1. Introduction

Pneumonia is a form of acute respiratory infection that is most commonly caused by viruses or bacteria. It can cause mild to life-threatening illness in people of all ages, however it is the single largest infectious cause of death in children worldwide. Pneumonia alone accounted for 14% of all deaths of children under 5 years in 2019 as reported by World Health Organization (WHO)[1].

Pneumonia detection is commonly performed through the examination of chest X-Ray radiographs (CXR) by highly-trained specialists. It usually manifests as an area or areas of increased opacity on CXR, the diagnosis is further confirmed through clinical history, vital signs and laboratory exams. However, computer-aided diagnosis using artificial intelligence based solutions can be made available to a large population residing in remote locations at a minimal cost.

2. Dataset

2.1. Training

The RSNA pneumonia detection challenge dataset[2] is composed of a subset of 30,000 exams from the original 112,120 frontal-view X-ray images of 30,805 unique patients (train + test) from the NIH ChestX-ray14 dataset released by Wang et al., 2017[3] using their original labels which were derived from radiology reports and, therefore with the understanding that they were not always accurate. The 30,000 selected exams were comprised of 15,000 exams with pneumonia-like labels ('Pneumonia', 'Infiltration', and 'Consolidation'), a random selection of 7,500 exams with a 'No Findings' label, and another random selection of 7,500 exams without the pneumonia-like labels and without the 'No Findings' label.

Six board-certified radiologists annotated all 30,000 chest radiographs distributed evenly to determine whether lung opacities suspicious for pneumonia were present on the image with their corresponding bounding box to specify the location.

The lung opacity class indicated a finding on chest radiograph that in a patient with cough and fever has a high likelihood of being pneumonia.

In the cases labeled Not Normal/No Lung Opacity, no lung opacity refers to no opacity suspicious for pneumonia. Other non-pneumonia opacities may be present.

We train on the official stage 1 train split of 25684 images provided by the RSNA pneumonia detection challenge of which 5659 images are Pneumonia cases and rest not Pneumonia.

2.2. Test

We test our model on the stage 1 test set from the challenge comprising 1000 Chest X-rays due to the annotations being publicly available. Of these 1000 images 353 are Pneumonia cases and the rest not Pneumonia. Also to verify the robustness of our model and that it does not suffer from geographic variance; We test our model on the Validation Set of CheXpert dataset by Irvin et al., 2019[4] comprising of ChestX-rays from 200 unique patients manually annotated by a group of 3 radiologists.

Fig. 1. Class distribution of RSNA Pneumonia Challenge Stage-1 train dataset

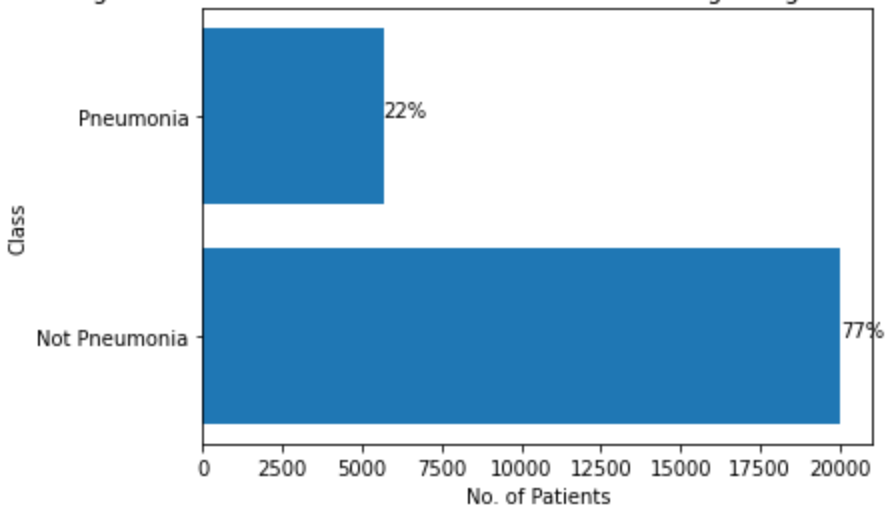


Fig. 2. Class distribution of RSNA Pneumonia Challenge Stage-1 test dataset

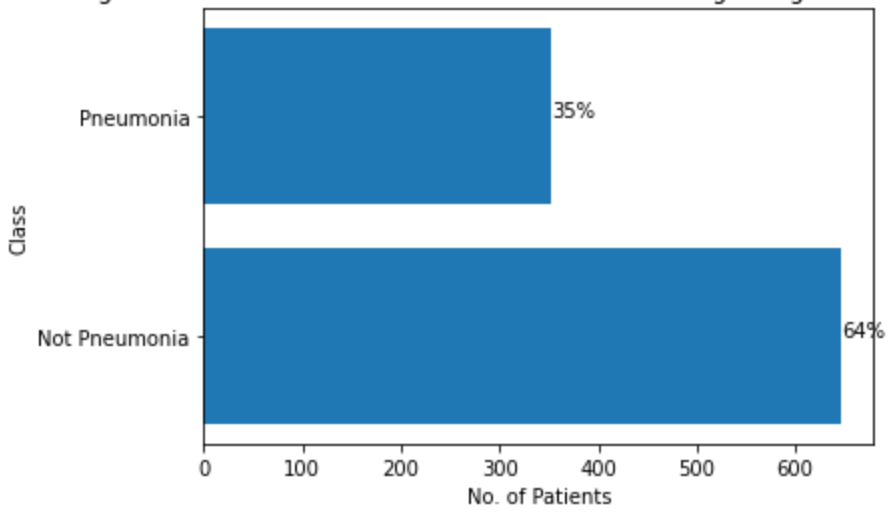


Fig. 3. Frequency of labels in the CheXpert validation set images

	Positive	%	Negative	%
No Finding	38	16.2	196	83.8
Fracture	0	0.0	234	100.0
Support Devices	107	45.7	127	54.3
Atelectasis	80	34.2	154	65.8
Cardiomegaly	68	29.1	166	70.9
Consolidation	33	14.1	201	85.9
Edema	45	19.2	189	80.8
Enlarged Card.	109	46.6	125	53.4
Lung Lesion	1	0.4	233	99.6
Lung Opacity	126	53.8	108	46.2
Pleural Effusion	67	28.6	167	71.4
Pleural Other	1	0.4	233	99.6
Pneumonia	8	3.4	226	96.6
Pneumothorax	8	3.4	226	96.6

3. Methodology

3.1. Problem Formulation

The pneumonia detection task is an object detection task where samples without bounding boxes are negative and contain no definitive evidence of pneumonia. Samples with bounding boxes indicate evidence of pneumonia. The input is a frontal-view chest Xray image and the output is zero or more bounding box annotations of the format

```
confidence x-min y-min width height
```

3.2. Model Architecture

We tested different architectures for consolidation detection and selected Retinanet by Lin et al., 2020[5] (one-stage) & DetectoRS by Qiao et al., 2021[6] (two-stage) detectors as our testing models for further analysis.

We use the MMDetection library by Chen et al., 2019[7] based on Pytorch by Paszke et al., 2019[8] to train our models.

3.2.1. Retinanet

One-stage detectors such as YOLO[9] and SSD[10] are fast and simple by using a fixed grid of boxes in substitute for region proposals but provide a relatively low accuracy of less than 10-40% over two-stage methods such as Mask R-CNN[11]. To address this problem caused by the foreground-background class imbalance, the one-stage RetinaNet suggests focal loss.

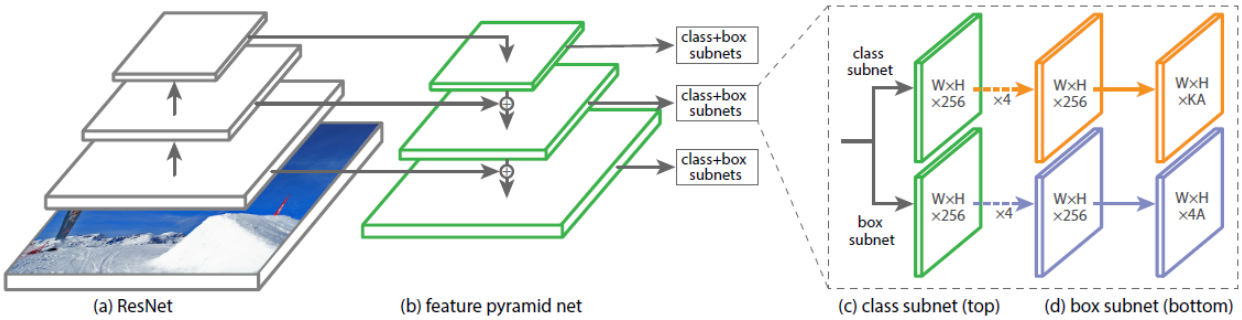
$$CE(p_t) = -\log(p_t)$$
$$FL(p_t) = -(1-p_t)^{\gamma}\log(p_t)(\gamma > 0)$$

This is a new loss function modified on standard cross entropy criterion. This function drastically reduced the scaling factor of cross entropy loss to almost zero as well-classified class increases. Therefore, it is able to concentrate on mis- classified examples during training and prevent learning from being overwhelmed by most negative samples.

As the RetinaNet network adopts Feature Pyramid Network (FPN) backbone[12] on top of the ResNet architecture[13], it generates multi-scale feature map layers with high resolution and rich

semantic information. Additionally, the model achieves dense coverage of boxes by using anchors of 3 scales and 3 aspect ratios at each pyramid level in FPN implying the higher potential that two-stage systems may not provide. Based on these factors, RetinaNet outperforms the accuracy of two-stage methods such as Mask R-CNN as a one-stage detector.

Fig. 4. RetinaNet Detector Architecture



We use 101 layers Residual Network (Resnet101) initialized with weights from a model pretrained on ImageNet [14] and Feature Pyramid Networks as backbone for feature extraction. We use the default presets for the network as described in the corresponding paper for detecting pneumonia.

We propose a novel training methodology that improves retinanet scores by 3-4%. The RSNA dataset only had annotations for pneumonia classes, so we had images and bounding box coordinates when pneumonia was present. A vanilla object detection model would learn the

features of pneumonia accurately but while predicting, it couldn't distinguish between conditions visually similar to pneumonia, so we have a lot of false positives.

To make the model discriminate between such false positive cases, we introduce a new class namely "similar to pneumonia but not it"(later referenced as "not pneumonia") and obtained bounding box annotations for this using the false positives. Then we retrain an object detection model with two classes: pneumonia and not pneumonia. Now, our model will predict overlapping boxes on the non pneumonia class but it will have higher confidence while predicting it as non pneumonia class so we can suppress these boxes using an empirical threshold.

3.2.2. DetectoRS

DetectoRS is a two-stage detector which includes a Recursive Feature Pyramid and Switchable Atrous Convolution. Recursive Feature Pyramid implements thinking twice at the macro level, where the outputs of FPN are brought back to each stage of the bottom-up backbone through feedback connections. Switchable Atrous Convolution instantiates looking twice at the micro-level, where the inputs are convolved with two different atrous rates.

Fig. 5. Two-stage structure of Recursive Feature Pyramid

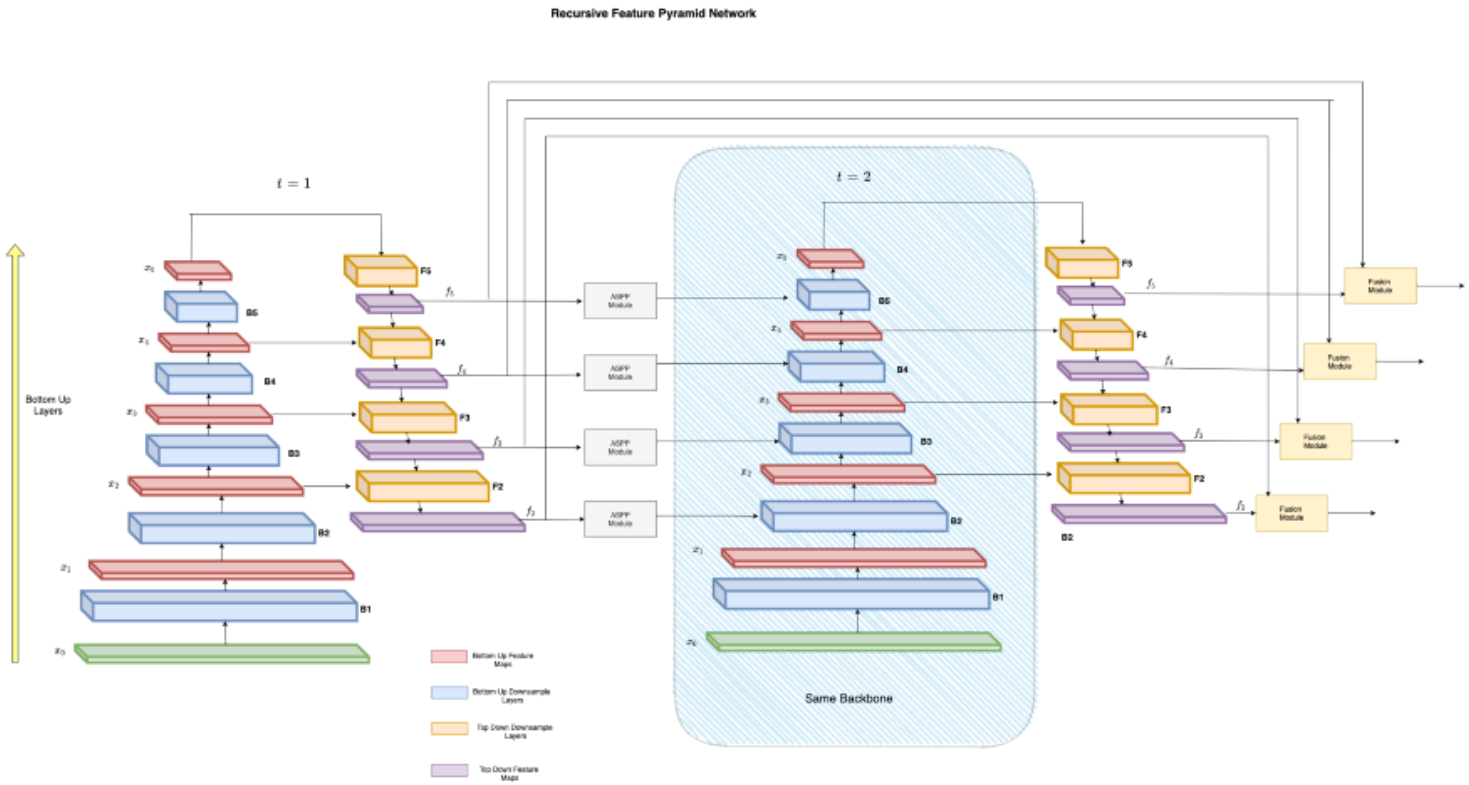
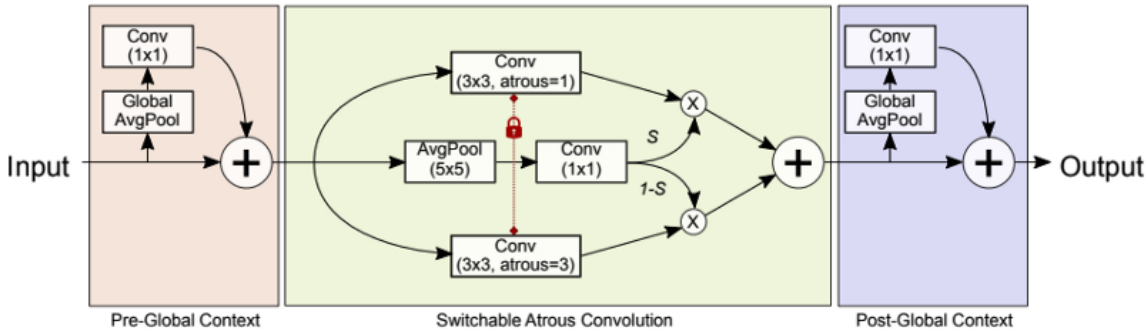


Fig. 6. Breaking of standard 3x3 convolution operation into switchable atrous convolutions of rates r=1 and r=3, respectively. Weights are shared between two convolutions(Locking mechanism)



We use Cascade RCNN[15] based DetectoRS model with 101 layers Residual Network (Resnet101) initialized with weights from a model pretrained on ImageNet and FPN as backbone for feature extraction. We use the default presets for the network as described in the corresponding paper.

3.3. Images preprocessing, augmentations and post-processing

Images were augmented using albumnetations library[16] with rotation, translation, scaling, HSV, RGB Shift, Random Brightness and contrast change, and horizontal flipping (shearing and vertical flipping were turned off).

The Retinanet model was trained with images resized to 512 X 512 px, while for the detectors model we utilized multiscale training the scale of the short edge was randomly sampled

from [512, 672], and the scale of the long edge is fixed as 672px. The images were normalized according to imagenet stats before feeding into the network.

We aggressively used Non max suppression to eliminate overlapping bounding boxes for both the models. For DetectoRS we use MultiScaleFlipAug as Test time augmentation to improve the results.

3.4. Training

We use SGD with momentum[17](0.9) and weight decay(1e-4) for training our models. We use a batch size of 8 for Retinanet and 2 for DetectoRS and an initial learning rate of 0.01 and 0.0025 respectively using the linear scaling rule[18]. We use linear warmup for the first 500 iterations with a warmup ratio of 1e-3. And a step-down policy after the third, sixth, and ninth epoch i.e the learning rate is decreased by a factor of 10 at each of the mentioned epochs. The code for training and inference can be found here.

4. Results

We evaluate for area under the receiver operating characteristic curve(**AUC**) for the RSNA test set and CheXpert Validation set. A receiver operating characteristic(**ROC**) curve is a graph showing the performance of a classification model at all classification thresholds. An ROC curve plots True Positive Rate(**TPR**) vs. False Positive Rate(**FPR**) at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. AUC provides an aggregate measure of performance across all possible classification thresholds. We extract image level labels from bounding box outputs for a

Chest X-ray Image as the maximum of Confidence score for all of the boxes predicted. For comparison of performance on the CheXpert dataset we train a densenet-121[16] baseline model as described in the CheXpert paper. We also evaluate an ensemble of Retinanet and DetectoRS models, since the output distribution of models are different we use ranking before ensembling the scores. The models score .90+ AUC on the RSNA stage 1 test data set, and also perform surprisingly well on the CheXpert Validation set, even outperforming the baseline model trained on the CheXpert train set. The detectoRS model is able to maintain the AUC score over datasets which is a remarkable result as models generally perform worse on external dataset.

- **AUC** provides an aggregate measure of performance across all possible classification thresholds.

Table 1. Results AUC score for all models on test sets

Model	RSNA stage 1 test set AUC	CheXpert Val set AUC
Retinanet	0.922	0.867
Detectors	0.920	0.921
Ensemble of Retinanet and DetectoRS	0.925	0.903
CheXpert-Baseline	0.849	0.778

- **Sensitivity** (true positive rate) refers to the probability of a positive test, conditioned on truly being positive.
- **Specificity** (true negative rate) refers to the probability of a negative test, conditioned on truly being negative.

Table 2. Specificities and Sensitivities for DetectoRS model on RSNA test set

threshold	sensitivity	specificity
0.05	0.980170	0.616692
0.1	0.923513	0.751159
0.15	0.849858	0.840804
0.20	0.770538	0.884080
0.25	0.696884	0.913447

Table 3. Specificities and Sensitivities for Retinanet model on RSNA test set

threshold	sensitivity	specificity
0.05	0.997167	0.075734
0.1	0.985836	0.531685
0.15	0.923513	0.743431
0.20	0.807365	0.892235
0.25	0.623229	0.821477

An **ROC** curve plots **TPR** vs. **FPR** at different classification thresholds. Lowering the classification threshold classifies more items as positive, thus increasing both False Positives and True Positives. Below figures are ROC curves for the DetectoRS and Retinanet model on the RSNA stage 1 test set respectively.

Fig. 7. ROC curve for DetectoRS model on RSNA stage 1 test set

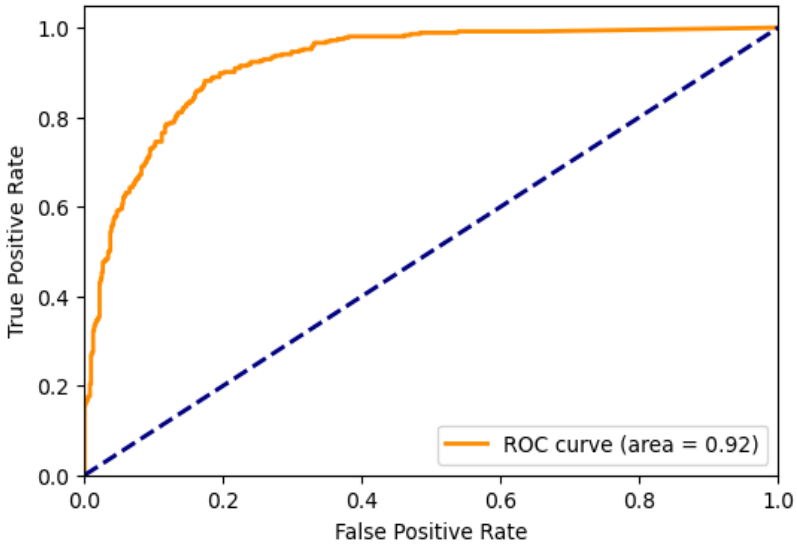
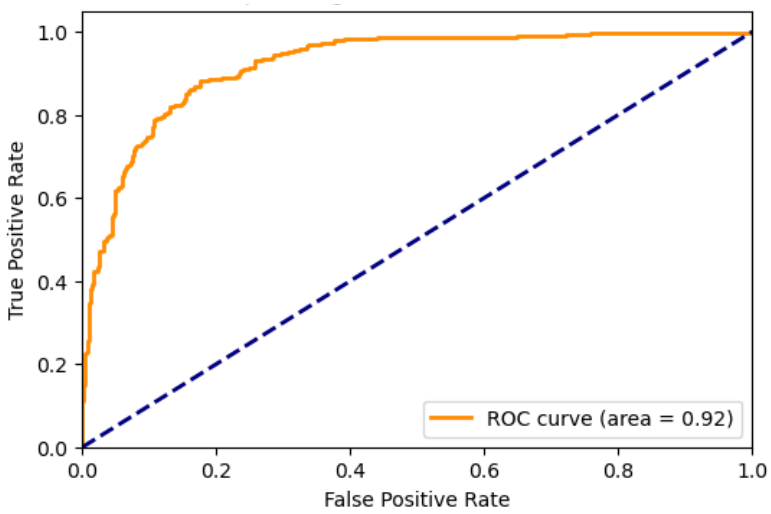


Fig. 8. ROC curve for Retinanet model on RSNA stage 1 test set



5. Future Scope

Expert radiologists' presence is the topmost necessity to properly diagnose thoracic diseases. The main aim of this paper is to improve the medical adeptness in remote areas where the radiologists availability is limited. This study can facilitate the early diagnosis of Pneumonia or more so reduce the workload on the available radiologists by having them inspect scans which are likely to be pneumonia according to our model. In this study we work on the RSNA dataset on a per image basis contrary to all the research that works on a bounding box basis. We also propose a novel approach to distinguish pneumonia and pneumonia-like occurrences using multistage training. This method seems to work well for single shot detectors, but not so much in the case of multistage detectors. In the future we can research for the exact details leading to this phenomenon.

6. References

[1] *Pneumonia*. (2021, November 11).

https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] Shih, G., Wu, C. C., Halabi, S. S., Kohli, M. D., Prevedello, L. M., Cook, T. S., ... & Stein, A. (2019). Augmenting the national institutes of health chest radiograph dataset with expert annotations of possible pneumonia. *Radiology. Artificial intelligence*, *1*(1).

[3] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).

[4] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., ... & Ng, A. Y. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 590-597).

[5] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).

[6] Qiao, S., Chen, L. C., & Yuille, A. (2021). Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10213-10224).

[7] Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., ... & Lin, D. (2019). MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*.

[8] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

[9] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[10] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

[11] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[12] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[14] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee.

[15] Cai, Z., & Vasconcelos, N. (2018). Cascade r-cnn: Delving into high quality object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6154-6162).

[16] Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: fast and flexible image augmentations. *Information*, *11*(2), 125.

[17] Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013, May). On the importance of initialization and momentum in deep learning. In *International conference on machine learning* (pp. 1139-1147). PMLR.

[18] Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., ... & He, K. (2017). Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*.