IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 April 2023, accepted 9 May 2023, date of publication 17 May 2023, date of current version 26 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3277029

RESEARCH ARTICLE

Classification of Polyps in Endoscopic Images Using Self-Supervised Structured Learning

QI-XIAN HUANG[1], (Member, IEEE), GUO-SHIANG LIN[2], (Member, IEEE), AND HUNG-MIN SUN[3], (Member, IEEE)

[1]Institute of Information Systems and Applications, National Tsing Hua University, Hsinchu 30013, Taiwan
[2]Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung 41170, Taiwan
[3]Department of Computer Science, National Tsing Hua University, Hsinchu 30013, Taiwan

Corresponding authors: Hung-Min Sun (hmsun@cs.nthu.edu.tw) and Guo-Shiang Lin (gslin@ncut.edu.tw)

The work of Guo-Shiang Lin was supported in part by the Chang Bing Show-Chwan Memorial Hospital Institute; and in part by the Ministry of Science and Technology, Taiwan, under Project MOST 111-2221-E-007-078-MY3 and Project 110-2221-E-007-040-MY3.

ABSTRACT This study uses a two-stage learning computer-aided diagnosis (CAD) scheme that has a convolutional neural network(CNN) with self-supervised learning(SSL) to classify polyps as either a hyperplastic polyp (HP) or a Tubular Adenoma (TA). The proposed model uses look-into-object (LIO) and contrastive learning in SimCLR to focus on the holistic polyp region and allows greater model performance. However, the LIO scheme relies on pretraining a model to provide basic representations so this model is modified using a warm-up scheme to improve the loss function. There are insufficient medical images to train efficient representation for polyp classification so another approach uses natural images, instead of polyp images, for the pretext task. The experimental results show that the proposed scheme which uses polyp object structure information and self-supervised learning produces a robust model that allows better classification as either HP or TA in the prediction head by transferring a backbone. The backbone model uses ResNet-18 effectively to concentrate on the holistic polyp using limited labeled polyp images. The proposed scheme outperforms an existing method with a 4% increase in accuracy and a 3% improvement in F1-score.

INDEX TERMS Computer-aided diagnosis, self-supervised learning, SimCLR, Polyp classification, look-into-object.

I. INTRODUCTION

Colorectal cancer accounts for a significant amount of cancer deaths in various countries, especially in Taiwan. These abnormal growths are called polyps and can become colorectal cancer [1], [2]. The current technique for detecting colorectal cancer involves stool tests and colonoscopy. Detecting and removing polyps by colonoscopy is used to prevent the spread of colorectal cancer.

White-light endoscopy images have been widely used for colorectal polyp examinations since the early 1960s [3]. In 2014, a novel image-enhanced endoscopy (IEE) approach that is called Blue Laser Imaging (BLI) was developed by the FUJIFILM Corporation (Tokyo, Japan) [4]. The BLI system uses illumination by two lasers and a white light phosphor to visually enhance surface vessels and structures [5]. BLI gives clearer visibility for the diagnosis of polyps than traditional image-enhanced endoscopy so BLI images allow a more accurate diagnosis of polyps.

There are two types of colorectal polyps: hyperplastic polyps (HP) and adenomatous polyps (Tubular Adenoma, TA). Figures 1(a) and 1(b) respectively show endoscopic images of HP and TA. A previous study [8] noted that the vascular lines on the surface of adenomatous polyps are often more obvious and irregular than those on hyperplastic polyps but it is difficult to distinguish adenomatous polyps from hyperplastic polyps using colonoscopy images and examining many endoscopic images to diagnose polyps is laborious and training endoscopists to diagnose polyps requires many years of work. A computer-aided diagnosis

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

(CAD) system to distinguish TA from HP polys decreases physicians' work-load and allows a consistent and beneficial assessment by analyzing BLI images for polyp classification.

Most recent studies of polyp classification using an endoscopic image use a supervised learning scheme [8], [9], [10], [11], [12], [49] It is difficult to obtain many medical images with annotated information and using only a small number of labelled images for training a deep network can result in overfitting.

Collection many images with annotations is expensive and time-consuming so model training using self-supervised learning (SSL) produces a more efficient method of diagnosis. SSL uses model training without the need for manual annotation and is used for computer vision and language processing. Some existing SSL-based applications are used for polyp classification. A previous study [14] used GANs to create unlabeled data sources, including photo-realistic images, and combines the SSL training and StyleGAN to learn polyp visual representations in the medical domain in the pretext task. Another study [15] used semi-supervised learning to obtain meaningful polyp representations using large quantities of unlabeled data. Another study [16] used an Auto-Encoder (AE) structure to learn useful polyp representations in the pretext task using SSL training. One study [17] used Multi-Instance Contrastive Learning (MICLe) as a generalization of existing contrastive learning [42], [43] methods for SSL to leverage multiple images per medical condition on chest X-rays and dermatology.

Previous studies show that SSL can be used to design a CAD system that requires only a small number of medical images. However, current methods do not use object structure information for feature extraction so the feature description for a foreground object must be enhanced, compared to the background. Look-into-Object (LIO) [32] allows more accurate object recognition and segmentation [52]. LIO uses two modules to extract object structure information: object extend learning (OEL), which captures the location of the holistic object, and spatial context learning (SCL), which determines the structure of the object. Using a SSL learning method with object structure information allows a better description of foreground object features.

This study proposes a polyp classification scheme that uses SSL and object structure information. The scheme learns the beneficial visual feature representations using a small number of medical images. The key findings and contributions include:

1) The proposed model uses natural images of a bird, instead of polyp medical images, and trains for classification in the pretext task if it is difficult to obtain sufficient medical images.
2) The proposed model uses the look into object scheme and SSL so the model concentrates on the holistic polyp foreground and ignores background interference using a small number unlabeled BLI medical images for pretraining in the pretext task.
3) The proposed model is robust and learns how to classify polyps as either HP or TA. Compared with existing methods [16] that use an Auto-Encoder (AE), the model performance is significantly better.

The remainder of this paper is organized as follows. Section II reviews related works on supervised learning and self-supervised learning. Section III describes the proposed method. Section IV presents the results of the experiments and compares them with those for existing methods. Conclusions are drawn in Section V.

II. RELATED WORK

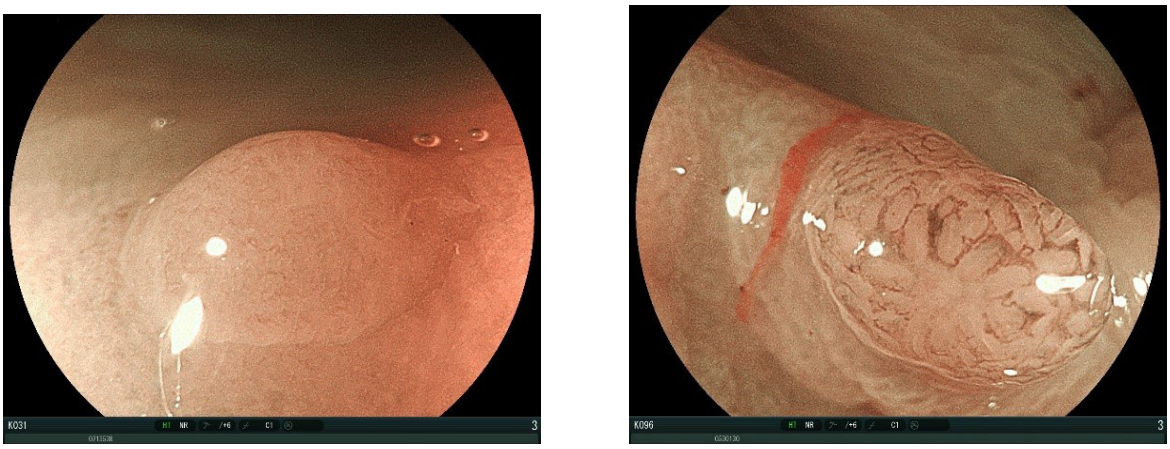
The study related to our work is recapped in this section. Our work draws on recent researches based on machine learning applied in Polyp medical fields, Supervised learning, Self-supervised learning and Model explanation.

A. SUPERVISED LEARNING

Some machine-learning (ML)-based algorithms are used to classify polyps in endoscopic images. Some methods [6], [7] use stereovision-based 3D object analysis to extract visual features using a Support Vector Machine (SVM) Classifier for Polyp detection and classification and correctly classifies polyps using a robust model.

Convolution neural networks (also known as CNNs/ConvNets) [18], [19], [20] are deep learning models and are commonly used to learn the multi-level visual features in images for image classification, object detection and semantic segmentation. Common CNNs for feature extraction are VGGNet(VGG), ResNet, GoogLeNet (also called Inception), CSPNet and DarkNet [19], [20], [21]. VGGNet features convolution layers with 3×3 filters to generate a deeper network. To reduce the effect of the gradient vanishing phenomenon, ResNet identifies a shortcut connection to give a deeper network. Classical CNN-based models are also used for other models for specific tasks. Resnet is often used as the backbone for FCN [22], U-Net [23] and DeepLab [24] for semantic segmentation. Some existing methods [8], [9], [10], [11] that use deep learning-based algorithms are used for the visual explanation, classification and detection of polyps.

One method [12] uses InceptionV3 model and transfer learning. After re-training InceptionV3 and fine-tuning a prediction head, a better classification well for polyp classification is created. One study [49] classifies colorectal polyps in BLI (Blue Laser Imaging) images using a deep neural network (DNN). Polyps are objects so a one-stage object detection network, YOLO (You Only Look Once), was used to develop a computer-aided diagnosis (CAD) system to detect and classify polyps. For a small dataset, one study [13] used a patch-based identification method for polyps that uses deep learning with a limited number of patients and used various image layers to emphasize different image information for polyp recognition and polyp classification.



(a) (b)

FIGURE 1. Endoscopic BLI images: (a) Hyperplastic Polyp (HP) and (b) Tubular Adenoma (TA).

B. SELF-SUPERVISED LEARNING

Supervised learning requires a large amount of data with label annotation and it is difficult to obtain sufficient suitable medical images. The process of collecting data that is labelled is time-consuming and expensive. Recent advances in contrastive self-supervised learning, such as instance discrimination [33], CPC [34], [35], Deep InfoMax [36], CMC [37], MoCo [38], [39], Simsiam [40], BYOL [41] and SimCLR [42], [43] are alternatives. Self-supervised learning uses unlabeled data and then fine-tunes the labeled subset in a task-specific task so self-supervised learning is used for standard image classification datasets.

It also has applications within the medical domain. One study [16] used an Auto-Encoder (AE) structure in the pretext task using SSL to learn useful polyp representations. Another study [17] used Multi-Instance Contrastive Learning (MICLe) as a generalization of existing contrastive learning [42], [43] methods to leverage multiple images per medical condition. It executes two distinct tasks: dermatological classification using digital camera images and multi-label chest X-ray classification. The results show that self-supervised learning using ImageNet, followed by more self-supervised learning using unlabeled domain-specific medical images significantly increases the accuracy of medical image classifiers. Another study [50] achieved more accurate polyp classification by fine-tuning a Network-in-Network (NIN) after applying a pre-trained model of the ImageNet database. Random shuffling is performed 20 times using 1000 colonoscopy images.

C. MODEL EXPLANATION

It is difficult to describe the prediction results for a CNN. CAM (Class Activation Mapping) [25] gives a visual explanation of a CNN but is limited by network architecture. Grad-CAM [26] shows attention heatmaps as a visual explanation for any CNN-based network architecture at that time. Other studies generate attention maps using CAM to produce a classification model that focuses on the most discriminative regions to allow the model to locate the attention map on objects [27], [28], [29], [30]. A 2018 study [31] used the model to obtain whole objects using Co-attention CNNs.

III. METHODOLOGY

This study proposes a CAD system that uses a CNN and SSL to classify polyps using BLI images. Determining the class of each polyp in a BLI image is a binary classification problem, so polyp is classified into one of two classes: HP or TA. Two issues are crucial to the design of a DL-based CAD system for the classification of HP infection.

It is necessary to determine a good feature representation to distinguish TA and HP polyps. A CNN can extract some feature maps from an endoscopic image but the measured feature maps may not be suitable for polyp classification. Usually, there are few endoscopic images for system training because the cost of annotating many medical images is great. The need for significant amounts of annotated medical data reduces the practicality of a CAD system.

To consider the above issues, an efficient CAD scheme is devised based on CNN and SSL to examine BLI images for polyp classification. The main feature of SSL is to use unlabeled images for model training [39], [42]. In fact, contrastive learning is an approach to achieve SSL and its key concept is to make similar images have similar features and let different images have very distinct features. Because of the advantages of contrastive learning, self-supervised contrastive learning can make the proposed scheme extract and learn distinguishable visual feature representations from unannotated BLI images. On the other hand, most SSL approaches seem not to consider the foreground object structure information. It is expected that a SSL approach with object structure information can extract a suitable and distinguishable feature representation from images. Then a SSL approach with object structure information is devised in the proposed CAD scheme.

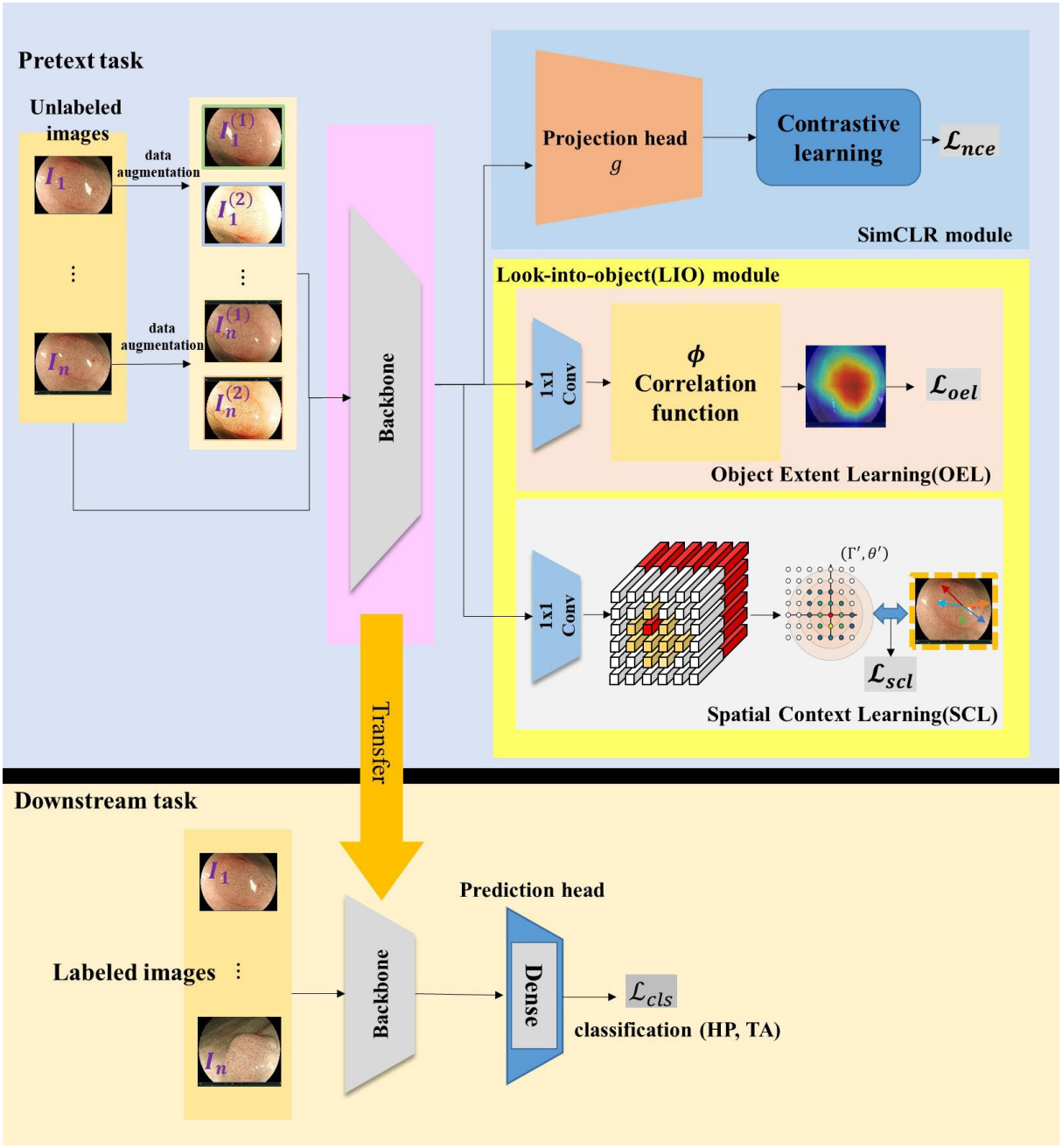


FIGURE 2. Network training for the proposed CAD scheme.

Figure 2 shows the network training for the proposed CAD scheme. Training the SSL-based CAD scheme involves two tasks: pretext and downstream. As shown in Figure 2, contrastive learning using object structure information is used in the pretext task. To allow the network to learn a distinguishable feature representation from the foreground objects, a two-stage learning strategy is used. During Stage 1 of the pretext task, contrastive learning is used to extract a feature representation from endoscopic images. In Stage 2, the object structure information is then learned using the result of Stage 1 to obtain a distinguishable feature representation from the foreground objects in images. After the pretext task, a limited number of annotated images are used to further train the proposed CAD scheme to classify polyps in the downstream task.

A. PRETEXT TASK: STAGE1

This study uses SimCLR [42], which is a self-supervised contrastive learning scheme. To learn visual representations of unlabeled polyp images, contrastive learning is used to learn representations by maximizing the agreement between differently augmented views of the same data using

contrastive loss. Contrastive loss distinguishes between similar and dissimilar data. This study uses it to distinguish HP from TA. For a batch of n polyp images $\{I_k\}_{k=1}^{n}$, each image I_k is transformed randomly into two related images, $I_k^{(1)}$ and $I_k^{(2)}$, using stochastic data augmentation. The two augmented images are entered into the backbone model f to give image representation features: $f_k^{(1)}$ and $f_k^{(2)}$. The formula is written as: $f_k^{(p)} = f(I_k^{(p)})$, the $p \in \{1, 2\}$.

As shown in Figure 3, the projection head g, is formed by a small neural network maps image representation to the vector space and then contrastive loss is applied as z_{2k-1} and z_{2k}, where $z_{2k-1} = g(f_k^{(1)})$, $z_{2k} = g(f_k^{(2)})$, and the $f_k^{(1)}$ and $f_k^{(2)}$ are the image representation features. The similarity between z_i and z_j is represented by a cosine similarity as:

$$s_{i,j} = \frac{z_i^T z_j}{\tau ||z_i|| * ||z_j||} \tag{1}$$

where τ is the temperature parameter, T is the transpose of matrix, $||z_i||$ and $||z_j||$ are the norm of z_i and z_j. We randomly sample a minibatch of N examples and define the contrastive task on pairs of augmented examples derived from the minibatch, resulting in $2N$ data points. $l_{[k\neq i]} \in \{0, 1\}$ is an indicator function evaluating to l iff $k \neq i$, $exp()$ means exponential function, and $log()$ is the logarithmic function. The contrastive loss function is called the Noise Contrastive Estimator(NCE) loss L_{nce}. The NCE loss is defined below:

$$\ell(i, j) = -log(\frac{exp(s_{i,j})}{\Sigma_{k=1}^{2N} l_{[k\neq i]} exp(s_{i,k})} \tag{2}$$

$$\mathcal{L}_{nce} = \frac{1}{2N} \sum_{k=1}^{N} [\ell(2k-1, 2k) + \ell(2k, 2k-1)] \tag{3}$$

The more similar the same images, the larger is the numerator in $\ell(2k-1, 2k)$ and the smaller is $\ell(2k-1, 2k)$. This increases the similarity between similar images and decreases the similarity between different images, in order to minimize $\mathcal{L}_{nce}$. Therefore, using $\mathcal{L}_{nce}$ for contrastive learning in the SimCLR allows the model to distinguish HP images from TA images.

B. PRETEXT TASK: STAGE2

Look-into object, (LIO) is shown in Figure 4. The LIO module produces image representations that include the object structure information to enhance the foreground of polyps. The LIO network has two modules:

- Object Extent Learning(OEL): This module learns the location of object in a given polyp image by calculating the correlation between two image feature representations.
- Spatial Context Learning(SCL): This module learns the internal structures of an object by predicting a relative polar coordinate position to strengthen the object connections between regions.

The original unlabeled image and two augmented images comprise a set of positive images. The original image I_k is denoted as $I_k^{(0)}$, which is input into the backbone model to produce the image feature for the original image $f_k^{(0)}$. The inputs for the LIO are three features $f_k^{(p)} = f(I_k^{(p)}) \in \mathbb{R}^{N\times N\times C}$, where $p \in \{0, 1, 2\}$. The LIO module then uses these three image features to learn the object structure information. In the SCL module, polar coordinates are used to represent the rotation of an object.

1) OBJECT EXTENT LEARNING

Object Extent Learning(OEL) learns the area in a specific polyp image by calculating the correlation function between two image feature vectors. The value of $(f_k^{(p)})_{i,j}$ for each feature vector pertains to a specific region in the input image $I_k^{(p)}$. If the regions are in the same category, they exhibit commonalities. Objects in an image in the same category exhibit commonality so the region-level semantic mask with a correlation between image I and I' is measured using the correlation function: $\phi(I, I') \rightarrow \mathbb{R}^{N\times N}$, where

$$\phi_{i,j}(I, I') = \frac{1}{C} \max_{1\leq i',j'\leq N} \langle f(I)_{i,j}, f(I')_{i',j'} \rangle \tag{4}$$

Maximum similarity occurs when the region in I and the region in I' are the same object, as shown in Figure 5. The ϕ correlation function discriminates between the main object area and the background in I and defines the correlation mask for I. Each polyp image can have two correlation masks from two other positive images so object localization $M_k^{(p)}$ is calculated as the average of these two correlation masks:

$$M_k^{(p)} = \frac{1}{2} \sum_{q=0}^{2} \mathbb{1}_{q\neq p} \phi(I_k^{(p)}, I_k^{(q)}) \tag{5}$$

where $M_k^{(p)}$ is the pseudo-label for Object Extent Learning, feature $f_k^{(p)}$ is processed by a 1×1 convolution to give the output $m_k^{(p)} \in \mathbb{R}^{N\times N}$ and $m_k^{(p)}$ is the prediction for the object location. The distance between the predicted object location $m_k^{(p)}$ and the pseudo label $M_k^{(p)}$ is the OEL loss, which is expressed as:

$$\mathcal{L}_{oel,k}^{(p)} = MSE(m_k^{(p)}, M_k^{(p)}) \tag{6}$$

$$\mathcal{L}_{oel} = \frac{1}{3n} \sum_{k=1}^{n} \sum_{p=0}^{2} \mathcal{L}_{oel,k}^{(p)} \tag{7}$$

This module minimizes $\mathcal{L}_{oel}$ and produces a predicted mask $m_k^{(p)}$ that is closer to the correlation mask for the pseudo label $M_k^{(p)}$, so it learns the object location information.

2) SPATIAL CONTEXT LEARNING MODULE

The Spatial Context Learning(SCL) module learns the internal structures of the object by predicting the relative polar coordinate positions using the object location $M_k^{(p)}$. The relative polar coordinates of $M_k^{(p)}$ become the pseudo label

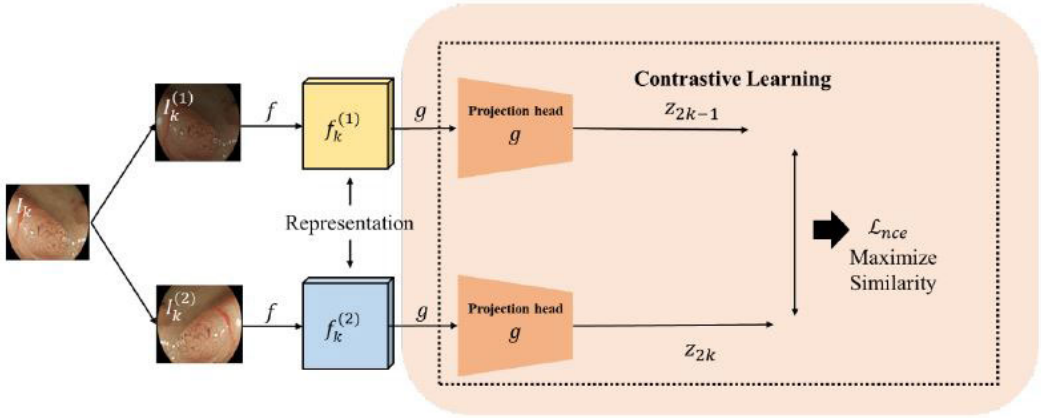


FIGURE 3. The illustration of contrastive learning.

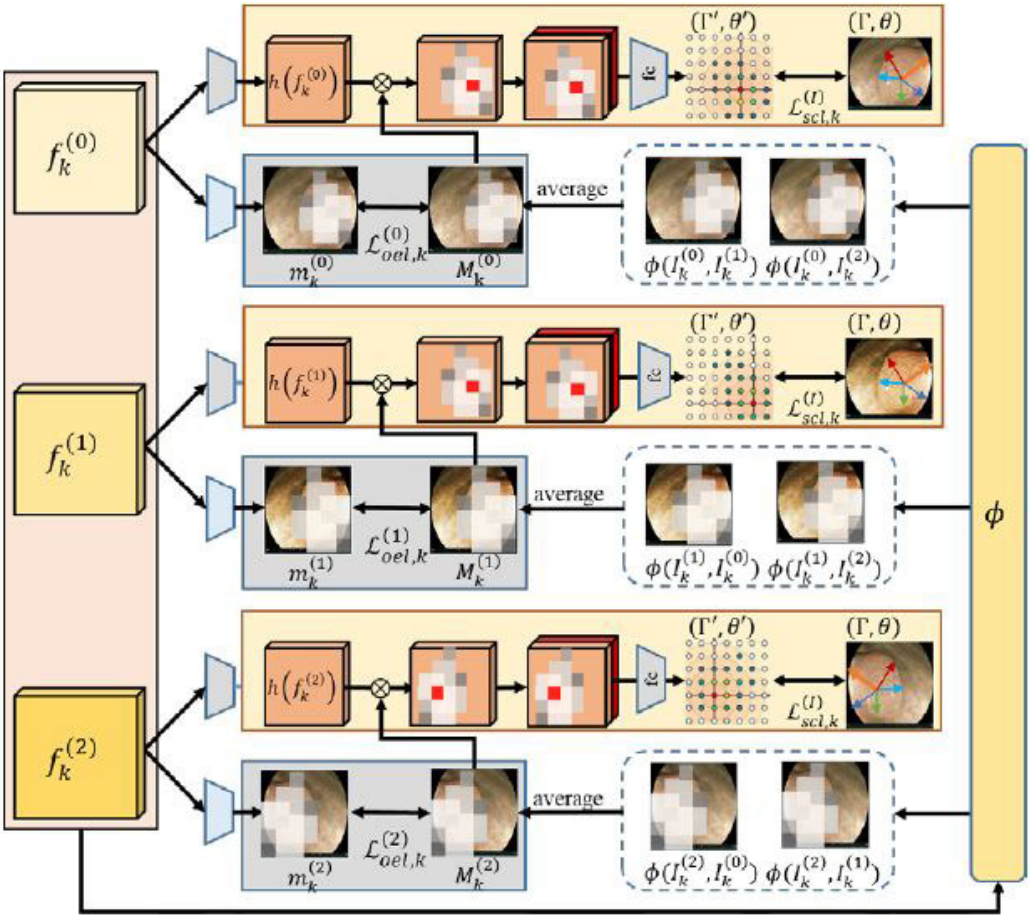


FIGURE 4. Structure of LIO module. The gray region is the Object Extent Learning Module (OEL), which learns the location of the object in a given polyp image, and the yellow region is the Spatial Context Learning Module (SCL), which learns the internal polyp structures.

for SCL. If $R_o = (x, y)$ is the origin for the polar coordinates, R_o is determined using the maximum value of $M_k^{(p)}$, where

$$R_o = (x, y) = \arg \max_{1 \le i,j \le N} (M_k^{(p)})_{i,j} \tag{8}$$

In Figure 6, the polar coordinate $(\Gamma_{i,j}, \theta_{i,j})$ for each region $R_{i,j}$ in the object location $M_k^{(p)}$ is calculated using the distance between the region and the origin R_o, and the angle $\theta_{i,j}$ between the horizontal line with the region. The region is

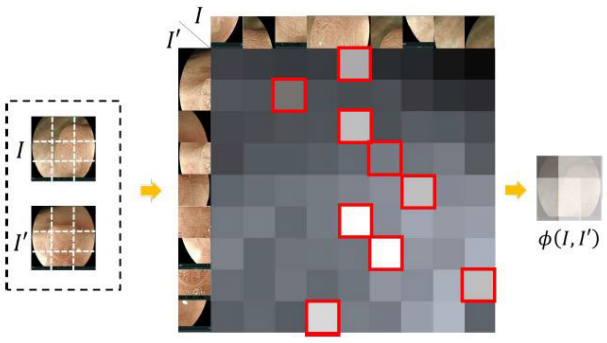


FIGURE 5. The correlation function is calculated using the region-level mask between images I and I'. Each square represents the similarity between two regions in a polyp image and the color expresses the degree of similarity 1, black represents 0 and Red squares represent the maximum value for the rows.

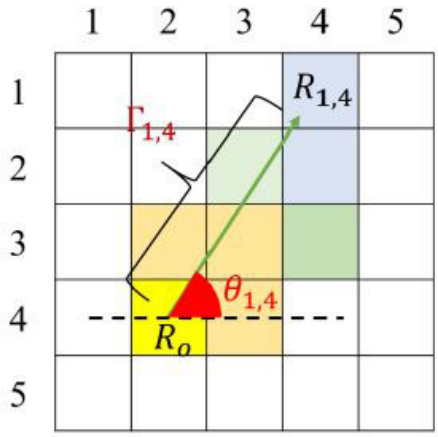


FIGURE 6. The expression for polar coordinate locations from the correlation mask $M_k^{(p)}$ in the colored grid. The angle is between line $\Gamma_{i,j}$ and the horizontal.

represented as:

$$\Gamma_{i,j} = \frac{1}{\sqrt{2}N}\sqrt{(x-i)^2 + (y-i)^2} \tag{9}$$

$$\theta_{i,j} = \frac{1}{2\pi}(\arctan(y-i, x-i) + \pi) \tag{10}$$

where $0 < \Gamma_{i,j} \leq 1$ and (Γ, θ) is the pseudo label for SCL. The SCL module estimates polar coordinates (Γ', θ') that match (Γ, θ). The SCL module inputs this feature structure representation into $h(f_k^{(p)})$ using a 1×1 convolution h, and then $h(f_k^{(p)})$ is multiplied by $M_k^{(p)}$ to extract the structure information. This structure information predicts the location of the polar coordinate and uses channel-wise concatenation for each region $h(f_k^{(p)})_{i,j}$ and the origin $h(f_k^{(p)})_{x,y}$. The region $R_{i,j}$ and the origin R_o are both used to predict the polar coordinates and the concatenated features act as a fully-connected layer to calculate the predicted polar coordinates (Γ', θ').

The two losses between (Γ, θ) and (Γ', θ') are used to compute the distance loss $\mathcal{L}_{dis,k}^{(p)}$ and angle loss $\mathcal{L}_{\angle,k}^{(p)}$ defined as:

$$\mathcal{L}_{dis,k}^{(p)} = \frac{1}{N^2}\sum_{1\leq i,j\leq N}(M_k^{(p)})_{i,j}(\Gamma'_{i,j} - \Gamma_{i,j})^2 \tag{11}$$

$$\mathcal{L}_{\angle,k}^{(p)} = \frac{1}{N^2}\sum_{1\leq i,j\leq N}(M_k^{(p)})_{i,j}(\theta_{\Delta_{i,j}} - \bar{\theta}_\Delta)^2 \tag{12}$$

where

$$\theta_{\Delta_{i,j}} = \begin{cases} \theta'_{i,j} - \theta_{i,j}, & \text{if } \theta_{i,j} \leq \theta'_{i,j} \\ 1 + \theta'_{i,j} - \theta_{i,j}, & \text{otherwise,} \end{cases} \qquad \bar{\theta}_\Delta = mean(\theta_\Delta) \tag{13}$$

The structure information for an object must be rotation-invariant, so the differences between predicted polar angles and pseudo polar angels are be the same. Therefore, $\mathcal{L}_{\angle,k}^{(p)}$ calculates the distance for the difference and the mean value. The variance for these two losses are multiplied by $M_k^{(p)}$. The more $M_k^{(p)}$ considers the place, the greater is its value, and the SCL module concentrates more on the structure inside $M_k^{(p)}$. The SCL loss $\mathcal{L}_{scl}$ sums all $\mathcal{L}_{dis,k}^{(p)}$ and $\mathcal{L}_{\angle,k}^{(p)}$:

$$\mathcal{L}_{scl,k}^{(p)} = \mathcal{L}_{dis,k}^{(p)} + \mathcal{L}_{\angle,k}^{(p)} \tag{14}$$

$$\mathcal{L}_{scl} = \frac{1}{3n}\sum_{k=1}^{n}\sum_{p=0}^{2}\mathcal{L}_{scl,k}^{(p)} \tag{15}$$

C. PRETEXT TASK WITH WARM-UP

The contrastive learning for the SimCLR module uses the value of $\mathcal{L}_{nce}$ to increase the accuracy with which polyps are classified as HP or TA and the LIO module learns the polyp structure information inside so the proposed network is trained by minimizing the total loss:

$$\mathcal{L} = \mathcal{L}_{nce} + \lambda(\mathcal{L}_{oel} + \mathcal{L}_{scl}) \tag{16}$$

where λ are the respective loss coefficients for the SCL and OEL modules. However, the weight of the backbone model is initialized randomly but the image features that are output by the backbone are random and meaningless, so the LIO module does not learn well and produce useful images features. A backbone with weights that are pretrained by the ImageNet classification task is improved by LIO, so the LIO module must have a pretrained model.

To address this dependence on a pretrained model, the loss coefficient λ is a 'warm-up', as shown in 7. A total of T epochs are trained as and the current epoch is t. The loss coefficient is warmed up by closing loss coefficients for T_{nce} epochs and increasing linearly to the original value after T_{nce} epochs:

$$\lambda(t) = \begin{cases} 0, & \text{if } t \leq T_{nce} \\ \dfrac{t - T_{nce}}{T - T_{nce}}\lambda_{max}, & \text{otherwise} \end{cases} \tag{17}$$

Therefore, the loss relies on epoch t in equation (16), which is modified as:

$$\mathcal{L} = \mathcal{L}_{nce} + \lambda(t)(\mathcal{L}_{oel} + \mathcal{L}_{scl}) \tag{18}$$

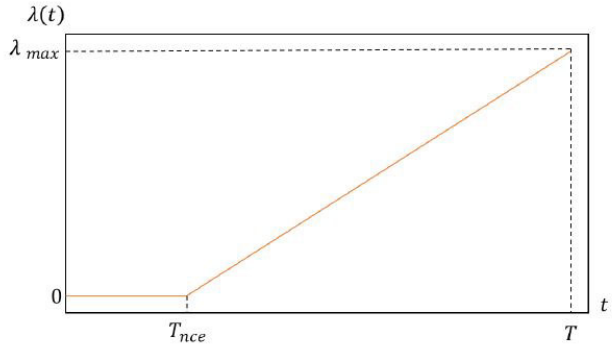


FIGURE 7. The loss coefficients warm-up scheme for the LIO using a linear increases. The x-axis represent the current epoch, the y-axis represents the value for the LIO loss coefficient and $\lambda_{max} = 0.1$.

The SCL and OEL loss coefficients are a function of time, so $T_{nce} = 100$ and $\lambda_{max} = 0.1$ for this study.

D. DOWNSTREAM TASK

The LIO an OEL modules are light-weight. After pretext task has finished training, the backbone that contains the foreground object for polyp structure information is derived and is not affected by background noise. The target model is fine-tuned using the backbone from the pretext task. The projection head g is replaced by the prediction head for polyp classification well in the downstream task. The classification loss function is written as

$$\mathcal{L}_{cls} = -\sum_{I \in \mathcal{I}} y^{(*)}(I) \cdot \log y(I) \tag{19}$$

where $\mathcal{I}$ is the image set for training, $y^{(*)}(I)$ is the ground truth and $y(I)$ is the probability vector from the classification network. The weights of the backbone and the prediction head are used for back-propagation.

IV. EXPERIMENTAL RESULT

In the experiments, ResNet-18 is used as the backbone model to classify polyp images as HP or TA. A small number of BLI images from the Cooperative Institute of Chang Bing Show Chwan Memorial Hospital are used. The BLI images were scaled from the original 1280 × 1024 pixels to 224 × 224 pixels. The initial weight of the model is initialized randomly. This model is implemented using a PyTorch Framework and trained on a NVIDIA GeForce RTX 3080Ti GPU with a 12GB memory.

During the pretext task, the dimension of the output of the projection head is 128. In the downstream task, the prediction head is the polyp classifier (one fully connected layer) behind the backbone Resnet-18 and the dimension of its output is equal to the number of classes in the dataset. The hyper-parameters for the pretext task are as shown below:

- Optimizer: SGD optimizer with momentum value 0.9 and weight decay 0.0005.
- Learning Learning Rate Scheduler: cosine decay schedule with the initial learning rate 0.0001. Initially, the linear warm-up is 100 epochs, and λ_{max} is 0.1.
- Batch size: 32
- Training epochs:512

In the downstream task, the training epoch is 128 and the learning rate is 0.005. The values of other hyper-parameters are the same as those for the pretext task.

Two types of datasets are used for training in the pretext task: 113 Polyp BLI medical images and natural images of birds, called CUB-200-2011 [48], which contains 11,788 images of 200 classes of birds. There are 5994 unlabeled bird images for training and 18 Polyp BLI images for testing. Data is augmented for both the pretext task and the downstream task to prevent overfitting. Six pre-processing stages are used for data augmentation:

1) Rotate the image by a random angle with angular range of $(-15°, 15°)$.
2) Resize using a random cropping location to give an output image of 224 × 224.
3) Adjust the Color jitter using the brightness, contrast, saturation and hue of an image.
4) Use a Gaussian Blur with a kernel size that is 10% of the image resolution randomly.
5) Randomly flip the image horizontally.
6) Convert the image to gray scale with a probability value of 0.2

A. PERFORMANCE INDEXES

To ensure an objective evaluation, performance indices are used to measure the performance in terms of image classification. Recall and precision are used for image retrieval, object detection and image segmentation. The recall rate is the ratio of correct predictions to the total number of images in the specific class. The precision rate is the ratio of correct predictions to the total number of detected images in a specific class. The F1-score is the weighted mean of the recall and precision for a specific class. The performances are defined as:

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{22}$$

where TP, FP, TN and FN are a true positive, a false positive, a true negative or a false negative for classifying the input into a specific class. A high recall and precision rate indicates good performance. The accuracy rate for the proposed scheme is calculated as:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{23}$$

the greater the value for the accuracy rate, the better the scheme performs.

TABLE 1. Precision, Recall and F1-socre rates for different schemes using unlabeled natural images of birds for the pretext task.

Pretext task cases(Natural images of bird)	Accuracy	Precision	Recall	F1-score
Stage1	0.828	0.828	0.833	0.828
Stage2: w/o warm-up	0.794	0.758	0.878	0.812
Stage2: w/ warm-up	0.861	0.860	0.867	0.862

TABLE 2. Precision, Recall and F1-socre rates for different schemes using unlabeled Polyp medical BLI images for the pretext task.

Pretext task cases(Polyp medical BLI images)	Accuracy	Precision	Recall	F1-score
Stage1	0.829	0.832	0.855	0.840
Stage2: w/o warm-up	0.811	0.816	0.800	0.805
Stage2: w/ warm-up	0.872	0.859	0.900	0.876

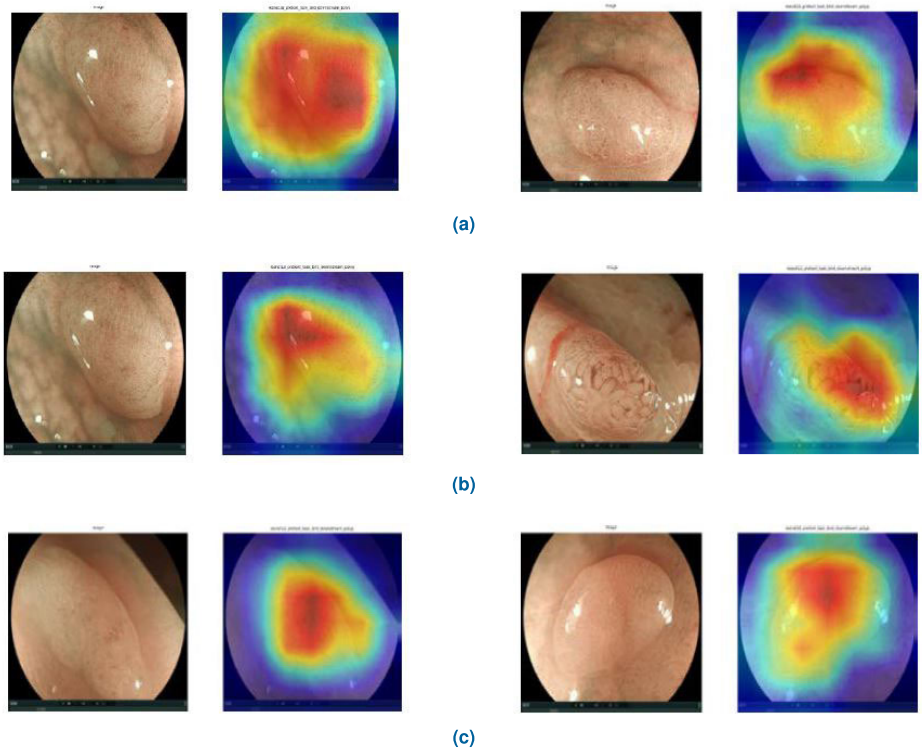


FIGURE 8. Heatmaps for three different schemes using unlabeled natural images of birds for training in the pretext task: (a) Stage1, (b) Stage2: w/o warm-up and (c) Stage2: w/ warm-up.

TABLE 3. 5-fold cross validation for Stage2: w/ warm-up scheme(K=5).

Fold K	Accuracy	Precision	Recall	F1-score
Fold 1	0.833	0.855	0.828	0.834
Fold 2	0.861	0.842	0.944	0.865
Fold 3	0.889	0.888	0.873	0.892
Fold 4	0.853	0.813	0.893	0.863
Fold 5	0.917	0.895	0.944	0.919
Mean value	**0.871**	**0.859**	**0.896**	**0.875**

B. ABLATION ANALYSIS

The performance is evaluated for three different schemes:

1) **Stage1 for contrastive learning**,
2) **Stage2: w/o warm-up**,
3) **Stage2: w/ warm-up**.

The LIO module learns well and captures useful polyp image features using foreground structure information and reduces background interference using a warm-up scheme that requires a pretrained model. T_{nce} is 100 so 100 epochs are required to train a suitable total loss. Stage2: w/ warm-up of the self-supervised learning learns the basic representation and the LIO module learns the polyp structure information correlation.

Stage2: w/ warm-up increases the model's performance. Two types of datasets are used for the pretext task, unlabeled natural images of birds and unlabeled Polyp BLI images.

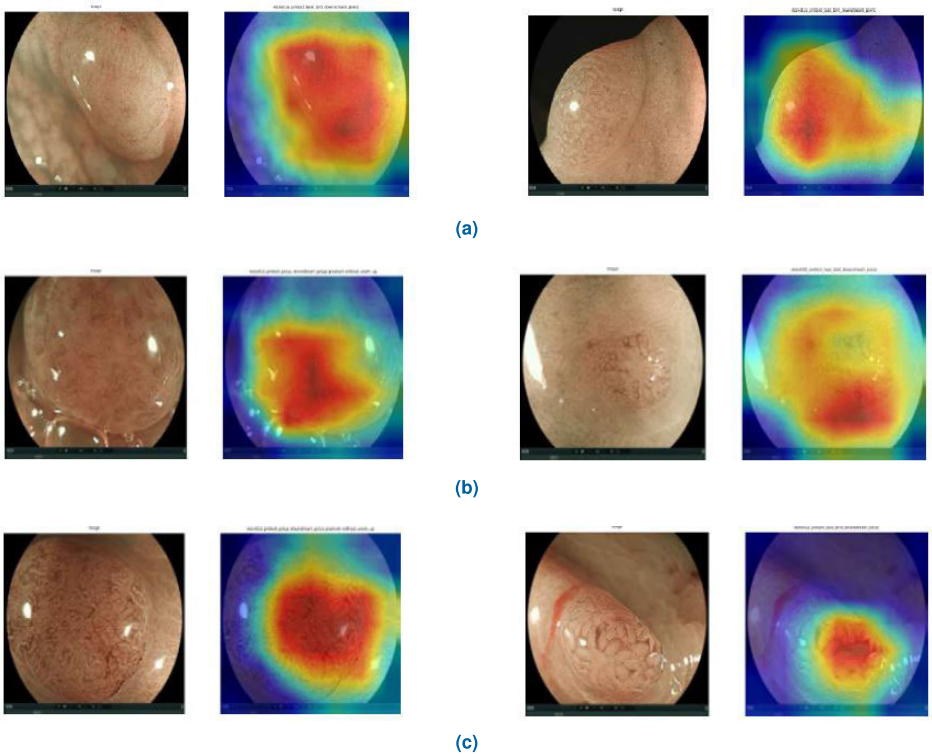


FIGURE 9. Heatmaps of three different schemes using unlabeled Polyp medical BLI images for training in the pretext task: (a) Stage1, (b) Stage2: w/o warm-up and (c) Stage2: w/ warm-up.

TABLE 4. Comparison of the Proposed scheme and an existing method.

Model scheme	Accuracy	Precision	Recall	F1-score
Proposed	0.871	0.859	0.896	0.875
[16]	0.861	0.817	0.833	0.865

The two target models in the downstream task use labeled polyp BLI images for testing, as shown in Table 1 and Table 2 for the three different schemes. The average F1-scores in Table 1 are 0.828, 0.812, and 0.862 for Stage1, Stage2: w/o warm-up, and Stage2: w/ warm-up, respectively. The proposed scheme in Stage2: w/ warm-up gives similar results to Stage1. The increments for the F1-score are 3%. For Stage2: w/o warm-up, the increments are 5%. The accuracy is 3% better than for Stage1 and 6% better than for Stage2: w/o warm-up. In Table 2, using the Polyp BLI medical images for the pretext task, the F1-score for Stage2: w/ warm-up is 0.876, which is 4% better than for Stage1, and it is 3% more accurate. Stage2: w/ warm-up is used as the backbone network for the proposed scheme for the pretext task.

1) SUBJECTIVE EVALUATION

To determine the effect of feature extraction on the proposed scheme, a heatmap is used to show where the patterns in the BLI images are useful for classification as HP or TA. In the heatmaps, the dark red and the dark blue areas represent the highest and lowest attention levels, respectively. As shown in Figure 8(a), the feature maps in the deeper layer contain more semantic information in Stage1. To increase the structure information on polyp foreground, Stage2: w/ warm-up is added to the model for commonality, as shown in Figure 8(c), and the Stage2: w/o warm-up results are shown in 8(b). Figures 8(a) to (c) show that the results for heatmaps that are trained using natural images of birds for the pretext task and Figures 9(a) to (c) show the results when they are trained using BLI images. The results show that if medical images are difficult to obtain, natural images such as ImageNet can be used for training, instead of medical images.

C. PERFORMANCE ANALYSIS

An objective evaluation is used for performance analysis.

1) OBJECTIVE EVALUATION

In order to validate the proposed scheme, K-fold cross validation is used for performance evaluation where K has a value of 5.

Table 3 shows the 5-fold results for the proposed scheme using Stage2: w/ warm-up for classification of polyps as HP or TA. The F1-score rate and the accuracy are at least 83% so the experimental results show that the proposed CAD system

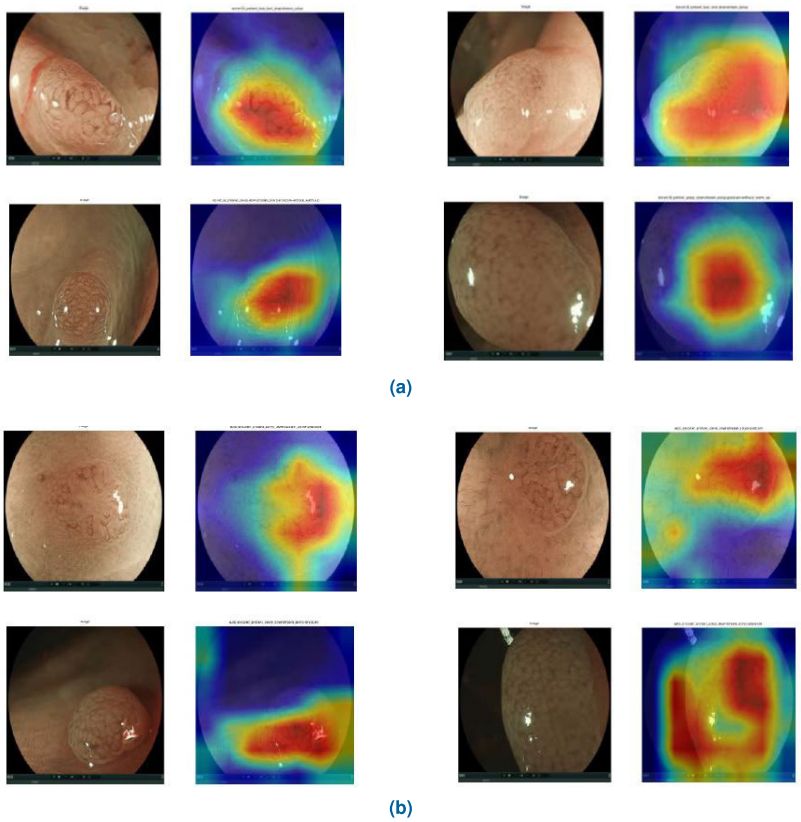


FIGURE 10. The Heatmaps of two schemes between proposed and [16] by using unlabeled Polyp medical BLI images for training in pretext task: (a) with Stage2: w/ warm-up and (b) with Auto-Encoder(AE).

that uses Stage1 and Stage2: w/ warm-up and the prediction head for the downstream task distinguishes polyps as either HP or TA images. The results also show that the proposed model is robust for BLI images.

D. COMPARISON WITH EXISTING METHOD

Current methods for polyp classification are described in Section II. Compared with an existing approach that uses an Auto-Encoder(AE) architecture for the pretext task from [16], as shown in Table 4, the proposed approach better distinguishes between HP and TA because the structure information is used, instead of whole images with background interference. The proposed scheme using SSL focuses on the foreground of polyp images. The warm-up scheme increases the accuracy by 1%, and the F1-score by 1%, the Precision and recall rates are superior to those for existing methods. The experimental results demonstrate that the proposed scheme extracts useful representations of foreground structure from BLI images for polyp classification. The heatmaps for subjective evaluation are shown in Figure 10.

V. CONCLUSION

This study uses a two-stage learning computer-aided diagnosis (CAD) scheme with a deep convolution neural network (CNN) and self-supervised learning (SSL) to classify polyps as either HP or TA. SimCLR using contrastive learning and Look-into object (LIO) using OEL and SCL modules are used to train the project head to concentrate on the foreground of the polyp structure and increase the model performance. If medical datasets are difficult to obtain, natural images from ImageNet can be used to train the projection head to classify different classes in the pretext task. The dataset of birds is used. This paper addresses the problem of insufficient medical images. Using these schemes, the prediction head classifies polyps as either HP or TA. In comparison to the existing auto-encoder (AE) method. The proposed model is also more accurate and has a higher F1-score.

REFERENCES

[1] T. Hamoudah, K. C. Vemulapalli, M. Alsayid, J. Van, K. Ma, S. Jakate, D. K. Rex, and J. Melson, "Risk of total metachronous advanced neoplasia in patients with both small tubular adenomas and serrated polyps," *Gastrointestinal Endoscopy*, vol. 96, no. 1, pp. 95–100, Jul. 2022, doi: 10.1016/j.gie.2022.02.015.

[2] Y. Dai, W. Chen, X. Xu, J. Chen, W. Mo, Y. Chen, and S. Xu, "Factors affecting adenoma risk level in patients with intestinal polyp and association analysis," *J. Healthcare Eng.*, vol. 2022, pp. 1–5, Jan. 2022, doi: 10.1155/2022/9479563.

[3] W. I. Wolff and H. Shinya, "Polypectomy via the fiberoptic colonoscope: Removal of neoplasms beyond reach of the sigmoidoscope," *New England J. Med.*, vol. 288, no. 7, pp. 329–332, Feb. 1973, doi: 10.1056/NEJM197302152880701.

[4] N. Yoshida, N. Yagi, Y. Inada, M. Kugai, T. Okayama, K. Kamada, K. Katada, K. Uchiyama, T. Ishikawa, O. Handa, T. Takagi, H. Konishi, S. Kokura, A. Yanagisawa, and Y. Naito, "Ability of a novel blue laser imaging system for the diagnosis of colorectal polyps," *Digestive Endoscopy, Off. J. Japan Gastroenterol. Endoscopy Soc.*, vol. 26, no. 2, pp. 250–258, Mar. 2014, doi: 10.1111/den.12127.

[5] K. Kaneko, Y. Oono, T. Yano, H. Ikematsu, T. Odagaki, Y. Yoda, A. Yagishita, A. Sato, and S. Nomura, "Effect of novel bright image enhanced endoscopy using blue laser imaging (BLI)," *Endoscopy Int. Open*, vol. 2, no. 4, pp. E212–E219, Oct. 2014.

[6] J. Ayoub, B. Granado, Y. Mhanna, and O. Romain, "SVM based colon polyps classifier in a wireless active stereo endoscope," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol.*, Aug. 2010, pp. 5585–5588, doi: 10.1109/IEMBS.2010.5626790.

[7] A. H. Chang and B. M. Case, "Attacks on image encryption schemes for privacy-preserving deep neural networks," 2020, *arXiv:2004.13263*.

[8] E. H. Jin, D. Lee, J. H. Bae, H. Y. Kang, M.-S. Kwak, J. Y. Seo, J. I. Yang, S. Y. Yang, S. H. Lim, J. Y. Yim, J. H. Lim, G. E. Chung, S. J. Chung, J. M. Choi, Y. M. Han, S. J. Kang, J. Lee, H. Chan Kim, and J. S. Kim, "Improved accuracy in optical diagnosis of colorectal polyps using convolutional neural networks with visual explanations," *Gastroenterology*, vol. 158, no. 8, pp. 2169–2179, Jun. 2020, doi: 10.1053/j.gastro.2020.02.036.

[9] R. Kader, A. V. Hadjinicolaou, F. Georgiades, D. Stoyanov, and L. B. Lovat, "Optical diagnosis of colorectal polyps using convolutional neural networks," *World J. Gastroenterol.*, vol. 27, no. 35, pp. 5908–5918, Sep. 2021, doi: 10.3748/wjg.v27.i35.5908.

[10] Y. Ma, Y. Li, J. Yao, B. Chen, J. Deng, and X. Yang, "Polyp location in colonoscopy based on deep learning," in *Proc. 8th Int. Symp. Next Gener. Electron. (ISNE)*, Oct. 2019, pp. 1–3, doi: 10.1109/ISNE.2019.8896576.

[11] X. Liu, Y. Li, J. Yao, B. Chen, J. Song, and X. Yang, "Classification of polyps and adenomas using deep learning model in screening colonoscopy," in *Proc. 8th Int. Symp. Next Gener. Electron. (ISNE)*, 2019, pp. 1–3, doi: 10.1109/ISNE.2019.8896649.

[12] I. Sima and K. Cincar, "Transfer learning-based classification of gastrointestinal polyps," in *Proc. IEEE 21st Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2021, pp. 1–4, doi: 10.1109/BIBE52308.2021.9635497.

[13] Y. Wang, X. Chen, H. Cai, and Z. Liang, "Effect of various image information in polyp classification by deeping learning with small dataset," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP)*, Dec. 2019, pp. 1–4, doi: 10.1109/ICSIDP47821.2019.9173377.

[14] J.-Y. Kim, G. Tangriberganov, W. Jung, D. S. Kim, H. S. Koo, S. Lee, and S. M. Kim, "Effective representation learning via the integrated self-supervised pre-training models of StyleGAN2-ADA and DINO for colonoscopy images," *BioRxiv*, 2022, doi: 10.1101/2022.06.15.496360.

[15] M. Golhar, T. L. Bobrow, M. P. Khoshknab, S. Jit, S. Ngamruengphong, and N. J. Durr, "Improving colonoscopy lesion classification using semi-supervised deep learning," *IEEE Access*, vol. 9, pp. 631–640, 2021, doi: 10.1109/ACCESS.2020.3047544.

[16] N. C. Thanh, "Colonoscopy image classification using self-supervised visual feature learning," *J. Mil. Sci. Technol.*, no. CSCE5, pp. 3–13, Dec. 2021, doi: 10.54939/1859-1043.j.mst.CSCE5.2021.3-13.

[17] S. Azizi, B. Mustafa, F. Ryan, Z. Beaver, J. Freyberg, J. Deaton, A. Loh, A. Karthikesalingam, S. Kornblith, T. Chen, V. Natarajan, and M. Norouzi, "Big self-supervised models advance medical image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3458–3468, doi: 10.1109/ICCV48922.2021.00346.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, 2012, pp. 1097–1105.

[20] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[24] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[25] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.

[26] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[27] Y. Wei, J. Feng, X. Liang, M.-M. Cheng, Y. Zhao, and S. Yan, "Object region mining with adversarial erasing: A simple classification to semantic segmentation approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1568–1576.

[28] F. Zhang, C. Gu, C. Zhang, and Y. Dai, "Complementary patch for weakly supervised semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7242–7251.

[29] Y. Wang, J. Zhang, M. Kan, S. Shan, and X. Chen, "Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12275–12284.

[30] Y. Su, G. Lin, Y. Hao, Y. Cao, W. Wang, and Q. Wu, "Self-supervised object localization with joint graph partition," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 2, pp. 2289–2297.

[31] K. J. Hsu, Y. Y. Lin, and Y. Y. Chuang, "Co-attention CNNs for unsupervised object co-segmentation," in *Proc. IJCAI*, vol. 1, Jul. 2018, p. 2.

[32] M. Zhou, Y. Bai, W. Zhang, T. Zhao, and T. Mei, "Look-into-object: Self-supervised structure modeling for object recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11774–11783.

[33] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3733–3742.

[34] O. J. Hénaff, A. Srinivas, J. De Fauw, A. Razavi, C. Doersch, S. M. A. Eslami, and A. van den Oord, "Data-efficient image recognition with contrastive predictive coding," 2019, *arXiv:1905.09272*.

[35] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[36] R. D. Hjelm, A. Fedorov, S. Lavoie-Marchildon, K. Grewal, P. Bachman, A. Trischler, and Y. Bengio, "Learning deep representations by mutual information estimation and maximization," 2019, *arXiv:1808.06670*.

[37] Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," 2019, *arXiv:1906.05849*.

[38] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," 2020, *arXiv:2003.04297*.

[39] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9729–9738.

[40] X. Chen and K. He, "Exploring simple Siamese representation learning," 2020, *arXiv:2011.10566*.

[41] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. H. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Daniel Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos, and M. Valko, "Bootstrap your own latent: A new approach to self-supervised learning," 2020, *arXiv:2006.07733*.

[42] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020, *arXiv:2002.05709*.

[43] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," 2020, *arXiv:2006.10029*.

[44] S. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 6707–6717.

[45] A. Srinivas, M. Laskin, and P. Abbeel, "CURL: Contrastive unsupervised representations for reinforcement learning," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 5639–5650.

[46] H. Hafidi, M. Ghogho, P. Ciblat, and A. Swami, "GraphCL: Contrastive self-supervised learning of graph representations," 2020, *arXiv:2007.08025*.

[47] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *CoRR*, vol. abs/1807.03748, pp. 1–13, Jul. 2018.

[48] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The Caltech-UCSD Birds-200–2011 dataset," California Inst. Technol., Pasadena, CA, USA, Tech. Rep. CNS-TR-2011-001, 2011.

[49] M.-H. Tsai, W.-J. Chen, J.-Y. Lin, G.-S. Lin, and S.-L. Yan, "Polyp classification based on deep neural network for colonoscopic images," in *Proc. 4th Int. Conf. Graph. Signal Process.* New York, NY, USA: Association for Computing Machinery, Jun. 2020, pp. 61–64, doi: 10.1145/3406971.3406977.
[50] Y. J. Kim, J. P. Bae, J.-W. Chung, D. K. Park, K. G. Kim, and Y. J. Kim, "New polyp image classification technique using transfer learning of network-in-network structure in endoscopic images," *Sci. Rep.*, vol. 11, no. 1, p. 3605, Feb. 2021, doi: 10.1038/s41598-021-83199-9.
[51] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 1–13.
[52] C.-Y. Chung et al., "Self-supervised object structure learning for image classification and segmentation," Dept. Comput. Sci., Nat. Tsing Hua Univ., Taiwan, 2022.

QI-XIAN HUANG (Member, IEEE) is currently pursuing the Ph.D. degree with the National Tsing Hua University, Hsinchu, Taiwan. He was an Intern with Taiwan Semiconductor Manufacturing Corporation (TSMC), and later with Qualcomm Semiconductor Corporation as a Senior Firmware Engineer. He was also a Researcher with the Artificial Intelligence E-learning Center, National Chengchi University, Taipei, Taiwan. His research interests include deep learning for computer vision algorithm and medical image applications, beyond 5G, 6G, blockchain networks, and the Internet of Things security. He is currently the Google Developer Technical Team Lead and has the great experiences to be an IEEE/ACM referee for journals of *Signal Processing System*, *Machine Learning*, and *Education E-learning*.

GUO-SHIANG LIN (Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan, in 2005. He joined the Department of Computer Science and Information Engineering, Da-Yeh University, Changhua County, Taiwan, in 2005, where he is currently an Associate Professor. He was a Visiting Professor with the Department of Mathematics and Computer Science, University of Münster, Germany, for four months, in 2009. Currently, he is employed as a Professor at the National Chin Yi University of Science and Technology, Taichung, Taiwan. His current research interests include image/video forensics, image/video processing and applications, 2-D-to-3-D image/video conversion, computer vision, and pattern recognition.

HUNG-MIN SUN (Member, IEEE) received the Ph.D. degree in computer science and information engineering from the National Chiao-Tung University, Hsinchu, in 1995. He was an Associate Professor with the Department of Information Management, Chaoyang University of Technology, from 1995 to 1999, the Department of Computer Science and Information Engineering, National Cheng-Kung University, from 2000 to 2002, and the Department of Computer Science, National Tsing Hua University, Hsinchu, from 2002 to 2008, where he is currently a Full Professor with the Department of Computer Science. He has published more than 200 international journals and conference papers. His research interests include deep learning in MRI image applications, privacy in AI security, network security, cryptography, blockchain, and automatic trading. He was the Program Co-Chair of 2001 National Information Security Conference, and the program committee members of many international conferences. He was the Honor Chair of 2009 International Conference on Computer and Automation Engineering, 2009 International Conference on Computer Research and Development, and 2009 International Conference on Telecom Technology and Applications. He is the General Chair of ACM AsiaCCS'2020. He serves as the editor members for many international journals. He won many best paper awards in academic journals and conferences, including the annual Best Paper Award from the *Journal of Information Science and Engineering*, in 2003, the Best Paper Award in MobiSys09, NSC05, NISC06, NISC07, CISC09, ICS2010, ICMS 2019, and IEEE ICKII 2019. He won the Y. Z. Hsu Scientific Paper Award from Far Eastern Y. Z. Hsu Science and Technology Memorial Foundation, in 2010, and the Outstanding Research Award from World Congress on Information Technology Applications and Services, in 2015. He won the Award of Outstanding Professor in Electrical Engineering from the Chinese Institute of Electrical Engineering, in 2014.