D., Chellappa, R.: Generate to adapt: Aligning domains using generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8503–8512 (2018)
9. Sun, B., Saenko, K.: Deep CORAL: correlation alignment for deep domain adaptation. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9915, pp. 443–450. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49409-8_35
10. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167–7176 (2017)
11. Usami, H., et al.: Colorectal polyp classification based on latent sharing features domain from multiple endoscopy images. Procedia Comput. Sci. **176**, 2507–2514 (2020)
12. Yang, Y.J., et al.: Automated classification of colorectal neoplasms in white-light colonoscopy images via deep learning. J. Clin. Med. **9**(5), 1593 (2020)
13. Zhang, W., Ouyang, W., Li, W., Xu, D.: Collaborative and adversarial network for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3801–3809 (2018)