

Paper List

CVPR2021

3D Graph Anatomy Geometry-Integrated Network for Pancreatic Mass Segmentation, Diagnosis, and Quantitative Patient Management

MICCAI2020

Voxel2Mesh: 3D Mesh Model Generation from Volumetric Data

Information

3D Graph Anatomy Geometry-Integrated Network for Pancreatic Mass Segmentation, Diagnosis, and Quantitative Patient Management

Tianyi Zhao^{*1,2}, Kai Cao^{*3}, Jiawen Yao¹, Isabella Nogues⁴, Le Lu¹, Lingyun Huang⁵, Jing Xiao⁵,
Zhaozheng Yin², Ling Zhang¹

¹PAII Inc., ²Stony Brook University, ³Changhai Hospital, ⁴Harvard University, ⁵Ping An Technology

Introduction

- Pancreatic cancer is the third leading cause of cancer-related deaths in the United States
- It has the poorest prognosis among all solid malignancies, with a 5-year survival rate ~ 10%
- Early diagnosis is crucial, as it can potentially increase the 5-year survival rate to ~ 50%
- Pancreatic masses often cannot be reached precisely via needle biopsy
- Therefore, reliable imaging-based diagnosis is critical to **patient management** in a timely fashion, while avoiding unnecessary iatrogenic morbidity

Cancer Risk of Masses (Tumors or Cysts)

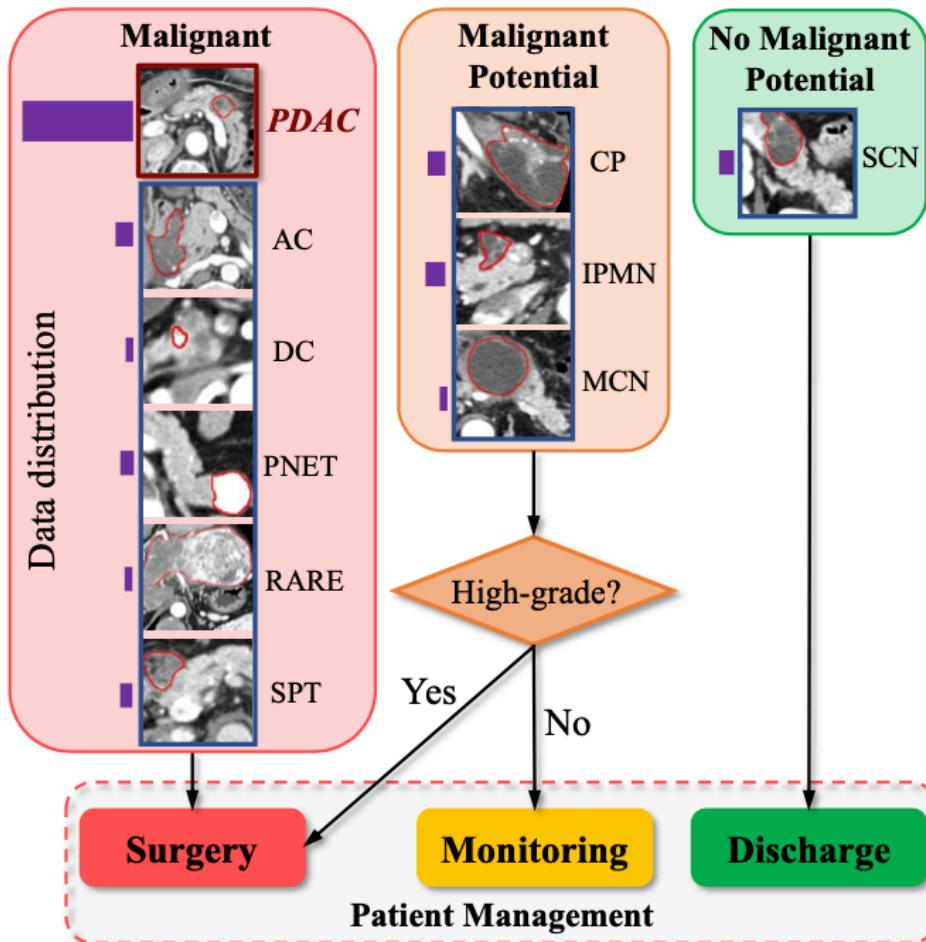


Figure 1. Disease taxonomy of the ten types of pancreatic masses (tumors, cysts). Mass type diagnosis determines the clinical malignancy indications that lead to the proper patient risk stratification and management. The purple histogram bars represent their relative frequencies. All images are in the arterial-late CT phase.

Difficulties

1. The same type of mass may appear in different textures, shapes, contrasts, and different enhancement patterns across CT phases.
2. Pancreatic ductal adenocarcinoma (PDAC) accounts for most cases in pancreatic cancer specialized hospitals, causing a long-tail distribution.
3. Masses, at times, are surrounded by inflamed tissues and thus cannot be easily identified.

Methodology

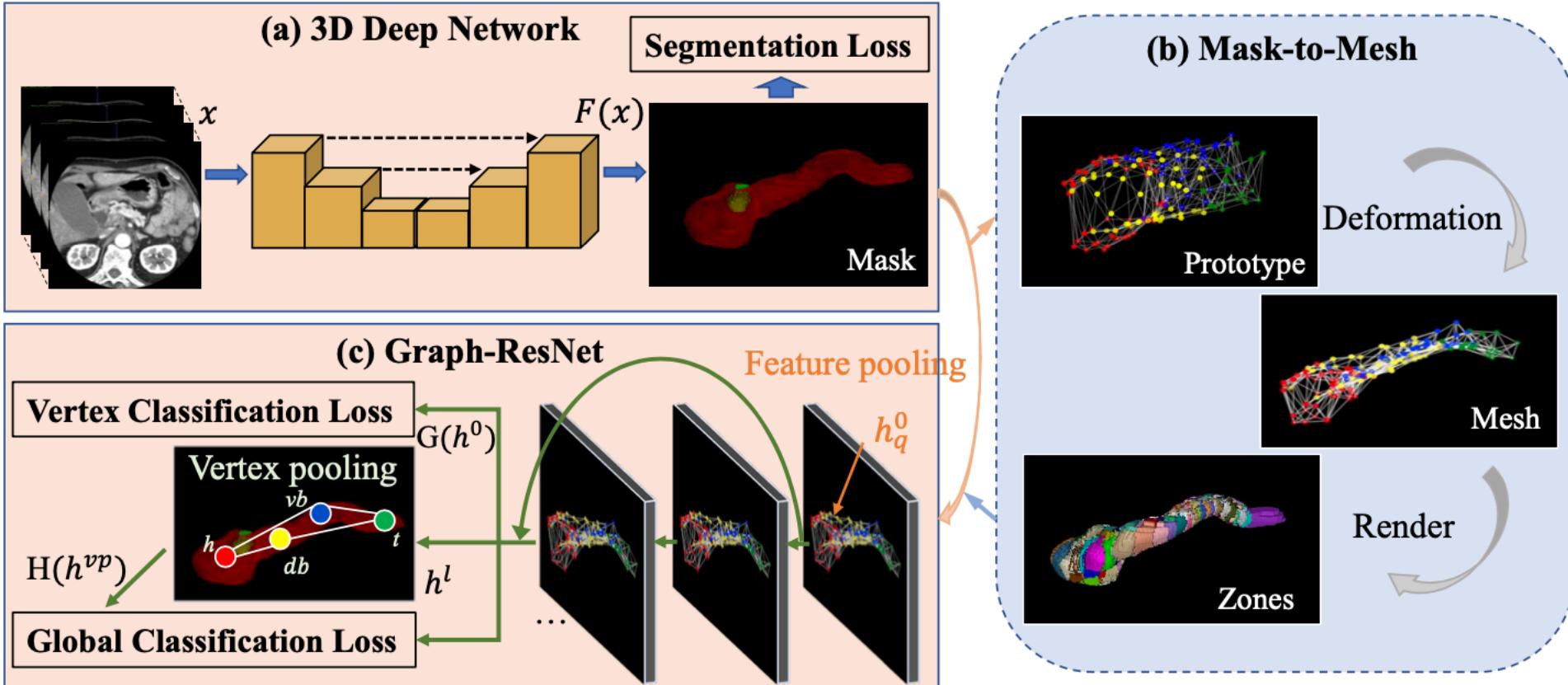
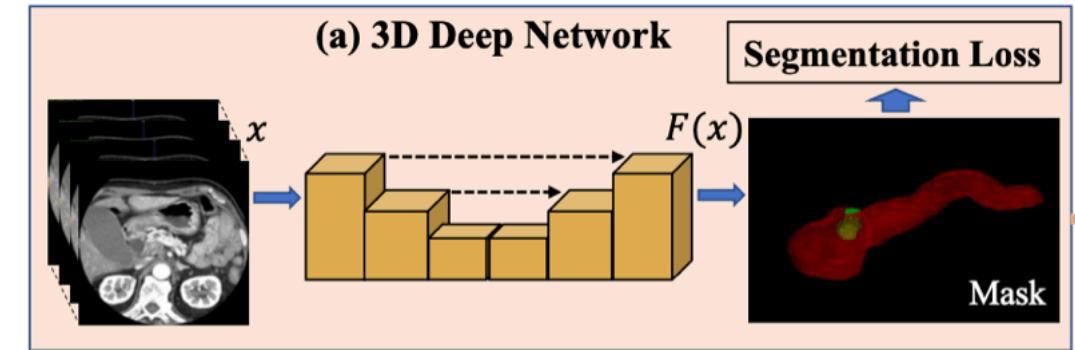


Figure 2. Flowchart of our proposed Segmentation-Mesh-Classification Network (SMCN). SMCN has three components: the pancreas-mass segmentation network, the mask-to-mesh 3D anatomy modeling, and the global mass classification network. The mesh model is the bridge between the segmentation network and the classification network, who pools the features from the segmentation network into the vertex feature vectors in the graph classification network.

Methodology

- Anatomy-Mass Segmentation Network

1. Input: multi(5)-phase CT scans, $X \in R^{5 \times W \times H \times D}$
2. Output: pancreas & masses mask,
3. Weak-supervised training: nonPDAC labels are usually lacking
4. Loss:



$$\begin{aligned} L_{CE} &= - \sum_{w,h,d} \sum_k^K y_{k,w,h,d} \log(F(x)_{k,w,h,d}), \\ L_{DC} &= -2 \sum_k^K \frac{\sum_{w,h,d} F(x)_{k,w,h,d} y_{k,w,h,d}}{\sum_{w,h,d} F(x)_{k,w,h,d} + \sum_{w,h,d} y_{k,w,h,d}}, \\ L_{Seg} &= L_{CE} + L_{DC}. \end{aligned} \tag{1}$$

Methodology

- 3D Mesh-based Anatomy Representation

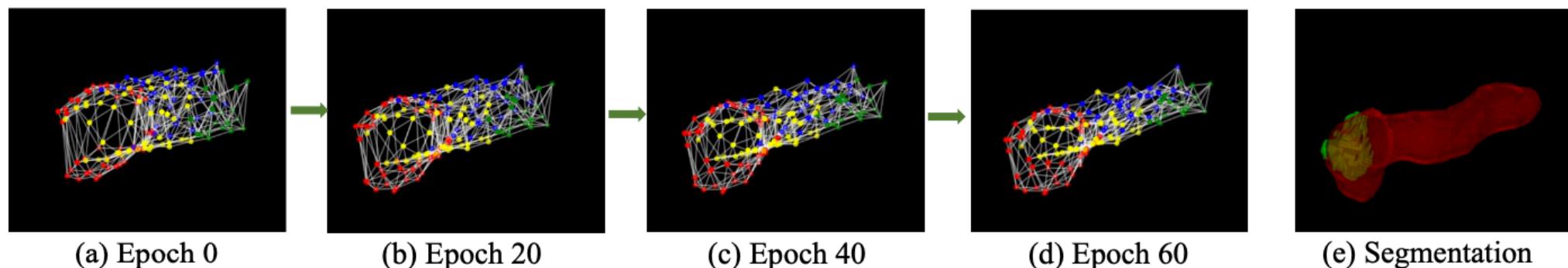


Figure 3. The 3D mesh deformation process. The mesh is initialized as the pancreas prototype (a). Given the pancreas-mass segmentation mask (e), the geometry of the mesh is deformed gradually to fit the surface of the segmentation mask, as in (b-d). Red: the head of pancreas; Blue: the ventral body of pancreas; Yellow: the dorsal body of pancreas; Green: the tail of pancreas.

Methodology

- 3D Mesh-based Anatomy Representation
 1. **Pancreas Anatomy:** Fig3(a).
 - create a prototype mesh based on the average pancreas shape from the training fold
 - 1 - 48th vertices: head, red
 - 49 - 90th vertices: ventral body, blue
 - 91-135th vertices: dorsal body, yellow
 - 136- 156th vertices: tail, green

Methodology

- 3D Mesh-based Anatomy Representation

1. Pancreas Anatomy

2. Mask-to-mesh Process: Fig. 3(a) - (d)

define p: the vertex in the mesh, q: the voxels of the surface in the segmentation mask

Point Loss: $L_{pt} = \sum_p \min_q \|p - q\|_2^2.$

First edge regularization term : preserve the geometry of the mesh

$$L_{e1} = \sum_e \|e - \text{mean}(e)\|_2^2, \quad e = \|p - p'\|_2, \quad p' \in N(p)$$

Second edge regularization losssimply minimize the edge length

$$L_{e2} = \sum_e e$$

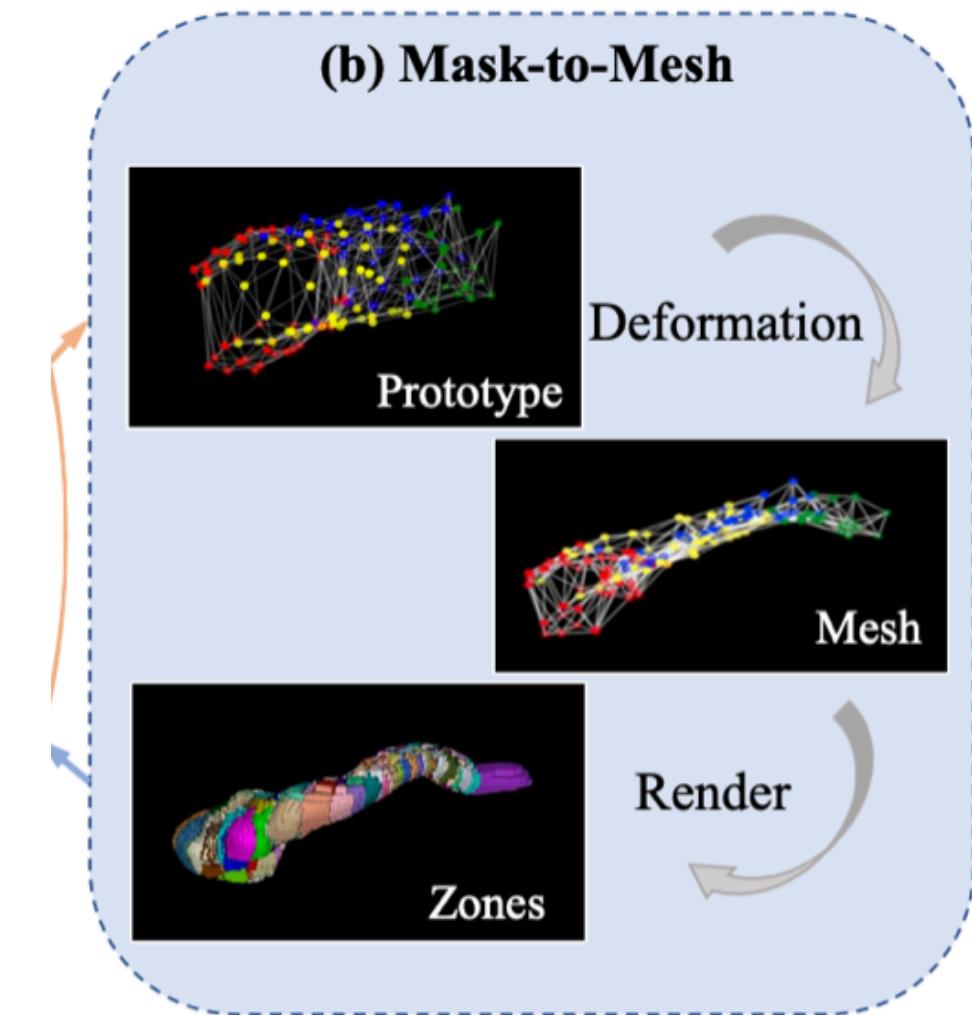
Finally,

$$L_{meshfit} = L_{pt} + \lambda_1 L_{e1} + \lambda_2 L_{e2}$$

Methodology

- 3D Mesh-based Anatomy Representation
 1. Pancreas Anatomy
 2. Mask-to-mesh Process
 3. **Rendering:**

Each voxel of the segmented pancreas volume is defined in a zone $[Z(p)]$ by its nearest vertex



Methodology

- Global Mass Classification Network

1. Vertex Feature Pooling:

Local feature vector: a list of

$$F(x)_{w,h,d} \in R^K, \text{ where } (w, h, d) \in Z(p).$$

Global feature vector: Global average pooling of $F(x)$

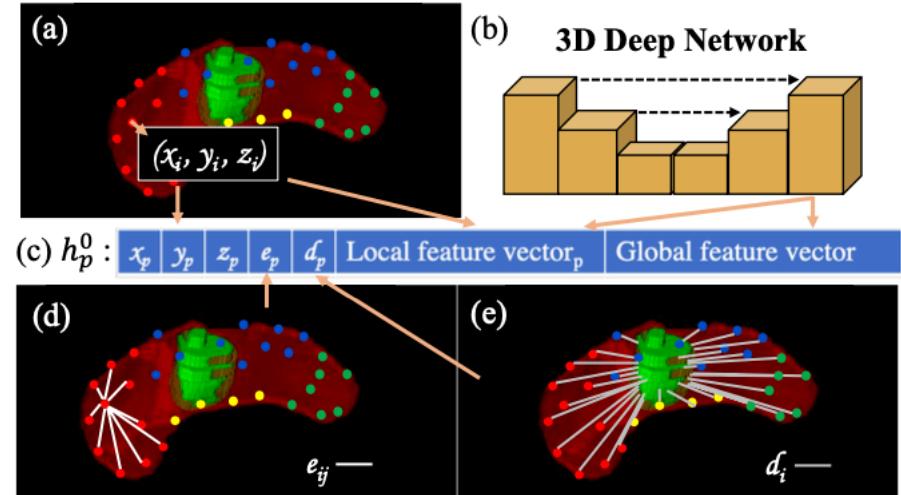


Figure 4. Illustration of the feature vector h_i for the i th vertex. x_i, y_i, z_i represents the location of the i th vertex. e_i is the average of all edge lengths e_{ij} (i.e., the distance from every neighboring vertex j of to i). d_i denotes the shortest distance from the i th vertex to the tumor surface.

Methodology

- Global Mass Classification Network

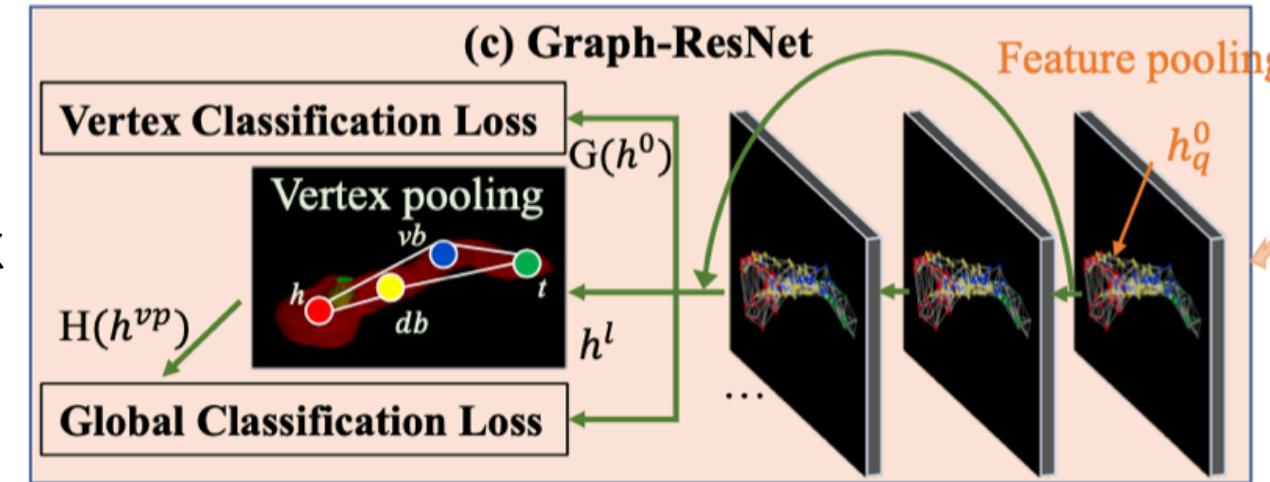
1. Vertex Feature Pooling
2. Graph-ResNet[1]:

- A smaller network than original one: 6 graph convolutional layers and shortcut connections between every 2 layers
- Each graph-based convolutional layer is defined as:

$$h_p^{l+1} = w_0 h_p^l + \sum_{p' \in N(p)} w_1 h_{p'}^l$$

- The vertex classification loss:

$$L_{Vertex} = - \sum_p \sum_k^K y_{k,p}^v \log(G(h^0)_{k,p}),$$



Methodology

- Global Mass Classification Network
 1. Vertex Feature Pooling
 2. Graph-ResNet
 3. Global Mass Classification
 1. Pool four features from all four pancreatic regions
 2. 4 global feature vectors and 156 local feature vectors are concatenated into one vector and fed into the final mass classification layer (FC)

Global Classification Loss :

$$L_{Global} = - \sum_k^K y_k^g \log(H(h^{vp})_k),$$

Methodology

- Anatomy-Mass Segmentation Network
- 3D Mesh-based Anatomy Representation
- Global Mass Classification Network
 1. Vertex Feature Pooling
 2. Graph-ResNet
 3. Global Mass Classification
- The overall loss function:

$$L = L_{Seg} + \eta_1 L_{Vertex} + \eta_2 L_{Global}.$$

Dataset

- 661 patients with surgical pathology-confirmed pancreatic masses:
 - 366 PDACs, 46 ACs, 12 DCs, 35 PNETs, 13 RAREs, 32 SPTs, 43 CPs, 61 IPMNs, 7 MCNs, and 46 SCNs
 - Each patient has 5-phase CT scans:
 - non-contrast (NC), arterial- early (AE), arterial-late (AL), venous (V), and delay (D).
 - The median voxel size is $0.419 \times 0.419 \times 3\text{mm}$

Results

- Visualization

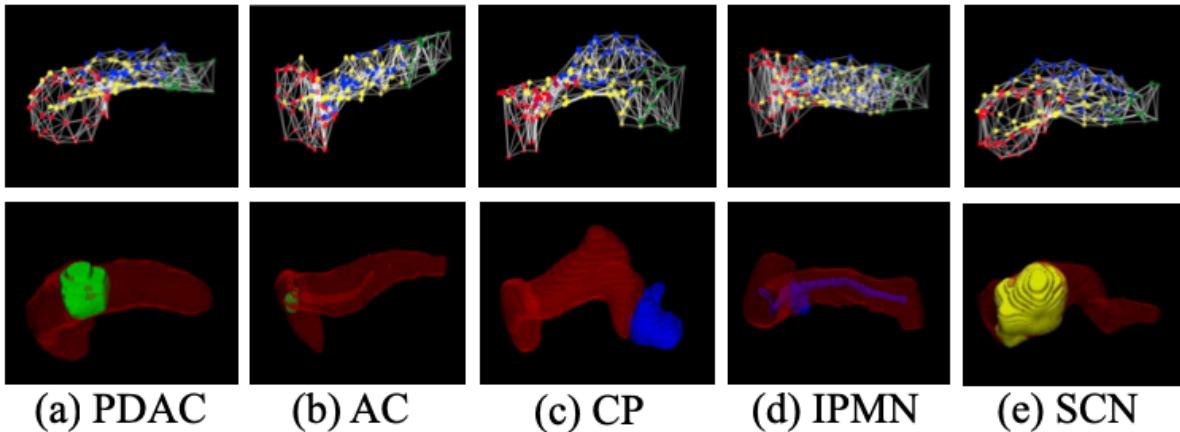


Figure 6. Examples of 3D meshes generated from pancreas segmentation under various scenarios. **Top:** red denotes the head of pancreas; blue the ventral body; yellow the dorsal body; and green the tail. **Bottom:** light red depicts the pancreas; green the surgery tumor; blue the monitoring tumor; and yellow the discharge tumor.

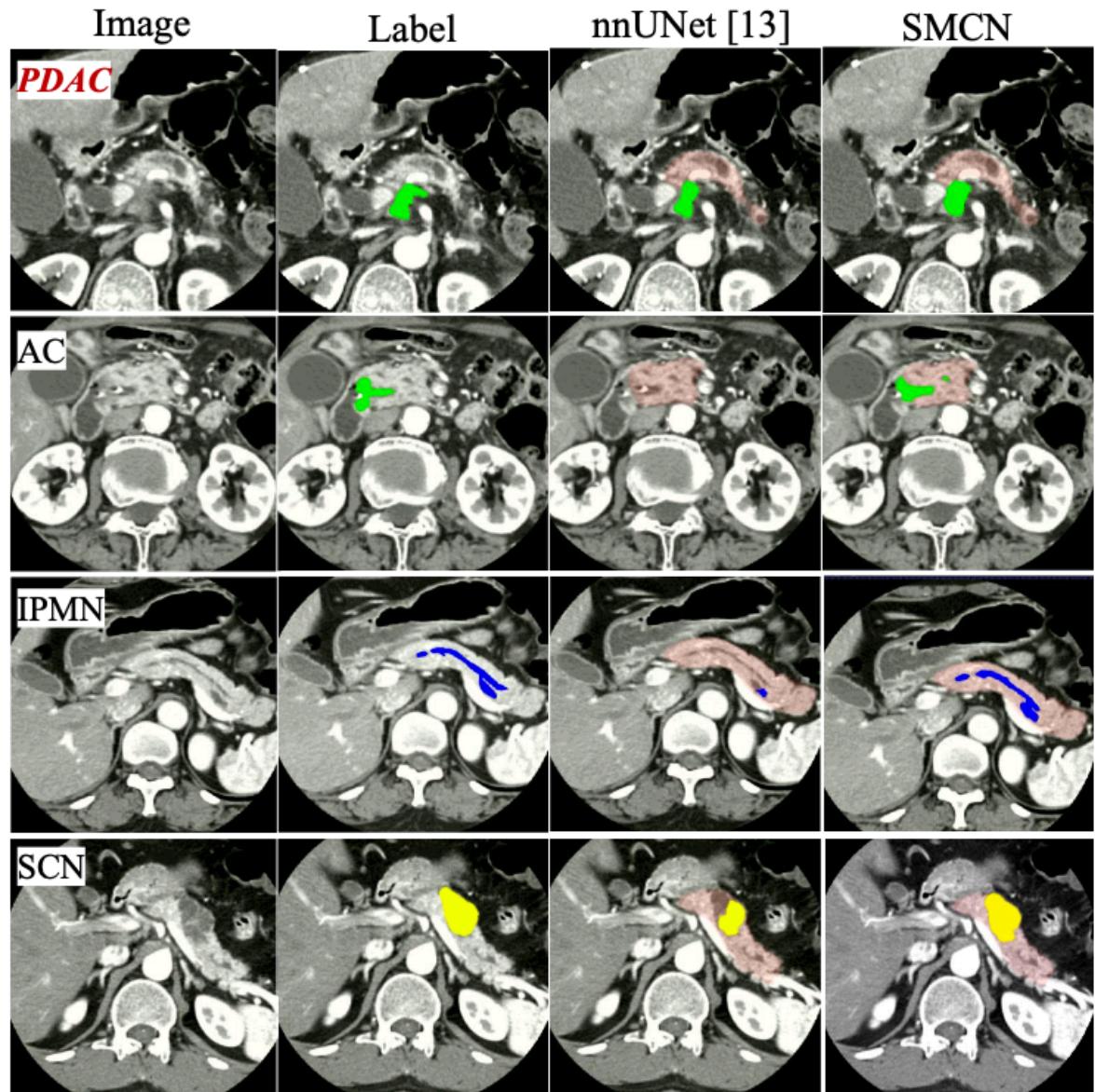


Figure 5. Examples of pancreas-mass segmentation results. Light red: pancreas. Green: surgery mass. Blue: monitoring mass. Yellow: discharge mass.

Results

Table 1. Segmentation and detection results over the ten types of pancreatic masses. Note that in this experiment, networks are trained with four labels: background, pancreas, PDAC, and nonPDAC. micro: result for all patients; macro: average of the metrics of the ten classes.

	Metric	PDAC	NonPDAC	AC	CP	DC	IPMN	MCN	PNET	RARE	SCN	SPT	micro	macro
nnUNet[13]	Average Dice	0.734	0.478	0.423	0.533	0.001	0.151	0.924	0.514	0.849	0.585	0.762	0.618	0.548
	Detection rate	0.959	0.693	0.7	0.778	0.0	0.308	1.0	0.833	1.0	0.8	1.0	0.838	0.738
(Ours)	Average Dice	0.738	0.611	0.438	0.668	0.006	0.602	0.924	0.514	0.824	0.666	0.801	0.681	0.618
	Detection rate	0.972	0.887	0.8	0.889	0.0	1.0	1.0	1.0	1.0	0.8	1.0	0.934	0.846

Table 2. PDAC segmentation under different CT phases: non-contrast (N), arterial (A), arterial-early (AE), arterial-late (AL), venous (V), and delay (D).

Methods	CT Phases	Average Dice	
		Reported from their original papers	
Zhang et al. [32]	N+A+V	0.709	
DDT [23]	V	0.634	
Xia et al. [27]	A+V	0.644	
Ours	AE	0.646	
Ours	AL	0.730	
Ours	V	0.675	
Ours	D	0.562	
Ours	N+AE+V+D	0.696	
Ours	N+AE+AL+V+D	0.738	

Table 3. Average values of PDAC vs. nonPDAC across 3 folds. PV: pixel voting; VV: vertices voting; GC: global classification.

Methods	Accuracy	Sensitivity	Specificity
Radiomics [21]	0.857	0.853	0.868
DeepTexture [28]	0.770	0.827	0.697
ResNet3D [9]	0.720	0.788	0.633
SMCN w PV	0.913	0.945	0.875
SMCN w VV	0.920	0.945	0.891
SMCN w GC	0.927	0.945	0.906
Expert radiologists	-	~0.94	~0.90

Results

Table 4. Classification confusion matrix of quantitative patient management and comparison our results against the clinical test [20].

CompCyst [20]	Discharge	Monitoring	Surgery	Ours	Discharge	Monitoring	Surgery
Discharge (n=53)	32 (60%)	14 (26%)	7 (13%)	(n=46)	28 (61%)	10 (22%)	8 (17%)
Monitor (n=140)	1 (1%)	68 (49%)	71 (51%)	(n=104)	11 (11%)	65 (63%)	27 (26%)
Surgery (n=152)	0 (0%)	14 (9%)	138 (91%)	(n=440)	6 (1%)	16 (4%)	418 (95%)

Table 5. Experimental Results for Patient Management. PV: pixel voting; VV: vertex-based voting; GC: global classification.

Methods	Accuracy	Average Recall
Radiomics [21]	0.854	0.667
DeepTexture [28]	0.743	0.557
ResNet3D [9]	0.737	0.607
SMCN w PV	0.832	0.735
SMCN w VV	0.839	0.656
SMCN w GC	0.865	0.746

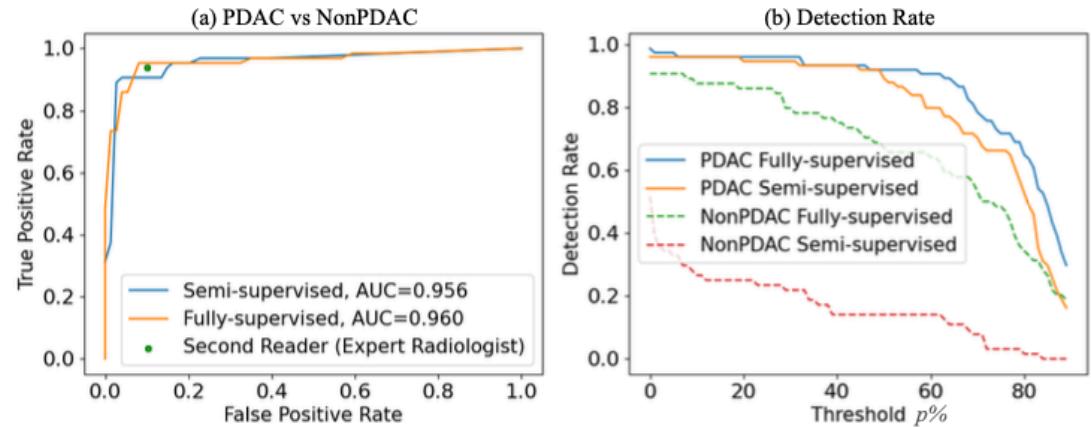


Figure 7. Comparison of (a) classification and (b) detection performances between fully-supervised and semi-supervised learning. The green dot depicts the performance of the mean second reader.

References

1. Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

Information

Voxel2Mesh: 3D Mesh Model Generation from Volumetric Data

Udaranga Wickramasinghe¹, Edoardo Remelli¹,
Graham Knott², and Pascal Fua¹

¹ Computer Vision Laboratory, École Polytechnique Fédérale de Lausanne,
Switzerland

udaranga.wickramasinghe@epfl.ch

² BioEM Laboratory, École Polytechnique Fédérale de Lausanne, Switzerland

Introduction

- Task: Volumetric images —> 3D surface meshes

Architecture

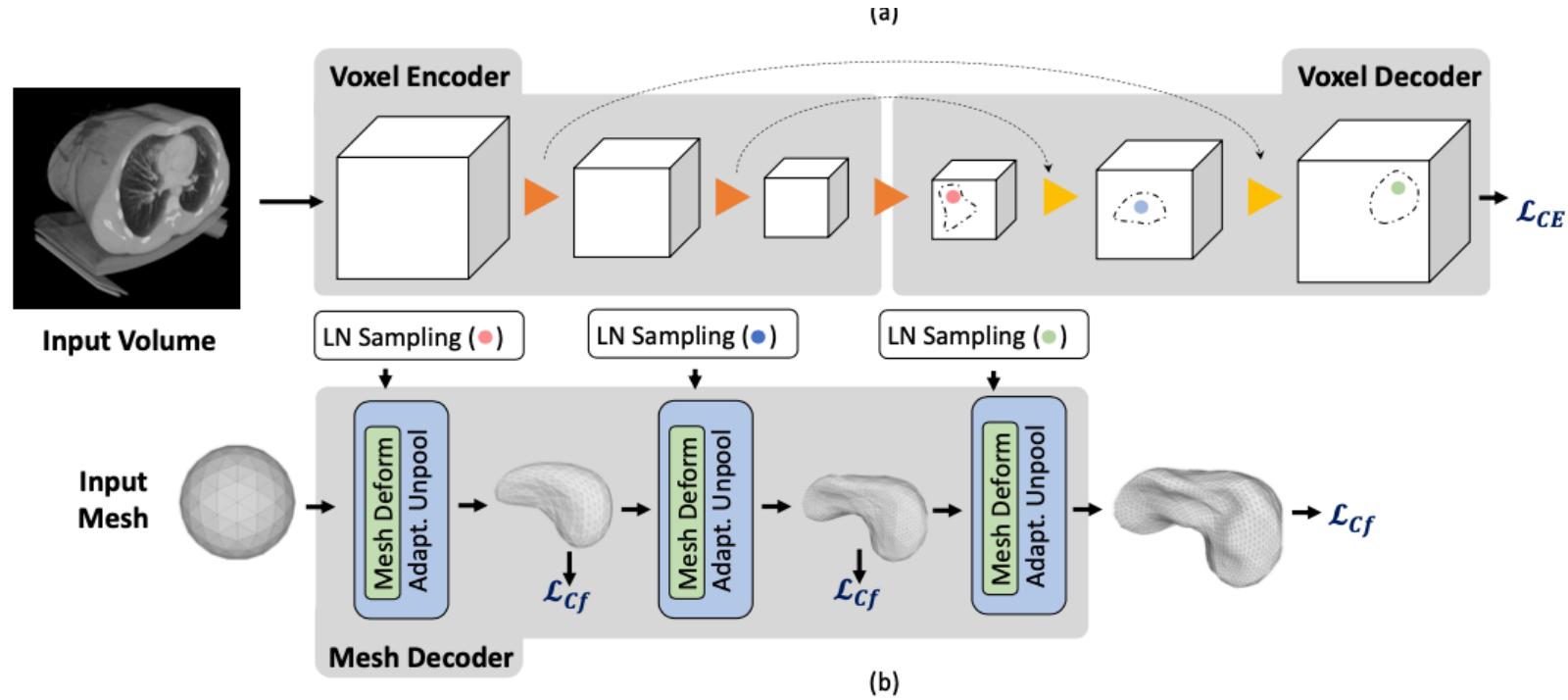


Fig. 1. Architectures (a) The **Pixel2Mesh-3D** architecture, a straightforward extension of [20], uses a surface decoder but no voxel decoder. (b) By contrast, our **Voxel2Mesh** architecture takes as input an image and spherical mesh. They are jointly encoded and then decoded into cubes and meshes of increasing resolution. At each mesh decoding stage, the decoder first receives as input the current mesh and a set of features sampled from the cube of corresponding resolution. Then the mesh is deformed and refined non-uniformly by adding vertices only where they are needed.

Architecture

- Learned Neighborhood Sampling (LNS):

- learns optimum sampling locations
- Avoid over-fit in smaller dataset

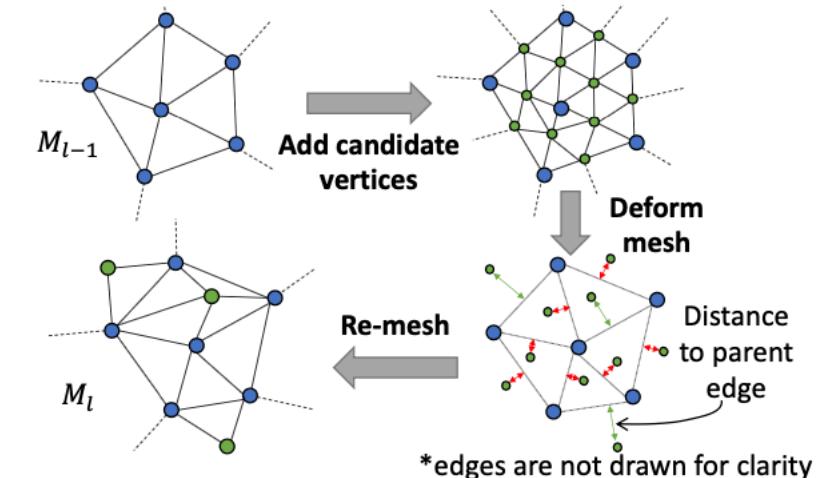
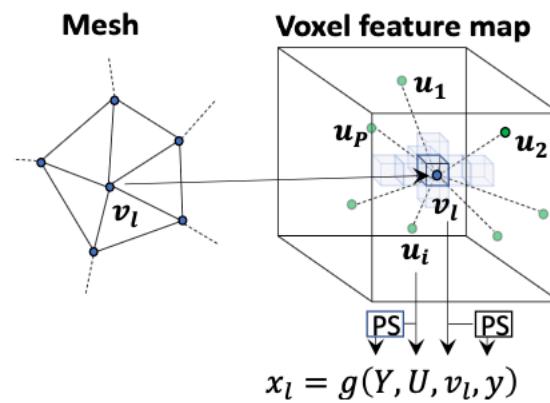


Fig. 2. Approach.(a) Learned Neighborhood Sampling. (b) Adaptive Mesh Unpooling.

- Adaptive Mesh Unpooling (AMU):

- Require densely sampled mesh vertices in high-curvature areas
- Eliminates the need for exponentially large amounts of memory that uniform unpooling(UMU) requires.

Results

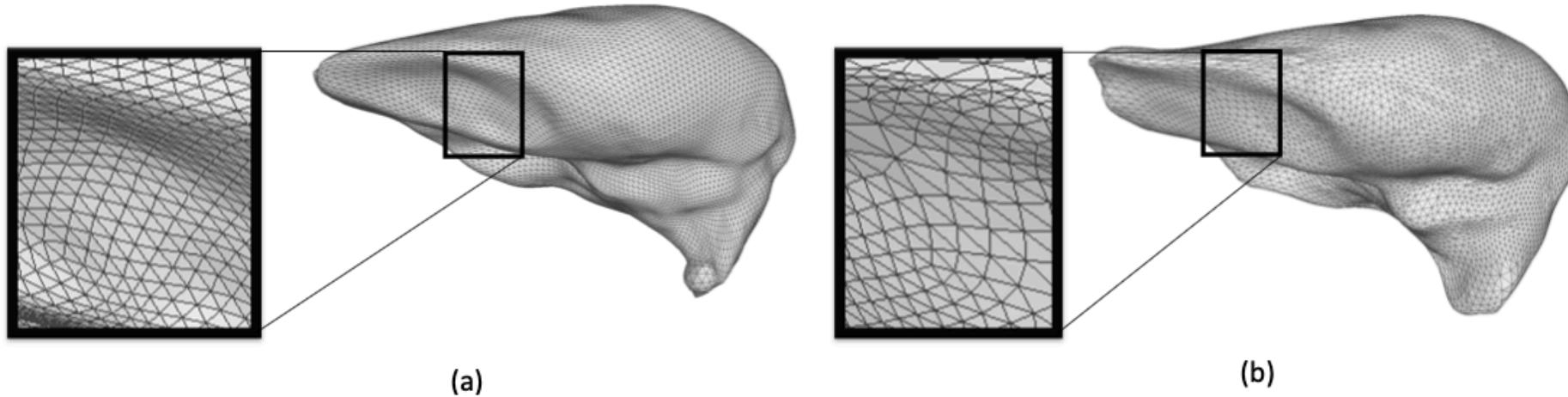


Fig. 3. Levels of resolution. (a) **Pixel2Mesh-3D** result (10422 vertices). (b) **Voxel2Mesh** result (7498 vertices). With our adaptative unpooling, we obtain better results with fewer vertices.

Results

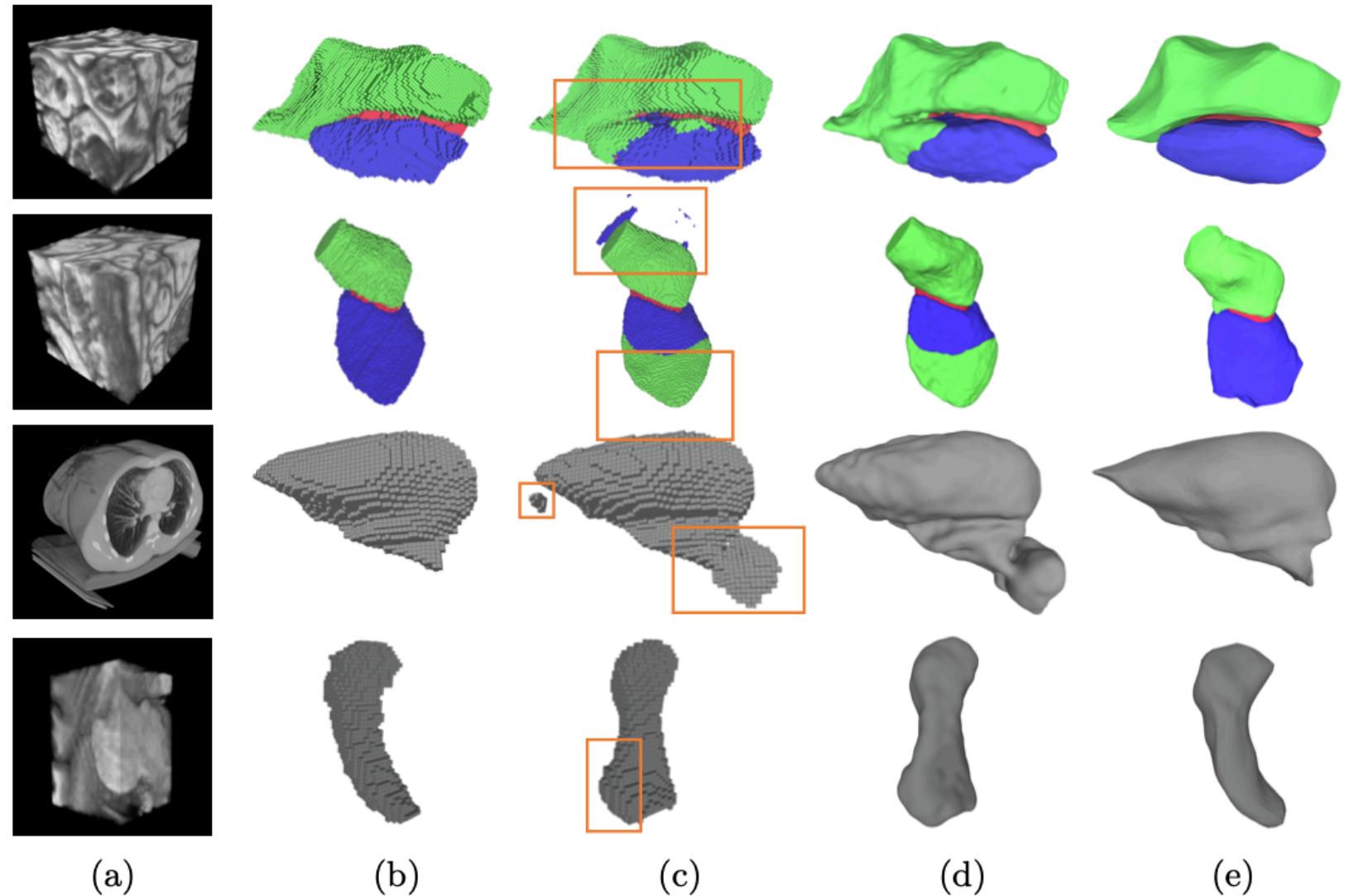


Fig. 4. Qualitative results. (a) Input volumes. EM (row 1,2), CT(row 3), MRI(row 4) (b) Ground truth (c) CNN baseline (d) CNN baseline + post processing (e) **Voxel2Mesh**. The orange boxes highlight false positive regions.

Results

Table 1. Comparative results on three datasets using the IoU metric.

	Liver	Hippo.	Synaptic Junction		
			Pre-Synap.	Synapse	Post-Synap.
TernausNet [7]	84.4 ± 1.3	78.4 ± 1.2	73.5 ± 1.3	64.4 ± 0.5	78.4 ± 1.3
LinkNet34 [3]	82.8 ± 1.4	79.4 ± 0.8	72.3 ± 0.5	63.2 ± 1.2	78.2 ± 1.1
ResNet50 [2]	82.1 ± 0.7	80.7 ± 0.2	70.3 ± 0.8	63.3 ± 0.6	76.2 ± 1.4
ResNet50-SE [2]	82.6 ± 1.2	80.5 ± 1.3	71.3 ± 0.6	63.6 ± 0.7	76.3 ± 0.9
V-NET [14]	81.5 ± 1.4	75.3 ± 1.4	64.3 ± 0.7	65.2 ± 1.3	74.1 ± 0.7
U-NET [4]	84.2 ± 1.6	80.9 ± 1.5	73.6 ± 1.3	67.2 ± 0.8	78.2 ± 0.9
<i>Best CNN + CLN</i>	84.6 ± 1.7	81.1 ± 1.5	74.5 ± 1.2	67.6 ± 0.8	79.5 ± 0.9
<i>Best CNN + FPP</i>	84.3 ± 1.7	80.8 ± 1.5	74.2 ± 1.2	67.4 ± 0.8	79.3 ± 0.9
Voxel2Mesh	86.9 ± 1.1	82.3 ± 0.9	77.3 ± 1.2	65.3 ± 1.2	83.2 ± 1.6

Ablation Study

Table 2. Comparative results against CNN based mesh deforming baselines.

	Liver		Hippocampus	
	IoU	Cf.	IoU	Cf.
PS + UMU	83.3 ± 0.8	3.3×10^{-3}	78.8 ± 1.1	2.9×10^{-3}
HS + UMU	84.2 ± 0.6	2.8×10^{-3}	79.9 ± 0.9	2.3×10^{-3}
LNS + UMU	85.6 ± 0.9	2.1×10^{-3}	81.2 ± 1.2	1.8×10^{-3}
LNS + AMU (Voxel2Mesh)	86.9 ± 1.1	1.3×10^{-3}	82.3 ± 0.9	1.1×10^{-3}