Correlation Matters: Multi-scale Fine-grained Contextual Information Extraction for Hepatic Tumor Segmentation*

Never Stand Still | Faculty of Engineering | A Talk in 9517

Shuchao Pang

Research Associate
School of Computer Science and
The University of New South Wales, Sydney, Australia
+61 480 108 078
shuchao.pang@unsw.edu.au

09 NOV 2020

* Pang *et al.* "Correlation Matters: Multi-scale Fine-Grained Contextual Information Extraction for Hepatic Tumor Segmentation." In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 462-474. Springer, Cham, 2020.

Agenda Layout

1. Background
2. Aims
3. Methods
4. Data
5. Results
6. Conclusions

Ingham Institute Applied Medical Research

UNSW THE UNIVERSITY OF NEW SOUTH WALES

Background

MEDICAL IMAGE SEGMENTATION VS GENERAL SEMANTIC SEGMENTATION

- Data characteristics

1) multiple modalities: X-ray, CT, MRI, Ultrasound, PET, etc. vs gray, colour images

2) pixel values: CT's Hounsfield (Hu): -1024-3071 vs 0-255

3) noise problems 4) artefacts

5) image size 6) severe class imbalance problem

- Segmentation targets

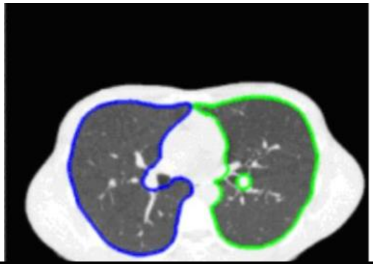
1) For medical image segmentation, we should know what we want to extract before building can start. For example, lesions, or organs, and which type.

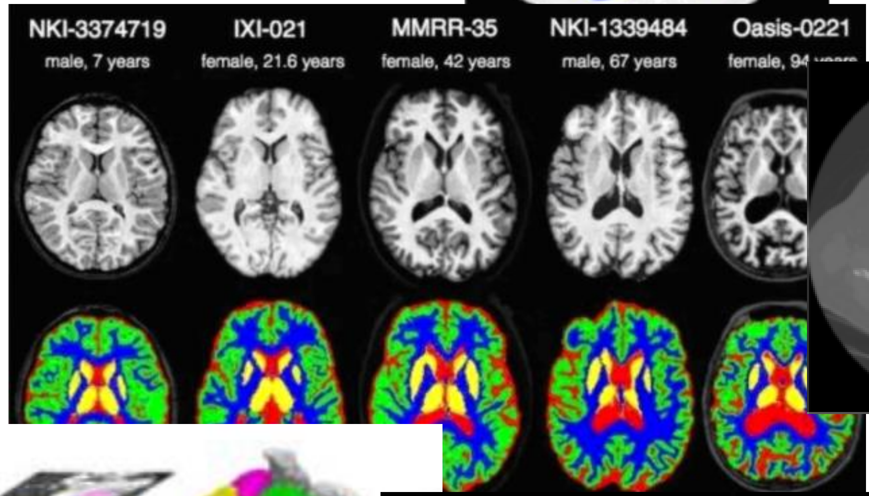
2) In other words, it's hard to design a unified model to segment all types of targets on medical images for different tasks. For example, limit the use of prior knowledge, which may lead to poor accuracy.

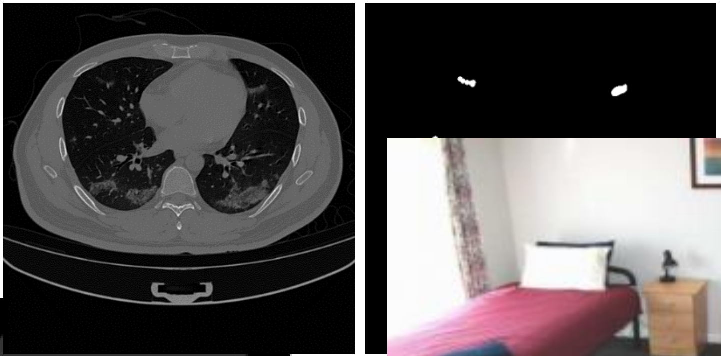
Background

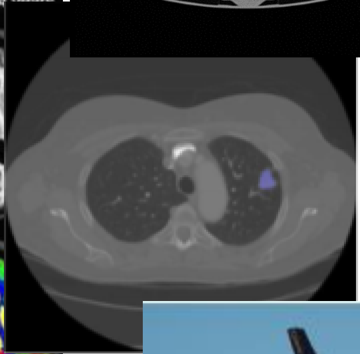
MEDICAL IMAGE SEGMENTATION VS GENERAL SEMANTIC SEGMENTATION

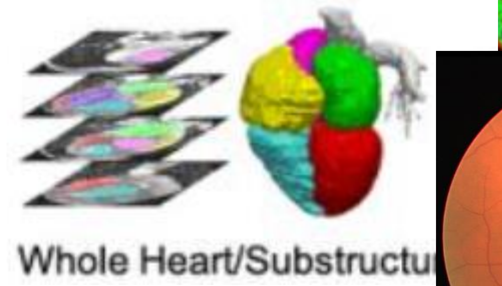
- Data scale
- Model design

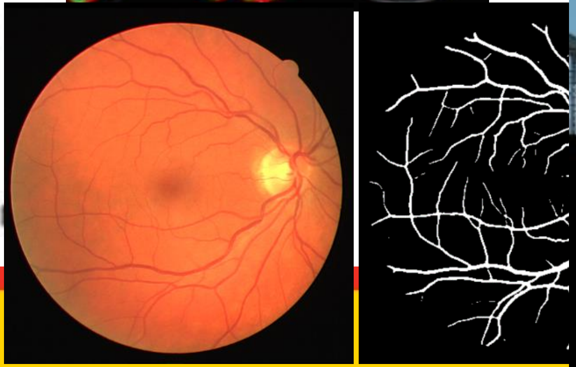


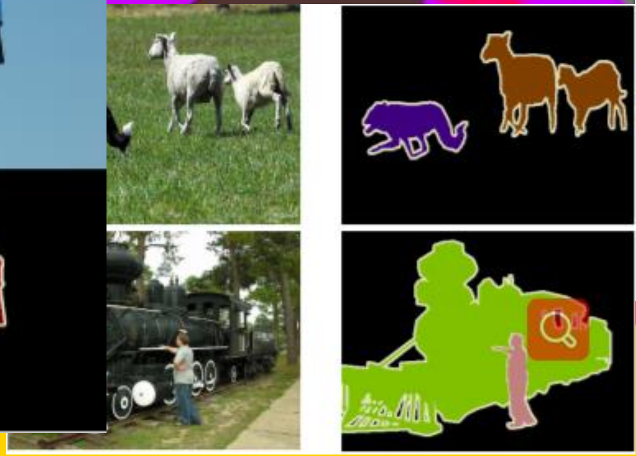










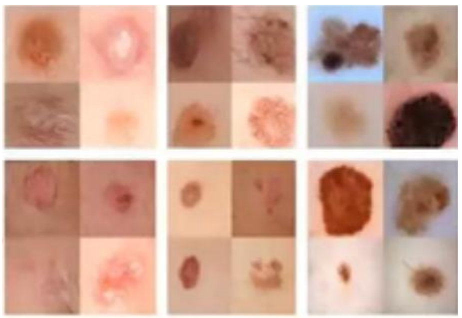


Ingham Institute
Applied Medical Research

Background

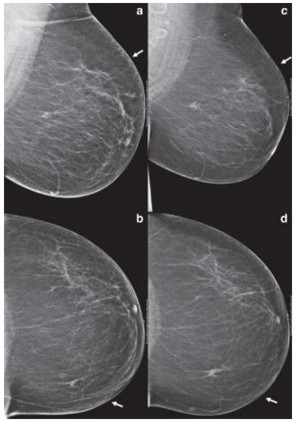
MEDICAL IMAGING DATA: 2D VS 3D

- 2D



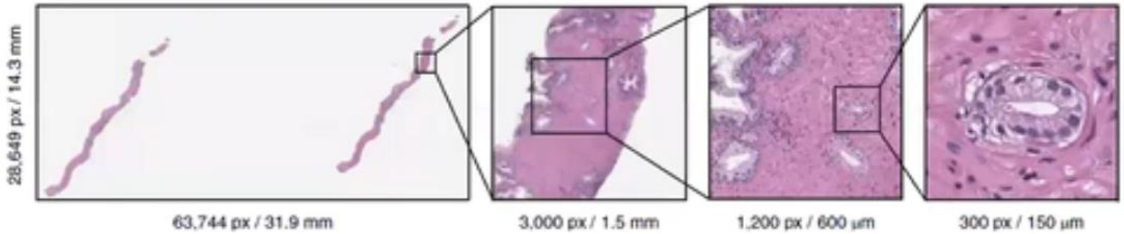
Skin cancer moles and lesions

Chest X-ray



Mammography

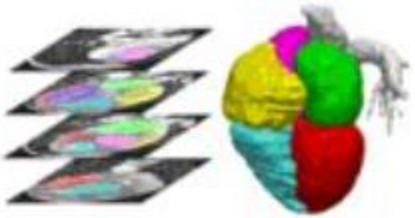
Fundus image



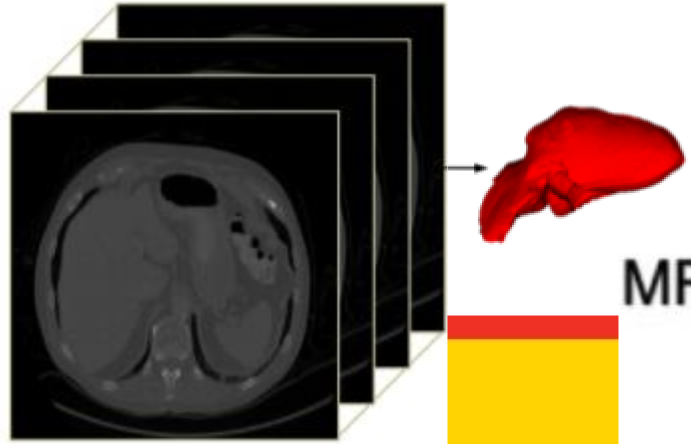
Pathology biopsy image for prostate cancer

- 3D

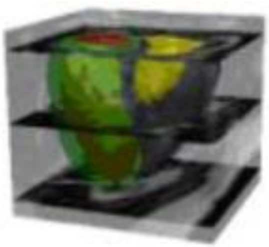
CT



Whole Heart/Substructures



MRI



Bi-ventricle (LV+RV)

Background

NORMAL TISSUES VS ABNORMAL TISSUES

	Normal tissues	vs	Abnormal tissues
1) **Scenarios**:	organs, vessels, cells	vs	tumours, lesions, cancerous cells
2) **Clinical relevance**:	quantification (volume), visualization, radiotherapy (OARs), clinical-oriented analysis.	vs	tumour quantification, diagnosis, prognosis, radiotherapy (GTV/CTV), monitoring.
3) **Challenges**:	low contrast, multi-class problems, variances in shape for some large organs.	vs	small scale, class imbalance, ambiguous boundaries, various context, irregular shape, limited visibility

4) **Common issues**: Data: high-dimensional, multi-modal, labour-intensive annotations

Goal: speed and accuracy

Background

TRADITIONAL MEDICAL IMAGE SEGMENTATION METHODS

- To deal with these challenges and issues, various algorithms have been extensively studied in the past decades.
- Previous algorithms mainly utilized statistical shape modeling, level sets, active contours, single/multi-atlas and graphical models, with hand-crafted features [1-3].
- In recent years, discriminative classifier based methods are very popular:

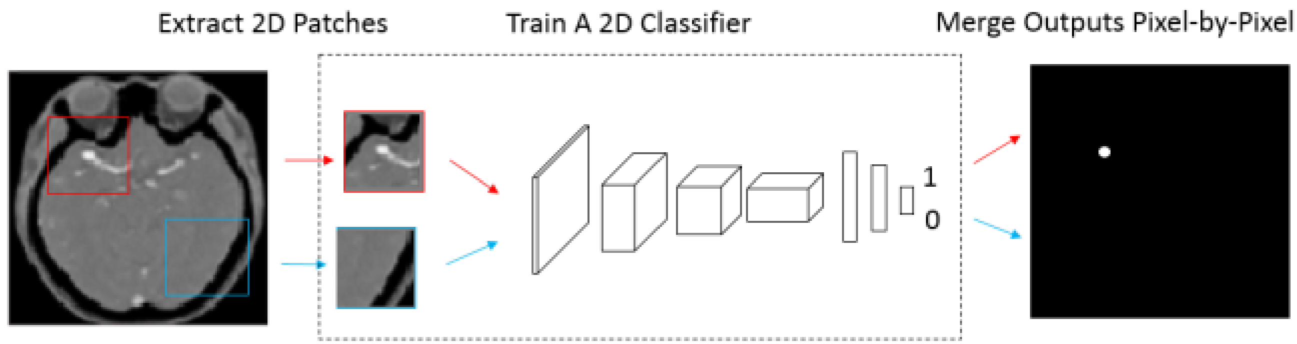
$$F_\theta(X) \rightarrow Z \in R^z; \;\; C_\phi(Z) \rightarrow Y \in R^c$$

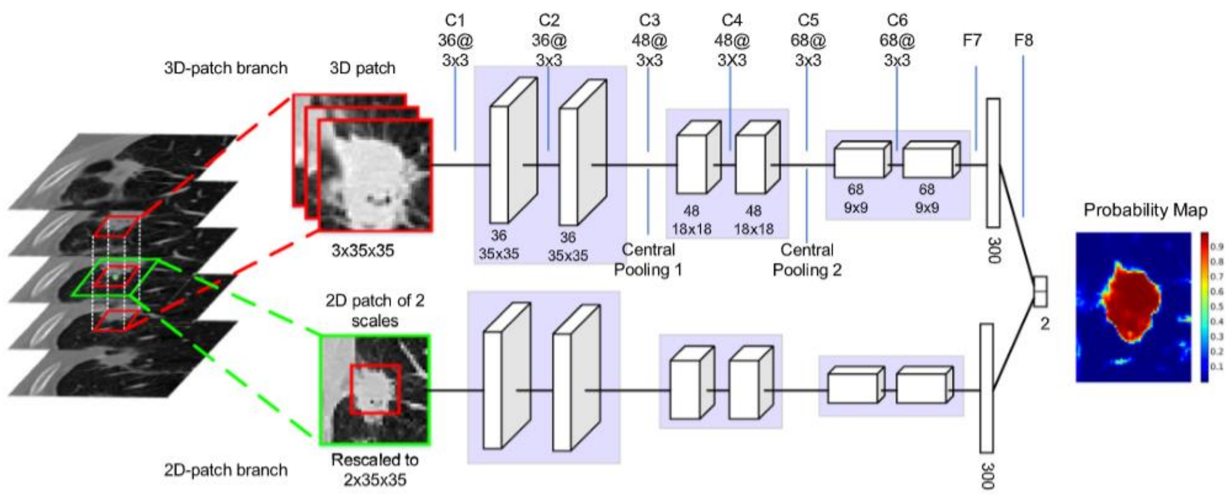
$$X \rightarrow \underbrace{F_\theta(\cdot) \circ C_\phi(\cdot)}_{\text{Deep Neural Networks based methods}} \rightarrow Y$$

[1] McInerney, T. and Terzopoulos, D., 1996. Deformable models in medical image analysis: a survey. Medical image analysis, 1(2), pp.91-108.
[2] Iglesias, J.E. and Sabuncu, M.R., 2015. Multi-atlas segmentation of biomedical images: a survey. Medical image analysis, 24(1), pp.205-219.
[3] Pham, D.L., Xu, C. and Prince, J.L., 2000. Current methods in medical image segmentation. Annual review of biomedical engineering, 2(1), pp.315-337.

Background

DEEP NEURAL NETWORKS (1) Patch-based Methods

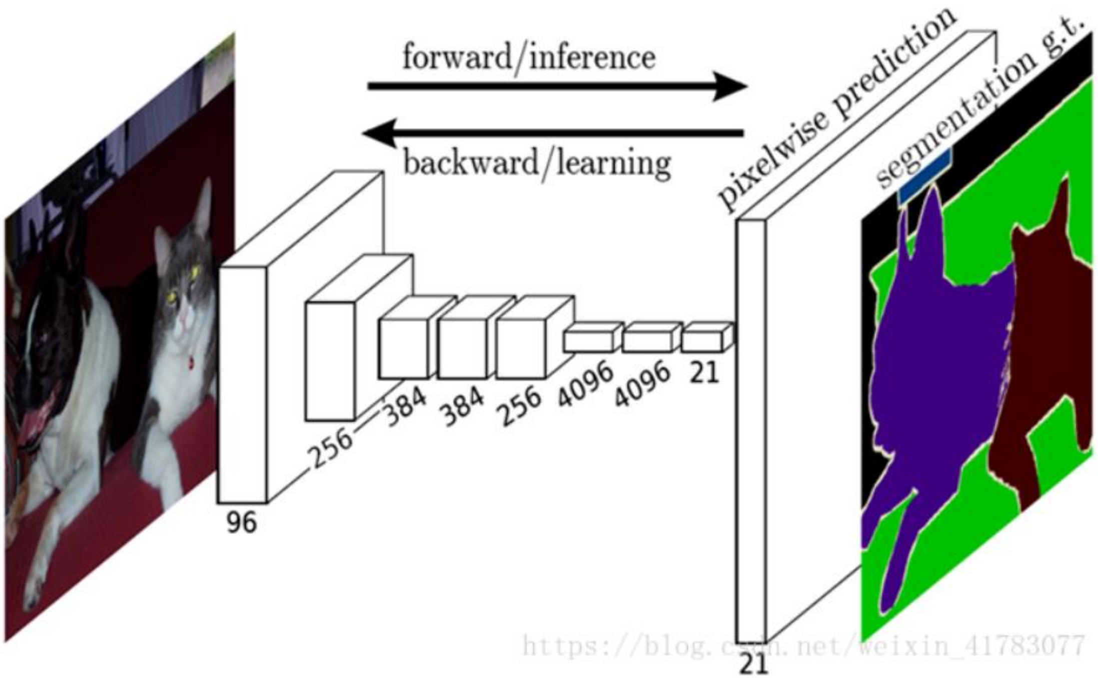


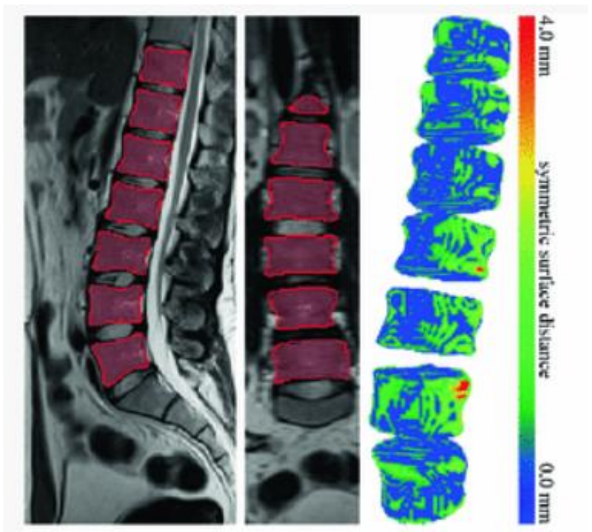


Wang, S., Zhou, M., Liu, Z., Liu, Z., Gu, D., Zang, Y., Dong, D., Gevaert, O. and Tian, J., 2017. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. Medical image analysis, 40, pp.172-183.

Background

DEEP NEURAL NETWORKS (2) FCN-based Methods

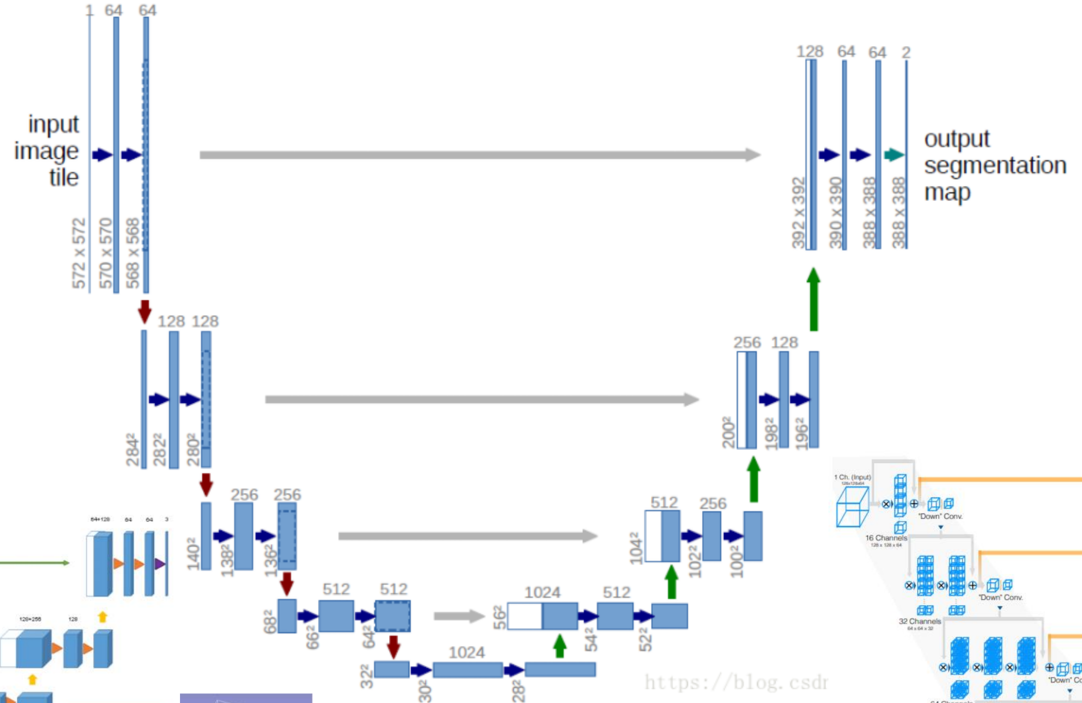


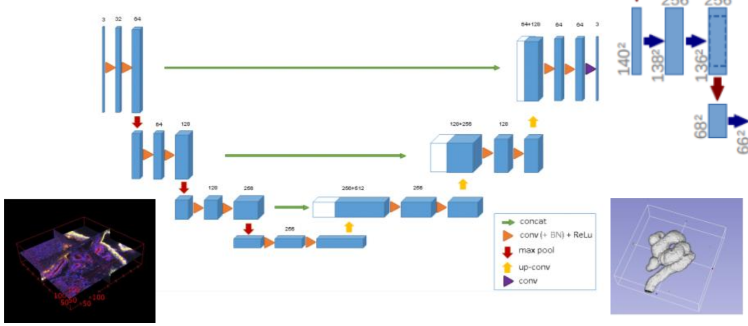


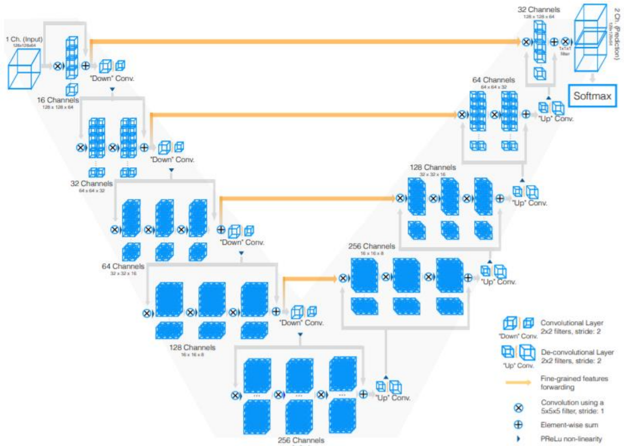
Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
Korez, R., Likar, B., Pernuš, F. and Vrtovec, T., 2016, October. Model-based segmentation of vertebral bodies from MR images with 3D CNNs. In International conference on medical image computing and computer-assisted intervention (pp. 433-441). Springer, Cham.

Background

DEEP NEURAL NETWORKS (3) UNet-based Methods







[1] Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham. [2] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O., 2016, October. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention (pp. 424-432). Springer, Cham. [3] Milletari, F., Navab, N. and Ahmadi, S.A., 2016, October. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV) (pp. 565-571). IEEE.

Ingham Institute Applied Medical Research

UNSW THE UNIVERSITY OF NEW SOUTH WALES

Aims

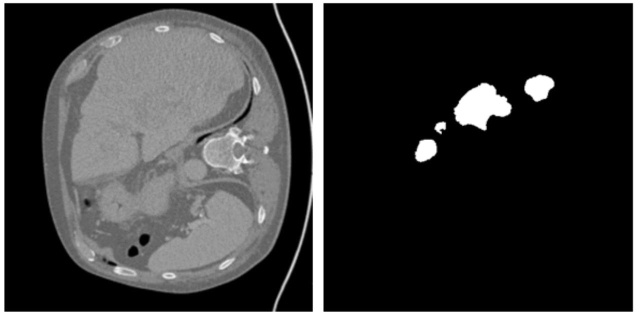
TUMOUR SEGMENTATION & CHALLENGES

- Among various clinical applications for medical image analysis, tumor localization and segmentation is still a more challenging task, compared to normal organ recognition and segmentation tasks.
- With hepatic tumors as an example based on CT scans, tough challenges are listed as follows:

  1) low tissue contrast;

  2) large variability in tumor shape, size and number;

  3) vague boundary between diseased and healthy regions.



- However, manually analyzing these examined images is time-consuming and error-prone for physicians and radiologists, plus some inter-observer variations.
- Therefore, an accurate and automatic hepatic lesions/tumors localization and segmentation approach is urgently required for early liver cancer detection.

Aims

MOTIVATION & EXISTING METHODS

- We aim to explicitly capture global context dependencies from multi-scale feature spaces for improving the performance of medical tumor segmentation.
- Although current FCNs- or U-Net-based approaches have been widely utilized in medical image segmentation tasks recently, the following limitations of existing methods for context information extraction should be noted:

1) the range of contextual information obtained from most methods is heavily limited by the depth of networks and the size of kernels used;

2) the multi-kernel context fusion is introduced from Inception series, but it cannot leverage the correlation between different objects in a global context;

3) although a new U-Net variant based on Recurrent Neutral Networks (RNNs) has been proposed to aggregate the context over local features, the implicitly captured global dependencies heavily rely on the learning outcome of the long-term memorization;

4) in addition, the relation between components from a feature map is not caught in CNNs.

Methods

FRAMEWORK

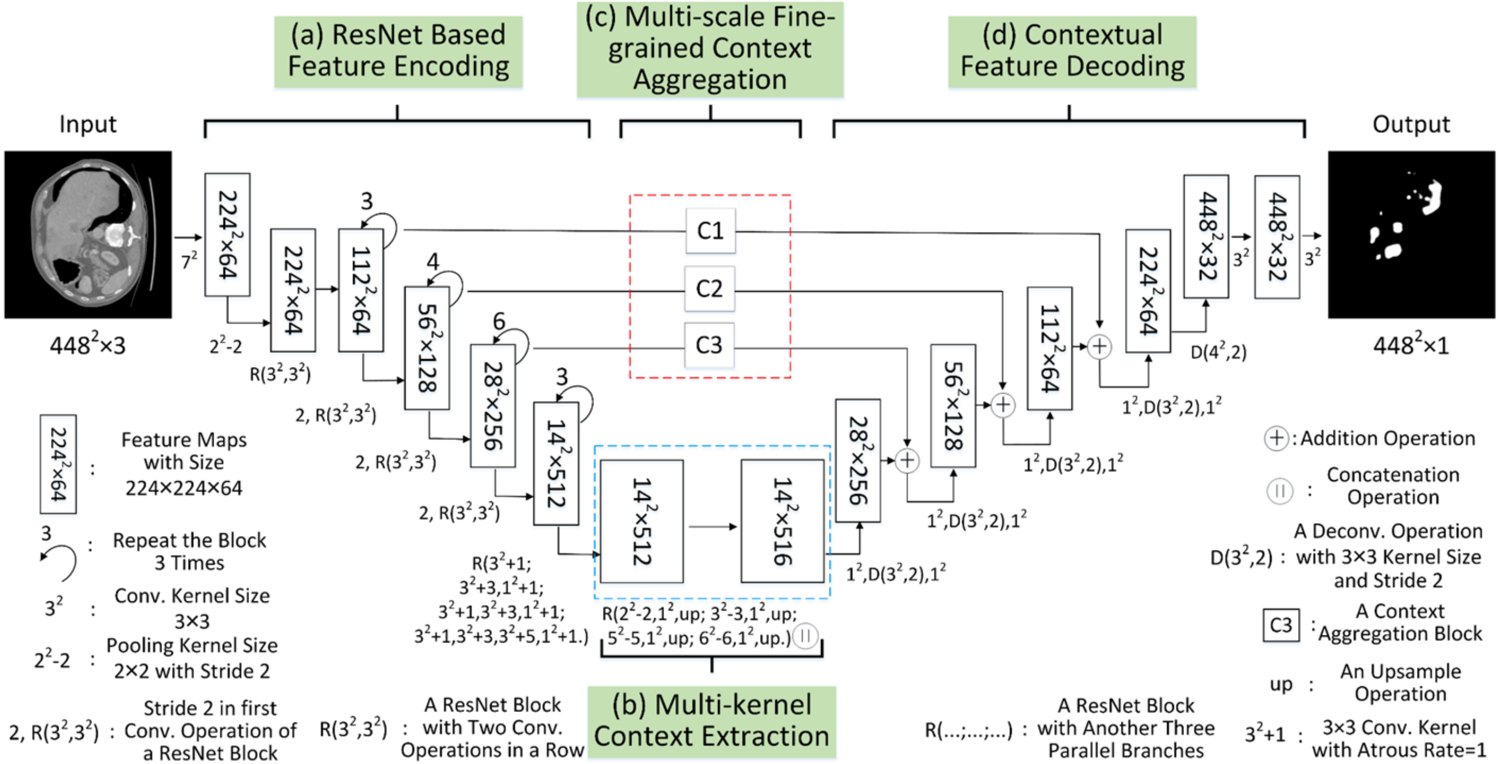


Fig. 2. Our multi-scale fine-grained contextual dependency framework for hepatic tumor segmentation, which consists of several main functional modules.

Methods

VISUALIZATION

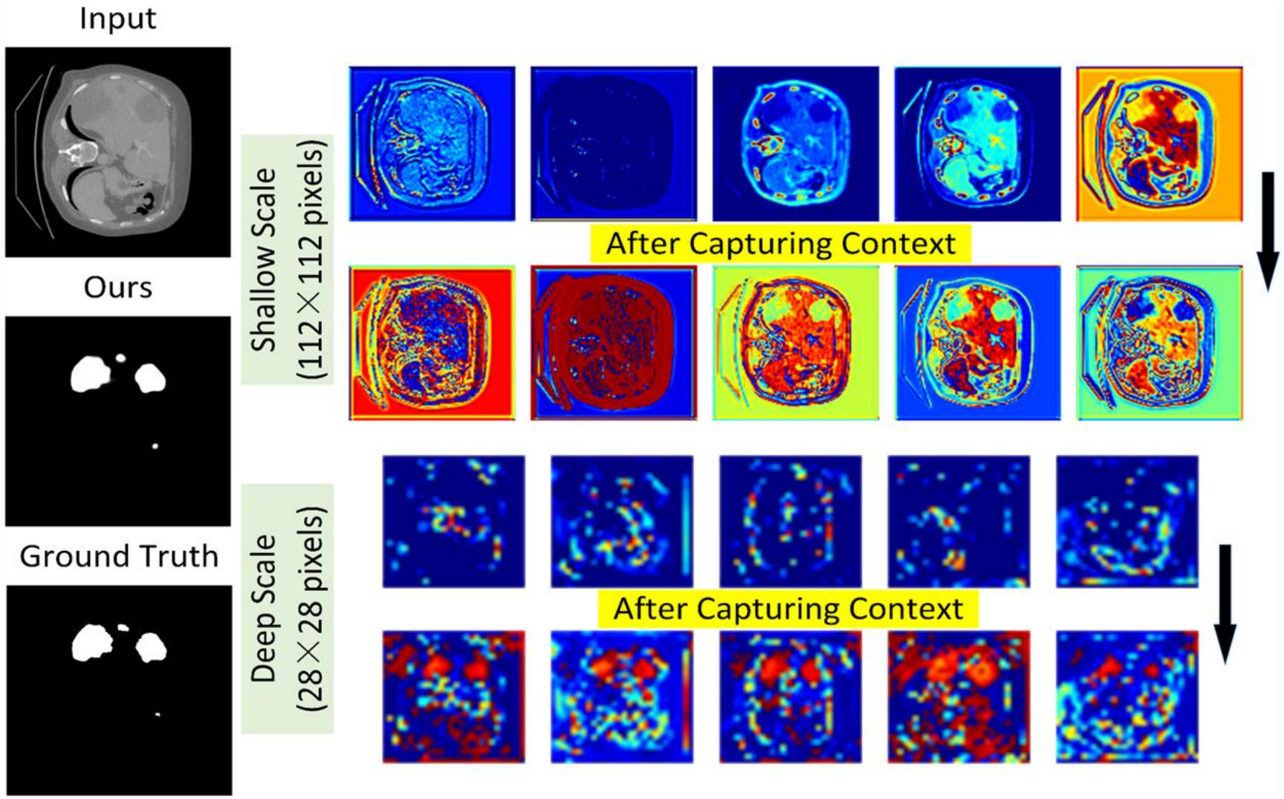
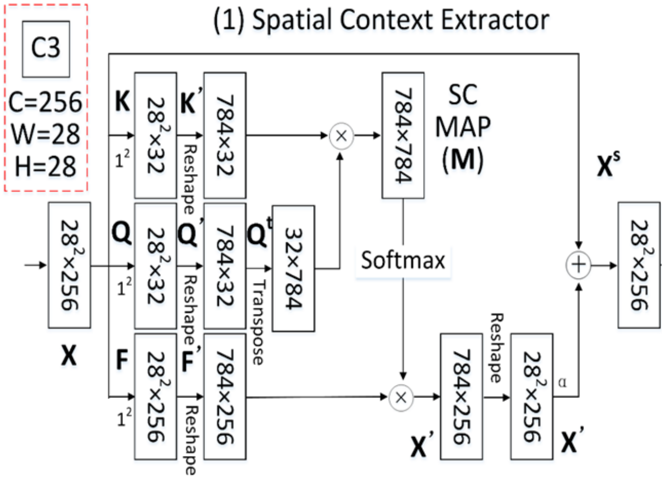


Fig. 1. A test example with our segmentation result and internal learned feature visualization comparison before and after using our multi-scale contextual dependency framework.

Methods

SPATIAL CONTEXT EXTRACTOR



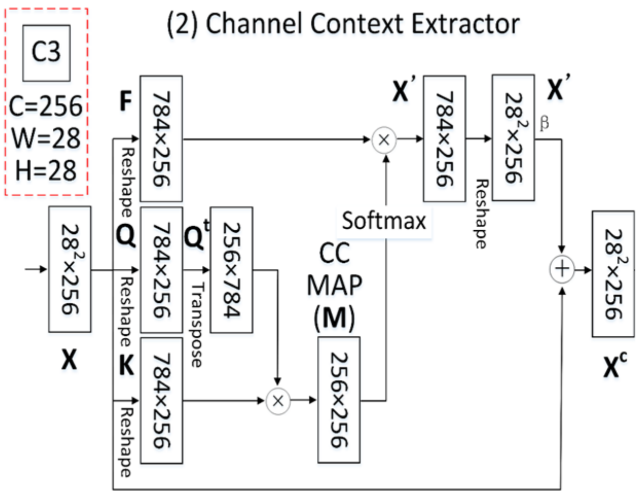
$V = \{v_i\}_{i=1:N}$ is the vertex set for each local contextual feature v_i at all 2D positions and N=W×H=28×28=784.

- Steps of Spatial Context Aggregation:

1) $Q^{(v_i)} = f\left(w_1 X^{(v_i)} + b_1\right)$; $K^{(v_i)} = f\left(w_2 X^{(v_i)} + b_2\right)$;
2) perform a 2D matrix multiplication between the $\boldsymbol{K'}$ and the transposed $\boldsymbol{Q^t}$, to calculate the mutual similarity of any two local contextual features $v_i \in \boldsymbol{K'}$ and $v_j \in \boldsymbol{Q^t}$;
3) update $\boldsymbol{M}$ with $M_{ij} = e^{v_i \cdot v_j} / \sum_{j=1}^{N} e^{v_i \cdot v_j}$;
4) generate the aggregated features $\boldsymbol{X'} \in \mathbb{R}^{(28\cdot 28)\times 256}$ by a matrix multiplication operation between $\boldsymbol{M}$ and $\boldsymbol{F'}$;
5) fuse local and global contexts: $\boldsymbol{X^s} = \alpha \boldsymbol{X'} + \boldsymbol{X}$.

Methods

CHANNEL CONTEXT EXTRACTOR



$V = \{v_i\}_{i=1:C}$ is the channel set for each channel contextual feature $v_i \in \mathbb{R}^{28 \cdot 28}$ at the third dimension.

◆ Steps of Channel Context Aggregation:

1) directly reshape $\boldsymbol{X}$ into $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{F} \in \mathbb{R}^{(28 \cdot 28) \times 256}$;
2) perform a 2D matrix multiplication between the transposed $\boldsymbol{Q}^t$ and $\boldsymbol{K}$ to calculate the channel similarity of any two channel maps over all the spatial positions: $\boldsymbol{M} = \boldsymbol{Q}^t \boldsymbol{K}$;
3) update $\boldsymbol{M}$ with $M_{ij} = e^{v_i \cdot v_j} / \sum_{j=1}^{C} e^{v_i \cdot v_j}, v_i \in \boldsymbol{Q}^t, v_j \in \boldsymbol{K}$;
4) obtain the global contextual dependency $\boldsymbol{X}'$ for each channel map by a matrix multiplication along the channel dimension: $\boldsymbol{X}' = \boldsymbol{F}\boldsymbol{M}$;
5) fuse local and global contexts: $\boldsymbol{X}^c = \beta \boldsymbol{X}' + \boldsymbol{X}$.

Methods

SPATIAL & CHANNEL CONTEXT AGGREGATIONS

- Based on an addition operation from these both context extraction steps, each context aggregation block can fully exploit contextual information in a global view from the spatial and the channel perspectives:

$$X^{\wedge} = X^{s} + X^{c}.$$

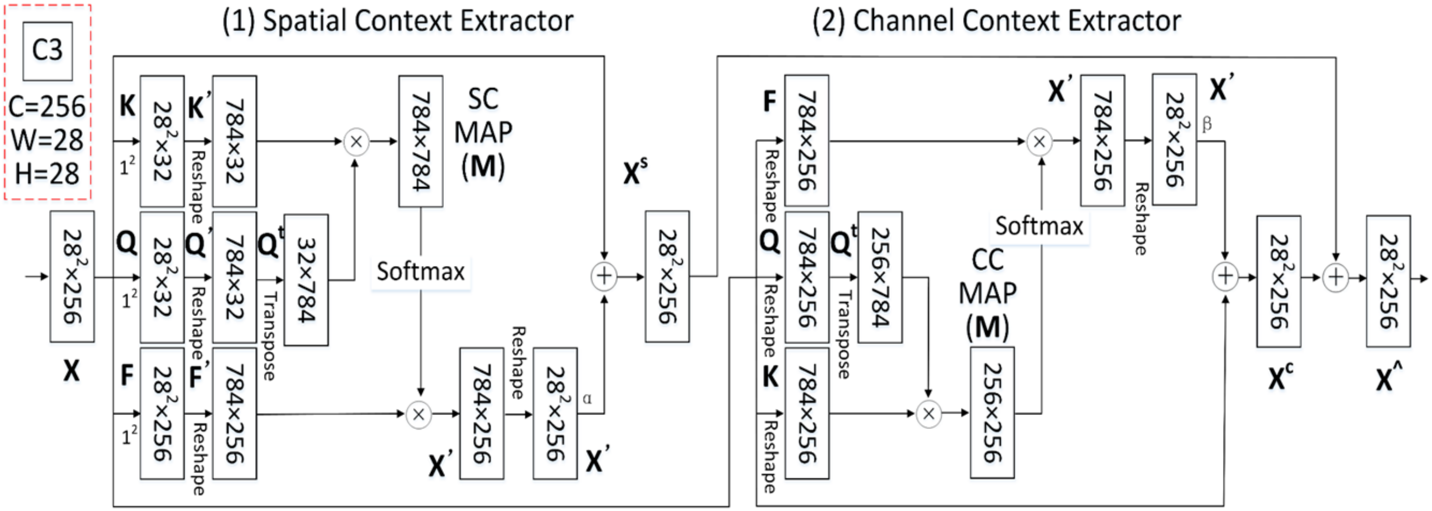


Fig. 3. A fine-grained contextual information aggregation block, taking C3 in Fig.2 as an example.

Data

HEPATIC TUMOUR DATASET & COMPARED METHODS

- This is a new challenge for hepatic tumor segmentation [5].
- The available dataset consists of 131 abdominal 3D CT scans acquired from 131 studied subjects with different types of liver tumor diseases, e.g., primary tumor diseases and secondary liver tumors.
- Here, we extracted and used 7190 CT slices with tumor annotations.
- Nine state-of-the-art segmentation methods were chosen in our experiments based on several representative models:

  1) U-Net Based Model: *U-Net [7], Attention UNet [9], Nested UNet [12]*

  2) Context Based Model: *R2U-Net [11], CE-Net [13], Self-attention [8, 10]*

  3) Attention Based Model: *SENet [15], DANet [17], CS-Net [18]*

  4) Fused Model: *Attention UNet [9], Self-attention [8, 10]*

[5] Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C.W., Han, X., Heng, P.A., Hesser, J., and Kadoury, S.: The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:1901.04056 (2019).

Results

QUANTITATIVE ANALYSIS

Table 1. Comparison results of the state-of-the-art segmentation methods with widely used evaluation metrics for hepatic tumor segmentation. Bold numbers represent the best results. Note that Hausdorff Distance uses pixel units and others %.

Methods/Metrics	Dice	Hausdorff Distance	Jaccard	Precision	Recall	Specificity	F1
U-Net [7]	73.62	52.65	63.67	67.46	86.70	99.76	75.88
Attention UNet [9]	78.70	37.13	69.32	72.93	89.65	99.83	80.43
R2U-Net [11]	74.55	46.04	64.46	68.38	87.27	99.77	76.68
Nested UNet [12]	73.58	46.89	63.55	67.39	86.95	99.76	75.93
CE-Net [13]	78.41	33.78	69.09	72.92	89.33	99.82	80.30
SENet [15]	77.88	39.09	68.55	72.39	89.16	99.82	79.90
Self-attention [8, 10]	76.49	38.78	66.80	70.82	88.72	99.79	78.76
DANet [17]	79.97	30.94	71.00	74.50	90.38	99.84	81.67
CS-Net [18]	78.90	32.90	69.45	73.03	90.03	99.83	80.64
Ours	**82.16**	**30.01**	**73.46**	**76.96**	**91.18**	**99.86**	**83.37**
Average Gain (↗)	*5.26*	*9.79*	*6.14*	*5.87*	*2.49*	*0.058*	*4.46*

Results

QUALITATIVE ANALYSIS

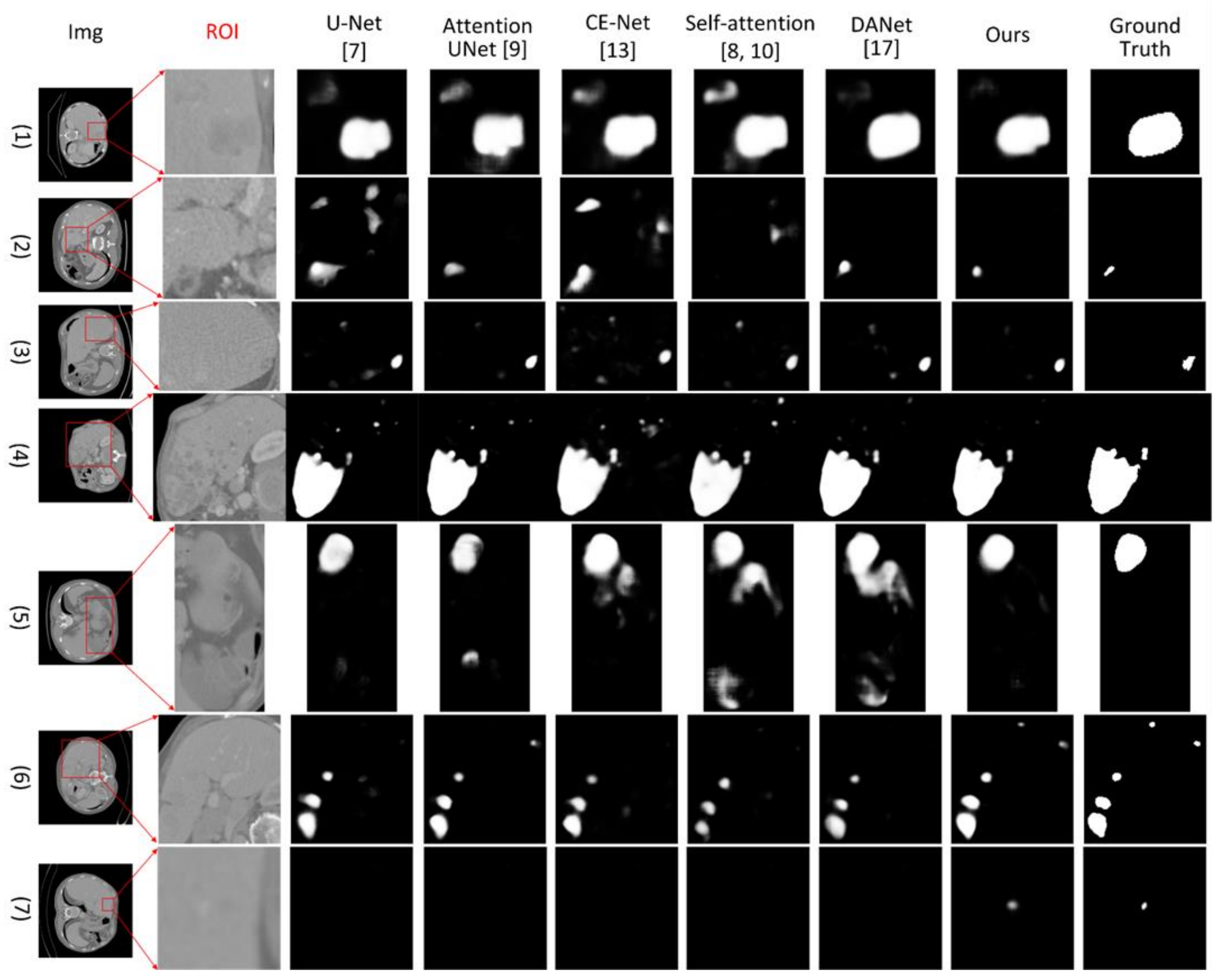


Fig. 4. Seven randomly selected samples with their segmentation results from the state-of-the-art methods. For clarity, we only report the regions of interest (ROI) of some of the compared methods due to space limitations.

Conclusions

- In this paper, we have proposed a multi-scale contextual dependency framework to explicitly capture fine-grained context correlations between tumor regions and enhance the discriminability of the learned features and hence to improve segmentation performance for hepatic tumors.
- In particular, we have modeled the semantic context dependencies over all the local features from both the spatial and channel dimensions.
- In addition, the proposed multi-scale framework features a light model with a very few additional parameters, and also its visualization capability significantly boosts networks' interpretability.
- Experimental results on a real-life liver tumor dataset have also demonstrated that our model outperforms nine compared state-of-the-art segmentation methods.
- Next, we plan to extend this framework into further clinical applications, and any popular segmentation methods seamlessly.

The End

Any Questions?