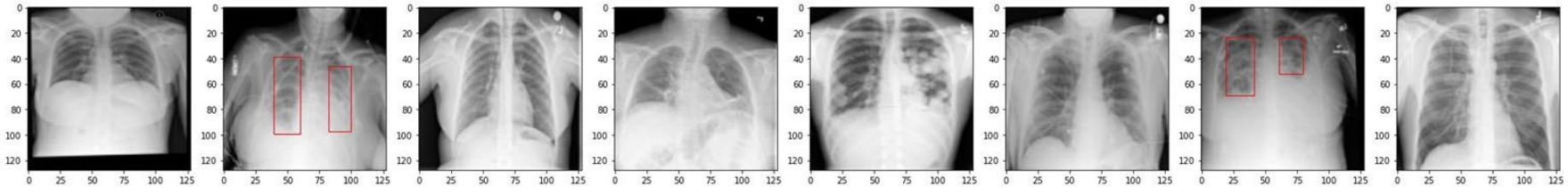
Pneumonia Detection

Steven Guo, Yucong Zhang
{hg138, yz646}@duke.edu

Background & Related Work



1. One-stage Detection
    a. YOLO [1]
2. Two-stage Detection
    a. Region-based CNN (RCNN) [2]

[1] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
[2] Girshick, Ross. "Fast r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2015.

Method - RetinaNet [3]

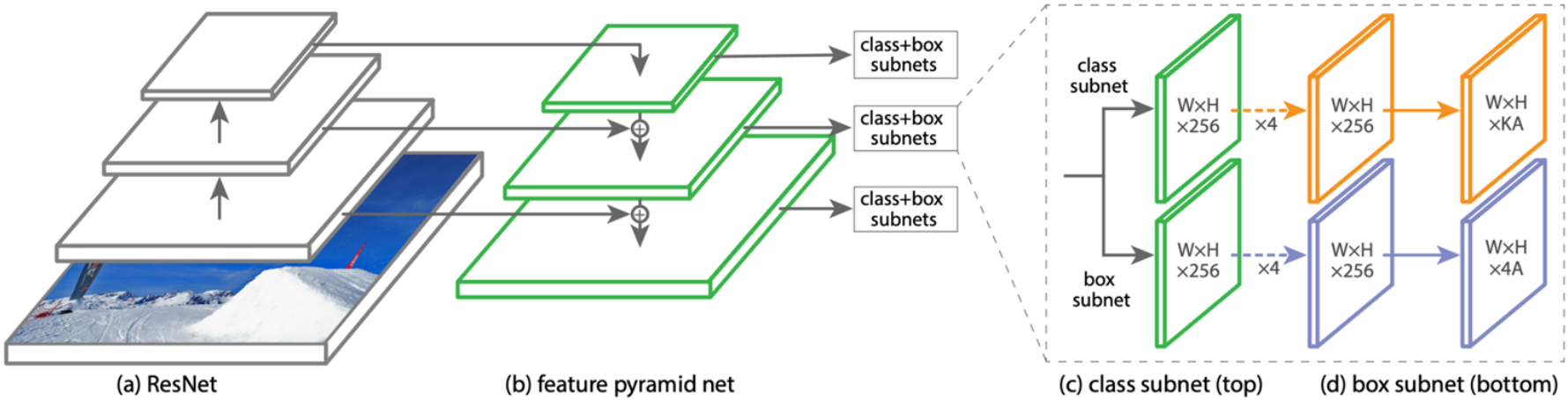


Figure 3. The one-stage **RetinaNet** network architecture uses a Feature Pyramid Network (FPN) [20] backbone on top of a feedforward ResNet architecture [16] (a) to generate a rich, multi-scale convolutional feature pyramid (b). To this backbone RetinaNet attaches two subnetworks, one for classifying anchor boxes (c) and one for regressing from anchor boxes to ground-truth object boxes (d). The network design is intentionally simple, which enables this work to focus on a novel focal loss function that eliminates the accuracy gap between our one-stage detector and state-of-the-art two-stage detectors like Faster R-CNN with FPN [20] while running at faster speeds.

[3] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.

Method - Backbone

- ResNet34 & ResNet50
- Pre-trained on two datasets
  - ImageNet dataset
  - RSNA kaggle dataset

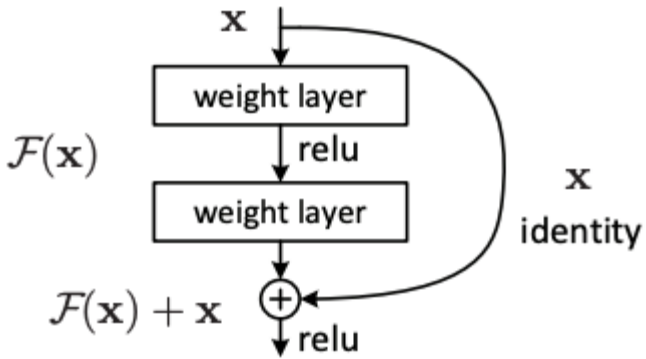


Fig 1 Residual Structure [2]

$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$
$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$
$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$

Fig 2 ResNet34 and ResNet50 Architecture [2]

[4] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Method - Feature Pyramid Network [5]

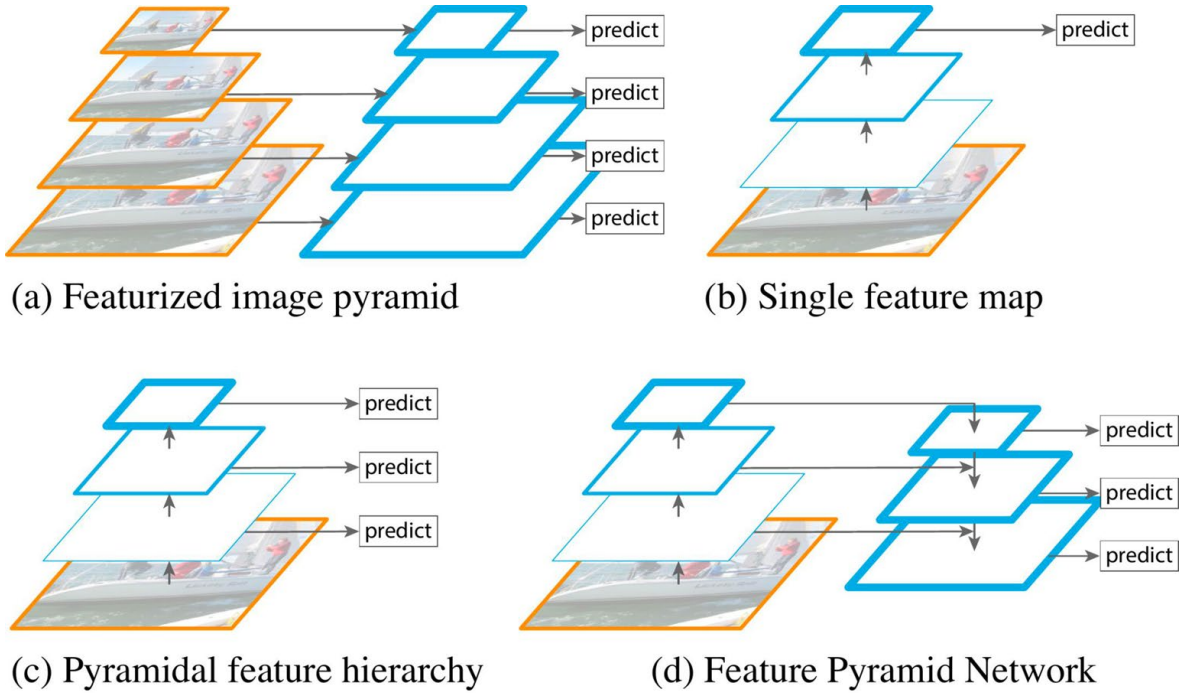


Fig 3 FPN Structure [1]

[5] Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Method - Classification & Regression Subnet

1. Classification Subnet
   a. CNN with 5 convolutional layers
   b. output channel = num_anchors * num_classes

1. Regression Subnet
   a. CNN with 5 convolutional layers
   b. output channel = num_anchors * 4

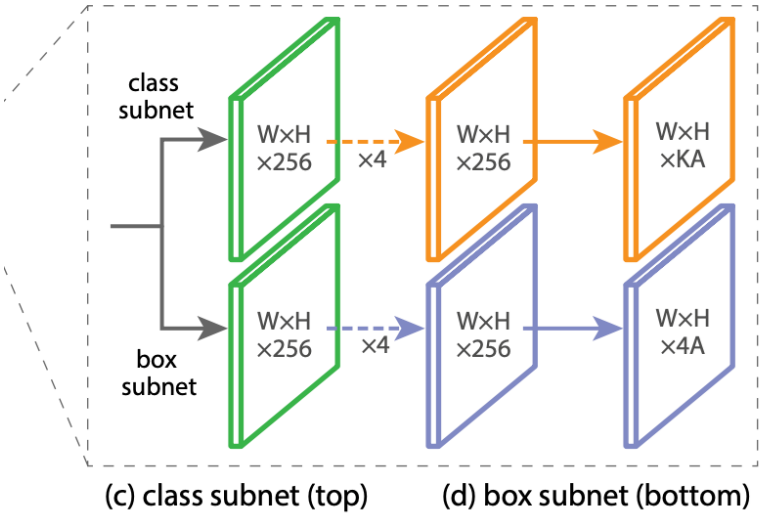


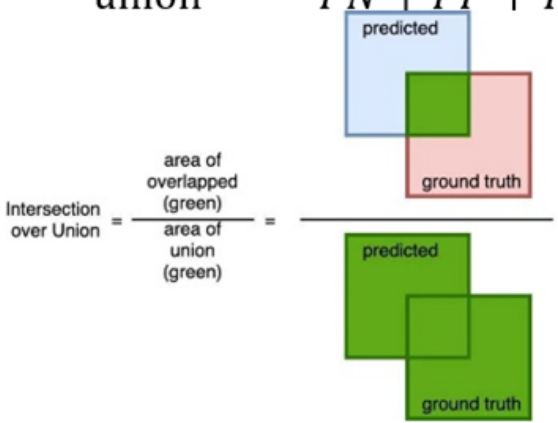
Fig 4 Subnets Structure [3]

[3] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *Proceedings of the IEEE international conference on computer vision*. 2017.

Method - Loss functions

1. Binary Cross Entropy
2. Focal Loss
3. Compound Loss

$$\mathrm{IoU} = \frac{\text{intersection}}{\text{union}} = \frac{TP}{FN + FP + TN}$$



$$\mathcal{L}_{\mathrm{BCE}(p,y)} = \mathrm{CE}\,(p_t) = -\log\,(p_t)$$

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

$$\mathcal{L}_{FLIoU} = \mathcal{L}_{FL} - IoU$$

[4]Ma, J. (2020). Segmentation loss odyssey. *arXiv preprint arXiv:2005.13449*

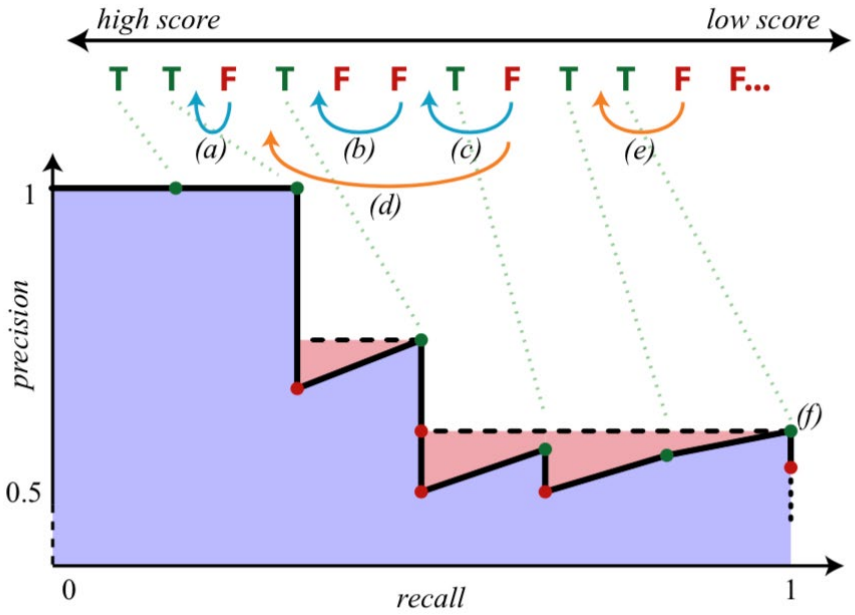
Experiment Settings

1. train test split
2. image downsizing 1024 => 128
3. hyperparameters
    a. epoch = 15
    b. lr = 0.0001
    c. confidence score > 0.05

Metric

1. Average Precision (AP)
    a. AUC of the precision and recall curve
    b. need smoothing

1. mean AP (mAP)

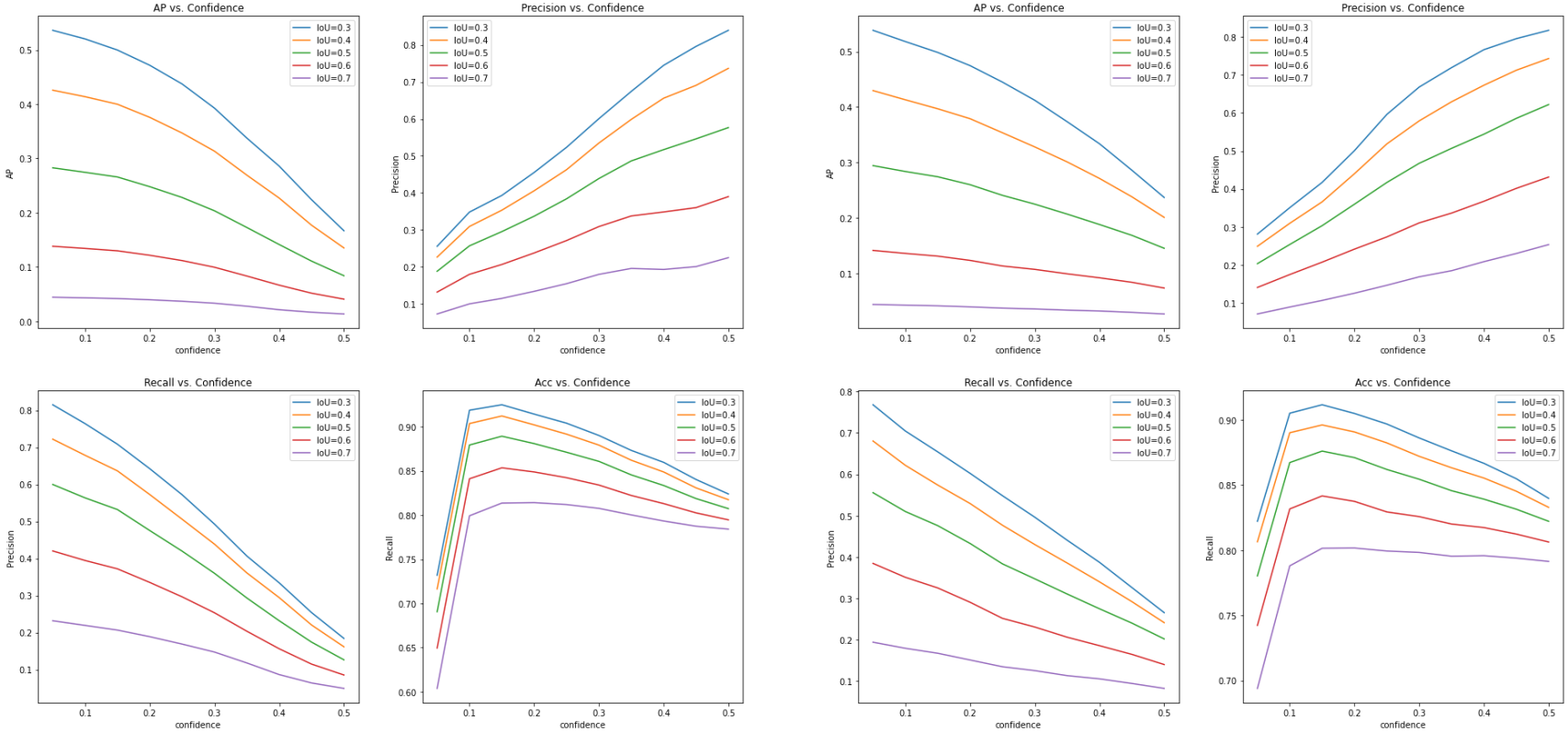


Results

Table 1: The average precision (AP) of all models regarding different IoU values. The Baseline model uses the ResNet34 model as the backbone without pre-trained on any dataset, and it is trained by binary cross entropy loss.

Model	Pretrained (CT)	Pretrained (Imagenet)	Focal Loss	Compound Loss	AP@0.3	AP@0.4	AP@0.5	AP@0.6	AP@0.7	mAP
Baseline	-	-	-	-	0.3300	0.2478	0.1441	0.0699	0.0201	0.1624
ResNet34	-	-	√	-	0.3481	0.2728	0.1801	0.0803	0.0207	0.1804
	-	√	√	-	0.3651	0.2851	0.1897	0.0992	0.0266	0.1931
	-	√	-	√	0.4108	0.3146	0.2011	0.0945	0.0284	0.2099
	√	-	√	-	0.5652	0.4406	0.2981	**0.1508**	**0.0439**	0.2973
	√	-	-	√	**0.5727**	**0.4662**	**0.3144**	0.1383	0.0377	**0.3058**
ResNet50	-	-	√	-	0.2840	0.2274	0.1535	0.0811	0.0246	0.1542
	-	√	√	-	0.4079	0.3207	0.2177	0.1011	0.0289	0.2153
	-	√	-	√	0.4142	0.3374	0.2280	0.1258	0.0424	0.2296
	√	-	√	-	0.5274	0.4260	**0.2883**	**0.1437**	**0.0415**	0.2854
	√	-	-	√	**0.5612**	**0.4452**	0.2861	0.1416	0.0399	**0.2948**

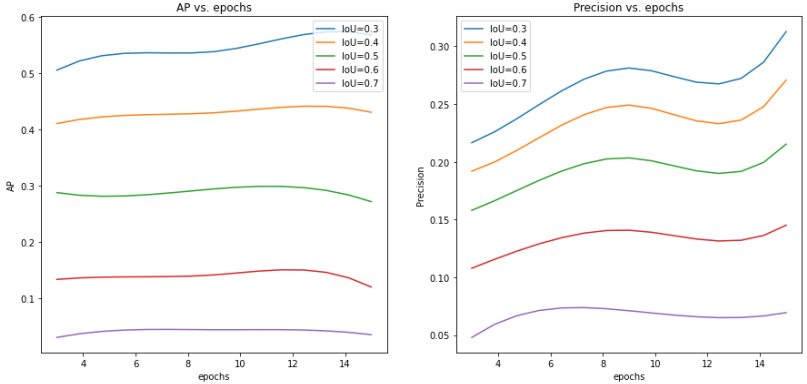
Discussions - Regarding the Confidence Level

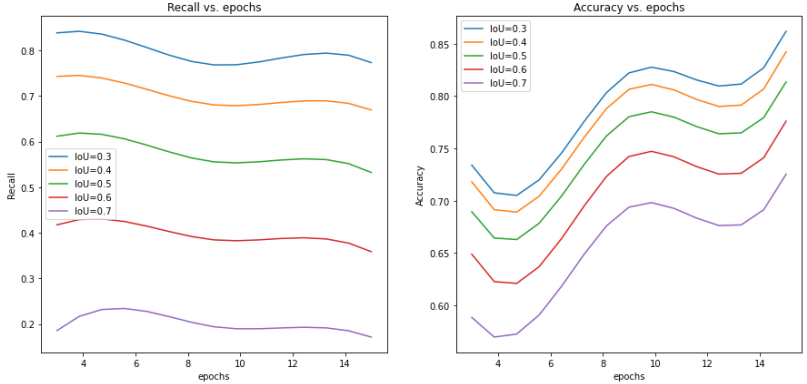


epoch=6

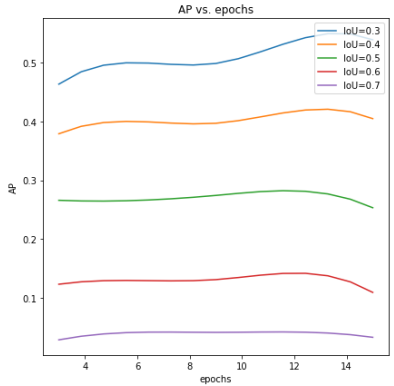
epoch=9

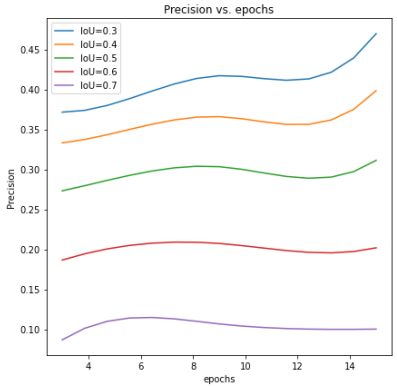
Discussions - Regarding the Training Epochs

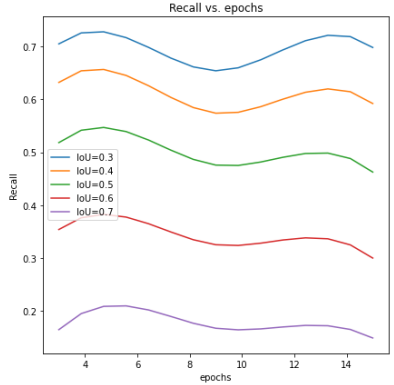


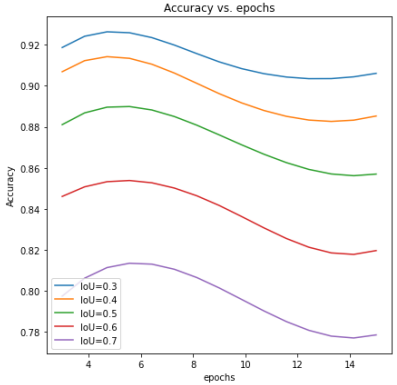


confidence > 0.05









confidence > 0.15

Q&A

Thank you!