A New Technique for Diagnosis of Dental Caries on the Children's First Permanent Molar

HONGBING YU[1], ZEBIN LIN[2], YISHU LIU[2], JING SU[1], BINGZHI CHEN[2],
AND GUANGMING LU[2], (Member, IEEE)
[1]Shenzhen Nanshan Center for Chronic Disease Control, Shenzhen 518054, China
[2]Shenzhen Medical Biometrics Perception and Analysis Engineering Laboratory, Harbin Institute of Technology, Shenzhen 518054, China

Corresponding authors: Yishu Liu (19S051057@stu.hit.edu.cn) and Bingzhi Chen (chenbingzhi@stu.hit.edu.cn)

ABSTRACT The current researches have been shown high prevalence and incidence of children's teeth caries, especially for the first permanent molar, which might do a lot of harm to their general health. Fortunately, early detection and protection can reduce the difficulty of treatment and protect children's oral health. However, traditional diagnostic methods such as dentist's visual inspection and radiographic imaging diagnosis are non-automatic and time-consuming. Given the COVID-19 epidemic, these methods should not be taken into consideration, since they fail to practice social distancing and further increase the risk of infection. To address these issues, in this paper we propose a novel caries detection and assessment (UCDA) framework to achieve a new technique for fully-automated diagnosis of dental caries on the children's first permanent molar. Inspired by an efficient in-network feature pyramid and anchor boxes, the proposed UCDA framework mainly contains a backbone network that is initialized with ResNet-FPN, and two parallel task-specific subnetworks for region regression and region classification. Due to the lack of the image database, we also present a novel children's oral image database, namely "Child-OID", which comprises 1, 368 primary school children's oral images with standard diagnostic annotations and labels, to evaluate the effectiveness of our UCDA method. Experiments on the Child-OID database demonstrate that commonly occurring caries on the first permanent molar can be more accurately detected via the proposed UCDA framework. Database and code are available at `https://github.com/GipinLinn/UCDA-and-Child-OID.git`.

INDEX TERMS Caries detection and assessment, first permanent molar, child-OID, oral images.

I. INTRODUCTION

As one of the most highly prevalent chronic diseases in children, dental caries can seriously affect the physical and mental health of children. In recent years, research shows that dental caries are generated by the long-term sophisticated interaction between many host factors and acid-producing bacteria that live in the plaque. Especially for the children's first permanent molar, it suffers from a high morbidity rate of dental caries. For the children, the first permanent molar is generally rooted in the deep part of the oral. Due to poor dental hygiene, food debris, saliva, and minerals will build

upon the ravines of the first permanent molar, leading to the development of plaque. Typically, the bacteria mix with carbohydrates from foods and create acids, which would break down areas of enamel cause dental caries, as shown in Fig. 1. To provide timely treatment and prevent further deterioration, early detection for dental caries on the children's first permanent molar are necessary. However, identifying and distinguishing dental caries on the first permanent molar still relies heavily on the availability of an expert dentist's visual inspection and radiographic imaging diagnosis, which are very labor-intensive and time-consuming. Moreover, in the current situation of the COVID-19 [1] epidemic, some child patients with dental caries need to practice social distancing to avoid escalation of infection, resulting in missing the

The associate editor coordinating the review of this manuscript and approving it for publication was Yongjie Li.
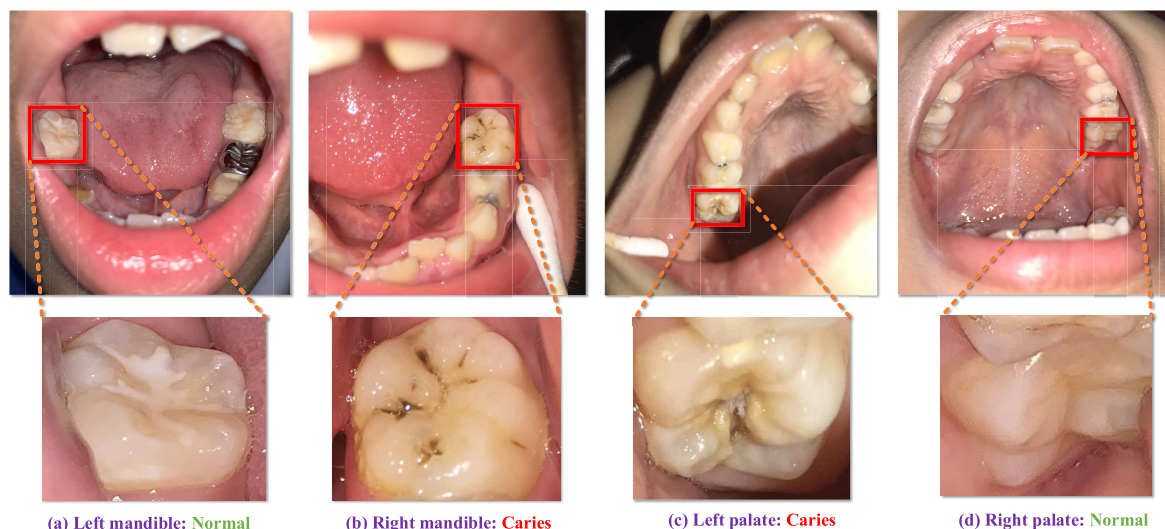
| (a) Left mandible: Normal | (b) Right mandible: Caries | (c) Left palate: Caries | (d) Right palate: Normal |

FIGURE 1. Different types of the first permanent molars, including left mandible (a), right mandible (b), left palate (c), and right palate (d). Compared with normal teeth, the cavities in the tooth may be various colors from yellow to black.

optimal treatment opportunities. Therefore, there are huge demands on developing fully-automated and high-precision caries detection and assessment system to assist the diagnosis of dental caries on children's first permanent molar.

Unlike general computer vision tasks, medical image annotation requires extensive clinical expertise. With the development of deep learning technologies and large-scale annotated image datasets, rapid and tremendous progress has been evidenced in a range of computer-aided detection/diagnosis (CADe/CADx) systems [2]–[4]. For example, Wang *et al.* [2] proposed a hospital–scale chest X-ray dataset, i.e., ChestX-ray14, and utilized Convolutional Neural Networks (CNN) [5]–[7] to boost the performance of multi-label chest X-ray image classification. Rajpurkar *et al.* [4] trained a 169-layer DenseNet [8] baseline model to detect and localize abnormalities based on their proposed large-scale dataset of musculoskeletal radiographs. Yan *et al.* [3] introduced a CT images dataset named DeepLesion, and further develop an automatic lesion detection algorithm to find all types of lesions with a unified CNN-based framework. Unlike the aforementioned CADe/CADx tasks in the medical image domain, to the best of our knowledge, automated caries diagnosis currently lacks an available annotated image dataset, which leads towards the failure of establishing a universal caries detection and assessment system.

As the fundamental topics in CADe/CADx, lesions detection and characterization such as skin lesions [9], lung nodules [10], and liver lesions [11] have attracted much interest to the application of deep learning approaches. However, vast infrequent types such as dental caries detection are ignored by most CADe programs. Moreover, dental caries on the first permanent molar is typically small and its characterization is easily affected by the light, background, and other conditions, which further increases the difficulty of detection and assessment. Therefore, it remains challenging to develop a universal

caries detection and assessment system, capable of detecting dental caries on the first permanent molars.

To address these challenges, in this paper we only focus on the first permanent molar and introduce a unified caries detection and assessment (UCDA) framework to perform the automated diagnosis of dental caries on the children's first permanent molar. Specifically, the proposed UCDA framework benefits from an efficient in-network feature pyramid and use of anchor boxes, and it is composed of a ResNet-FPN [12] based backbone network and two parallel task-specific subnetworks, i.e., region regression and region classification subnetworks. The region regression subnetwork is designed to perform convolutional bounding box regression, while the region classification subnetwork aims to classify the object from the output of the backbone. Importantly, a novel children's oral image database, namely ''Child-OID'', is proposed to evaluate the effectiveness of the proposed UCDA method. Our Child-OID, supported by Shenzhen Chronic Disease Prevention and Treatment Center, is composed of 1, 368 primary school children's oral images with standard diagnostic annotations and labels. Note that, all the annotations and labels are marked by two expert dentists according to the WHO's basic methods and the diagnostic standards. To mitigate the class imbalance problem, we introduce a novel weighted Cross-Entropy (W-CE) loss to further optimize our UCDA framework in the training phase. The main contributions of this paper are summarized as follows:

- We propose a unified caries detection and assessment (UCDA) framework for dental caries diagnosis on the children's first permanent molar, which is crucial for building an automatic dental caries diagnosis system.
- We also propose a novel children's oral image database (i.e., Child-OID) with standard diagnostic annotations and labels. To the best of our knowledge, this is the

first benchmark image dataset that is appropriate for children's dental caries diagnosis.

- Experimental results on the proposed benchmark image datasets demonstrate that our UCDA framework can achieve a diagnostic accuracy of over 95%, which fully meets the clinical demand for dental caries detection and assessment.

The rest of the article is arranged as follows. Section II mainly discusses the related work including existing caries detection methods and common object detection algorithms. Section III explores the proposed UCDA framework, while Section IV covers the proposed Child-OID dataset. Next, the comprehensive experiments are conducted in Section V. Finally, Section VI briefly concludes the whole work.

II. RELATED WORK

In this part, we first make a brief review of existing diagnostic methods for dental caries diagnosis. In addition, we also discuss the topic of existing object detection algorithms in detail.

A. CARIES DIAGNOSIS METHODS

1) DENTIST'S VISUAL INSPECTION APPROACHES

With the aid of the artificial light, air drying, mouth mirror, and probe, dental caries diagnosis in daily dental practice is prevailingly performed by the expert dentist's visual inspection. For example, Lino *et al.* [13] utilized visual examination to investigate the characteristics of the margins of restoration in 88 permanent molars and premolars from 18 patients. To provide flexibility for clinicians and researchers to choose the stage of caries process and other features that fit the needs of their research or practice, Ismail *et al.* [14] proposed a new system which was named the International Caries Detection and assessment system (ICDAS), which makes up of the lack of consistency among the contemporary criteria systems. Moreover, Jan *et al.* [15] leveraged a Universal Visual Scoring System (UniViSS) to address the problem of occlusal and smooth surface lesions. Besides, some population-based oral health researchers utilize the DMFT method [16], [17] such as Significant Caries Index (SiC Index) [18], [19] to calculate and present the caries prevalence for a specific area. However, these operations rely heavily on dentists' subjective perceptions and clinical experiences. Especially in the current situation of the COVID-19 epidemic, visual inspection approaches may cause cross-infection and would be prohibited. Therefore, it is urgent to develop an automated caries detection and assessment system, avoiding delay children's caries diagnosis opportunities.

2) RADIOGRAPHIC IMAGING AND OPTICAL-BASED APPROACHES

Radiographic imaging is one of the common means used in diagnosis of occlusion, periapical, and panoramic [20]. Some researches [20]–[22] utilized the bitewing radiographic images to give dentists precise areas affected by dental caries.

Bhan *et al.* [23] have proposed a method including the preprocessing of bitewing radiographic images, edge recognition, thresholding, and connected component labeling. Lee *et al.* [24] trained a periapical radiographic diagnosis model with 3, 000 periapical radiographic images based on the pre-trained Inception v3 [25] network. However, radiography imaging approaches might suffer from the overlapping occlusion of teeth, bones, and surrounding soft tissues. Moreover, such operations might cause harmful radiation in the human body. To avoid the disadvantages of radiography, other optical-based approaches have received much attention for dental caries diagnosis. For instance, near-infrared light transillumination [26]–[28]was often used to assist dentist diagnosis.

B. COMMON OBJECT DETECTION ALGORITHMS

1) TWO-STAGE APPROACHES

Recently, the dominant paradigm in modern object detection is based on a two-stage approach. For example, Girshick *et al.* [29] introduced the Selective Search algorithm to propose a new region-based framework named R-CNN, which has achieved breakthrough performance in the field of object detection. He *et al.* [30] presented SPP-Net to solve the disadvantages of R-CNN, which allowed input images of any size to feed into feature extractor by adding a spatial pyramid pooling layer before the fully-connected layer. Inspired by the strength of SPP-Net, Girshick *et al.* [31] proposed the Fast R-CNN that replaced the initial spatial pyramid pooling layer with the RoI Pooling Layer and utilized multi-task loss for the end-to-end training, which has further enriched the detection information and improved the detection speed. Then Ren *et al.* [32] presented the classical Faster R-CNN, which was extended from Fast R-CNN and has achieved excellent progress in many tasks of object detection challenges. Although the existing highest accuracy object detectors are based on the two-stage approaches, they only focus on accuracy while ignoring the detection efficiency.

2) ONE-STAGE APPROACHES

By contrast, the one-stage algorithms can reduce the time lost due to global regression instead of two-stages partial regression, which generates the bounding box coordinates and class probabilities directly from input image pixels. Recently, YOLO [33] and SSD [34] have renewed interest in one-stage methods. For instance, Redmon *et al.* [33], [35] presented a series of one-stage detectors, i.e., YOLO and YOLOv2, to focus on an even more extreme speed/accuracy trade-off. Though YOLO and YOLOv2 have gained impressive performance on efficiency and detection accuracy, it still has some limitations. For example, YOLO is hard to detect small objects and generates relatively coarse features. Liu *et al.* [34] presented the SSD that efficiently integrates the advantages of anchors and multi-scale representation.

Unlike the above works, the proposed UCDA framework is initialized with a ResNet-FPN based backbone and two
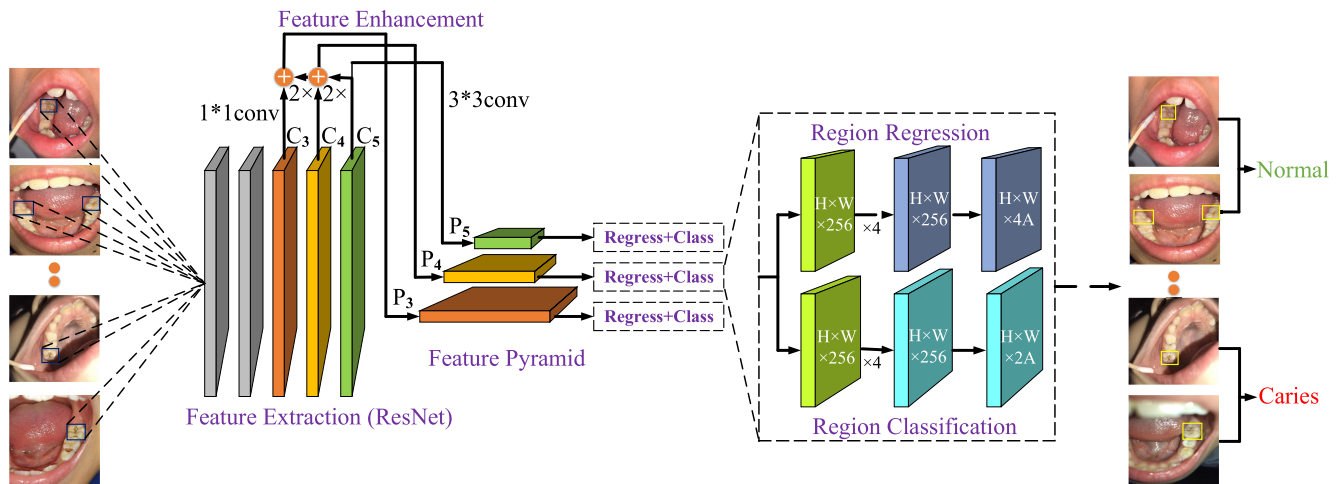
FIGURE 2. Illustration of the proposed caries detection and assessment (UCDA) framework. The proposed UCDA framework is composed of a ResNet-FPN based backbone network and two task-specific subnetworks, i.e., region regression and region classification subnetworks.

task-specific subnetworks for region regression and region classification, in order to make full use of the advantages of the efficient in-network feature pyramid and an-chor boxes.

III. THE PROPOSED UCDA FRAMEWORK

Firstly, we introduce an overview of the proposed UCDA framework. Then, a detailed description of each component in the proposed UCDA framework is given. Finally, we make a definition of W-CE loss used in this work.

A. OVERVIEW

Inspired by the idea of the efficient in-network feature pyramid and use of anchor boxes, we construct a novel unified caries detection and assessment (UCDA) framework for automated diagnosis of dental caries on the children's first permanent molar. The architecture of the proposed UCDA framework is illustrated in Fig. 2. Given an oral image, our ResNet-FPN backbone first utilizes a feature extractor to generate coarse feature maps of the first permanent molar. To reduce the noise regions of the initial feature map, it also adopts a feature enhancement module, which will select the fine-grained feature maps from the last few convolutional layers of the feature extractor. Then these fine-grained feature maps will be fed into a region regression and region classification subnetworks, respectively. Finally, the predicted box coordinates of the first permanent molar and its corresponding category can be acquired and merged to output the final results. In addition, for the class-imbalance of our Child-OID dataset, we utilize the W-CE loss to optimize our UCDA framework. We will explain these components in the following subsections.

B. ResNet-FPN BACKBONE

As mentioned above, the task of caries diagnosis on the first permanent can be converted into a small target detection problem, which might suffer from noisy regions. To increase the precision of caries detection, the main core of our ResNet-FPN backbone is designed to generate a multi-scale feature pyramid from a single resolution oral image. Specifically, it is constructed with a feature extractor and a feature enhancement module, which are initialized with the pretrained ResNet module and Feature Pyramid Network (FPN), respectively. That is, we build the FPN on top of the ResNet architecture. In general, low-level feature maps have rich detailed information while high-level feature maps present stronger semantic information. Each level of the pyramid leverages the feature maps from different convolution layers to detect multi-scale objects. In this way, we can aggregate these proposal results from the pyramid and apply them to the region regression and region classification subnetworks.

In our work, the output feature maps of the last three residual blocks of ResNet $\{C_3, C_4, C_5\}$ are consider to generate the proposed feature pyramid. Firstly, we attach a $1 * 1$ convolutional layer on $\{C_3, C_4, C_5\}$ to obtain feature maps with same number of dimensions, which are denoted as $\{D_3, D_4, D_5\}$. Since the low-level feature map (D_3) has higher resolution, we upsample the resolution of the high-level feature maps of $\{D_4, D_5\}$. Then we utilize the nearest neighbor upsampling on $\{D_4, D_5\}$ to obtain upsampled feature maps $\{E_4, E_5\}$. After that, the upsampled feature maps are integrated with the corresponding low-level feature maps by element-wise addition. In this way, we obtain the fused feature maps denoted as $\{F_3, F_4, F_5\}$. Finally, in order to alleviate the overlapping interference of fused feature maps, we apply a $3 * 3$ convolutional layer on each merged map to obtain final feature maps $\{P_3, P_4, P_5\}$.

C. REGION CLASSIFICATION SUBNETWORK

The region classification subnetwork predicts the probability of caries presence at each oral image. By taking an input feature maps from a given pyramid level, our region regression module is built on a fully convolutional network (FCN),

TABLE 1. Data statistics of the proposed Child-OID dataset.

Child-OID	Train		Test		Total Number
	Normal	Abnormal	Normal	Abnormal	
Left mandible	325	44	145	25	539
Right mandible	189	25	52	13	279
Left palate	302	34	90	10	436
Right palate	254	27	94	9	384
Total Number	1,070	130	381	57	1,638

which consists of four $3 * 3$ convolutional layers, each followed by ReLU activation function. Finally, a softmax layer is attached to output the corresponding confidence scores. Noted that, our region classification subnetwork does not share parameters with the region regression subnet.

D. REGION REGRESSION SUBNETWORK
Our region regression subnetwork leverages another parallel FCN to connect each layer of the pyramid for regressing the offset from each anchor box to a nearby ground-truth object. To meet the needs of multi-scale of input image, we assign the training labels to anchors based on their Intersection-over-Union(IoU) ratios with ground-truth, and the areas of anchors are defined as $\{64^2, 128^2, 256^2\}$, respectively. In our work, an anchor is assigned as a positive label if it has an IoU over 0.7 with any ground-truth box, or has the highest IoU for a specific ground-truth box. And a negative label is allocated to the anchor if all IoUs with all ground-truth boxes are lower than 0.3. It is noted that only the positive anchors would be considered to the loss of region regression, since the negative anchors have serious deviation with all ground-truth boxes.

E. WEIGHTED CROSS ENTROPY LOSS
In the training phase, vast easily-classified negatives can comprise the majority of loss and dominate the gradient, both of which result in performance degradation [36]. To reduce the overload of vast easily-classified negatives, our proposed UCDA framework requires to focus on the hard-classified samples. Therefore, we introduce a pair of modulating factors p^γ and $(1-p)^\gamma$ into our W-CE loss function to balance the influence between the hard-classified and easily-classified samples in the training phase. Here, the objective function of the proposed W-CE loss takes the following form:

$$L(p, l) = -\alpha \sum_{y=1} (1-p)^\gamma \log(p)$$
$$-(1-\alpha) \sum_{y=0} p^\gamma \log(1-p) \quad (1)$$

where α and $(1-\alpha)$ are weights of abnormal and normal samples, l represents the ground truth label and p is the corresponding confidence score. When the proposal region is difficult to distinguish, p would close 0 and $(1-p)^\gamma$ is near

1.0, thus its loss contribution would be strengthened, and vice versa.

IV. THE PROPOSED CHILD-OID DATABASE
In this section, the proposed Child-OID database will be introduced in detail from the following aspects: data collection, statistics, and augmentation, as well as the challenges of Child-OID.

A. DATA COLLECTION AND STATISTICS
Shenzhen Chronic Disease Prevention and Treatment Center approved the study collected children's oral images from 4 public primary schools for caries diagnosis on the first permanent molar. We assemble a children's oral images database consisting of a total of 1,368 multi-view oral images from 342 children. Each belongs to one of four standard image types: left mandible, right mandible, left palate, and right palate, as shown in Fig. 1. Note that, some images might contain multiple first permanent molars. To facilitate applications such as caries detection and assessment system, we converted the diameters of the first permanent molars into bounding-boxes. Based on the WHO's basic methods and the diagnostic standards, each first permanent molar was manually labeled as normal or abnormal by board-certified dentists from our institution. To investigate the types of observations present in the dataset, we reviewed the collected image to manually label 1,451 normal findings and 187 abnormal findings. The collected images vary in resolution and aspect ratios. We split the dataset into training set (242 children, 1,052 images) and test set (100 children, 316 images). There is no overlap in children between any of the sets. Table 1 summarizes the distribution of normal and abnormal findings in the proposed Child-OID database.

B. DATA AUGMENTATION
CNN-based models have performed remarkably well on many computer vision tasks. However, they are heavily reliant on large amounts of data to avoid overfitting, especially for the small sample size of data. In our work, we also perform data augmentation, a data-space solution to the problem of limited data. Specifically, we consider some simple strategies such as up-down and left-right flip operations to triple the size of training set such that a better diagnostic model can be built using them.

C. THE CHALLENGES OF CHILD-OID

As shown in Fig. 3, our Child-OID database inevitably suffers from the problem of shadow and occlusion, which further increases the difficulty of detection. In detail, about 8 percent of first permanent molars are covered by shadow since they completely located deep within oral. Besides, about 5 percent of first permanent molars occluded-with the tongue or the saliva. The sample images of shadow and occlusion in our Child-OID database are shown in Fig. 4.
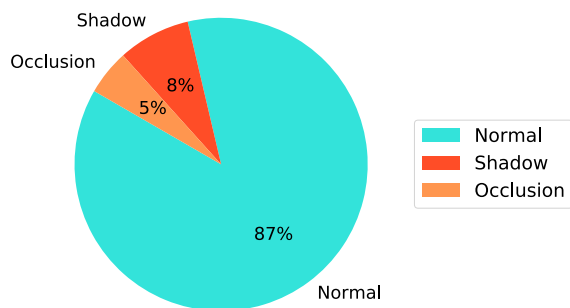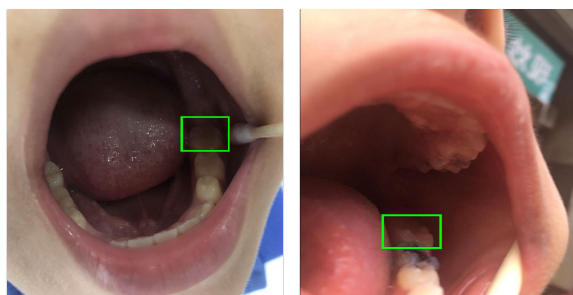

FIGURE 3. The proportion of normal, shadow, and occlusion in the proposed Child-OID database.

(a) Shadow

(b) Occlusion

FIGURE 4. Illustration of examples of shadow (a) and occlusion (b) images the proposed Child-OID database.

V. EXPERIMENTS

In this section, we first describe the implementation details and evaluation metrics used in our experiments. Then we evaluate the performance of the proposed UCDA framework on our Child-OID database compared to current state-of-the-art object detection algorithms. Finally, some qualitative results are presented.

A. IMPLEMENT DETAILS

In our experiments, the feature extractor is initialized as the pretrained ResNet-50 model. The proposed UCDA framework is conducted by using the deep learning toolbox PyTorch and runs on one Nvidia Titan XP GPU with 12 GB memory. The biases and weights of region regression sub-network are initialed with 0. And the biases of region classification sub-network are the same as the region regression module. In the training phase, we set the min-batch size to 1 and the initial learning rate is set to 1e-6, which is decreased 10 times every 10 epochs. And we use Adaptive Moment Estimation (Adam) as our parameter optimizer. In reference, we empirically set the non-maximum suppression threshold to 0.05 and our framework mainly outputs predicted bounding boxes and corresponding scores.

B. EVALUATION METRICS

Follow the evaluation indicators in the medical image domain, in this paper we consider the detection accuracy, sensitivity, and specificity as our evaluation metrics to verify the performance of the proposed UCDA framework. These metrics are associated with four values, i.e., true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN), which are defined as below:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$
$$Sensitivity = \frac{TP}{TP + FN}$$
$$Specificity = \frac{TN}{FP + TN} \qquad (2)$$

C. PARAMETER ANALYSIS

In this section, we mainly analyze the impact of parameters on our framework's performance. We perform evaluations of parameter analysis from three aspects, including the influence of the depth of ResNet, and the sensitivity parameters α and γ in the loss function.

1) EVALUTAION ON THE DEPTH OF ResNet

The depth of ResNet is the key to feature extractor in our proposed UCDA framework and it determines the quality of the feature map. The deeper network will generally obtain higher quality feature map but might suffer from the problem of under-fitting. In our experiments, we conduct comparisons between different ResNet layers for teeth feature extraction, including ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. As shown in Fig. 5, our proposed UCDA framework achieves the best performance on Child-OID when we adopt the ResNet-50 as the feature extractor.

TABLE 2. The comparison results of different caries detection methods in the proposed Child-OID database. All results were analysed in percentage (%) terms.

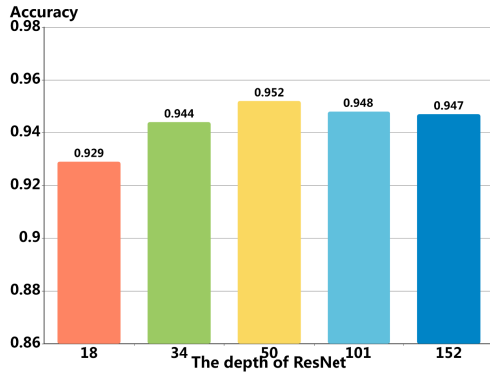|	Accuracy(%)	Sentivity(%)	Specificity(%)
YOLOv1 [33]	89.28	85.14	91.98
YOLOv2 [35]	90.35	85.69	92.28
SSD [34]	91.56	86.78	93.41
DSSD [37]	93.12	87.39	94.34
Our UCDA	**95.25**	**89.83**	**96.10**

FIGURE 5. Comparison of detection accuracy with different depths of feature extractor.
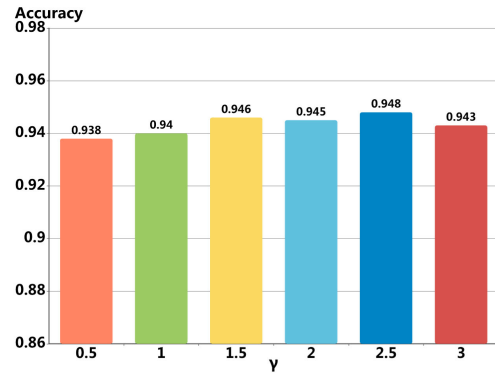
FIGURE 7. Comparison of detection accuracy with different values γ.

2) EVALUATION ON PARAMETER α

The value α in (1) aims to control the weights of abnormal and normal samples. By fixing other parameters, we conduct our loss function with a range of different value α, i.e., $\alpha \in \{0.1, 0.15, 0.2, 0.25, 0.3, 0.35\}$. It is clear that the performance of the proposed UCDA framework is limited by the weights of abnormal samples when the threshold α is too small. As shown in Fig. 6, the experimental results evaluated on Child-OID show the best setting of α for our framework is 0.25.
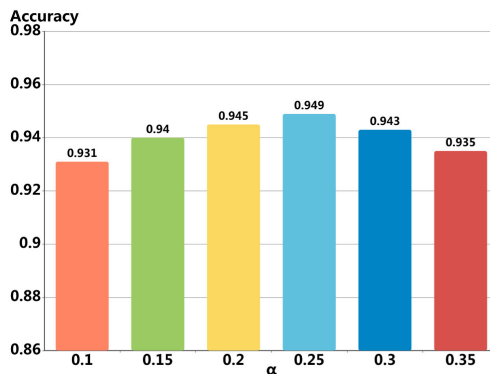

FIGURE 6. Comparison of detection accuracy with different values α.

3) EVALUATION ON PARAMETER γ

The value γ in (1) is designed to balance the influence between the hard-classified and easily-classified samples.

To determine the optimal value of γ, we first fix other parameters and then study how the accuracy score of the proposed UCDA framework changes with a range of different values of γ, i.e., $\gamma \in \{0.5, 1, 1.5, 2, 2.5, 3.0\}$. As shown in Fig. 7, we can see that the proposed UCDA framework achieves the best performance when the value of γ is set to be 2.5.

D. COMPARISON WITH STATE-OF-THE-ART BASELINES

To evaluate the performance of our UCDA method on the diagnosis of dental caries task, we further perform the experiments on our newly introduced Child-OID dataset. In the experiments, we compared the proposed UCDA method with some state-of-the-art baselines, including YOLO [33], [35] and SSD [34], [37]. The comparative results are presented in Table 2.

From Table 2, we have the following observations. 1) The proposed UCDA framework comprehensively improves the caries detection performance on the first permanent molar of children in terms of accuracy, sensitivity, and specificity. In particular, it can achieve the highest accuracy of 95.25% on our Child-OID dataset, which demonstrates the effectiveness of our method. 2) Moreover, the proposed UCDA framework is obviously superiors to the YOLO networks, improving the accuracy score from 90.35% to 95.25%. 3) Benefitting from the advantages of one-stage object detection algorithms, DSSD can achieve reliable performance, surpassing the SSD by 1.56% in terms of accuracy. Although making some progress, our method consistently outperforms these aggregated results, especially for detection accuracy
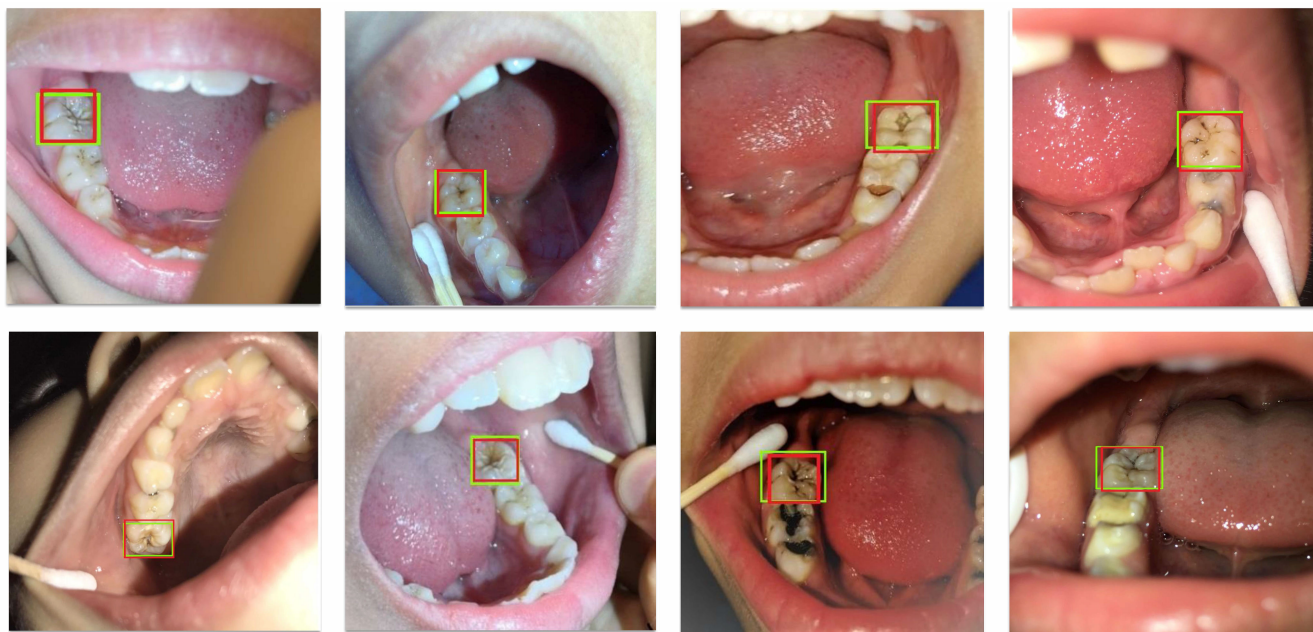
FIGURE 8. The diagnostic results of the proposed UCDA framework for dental caries detection on the children's first permanent molar. In this figure, the red boxes denote the predicted bounding boxes and the green boxes indicate the ground-truth boxes/labels.

(95.25% vs. 93.12%), the sensitivity (89.83% vs. 87.39%), and specificity (96.10% vs. 94.34%) with improvements of over 2.0 percent. Given the above comparative results, we can conclude that the proposed UCDA framework contributes a new technique to the diagnosis of dental caries on the Children's first permanent molar.

E. QUALITATIVE RESULTS

Fig. 8 illustrates the intuitive presentations of the first permanent molar caries detection. As shown, the proposed UCDA framework shows a satisfactory diagnosis effect in the diagnosis of dental caries on the children's first permanent molar. In particular, our method can effectively distinguish the different types of first permanent molars from the global oral image and correctly judge dental caries, which again verifies the feasibility of our method.

VI. CONCLUSION AND FURTHER WORK

In this paper, we propose a unified caries detection and assessment (UCDA) framework that is of the advantages of low cost and high performance for the diagnosis of dental caries on the children's first permanent molar. Moreover, we also present a novel children's oral image database with standard diagnostic annotations and labels, which we term the ''Child-OID''. Our Child-OID is the first benchmark image dataset for children's dental caries diagnosis. Extensive experiments on Child-OID demonstrate the effectiveness of the proposed UCDA framework in comparison with some state-of-the-art baselines. In the current epidemic of COVID-19, our work undoubtedly facilitates the application of automated caries diagnosis system and improves the physical and mental

health of children. In our future work, we would expand the proposed UCDA method toward the caries detection for all the teeth, which can further boost the development of the automated caries diagnosis system.

REFERENCES

[1] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and C. Zheng, "A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2615–2625, Aug. 2020.

[2] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[3] K. Yan, X. Wang, L. Lu, and R. M. Summers, "DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *J. Med. Imag.*, vol. 5, no. 3, p. 1, Jul. 2018.

[4] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "MURA: Large dataset for abnormality detection in musculoskeletal radiographs," 2017, *arXiv:1712.06957*. [Online]. Available: http://arxiv.org/abs/1712.06957

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–14.

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

[8] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[9] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[10] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, and W. Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 991–1004, Apr. 2019.

[11] A. Ben-Cohen, E. Klang, A. Kerpel, E. Konen, M. M. Amitai, and H. Greenspan, "Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations," *Neurocomputing*, vol. 275, pp. 1585–1594, Jan. 2018.

[12] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[13] J. R. Lino, J. Ramos-Jorge, V. S. Coelho, M. L. Ramos-Jorge, M. R. Moysés, and J. C. R. Ribeiro, "Association and comparison between visual inspection and bitewing radiography for the detection of recurrent dental caries under restorations," *Int. Dental J.*, vol. 65, no. 4, pp. 178–181, Aug. 2015.

[14] A. I. Ismail, W. Sohn, M. Tellez, A. Amaya, A. Sen, H. Hasson, and N. B. Pitts, "The international caries detection and assessment system (ICDAS): An integrated system for measuring dental caries," *Community Dentistry Oral Epidemiology*, vol. 35, no. 3, pp. 170–178, Jun. 2007.

[15] J. Kühnisch, I. Goddon, S. Berger, H. Senkel, K. Bücher, T. Oehme, R. Hickel, and R. Heinrich-Weltzien, "Development, methodology and potential of the new universal visual scoring system (UniViSS) for caries detection and diagnosis," *Int. J. Environ. Res. Public Health*, vol. 6, no. 9, pp. 2500–2509, Sep. 2009.

[16] J. Z. Anaise, "Measurement of dental caries experience-modification of the DMFT index," *Community Dentistry Oral Epidemiology*, vol. 12, no. 1, pp. 43–46, Feb. 1984.

[17] S. Jarvinen, "Epidemiologic characteristics of dental caries: Relation of DMFS to DMFT," *Community Dentistry Oral Epidemiology*, vol. 11, no. 6, pp. 363–366, Dec. 1983.

[18] M. Nishi, J. Stjernswärd, P. Carlsson, and D. Bratthall, "Caries experience of some countries and areas expressed by the significant caries index," *Community Dentistry Oral Epidemiology*, vol. 30, no. 4, pp. 296–301, Aug. 2002.

[19] L. Rezaei-Soufi, S. Kasraei, and M. Jazaeri, "Dental caries experience in 13-19-year-old iranian students expressed by DMFT and significant caries index," *Avicenna J. Dental Res.*, vol. 3, no. 2, pp. 102–109, 2018.

[20] J. Oliveira and H. Proença, "Caries detection in panoramic dental X-ray images," in *Computational Vision and Medical Image Processing*, vol. 19. Dordrecht, The Netherlands: Springer, 2011, pp. 175–190.

[21] A. Wenzel, "Bitewing and digital bitewing radiography for detection of caries lesions," *J. Dental Res.*, vol. 83, pp. 72–75, Jul. 2004.

[22] B. Newman, W. Seow, S. Kazoullis, D. Ford, and T. Holcombe, "Clinical detection of caries in the primary dentition with and without bitewing radiography," *Austral. Dental J.*, vol. 54, no. 1, pp. 23–30, Mar. 2009.

[23] A. Bhan, A. Goyal, Harsh, N. Chauhan, and C.-W. Wang, "Feature line profile based automatic detection of dental caries in bitewing radiography," in *Proc. Int. Conf. Micro-Electron. Telecommun. Eng. (ICMETE)*, Sep. 2016, pp. 635–640.

[24] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *J. Dentistry*, vol. 77, pp. 106–111, Oct. 2018.

[25] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *Proc. 2nd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2017, pp. 783–787.

[26] F. Söchtig, R. Hickel, and J. Kühnisch, "Caries detection and diagnostics with near-infrared light transillumination: Clinical experiences," *Quintessence Int.*, vol. 45, no. 6, pp. 531–538, 2014.

[27] J. Kühnisch, F. Söchtig, V. Pitchika, R. Laubender, and R. Hickel, "*In vivo* validation of near-infrared light transillumination for interproximal dentin caries detection," *Clin. Oral Investigations*, vol. 20, no. 4, p. 821, 2015.

[28] G. Ozkan and K. G. U. Guzel, "Clinical evaluation of near-infrared light transillumination in approximal dentin caries detection," *Lasers Med. Sci.*, vol. 32, no. 6, pp. 1417–1422, 2017.

[29] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Jan. 2014.

[31] R. Girshick, "Fast R-CNN," 2015, *arXiv:1504.08083*. [Online]. Available: https://arxiv.org/abs/1504.08083

[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[34] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[35] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 7263–7271.

[36] 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, pp. 2999–3007.

[37] C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "DSSD: Deconvolutional single shot detector," 2017, *arXiv:1701.06659*. [Online]. Available: https://arxiv.org/abs/1701.06659

HONGBING YU received the M.S. degree in oral clinical medicine from Guizhou Medical University, in 2010. He is currently an Associate Chief Physician with the Nanshan District Chronic Disease Prevention and Control Hospital, Shenzhen. His research interests include oral prevention, computerized medical diagnosis, and medical biometrics.

ZEBIN LIN received the B.S. degree in network engineering from South China Normal University, in 2018. He is currently pursuing the M.S. degree from the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. His current research interests include object detection, medical diagnosis, pattern recognition, and deep learning.

YISHU LIU received the B.S. degree in information security from the Harbin Institute of Technology, Weihai, in 2019. She is currently pursuing the M.S. degree with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. Her current research interests include computerized medical diagnosis, pattern recognition, deep learning, and machine learning.

JING SU received the M.Sc. degree in epidemiology and biostatistics from the Chinese University of Hong Kong, in 2016. She is currently working with the Shenzhen Nanshan District of Chronic Disease Prevention and Control Institute, and involves in project management of chronic diseases, such as oral diseases and so on.

BINGZHI CHEN received the B.S. degree in software engineering from South China Normal University, in 2017. He is currently pursuing the Ph.D. degree from the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. His current research interests include computerized medical diagnosis, pattern recognition, deep learning, and machine learning.

GUANGMING LU (Member, IEEE) received the B.S. degree in electrical engineering, the M.S. degree in control theory and control engineering, and the Ph.D. degree in computer science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1998, 2000, and 2005, respectively. He was a Postdoctoral Fellow with Tsinghua University, Beijing, China, from 2005 to 2007. He is currently a Professor with the Biocomputing Research Center, Harbin Institute of Technology, Shenzhen, China. He has published more than 120 technical papers at prestigious international journals and conferences, including TIP, TNNLS, TCYB, TCSVT, CVPR, AAAI, ACMM, and IJCAI. His current research interests include pattern recognition, image processing, and automated biometric technologies, and applications.

• • •