

O where the party at?

Project 1.2

Deep Learning in Computer Vision

June 2021

You've been invited to a party by a friend and have been given the address. However, you suffer from severe Dyscalculia and cannot read house numbers. You have therefore decided to create an algorithm that automatic can detect and read house numbers. Fortunately, you're the world champion of image detection using deep learning and know exactly what to do!

To train your model you should use the Street View House Numbers (SVHN) dataset <http://ufldl.stanford.edu/housenumbers/>. PyTorch has a version of this dataset built in where all the digits have been cropped and re-sized to 32x32 images. This is perfect for training! The labels in this dataset are integers between 0 and 9. You also need to train your classifier on images that do not contain



Figure 1: Examples from the SVHN dataset. Left: Full images with bounding boxes. Right: cropped images centered around a single digit.

digits. Otherwise you will always detect a digit, which is not very useful. You have a free choice on which images you use for this. You can for example use other built-in datasets such as CIFAR10 and/or crop parts from the full size training images where there is no digits. This is an important step to avoid false

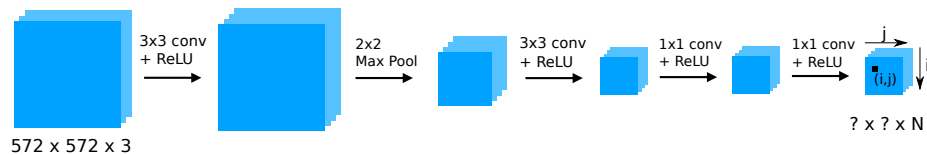


Figure 2: A CNN. All convolutions and poolings are without padding.

positives. (Hint: create another class for negative samples meaning 'no digit')

For testing purposes you should use the full non-cropped test images in a sliding window way. These can be found on: <http://ufldl.stanford.edu/housenumbers/>. We have also uploaded the full size training and test images to https://drive.google.com/uc?id=10N__ctmFgHKijCla8f-FvPxDCrUImKft. Here, you will also find the bounding box information and labels stored in csv files.

Warmup

In the task below, you will find bounding boxes which are the receptive field of individual pixels of the output of a CNN.

- For the network shown in Fig. 2, please compute the size "?" in the figure. You do not need to consider the number of channels.
- For the pixel (i, j) Fig. 2, what are the coordinates of its receptive field in the original, 572×572 image?

The task

Your tasks are:

1. Train a convolutional neural that can classify single digits
2. Use your network with a convolutional implementation of the sliding window algorithm to detect bounding boxes around digits.
3. Keep only the best bounding boxes by using non-maximum suppression and Intersection over Union (IoU).
4. Use IoU with a suitable threshold to validate whether you have correctly identified the digits.
5. Explain how you have used the training- and test datasets to set your threshold and carry out your evaluation.

Your process, performance evaluation and results should be documented and discussed in a PDF poster to be uploaded on DTU Learn. Project 1.1 and Project 1.2 should be described on the same poster.

Additional task

You should give peer feedback on 3 other posters after handing in. This should be done at latest two work days after your hand in deadline. You find these assigned to you in Peergrade.