Smart Roads:
Real-Time Pothole Detection

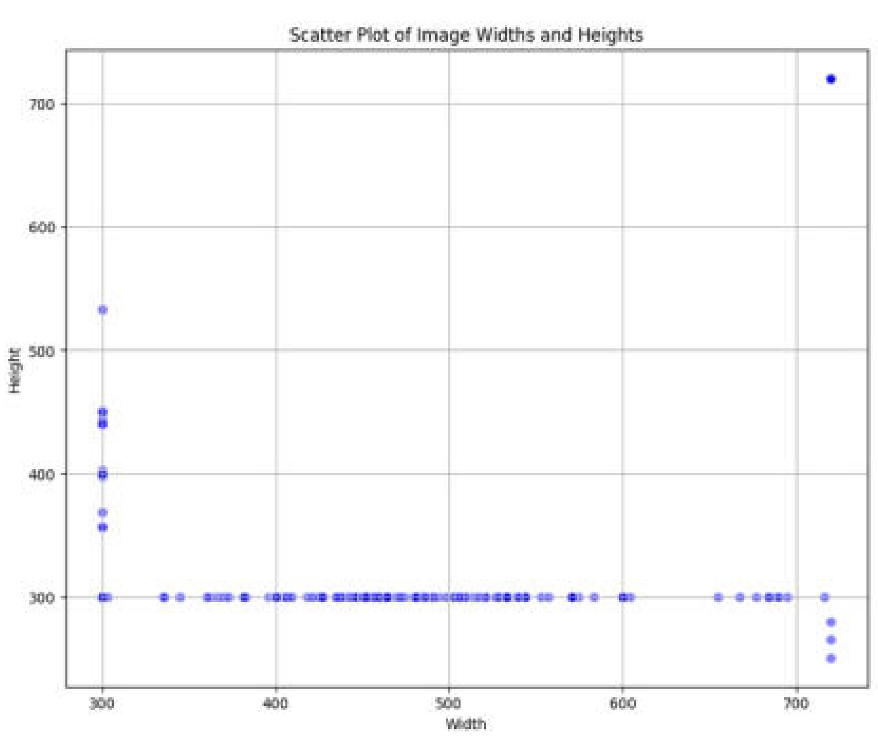
Samer Bujana, Rasmus Clausen, Alba Gonzalo, Robert Jarvi, Leonardo Rodovero

INTRODUCTION

Potholes pose a serious threat to road safety, increasing the risk of accidents and causing significant damage to vehicles. In this project, we aim to develop an advanced deep-learning object detection system capable of automatically identifying potholes in images captured from real-world environments. To achieve this, we will leverage a Region-based Convolutional Neural Network (RCNN) architecture, utilizing Selective Search (SS) and Edge Boxes as region proposal methods. This innovative approach addresses the challenges of detecting and mitigating potholes, ultimately contributing to safer and more reliable road conditions.

ENCODER-DECODER PROPOSAL EXTRACTION

When analysing the resolutions of the images we found a big difference between individual images, With the largest images being 720x720, while the smallest are around 300 in width or lower. This means that rescaling the pictures to the same resolution breaks a lot of details. We choose instead to not rescale the images before running Selective Search.



Threshold used to determine a match for recall and ABO calculation: k = 0.5.

$$Recall = \frac{NumberofTruePositives}{TotalNumberofGroundTruthObjects}$$

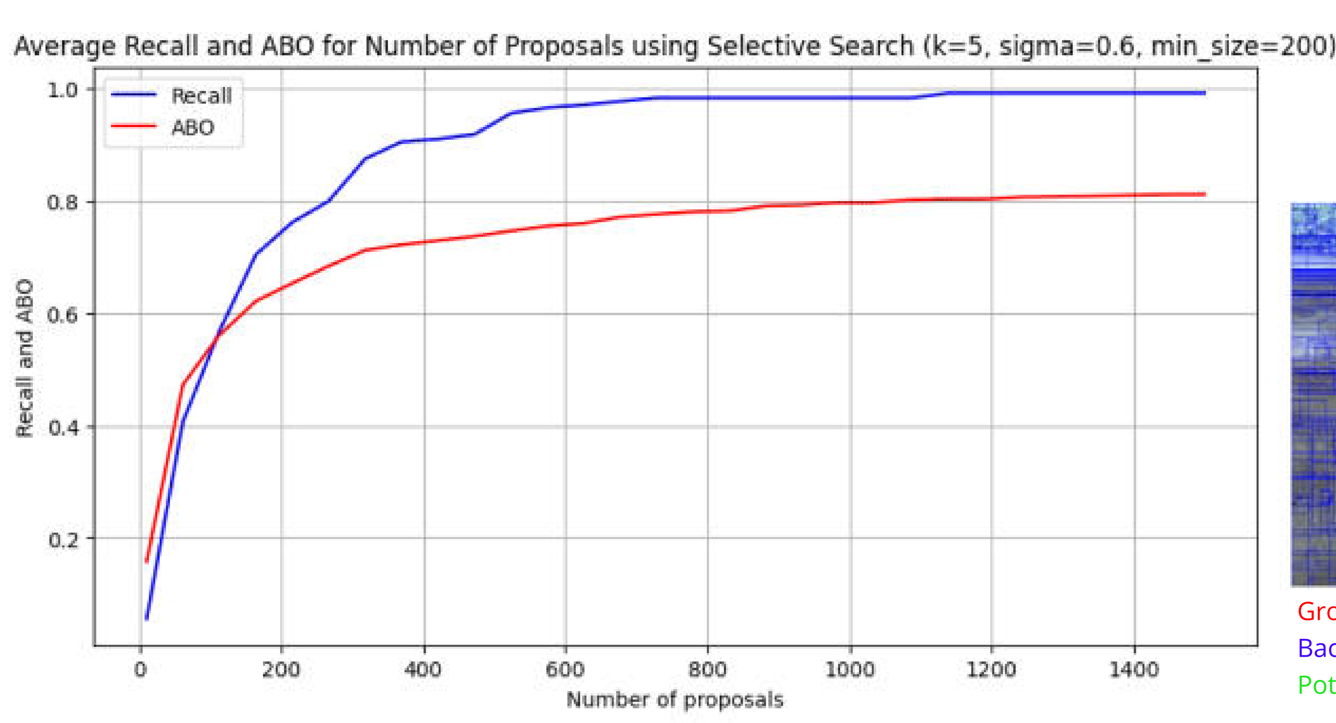
$$IoU = \frac{AreaofIntersection}{AreaofUnion}$$

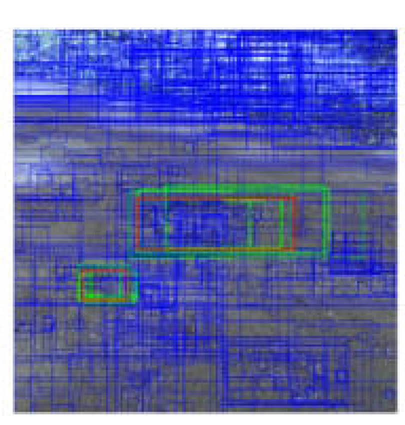
$$ABO = \frac{1}{N}\sum_{g_i{}^c \in G^c} max_{l_j \in L}\left(IoU\left(g_i^c, l_j\right)\right)$$

SELECTIVE SEARCH

Used the fast mode of SS, and tune-up the proposals by changing the different parameters:

- **MinSize:** Minimum size of segments.
- **Scale (k):** Controls the level of segmentation for selective search.
- **Sigma:** Smoothness parameter for the selective search segmentation.



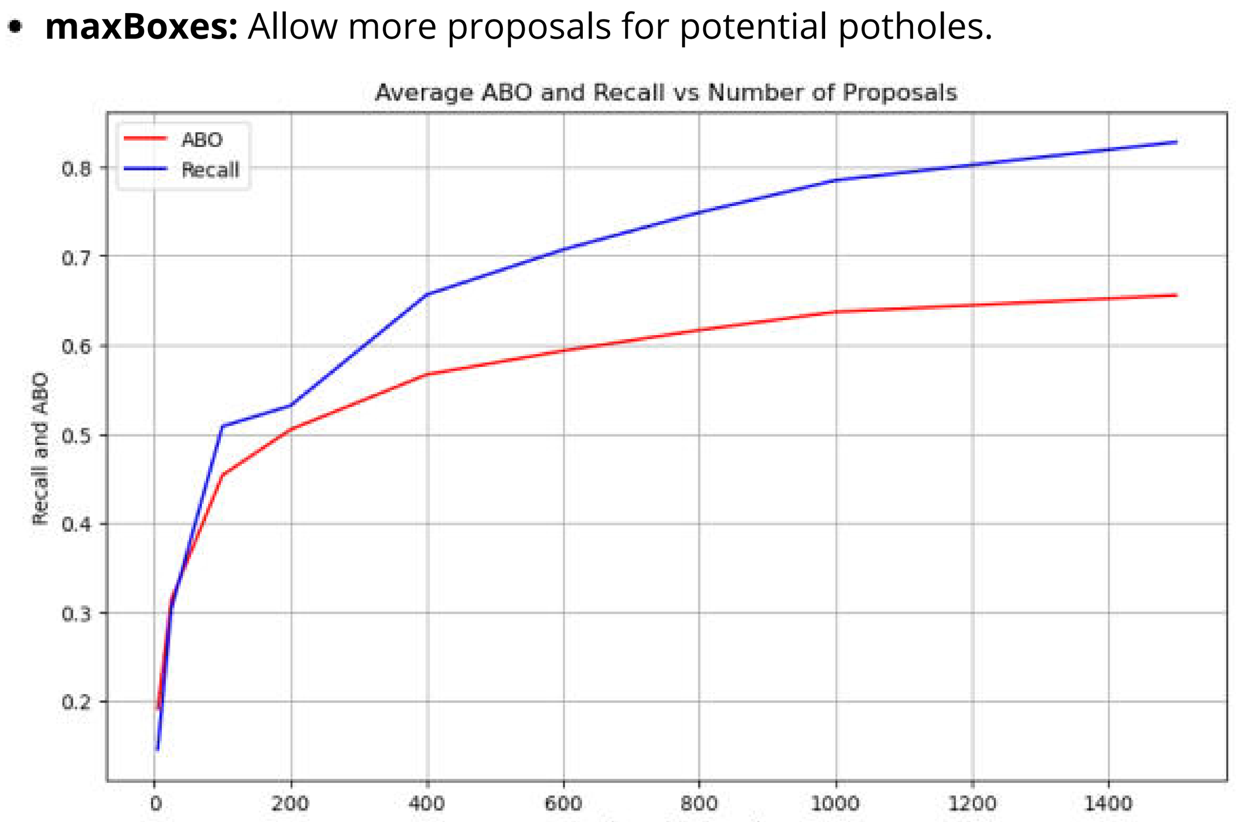


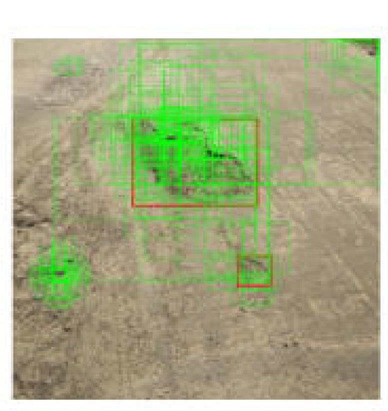
Ground truth
Background
Pothole

EDGE BOXES

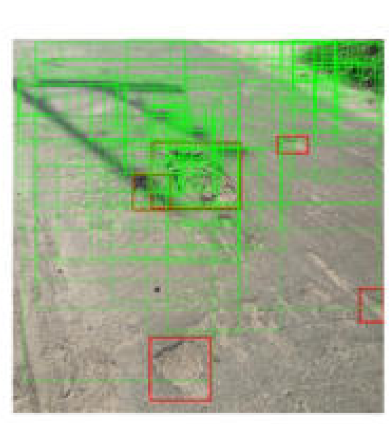
Tune-up the proposals by changing the different parameters:

- **Alpha:** Finer granularity in sliding window. Set to 0.8.
- **Beta:** Increased sensitivity to edge density. Set to 0.6.
- **minScore:** Minimum score for boxes. Set to 0.05.
- **maxBoxes:** Allow more proposals for potential potholes.





Ground truth
Proposal



Ground truth
Proposal

Better proposals were obtained using SS. The maximum Recall and ABO are achieved with **1200** proposals, beyond which increasing the number of proposals appears to yield no significant improvement.

DATA AND CLASS IMBALANCE

Data attributes:

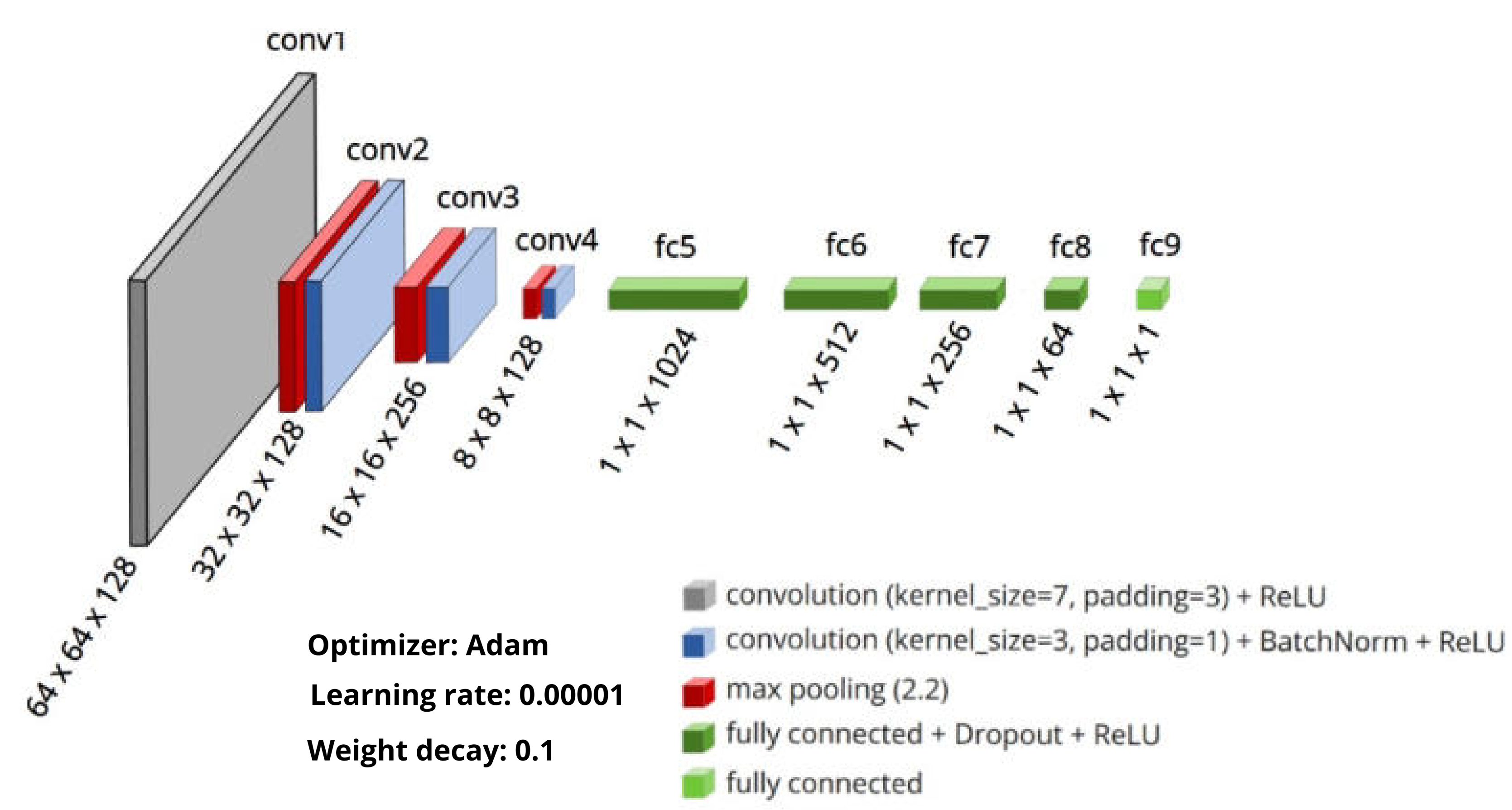
- 665 images, not already split .
- Image and grounded boxes provided.
- Original image resolution 720×720. The different proposals are resized to 64x64.
- 70-15-15 Train / Validation / Test split.

Data transformations (Data Augmentation):

- RandomGrayScale
  - p: 0.1
- ColorJitter
  - Brightness: 0.2, Contrast: 0.2
  - Saturation: 0.2, Hue: 0.1
- Normalization
  - mean: (0.485, 0.456, 0.406)
  - std: (0.229, 0.224, 0.225)

Class imbalance was addressed through batch sampling, ensuring that each batch contains a sufficient number of foreground objects, with a 50-50 split. Instead of using all background proposals for training, random background proposals are selected for each batch.

CNN ARCHITECTURE



Optimizer: Adam
Learning rate: 0.00001
Weight decay: 0.1

CLASSIFICATION ACCURACY

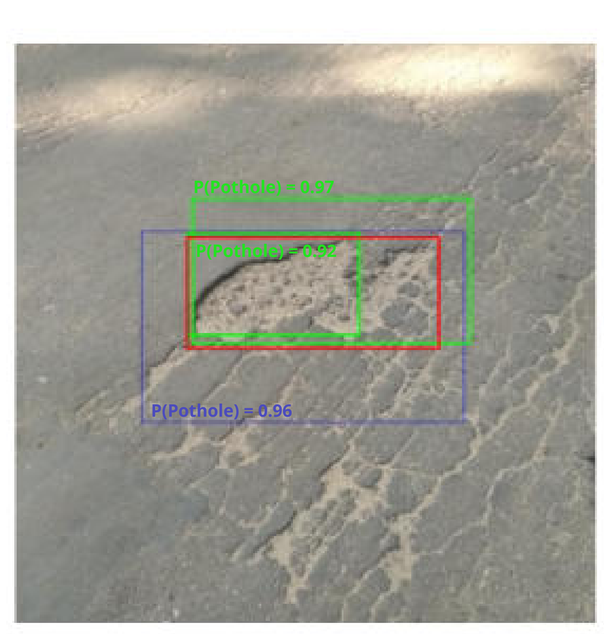
Average Validation Accuracy 86.9 %: Indicates that the model correctly predicts the presence or absence of potholes for most proposals.

POTHOLE DETECTION EVALUATION

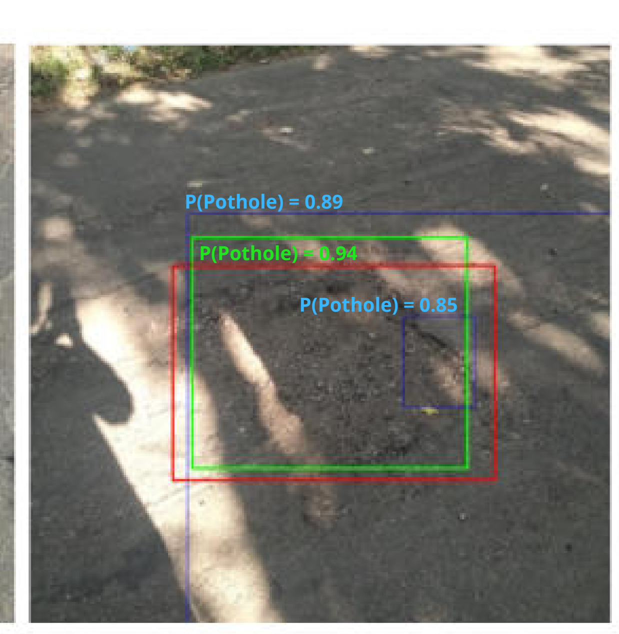
TP 399879	**FN** 99851
FP 2553	**TN** 10107

After evaluating the CNN on the test set (composed of **512,390** proposals), the following confusion matrix is presented to show how the proposals are being evaluated for each image in the test set.

Non-Maximum Suppression (NMS) is implemented to discard overlapping bounding boxes, keeping only the most relevant ones. The model's performance is then evaluated using the Average Precision (AP) metric, which provides a comprehensive measure of detection accuracy.



Ground truth
Background
Pothole



Ground truth
Background
Pothole

Average Precision AP = 0.4078: show room for improvement in distinguishing true positives from false positives.

References and use of AI:
.Uijlings et al, “Selective Search for Object Recognition”, IJCV 2013.
Use of AI: Aided in debugging and code generation, explaination of topics.