Object Detection

Lin Zhu(s232291), Cong Jin(s232254), Dong Yun(s232293), Yu Zhang(s230000)

Problem description

1. Develop a CNN-based **object detection system** using object proposals on the **Potholes dataset**, addressing class imbalance, and evaluate its classification accuracy on the validation set.
2. Apply the trained detector to test images, implement **Non-Maximum Suppression** to refine detections, and evaluate the object detection performance using the **Average Precision (AP)** metric.

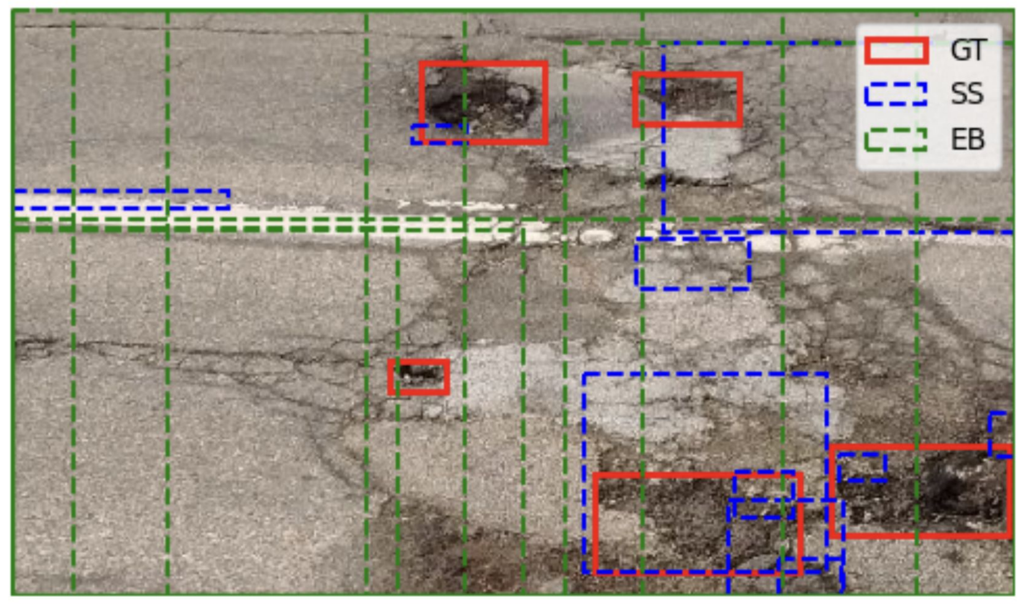
Abstract

Dataset: Potholes dataset - a fully annotated image dataset is utilized, comprising 665 images with corresponding annotations in Pascal VOC XML format.
TRAIN : VAL : TEST = 425 : 107 : 133 = 64% : 16% : 20%

Region Proposals: Selective Search(optimal proposals 551) & Edge boxes(optimal 1775)
Object detector: **R-CNN** & **R-CNN(with box regression)** & **Fast R-CNN**
Result: We found that when using **NMS** to filter proposals, setting the IoU threshold to 0.1 resulted in the best Average Precision performance. Fast R-CNN significantly outperforms R-CNN, offering more precise localization and much higher computational efficiency. However, in terms of classification accuracy, the performance is similar.

Region proposals methods

1. We evaluated two object proposal methods: **Selective Search** and **Edge Boxes**.
2. Three key metrics are used to assess their performance: **Recall / MABO / Image Coverage.**
3. Besides, we used an efficiency metric (**recall/log(proposals)**) to balance performance and computational cost to determine the **optimal** number of proposals.

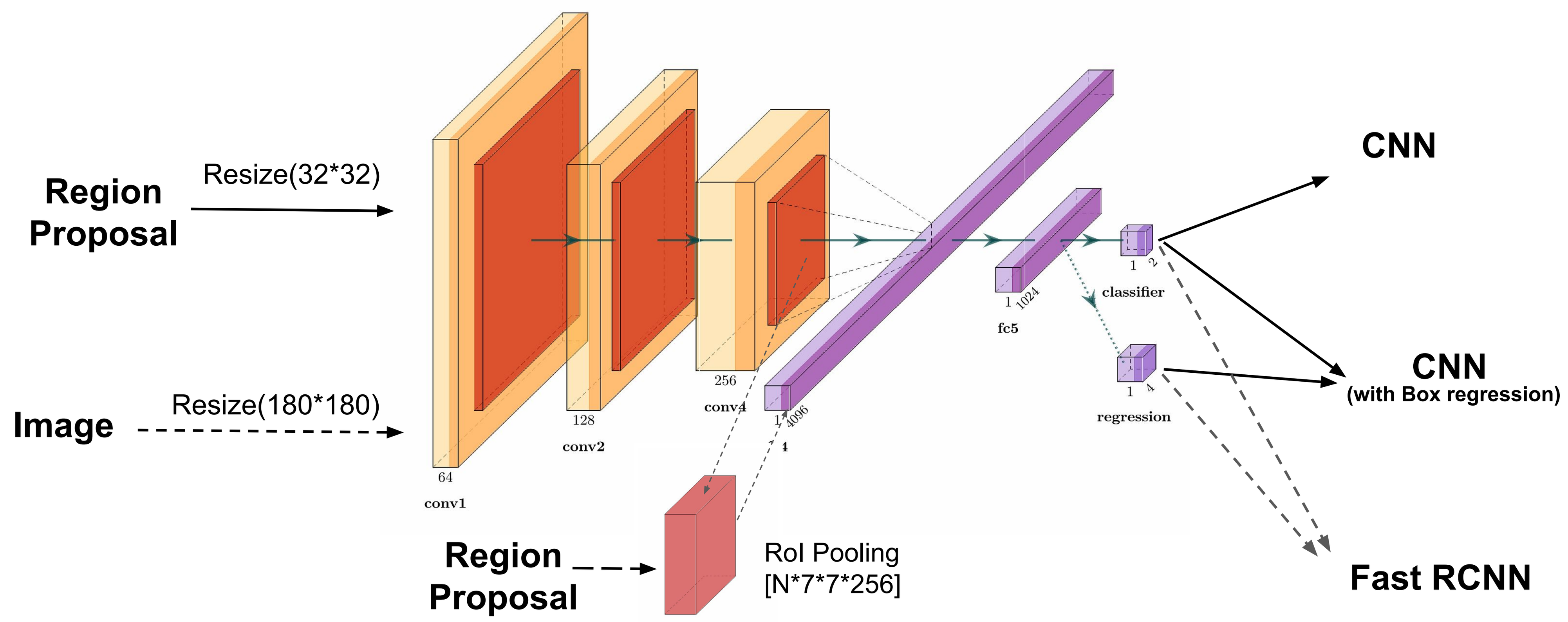


Proposals using SS and EB

Table1: Performance between SS and Edge Boxes

Metrics	SS Fast	Edge Boxes
Optimal proposals	**551**	1775
Recall	**0.863**	0.589
MABO	**0.687**	0.538
Image Coverage	**0.856**	0.499

Architecture



Object detector - Train & val

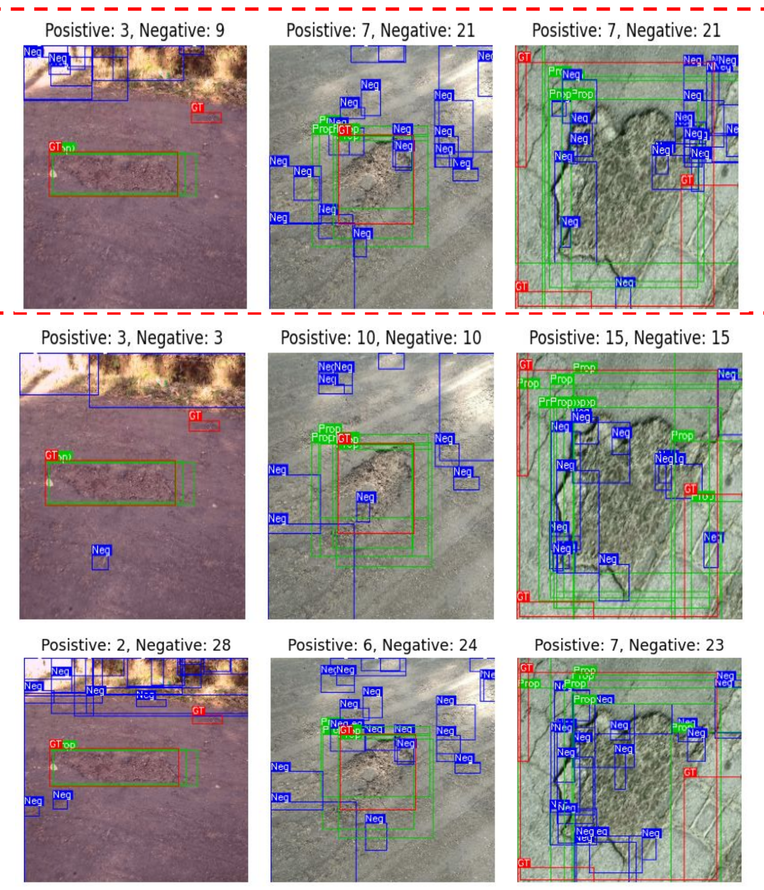
To address the class imbalance issue of the background proposals,

we reduced the number of background proposals, limiting their proportion to only **75%.**

25%:75%

50%:50%

Random



All proposals

CNN_pred

CNN(with box regression)

Fast RCNN

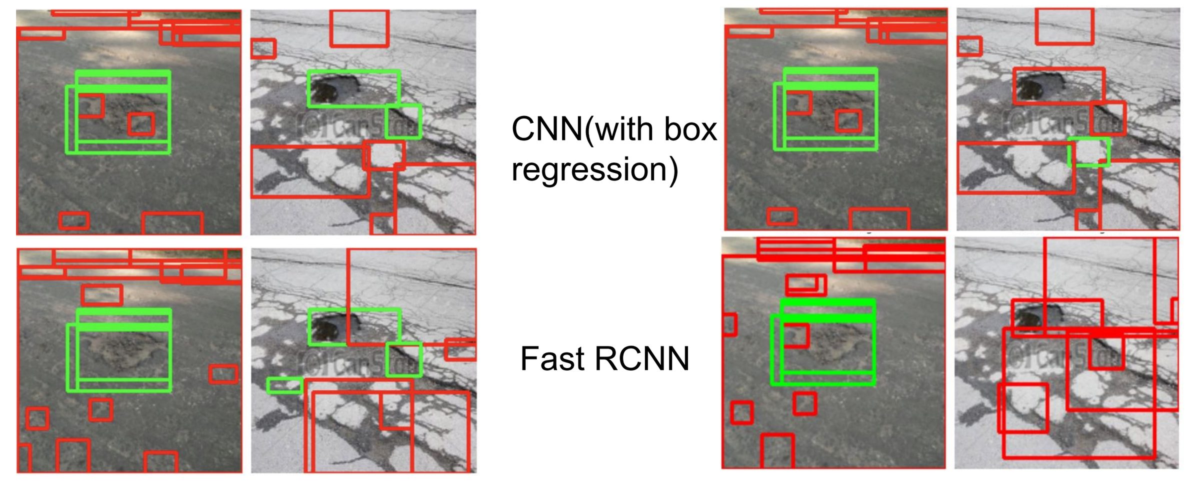
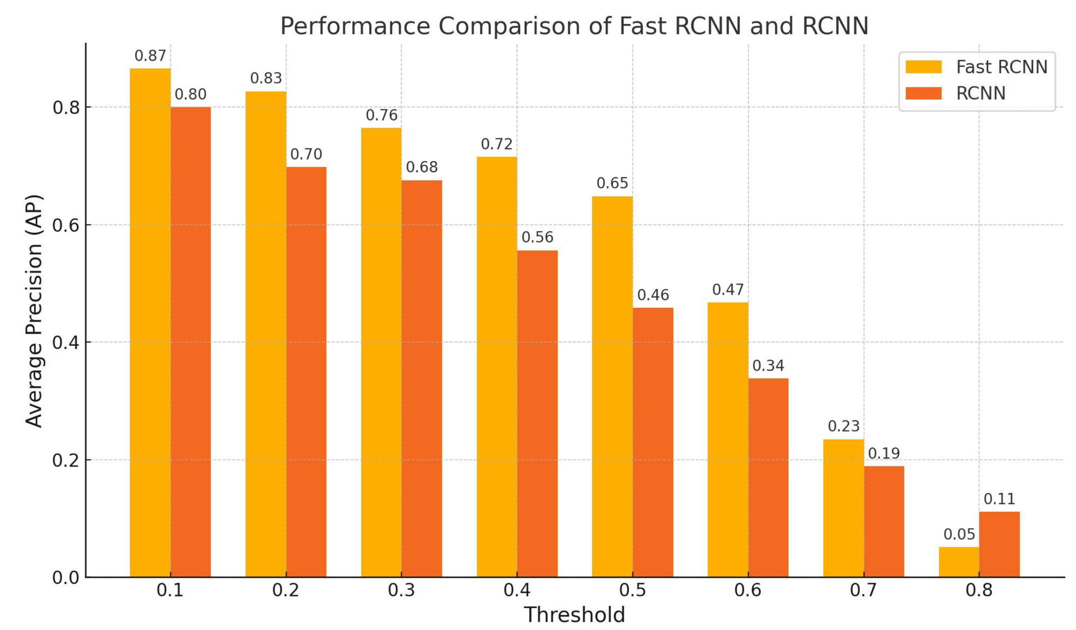


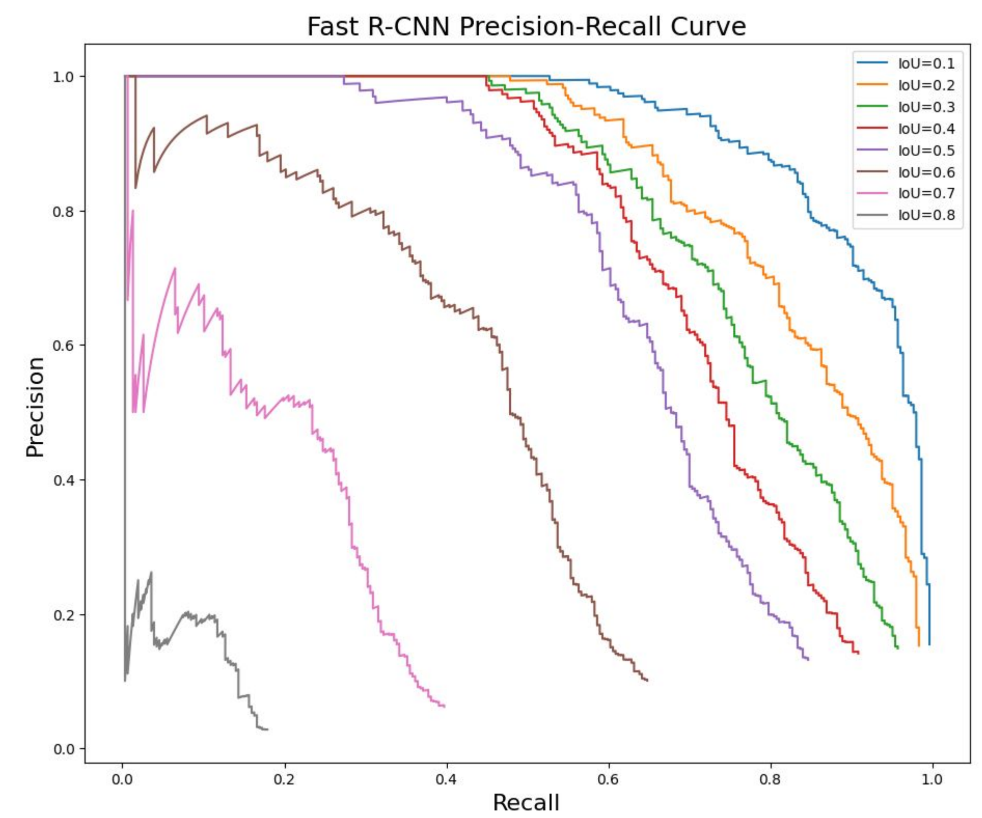
Table2: Classification performance of each model

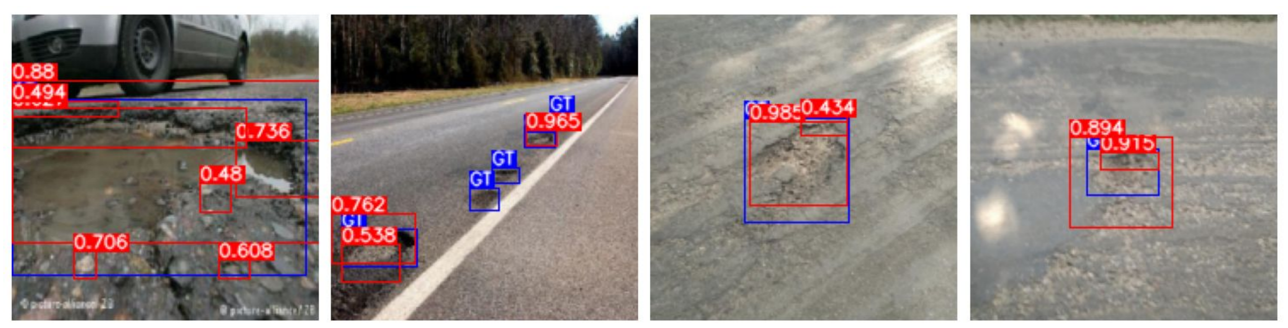
	RCNN	RCNN (with box regression)	Fast RCNN	Fast RCNN (with VGG16)
Accuracy	87.9%	89.4%	87.5%	95.11%
Recall(positive proposals)	71.0%	63.8%	62.4%	62.9%
Precision(positive proposals)	79.5%	81.1%	83.5%	84.7%
F1_scole	75.0%	71.4%	71.5%	72.2%

Object detector - Test

When using the **NMS**, we find that the lower the IoU threshold, the higher the Average Precision (below bar chart). This might be due to the low quality and sparsity of the generated proposals.
It is evident that Fast RCNN demonstrates superior accuracy and quality in object detection. We measured the computation time for both on an A100 GPU. Fast RCNN takes an average of 0.015 seconds per image, while RCNN is approximately 20 times slower, taking about 0.283 seconds per image.







Result from R-CNN

Result from Fast R-CNN

References

[1] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv preprint* arXiv:1311.2524.
[2] Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T., & Smeulders, A. W. M. (2013). Selective search for object recognition. *arXiv preprint* arXiv:1305.0799.
[3] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR).
[4] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *arXiv preprint* arXiv:1506.01497.
[5] Zitnick, C. L., & Dollár, P. (2014). Edge Boxes: Locating object proposals from edges. European Conference on Computer Vision (ECCV), 391–405.