Multitask Universal Lesion Analysis Network

for Joint Lesion Detection, Tagging, and Segmentation

MULAN

2020. 06. 06. SAT
KIM TAEWOO

CONTENTS

Abstract

CT Scan Image

Lesion

Detection

Tagging

Segmentation

CT (Computed Tomography)

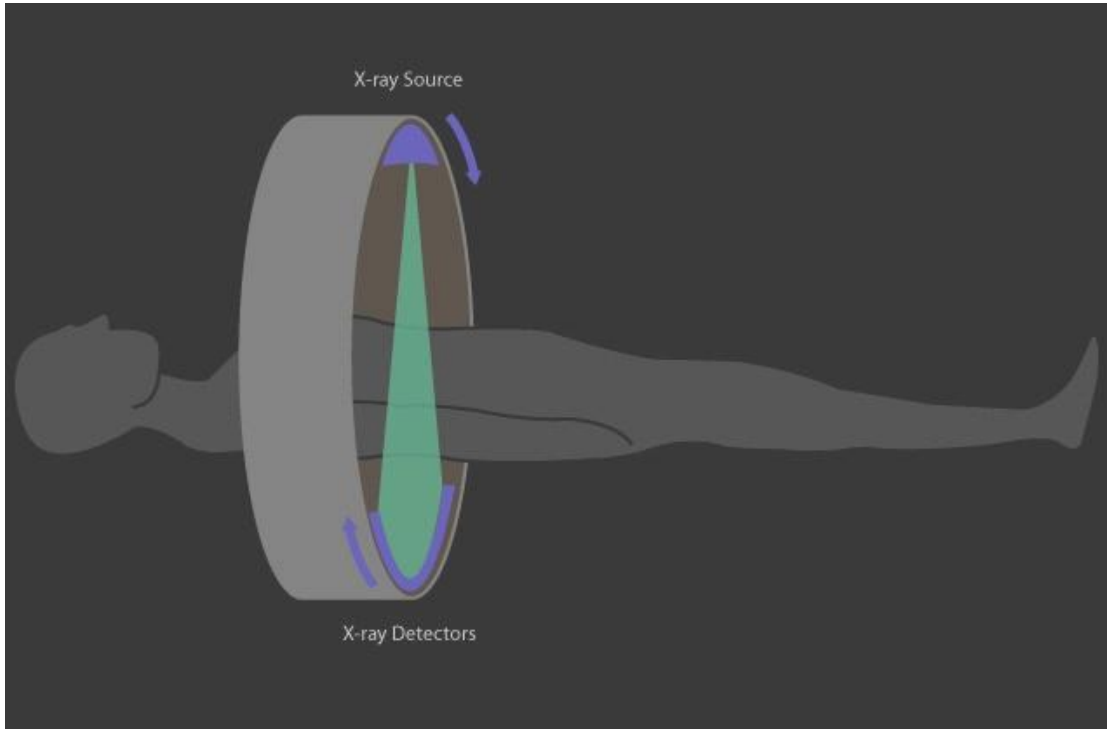
CT Scan Image

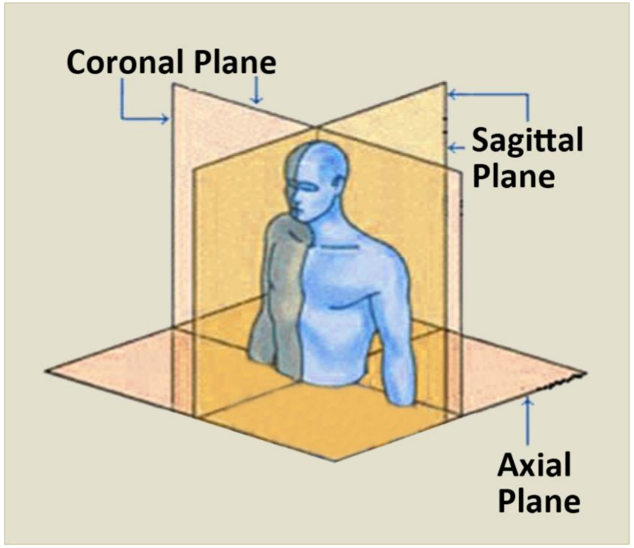
Lesion

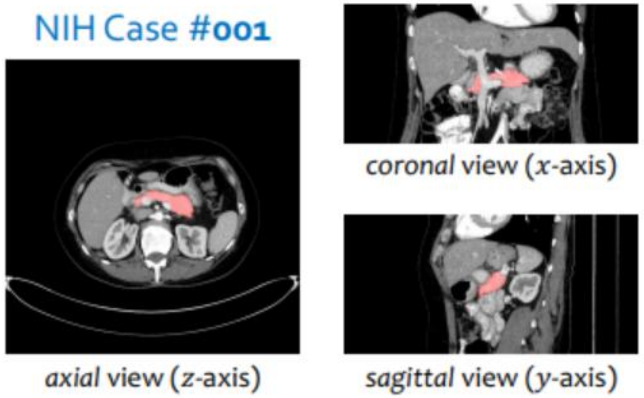
Detection

Tagging

Segmentation







CT란 X-선을 인체에 투과해 그 흡수차이를 컴퓨터로 재구성하여 인체의 단면영상을 얻거나 3차원적인 입체영상을 얻는 영상진단법

A lesion is any damage or abnormal change in the tissue of an organism

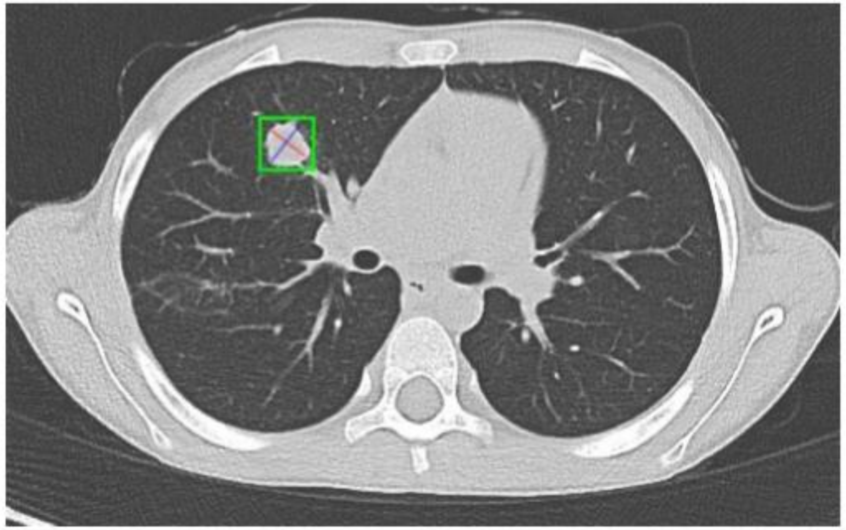
CT Scan Image

Lesion

Detection

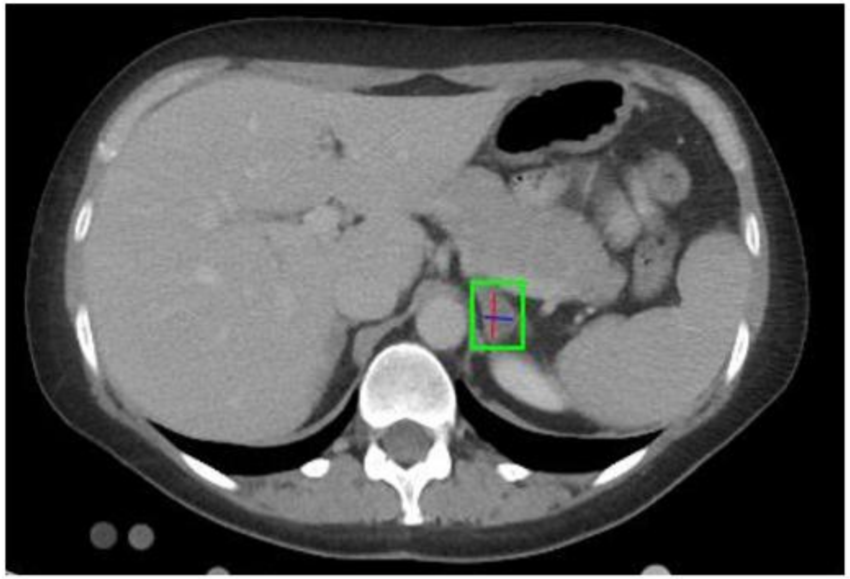
Tagging

Segmentation



Sentence: Within the right middle lobe there is a stable nodule that measures BOOKMARK.

Tags: Right mid lung, nodule



Sentence: Low density left adrenal nodule BOOKMARK, likely adenoma.

Tags: Hypoattenuation, left adrenal gland, nodule, adenoma

DeepLesion Dataset

CT Scan Image

Lesion

Detection

Tagging

Segmentation

DeepLesion Dataset

- National Institutes of Health(NIH) Clinical Center, USA

- Contains 32,735 lesions on 32,120 **axial** computed tomography (CT) slices from 10,594 CT scans (studies) of 4,427 unique patients.

- lesions in each image with accompanying bounding boxes and size measurements, adding up to 32,735 lesions altogether.

- Semantic labels (tags) for the lesions mined from radiological reports. The labels describe the lesions' body part, type, and attributes. (2019. 4. added, the label with tags = Train : 20,266 / Valid : 1,976 / Test : 1879)

- Lesion size : 0.2mm ~ about 340 mm

DeepLesion Dataset

CT Scan Image

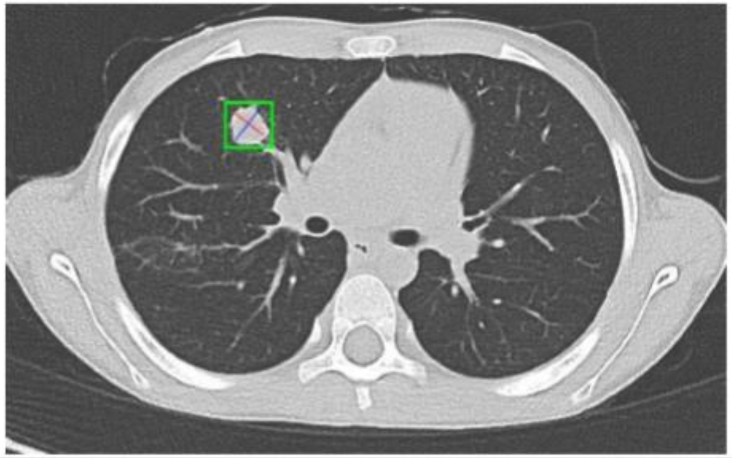
Lesion

Detection

Tagging

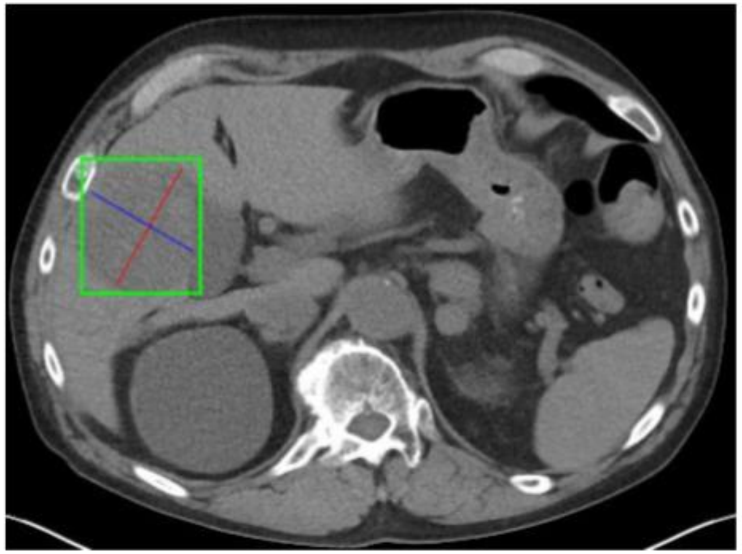
Segmentation

DeepLesion Dataset



Sentence: Within the right middle lobe there is a stable nodule that measures BOOKMARK.

Tags: Right mid lung, nodule



Sentence: A large right hepatic mass, incompletely characterized BOOKMARK.

Tags: Large, liver, liver mass, mass

Abstract

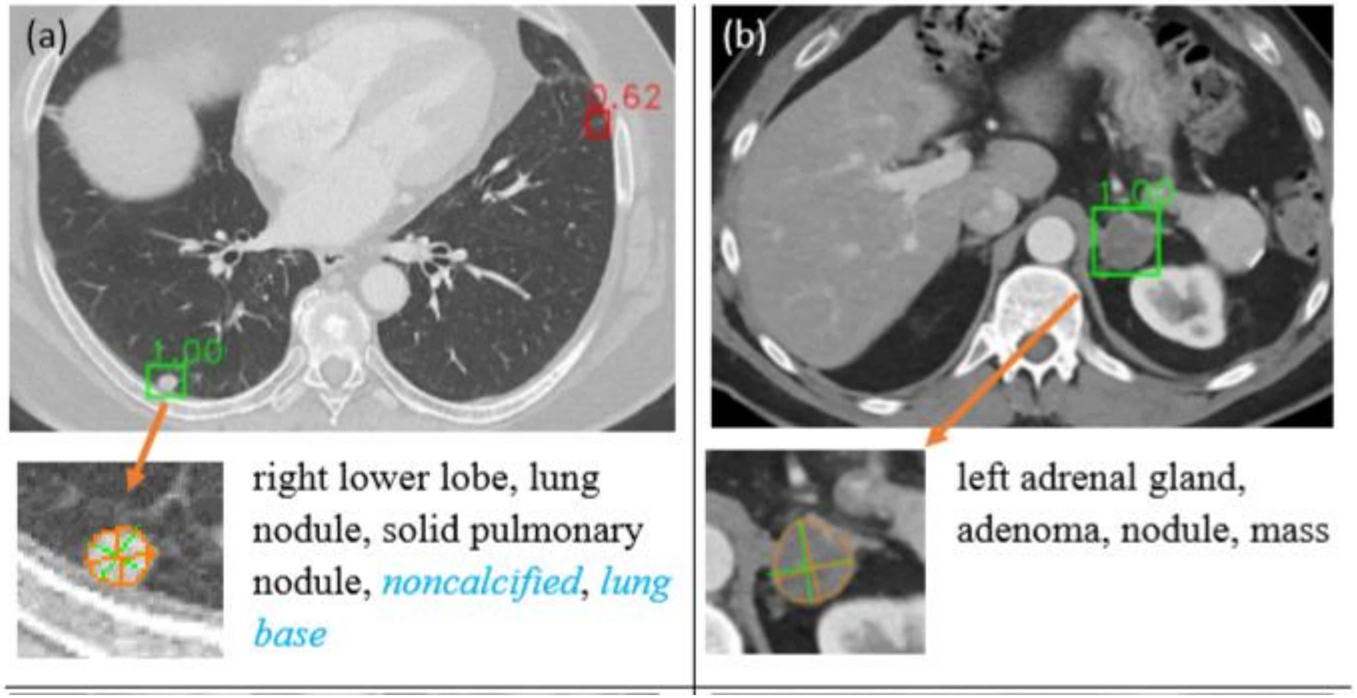
CT Scan Image

Lesion

Detection

Tagging

Segmentation



Whole Framework

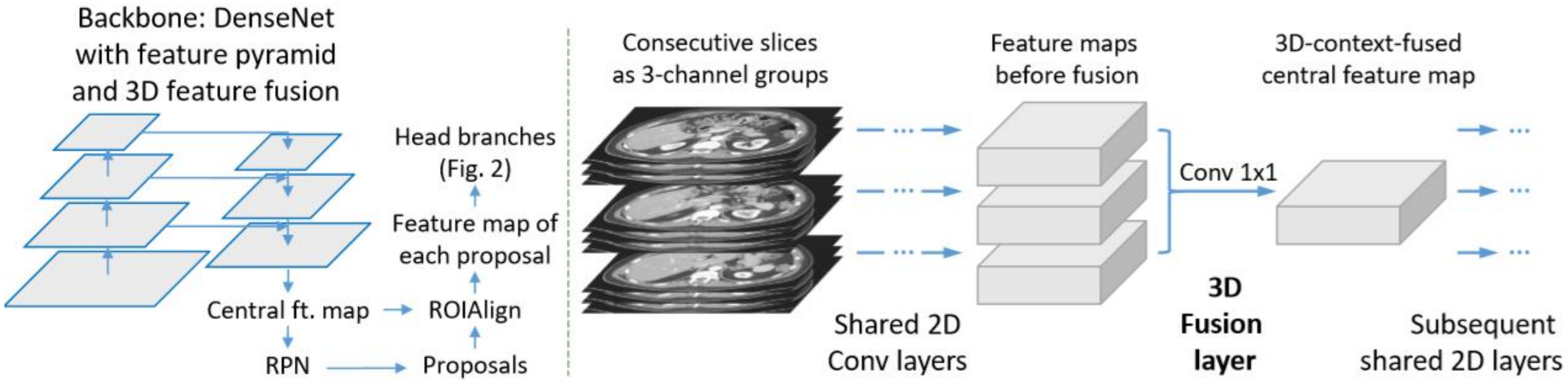
[1] Backbone

[2] 3D feature fusion

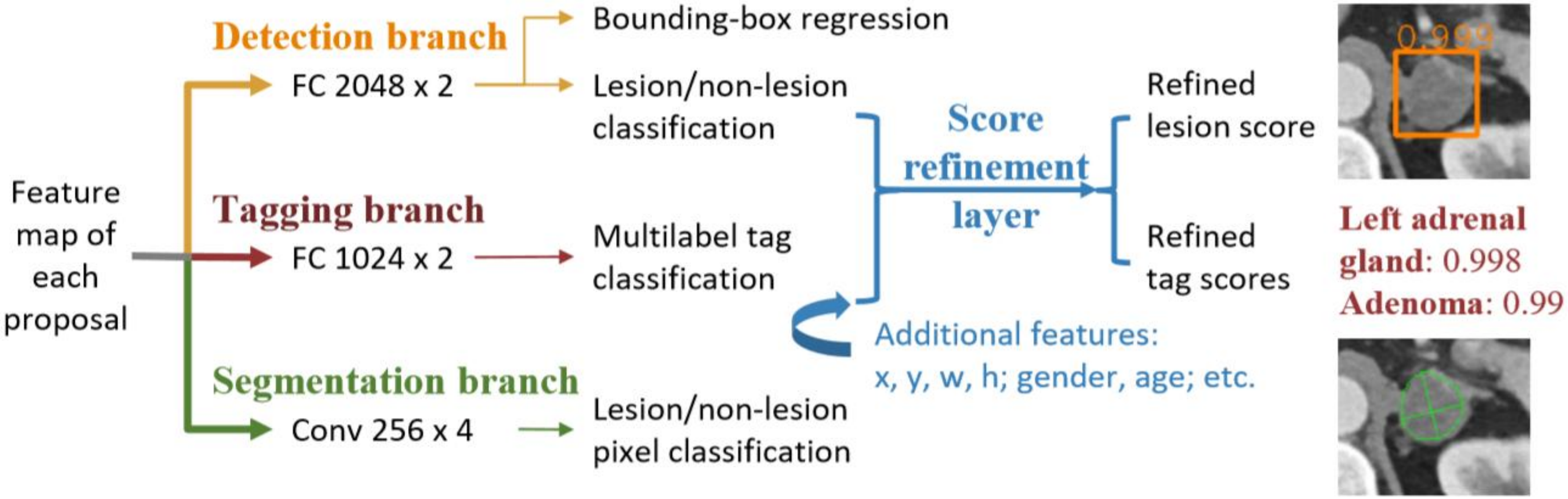
[3] Mask R-CNN

[4] 3 branches

[5] Refinement



(a) Flowchart **(b) 3D feature fusion**



Method

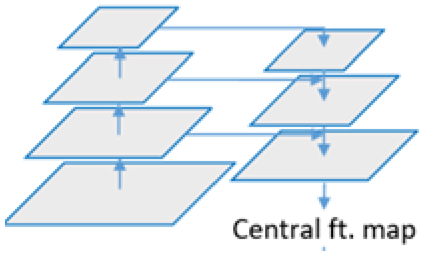
[1] Backbone

[2] 3D feature fusion

[3] Mask R-CNN

[4] 3 branches

[5] Refinement



- DenseNet-121 with the last dense block and transition layer removed
- ImageNet pretrained
- Feature pyramid strategy (increase the size of final feature map, which will benefit the detection and segmentation of small lesions)
- Attach the head branches only to the finest level

Method

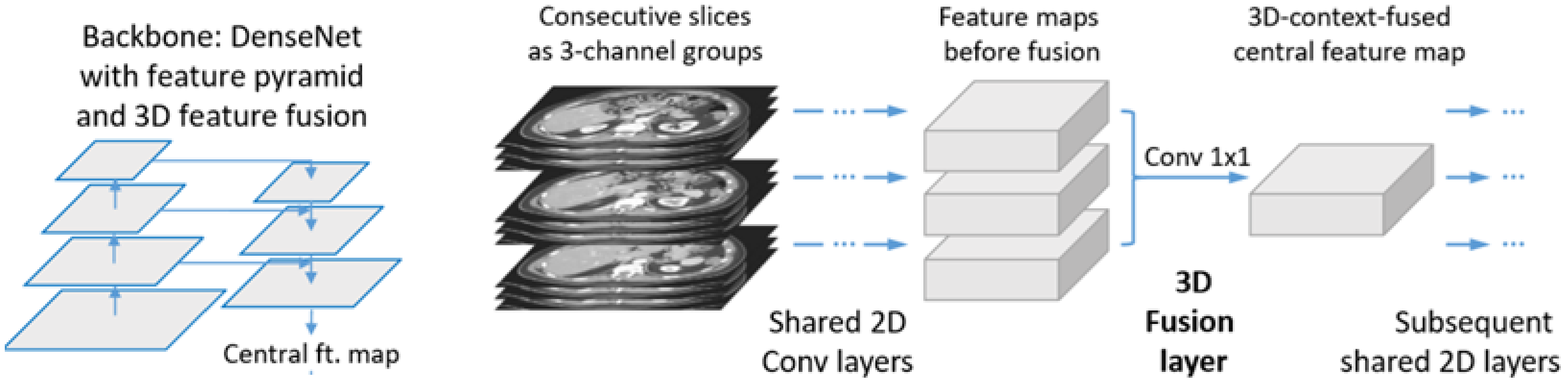
[1] Backbone

[2] 3D feature fusion

[3] Mask R-CNN

[4] 3 branches

[5] Refinement



(b) 3D feature fusion

- **3D context information is very important!!**
- First, group consecutive axial slices in a CT volume into 3-channel
- With a 1x1 Conv layer (3D fusion layer)
- Fuse features of multiple slices(9) in eariler Conv layers. (after dense block 2 and the last layer of the feature pyramid)
- All FMs are fed to subsequent Conv layer, with Central FM is replaced by the FM(feature map)s are fused with the 3D fusion layer, upper FM and lower FM.

Method

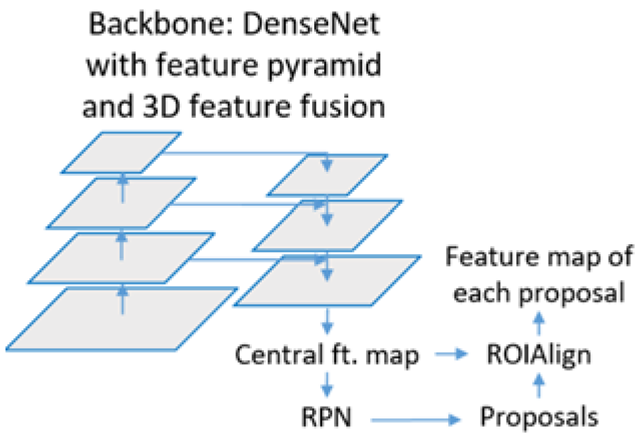
[1] Backbone

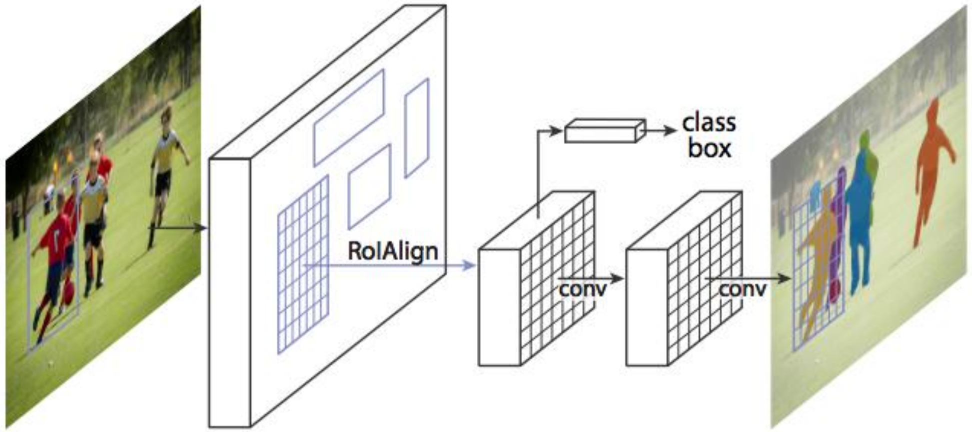
[2] 3D feature fusion

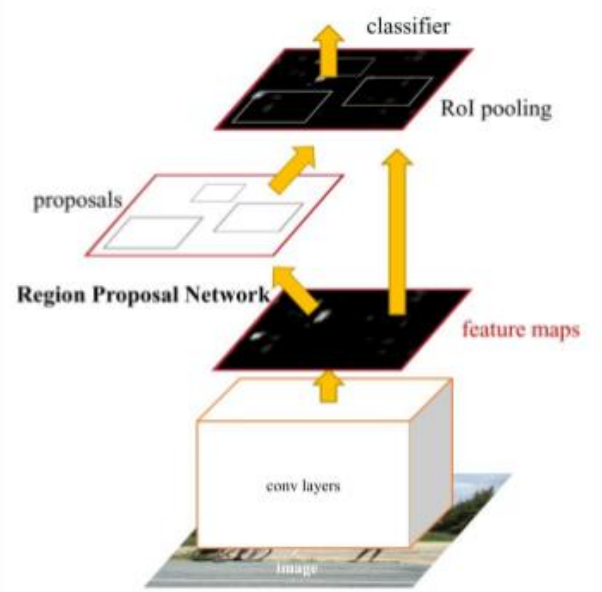
[3] Mask R-CNN

[4] 3 branches

[5] Refinement







1. Faster R-CNN : bbox + class → bbox + class + seg
2. ROI Pooling 대신 ROI Align 을 사용함
3. Mask prediction 과 class prediction 을 decouple 함 (클래스 상관없이 masking)

출처: https://mylifemystudy.tistory.com/82 [ENCAPSULATION]

https://www.youtube.com/watch?v=RtSZALC9DlU

Method

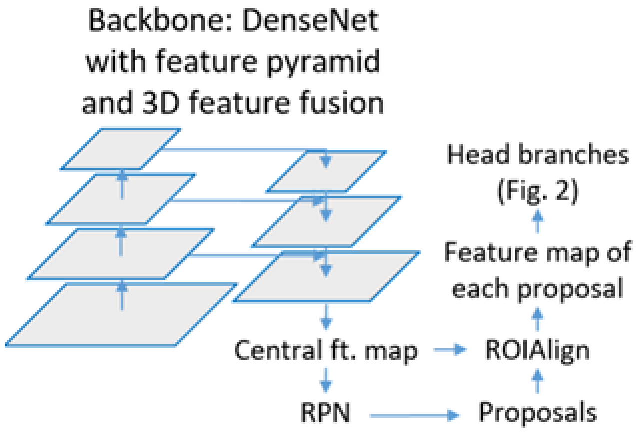
[1] Backbone

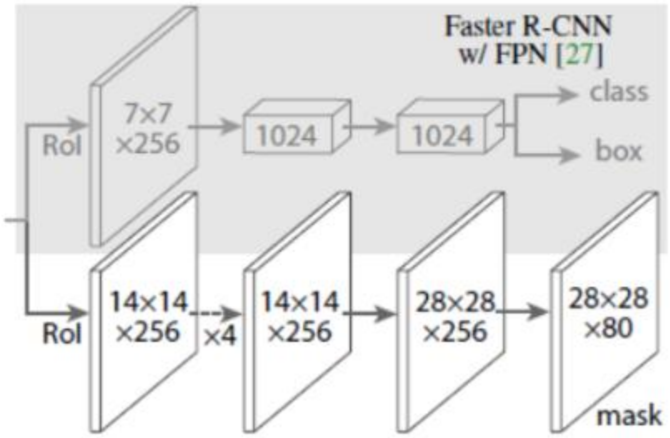
[2] 3D feature fusion

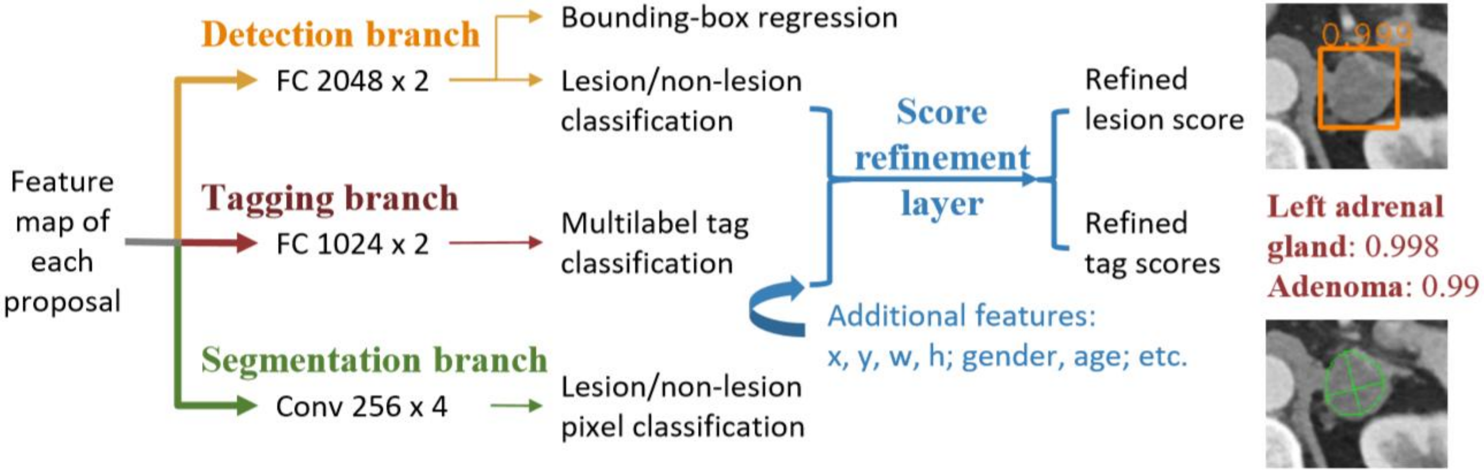
[3] Mask R-CNN

[4] 3 branches

[5] Refinement







Method

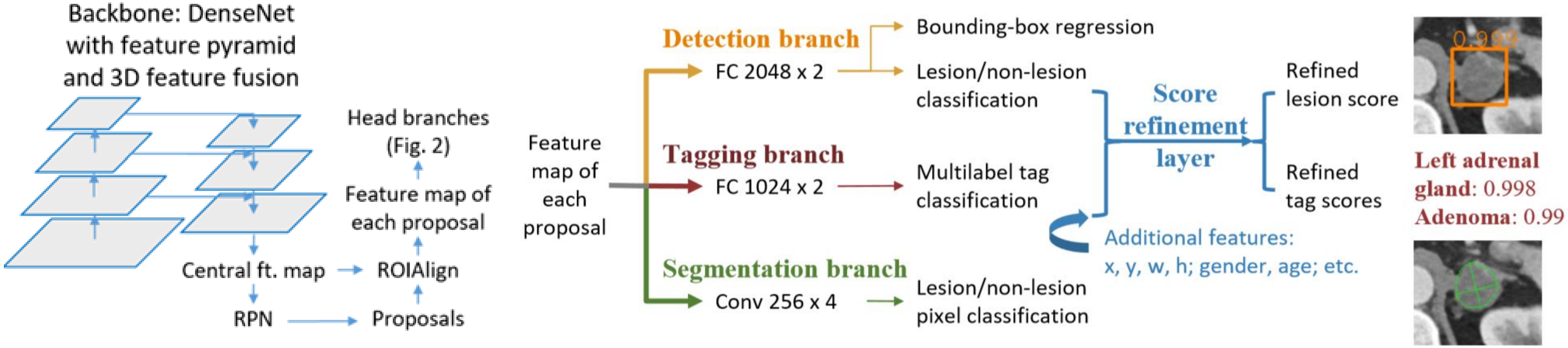
[1] Backbone

[2] 3D feature fusion

[3] Mask R-CNN

[4] 3 branches

[5] Refinement



$$L = L_{\mathrm{det,seg}} + L_{\mathrm{tag,WCE}} + L_{\mathrm{tag,RHEM}} + L_{\mathrm{cls,SRL}} + L_{\mathrm{tag,WCE,SRL}}$$

$$L_{\mathrm{det,seg}} = L_{\mathrm{RPN,cls}} + L_{\mathrm{RPN,box}} + L_{\mathrm{det,cls}} + 10L_{\mathrm{det,box}} + L_{\mathrm{seg,dice}}$$

$$L_{\mathrm{tag,WCE}} = \sum_{i=1}^{B}\sum_{c=1}^{C}\left(\beta_c^{\mathrm{p}} y_{i,c}\log\sigma_{i,c} + \beta_c^{\mathrm{n}}(1-y_{i,c})\log(1-\sigma_{i,c})\right)$$

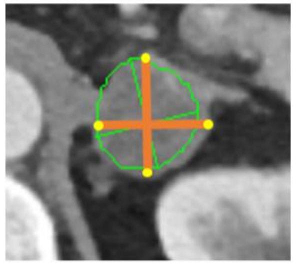
*relational hard example mining (RHEM)

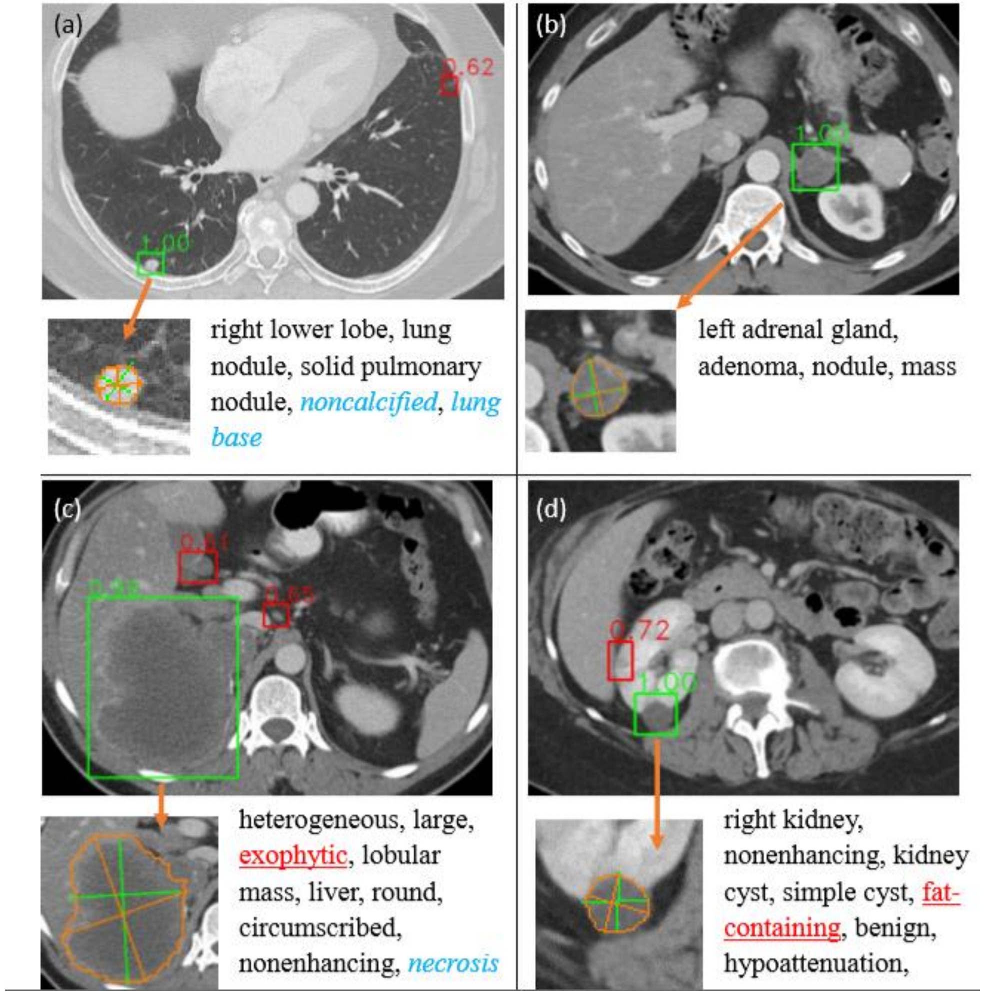
Result

	Detection (%)	Tagging (%)		Segmentation (mm)	
	Avg. sensitivity	AUC	F1	Distance	Diam. err.
ULDor [13]	69.22	–	–	–	–
3DCE [15]	75.55	–	–	–	–
LesaNet [16] (rerun)	–	95.12	43.17	–	–
Auto RECIST [11]	–	–	–	–	**1.7088**
MULAN	**86.12**	**96.01**	**45.53**	1.4138	1.9660
(a) w/o feature pyramid	79.73	95.51	43.44	1.6634	2.3780
(b) w/o 3D fusion	79.57	95.88	44.28	1.4120	1.9756
(c) w/o detection branch	–	95.16	40.03	**1.2445**	1.7837
(d) w/o tagging branch	84.79	–	–	1.4230	1.9589
(e) w/o mask branch	85.21	95.87	43.76	–	–
(f) w/o score refine. layer	84.24	95.65	44.59	1.4260	1.9687

Table 3. Sensitivity (%) at various FPs per image on the test set of DeepLesion (the 171 tags were used for training MULAN).

FPs per image	0.5	1	2	4	8	16	Avg. of [0.5,1,2,4]
3DCE [15]	62.48	73.37	80.70	85.65	89.09	91.06	75.55
ULDor [13]	52.86	64.80	74.84	84.38	87.17	91.80	69.22
MULAN	**76.12**	**83.69**	**88.76**	**92.30**	**94.71**	**95.64**	**85.22**





Conclusion

2 stage Object detection algorithm **– MASK-RCNN**

An effective **3D feature fusion** strategy

Jointly learning to automate radiologists' process

!!!

THANK YOU