

Medical Vision Seminar

Yujin Tang

2022.01.26

- (MICCAI2021) Point-UNet: A Context-Aware Point-Based Neural Network for Volumetric Segmentation
- (MICCAI2021) RibSeg Dataset and Strong Point Cloud Baselines for Rib Segmentation from CT Scans

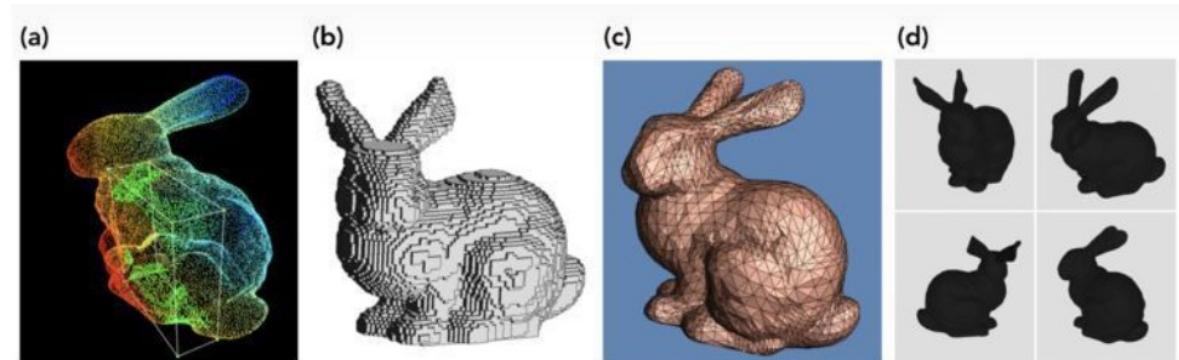
3D Data Representation

(a)点云 (Point clouds) : 点云是三维空间用xyz坐标表示的点的集合

(b)体素网格(Voxel grids) : 体素是3D空间的像素。量化的，大小固定的点云。体素网格是用固定大小的立方块作为最小单元，来表示三维物体的一种数据结构

(c)多边形网格(Polygon meshes) : mesh是面片的集合

(d) 多视图表示(Multi-view representations) :
多视图表示是从不同模拟视点渲染的2D图像集合





Point-Unet: A Context-Aware Point-Based Neural Network for Volumetric Segmentation

Ngoc-Vuong Ho^{1(✉)}, Tan Nguyen¹, Gia-Han Diep³, Ngan Le⁴,
and Binh-Son Hua^{1,2}

¹ VinAI Research, Hanoi, Vietnam

{v.vuonghn,v.tannh10,v.sonhb}@vinai.io

² VinUniversity, Hanoi, Vietnam

³ University of Science, VNU-HCM, Ho Chi Minh, Vietnam
han.diep@ict.jvn.edu.vn

⁴ Department of Computer Science and Computer Engineering,
University of Arkansas, Fayetteville 72701, USA
thile@uark.edu

Introduction

- Input : 3D MRI
- Task : Volumetric Segmentation
- Previous Study:
 - ✓ Existing voxel-based neural networks for volumetric segmentation have **prohibitive memory requirements**
 - ✓ Using smaller grid for computation, resulting in degraded performance.

Motivation

- Leverage the 3D PC representation for the problem of medical volumetric segmentation as inspired by recent success in 3D point cloud (PC) analysis
- PCs are also suitable for capturing global features that are challenging and costly to have with a regular voxel grid

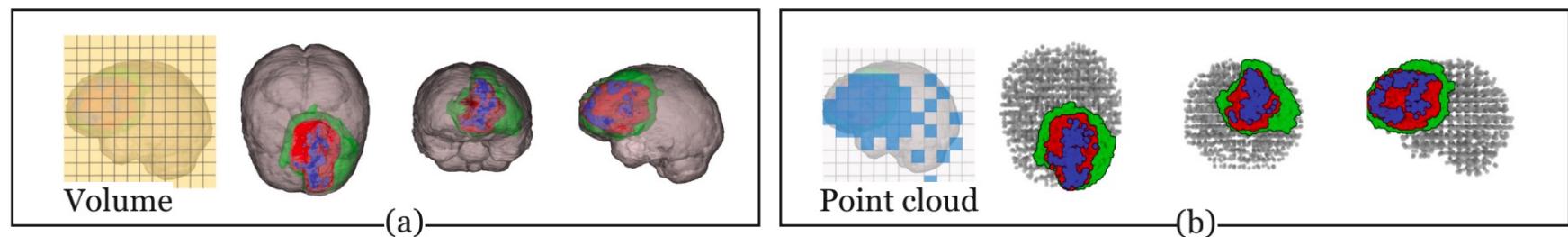


Fig. 1. (a): a 2D voxel grid and a segmentation rendered by volume rendering; (b): a PC from the grid and the point-based segmentation results.

Contribution

- ✓ Point-UNet, a new perspective and formulation to solve medical volumetric segmentation using a **PC representation**
- ✓ A **saliency proposal network** to extract an attentional probability map which emphasizes the regions of interests in a volume
- ✓ An efficient **context-aware point sampling mechanism** for capturing better local dependencies within regions of interest while maintaining global relations
- ✓ A **comprehensive benchmark** that demonstrates the advantage of our point-based method over other SOTA voxel-based 3D networks at both accuracies, memory usage during training, and inference time

Related Work

Volumetric Segmentation

- ✓ **Previous:** Unet, Vnet, DeepMedic
- ✓ **Recent:** nnUNet(2nd in BraTS18), aeUnet(1st in BraTS18)

Point Cloud Segmentation

- ✓ **PointNet:** used MLPs to learn the representation of each point
- ✓ **PointNet++:** developed to address the lack of local features by using a hierarchy of PointNet
- ✓ **PointCNN:** used a X -transformation to learn features from unstructured PCs
- ✓ **RandLA-Net:** achieved SOTA performance on semantic segmentation of large point clouds by leveraging random sampling at inference

Methods-Overview

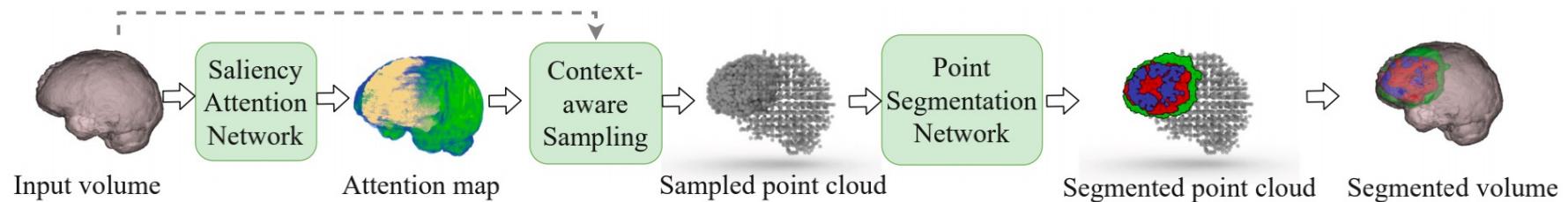


Fig. 2. Point-Unet takes a volume as input and consists of 3 modules: saliency attention network, context-aware sampling and point segmentation network.

- ✓ saliency attention module
- ✓ context-aware sampling
- ✓ point-based segmentation module

Methods: saliency attention module

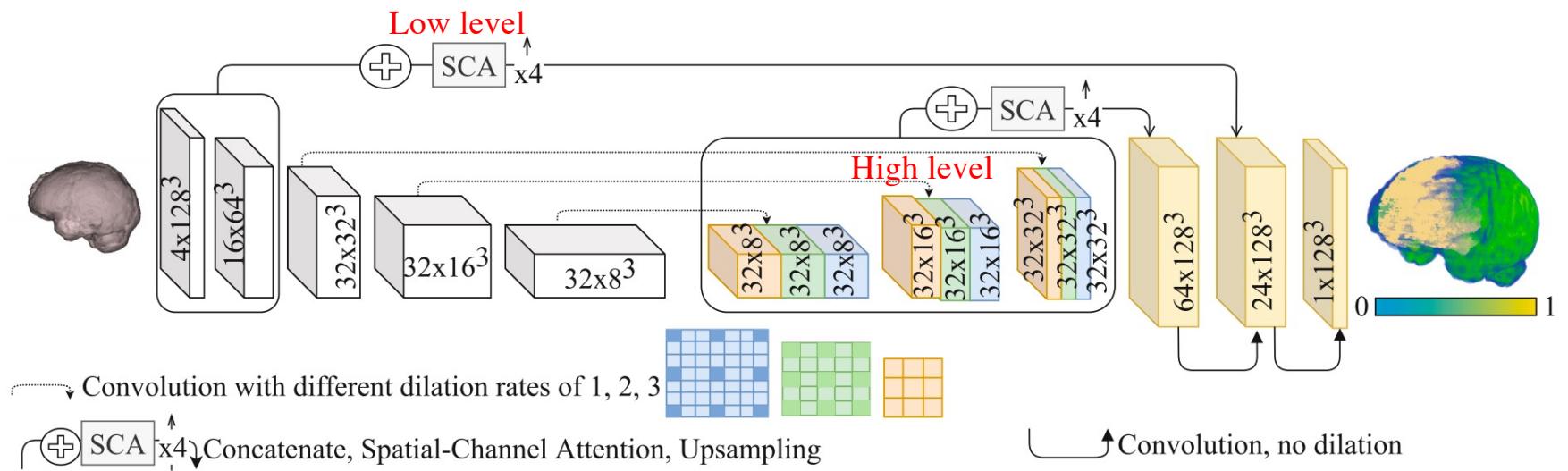


Fig. 3. Our proposed saliency attention network.

Methods: context-aware sampling

- Sample the points identified by the **attentional probability map** obtained by the saliency attention module.
- On the **salient region** where the probabilities are higher, we **densely sample points** to better learn contextual local representation.
- In the **non-salient region**, apply **random sampling** to keep global relations.

Methods: point-based segmentation module

Redesign RandLA-Net under Unet architecture

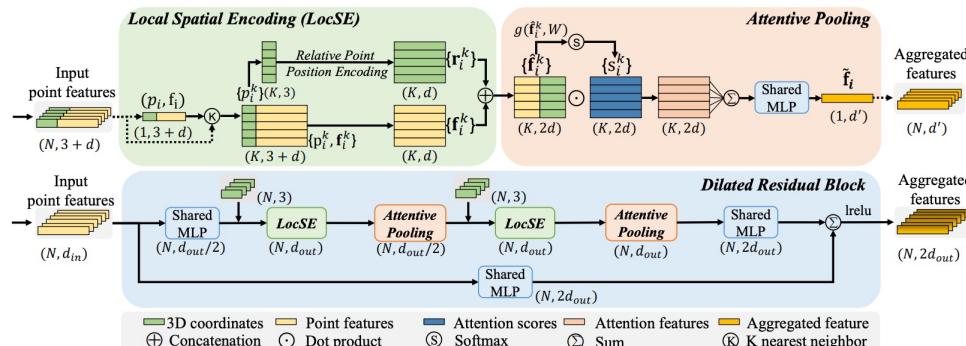


Figure 3. The proposed local feature aggregation module. The top panel shows the location spatial encoding block that extracts features, and the attentive pooling mechanism that weights the most important neighbouring features, based on the local context and geometry. The bottom panel shows how two of these components are chained together, to increase the receptive field size, within a residual block.

RandLA-Net

Input: $\pi = \{x_i, y_i\}$
 x_i : coordinate
 y_i : point features

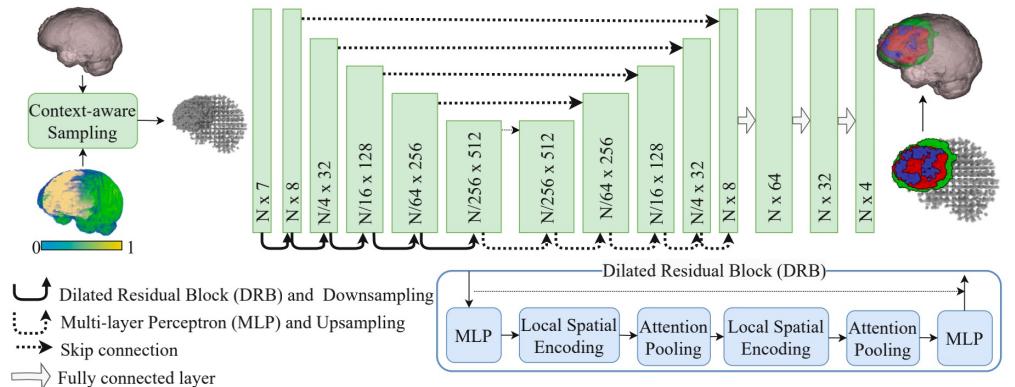


Fig. 4. Our proposed Point-Unet segmentation on volumetric data.

Replace CE loss in RandLa-Net with Generalized Dice Loss

Experiment: BraTS18/19/20 and Pancreas

Table 1. Comparison on BraTS18. The **best**, second best and *third best* are highlighted.

BraTS18	Offline validation set			Online validation set			
Methods	Dice score ↑		HD95 ↓	Methods	Dice score ↑		HD95 ↓
	ET/WT/TC	AVG	AVG		ET/WT/TC	AVG	AVG
3DUnet [15]	66.82/81.19/77.58	75.20	9.32	3DUnet [35]	72.05/84.24/76.41	77.56	17.83
3DUNet [11]	68.43/ <u>89.91</u> / 86.77	81.70	5.33	KD-Net [7]	71.67/81.45/76.98	76.70	—
h-Dense [19]	70.69/89.51/82.76	80.99	6.41	DenseNet [39]	<u>80.00</u> /90.00/82.00	84.00	—
DMF [2]	76.35/89.09/82.70	82.71	—	aeUnet [25] ^b	81.45 /90.42/ 85.96	85.94	5.52
aeUnet [25]	75.31/85.69/81.98	80.99	8.64	aeUnet [25] ^a	72.60/85.02/77.33	78.32	18.58
S3D [4]	73.95/88.81/84.42	82.39	5.40	S3D [4]	74.93/89.35/83.09	82.56	5.63
nnNet [12]	76.65/81.57/84.21	80.81	12.59	nnNet [12] ^b	79.59 / 90.80 /84.32	84.90	5.36
KaoNet [16]	73.50/ <u>90.20</u> /81.30	81.67	<u>5.92</u>	nnNet [12] ^a	75.01/82.23/81.84	79.69	7.14
RandLA [8]	70.05/88.13/80.32	79.50	6.36	RandLA [8]	73.05/87.30/76.94	79.10	5.79
Ours	80.76 / 90.55 / 87.09	86.13	6.01	Ours	80.97/90.50/84.11	85.19	6.30

^a Reproduce the results on the network trained with 100 epochs.

^b The results reported in the paper.

Sampling: 180000 points in Pancreas, 350000 points in BraTS(train and inference)

Table 2. Comparison on BraTS19. The **best**, second best and *third best* are highlighted.

BraTS19	Offline validation set			Online validation set			
	Methods	Dice score ↑		HD95 ↓	Methods	Dice score ↑	
		ET/WT/TC	AVG			ET/WT/TC	AVG
3DUnet [15]	67.74/80.17/78.92	75.61	11.69	3DUnet [35]	64.26/79.65/72.07	72.00	23.68
nnNet [12]	79.46/81.13/ <u>87.08</u>	82.67	6.75	nnNet [12] ^a	70.42/81.53/78.22	76.72	8.30
aeUnet [25]	80.55/86.26/85.78	84.19	10.94	aeUnet [25] ^a	64.81/83.02/74.48	74.10	21.75
HNF [13]	<u>80.96</u> / <u>91.12</u> /86.40	<u>86.16</u>	—	HNF [13]	81.16 / 91.12 / <u>84.52</u>	85.60	<u>3.81</u>
N3D [34]	83.0/ 91.60 / <u>88.80</u>	87.35	3.58	Synth [6]	76.65/ <u>89.65</u> /79.01	81.77	5.75
				2stage [14] ^b	79.67/90.80/ 85.89	85.45	3.74
				3Unet [34]	73.70/89.40/80.70	81.27	5.84
				3DSe [26]	80.00/89.40/ <u>83.40</u>	<u>84.27</u>	<u>4.91</u>
				Bag-trick [42] ^c	70.20/88.30/80.00	79.50	4.93
RandLA [8]	76.68/89.01/84.81	83.50	4.45	RandLA [8]	70.77/86.95/74.27	70.77	7.09
Ours	85.67 / <u>91.18</u> / 90.10	88.98	<u>4.92</u>	Ours	79.01/87.63/79.70	82.11	10.39

^a Reproduce the results on the network trained with 100 epochs.

^b The first place of BraTS19. We choose the Ensemble of 5-fold.

^c The second place on BraTS19.

Reproduced result lower than reported result(with postprocessing)

Experiment: BraTS18/19/20 and Pancreas

Table 3. Comparison on BraTS20. The **best**, second best and *third best* are highlighted.

BraTS20	Offline validation set			Online validation set			
Methods	Dice score ↑		HD95 ↓	Methods	Dice score ↑		HD95 ↓
	ET/WT/TC	AVG	AVG		ET/WT/TC	AVG	AVG
3DUNet [15]	66.92/82.86/72.98	74.25	30.19	3DUNet [35]	67.66/87.35/79.30	78.10	21.16
nnNet [12]	<u>73.64</u> /80.99/ <u>81.60</u>	<u>78.74</u>	<u>14.33</u>	nnNet [12] ^a	68.69/81.34/78.06	78.03	24.30
aeUnet [25]	<u>71.31</u> / <u>84.72</u> / <u>79.02</u>	<u>78.35</u>	15.43	nnUNet [10] ^c	77.67 / 90.60 / 84.26	84.18	15.30
				aeNet [25] ^a	64.00/83.16/74.66	73.95	33.91
				Cascade [21] ^b	<u>78.81</u> / <u>89.92</u> / <u>82.06</u>	<u>83.60</u>	<u>12.00</u>
				KiUNet [33]	73.21/87.60/73.92	78.24	8.38
RandLA [8]	67.40/ <u>87.74</u> /76.85	77.33	7.03	RandLA [8]	66.31/88.01/77.03	77.17	16.65
Ours	76.43 / 89.67 / 82.97	83.02	8.26	Ours	78.98 / <u>89.71</u> / <u>82.75</u>	83.81	11.73

^a Reproduce the results on the network trained with 100 epochs.

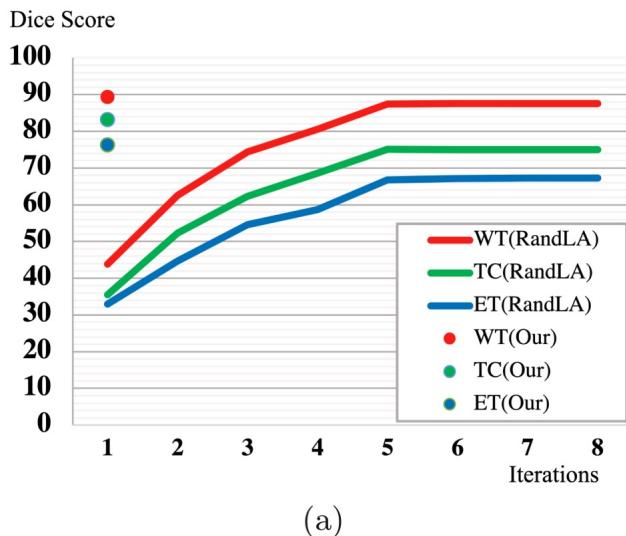
^b We choose the model with as similar batch size as ours.

^c We choose single model with 190 epoches, stage 1. The best model at Brats2020.

Table 4. Dice score comparison on Pancreas dataset.

Method	Average ↑	Method	Average ↑
Oktay et al. [27]	83.10 ± 3.80	Yu et al. [40]	84.50 ± 4.97
Zhu et al. [43]	84.59 ± 4.86	Ours	85.68 ± 5.96

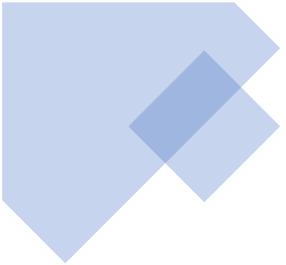
Experiment: Performance Analysis



Networks	Patch size	Memory ↓	Inference ↓
3DUnet	128 × 128 × 128	8.75 GB	7.80 s
baseline [15]	160 × 192 × 128	16.70 GB	0.28 s
	240 × 240 × 144	32.00 GB	0.23 s
nnNet [12]	128 × 128 × 128	7.20 GB	55.30 s
	160 × 192 × 128	11.10 GB	26.50 s
	240 × 240 × 144	21.70 GB	2.30 s
aeUnet [25]	128 × 128 × 128	17.21 GB	110.40 s
	160 × 192 × 128	31.42 GB	78.80 s
	240 × 240 × 144	>48.00 GB	7.10 s
RandLA[8]	240 × 240 × 155	15.98 GB	8.00 s
(Ours)	240 × 240 × 155	17.22 GB	1.24 s

(b)

Fig. 5. Performance analysis. (a) With a single inference, our Point-Unet outperforms RandLA-Net, which requires multiple iterations at inference. (b) Memory requirement for training with batch size 1 and inference time with different volume patch sizes.



RibSeg Dataset and Strong Point Cloud Baselines for Rib Segmentation from CT Scans

Jiancheng Yang^{1,2,*}, Shixuan Gu^{1,*}, Donglai Wei³,
Hanspeter Pfister³, and Bingbing Ni^{1†}

¹ Shanghai Jiao Tong University, Shanghai, China
nibingbing@sjtu.edu.cn

² Dianei Technology, Shanghai, China

³ Harvard University, Cambridge MA, USA

Introduction

- Input : 3D CT
- Task : rib segmentation
- Previous Study:
 - ✓ Mostly use in-house labeled datasets that are publicly unavailable
 - ✓ Work on 3D Volume that are computationally inefficient

Contribution

- The **first public benchmark** for rib segmentation, which enables downstream applications and method comparison.
- A novel point-based perspective on modeling 3D medical images beyond voxel grids
- A point cloud-based rib segmentation baseline with high efficiency and accuracy.

Dataset

- CT scans: saved in NIFTI (.nii) format volume sizes of $N \times 512 \times 512$, where 512×512 is the size of CT slices, and N is the number of CT slices (typically $300 \sim 400$)
- Semi-Automatic Annotation procedure:

Rib Segmentation->Centerline Extraction->Manual Checking and Refinement

Table 1: Overview of RibSeg dataest.

Subset	No. of CT Scans	No. of Individual Ribs
Training / Development / Test	320 / 50 / 120	7,670 / 1,187 / 2,862

Methods: from a Viewpoint of Point Clouds

- Can **avoid heavy computation** on dense voxels with sparse point clouds instead.
- Use geometric information directly, **reducing the texture bias** of pixel/voxel-based CNNs.
- **Deep learning for point cloud analysis** has been an emerging, only a few studies have applied deep shape analysis in medical imaging scenarios.

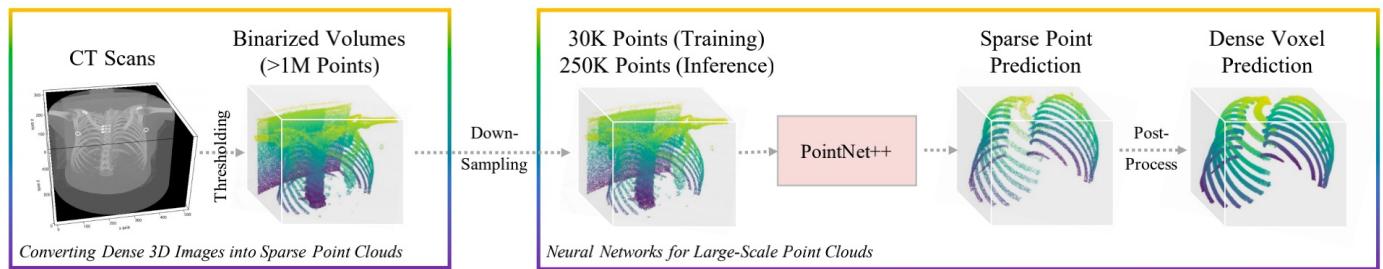


Fig. 2: **Rib Segmentation from a Viewpoint of Point Clouds.** The CT volumes are first binarized to obtain candidate bone voxels as inputs, then a point cloud neural network (*e.g.*, PointNet++ [13]) is used to segment each point in downsampled input point clouds. Note that the downsampling scale is different during training (30K points) and inference (250K points).

Experiment

- ✓ All point-based methods significantly outperform voxel-based 3D UNet
- ✓ More efficient(forward time)
- ✓ Augmentation->higher performance
- ✓ Large-scale input-richer details and more time

Table 2: Quantitative metrics on RibSeg test set, including Dice over sparse points ($Dice^{(P)}$), Dice over dense voxels after post-processing ($Dice^{(V)}$), ratio of missed all/first/intermediate/twelfth rib pairs (A/F/I/T) at recall > 0.5 , and the model forward time in second. Post-processing time is not included as it heavily depends on the implementation.

Methods	$Dice^{(P)}$	$Dice^{(V)}$	Missed Ribs (A/F/I/T)	Forward (s)
Voxel-Based 3D UNet [2,7]	-	86.3%	4.6%/7.9%/2.3%/24.6%	30.63
PN++ [13] (30K)	92.3%	91.0%	1.6%/2.9%/0.7%/10.4%	0.32
PN++ [13] (250K)	91.5%	92.3%	0.9%/3.3%/0.3%/4.7%	1.12
PN++ [13] (30K) + aug.	94.9%	94.3%	1.1%/0.8%/0.4%/9.0%	0.32
PN++ [13] (250K) + aug.	94.6%	95.2%	0.6%/0.4%/0.2%/5.2%	1.12

Visualization

- ✓ CT scans of incomplete rib cages
- ✓ CT scans of serious spinal pathology
- ✓ CT scans containing metal objects inseparable from ribs

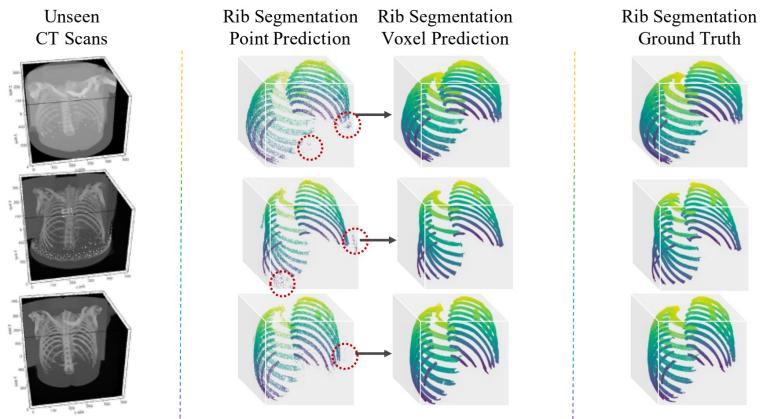


Fig. 3: **Visualization of Predicted Rib Segmentation.** Red circles denote imperfect (sparse) point prediction.

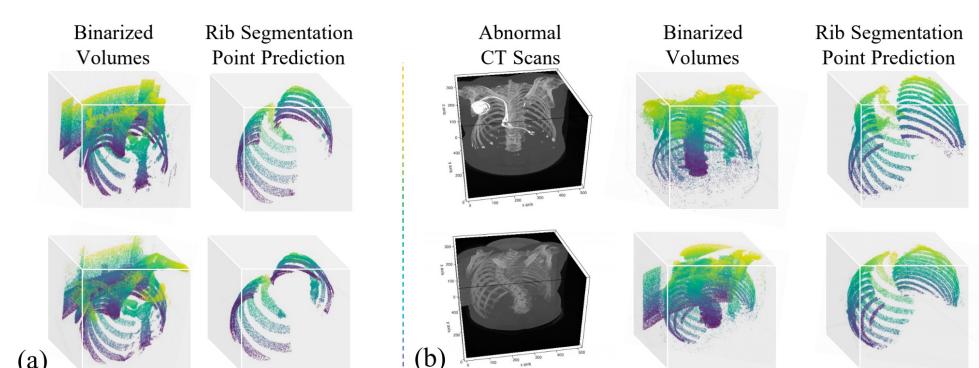


Fig. 4: **Robustness Test on Extreme Cases.** (a) Point prediction on unseen incomplete rib cages. (b) Point prediction on unseen abnormal CT scans.

Thank you!