Colonoscopy image classification using self-supervised visual feature learning

Le Thi Thu Hong[1], Nguyen Chi Thanh[1*], Nguyen Duc Hanh[1],
Trinh Tien Luong[1], Pham Duy Thai[2], Ngo Van Quan[3]

[1]Institute of Information Technology, AMST.
[2]Le Quy Don Technical University.
[3]Air Defense - Air Force Academy.
*****Corresponding author: thanhnc80@gmail.com.
Received 11 September 2021; Revised 12 December 2021; Accepted 15December 2021.
DOI: https://doi.org/10.54939/1859-1043.j.mst.csce5.2021.3-13

ABSTRACT

Colonoscopy image classification is an image classification task that predicts whether colonoscopy images contain polyps or not. It is an important task input for an automatic polyp detection system. Recently, deep neural networks have been widely used for colonoscopy image classification due to the automatic feature extraction with high accuracy. However, training these networks requires a large amount of manually annotated data, which is expensive to acquire and limited by the available resources of endoscopy specialists. We propose a novel method for training colonoscopy image classification networks by using self-supervised visual feature learning to overcome this challenge. We adapt image denoising as a pretext task for self-supervised visual feature learning from unlabeled colonoscopy image dataset, where noise is added to the image for input, and the original image serves as the label. We use an unlabeled colonoscopy image dataset containing 8,500 images collected from the PACS system of Hospital 103 to train the pretext network. The feature exactor of the pretext network trained in a self-supervised way is used for colonoscopy image classification. A small labeled dataset from the public colonoscopy image dataset Kvasir is used to fine-tune the classifier. Our experiments demonstrate that the proposed self-supervised learning method can achieve a high colonoscopy image classification accuracy better than the classifier trained from scratch, especially at a small training dataset. When a dataset with only annotated 200 images is used for training classifiers, the proposed method improves accuracy from 72,16% to 93,15% compared to the baseline classifier.

Keywords: Self-Supervised Visual Feature Learning; Transfer Learning; Colonoscopy Image Classification; Polyp Recognization.

1. INTRODUCTION

Colonoscopy is the primary method for colorectal cancer screening. Automatic polyp detection is highly desirable for colon screening due to the polyp miss rate by physicians during colonoscopy[1]. Colonoscopy image classification is an important task for automatic polyp detection systems. It is an image classification task that predicts whether colonoscopy images contain polyps or not. Figure 1 shows the example of colonoscopy images with polyps and normal colonoscopy without polyps.

Over the past years, deep learning has boosted medical image analysis, including polyp recognization. Training a good model based on deep learning requires a large amount of labeled data. However, it is often difficult to obtain a sufficient number of annotated colonoscopy images for training because annotation usually requires expert knowledge of endoscopists. Therefore, boosting the performance of colonoscopy image classification networks by using unlabelled and labeled data is a crucial problem.

(a)

(b)

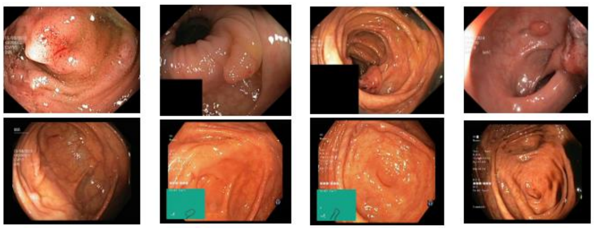

Figure 1. *The examples of colonoscopy images: (a)- Colonoscopy images with polyps, (b) - Colonoscopy images without polyps.*

Self-supervised visual feature learning [3] allows visual feature learning from large-scale unlabeled images. It is a good direction to use unlabelled images to boost the performance of the deep neural network when only limited labeled data is available. Generally, the framework that employs self-supervised feature learning involves performing two tasks, a pretext task and a real (downstream) task [3]. The real task can be any object recognition task like classification, detection, or segmentation, with insufficient annotated data samples. The pretext task is the self-supervised learning task for learning visual representations. Then, the learned representations or model weights are transferred to the downstream task.

This work proposes a novel method for training a colonoscopy images classification network, which formulates a self-supervised task for visual feature learning. Image denoising is offered for the pretext task to boost the performance of the colonoscopy images classification network. The visual features of colonoscopy images are learned by training the denoising autoencoder (DAE) [8]. We use an unlabeled colonoscopy image dataset containing 8,500 images collected from the PACS System of Hospital 103 to train the DAE. The Gaussian noise is added to the image to generate the input, and the original image serves as the label. After the self-supervised pretext model training is finished, the learned parameters serve as a pre-trained model and are transferred to the downstream model. The labeled classification colonoscopy image dataset, which contains 200 labeled images from the public Kvarsir dataset [9], was used to train the colonoscopy image classification network. We use MobileNetV2 [10] architectures as a backbone for both the encoder of the pretext network and feature extractor of the downstream network because our goal is to build a classifier that can infer results with maximizing accuracy while being used on the restricted resources for an on-device or embedded application. Our experiment shows that the proposed method improves the accuracy significantly for colonoscopy image classification compared to the baseline classification network.

The rest of the article is organized as Section 2 reviews related research. In Section 3, we describe our proposed method of self-supervised visual features learning for colonoscopy image classification in detail. Section 4 outlines our experiment settings, experimental results, and discussion. Finally, in Section 5, we summarize and conclude this work.

2. RELATED WORK

2.1. Colonoscopy image classification based on deep learning methods

Colonoscopy image classification is an important task for automatic polyp detection systems. It is an image classification task that predicts whether colonoscopy images contain polyps or not. Recently published papers dealt with polyp classification based on deep learning methods in endoscopic images [14-17]. Kim, Young Jae, et al. [14] proposed to improve the performance of polyp classification through the fine-tuning of Network-in-Network (NIN) after applying a pre-trained model of the ImageNet database. They fine-tune NIN with a dataset of 1000 labeled colonoscopy images, and all data were acquired from the children's medicine department of Gachon University Gil Hospital in South Korea. Hsu, Chen-Ming, et al. [15] proposed the work on detecting and classifying polyps on colonoscopy images. The authors used the technique of converting color endoscopic images to gray images and using deep learning networks to detect and classify polyps on images with the highest polyp classification accuracy of about 95.2%. They used the Clinic-DB set of 612 images and the dataset of 1000 endoscopic images containing polyps from Linkou Chang Gung Medical Hospital for training and evaluation. Đây là những bộ dữ liệu có gán nhãn phân vùng polyp trên ảnh nội soi đại tràng và mỗi ảnh nội soi trong bộ dữ liệu này đều chứa ít nhất một polyp. These are datasets labeled with segmented polyps, and each colonoscopy image in this dataset contains at least one polyp. Wang, Yan, et al. [16] proposed the works on classifying polyps on colonoscopy images with four different types of polyps using deep learning networks and transfer learning methods from the ImageNet natural image dataset. They used labeled datasets with polyp classification (CVC-ClinicDB containing 612 images and Kvasir- Seg containing 1000 images) to train the model and test on the dataset with 430 images. The results achieved the highest average polyp classification accuracy of 86.4%. Taha, Dima, et al. [17] proposed a study on detecting and classifying polyps on colonoscopy images. The authors used some image preprocessing techniques to extract polyp-containing areas and used deep learning networks to classify polyps with high accuracy. The highest accuracy is about 98.4%.

These methods used the deep neural network, which requires a large amount of manually annotated data for training. This is limited by the available resources of endoscopy specialists.

2.2. Self-supervised learning in the medical imaging domain

In computer vision, various types of pretext tasks have been proposed, such as Patch relative positions [4], local context prediction [5], colorization [6], and predicting image rotations [7]. Self-supervised learning has also been explored for medical imaging but to a less extent. Jamaludin et al. [11] proposed a self-supervised learning method to predict the level of vertebral bodies, e.g., classify whether two spinal MR images came from the same patient or not. They used to recognize the recognization of patients' MR scans as a pretext task and prediction the level of vertebral bodies as a real task. Tajbakhsh et al. [12] proposed a self-supervised learning method for lung lobe segmentation and nodule detection tasks by using rotation prediction as a pretext task. Ross et al. [13] proposed the method to exploit the potential of unlabeled endoscopic video data for surgical instrument segmentation. They defined the decolorization of surgical videos as a pretext task and used the pre-trained features to initialize a surgical instrument segmentation network.

3. PROPOSED METHOD

3.1. Overview of the proposed method

The overall proposed method, which adapts self-supervised visual feature learning for colonoscopy image classification, is depicted in Figure 2.

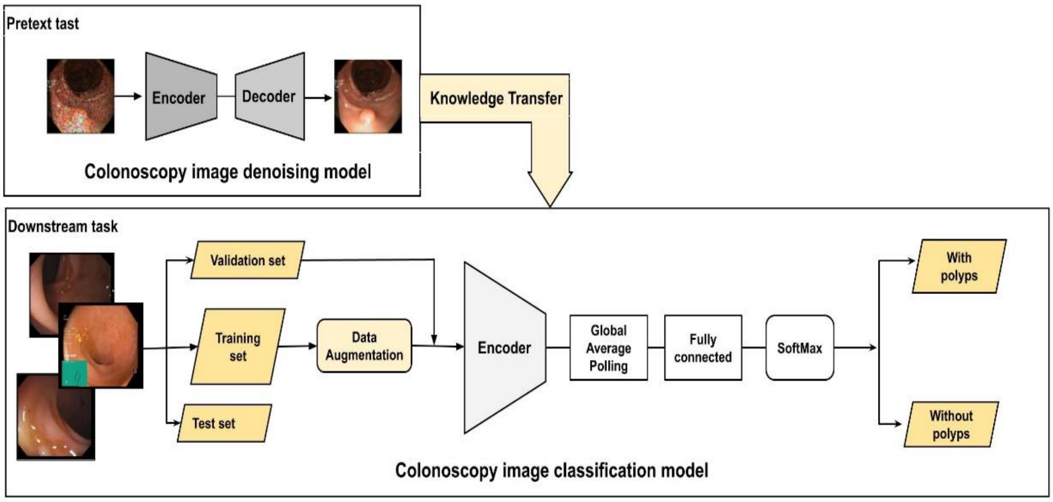

Figure 2. *Flowchart of the proposed self-supervised visual features learning method for colonoscopy image classification.*

The image denoising is used as a pretext task. We use a Denoising Autoencoder [9] (DAE) for the pretext task. DAE is a kind of unsupervised neural network, which comprises two parts encoder and decoder. Encoder defines a parameterized function to extract features, while decoder attempts to reconstruct original data from encoded features. Adding noise to the inputs forces the model to learn important features from data. The model will learn the general-purpose features that capture the salient characteristics of the colonoscopy images. Once the objective is achieved, we use the learned features to transfer the knowledge to the colonoscopy image classification downstream task. The encoder of DAE is used as a feature extractor of the colonoscopy image classifier. We add Global average poling, Fully connected, and Softmax layers on top of the DAE'encoder to generate the classifier. The classifier is fine-tuned by the colonoscopy image classification dataset to classify colonoscopy images with or without polyp.

In this study, we use MobileNetV2 [10] architectures as the encoder of DAE because MobileNetV2 is designed to effectively maximize accuracy while being mindful of the restricted resources for an on-device or embedded application. The decoder is symmetric to the encoder, and it enables precise localization using transposed convolution. MobileNetV2 is also used as a classifier for colonoscopy image classification tasks.

3.2. Denoising Autoencoder for self-supervised visual features learning

We use colonoscopy image denoising as the pretext task for self-supervised visual features learning. Denoising Autoencoder is used for image denoising tasks. The colonoscopy image is partially corrupted by noises added to the input image in a stochastic manner. Then, the model is trained to predict the original image as its output. Figure 3 presents the DAE architechture. An image x is sampled from an unlabeled colonoscopy image dataset. A corrupted version of this image is sampled from a stochastic mapping $\boldsymbol{M}(\boldsymbol{x}|\widetilde{\boldsymbol{x}})$, $(\widetilde{\boldsymbol{x}}, \boldsymbol{x})$ is used as a training example. The encoder takes in parameter the input data $\boldsymbol{x}$ and compresses it. The encoder can be considered as a function like this $\boldsymbol{E}(\widetilde{\boldsymbol{x}}) = \boldsymbol{c}$, where $\widetilde{\boldsymbol{x}}$ is the input data, *c* is the latent representation, and *E* is the encoding function. The decoder takes in parameter the latent representation c and tries to reconstruct the original input. *D(c) = x′* where *x′* is the output of the decoder and *D* is the decoding function.

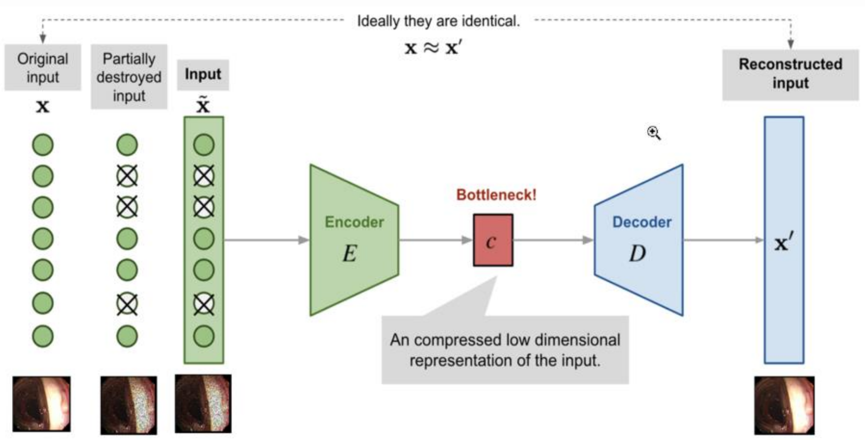

***Figure 3.** Denoising Autoenconder architechture**.*

In this study, we used Gaussian noise as the corruption mapping distribution. Gaussian noise is a statistical noise having a probability density function equal to the normal distribution, also known as Gaussian distribution. The probability density function of the normal distribution is defined as follow:

$$\boldsymbol{f}(\boldsymbol{x}) = \frac{\mathbf{1}}{\boldsymbol{\sigma}\sqrt{\mathbf{2}\boldsymbol{\pi}}}\boldsymbol{e}^{\frac{-\mathbf{1}}{\mathbf{2}}\left(\frac{\boldsymbol{x}-\boldsymbol{\mu}}{\boldsymbol{\sigma}}\right)^{\mathbf{2}}} \tag{1}$$

Where $\boldsymbol{\mu}$ is mean, $\boldsymbol{\sigma}$ is the standard deviation. The magnitude of Gaussian noise depends on the standard deviation $\boldsymbol{\sigma}$. Noise magnitude is directly proportional to the $\boldsymbol{\sigma}$ value. We also apply three methods to add Gaussian noise to an image: "noise overlaid over the image" where the Gaussian noise is added over the image, and noise is spread throughout; "noise multiplied by image" where the Gaussian noise multiplied then added over the image, noise increases with image value; "noise multiplied by bottom and top half image" where the image is folded over, and gaussian noise is multiplied and added to it, peak noise affects mid values, white and black receiving little noise. Figure 4 presents examples of transformed images with different methods and different $\boldsymbol{\sigma}$ values.

For training DAE, we use L2 loss or mean squared error (MSE) loss. L2 loss is computed as follows:

$$L_2 = \frac{1}{N}\sum_{i=1}^{N}\|y_i - \widehat{y_i}\|_2^2 \tag{2}$$

Where ***N*** is the total number of pixels in an image, *i* is an index of a pixel, y_i is the ground truth value of the pixel index ***i***, $\widehat{y_i}$ is the predicted value of the pixel index ***i***. The DAE was trained by colonoscopy images unlabeled dataset. Throughout the training phase, the goal is driving DAE to be able to learn how to reconstruct the original colonoscopy image. However, it is not the output of the decoder that is interested but rather the latent space representation. Training the DAE with an unlabeled colonoscopy image dataset will allow the encoder to find useful features in the colonoscopy image dataset. The DAE will learn the general features that capture the salient characteristics of the colonoscopy image data. The learned features are transferred to the downstream network for colonoscopy image classification.

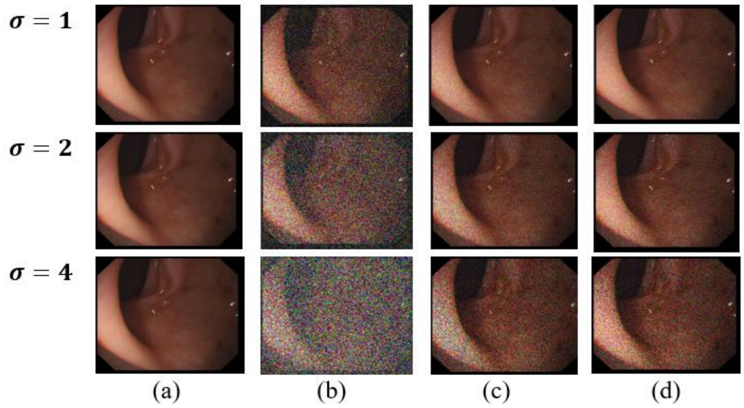

Figure 4. *Transformed images with different adding Gaussian noise methods: (a) original image, (b) noise overlaid over the image, (c) noise multiplied by image, (d) noise multiplied by bottom and top half image.*

3.3. Colonoscopy image classification using knowledge transferred from self-supervised visual features learning

After the network is self-trained on the pretext task (colonoscopy image denoising), it is transferred to the downstream task (colonoscopy image classification). By adding Global average poling, Fully connected, and Softmax layers on top of the DAE's encoder, we re-purpose the encoder of colonoscopy image denoising network for colonoscopy image classification task as a feature extractor. We investigate two ways of transfer learning for the colonoscopy image classification task. The first way is to freeze the weights learned at all layers of the feature extractor and only fine-tune the last layer on top. The second way is to fine-tune all the weights of the classifier using the learned weights of the feature extractor as initiations. To evaluate the effect of self-supervised visual features learning for colonoscopy image classification, we train the classifier from scratch and compare the received results with the result from the transfer learned classifier. For training the colonoscopy image classification network, we use a public polyp segmentation dataset, which consists of colonoscopy images and their corresponding label with polyps or without polyps. The binary cross-entropy loss (log loss) is used for training the classifier. Binary cross-entropy loss is computed as follows:

$$L_{log} = \frac{-1}{N}\sum_{i=1}^{N} y_i \cdot log\big(p(y_i)\big) + (1 - y_i)log\big(1 - p(y_i)\big) \quad (3)$$

Where y is the label (1 for image with polyps and 0 for the image without polyps) and p(y) is the predicted probability of the image with polyps for all N images.

4. EXPERIMENTS AND RESULTS

4.1. Dataset

For self-supervised visual feature learning, the unlabeled colonoscopy image dataset acquired from PACS Systems of Hospital 103 is used. The dataset includes images that were extracted from colonoscopy videos of patients who may or may not have polyps in the colon. After collecting and standardizing data, we obtained 8,500 colonoscopy images with size (384× 288).

The classifier for colonoscopy image classification is trained and evaluated on the dataset, which is selected from the Kvasir dataset [9]. Kvasir dataset, publicized by Simula Research Laboratory, consists of images annotated and verified by medical doctors (experienced endoscopists), including several classes Z-line, pylorus, cecum, polyps, ulcerative colitis. The dataset consists of the images organized in a way where they are sorted in separate folders named accordingly to the content. We selected 300 images with or without polyps from the Kvasir dataset for training, validating, and testing the classifier.

4.2. Implementation

The proposed models are implemented using Keras and Tensorflow backend. All algorithms have been programmed/trained on a PC with a GeForce GTX 1080 Ti GPU. Both pretext network and downstream network are updated via Adam optimizer, the learning rate of Adam is set to 0.0001. All the training data is divided into mini-batches for network training, and the mini-batch size is set as four during the training stage. Data augmentation was performed on the fly, including vertical flipping, horizontal flipping, random rotation, random scaling, random shearing, random Gaussian blurring, random brightness, and random cropping and padding for training colonoscopy image classifier. The model generated at the epoch with a min loss value on the validation set is the final self-supervised learning model.

4.3. Colonoscopy image denoising results

We implement DAE for colonoscopy image denoising with backbone MobileNetV2 as described in section 2.2. The unlabeled colonoscopy image dataset with 8,500 images from Hospital 103 was used for training the denoising network. We use three methods to add gaussian noise to an image: noise overlaid over image, noise multiplied by image, noise multiplied by bottom and top half image, to generate self-supervised labels with equal probability. With each method, we conduct experiments with different σ values of Gaussian noise to explore the impact of gaussian noise's magnitude on the model's performance.

Table 1. *Accuracy of image denoising with different Gaussian noise adding methods.*

Noise overlaid over image		Noise multiplied by image		Noise multiplied by bottom and top half image	
Sigma	**Accuracy**	**Sigma**	**Accuracy**	**Sigma**	**Accuracy**
1	86.15%	1	89.16%	1	93.12%
2	85.18%	2	89.45%	**2**	**94.41%**
4	84.46%	4	87.25%	4	92.65%

Table 1 presents the accuracy of image denoising with different drop σ values of Gaussian noise on the test set. This table shows that when Gaussian noise is added to the image by "Noise multiplied by bottom and top half image" with the σ values of Gaussian noise equals 2, image denoising accuracy is highest with 94.41%. Moreover, Figure 5 visualizes some examples of denoising network predictions with "Noise multiplied by bottom and top half image" and the σ values of Gaussian noise equals 2.

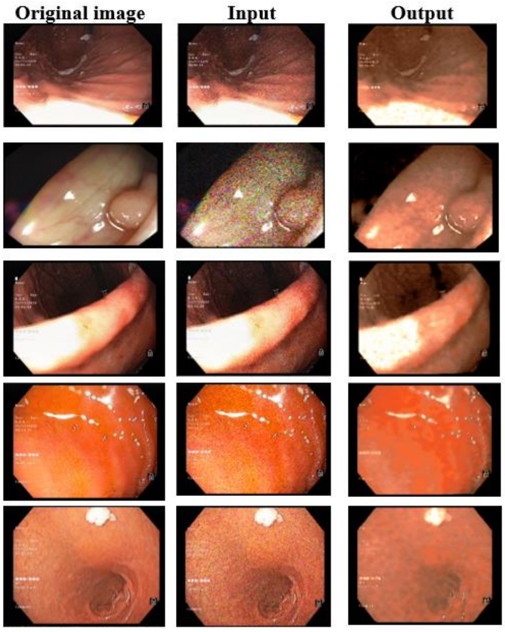

Figure 5. *Examples of denoising network prediction.*

4.4. Colonoscopy image classification results

For colonoscopy image classification, we apply transfer learning on the pretext network. We re-purpose the encoder of colonoscopy image denoising network for colonoscopy image classification task as a feature extractor. The colonoscopy image classification is fine-tuned using a small labeled dataset. First, we conducted an experiment that used transfer learning MobilnetV2 from pretext task with three methods to add Gaussian noise to image: noise overlaid over image, noise multiplied by image, noise multiplied by bottom and top half image. Each method uses different σ values of Gaussian noise to explore the impact of Gaussian noise's magnitude on classification tasks. We use a labeled dataset that consists of 300 colonoscopy images Kvasir dataset that is annotated with the label "with polyp" or "without polyp" for fine-tuning and testing the classifier. The dataset is split into 200 images for training and validation, 100 images for testing. Table 2 presents the accuracy of colonoscopy image classification on the test set. The table shows that the accuracy of colonoscopy image classification is highest when Gaussian noise is added to the image by the "Noise multiplied by bottom and top half image" method with the σ=2.

Next, we evaluated the colonoscopy image classification network's performance using transfer learning from the pretext task with "Noise multiplied by bottom and top half image" adding method and σ =2. We trained MobilnetV2 from scratch and fine-tuned MobilnetV2 with a transfer learning method for colonoscopy image classification. Table 3 compares the performance of colonoscopy image classification between the MobilnetV2 trained from scratch and self-supervised learning methods on the test set. As it shows, the accuracy of the self-supervised learning method outperforms MobilnetV2 trained from scratch with an improvement of 20.69%. Even if we freeze the encoder and only fine-tune the top layer of the classifier, we can achieve a high accuracy comparable to training a classifier from scratch. This indicates that self-supervised learning can learn good features at the encoder that are transferable for the colonoscopy image classification task.

Table 2. *Colonoscopy image classification results with different Gaussian noise adding methods.*

Noise overlaid over image		Noise multiplied by image		Noise multiplied by bottom and top half image	
Sigma	Accuracy of image classification	Sigma	Accuracy of image classification	Sigma	Accuracy of image classification
1	89.15%	1	91.53%	1	91.15%
2	87.18%	2	89.45%	**2**	**93.1%**
4	85.41%	4	88.26%	4	90.65%

Moreover, we have compared the proposed method with [14], which is proposed by Guo, Zhe, et al. In this works, they use improved the performance of polyp image classification through the fine-tuning of Network-in-Network (NIN) after applying a pre-trained model of the ImageNet database. The colonoscopy image dataset for fine-tuning includes 1000 images. The comparison is presented in Table 4. We achieve an image classification accuracy of 93.1% with the training dataset of only 200 images, while the work [14] achieves an image classification accuracy of 82.85% with a training dataset of 1000 images. This indicates that the outstanding advantage of the proposed method is that it has significantly improved the accuracy of endoscopic image classification, especially in the case where there is very little labeled data for training.

Table 3. *Colonoscopy image classification results with different transfer learning methods.*

Transfer learning method	Accuracy of image classification
Training MobileNetV2 from scratch	72.41%
Finetuning the top layer of MobileNetV2	86.21%
Finetuning all layers of MobileNetV2	**93.1%**

Table 4. *Comparison of colonoscopy image classification results with other methods.*

Method	Number of training labeled images	Accuracy of image classification
Network-in-Network [14]	1000	82,85%
Proposed method	**200**	**93.1%**

5. CONCLUSION

In this paper, we presented self-supervised visual feature learning for colonoscopy images classification to predict whether colonoscopy image contains polyps. We adapted self-supervised visual features learning with image denoising as a pretext task and images classification as a downstream task. Autoencoder with MobinetV2 as encoder and the decoder is symmetric to the encoder was used for pretext task. MobinetV2 is also used for colonoscopy images classification tasks. An unlabeled colonoscopy image dataset with 8,500 images collected from Hospital 103 is used to train the pretext network. The small dataset with 200 colonoscopy images from the Kvasir dataset that be annotated with the label "with polyp" or "without polyp" was used for fine-tuning the downstream network. We also presented a self-supervised label generation method that adds Gaussian noise to the image as input, and the original image serves as the label. We experimented with different σ values of Gaussian noise to explore the impact of gaussian noise's magnitude on classification tasks and different ways of transfer learning for the downstream task. Our experimental results show that the colonoscopy image classification accuracy of the proposed method outperforms the classifier trained from scratch. This indicates that self-supervised learning can learn good features at the encoder that are transferrable for the colonoscopy image classification task. In this work, we use Gaussian noise for self-supervised label generation. In the future, we will focus on studying other adding noises methods for self-supervised label generation to improve the performance of the proposed method. In addition, it would be interesting to extend this work for polyp detection on colonoscopy images since they are also important tasks in colonoscopy image analysis.

REFERENCES

[1]. A. M. Leufkens, M. G. H. van Oijen, F. P. Vleggaar, and P. D. Siersema. *"Factors influencing the miss rate of polyps in a back-to-back colonoscopy study,"* Endoscopy, 44(05):470475, 2012.

[2]. D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdzal, and A. Courville, *"A benchmark for endoluminal scene segmentation of colonoscopy images,"* Journal of healthcare engineering, vol. 2017, 2017.

[3]. Jing, Longlong, and Yingli Tian. *"Self-supervised visual feature learning with deep neural networks: A survey."* IEEE transactions on pattern analysis and machine intelligence (2020).

[4]. Doersch, Carl, Abhinav Gupta, and Alexei A. Efros. *"Unsupervised visual representation learning by context prediction."* Proceedings of the IEEE international conference on computer vision. 2015.

[5]. Pathak, Deepak, et al. *"Context encoders: Feature learning by inpainting."* Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[6]. Zhang, Richard, Phillip Isola, and Alexei A. Efros. *"Colorful image colorization."* European conference on computer vision. Springer, Cham, 2016.

[7]. Komodakis, Nikos, and Spyros Gidaris. *"Unsupervised representation learning by predicting image rotations."* International Conference on Learning Representations (ICLR). 2018.

[8]. Bengio, Yoshua, et al. *"Generalized denoising auto-encoders as generative models."* arXiv preprint arXiv:1305.6663 (2013).

[9]. D. Jha et al., *"Kvasir-seg: A segmented polyp dataset"* in Proc. Int. Conf. Multimedia Model., 2020, pp. 451–462.

[10]. Sandler, Mark, et al. *"Mobilenetv2: Inverted residuals and linear bottlenecks."* Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[11]. Jamaludin, Amir, Timor Kadir, and Andrew Zisserman. *"Self-supervised learning for spinal MRIs."* Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2017. 294-302.

[12]. Tajbakhsh, Nima, et al. *"Surrogate supervision for medical image analysis: Effective deep learning from limited quantities of labeled data."* 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019.

[13]. Ross, Tobias, et al. *"Exploiting the potential of unlabeled endoscopic video data with self-supervised learning."* International journal of computer assisted radiology and surgery 13.6 (2018): 925-933.

[14]. Kim, Young Jae, et al. *"New polyp image classification technique using transfer learning of network-in-network structure in endoscopic images."* Scientific Reports 11.1 (2021): 1-8.

[15]. Hsu, Chen-Ming, et al. *"Colorectal Polyp Image Detection and Classification through Grayscale Images and Deep Learning."* Sensors 21.18 (2021): 5995.

[16]. Wang, Yan, et al. *"Multiclassification of endoscopic colonoscopy images based on deep transfer learning."* Computational and Mathematical Methods in Medicine 2021 (2021).

[17]. Taha, Dima, et al. *"Automated Colorectal Polyp Classification Using Deep Neural Networks with Colonoscopy Images."* International Journal of Fuzzy Systems (2021): 1-

TÓM TẮT

PHÂN LOẠI ẢNH NỘI SOI ĐẠI TRÀNG SỬ DỤNG PHƯƠNG PHÁP TỰ HỌC CÁC ĐẶC TRƯNG THỊ GIÁC

Phân loại ảnh nội soi đại tràng là tác vụ nhằm dự đoán ảnh nội soi đại tràng có hoặc không có polyp. Đây là một tác vụ quan trọng trong hệ thống tự động phát hiện polyp trên ảnh nội soi đại tràng. Trong những năm gần đây, các phương pháp học sâu thường được sử dụng cho tác vụ này vì chúng có khả năng tự động trích rút các đặc trưng của ảnh nội soi sử dụng để phân loại ảnh với độ chính xác vượt trội. Tuy nhiên, để huấn luyện được các mô hình học sâu với độ chính xác cao cần một số lượng lớn các dữ liệu được gán nhãn thủ công bởi các chuyên gia y tế, việc này đòi hỏi chi phí cao và tốn rất nhiều công sức. Nhằm khắc phục khó khăn này, chúng tôi đề xuất một phương pháp mới để huấn luyện mạng học sâu cho phân vùng ảnh nội soi sử dụng phương pháp tự học các đặc trưng thị giác của ảnh. Chúng tôi đã dùng tác vụ khử nhiễu của ảnh (image denoising) làm tác vụ giả định (pretext task) cho hệ thống tự học các đặc trưng thị giác của ảnh nội soi đại tràng từ dữ liệu không được gán nhãn. Mạng học sâu khử nhiễu được huấn luyện bằng một bộ dữ liệu không được gán nhãn gồm 8.500 ảnh nội soi đại tràng được thu thập từ hệ thống lưu trữ và truyền dữ liệu hình ảnh (PACS) của Bệnh viện Quân y 103. Các ảnh ban đầu được thêm nhiễu để trở thành đầu vào cho mạng khử nhiễu, nhãn đầu ra chính là các ảnh ban đầu. Bộ trích xuất đặc trưng của mạng khử nhiễu sau đó được sử dụng lại cho bộ phân loại ảnh bằng phương pháp học chuyển giao. Để huấn luyện mạng phân loại ảnh nội soi, chúng tôi đã sử dụng một bộ dữ liệu có gán nhãn nhỏ được trích xuất từ bộ dữ liệu ảnh nội soi Kvarsir (là bộ dữ liệu ảnh soi chuẩn được công bố cho các mô hình học máy). Kết quả thí nghiệm cho thấy bộ phân loại được huấn luyện theo phương pháp tự học của chúng tôi đã cho độ chính xác vượt trội so với bộ phân loại cơ sở, đặc biệt với bộ dữ liệu huấn luyện nhỏ. Cụ thể với bộ dữ liệu huấn luyện chỉ gồm 200 ảnh được gán nhãn, phương pháp đề xuất đã nâng cao được độ chính xác của bộ phân loại từ 72,16% lên 93,15%.

Từ khóa: Tự học các đặc trưng hình ảnh trực quan; Học chuyển giao; Phân loại ảnh nội soi đại tràng; Nhận dạng polyp.