

# Technical Report

## Backbone

The backbone of a regular YoloV4 network is CSPDarkNet-53. It has five downsampling (pooling) layers, meaning the stride would be 32. If we have our input as approx. 200x100, there will be not enough feature information (low resolution) in the last layer. A potential solution should be delete the last downsampling layer.

## Global Non Maximum Suppression

In addition to ordinary NMS, we found that a digit may be recognized as several different numbers and for a two-digit street number, there is possibility to be detected as three-digit one. In this case, a [global non maximum suppression](#) method was applied. But before implementation, think of a scenario where there is noise on the image. The noise is visually separable but it may look like a digit and may be detected as a digit. Since it often happens at the border of the image, we implemented a algorithm which looks like below:

---

### Digit Detection Enhancement

---

if minus representation in coordinate: mark it as a warning and record its area and index

if warning in detections: calculate other detections' average area and compare to warning

if warning area is too small: recognize it as noise and remove it from detections

---

## Voting Discriminator

For this challenge, the outlook of the each image is a sequential numbers from left to right. To make full use of detection results from multiple models (YoloV4-tiny, YoloV4-tiny with mosaic, cosine scheduler and label smooth (I don't know why this could be worse), CRNN (I think I can try auto-augmentation instead of regular augmentations?)), we can design a [naive voting algorithm](#). We denote three results as  $r_1$ ,  $r_2$  and  $r_3$ . The algorithm is as follows:

---

#### Voting algorithm

---

if  $r_1$  and  $r_2$ : choose the longer string if both lengths not equal, otherwise randomly pick

elif not  $r_1$  or  $r_2$ : choose the one with valid prediction

else: choose  $r_3$

if  $r_3.\text{length} > \text{cur.length}$ : if it's just a shift (like 1159 and 159), ignore the inconsistency,  
otherwise, follow the longer one

if  $r_3 \cap \text{cur}$  is empty, concatenate them

---