Polyp Detection via Imbalanced Learning and Discriminative Feature Learning

Seung-Hwan Bae, *Student Member, IEEE*, and Kuk-Jin Yoon*, *Member, IEEE*

***Abstract*—Recent achievement of the learning-based classification leads to the noticeable performance improvement in automatic polyp detection. Here, building large good datasets is very crucial for learning a reliable detector. However, it is practically challenging due to the diversity of polyp types, expensive inspection, and labor-intensive labeling tasks. For this reason, the polyp datasets usually tend to be imbalanced, i.e., the number of non-polyp samples is much larger than that of polyp samples, and learning with those imbalanced datasets results in a detector biased toward a non-polyp class. In this paper, we propose a data sampling-based boosting framework to learn an unbiased polyp detector from the imbalanced datasets. In our learning scheme, we learn multiple weak classifiers with the datasets rebalanced by up/down sampling, and generate a polyp detector by combining them. In addition, for enhancing discriminability between polyps and non-polyps that have similar appearances, we propose an effective feature learning method using partial least square analysis, and use it for learning compact and discriminative features. Experimental results using challenging datasets show obvious performance improvement over other detectors. We further prove effectiveness and usefulness of the proposed methods with extensive evaluation.**

***Index Terms*—Endoscopy, colonoscopy, computer aided detection (CAD), polyp detection, imbalanced learning, feature learning, partial least square analysis, medical imaging system.**

I. INTRODUCTION

AUTOMATED polyp detection is to find locations and sizes of polyps in endoscopic or colonoscopic images automatically. As listed in Table I, many polyp detection methods [1]–[6] have been developed and achieved the impressive performance improvement based on elegant machine learning algorithms and well-established features. However, this problem is still challenging, and frequent failures occur in many practical situations as in Fig. 1.

The main difficulty of the detection is to handle the diversity of polyp appearance. Indeed, many different types of polyps exist, and each polyp contains significantly different colors, textures and shapes as in Fig. 2. In addition, appearance variations frequently occur by viewpoint changes and partial occlusions during endoscopy or colonoscopy.

Therefore, in order to achieve high performance in real applications, learning with large training datasets containing great variability of polyp appearances is required. As discussed in [7], [8], detectors learned with large datasets handling great variability of object appearances improve detection performance in general. Building the large datasets, however, is practically difficult in medical imaging since expensive endoscopy (*or* colonoscopy) of patients and labeling by human supervision are needed. For these reasons, datasets [1]–[3], [5] are not large enough to cover all the challenges of the polyp detection, and they are also often imbalanced where a larger number of non-polyp samples represent a majority (i.e., non-polyp or negative) class while a small number of polyp samples only represent a minority (i.e., polyp or positive) class.

In fact, the class distribution, the proportion of samples of each class in a dataset, plays an important role in a classification [10]. Learning with imbalanced sets usually results in biased classifiers prone to yield higher detection accuracy over a majority class, but poor accuracy over a minority class[1]. However, it is more important to improve the performance for the minority class for early detection.

In this paper, we tackle this problem of learning detectors with imbalanced datasets and propose a novel data sampling-based boosting framework to resolve the problem. The proposed framework is basically based on Adaboosting and up/down data sampling. At each round, we collect hard samples, which are not correctly classified by the preceding classifiers. By up-sampling, we generate synthetic samples for the selected hard samples of the minority class. We then remove hard samples of the majority class that surround the minority class samples using down-sampling. By generating more balanced datasets using up/down sampling in this manner, we prevent weak classifiers from biasing toward the majority class.

In addition, we propose discriminative feature learning to improve the detection performance. In fact, the appearances of polyp samples and non-polyp samples are quite similar as shown in Fig. 3. Thus, it is difficult to discriminate them just using conventional color- and texture-based features[2]. To resolve this problem, we learn more compact and discriminative features by projecting the high-dimensional histogram of oriented gradient (HOG) features [9] onto the weight matrix learned by partial least square (PLS) analysis [11]. Based on

Manuscript received April 09, 2015; revised May 07, 2015; accepted May 10, 2015. Date of publication May 18, 2015; date of current version October 28, 2015. *Asterisk indicates corresponding author.*

S.-H. Bae is with School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju 500-712, South Korea (e-mail: bshwan@gist.ac.kr).

*K.-J. Yoon is with School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju 500-712, South Korea (e-mail: kjyoon@gist.ac.kr).

Digital Object Identifier 10.1109/TMI.2015.2434398

---

[1] In Section V.D, we provided the performance comparison of both classes.

[2] Performances of detectors learned by different features are given in Table III and Fig. 14(a).

0278-0062 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DESCRIPTION OF THE SEGMENTATION- AND PATCH-BASED POLYP DETECTION METHODS

Method	Feature	Classification	Datasets	Sensitivity	ROC-AUC
[1]	Color wavelets	LDA	1380 images	93.60%	—
[2]	PCA-based feature	Neural network	64 images	97.72%	—
[3]	HSV color & Watershed	Two-stage segmentation	100 images	96.00%	—
[4]	MPEG-7 descriptors	Descriptor distance	899 images	90.00%	—
[5]	RGB color	SVM	35 images	—	93.16%
[6]	Intensity valleys	Segmentation	15 polyps, 380 images	89.00%	76.00%

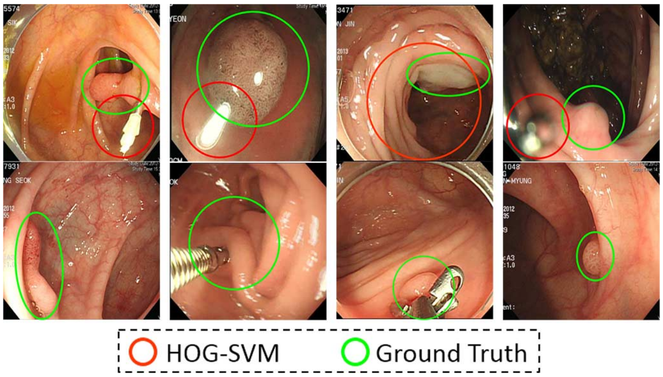


Fig. 1. Polyp detection results using the HOG and SVM detector [9]. False positive detections (top) and missing (bottom) frequently occur when applying the conventional detector.

the proposed imbalanced learning and discriminative feature learning, we build an ensemble framework for polyp detection as shown in Fig. 4.

The rest of the paper is organized as follows. We first discuss related works in Section II. Then, we propose feature learning and imbalanced learning for polyp detection in Sections III and IV, respectively. We provide some experimental results in Section V, and discuss the difference of Adaboost and cascade schemes in Section VI. We finally conclude the paper in Section VII.

II. RELATED WORKS

We discuss some previous works on polyp detection and imbalanced learning, which are related to our study.

A. Polyp Detection

In recent years, automated polyp detection has been extensively studied for reducing costs and time of colonoscopy and endoscopy. Since the task has many difficulties owing to the diversity of polyp types, appearance variations by camera viewpoint changes, similar colors and textures between polyps and non-polyps, and severe occlusions, a large number of polyp detection methods have been proposed using elegant machine learning and well-established features to deal with the difficulties.

For distinguishing abnormal regions (containing polyps or ulcers) from normal regions, a segmentation-based approach has flourished. It usually exploits color and/or texture differences between polyp regions and their surrounding regions. In [3], [12], polyp regions are iteratively segmented based on color and watershed segmentations. Bernal *et al.* [6] present a polyp region descriptor using the depth of a valleys image. However, these methods are unsuitable for real-time applications because they need expensive computation for the iterative segmentation procedures.

On the other hand, a patch-based approach learns a classifier (*or* detector) beforehand and then detects abnormal regions for the test image using the learned classifier. For improving discriminability, the color wavelet covariance [1], dimension-reduced features using principal component analysis (PCA) [2], MPEG-7 descriptors [4], and the support vector machine (SVM)-based polyp detector with high-dimensional features [13] have been proposed. These features are exploited for the tasks of various medical symptom detections such as bleeding, ulcers, and polyps. Since these patch-based methods need to learn detectors with training sets beforehand, their detection results highly rely on the quality of training datasets. Actually, they tend to generate biased classifiers from imbalanced datasets since the imbalanced data distribution problem between different classes is not appropriately treated.

We summarize the performance of the previous studies for polyp detection in Table I. Here, it should be noted that there is a discrepancy between the performance shown in Table I and the expected performance in real applications. The reason is that most of the works [1]–[5] evaluate the performance using a per-window measure rather than a per-image measure. However, as discussed in [14], the per-window measure typically leads to higher sensitivity and specificity since detections with incorrect scales or positions are not counted in the evaluation after non-maximal suppression or other post-processing.

In this paper, we propose a patch-based method for polyp detection. To effectively resolve the imbalanced learning problem, we learn a polyp detector based on a novel ensemble framework with feature learning and imbalanced learning. In addition, we discuss the discrepancy of polyp detection performance using different measures in Section V.D, and then we rigorously evaluate the proposed and existing methods using the more strict per-image measure in Sections V.E and V.F.

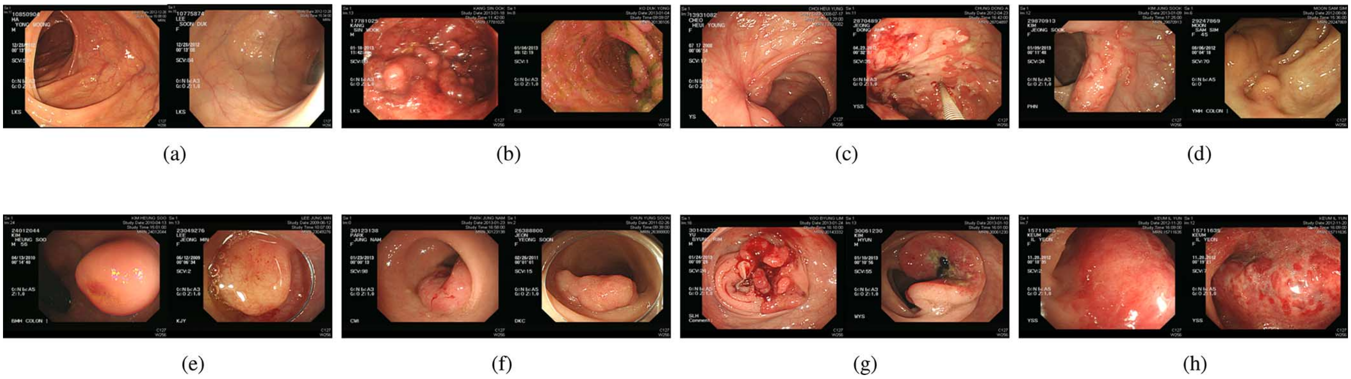
B. Imbalanced Learning

Previous methods[3] on imbalanced learning can be roughly categorized into two groups: reweighting-based and resampling-based approaches. The former enforces the classifier to be less biased toward the majority class by adaptively re-weighting the samples of different classes in consideration of the class importance. Here, since it is required to reduce critical false negatives rather than false positives in many cases, the cost for false negatives is higher than that for false positive in general. To this end, some cost-sensitive boosting (CSB) algorithms such as AdaBoost [16], CSB1/CSB2 [17],

---

[3] For more detailed review of imbalanced learning, refer to [10], [15].

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.


(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 2. (a)–(h): The diversity of polyp types: colors, textures and shapes of the polyps are different from each other. (a) Normal rectal tissue, (b) Inflammatory bowel disease (Ulcerative Colitis: UC), (c) Inflammatory bowel disease (Crohn's Disease: CD), (d) Tuberculosis (TB), (e) Rectal carcinoid, (f) Early colorectal cancer, (g) Progression of colonrectal cancer, (h) Colorectal other-related disease.

and AdaC1/AdaC2/AdaC3 [18] have been presented. Recently, the optimal cost-sensitive boosting [19] has been proposed by modeling and minimizing expected and empirical losses.

On the other hand, the latter approach concentrates on rebalancing the imbalanced distribution using resampling methods. To safely remove noisy and redundant samples belonging to the majority class, the one-sided selection method has been proposed [20]. SMOTE [21] and its extension [22] have been also proposed for generating new samples synthetically with the samples in a minority class. Guo [23] designs the boosting-based imbalanced learning algorithm with data generation. As discussed in [10], these resampling-based methods are usually independent of the base classifier.

Recently, in an attempt to promote the performance of the imbalanced classification, asymmetric classifiers have been developed for binary [24], [25] and multi-class classifications [26], [27]. In particular, PLS-based classifiers [25]–[27] show the high performance for skewed datasets. The main reason is that learning a PLS-based classifier is only related to an eigen-decomposition problem of the between-class scatter matrix without consideration of the number of data in each class [24], [25]. Therefore, it is less affected by the class distribution.

Inspired by recent advances in imbalanced learning, we build a novel ensemble framework, named *data sampling-based boosting*. Compared to previous works, the proposed framework has two obvious benefits. Firstly, our framework is an ensemble of the existing powerful imbalanced learning methods such as Adaboost, resampling, and PLS. By combining them, we can train an unbiased polyp detector even under extremely skewed distribution, and achieve the better detection performance than other imbalanced learning methods as shown in Table III. Secondly, different from [25]–[27], we exploit the PLS method for feature learning and dimension reduction rather than directly employ it as a classifier. As a result, we can greatly reduce the complexity caused by data re-sampling, classifier learning, and classification since we perform all the tasks with the low-dimensional features.

In terms of experimental validation, Galar *et al.* [10] pointed out that the existing imbalanced learning methods have been validated with relatively small-sized datasets [16], [20], [21] or the medical diagnosis dataset [18], [23], [25] with fixed and low-dimensional feature vectors. This brings the performance of the previous learning methods into question: can those methods work successfully for imbalanced large datasets obtained from real medical images? We think that the experimental results of the proposed and other imbalanced learning methods can suggest the answer to the question and worthy conclusion for medical imaging detection and classification with imbalanced datasets.

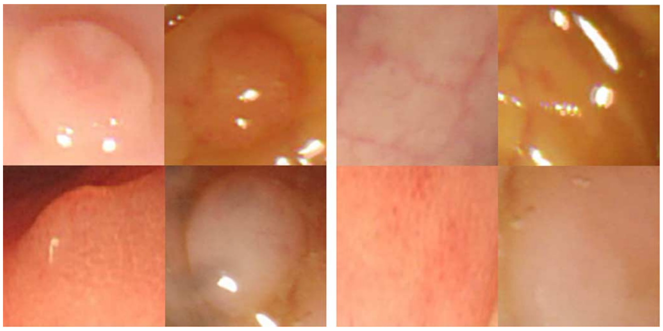


Fig. 3. Similar colors and textures between polyps (left) and non-polyps (right) samples.

III. DISCRIMINATIVE COMPACT FEATURE LEARNING

In this section, we provide the details of proposed feature learning using PLS analysis.

A. Feature Extraction

We use the HOG feature [9] for describing samples due to its discrimination power and robustness over noises or illumination changes [28], [29].

Given normalized image patches with resolution $w \times h$ containing polyps or non-polyps, HOG features [9] can be extracted for each image. The first step of HOG extraction is to compute gradients of pixels $\nabla f(x, y)$ by convolving the image with a filter $[-1, 0, +1]$ and its transpose. We then compute the magnitude $\|\nabla f(x, y)\|$ and orientation $\theta(x, y)$ of the gradient $\nabla f(x, y)$. At each pixel, the gradient orientation is discretized into one of p values within ranges of $[0, \pi]$ or $[0, 2\pi]$. The

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

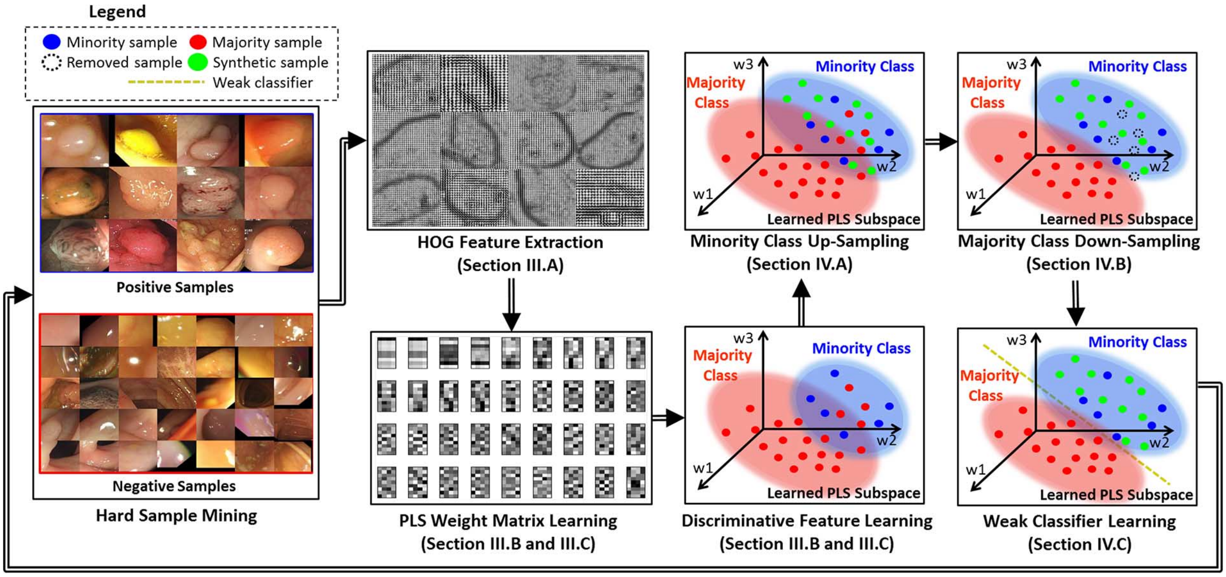


Fig. 4. A data sampling-based boosting framework for learning a polyp detector from imbalanced datasets.

contrast insensitive (B_1) and contrast sensitive (B_2) values are computed with the undirected and directed gradients as follows:

$$B_1(x,y) = \left\lfloor \frac{p\theta(x,y)}{\pi} \right\rceil \bmod p$$
$$B_2(x,y) = \left\lfloor \frac{p\theta(x,y)}{2\pi} \right\rceil \bmod p, \tag{1}$$

where $\lfloor\cdot\rceil$ means a round function.

A sparse orientation histogram $F(x,y)_b$, $b \in \{0,\ldots,p-1\}$ with p channels is generated with the discretized orientation and its magnitude.

$$F(x,y)_b = \begin{cases} \|\nabla f(x,y)\|, & b = B(x,y), \\ 0, & \text{otherwise}. \end{cases} \tag{2}$$

The sparse histogram at each pixel can be considered as a pixel-based feature map. We aggregate $F(x,y)_b$ within a square cell, where each cell region is determined by the parameter s representing the side length of a square area. A cell-based feature vector (i.e., gradient histogram) $C(i,j)$ is constructed by aggregating the pixel-based features, where $0 \le i \le \lfloor(w-1)/s\rfloor$ and $0 \le j \le \lfloor(h-1)/s\rfloor$. In general, this aggregation not only improves the robustness to small deformation but also reduces the size of a feature map [29]. When aggregating the pixel-based features, we further exploit the soft binning approach [9], [29] using trilinear interpolation for reducing the aliasing effect. Thus, instead of directly mapping each pixel (x,y) into the corresponding cell $\lfloor x/s\rfloor$ and $\lfloor y/s\rfloor$, we make each pixel contribute to the neighboring four cells for histogram smoothing.

Let define a square block with 2 × 2 cells and the histograms of overlapped cells between blocks are normalized with gradient energies of the blocks[4]. This normalization process enhances invariance to the illumination changes since the effect of gain (*or* scaling) are removed. In the similar manner as [9], [29], we utilize four normalization factors with $\delta,\gamma \in \{-1,+1\}$ as

$$N_{\delta,\gamma}(i,j) = \Big(\|C(i,j)\|^2 + \|C(i+\delta,j)\|^2 + \|C(i,j+\gamma)\|^2 + \|C(i+\delta,j+\gamma)\|^2\Big)^{\frac{1}{2}}, \tag{3}$$

Each factor $N_{\delta,\gamma}(i,j)$ evaluates the gradient energy (i.e., the sum of the squares of the gradient histogram components) in a block containing the cell (i,j).

For building a HOG map $H(i,j)$, we concatenate different normalization results for $C(i,j)$, normalized by the gradient energies of four blocks containing the cell (i,j):

$$H(i,j) = \begin{pmatrix} T_\alpha\left(C(i,j)/N_{-1,-1}(i,j)\right) \\ T_\alpha\left(C(i,j)/N_{+1,-1}(i,j)\right) \\ T_\alpha\left(C(i,j)/N_{-1,+1}(i,j)\right) \\ T_\alpha\left(C(i,j)/N_{+1,+1}(i,j)\right) \end{pmatrix}, \tag{4}$$

where a truncation function $T_\alpha(\cdot)$ with the lower bound α is used for removing noisy features.

In the case of using 9 orientation features ($b = 9$), we can extract a commonly used 36-dimensional HOG feature for each cell. However, for improving the detection performance as discussed in [29], we extract the 31-dimensional HOG feature for each cell instead of using the original HOG feature [9]. The main difference of both features is that the original one is a 36-dimensional feature with undirected or directed gradients using four normalization factors as in (4), while its variant is a combined feature with both directed gradients (18 dimensions) and undirected gradients (9 dimensions) as well as gradient energy features of the four neighboring blocks (4 dimensions)[5].

Fig. 5(a)–(d) show (positive and negative) samples and extracted HOG feature maps. Although colors and textures of polyps are different from each other in Fig. 5(a), their HOG

---

[4] A cell of each block is overlapped with its four neighborhood blocks. Therefore, the size of the overlapped region between four blocks is the same as the cell size $s \times s$. Given $N_h \times N_w$ cells, we obtain $(N_h-1)\times(N_w-1)$ blocks. For example, if the sizes of an image patch and a cell are $128(w) \times 128(h)$ and $s = 8$, we obtain $16(N_h) \times 16(N_w)$ cells and 15×15 overlapping blocks since $N_h = \lfloor h/s\rfloor$ and $N_w = \lfloor w/s\rfloor$.

[5] In our experiment, we found that the variant of the HOG feature significantly improves the performance of polyp detection as in Table III.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

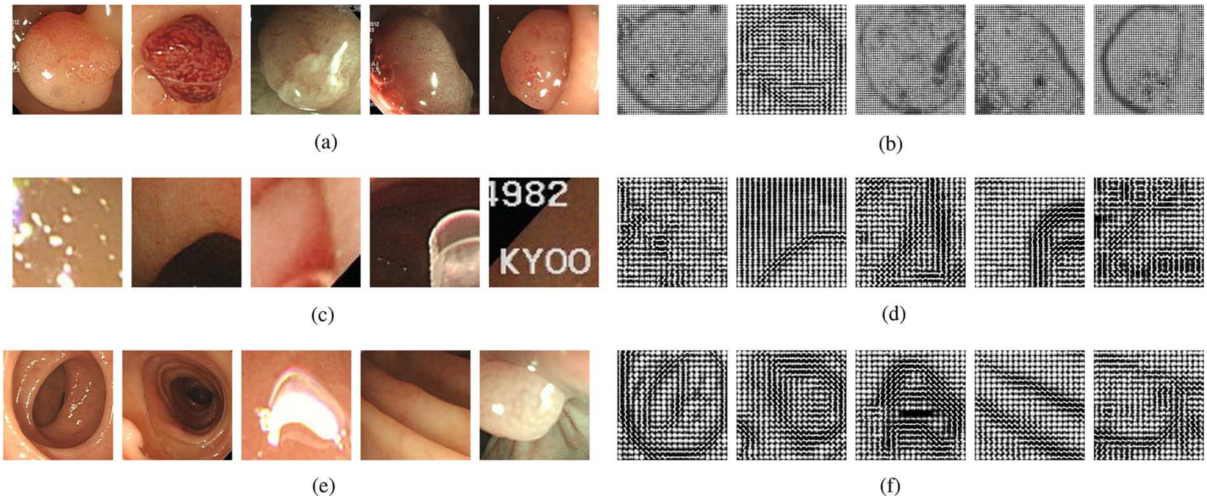

(a) (b) (c) (d) (e) (f)

Fig. 5. Polyp and non-polyp patches (left) and their HOG feature maps (right). (a) Positive training samples (polyp patches). (b) Extracted HOG features from positive training samples in (a). (c) Negative training samples (non-polyp patches). (d) Extracted HOG features from negative training samples in (c). (e) False positive samples that do not contain polyps. (f) Extracted HOG features from false positive samples in (e).

features usually have spherical shapes as shown in Fig. 5(b). On the other hand, Fig. 5(c) shows negative training samples consisting of non-polyps (i.e., normal tissues, colon holes, polyp fragments, medical equipments and patient names). From Fig. 5(d), we know that their HOG shapes are non-spherical and different from the polyp ones. In Fig. 5(e), we further show false positive samples when applying the conventional HOG and SVM detector [9].

In real cases, these false positives are frequently caused by colon holes, specular reflection, colon wrinkles and polyp fragments. Note that the HOG shapes of false positives in Fig. 5(f) are (approximately) spherical and rather similar to HOG shapes of polyps in Fig. 5(b). It means the challenge of discriminating between polyps and false positives using the HOG feature. In addition, HOG is a very high-dimensional feature. In general, this high dimensionality leads to models with lots of parameters and increases complexities of learning and detecting phases. To resolve these problems, we propose a feature learning method using PLS in next section.

B. Partial Least Square Analysis

PLS analysis models relations between sets of observed variables with latent variables. The common assumption is that observed data is generated by the system and process driven by a few number of latent variables. It aims at creating orthogonal score vectors (*or* latent vectors) by maximizing the covariance between different variable sets. Due to its effectiveness, PLS analysis has been exploited for many other applications [30]–[34]. We briefly introduce the main ideas and process of PLS analysis. For more details of PLS analysis, refer to [11], [30].

We denote a d-dimensional space of variables and a m-dimensional space of other variables as $\mathcal{X} \in \Re^d$ and $\mathcal{Y} \in \Re^m$. Given n data samples of each variable, we denote matrices of the two zero-mean variables as $\mathbf{X}(n \times d)$ and $\mathbf{Y}(n \times m)$, respectively. For preserving variations of each matrix and maximizing correlation between both variables, PLS decomposes the matrices as follows:

$$\begin{aligned} \mathbf{X} &= \mathbf{Z}\mathbf{P}^{\mathrm{T}} + \mathbf{E}, \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^{\mathrm{T}} + \mathbf{R}, \end{aligned} \tag{5}$$

where $\mathbf{Z}$ and $\mathbf{U}$ are $(n \times p)$ matrices of the p extracted latent vectors, the matrix $\mathbf{P}(d \times p)$ and the matrix $\mathbf{Q}(m \times p)$ represent the loadings, and the matrix $\mathbf{E}(n \times d)$ and the matrix $\mathbf{R}(n \times m)$ are residuals. Then, the PLS method finds weigh vectors $\mathbf{w}_1$ and $\mathbf{c}_1$ maximizing the covariance between the two latent vectors as follows:

$$\begin{aligned} &\underset{\|\mathbf{w}_1\|=1,\|\mathbf{c}_1\|=1}{\operatorname{argmax}} \quad \mathrm{Cov}(\mathbf{z}_1, \mathbf{u}_1) \\ &\qquad = \underset{\|\mathbf{w}_1\|=1,\|\mathbf{c}_1\|=1}{\operatorname{argmax}} \quad \rho(\mathbf{z}_1, \mathbf{u}_1)\sqrt{\mathrm{Var}(\mathbf{z}_1)\mathrm{Var}(\mathbf{u}_1)}. \end{aligned} \tag{6}$$

where $\mathbf{z}_1 = \mathbf{X}\mathbf{w}_1$ and $\mathbf{u}_1 = \mathbf{Y}\mathbf{c}_1$ are the first columns of $\mathbf{Z}$ and $\mathbf{U}$, respectively.

As discussed in [11], [30], the above optimization problem can be reformulated as the following eigenproblem:

$$\begin{aligned} \mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{X}\mathbf{w}_1 &= \lambda\mathbf{w}_1, \\ \mathbf{Y}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{c}_1 &= \lambda\mathbf{c}_1. \end{aligned} \tag{7}$$

Once we obtain $\mathbf{w}_1$ and $\mathbf{c}_1$ by solving (7), other weight vectors for the two variables are iteratively evaluated using the nonlinear iterative partial least squares (NIPALS) algorithm [11]:

- The score vectors $\mathbf{z}_l$ and $\mathbf{c}_l$ are evaluated by $\mathbf{z}_l = \mathbf{X}_l\mathbf{w}_l$ and $\mathbf{u}_l = \mathbf{Y}_l\mathbf{c}_l$, where $l = 1, 2, \ldots$ and $\mathbf{X}_l = \mathbf{X}$ and $\mathbf{Y}_l = \mathbf{Y}$ when $l = 1$.
- The vectors of loading $\mathbf{p}_l$ and $\mathbf{q}_l$ are computed by $\mathbf{p}_l = \mathbf{X}_l^{\mathrm{T}}\mathbf{z}_l/(\mathbf{z}_l^{\mathrm{T}}\mathbf{z}_l)$ and $\mathbf{q}_l = \mathbf{Y}_l^{\mathrm{T}}\mathbf{u}_l/(\mathbf{u}_l^{\mathrm{T}}\mathbf{u}_l)$.
- The sample data matrices $\mathbf{X}_l$ and $\mathbf{Y}_l$ are deflated by $\mathbf{X}_{l+1} = \mathbf{X}_l - \mathbf{z}_l\mathbf{p}_l^{\mathrm{T}}$ and $\mathbf{Y}_{l+1} = \mathbf{Y}_l - \mathbf{u}_l\mathbf{q}_l^{\mathrm{T}}$.
- The new weight vectors $\mathbf{w}_{l+1} = \mathbf{X}_{l+1}^{\mathrm{T}}\mathbf{u}_l/(\mathbf{u}_l^T\mathbf{u}_l)$ and $\mathbf{c}_{l+1} = \mathbf{Y}_{l+1}^{\mathrm{T}}\mathbf{z}_l/(\mathbf{z}_l^T\mathbf{z}_l)$ of the deflated matrices are computed and normalized to be $\|\mathbf{w}_{l+1}\| = 1$ and $\|\mathbf{c}_{l+1}\| = 1$.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

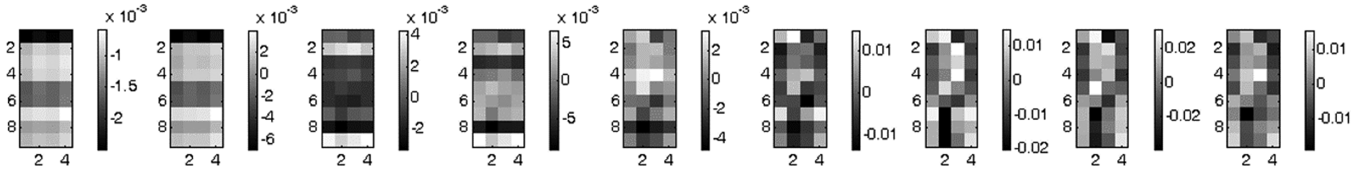


Fig. 6. Weight vectors of HOG features [9] using PLS. From left to right, first learned weight vectors are shown. Each weight vector is displayed as 9 by 4 matrix. Since we use 9 orientation bins and 4 normalization factors for extracting HOG features, each row corresponds to one of the orientation bins and each column to one of the normalization factors.

We repeat the above procedure until sets of the p weight vectors, $\mathbf{W} = \{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_p}\}$ and $\mathbf{C} = \{\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_p}\}$, are obtained. In next section, we apply the PLS analysis for learning features for polyp detection.

C. Discriminative and Compact Space Learning

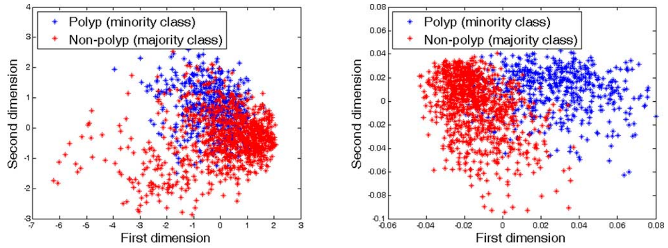
For better discrimination between polyps and non-polyps, we learn a discriminative and compact polyp feature using the PLS method. With the same notations as in Section III.B, we denote a set of d-dimensional HOG feature vectors from n polyp and non-polyp images as $\mathbf{X}(n \times d)$, and their labels $\mathbf{y}(n \times 1)$, where each label $y_i \in \{-1, +1\}$. We can obtain $\bar{\mathbf{X}}$ and $\bar{\mathbf{y}}$ with the zero mean by subtracting $\mathbf{X}$ and $\mathbf{y}$ with their means, $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$. We then decompose $\bar{\mathbf{X}}$ and $\bar{\mathbf{y}}$ into the PLS form as (5).

Note that $\bar{\mathbf{y}}$ has only one variable. Therefore, the loadings and the residuals are denoted as vectors $\mathbf{q}(1 \times p)$ and $\mathbf{r}(n \times 1)$, respectively. It also means $\mathbf{c}_1$ to be a scalar due to $\mathbf{u}_1 = \bar{\mathbf{y}}\mathbf{c}_1$ and $\|\mathbf{c}_1\| = 1$. As a result, we only evaluate latent vectors of HOG features $\bar{\mathbf{X}}$ since $\mathbf{c}_1 = 1$ and $\mathbf{u}_1 = \bar{\mathbf{y}}$. As provided in previous Section III.B, we iteratively compute weight vectors $\mathbf{w}_l$ using the NIPALS algorithm. In the first iteration, we can compute $\mathbf{w}_1 = \bar{\mathbf{X}}_\mathbf{1}^{\mathrm{T}}\bar{\mathbf{y}}/\|\bar{\mathbf{y}}^{\mathrm{T}}\bar{\mathbf{y}}\|$ due to $\mathbf{u}_1 = \bar{\mathbf{y}}$ where $\bar{\mathbf{X}}_1 = \bar{\mathbf{X}}$. In the following l-th iteration $(l > 1)$, we obtain the deflated matrix $\bar{\mathbf{X}}_{l+1} = \bar{\mathbf{X}}_l - \mathbf{z}_l\mathbf{p}_l^{\mathrm{T}}$ and the weight vector $\mathbf{w}_{l+1} = \bar{\mathbf{X}}_{l+1}^{\mathrm{T}}\bar{\mathbf{y}}/(\bar{\mathbf{y}}^T\bar{\mathbf{y}})$.

Fig. 6 shows the learned weight vectors and values using PLS analysis, where higher weights (more importance) are marked with white but lower weights with black. Note that top weight vectors are (approximately) constant along each row and column of the matrix representation. As proved in [29], these results indicate that the top weight vectors can be considered as a two-dimensional separable Fourier basis because each weight vector seems like a sine and cosine function of one variable. Furthermore, the appearance of the Fourier basis proves the two-dimensional rotation invariance of the learned weight vectors.

Once the weight matrix $\mathbf{W}(d \times p)$ is learned, we project each feature vector $\mathbf{x}(1 \times d)$ onto the $\mathbf{W}$ for generating the compact and discriminative feature $(d \gg p)$. As a result, we can obtain the reduced feature vector $\mathbf{f}_i(1 \times p)$ and employ it in the classification.

In a similar manner, PCA is also used as a subspace learning method. However, PCA is different from PLS in some senses. When creating orthogonal weight vectors, PLS considers the covariance of feature vectors $\mathbf{X}$ and class labels $\mathbf{y}$, but PCA only considers the variances of the features. In addition, PLS is less affected by the class distribution [24], [25] since learning subspaces using PLS is only related to the eigen-decomposition problem of the between-class scatter matrix without consideration of the number of data in each class (*cf.* PCA). Thus, the former is much more appropriate for polyp detection since it is more capable of distinguishing class variability and handling the imbalanced dataset. Fig. 7 shows the first two-dimensional components after the dimension reduction using PCA and PLS. It clearly shows that polyps and non-polyps are more separated by PLS rather than by PCA.



(a) (b)

Fig. 7. Feature distribution after applying PCA (a) and PLS analysis (b).

LDA also generates discriminative subspaces for distinguishing different class samples similar to PLS. However, LDA is not suitable for the dimension reduction of high-dimensional features since it can only produce $\xi - 1$ projections, where ξ is the number of classes. In addition, the covariances cannot be full rank and the weight vectors cannot be extracted when the feature dimension exceeds the amount of training samples.

IV. DATA SAMPLING BASED BOOSTING FOR IMBALANCED LEARNING

In this section, we present re-sampling methods for balancing imbalanced datasets and a data sampling-based boosting framework based on the proposed data sampling.

A. Up-Sampling: Synthetic Sample Generation

Even though the discriminative features learned by PLS can improve the classification performance with imbalanced datasets, it is not enough to fully consider the imbalance problem. To deal with the imbalance problem more appropriately, we also introduce a method to generate synthetic samples from given training samples. Here, although our method can generate synthetic samples for both classes, samples of the polyp class are only considered for balancing datasets. In addition, it is worthy of notice that, inspired by the SMOTE algorithm [21], we generate samples in the feature space rather than the data space since generating synthetic images is much more computationally expensive.

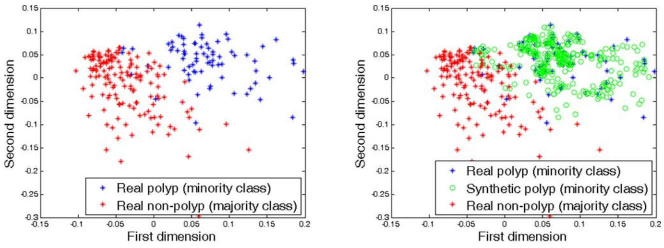
Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Up-sampling

```
Input : A training set $\mathbf{X} = \{\mathbf{x}_i\}, \mathbf{x}_i \in \Re^d$, amount of
        up-sampling $L$, number of nearest neighbors $k$
Output: A synthetic sample set $\mathbf{A}$
Compute $k$-nearest neighbors in $\mathbf{X}$ and save the indices in the $knnArray$;
$ct = 0$, while $L \neq 0$ do
    Randomly choose a sample $i$ and its neighbor $r$;
    for $j \leftarrow 1$ to $d$ do
        $dif \longleftarrow x[knnArray[r]][j] - x[i][j]$;
        $gap \longleftarrow$ random number between 0 and 1 ;
        $\mathbf{A}[c][j] \longleftarrow x[i][j] + gap * dif$;
    end
    $ct \longleftarrow ct + 1; L \longleftarrow L - 1$;
end
```

Algorithm 2: Down-sampling

```
Input : A training set $\mathbf{S} = \{\mathbf{x}_i, y_i\}$ (sample, label), $i = 1, ..., n$
Output: A down-sampled set $\mathbf{O}$
while $\mathbf{T} \neq \emptyset$ do
    Generate a subset $\mathbf{C} \subset \mathbf{S}$ containing all minority class
    samples and one randomly selected majority class sample
    from $\mathbf{S}$;
    Classify the samples in $\mathbf{S}$ with the 1-NN rule using the
    samples in $\mathbf{C}$;
    Move all misclassified samples into $\mathbf{C}$;
    Find Tomek-linked pairs from $\mathbf{C}$ and move majority class
    samples participating in the pairs to $\mathbf{T}$;
    $\mathbf{S} \longleftarrow \mathbf{S} - \mathbf{T}$;
end
$\mathbf{O} \longleftarrow \mathbf{S}$
```



(a) (b)

Fig. 8. Feature distribution before (a) and after (b) up-sampling.

For balancing an imbalanced dataset, up-sampling generates synthetic polyp samples from real polyp samples and their neighbors, which are determined by a k-nearest neighbor algorithm. Depending on the amount of up-sampling, some neighbors among the k-nearest neighbors are randomly selected. When the number of real samples is 100 and the amount of up-sampling needed is 300%, 300 synthetic samples are generated from the randomly chosen samples and their neighbors. Given one sample and its neighbors are provided, synthetic polyp samples are generated as follows: (i) calculate a difference between each feature vector element of the sample and one of its nearest neighbors, (ii) multiply the difference by a random number between 0 and 1, and (iii) add it to the feature vector element under consideration. This steps are performed until all feature elements are computed. The detailed up-sampling algorithm for generating synthetic samples is presented in Algorithm 1.

Fig. 8 shows sample distributions of polyp and non-polyp classes before and after up-sampling. We can see that the region of the polyp class is relatively expanded. This up-sampling allows us to create larger and less specific decision regions when learning polyp classifiers

B. Down-Sampling: Noisy Sample Elimination

Another simple way for re-balancing datasets is to eliminate some unimportant samples in the non-polyp class. In some cases, abundant majority class samples in an imbalanced dataset make classification more difficult. For instance, assume that polyp samples are surrounded by non-polyp samples. Then, the non-polyp ones can be the nearest neighbors of polyp samples. Under this circumstance, a classifier is likely to be biased to the non-polyp class in order to minimize the classification error. Many polyp samples can be then misclassified. To address this problem, we just remove some portion of non-polyp samples with the random selection.

This simple way, however, is likely to remove non-polyps needed to be kept for correct classification (e.g., non-polyps to be support vectors in SVM). For removing noisy non-polyps while keeping important ones, we investigate Tomek links [20] within training samples, which can be defined as

Definition 1: Consider two samples, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, with different labels, and denote their distance as $\psi(\boldsymbol{x}_1, \boldsymbol{x}_2)$. The pair $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is then defined as a Tomek link if no sample $\boldsymbol{z}$ exists such that $\psi(\boldsymbol{x}_1, \boldsymbol{z}) < \psi(\boldsymbol{x}_1, \boldsymbol{x}_2)$ or $\psi(\boldsymbol{x}_2, \boldsymbol{z}) < \psi(\boldsymbol{x}_2, \boldsymbol{x}_1)$.

We consider samples participating in Tomek links as noisy samples and remove them from the original training set since they are not informative in learning at all. However, finding Tomek linked pairs for whole samples is an expensive and time-consuming process. Therefore, we first classify the original set using 1-nearest neighbor (NN) classification with all minority class samples and one randomly selected majority class samples. Then, we only consider the linked pairs between all minority class samples and misclassified majority class samples. The detailed down-sampling procedures are given in Algorithm 2.

Fig. 9 demonstrates the results of the down-sampling for the polyp dataset used for our experiments. We can see that our down sampling removes noisy non-polyp samples, which are too close to polyp ones, while keeping crucial samples around the class boundary. Thus, the dataset becomes more balanced after down-sampling without loss of information.

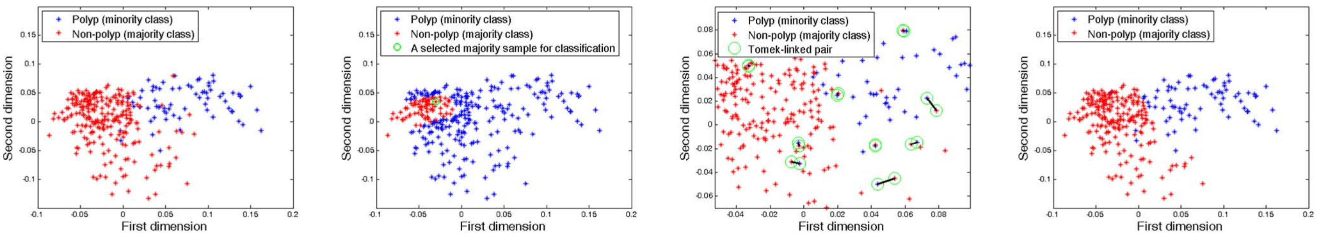
Note that the amount of the down-sampling highly relies on the data distribution of classes and it is automatically determined by Tomek-linked pairs. When the polyp and non-polyp samples are closely distributed, the large amount of non-polyp samples is down-sampled. On the other hand, only the small amount of non-polyp samples is down-sampled when the distributions are well separated.

C. Data Sampling-Based Boosting

To cover the diversity of polyp appearances with imbalanced polyp datasets, we design a data sampling-based boosting framework by combining the proposed feature learning and data sampling with the AdaBoost.M1 algorithm [35], [36].

Given a training dataset $\mathbf{S} = \{\mathbf{f}_i, y_i\}_{i=1}^n$ with n image patches $\mathbf{f}_i$ (i.e., learned features using PLS as discussed in Section III.C) and labels $y_i \in \{-1, +1\}$, we aim to learn

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c) (d)

Fig. 9. Down-sampling results: (a) original sample distribution, (b) distribution after 1-NN classification (c) Tomek-linked pairs (d) distribution after removing non-polyps participating in Tomek-links for 5 iterations.

K multiple weak classifiers. At each training round k, we collect hard samples[6] in $\mathbf{S}$, which are selected from the sample distribution $D_k(i)$ updated by the classification results of the preceding weak classifiers[7].

We re-balance the dataset using up/down sampling: The up-sampling generates synthetic samples from the hard samples of the polyp class, whereas the down-sampling removes hard samples of the non-polyp class when their Tomek link pairs are polyps. A classifier h_k is then trained with the re-balanced dataset. The classification error ϵ_k of h_k is evaluated with the training samples as $\epsilon_k = \sum_{i:h_k(\mathbf{f}_i)\neq y_i} D_k(i)$, and the classifier is accepted as a weak classifier only when $\epsilon_k > 1/2$. Once a weight β_k of the learned classifier is evaluated by $\beta_k = \epsilon_k/(1-\epsilon_k)$, the sample distribution is updated as

$$D_k : D_{k+1}(i) = \frac{D_k(i)}{Z_k} \times \begin{cases} \beta_k, & \text{if } h_k(\mathbf{f}_i) = y_i \\ 1, & \text{otherwise} \end{cases}, \tag{8}$$

where $Z_k = \sum_i D_k(i)$ is a normalization factor to make $D_{k+1}(i)$ a proper distribution. The updated distribution ensures that hard samples misclassified by the preceding classifiers are more likely to be selected for the subsequent weak classifier learning. These procedures are repeated during M iterations. As a result, $K(K \leq M)$ polyp classifiers and their weight matrices $\mathbb{W} = \{\mathbf{W}_k\}_{k=1}^{K}$ are learned.

Given a test feature $\mathbf{x}_j$, we obtain the final score $H(\mathbf{x}_j)$ with the learned h_k and $\mathbf{W}_k$ based on weight voting:

$$H(\mathbf{x}_j) = \sum_{k=1}^{K} \log(1/\beta_k) h_k(\mathbf{x}_j \mathbf{W}_k). \tag{9}$$

The pseudo code for learning a polyp detector based on the proposed scheme is provided in Algorithm 3.

Algorithm 3: Data sampling-based boosting

```
Input : A training set $\mathbf{S} = \{\mathbf{s}_i, y_i\}_{i=1}^{n}$, integer $M$ specifying
        number of iterations and $K = M$.
Output: A strong classifier $H(\mathbf{x}) = \sum_{k=1}^{K} \log(1/\beta_k) h_k(\mathbf{x})$
        and a set of PLS weight matrices $\mathbb{W} = \{\mathbf{W}_k\}_{k=1}^{K}$
Initialize $D_1(i) = 1/n, i = 1, ..., n$;
for $k \leftarrow 1$ to $M$ do
    Select a training data subset $\mathbf{S}_k$, drawn from the distribution $D_k$;
    Extract a set of HOG features $\mathbf{X} = \{\mathbf{x}_i\}$ from the selected subset $\mathbf{S}_k$ (given in Sec. III-A);
    Generate a projection matrix $\mathbf{W}_k$ for $\mathbf{X}$ using PLS (given in Sec. III-C);
    Produce discriminative features $\mathbf{f}_i$ by projecting all features $\mathbf{x}_i \in \mathbf{S}_k$, $\mathbf{f}_i = \mathbf{x}_i \mathbf{W_k}$ ;
    $\mathbf{S}_k \longleftarrow$ Do Up-sampling $(\mathbf{S}_k)$ (given in Algorithm 1);
    $\mathbf{S}_k \longleftarrow$ Do Down-sampling $(\mathbf{S}_k)$ (given in Algorithm 2);
    $h_k \longleftarrow$ Do Weak-classifier-learning $(\mathbf{S}_k)$;
    Calculate the error of $h_k$ : $\epsilon_k = \sum_{i:h_k(\mathbf{f}_i)\neq y_i} D_k(i)$ ;
    If $\epsilon_k > 1/2$, then $K \leftarrow K - 1$ and continue loop;
    $\beta_k \leftarrow \epsilon_k/(1-\epsilon_k)$ ;
    Update distribution $D_k : D_{k+1}(i) = \frac{D_k(i)}{Z_k} \times \begin{cases} \beta_k, & \text{if } h_k(\mathbf{f}_i) = y_i, \\ 1, & \text{otherwise,} \end{cases}$ where $Z_k = \sum_i D_k(i)$ is a normalization factor to make $D_{k+1}$ a proper distribution;
end
```

V. EXPERIMENTS

As given in Algorithm 3, the proposed polyp detector has been implemented in MATLAB. More detailed explanation of the proposed detector is given in Section V.A. The evaluation metrics and polyp datasets for performance evaluation are discussed in Sections V.B and V.C. We first show the performance discrepancy by exploiting different evaluation measures (i.e., per-window and per-image measure) in Section V.D. Using the more strict per-image measure, we compare the performance of the proposed detector with other detectors in Section V.E. We then investigate how the proposed methods (i.e., feature learning and up/down sampling) affect the overall performance of our detector in Section V.F.

A. Implementation

We have implemented polyp detectors as provided in Algorithm 3. From polyp datasets described in Section V.C, we first collect positive and negative training samples (i.e., image patches). As shown in Fig. 5, the positive patches are extracted to be contained with polyps using the ground truth while the negative patches are randomly extracted in images but not too much overlapped with polyps[8].

We resize the different-sized image patches to 128×128 (pixels), and then extract 31-dimensional HOG features for each cell in the resized patches as discussed in Section III.A[9]. The cell size and the number of undirected orientation bins are set to $s = 16$ and $b = 9$. As a result, we can extract a

---

[6] In our experiment, the hard samples are mainly extracted from segments of polyps, colon wrinkles, and holes that have similar appearances with polyps as shown in Fig. 5(e).

[7] When $k = 1$, the initial sample distribution for n samples is $D_1(i) = 1/n, i = 1, \ldots, n$.

[8] The results with different amounts of overlap areas are provided in Fig. 14(c).

[9] In our implementation, we use the code [37] open to the public.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

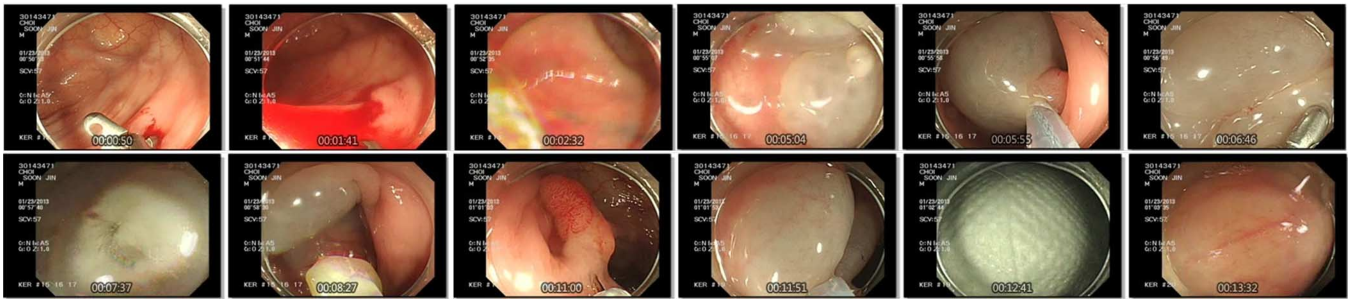


Fig. 10. Sample patches in our own datasets. We construct our own polyp datasets from 146 endoscopic videos of 141 patients.

1984-dimensional feature for each patch by concatenating 31-dimensional histograms of the 64 cells[10].

Given the HOG features and the corresponding labels (indicating a polyp or not), we learn PLS weight vectors as described in Section III.C. In our experiment, we learn a weight matrix (1984 × 7) and then project the high-dimensional HOG features onto the weight matrix. As a result, we can obtain the 7-dimensional features and use the dimension-reduced features for training and testing[11]. For ensemble learning, we use a linear SVM as a weak classifier. The number of training rounds (M) is set to 10.

For detecting various sizes of polyps, we build the image pyramid with 7 scale levels and extract a HOG feature map at each level. To determine the scaling factor of each level, we analyze the minimum and maximum sizes of polyps in endoscopic images. We found out the minimum and maximum sizes of them are almost 100 × 100 and 320 × 320 (pixels). Based on this analysis, we determine the 7 scaling factors from 0.4 to 1.0.

B. Evaluation Metric

We evaluate the detection performance using the per-image measure [14]. Given detection boxes BB_{dt} (with detection scores) and ground truth boxes BB_{gt}, we consider two boxes as matched if the ratio of an overlap area over an union of them exceeds 0.5: $\text{Area}(BB_{dt} \cap BB_{gt})/\text{Area}(BB_{dt} \cup BB_{gt}) > 0.5$. When two or more boxes BB_{dt} are matched with one BB_{gt}, we select the detection with the highest score. While the matched BB_{dt} is counted as true positives $(\mathbf{TP})$, unmatched BB_{dt} counted as false positives $(\mathbf{FP})$ and unmatched BB_{gt} as false negatives $(\mathbf{FN})$ (*or* missed detections). We then plot the **miss rate** against the false positive per image $(\mathbf{FPPI})$ in log-log scale by changing the threshold on detection scores. Here, lower curves indicate better performance.

For more comparisons, we further use the following metrics commonly used for detection performance evaluation:

- **Precision**: The number of correctly matched detections (TP) divided by the total number of output detections, $\text{Precision} = (\text{TP})/(\text{TP} + \text{FP})$.
- **Recall**: The number of correctly matched detections (TP) divided by the total number of detections in ground truth, $\text{Recall} = (\text{TP})/(\text{TP} + \text{FN})$.

We also compute an area under the precision-recall (PR) curve, where the larger area under the curve (AUC) means better performance.

C. Polyp Dataset

We used the open CVC-ColonDB [6] containing 379 polyp images with 500 × 574 pixel resolution captured in a colonoscope. All the images contain polyps, and the central potions of the images are cropped to remove black borders. In order to increase variability of the polyp appearances, polyp images are captured at different points of views. However, the diversity of polyp types is limited since the dataset only contains 15 different polyps.

Since the dataset size is rather small, we created our own polyp dataset. From 146 endoscopic videos of the 141 patients, we captured 2365 images with 520 × 540 pixel resolution and removed 1102 images captured in similar frames. Therefore, our dataset consists of 1263 images. Since our work is focused on evaluating detection performance, each image could contain a polyp or not. Figs. 10 and 15 show some examples in the dataset. We observe that the dataset contains many challenges; diversity of polyp types, appearance variability due to viewpoint changes, occlusions by other polyps or medical equipments, and illumination changes by specular reflection.

D. Evaluation Using Per-Window and Per-Image Measures

To show the discrepancy of detection performance when using different evaluation measures, we evaluate the detection performance of each class using the proposed detector[12]. As described in [14], the per-window measure evaluates a detector by classifying cropped windows (with/without interesting objects), whereas the per-image measure evaluates it by considering overlapped areas between detection and ground truth boxes.

For the per-window evaluation, we collected 1000 polyp patches and 100 000 non-polyp patches from our polyp dataset. Based on the 5-fold cross-validation, we trained and tested

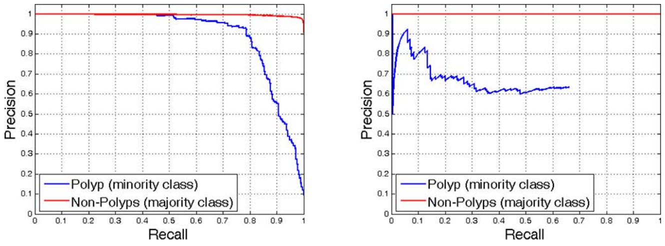
---

[10] When using the commonly used cell size $s = 8$, a 7936-dimensional HOG feature can be extracted by concatenating 31-dimensional features of 256 cells. However, from the extensive evaluation, we found that using the low-dimensional HOG feature improves the detection speed without performance degradation as proved in Table III.

[11] We determine the dimension of PLS from the performance evaluation shown in Fig. 14(b).

[12] This detector was implemented using Algorithm 3 without up/down sampling to clearly see the performance discrepancy.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 11. (a)–(b): Using different measures, precision and recall scores for an imbalanced polyp dataset are evaluated. (a) Per-window measure, (b) Per-image measure.

a detector to evaluate precision and recall curves for image patches of both classes as shown in Fig. 11(a).

For 1000 polyp images, we also evaluate the performance using the per-image measure based on the 5-fold cross-validation. We collected 100 non-polyp image patches in each image for training. After non-maximal suppression, we determined the areas (*or* boxes) of polyps for the 7-scaled image pyramid. Fig. 11(b) demonstrates precision and recall curves of both classes by matching the detection boxes with ground truth boxes as provided in Section V.B.

Although it achieves high performance for both classes for the per-window evaluation as in Fig. 11(a), the performance for the minority class falls far behind that for the majority class for the per-image evaluation as shown in Fig. 11(b). The reason is that the per-window measure typically leads to higher scores since detections with incorrect scales or positions are not counted after non-maximal suppression or other post processing. Therefore, we evaluate the performance of the polyp detectors using the more strict per-image measure in the next sections.

E. Comparison With Other Detectors

Even though many methods reported the performance as in Table I, it is not appropriate to compare our detector with them since they used their own datasets (not available to the public) and the per-window measure for evaluation.

Since [6] evaluated the performance using the per-image measure and opened the CVC colon dataset to the public, we compared our detector with [6]. However, it is not reasonable to directly compare our method with [6] since both methods are based on different approaches (i.e., ours: patch-based approach, [6]: segmentation-based approach).

To compare both methods, in a similar manner to [6][13], we investigate whether the maximum score points fall inside the polyp mask (as shown in Fig. 12) after non-maximal suppression of the confidence score map of the image pyramid. We consider the point inside and outside of the mask as **TP** and **FP**. When any point is not inside the mask, we consider it as **FN**. We also compute the **precision, recall, PR-AUC**, and **speed** of the detector. For a fair comparison, we divided the 1263 images of our dataset into 5 subsets (consisting of roughly 250 images) and trained 5 detectors for each subset. Subsequently, we evaluated each performance for the first 300 frames of the CVC dataset and provide the performance of the detector with the highest PR-AUC score in Table III. Note that different datasets were used for training (i.e., our polyp dataset) and testing (i.e., CVC-colon dataset) separately, and for testing we only used the polyp samples that had not been used for training to avoid performance bias.

For more comparisons, we have implemented a variety of polyp detectors based on different features and imbalanced learning algorithms, and compared their performance in Table III. Since the detector performance is affected by changing the detection score threshold τ and the suppression window size $[128 \times \sigma, 128 \times \sigma]$, we provide the optimal parameters of each detector in the last column of Table III. We also report the number of the trained weak classifiers K for boosting, and the best cost ratio C of the minority class to the majority class of cost-sensitive boosting algorithms (AdaC1-AdaC3) [18]. Table II shows the range of the parameter value and its interval used for evaluation.

1) Evaluation Using Different Features: In Table III.A, we compare the detection performance using the same classifier but different features: when extracting the color histogram (A2), local binary pattern (LBP) [38] (A3), original HOG [9] (A4), and the variant of HOG (VHOG) [29] (A5-A6), we can extract 24-, 58-, 36- and 31-dimensional features for each cell with the size $s \times s$, respectively. Using the same cell size $s = 8$ as [9], we extracted 9216-dimensional HOG and 7936-dimensional VHOG by concatenating all histograms of 256 cells. We further extracted the color histogram, LPB, and VHOG with 1536, 3712, and 1984 dimension using $s = 16$. When extracting these features, we exploited the integral histogram technique [39] to accelerate detection process. In addition, we extracted the color wavelet covariance features (A1) with 72 dimension [1]. We learned the compressed VHOG-PCA (A8) and VHOG-PLS (A7 and A9) features by projecting the VHOG onto PCA and PLS matrices, respectively.

We confirm that the VHOG (A6) greatly improves the metric scores, compared to other features (A1-A4). Note that original HOG [9] (A4) using only the undirected gradient or directed gradient is not suitable for polyp detections. In addition, exploiting the large cell size ($s = 16$) is more effective rather than using the small cell size ($s = 8$) when comparing (A5) and (A6) results. Using the proposed feature learning (A9), we can obtain the best results in terms of both detection performance and complexity.

2) Evaluation Using Imbalanced Learning Algorithm: Tables III.B and III.C show the quantitative comparison results of detectors using different imbalanced learning methods. Based on the Adaboost scheme in Algorithm 3, we implemented the following detectors using resampling-based imbalanced learning methods in Table III.B:

- **SVM:** Single linear SVM classifier;
- **ESM:** Ensemble classifier *without data-sampling*;
- **ESM-DW:** Ensemble classifier *with down-sampling only*;
- **ESM-UP:** Ensemble classifier *with up-sampling only*;
- **ESM-UP-DW:** Ensemble classifier *with up/down sampling*,

where the amount of the up-sampling is set to $L = 200$ for all the detectors using the proposed up-sampling. As mentioned

---

[13] Bernal *et al.* [6] investigate whether the global maximum of the energy map falls inside the polyp region or not.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

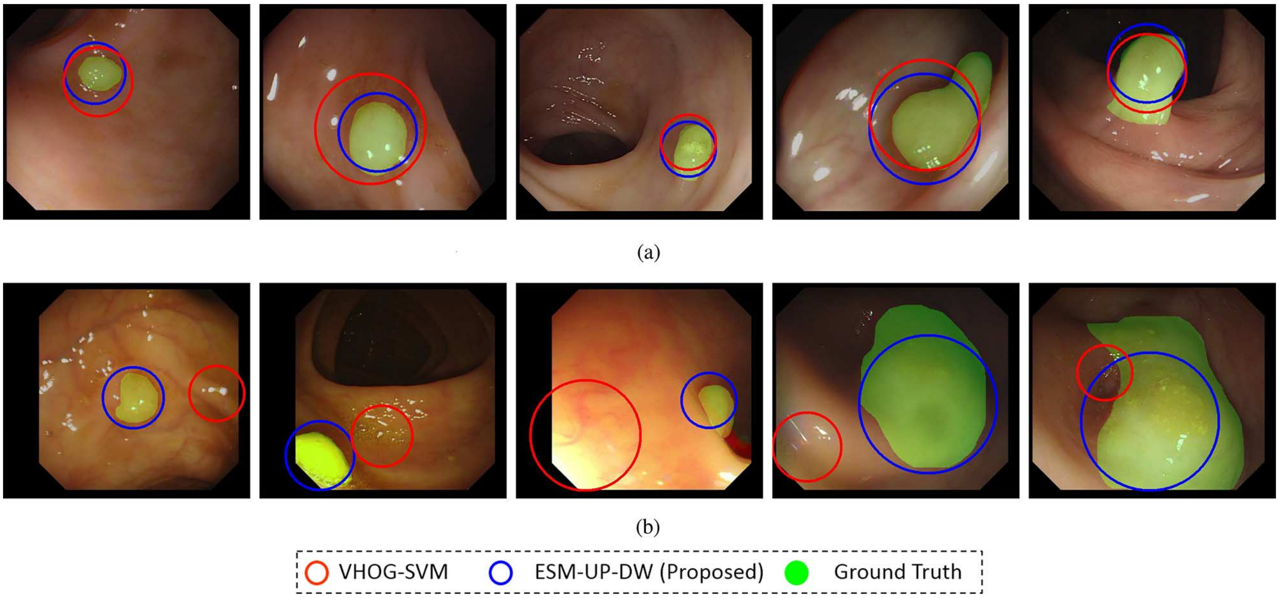


Fig. 12. For the CVC-Colon dataset, ground truth and detection regions of the VHOG [29]-SVM (A6) and the proposed (B7) detectors are depicted. (a) Accurately detected polyps by both detectors. (b) Inaccurately detected polyps by the VHOG [29]-SVM (A6) detector.

TABLE II
PARAMETERS USED FOR EVALUATION

Parameter	Range	Interval
K: Number of weak classifiers	[1.0 10.0]	1.0
τ: Score threshold	[-1.1 0.3]	0.2
σ: Window size scaling factor	[0.1 2.0]	0.1
C: Cost ratio	[0.1 0.9]	0.1

in Section IV.B, the amount of down-sampling is automatically determined by the Tomek-linked pairs but it deepens on the data distribution[14]. To show the effect of the proposed data sampling, we further implemented a detector based on random sampling: RDW randomly selects samples of the majority class and remove them, whereas RUP randomly selects samples of the minority class and copy them. In RUP, the amount of up-sampling is set to 200% as used in the proposed up-sampling. Using RDW we make the sample ratio between both classes classes 1.

In addition, we evaluated the detectors using reweighting-based imbalanced learning in Table III.C. Their common way is to assign the higher weights to the samples of the minority class for balancing the dataset. To this end, in ESM-Weighting, we multiply the weight of each minority sample by $\sum_{i,y_i=-1} D_k(i)/\sum_{i,y_i=1} D_k(i)$, and make the total sample weight of both classes the same (i.e., $\sum_{i,y_i=1} D_k(i) = \sum_{i,y_i=-1} D_k(i)$). By inserting the cost item C_i into the weight update formula of Adaboost with different ways, we implemented three cost-sensitive boosting [18]: AdaC1, AdaC2 and AdaC3.

From the results in Tables III.B and III.C, we confirm that the performance of the detectors using the proposed data sampling (B5-B7) is superior to that of other detectors without sampling (B1) and with random sampling (B2-B4). In particular, we can see that the random up/down sampling degrades the detection performance from the scores of (B1) and (B2-B4). Compared to other reweighting-based detectors (C1-C4), the proposed detector (B7) shows much better performance in terms of all the metrics. Furthermore, the proposed detector (B7) greatly improves the precision rate while producing a similar recall rate, compared to the segmentation-based detector [6]. Notably, the speed of our detector is roughly 30 times faster than [6]. The performance improvement proves the effectiveness and robustness of the proposed polyp detector.

Fig. 12 shows detection results using the VHOG [29]-SVM (A6) and the proposed detector (B7). We further illustrate ground truth regions of polyps for the better comparison. As shown in Fig. 12(a), both detectors can successfully detect locations and sizes of polyps in endoscopic images. However, false detections are occurred by the VHOG [29]-SVM (A6) when specular reflection by camera lights exists, polyps are occluded, and shapes of polyps are nonspherical as in Fig. 12(b). On the other hand, the proposed detector (B7) produces accurate detection results even in such cases.

F. Evaluation of Proposed Methods

For verifying the effect of the proposed feature learning and imbalanced learning methods, we implemented the SVM, ESM, ESM-DW, ESM-UP and ESM-UP-DW detectors (described in Section V.E2) with CVC-Colon and our datasets. From the combined dataset, total 1642 polyp images were obtained. Based on the 3-fold cross-validation, we evaluated the performance of detectors in Figs. 13 and 14 (all with VHOG($s = 16$)-PLS except for Fig. 14(a)).

We use the same bandwidth factor $\sigma = 1$ for all the experiments. Since the distribution of the confidence scores of a detector is usually different for each test image, we use the non-fixed $\tau = \{\eta | \eta \geq -0.5\}$, where η is the minimum value of the confidence scores of the image pyramid.

---

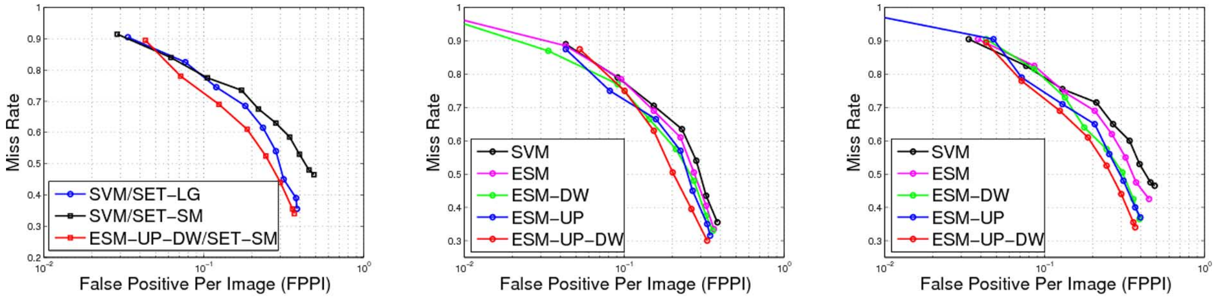
[14] In our experiment, the means of the removed samples are almost 1.19%, 1.65% and 5.97% at each iteration when using VHOG-PLS, VHOG-PCA, and VHOG, respectively.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

TABLE III

PERFORMANCE COMPARISON WITH DIFFERENT FEATURES AND IMBALANCED LEARNING METHODS. THE PROPOSED FEATURE LEARNING AND IMBALANCED LEARNING METHODS ARE MARKED WITH BLUE. FOR EACH METRIC, THE BEST RESULTS ARE MARKED WITH RED. IN EACH SUB-TABLE (A-D), THE PERFORMANCE OF THE DETECTOR WITH THE HIGHEST PR-AUC IS HIGHLIGHTED

Detector	**Feature**	**Classifier**	**GT**	**TP** ↑	**FP** ↓	**FN** ↓	**Precision** ↑	**Recall** ↑	**PR-AUC** ↑	**Speed** (sec/frame)	**Parameter** ($K/\tau/\sigma/C$)
A. Multiple features with a single SVM classifier											
A1	Wavelet[1]	SVM	300	104	1426	196	6.80%	34.67%	2.56%	41.33	1/-0.9/0.3
A2	Color $(8 \times 8 \times 24 = 1536)$	SVM	300	100	975	200	9.30%	33.33%	3.61%	2.867	1/-0.5/0.5
A3	LBP [38] $(8 \times 8 \times 58 = 3712)$	SVM	300	165	160	135	50.77%	55.00%	41.85%	2.922	1/0.3/1.1
A4	HOG [9] $(16 \times 16 \times 36 = 9216)$	SVM	300	116	317	184	26.79%	38.67%	14.12%	26.696	1/-0.1/0.9
A5	VHOG(s = 8) [29] $(16 \times 16 \times 31 = 7936)$	SVM	300	182	137	118	57.05%	60.67%	45.34%	22.68	1/-0.3/1.1
A6	VHOG(s = 16) [29] $(8 \times 8 \times 31 = 1984)$	SVM	300	192	131	108	59.44%	64.00%	49.09%	1.221	1/-0.1/1.3
A7	VHOG(s = 8)-PLS $(7936 \rightarrow 7)$	SVM	300	194	151	106	56.23%	64.67%	53.68%	1.686	1/0.1/1.1
A8	VHOG(s = 16)-PCA $(1984 \rightarrow 25)$	SVM	300	183	126	117	59.22%	61.00%	50.38%	0.566	1/-0.7/1.5
A9 (*with proposed feature*)	VHOG(s = 16)-PLS $(1984 \rightarrow 7)$	SVM	300	194	111	106	63.61%	64.67%	54.03%	0.330	1/0.1/0.9
B. Resampling-based imbalanced learning methods (with the proposed feature VHOG(s=16)-PLS)											
B1	VHOG-PLS	ESM	300	208	102	92	67.10%	69.33%	61.35%	0.761	10/-0.7/1.1
B2	VHOG-PLS	ESM-RDW	300	186	113	114	62.21%	62.00%	53.29%	0.556	9/-0.7/1.7
B3	VHOG-PLS	ESM-RUP	300	182	119	118	60.47%	60.67%	53.60%	0.816	10/-0.7/1.7
B4	VHOG-PLS	ESM-RUP-RDW	300	197	103	103	65.67%	65.67%	57.37%	0.700	10/-0.7/0.7
B5	VHOG-PLS	ESM-DW	300	204	96	96	68.00%	68.00%	61.76%	0.654	10/-0.5/0.7
B6	VHOG-PLS	ESM-UP	300	209	96	91	68.52%	69.67%	63.27%	0.657	10/-0.5/1.1
B7 (*with all*)	VHOG-PLS	ESM-UP-DW	300	212	88	88	70.67%	70.67%	65.43%	0.652	10/-0.1/1.3
C. Reweighting-based imbalanced learning methods (with the proposed feature VHOG(s=16)-PLS)											
C1	VHOG-PLS	ESM-Weighting	300	204	150	96	57.63%	68.00%	56.65%	0.568	7/0.1/0.9
C2	VHOG-PLS	AdaC1[18]	300	210	125	90	62.69%	70.00%	62.18%	0.909	10/-0.1/1.1/0.1
C3	VHOG-PLS	AdaC2[18]	300	196	117	104	62.62%	65.33%	56.52%	0.574	8/-0.1/1.3/0.4
C4	VHOG-PLS	AdaC3[18]	300	197	103	103	65.67%	65.67%	58.34%	0.656	6/-0.7/1.5/0.6
D. Segmentation-based detector											
Bernal *et al.* [6]	SA-DOVA	Watersheds	300	215	241	85	47.15%	71.67%	—	19	—



(a) (b) (c)

Fig. 13. Performance evaluation for different imbalanced polyp datasets. Miss rates against FPPIs of the detectors are plotted on a log-log scale. (a) With both datasets, (b) With SET-LG, (c) With SET-SM.

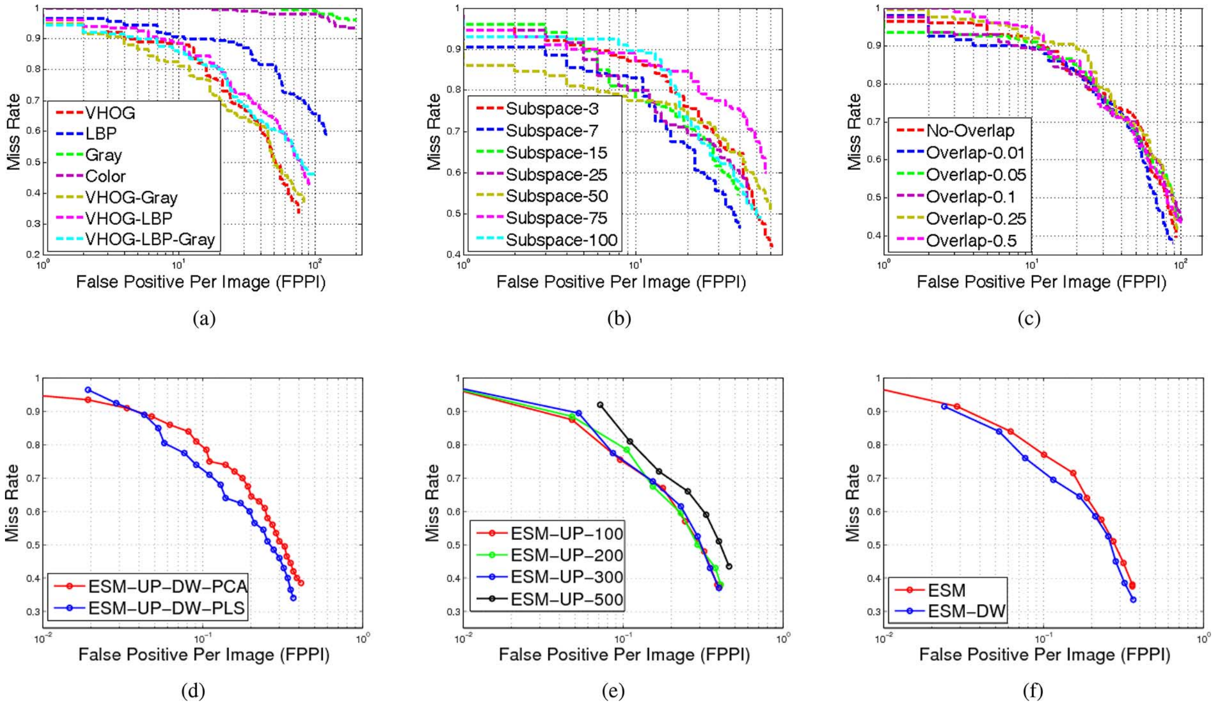
1) Evaluation Using Different Datasets: Since we focus on the classification with imbalanced datasets, we generated two training sets having different imbalanced ratios between positive and negative samples at each cross-validation step:

- SET-LG: The imbalanced ratio is 1:10 (1000 positive (i.e., polyps) and 10000 negative (i.e., non-polyps) samples).
- SET-SM: The ratio is 1:100 (100 positive and 10000 negative samples). Note this dataset is more imbalanced and contains fewer positive samples.

The performance of the detectors using different imbalanced learning is compared in Fig. 13. Remarkably, the proposed ESM-UP-DW achieves the best performance with the more imbalanced SET-SM as shown in Fig. 13(a). Fig. 13(b) and (c) also affirm that the performance of the ESM-UP-DW is superior to others regardless of dataset sizes. Especially, Fig. 13(c) clearly shows the performance gaps of the detectors when using small positive samples (SET-SM). These evaluation results prove that the proposed (up- and down-) sampling methods make the imbalanced sets be rebalanced, and can significantly improve the performance with only a tiny number of positive samples.

2) Method Evaluation: To prove feasibility of the used VHOG, we trained several SVM-based detectors using different features, and compared their performance in Fig. 14(a). When extracting the VHOG [29], LBP [38], gray and color

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c)

(d) (e) (f)

Fig. 14. For performance analysis of the proposed methods, we plot miss rates against FPPIs on a log-log scale. (a) Multi-feature, (b) The dimension of subspaces of PLS, (c) The amount of overlap area, (d) PCA vs PLS, (e) Up-sampling, (f) Down-sampling.

histograms, we use the same cell size $s = 16$[15]. For more evaluation, we further generated combined features with different features and learned detectors using them.

As can be seen, the detector using the VHOG feature achieves the much better performance than using LBP, gray and color features. By using the VHOG-Gray feature combination, we can slightly improve detector performance compared when using only the VHOG feature. However, extracting the multiple features also increases computational complexity[16]. These comparison results affirm that the VHOG feature is the most suitable for polyp detection in consideration of both performance and computation.

For determining the optimal dimension of the PLS matrix, we trained several VHOG-PLS/SVM detectors with different PLS dimensions and compared them in Fig. 14(b). We obtain the best rate using the 7 dimension of PLS.

Fig. 14(c) compares the performance of VHOG-PLS/SVM detectors trained with the different amount of overlap areas between positive and negative samples. When using negative samples overlapped with positive one less than 0.01, the detector shows the best performance.

In Fig. 14(d), we use different subspace learning methods (PCA and PLS) for ESM-UP-DW detectors. We also see that employing PLS provides the better performance than PCA. Fig. 14(e) compares performance of ESM-UP detectors learned with different L, the amount of up-sampling. The performance is almost the same for $L = 100, 200$, and 300 but decreased for $L = 500$. It implies the proposed up-sampling is not sensitive to L, but too much oversampling can degrade the performance since a detector is likely to be over-fitted. In Fig. 14(f), the ESM-DW shows the slight performance improvement, compared to ESM. However, exploiting down-sampling together with up-sampling can achieve the better performance when comparing ESM-UP and ESM-UP-DW in Table III and Fig. 13(c).

Some polyp detection results are shown in Fig. 15. There exist many types of polyps in which they have different shapes. In addition, some polyps are occluded and have very similar colors and textures with other non-polyps. However, the proposed ESM-UP-DW successfully detects the polyps. These results also support the performance improvements of our methods. Our polyp detector was implemented using MATLAB on a PC with 3.07 GHZ CPU without parallel programming. For each image, the total running time of the ESM-UP-DW detector is about 0.6375 (sec/frame). When using one single classifier, the computation cost can be greatly reduced by almost 50%.

VI. DISCUSSION OF ADABOOST AND CASCADE SCHEME

It is worthy of note that a cascade scheme [40] can be also applied to ensemble learning with imbalanced datasets instead of Adaboost used for our work. In general, both schemes aim at reducing the false positive rate while maintaining the high detection rates. We compared their performance using the combined dataset based on the 3-fold cross-validation as described in Section V.F. We implemented a cascade polyp detector consisting of 3 layers. At each layer, we trained a SVM classifier using VHOG-PLS and then found its optimal threshold τ and window size σ, which make the recall greater than 0.90%.

As shown in Table IV, Adaboost achieves a much higher rate. The main reason is that they have different ways of aggregating

---

[15] VHOG, LBP and color histogram feature have the 1984, 3712, and 1536 dimensions as described in Section V.E1. We extract a 512-dimensional gray feature after concatenating 8-dimensional histograms of 64 cells.

[16] In our experiment, running time of the detector using the VHOG feature is almost two times faster that using the VHOG-Gray.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

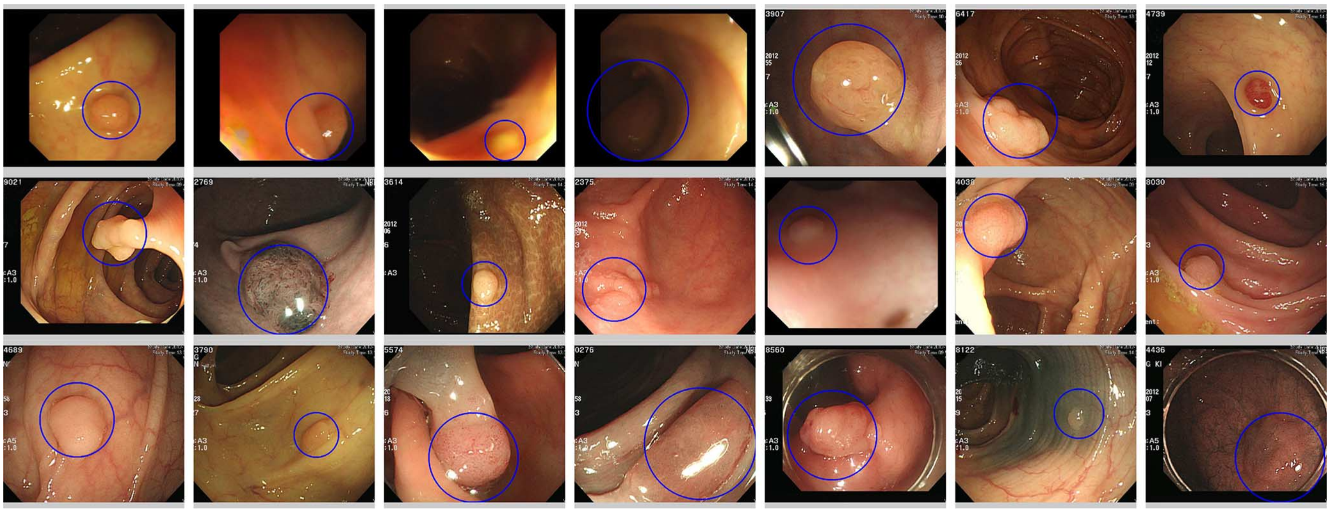


Fig. 15. For CVC-Colon and our datasets, polyp detection results using the ESM-UP-DW detector trained with SET-SM are shown. More results can be found in our supplementary material.

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT SCHEMES

Scheme	Layer	PR-AUC	Parameter (τ/σ)
Cascade	1	1.80%	-1.0/7.5
	2	11.03%	-0.5/3.5
	3	34.55%	0.0/1.5
Adaboost	(3 weak classifier)	50.49%	-0.05/1.5

scores of weak classifiers: In Adaboost, all test samples are evaluated by the learned classifiers and all the scores of them are used when computing the final detection scores. On the other hand, in the cascade, test samples are sequentially filtered out by each classifier at each level. In other words, only remaining samples after filtering out at the previous levels are evaluated at the current level. Therefore, by using AdaBoost, we can enhance the overall accuracy by considering more decisions of classifiers when aggregating scores.

VII. CONCLUSION

In this paper, we tackled an automated polyp detection problem in endoscopic images. This problem is still challenging because only a small amount of polyp samples is available in spite of the diversity of polyps. For this reason, a polyp dataset is usually prone to be imbalanced and learning with the imbalanced dataset generates a classifier biased toward a majority class. Another difficulty of the polyp detection is that polyps and non-polyps have very similar colors and textures. Therefore, it is difficult to discriminate between them using the conventional features.

For learning an unbiased detector from the imbalanced dataset, we proposed the data sampling-based boosting framework with up/down sampling. As a result, we can learn ensemble classifiers with the rebalanced dataset and use them in order to detect different types of polyps. In addition, for better discrimination, we proposed the effective feature learning method using PLS analysis. Using the learned compact and discriminative feature, we can improve the detection performance while reducing the complexity.

The experimental results showed the enhanced performance of the proposed detector, compared to other state-of-the-art detectors. We further proved usefulness and effectiveness of the proposed methods by implementing and comparing several versions of detectors with different features and imbalanced learning. We believe that the proposed methods can be applicable to other detection problems and other applications such as object tracking and classifications.

REFERENCES

[1] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, “Computer-aided tumor detection in endoscopic video using color wavelet features,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 3, pp. 141–152, Sep. 2003.

[2] M. P. Tjoa and S. M. K. , “Feature extraction for the analysis of colon status from the endoscopic images,” *Biomed. Eng. Online*, vol. 2, no. 9, 2003.

[3] B. V. Dhandra, R. Hegadi, M. Hangarge, and V. S. Malemath, “Analysis of abnormality in endoscopic images using combined HSI color space and watershed segmentation,” in *Int. Conf. Pattern Recognit.* , 2006, pp. 695–698.

[4] M. T. Coimbra and J. P. da Silva Cunha, “MPEG-7 visual descriptors—Contributions for automated feature extraction in capsule endoscopy,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 16, no. 5, pp. 628–637, May 2006.

[5] L. A. Alexandre, J. Casteleiro, and N. Nobreinst, “Polyp detection in endoscopic video using svms,” in *PKDD*, 2007, pp. 358–365.

[6] J. Bernal, F. J. Sánchez, and F. Vilariño, “Towards automatic polyp detection with a polyp appearance model,” *Pattern Recognit.*, vol. 45, no. 9, pp. 3166–3182, 2012.

[7] J. Marín, D. Vázquez, D. Gerónimo, and A. M. López, “Learning appearance in virtual scenarios for pedestrian detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 137–144.

[8] S. Munder and D. M. Gavrila, “An experimental study on pedestrian classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1863–1868, Nov. 2006.

[9] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Comput. Vis. Pattern Recognit.*, 2005, pp. 886–893.

[10] M. Galar, A. Fernández, E. B. Tartas, H. B. Sola, and F. Herrera, “A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches,” *IEEE Trans. Syst., Man, Cybern., Part C*, vol. 42, no. 4, pp. 463–484, Jul. 2012.

[11] R. Rosipal and N. Krämer, “Overview and recent advances in partial least squares,” in *SLSFS*, 2005, pp. 34–51.

[12] S. Hwang, J.-H. Oh, W. Tavanapong, J. Wong, and P. C. de Groen, “Polyp detection in colonoscopy video using elliptical shape feature,” in *Proc. IEEE Int. Conf. Imag. Proess.*, 2007, pp. 465–468.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.

[13] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[14] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Comput. Vis. Pattern Recongnit.*, 2009, pp. 304–311.

[15] T. R. Hoens, N. V. Chawla, H. He, and Y. Ma, *Imbalanced Datasets: From Sampling to Classifiers*. New York: Wiley-IEEE Press, 2013.

[16] W. Fan, S. J. Stolfo, J. Zhang, and P. K. Chan, "AdaCost: Misclassification cost-sensitive boosting," in *Int. Conf. Mach. Learn.*, 1999, pp. 97–105.

[17] K. M. Ting, "A comparative study of cost-sensitive boosting algorithms," in *Int. Conf. Mach. Learn.*, 2000, pp. 983–990.

[18] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognit.*, vol. 40, no. 12, pp. 3358–3378, 2007.

[19] H. Masnadi-Shirazi and N. Vasconcelos, "Cost-sensitive boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 294–309, Feb. 2011.

[20] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," in *Int. Conf. Mach. Learn.*, 1997, pp. 179–186.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[22] S. Hu, Y. Liang, L. Ma, and Y. He, "MSMOTE: Improving classification performance when training data is imbalanced," in *Int. Workshop Comput. Sci. Eng.*.

[23] H. Guo, "Learning from imbalanced data sets with boosting and data generation: The databoost-IM approach," *SIGKDD Explorations*, vol. 6, 2004.

[24] X. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 931–937, May 2009.

[25] H. Qu, G. Li, and W. Xu, "An asymmetric classifier based on partial least squares," *Pattern Recognit.*, vol. 43, no. 10, pp. 3448–3457, 2010.

[26] M. You, R. Zhao, G. Li, and X. Hu, "MAPLSC: A novel multi-class classifier for medical diagnosis," *Int. J. Data Min. Bioinform.* , vol. 5, no. 4, pp. 383–401, 2011.

[27] D. Chung and S. Keles, "Sparse partial least squares classification for high dimensional data," *Stat. Appl. Genetics Mol. Biol.*, vol. 9, no. 1, pp. 1–32, 2010.

[28] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Int. Conf. Comput. Vis.*, 2009, pp. 32–39.

[29] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[30] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemom.*, vol. 17, no. 3, pp. 166–173, 2003.

[31] D. V. Nguyen and D. M. Rocke, "Tumor classification by partial least squares using microarray gene expression data," *Bioinformatics*, vol. 18, no. 1, pp. 39–50, 2002.

[32] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, "Object tracking via partial least squares analysis," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4454–4465, Oct. 2012.

[33] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *Int. Conf. Comput. Vis.*, 2009, pp. 24–31.

[34] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel Hilbert space," *J. Mach. Learn. Res.*, vol. 2, pp. 97–123, 2001.

[35] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 2006.

[36] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, no. 1–2, pp. 1–39, 2010.

[37] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," [Online]. Available: http://www.vlfeat.org/ 2008

[38] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[39] F. M. Porikli, "Integral histogram: A fast way to extract histograms in Cartesian spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 829–836.

[40] P. A. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

Authorized licensed use limited to: Inha University. Downloaded on November 06,2020 at 08:32:24 UTC from IEEE Xplore. Restrictions apply.