ICTgroup (ICT.eu): Detecting and classifying damage to traffic signs from images

Participants: Jeroen Delcour (company representative), Bernardo Marques (company representative), Efstratios Gavves (academic leader), Tom Viering, ...

Context and problem

At 139,294 km long, the Netherlands has one of the world's densest road networks. Maintaining such a large network of not only roads but also surrounding infrastructure such as traffic signs can be very time-consuming and expensive if not done efficiently. In the case of traffic signs, video along with the camera's position in the world is recorded from a car. Trained inspectors look through these videos, annotating any damage to the traffic signs they see. Needless to say, this is a very arduous task. Additionally, there is a constant shortage of trained inspectors.

The goal is to automate the detection of damaged traffic signs from video. It is a challenging case since it involves a wide range of possible damages. The signs could be poorly legible: e.g. obscured by vegetation, covered in stickers, or faded over time. Additionally, some types of damage aren't visible on the sign itself, but involve the angle of the sign relative to the road it's associated with (i.e. it should face the road) or even the angle of its supporting pole (such as after a collision with a car). Perhaps most challenging to detect is deformation of the sign, often as a result of fireworks.

Figure 1: Examples of damaged traffic signs.

We believe reliably detecting this wide range of damage in a real-world scenario to be a challenge involving a number of sub-disciplines. Most notable are computer vision and machine learning, but we are open to creative approaches from other fields. As such, the lessons learnt from this challenge may be beneficial to an equally large number of sub-disciplines and could provide inspiration for further research into solving several computer vision tasks, such as classification problems where classes are very similar or their differences are poorly defined, and orientation estimation both in 2D and 3D.

Research approach

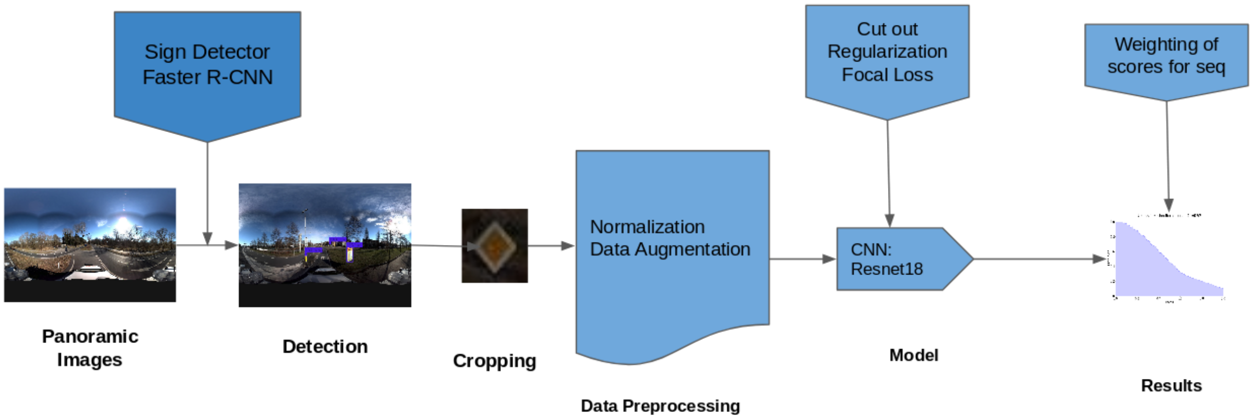

Fig 2: Pipeline for the Method

In this work we proposed a method for traffic sign detection and their classification into damaged and undamaged categories using panoramic images, complete pipe line for the proposed method is shown in Fig 1.

1. For this purpose at first we performed traffic sign detection using Faster-RCNN model. Faster-RCNN is a model for object detection in images. In this work we used this model and detected traffic signs.
2. In the second step we used these traffic sign detections and cropped them from panoramic images.
3. In the third step we used following data preprocessing for training the classification network
    a. Data Normalization
    b. Data Augmentation using synthetic images
    c. Data serialization
4. After data preprocessing we trained Resnet18 for binary classification into damaged and undamaged classes. Resnet is one of the state of the art convolutional neural network with residual learning framework for the easy training of the deep convolutional neural networks.
5. As the given BAM traffic sign dataset was imbalanced i.e. there were a lot of images of undamaged traffic signs but less images of damaged signs therefore, we used some techniques from computer vision literature to tackle this problem one of which was to use Focal loss.
    a. **Focal loss** is the technique which modifies the cross entropy loss in such a way that the correctly/well classified samples are down weighted.
    b. Another Technique which we used to avoid overfitting and balancing our dataset was **Cut out** regularization. In this regularization technique some parts of the input images are masked randomly while training the network.
6. Finally, we performed the evaluation of our results using Mean average precision criterion and some visualizations. We used K-fold cross validation while calculating our mean average precision in order to avoid bias in our model.

Results

Detection of the traffic signs turned out to be a relatively easy problem solved with Faster R-CNN.

Because of the great imbalance problem of our data we decided to use Mean Average Precision as our main performance metric. We carefully decided on the best model, which should be able to counter heavy overfitting *and* still be capable of working with the class imbalance. In this week, a ResNet18 with Focal Loss regularized with Cutout worked best.

Training and testing on the German GTSRB dataset turned out to be a relatively easy problem due to the cleanliness of the data. Unfortunately, the model did not generalize to the Dutch data. Even training on a mix of Dutch and German data resulted in weak performance.

At the time of writing, only training and testing on the Dutch BAM dataset yielded the best mAP: 0.52. However, as the observations come in sequences we can combine the predictions resulting in quite an improved mAP of 0.62. The 5-fold cross validation precision recall curve is shown in Fig 2.

Experiments with training on German to extract the most important features in damage classification and fine-tuning (perhaps by freezing some low-level layers) these on Dutch data are still running.

Overall, the model is great at detecting true negatives but returns relatively many false negatives. This also has to do with many noisy examples in the data.

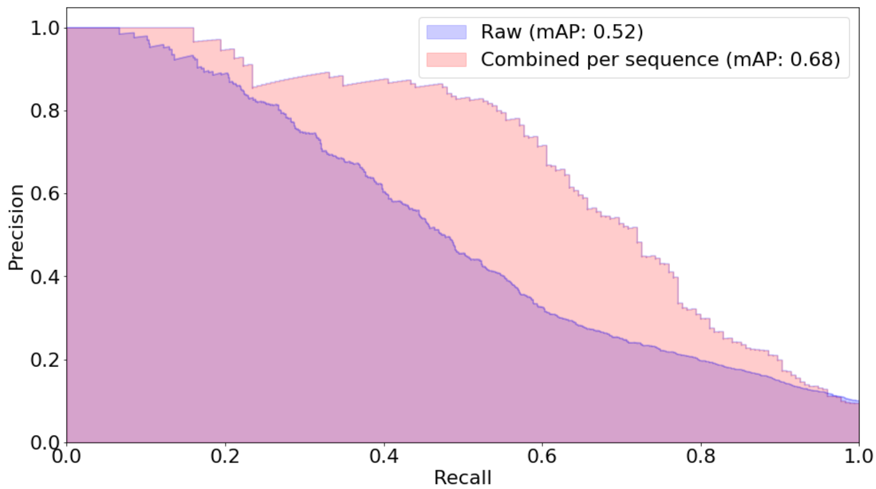

Fig 3: 5-fold CV precision-recall curve

Future work

As we obtained our best results by using the predictions on a whole sequence of images of a single traffic sign, using a model that can exploit this sequentiality (RNN) might result in better results. Also, BAM should focus on obtaining images of the highest quality possible to ensure the damage is detectable. Countering the class imbalance by generating artificial damage seems promising.