Medical Image Analysis 107 (2026) 103825

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Nested resolution mesh-graph CNN for automated extraction of liver surface anatomical landmarks

Xukun Zhang [a], Jinghui Feng [a], Peng Liu [b], Minghao Han [a], Yanlan Kang [a], Jingyi Zhu [a], Le Wang [a], Xiaoying Wang [c,*], Sharib Ali [d,*], Lihua Zhang [a,e,*]

[a] Academy for Engineering and Technology, Fudan University, 200082, Shanghai, China
[b] Department of Translational Surgical Oncology, National Center for Tumor Diseases (NCT/UCC Dresden), 01307, Fetscherstraße 74, Dresden, Germany
[c] Liver Cancer Institute, Zhongshan Hospital, Fudan University, 200082, Shanghai, China
[d] School of Computing, University of Leeds, LS2 9JT, Leeds, United Kingdom
[e] Meta-medical Center, Institute of Intelligent Medicine, Fudan University, 200032, Shanghai, China

ARTICLE INFO

Dataset link: Dataset Link

Keywords:
Liver surface
Mesh segmentation
Anatomical landmarks
Attention mechanism
3D–2D image fusion

ABSTRACT

The anatomical landmarks on the liver (mesh) surface, including the falciform ligament and liver ridge, are composed of triangular meshes of varying shapes, sizes, and positions, making them highly complex. Extracting and segmenting these landmarks is critical for augmented reality-based intraoperative navigation and monitoring. The key to this task lies in comprehensively understanding the overall geometric shape and local topological information of the liver mesh. However, due to the liver's variations in shape and appearance, coupled with limited data, deep learning methods often struggle with automatic liver landmark segmentation. To address this, we propose a two-stage automatic framework combining mesh-CNN and graph-CNN. In the first stage, dynamic graph convolution (DGCNN) is employed on low-resolution meshes to achieve rapid global understanding, generating initial landmark proposals at two levels, "dilation" and "erosion", and mapping them onto the original high-resolution surface. Subsequently, a refinement network based on mesh convolution fuses these landmark proposals from edge features along the local topology of the high-resolution mesh surface, producing refined segmentation results. Additionally, we incorporate an anatomy-aware Dice loss to address resolution imbalance and better handle sparse anatomical regions. Extensive experiments on two liver datasets, both in-distribution and out-of-distribution, demonstrate that our method accurately processes liver meshes of different resolutions, outperforming state-of-the-art methods. The reconstructed liver mesh dataset and the source code are available at https://github.com/xukun-zhang/MeshGraphCNN.

1. Introduction

Augmented Reality (AR)-based navigational guidance for laparoscopic hepatectomy introduces a groundbreaking visualization approach, central to which is the alignment of a preoperative 3D liver model (mesh) with intraoperative 2D laparoscopic images (Ali et al., 2025; Lopez, 2022; Pfeiffer et al., 2018). This technique enables surgeons to accurately identify internal structures by overlaying the liver model onto the laparoscopic view (Fig. 1A). The success of this 3D–2D fusion heavily relies on using anatomical landmarks as registration constraints (Robu et al., 2018; Mhiri et al., 2024), such as the liver's ridge and the falciform ligament. Currently, these anatomical landmarks are manually annotated on liver meshes during the preoperative phase (Koo et al., 2017, 2022). Although technically feasible, this process is time-consuming – often requiring several hours per case (Plantefeve et al., 2016) – and demands substantial anatomical expertise. Furthermore, consistent identification across diverse anatomical shapes remains challenging, especially when subtle landmarks such as the falciform ligament are involved. These limitations constrain the scalability and standardization of AR-assisted navigation. Therefore, automating the extraction (i.e., segmentation) of these anatomical features is not only desirable, but also essential for streamlining surgical workflows and promoting wider clinical adoption of AR technology.

However, automatically segmenting key anatomical landmarks from the liver mesh is challenging, as it requires a comprehensive understanding of both the global geometric structure (spatial relationships) and the local topological information (mesh unit shape). As illustrated in Fig. 1(A), two such landmarks are commonly used for visual navigation: the falciform ligament and the liver ridge. The falciform ligament

* Corresponding authors.
E-mail addresses: xiaoyingwang@fudan.edu.cn (X. Wang), S.S.Ali@leeds.ac.uk (S. Ali), lihuazhang@fudan.edu.cn (L. Zhang).

https://doi.org/10.1016/j.media.2025.103825
Received 18 September 2024; Received in revised form 20 September 2025; Accepted 26 September 2025
Available online 4 October 2025

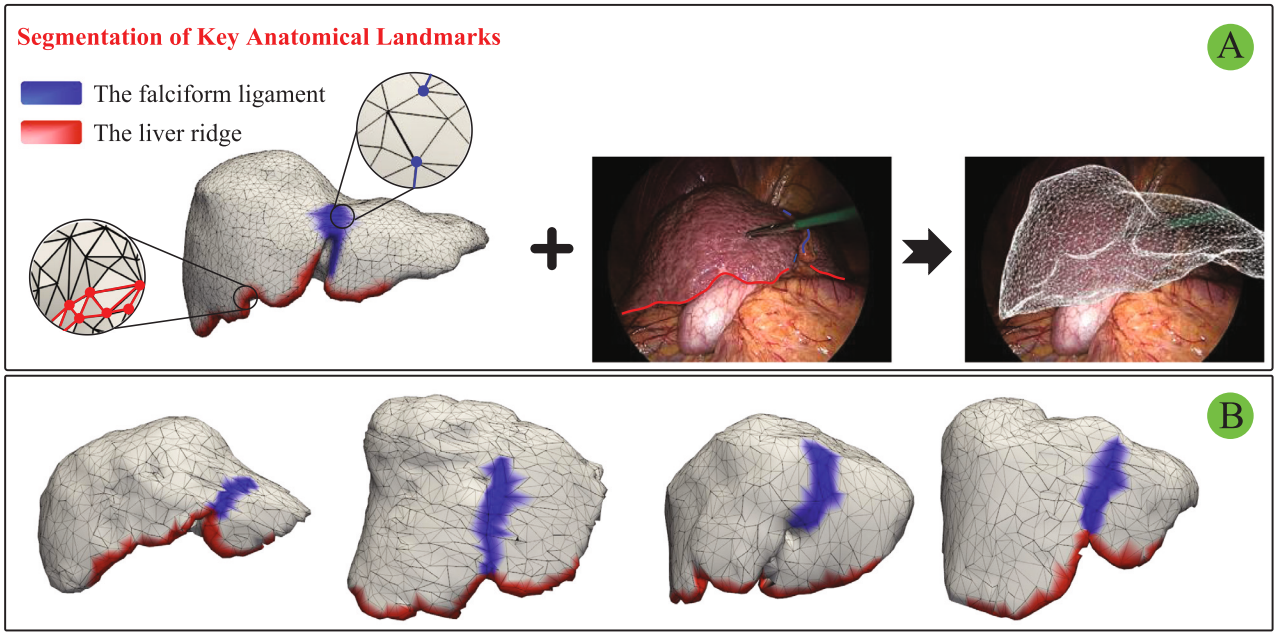

Fig. 1. (A) Accurate segmentation of anatomical landmarks on the liver surface is crucial for developing intraoperative AR navigation systems. The magnified inset highlights landmark vertices and their shared edges, which define the segmentation regions studied in this work. (B) The liver mesh surface lacks texture and exhibits significant variations in shape and appearance, making landmark segmentation particularly challenging, especially for the falciform ligament.

is a thin fibrous structure that connects the anterior surface of the liver to the abdominal wall, typically located near the midline-left region of the liver surface. It lies in a relatively flat, low-curvature area composed of large, evenly distributed triangular faces. However, the local surface cues that define this structure – such as directional edge alignment or subtle curvature transitions – are often too subtle to be reliably recognized by human annotators. This motivates the development of automated methods capable of learning these anatomically meaningful yet visually ambiguous patterns. In contrast, the liver ridge, positioned on the anterior base of the liver and extending laterally across both lobes, appears in a high-curvature region characterized by small, densely clustered triangles and sharper local topology.

The Preoperative to Intraoperative Laparoscopy Fusion (P2ILF) challenge (Ali et al., 2025) marked a pioneering effort in addressing this segmentation task within the scope of preoperative to intraoperative liver image fusion. This challenge provided a dataset of 11 liver meshes, sparking various innovative solutions from the global research community. The methods proposed by participants included point-based approaches like Pointnet++ (Qi et al., 2017b), graph-based methods (Kipf and Welling, 2016), and mesh-specific strategies such as MeshCNN (Hanocka et al., 2019) (the winning method). Pointnet++ (Qi et al., 2017b) and Graph Convolutional Network (GCN) methods (Kipf and Welling, 2016) excel in global understanding by aggregating point cloud features but often overlook the rich details of the mesh surface. Conversely, MeshCNN-based methods (Hanocka et al., 2019) perform feature computation along the mesh surface topology through mesh convolutions (MeshConv), offering robust local topological learning. However, when processing meshes with high and inconsistent resolutions (Fig. 1(B)), MeshCNN (Hanocka et al., 2019) suffers from issues in computational efficiency and consistent global shape representation due to its sequential pooling and edge-collapse strategy. These problems become particularly challenging when learning anatomical structures across samples with diverse liver shapes and mesh complexities (see Section 3.2.1 for details).

In this paper, to address the aforementioned challenges, we propose a novel geometric deep learning framework that combines the strengths of graph-based and mesh-based methods while mitigating their respective weaknesses. Specifically, our Nested Resolution Mesh-Graph CNN framework is designed to accurately extract key liver anatomical landmarks, such as the falciform ligament and liver ridge, which are represented as combinations of vertices and edges on the 3D liver mesh. The task is thus reformulated as a vertex or edge segmentation problem on the mesh. Our approach operates on two mesh resolution levels: compressed low-resolution meshes and the original high-resolution meshes. For low-resolution meshes, we utilize Dynamic Graph Convolutional Networks (DGCNN) (Wang et al., 2019b) to quickly learn the liver's overall shape and appearance, generating initial anatomical landmark segmentations. These segmentations are mapped onto the high-resolution mesh surface through different propagation methods, obtaining landmark proposals at "dilation" and "erosion" levels. Subsequently, we design an anatomical refining network built on MeshConv, which integrates these landmark proposals with the detailed topology of the original high-resolution mesh. The network employs a fine-grained aggregation-based attention mechanism that effectively balances and combines the potential correct priors from the different proposals, producing refined segmentation results sensitive to the complex topological structures of the liver surface. Additionally, we introduce an anatomy-aware Dice loss that addresses the mesh surface's unevenness and captures the relative relationships of ligaments and ridges, further enhancing the segmentation performance.

In summary, the main contributions of this work are:

- We propose a nested-resolution framework that integrates dynamic graph convolution (DGCNN) with MeshConv-based refinement to enable accurate liver landmark segmentation across meshes with varying resolutions.
- We introduce an anatomical refinement strategy that adaptively integrates coarse landmark proposals through attention fusion and auxiliary supervision, and is further enhanced by an anatomy-aware Dice loss that mitigates resolution imbalance and improves segmentation quality.
- We provide a manually annotated dataset of 200 liver meshes from public CT datasets and validate the method on the clinically relevant P2ILF challenge cases, demonstrating robust generalizability across imaging modalities and mesh reconstruction pipelines.

The organization of our paper is as follows. In Section 2, we comprehensively review related applications and work in this field. In Section 3, we describe the data and proposed framework in detail. In Section 4, we present our experimental settings and results. Finally, we discuss and conclude this work in Section 5.

2. Related work

Our research focuses on the application of augmented reality-assisted laparoscopic liver resection (AR-LLR), emphasizing 3D mesh segmentation techniques, particularly the automated extraction of critical anatomical landmarks on the liver surface.

2.1. 3D-2D registration-based LLR navigation

Augmented reality technology (AR) addresses the challenge of internal structure invisibility in LLR by integrating 3D liver models derived from CT or MR imaging with laparoscopic images. Existing methods register preoperative 3D data with intraoperative 2D images or 3D data (Modrzejewski et al., 2019; Adagolodjo et al., 2017; Labrunie et al., 2022, 2023; Oya et al., 2024). Accurate extraction and segmentation of common landmarks between laparoscopic images and preoperative 3D models are essential for achieving precise 3D-to-2D registration. Studies (Koo et al., 2017; Espinel et al., 2021) highlight the liver ridge, falciform ligament, and top boundary contours as potential 3D–2D registration landmarks. For instance, by combining biomechanical models with these landmarks, intraoperative dynamic tracking of the liver is achieved (Koo et al., 2017). Other works (Espinel et al., 2021) improve navigation accuracy by combining anatomical markers in intraoperative images with preoperative 3D models. Similar methods (Koo et al., 2022; Labrunie et al., 2022; Pei et al., 2024) using CASENet, UNet and SAM (Kirillov et al., 2023) for 2D detection still require manual annotation of 3D landmarks. These studies underscore the importance of automated anatomical landmark extraction algorithms to reduce user interaction and enhance the clinical utility of AR-LLR navigation.

2.2. Mesh segmentation

The segmentation of liver surface landmarks falls under the broader category of mesh segmentation in 3D shape processing. Current mesh segmentation methods can be broadly categorized as follows:

- 2D Projections: Early approaches (Chen et al., 2017; Dai and Nießner, 2018; Le et al., 2017; Pang and Neumann, 2016) typically project 3D geometric data onto 2D images from predefined viewpoints and process the projections using 2D CNNs. However, 2D projections inevitably lose spatial information, limiting their performance in fine-grained segmentation tasks.
- Volumetric Methods: Voxel-based methods (Graham et al., 2018; Le and Duan, 2018; Riegler et al., 2017; Wang and Lu, 2019) discretize 3D space into regular volumetric grids and segment using 3D CNNs. These methods suffer from high memory and computational costs, with the resolution scaling exponentially with each dimension (Liu et al., 2019).
- Point Cloud Methods: Point-based methods directly process 3D geometric data using deep learning architectures. Pointnet (Qi et al., 2017a) achieves permutation invariance of point clouds by using symmetric max-pooling operations to aggregate features. Pointnet++ (Qi et al., 2017b) enhances local spatial relationship learning by hierarchically applying Pointnet. Subsequent works integrate attention modules (Wu et al., 2019), geometric sharing modules (Xu et al., 2020), and edge branches (Jiang et al., 2019) to extend Pointnet++ for finer local detail learning.
- Graph Convolutional Networks (GCNs): Given the inherent spatial relationships in 3D meshes, GCN-based methods have been proposed for 3D mesh recognition and segmentation tasks (Liang et al., 2020; Wang et al., 2019a). These methods (Wang et al., 2018; Xie et al., 2020) represent 3D mesh data as graph structures and use spectral or spatial graph convolutions to aggregate local information for each node. For instance, Wang et al. (2019b) combines multi-scale strategies and introduce dynamic graph convolutional networks (DGCNN) to handle dynamic graph structures, significantly expanding the application range of graph convolution methods to more complex and variable real-world problems.
- MeshCNN: Traditional methods struggle with the non-uniform and irregular topography of 3D mesh surfaces. Hanocka et al. (2019) proposed MeshCNN, a neural network architecture specifically designed for meshes. MeshCNN operates directly on irregular triangular meshes, performing tailored convolution and pooling operations. In this framework, mesh edges are analogous to pixels in images, forming a fixed-size convolution neighborhood containing four edges. Additionally, MeshCNN's pooling operation compresses edges in a task-driven manner, achieving downsampling similar to size reduction in CNNs. Recent studies (Schneider et al., 2021; Chen et al., 2023) have adopted the MeshCNN architecture for medical applications, such as vascular mesh and dental surface segmentation tasks, demonstrating performance comparable to state-of-the-art models.

2.3. Landmark extraction on mesh surfaces

Unlike typical mesh segmentation tasks, landmark extraction (or segmentation) on mesh data is highly class-imbalanced. For example, anatomical landmark segmentation on the liver surface involves irregular curves like ligaments and ridges that occupy a small mesh area. The P2ILF challenge at MICCAI'2022 (Ali et al., 2025) marked a milestone in liver mesh segmentation, collecting and annotating data from 11 patients and attracting six international teams. Four teams employed Pointnet++, GCN, and MeshCNN technologies, with the MeshCNN-based team emerging as the winner. Similar to this task, Chen et al. (2023) designed an improved MeshCNN with residual learning and multi-scale attention mechanisms for gingiva line detection/segmentation on 3D dental surfaces. Zhao et al. (2021) developed a dual-stream graph convolutional network tailored for the segmentation of 3D tooth meshes derived from oral scans, demonstrating state-of-the-art performance in the field. However, unlike the rigid dental landmark detection task (Wu et al., 2019), liver data exhibits significant deformation, making direct application of these methods ineffective. The P2ILF challenge results further confirmed this, showing that GCN-based liver mesh segmentation was less effective than Pointnet++ and MeshCNN. Additionally, in contrast to the well-studied dental models, the liver domain lacks a larger-scale publicly available dataset, which is crucial for advancing research and development in this field.

By comprehensively reviewing related applications and work in the field, we aim to demonstrate the significant advancements and remaining challenges in AR-assisted LLR and 3D mesh segmentation, particularly for liver surface landmark extraction.

3. Materials and methods

3.1. Datasets

In this study, we manually annotated 200 liver meshes to evaluate the proposed method and to promote further research in this field. These meshes were derived from three publicly available liver datasets: 3Dircadb (Soler et al., 2010), LiTS (Bilic et al., 2023), and Amos (Ji et al., 2022). The surface models were first extracted from ground truth

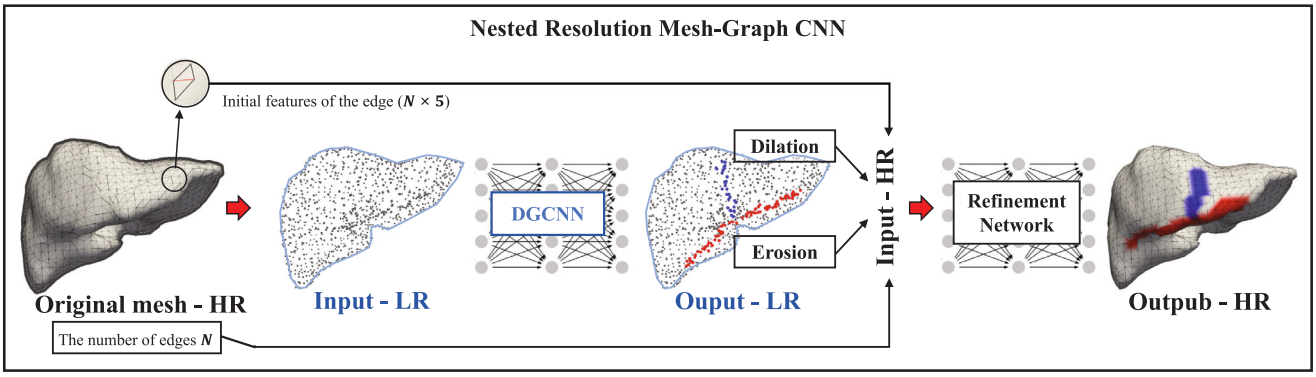

Fig. 2. The overall framework of our proposed nested resolution Mesh-Graph CNN for segmenting key anatomical landmarks on the surface of the liver, i.e. the falciform ligament (in blue) and liver ridge (in red). Here, both sparse representation (denoted as LR for low resolution) and dense representation (denoted as HR for high resolution) of the point cloud are used. N represents the number of edges in the original mesh.

segmentations using the Marching Cubes algorithm in 3D Slicer (Fedorov et al., 2012), and further processed using MeshLab (Cignoni et al., 2008) to ensure manifold simplification, compression, and watertightness. The falciform ligament and liver ridge were manually annotated as vertex-level regions on the liver mesh using Blender software. Their annotation extents were defined with reference to anatomical landmarks: the falciform ligament extended upward from the midline fissure toward the superior surface, terminating approximately 1–1.5 cm below the highest point of the liver. The liver ridge extended bilaterally from the fissure toward the lateral boundaries, ending approximately 3–5 cm inward from the endpoints of the liver's longest transverse diameter. While the liver meshes underwent nearly identical processing steps, the resolution of the resulting meshes, measured in edge counts, varied significantly from 3000 to 20000 edges. This variation was not intentionally introduced but rather arose from the inherent differences in liver shapes and sizes across the datasets. These natural variations in liver morphology present a primary challenge in this task, highlighting the need for a robust segmentation approach.

All annotations were meticulously reviewed by clinical experts to ensure accuracy and fairness in evaluation. Additionally, to further assess the generalizability of our proposed method, we included 9 training cases from the P2ILF challenge (Ali et al., 2025) as an external evaluation cohort. These meshes were generated through heterogeneous reconstruction pipelines (including CT and MRI modalities) with unknown meshing strategies, often resulting in distinct local surface characteristics compared to our internal dataset.

3.2. Nested resolution mesh-graph architecture for anatomical segmentation

Our proposed framework for anatomical segmentation of the liver surface employs a nested resolution mesh-graph architecture. The simplified process is illustrated in Fig. 2. The compressed resolution liver mesh is processed by a dynamic graph network (DGCNN) to capture the global structural features of the liver, obtaining preliminary segmentation results. These initial results are then mapped onto the original high-resolution mesh surface through two levels of propagation, "dilation" and "erosion". Subsequently, combining the topological details present in the high-resolution mesh, we propose an anatomical refining network based on MeshConv, which effectively balances and integrates different levels of priors to produce accurate landmark segmentation.

3.2.1. Theoretical preliminaries in MeshCNN

In this section, we provide the foundational background of MeshCNN. Understanding the key components of MeshCNN — mesh convolution (MeshConv), mesh pooling, and unpooling is crucial for grasping the innovative aspects of our method.

Mesh Convolution (MeshConv). As illustrated in Fig. 3(A), taking the liver mesh as an example, any randomly selected edge (marked in red) on the liver surface lies within a "one-ring" structure, where this edge and its four neighboring edges constitute two faces. This local structure is invariant on the liver surface, providing a robust basis for performing convolution on edges. We show a close-up view of the one-ring structure in Fig. 3(B) and denote the target edge and its neighboring edges as (e) and (a, b, c, d), respectively. For each edge, MeshCNN defines its input feature as a 5-dimensional vector consisting of the dihedral angle, two inner angles, and two edge-length ratios for each adjacent face. Convolution is the dot product between a kernel k and a neighborhood. Thus, the convolution for an edge feature e and its four adjacent edges is:

$$\begin{aligned} f'(e) &= f(e) \cdot k_0 + \sum_{j=1}^{4} k_j \cdot f(e_j) \\ &= k_0 \cdot f(e) + k_1 \cdot f(a) + k_2 \cdot f(b) + k_3 \cdot f(c) + k_4 \cdot f(d), \end{aligned} \tag{1}$$

where $f(a)$, $f(b)$, $f(c)$, $f(d)$ are the features of the neighboring edges and $\{k_j \mid j = 0, 1, \ldots, 4\}$ are the learnable weights. To ensure convolution invariance to the ordering of the input data, MeshCNN applies a set of symmetric functions before convolution, thus:

$$\{e_1, e_2, e_3, e_4\} = \{|a - c|, a + c, |b - d|, b + d\}. \tag{2}$$

This guarantees that the convolution operation is invariant to the initial ordering of the mesh elements. MeshConv always operates on five edges (including the target edge and its four neighbors), maintaining a fixed feature dimension of $N \times C \times 5$, where N is the number of edges, C is the feature dimension, and 5 includes the edge and its neighbors. This convolution can be efficiently implemented using general matrix multiplication (GEMM), with the convolution kernel size set to $(1, 5)$. Thus, MeshConv directly learns fine-grained information along the mesh surface topology, which is crucial for understanding the local details of the mesh. However, due to its strict reliance on local 1-ring neighborhoods, MeshConv captures only limited context within each layer. This limitation becomes more pronounced on high-resolution or resolution-varying meshes, where shallow MeshConv stacks struggle to aggregate stable global features. To mitigate this, traditional MeshCNN pipelines typically introduce mesh pooling to unify resolution and progressively enlarge the receptive field for global context learning.

Mesh Pooling. Unlike pooling operations in 2D vision, mesh pooling in MeshCNN is based on an edge collapse framework, where edges with less important features are collapsed or removed. The pooled mesh retains the significant features while reducing the total number of edges. As shown in Fig. 3(B), the edge collapse operation removes the target edge (e) and converts five original edges (e, a, b, c, d) into two new edges (p, q). The features of (p, q) are defined as:

$$p = \text{avg}(a, b, e) \quad \text{and} \quad q = \text{avg}(c, d, e). \tag{3}$$

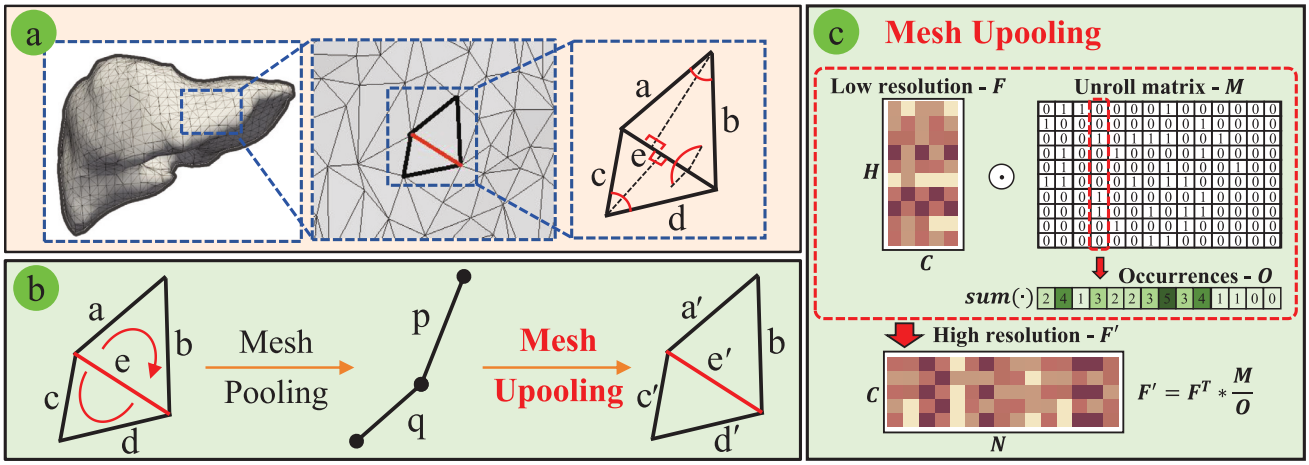

Fig. 3. (A) A close-up view of the single-ring structure of the liver surface and the input edge features, where a, b, c, d, e represent the edges in the single-ring structure. (B) Mesh convolution and the process of pooling and unpooling, resulting in new edges a', b', c', d', e'. (C) The feature propagation process during unpooling in MeshCNN. Here, $F \in \mathbb{R}^{H \times C}$ represents the feature matrix of the low-resolution mesh (with H edges). Using the Unroll matrix $M \in \mathbb{R}^{H \times N}$, these features are propagated to the high-resolution mesh level, yielding the corresponding features $F' \in \mathbb{R}^{N \times C}$ (with N edges, where $N > H$).

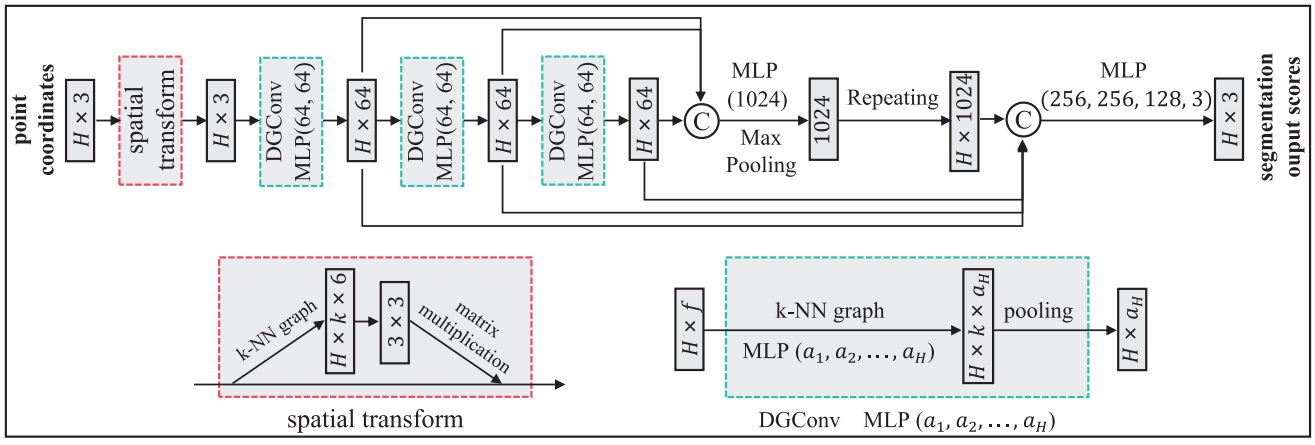

Fig. 4. Dynamic Graph Convolutional Neural Network (DGCNN) architecture used for the initial segmentation of anatomical landmarks on compressed low-resolution meshes. The network includes a spatial transform module for normalizing input edge features, followed by dynamic graph convolution (DGConv) layers that extract multi-level features. Finally, Multi-Layer Perceptron (MLP) layers aggregate these features to produce initial segmentation scores. k-Nearest Neighborhood (k-NN) is used in both spatial and DGConv layers.

However, mesh pooling in MeshCNN is a sequential process that collapses edges one at a time based on feature ranking, making it computationally inefficient for high-resolution inputs. Moreover, due to the anatomical variability and resolution differences of liver meshes, pooling often collapses structurally dissimilar edges, resulting in inconsistent appearances of the pooled meshes and compromising the stability of global representation learning.

Mesh Unpooling. Unpooling is implemented as the inverse process of pooling. Specifically, MeshCNN dynamically memorizes a 0 or 1-encoded unroll matrix M (size $H \times N$) during pooling. As shown in Fig. 3(C), if the value of $M[h, n]$ is 1, it means that the feature of the nth edge in the original mesh participated in the feature calculation of edge-h in the low-resolution mesh. During unpooling, the feature of each high-resolution edge is reconstructed by averaging the features of all low-resolution edges that contributed to its pooling path. While this enables coarse predictions to be propagated back to the original mesh, large resolution gaps may dilute the resulting feature responses. This property is later leveraged to construct soft landmark priors on high-resolution surfaces in our refinement framework.

3.2.2. Low-resolution stage: Global structure modeling with DGCNN

In clinical practice, the appearance and shape of the liver vary significantly, which is also reflected in the liver meshes with different levels of resolution, i.e., the number of edges in the mesh. To address this issue, neural network methods for 2D image tasks typically adjust the input size when handling inputs of varying resolutions to minimize performance degradation. Similarly, for anatomical segmentation tasks of the liver surface, we propose first introducing a modified Dynamic Graph Convolutional Neural Network (DGCNN) to perform rapid global understanding on a unified low-resolution network, producing initial landmark segmentations.

Network Structure. Fig. 4 details our DGCNN network structure, designed to grasp the liver's global structure and generate initial anatomical landmark segmentation. The input to DGCNN consists of the edge features (size $H \times 3$, $H = 3000$) of the low-resolution mesh (also see Fig. 2), where H represents the number of edges at the low-resolution level, and 3 corresponds to the spatial coordinates (x, y, z) of the edge midpoints. These coordinates intuitively reflect spatial relationships, capturing global information.

First, the input mesh features pass through a spatial transform module for normalization and alignment, applying local perturbations. This

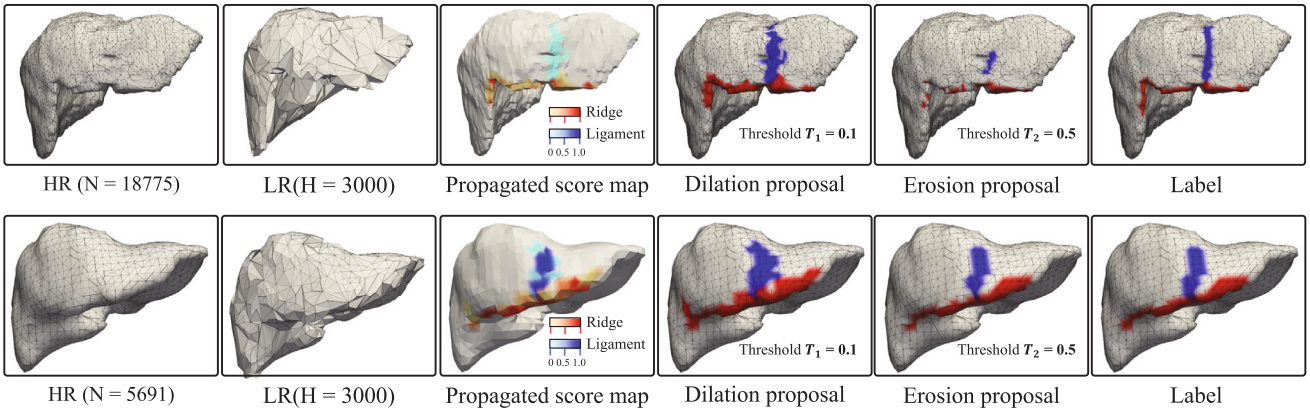

Fig. 5. Comparison with two original meshes with different resolutions (from left to right): the appearance of the (original) high-resolution mesh (HR), the appearance of the compressed low-resolution (LR) mesh, the propagation score map (PSM) mapping the initial segmentation onto the surface of the original high-resolution mesh, the "dilation" landmark proposal based on threshold T_1, the "erosion" landmark proposal based on threshold T_2, and the landmark annotation on the original high-resolution mesh referred as "label".

module constructs a k-nearest neighbor (k-NN) graph to extract features of adjacent points and generates a 3×3 transformation matrix to ensure data alignment. Following normalization, the data is processed through multiple Dynamic Graph Convolution (DGConv) layers.

DGConv layers capture local geometric features, dynamically update adjacency relationships, and perform multi-level feature extraction, enriching feature representations. These layers iteratively extract higher-level graph feature representations using Multi-Layer Perceptron (MLP) layers, further enhancing feature representations. The output of each layer undergoes max pooling to ensure the invariance of feature arrangement.

Ultimately, the features are aggregated through max pooling, resulting in a 1024-dimensional feature vector, which is then concatenated with the original feature matrix to form an $H \times 1220$ matrix. This concatenated matrix is passed through a sequential MLP with layers of $(256, 256, 128, 3)$, producing an $H \times 3$ matrix where each vertex is assigned a score corresponding to one of three classes: falciform ligament, hepatic ridge, or other regions.

This design effectively extracts global information from low-resolution liver meshes, generating initial anatomical landmark segmentations and laying a foundation for refinement on high-resolution meshes.

3.2.3. High-resolution stage: Anatomical landmark refinement

The liver surface is represented by complex triangular units of varying shapes, sizes, and positions. High-resolution meshes capture more local details, which are crucial for accurate landmark segmentation. A straightforward strategy is to map the initial landmark segmentation from the low-resolution mesh onto the high-resolution surface to provide priors, but this approach faces two challenges: (a) mapping from a compressed mesh to the original high-resolution mesh is a reverse operation, making it difficult to achieve high-precision segmentation priors; (b) the number of edges in the original high-resolution mesh varies significantly (as shown in Fig. 5), complicating the neural network's ability to effectively integrate local topology with the segmentation priors, which can lead to suboptimal segmentation performance on the high-resolution mesh. To address these issues, inspired by the unpooling operation in MeshCNN, we propose two levels of mapping strategies: "dilation" and "erosion", to propagate the initial landmark segmentations onto the high-resolution mesh surface. Following this, we design an anatomical refining network based on MeshConv, with the aim of effectively integrating and adjusting multi-level priors on high-resolution meshes, so as to better capture fine-grained anatomical boundaries under varying topological conditions.

Dilation and Erosion Landmark Proposals. Building on the earlier descriptions, the unpooling operation in MeshCNN allows for the lossless restoration of high-resolution meshes while propagating edge features onto the high-resolution surface. We propose to map the landmark segmentations from the low-resolution mesh as edge features onto the original high-resolution mesh surface, generating segmentation priors. Specifically, the low-resolution segmentation results are one-hot encoded into $H \times 3$ matrices ($H = 3000$). Through unpooling, these encoded features are propagated to the original resolution, resulting in score maps on the high-resolution mesh. Fig. 5 displays these propagated score maps for two samples (third column). In the first row, for the original mesh with a significantly higher number of edges ($N = 18775$), the difference in edge counts compared to the compressed mesh ($H = 3000$) leads to feature "dilution" during propagation due to averaging (see Fig. 3(C) for details). In contrast, for the original mesh with an edge count closer to that of the low-resolution mesh (second row), the landmarks are less diluted during propagation, resulting in scores closer to the ground truth.

When comparing these propagation score maps with the corresponding labels, an interesting observation emerges: for higher-resolution original meshes, using a lower confidence threshold on the propagation score map produces landmark proposals closer to the true labels (see Fig. 5, columns 4 and 6). Conversely, for original meshes with resolutions close to the compressed mesh, setting a higher confidence threshold yields more accurate landmark proposals (see Fig. 5, columns 5 and 6). This observation suggests that no single threshold is universally optimal across mesh resolutions. To achieve complementary coverage and boundary sensitivity, we introduce two landmark mapping strategies: "dilation" using a lower threshold (T_1) for broader recall, and "erosion" using a higher threshold (T_2) for confident boundary cues. Specifically, based on the initial landmark segmentation obtained from the DGCNN model, we apply two confidence thresholds to obtain two levels of landmark proposals on the original mesh surface:

$$Dilation = \begin{cases} 1 & \text{if } \mathbf{PSM} \geq T_1 \\ 0 & \text{if } \mathbf{PSM} < T_1, \end{cases} \tag{4}$$

$$Erosion = \begin{cases} 1 & \text{if } \mathbf{PSM} \geq T_2 \\ 0 & \text{if } \mathbf{PSM} < T_2, \end{cases} \tag{5}$$

where **PSM** represents the propagation score map. The thresholds $T_1 = 0.1$ and $T_2 = 0.5$ were empirically selected ($T_1 < T_2$) based on grid search experiments to balance recall and precision across mesh resolutions. Fig. 5 shows the "dilation" and "erosion" landmark proposals for two different resolution original meshes. It can be seen that "erosion"

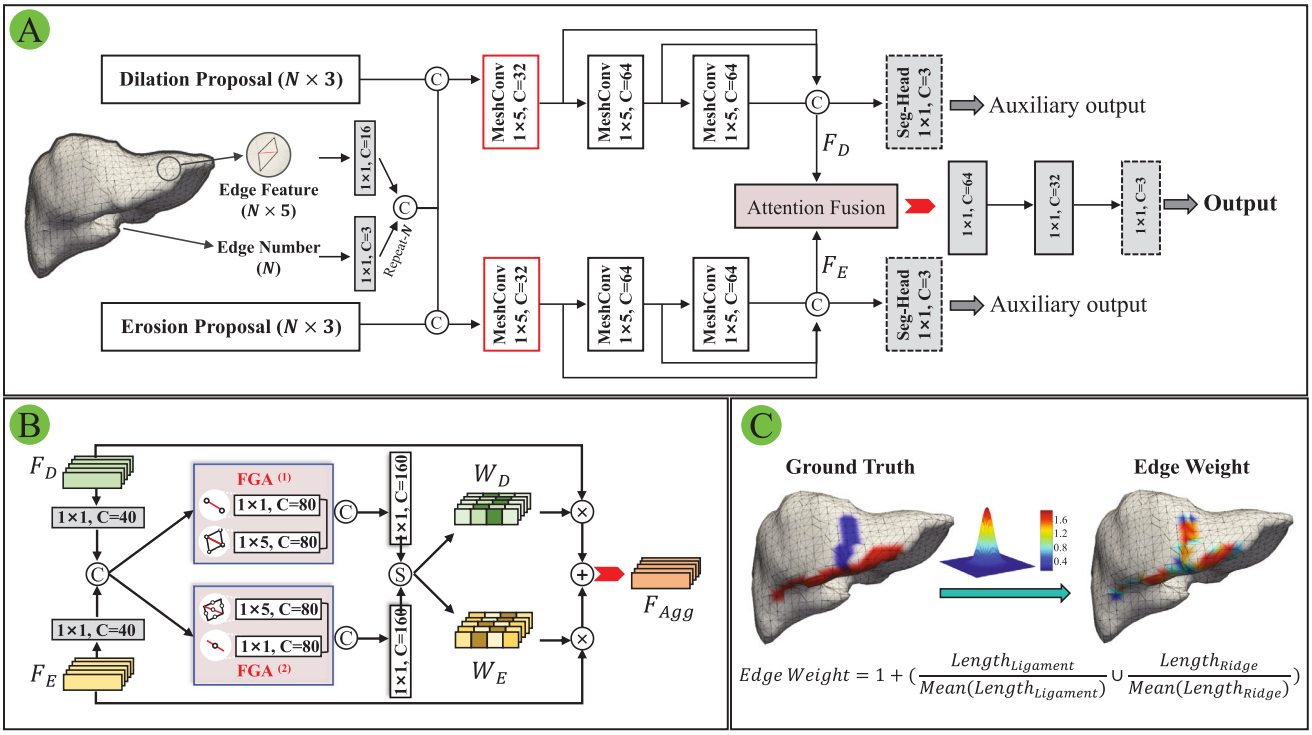

Fig. 6. (A) Architecture of the anatomical refining network based on MeshConv. Auxiliary segmentation heads (ASH) promote reliable prior extraction, while the attention fusion module (AFM) integrates these priors into refined landmark predictions. The first MeshConv (in red) omits symmetric functions (i.e., Eq. (2)) due to resolution-encoded inputs. (B) Attention fusion module (AFM) based on fine-grained aggregation (FGA), where superscripts (1) and (2) indicate separate instances used to enhance cross-branch communication. (C) Edge weights incorporated in our anatomy-aware Dice (AAD) losses.

proposals represent higher segmentation confidence but result in under-segmentation for higher-resolution meshes with more edges (see Fig. 5, row 1, column 5). Although "dilation" proposals are closer to the ground truth in high-resolution meshes, they may contain more errors for original meshes with fewer edges (see Fig. 5, row 2, column 4). Therefore, effectively incorporating the topological characteristics of the original mesh surface and selectively extracting reliable priors from the dilation and erosion proposals are key to refining anatomical landmark segmentation on liver surfaces with varying morphologies.

Refinement Network Architecture for Anatomical Segmentation. As previously described, MeshConv in MeshCNN performs edge feature calculations along the mesh surface topology, effectively capturing the topological details of high-resolution meshes. Building on this, we propose an anatomical refining network based on MeshConv. This network is designed to selectively incorporate the dilation and erosion proposals, with the goal of capturing reliable anatomical cues across meshes with varying shapes, appearances, and resolutions. The architecture of the refinement network is illustrated in Fig. 6, and the inputs include: (1) the edge features (size $N \times 5$) edge features, which directly represent the topology of the mesh units; (2) the number of edges (N), corresponding to the resolution of the original mesh; and (3) the "dilation" and (4) "erosion" landmark proposals, both encoded as the one-hot vectors of size $N \times 3$.

First, the edge features (size $N \times 5$) and mesh resolution (N) pass through an MLP, extracting new feature vectors of size $N \times 16$ and $N \times 3$, respectively, which are then combined. The combined features (size $N \times 19$) are further concatenated with the different levels of landmark proposals, integrating the topological details of the original mesh with the landmark priors from the low-resolution mesh. Our method generates two sets of fused features, each with size $N \times 22$, which are processed by independent branches based on MeshConv. The resolution encoding is introduced to guide each branch in learning plausible priors from the corresponding landmark proposals. Specifically, when the original mesh has a relatively high resolution, the branch fused with the dilation proposal benefits from broader coverage, whereas for meshes closer in resolution to the compressed input, the erosion-guided branch is more effective in preserving segmentation precision. This design enables each branch to specialize in extracting reliable cues under different mesh configurations. Each branch in our network consists of three consecutive MeshConv layers, which gradually extract higher-level features. The multi-scale features extracted by each branch are then concatenated and passed through an attention fusion module (AFM), which is designed to adaptively integrate the features from different branches and enhance the representation of anatomical cues.

These comprehensive features are subsequently passed through sequential MLP layers to output the final refined landmark segmentation. It is also important to note that to facilitate the effective extraction of reliable priors from different proposals, each branch in our model is connected to an auxiliary segmentation head (ASH), providing direct supervision to encourage branch-specific learning before feature fusion. To train the refinement network, we adopt a combination of cross-entropy loss and anatomy-aware Dice (AAD) loss.

Attention Fusion Module (AFM). In our anatomical refining network, the attention fusion module (AFM) is introduced to adaptively balance and integrate informative cues from different branches, aiming to enhance landmark prediction across meshes with varying morphologies. As illustrated in Fig. 6(B), the module begins by using a 1×1 convolution to compress the multi-scale features (F_D and F_E) from each branch, generating feature vectors of size $N \times 40$, which are then concatenated. This combined feature vector is subsequently processed by a Fine-Grained Aggregation (FGA) module. The FGA module computes two types of response maps, W_D (size $N \times 160$) and W_E (size $N \times 160$), to modulate and integrate F_D and F_E, respectively. The formula for the cross-branch feature aggregation is as follows:

$$F_{Aggr} = W_D \otimes F_D + W_E \otimes F_E, \tag{6}$$

where F_{Aggr} (size $N \times 160$) represents the aggregated features, and $\otimes$ denotes element-wise multiplication.

The FGA module excels at generating response maps that capture cross-branch relationships at two granularity levels: points (or edges) and triangular elements. To handle this complex task, we deploy two instances of the FGA, namely $FGA^{(1)}$ and $FGA^{(2)}$, to model these cross-branch correspondences. Specifically, both FGAs leverage dual MeshConv layers with different kernel sizes. The process begins with a (1×1) convolution layer for primary feature distillation, followed by a (1×5) convolution layer that further processes the features, generating the outputs for $FGA^{(1)}$ and $FGA^{(2)}$ to form the response maps W_D and W_E. These response maps are obtained through a series of concatenation, convolution, Softmax layers, and weighted summation operations. Together, these components define the attention-based fusion process within our anatomical refinement framework.

Anatomy-Aware Dice (AAD) Loss. In vertex- or edge-based landmark segmentation tasks on meshes, traditional Dice loss often underperforms due to the uneven distribution of edge lengths. This issue is particularly prominent in regions like the falciform ligament, where sparse and elongated triangles can lead to underrepresented gradient signals during optimization.

To address this problem, we propose an anatomy-aware Dice loss with two enhancements tailored to the liver surface mesh:

- **Joint Modeling of Anatomical Regions:** The falciform ligament and hepatic ridge exhibit consistent spatial correlations and often form a continuous anatomical boundary. To leverage this, we treat them as a unified region and compute a joint Dice loss, encouraging the model to learn their shared topological structure more effectively.
- **Incorporation of Edge Weights:** To mitigate edge length imbalance, we apply a weighting factor to each edge based on its relative geometric length. Specifically, the edge weights w_i (denoted as *Edge Weight* in Fig. 6(C)) are used to compute a weighted Dice loss:

$$\mathcal{L}_{\text{weighted-Dice}} = 1 - \frac{2\sum_i w_i p_i g_i}{\sum_i w_i (p_i + g_i)}, \tag{7}$$

where p_i and g_i are the predicted and ground truth scores for edge i, and w_i is its edge weight.

3.3. Two-stage training and inference for nested learning

As described in the previous sections, our nested-resolution architecture consists of two stages: a low-resolution stage for global landmark localization, and a high-resolution stage for refining anatomical details based on mesh topology. In this section, we detail the training pipeline and inference strategy that underpin this design.

Stage 1: Training the low-resolution DGCNN. To learn global geometric structures efficiently, we train the DGCNN model using randomly sampled edge midpoints from the original high-resolution mesh. In each training epoch, a set of $H = 3000$ edge midpoints is sampled to form a point cloud, which serves as input to the DGCNN for initial landmark segmentation. The model is supervised with a combined loss function:

$$\mathcal{L}_{low-res} = \mathcal{L}_{CE} + \mathcal{L}_{Dice}. \tag{8}$$

This stage is designed to facilitate the learning of coarse but globally consistent anatomical priors across meshes of varying shapes and scales.

Stage 2: Training the high-resolution refinement network. In the second stage, the pretrained DGCNN is fixed. Given an original high-resolution liver mesh, we apply mesh pooling to compress it to a resolution of $H = 3000$. The resulting mesh is then processed by the DGCNN to produce initial landmark predictions, which are mapped back onto the original high-resolution surface using unpooling and confidence-based thresholding, generating both "dilation" and "erosion" proposals. These proposals, along with the mesh's edge features and resolution encoding, are fed into the anatomical refinement network. Although only the refinement network is updated during this stage, the input generation depends on the full nested-resolution pipeline. Therefore, we refer to this approach as nested-resolution training.

The loss function for this stage aggregates supervision from three outputs, including the two branch-specific auxiliary segmentation heads (ASH) and the final fused prediction. The overall loss is formulated as:

$$\mathcal{L}_{high-res} = \mathcal{L}_{Dilation} + \mathcal{L}_{Erosion} + \mathcal{L}_{Fusion}, \tag{9}$$

where each term combines cross-entropy loss and the anatomy-aware Dice loss defined in Section 3.2.3.

Inference. During testing, our model operates in a fully automatic and end-to-end manner. Starting from an input high-resolution liver mesh, we first apply mesh pooling to obtain a compressed version, which is passed through the DGCNN to produce initial landmark predictions. These predictions are propagated back to the original surface to construct the "dilation" and "erosion" proposals, which – along with other topological features – are used by the refinement network to generate the final anatomical segmentation. The entire process is completed within a single forward pass, without requiring any manual post-processing.

4. Experiments and results

4.1. Competing methods and experimental setup

To thoroughly validate our method, we randomly divided the manually annotated liver mesh data into training, validation, and test sets in a 10:3:7 ratio, reporting the average performance on the test set across multiple experiments.

Currently, no publicly available algorithms specifically address the segmentation of liver surface landmarks. Therefore, we selected four state-of-the-art (SOTA) deep learning methods for mesh data processing: Pointnet++ (point-based), DGCNN (graph-based), MeshCNN (edge-based), and TSGCN (designed for dental meshes). Pointnet++ and MeshCNN were employed by multiple teams in the P2ILF challenge (Ali et al., 2025), demonstrating their capability on liver meshes. DGCNN, known for its strong local and global learning abilities, was also chosen for processing the compressed low-resolution meshes in our study. Additionally, the original MeshCNN, which takes five initial edge features as input, excels in local understanding but struggles with global information. To mitigate this limitation, we computed the center point coordinates of the edges as an additional input, enhancing MeshCNN and designating this variant as MeshCNN-1 for comparison. Furthermore, TSGCN, which incorporates dual parallel graph networks and leverages attention mechanisms for feature fusion, is specialized for dental mesh segmentation tasks, offering an additional perspective in our comparative analysis.

All methods, including ours, were implemented in PyTorch and executed on an RTX 8000 GPU. For Pointnet++ and DGCNN, the training involved random sampling of spatial coordinates (size 3000×3) while testing inputs on all edges to obtain segmentation results on the original resolution mesh. MeshCNN and MeshCNN-1 employed a UNet-like architecture with three levels of down-sampling, processing fixed-size input data of edges (size 20000×5). We also used the official implementation of TSGCN for training and evaluation, modifying the number of input edges to 3000, matching the DGCNN setting. Finally, in our method, we first obtained different levels of landmark proposals using a trained DGCNN model, followed by training the anatomical refining network with original meshes at varying resolutions. All experiments were optimized using the Adam optimizer with a learning rate of 0.01, and models with a minimum loss of over 600 epochs were selected for testing.

Table 1
Quantitative comparison with different segmentation methods. "Overall" refers to the combined evaluation of both anatomical landmarks – falciform ligament and liver ridge – as a single class, serving as a complementary summary metric.

	Ligament			Ridge			Overall		
	Dice(%)	CD(mm)	HD(mm)	Dice(%)	CD(mm)	HD(mm)	Dice(%)	CD(mm)	HD(mm)
Pointnet++	0.0 ± 0.0	60.3 ± 18.2[a]	47.8 ± 11.4[a]	20.7 ± 16.5	33.4 ± 24.2	54.8 ± 29.4	19.1 ± 15.2	37.1 ± 22.5	61.9 ± 24.1
DGCNN	21.2 ± 23.5	22.1 ± 18.7	25.8 ± 14.4	57.2 ± 12.8	5.6 ± 6.0	21.0 ± 15.3	50.3 ± 15.1	8.0 ± 6.8	25.8 ± 15.1
MeshCNN	0.0 ± 0.1	165.1 ± 55.6[a]	118.8 ± 27.1[a]	0.6 ± 2.0	85.0 ± 51.4	81.3 ± 37.1	0.8 ± 2.3	86.1 ± 49.8	82.7 ± 35.9
MeshCNN-1	16.2 ± 13.8	33.6 ± 19.5	56.3 ± 36.4	49.7 ± 16.9	12.2 ± 12.1	34.3 ± 23.5	42.5 ± 14.2	15.0 ± 9.2	43.3 ± 22.4
TSGCN	1.8 ± 7.2	34.8 ± 25.8[a]	34.2 ± 13.1[a]	35.9 ± 19.0	22.7 ± 23.0	49.6 ± 31.6	33.3 ± 17.7	27.1 ± 24.9	55.6 ± 28.6
Ours	**32.6 ± 17.9**	**14.9 ± 12.3**	**22.2 ± 14.0**	**59.6 ± 13.8**	**5.5 ± 4.8**	**16.4 ± 9.5**	**54.9 ± 13.5**	**6.9 ± 4.8**	**18.4 ± 8.5**

[a] Indicates that the predicted landmark region is empty (i.e., $|\mathbf{v}| = 0$ in Eq. (11)), and thus the distance metrics (CD and HD95) are calculated only on the subset of test cases with non-empty predictions. The corresponding prediction rates are reported in Table A.1.

4.2. Metrics for comparison

The performance of our method was rigorously evaluated using three standard metrics: Dice similarity coefficient (Dice), 3D Chamfer Distance (CD) (Wu et al., 2021), and Hausdorff Distance (HD). These metrics respectively capture segmentation accuracy, average spatial deviation, and worst-case boundary errors, offering a comprehensive assessment of anatomical landmark prediction quality.

Given the high class imbalance in this task – where landmarks occupy only a small fraction of the liver mesh – we adopt the Dice coefficient to evaluate the model's ability to detect foreground (i.e., landmark) regions:

$$Dice = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \tag{10}$$

where TP, FP, and FN denote the numbers of true positives, false positives, and false negatives, respectively.

To assess geometric alignment in 3D space, we also employed the Chamfer Distance (CD, mm), which measures the average bidirectional distance between predicted and ground truth landmark surfaces:

$$CD(V, W) = \frac{\sum_{\mathbf{v}\in V} \min_{\mathbf{w}\in W} \|\mathbf{v} - \mathbf{w}\|^2}{|V|} + \frac{\sum_{\mathbf{w}\in W} \min_{\mathbf{v}\in V} \|\mathbf{w} - \mathbf{v}\|^2}{|W|}, \tag{11}$$

where V and W denote the sets of vertex coordinates from the predicted and ground truth landmark regions, and $|\cdot|$ denotes set cardinality.

To further capture local outlier deviations, we additionally introduce the Hausdorff Distance (HD, mm), defined as:

$$HD(V, W) = \max\left\{ \sup_{\mathbf{v}\in V} \inf_{\mathbf{w}\in W} \|\mathbf{v} - \mathbf{w}\|, \sup_{\mathbf{w}\in W} \inf_{\mathbf{v}\in V} \|\mathbf{w} - \mathbf{v}\| \right\}. \tag{12}$$

Here, sup and inf denote the supremum (maximum) and infimum (minimum) over the respective point sets. In practice, we report the 95th-percentile Hausdorff Distance (HD95) to reduce sensitivity to extreme outliers. Unlike CD, which captures average proximity, HD95 reflects the worst-case deviation between surfaces, offering a complementary perspective on boundary quality in challenging regions.

All metrics were averaged across the test set. For methods with empty predictions on the falciform ligament, CD and HD95 were calculated only on non-empty cases, and the corresponding prediction rates are reported in Appendix (Tables A.1–A.2). Higher Dice scores indicate better segmentation accuracy, while lower CD and HD values reflect closer geometric alignment with ground truth.

4.3. Comparison of results by different methods

Table 1 reports the Dice similarity coefficient (Dice), 3D Chamfer Distance (CD), and Hausdorff Distance (HD) values for each anatomical landmark. Consistently higher scores on the ridge suggest that the ligament is more challenging to segment, likely due to its less distinctive surface geometry. Among the baselines, MeshCNN and Pointnet++ both performed poorly on the ligament, with Dice scores near zero. While Pointnet++ lacks topological modeling, MeshCNN focuses on local edge structures without positional encoding. Notably, Pointnet++ slightly outperformed MeshCNN on the ridge, indicating that topology alone may not suffice. By incorporating spatial coordinates, MeshCNN-1 yielded substantial improvements—for instance, ligament Dice rose from 0.0 to 16.2-underscoring the benefit of geometric cues. TSGCN, tailored for rigid dental meshes, showed moderate performance but struggled on the ligament (Dice: 1.8), likely due to the liver's deformable anatomy and irregular mesh topology. DGCNN delivered more balanced results, particularly on the ridge, with a Dice of 57.2 and relatively low CD and HD values (5.6 mm and 21.0 mm, respectively), benefiting from its ability to capture both local and global geometry via graph-based learning. Building on DGCNN, our method introduces a nested-resolution architecture with anatomical prior refinement. It consistently outperformed all baselines across metrics, achieving Dice scores of 32.6 and 59.6 for the ligament and ridge, respectively, and further reducing CD on the ligament from 22.1 mm to 14.9 mm—highlighting its improved performance, especially in anatomically complex regions such as the falciform ligament. Notably, the reduced Chamfer Distance approaches the clinically acceptable margin for intraoperative registration (Zhong et al., 2017), supporting its potential utility in surgical navigation.

To further assess model performance, Fig. 7 presents a qualitative comparison across liver meshes of varying resolutions (edge count **N**). Pointnet++, which operates solely on spatial coordinates without topological context, failed to detect the falciform ligament and produced notable false positives along the ridge. MeshCNN, despite leveraging edge-based topology, performed worst overall, often missing both landmarks-a result consistent with its near-zero Dice and high HD in Table 1. By incorporating spatial coordinates, MeshCNN-1 showed noticeable improvements, particularly in ridge localization, though frequent over- and under-segmentation persisted. TSGCN, originally designed for dental meshes, achieved modest results: it successfully captured the ridge in certain cases but introduced large segmentation errors under higher-resolution conditions, likely due to the liver's non-rigid morphology and irregular surface patterns. DGCNN offered more stable outputs and consistently identified the approximate locations of both landmarks, especially the ridge, demonstrating strong global representation capabilities. However, its predictions for the ligament remained coarse, likely due to resolution variance and the sparsity of the target region. Our method extends DGCNN by introducing anatomical prior refinement and nested-resolution processing, aimed at improving landmark localization through enhanced structural awareness. As shown in the final column, the resulting segmentations appear more consistent with ground truth, particularly in regions with high surface complexity or lower landmark visibility, aligning with the improvements observed in quantitative metrics.

4.4. Ablation and component analysis

We conducted a comprehensive and thorough ablation study using the DGCNN model architecture as a baseline to evaluate the effectiveness of the key components in the proposed nested resolution Mesh-Graph CNN. The following eight ablation settings were included:

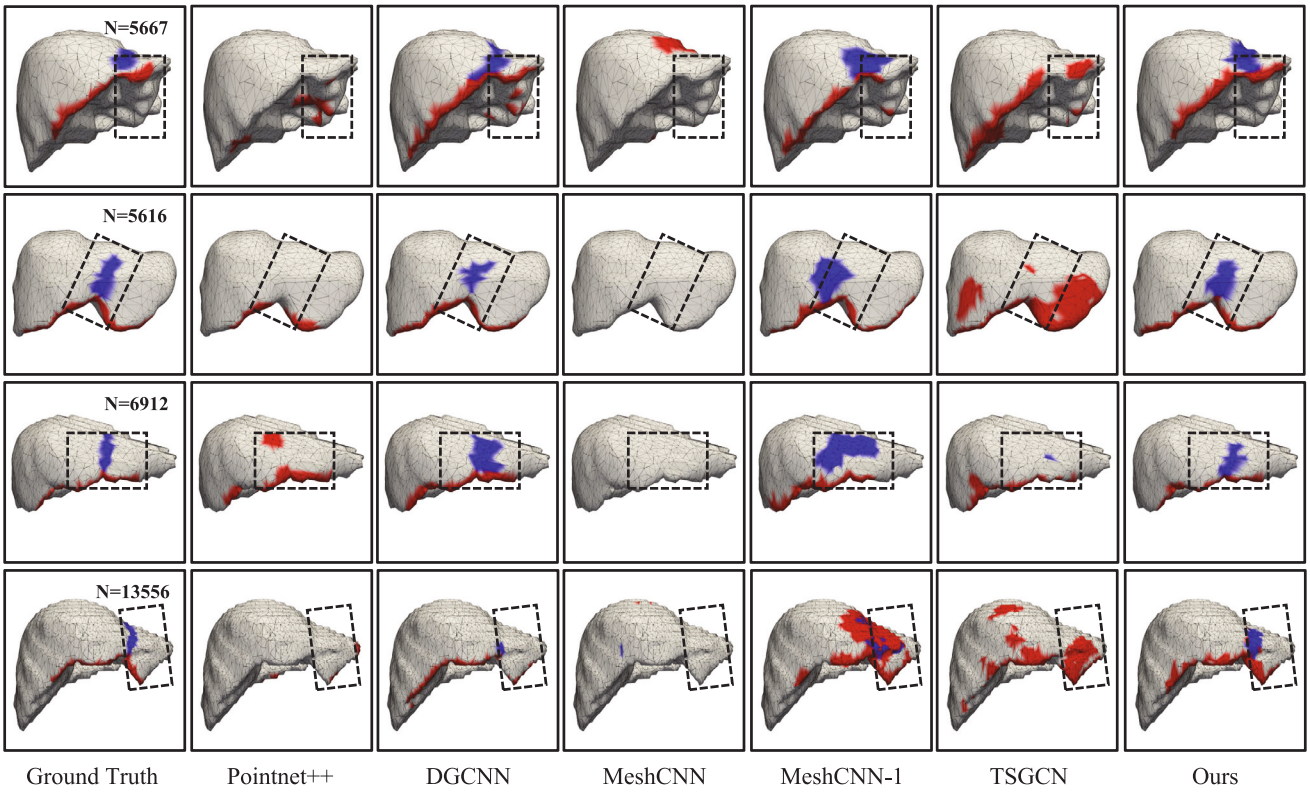

Fig. 7. Comparison of segmentation results of different segmentation methods. The falciform ligament on the surface of the liver is shown in blue, and the liver ridge is shown in red. In addition, the dotted boxes in the figure indicate areas to focus on.

Table 2
Ablation study on different key components. Wilcoxon signed-rank test and effect size (r) were applied to the last four variants to evaluate statistical significance and effect magnitude.

	Ligament			Ridge			Overall		
	Dice(%)	CD(mm)	HD(mm)	Dice(%)	CD(mm)	HD(mm)	Dice(%)	CD(mm)	HD(mm)
DGCNN	21.2 ± 23.5	22.1 ± 18.7	25.8 ± 14.4	57.2 ± 12.8	5.6 ± 6.0	21.0 ± 15.3	50.3 ± 15.1	8.0 ± 6.8	25.8 ± 15.1
DGCNN-D	25.3 ± 20.6	17.1 ± 13.8	22.8 ± 11.8	54.2 ± 14.8	5.5 ± 5.0	16.0 ± 10.8	49.9 ± 14.3	7.1 ± 4.9	19.2 ± 10.3
DGCNN-E	18.3 ± 16.4	18.5 ± 13.8	23.0 ± 11.0	41.8 ± 14.5	6.3 ± 5.0	**15.8 ± 10.4**	38.3 ± 14.0	7.8 ± 4.9	19.0 ± 10.1
DGCNN-M	27.7 ± 19.0	29.6 ± 39.3	41.3 ± 37.6	53.2 ± 13.3	8.5 ± 8.8	31.7 ± 32.4	49.7 ± 12.2	9.8 ± 8.5	32.0 ± 26.5
Ours-ADD	31.0 ± 20.4**§	17.6 ± 14.5**§	30.0 ± 27.7**§	58.3 ± 14.2**§	5.6 ± 4.8**§	27.1 ± 25.9**§	54.0 ± 13.7**§	7.2 ± 4.7**§	26.6 ± 19.8**§
Ours-AFM	31.1 ± 20.3†	16.9 ± 13.8*‡	25.9 ± 26.6**§	58.8 ± 14.0**‡	5.9 ± 5.0	20.0 ± 17.0**§	54.4 ± 13.4*‡	7.1 ± 4.7†	22.0 ± 14.3**§
Ours-ASH	31.7 ± 19.0*§	15.8 ± 13.4*‡	24.4 ± 22.1**‡	59.0 ± 14.4†	5.6 ± 5.0†	18.2 ± 13.0**§	54.7 ± 13.5*‡	7.0 ± 4.7	21.3 ± 12.7†
Ours-AAD	**32.6 ± 17.9****§	**14.9 ± 12.3****‡	**22.2 ± 14.0****§	**59.6 ± 13.8****§	**5.5 ± 4.8**†	16.4 ± 9.5**§	**54.9 ± 13.5**†	**6.9 ± 4.8**	**18.4 ± 8.5****§

* $p < 0.1$.
** $p < 0.05$.
† Small effect ($r \geq 0.1$).
‡ Medium effect ($r \geq 0.3$).
§ Large effect ($r \geq 0.5$).

(1) The landmark segmentation result of the baseline DGCNN is shown in Table 1 (denoted as "DGCNN").
(2) Baseline DGCNN's "dilation" proposals on the original high-resolution mesh (denoted as "DGCNN-D").
(3) Baseline DGCNN's "erosion" proposals on the original high-resolution mesh (denoted as "DGCNN-E").
(4) Combining the edge features of the high-resolution mesh with the results from settings (1) to (3), the segmentation results on the original resolution mesh were generated using a two-layer MeshConv (denoted as "DGCNN-M").
(5) Replacing the two-layer MeshConv in setting (4) with the proposed anatomical refining network, but substituting the attention fusion module with simple feature addition (denoted as "Ours-ADD").
(6) The proposed anatomical refining network with attention fusion module was used to achieve landmark segmentation on the original mesh (denoted as "Ours-AFM").
(7) Based on setting (6), an auxiliary segmentation head was added to each MeshConv-based branch of the anatomical refining network (denoted as "Ours-ASH").
(8) The full Mesh-Graph CNN, where the refinement network is trained using a combination of CE loss and anatomy-aware Dice loss (denoted as "Ours-AAD"). Note that settings (2) to (7) used only CE loss for training.

Table 2 presents the quantitative results of our ablation study. The first three configurations – baseline DGCNN and its propagated proposals (DGCNN-D and DGCNN-E) – show that relying solely on

Table 3
Dice value comparison of different loss function setups for Falciform Ligament and Liver Ridge, considering edge length distributions. The threshold t is defined as a multiple of the mean edge length ($Mean$), where edges longer than t are considered "Long Edges" (which are sparse and important on the liver surface), and edges shorter than t are "Short Edges" (which are denser and more frequent). The last three rows (using Anatomy-Aware Dice) were further evaluated using Wilcoxon signed-rank tests and effect size (r) to assess significance and effect strength.

		Ligament		Ridge	
		Long edges (>t)	Short edges (<t)	Long edges (>t)	Short edges (<t)
CE	$t = 2.0 \times Mean$	27.1 ± 37.0	31.9 ± 19.1	77.7 ± 36.3	**95.0 ± 12.8**
	$t = 1.5 \times Mean$	28.7 ± 28.4	32.2 ± 19.8	94.1 ± 15.6	94.6 ± 13.3
	$t = 1.0 \times Mean$	38.0 ± 22.4	28.5 ± 19.4	98.2 ± 9.2	90.8 ± 17.3
CE + Basic Dice	$t = 2.0 \times Mean$	30.1 ± 39.8	30.9 ± 19.0	78.9 ± 34.1	94.6 ± 13.7
	$t = 1.5 \times Mean$	28.1 ± 27.0	31.0 ± 19.7	95.5 ± 14.9	94.3 ± 13.9
	$t = 1.0 \times Mean$	37.7 ± 22.8	27.3 ± 18.9	97.9 ± 9.6	90.5 ± 17.0
CE + Anatomy-Aware Dice	$t = 2.0 \times Mean$	**33.3 ± 38.0****§	**32.7 ± 18.1***‡	**79.4 ± 33.1***†	94.9 ± 12.9*†
	$t = 1.5 \times Mean$	**30.0 ± 26.7***‡	**32.9 ± 18.6****§	**96.0 ± 13.6***†	**94.8 ± 12.9**†
	$t = 1.0 \times Mean$	**41.2 ± 21.8****§	**28.9 ± 17.9****‡	**98.5 ± 8.0***†	**91.4 ± 16.5**†

* $p < 0.1$.
** $p < 0.05$.
† Small effect ($r \geq 0.1$).
‡ Medium effect ($r \geq 0.3$).
§ Large effect ($r \geq 0.5$).

low-resolution predictions results in suboptimal performance on high-resolution meshes, with modest Dice scores and CD values, and relatively stable HD, likely due to the absence of any refinement process. In contrast, directly applying a two-layer MeshConv for refinement in DGCNN-M introduced notable noise, leading to worse HD scores-for instance, on the ligament, HD increased from 25.8 mm (DGCNN) to 41.3 mm—without improvement in Dice. To address this, Ours-ADD introduces a dual-branch design that processes "dilation" and "erosion" proposals separately before feature fusion via simple addition. Compared to DGCNN-M, this approach improved ligament Dice by 3.3 points and reduced HD by over 11 mm, with statistically significant improvements ($p < 0.05$) and large effect sizes ($r \geq 0.5$), supporting the benefit of preserving distinct proposal pathways. Replacing feature addition with an attention fusion module (AFM), which adaptively weighs branch-specific features, yielded further gains—for example, ridge HD dropped from 27.1 mm to 20.0 mm in Ours-AFM. Adding auxiliary segmentation heads (ASH) in Ours-ASH provided branch-level supervision, resulting in small yet consistent improvements, such as reducing ligament HD from 25.9 mm to 24.4 mm. Finally, applying anatomy-aware Dice loss (AAD) across all segmentation heads led to the most stable results overall; compared to Ours-ASH, it achieved higher Dice on the ligament (32.6 vs. 31.7), reduced CD and HD, and notably lowered standard deviations across all metrics.

Effectiveness of the Attention Fusion Module. To assess the impact of the attention fusion module (AFM), we visualize the attention weights W_D and W_E from Eq. (6), averaged across channels, on three representative liver meshes with varying resolutions (Fig. 8, AFM columns). In the first mesh (**N** = 5613), although both proposals capture the general landmark regions, AFM successfully suppresses false positives along the ridge caused by the dilation branch, leading to sharper boundaries. In the second, higher-resolution mesh, the dilation proposal broadly covers the ligament, while the erosion proposal offers tighter segmentation but introduces uncertainty in the ligament region. Here, AFM assigns greater weights to the dilation branch around the ligament and to the erosion branch around the ridge, adaptively combining their strengths. On the third mesh, the erosion proposal undersegments the landmark, while the dilation branch introduces extensive false positives. AFM selectively enhances confident regions while down-weighting unreliable predictions, resulting in more complete and precise outputs. These visual patterns illustrate how AFM enables the model to dynamically integrate complementary priors under varying anatomical and resolution conditions.

Role of Auxiliary Segmentation Heads (ASH). To improve the reliability of each proposal branch, auxiliary segmentation heads (ASH) are added to the ends of the dilation and erosion paths, providing branch-specific supervision. As illustrated in Fig. 8 (ASH columns), these outputs partially correct errors in the initial proposals – such as over-segmentation or missed regions – and yield more structured predictions. Though not final outputs, they help each branch preserve complementary anatomical cues during feature learning, supporting effective AFM fusion in final landmark refinement.

Comprehensive Analysis of Anatomy-Aware Dice Loss. To further examine the impact of the anatomy-aware Dice loss (AAD), we evaluated three loss configurations-CE only, CE + basic Dice, and CE + AAD-across different edge-length thresholds (t = 1.0, 1.5, 2.0× mean). Table 3 reports the Dice scores for short and long edges, separately for the falciform ligament and liver ridge. Long edges, though sparse, are critical for anatomical completeness yet more difficult to segment accurately. As t increases, segmentation performance on long edges declines notably with CE loss alone—for instance, Dice for ligament long edges drops from 38.0 at $t = 1.0$ to 27.1 at $t = 2.0$. Adding a standard Dice term yields limited and inconsistent improvement, and may compromise short-edge accuracy, likely due to its uniform weighting across spatial structures. In contrast, incorporating the anatomy-aware Dice loss leads to consistent gains across all thresholds. At $t = 2.0$, Dice scores for long edges improve to 33.3 (ligament) and 79.4 (ridge), with concurrent improvements on short edges. Many of these gains are statistically significant ($p < 0.05$), with medium to large effect sizes ($r \geq 0.3$), as indicated in Table 3. These results suggest that the anatomy-aware Dice loss helps achieve more balanced segmentation across heterogeneous edge types.

4.5. Generalizability validation on external dataset

To validate generalizability, our method was applied to 9 external liver mesh datasets from the P2ILF challenge at MICCAI'2022, with mesh resolutions ranging from 3000 to 13000 edges. It is important to note that the 3D mesh annotations in the P2ILF dataset are based on actual intraoperative observations, which are limited by the camera's field of view, making it impossible to observe the complete ligament and hepatic ridge areas. As a result, compared to the 200 liver meshes annotated in our study, the 9 liver meshes provided by the P2ILF challenge – generated using heterogeneous reconstruction pipelines – can be considered partially annotated and topologically distinct. Furthermore, due to the real clinical scenario of preoperative 3D scans of patients in the P2ILF dataset, there are significant differences in appearance, shape, and spatial position of the liver meshes, presenting a greater challenge for liver mesh landmark extraction methods.

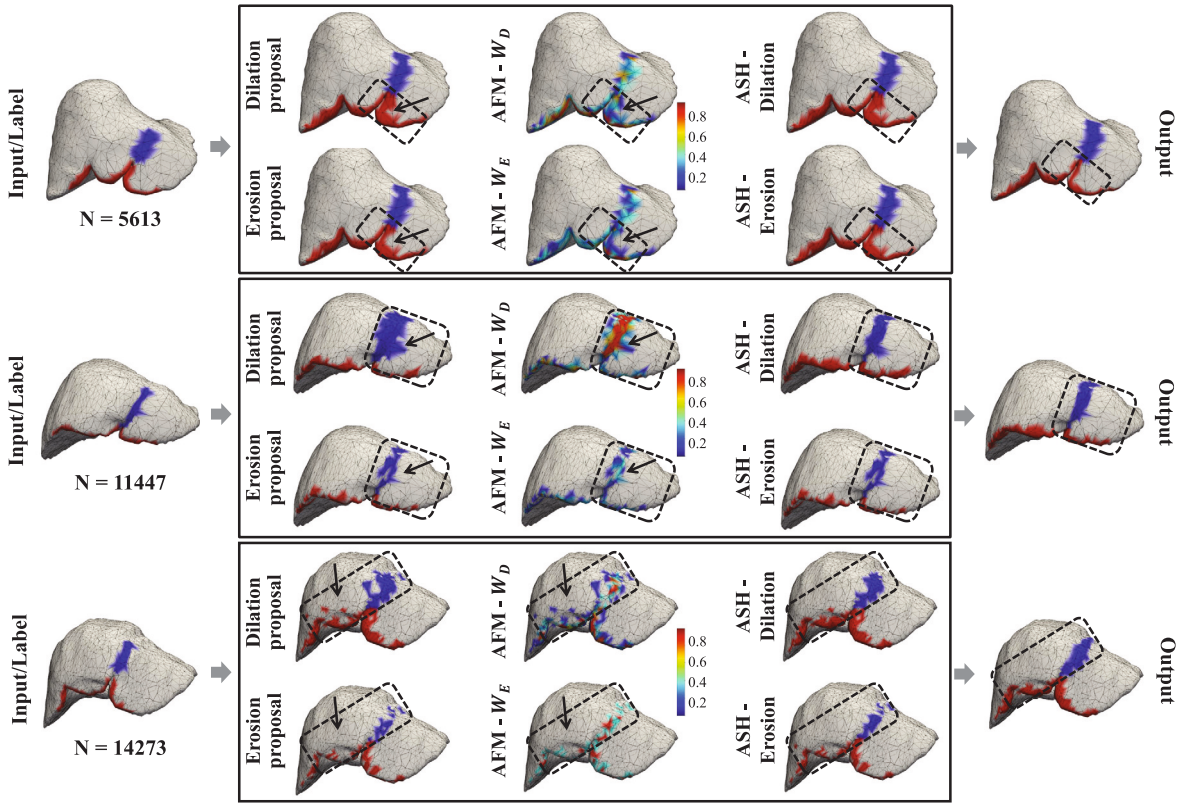

Fig. 8. Visual analysis of the attention fusion module (AFM) and auxiliary segmentation heads (ASH). Three representative liver meshes with varying resolutions are shown in the first column, overlaid with ground truth landmarks. The next two columns show coarse landmark proposals obtained by applying different confidence thresholds ($T_1 < T_2$) to the propagated score maps. AFM columns visualize attention weights from both branches, while ASH outputs illustrate how auxiliary supervision improves proposal quality. The final column shows refined predictions after integration.

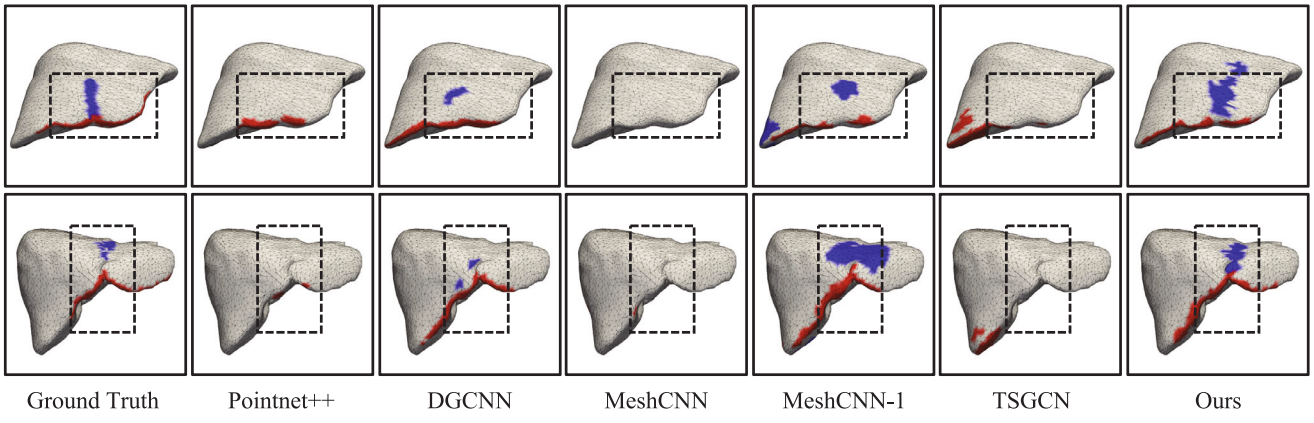

Fig. 9. Generalizability analysis on external dataset — visual results from P2ILF challenge dataset. In addition, the dotted boxes in the figure indicate areas to focus on.

Table 4 presents the quantitative results of the compared methods on the P2ILF data for external testing. Given the characteristics of the P2ILF data, especially the partial annotations, we primarily present the 3D Chamfer Distance, a key official metric of the P2ILF challenge. The results in Table 4 indicate that the basic Pointnet++, DGCNN, and MeshCNN experienced varying degrees of performance decline, highlighting the challenges of applying existing methods to meshes with significant differences in appearance and shape—a common issue in clinical practice. Similarly, TSGCN, which is designed for dental surfaces, failed to achieve effective segmentation in the generalization test on liver meshes. It is noteworthy that although these methods did not perform satisfactorily, DGCNN achieved a relatively low 3D Chamfer Distance of 27.1 mm overall. This may suggest that DGCNN's results contain errors but can still roughly identify the position and orientation of landmarks, producing a coarse segmentation result in external testing. The visual results of DGCNN in Fig. 9 confirm this hypothesis.

Satisfactorily, our method effectively combines the strengths of DGCNN and MeshCNN across different resolution levels, achieving initial segmentation on compressed meshes and refining it through the anatomical refining network, resulting in superior generalization performance. Fig. 9 shows the visual results for two samples from the P2ILF challenge, demonstrating our method's generalizability and adaptability in liver surface landmark extraction.

Table 4
External test results on P2ILF dataset — 3D Chamfer Distance (CD).

	Ligament	Ridge	Overall
Pointnet++	–	60.5 ± 41.8	65.7 ± 43.2
DGCNN	60.2 ± 36.4	24.3 ± 13.3	27.1 ± 12.3
MeshCNN	–	88.5 ± 53.7	80.7 ± 42.9
MeshCNN-1	67.8 ± 46.1	38.3 ± 19.4	42.9 ± 14.3
TSGCN	45.8±0.0[a]	70.5 ± 46.7	74.5 ± 46.3
Ours	**36.3 ± 22.7**	**19.0 ± 12.5**	**18.8 ± 8.6**

"–" indicates that the predicted landmark region is empty, so CD cannot be calculated.
[a] CD is computed only on cases with non-empty predictions; the corresponding prediction rates are given in Table A.2.

Table 5
Efficiency comparison of segmentation methods in terms of Parameters and Inference Time per sample.

	Parameters (M)	Inference time (s/sample)
Pointnet++	1.735	0.107
DGCNN	1.454	**0.007**
MeshCNN	**0.982**	0.493
MeshCNN-1	**0.982**	0.574
TSGCN	4.128	0.083
Ours	1.454 + 0.224	0.267

5. Discussion and conclusion

This study addresses the critical need for effective 3D landmark extraction algorithms for AR-assisted laparoscopic navigation by developing a deep learning-based segmentation algorithm for two critical anatomical landmarks on the liver surface: the falciform ligament and the liver ridge. Our proposed algorithm employs a nested resolution network architecture that combines a Dynamic Graph CNN (DGCNN) and a MeshConv-based anatomical refining network to tackle the challenges posed by the significant variability in shape and appearance of the liver surface. Initially, the liver mesh at its original resolution is pooled into a low-resolution mesh with a fixed number of edges through a random edge collapse strategy, and the first-stage landmark segmentation is generated by the DGCNN model. The landmark segmentation results at low resolution are then propagated to the high-resolution mesh surface using the unpooling operation in MeshCNN, resulting in "dilation" and "erosion" landmark proposals. Subsequently, we design an anatomical refining network that integrates the landmark proposals, edge features from the high-resolution mesh, and resolution encoding. This network, leveraging MeshConv-based specialized branches and an attention fusion module, extracts the correct priors from the previous stage, ultimately achieving accurate landmark segmentation on original meshes of varying resolutions. Empirical evaluations on two liver mesh datasets demonstrate that our framework consistently outperforms existing methods in terms of both accuracy and robustness.

As hepatobiliary surgeries increasingly move toward minimally invasive and precision approaches, the design and development of AR navigation systems for laparoscopic procedures have garnered significant attention. The 3D–2D registration (Sun, 2023) constrained by anatomical landmarks is foundational to realizing this groundbreaking visualization technique. However, most related studies still rely on manual annotation of key anatomical landmarks on 3D liver meshes by surgeons, largely due to the considerable variability in liver appearance and shape, as well as the lack of mesh data annotated with landmarks. This has presented significant challenges for the development of deep learning-based automated landmark segmentation methods. In response, we reconstructed and annotated 200 watertight meshes of varying shapes, appearances, and resolutions from three publicly available liver CT datasets to develop and validate our liver landmark segmentation algorithm. Additionally, we applied the proposed method to the external P2ILF dataset collected from real patients for further validation. The experimental results in Table 4 affirm the superiority of our approach, particularly in the extraction of the falciform ligament, where it reduced the 3D Chamfer distance by approximately 24 mm compared to DGCNN.

Significance of the Falciform Ligament in Surgical Navigation. Accurately identifying surface anatomical landmarks on the liver mesh presents a unique challenge due to the absence of true color and texture information and the subtlety of local geometric cues. In particular, failure to detect key landmarks or predicting fragmented, disconnected regions can directly undermine the reliability of downstream tasks such as 3D–2D registration. To simulate real-world registration performance, we adopted the differentiable PyTorch3D framework (Ravi et al., 2020) following the protocol established by the NCT team – the P2ILF challenge winner (Ali et al., 2025) – to visualize rigid alignment results using our predicted landmarks. Fig. 10 illustrates two representative cases from the P2ILF dataset, comparing three segmentation settings: (i) our method, (ii) a setting with missing ligament prediction, and (iii) a setting with large fragmented ligament segments. In both examples, our model produced continuous and anatomically coherent landmark predictions on high-resolution liver meshes, enabling robust alignment to laparoscopic keyframes—even when minor discontinuities were present in the liver ridge. This is because our predictions consistently contain a dominant, spatially connected region that overlaps reliably with visible anatomy. In contrast, omitting the falciform ligament – a failure mode occasionally observed in local-feature-based models such as MeshCNN – prevents meaningful registration due to the deformation-prone nature of the liver ridge alone. Likewise, predictions containing multiple disconnected landmark segments introduce ambiguity in geometric correspondence, which interferes with the registration optimization and may lead to visibly incorrect alignment—such as spatial distortions or flipped liver meshes. These outcomes are shown in the supplementary videos accompanying Fig. 10. Overall, the falciform ligament offers a more stable surface anchor for intraoperative registration, given its anatomical position and resistance to surgical deformation. Our method's ability to consistently produce contiguous and anatomically valid ligament predictions – even under varying mesh topologies – makes it better suited for real-world AR navigation and clinical integration.

Method Efficiency: Parameters and Inference Time. Compared to the previous labor intensive manual annotation of liver mesh landmarks, the deep learning-based automated segmentation algorithm offers a more efficient solution for handling meshes of varying resolutions. In this study, the proposed nested resolution Mesh-Graph CNN includes a foundational DGCNN and an anatomical refining network composed of MeshConv layers. With convolution channels set to 32, 64, and 64 in the anatomical refining network, the overall parameter count is comparable to that of DGCNN ($\uparrow$ $0.25M$, in Table 5). In terms of inference time, the original MeshCNN's reliance on multiple pooling operations to achieve a global receptive field posed a significant efficiency bottleneck. In contrast, our proposed Mesh-Graph CNN achieves high inference efficiency ($0.493s \rightarrow 0.267s$, in Table 5) by combining global understanding from DGCNN with local refinement via MeshConv, requiring only one pooling and unpooling operation to complete the forward pass. Unlike MeshCNN's iterative edge-collapse strategy, our method adopts random pooling, which may enable faster computation through matrix-based operations. This leads to inference speeds close to DGCNN (0.007 s), while avoiding the sequential bottlenecks of MeshCNN (Table 5). Although efficiency is not the primary focus of this work, we plan to further optimize the design toward real-time intraoperative segmentation in future studies.

Resource Sharing, Limitations, and Future Work. To support future research, we have publicly released a dataset of 200 manually annotated liver surface meshes, along with the training and inference code, at GithubLink. This curated dataset provides a valuable resource for developing and evaluating liver landmark segmentation algorithms, particularly for teams with limited access to clinical mesh data. While

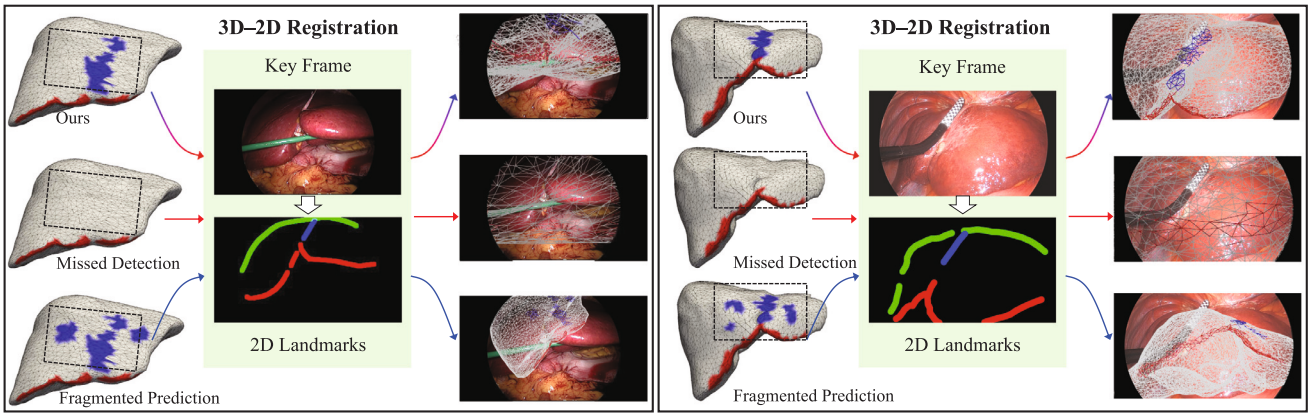

Fig. 10. Visual comparison of 3D–2D registration outcomes across three types of landmark segmentations on two P2ILF cases. Our method yields more complete and spatially coherent predictions for both ligament and ridge, enabling successful overlay, while the other two settings – with missing or fragmented landmark regions – lead to failed or unstable alignment. The green curve indicates the extracted liver silhouette from the laparoscopic key frame. Registration was performed using our PyTorch3D-basedpipeline. Supplementary videos demonstrate the dynamic rigid alignment process.

our framework has shown robust performance across heterogeneous mesh resolutions and reconstruction pipelines, several limitations remain. First, each liver mesh was annotated by a single rater, precluding direct interobserver variability analysis. Given the labor-intensive nature of this task, future work will focus on multi-annotator labeling of representative cases to assess annotation consistency and compare it with automated outputs. Second, although our current formulation relies solely on surface geometry, future extensions may explore incorporating image-derived features (e.g., from CT or MRI) to enrich landmark detection with cross-modal cues. Finally, to further bridge the gap between preoperative modeling and intraoperative deployment, we plan to curate paired datasets combining volumetric scans and laparoscopic views with anatomical landmarks and silhouettes. This will enable refinement of our algorithm's performance in clinically realistic AR navigation scenarios.

CRediT authorship contribution statement

Xukun Zhang: Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Jinghui Feng:** Writing – original draft, Validation. **Peng Liu:** Validation, Formal analysis. **Minghao Han:** Validation, Formal analysis. **Yanlan Kang:** Validation, Formal analysis. **Jingyi Zhu:** Validation, Formal analysis. **Le Wang:** Software, Data curation. **Xiaoying Wang:** Supervision, Data curation. **Sharib Ali:** Writing – review & editing, Supervision, Methodology. **Lihua Zhang:** Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project was funded by the National Natural Science Foundation of China (82090052, 82090054, 82001917 and 81930053), the Clinical Research Plan of Shanghai Hospital Development Center (No. 2020CR3004A), and the National Key Research and Development Program of China under Grant (2021YFC2500402). The work was also supported by the Engineering and Physical Sciences Research Council [grant number UKRI914]; and by the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Thanks to the pump-priming funding initiated by the NIHR Leeds BRC.

Table A.1
Prediction rates (%) for the falciform ligament landmark on the internal test set (Number of test cases = 70). Prediction rate is defined as the percentage of test cases with non-empty predictions.

Method	Prediction rate (%)
Pointnet++	11.4
MeshCNN	18.6
TSGCN	22.9

Table A.2
Prediction rates (%) for the falciform ligament landmark on the P2ILF data (9 cases used for external evaluation). Prediction rate is defined as the percentage of test cases with non-empty predictions.

Method	Prediction rate (%)
Pointnet++	0.0
MeshCNN	0.0
TSGCN	11.1

Appendix. Prediction rates

See Tables A.1 and A.2

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2025.103825.

Data availability

The 200 manually annotated liver meshes used in this study, along with a concise annotation instruction document, are available at: DatasetLink.

References

Adagolodjo, Y., Trivisonne, R., Haouchine, N., Cotin, S., Courtecuisse, H., 2017. Silhouette-based pose estimation for deformable organs application to surgical augmented reality. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 539–544.

Ali, S., Espinel, Y., Jin, Y., Liu, P., Güttner, B., Zhang, X., Zhang, L., Dowrick, T., Clarkson, M.J., Xiao, S., et al., 2025. An objective comparison of methods for augmented reality in laparoscopic liver resection by preoperative-to-intraoperative image fusion from the MICCAI2022 challenge. Med. Image Anal. 99, 103371.

Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., et al., 2023. The liver tumor segmentation benchmark (lits). Med. Image Anal. 84, 102680.

Chen, X., Ma, H., Wan, J., Li, B., Xia, T., 2017. Multi-view 3d object detection network for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1907–1915.

Chen, G., Qin, J., Amor, B.B., Zhou, W., Dai, H., Zhou, T., Huang, H., Shao, L., 2023. Automatic detection of tooth-gingiva trim lines on Dental Surfaces. IEEE Trans. Med. Imaging 42 (11), 3194–3204.

Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G., et al., 2008. Meshlab: an open-source mesh processing tool. In: Eurographics Italian Chapter Conference, vol.2008, Salerno, Italy, pp. 129–136.

Dai, A., Nießner, M., 2018. 3Dmv: Joint 3d-multi-view prediction for 3d semantic scene segmentation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 452–468.

Espinel, Y., Calvet, L., Botros, K., Buc, E., Tilmant, C., Bartoli, A., 2021. Using multiple images and contours for deformable 3d-2d registration of a preoperative ct in laparoscopic liver surgery. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24. Springer, pp. 657–666.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al., 2012. 3D slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging 30 (9), 1323–1341.

Graham, B., Engelcke, M., Van Der Maaten, L., 2018. 3D semantic segmentation with submanifold sparse convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9224–9232.

Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., Cohen-Or, D., 2019. MeshCNN: a network with an edge. ACM Trans. Graph. 38 (4), 90:1–90:12.

Ji, Y., Bai, H., G.E., C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., Luo, P., 2022. AMOS: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (Eds.), In: Advances in Neural Information Processing Systems, vol. 35, Curran Associates, Inc., pp. 36722–36732.

Jiang, L., Zhao, H., Liu, S., Shen, X., Fu, C.W., Jia, J., 2019. Hierarchical point-edge interaction network for point cloud semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10433–10441.

Kipf, T.N., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al., 2023. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026.

Koo, B., Ozgur, E., Roy, B.L., Buc, E., Bartoli, A., 2017. Deformable registration of a preoperative 3D liver volume to a laparoscopy image using contour and shading cues. In: Descoteaux, M., Maier-Hein, L., Franz, A.M., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I. In: Lecture Notes in Computer Science, vol. 10433, Springer, pp. 326–334.

Koo, B., Robu, M.R., Allam, M., Pfeiffer, M., Thompson, S.A., Gurusamy, K., Davidson, B.R., Speidel, S., Hawkes, D.J., Stoyanov, D., Clarkson, M.J., 2022. Automatic, global registration in laparoscopic liver surgery. Int. J. Comput. Assist. Radiol. Surg. 17 (1), 167–176.

Labrunie, M., Pizarro, D., Tilmant, C., Bartoli, A., 2023. Automatic 3D/2D deformable registration in minimally invasive liver resection using a mesh recovery network. In: MIDL. pp. 1104–1123.

Labrunie, M., Ribeiro, M., Mourthadhoi, F., Tilmant, C., Le Roy, B., Buc, E., Bartoli, A., 2022. Automatic preoperative 3d model registration in laparoscopic liver resection. Int. J. Comput. Assist. Radiol. Surg. 17 (8), 1429–1436.

Le, T., Bui, G., Duan, Y., 2017. A multi-view recurrent neural network for 3D mesh segmentation. Comput. Graph..

Le, T., Duan, Y., 2018. Pointgrid: A deep network for 3d shape understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9204–9214.

Liang, Z., Yang, M., Li, H., Wang, C., 2020. 3D instance embedding learning with a structure-aware loss function for point cloud segmentation. IEEE Robot. Autom. Lett. 5 (3), 4915–4922.

Liu, Z., Tang, H., Lin, Y., Han, S., 2019. Point-voxel cnn for efficient 3d deep learning. In: Advances in neural information processing systems, vol. 32.

Lopez, Y.E., 2022. Automatic Registration of CT/MRI Data on Laparoscopic Liver Images for Surgical Gesture Assistance by Augmented Reality (Ph.D. thesis). Université Clermont Auvergne.

Mhiri, I., Pizarro, D., Bartoli, A., 2024. Neural patient-specific 3D–2D registration in laparoscopic liver resection. Int. J. Comput. Assist. Radiol. Surg. 1–8.

Modrzejewski, R., Collins, T., Seeliger, B., Bartoli, A., Hostettler, A., Marescaux, J., 2019. An in vivo porcine dataset and evaluation methodology to measure soft-body laparoscopic liver registration accuracy with an extended algorithm that handles collisions. Int. J. Comput. Assist. Radiol. Surg. 14, 1237–1245.

Oya, T., Kadomatsu, Y., Chen-Yoshikawa, T.F., Nakao, M., 2024. 2D/3D deformable registration for endoscopic camera images using self-supervised offline learning of intraoperative pneumothorax deformation. Comput. Med. Imaging Graph. 116, 102418.

Pang, G., Neumann, U., 2016. 3D point cloud object detection with multi-view convolutional neural network. In: 2016 23rd International Conference on Pattern Recognition. ICPR, IEEE, pp. 585–590.

Pei, J., Cui, R., Li, Y., Si, W., Qin, J., Heng, P.A., 2024. Depth-driven geometric prompt learning for laparoscopic liver landmark detection. arXiv preprint arXiv:2406.17858.

Pfeiffer, M., Kenngott, H., Preukschas, A., Huber, M., Bettscheider, L., Müller-Stich, B., Speidel, S., 2018. IMHOTEP: virtual reality framework for surgical applications. Int. J. Comput. Assist. Radiol. Surg. 13, 741–748.

Plantefeve, R., Peterlik, I., Haouchine, N., Cotin, S., 2016. Patient-specific biomechanical modeling for guidance during minimally-invasive hepatic surgery. Ann. Biomed. Eng. 44, 139–153.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 652–660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA. pp. 5099–5108.

Ravi, N., Reizenstein, J., Novotny, D., Gordon, T., Lo, W.Y., Johnson, J., Gkioxari, G., 2020. Accelerating 3d deep learning with pytorch3d. arXiv preprint arXiv:2007.08501.

Riegler, G., Osman Ulusoy, A., Geiger, A., 2017. Octnet: Learning deep 3d representations at high resolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3577–3586.

Robu, M.R., Ramalhinho, J., Thompson, S.A., Gurusamy, K., Davidson, B.R., Hawkes, D.J., Stoyanov, D., Clarkson, M.J., 2018. Global rigid registration of CT to video in laparoscopic liver surgery. Int. J. Comput. Assist. Radiol. Surg. 13 (6), 947–956.

Schneider, L., Niemann, A., Beuing, O., Preim, B., Saalfeld, S., 2021. MedmeshCNN-enabling meshcnn for medical surface models. Comput. Methods Programs Biomed. 210, 106372.

Soler, L., Hostettler, A., Agnus, V., Charnoz, A., Fasquel, J., Moreau, J., Osswald, A., Bouhadjar, M., Marescaux, J., 2010. 3D Image Reconstruction for Comparison of Algorithm Database: A Patient Specific Anatomical and Medical Image Database. Tech. Rep 1, (1), IRCAD, Strasbourg, France.

Sun, H., 2023. A review of 3D-2D registration methods and applications based on medical images. Highlights Sci. Eng. Technol. 35, 200–224.

Wang, L., Huang, Y., Hou, Y., Zhang, S., Shan, J., 2019a. Graph attention convolution for point cloud semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10296–10305.

Wang, Z., Lu, F., 2019. Voxsegnet: Volumetric cnns for semantic part segmentation of 3d shapes. IEEE Trans. Vis. Comput. Graphics 26 (9), 2919–2930.

Wang, C., Samari, B., Siddiqi, K., 2018. Local spectral graph convolution for point set feature learning. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 52–66.

Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019b. Dynamic graph cnn for learning on point clouds. ACM Trans. Graph. (Tog) 38 (5), 1–12.

Wu, T., Pan, L., Zhang, J., Wang, T., Liu, Z., Lin, D., 2021. Balanced chamfer distance as a comprehensive metric for point cloud completion. In: Advances in Neural Information Processing Systems, vol. 34, pp. 29088–29100.

Wu, W., Qi, Z., Fuxin, L., 2019. Pointconv: Deep convolutional networks on 3d point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9621–9630.

Xie, Z., Chen, J., Peng, B., 2020. Point clouds learning with attention-based graph convolution networks. Neurocomputing 402, 245–255.

Xu, M., Zhou, Z., Qiao, Y., 2020. Geometry sharing network for 3d point cloud classification and segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, (07), pp. 12500–12507.

Zhao, Y., Zhang, L., Liu, Y., Meng, D., Cui, Z., Gao, C., Gao, X., Lian, C., Shen, D., 2021. Two-stream graph convolutional network for intra-oral scanner image segmentation. IEEE Trans. Med. Imaging 41 (4), 826–835.

Zhong, J.-H., Peng, N.-F., You, X.-M., Ma, L., Xiang, X., Wang, Y.-Y., Gong, W.-F., Wu, F.-X., Xiang, B.-D., Li, L.-Q., 2017. Tumor stage and primary treatment of hepatocellular carcinoma at a large tertiary hospital in China: A real-world study. Oncotarget 8 (11), 18296.