Deep Esophageal Clinical Target Volume Delineation using Encoded 3D Spatial Context of Tumors, Lymph Nodes, and Organs At Risk

PING AN

CHANG GUNG MEDICAL FOUNDATION
Chang Gung Memorial Hospital, Linkou

MICCAI 2019
13-17 OCTOBER 2019
SHENZHEN. CHINA

Dakai Jin[1], Dazhou Guo[1], Tsung-Ying Ho[2], Adam P. Harrison[1], Jing Xiao[3], Chen-Kan Tseng[2], Le Lu[1]

1. PAII Inc., Bethesda, MD, USA 2. Chang Gung Memorial Hospital, Linkou, Taiwan, ROC 3. Ping An Technology, Shenzhen, China

Motivation and Objective

Motivation

- **Esophageal clinical target volume (CTV) delineation**
  - Essential yet extreme difficult task in radiotherapy treatment planning
  - Spatially encompass tumor(s), i.e., GTV, regional lymph nodes (LNs) and sub-clinical disease regions
  - Receives similar radiation dose as GTV
  - High intra- and inter-user variability (0.51~0.81 IoU)
- **CTV Delineation challenges**
  - Not a uniform margin expansion from GTV
  - Mixture of predefined & judgment-based margins
  - Encompass GTV + regional lymph nodes (LNs)
  - Avoiding excessive radiation to organs at risk (OARs)
- **Prior arts on CTV delineation**
  - ill-posed problem formulation
  - Mostly operate based on *pure CT appearance*, while missing spatial context information used by oncologists[1,2]
  - *Performance: low Dice score (<75%) and large average distance errors (>10 mm)*

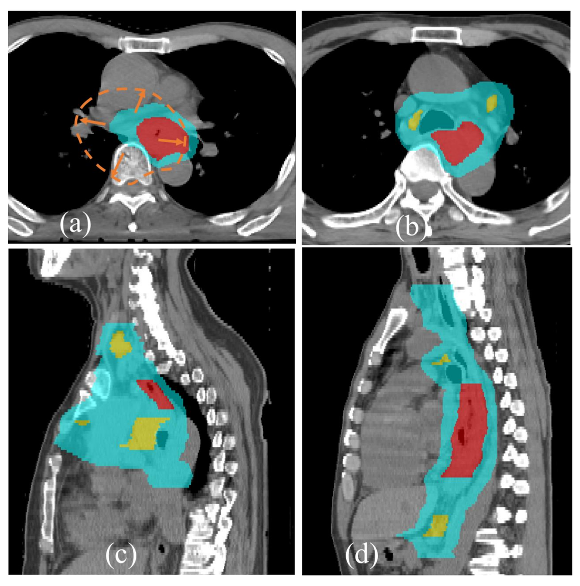

Fig. 1 Esophageal CTV examples

Aim

- ***Develop an accurate and robust 3D esophageal CTV delineation method:***
  - Formulate as a deep contextual appearance-based problem using encoded spatial contexts of GTV, regional LNs, and OARs
  - Allows deep network to better learn from and emulate the margin- and appearance-based delineation performed by oncologists
  - 3-fold cross-evaluate the proposed method on RTCT from 135 patients

Methods

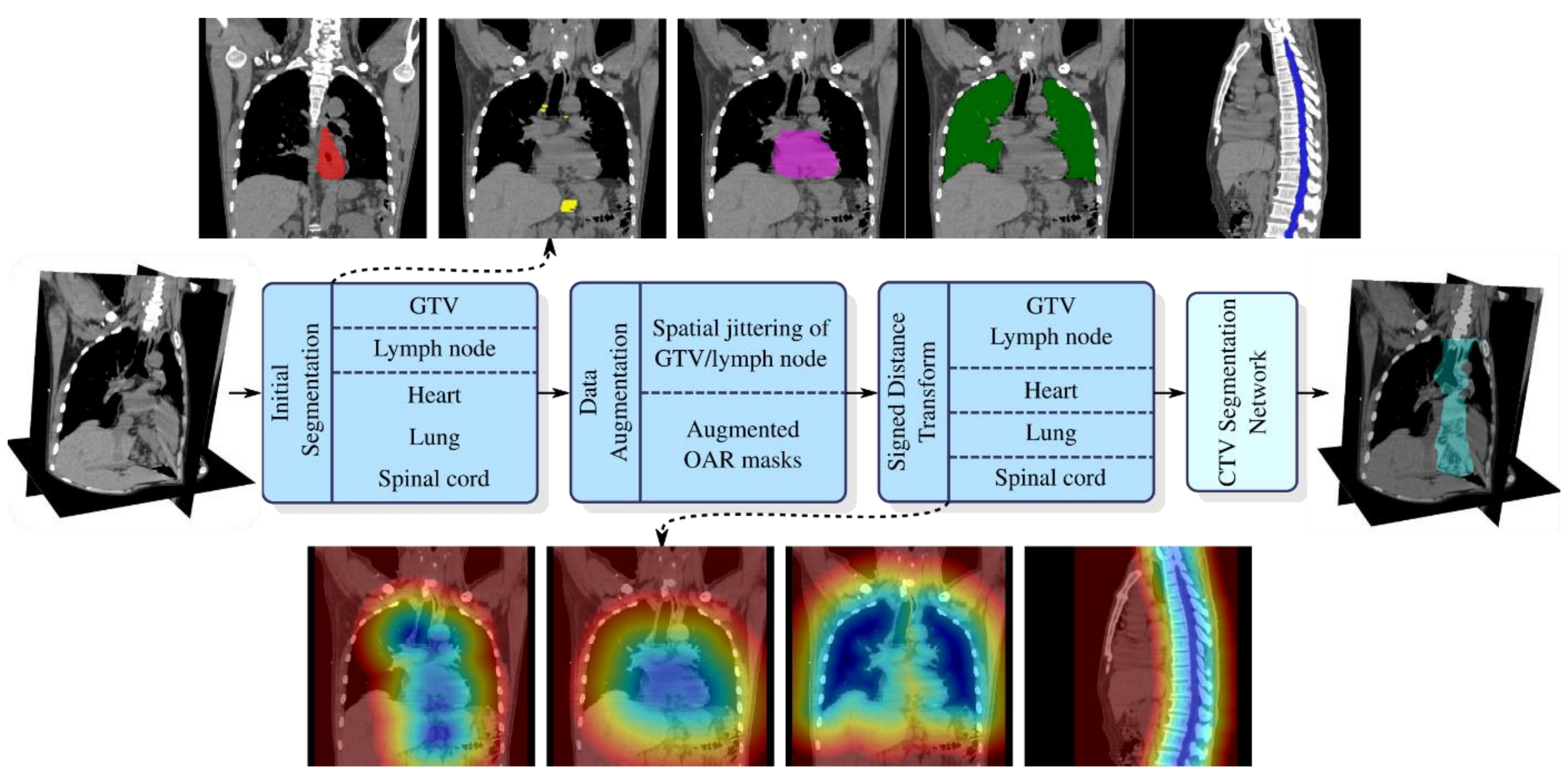

Fig. 2 Overall workflow of the prosed spatial context encoded deep CTV delineation framework

Prerequisite Region Segmentation

- Manual segmentations of esophageal GTV and regional LNs
- Three major OARs: lung, heart and spinal canal
- Automatic OARs segmentation by training 2D PHNN[3]
- Validation Dice score for lung, heart and spinal canal: 97%, 95% and 78%

3D Signed Distance Transform

- Signed distance transform maps (SDMs) of GTV, reginal LNs and OARs as the encoded spatial context
- Efficient algorithm to compute 3D the Euclidean distance[4]

$$\mathrm{SDM}_{\Gamma(\mathcal{O}_i)}(p) = \begin{cases} \min\limits_{q \in \Gamma(\mathcal{O}_i)} d(p,q) & if\ p \notin \mathcal{O}_i \\ -\min\limits_{q \in \Gamma(\mathcal{O}_i)} d(p,q) & if\ p \in \mathcal{O}_i \end{cases}$$

Domain-specific Data Augmentation

- Increase the robustness of the training and harden the network to noise from prerequisite segmentations
  - Spatially jittering of GTV and regional LNs masks (within 4 x 4 x 4 mm^3)
  - Calculation of SDMs of OARs using both manual and automatic segmentations

CTV Delineation Network

- Symmetric encoder-decoder networks, e.g. 3D U-Net[5], not always effective
  - computationally heavy and memory-consuming
  - half of its computation consumed on decoding path
- Adopt a simple 3D PHNN[3]
  - More balanced input image size vs. computational/memory burden

Experiments and Results

Datasets & Evaluation Metrics

- 135 esophageal cancer patients diagnosed at stage II or later undergoing RT
- Evaluation metrics: Dice score, Hausdorff distance (HD) in mm, and average surface distance (ASD) in mm

Training Data Generation and Training Parameters

- 96x96x64 training VOI near ground truth CTV or randomly sampled from background
- average ~80 training VOI per patient
- Adam solver with momentum 0.99 and a weight decay of 0.005, train for ~30 epochs

Qualitative compare:

- ❑ 1st row: Pure CT setup[1]
- ❑ 2nd row: CT + GTV/LNs binary mask setup[2]
- ❑ 3rd row: CT + GTV/LNs SMDs setup
- ❑ 4th row: CT + GTV/LNs /OARs SDMs setup
- ✓ Pure CT setups fail to include the regional LNs, while (c) to (e) depict severe over-segmentations
- ✓ GTV/LNs mask setup still suffer from inaccurate CTV margin or over-segmentation

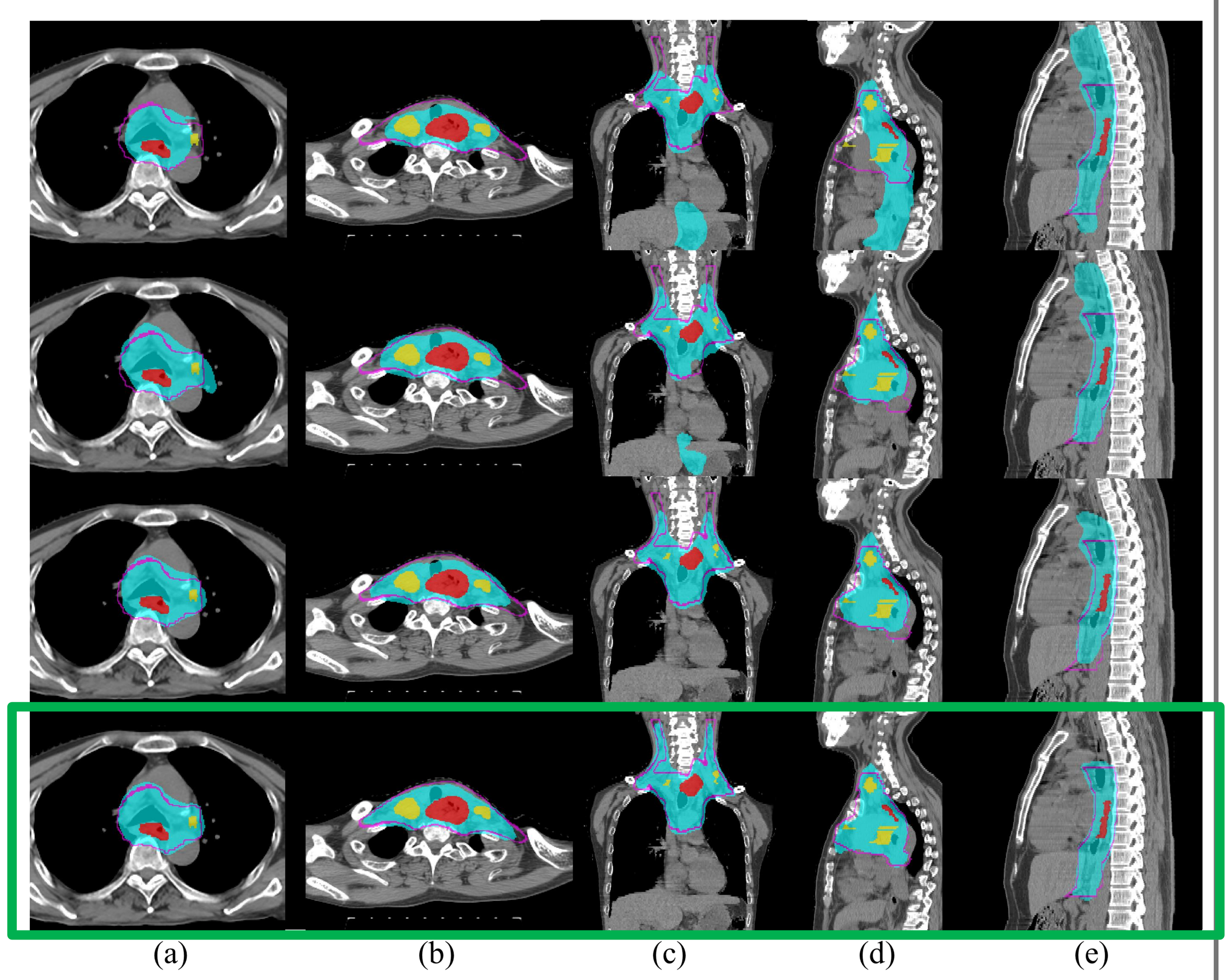

Table. 1 Quantitative results for the esophageal cancer CTV delineation. The last, starred row represents performance when using automatically generated OAR SDMs.

Models	Setups	Dice	HD (mm)	ASD (mm)
U-Net	CT	0.739±0.126	69.5±42.7	10.1±9.4
	CT + GTV/LN/OAR SDMs	0.829±0.061	36.9±23.8	4.6±3.0
PHNN	CT	0.739±0.117	68.5±43.8	10.6±9.2
	CT + GTV/LN masks	0.801±0.075	56.3±35.4	6.6±5.3
	CT + GTV/LN SDMs	0.816±0.067	44.7±25.1	5.4±4.1
	CT + GTV/LN/OAR SDMs	**0.839±0.054**	**35.4±23.7**	**4.2±2.7**
	CT + GTV/LN/OAR SDMs*	0.823±0.059	43.6±26.4	5.1±3.3

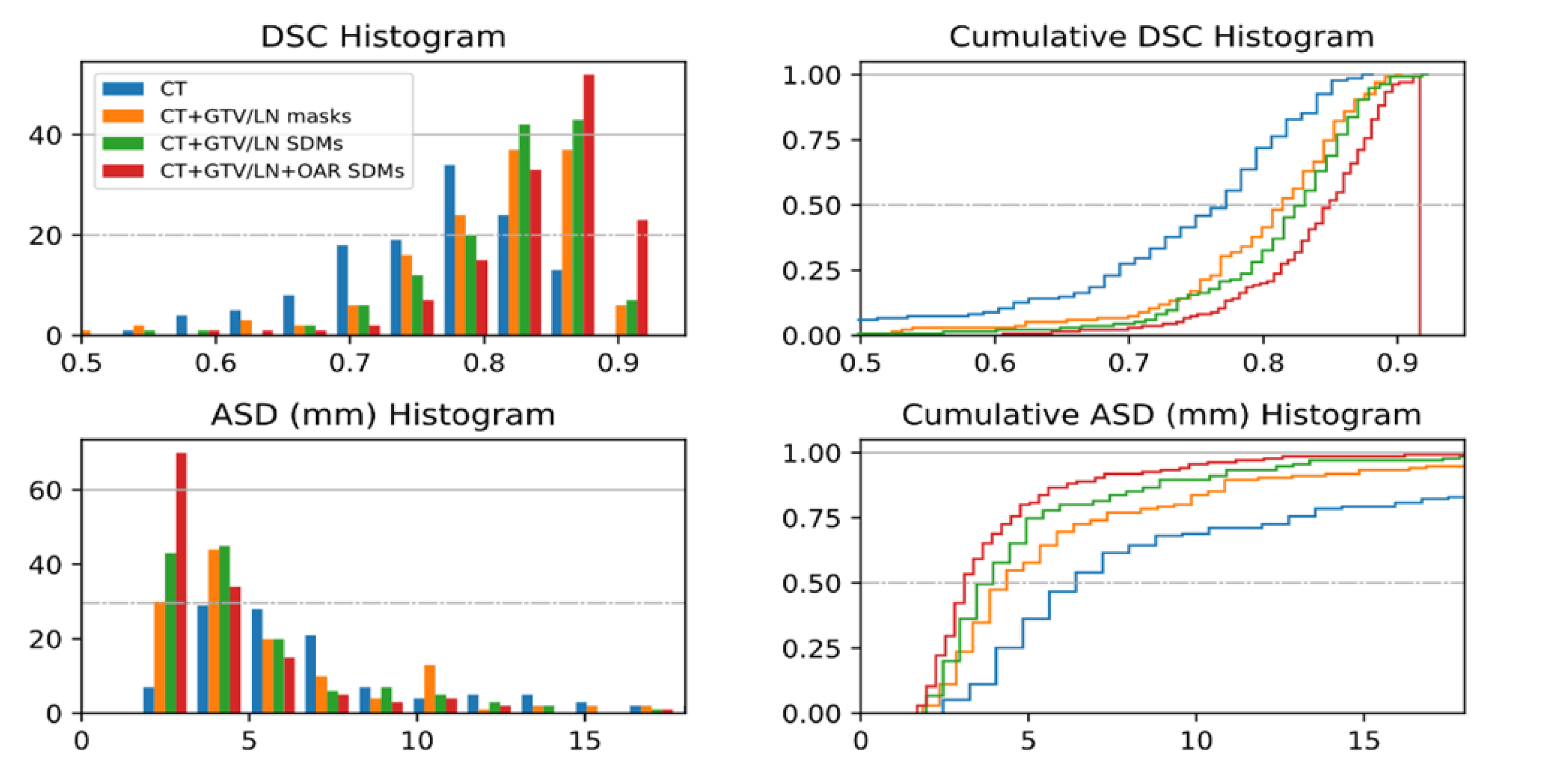

Fig. 3 Cumulative histograms of the CTV delineation performance under 4 setups using 3D PHNN. The proposed method*: > 77% patients have Dice score > 0.80, and > 55% patients have Dice score > 0.85*. These indicate that, for a high percentage of the studied patient population, little to no additional manual revision is needed on the automatically delineated CTVs.

Conclusion

- Introduced a spatial-context encoded deep esophageal CTV delineation framework designed to produce superior margin-based CTV boundaries
- Consider both appearance & distance-based information for CTV delineation by computing the SDMs of GTV, LNs and OARs and feeds together with CT into 3D CNN
- Demonstrate that the proposed method outperform the state-of-the-art CTV alternatives by wide margins in Dice score, HD, and ASD

References:

[1] K Men, et al. "Fully automatic and robust segmentation of CTV for radiotherapy of breast cancer using big data deep learning," Physica Medica 2018

[2] C. E. Cardenas, et al, "Auto-delineation of oropharyngeal clinical target volumes using 3d convolutional neural networks." Physics in Medicine & Biology, 2018

[3] A. P. Harrison, *et al.*: Progressive and Multi-Path Holistically Nested Neural Networks for Pathological Lung Segmentation from CT Images, MICCAI, 2017

[4] C. R. Maurer, *et al.*: A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. IEEE Trans. Pattern Anal. Mach. Intell, 2003

[5] Cicek, O., et al.: "3d u-net: Learning dense volumetric segmentation from sparse annotation". In: MICCAI, 2016