RESEARCH ARTICLE

Distribution-Aware Loss for Lesions Detection Using White-Light Endoscopy in Colorectal Region

WEI SUN[1], RUIQI ZHAO[1], KUNPENG ZHANG[1], JUNBO GAO[1], AND GUOQIANG QU[2]

[1]College of Information Engineering, Shanghai Maritime University, Shanghai 201308, China
[2]Shanghai Sixth People's Hospital, Shanghai 200235, China

Corresponding author: Wei Sun (weisun@shmtu.edu.cn)

ABSTRACT Computer-aided diagnostic systems have evolved into critical tools for endoscopists in diagnosing and reducing missed diagnoses. However, due to the lower incidence of rare diseases in medical images compared to common diseases, there exists an imbalance in sample distribution for lesion classification. This imbalance results in reduced detection and classification accuracy. In the realm of deep learning, detection accuracy is influenced not only by the model but also by the choice of the loss function.This study introduces a novel solution to address the imbalance issue of colorectal lesions in white light endoscopy by proposing a loss function named Label Distribution and Scale Distribution Aware Loss (LSDA-Loss). Our innovative loss function resolves the category imbalance problem by considering sample distribution and employing Bayesian equations to quantify the degree of imbalance. Furthermore, we adopt proportional distribution to evaluate the complexity of categorizing each sample. Experimental results from three independent datasets demonstrate that: 1) the integration of the proposed loss function with three typical FPN models significantly enhances detection accuracy, achieving improvements of up to 94.56%. 2) Our loss function effectively balances detection accuracies across the three categories, surpassing the performance of the original loss function.

INDEX TERMS Colorectal cancer, deep learning, medical image process, white-light endoscopy.

I. INTRODUCTION

Colorectal cancer is ranked third globally in cancer incidence rates and is associated with high mortality. Recent trends indicate a progressive decrease in the age of onset of colorectal cancer, imposing significant health and economic burdens on populations. Approximately 95% of colorectal cancer cases originate from polyps in the rectum or colon lining, with certain types like adenomatous polyps carrying the potential to advance to colorectal cancer. Early detection of colorectal cancer yields a cure rate as high as 90% [1] Notably, early-stage patients often present no overt clinical symptoms, prompting the reliance on colorectal cancer screening and the expertise of seasoned endoscopists for detection. Presently, white light endoscopy serves as the predominant method for colorectal cancer screening, with multiple studies illustrating its efficacy in reducing morbidity and mortality rates associated with the disease. High-quality white light endoscopy is linked to improved prognostic outcomes for patients [2] However, disparities exist in the quality of white light endoscopy among providers and endoscopists, as evidenced by studies reporting false-negative rates for colonoscopy ranging from 22% to 28%. Such discrepancies can result in delayed colon cancer diagnoses and alarming survival rates as low as 10% [3] Thus, enhancing early screening for colorectal cancer necessitates the integration of innovative technological tools to facilitate lesion detection and mitigate the risk of delayed diagnoses.

In recent years, deep learning models have achieved significant advancements in various medical imaging domains, surpassing conventional diagnostic methods in tasks like image detection and classification. Within the realm of

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

intelligent medicine [4], there is a growing acceptance and utilization of AI-assisted diagnosis by doctors, patients, and technicians. Leveraging deep learning models to support medical diagnosis can enhance detection accuracy, addressing missed diagnoses attributed to the inexperience and visual fatigue commonly observed among colorectaloscopists. Notably, the field has seen enhanced detection and classification accuracy following pivotal works such as the utilization of a convolutional network (CNN) based deep learning model for polyp detection by Kumar et al. [5] Pogorelov et al. [6] pioneered the integration of deep neural networks, information retrieval, and global and local image characterization for multi-class classification, detection, and localization of various colorectal lesions. Yhandah et al. [7] achieved a notable 75.1% accuracy in classifying adenomatous and non-adenomatous polyps using CNN models with ten-fold cross-validation. Wimmer et al. [8] applied three pre-trained CNN architectures to endorse the classification of colonic polyps using SVM techniques. Byrne et al. [9] also distinguished adenomas from hyperplastic polyps with high precision based on a CNN model with 83% specificity and 90% positive predictive value (PPV).Gao et al. [10] effectively employed ResNet50 to distinguish colonoscopy images containing lesions from those without lesions, subsequently categorizing specific lesions into adenomas, carcinomas, and polyps with an impressive AP50 detection value of 0.903. Yuan et al. [11] introduced the RIS-DenseNet network, enhancing polyp image recognition accuracy by incorporating rotational invariance and image similarity principles into the algorithm.

Deep learning datasets must be sufficiently large to encompass a diverse array of scenarios and variations pertinent to the task, ensuring effective generalization of the model to novel data. Public datasets commonly comprise hundreds of thousands to millions of samples. Unlike generic images, acquiring medical images is challenging due to ethical constraints governing the image-capture process, resulting in a scarcity of medical image datasets. Moreover, the prevalence of rare diseases is lower compared to common ailments. For instance, in white light endoscopic images of colorectal lesions, the quantity of samples for polyps and carcinomas is notably fewer than for adenomas. This imbalance in sample distribution across categories poses a challenge by biasing the model's detection accuracy towards the majority categories, often neglecting the minority [12], [13]. However, in real-world clinical settings, a low polyp detection rate within any category could impede the accurate diagnosis of colorectal cancer. Analyzing white light endoscopic images of colorectal lesions reveals that polyps and adenomas typically constitute a small portion of the entire lesion image, leading to a substantial imbalance in pixels. This foreground-background disparity prompts trained models to inaccurately favor predicting background categories over actual lesions, presenting a significant hurdle for deep learning in medical image detection.Previous studies have highlighted the potential of deep learning in detecting col-

orectal polyps in white-light endoscopic images [14]. Now, the quest for effective methods to enhance the performance of deep models on imbalanced datasets has garnered increasing attention from researchers.

Neural networks in deep learning are utilized for learning and predicting input data through multi-level nonlinear transformations and weight updates, guided by feedback generated from the loss function to optimize parameters. The most widely used loss function in multiclassification is cross-entropy due to its feasibility for micro-optimizable optimization. However, the traditional cross-entropy loss function treats each data instance equally, overlooking the imbalance in category categorization, specifically in white light endoscopic images of colorectal lesions. As a result, a dynamic weighting strategy is proposed in this study, assigning weights to different categories with a focus on challenging samples, alongside introducing a novel loss function termed LSDA-Loss. This new approach addresses the category imbalance issue by considering sample distribution and utilizing a proportional distribution to assess the classification difficulty of each sample. To showcase the effectiveness of the proposed method, experiments are conducted using three prominent single-stage deep learning models.

The remainder of the paper is structured as follows: Section II provides a concise overview of prior research efforts addressing the category imbalance issue. Section III elaborates on the LSDA-Loss function and the network component of the experimental model. Section IV assesses the method's efficacy and performance through experimental evaluation, while Section V offers the paper's concluding remarks.

II. RELATED WORK
A. CLASS IMBLANCE ON IMAGES
The concept of category imbalance pertains to the unequal distribution of samples across various target categories within the training dataset. Such imbalanced datasets pose challenges throughout all phases of target detection and significantly impact its performance. Currently, two primary types of category imbalance issues exist:

PN-class Imbalance: This imbalance concerns the unequal representation of positive samples (samples containing the target) and negative samples (samples without the target). For target detection tasks, negative samples typically outnumber positive samples as the background region exceeds the target region in size. This disproportion can lead the model to excel in classifying negative samples but struggle in accurately identifying positive samples, thereby elevating the likelihood of false detections.

PP-class Imbalance: This form of imbalance reflects a substantial variance in the sample quantities across distinct target categories. Certain target categories may be overrepresented compared to others, creating an uneven sample distribution. This imbalance obstructs a comprehensive grasp of targets within specific categories during the learning phase,

ultimately diminishing the model's detection accuracy. Consequently, the overall performance of the target detection model is impacted by the disparity in categorizing sample difficulty.

Sample classification difficulty imbalance is a common issue in target detection, whereby there exists a notable disparity in the distribution of easy and challenging samples within the training dataset. When the proportion of easy samples surpasses that of difficult ones in the training data, the model tends to overly focus on learning and optimizing easy samples at the expense of neglecting the difficult ones. Consequently, this compromises the model's capacity to generalize in real-world scenarios. Recognizing the complexity of sample classification aids in assessing model performance and devising suitable optimization strategies to enhance target detection accuracy.

B. SOLUTION METHODS

Currently, the primary approaches to address the issue of category imbalance in deep learning models are data-level techniques and algorithmic strategies.

There are various techniques at the data level. Undersampling [15] techniques can be used to construct a relatively sample-balanced dataset by filtering the majority of the samples.There is also the option of oversampling [16] a small number of samples, i.e., selectively copying some of the samples to make the number balanced, but this can lead to overfitting of some of the samples after training.Nowadays, the commonly used method is resampling [17], and SMOTE [18] is one of the state-of-the-art methods. It is a dataset balancing technique that effectively solves the data imbalance problem by generating new samples between a few classes of samples through an innovative interpolation method.Unlike the methods described above, SMOTE does not simply replicate, but algorithmically generates new samples. These additional samples play a crucial role in enhancing the representation of minorities, ultimately improving the performance of the model.

Algorithm level based methods are mainly used to solve the category imbalance problem effectively by adapting the modeling process of the classifier and enhancing some of the samples through the loss function.

Target detection produces a large number of anchor frames but few positive image samples, resulting in an extreme imbalance between positive and negative samples.To address this challenge, Lin et al. [19] introduced the Focal Loss method in 2017. This technique alleviates the imbalance by augmenting the weight assigned to positive example samples within the cross-entropy loss function. Li et al. [20] devised the GHM loss function, which dynamically adjusts sample weights by applying gradients proportionally to the sample distribution. Specifically, easy samples with low gradients receive reduced weights, while difficult samples with moderate gradients see weight increments, and outlying samples with high gradients have their weights decreased.Unlike

Focal Loss and GHM, the PISA [21] method uses a ranking method called IoU-HLR to evaluate the sample difficulty based on Mean Average Precision (MAP). It effectively suppresses the scores of unimportant samples, thereby enhancing the emphasis on crucial samples. Qian et al. [22] introduced the DRLoss loss function, which categorizes foreground elements by analyzing the candidate background frames' distribution and adjusts their weights accordingly to achieve a distribution associated with the worst-case loss scenario. This approach concentrates the loss function on the decision boundaries between foreground and background distributions, diminishing logarithmic sorting and enhancing computational efficiency. Li et al. [23] proposed the Generalized Focal Loss (GFL) to tackle the mismatch between training and prediction processes. They further developed two variants: Quality Focal Loss (QFL), which targets a sparse set of challenging samples and generates continuous quality estimates ranging between 0 and 1 for the relevant categories, and Distribution Focal Loss (DFL), which enables the network to concentrate on learning probability values in the proximity of continuous target bounding box positions across flexible distributions.Chen et al. [24] put forward the AP-loss as a novel ranking method instead of a categorization loss to boost the performance enhancement in single-stage target detection. Zhang et al. [25] introduced the IOU perceptual function of categorization score (IACS) for Non-Maximum Suppression (NMS) sorting, indicating object category presence and enhancing bounding box localization accuracy. Yu et al. [26] devised a double balanced loss function based on Bayesian principles to address the data imbalance issue among positive example samples. Sadi et al. [27] proposed the LDAM loss function to reduce model generalization error by imposing stronger regularization on the minority class compared to the majority class. Ozan et al. [28] introduced the Omni-Comprehensive Loss, a new loss function that leverages NCC and Tversky's index to mitigate dataset imbalance and the issue of gradient vanishing. Saiji et al. [29] introduced the DLINEX loss function, which innovatively incorporates the exponential loss function from statistics into a deep learning framework to combat category imbalance by fine-tuning parameters to focus more on scarce and challenging-to-categorize samples.

In summary, there are two main approaches to solving the class imbalance problem. As mentioned earlier, classifier-level techniques aim to enhance the loss function. These techniques can be divided into two categories: assessing the complexity of sample classification (sample-weight) and balancing the representativeness of a small number of classes (class-weight), as shown in Table 1.

III. METHODS

This section focuses on addressing the imbalance issue observed in white-light endoscopic images of colorectal lesions. We provide a brief overview of the Feature Pyramid Network (FPN) framework for feature extraction in lesion

TABLE 1. Innovative improvements of loss function by scholars over the years.

Loss function	Years	Class imbalance category	Strategy category	Innovations
Focal Loss [19]	2017	PN-class imbalance	sample-weight	Judgment by confidence level as a criterion for hard and easy samples
GHM [20]	2018	PN-class imbalance	sample-weight	Judgement by the sample gradient as a criterion for hard and easy samples
PISA [21]	2019	PN-class imbalance	sample-weight	Use Map to measure the difficulty of the sample
DRLoss [22]	2019	PP- class imbalance	sample-weight	Use the sample on the decision boundary as a hard sample
GFocal Loss [23]	2020	PP- class imbalance	sample-weight	Extending the discrete version of Focal Loss to a continuous variant
AP-Loss [24]	2020	PP- class imbalance	sample-weight	Use ranking method instead of classification loss
Varifocal loss [25]	2021	PP- class imbalance	sample-weight	Propose a classification score for Iou perception function
DBLoss [26]	2022	PP- class imbalance	class-weight	A double equilibrium loss function based on Bayesian ideas
LDAM [27]	2022	PP- class imbalance	class-weight	Introduce stronger regularization for minority classes than for majority classes
Omni-Comprehensive Loss[28]	2022	PN-class imbalance	sample-weight	Handling polyp shape-awareness situations as well as improving recall through parameterization
DLINEX[29]	2023	PP- class imbalance	class-weight	It is possible to adaptively focus more on the few and difficult to categorize instances, and it is possible to achieve both inter- and intra-class diversity.

detection and introduce a novel loss function named Labeled Distribution and Scale-Distribution Perceptual Loss (LSDA-Loss). LSDA-Loss combines a labeling distribution and a scale distribution to assign different weights to the samples in each category and uses the target scale to identify difficult samples.

A. FPN
Convolutional neural network models in the field of deep learning are extensively employed for detecting lesions in medical images. These networks extract intricate hierarchical features by amalgamating local features through convolutional learning and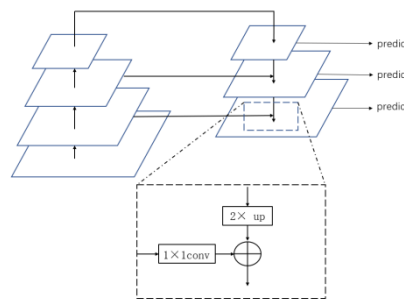 spatial pooling operations. During feature extraction, variations in features across layers occur. The Feature Pyramid Network (FPN) [30] addresses this by merging feature maps from lower to higher layers, thereby maximizing the utilization of features from various network layers and enhancing the detection of small targets in target identification tasks.

FIGURE 1. FPN feature fusion.

The feature fusion operation of FPN, illustrated in Figure. 1, entails upsampling the low-resolution feature maps by 2x using the nearest neighbor method and summing them element-wise. Subsequently, the corresponding low-resolution feature maps undergo a 1×1 convolutional layer to adjust the channel count to align with the upsampled feature maps, and these enhanced feature maps are combined to form a high-resolution feature map. This process is iterated multiple times until the desired new feature maps are generated. Finally, each fused feature map undergoes a 3×3 convolution operation to yield the ultimate feature map.

B. DIFFICULTY OF SAMPLE CLASSIFICATION
In previous research on addressing imbalanced problems, the prevalent loss function utilized is the Focal Loss. The Focal Loss assesses difficult samples based on the confidence level magnitude. Specifically, for positive samples, the model assigns samples with low confidence as difficult, while samples with high confidence (easy samples) have a negligible impact on enhancing model performance. A modulation factor, denoted as 't', diversely influences the contribution of various samples to the loss, thereby enabling the model training to prioritize difficult, misclassified samples. The complexity of samples is evaluated by examining the scales of the three categories of real white light endoscopic colorectal lesions (as depicted in Figure. 2, with the scale distribution utilized as a criterion for assessing sample difficulty.

After analyzing data on the distribution of lesions in various white light endoscopy image databases, the results are presented in Figure 3. The samples were categorized into small-scale and large-scale samples using a threshold of 0.3 based on the lesion area percentage relative to the whole image. In the large-scale samples, cancerous structures accounted for over 80%, adenomas for 10%, and polyps for the lowest percentage. Therefore, the detection and classification of cancerous structure samples are more prominent in the test set. Conversely, in the small-scale samples, polyps
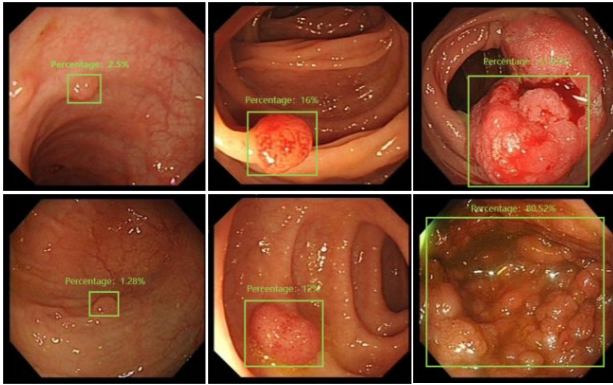
FIGURE 2. Labeled percentage of polyp, adenoma, and carcinoma-like lesions in the overall image in the dataset.
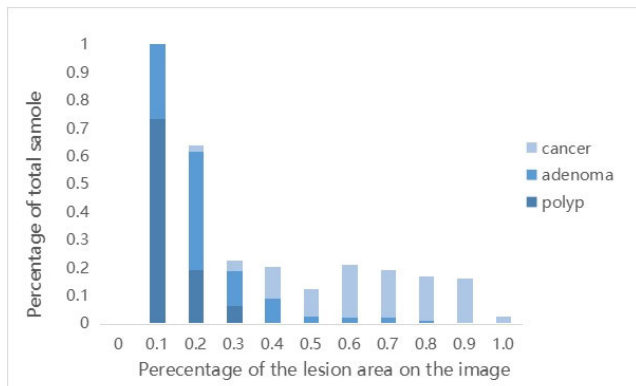

FIGURE 3. Proportion of scale occupied by polyps, adenomas, and cancer lesions.

constituted 40%, adenomas 30%, and cancers 10%, with each category being relatively balanced, posing challenges for classification. It is crucial to highlight that inaccuracies in classification and diagnosis of medical images may result in treatment delays, potentially impacting the patient's condition if not addressed promptly.

Based on the preceding analysis, we employ the scale distribution as a criterion for evaluating sample complexity and introduce a computational approach, demonstrated in Eq. (1). The variable t in this equation signifies the ratio of the prediction frame to the image area, where a lower ratio indicates higher sample complexity. To address the imbalance issue, each object is assigned a weight based on its proportion, enhancing the influence of smaller samples. Furthermore, a fixed parameter β is introduced to ensure equilibrium among diverse objects. The weights, denoted as ω_i and determined from Eq. (1), are integrated into the loss computation process, effectively alleviating the imbalance in positive samples during detection by assigning appropriate weights to the losses of distinct targets.

$$\omega_i = 1 + \frac{t + \beta}{\log(t)} \qquad (1)$$

By weighting the original loss function, the classified small-scale targets are considered as difficult samples, increasing the contribution of the small percentage of samples to the total loss. In this case, the network can extract features that are more meaningful for small target detection. Simultaneously, as extracting features for large targets poses no significant challenge, their detection accuracy remains unaffected. Hence, this approach significantly enhances the overall detection accuracy.

C. DATA IMBALANCE IN THE POSITIVE EXAMPLE

Commonly used public datasets like MS COCO [31] and ImageNet [32] contain vast amounts of data in each category, leading to a scarcity of consideration for category imbalance in traditional loss functions. This scenario contrasts with medical image datasets, where significant variations exist in sample quantities across categories. For instance, within white-light endoscopic images of colorectal lesions, the number of cancerous structures constitutes only 30% of the adenoma samples. However, in actual clinical diagnosis, low detection of any type of polyp hinders accurate diagnosis of colorectal cancer. In order to address the issue of imbalanced positive samples, this study utilized the Bayesian approach. Initially, the frequency of occurrence for each sample category was calculated to establish the prior probabilities. The model is guided to prioritize certain categories in the initial phase by the value of the prior probability. Subsequently, introducing weighting factors, optimal parameter values were determined through numerous experiments to establish classification difficulty weights. These weights, as depicted in Eq. (2), were customized for each sample category to enhance the model's focus on specific categories during prediction, thereby improving classification accuracy and effectively tackling the challenge of sample imbalance.

$$p_p(c_i) = \frac{\left(1/n_i\right)^{\rho}}{\sum_{j=1}^{k} \left(1/n_j\right)^{\rho}} \qquad (2)$$

where n_i denotes the number of categories c_i, $p_p(c_i)$ denotes the corresponding prior probability of the category, and ρ is used as a parameter. the smaller the number of n_i, the more difficult it is to categorize, and the larger the value of $p_p(c_i)$. By calculating the prior probability of each category, we can obtain the prior distribution of the training samples as shown in Eq. (3).

$$p_p = [p_p(c_1), p_p(c_2), \ldots, p_p(c_k)] \qquad (3)$$

Since categories in medical images are mutually exclusive, we solely focus on computing the prior probabilities and predicted probabilities associated with the correct categories. The actual vector pertaining to category c_i is represented by q. (4).

$$true_{c_i} = [0, 0, \ldots, 1, \ldots, 0]^T \qquad (4)$$

In this vector, the positions corresponding to the real categories are labeled with a 1, and all the rest are 0.

$$p_p * true_{c_i} = \left[p_p\left(c_1\right), p_p\left(c_2\right), \ldots, p_p\left(c_k\right) \right]$$
$$* [0, 0, \ldots, 1, \ldots, 0]^T = p_p\left(c_i\right) \quad (5)$$

$$p_t * true_{c_i} = \left[p_t\left(c_1\right), p_t\left(c_2\right), \ldots, p_t\left(c_k\right) \right]$$
$$* [0, 0, \ldots, 1, \ldots, 0]^T = p_t\left(c_i\right) \quad (6)$$

$$Loss\left(p_t\right) = -p_p\left(c_i\right) * \log\left(p_t\left(c_i\right)\right) \quad (7)$$

In Eq. (5) and (6), $p_p\left(c_i\right)$ represents the a priori probability of object categorization into the correct category, and $p_t\left(c_i\right)$ signifies the predicted probability of object categorization into the accurate category. To mitigate the impact of misclassification on outcomes, solely the difficulty weights of the correct categories are computed in Eq. (7).By calculating the prior probability $p_p\left(c_i\right)$ of each category, the weight of a small number of categories to the overall loss can be increased, allowing the model to focus more on a small number of categories of samples when it is being trained, and the small number of categories can be prioritized over the large number of categories.

D. LSDA-LOSS

In this study, we propose a solution to improve the detection performance of unbalanced datasets by introducing a weighting term based on Focal Loss while considering the imbalance of samples and the classification difficulty of each sample. First, a concise expression is obtained by Focalloss:

$$FL\left(p_t\right) = \alpha_t \left(1 - p_t\right)^\gamma CE\left(p, y\right) \quad (8)$$

The weight coefficient, denoted as α_t in Eq. (8), varies between 0 and 1 to fine-tune the weights of positive and negative samples. Furthermore, the weight coefficient $(1 - p_t)^\gamma$ assesses the difficult classification based on the confidence level of the samples and makes corresponding weight adjustments. In Eq. (8), CE(p,y) represents the cross-entropy loss function. After augmenting the positive samples using the focal loss function to mitigate the foreground-background imbalance issue, we substitute Eq. (8) into Eq. (7). Eq. (9) is obtained by elevating the weight of a small number categories in the overall loss via a Bayesian approach. This adjustment enables the model to prioritize specific categories of samples during training, thereby rectifying the categorization imbalance within the positive samples.

$$Loss(p_t) = -\alpha_t(1 - p_t(c_i))^\gamma p_p(c_i)log(p_t(c_i))) \quad (9)$$

In our methodology, the weight ω_i plays a critical role in assessing the difficulty level, obtained from Eq. (1), and integrates into the loss computation in Eq. (11) to ascertain the ultimate loss value. The parameters β and ρ are utilized for this purpose. Moreover, k represents the quantity of categories, and n_i denotes the sample count in the ith category. Through the proportional distribution of the sample, thus enhancing the impact of smaller samples. Consequently, our proposed LSDA-Loss is formally articulated in Eq. (12).

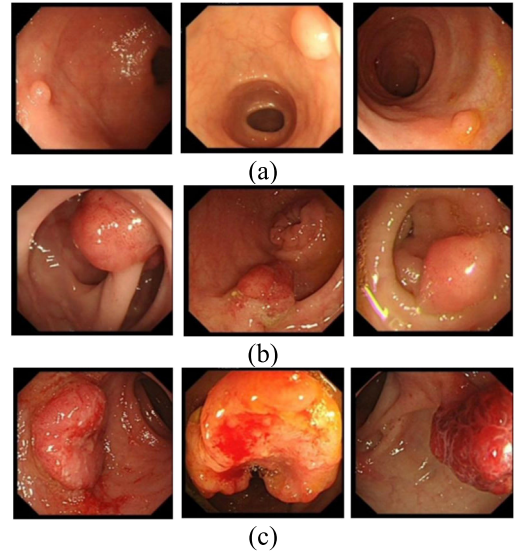$$Loss' = (1 + \omega_i) Loss \quad (10)$$

(a)

(b)

(c)

FIGURE 4. White light endoscopic image (a) polyp, (b) adenoma and (c) cancer.

$$LSDA-Loss = -\alpha_t \left(2 + \frac{t + \beta}{\log(t)} \right) (1 - p_t\left(c_i\right))^\gamma$$
$$\frac{(1/n_i)^\rho}{\sum_{j=1}^t \left(1/n_j\right)^\rho} \log\left(p_t\left(c_i\right)\right) \quad (11)$$

IV. EXPERIMENTAL AND RESULT ANALYSIS
A. DATASET

This research utilized the SSPH_WL dataset, comprising white light endoscopic images captured during patient colonoscopies at the Gastrointestinal Endoscopy Center of the Eastern Hospital within the Shanghai Sixth People's Hospital. The dataset collection process received ethical approval and involved seasoned physicians. Gastroenterologists with over 5 years of clinical diagnostic expertise conducted the dataset labeling. Colorectal lesions were classified into three categories—polyps, adenomas, and cancers—in alignment with clinical diagnoses. The Shanghai Sixth People's Hospital (East) supplied the dataset with strict adherence to privacy protocols, ensuring the removal of personal sensitive information and restricting its usage solely for academic research purposes. Figure 4 showcases sample images depicting each category.

The SSPH_WL dataset comprises 1,709 white light endoscopic lesion images gathered from June 2015 to September 2019, encompassing 1048 adenoma cases, 381 polyp cases, and 280 cancer cases. In this research, the SSPH_WL dataset was split into the training set (SSPH_WL-I) and the test set (SSPH_WL-II). During dataset partitioning, stratification based on lesion type was conducted, followed by independent sampling within each stratum to maintain an 8:2 ratio for the training and test sets. Additionally, publicly accessible colorectal white light endoscopic images from CVC_ClinicDB [33], CVC_ColonDB [34], and Kvasir [35] were utilized as supplemental test sets to assess the

TABLE 2. Training dataset and testing dataset used in this study.

Datasets	Used for	Description	Resolution(w*h)	Public
SSPH_WL_I	Train	1367	375*347	No
SSPH_WL_II	Test	342	375*347	No
Kvasir	Test	428	388*284	Yes
CVC	Test	95	574*500	Yes

classification models' generalization capabilities. A total of 428 colonoscopy images were selected from the Kvasir dataset, featuring 180 adenomas, 73 cancers, and 175 polyps. Furthermore, a subset of 95 images from CVC_ClinicDB, CVC_ColonDB, and an in-house video collection, comprising 36 polyp images, 40 adenoma images, and 19 cancer images, was assembled into a new dataset termed CVC. For further details on these datasets, refer to Table 2.

B. EXPERIMENTAL SETTINGS

The model was trained with a momentum parameter of 0.9 and a learning rate of 0.0001, while the other parameters were kept at their default values. The FPN feature pyramid was chosen as the backbone network framework. Three different FPN models were selected for experimentation to enhance the performance of the loss function across various models. Due to the presence of unbalanced data in the multi-label categorization task, accuracy is deemed inadequate as an evaluation metric for individual category models. Instead, Average Precision (AP) and Average Recall (AR) are utilized. AP evaluates the area under the precision-recall curve, providing a category-specific performance assessment, while AR denotes the average recall across categories. Extensions of these metrics include AP at IoU = 0.50, AP at IoU = 0.75, AR given 1 detection per image, and AR given 10 detections per image. Moreover, and are employed to gauge the accuracy of positive predictions, and F1 is calculated as the harmonic mean of precision and recall.

$$Presion = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$F1 = \frac{Presion \times Recall}{Presion + Recall} \times 2 \quad (14)$$

where TP (True Positive) means that the model correctly predicted a positive category. TN (True Negative) means that the model accurately predicted a negative category. FP (False Positive) means that the model incorrectly predicted a positive category. Finally, FN (false negative) is when the model incorrectly predicts a negative category.

C. HYPER-PARAMETER EXPERIMENTS

The LSDA-Loss function comprises four parameters: ρ controls the prior probability of the category, β regulates the sample size difficulty parameter, while γ and α are fixed at 2 and 0.25, respectively. To optimize the hyperparame-

TABLE 3. AP and AR of RetinaNet with the hyper-parameters.

ρ	β	AP_{50}	AP_{75}	AR	AR_{10}
0.25	0.25	94.27%	87.05%	92.57%	92.56%
0.25	**0.5**	94.56%	**88.85%**	**93.18%**	**93.19%**
0.25	0.75	94.17%	87.94%	91.58%	91.58%
0.5	0.25	93.69%	87.18%	92.60%	92.61%
0.5	0.5	94.31%	86.93%	91.95%	91.95%
0.5	0.75	94.20%	88.47%	92.92%	92.90%
0.75	0.25	94.50%	86.96%	92.58%	92.57%
0.75	0.5	**95.00%**	87.21%	91.83%	91.83%
0.75	0.75	94.62%	86.71%	92.77%	92.78%

ters ρ and β, a series of experiments were conducted to refine their values throughout the training process, ensuring a more effective update of the model parameters. The performance evaluation was carried out on the test set SSPH_WL-II. The tuning strategy adopted here involved maintaining and exploring three key values: 0.25, 0.5, and 0.75, with subsequent adjustments made in increments of 0.25. The outcomes from various parameter combinations are presented in Table 3.

In terms of overall performance, the test set SSPH_WL-II showed that the best APs and ARs were detected when $\beta = 0.5$ and $\rho = 0.25$. Therefore, subsequent experiments were used at this parameter combination.

D. RESULT AND ANALYSIS

To assess the suitability of LSDA-Loss, we conducted comparative experiments using three typical FPN single-stage models: Retinanet, YOLOV7, and SSD. Additionally, to further validate the efficacy of the proposed LSDA-Loss function in classifying white-light endoscopic images of colorectal lesions, we compared it with four common loss functions used to address similar imbalance problems in the same model. These loss functions include cross-entropy loss (CE), focal length loss (FL), varifocal loss (VFL), double balance loss (DB), and our LSDA loss function. The detection results obtained by integrating these five loss functions into the model and the evaluation of Average Precision (AP) are presented in Table 4.

Comparing the results of the above data, for colorectal lesion detection under white light endoscopy, Retinanet outperforms Yolov7 and SSD with a detection accuracy of 94.56% and a 12.27% improvement over the original Focalloss using LSDA-Loss. In YOLOv7, the detection accuracy of LSDA-Loss is 7% better than CE and 7.4% better than FL. In SSD, the loss rate of LSDA-Loss reaches 87.09%, which is 5.46%, 2.13% and 1.29% higher than the other three loss functions, respectively. Moreover, LSDA-Loss effectively balances the APs of the three categories in all three models. In all three models, LSDA-Loss's can significantly

TABLE 4. Comparison of CE, FL, VFL, DB, and LSDA-Loss loss functions under different models.

Models		polyp	adenoma	cancer	AP_{50}	AP_{75}
RetinaNet	CE	89.60%	85.68%	88.74%	88.01%	82.64%
	FL	77.11%	87.49%	82.25%	82.29%	75.95%
	VFL	95.08%	80.42%	69.35%	80.34%	77.86%
	DB	94.26%	91.92%	**95.61%**	93.93%	83.94%
	LSDA-Loss	**95.82%**	**94.36%**	92.48%	**94.56%**	**88.85%**
Yolov7	CE	71.5%	79.3%	82.6%	77.8%	69.4%
	FL	67.2%	82.6%	82.4%	77.4%	62.7%
	VFL	82.3%	86.7%	**88.4%**	**85.8%**	79.8%
	DB	80.6%	**90.1%**	86.5%	85.7%	**80.9%**
	LSDA-Loss	**86.8%**	85%	82.6%	84.8%	78.5%
SSD	CE	79.16%	83.81%	81.94%	81.63%	76.48%
	FL	80.69%	**90.15%**	84.03%	84.96%	76.95%
	DB	82.58%	82.51%	**86.22%**	85.8%	77.85%
	LSDA-Loss	**87.89%**	87.20%	86.19%	**87.09%**	**80.64%**

FIGURE 5. Results on the loss function is LSDA-Loss, CE, FL, DB.
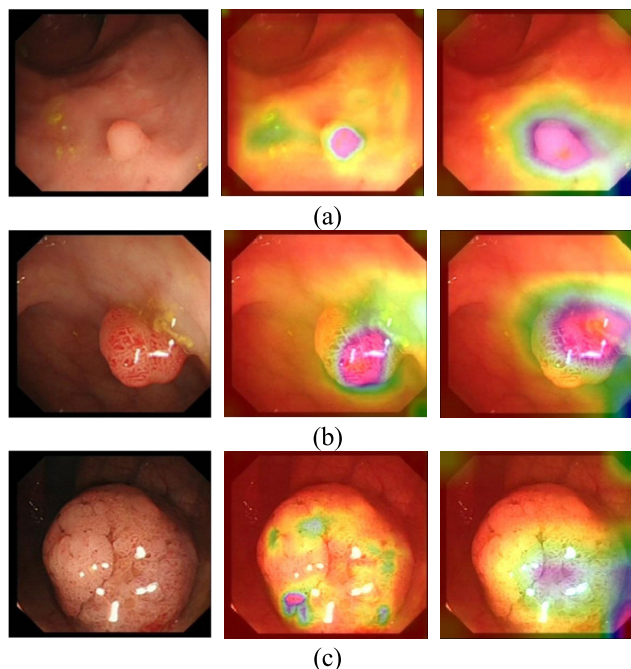
FIGURE 6. CAM images of polyps, adenomas, and carcinomas by RetinaNet with FL and LSDA-Loss.

improve the detection accuracy of the models and balance the classification accuracy of the three categories.

The results depicted in Figure 5 illustrate the enhanced detection performance achieved with LSDA-Loss. The results, shown from left to right, compare the effectiveness of four different loss functions LSDA-Loss, CE, FL, and DB in detecting similar lesion images. Picture (a) displays an adenoma image, demonstrating the improved classification confidence provided by LSDA-Loss. Meanwhile, picture (b) exhibits a polyp image, showcasing the enhanced classification accuracy. This advancement enables the correct classification of previously misclassified or unclassifiable diseases.This improvement allows previously misclassified or unclassifiable diseases to be correctly classified. Moreover, LSDA-Loss has better classification performance for small targets.As shown in (c), which is a cancer image, it can be seen that the classification accuracy is significantly improved.Hence, the comparison unequivocally demonstrates the significant enhancement in detection performance facilitated by LSDA-Loss.

In Figure. 6, the feature maps extracted from the backbone network are visualized using CAM. In Figure 6, the lesion images of polypoid adenoma and cancer are displa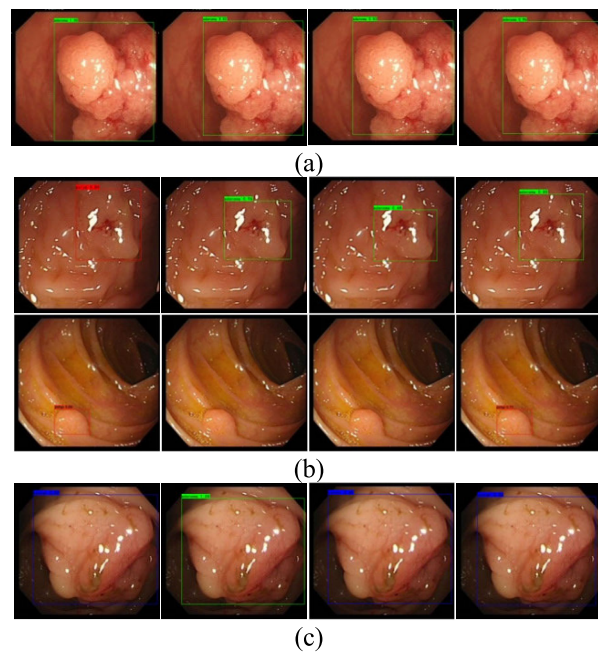yed from left to right. Panel (a) presents the input white light endoscopic image, followed by panel (b) showing the visualization image generated post-detection using FL, and panel (c) displaying the visualization image produced after detection using LSDA-Loss. A comparison between the feature maps obtained with LSDA-Loss and FL reveals that LSDA-Loss demonstrates superior classification capabilities for white-light endoscopic images of colorectal lesions.

TABLE 5. Comparison of the two parts of LSDA-Loss on the three test sets.

| Datasets	omega	prior	Precision	AP_{50}	AR	F1
SSPH_WL_II	×	×	76.82%	82.29%	82.03%	0.77
	√	×	89.84%	94.29%	**93.88%**	**0.91**
	×	√	**90.03%**	93.88%	91.64%	0.90
	√	√	88.98%	**94.56%**	93.18%	**0.91**
Kvasir	×	×	86.26%	92.88%	86.38%	0.86
	√	×	90.11%	94.04%	91.95%	**0.91**
	×	√	88.53%	93.67%	91.40%	**0.91**
	√	√	**90.23%**	**94.14%**	**92.31%**	0.9
CVC	×	×	89.89%	93.91%	91.31%	**0.91**
	√	×	89.99%	93.31%	91.79%	**0.91**
	×	√	90.05%	93.69%	91%	0.90
	√	√	**90.30%**	**94.03%**	**92.31%**	0.90
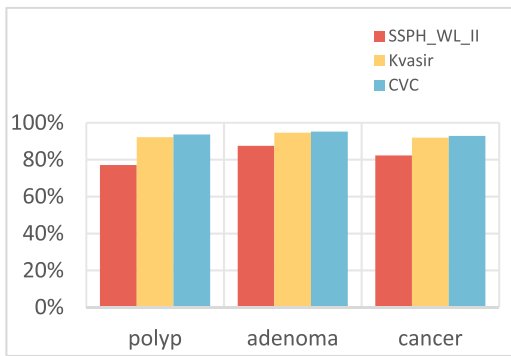
(a)Focal Loss
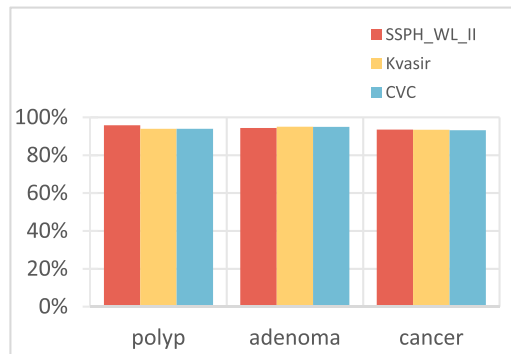
(b)LSDA-Loss

FIGURE 7. Comparison of two loss functions in three test sets.

E. ABLATION EXPERIMENTS

This study conducted experiments on the RetinaNet model using three separate test sets to validate the effectiveness and necessity of each component of LSDA-Loss. Table 5 presents the performance metrics of LSDA-Loss on the SSPH_WL_II, Kvasir, and CVC test sets. The "omega" labels signify classification difficulty, while the "priority" labels indicate data imbalance. Introducing only the "omega" component led to an increase in recall to approximately 90% across all three datasets by 13.02%, 3.85%, and 0.1% respectively, demonstrating excellent performance. This highlights the efficacy of utilizing "omega" for scaling the distribu-
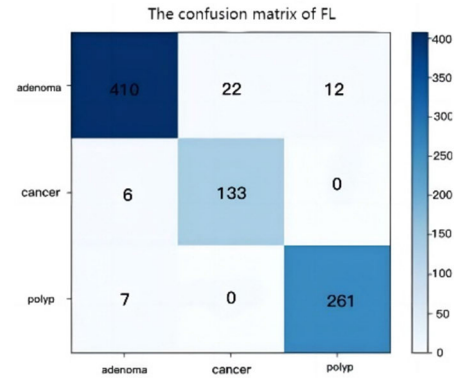
(a) Focal Loss
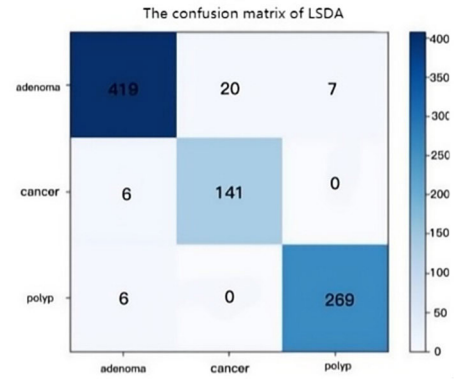
(b) LSDA-Loss

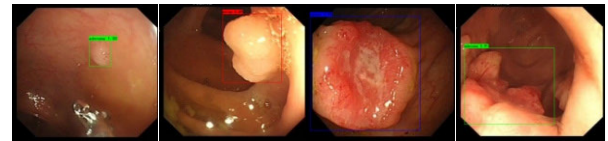**FIGURE 8.** Comparison of confusion matrix results for the two loss functions.

FIGURE 9. Misclassified samples.

tion weights. Similarly, adding only the "priority" element resulted in stable evaluation metrics of around 90% across the datasets, confirming the effectiveness of the "priority" component. When both weightings were incorporated in the experiment, the combined evaluation across the four metrics achieved the highest values, with AP50 accuracy improving by 12.27%, 1.26%, and 0.12% respectively. These experiments underscore that LSDA-Loss exhibits robust generalization capabilities on the test set, enhancing model detection performance and overall classification accuracy.

F. EFFECTIVENESS OF CATEGORY CLASSIFICATION

This study not only assesses the overall classification performance of the model but also evaluates its ability to differentiate between the three categories. Figure 7 illustrates the average accuracy results of the Focal Loss and LSDA-Loss functions on the three test sets. Notably, the accuracy in classifying polyps using LSDA-Loss notably improved in the SSPH_WL_II dataset, with a significant increase of 18.71%. Both the Kvasir dataset and the CVC dataset showed lower accuracy in classifying polyps and

cancers; however, after the enhancement of the loss function, their accuracies were improved, leading to more balanced classification accuracies across all three categories. Analysis of the histograms indicates that the use of LSDA-Loss enables the model to prioritize enhancing the performance of the initially underperforming categories, consequently achieving more balanced classification accuracies for all three categories.

False detection occurs when the model successfully localizes the lesion but fails to classify it accurately due to a lack of classification ability. The confusion matrix in Figure 8 presents a detailed analysis of the categories of false detections. When utilizing the Focal Loss function (a), 22 adenomas were erroneously classified as cancers, 6 cancers as adenomas, 12 adenomas as polyps, and 7 polyps as glands. In contrast, with the LSDA-Loss (b), 20 adenomas were misclassified as cancers, 6 cancers as adenomas, 7 adenomas as polyps, and 6 polyps as adenomas. Polyps, characterized by their small size and similar color and lining, exhibit distinctions from cancers, which are typically large, darker in color, and may appear blood-tinged. These marked differences in lesion characteristics account for all detection errors. The variable shapes of adenomas contribute to misclassification; some small adenomas resemble polyps, while larger adenomas share features with cancers, making them prone to misidentification.

Figure 9 lists some examples of images that were incorrectly detected as described above. From left to right, polyps were misclassified as adenomas, adenomas were misclassified as polyps and cancers, and cancers were misclassified as adenomas.

V. CONCLUSION

Aiming to address the problem of class imbalance with lesion samples in medical images, this paper introduces a novel loss function known as LSDA-Loss. The proposed loss function utilizes Bayesian principles to prioritize specific sample categories, thereby achieving balanced classification accuracy across all categories. Furthermore, the target scale is utilized to distinguish between difficult and easy samples, leading to improved classification accuracy. Through experiments in three typical FPN networks, the evaluation metrics of LSDA-loss outperform the other loss functions, with a detection accuracy of up to 94.56%, and balances the detection accuracy of the three categories.Through parameter tuning on three distinct datasets, the findings demonstrate that the model attains optimal detection performance when $\rho = 0.25$ and $\beta = 0.5$, resulting in significant increases in the AP value of 18.71%, 6.87%, and 10.23% respectively.These findings confirm the effectiveness of the proposed loss function in enhancing accuracy and achieving desired outcomes in the classification of white light endoscopy images.

The loss function is mainly based on the detection of colorectal images under white light endoscopy, and there is a great deal of generalization in other medical images. In future research, it can be further generalized in other medical images, such as tissue slices, to continuously improve the loss function to achieve higher detection accuracy.

REFERENCES

[1] O. Chuquimia, A. Pinna, X. Dray, and B. Granado, "Polyp follow-up in an intelligent wireless capsule endoscopy," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2019, pp. 1–4.

[2] P. Wieszczy, E. Waldmann, M. Løberg, J. Regula, M. Rupinski, M. Bugajski, K. Gray, M. Kalager, M. Ferlitsch, M. F. Kaminski, and M. Bretthauer, "Colonoscopist performance and colorectal cancer risk after adenoma removal to stratify surveillance: Two nationwide observational studies," *Gastroenterology*, vol. 160, no. 4, pp. 1067–1074, Mar. 2021.

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

[4] Y. Yang and D. Yu, "Development of intelligent medicine and construction of intelligent medical engineering specialty," *Int. J. Learn. Teaching*, vol. 9, no. 1, pp. 31–32, 2023.

[5] T. K. Dutta, D. R. Nayak, and Y. D. Zhang, "ARM-Net: Attention-guided residual multiscale CNN for multiclass brain tumor classification using MR images," *Biomed. Signal Process. Control*, vol. 87, Jan. 2024, Art. no. 105421.

[6] K. Pogorelov, M. Riegler, S. L. Eskeland, T. de Lange, D. Johansen, C. Griwodz, P. T. Schmidt, and P. Halvorsen, "Efficient disease detection in gastrointestinal videos—Global features versus neural networks," *Multimedia Tools Appl.*, vol. 76, no. 21, pp. 22493–22525, Nov. 2017.

[7] Y. Komeda, H. Handa, T. Watanabe, T. Nomura, M. Kitahashi, T. Sakurai, A. Okamoto, T. Minami, M. Kono, T. Arizumi, M. Takenaka, S. Hagiwara, S. Matsui, N. Nishida, H. Kashida, and M. Kudo, "Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: Preliminary experience," *Oncology*, vol. 93, pp. 30–34, Dec. 2017.

[8] G. Wimmer, A. Vécsei, and A. Uhl, "CNN transfer learning for the automated diagnosis of celiac disease," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Dec. 2016, pp. 1–6.

[9] M. F. Byrne, N. Chapados, F. Soudan, C. Oertel, M. L. Pérez, R. Kelly, N. Iqbal, F. Chandelier, and D. K. Rex, "Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model," *Gut*, vol. 68, no. 1, pp. 94–100, Jan. 2019.

[10] J. Gao, Y. Guo, Y. Sun, and G. Qu, "Application of deep learning for early screening of colorectal precancerous lesions under white light endoscopy," *Comput. Math. Methods Med.*, vol. 2020, Aug. 2020, Art. no. 8374317.

[11] Y. Yuan, W. Qin, B. Ibragimov, B. Han, and L. Xing, "RIIS-DenseNet: Rotation-invariant and image similarity constrained densely connected convolutional network for polyp detection," in *Proc. Int. Conf. Med. lmage Comput. Comput.-Assist. Intervent.* Leeds, U.K.: Springer, 2018, pp. 620–628.

[12] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning From Imbalanced Data Sets*, vol. 10. Springer, 2018.

[13] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–36, Jul. 2020.

[14] P. Wang, T. M. Berzin, J. R. G. Brown, S. Bharadwaj, A. Becq, X. Xiao, P. Liu, L. Li, Y. Song, D. Zhang, Y. Li, G. Xu, M. Tu, and X. Liu, "Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study," *Gut*, vol. 68, no. 10, pp. 1813–1819, Oct. 2019.

[15] S. Mohaoui, A. Hakim, and S. Raghay, "Bi-dictionary learning model for medical image reconstruction from undersampled data," *IET Image Process.*, vol. 14, no. 10, pp. 2130–2139, Aug. 2020.

[16] A. Viloria, O. B. P. Lezama, and N. Mercado-Caruzo, "Unbalanced data processing using oversampling: Machine learning," *Proc. Comput. Sci.*, vol. 175, pp. 108–113, Jan. 2020.

[17] P. Yang, L. Xu, Y. Wan, J. Yang, Y. Xue, Y. Jiang, C. Luo, J. Wang, and T. Niu, "Deep neural network-based approach to improving radiomics analysis reproducibility in liver cancer: Effect on image resampling," *Phys. Med. Biol.*, vol. 66, no. 16, Aug. 2021, Art. no. 165009.

[18] H. Guan, L. Zhao, X. Dong, and C. Chen, "Extended natural neighborhood for SMOTE and its variants in imbalanced classification," *Eng. Appl. Artif. Intell.*, vol. 124, Sep. 2023, Art. no. 106570.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, p. 99.

[20] B. Li, Y. Liu, and X. Wang, "Gradient harmonized single-stage detector," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 8577–8584.

[21] Y. Cao, K. Chen, C. C. Loy, and D. Lin, "Prime sample attention in object detection," 2019, *arXiv:1904.04821*.

[22] Q. Qian, L. Chen, H. Li, and R. Jin, "DR loss: Improving object detection by distributional ranking," 2019, *arXiv:1907.10156*.

[23] X. Li, W. Wang, L. Wu, S. Chen, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection," in *Proc. Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 33, Dec. 2020, pp. 21002–21012.

[24] K. Chen, J. Li, W. Lin, J. See, J. Wang, L. Duan, Z. Chen, C. He, and J. Zou, "Towards accurate one-stage object detection with AP-loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5114–5122.

[25] H. Zhang, Y. Wang, F. Dayoub, and N. Sünderhauf, "VarifocalNet: An IoU-aware dense object detector," 2020, *arXiv:2008.13367*.

[26] C. Yu, W. Sun, Q. Xiong, J. Gao, and G. Qu, "Double-balanced loss for imbalanced colorectal lesion classification," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–14, Aug. 2022.

[27] A. A. Sadi, L. Chowdhury, N. Jahan, M. N. S. Rafi, R. Chowdhury, F. A. Khan, and N. Mohammed, "LMFLOSS: A hybrid loss for imbalanced medical image classification," 2022, *arXiv:2212.12741*.

[28] O. Gökkan and M. Kuntalp, "A new imbalance-aware loss function to be used in a deep neural network for colorectal polyp segmentation," *Comput. Biol. Med.*, vol. 151, Dec. 2022, Art. no. 106205.

[29] S. Fu, D. Su, S. Li, S. Sun, and Y. Tian, "Linear-exponential loss incorporated deep learning for imbalanced classification," *ISA Trans.*, vol. 140, pp. 279–292, Sep. 2023, doi: 10.1016/j.isatra.2023.06.016.

[30] Q. Yang, C. Zhang, H. Wang, Q. He, and L. Huo, "SV-FPN: Small object feature enhancement and variance-guided RoI fusion for feature pyramid networks," *Electronics*, vol. 11, no. 13, p. 2028, Jun. 2022.

[31] K. Tong and Y. Wu, "Rethinking PASCAL-VOC and MS-COCO dataset for small object detection," *J. Vis. Commun. Image Represent.*, vol. 93, May 2023, Art. no. 103830.

[32] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104115.

[33] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Computerized Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.

[34] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdzal, and A. Courville, "A benchmark for endoluminal scene segmentation of colonoscopy images," *J. Healthcare Eng.*, vol. 2017, Jun. 2017, Art. no. 4037190.

[35] D. Jha, P. H. Smedsrud, M. A. Riegler, P. Halvorsen, T. de Lange, D. Johansen, and H. D. Johansen, "Kvasir-SEG: A segmented polyp dataset," 2019, *arXiv:1911.07069*.

RUIQI ZHAO was born in Bengbu, Anhui, China, in 1999. She received the B.S. degree in network engineering from Tianjin University of Science and Technology, in 2017. Currently pursuing the M.S. degree in computer science and technology with Shanghai Maritime University, China. Her current research interests include computer vision and deep learning.

KUNPENG ZHANG is currently pursuing the master's degree in computer science and technology with Shanghai Maritime University, China. His current research interests include computer vision and deep learning.

JUNBO GAO is currently an Associate Professor and the master's Supervisor of Shanghai Maritime University, focusing on artificial intelligence, emotional disposition analysis, and opportunity discovery. He has hosted and participated in more than 20 projects of the National Natural Science Foundation of China, 863 projects, Shanghai Natural Science Foundation, and Shanghai Municipal Education Commission, and projects entrusted by enterprises and institutions.

WEI SUN received the B.Sc., M.Eng., and Ph.D. degrees from Harbin Engineering University, in 2006.

He became a Lecturer and an Associate Professor with Shanghai Maritime University, in 2007, where he has been an Associate Professor of computer science and technology, since 2011. His academic career also includes experience with Victoria University, Australia, from March 2011 to August 2011. He currently leads a number of research projects funded by industry and the university, in applying artificial intelligence (cognitive modeling, knowledge modeling, reasoning, and learning) in construction of smart software agents, which are applied in health/medical services. He was awarded the Science and Technology Progress Award by China Government, in 2006, 2011, and 2013.

GUOQIANG QU is currently the Associate Chief Physician of the Department of Gastroenterology, Shanghai Sixth People's Hospital, specializing in gastrointestinal, hepatobiliary, and other diseases. He specializes in gastroscopy, enteroscopy, and minimally invasive endoscopic treatment.

• • •