# Medical Image Seminar

Luyue Shi
May 21, 2020

Outline:

Normal Appearance Auto-encoder for Lung Cancer Detection and Segmentation (MICCAI2019)

Spatial-Frequency Non-local Convolutional LSTM Network for pRCC Classification (MICCAI2019)

MVP-Net: Multi-view FPN with Position-Aware Attention for Deep Universal Lesion Detection (MICCAI2019)

(MICCAI2019)

# Normal Appearance Auto-encoder for Lung Cancer Detection and Segmentation

Mehdi Astaraki[1,2], Iuliana Toma-Dasu[2], Örjan Smedby[1], and Chunliang Wang[1]

[1] Department of Biomedical Engineering and Health Systems,
KTH Royal Institute of Technology, Hälsovägen 11C, 14157 Huddinge, Sweden {mehast,chunwan}@kth.se

[2] Department of Oncology-Pathology, Karolinska Institutet, Karolinska Universitetssjukhuset, Solna, 17176 Stockholm, Sweden

# Introduction

- Topic

  Lung cancer detection and segmentation using CT images

- Problem

  One of the major differences between medical doctor training and machine learning is that doctors are trained to recognize normal/healthy anatomy first. Knowing the healthy appearance of anatomy structures helps doctors to make better judgement when some abnormality shows up in an image

- Key Idea

  Proposed a normal appearance autoencoder (NAA), that removes abnormalities from a diseased image

# Model

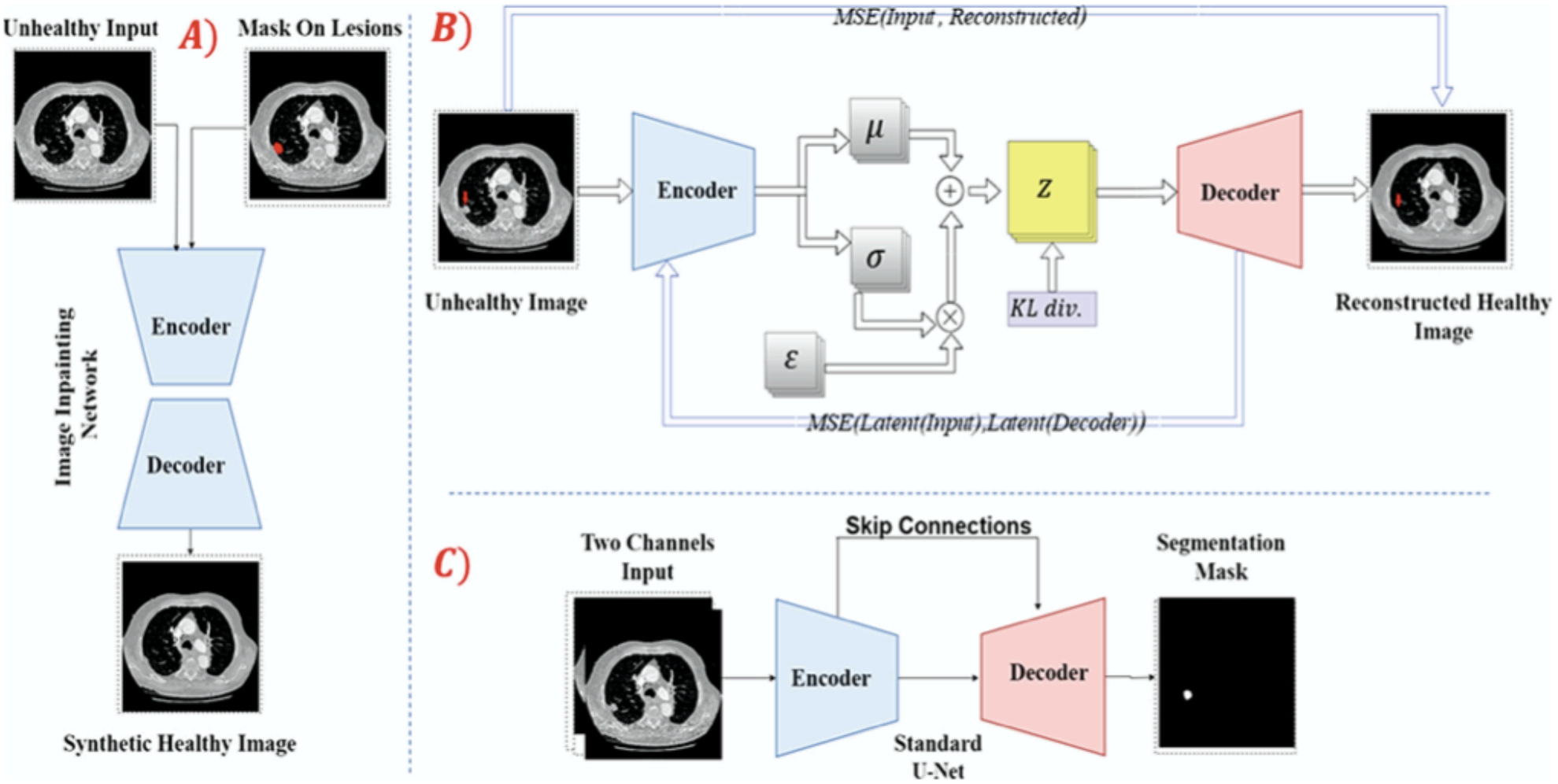
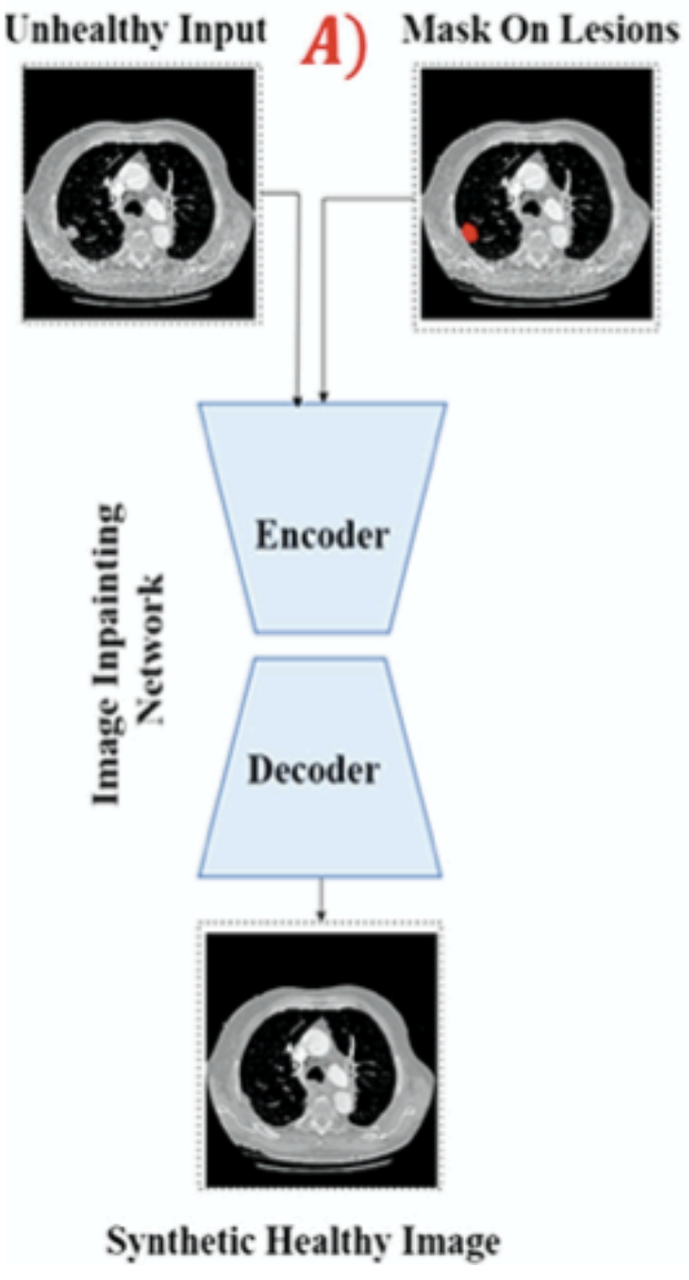



**Fig. 1.** Schematic illustration of the pipeline: (A) Partial-convolutional network for inpainting the nodules, (B) Semi-supervised fully convolutional variational autoencoder for reconstructing the unhealthy images without the nodules, (C) Two-channels U-Net for nodule segmentation.

- Inpainting model for inpainting nodules
- VAE model for reconstructing healthy images
- U-Net for lung cancer segmentation

# Model

- Inpainting model




(NVIDIA: Image Inpainting for Irregular Holes Using Partial Convolutions)

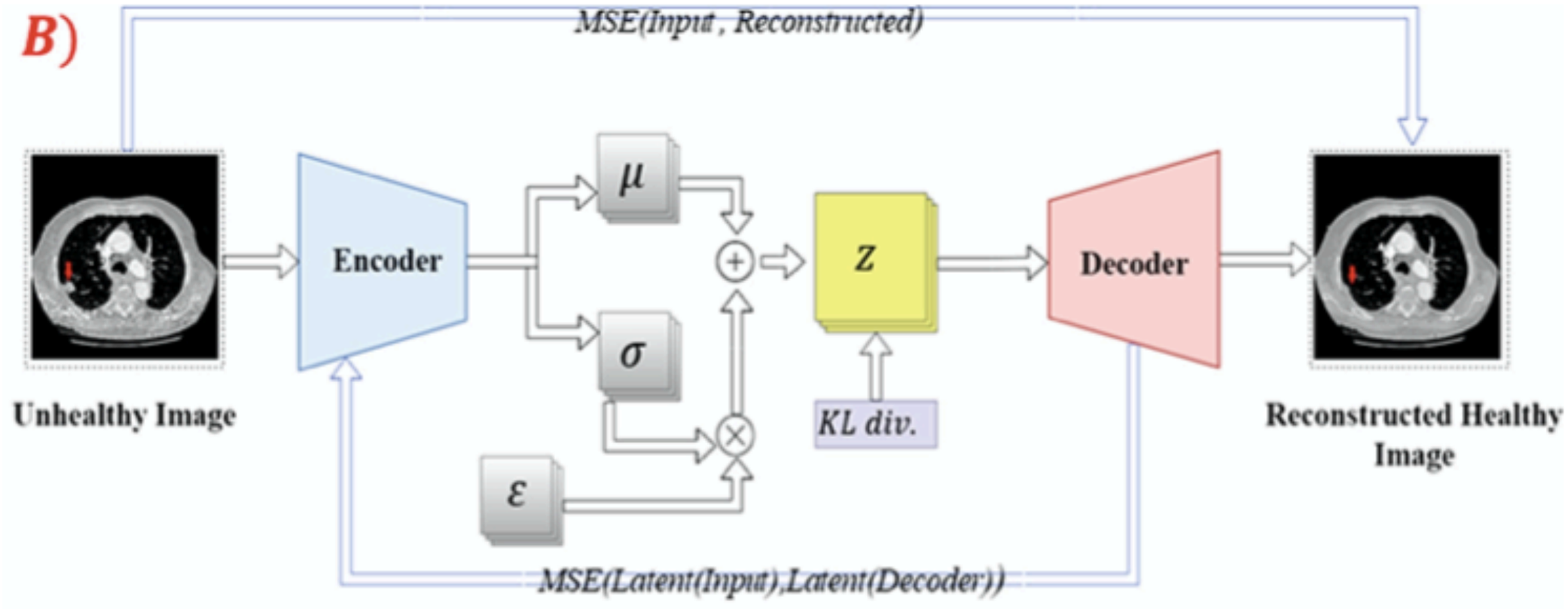
Partial Convolutional Layer:

$$x' = \begin{cases} \mathbf{W}^T(\mathbf{X} \odot \mathbf{M}) \frac{\text{sum}(\mathbf{1})}{\text{sum}(\mathbf{M})} + b, & \text{if sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$m' = \begin{cases} 1, & \text{if sum}(\mathbf{M}) > 0 \\ 0, & \text{otherwise} \end{cases}$$

# Model

- VAE model




Modified VAE:

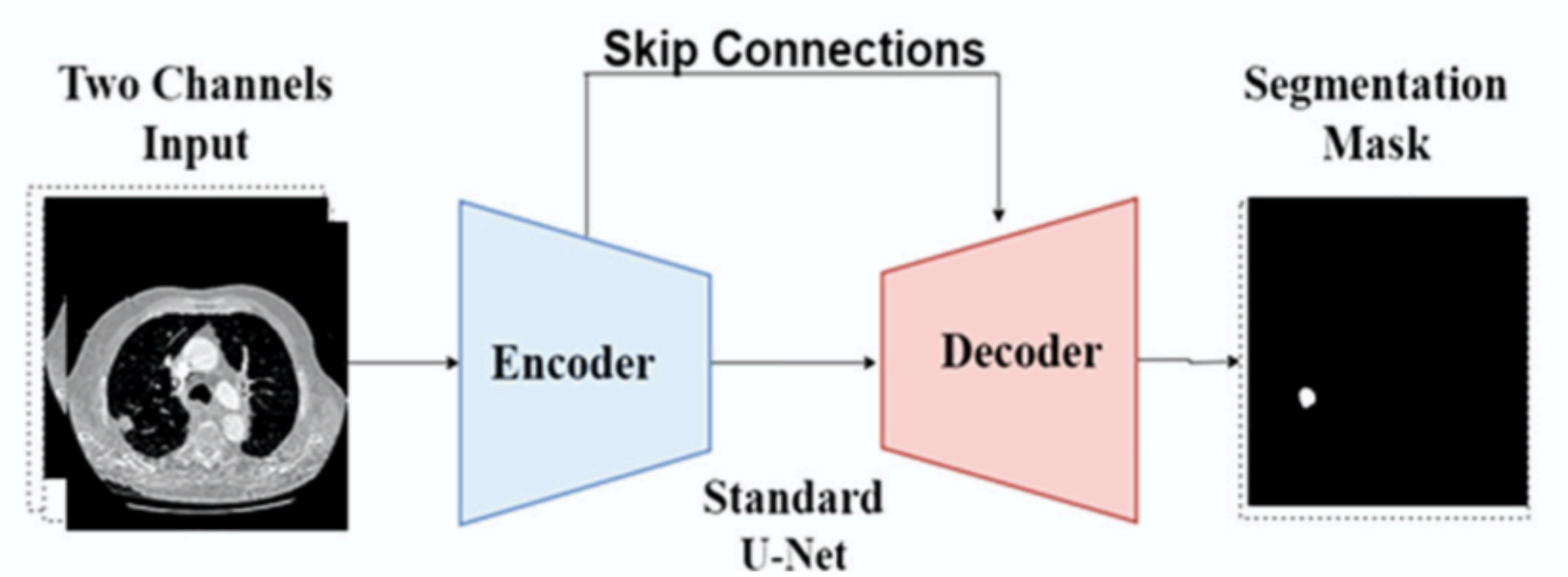
a. Using fully convolutional layers

b. Regularization term, which minimizes the differences between the latent of the input and latent of the reconstructed images

c. Weighting map, assigns lower weight in loss for the regions outside the lungs

# Model

- Segmentation model (U-Net)




Inputs: a. Original CT Image;

b. Differences between the input and reconstructed images

# Results

- Dataset

  Selected from LIDC-IDRI: 1175*3 CT slices

- Segmentation

**Table 1.** Segmentation accuracy over the test-sets. * denotes the results reported at [9] and ** refers to the reference [10]. For each case, the highest value is marked in bold.

| Method | Criteria (Mean ± SD) | | |
|---|---|---|---|
| | Dice | Sensitivity | Positive predictive value |
| A priori-based U-Net | **85.86 ± 1.24** | 83.17 ± 2.67 | **90.32 ± 2.38** |
| Single-channel U-Net | 66.69 ± 2.85 | 59.22 ± 12.72 | 78.30 ± 10.78 |
| Level Set* | 60.63 ± 17.39 | 64.38 ± 22.75 | 71.03 ± 24.35 |
| 2-D Patch Branch* | 80.47 ± 11.23 | 91.36 ± 14.40 | 74.64 ± 13.16 |
| CF-CNN* | 82.15 ± 10.76 | **92.75 ± 12.83** | 75.84 ± 13.14 |
| TLBO** | 82.73 ± 5.41 | 87.29 ± 9.52 | 86.03 ± 18.84 |

# Results

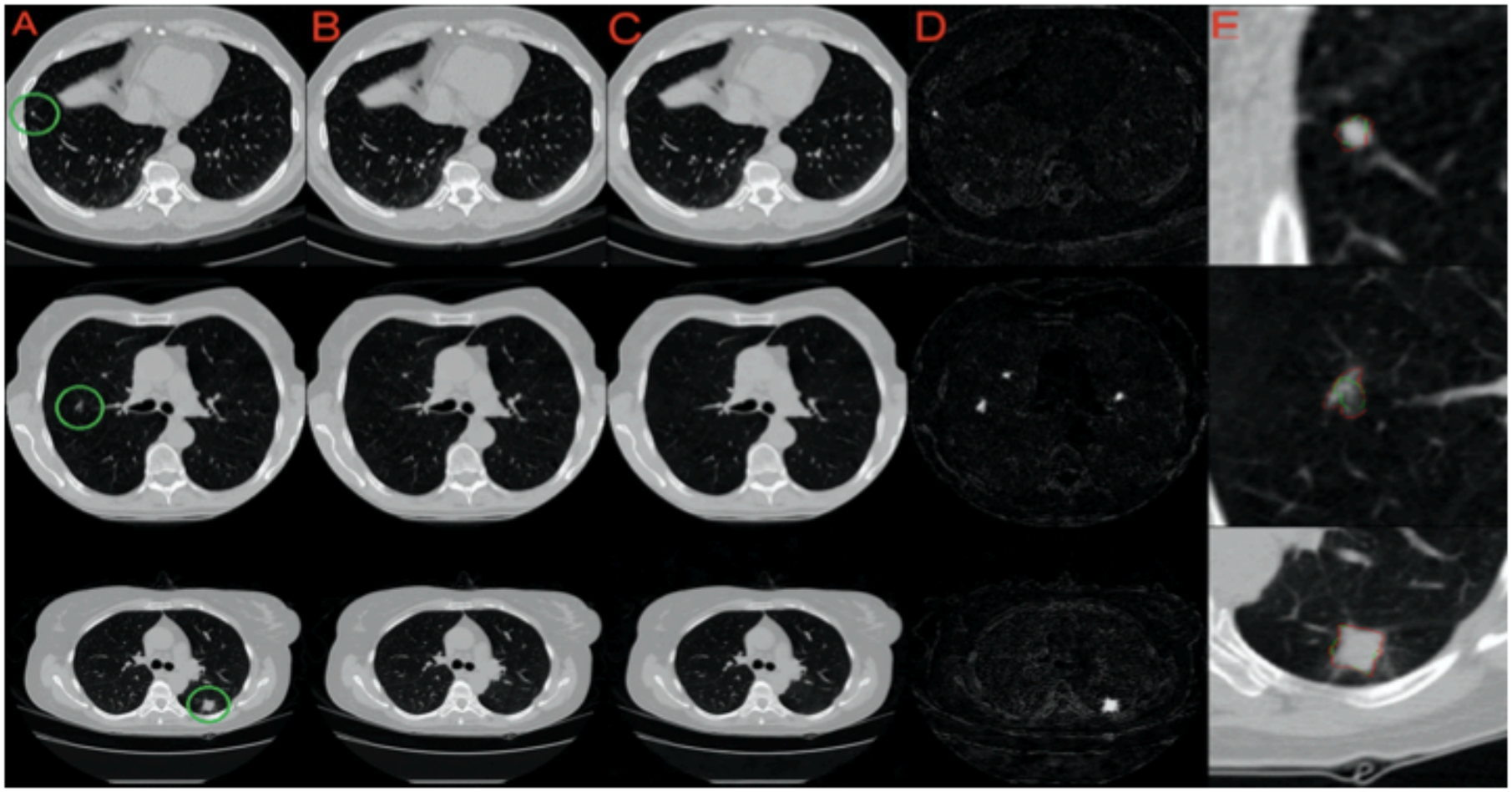



**Fig. 2.** Step by step examples of the results. Column A: represents the original input images containing nodules. Column B: shows the results of synthetic healthy images from inpainting model. Column C: represents the high-quality reconstructed images without the nodules from the VAE model. Column D: shows the differences between the input images and reconstructed images. Column E: segmentations resulted from two-channels U-Net model; the images are magnified around the nodules to better visualize the agreement between the manual (green) and algorithm (red) contours. (Color figure online)

(MICCAI2019)

# Spatial-Frequency Non-local Convolutional LSTM Network for pRCC Classification

Yu Zhao[1], Yuan Liu[1], Yansheng Kan[2], Anjany Sekuboyina[1], Diana Waldmannstetter[1], Hongwei Li[1], Xiaobin Hu[1(B)], Xiaozhi Zhao[2], Kuangyu Shi[3], and Bjoern Menze[1]

[1] Department of Computer Science, Technische Universit¨at Mu¨nchen, Munich, Germany
xiaobin.hu@tum.de

[2] Urology Department, The Affiliated Nanjing Drum Tower Hospital, Nanjing University Medical School, Nanjing, Jiangsu, China

Department of Nuclear Medicine, University of Bern, Bern, Switzerland

# Introduction

- Topic

  Classification of papillary renal cell carcinoma (pRCC) using CT images

- Key Idea

  a. Extract features from both the spatial and frequency domains
  b. LSTM for learning inter-slice 3D context information
  c. Non-local block for capturing long-range dependencies

# Model

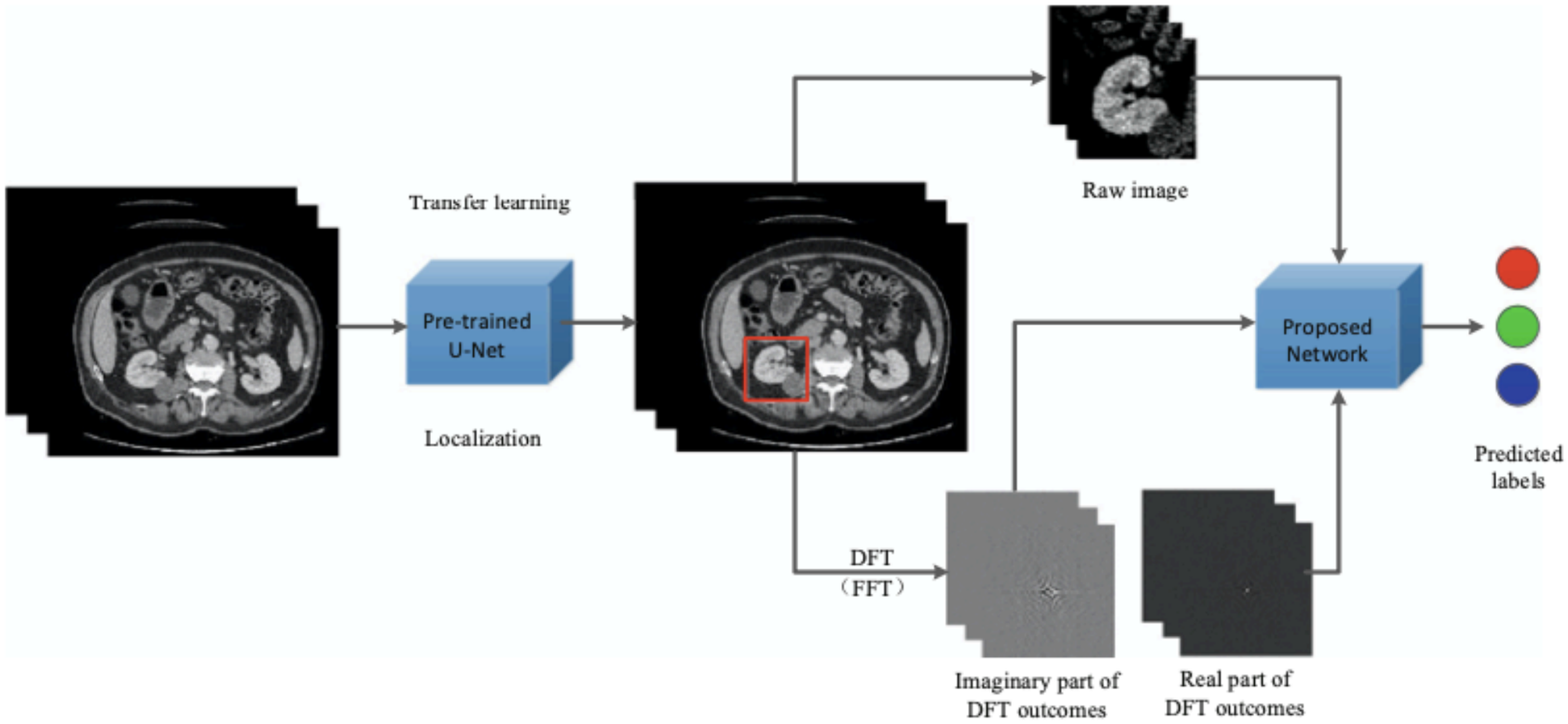



**Fig. 1.** The overall framework of the proposed approach. It consists of a localization step and a classification step. The localization stage leverages the idea of transfer learning to propose the ROI for the following classification stage. While, in the classification stage, the proposed network synthesizes features from both the spatial and frequency domains for classification.

- U-Net for predicting the kidney region
- Proposed spatial-frequency non-local convolutional LSTM network for predicting the pRCC subtype

# Model

- Proposed spatial-frequency non-local convolutional LSTM network

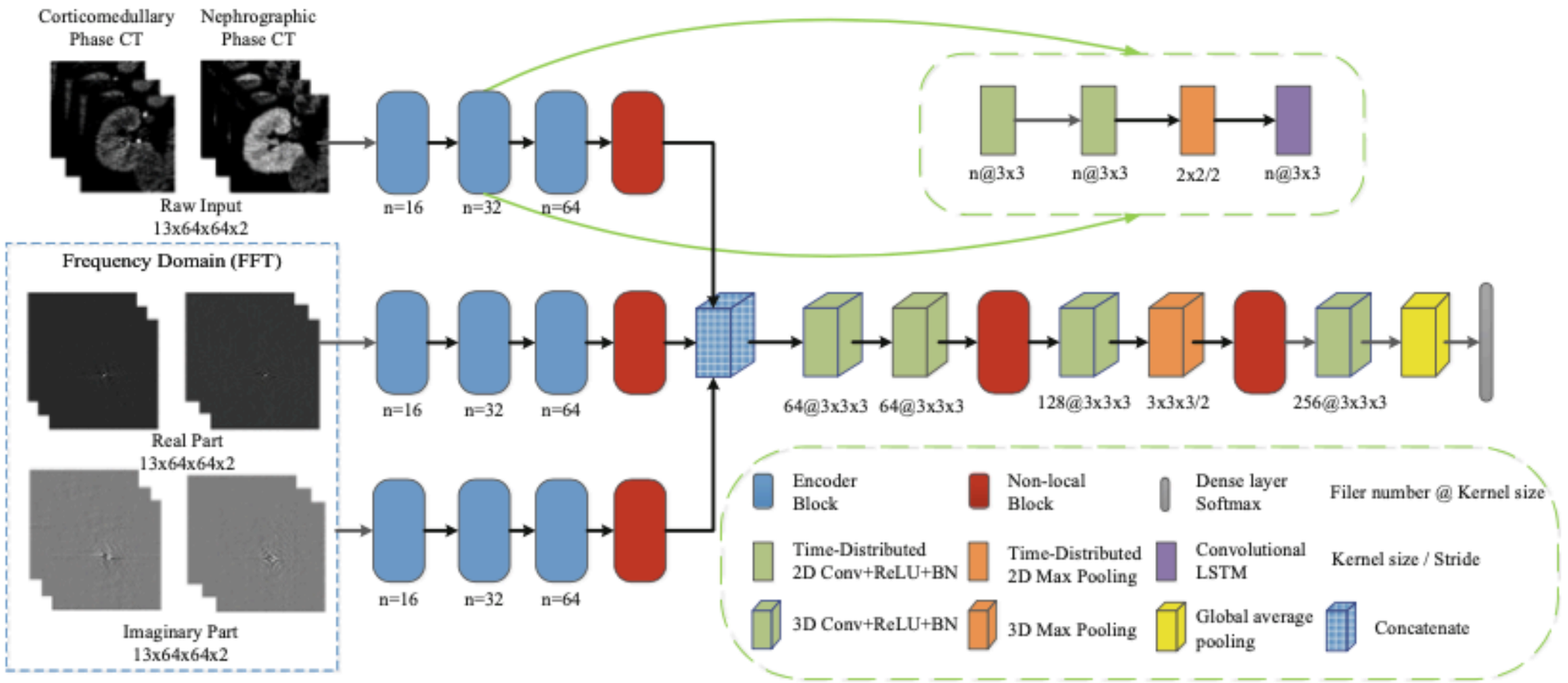



**Fig. 2.** Proposed spatial-frequency non-local convolutional LSTM network. It consists of a feature-extraction component and a fusion component, which are combined by the concatenation layer. Different operations are denoted by different colors. The details of each operation such as kernel size and filter number are shown below the graph (Conv: Convolutional layer, BN: batch normalization, ReLU: Rectified Linear Unit). (Color figure online)

# Model

- Proposed spatial-frequency non-local convolutional LSTM network

Non-local Block (CVPR2018: Non-local Neural Networks):

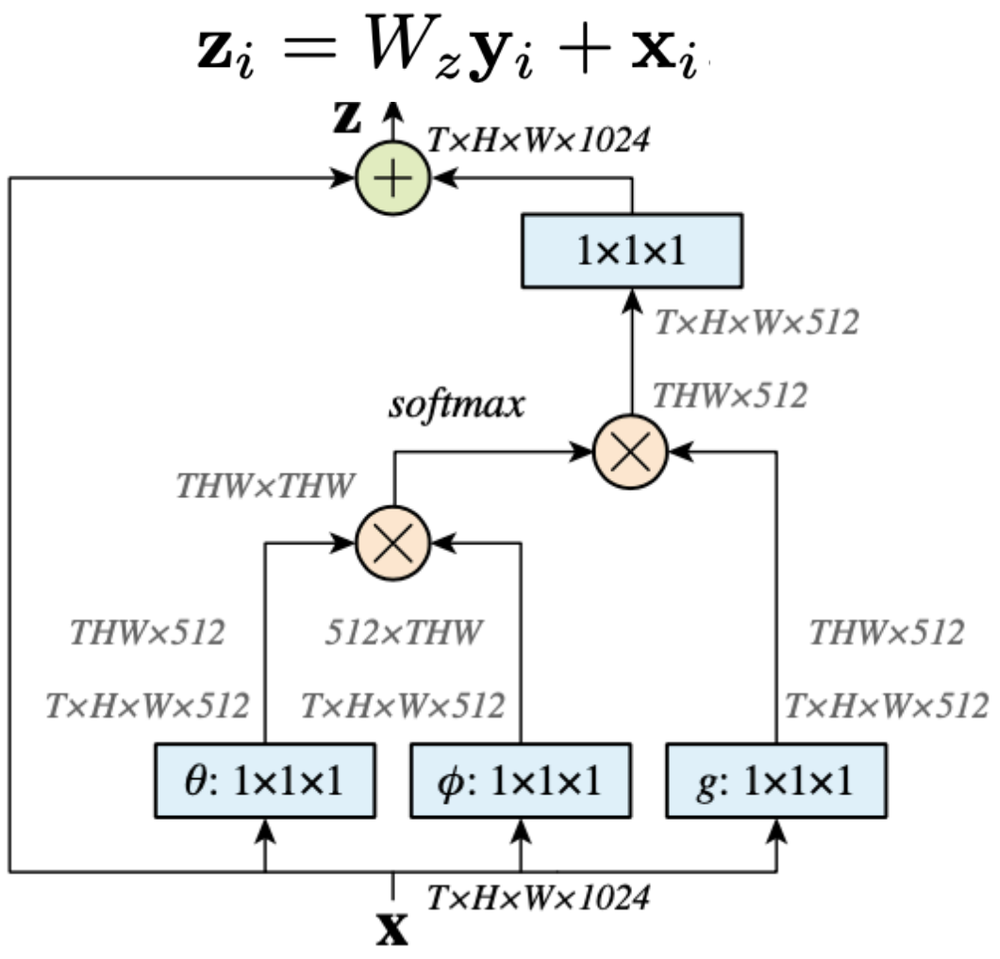
$$\mathbf{y}_i = \frac{1}{\mathcal{C}(\mathbf{x})} \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_j)$$

$$g(\mathbf{x}_j) = W_g \mathbf{x}_j$$

Embedded Gaussian:

$$f(\mathbf{x}_i, \mathbf{x}_j) = e^{\theta(\mathbf{x}_i)^T \phi(\mathbf{x}_j)}$$

$$\mathbf{y} = softmax(\mathbf{x}^T W_\theta^T W_\phi \mathbf{x}) g(\mathbf{x})$$



# Results

- Dataset

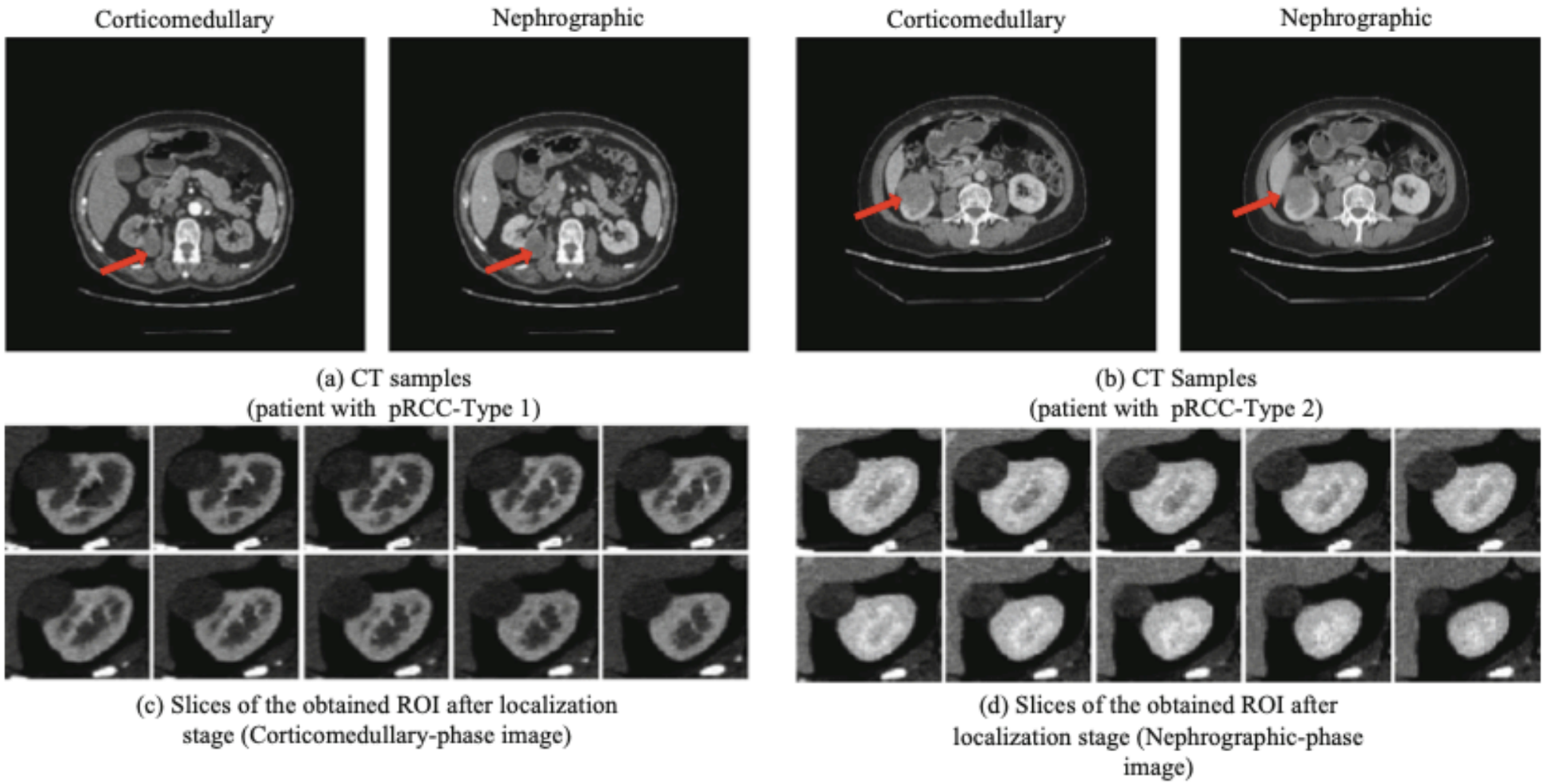



**Fig. 3.** Examples illustrating the raw CT images (corticomedullary phase and nephrographic phase) as well as the obtained ROIs after localization stage.

# Results

- Classification

**Table 1.** Ablation study results and comparison between the proposed network and sate-of-the-art methods

| Method | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|
| Our | **84.2** | **80.3** | **85.7** |
| Spatial-Frequency C-LSTM | 80.7 | 71.2 | 81.6 |
| Spatial nonlocal C-LSTM | 77.2 | 66.7 | 78.6 |
| Spatial C-LSTM | 70.2 | 63.6 | 71.4 |
| 3D ResNet | 61.4 | 46.7 | 62.5 |
| 3D DenseNet | 66.7 | 60.3 | 69.0 |

(MICCAI2019)

# MVP-Net: Multi-view FPN with Position-Aware Attention for Deep Universal Lesion Detection

Zihao Li[1,2], Shu Zhang[3], Junge Zhang[1], Kaiqi Huang[1], Yizhou Wang[2,3,4], and Yizhou Yu[2(B)]

[1] Institute of Automation, Chinese Academy of Sciences, Beijing, China [2] Deepwise AI Lab, Beijing, China

yuyizhou@deepwise.com

[3] Department of Computer Science, Peking University, Haidian District, China [4] Peng Cheng Laboratory, Shenzhen, China

# Introduction

- Topic

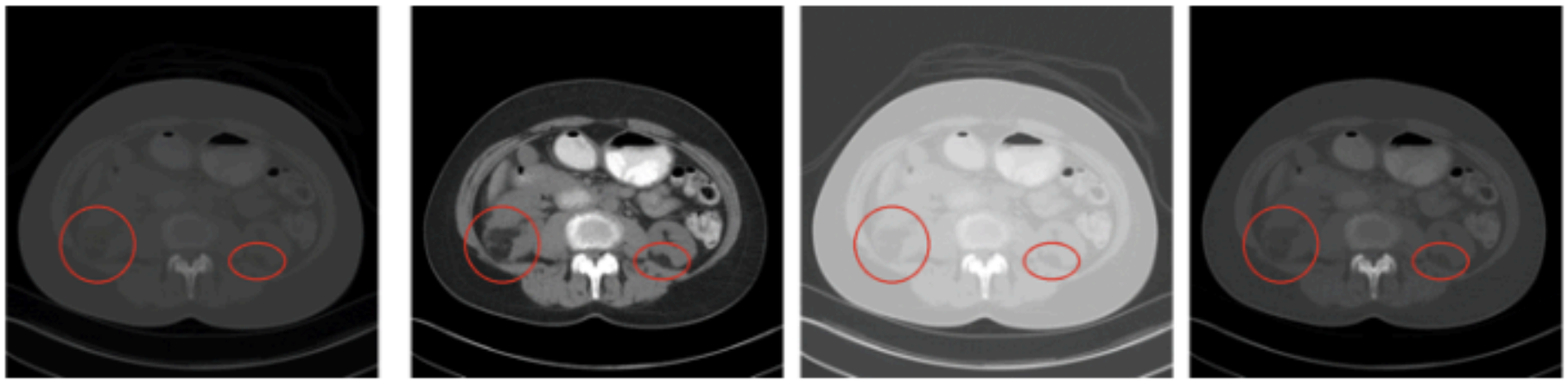
  Deep Universal Lesion Detection using CT Images

- Key Idea

  a. As radiologists tend to inspect multiple windows for an accurate diagnosis, proposed a multi-view FPN.

  b. Proposed a position-aware attention module to effectively combine the multi-view information

# Introduction



(a) [1024, 4096] (b) [50, 449] (c) [-505, 1980] (d) [446, 1960]

**Fig. 1.** CT images under different window level and window width. (a) is the image used in 3DCE. (b), (c), (d) are the multi-view images used in our MVP-Net.

# Model

- Part A: Multi-view FPN
- Part B: Position(z-axis) prediction
- Part C: Feature aggregation
- Part D: RPN+RCNN

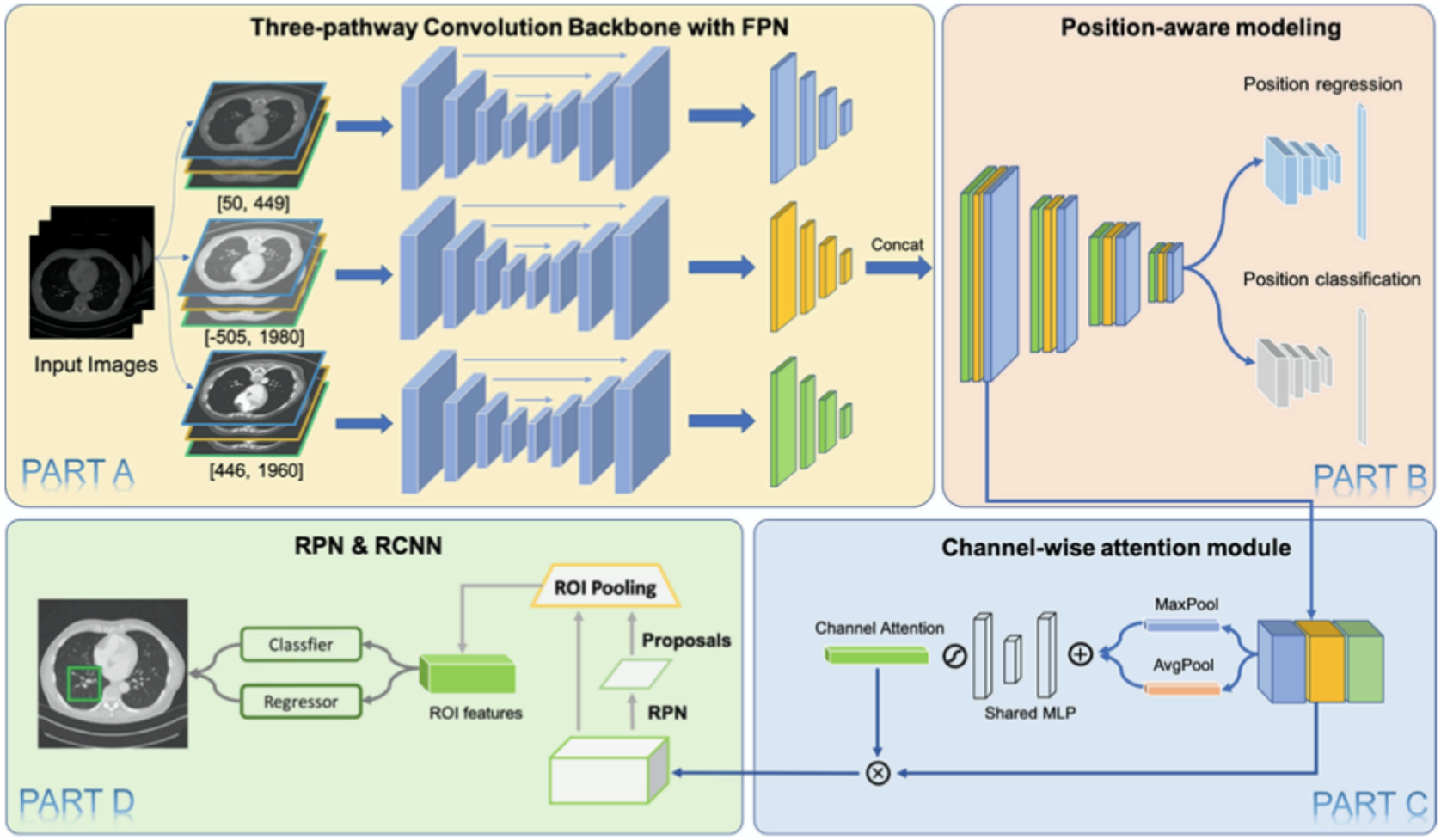



**Fig. 2.** Overview of our proposed MVP-Net. Coarser feature maps of FPN are omitted in part C and D for clarity, they use the same attention module with shared parameters for feature aggregation.

# Model

- Part A: Multi-view FPN

  K-means algorithm to cluster the recommended windows
  Multi-view FPN

- Part C: Feature aggregation

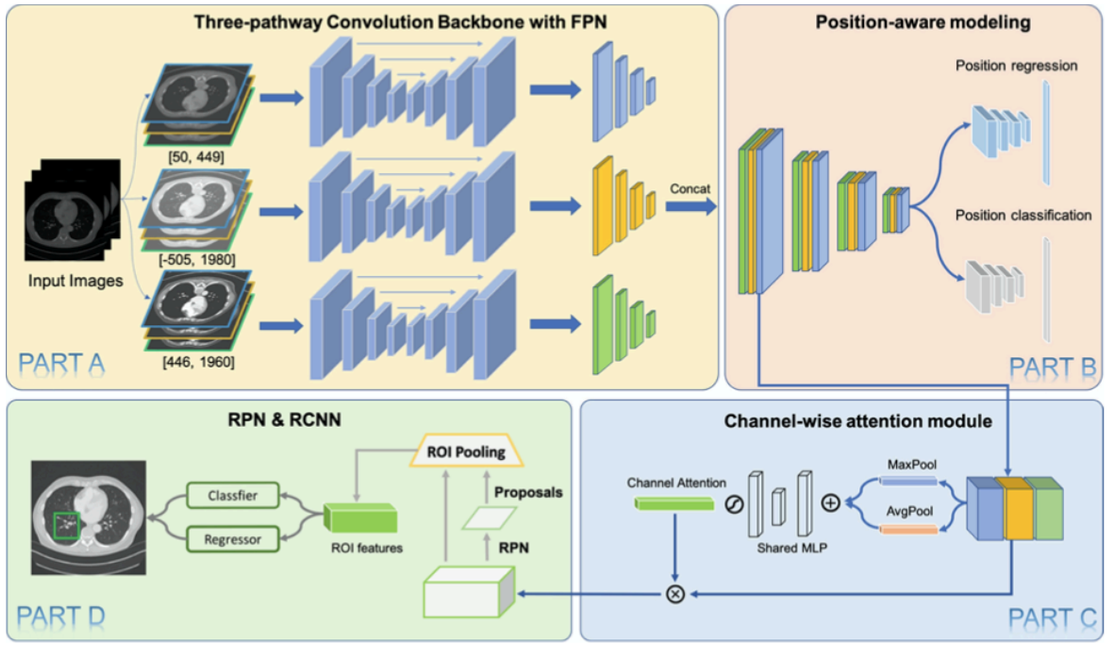
$$\boldsymbol{F_c} = \boldsymbol{F} \cdot \sigma(\theta(P_{avg}(\boldsymbol{F}) + P_{max}(\boldsymbol{F})))$$

- Part D: RPN+RCNN

- Part C: Feature aggregation

  Continuous position regressor

  Discrete position classification(chest, abdomen, and pelvis)



# Results

- Dataset

  NIH DeepLesion, 32,735 lesion instances on 32,120 axial CT slices

- Comparison with SOTA

**Table 1.** Sensitivity (%) at various FPs per image on the testing set of DeepLesion. We don't provide results with 27 slices due to memory limitation. * indicates re-implementation of 3DCE with FPN as backbone.

| FPs per image | 0.5 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| ULDOR [5] | 52.86 | 64.80 | 74.84 | - | 84.38 |
| 3DCE, 3 slices [6] | 55.70 | 67.26 | 75.37 | - | 82.21 |
| 3DCE, 9 slices [6] | 59.32 | 70.68 | 79.09 | - | 84.34 |
| 3DCE, 27 slices [6] | 62.48 | 73.37 | 80.70 | - | 85.65 |
| FPN+3DCE, 3 slices* | 58.06 | 68.85 | 77.48 | 81.03 | 83.27 |
| FPN+3DCE, 9 slices* | 64.25 | 74.41 | 81.90 | 85.02 | 87.21 |
| FPN+3DCE, 27 slices* | 67.32 | 76.34 | 82.90 | 85.67 | 87.60 |
| **Ours, 3 slices** | **70.01** | **78.77** | **84.71** | **87.58** | **89.03** |
| **Ours, 9 slices** | **73.83** | **81.82** | **87.60** | **89.57** | **91.30** |
| **Imp over** 3DCE, 27slices [6] | ↑ **11.35** | ↑8.45 | ↑6.90 | **-** | ↑5.65 |

# Results

- Ablation Study

**Table 2.** Ablation study of our approach on the DeepLesion dataset.

| FPN | Multi-view | Attention | Position | 9 slices | **FPs@2.0** | **FPs@4.0** |
|---|---|---|---|---|---|---|
| ✓ | | | | | 77.48 | 83.27 |
| ✓ | ✓ | | | | 81.29 | 86.18 |
| ✓ | ✓ | ✓ | | | 84.18 | 87.89 |
| ✓ | ✓ | ✓ | ✓ | | 84.71 | 89.03 |
| ✓ | ✓ | ✓ | ✓ | ✓ | **87.60** | **91.30** |

# Results

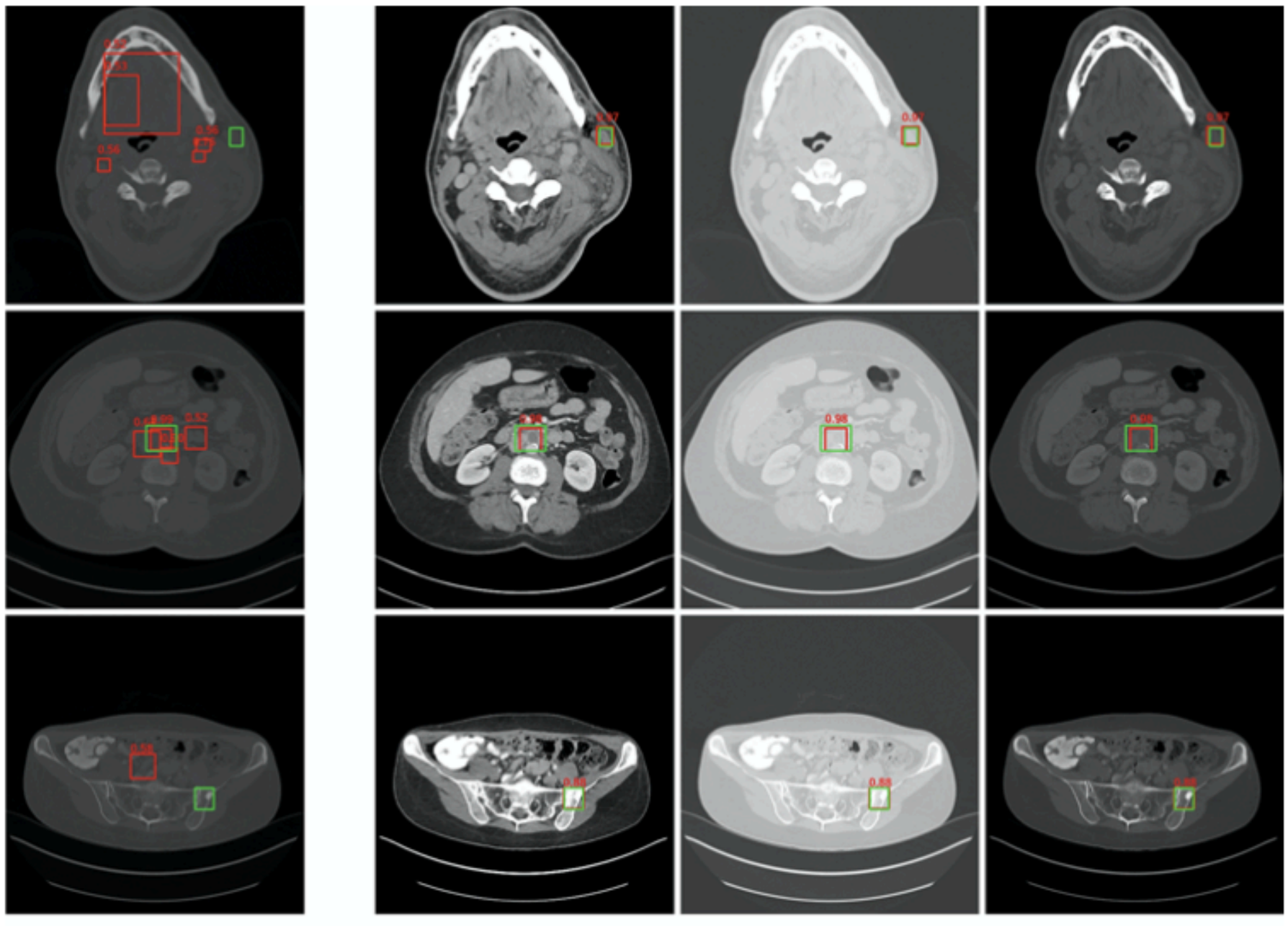



**Fig. 3.** Case study for 3DCE (left-most column) and attention based multi-view modeling (the other three columns). Green and red boxes correspond to ground-truths and predictions respectively. (Color figure online)