HLCV Project Proposal

Júlia Bozzino (7011915)
Ganesh Gopalkrishna Hegde (7001558)
Robin Kneepkens (7010559)

June 11, 2021

Introduction

This will be a short overview of the contents of the slides containing our project proposal. This report will be split in the same sections as the slides.

Task and motivation

Our goal will be to make a model that can successfully classify bird species from North America. We have found a good dataset for this subject, however, this dataset has the issue that the input images have a wide variety of dimensions. While convolutional layers have no issues with arbitrary sizes of input images, the fully connected layers deeper in the CNN's often expect a preset input size. This cannot be guaranteed by the convolutional layers if the input image sizes differ.

This first lead us to the idea of having a two network architecture. One BirdSpotter network that crops the bird from the image so the Birder network can use the cropped image to classify. This still runs into the input size problem for both networks, so we then wanted to find some way to handle that as well. This lead to finding the Spatial Pyramid Pooling method that allows CNN's to handle arbitrary input dimensions. We still liked the idea of having two networks, so we decided to implement that as well, as it might lead to better results. This architecture can be seen as a hard attention model.

Our motivation is that fine grained image classification seems hot right now. The same is the case for attention based systems. It also seems like a very fun challenge and an intuitive architecture for approaching such a task. Finally, one of our group members also really likes birds.

Goals

Of course it would be very nice if we manage to get a network that can compete with the cutting edge for this dataset, even though that is unlikely. This leaderboard will be our main performance metric. We also hope to successfully implement the SPP technique, as handling arbitrary image dimensions seems like such a natural feature for CNN's to have.

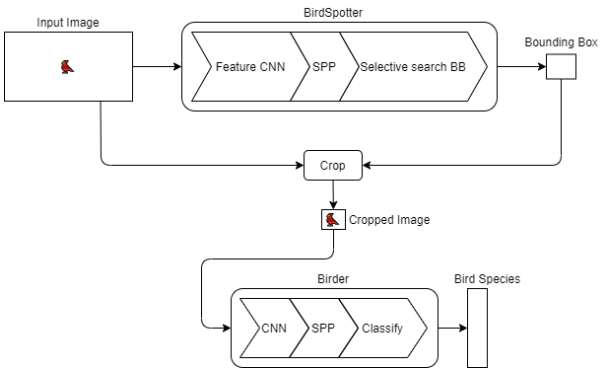

Figure 1: Rough overview of the model architecture.

For the mid term, we would like to have a solid grip on the dataset: solid handles for loading and interacting with the images and annotations, and some exploratory data analysis and visualizations. It would also be very nice if we can have an initial implementation of the two networks, at least the BirdSpotter. This would give us some leeway to experiment with different pretrained feature extractors and maybe even give us the chance to train a feature extractor ourselves.

Methods

Figure 1 shows a rough outline of the architecture. The SPP layers allow the Selective Search Bounding Box generator and the classifier to work on arbitrary sized input images. We decided to also add an SPP layer in the Birder network, as the bounding boxes can also be arbitrarily sized. Even though it is a common practise to resize inputs for a CNN, this is often done by cropping or stretching, with both influence the original data in the image. With cropping, important parts of the bird have the chance of not reaching the Birder network. With stretching, proportions of the birds may be off, which could negatively influence the classification step.

We can use the convolutional layers from pretrained CNN's for the feature extraction step (the ones labeled with CNN). This allows us to focus mostly on implementing the SPP technique and tweaking the hyperparameters of the networks for better performance.

Dataset

Our dataset is fairly extensive. It contains about 48000 images of birds from North America. There are 400 species and each species has at least 100 photos. There are also annotations for each image containing information such as bounding boxes around the bird (perfect for training the BirdSpotter), as well as some other, less important annotations.

Evaluation

Our main metric will be accuracy of the Birder network. This is also the metric that is displayed on the leaderboard. Because we have a sort of modular design, we can also train and evaluate the BirdSpotter individually. This can be done with the mean average precision metric, which looks at how much the suggested bounding box overlaps the given one.