GRAB-Net: Graph-Based Boundary-Aware Network for Medical Point Cloud Segmentation

Yifan Liu, Wuyang Li, Jie Liu, Hui Chen, and Yixuan Yuan, *Member, IEEE*

***Abstract*— Point cloud segmentation is fundamental in many medical applications, such as aneurysm clipping and orthodontic planning. Recent methods mainly focus on designing powerful local feature extractors and generally overlook the segmentation around the boundaries between objects, which is extremely harmful to the clinical practice and degenerates the overall segmentation performance. To remedy this problem, we propose a GRAph-based Boundary-aware Network (GRAB-Net) with three paradigms, Graph-based Boundary-perception Module (GBM), Outer-boundary Context-assignment Module (OCM), and Inner-boundary Feature-rectification Module (IFM), for medical point cloud segmentation. Aiming to improve the segmentation performance around boundaries, GBM is designed to detect boundaries and interchange complementary information inside semantic and boundary features in the graph domain, where semantics-boundary correlations are modelled globally and informative clues are exchanged by graph reasoning. Furthermore, to reduce the context confusion that degenerates the segmentation performance outside the boundaries, OCM is proposed to construct the contextual graph, where dissimilar contexts are assigned to points of different categories guided by geometrical landmarks. In addition, we advance IFM to distinguish ambiguous features inside boundaries in a contrastive manner, where boundary-aware contrast strategies are proposed to facilitate the discriminative representation learning. Extensive experiments on two public datasets, IntrA and 3DTeethSeg, demonstrate the superiority of our method over state-of-the-art methods.**

***Index Terms*— Point cloud segmentation, graph-based framework, boundary-aware segmentation.**

I. Introduction

POINT cloud segmentation is a fundamental technique in a wide range of medical applications. For instance, segmentation on 3D scanned data of dental models is beneficial for dentists to simulate teeth extraction, deletion, and rearrangement, easing the prediction of treatment outcomes [1], [2], [3], [4]. Another instance is the intracranial segmentation on 3D vessel surfaces, which provides informative boundary clues for the aneurysm clipping surgery process [5], [6]. In the clinical practice, one feasible way is to manually segmenting objects, however, it is labor-intensive and prone to inter-observer variability. Hence, there is a high demand for accurate and reliable automatic point cloud segmentation methods that can derive quantitative assessments.

In recent years, many point cloud segmentation methods have been proposed and they can be divided into three categories according to the design philosophy. *Extractor-focused* methods [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19] exhaustively design various local feature extractors to extract informative representations. *Semantics-enhanced* methods [20], [21], [22] instead try to introduce additional information to enhance the original semantic features. Though incremental performance achieved, they can hardly perform well around boundaries between objects, which is harmful to the clinical practice since many operations, e.g, aneurysm clipping [6] and teeth extraction [4], are performed along the boundary lines. To address this problem, *boundary-aware* methods [23], [24], [25] are proposed with boundary perception and boundary-aware contrastive strategies. The former strategy used in [23] and [24] is generally to perceive boundaries using an extra branch and incorporate the predicted boundary masks into the semantic features, while the latter one used in [25] aims to distinguish ambiguous features around boundaries in a contrastive manner [25].

Though improvements achieved around the boundary, there are still two challenges in existing boundary-aware frameworks [23], [24], [25]. Firstly, current methods directly combine semantic and boundary features and model their relations by the local feature extraction, which overlooks the global semantic and shape clues, causing *insufficient duality constraints*. More specifically, semantic and shape clues are hidden in respective features, and we expect the network to globally perceive the two kinds of information, i.e., the whole semantic distribution and the complete boundary shape, which can provide sufficient constraints for the network to produce appropriate features for segmentation. However, existing works merely perceive partial hidden information due to the locality introduced in the feature extraction process. Besides, directly combining two types of features adopted in current methods is rather coarse, which increases the difficulty for the

Manuscript received 30 January 2023; revised 27 March 2023; accepted 30 March 2023. Date of publication 6 April 2023; date of current version 31 August 2023. This work was supported in part by the Hong Kong Research Grants Council (RGC) Collaborative Research Fund under Grant C4063-18G and in part by the Innovation and Technology Commission-Innovation and Technology Fund under Grant ITS/100/20. *(Corresponding author: Yixuan Yuan.)*

Yifan Liu and Yixuan Yuan are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, SAR, China (e-mail: 1155195605@link.cuhk.edu.hk; yxyuan@ee.cuhk.edu.hk).

Wuyang Li and Jie Liu are with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong, SAR, China (e-mail: wuyangli2-c@my.cityu.edu.hk; jliu.ee@my.cityu.edu.hk).

Hui Chen is with the Faculty of Dentistry, The University of Hong Kong, Hong Kong, SAR, China (e-mail: amyhchen@hku.hk).

Digital Object Identifier 10.1109/TMI.2023.3265000

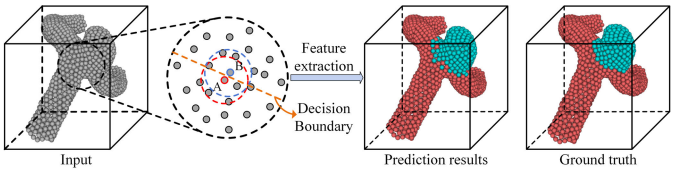

Fig. 1. Illustration of the context confusion during the convolution process nearby the boundary.

network to capture the mutual association. To solve these limitations, we use graph techniques to globally and finely model the semantic-boundary correlations. In the graph domain, duality correspondences can be represented by dynamically constructed graph connections [26], [27], which can capture global range dependencies. Based on that, graph convolution is performed to adaptively exchange desired messages, providing sufficient constraints to generate features equipped with finer details.

The second challenge lies in the *context confusion* existing in the feature extraction around boundary areas, which further causes ambiguous representations, i.e., the feature ambiguity problem. For example, as shown in Fig.1, points A and B are close in the coordinate space, therefore their contexts, i.e, the participants of the feature extraction, would be confused with high similarities. This context confusion can induce the network to produce similar features and further the same category predictions, while the two points should belong to different categories. To remedy this issue, for features outside the predicted boundaries, we are committed to assigning dis-similar contexts to points of different categories. Considering boundaries are natural geometrical landmarks for different categories, we can reformulate contexts as graph connections and cut off connections across the predicted boundaries, which ensures that points outside the predicted boundaries are only connected to the same-sided, i.e., the same-category points, and thus contexts of different categories are differentiated. While for the features inside the boundaries that have no side concept, we design a boundary-aware contrastive learning paradigm to reduce the feature ambiguity, where two types of contrast are advanced to achieve the discriminative representation learning.

To sum up, we propose a GRAph-based Boundary-aware Network (GRAB-Net) for medical point cloud segmentation with three novel paradigms. Firstly, We design a *Graph-based Boundary-perception Module (GBM)* to model semantics-boundary correlations in the graph domain, where relations can be modeled and complementary information can be inter-exchanged in a global manner. Secondly, to solve the context confusion problem outside the boundaries, we establish an *Outer-boundary Context-assignment Module (OCM)*, where a contextual graph is constructed to assign appropriate contexts to each point, refining the confused features with the following graph reasoning layers. Finally, to further reduce the feature ambiguity inside the boundary areas, an *Inner-boundary Feature-rectification Module (IFM)* is proposed to differentiate these features in a contrastive manner. To be summarized, our contributions are as follows.

- We propose a GRAph-based Boundary-aware Network (GRAB-Net) for point cloud segmentation. To the best of our knowledge, this work represents the first effort to improve the segmentation performance around boundaries in the medical point cloud segmentation task.
- We propose a novel graph-based boundary perception module GBM, which globally builds duality correlations and inter-exchange complementary information in the graph domain, producing features of better quality.
- We advance two novel paradigms OCM and IFM, which solves the context confusion and reduces the feature ambiguity outside and inside boundaries, respectively. OCM assigns appropriate context to points by leveraging conditional graph connections, and IFM distinguishes features inside boundaries in a contrastive manner.
- We validate the effectiveness of our proposed GRAB-Net in different medical point cloud segmentation tasks, including 3D intracranial aneurysm segmentation and 3D teeth segmentation. Comprehensive experiments testify that our method can surpass previous methods with large improvements.

II. Related Work

A. Surface Model Segmentation in Medical Domain

Recently, many automatic methods [3], [4], [6], [28], [29] for 3D surface model segmentation in medical scenarios have been investigated. Generally, they can be divided into mesh-based and point-based based on the input 3D data format [4].

Mesh-based methods [3], [28], [29] take triangle meshes as input, and produce corresponding labels for each mesh. Reference [3] proposes to extract multi-scale local contextual features via cascaded graph-constrained learning modules. Reference [28] adopts two graph-learning streams to extract discriminative feature representations from coordinates and normals space. Recently, [29] further advances a two-stage framework to segment meshes and regress anatomical landmarks simultaneously. The limitation of mesh-based methods lies in that triangle meshes are required besides raw point clouds, and they may also introduce inappropriate shape biases since triangle meshes are only local planar approximations of the real surface model.

In contrast, point-based methods [4], [6], i.e., point cloud segmentation methods, can directly process the raw point clouds and excavate the hidden topology information in a learnable manner, thus we follow this line of works in this paper. In [6], a public 3D aneurysm segmentation dataset IntrA is proposed and many representative works [7], [8], [9], [11], [30], [31] in general 3D vision are implemented on this dataset, revealing the transfer-ability of 3D segmentation methods from general domain to medical domain. Then [4] proposes a two-stage instance segmentation framework for intra-oral scan segmentation, which detects teeth centroids in the first stage, and then classifies the foreground and background in the following stage.

Though impressive performances achieved, these methods overlook the segmentation around boundaries, which is crucial

to the segmentation in medical scenarios. Besides, these methods are limited to the specific application, while our proposed GRAB-Net is a general boundary-aware framework for point cloud segmentation in medical scenarios.

B. Point Cloud Segmentation

Recently, various methods [7], [8], [11], [13], [14], [18], [20], [21], [22], [23], [24], [25], [32] for point cloud segmentation have been proposed, and they can be grouped into three lines according to the design philosophies.

Extractor-focused methods [7], [8], [9], [10], [11], [12], [13], [14], [16], [17], [18], [33] focus on designing efficient and effective local feature extractors. This line of work is pioneered by PointNet [7], which firstly utilizes permutation-invariant MLPs and MaxPoolings to directly process the raw point clouds. To further capture local patterns, PointNet++ [8] advances to build a hierarchical structure by applying PointNet recursively. Afther that, PointConv [11] and KPConv [13] borrow the idea from 2D convolutions, successfully applying discrete 3D convolutions on the point cloud. Besides, graph-based method DGCNN [32] is also proposed to build local graphs dynamically across layers. More recently, inspired by the success of transformers in 2D vision community [34], [35], [36], PointTransformer [18] is advocated for point-based 3D segmentation with modified local transformers.

Semantics-enhanced methods [20], [21], [22] expect to enhance the original semantic features by introducing additional information. Reference [20] advances a bilateral augmentation strategy to enhance the local geometrical and semantic context. Reference [21] proposes to incorporate global context during the local aggregation process, and [22] designs a category-guided aggregation module, utilizing different aggregation strategies between the same categories and different categories.

Boundary-aware methods [23], [24], [25] aim at improving the segmentation performance around boundaries, which are inspired by many works in 2D image processing [37], [38], [39], [40]. Reference [23] tries to improve the segmentation performance by jointly optimizing the semantics and boundary branches with shared encoders. Reference [24] proposes to incorporate the predicted boundary masks into semantic features during the feature extraction process. Recently, [25] is designed to distinguish ambiguous features around boundaries by leveraging contrastive learning techniques.

Though improvements around the boundary, these methods overlook the global semantics-boundary correlations and thus are insufficient in duality constraints. Besides, they ignore the context confusion problem during the feature extraction, which inevitably leads to ambiguous features. In this work, we propose a GRAB-Net framework to tackle these challenges, which models semantics-boundary correlations globally in the graph domain, and assign different contexts to points of different categories to reduce the feature ambiguity.

III. Method

In this work, we propose GRAB-Net for medical point cloud segmentation. It consists of three key components: Graph-based Boundary-perception Module (GBM), Outer-boundary Context-assignment Module (OCM), and Inner-boundary Feature-rectification Module (IFM). The whole framework workflow is illustrated in Fig. 2. Specifically, given the input point cloud $I \in \mathcal{R}^{N\times C}$ composed of coordinates $P \in \mathcal{R}^{N\times 3}$ and other attributes like normals, where N is the number of points, we first utilize a dual-branch backbone with shared encoders to extract semantic feature $X^s \in \mathcal{R}^{N\times D}$ and boundary feature $X^b \in \mathcal{R}^{N\times D}$, where D is the feature dimension. To explicitly model semantics-boundary relations in the graph domain, GBM (§ III-A) projects point-level features X^s and X^b into the semantic nodes V^s and boundary nodes V^b, and then relations across V^s and V^b are globally modelled and duality information are exchanged to obtain the updated graph nodes $\widehat{V}^s$ and $\widehat{V}^b$, which are further re-projected into point-level features $\widehat{X}^s$ and $\widehat{X}^b$. Then OCM (§ III-B) is proposed to solve the context confusion problem by assigning contexts of each point. In particular, the semantic feature $\widehat{X}^s$ and the boundary indicators $\widehat{Y}^b$ generated from $\widehat{X}^b$ are used to build the contextual graph G^c outside the predicted boundaries, based on which graph convolution is performed to generate the refined feature $\widetilde{X}^s$. Finally, to further reduce the feature ambiguity inside the boundaries, IFM (§ III-C) is proposed to project the semantic feature $\widetilde{X}^s$ into the embedding space E, distinguishing these embeddings in a contrastive manner with a dynamically updated memory bank M.

A. Graph-Based Boundary-Perception Module

Sufficiently leveraging duality constraints hidden in semantic and boundary features is necessary for accurate segmentation around the boundaries, which is crucial to several medical applications, e.g, aneurysm clipping [6] and teeth extraction [4]. To fulfill this need, we propose GBM to globally explore the duality constraints in the graph domain, where point-level features X are first projected into coherent nodes V, then correlations are modelled globally to generate updated nodes $\widehat{V}$, and finally, nodes are re-projected to obtain point-level features $\widehat{X}$.

1) Coherent Node Projection: Instead of directly constructing a graph on point-level features, which is computationally intensive and not robust to noise, we design the coherent node projection to project point-level features $X^s, X^b \in \mathcal{R}^{N\times D}$ from the backbone into more compact features $V^s, V^b \in \mathcal{R}^{M\times D} (M < N)$, and treat them as semantic and boundary graph nodes, respectively.

In particular, a subset $P^{sub} \in \mathcal{R}^{M\times 3}$ (red points in Fig. 3) are first uniformly sampled from the original point cloud $P \in \mathcal{R}^{N\times 3}$ by using the farthest point sampling (FPS) algorithm. Then, for each sampled point P_i^{sub}, k_1 nearest neighbors (including P_i^{sub}) are searched in P to compose the neighborhood $\mathcal{N}(P_i^{sub})$. Finally, the features of these neighbors are aggregated in a permutation-invariant way:

$$V_i = \gamma(MAX(h(X_{P_j}))), P_j \in \mathcal{N}_{k_1}(P_i^{sub}) \quad (1)$$

where $X_{P_j} \in X$ is the corresponding feature of point P_j, γ, h are multi-layer perceptron (MLP) layers, MAX is conducted along the point dimension, and $V_i \in \mathcal{R}^D$ is the aggregated

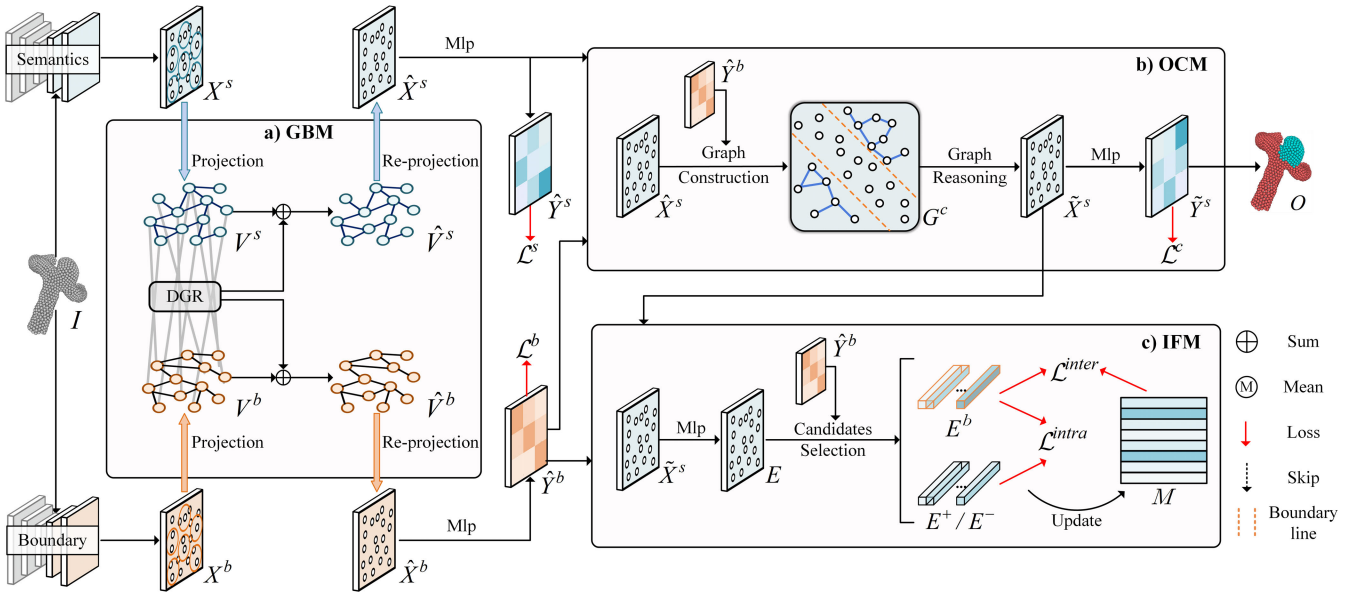

Fig. 2. Illustration of the proposed framework, composed of (a) Graph-based Boundary-perception Module (GBM), (b) Outer-boundary Context-assignment Module (OCM), and (c) Inner-boundary Feature-rectification Module (IFM). Input point clouds are fed into the dual-branch network to extract the semantic feature and boundary feature respectively. GBM models and inter-exchanges correlations between point and boundary features via Duality Graph Reasoning (DGR). Finally, IFM further reduce the feature ambiguity inside boundaries in a contrastive manner.

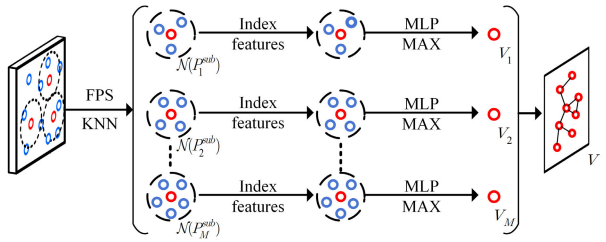

Fig. 3. Illustration of the coherent node projection process.

graph node of sampled point P_i^{sub}. By aggregating neighboring semantic and boundary features X^s and X^b of each sampled point P_i^{sub}, we can obtain the semantic and boundary graph nodes V^s and V^b, respectively.

2) Duality Graph Reasoning (DGR): Given projected semantic nodes V^s and boundary nodes V^b, we can build the duality graph $G^d = (V, A)$ to facilitate the information exchange. Specifically, the vertice set $V \in \mathcal{R}^{2M\times D}$ is the combination of V^s and V^b: $V = [(V^s)^T, (V^b)^T]^T$ and $A \in \mathcal{R}^{2M\times 2M}$ is the affinity matrix that represents the correlations between node features:

$$A = \begin{pmatrix} 0 & A^{b\to s} \\ A^{s\to b} & 0 \end{pmatrix}, \tag{2}$$

where $A^{b\to s} = \{A_{i,j}^{b\to s}\} \in \mathcal{R}^{M\times M}$ assembles the correlation weight from j-th node of V^b to i-th node of V^s, and $A^{s\to b}$ is explained reversely. To obtain the coefficient $A_{i,j}^{b\to s}$, attention mechanism that can capture long-range dependencies is used:

$$A_{i,j}^{b\to s} = \rho(\phi(V_i^s) - \psi(V_j^b) + \delta(P_i - P_j))), \tag{3}$$

where ϕ, ψ are single MLP layers, and ρ, δ are MLP mapping functions with two linear layers and one ReLU nonlinearity. The subtraction between $\phi(V_i^s)$ and $\psi(V_j^b)$ encodes feature relations between semantic and boundary nodes, and $\delta(P_i - P_j)$ reflects the relative geometrical correspondences. Analogously, $A^{s\to b}$ can also be obtained. With computed adjacency matrix A, we can incorporate boundary/semantics information into the semantics/boundary feature by:

$$\widehat{V} = A \cdot \sigma(V) + V, \tag{4}$$

where $\widehat{V} = [(\widehat{V}^s)^T, (\widehat{V}^b)^T]^T \in \mathcal{R}^{2M\times D}$ is the updated graph nodes, and σ is a single MLP layer. In doing so, the complementary information is inter-exchanged between semantic and boundary nodes globally, providing sufficient duality constraints for the feature generation.

3) Distance-Aware Node Re-Projection: To obtain point-level predictions, the three-nearest distance-aware interpolation strategy [7], [8] is adopted to re-project node-level features $\widehat{V} \in \mathcal{R}^{M\times D}$ into point-level features $\widehat{X} \in \mathcal{R}^{N\times D}$. The re-projected point-level semantic feature $\widehat{X}^s$ and boundary feature $\widehat{X}^b$ are further passed to MLP layers to obtain logits $\widehat{Y}^s$ and $\widehat{Y}^b$, supervised by cross-entropy loss $\mathcal{L}_s = -\frac{1}{N}\sum_{i=1}^{N} CE(\widehat{Y}_i^s, Y_i^{sgt})$ and $\mathcal{L}_b = -\frac{1}{N}\sum_{i=1}^{N} CE(\widehat{Y}_i^b, Y_i^{bgt})$, where Y^{sgt} and Y^{bgt} are semantic and boundary annotations, respectively. Guided by the two losses, the informative correlations between semantics and boundaries can be fully excavated in a learnable way, and such correlations are then used to generate enhanced semantic and boundary features.

B. Outer-Boundary Context-Assignment Module

The context confusion nearby boundaries during the feature extraction could confuse the network to produce ambiguous features, degenerating the medical point cloud segmentation performance around boundaries. To tackle this problem, we propose OCM to assign contexts of each point nearby boundaries by leveraging graph techniques. As shown in

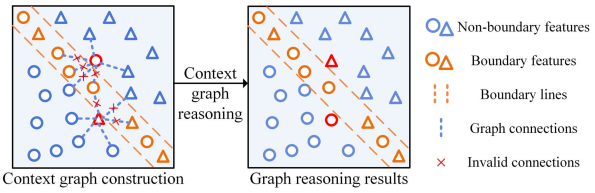

Fig. 4. Illustration of the context graph construction and reasoning procedure.

Fig. 2(b), we first construct a contextual graph to reformulate the context of each point as graph connections, and conditionally establish the connections with the guidance of the predicted boundaries. Based on the constructed contextual graph, graph reasoning is performed to rectify the original ambiguous features with newly clarified context.

1) Contextual Graph Construction: To assign contexts of points outside the boundaries, we treat each point as a graph node and represent the context of a certain node as graph adjacency with other nodes. In this way, the context assignment can be achieved by controlling graph connections.

Inspired by this observation, a contextual graph $G^c = (V^c, E^c)$ is constructed, where $V^c = \widehat{X}^s$ since we expect to refine these semantic features, and $E^c \in \{0,1\}^{|V^c|\times|V^c|}$ is the adjacency matrix that describes the connections of nodes. Then, considering boundaries are natural geometrical landmarks for different categories, we can separate contexts of points of different categories by cutting off connections across the predicted boundary points B, which is obtained from boundary logits $\widehat{Y}^b$:

$$B = \{P_i | argmax(\widehat{Y}_i^b) = 1, i \in \{1, \ldots, N\}\}, \tag{5}$$

and points in B are treated as inner-boundary points P_{in}, while the rest points are defined as outer-boundary points P_{out}:

$$P_{in} = B, P_{out} = P - B, \tag{6}$$

With obtained inner-boundary points P_{in}, the connections between i-th and j-th point $E^c_{i,j}$ is defined as below:

$$E^c_{i,j} = \begin{cases} 1, & (P_j \in \mathcal{N}_{k_2}(P_i)) \wedge (d(P_i, P_j) < \min\limits_{P_k \in P_{in}} d(P_i, P_k)), \\ 0, & otherwise, \end{cases}$$

where $\wedge$ refers to the logical and. As shown in Fig. 4, for the i-th point (red circles or triangles), the first term $P_j \in \mathcal{N}_{k_2}(P_i)$ provides basic contexts of k_2 nearest neighbors (blue dashes), and the second term constrains the distance between i-th point and j-th neighbor, ensuring it can not equal or exceed the minimum distance between i-th point and boundaries points in B (invalid connections are cut off by red cross). In this way, connections across the boundaries are prohibited, and points outside the boundary P_{out} with the context confusion problem, are assigned with appropriate contexts.

2) Contextual Graph Reasoning: With constructed contextual graph G^c, graph reasoning is performed to produce features with clarified contexts. To avoid the imbalance connections, self-loop and degree normalization are first applied on adjacency matrix E^c to get $A = D^{-\frac{1}{2}}(E + I)D^{-\frac{1}{2}}$, where I is the identity matrix, $D_{i,i} = \sum_j E_{i,j}$ and $D_{i,j\neq i} = 0$. And then graph reasoning is performed based on the normalized adjacency A:

$$\widetilde{X}^s = ReLU(AV^cW), \tag{7}$$

where $W \in \mathcal{R}^{D\times D}$ is trainable parameters and $\widetilde{X}^s$ is the reasoned semantic feature. In this way, original ambiguous features outside boundaries generated from confused contexts can be refined with newly assigned contexts. At last, the clarified feature $\widetilde{X}^s$ is passed to MLP layers, and resulting logits $\widetilde{Y}^s$ are supervised by cross-entropy loss $\mathcal{L}_c = -\frac{1}{N}\sum_{i=1}^{N} CE(\widetilde{Y}_i^s, Y_i^{sgt})$, where Y_i^{sgt} is the category annotation of point i. In summary, OCM adopts a novel contextual graph construction strategy, assigning different contexts to points of different categories, based on which graph reasoning is performed to produce less ambiguous representations.

C. Inner-Boundary Feature-Rectification Module

OCM can overcome the context confusion outside the predicted boundaries, however, features inside boundaries remain ambiguous due to the lack of geometrical landmarks, leading to inaccurate segmentation predictions. To overcome this bottleneck, we devise IFM to differentiate confused features inside boundaries with the delicately designed intra-sample contrast and inter-sample boundary-aware contrast.

1) Intra-Sample Boundary-Aware Contrast: For intra-sample contrast, our target is to make category-specific embeddings inside boundaries similar as embeddings outside boundaries of the same categories, and distinct from embeddings of the different categories. In doing so, the feature ambiguity can be effectively reduced as the confused features are optimized towards the correct direction in the feature space.

We first project features with two subsequent MLP layers into feature embeddings E and make the contrast in the embedding space following [34]. Then, given a category-specific anchor embedding $E_i \in \{E_u | P_u \in P_{in}\}$ belonging to inner-boundary points, we regard embeddings in $E_i^+ = \{E_u | Y_u^{gt} = Y_i^{gt}, P_u \in \mathcal{N}_{k_3}^{P_{out}}(P_i)\}$ that belongs to k_3 nearest non-boundary neighbors with the same category as positive embeddings, and embeddings in $E_i^- = \{E_u | Y_u^{gt} \neq Y_i^{gt}, P_u \in \mathcal{N}_{k_3}^{P_{out}}(P_i)\}$ that belongs to k_3 nearest non-boundary neighbors but with different categories as negative ones. The reason that we select k_3 nearest neighbors rather than all non-boundary points is that these geometrically neighboring embeddings are harder embeddings and hard positives/negatives have been demonstrated to be more beneficial to the contrastive learning [34] compared to the easy ones.

To restrain the anchor embeddings and positive/negative embeddings via measuring their similarities, the intra-sample boundary-aware contrastive loss $\mathcal{L}_{intra}$ is formulated as:

$$\mathcal{L}_{intra} = -\frac{1}{|B|}\sum_{E_i}\frac{1}{|K_i^+|}\sum_{E_i \in K_i^+} log\frac{h_\theta(E_i, E_j)}{\sum_{E_k \in K_i^-} h_\theta(E_i, E_k)}, \tag{8}$$

where $h_\theta(\cdot)$ denotes the affinity function and we adopt exponential cosine similarity as: $h_\theta(p, q) = exp(\frac{p\cdot q}{|p||q|}\cdot\frac{1}{\tau})$, where τ is the temperate factor. Optimized by this loss, the model can distinguish ambiguous features inside the boundaries.

2) Inter-Sample Boundary-Aware Contrast: As revealed by recent studies [34], [41], a large set of effective negatives is critical to the contrastive representation learning. While in the intra-sample contrast, the number of effective negative embeddings is limited by the size of the current point cloud sample. Therefore, we propose inter-sample boundary-aware contrast to incorporate embeddings from other samples.

To fulfill this need, we maintain an external memory bank $M \in \mathcal{R}^{L\times C\times D}$ during the training process, where L is the length of training samples, C is the category number, and D is the feature dimension. During the training, we first warm up the network for T epochs to generate reasonable embeddings and then initialize the memory bank M_T. Considering saving all point-level embeddings requires too much memory usage, we instead save averaged embedding for each category (the median embedding is a viable alternative, and it can be robust to noisy samples). Then at epochs $t = T + 1, \ldots$, for each sample, we can compute the inter-sample contrastive loss $\mathcal{L}_{inter}$ similarly as Eq. 8, but differently replacing the positive/negative embeddings with embeddings from the memory bank M saved in epoch $t-1$: $E_i^+ = \{E_u | Y_u^{gt} = Y_i^{gt}, E_u \in M_{t-1}\}$, $E_i^- = \{E_u | Y_u^{gt} = Y_i^{gt}, E_u \in M_{t-1}\}$. After computing $\mathcal{L}_{inter}$, the memory bank is updated in a momentum way:

$$M_{i,c}^t = \alpha M_{i,c}^{t-1} + (1-\alpha) E_{i,c} \tag{9}$$

where α is the balanced weight of previous and current embeddings, and $E_{i,c}$ is the averaged embedding of the c category in the i-th training sample. In this way, the memory bank can be updated to retain appropriate candidate embeddings for the effective contrast with anchor embeddings, encouraging confused anchor features to optimize towards the correct direction in the feature space.

D. Optimization

During the training procedure, we jointly optimize the loss in GBM, OCM, and IFM. First in GBM, cross-entropy loss for semantics $\mathcal{L}_s$ and boundaries $\mathcal{L}_b$ are used to guide the network in generating appropriate representations. Then in OCM, to optimize the graph reasoning layers, cross-entropy loss $\mathcal{L}_c$ is used to supervise the refined features outside the predicted boundaries. Finally, in IFM, intra-sample loss $\mathcal{L}_{intra}$ and inter-sample loss $\mathcal{L}_{inter}$ are proposed to distinguish ambiguous features inside the predicted boundaries. In summary, the overall objective function of the proposed GRAB-Net is:

$$\mathcal{L}_{overall} = \lambda_1(\mathcal{L}_s + \mathcal{L}_b) + \lambda_2 \mathcal{L}_c + \lambda_3(\mathcal{L}_{intra} + \mathcal{L}_{inter}), \tag{10}$$

where λ_1, λ_2, and λ_3 are trade-off factors to balance the contribution of each term.

IV. Experiment

A. Experimental Setup

1) Datasets: To evaluate our proposed method, we conduct experiments on two 3D segmentation datasets as follows:

IntrA is an open-access 3D intracranial aneurysm dataset [6]. It consists of 1,909 blood vessel point cloud segments extracted from 3D surface models of real patients, including 1,694 healthy vessel segments and 215 aneurysm segments. Following the segmentation setup in [6], 115 aneurysm segments with point-wise annotations are used to evaluate the binary segmentation performance. To obtain the boundary labels from category annotations, we follow previous boundary-aware works [23], [24], [25] to identify points whose eight nearest neighbors have different categories as boundaries.

3DTeethSeg is a publicly available 3D teeth segmentation dataset proposed in the MICCAI'22 3DTeethSeg challenge (https://3dteethseg.grand-challenge.org/). It contains 600 lower and 600 upper 3D surface models scanned by advanced intraoral scanners (IOS), where each dental surface model contains about 100,000 points. We randomly split the lower dataset into three subsets, 450 models for training, 30 models for validating, and 120 for testing. The tooth identification is based on the notation system (ISO-3950). The boundary annotations are obtained analogously to the IntrA dataset.

For both dataset, the projected node number M is set to 2^5. The number of nearest neighbors k_1, k_2, and k_3 are set to 16, 16, 32. The temperature τ and iteration weight α are set to 0.1 and 0.9. To balance three kinds of losses, λ_1, λ_2, and λ_3 are set as 1, 0.1, and 0.1, repectively.

2) Evaluation Metrics: The performance of our method is assessed by four metrics, including the Jaccard Index (also known as IoU), the Dice Similarity Coefficient (DSC), the class-wise mean accuracy (mAcc), and the boundary IoU (B-IoU). The former three metrics measure the similarity between the predicted results and ground truth annotations while B-IoU is used to evaluate the performance of the segmentation around boundaries [25].

3) Implementation Details: Our methods are implemented on a single RTX2080Ti GPU with the PyTorch library [42]. The baseline model is PointTransformer [18]. For the **IntrA** dataset, a fixed number of 512, 1024, and 2048 points are separately sampled from inputs as in [6]. Furthermore, on-the-fly data augmentation is used to enlarge the training data, including random scale of [0.8, 1.25], random rotation angle of $[-\pi, \pi]$ around Z axis, random translation of $[-0.1, 0.1]$, and random point jitter with 0 mean and 0.01 standard deviation. The Adam optimizer [43] is adopted with the initial learning rate as 0.001, and the cosine annealing learning rate scheduler is attached to gradually reduce the learning rate. We choose a batch size of 8 and the maximum epoch number of 400, and the warmup epoch T is set to 200. During the evaluation process, predictions on sampled points are re-projected into the original point cloud via three-nearest interpolation, and computing metrics on which is more reasonable than the sampled points since the latter will introduce randomness. For the **3DTeethSeg** dataset, we first uniformly sample 16,000 points from inputs as in [4]. Then, several online augmentation including random scale of [0.8, 1.25], random rotation angle of $[-\pi, \pi]$ around Z axis, random translation of $[-0.1, 0.1]$ are used. The Adam optimizer [43] and cosine annealing scheduler are also adopted with an initial learning rate of 0.002. The batch size is set to 1 due to the memory limitation. The maximum training epoch is 200 and the warmup epoch T

TABLE I
COMPARISON RESULTS ON THE INTRA DATASET UNDER 512, 1024, AND 2048 SAMPLING SCHEMES

Methods	Sampled points	IoU%			DSC%			mAcc%	B-IoU%
		Vessel	Aneurysm	Mean	Vessel	Aneurysm	Mean		
PointNet++ (17') [8]	512	93.44	77.51	85.47	96.61	87.33	92.17	91.59	27.59
	1024	94.90	80.57	87.74	97.38	89.23	93.47	92.97	29.53
	2048	94.95	81.40	88.17	97.41	89.75	93.71	93.28	30.86
PointConv (19') [11]	512	94.22	80.21	87.22	97.02	89.02	93.17	92.64	31.54
	1024	94.28	80.24	87.26	97.06	89.03	93.20	92.79	33.95
	2048	94.13	81.33	87.73	96.98	89.70	93.46	93.38	34.62
PCT (21') [16]	512	94.25	82.13	88.19	97.04	90.19	93.72	93.45	34.86
	1024	94.33	82.78	88.55	97.08	90.58	93.93	94.08	36.21
	2048	94.45	82.01	88.23	97.15	90.12	93.75	93.25	35.06
PAConv (21') [17]	512	94.56	81.87	88.22	97.20	90.03	93.74	93.05	34.87
	1024	94.43	81.05	87.74	97.14	89.53	93.47	92.77	38.33
	2048	94.18	81.69	87.94	97.00	89.92	93.58	93.00	37.79
PointTransformer (21') [18]	512	95.91	83.25	89.58	97.91	90.86	94.50	94.02	38.87
	1024	95.78	84.86	90.32	97.84	91.81	94.91	94.53	43.39
	2048	95.12	83.58	89.35	97.50	91.05	94.38	93.81	39.52
CBL (22') [25]	512	95.51	83.30	89.41	97.70	90.89	94.41	93.92	38.25
	1024	95.67	83.84	89.75	97.79	91.21	94.60	94.15	42.94
	2048	96.14	85.49	90.81	98.03	92.18	95.18	94.86	42.58
Ours	512	96.20	86.63	91.42	98.06	92.84	95.52	95.21	44.40
	1024	96.35	86.79	91.57	98.14	92.93	95.60	95.74	45.04
	2048	96.64	89.19	92.91	98.29	94.29	96.32	96.02	47.21

is set to 50. During the evaluation period, the interpolation strategy is also used.

B. Results on IntrA Dataset

We first validate the effectiveness of the proposed approach on the IntrA dataset, by comparing with state-of-the-art point-based 3D segmentation methods [8], [11], [16], [17], [18], [25], which are reproduced based on their official code repositories. As shown in Table I, our method shows superior results under all sampling scenarios. In particular, take 2048 sampling scheme for instance, the proposed method possesses the superior capability for 3D intracranial segmentation with increments of 4.74%, 5.18%, 4.68%, 4.97%, 3.56%, 2.10% in mean *IoU*, 2.61%, 2.86%, 2.57%, 2.74%, 1.94%, 1.14% in mean *DSC*, 2.74%, 2.64%, 3.67%, 3.02%, 2.21%, 1.15% in *mAcc*, and 16.35%, 12.59%, 12.15%, 9.42%, 7.69%, 4.63% in *B-IoU* compared with state-of-the-art methods [8], [11], [16], [17], [18], [25], respectively. It is worth noting that methods [8], [11], [16], [17], [18] focus on designing sophisticated local feature extractors and ignore the boundary segmentation, leading to inferior performance to ours. Although the latest boundary-aware method [25] utilizes contrastive learning to regularize features around boundaries, yielding better results compared to previous methods, it overlooks the global duality correspondences between semantics and boundaries as well as the context confusion problem. On the contrary, our proposed GRAB-Net can model duality correlations globally and clarify the confused contexts, contributing to substantial performance increases around boundary areas. To provide intuitive demonstrations, intracranial segmentation results of various shapes are illustrated in Fig. 5. It is evident that our proposed method outperforms the state-of-the-art 3D segmentation methods [8], [11], [16], [17], [18], [25] from the qualitative perspective, especially around boundaries areas between vessels and aneurysms.

C. Results on 3DTeethSeg Dataset

To verify the effectiveness of the proposed GRAB-Net on 3D teeth segmentation task, we evaluate the performance of our method and existing state-of-the-art methods [8], [11], [16], [17], [18], [25]. For a fair comparison, we implement network architectures based on their official code repositories and use the same data processing pipelines and training strategies as our proposed method. The quantitative results of various methods on 3DTeethSeg are listed in Table II. It is observed that the proposed GRAB-Net achieves supreme performance over the other point-based 3D segmentation methods [8], [11], [16], [17], [18], [25] with increments of 9.13%, 4.56%, 1.70%, 2.11%, 1.43%, 0.53% in mean *IoU*, 5.80%, 2.82%, 1.03%, 1.28%, 0.87%, 0.32% in mean *DSC*, 5.59%, 3.07%, 1.14%, 1.25%, 0.92%, 0.66% in *mAcc*, and 21.12%, 18.18%, 7.60%, 8.71%, 3.88%, 3.14% in *B-IoU*. Fig. 6 shows three qualitative comparison results between our methods and other models. It is obvious that the proposed method can segment more precisely around boundary areas in red circles, showing superior performance in visualization.

D. Ablation Analysis of Key Components

1) Effectiveness of GBM, OCM, and IFM: For an in-depth analysis of the proposed three modules, we conduct ablation studies under the 2048 sampling schemes on IntrA dataset. The comparison results listed in Table. III show that our proposed GRAB-Net (8^{th} *row*) achieves superior performance against the baseline model (1^{st} *row*) with increments of 2.68% in mean *IoU* and 6.26% in *B-IoU*, demonstrating the good capability of segmentation around boundary areas. We then quantify the contribution of GBM, OCM, and IFM ($2^{nd}-4^{th}$ *row*) by comparing them against the baseline model (1^{st} *row*). It is observed that adding GBM, OCM, and IFM show increments of 0.51%, 0.66%, 0.17% in mean *IoU* and

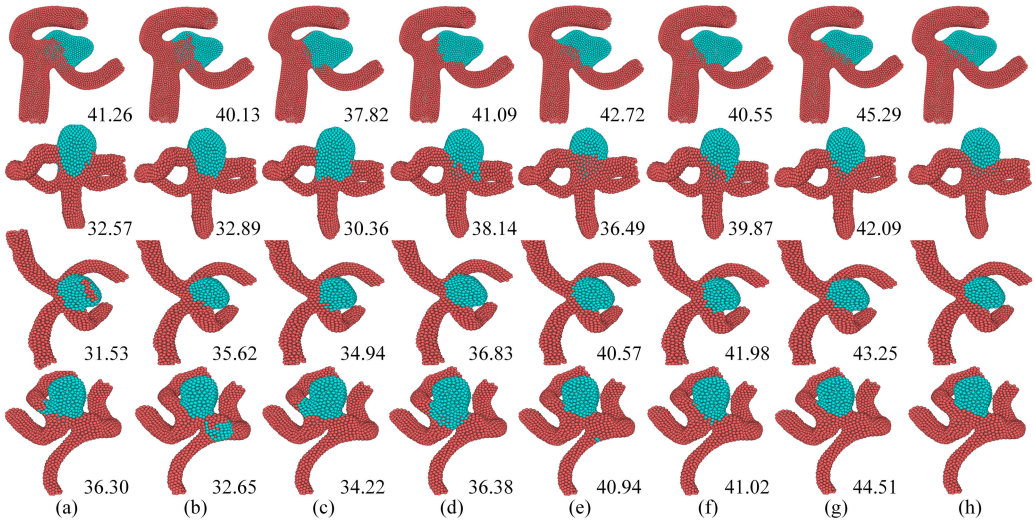

Fig. 5. Several typical examples of intracranial segmentation results. Each row represents the results of (a) PointNet++ [8], (b) PointConv [11], (c) PCT [16], (d) PAConv [17] (e) PointTransformer [18], (f) CBL [25], (g) ours, and (h) ground truth. Note that the scalar in the lower right corner represents *B-IoU* quantity that measures the segmentation performance around boundary areas.

TABLE II
COMPARISON RESULTS ON THE 3DTEETHSEG DATASET

Methods	IoU%					DSC%					mAcc%	B-IoU%
	Incisor	Canine	Premolar	Molar	Mean	Incisor	Canine	Premolar	Molar	Mean		
PointNet++ (17') [8]	66.96	74.74	75.26	74.70	72.91	80.21	85.54	85.88	85.52	84.33	84.70	21.68
PointConv (19') [11]	72.83	79.53	81.04	76.50	77.48	84.28	88.60	89.53	86.69	87.31	87.22	24.62
PCT (21') [16]	79.51	82.39	83.06	76.38	80.34	88.59	90.34	90.75	86.61	89.10	89.15	35.20
PAConv (21') [17]	78.25	80.66	83.53	77.26	79.93	87.79	89.29	91.03	87.17	88.85	89.04	34.09
PointTransformer (21') [18]	83.27	86.14	87.54	83.28	80.61	90.87	92.55	93.36	90.88	89.26	89.37	38.92
CBL (22') [25]	86.03	87.55	89.24	86.64	81.51	92.49	93.36	94.31	92.84	89.81	89.63	39.66
Ours	86.56	87.99	89.35	87.74	82.04	92.80	93.61	94.38	93.47	90.13	90.29	42.80

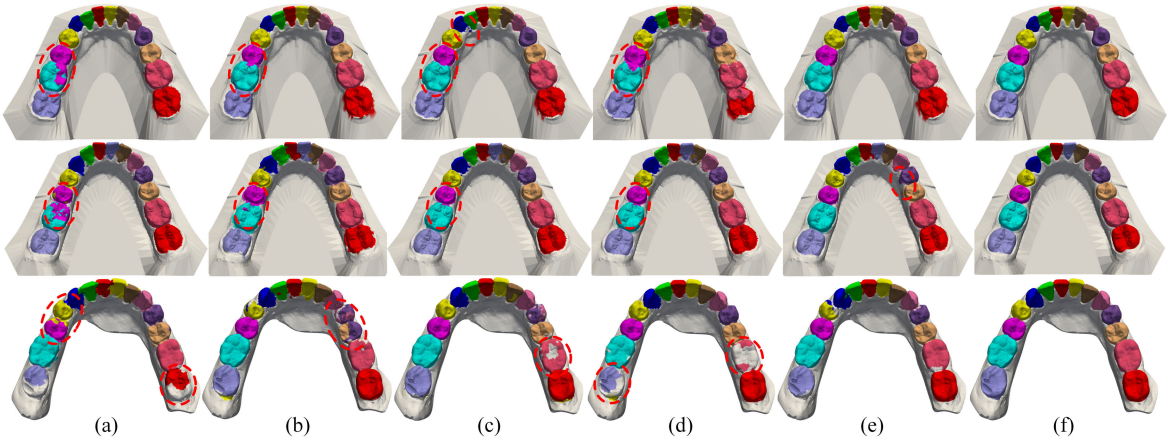

Fig. 6. Two typical examples of teeth segmentation results. Each row represents the results of (a) PCT [16], (b) PAConv [17] (c) PointTransformer [18], (d) CBL [25], (e) Ours, and (f) ground truth.

3.07%, 2.82%, 2.32% in *B-IoU*. These results demonstrate the advantage of the proposed three modules, which model the semantics-boundary correlations globally, rectify the confused contexts outside boundaries, and distinguish ambiguous features inside boundaries. In addition, combining any two of the proposed modules and all modules ($5^{th} - 8^{th}$ *row*) performs better than using only a single module ($2^{nd} - 4^{th}$ *row*), which further verify the compatibility of the proposed modules.

2) Analysis on Projected Node Numbers in GBM: In GBM, point-level features are projected into M nodes in the graph domain, which encodes local coherent features to avoid redundancy. To investigate the influence of the projected node numbers, we conduct ablation experiments on IntrA dataset under different node settings, where M equals 2^5, 2^6, 2^7, 2^8, and 2^9, respectively. As shown in Fig. 7, it is obvious that the performance is poor if the node number is too small, e.g.,

TABLE III
ABLATION RESULTS OF GBM, OCM, AND IFM. THE BEST AND SECOND BEST RESULTS ARE **HIGHLIGHTED** AND UNDERLINED

Baseline	GBM	OCM	IFM	IoU%			B-IoU%
				Vessel	Aneurysm	Mean	
✓				95.39	85.06	90.23	40.95
✓	✓			95.32	86.15	90.74	44.02
✓		✓		95.68	86.09	90.89	43.78
✓			✓	95.18	85.62	90.40	43.27
✓	✓	✓		96.38	88.75	92.57	46.08
✓	✓		✓	96.23	87.45	91.84	46.82
✓		✓	✓	95.94	87.83	91.88	46.33
✓	✓	✓	✓	**96.64**	**89.19**	**92.91**	**47.21**

TABLE IV
EXPERIMENTAL RESULTS OF GRAPH REASONING LAYERS. BEST AND SECOND BEST RESULTS ARE **HIGHLIGHTED** AND UNDERLINED

Number	IoU%			DSC%			B-IoU%
of layers	V.	A.	Mean	V.	A.	Mean	
1	96.47	87.75	92.11	98.20	93.48	95.89	46.23
2	**96.64**	**89.19**	**92.91**	**98.29**	**94.29**	**96.32**	47.21
3	96.55	88.15	92.35	98.24	93.70	96.02	45.89
4	96.57	88.41	92.49	98.26	93.85	96.10	**47.81**
5	96.41	87.79	92.10	98.17	93.50	95.89	47.07

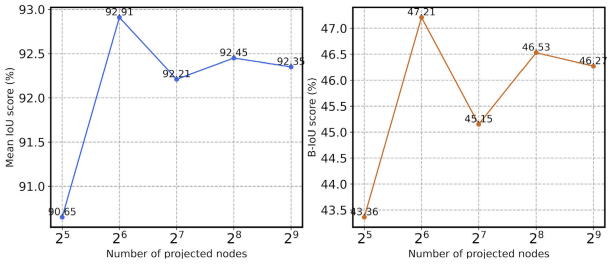

Fig. 7. Mean IoU score and B-IoU score under different projected node number settings.

$M = 2^5$, mainly caused by the information lost during the projection procedure. With the node number getting larger, the performance tends to increase but would saturate and even decrease when the numbers are too large, e.g, when M is larger than 2^7. Too many node numbers may introduce information redundancy, obscuring the module to explore the hidden correlations. When M is set to 2^6, the segmentation performance achieves the best trade-off between information reservation and clarification.

3) Analysis on the Number of Graph Reasoning Layers in OCM: The graph reasoning in OCM can propagate desired messages among adjacent nodes, refining ambiguous features with specified contexts, in which the number of graph reasoning layers determines the exchange extent. To investigate the influence of the number of layers l, we experiment with $l = 1, 2, 3, 4, 5$, respectively. Comparison results in Table IV show that $l = 2$ performs favorably against $l = 1, 3, 4, 5$ settings with improvements 0.80%, 0.56%, 0.42%, and 0.81% in mean IoU, respectively. This result reveals that less reasoning layers, e.g, $l = 1$ can hardly propagate desired messages sufficiently, while too many reasoning layers, e.g, $l = 3, 4, 5$ would risk the model of over-fitting.

TABLE V
RESULTS OF THE PROPOSED METHOD WITH OR WITHOUT UNCERTAINTIES IN THE GROUND TRUTH

K	IoU%			DSC%			B-IoU%
	V.	A.	Mean	V.	A.	Mean	
None	96.64	89.19	92.91	98.29	94.29	96.32	47.21
8	96.35	89.14	92.75	98.14	94.26	96.20	46.84
16	95.68	89.01	92.35	97.79	94.19	95.99	46.77
32	95.87	89.35	92.61	97.89	94.38	96.13	46.87

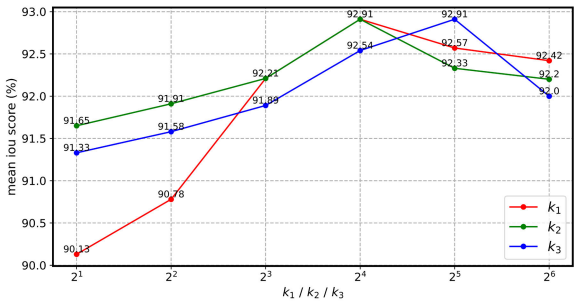

Fig. 8. Mean IoU score of different k settings under the 2048 sampling scheme on IntrA dataset.

4) Analysis on the Number of Nearest Neighbors k_1, k_2, and k_3: k_1, k_2, and k_3 is the number of nearest neighbors used in GBM, OCM, and IFM, and their values affect the process of coherent node projection, contextual graph construction, and intra-sample boundary-aware contrast, respectively. To explore the influence of the value of k on the model performance, we conduct more experiments where k_1, k_2, and k_3 are set to 2, 4, 8, 16, 32, 64. Results in Fig. 8 reveal the performance curve of k_1, k_2, and k_3 have the same tendency. Too small k value would obtain a low score, while the performance of too large k value would saturate, and the peek score is achieved when the value of k is medium. Though the tendencies of k_1, k_2, and k_3 are similar, the reason behind the phenomenon is different. For k_1, small k_1 would lead to non-robust node features, while large k_1 can cause feature overlap between different nodes. For k_2, small k_2 can cause sparse graph connections and hinder the long-range propagation, while large k_2 can incur redundant graph connections. For k_3, small k_3 can hardly provide enough hard embeddings for contrast, and large k_3 would introduce extra easy embeddings, which would confuse the network training.

5) Analysis on the Influence of Uncertainties in the Ground Truth: In real applications, the annotations of the collected dataset are usually performed by multiple annotators, and the resulting labels may have uncertainties. To investigate the influence of uncertainties in the ground truth to the proposed method, we design more experiments on IntrA dataset under the 2048 sampling scheme. Considering the ground truth of IntrA dataset is either 0 for vessel or 1 for aneurysm, we generate uncertainties by sliding a local window across point clouds. More specifically, for each point, we find its K nearest neighbors and average the ground truth value of these neighbors. In this way, the generated ground truth of points nearby boundaries is close to 0.5, while points far from the boundary are close to either 0 or 1. The resulting

TABLE VI
RESULTS OF VARIOUS CONTRASTIVE STRATEGIES. BEST AND SECOND BEST RESULTS ARE **HIGHLIGHTED** AND UNDERLINED

Methods	IoU%					B-IoU%
	Incisor	Canine	Premolar	Molar	Mean	
baseline	84.04	86.43	87.90	83.56	80.79	39.26
w/ intra	**85.93**	86.74	87.25	**85.22**	81.49	40.52
w/ inter	84.50	87.34	87.51	84.73	81.53	39.98
w/ both	85.62	**87.53**	**88.30**	**85.22**	**82.09**	**41.30**

soft labels can be treated as ground truth with uncertainties, and experimental results are shown in Table. V. Compared with using the original ground truth with soft labels, i.e., 'None' in Table. V, using labels with uncertainties where K is set to 8, 16, and 32 only cause slight performance drop of 0.16%, 0.56%, and 0.30% mIoU, respectively, revealing that our method is robust to the uncertainties existed in the ground truth.

6) Effectiveness of the Intra- and Inter-Sample Contrast in IFM: In IFM, we expect to utilize contrastive learning to distinguish ambiguous features inside the boundary areas. To demonstrate the effectiveness of the intra-sample and inter-sample contrast, we conduct four experiments on 3DTeethSeg dataset, which are 'baseline', 'w/ intra' and 'w/ inter' that only equips the baseline with two types of contrast, respectively, and 'w/ both' that adopts both contrasts. As shown in Table VI, adding intra-sample or inter-sample contrast could bring performance improvement of 0.70%, 0.74% in mean IoU, and adopting both achieves the largest improvement of 1.50% in mean IoU. These results reveal that the two types of contrast are both necessary for effective representation learning and they can be well incorporated together into the baseline model. Besides, to qualitatively verify the effectiveness of two types of contrasts, the nonlinear dimensional reduction algorithm, t-SNE [44], is performed to scatter the feature space of 'baseline', 'w/ intra', 'w/ inter', and 'w/ both', as illustrated in Fig 9. In 'baseline' (Fig. 9 (a)), it is obvious that features of some categories are mixed with high overlap (regions in red circles), which are ambiguous for the network to produce correct predictions. On the contrary, the feature distribution of 'w/ intra' (Fig. 9 (b)) and 'w/ inter' (Fig. 9 (c)) show less overlap compared to the 'baseline', demonstrating the effectiveness of the proposed two types of contrast. Furthermore, the feature space of adding both contrasts (Fig. 9 (d)) is well separated with nearly no overlap, which reflects the effectiveness of the two types of contrast from the qualitative perspective.

E. Limitation

Although the proposed method has achieved remarkable results on medical point cloud segmentation tasks, outperforming previous state-of-the-art methods, it still presents limitations to be considered. One typical example lies in the aneurysm segmentation task, where inputs with two more aneurysms can be hardly segmented well, as shown in Fig.10. One possible reason is that these cases are rare, since most inputs in IntrA dataset contain only one aneurysm each, and thus the case that one segment with multiple aneurysms is seldom seen by the network during the training process. In the future, we would like to explore more effective methods to handle datasets with imbalanced data type distribution.

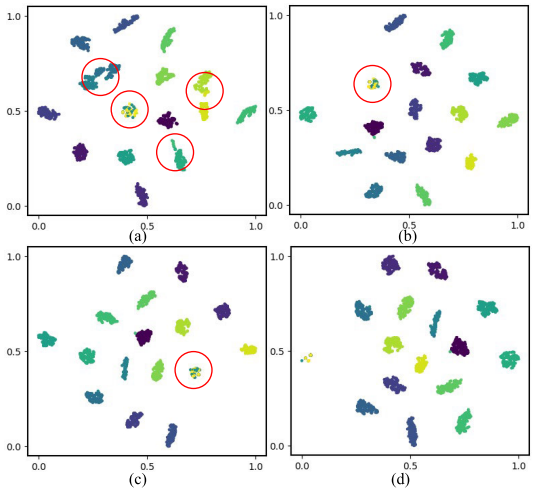

Fig. 9. Illustration of the normalized feature space of (a) baseline, (b) w/ intra-sample contrast, (c) w/ inter-sample contrast, and (d) w/ both. Note that different colors represent different categories, and red circles highlight regions that are not separated well.

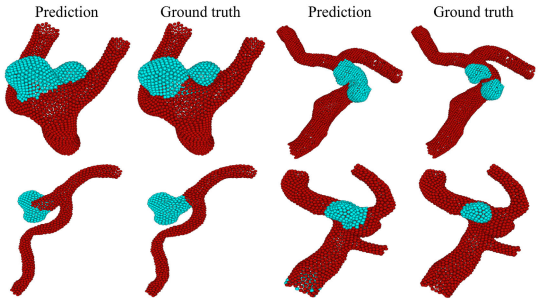

Fig. 10. Illustration of failure cases on IntrA dataset.

V. CONCLUSION

Point cloud segmentation is crucial in many medical applications, and it is challenging to segment well around the boundaries due to the insufficient duality constraints and context confusion. In this paper, we propose a GRAB-Net framework with GBM, OCM, and IFM to tackle the aforementioned issues. GBM proposes to build global relations between semantics and boundaries in the graph domain, providing sufficient duality information for producing features of high quality. The proposed OCM leverages the contextual graph to assign appropriate contexts outside boundaries. Furthermore, IFM is designed to reduce the feature ambiguity inside boundaries with intra- and inter-sample contrast. Extensive experiments on two benchmark datasets, IntrA and 3DTeethSeg, verify the superiority of the proposed method. In addition, the comprehensive experiments demonstrate the effectiveness of each proposed component.

REFERENCES

[1] M. Y. Hajeer, D. Millett, A. Ayoub, and J. Siebert, "Applications of 3D imaging in orthodontics: Part I," *J. Orthodontics*, vol. 31, no. 1, pp. 62–70, 2004.

[2] F. G. Zanjani et al., "Mask-MCNet: Instance segmentation in 3D point cloud of intra-oral scans," in *Proc. MICCAI*. Cham, Switzerland: Springer, 2019, pp. 128–136.

[3] C. Lian et al., "Deep multi-scale mesh feature learning for automated labeling of raw dental surfaces from 3D intraoral scanners," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2440–2450, Jul. 2020.

[4] Z. Cui et al., "TSegNet: An efficient and accurate tooth segmentation network on 3D dental model," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101949.

[5] A. Alaraj et al., "Virtual reality cerebral aneurysm clipping simulation with real-time haptic feedback," *Operative Neurosurg.*, vol. 11, no. 1, pp. 52–58, 2015.

[6] X. Yang, D. Xia, T. Kin, and T. Igarashi, "IntrA: 3D intracranial aneurysm dataset for deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2656–2666.

[7] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.

[8] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 30, 2017, pp. 1–14.

[9] J. Li, B. M. Chen, and G. H. Lee, "SO-Net: Self-organizing network for point cloud analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9397–9406.

[10] Y. Liu, B. Fan, S. Xiang, and C. Pan, "Relation-shape convolutional neural network for point cloud analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8895–8904.

[11] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep convolutional networks on 3D point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9621–9630.

[12] H. Zhao, L. Jiang, C.-W. Fu, and J. Jia, "PointWeb: Enhancing local neighborhood features for point cloud processing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5565–5573.

[13] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6411–6420.

[14] Q. Hu et al., "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11108–11117.

[15] C. Xu et al., "SqueezeSegV3: Spatially-adaptive convolution for efficient point-cloud segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2020, pp. 1–19.

[16] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R. R. Martin, and S.-M. Hu, "PCT: Point cloud transformer," *Comput. Vis. Media*, vol. 7, no. 2, pp. 187–199, 2021.

[17] M. Xu, R. Ding, H. Zhao, and X. Qi, "PAConv: Position adaptive convolution with dynamic kernel assembling on point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3173–3182.

[18] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16259–16268.

[19] R. Cheng, R. Razani, E. Taghavi, E. Li, and B. Liu, "$(AF)^2$-S3Net: Attentive feature fusion with adaptive feature selection for sparse semantic segmentation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12547–12556.

[20] S. Qiu, S. Anwar, and N. Barnes, "Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1757–1767.

[21] S. Fan, Q. Dong, F. Zhu, Y. Lv, P. Ye, and F.-Y. Wang, "SCF-Net: Learning spatial contextual features for large-scale point cloud segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14504–14513.

[22] T. Lu, L. Wang, and G. Wu, "CGA-Net: Category guided aggregation for point cloud semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11693–11702.

[23] Z. Hu, M. Zhen, X. Bai, H. Fu, and C.-L. Tai, "JSENet: Joint semantic segmentation and edge detection network for 3D point clouds," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 222–239.

[24] J. Gong et al., "Boundary-aware geometric encoding for semantic segmentation of point clouds," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 2, pp. 1424–1432.

[25] L. Tang, Y. Zhan, Z. Chen, B. Yu, and D. Tao, "Contrastive boundary learning for point cloud segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8489–8499.

[26] W. Li, X. Liu, and Y. Yuan, "SIGMA: Semantic-complete graph matching for domain adaptive object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5291–5300.

[27] Y. Wu et al., "Bidirectional graph reasoning network for panoptic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9080–9089.

[28] L. Zhang et al., "TSGCNet: Discriminative geometric feature learning with two-stream graph convolutional network for 3D dental model segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6699–6708.

[29] T.-H. Wu et al., "Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3D intraoral scans," *IEEE Trans. Med. Imag.*, vol. 41, no. 11, pp. 3158–3166, Nov. 2022.

[30] Y. Xu, T. Fan, M. Xu, L. Zeng, and Y. Qiao, "SpiderCNN: Deep learning on point sets with parameterized convolutional filters," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 87–102.

[31] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on χ-transformed points," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 31, 2018, pp. 1–11.

[32] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, Nov. 2019.

[33] X. Lai et al., "Stratified transformer for 3D point cloud segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8500–8509.

[34] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[35] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2020, pp. 213–229.

[36] B. Cheng, A. Schwing, and A. Kirillov, "Per-pixel classification is not all you need for semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 17864–17875.

[37] H. Ding, X. Jiang, A. Q. Liu, N. M. Thalmann, and G. Wang, "Boundary-aware feature propagation for scene segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6819–6829.

[38] D. Marin et al., "Efficient segmentation: Learning downsampling near semantic boundaries," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2131–2141.

[39] H. J. Lee, J. U. Kim, S. Lee, H. G. Kim, and Y. M. Ro, "Structure boundary preserving segmentation for medical image with ambiguous boundary," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4817–4826.

[40] B. Cheng, R. Girshick, P. Dollar, A. C. Berg, and A. Kirillov, "Boundary IoU: Improving object-centric image segmentation evaluation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15334–15342.

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

[42] A. Paszke et al., "Automatic differentiation in PyTorch," in *Proc. NIPS Workshop*, 2017.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[44] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–10, 2008.