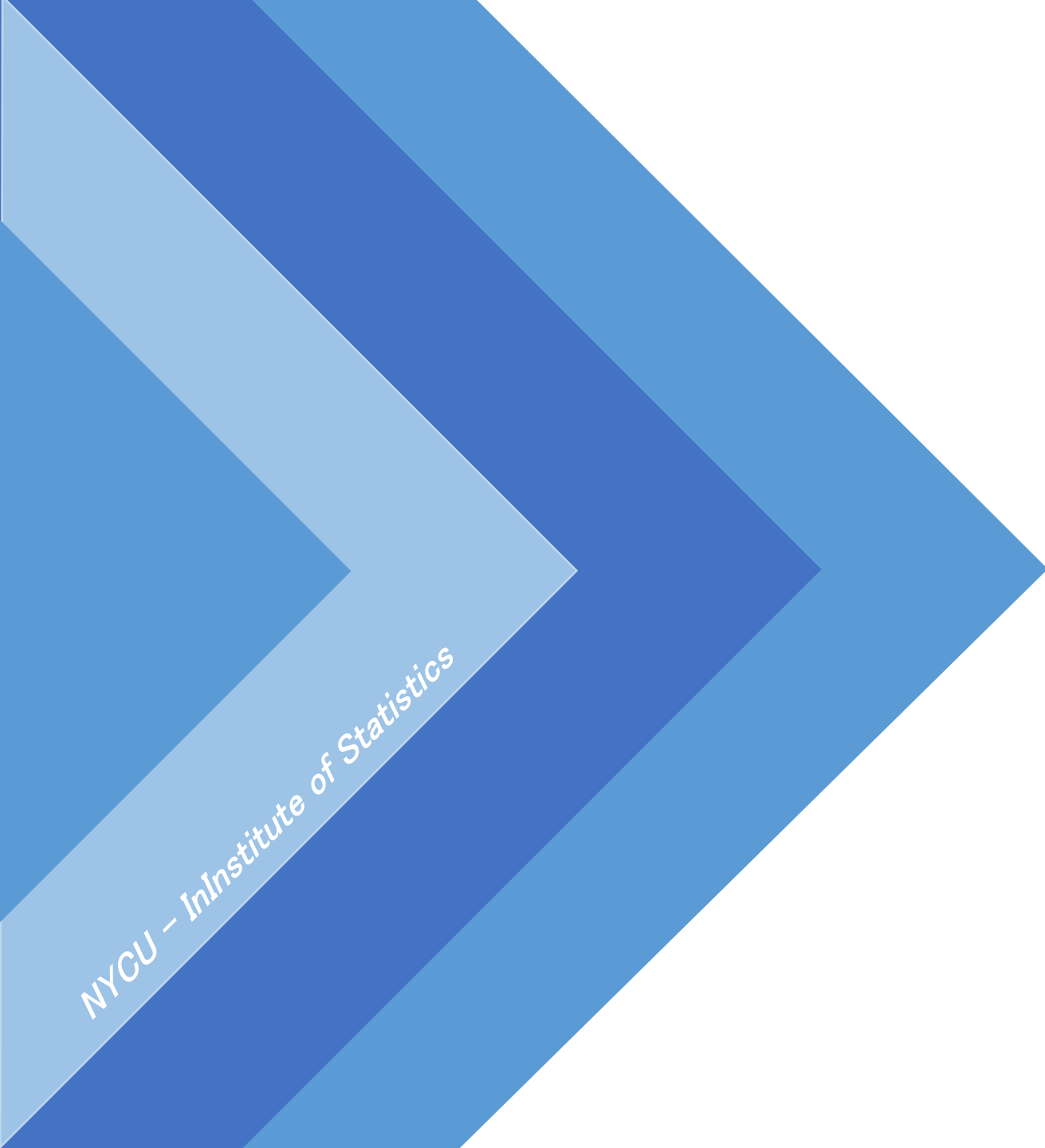


Deep learning-based object detection approaches for common thorax disease detection in chest X-rays images

Yi-Chieh Hung and Guan-Hua Huang

Institute of Statistics, National Yang Ming Chiao Tung University, Taiwan

2023/07/12



NYCU – Institute of Statistics

Introduction

Motivation

- Clinical medical images play an essential role in diagnosing and prognosis of clinical diseases by reflecting the shape and function of human tissues and organs.
- High accuracy of image classification and object detection with the maturity of computer vision

Objective

- When we apply deep learning technology to chest abnormality detection, although the problem of **low contrast** in medical images needs to be improved, we must also consider the viewpoint of clinical medical imaging, and play a fundamental role in diagnosing and prognoses clinical diseases. The rational use of artificial intelligence technology will effectively improve the efficiency and accuracy of clinical diagnosis.
- Discuss the impact of **data imbalance** on the object detection model.

Dataset Description

Categories	Categories English	Sample size
正常	normal	1314
主動脈弓粥樣硬化斑塊	aortic arch atherosclerotic plaque	28
主動脈弓鈣化	aortic arch calcification	16
主動脈粥樣硬化	aortic atherosclerosis	25
主動脈壁鈣化	aortic wall calcification	22
主動脈彎曲	Aortic curvature	67
胸椎動脈彎曲	Thoracic vertebral artery curvature	29
肺小結節	small pulmonary nodules	5
肺野結節陰影	shadows of pulmonary nodules	8
肺結核狀病變	tuberculosis	5
肺纖維化	pulmonary fibrosis	15
肺紋增加	increased lung streak	24
肺野浸潤增加	lung field infiltration	85
肺門明顯	obvious hilar	45
胸椎退化性關節病變	degenerative joint disease of the thoracic spine	76
脊椎側彎	scoliosis	75
肋間肋膜增厚	intercostal pleural thickening	36
心臟肥大	cardiac hypertrophy	42
心臟節律器置放	heart pacemaker placement	7

Source: E-Da hospital

Sample size: 1924

Sample category: 19

Image format: DICOM

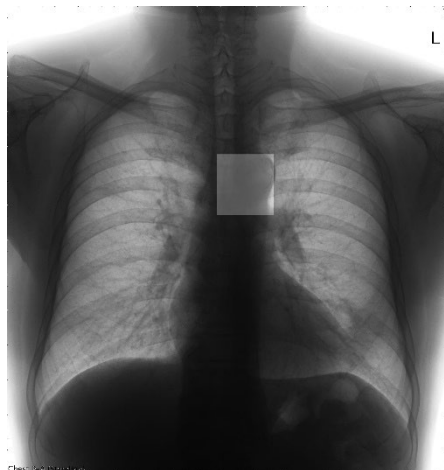
Image resolution: 0.16 mm per pixel

Image size: 1824~2688 pixels in length
1536~2680 pixels in width

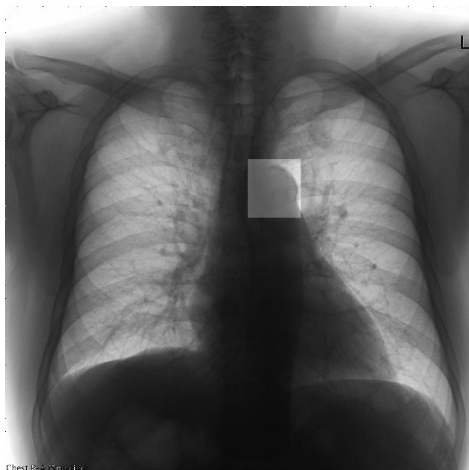
The dataset did not include patient personal information, such as name, age, and medical record number.

主動脈壁硬化(鈣化)

主動脈弓粥樣硬化斑塊



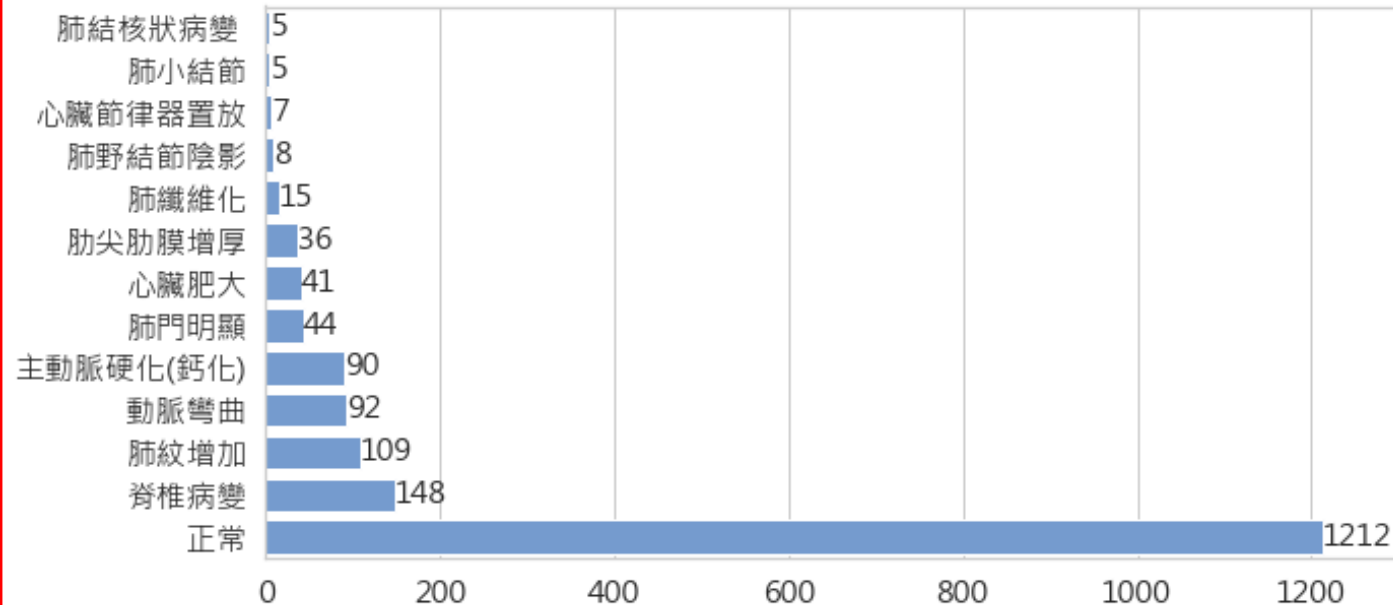
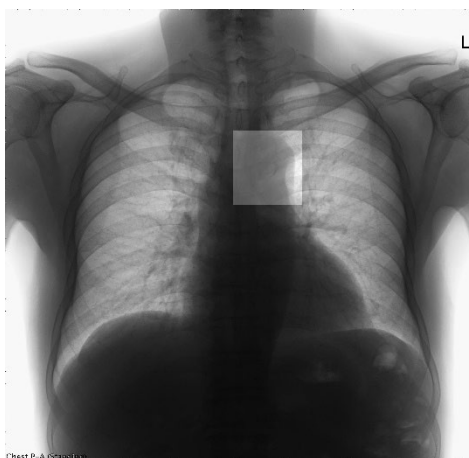
主動脈弓鈣化



主動脈粥樣硬化



主動脈壁鈣化



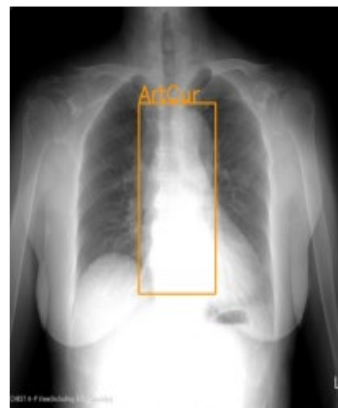
1924 – 110 (duplicate) – 1 (not displayed correctly) – 1 (without BBox)
= 1812

Bounding Boxes

主動脈硬化(鈣化)



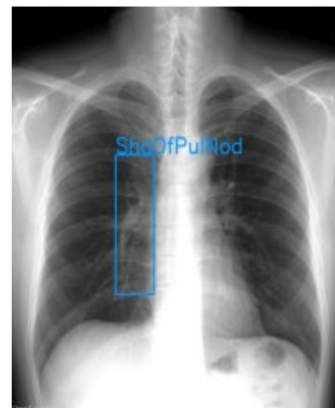
動脈彎曲



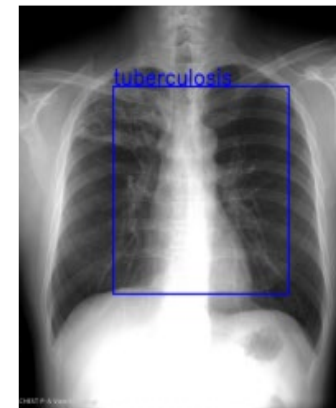
肺小結節



肺野結節陰影



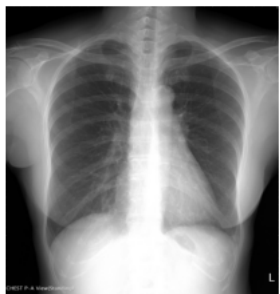
肺結核狀病變



肺纖維化



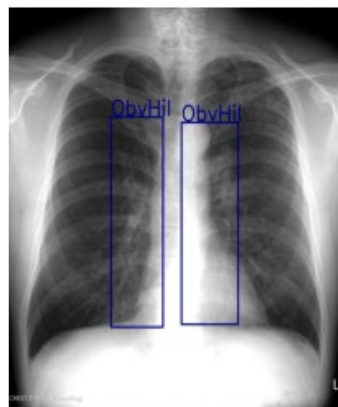
正常



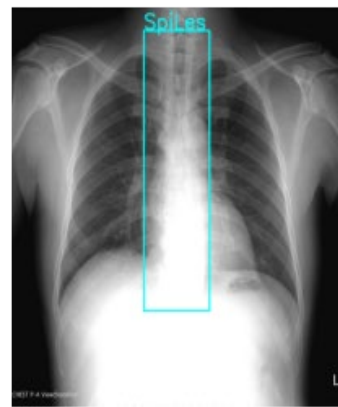
肺紋增加



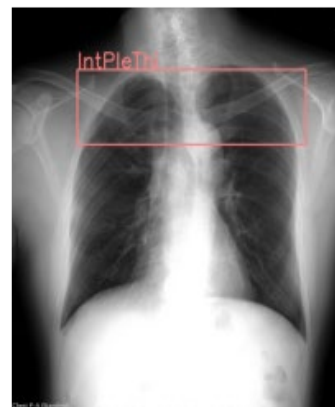
肺門明顯



脊椎病變



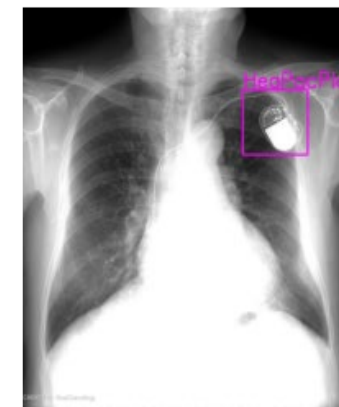
肺尖肋膜增厚



心臟肥大

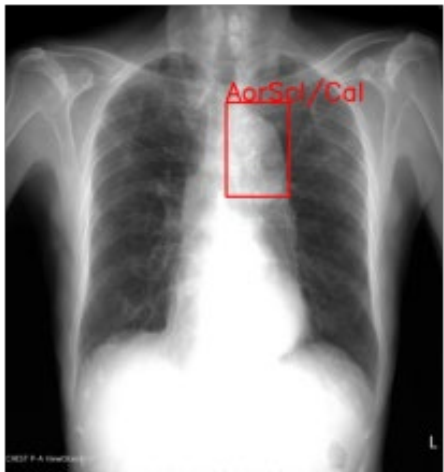


心臟節律器置放



ID = 'TDR0220161117153038'

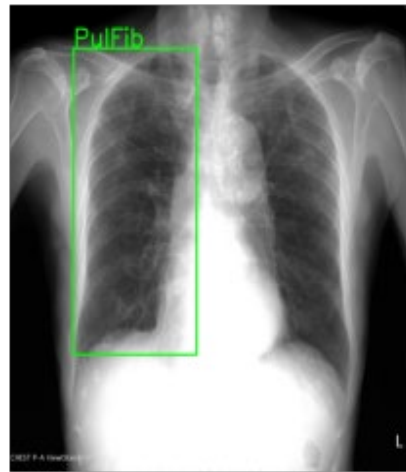
主動脈硬化(鈣化)



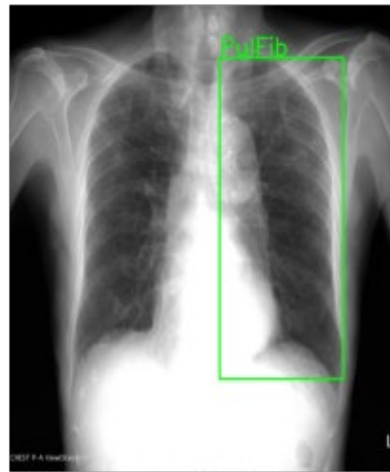
Label = 1



肺纖維化



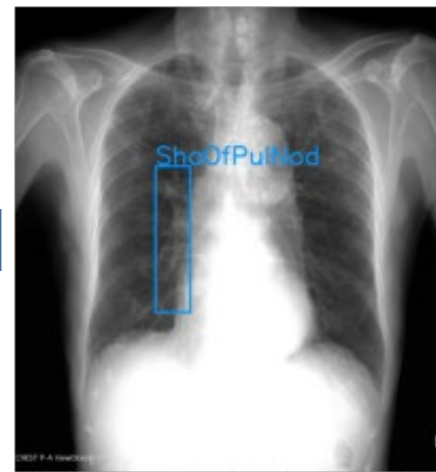
Label = 6



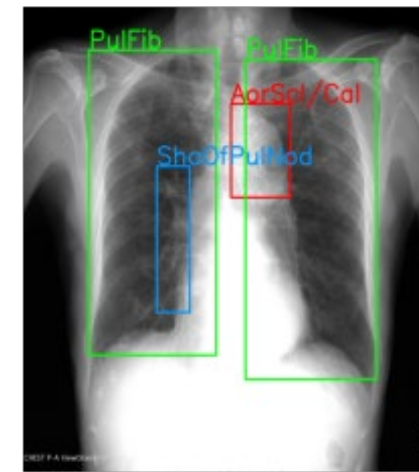
Label = 6



肺野結節陰影



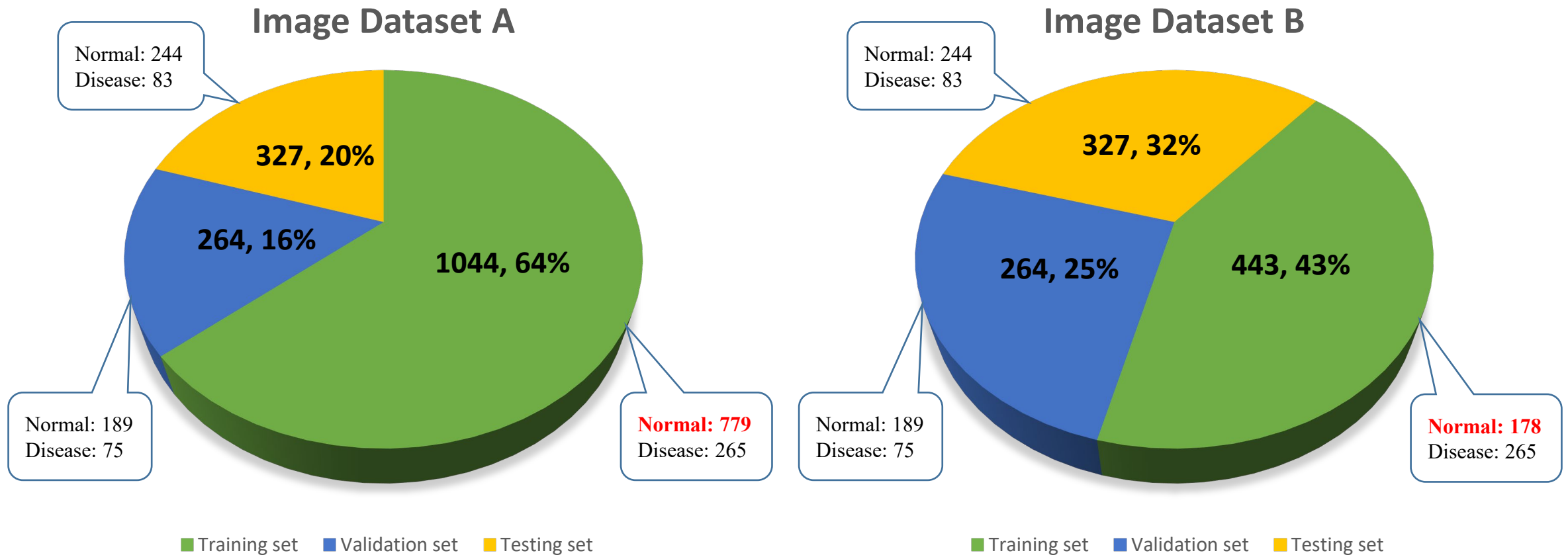
Label = 4

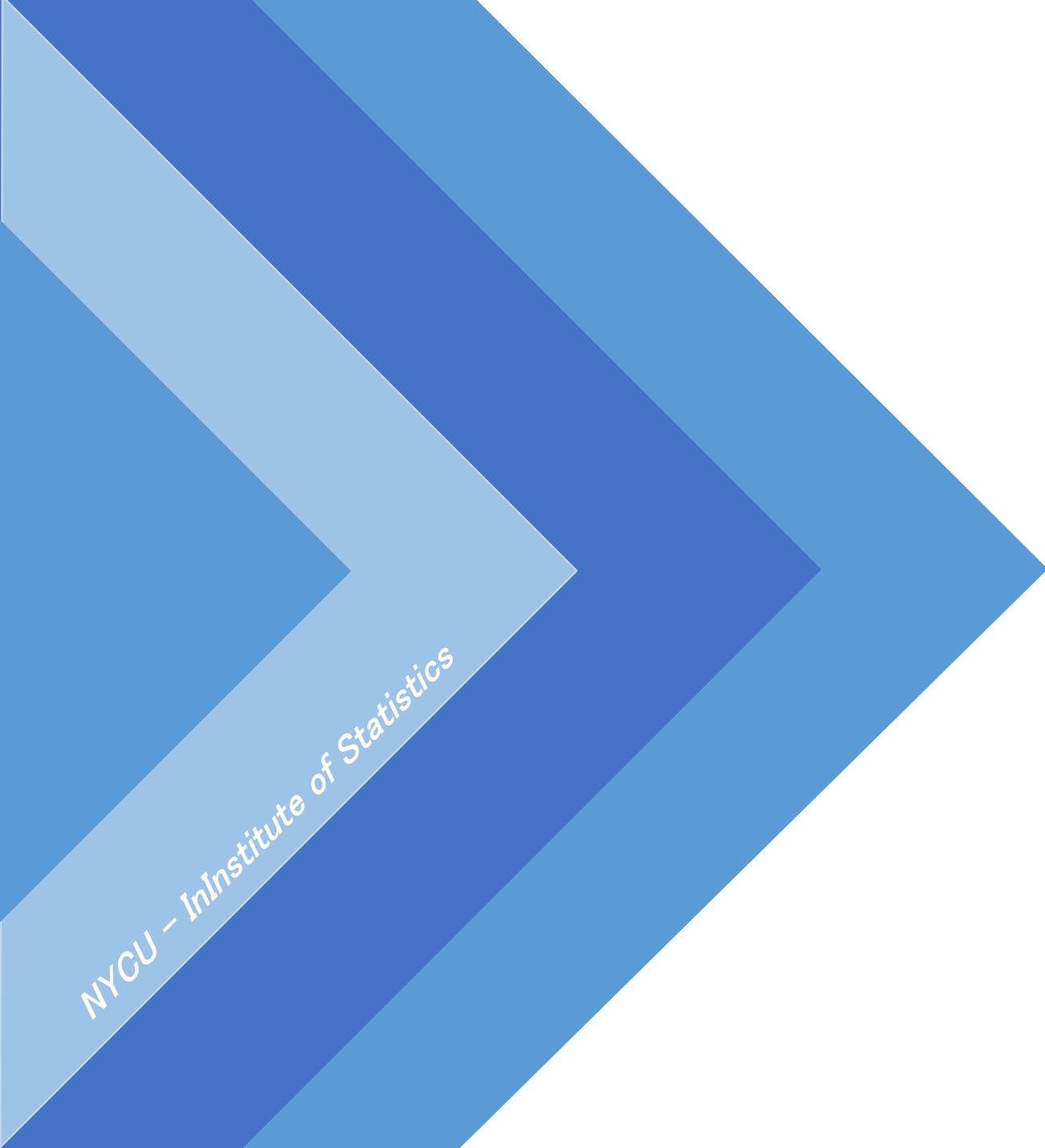


Label = [1,6,6,4]

By setting patient ID, $1635 = 1491$ (one-label) + 144 (multi-label)

Two Different Dataset





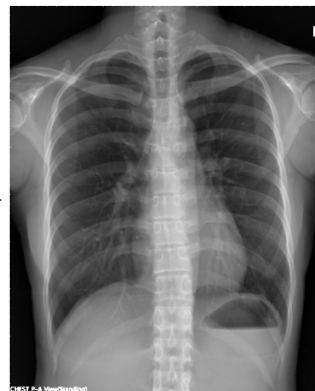
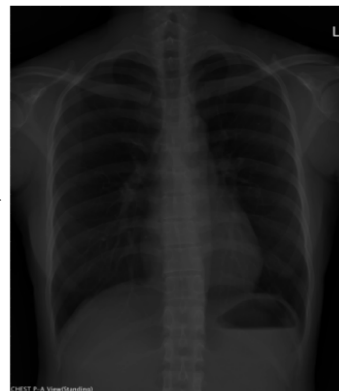
Materials and Implement

1.

Original DICOM
image

(vmin = 0, vmax = iMax)

(vmin = 0, vmax = 2.5)

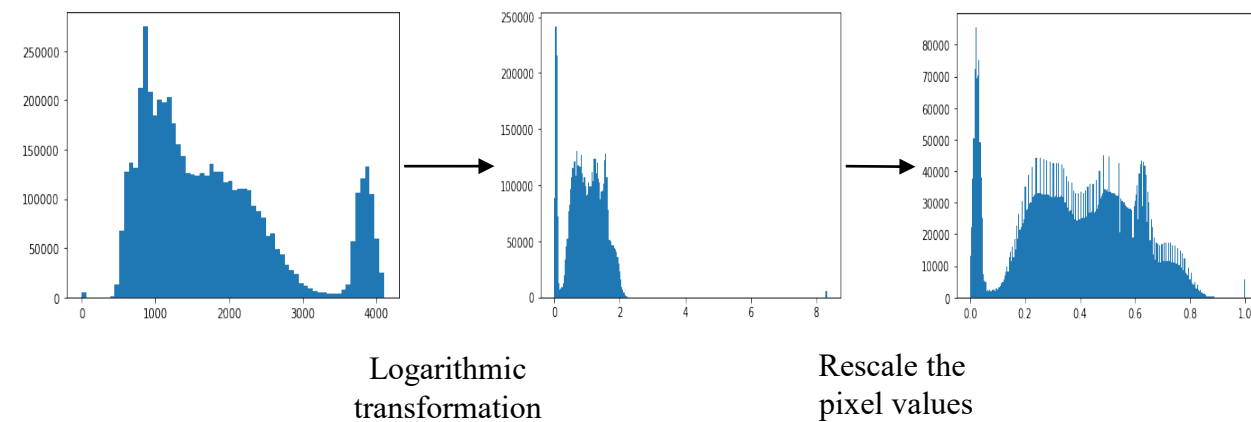
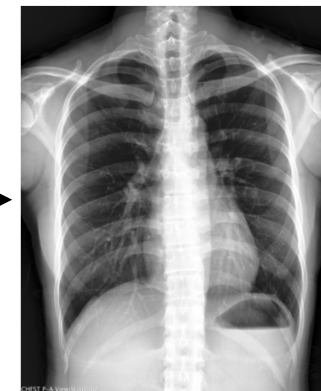
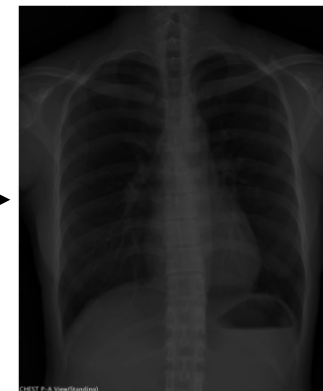
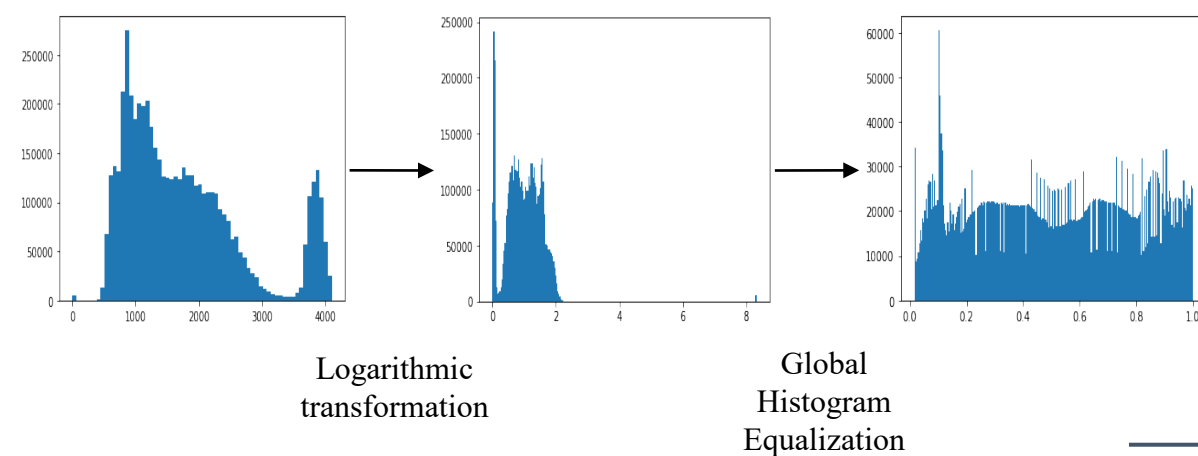


2.

Original DICOM
image

(vmin = 0, vmax = iMax)

(vmin = 0, vmax = 2.5)

Logarithmic
transformationRescale the
pixel valuesLogarithmic
transformationGlobal
Histogram
Equalization

Q: attenuation effect

A:

Algorithm 1: Logarithmic transformation algorithm for chest X-ray image.

Input: x

for $i = 0, \dots, N - 1$ **do**

if $x[i] < iMin$, **then** $x[i] = iMin$

if $x[i] > iMax$, **then** $x[i] = iMax$

$Z[i] = -\log((1 + x[i]) / 2^{\text{BitDepth}})$

Output: Z

Q: contrast enhancement

A:

1. Rescale the pixels

Algorithm 2: Rescale the pixels, which refers the Simplest Color Balance algorithm.

Input: Z

for $i = 0, \dots, N - 1$ **do**

$c[i] = (Z[i] - v_{min}) / (v_{max} - v_{min})$

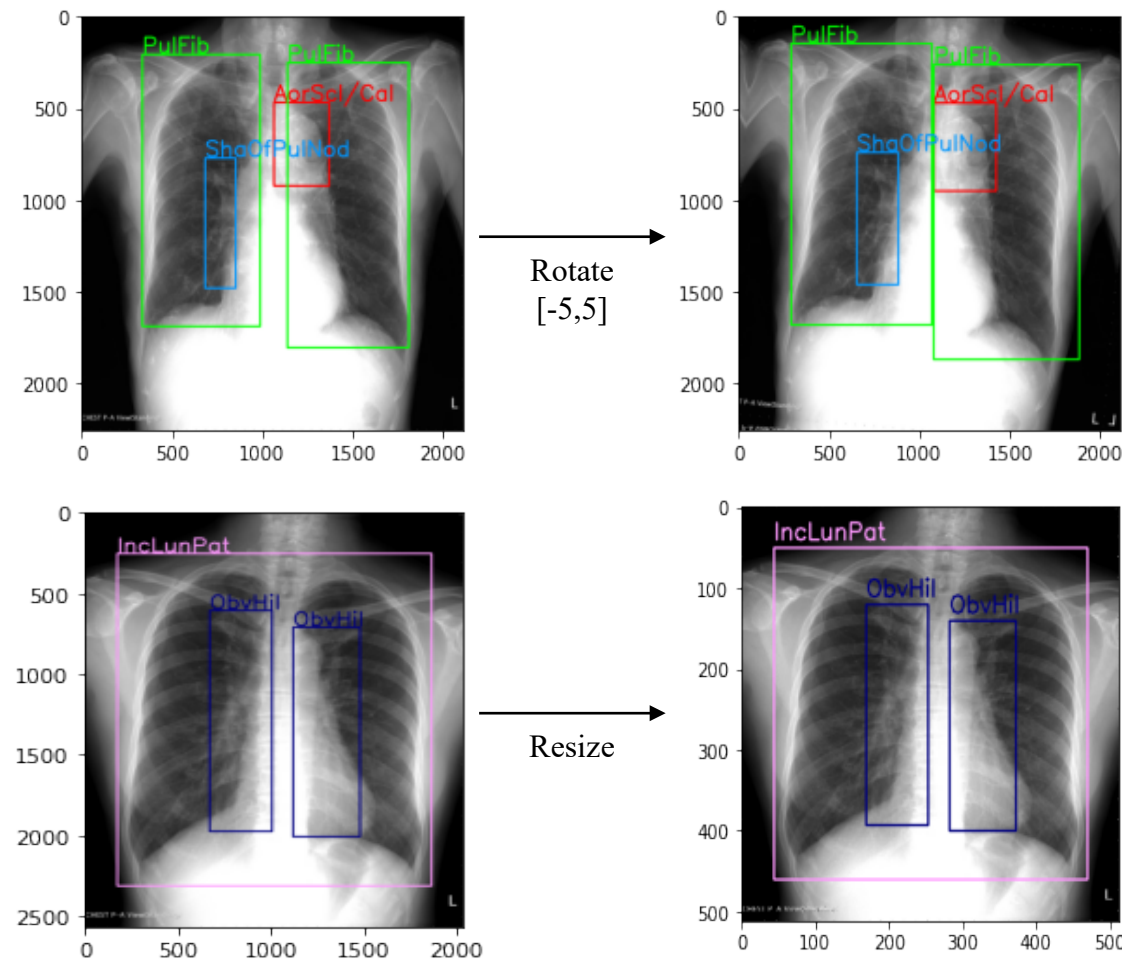
if $c[i] < 0$, **then** $c[i] = 0$

if $c[i] > 1$, **then** $c[i] = 1$

Output: c

2. Global Histogram Equalization

Augmentation	Clinical perspective	Reason
Horizontal Flip	not recommended	- Horizontal Flip does not change posterior-anterior (PA) to anterior-posterior (AP), and no existing data augmentation technique can simulate the differences between PA and AP CXR images. - It would lead to non-physiologic images (e.g., heart in the right thorax rather than the left thorax), which might confound learning.
Vertical Flip	not recommended	In medicine, there are no existing upside-down x-ray images.
Rotation	acceptable	Slight rotations such as between -5 and 5 are seen in clinical practice. However, 90-degree rotations are “not recommended.”
Resize	acceptable	An equal scaling is “possible,” however, resizing in only the x-axis or y-axis can be considered “not recommended” clinically.



Loss Function

	1 label	...	...	k label
1	1	0	...	0
⋮	0	0	...	0
⋮	⋮	⋮	⋮	⋮
m	0	1	...	1

$$Loss\ function_{mW-BCE} = -\frac{1}{M} \sum_{m=1}^M \left\{ \sum_{k: y_{mk}=1} \beta_{Pk} [\log(h_{\theta}(x_m))] + \beta_N \sum_{k: y_{mk}=0} [\log(1 - h_{\theta}(x_m))] \right\}$$

where β_{Pk} is set to $\frac{|P|+|N|}{|P_k|}$ while β_N is set to $\frac{|P|+|N|}{|N|}$. Here, $|P|$ and $|N|$ are the number of 1's and 0's in the label vectors of the training dataset, respectively, and $|P_k| = \sum_{m=1}^M y_{mk}$ is the number of 1's in the k th label. Notice that our weighting scheme gives a different weight for each disease label.

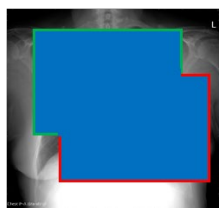
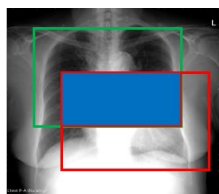
Classification

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

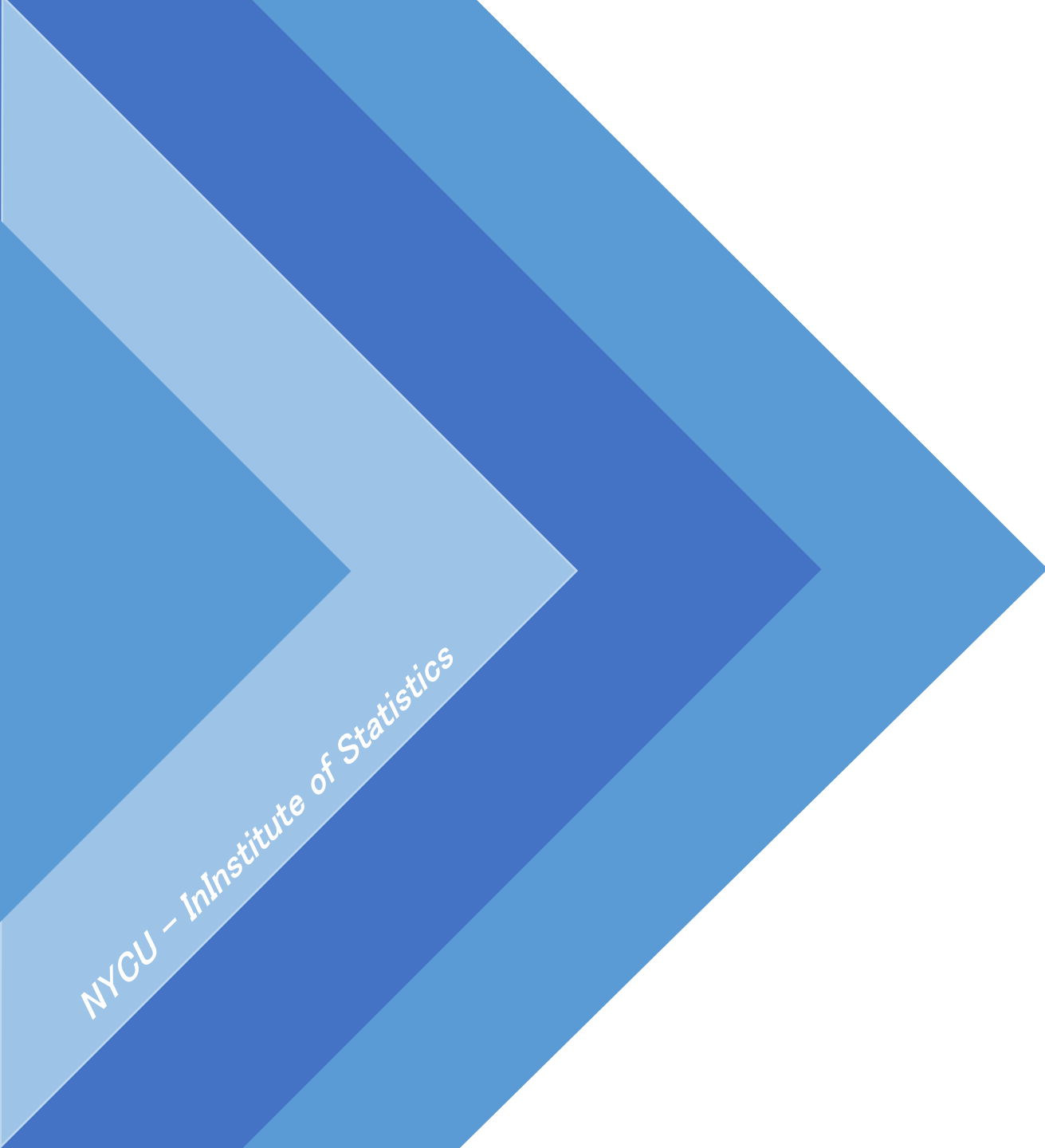
Metrics	Formula	Definition
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$	It generally measures the ratio of correct predictions to the total number of input samples.
Precision	$\frac{TP}{TP + FP}$	The proportion of all the positive predicted samples that were predicted as positive.
Recall	$\frac{TP}{TP + FN}$	The proportion of all the actual positive samples that were predicted as positive.
Sensitivity	$\frac{TP}{TP + FN}$	The proportion among those with the disease is predicted to be positive.
Specificity	$\frac{TN}{TN + FP}$	The proportion among those without the disease is predicted to be negative.

Object detection

$$IoU = \frac{area(\mathcal{B}_p \cap \mathcal{B}_{gt})}{area(\mathcal{B}_p \cup \mathcal{B}_{gt})}$$



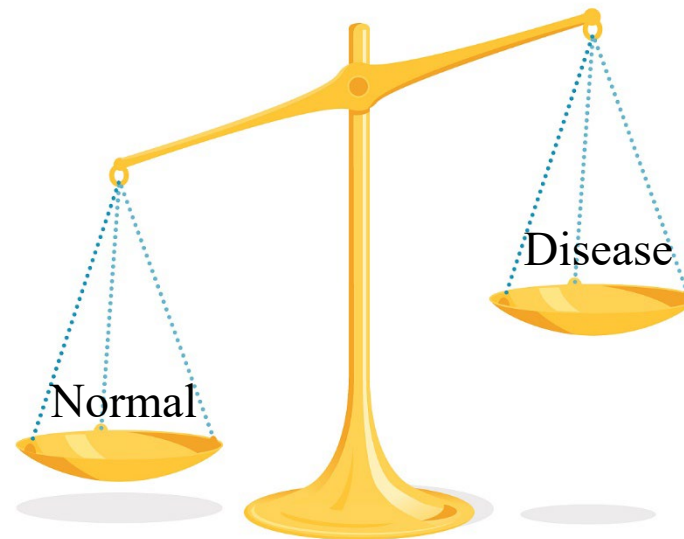
Metrics	Definition
True Positive (TP)	$IoU \geq 0.5$
False Positive (FP)	$IoU < 0.5$
False Negative (FN)	Failing to detect BBox
True Negative (TN)	No BBox
Precision (P)	TP/all detections
Recall (R)	TP/all ground truths BBox
Average Precision (AP)	Under the area of Precision-Recall curve
Average Recall (AR)	Twice the under area of R-IOU curve



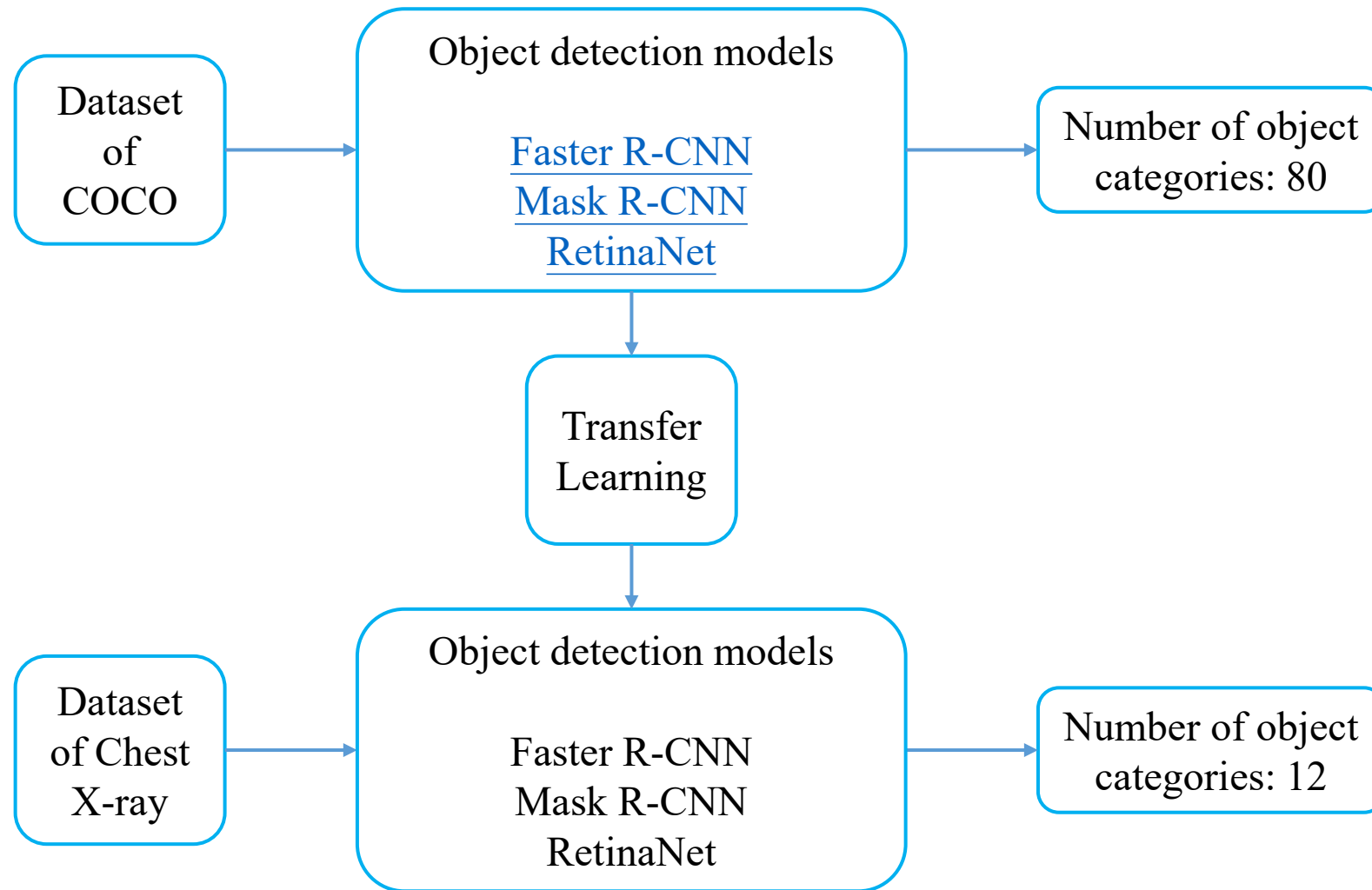
NYCU – Institute of Statistics

Methods

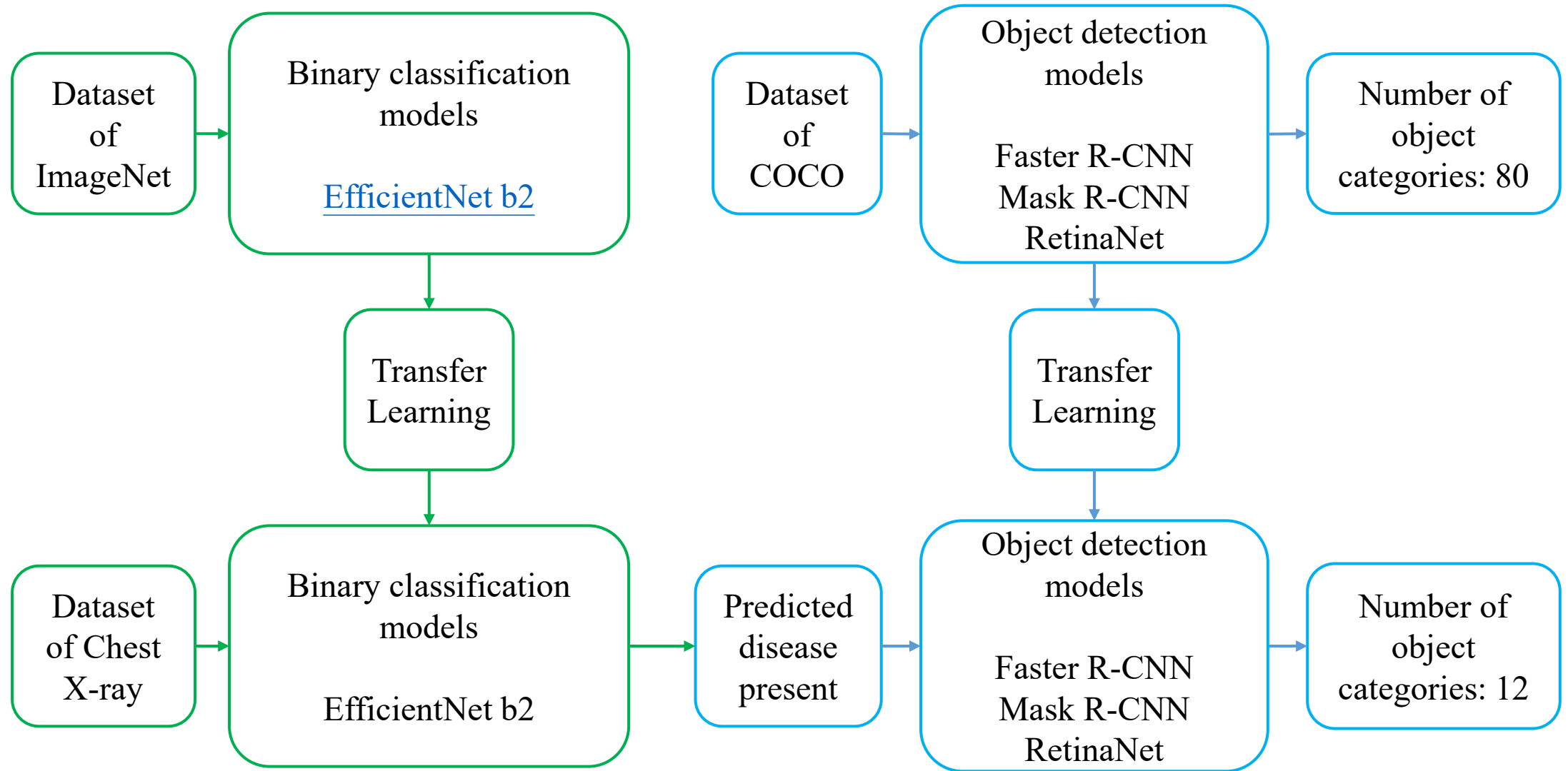
We design four different methods for chest anomaly detection. Because **the Normal category in the dataset has more images than any other disease categories**, resulting in insufficient images to be learned for detection the diseases, we are facing a considerable class imbalance difficulty. Therefore, **we are interested in whether it can effectively improve the performance of the model or not if the images are first classified as diseased, and then the predicted diseased images are input into the object detection model**. This approach distinguishes images with essential features and further solves the impact of class imbalance on the object detection model.



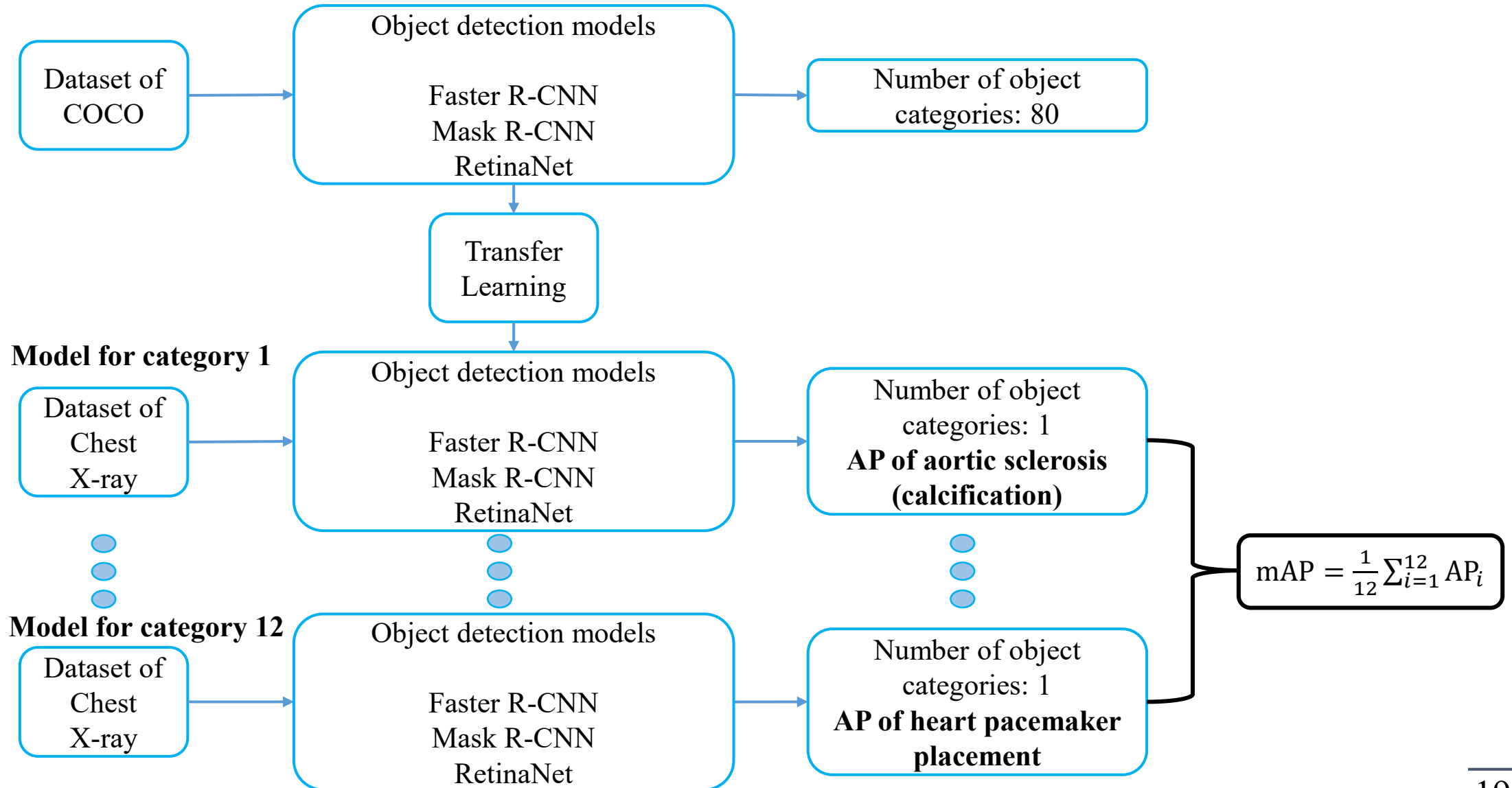
Object Detection Scheme 1



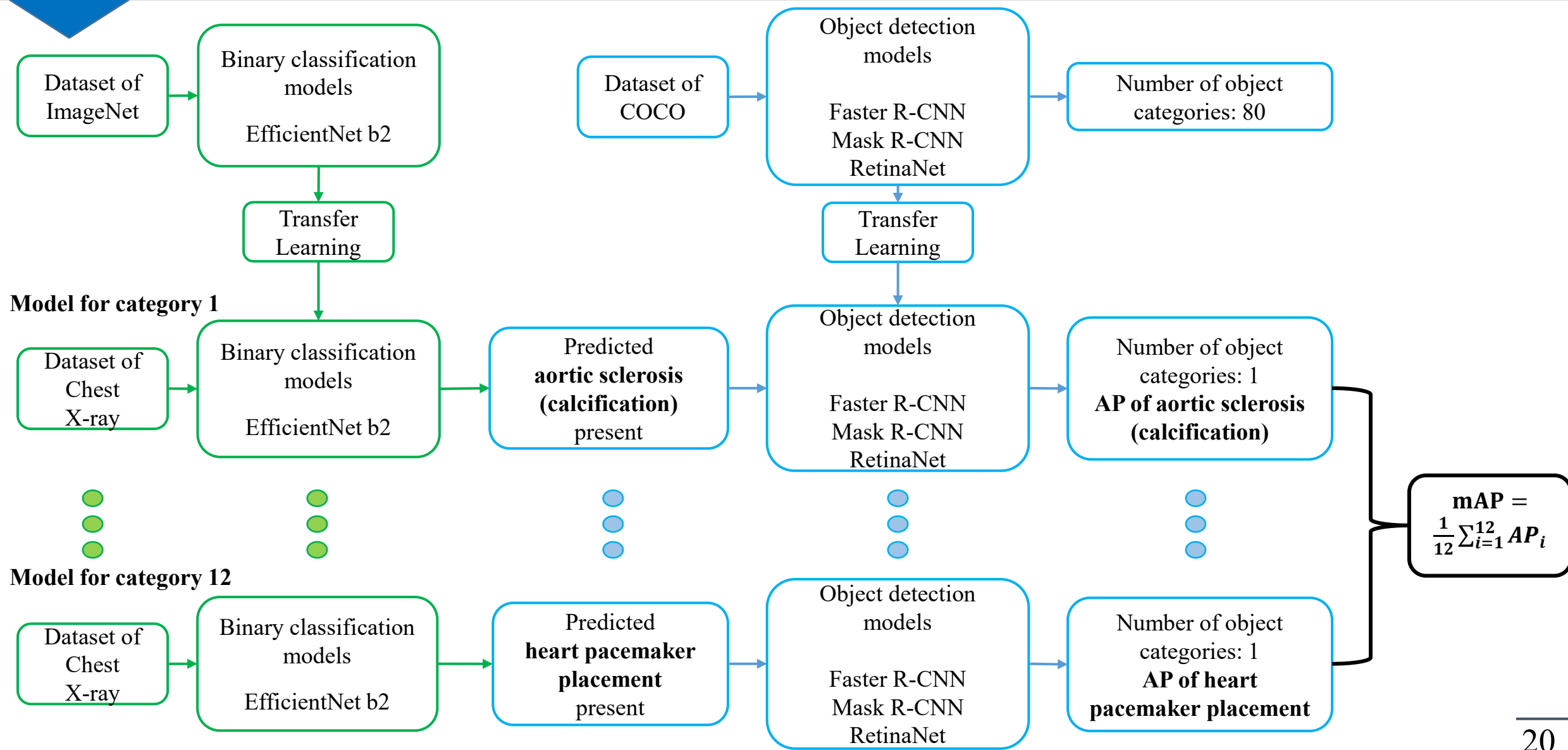
Object Detection Scheme 2

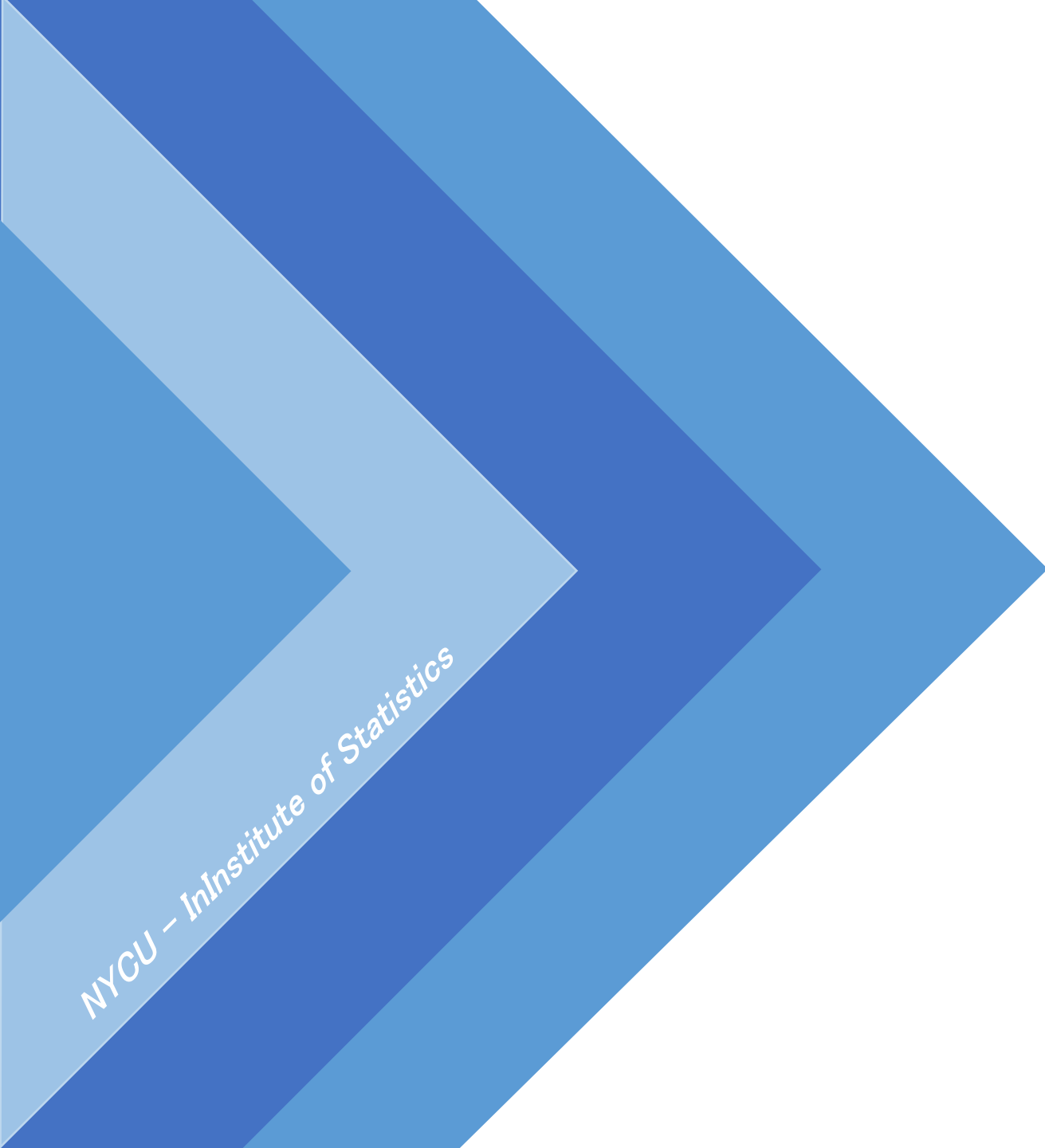


Object Detection Scheme 3



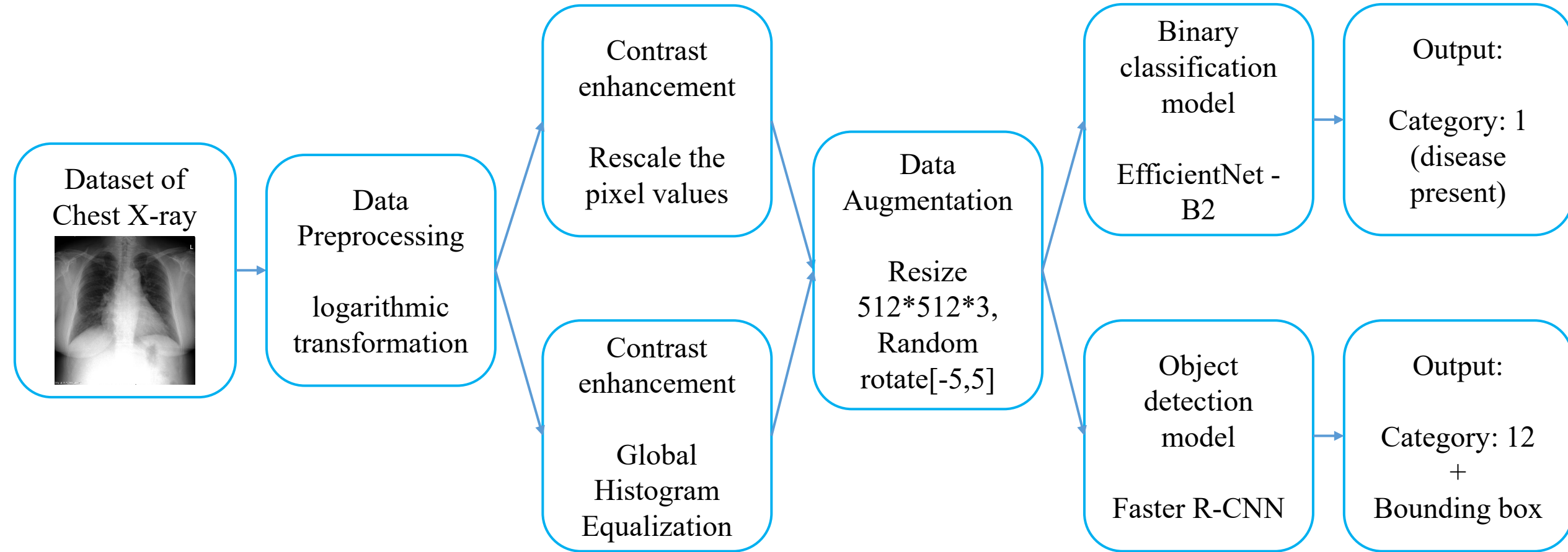
Object Detection Scheme 4





Experiment Results

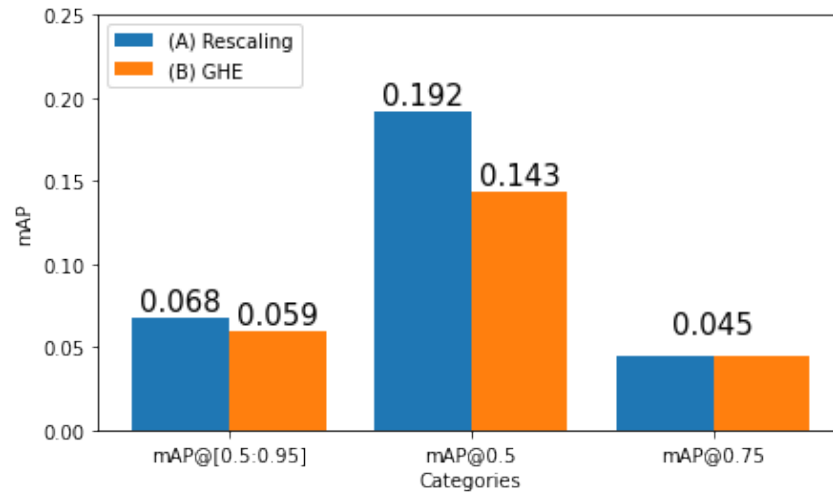
Pre-Processing Comparison



Classification

Data Preprocessing	Accuracy	Precision	Recall	Sensitivity	Specificity
(A) Rescale the pixel values	88.69	80.26	73.49	73.49	93.85
(B) GHE	85.02	78.33	56.63	56.63	94.67

Object detection

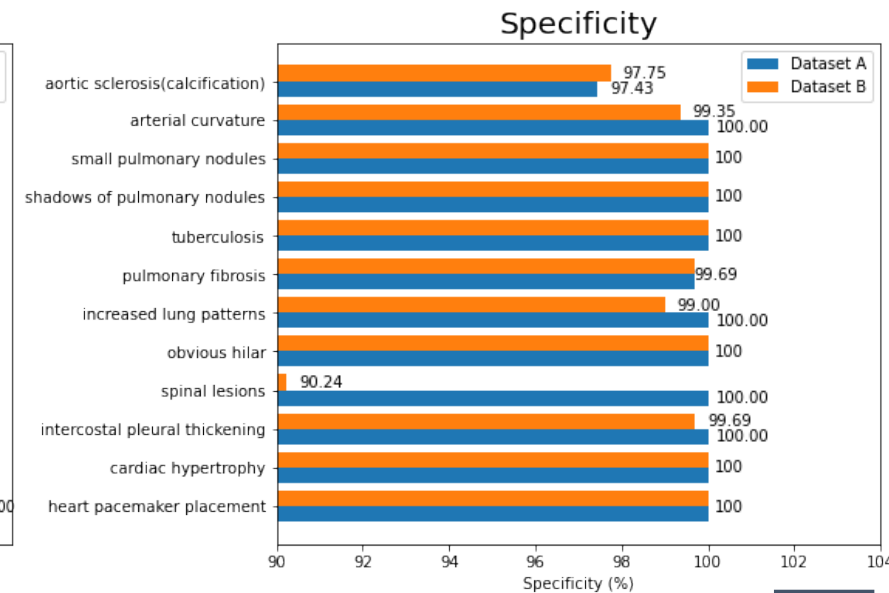
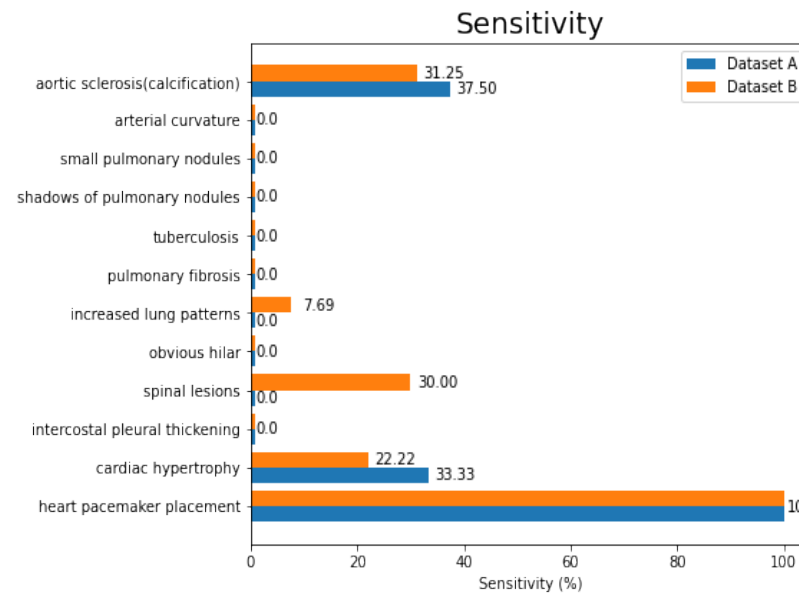
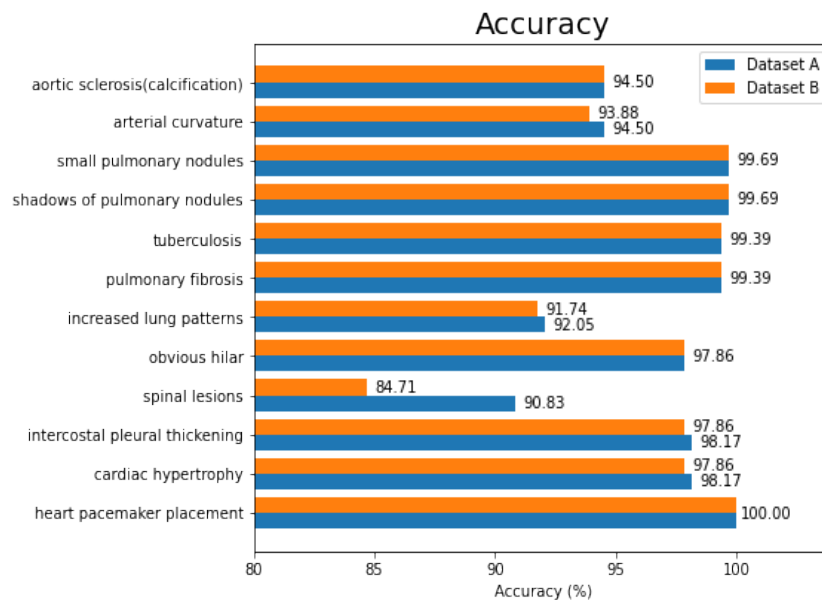


We selected to rescale the pixel values after logarithmic transformation to preprocess the images for the subsequent modeling.

Prediction on All Disease Categories Combined

Training set	Accuracy	Precision	Recall	Sensitivity	Specificity
Dataset A	88.69	80.26	73.49	73.49	93.85
Dataset B	81.35	59.82	80.72	80.72	81.56

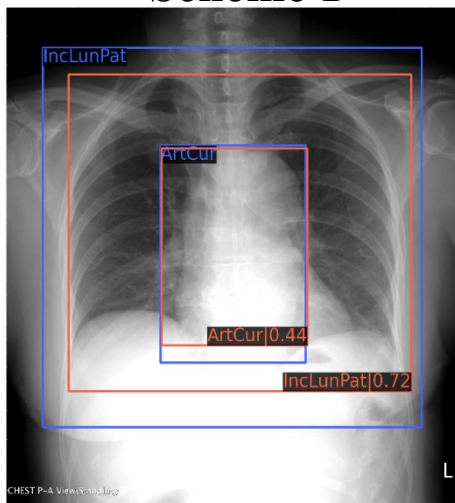
Individual Prediction for 12 Disease Categories



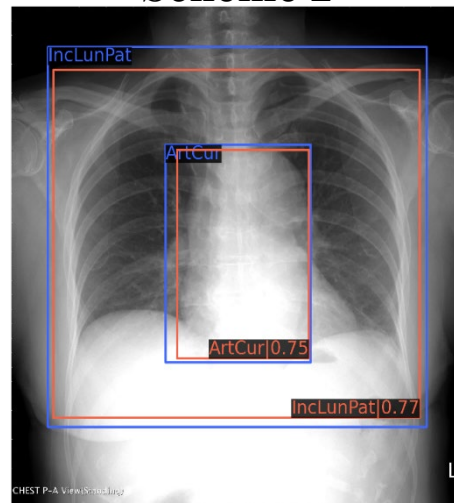
Scheme 1 vs. Scheme 2 vs. Scheme 3 vs. Scheme 4

Faster R-CNN

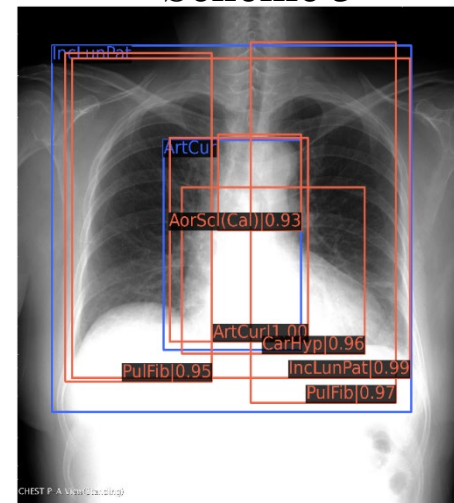
Scheme 1



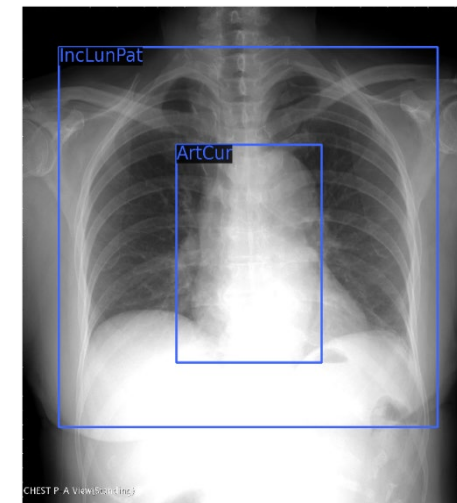
Scheme 2



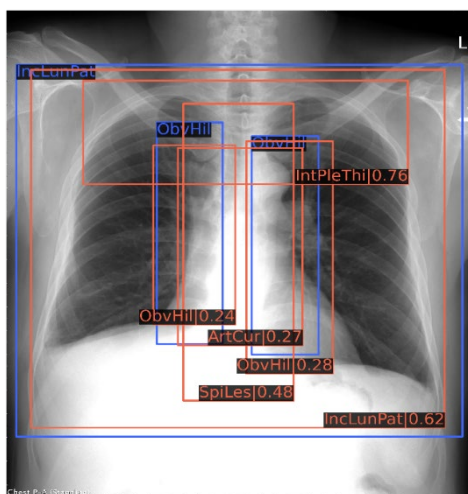
Scheme 3



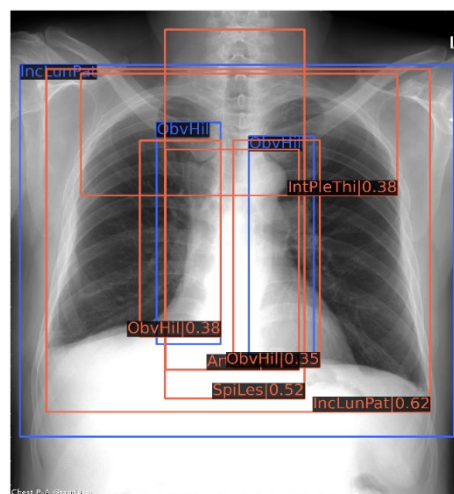
Scheme 4



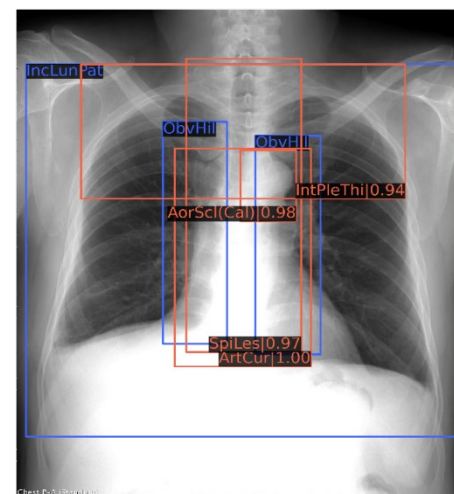
Scheme 1



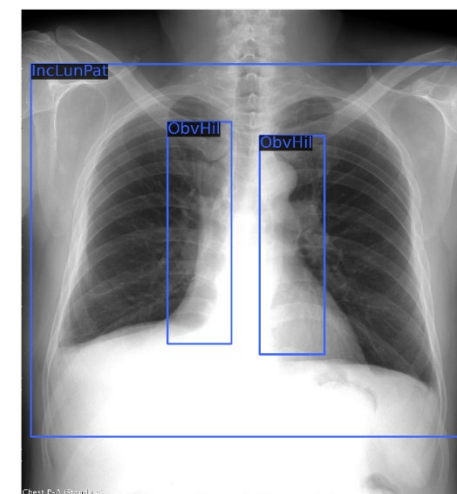
Scheme 2



Scheme 3



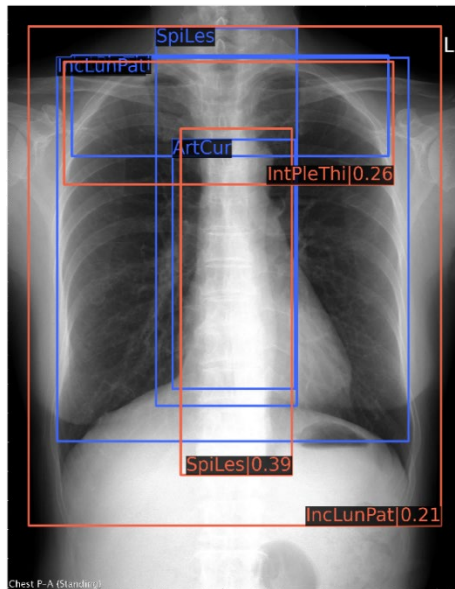
Scheme 4



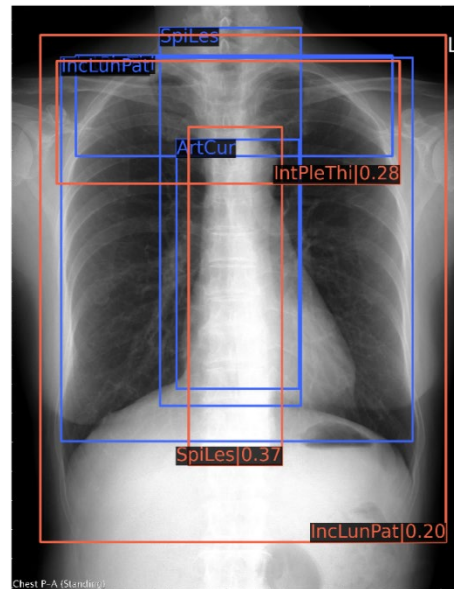
Mask R-CNN

RetinaNet

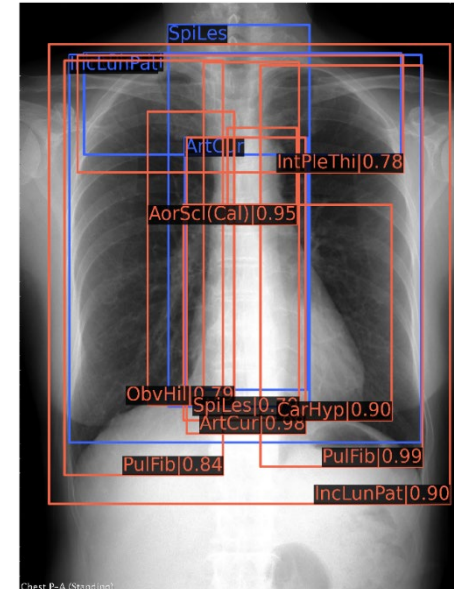
Scheme 1



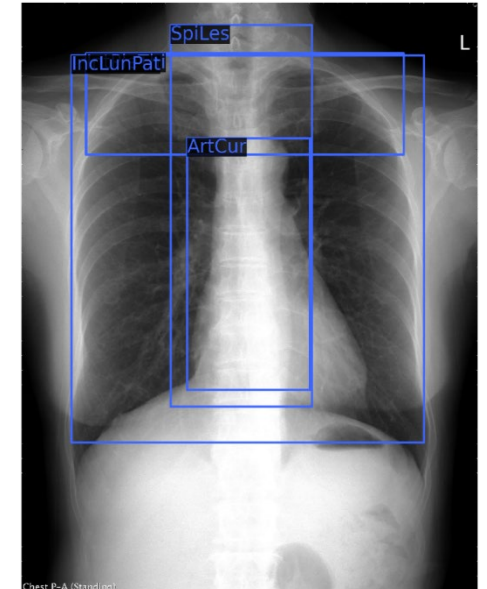
Scheme 2

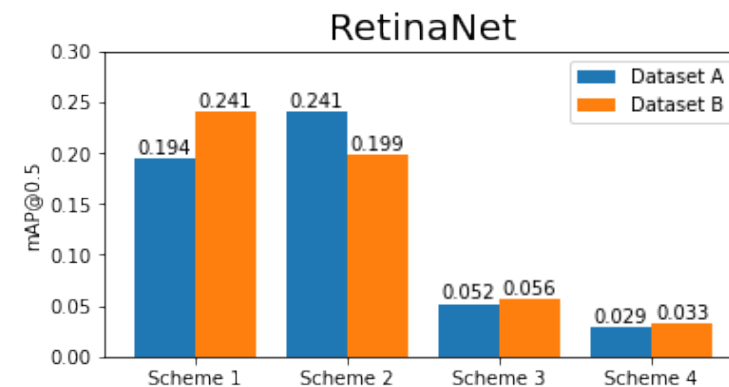
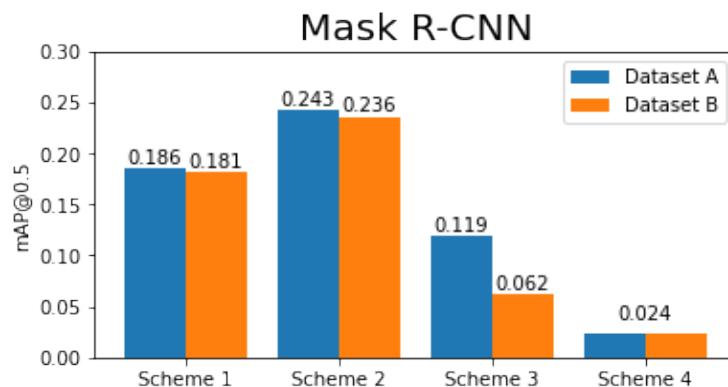
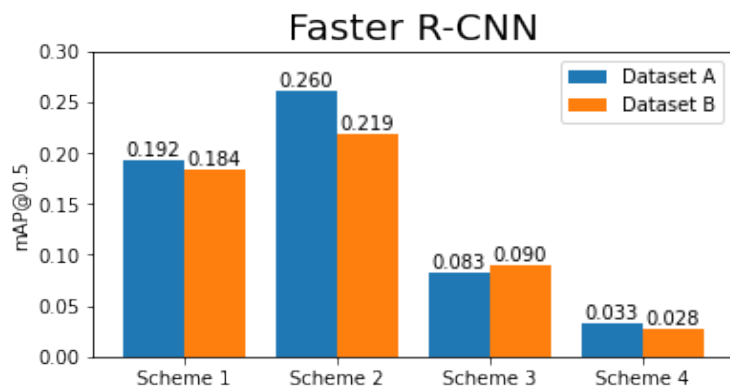


Scheme 3



Scheme 4





Scheme 1 vs. Scheme 2



Scheme 3 vs. Scheme 4

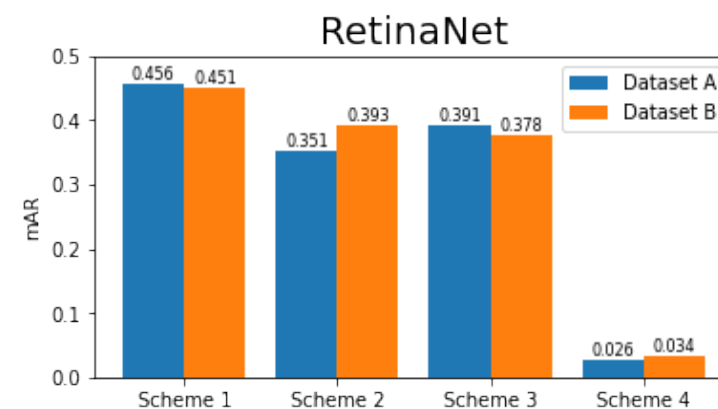
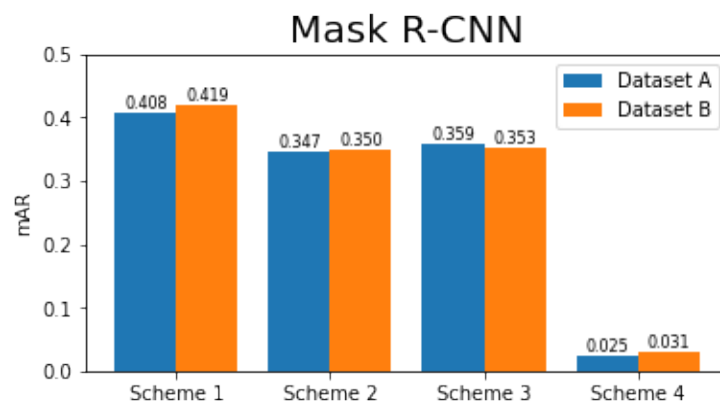
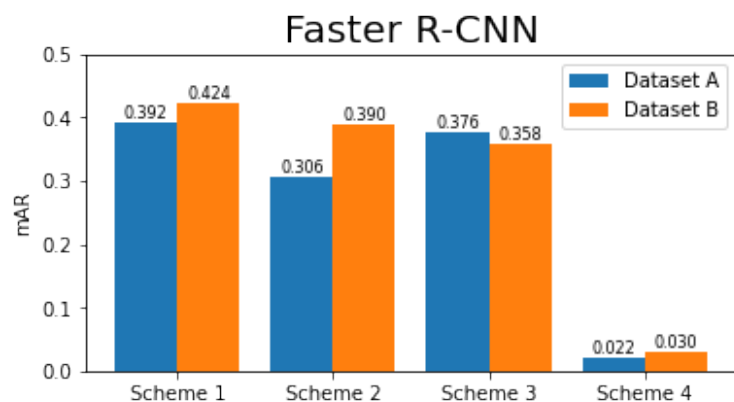


Scheme 1 vs. Scheme 3

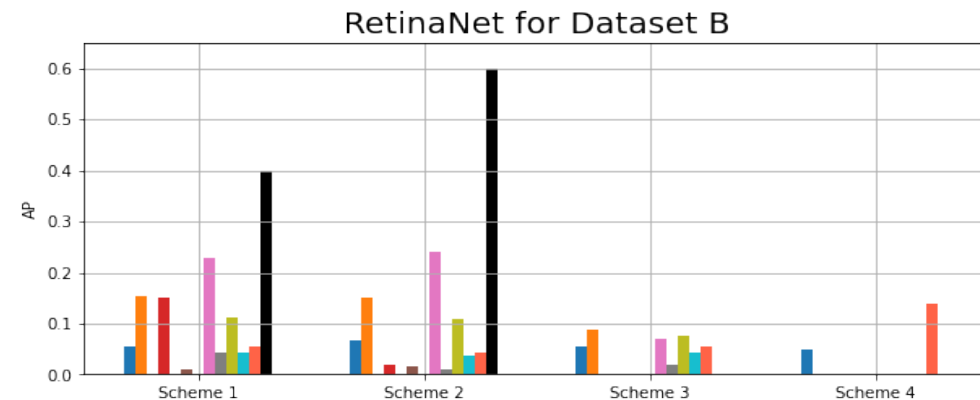
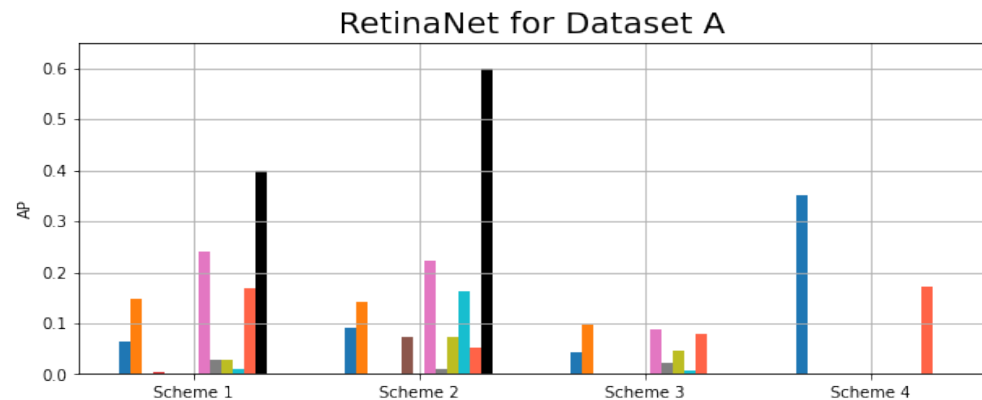
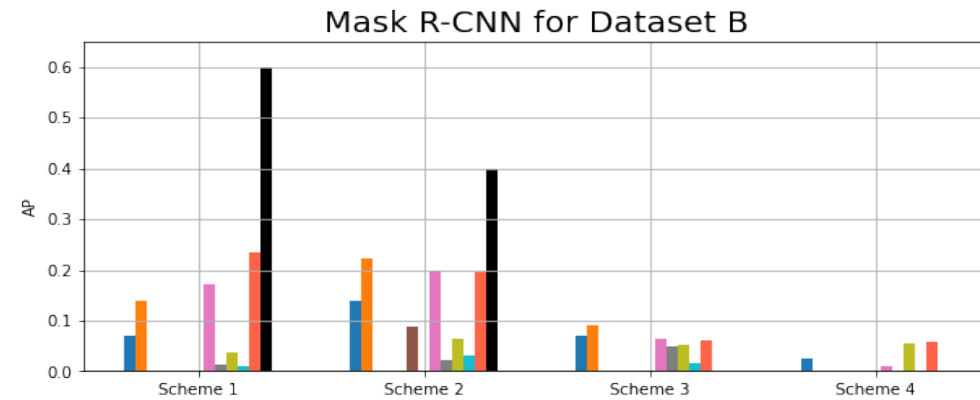
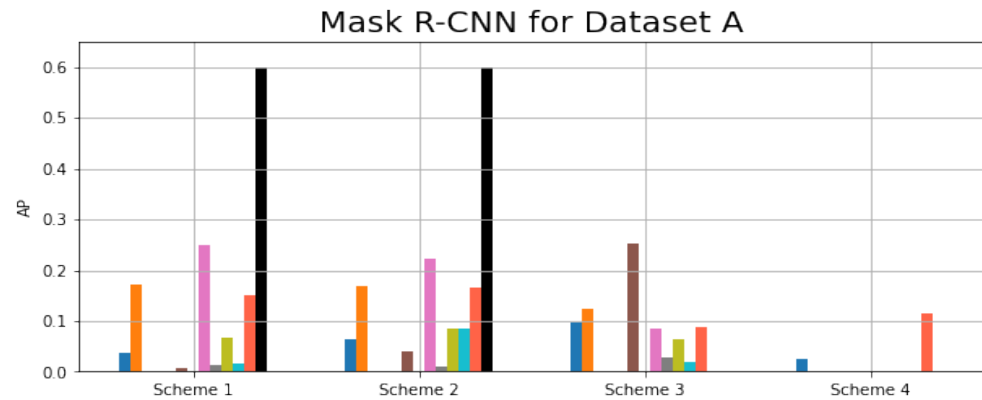
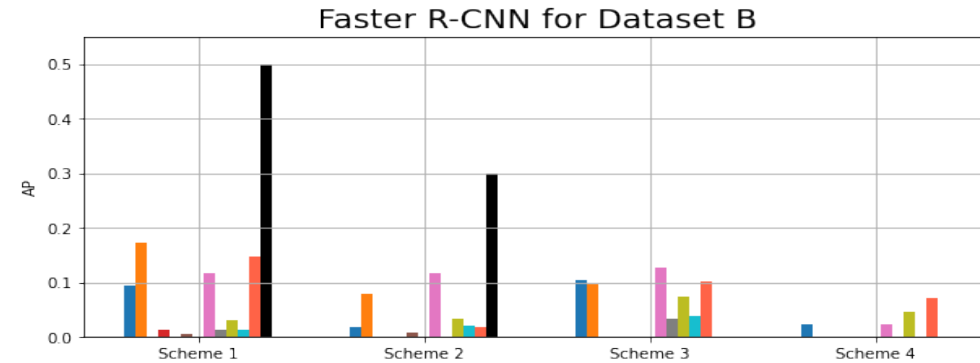
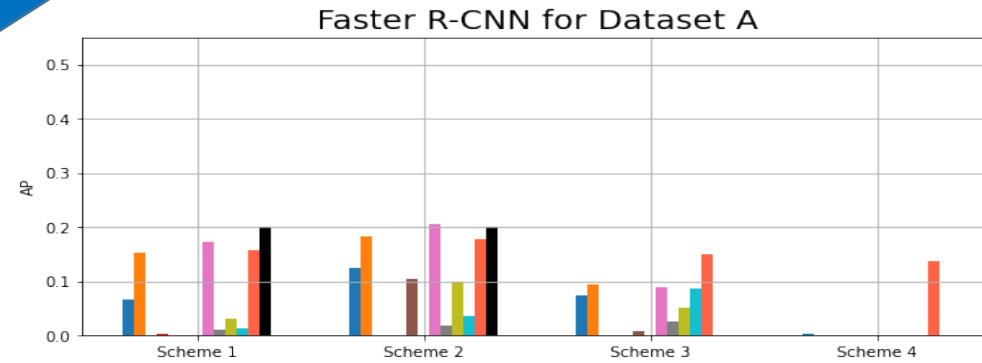


Dataset A vs. Dataset B





The average precision (AP of 12 categories)



- aortic sclerosis(calcification)
- arterial curvature
- small pulmonary nodules
- shadows of pulmonary nodules
- tuberculosis
- pulmonary fibrosis
- increased lung patterns
- obvious hilar
- spinal lesions
- intercostal pleural thickening
- cardiac hypertrophy
- heart pacemaker placement

Conclusion

Results

- After log-transforming images' pixel values, **the rescaling pixel value contrast enhancement method** results in better classification and object detection performance than the global histogram equalization method does.
- **The object detection Scheme 2** best among the four schemes.

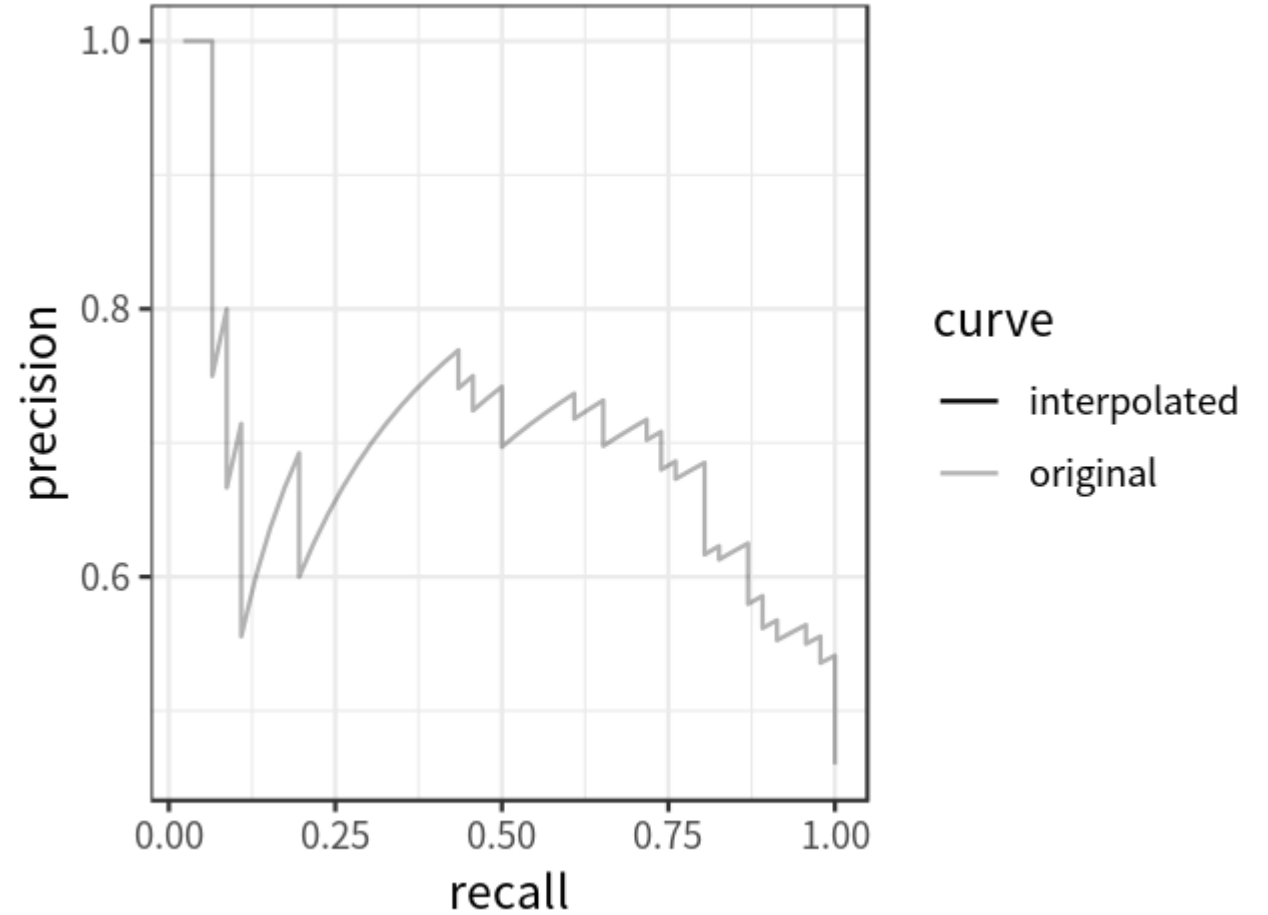
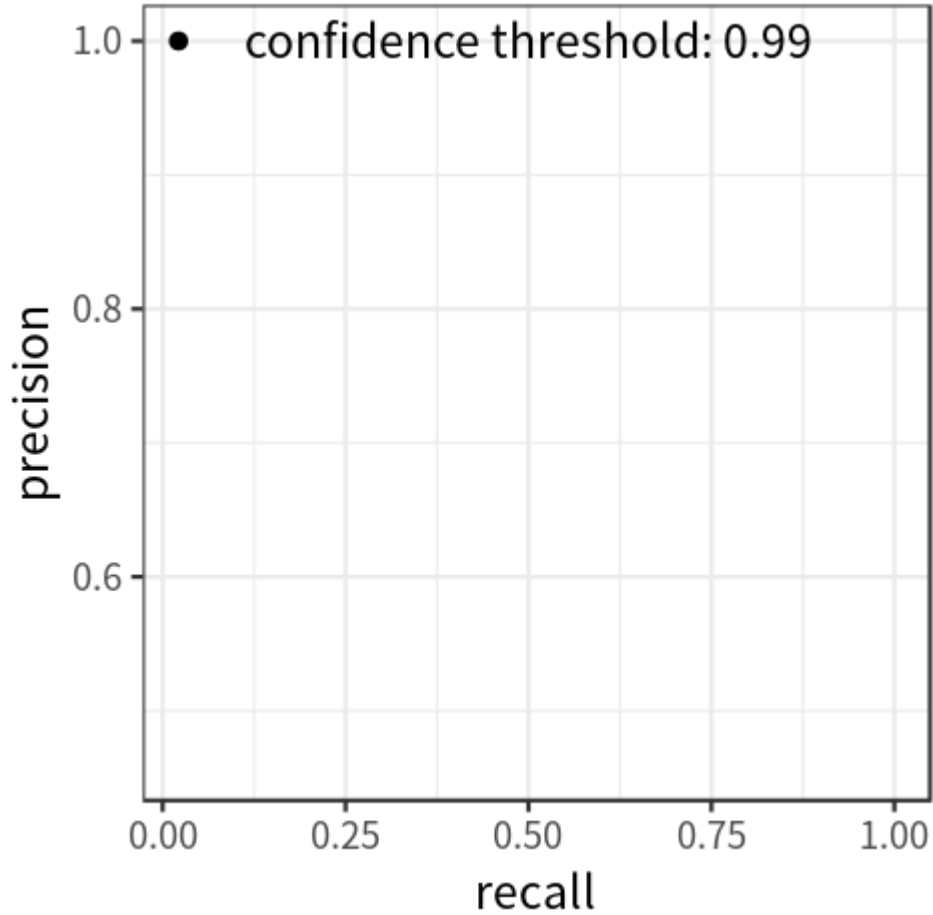
Future work

- In Scheme 4, because the problem of class imbalance has not been fully solved, the detection results of a single target disease of a small number of samples are not ideal
- Establishing targeted deep learning models based on the characteristics of different medical images is a significant future development direction of artificial intelligence medical image analysis.

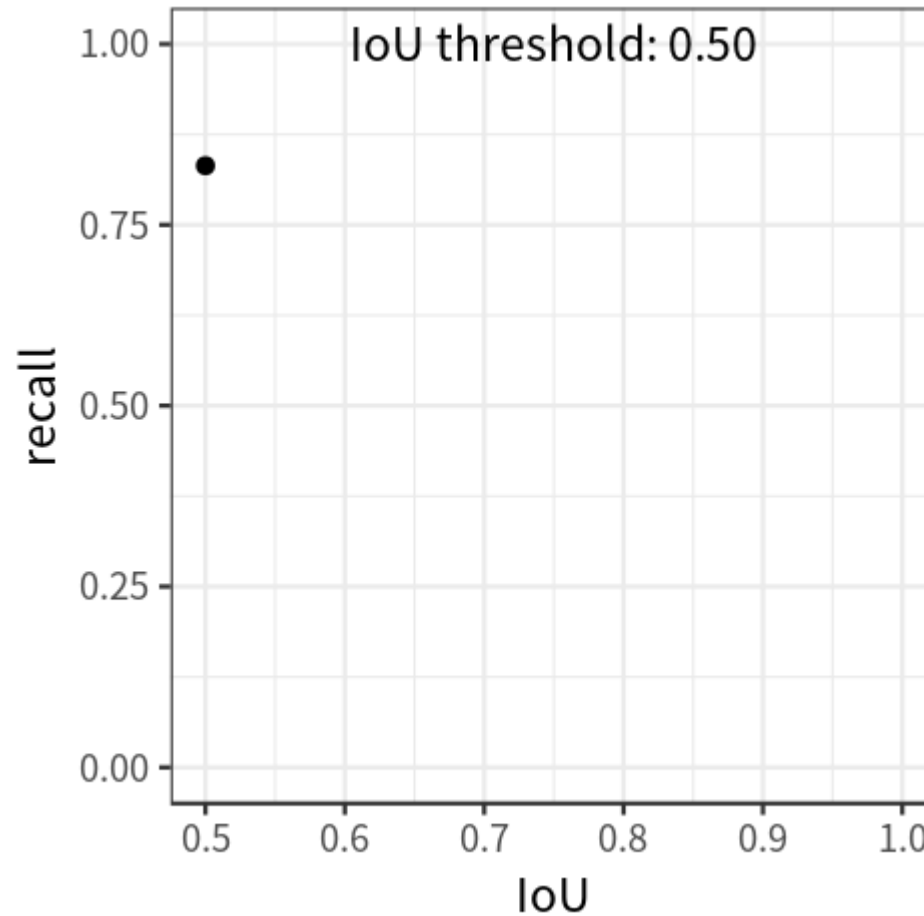
Thank You

國立陽明交通大學
National Yang Ming Chiao Tung University

Precision x Recall Curve

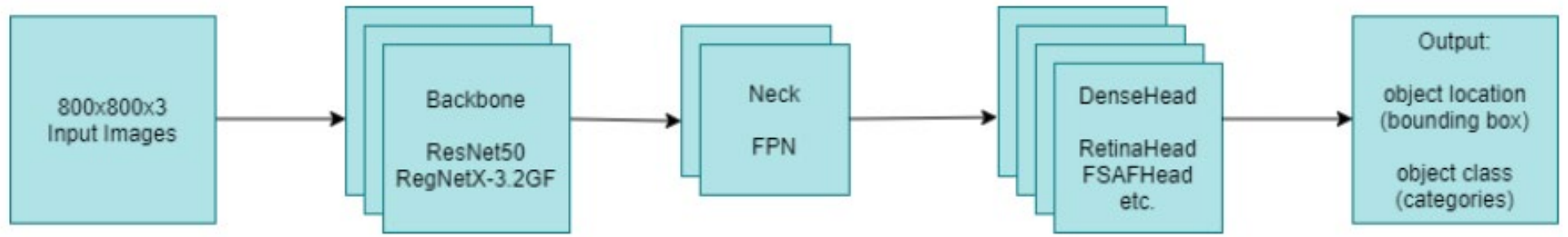


Recall-IoU Curve

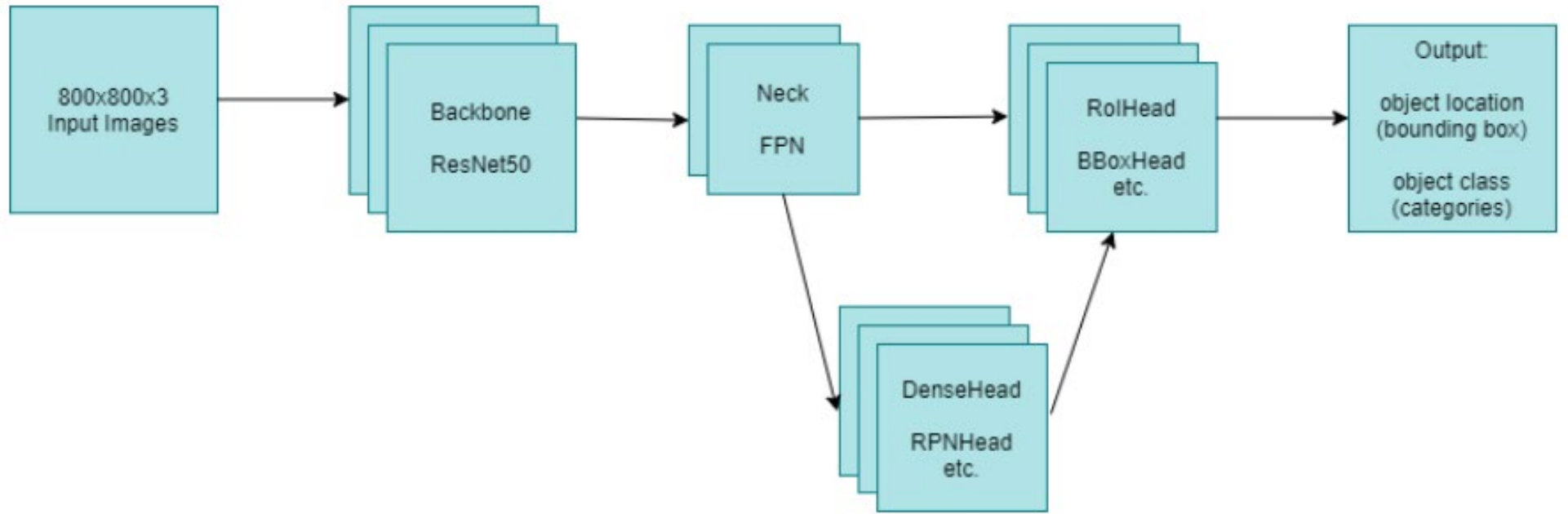


Object Detection

One-stage



Two-stage



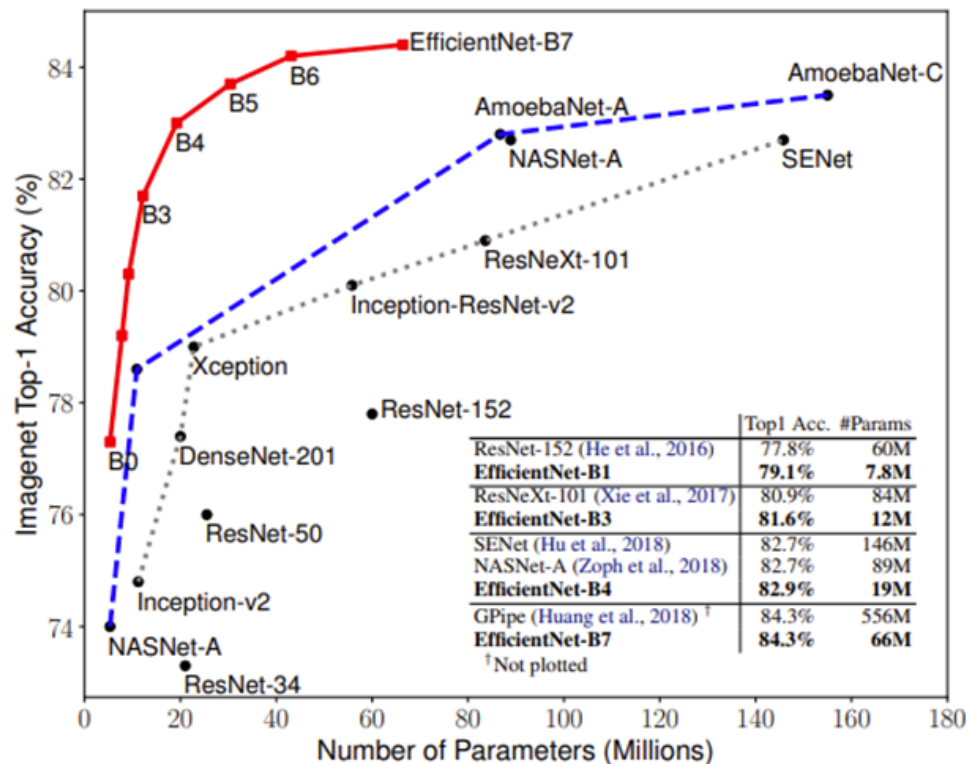


Figure 1. Model Size vs. ImageNet Accuracy. All numbers are for single-crop, single-model. Our EfficientNets significantly outperform other ConvNets. In particular, EfficientNet-B7 achieves new state-of-the-art 84.3% top-1 accuracy but being 8.4x smaller and 6.1x faster than GPipe. EfficientNet-B1 is 7.6x smaller and 5.7x faster than ResNet-152. Details are in Table 2 and 4.

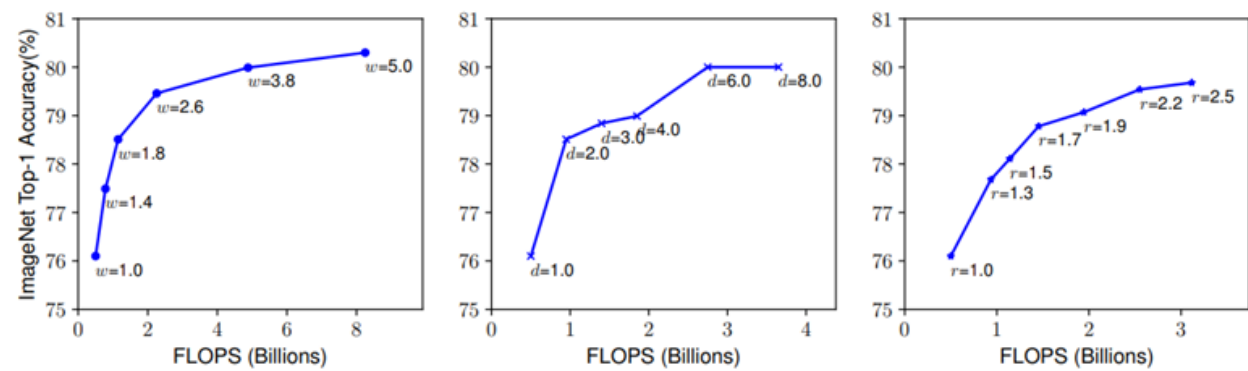


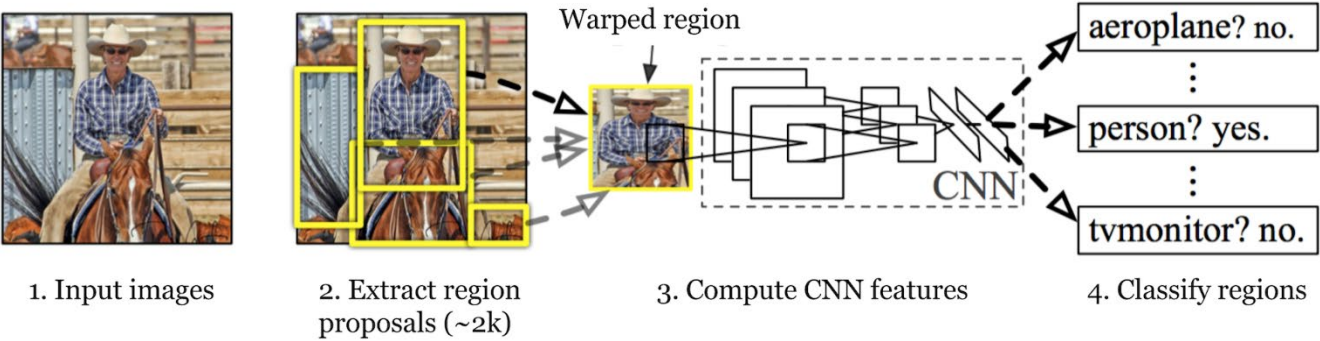
Figure 3. Scaling Up a Baseline Model with Different Network Width (w), Depth (d), and Resolution (r) Coefficients. Bigger networks with larger width, depth, or resolution tend to achieve higher accuracy, but the accuracy gain quickly saturate after reaching 80%, demonstrating the limitation of single dimension scaling. Baseline network is described in Table 1.

複合縮放 (Compound Scaling)

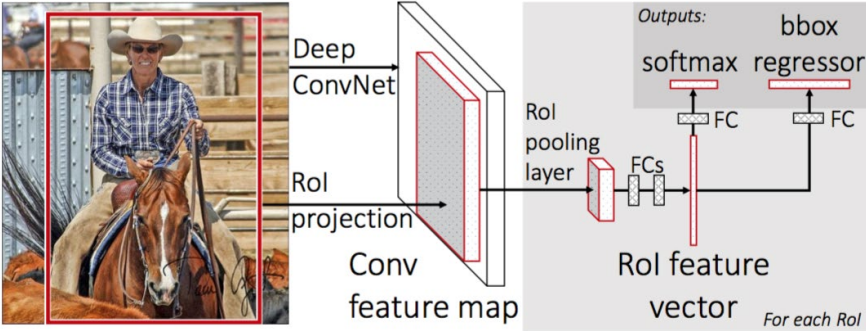
$$\begin{aligned}
 &\text{depth: } d = \alpha^\phi \\
 &\text{width: } w = \beta^\phi \\
 &\text{resolution: } r = \gamma^\phi \\
 &\text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\
 &\alpha \geq 1, \beta \geq 1, \gamma \geq 1
 \end{aligned}
 \tag{3}$$

Faster R-CNN

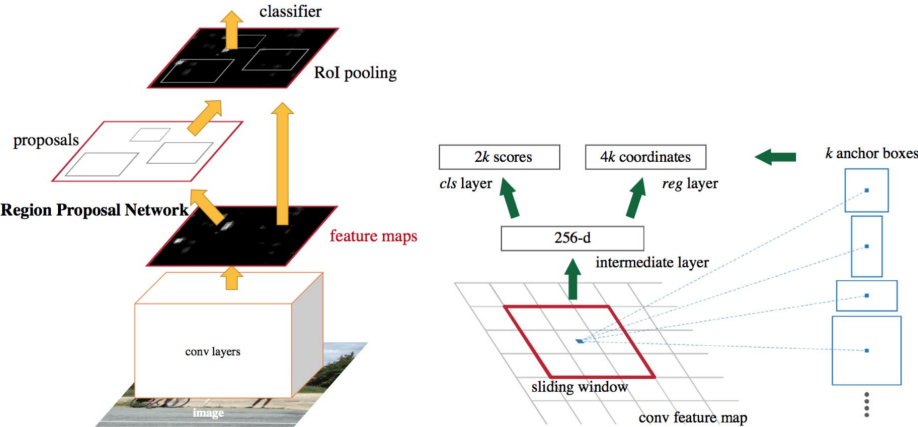
R-CNN



Fast R-CNN



Faster R-CNN



$$L = L_{cls} + L_{box} + L_{mask}$$

Cross Entropy

$$\text{CE}(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1 - p) & \text{otherwise.} \end{cases} \quad (1)$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise,} \end{cases} \quad (2)$$

$$\text{CE}(p, y) = \text{CE}(p_t) = -\log(p_t).$$

Balanced Cross Entropy

$$\text{CE}(p_t) = -\alpha_t \log(p_t). \quad (3)$$

Focal loss

$$\text{FL}(p_t) = -(1 - p_t)^\gamma \log(p_t). \quad (4)$$

$$\text{FL}(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t). \quad (5)$$

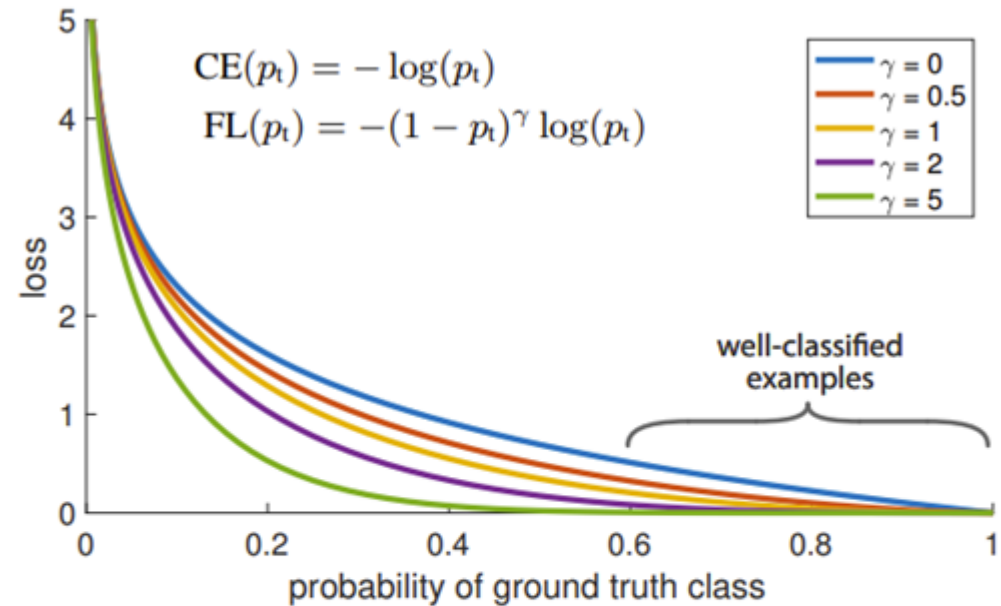


Figure 1. We propose a novel loss we term the *Focal Loss* that adds a factor $(1 - p_t)^\gamma$ to the standard cross entropy criterion. Setting $\gamma > 0$ reduces the relative loss for well-classified examples ($p_t > .5$), putting more focus on hard, misclassified examples. As our experiments will demonstrate, the proposed focal loss enables training highly accurate dense object detectors in the presence of vast numbers of easy background examples.