Medical Image Seminar

Luyue Shi
June 11, 2020

Outline:

Attentive CT Lesion Detection Using Deep Pyramid Inference with Multi-scale Booster (MICCAI2019)

Improving Deep Lesion Detection Using 3D Contextual and Spatial Attention (MICCAI2019)

Probabilistic Radiomics: Ambiguous Diagnosis with Controllable Shape Analysis (MICCAI2019)

(MICCAI2019)

Attentive CT Lesion Detection Using Deep Pyramid Inference with Multi-scale Booster

Qingbin Shao[1,2], Lijun Gong[1(B)], Kai Ma[1], Hualuo Liu[2], and Yefeng Zheng[1]

[1] Tencent Youtu Lab, Shenzhen, China lijungong@tencent.com

[2] Jilin University, Changchun, China

Introduction

- Topic

  Universal CT Lesion Detection

- Problem

  Various lesion scales(In the DeepLesion dataset, the lesion size ranges from 0.21 mm to 342.5 mm)

Introduction

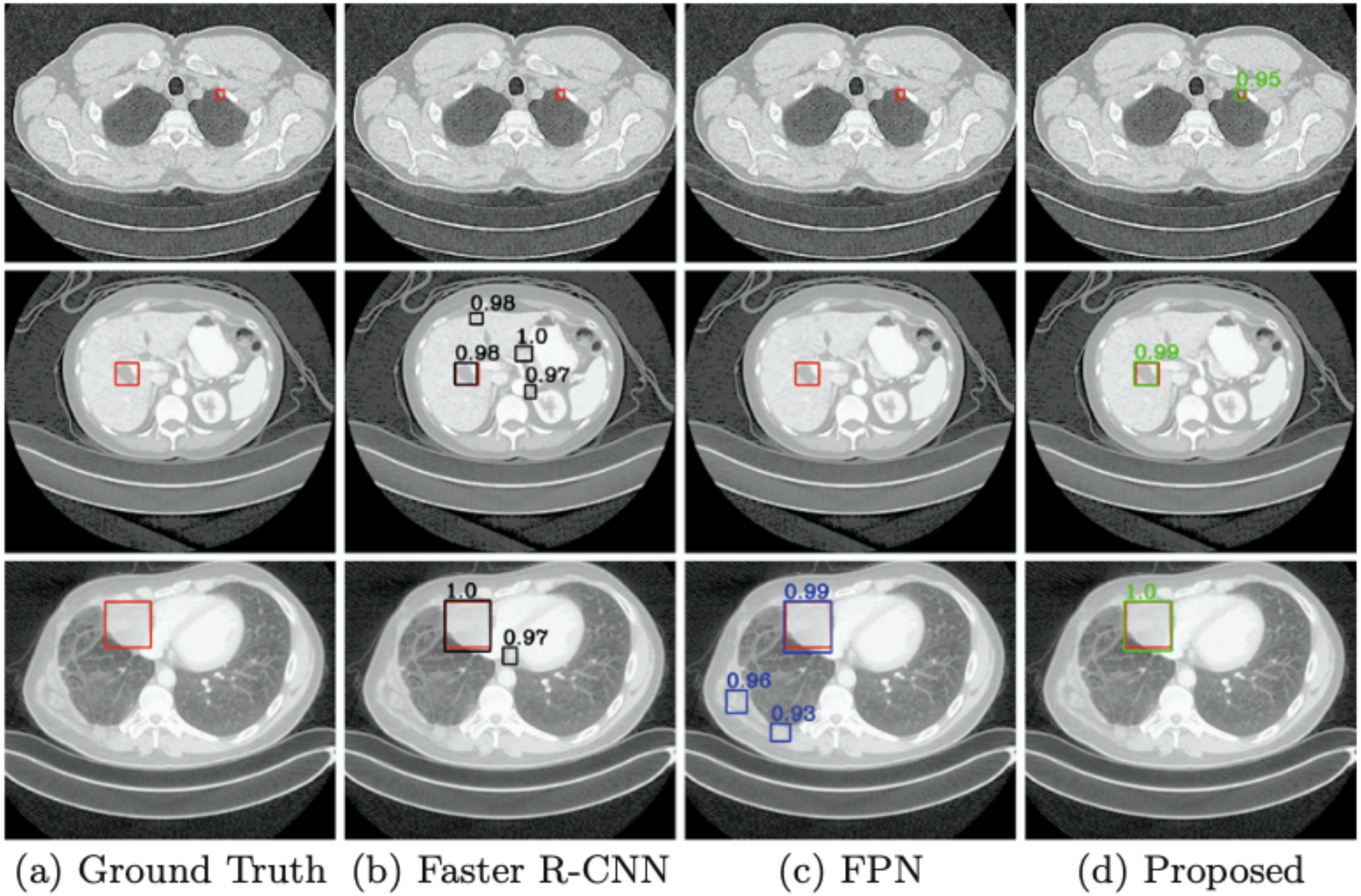


Fig. 1. Lesion detection results. The red bounding boxes represent ground truth annotations. The black, blue and green bounding boxes are the predicted results by Faster R-CNN [8], FPN [7] and the proposed method, respectively. (Color figure online)

Model

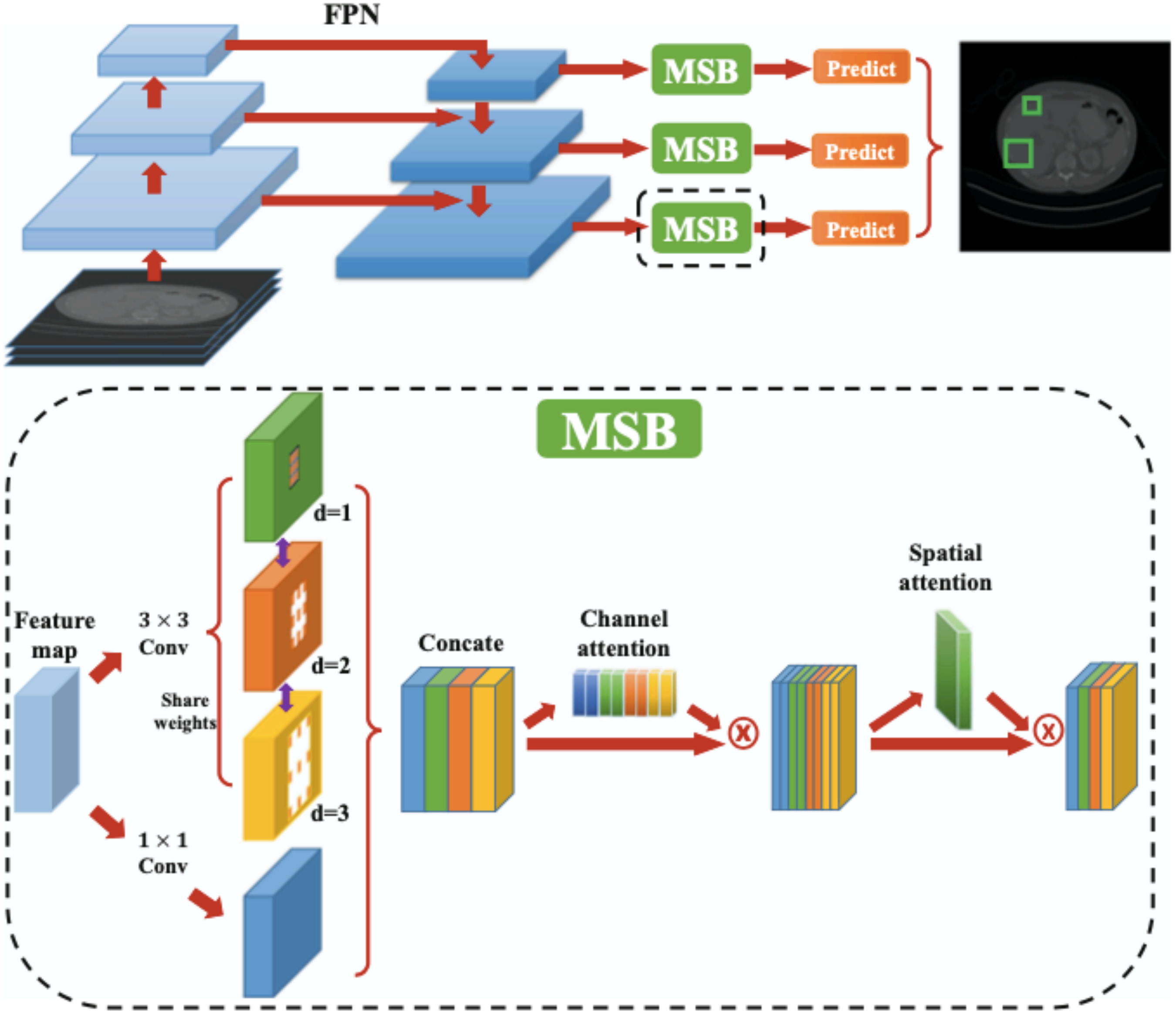


Fig. 2. Frameworks of the proposed approach. The detailed architecture of the Multi-Scale Booster (MSB) module is shown in the second row.

- FPN
- MSB(Dilated Convolution + Channel Attention + Spatial Attention)

Model

- MSB

A. Hierarchically Dilated Convolution

B. Channel Attention (CVPR2018 Squeeze-and-Excitation Networks)

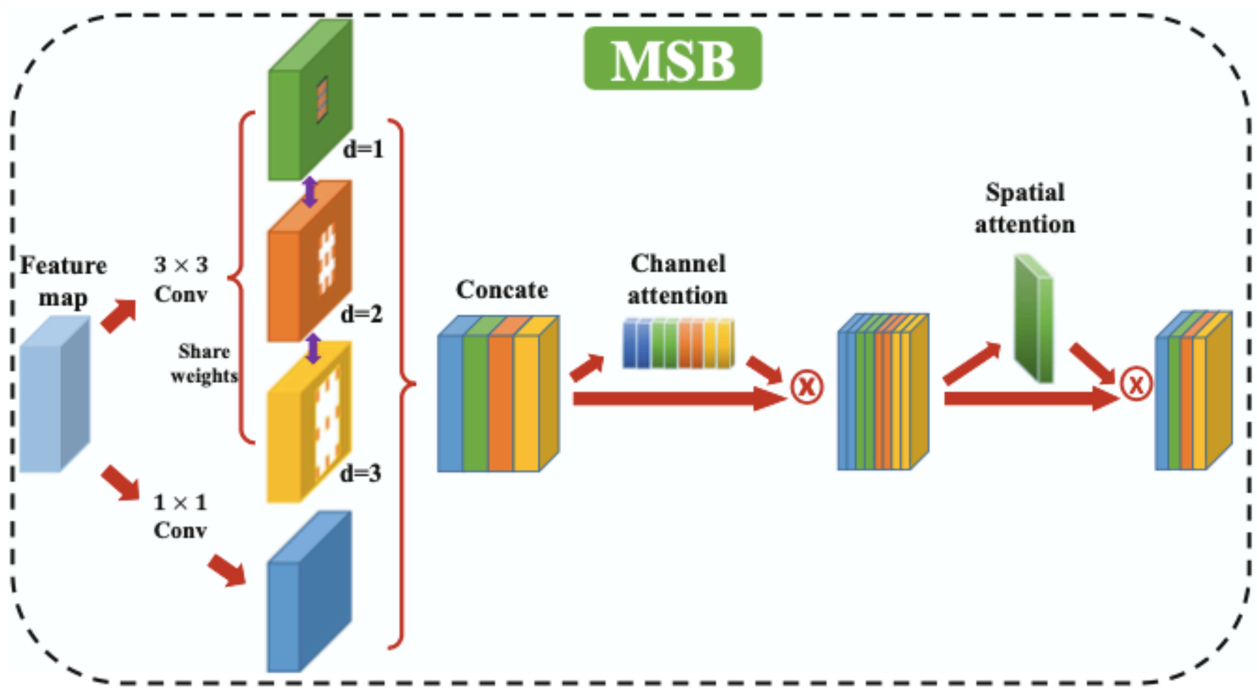
$$\mathcal{F}_{ch}(H_i^D) = \mathcal{P}_{avg}(H_i^D) * W_{1\times 1}$$

$$H_i^{Dch} = \mathcal{F}_{ch}(H_i^D) \otimes H_i^D$$

C. Spatial attention

$$\mathcal{F}_{sp}(H_i^{Dch}) = \mathcal{P}_{max}(H_i^{Dch})$$

$$\hat{P}_i = \mathcal{F}_{sp}(H_i^{Dch}) \otimes H_i^{Dch}$$



Results

- Dataset

  DeepLesion, 32735 lesions from 32120 CT slices

- Ablation Study

Table 1. An ablation study with various configurations of the proposed modules. Lesion detection sensitivity is reported at different false positive (FP) rates on the DeepLesion [14] test set.

Method	Backbone	FPs per image				
		0.5	1	2	4	8
FPN	ResNet-50	0.621	0.728	0.807	0.864	0.890
FPN+HDC (weights sharing)	ResNet-50	0.622	0.734	0.818	0.873	0.910
FPN+HDC+CH (weights sharing)	ResNet-50	0.645	0.746	0.820	0.880	0.911
FPN+HDC+SP (weights sharing)	ResNet-50	0.629	0.743	0.821	0.881	0.914
FPN+MSB	ResNet-50	0.637	0.748	0.819	0.871	0.917
FPN+MSB (weights sharing)	ResNet-50	**0.670**	**0.768**	**0.837**	**0.890**	**0.920**

Results

- Comparison with SOTA

Table 2. Comparison of the proposed method (FPN + MSB) with state-of-the-art methods on the DeepLesion [14] test set. Lesion detection sensitivity values are reported at different false positive (FP) rates.

Method	Backbone	Number of slices	FPs per image				
			0.5	1	2	4	8
3DCE [13]	VGG-16	3	0.569	0.673	0.756	0.816	0.858
	VGG-16	9	0.593	0.707	0.791	0.843	0.878
	VGG-16	27	0.625	0.737	0.807	0.857	0.891
Faster R-CNN [8]	ResNet-50	3	0.560	0.677	0.763	0.832	0.867
FPN [7]	ResNet-50	3	0.621	0.728	0.807	0.864	0.890
FPN+MSB (weights sharing)	ResNet-50	3	**0.670**	**0.768**	**0.837**	**0.890**	**0.920**

Results

- Comparison with SOTA

Table 3. Sensitivity values at four false positives per image on five test subsets categorized by different lesion size.

Method	Backbone	Number of slices	Lesion diameters (mm)				
			<10	10–30	30–60	60–100	>100
3DCE [13]	VGG-16	27	0.78	0.86	0.84		
Faster R-CNN [8]	ResNet-50	3	0.77	0.86	0.81	0.88	0.72
FPN [7]	ResNet-50	3	0.83	0.88	0.82	0.91	0.77
FPN+HDC (weights sharing)	ResNet-50	3	0.85	0.89	**0.88**	**0.93**	0.79
FPN+MSB (weights sharing)	ResNet-50	3	**0.86**	**0.91**	0.86	**0.93**	**0.86**

(MICCAI2019)

Improving Deep Lesion Detection Using 3D Contextual and Spatial Attention

Qingyi Tao[1,2(B)], Zongyuan Ge[1,3], Jianfei Cai[2], Jianxiong Yin[1], and Simon See[1]

[1] NVIDIA AI Technology Center, Santa Clara, USA

[2] Nanyang Technological University, Singapore, Singapore qtao002@e.ntu.edu.sg

[3] Monash University, Melbourne, Australia

Introduction

- Topic

  Universal CT Lesion Detection

- Problem

  a. The lesion size can be extremely small compared to natural objects
  b. The inter-class variance is small, i.e., lesions and non-lesions often have very similar appearances.

- Key Idea

  Contextual Attention + Spatial Attention

Model

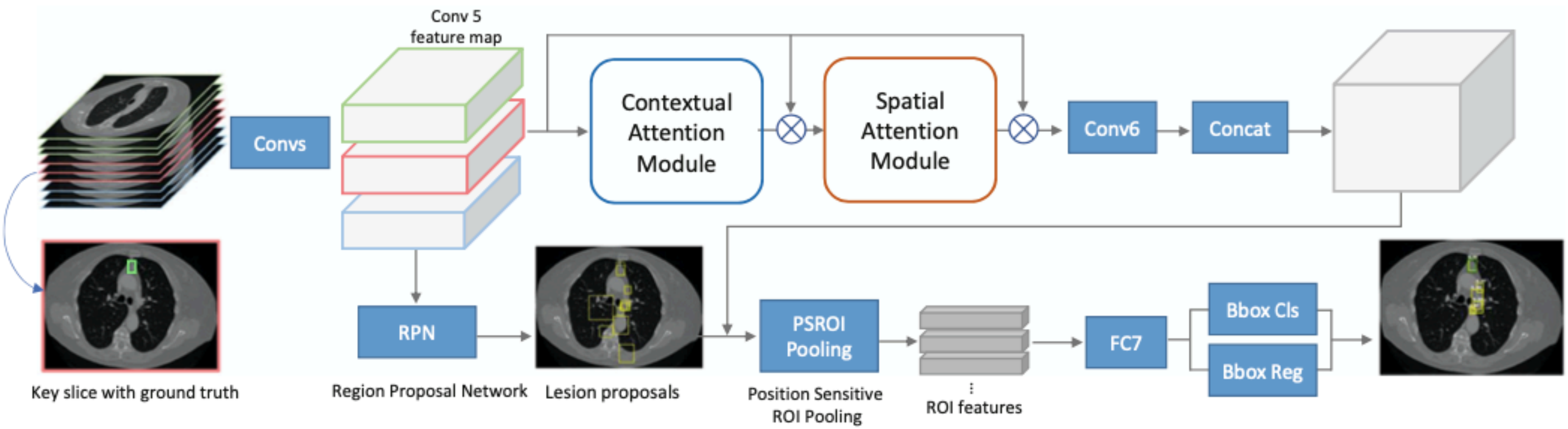
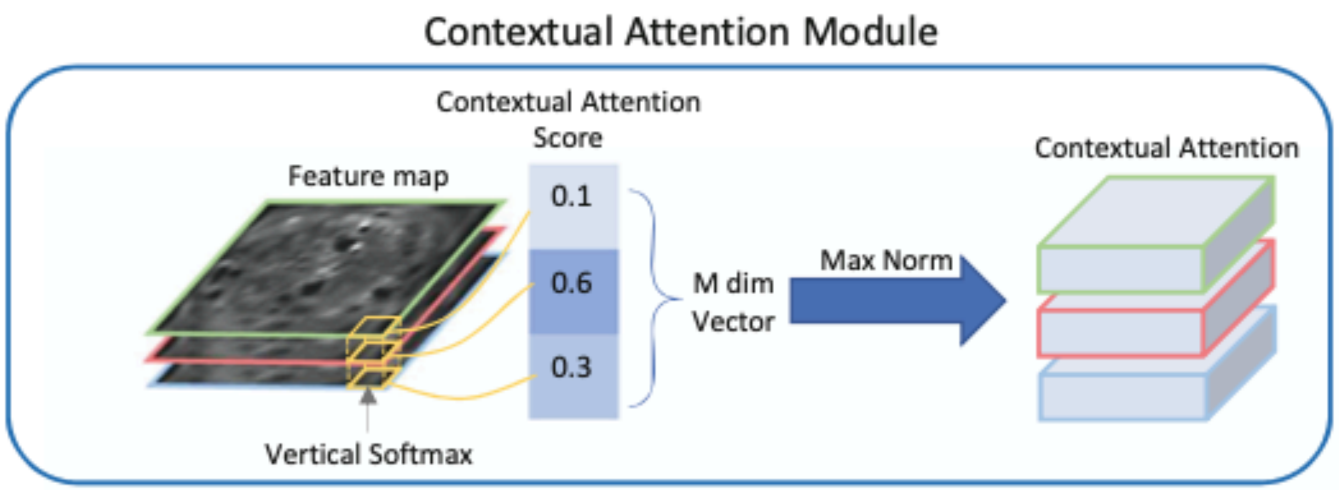


Fig. 1. Overview of network architecture: using 3DCE [7] as the base framework, we introduce two attention modules: (a) contextual attention module to re-weight the feature importance across all input slices; (b) spatial attention module to amplify the learning of the most prominent regions within each feature map.

Model

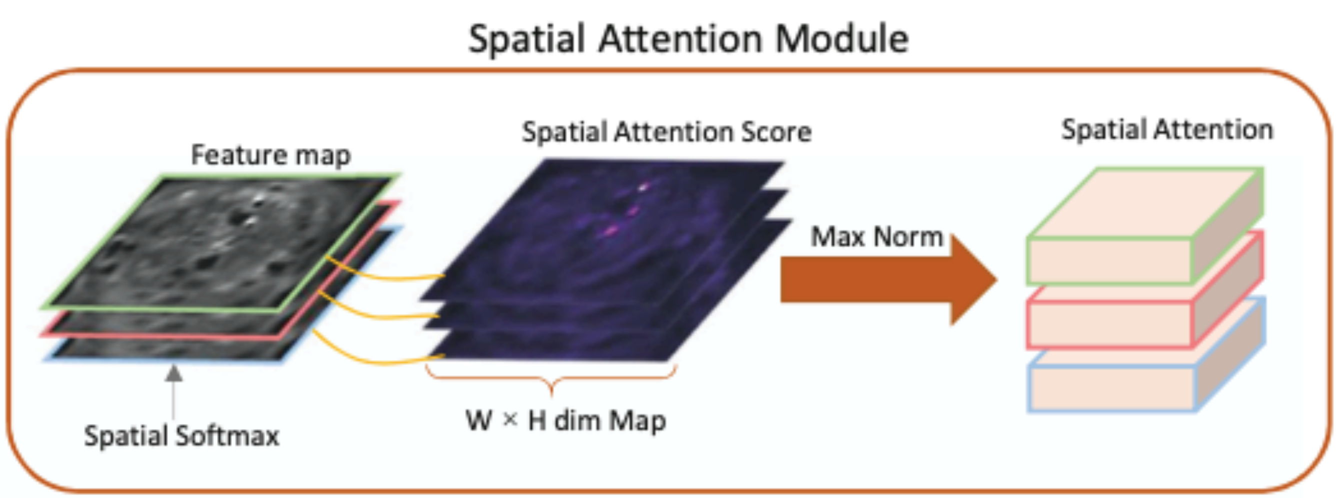


$$C_i = \phi_C(X_i)$$

$$C'^{w,h}_{i,d} = \frac{exp(C^{w,h}_{i,d})}{\sum_{i=1}^{M} exp(C^{w,h}_{i,d})}$$

$$C''^{w,h}_{i,d} = \frac{C'^{w,h}_{i,d}}{\max_i |C'^{w,h}_{i,d}|}$$

$$X'_i = C''_i \otimes X_i.$$



$$S_i = \phi_S(X'_i)$$

$$S'^{w,h}_{i,d} = \frac{exp(S^{w,h}_{i,d})}{\sum_{w=1}^{W}\sum_{h=1}^{H} exp(S^{w,h}_{i,d})}$$

$$S''^{w,h}_{i,d} = \frac{S'^{w,h}_{i,d}}{\max_{w,h} |S'^{w,h}_{i,d}|}$$

$$X''_i = S''_i \otimes X'_i.$$

Results

- Dataset

  DeepLesion, 22k CT slices for training

- Ablation Study

Table 3. Sensitivity (%) at various FPs per image on the test set of the official data split of DeepLesion using different attention components with 15 slices.

C_Att	S_Att	0.5	1	2	4	8	16
		63.0	73.1	80.2	85.2	87.8	89.7
✓		64.0	74.0	81.4	86.0	88.6	90.5
	✓	69.0	77.4	83.1	86.7	89.1	90.8
✓	✓	**70.8**	**78.6**	**83.9**	**87.5**	**89.9**	**91.4**

Results

- Comparison with SOTA

Table 1. Sensitivity (%) at different false false positives (FPs) per image on the test set of the official data split of DeepLesion. Note that the results of 3DCE are obtained from our experiments which are higher than the reported results in the original paper.

Sensitivity @	0.5	1	2	4	8	16
Improved R-FCN, 3 Slices [7]	56.5	67.7	76.9	82.8	87.0	89.8
3DCE, 9 Slices [7]	61.7	71.9	79.2	84.3	87.8	89.7
3DCE, 15 Slices [7]	63.0	73.1	80.2	85.2	87.8	89.7
3DCE, 21 Slices [7]	63.2	73.4	80.9	85.6	88.4	90.2
3DCE_CS_Att, 9 Slices (Ours)	67.8	76.3	82.9	86.6	89.3	90.7
3DCE_CS_Att, 15 Slices (Ours)	70.8	78.6	83.9	87.5	89.9	**91.4**
3DCE_CS_Att, 21 Slices (Ours)	**71.4**	**78.5**	**84.0**	**87.6**	**90.2**	**91.4**

Table 2. Sensitivity (%) at 4 FPs per image on the test set of DeepLesion using the baseline model (3DCE) and the proposed model (3DCE_CS_Att), both using 15 slices.

	Lesion type								Lesion diameter			Slice interval	
	LU	ME	LV	ST	PV	AB	KD	BN	<10	10–30	>30	<2.5	>2.5
3DCE	90.9	88.1	90.4	73.5	82.1	81.3	82.1	**75.0**	80.9	87.8	82.9	85.8	85.1
3DCE_CS_Att	**92.0**	**88.5**	**91.4**	**80.3**	**85.0**	**84.4**	**84.3**	**75.0**	**82.3**	**90.0**	**85.0**	**87.6**	**87.6**

Results

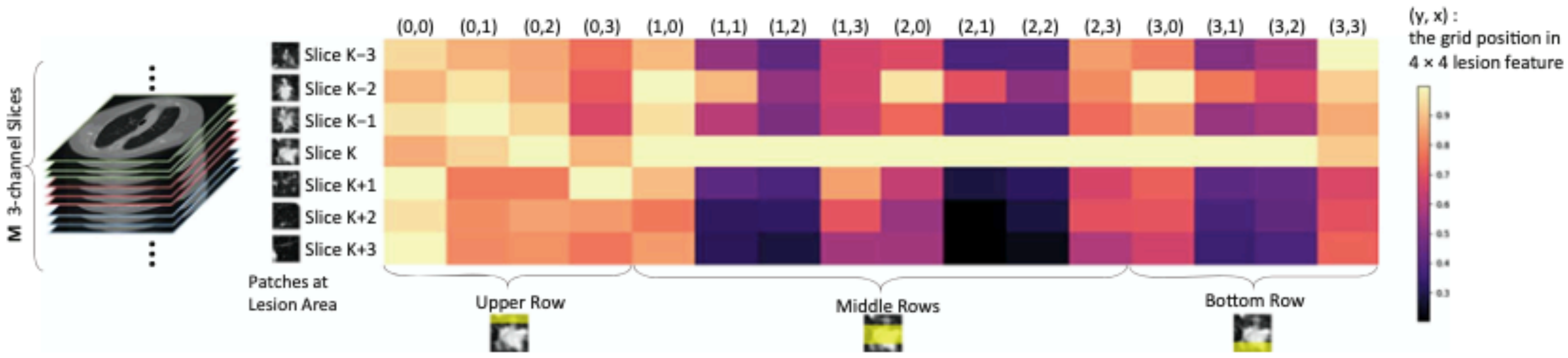


Fig. 2. Visualization of cross-slice contextual attention vectors based on our model with 7 three-channel images ($M = 7$). We visualize the slice patches corresponding to the lesion area in the key slice K as well as its previous and subsequent slices. The lesion region corresponds to 4×4 grids in Conv 5 feature. Therefore, we obtain 16 attention vectors for each feature grid from the contextual attention module. The vectors are visualized as a heatmap where each column (with a (y, x) coordinate in sub feature map of lesion patch) shows a normalized cross-slice attention vector.

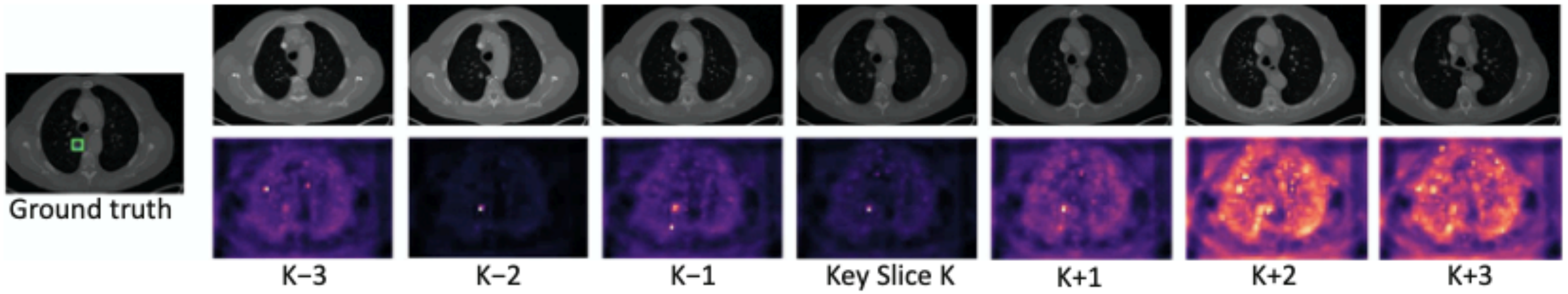


Fig. 3. Visualization of spatial attention map based on our model with 7 three-channel images. We can obtain 7 attention maps that are self-normalized to re-weight the feature importance within each feature map.

(MICCAI2019)

Probabilistic Radiomics: Ambiguous Diagnosis with Controllable Shape Analysis

Jiancheng Yang[1,2,3], Rongyao Fang[1], Bingbing Ni[1,2,3(B)], Yamin Li[1], Yi Xu[1], and Linguo Li[1]

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

[2] MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, Shanghai, China

[3] Shanghai Institute for Advanced Communication and Data Science, Shanghai, China

Introduction

- Topic

Pulmonary Nodules Diagnosis using CT images

- Problems

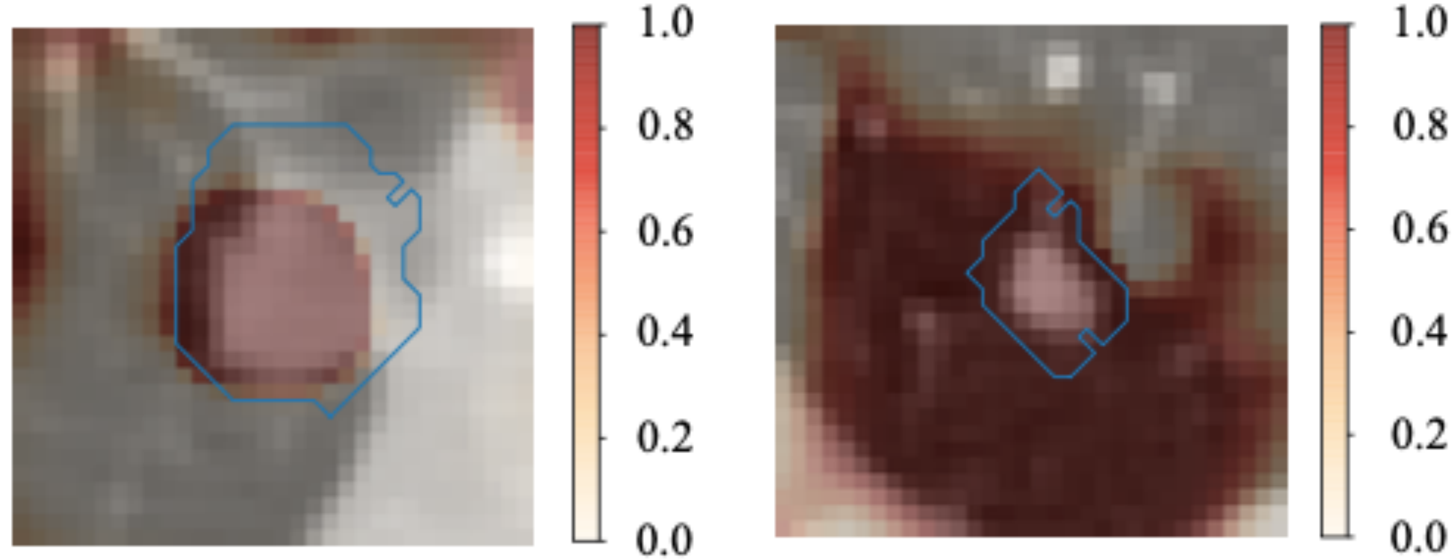
a. CNNs are “diagnosing” voxels (or pixels), rather than lesions; in other words, visual saliency from a trained CNN is not necessarily concentrated on the lesions.

b. Classification in clinical applications suffers from inherent ambiguities: radiologists may produce diverse diagnosis on challenging cases.

Introduction

- Example

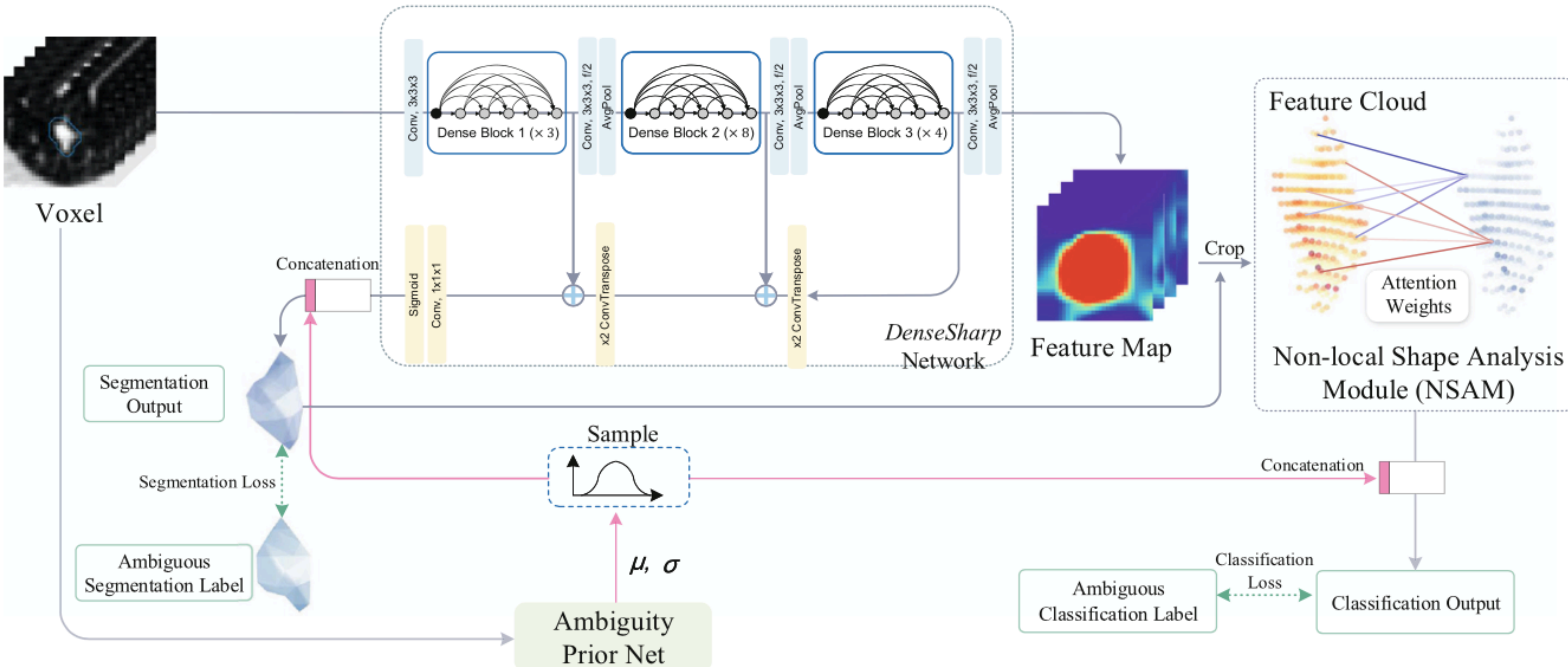
Malignant CAM



Failure (a) Correct
Label: Benign
Prediction: Benign

Failure (b) Incorrect
Label: Malignant
Prediction: Benign

Model



$$Attn(X) = softmax(XX^T/\sqrt{c}) \cdot \sigma(X)$$

$$NSAM(X) = concat\{Attn(X_i)|X_i = XW_i\}_{i=1,..,g}) + X$$

Results

- Dataset

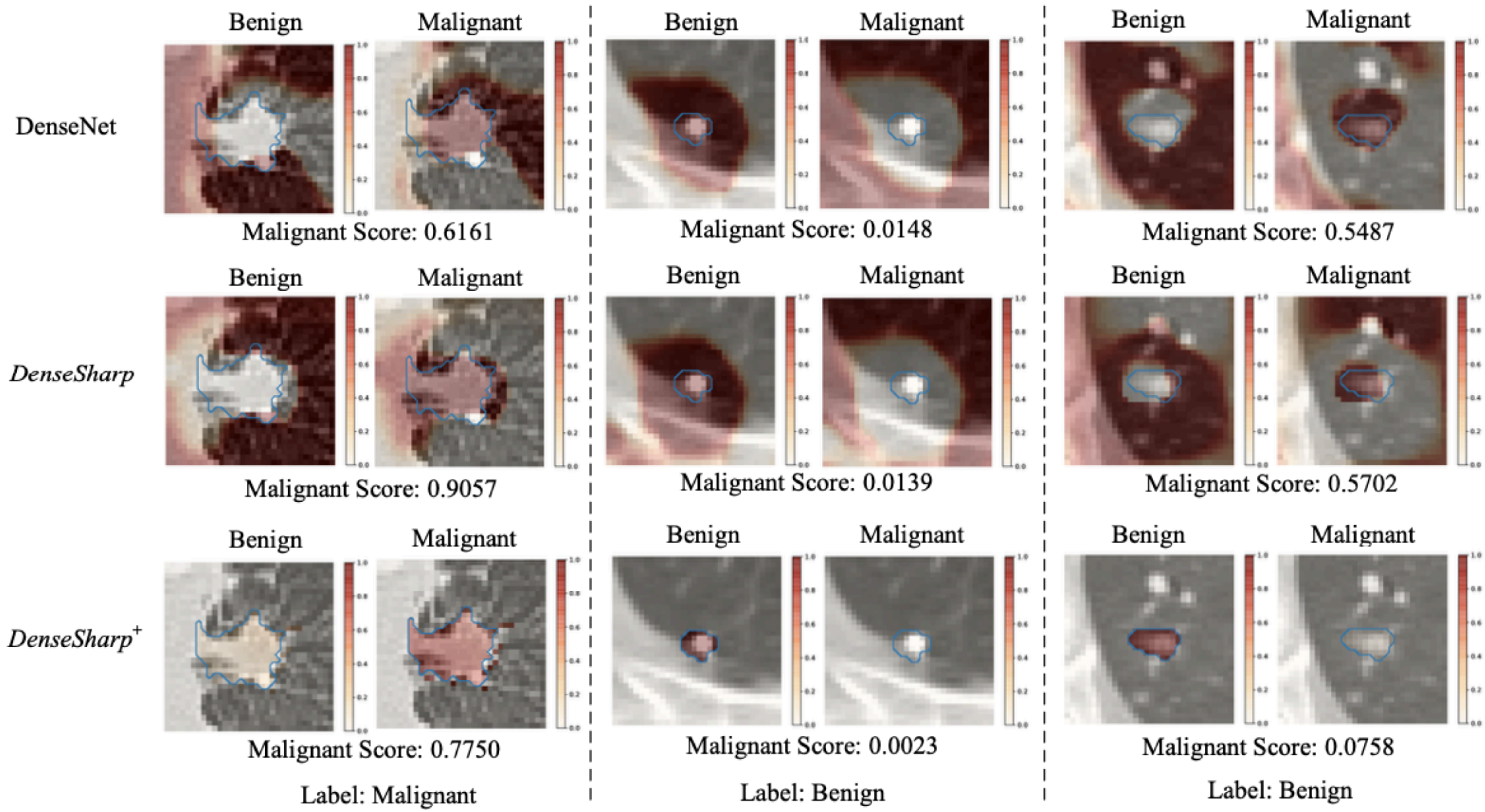
  LIDC-IDRI, 2635 nodules as HighAmbig, 656 benign and 527 malignant as LowAmbig

- Results

Table 1. AUC and accuracy of DenseNet, *DenseSharp*, *DenseSharp*$^+$, and prior studies. The performance of our models is evaluated on *LowAmbig* LIDC-IDRI [1] dataset (see Sect. 2.1) with 5-fold cross validation.

Method	AUC	Accuracy (%)
3D DPN [14]	–	88.28
3D DPN ensemble [14]	–	90.44
3D CNN w. MTL [5]	–	80.08
3D CNN w. sparse MTL [5]	–	91.26
3D DenseNet (our implementation)	0.9218	87.82
DenseSharp [11] (our implementation)	0.9393	89.26
DenseSharp$^+$ (LowAmbig)	0.9480	90.87
DenseSharp$^+$ (HighAmbig)	**0.9566**	**91.52**

Results



Benign
Malignant
DenseNet
Malignant Score: 0.6161
Malignant Score: 0.0148
Malignant Score: 0.5487
DenseSharp
Malignant Score: 0.9057
Malignant Score: 0.0139
Malignant Score: 0.5702
DenseSharp+
Malignant Score: 0.7750
Malignant Score: 0.0023
Malignant Score: 0.0758
Label: Malignant
Label: Benign
Label: Benign