Semantic Decomposition Network With Contrastive and Structural Constraints for Dental Plaque Segmentation

Jian Shi, Baoli Sun, Xinchen Ye, Zhihui Wang, Member, IEEE, Xiaolong Luo, Jin Liu, Heli Gao, and Haojie Li

Abstract—Segmenting dental plaque from images of medical reagent staining provides valuable information for diagnosis and the determination of follow-up treatment plan. However, accurate dental plaque segmentation is a challenging task that requires identifying teeth and dental plaque subjected to semantic-blur regions (i.e., confused boundaries in border regions between teeth and dental plaque) and complex variations of instance shapes, which are not fully addressed by existing methods. Therefore, we propose a semantic decomposition network (SDNet) that introduces two single-task branches to separately address the segmentation of teeth and dental plaque and designs additional constraints to learn category-specific features for each branch, thus facilitating the semantic decomposition and improving the performance of dental plaque segmentation. Specifically, SDNet learns two separate segmentation branches for teeth and dental plaque in a divide-and-conquer manner to decouple the entangled relation between them. Each branch that specifies a category tends to yield accurate segmentation. To help these two branches better focus on category-specific features, two constraint modules are further proposed: 1) contrastive constraint module (CCM) to learn discriminative feature representations by maximizing the distance between different category representations, so as to reduce the negative impact of semantic-blur regions on feature extraction; 2) structural constraint module (SCM) to provide complete structural information of dental plaque of various shapes by the supervision of an boundary-aware geometric constraint. Besides, we construct a large-scale open-source Stained Dental Plaque Segmentation dataset (SDPSeg), which provides high-quality annotations for teeth and dental plaque. Experimental results on SDPSeg datasets show SDNet achieves state-of-the-art performance.

Index Terms—Dental plaque segmentation, semantic decomposition, contrastive constraint, structural constraint.

I. INTRODUCTION

DENTAL plaque is a soft, unmineralized biofilm strongly adhering to tooth surface which leads to various lesions such as dental caries and periodontitis [1], [2], [3]. Plaque stain reagent is commonly used in the clinical practice to stain dental plaque and the Quigley-Hein index (Q-H/TPI) [4], [5] is used to assess the oral hygiene state by scoring the distribution of dental plaque. However, estimation of the distribution of dental plaque depends on the clinician’s experience, leads to a high degree of subjectivity. The accuracy of evaluation varies dramatically among different physicians which drives us to using computer-aided diagnosis (CAD) systems for automatic diagnosis of dental plaque to get a more objective evaluation result. Segmentation of dental plaque is a fundamental step to build such CAD systems and crucial for accurate quantification of dental plaque distribution.

Recently, there have been some works on the classification and segmentation of dental plaque. Reference [6] presented an automated dental plaque image classification model based on Convolutional Neural Networks (CNN) on Quantitative Light-induced Fluorescence images. Reference [7] conducted low-shot learning at the super-pixel level of automatic dental plaque segmentation based on local-to-global feature fusion. Although existing dental plaque segmentation method has demonstrated remarkable progress, the discriminative ability in semantic-blur regions (i.e., confused boundaries in border regions between teeth and dental plaque) is still remained to be poor, as [7] neglected to distinguish the contextual dependencies of different categories, which may result in a less reliable context especially in semantic-blur regions. There are still several limitations in accurate segmentation of teeth and dental plaque: (1) as Fig. 1 shown, the semantic-blur regions with low contrast between dental plaque and teeth have lower confidence in classification. Therefore, how to reduce the negative impact of semantic-blur regions on feature extraction of teeth and dental plaque still needs to be studied; (2) complex structure information caused by high frequency variations in shape and size of the dental plaque instances increases the difficulty of segmentation.

Motivated by the above analysis, this paper introduces a novel semantic decomposition network (SDNet) which employs divide-and-conquer strategy for the task of dental plaque segmentation, as shown in Fig. 2. We explore for...
Fig. 1. Examples of oral images and corresponding labels in our proposed SDPSeg dataset. (a) and (b) show examples obtained by high-resolution laser scanner (SDPSeg-S) and high-definition camera (SDPSeg-C), respectively. Two major challenges in dental plaque segmentation can be observed from these examples: the varied shapes and sizes of dental plaque and the blurred boundaries between teeth and dental plaque.

Fig. 2. The Overall network architecture. The “feature extractor” is shared by the decoupled two-branch subnetworks which produce dental plaque prediction and teeth prediction in plaque branch and teeth branch, respectively. Contrastive constraint module (CCM) maximizes the distance between category representations in the semantic-blur regions. Structure constraint module (SCM) improves the structural integrity of instance by the supervision of an boundary-aware geometric constraint.

In order to learn category-specific feature representations for two separated branches, we introduce additional constraints from the space of latent features to clearly distinguish instances of different categories. Motivated by our observation that the regions around the interaction of teeth and dental plaque are more category-uncertainty than others, we devise a contrastive constraint module (CCM) based on contrastive learning [8], [9], [10] to enlarge the distance between the teeth and dental plaque category in the latent space, especially in the semantic-blur regions, which provides a more discriminative feature representation for semantic information decomposition. Secondly, to reduce the structural uncertainty caused by irregular variations of shape and size in different dental plaque instances, we further propose a structural constraint module (SCM) that provides holistic structure information for dental
plaque by an additional pixel-level supervision of a boundary-aware geometric constraint, so as to guide the learning of spatial details especially in the semantic-blur regions.

Besides, deep learning methods of semantic segmentation achieve excellent performance by training deep models with massive labeled data, but they still suffer from unreliable adaptation and overfitting when training on a small-scale dataset. As a result, a large-scale and high-quality dental plaque segmentation dataset is needed for the development of automatic diagnosis of dental plaque. However, stained dental plaque data is rare as the collection limited by the privacy of patients, and the large quantities of annotated data is also expensive. To the best of our knowledge, there is no specially designed stained dental plaque public dataset for CAD, which impedes the development of this research area. Therefore, we construct a large-scale open-source Stained Dental Plaque Segmentation dataset (SDPSeg) for research and assessment of the task. In SDPSeg, we present two high-quality public datasets, SDPSeg-S and SDPSeg-C which contain 565 and 1304 images captured by high-resolution laser scanner and HD camera, respectively. Both datasets contain three categories: teeth, dental plaque and background, each image contains a teeth instance and zero or several dental plaque instances. The labels in two datasets are meticulously annotated at pixel level by experienced experts, fine grained dental plaque and complex contours are also annotated in detail.

The key contributions are summarized as follows:

- We propose a novel semantic decomposition network that employs a divide-and-conquer strategy to effectively segment the dental plaque on teeth with two decomposed branches.
- We further devise two constraint modules (CCM and SCM) to address the problem of semantic-blur and instance shape variance by explicitly modeling the dissimilarity of the features of different categories through contrastive learning and providing holistic structural information for dental plaque with variant shapes and sizes through a boundary-aware geometric constraint.
- We develop two large-scale datasets (SDPSeg-S, SDPSeg-C) that cover a wide range of age groups and conform to the real situation of dental plaque distribution in clinical diagnosis, which is suitable for the real-world CAD of dental plaque.
- Extensive experiments on our proposed SDPSeg datasets show that our SDNet outperforms previous state-of-the-art methods without introducing additional computation during inference time, since both additional constraints are only required in training stage. Our method is more accurate in estimating the ratio of dental plaque to teeth compared to clinicians, which indicates that our effectiveness in CAD.

II. RELATED WORK

A. Medical Image Segmentation

Most semantic segmentation frameworks based on convolutional neural networks are either full convolutional networks (FCN) [11] or U-shaped networks based on an encoder-decoder architecture like UNet [12] designed for HeLa cells and neuronal structures of electron microscopic stacks. Various modifications based on these two forms have been proposed for natural image segmentation [16], [17] and medical image segmentation [23], [24]. The purpose of polyp segmentation is to accurately segment polyps from a given colonoscopy image. UACANet [25] proposed uncertainty augmented context attention network to augment the uncertainty of the boundary area. Parallel reverse attention network [26] utilized a parallel partial decoder to generate the high-level semantic global map and three reverse attention modules to mine complementary regions and details from high-level features. In the field of retinal vessel segmentation, the semantic aggregation module and multi-scale aggregation module were used in [27] to obtain more semantic feature representation and extract multi-scale information, respectively. CAG-Net [28] employed prediction module and refinement module to generate more accurate results for retinal vessel segmentation. A Study Group Learning (SGL) scheme [29] was proposed to improve the robustness of the model which trained on noisy labels. In 3D prostate segmentation, MS-Net [30] utilized a multi-site network which learned robust representation and leveraged multiple sources of data for improving segmentation results. Jia et al. [31] proposed a 3D adversarial pyramid convolutional deep neural network which consisted of a generator performing image segmentation and a discriminator that distinguished segmentation result and its corresponding groundtruth. Medical image segmentation plays a great role in the process of CAD. However, in the clinical diagnosis of dental plaque, the application of CAD is still insufficient.

B. Dental Plaque Segmentation

Dental plaque segmentation task aims to accurately segment the dental plaque region and teeth region with medical reagent staining. In this paper, we first propose a stained dental plaque segmentation task and an open-source stained dental plaque segmentation dataset. In other tasks related to dental plaque, such as dental plaque classification and dental plaque segmentation tasks, [6] proposed a dental plaque image classification dataset containing 427 images obtained by Quantitative Light-induced Fluorescence (QLF) camera where dental plaque fluoresces red, and presented an automated dental plaque image classification model based on Convolutional Neural Networks (CNN). Li et al. [7] proposed a dental plaque segmentation dataset containing 607 images without staining obtained by oral endoscopes, the dataset is private, and the annotation method that dentists mark the regions of plaque referring to the post-stained images leads to inaccurate annotation. [7] also proposed an automatic dental plaque segmentation network which fused multi-scale contextual features using traditional algorithms like heat kernel signature (HKS) [32] and local binary pattern (LBP) [33] for the daily monitoring of the patient with no reagent staining. Unfortunately, there is a lack of a public dataset of stained dental plaque and research studies in the clinical diagnostic application of dental plaque segmentation. Therefore, we present an open-source dental plaque staining dataset and a semantic decomposition framework utilizing divide-and-conquer strategy to improve the performance of dental plaque segmentation.
C. Divide-and-Conquer Strategy

Lin et al. [34], Kim et al. [35], Huang et al. [36] and Xu et al. [37] proposed adversarial learning network based on divide-and-conquer strategy for generation, super-resolution, enhancement of images and robust semantic segmentation, respectively. Lin et al. [34] reduced memory requisition during training and inference by dividing image generation into separated parallel sub-procedures. Kim et al. [35] designed a novel GAN-based joint SR-ITM network with a divide-and-conquer approach which was divided into three task-specific subnets: an image reconstruction subnet, a detail restoration (DR) subnet and a local contrast enhancement (LCE) subnet. Huang et al. [36] decomposed the photo enhancement process at three levels of perception, frequency and dimension. Xu et al. [37] employed dynamical division mechanism and automatically divided pixels into multiple branches. In the field of medical image segmentation, we innovatively utilize the divide-and-conquer strategy to divide the plaque segmentation task into two sub-problems, which uses two category-specific branches to solve the segmentation of teeth and plaque, respectively.

D. Contrastive Learning

Contrastive learning [38], [39], [40], a successful variant of self-supervised learning (SSL), forcing the embedding features of similar images to be close in the latent space and those of dissimilar ones to be apart, is often used in semi-supervised or self-supervised medical image segmentation with limited annotations. Krishna et al. [41] proposed a novel domain-specific contrastive strategies leveraging the structural similarity across volumetric medical images. Hu et al. [42] proposed a supervised local contrastive loss which forced pixels with the same label to gather in the embedded space with limited pixel-level annotations. Bai et al. [43] introduced contrastive learning to enhance contextual relationships between pixels in a dataset rather than just in an image. However, in this paper, we employ CCM which employs contrastive loss to construct positive and negative pairs of teeth and dental plaque category in an image to enlarge the differences between different categories as an additional constraint in the supervised learning.

E. Boundary-Aware Information

The boundary of object is often used in medical image segmentation as guidance information. Wang et al. [57] proposed a novel boundary coding network (BCNet) to learn a discriminative representation for organ boundary and used it as the context information to guide the segmentation. Wen et al. [58] presented two knowledge distillation modules, namely boundary-guided deep supervision and output space boundary embedding alignment, to explicitly transfer boundary information. Zhang et al. [59] and et al. [60] shared a similar boundary attention idea, where the object boundary was implicitly extracted from region predictions with a foreground erasing mechanism. Different from the above methods, we integrate the boundary-aware geometry constraint CCM into the loss function to provide holistic structure information for plaque segmentation.

III. PROPOSED METHOD

A. Overview

The overview of our proposed semantic decomposition network (SDNet) is illustrated in Fig. 2. The whole network is mainly divided into three components: (1) a shared feature encoder for spatial feature extraction; (2) two category-specific segmentation branches, i.e., dental plaque branch and teeth branch, for predicting teeth and dental plaque segmentation masks, respectively; (3) two constraint modules consisting of contrastive constraint module (CCM) and structure constraint module (SCM), for introducing additional constraints from the space of latent feature to clearly distinguish instances of different categories in the semantic-blur regions between teeth and dental plaque and provide complete structural information for dental plaque with various shapes.

The goal of our SDNet is to improve the segmentation accuracy of dental plaque by decoupling the entangled relation between dental plaque and teeth in a divide-and-conquer strategy. Given an input image \( I \), the shared feature encoder first extracts the intermediate features \( F \) for latter segmentation branches through continuous convolution layers and downsampling operations. Then the prediction is generated by decoding the required category-specific features from the shared encoder through different supervisions in each branch. CCM is introduced in the front stage of the decoder that can model the irrelevance of the category-specific features in high level, which enlarges the difference between categories and obtain more discriminant features for decomposition. SCM is introduced in the last stage of the decoder that can improve the structural integrity of instances by introducing low-level features to recover the boundary detailed information.

B. Shared Feature Encoder

As shown in Fig. 2, we employ a shared encoder \( E_s \) to extract both teeth and dental plaque features from an input oral image \( I \). The encoder network \( E_s \) contains 5 stacked convolution blocks. Each convolution block consists of the typical architecture of two \( 3 \times 3 \) convolution layers, a rectified linear unit (ReLU) activation function, and each convolution block is followed by a \( 2 \times 2 \) max pooling operation with stride 2 for feature downsampling except for the last one.

C. Semantic Decomposition

As we observed, many segmentation errors generally occur in the conflict areas between categories, especially the junction areas or overlapping areas. Previous approaches resolved the prediction conflicts between neighboring objects through context prior information or additional post-processing [44]. Consequently, their results were over-smooth along boundaries or exhibited small gaps between neighboring objects. Motivated by the divide-and-conquer strategy, we design two independent branches to decouple the entangled relation of teeth and dental plaque, which employ unique feature decoders...
to decode category-specific features as required according to the difference of semantic information between categories. Each different feature decoder aims to generate category-specific prediction for teeth and dental plaque with the guidance of corresponding supervision.

Specifically, given the feature map $F$ from the shard encoder $E_s$, the dental plaque decoder $D_p$ and teeth decoder $D_t$ produce dental plaque prediction and teeth prediction, respectively. The branch for dental plaque prediction is designed in a simple yet effective way: a decoder with four stacked blocks where each block contains a upsampling layer and a convolution block. And finally, an output layer with a series of convolution layers is applied to obtain mask predictions. The structure of teeth prediction branch is the same as that of dental plaque prediction. The segmentation loss $L_S^p$ for the dental plaque mask prediction is defined as:

$$L_S^p = L_{CE}(P_p, Y_p),$$  \tag{1}

and the loss for teeth mask prediction is defined as:

$$L_S^t = L_{CE}(P_t, Y_t),$$  \tag{2}

where $L_{CE}$ denotes the cross-entropy loss function. $P_p$, $P_t$ denote the dental plaque prediction and teeth prediction, and $Y_p$, $Y_t$ denote the binary ground truths for dental plaque and teeth obtained by channel separation from the original masks.

### D. Contrastive Constraint Module (CCM)

Contrastive Learning aims to learn discriminative representation by comparing positive and negative pairs which is usually used in the field of self-supervised representation learning to keep the distribution of unlabeled data consistent with that of labeled data. Specifically, our key idea innovatively employs contrastive learning to maximize the divergence among different categories and strengthen category-specific features in the supervised learning.

As shown in Fig. 2, we introduce CCM before the first layer of each separated decoder to push two prediction maps from each other in high-level feature space. Given two feature maps $F_p$, $F_t$ with the spatial resolution $h \times w$, and $C$ feature channels from the convolution layer before the first stage of dental plaque decoder $D_{p}$ and teeth decoder $D_{t}$, respectively, two projection heads $proj_{p}$, $proj_{t}$ are introduced to map two feature maps $F_p$ and $F_t$ into the vector spaces $\{f_i^{p}\}_{i=1}^{wh}$ and $\{f_i^{t}\}_{i=1}^{wh}$ where each pixel-level feature $f_i^{p}$ or $f_i^{t}$ is regarded as an independent class-wise representation for contrastive constraint. The projection head can be instantiated as any of the existing projection heads such as the ones in [8] and [39], which takes the dense feature map as input and generates multiple feature vectors for each class representation. In this paper, each projection head consists of a series of convolution layers used for reducing channel dimensions and three fully connected layers. The latent dental plaque feature representation is obtained by:

$$\{f_i^{p}\}_{i=1}^{wh} = proj_{p}(F_p).$$  \tag{3}

Similarly, we obtain the teeth representation:

$$\{f_i^{t}\}_{i=1}^{wh} = proj_{t}(F_t).$$  \tag{4}

Then, the cosine similarity is used to measure the difference between dental plaque and teeth category-specific representation in latent feature space:

$$L_{CCM} = \sum_{i=1}^{wh} \frac{f_i^{p} \cdot f_i^{t}}{\| f_i^{p} \|_2 \cdot \| f_i^{t} \|_2},$$  \tag{5}

where $\| \cdot \|_2$ is L2-normalized of corresponding feature representation. The contrastive constraint loss is devised based on contrastive learning [45], and the similarity between teeth feature $f_i^{t}$ and dental plaque feature $f_i^{p}$ gradually decreases as the loss function gradually tends to 0 during the training process, which enlarges the distance between the two representations in the high-level feature space, so the two decoders focus on more category-specific feature and thus each branch can get more accurate segmentation result in semantic-blur regions.

### E. Structure Constraint Module (SCM)

For dental plaque with varied shapes and sizes, the complete structure of dental plaque instance is difficult to accurately predict. In order to get more structure information, we introduce the structure constraint module (SCM) which employs the supervision of an boundary-aware geometric constraint. Specifically, SCM is designed to predict boundary-aware mask model in both two branches.

We employ the Canny operator (Fig. 3 shows the samples generated by Canny operator, the first row shows the results on SDPseg-S dataset, and the second row shows the results on SDPseg-C dataset) to generate the binary boundary maps $Y_{p}^{e}$ and $Y_{t}^{e}$ for dental plaque and teeth from the binary mask ground truths $Y_p$ and $Y_t$. Structure constraint module (SCM) updates the decoder parameters shared with mask prediction branch through back propagation, achieving the purpose of refining segmentation output. Therefore, it is particularly important to use an effective loss function which can constrain convolution network to generate an accurate and sharp boundary prediction. Highly imbalanced categories of boundary versus background in training data leads that the boundary constraint is not robust enough. Followed [46], we employ compound boundary loss that contains binary cross-entropy loss and dice loss [47] to optimize the imbalanced boundary prediction.

Dice loss only measures the distance between prediction and ground truth at pixel level regardless of which category
the pixel belongs to, thus reducing the imbalance between foreground and background pixels. Our loss $L_{SCM}^p$ for dental plaque structure constraint is:

$$L_{SCM}^p = \alpha L_{BCE}(E_p, Y_{p}^c) + \beta L_{Dice}(E_p, Y_{p}^c),$$

where $E_p$ denotes the predicted boundary for dental plaque and $Y_p^c$ denotes the corresponding boundary groundtruth. $\alpha$, $\beta$ are hyper-parameters to adjust the weight of BCE loss and Dice loss. Dice loss is formulated as follows:

$$L_{Dice} = 1 - \frac{2 \sum_{i}^{H \times W} (p_{i})^2 + \tau}{\sum_{i}^{H \times W} (p_{i})^2 + \sum_{i}^{H \times W} (y_{i})^2 + \tau},$$

where $i$ denotes the $i$-th pixel of the feature map, $\tau$ is a smooth term to avoid zero division.

The design of SCM in teeth branch is the same as the dental plaque branch. And $L_{SCM}^t$ and $L_{SCM}^p$ have the same form of computation.

$$L_{SCM}^t = \alpha L_{BCE}(E_t, Y_{t}^c) + \beta L_{Dice}(E_t, Y_{t}^c),$$

where $E_t$ denotes the predicted boundary for teeth and $Y_t^c$ denotes the corresponding boundary ground truth. $\alpha$, $\beta$ are hyper-parameters to adjust the weight of BCE loss and Dice loss, which have the same setup as dental plaque branch.

During the training process, SCM feeds back to the decoder by optimizing the output of the boundary prediction branch. In particular, the output of the SCM module is a fine plaque boundary map or teeth boundary map, and the accurate boundary information provided by boundary map can model the spatial variation trend along the side of boundary contours for the decoder of teeth and dental plaque to identify the geometry of the entire instances, which significantly boosts performance on thinner and smaller objects.

F. Network Training

The whole architecture can be trained in an end-to-end manner using the total loss function:

$$L = L_S^p + L_S^t + L_{SCM}^p + L_{SCM}^t + L_{CCM},$$

where $L_S$ is the standard implementations of the segmentation loss function for dental plaque or teeth instances, $L_{SCM}$ is the loss function in SCM. $L_{CCM}$ is the loss in CCM. Since boundary and mask are crossed linked by sharing one decoder, optimizing the loss for boundary also enhances the feature representation for mask prediction. Moreover, optimizing the loss in CCM can push confused regions belonging to different categories away from each other which also benefits the boundary prediction.

IV. THE STAINED DENTAL PLAQUE DATASET

The lack of high-quality, open-source and finely annotated datasets is an important factor that has hindered the vigorous development of the field of dental plaque segmentation. Our goal is to introduce a worthy benchmark with high-quality annotations to the community of dental plaque segmentation and improve the accuracy of computer-aided diagnosis.

A. Dataset Collection and Annotation

Our datasets are collected by the dentists at Shanghai Shanda dental clinic. According to the different oral digital imaging devices, i.e., 3-shape intraoral scanner and HD camera, we put forward two datasets, which are referred as SDPSeg-S (stained dental plaque segmentation dataset obtained by scanner) and SDPSeg-C (stained dental plaque segmentation dataset obtained by camera), respectively.

The oral images in the SDPSeg-S dataset are obtained by a high-resolution laser scanner (iTero, America). 3-shape intraoral scanner is an advanced oral digital impression system in the world which uses ultrafast optical cutting technology and confocal microscopy technology to capture multi-frame 2D images to create 3D impressions in real time. 3-shape oral scan is comfortable, accurate and efficient, and will definitely become the trend of the future. Therefore, we collect 2D images of each patient’s left view, right view, front view, maxillary and mandibular occlusal view generated by 3D oral scanner to construct the dataset. The oral images in SDPSeg-C dataset are taken by an HD camera (Cannon 750D, Japan) and also collected from each patient’s left view, right view and front view, maxillary and mandibular occlusal view. This collection method is more convenient and is one of the main collection methods at present.

To get a more fine dental plaque contour, for both datasets, we first crop out every teeth visible in the field of vision, normalize its size, and then annotate the dental plaque and teeth areas pixel by pixel using annotation tool. All annotations are performed by 6 people guided by professional dentists, and the annotation results are rechecked under strict quality control of the experienced dentists.

B. Dataset Statistics

1) Dataset Splitting: Totally, we collect 565 and 1304 images for SDPSeg-S and SDPSeg-C datasets, respectively. The two datasets are both randomly split into training set, test set and validation set with a ratio of 8:1:1, which contain 455, 57, 57 images and 1043, 131, 130 images, respectively.

2) The Proportion of Plaque: Both of the two datasets contain three categories: teeth, dental plaque and background. In this section we describe the complex distribution of dental plaque. The SDPSeg-S and SDPSeg-C are divided into 11 severity levels from 0%, 1-10% to 91-100% by calculating the ratio of dental plaque area to teeth area in each oral image. Fig. 4 (a) shows the dental plaque ratio on SDPSeg-S dataset, where oral
images with and without dental plaque account for 46.9% and 53.1%, respectively. Fig. 4 (b) shows the dental plaque ratio in SDPSeg-C dataset, where oral images with and without dental plaque account for 81.1% and 18.9%, respectively. There are significant differences in dental plaque distribution among different ages, oral sites and disease states. The oral images in SDPSeg-S and SDPSeg-C datasets were collected from people with different degrees of oral disease and age distribution, so the distribution of dental plaque varies greatly.

Observing from the statistical result of images with dental plaque, the data of mild dental plaque, i.e., 1-30%, accounts for majority on both SDPSeg-S and SDPSeg-C. On SDPSeg-S dataset, the proportion of dental plaque in all data is less than 90%, while on the SDPSeg-C dataset with oral problems, the data with dental plaque accounting for more than 90% of teeth occupy 2% of the dataset. Overall, the distribution of stained dental plaque data on both datasets conforms to the real situation of dental plaque distribution in clinical diagnosis.

3) The Distribution of Age: All oral data in the two datasets are both provided by adults. Fig. 5 illustrates the distribution of age on SDPSeg-S and SDPSeg-C datasets. SDPSeg-S dataset covers three age groups: 19-29, 29-39, and 49-59 years old, among which the age group 29-39 accounts for the largest proportion, 57.1%. SDPSeg-C dataset includes the data of all ages from the adolescent to the elderly, among which the age group 29-39 accounts for the largest proportion, 57.1%, the age group 39-49 accounts for 40.9%, and SDPSeg-C dataset also contains stained dental plaque images of some older groups, i.e., the age group 59-69 accounts for 9.1%.

V. EXPERIMENTAL RESULTS

A. Evaluation Metric

In this paper, we employ the commonly used evaluation metrics MIoU and Dice in medical image segmentation. In addition, to provide a more objective reference for the clinical diagnosis of doctors, we propose a new metric pixel ratio (PR). PR is the ratio of dental plaque area to teeth area which is a major index for doctors to make a decision on the severity of dental plaque, and for a testing image, the PR is defined as follows:

$$PR = \frac{\sum_{i=0}^{H} \sum_{j=0}^{W} P_{ij}^p}{\sum_{i=0}^{H} \sum_{j=0}^{W} P_{ij}^p + P_{ij}^t}$$

where $H$ and $W$ represent the height and width of the image, respectively. $P_{ij}^p$ and $P_{ij}^t$ represent the pixels belonging to the dental plaque category and teeth category, respectively. Here, the PR value of the ground truth marked by the doctor is defined as $PR_{gt}$, and $PR_{pre}$ indicates the PR value of the estimated results of our method or doctors/nurses. Therefore, we define $ERR$ as an evaluation metric to measure the accuracy of different approaches in estimating the plaque ratio for the testing images as follows.

$$ERR = \frac{1}{N} \sum_{i=0}^{N} \| PR_{gt} - PR_{pre} \|$$

where $N$ denotes all images of the test dataset, and $ERR$ represents the average error between the estimation and ground truth of plaque ratio.

B. Implementation Details

During the training stage, we use the data division method of the proposed dataset, and the image dimensions used for training is 128 × 128 which is the original dimensions of the proposed dataset. We use Adam optimizer [55] with momentum = 0.9, $\beta_1 = 0.9$, $\beta_2 = 0.99$ and $\epsilon = 10^{-8}$ to train our model. The batch size is set to 16 on both datasets. The total epochs is set to 120, and the initial learning rate is set to $1e^{-4}$ and decreased by multiplying 0.1 every 40 epochs on SDPSeg-S dataset. And the total epochs is set to 300, and the initial learning rate is set to be $1e^{-4}$ and decreased by multiplying 0.1 every 100 epochs on SDPSeg-C dataset. The hyper-parameters $\alpha$, $\beta$ in compound loss function $L_{SCM}$ are set to 0.1, 1 on both two datasets. We implement our method using PyTorch with a single Tesla V100-SWM2 GPU.

C. Results and Analysis

1) Results on SDPSeg-S Datasets: Tab. I shows a comparison of our method on SDPSeg-S dataset with some recent methods including UNet and UNet variants in the field of medical segmentation. To be fair, all the comparison methods in Tab. I use the same training methods and training parameters as ours including the initial learning rate, batch size, delay method of learning rate, and the data argument methods are the same (including horizontal flip and vertical flip), too. SDNet outperforms all the methods on the this dataset and the improvement is particularly significant on the category of dental plaque. SDNet achieves a MIoU of 90.35%, Dice of 95.58% on teeth category, and a MIoU of 80.08%, Dice of 77.02% on dental plaque category. We can observe an improvement of 1.28% and 1.68% over UNet on the teeth category in MIoU and Dice, respectively, and the improvement is particularly large on the more complex dental plaque category, which reaches 9.84% and 9.62% in MIoU and Dice, respectively. Compared with TransUNet [53], which takes advantage of both transformers [56] and UNet, SDNet also outperforms it by 1.27% in MIoU and 1.28% in Dice on the teeth category, and 8.30% and 6.37% on the dental plaque category. As shown in the 1st and 2nd rows of Fig. 6, we present the visual comparison on SDPSeg-S. All of the above methods can obtain outstanding segmentation results for the simple segmentation samples containing only
Fig. 6. Qualitative visual results on dental plaque segmentation task compared with other methods. The first two rows are the results on SDPSeg-S dataset, and the last two rows are the results on SDPSeg-C dataset. (a) Original images; (b) UNet [12]; (c) Attention UNet [49]; (d) Channel UNet [51]; (e) TransUNet [53]; (f) SDNet (Ours); (g) Ground truth. Our method can segment relatively complete dental plaque regions and has a stronger ability to recognize details.

| Methods                  | SDPSeg-S | SDPSeg-C | Time (ms) |
|--------------------------|----------|----------|-----------|
|                           | Teeth MIoU | Dice | Teeth MIoU | Dice | Teeth MIoU | Dice |              |
| UNet [12]                 | 89.07     | 93.90   | 70.24     | 67.40 | 86.80     | 91.15 | 75.44 | 75.78 | 129.21 |
| ResUNet [48]              | 86.38     | 93.08   | 67.33     | 65.72 | 86.00     | 90.04 | 72.54 | 70.94 | 144.59 |
| Attention UNet [49]       | 89.61     | 94.46   | 70.66     | 68.09 | 87.56     | 91.01 | 76.75 | 76.45 | 157.94 |
| R2 UNet [50]              | 71.31     | 86.88   | 51.39     | 53.97 | 79.96     | 87.18 | 64.10 | 62.59 | 183.12 |
| Channel UNet [51]         | 89.76     | 94.35   | 68.24     | 63.26 | 87.33     | 90.98 | 75.97 | 75.56 | 155.69 |
| KiUNet [52]               | 85.54     | 92.94   | 66.04     | 64.70 | 84.61     | 90.50 | 70.41 | 70.34 | 160.75 |
| TransUNet [53]            | 89.08     | 94.30   | 71.78     | 70.65 | 87.38     | 91.42 | 75.95 | 76.05 | 236.57 |
| MedT [54]                 | 70.02     | 85.80   | 46.27     | 43.29 | 77.29     | 83.05 | 62.82 | 59.23 | 447.90 |
| SDNet (Ours)              | 90.35     | 95.58   | 80.08     | 77.02 | 87.55     | 94.94 | 82.15 | 83.08 | 57.43  |

2) Results on SDPSeg-C Dataset: The quantitative results of our SDNet and other methods on SDPSeg-C dataset are also shown in Tab. I. Our SDNet almost outperforms the other methods on all metrics, obtains a MIoU of 87.55% and a Dice of 94.94% on the teeth category, and a MIoU of 82.15% and a Dice of 83.08% on the dental plaque category. The improvement in Dice can be 3.79% and 7.30% on the dental plaque category compared with UNet.

Our SDNet also achieves an improvement of 20.49% in Dice on dental plaque category as compared to R2UNet [50], an improvement 23.85% in Dice as compared MTNet [54] which uses transformer mechanism to encode long-range dependencies and learn representations that are highly expressive. The visual comparison on SDPSeg-C dataset can be observed in Fig. 6 (3rd – 4th rows). Observing the 3rd and 4th rows in Fig. 6, when the distribution of dental plaque becomes more random, our segmentation results perform better than other methods, which produces prediction masks with nearly the same structures and shapes of dental plaque as compared to the mask ground truth, especially on the areas with fine boundaries.

3) Inference Time: Our method also achieves the optimal result in inference time, which has a shorter prediction time for an input image, just 57.43ms, compared with other methods.
As shown in Tab. I, we can observe that the inference time of our method is less than half compared with the simplest UNet, and the inference time of the more complex method MedT is nearly 8 times than our method. Note that, our proposed two constraint modules, i.e., CCM and SCM, are both utilized only in training stage to help two branches better focus on category-specific feature which can improve segmentation performance and do not introduce additional computation and parameters during inference.

4) Application in Computer Aided Diagnosis: The ratio of dental plaque to teeth is an important reference for the dental plaque index to evaluate the severity of dental plaque. Tab. II shows the average error of our method in predicting the ratio of dental plaque to teeth compared with doctors Dr. and nurses N. (2 doctors and 2 nurses are from the Shanghai Shanda dental clinic). ERR is an index introduced in section V-A to evaluate the accuracy of the ratio of dental plaque to teeth. We find that the accuracy of estimation varies obviously among different doctors and nurses because of the different experience levels of physicians. Our average error ERR reaches 2.41% on the SDPSeg-S dataset, which is 3.67% and 12.54% lower than experienced doctor and nurse, respectively. And on the SDPSeg-C dataset, our method also achieves optimal results, decreasing 5.29% and 6.02% compared with the experienced doctor and nurse. The above results shows that our method is more objective than the results roughly estimated and inferred by doctors/nurses which indicates that our method possesses promising application prospect in computer aided diagnosis.

5) Generalizability Experiment on Other Datasets: We conduct generalization experiment on the polyp segmentation datasets CVC-ClinicDB [61] and CVC-ColonDB [62] to verify the generalization of our method on other task. CVC-ClinicDB and CVC-ColonDB are public polyp datasets including polyp and background categories annotated pixel by pixel and contain 612 and 380 cases, respectively. Tab. III shows the performance of our method compared with other state-of-the-art methods. To keep the fairness of the experiments, all methods in the Tab. III follow the advise of [26] and take exactly the same training and testing dataset division. Compared with UNet [12], our method improves greatly in both datasets. On CVC-ClinicDB, the improvement of our method in MIOU is as high as 10.7%, and on CVC-ColonDB, the improvement reaches 31.4% and 29.1% in MDice and MIOU. We can observe an improvement of 1.9% and 0.8% over the representative method PraNet [26] in MDice and MIOU on CVC-ClinicDB dataset, and the improvement is particularly large on the more difficult dataset CVC-ColonDB, which could be achieved 11.4% and 9.5% in MDice and MIOU. Compared with the latest method [63] for polyp segmentation, our method achieves similar results on CVC-ClinicDB dataset, and on CVC-ColonDB dataset, our method improves by 7.3% and 6.5% on MDice and MIOU, respectively. The proposed method SDNet achieves comparable or much higher accuracy on the polyp segmentation datasets CVC-ClinicDB and CVC-ColonDB compared with the state-of-the-art methods, indicating that our method is not only designed for dental plaque segmentation task, but also effective on other tasks.

D. Ablation Study

We conduct ablation studies to investigate five questions regarding our SDNet: 1) the contribution of each key designed component to our model performance; 2) the impact of different loss functions for constraining the boundary information introduced by SCM; 3) the impact of different integration positions for CCM; 4) the impact of various edge operators which generate slight different boundaries for the model performance; 5) the impact of hyper-parameters in the training process.

1) Impact of Each Component: Tab. IV and Tab. V show the performance comparison in terms of MIOU and Dice among different configurations, including semantic decomposition (SD) structure for decoupling dental plaque and teeth instances, CCM and SCM on SDPSeg-C and SDPSeg-S datasets, respectively. We investigate the impact of above three key components by gradually integrating each component into

---

**TABLE II**

| Datasets       | Dr.1 | Dr.2 | N.1  | N.2  | Ours  |
|----------------|------|------|------|------|-------|
| **SDPSeg-S** (ERR$_a$) | 6.08 | 15.78| 14.95| 17.68| **2.41** |
| **SDPSeg-C** (ERR$_a$) | 9.77 | 10.34| 14.85| 10.50| **4.48** |

**TABLE III**

| Models       | ClinicDB | ColonDB |
|--------------|----------|---------|
| UNet [12]    | 0.823    | 0.750   |
| ResUNet [50] | 0.779    | -       |
| PraNet [26]  | 0.899    | 0.849   |
| SANet [63]   | 0.916    | **0.859** |
| SDNet(ours)  | **0.918** | 0.857   |

**TABLE IV**

| Models          | Teeth | Plaque |
|-----------------|-------|--------|
| UNet            | 89.07 | 70.23  |
| UNet + SD       | 89.33 | 75.86  |
| UNet + SD + SCM | 89.77 | 78.07  |
| UNet + SD + CCM | 89.92 | 77.34  |
| UNet + SD + SCM + CCM (SDNet) | **90.35** | **80.08** |

**TABLE V**

| Models          | Teeth | Plaque |
|-----------------|-------|--------|
| UNet            | 86.80 | 75.44  |
| UNet + SD       | 86.81 | 80.15  |
| UNet + SD + SCM | 87.06 | 81.19  |
| UNet + SD + CCM | 86.83 | 81.85  |
| UNet + SD + SCM + CCM (SDNet) | **87.55** | **82.15** |
our framework. (1) Baseline (UNet): the baseline starts from a UNet [12] network by feeding each oral image into it for teeth and dental plaque segmentation. (2) UNet + SD: the baseline is upgraded with a semantic decomposition (SD) structure. Compared with the baseline UNet, the performance is improved obviously on both two datasets, especially on the dental plaque category. (3) UNet + SD + SCM: the baseline is extended with SCM. By adding SCM on the setting (2), the performance gets a significant improvement on the results of dental plaque category compared with on the teeth category, which presents that our SCM can focus on the fine boundaries of dental plaque with various shapes and improve the dental plaque segmentation. (4) UNet + SD + CCM: the baseline is extended with CCM. By simply adding CCM on the setting (2), the performance also gets an obvious improvement on the results of dental plaque category. (5) UNet + SD + SCM + CCM: Compared with setting (2), we simultaneously add SCM and CCM to introduce stronger constraints to distinguish the features in the semantic-blur regions and provide complete structural information for dental plaque segmentation. Our final configuration gains the best performance, which demonstrates the effectiveness of category-specific semantic decomposition strategy and additional constraints.

### Table VI

**The Ablation of Different Loss Functions Introduced by SCM in Teeth and Dental Plaque Categories on SDPSeg-S and SDPSeg-C Datasets in Dice(%)**

|                | SDPSeg-S | SDPSeg-C |
|----------------|----------|----------|
|                | Teeth    | Plaque   | Teeth    | Plaque   |
| SD + CCM + BCE | 95.47    | 76.09    | 95.38    | 76.29    |
| SD + CCM + Dice| 95.33    | 75.33    | 95.33    | 75.33    |
| SD + CCM       | 95.47    | 76.09    | 95.38    | 76.29    |
| SD + Dice      | 95.33    | 75.33    | 95.33    | 75.33    |
| SD + BCE       | 95.47    | 76.09    | 95.38    | 76.29    |
| BCE            | 95.33    | 75.33    | 95.33    | 75.33    |
| Dice           | 95.47    | 76.09    | 95.38    | 76.29    |

### Table VII

**The Ablation of the Position of CCM Between the Teeth and Dental Plaque Branches on SDPSeg-S and SDPSeg-C Datasets in Dice(%)**

|                  | SDPSeg-S | SDPSeg-C |
|------------------|----------|----------|
|                  | Teeth    | Plaque   | Teeth    | Plaque   |
| After f1         | 95.47    | 76.09    | 95.38    | 76.29    |
| After f2         | 95.47    | 76.09    | 95.38    | 76.29    |
| After f3         | 95.47    | 76.09    | 95.38    | 76.29    |
| Our              | 95.47    | 76.09    | 95.38    | 76.29    |

Fig. 7 further demonstrates the visual performance of our proposed method. As shown in the segmentation results, SD structure which decouples teeth category and dental plaque category into two layers can pay more attention to dental plaque regions than UNet, the dental plaque instances can be more detailed and complete after adding CCM which models the difference between the two branches and SCM which can enhance the supervision of entire instance structure. Fig. 8 shows the boundary prediction result in our SCM on SDPSeg-S and SDPSeg-C datasets. It can be observed that SCM gets a relatively complete dental plaque contours, which provides complete structure information for accurately segmenting the whole instance.

**2) Impact of Each Loss Function for SCM:** In order to verify the ability of the compound function to supervise the boundary information introduced by the SCM, we reduce one of the loss functions each time to conduct experiments, i.e. only BCE loss or only Dice loss. The quantitative result shown in Tab. VI indicates that when gradually removing each loss from SCM, it brings a performance degradation from 83.08% to 82.51% (removing Dice loss) and 81.49% (removing BCE loss) for dental plaque category on SDPSeg-C. And the degradation trend is the same on SDPSeg-S, a decline of 2.21% and 0.43% when removing Dice loss removing BCE loss on dental plaque category, respectively. The experimental result proves that when using Dice loss and BCE loss together to supervise the boundary prediction in SCM, the performance achieves the best.

**3) Optimum Position of CCM:** We also set several integration positions of CCM, including after the first layer of decoder (After f1), after the second layer of decoder (After f2) and after the third layer of decoder (After f3), to compare with our integration position before the first layer of decoder. As shown in Tab. VII, the position we choose to insert CCM achieves the best performances on both datasets in dental plaque category. Although the Dice of the position After f2 is higher than our method in teeth category on SDPSeg-C dataset, the size of feature map increases with the upsampling operation in the decoder which leads to the exponentially increasing of computing cost. And more importantly, the dice on more complex category dental plaque is nowhere near as good as...
ours. Our method is the optimal choice of balancing computing cost and accuracy.

4) Impact of Training Hyper-Parameters: In this part, we conduct ablation experiments on hyper-parameters $\alpha$ and $\beta$ in the losses ($L_{\text{SCM}}^p$ and $L_{\text{SCM}}$) of the SCM during the training process. We set up the ablation experiment for the hyper-parameters according to the boundary perception ability of the loss function BCE loss and Dice loss. Dice Loss is more advantageous to solve the problem of category imbalance caused by the large difference between foreground and background pixels. Therefore, we fix the hyper-parameter $\beta$ in front of Dice Loss as 1, and transform $\alpha$ to verify the influence of different hyper-parameter ratios on the loss function. As can be seen in Table VIII, in the two datasets, the accuracy of both teeth category and dental plaque category generally presents a declining trend with the increase of $\alpha$. So we set $\alpha$ as 0.1 and $\beta$ as 1.

5) Impact of Edge Operator: To verify the robustness of our method in dealing with incorrect boundary labels, we conduct ablation experiments on SDPSeg-S dataset by using different edge operators to generate boundary ground truths for training. As shown in the Table IX, both on the plaque category and the teeth category, the Sobel and Roberts edge operators can achieve comparable performance to the Canny edge operator (Ours), which shows that our method is robust to the boundary ground truths generated by various edge operators with slight different boundaries.

VI. Conclusion

In this paper, we propose a semantic decomposition network (SDNet) which employs divide-and-conquer strategy for the task of dental plaque segmentation. Specifically, we employ semantic decomposition structure to focus on category-specific features in each branch, and then propose a contrastive constraint module (CCM) to maximize the distance between different category representations, a structural constraint module (SCM) to provide complete structure information for dental plaques with various shapes. Besides, We construct a large-scale open-source dataset SDPSeg consisting of stained dental plaque images for research and assessment of dental plaque segmentation. Extensive experiments demonstrate the effectiveness of our proposed method. More importantly, our method is more accurate than experienced doctors in evaluating the ratio of dental plaque to teeth, which indicates that our method possesses promising application prospect in computer-aided diagnosis.

References

[1] P. D. Marsh, “Microbial ecology of dental plaque and its significance in health and disease,” Adv. Dental Res., vol. 8, no. 2, pp. 263–271, 1994.
[2] J. M. Lovegrove, “Dental plaque revisited: Bacteria associated with periodontal disease,” J. New Zealand Soc. Periodontol., vol. 87, pp. 7–21, Jan. 2004.
[3] N. Takahashi and B. Nyvad, “The role of bacteria in the caries process: Ecological perspectives,” J. Dental Res., vol. 90, no. 3, pp. 294–303, Mar. 2011.
[4] G. A. Quigley and J. W. Hein, “Comparative cleansing efficiency of manual and power brushing,” J. Amer. Dental Assoc., vol. 65, no. 1, pp. 26–29, Jul. 1962.
[5] S. Turesky, N. D. Gilmore, and I. Glickman, “Reduced plaque formation by the chloromethyl analogue of victamine c,” J. Periodontol., vol. 41, no. 41, pp. 41–43, Sep. 1970.
[6] S. Imanguliyev, M. V. D. Veen, C. M. C. Volgenant, B. J. F. Keijzer, W. Crielard, and E. Levin, “Deep learning for classification of dental plaque images,” in Proc. MOD, vol. 10122, 2016, pp. 407–410.
[7] S. Li et al., “Low-shot learning of automatic dental plaque segmentation based on local-to-global feature fusion,” in Proc. IEEE Int. Symp. Biomed. Imaging, Apr. 2020, pp. 664–668.
[8] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 9726–9735.
[9] J. Grill et al, “Bootstrap your own latent—A new approach to self-supervised learning,” in Proc. Adv. Neural Inf. Process. Syst., 2020, pp. 1–14.
[10] X. Chen and K. He, “Exploring simplesiamese representation learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 15750–15758.
[11] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3431–3440.
[12] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), vol. 9351. Springer, 2015, pp. 234–241.
[13] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, Oct. 2017.
[14] H. Ding, X. Jiang, B. Shuai, A. Q. Liu, and G. Wang, “Semantic correlation promoted shape-variant context for segmentation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 8885–8894.
[15] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, “CCNet: Criss-cross attention for semantic segmentation,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 603–612.
[16] C. Yu, J. Wang, C. Gao, G. Yu, C. Shen, and N. Sang, “Context prior for scene segmentation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 12413–12422.
[17] H. Zhao et al, “PSANet: Point-wise spatial attention network for scene parsing,” in Proc. Eur. Conf. Comput. Vis. (ECCV), vol. 11213, Sep. 2018, pp. 270–286.
[18] Z. Gu et al., “CE-Net: Context encoder network for 2D medical image segmentation,” IEEE Trans. Med. Imaging., vol. 38, no. 10, pp. 2281–2292, Mar. 2019.
R. Xu, X. Ye, G. Jiang, T. Liu, L. Li, and S. Tanaka, "Retinal vessel segmentation trained with noisy labels," in *Proc. Med. Image Comput.* Comput.-Assist. Intervent. (MICCAI), vol. 11766. Springer, 2015, pp. 848–855.

W. Ji et al., “Learning calibrated medical image segmentation via multitask agreement modeling,” in *Proc. IEEE/CVF Int. Conf. Pattern Recognit.* (CVPR), Jun. 2021, pp. 12341–12351.

T. Kim, H. Lee, and D. Kim, “UACANet: Uncertainty augmented context attention for polyp segmentation,” in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 2167–2175.

D. Fan et al., “PrANet: Parallel reverse attention network for polyp segmentation,” in *Medical Image Computing and Computer Assisted Intervention (Lecture Notes in Computer Science)*, vol. 12266. Cham, Switzerland: Springer, 2020, pp. 263–273.

R. Xu, X. Ye, G. Jiang, T. Liu, L. Li, and S. Tanaka, “Retinal vessel segmentation via a semantics and multi-scale aggregation network,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1085–1089.

M. Li, Y. Zhang, Z. Xiong, and D. Liu, “Cascade attention guided network for retinal vessel segmentation,” in *Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 12069. Springer, 2020, pp. 62–71.

Y. Zhou, H. Yu, and H. Shi, “Study group learning: Improving retinal vessel segmentation trained with noisy labels,” in *Proc. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 12901. Springer, 2021, pp. 57–67.

Q. Liu, Q. Dou, L. Yu, and P. A. Heng, “MS-Net: Multi-site network for improving prostate segmentation with heterogeneous MRI data,” *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2713–2724, Sep. 2020.

M. M. Bronstein and I. Kokkinos, “Scale-invariant heat kernel signatures for non-rigid shape recognition,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1704–1711.

T. Ojala, M. Pietikainen, and T. Maenpaa, “Multi resolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Aug. 2002.

C. H. Lin, C.-C. Chang, Y.-S. Chen, D.-C. Juan, W. Wei, and H.-T. Chen, “COCO-GAN: Generation by parts via conditional coordinating,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.* (CVPR), Oct. 2019, pp. 4511–4520.

S. Y. Kim, J. Oh, and M. Kim, “JSI-GAN: GAN-based joint super-resolution and inverse tone-mapping with pixel-wise task-specific filters for UHD HDR video,” in *Proc. AAI Conf. Artif. Intell.*, vol. 34, no. 7, Apr. 2020, pp. 11287–11295.

Z. Huang, D. P. Paudel, G. Li, J. Wu, R. Timofte, and L. Van Gool, “Divide- and-conquer adversarial learning for high-resolution image and video enhancement,” 2019, arXiv:1910.10455.

X. Xu, H. Zhao, and J. Jia, “Dynamic divide- and-conque adversarial training for robust semantic segmentation,” 2020, arXiv:2003.06555.

P. Bachman, R. D. Hjelm, and W. Buch-Walter, “Learning representations by maximizing mutual information across views,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 15509–15519.

T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, “A simple framework for contrastive learning of visual representations,” in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

I. Misra and L. van der Maaten, “Self-supervised learning of pretext-invariant representations,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (CVPR), Jun. 2020, pp. 6706–6716.