Yoav Raytsfeld
Danel Kornis

Face mask detection project

Training Background

Problem definition: The problem is object identification in an image. In this type of problem we are required to perform two tasks from the world of computer vision, both a classification problem and a localization problem. As part of this project we were required to identify whether a person is wearing a face mask. We had about 200 pictures at our disposal and after examining them we discovered that 187 pictures contained at least one mask and the rest (13) only contained people without a mask. To tag the masks we used the tool coco-annotator. It is a simple and efficient tool that eventually produces a Json file with the mask locations in the COCO format image.

The challenge is to identify the mask object in the most comprehensive way which cannot be defined simply. There is a rich and wide variety that face masks can appear in, and the different locations and angles do not allow us to solve the problem as a simple classification problem and in order to be able to identify these, one can study the properties of the mask, this can be achieved by Neuron network training.

How we decided on the types of labels:

We want to solve these problems with the help of the model built, but in order to proceed to this stage in the project we must decide on the possible types of labeling. Apparently there are 3 categories for data labeling:

0. Background
1. Without mask
2. Wearing a mask

But the very addition of the first label - "without a mask" would obliges us to label all the entities in the image even if they are distant and blurry. So to avoid impairing the learning of the model (for example if I do not mark the person and the model does recognize the penalty will be large). We marked only the visible masks we can identify, as long as they are not too low resolution or too blurred.

How we tagged BBOX:

At this point we added another working premise, sometimes the mask hides most of the face and the exposed part is hidden by another object (hair, sunglasses, hat, hands), this was tolerable as long as the mask and some of the face was recognizable, but creating too wide of a BBOX may force learning excessive features related to the human faces and this may cause the model to give them greater importance and the network may not recognize masks without facial features bais. We've decided on marking only the masks consistently and ignore the mask strips.

How we chose the architecture:

After examining the existing options we chose the FASTERRCNN architecture. This is a network that works in two stages where in the first part feature extraction is performed by a separate NN (backbone) and in the second stage it is determined the object and its location. This architecture allows us to freely change the backbone type in the first stage .

We chose this network because it is a well-known and flexible network with a good balance between the training time and the quality of the results returned.

Preparing the tools for training

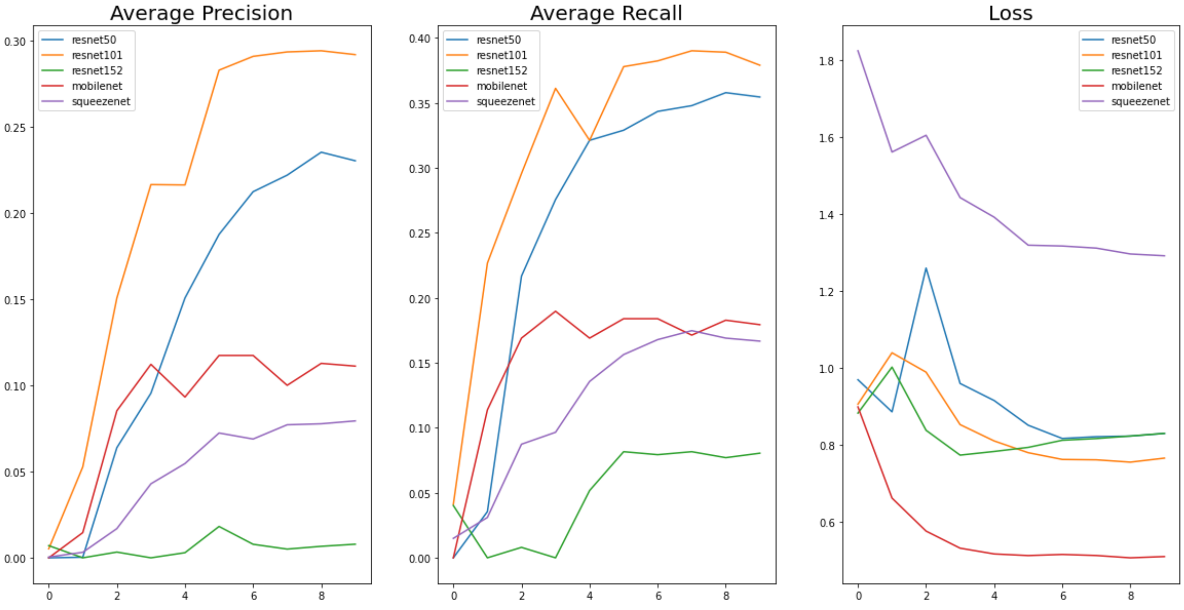
Implementation of Torch Dataset and DataLoader.
For the sake of working with the images we have set up a custom dataset. Its job is both to pull an image from the image folder and also to match them to the relevant annotations format from the json file. In addition, a data loader was set up with which we will pull the images during training.

Choice of backbone and apply “Dirty Implementation”

As mentioned earlier, the network has two steps to perform the identification and we are given the right to choose which network the backbone will form. You can select one of the networks from Torchvision.models, and you can select a network pretrained on the imagenet dataset.

To decide which network is better to use as a backbone We replaced the networks backbone several times and ran training in methodology in dirty implementation. Each backbone was trained for about 10 epoches. From resnet50, resnet101, resnet152, mobilenet, squeezenet, which we examined, the model that achieved the highest Average Precision result was selected.
The model chosen was resmet101.

Graph during dirty implementation



Improving results:

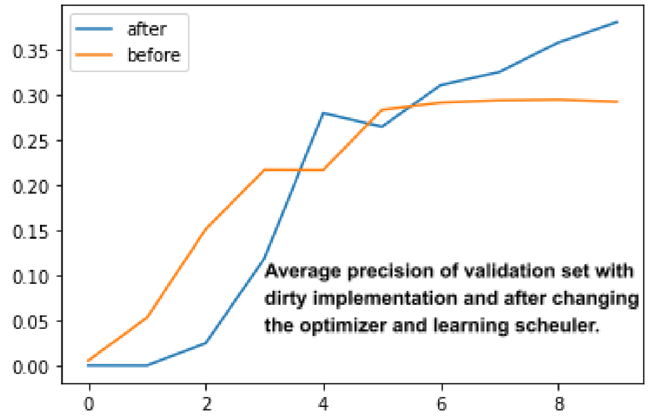
In order to improve the training results we chose to change the optimizer, scheduler and initial learning rate.

We trained the network with an **AdaMod optimizer** This is an optimizer that limits the highest and lowest learning rate and makes learning rate changes more smoothly, it is based on adam, less sensitive to hyperparameters and does not require warmup unlike similar optimizers.

For the learning rate schuler we used **Cosine scheduler** , this is a scheduler that allows you to lower the learning rate in a linear and graded manner. In practice the use of this tool is preferable because it has fewer hyper parameters that need to be adapted.
The learning rate dropped to 0.0001

Results:

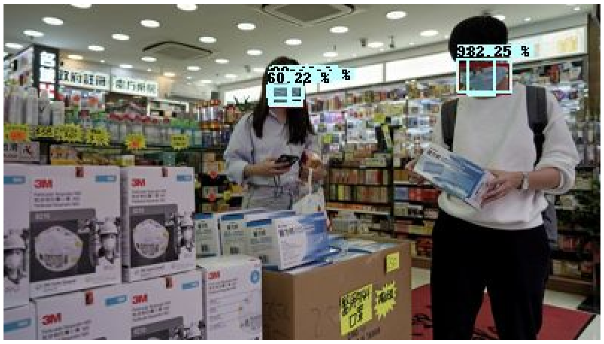
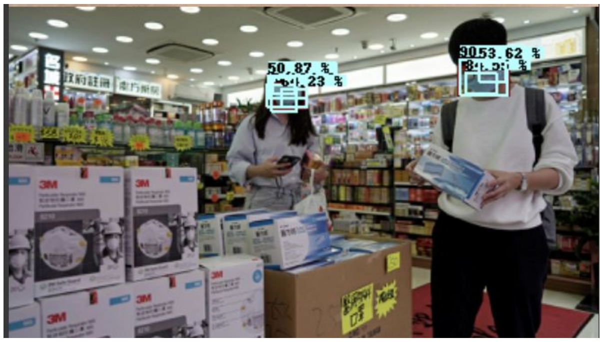
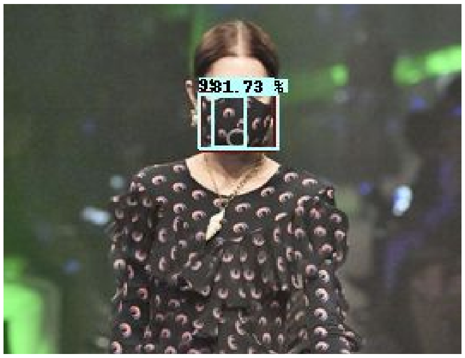
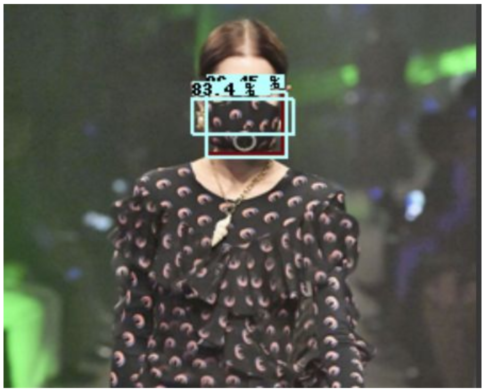
The results did improve after the change and it can be seen that the training results are quite good, the network consistently identifies masks in different colors and lighting conditions with relatively few false negatives, and with tight markings around the masks. The weaknesses of the network are the failure to identify masks at extreme angles, as well as the multiplicity of bounding boxes around a single mask.

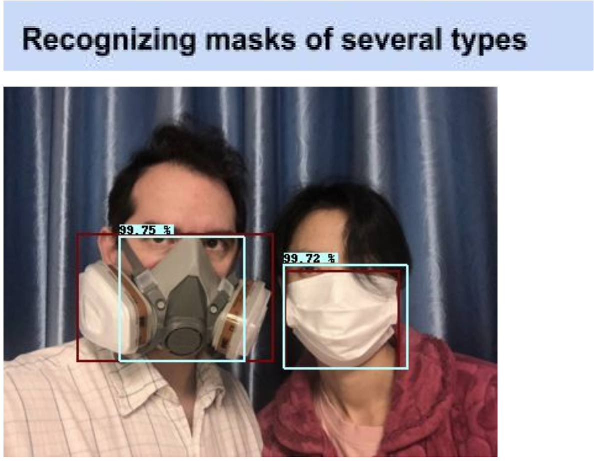
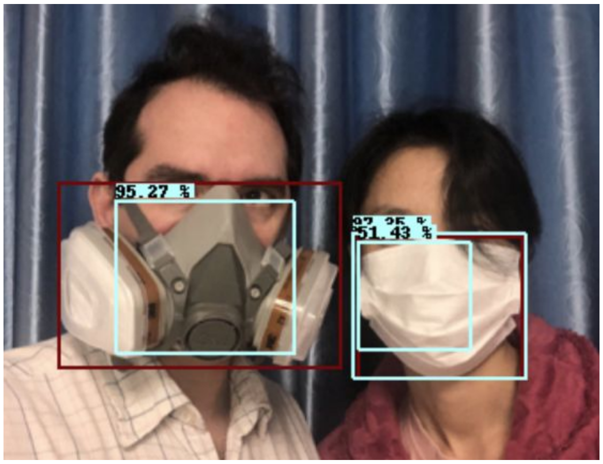
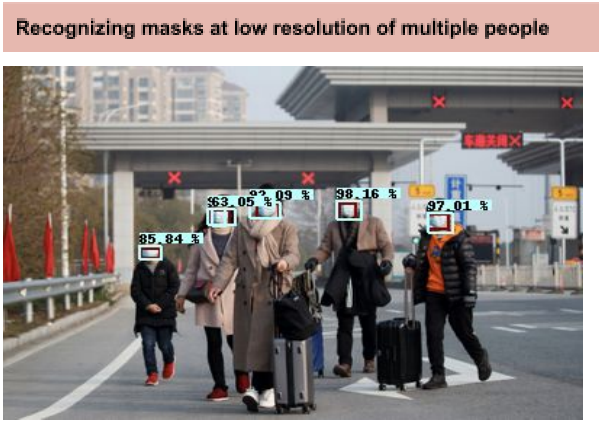
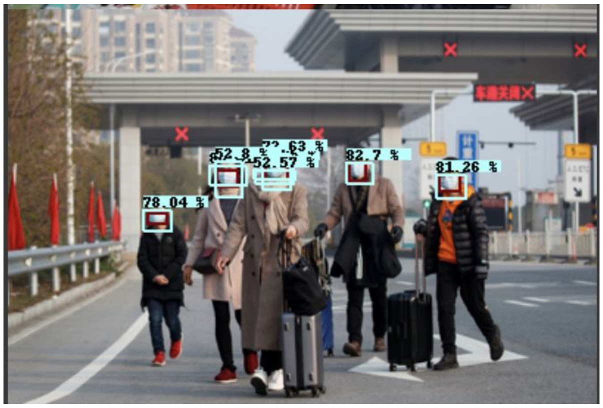


Average precision of validation set with dirty implementation and after changing the optimizer and learning scheuler.

Things to improve:

Using the methods to unify predicted bboxes (like Weighted Boxes Fusion (WBF) or non-maximum suppression (NMS)), testing which threshold results in more accurate products, attempting to use new and more modern models for the task, testing more random learning rates and optimizers for the tasks, and adding image augmentations during training.

<table>
<tr><td colspan="2"><b><u>So-so Results</u></b>
Multiple bounding boxes on the same people</td></tr>
<tr><td>After improvements</td><td>Before improvements</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Difficult textures on the face, yet the model recognizes the mask.

</td><td>
</td></tr>
</table>

<table>
<tr><th colspan="2"><u><b>Good Results</b></u>
One bounding boxes on the same people</th></tr>
<tr><td>After improvements</td><td>Before improvemenst</td></tr>
<tr><td>
Recognizing masks of several types
 </td><td>
 </td></tr>
<tr><td>
Recognizing masks at low resolution of multiple people
 </td><td>
 </td></tr>
</table>

Yoav Raytsfeld
Danel Kornis

From undetected to detect	
This person is wearing a hat that almost covers her eyes, yet is still recognized as wearing a mask