Object Tracking using Neural Networks

Alexander Selberg

June 2020

Machine Learning has had an enormous upswing in recent years, revolutionizing industries and paving the way for an era of Artificial Intelligence. This thesis has investigated how Machine Learning can be used to teach a computer how to track any object. The tracker is then tested on several videos and also applied to a small robot.

Machine Learning can be defined as a learning process where the computer through repeated exposure and gained experience learns to recognize patterns and important features for different kinds of problems. This can be anything from learning to differentiate between spam and non-spam emails, creating an automatic customer system, or identifying cats in images. The task in this thesis is to teach the computer to follow any object, given only its initial starting position.

We successfully managed to create a tracker that could track several different objects such as a car driving down a road, a ball being kicked back and forth or a woman biking down a road. The size of the tracker became a problem when applying it to the robot, requiring significant size reduction, unfortunately resulting in less-than-stellar tracking performances.

The tracker consisted of a large neural network with millions of parameters that had to be trained and adjusted. The idea is to expose the tracker to a cropped image pair from an image sequence, consisting of an object at time t and t + 1, both centered at the object at time t. The object's position at time t is known and it is up to the tracker to tell us where the object is in the image at time t + 1. The process together with the network architecture can be seen in Figure 1, where a skier is being tracked.

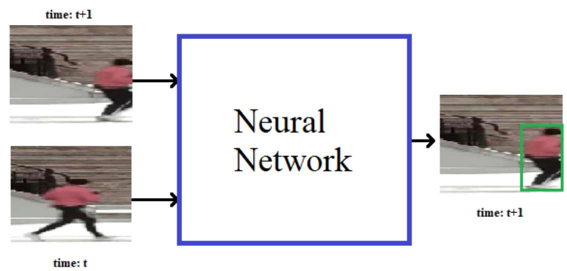

Figure 1: Network architecture with two cropped images as input and bounding box coordinates as output

The first part consisted of a training phase, where the network was exposed to hundreds of thousands of image pairs, all centered at the object at time t. The premature tracker guessed the object's position at the second image and was then given the correct position. Using the difference between the predicted and the true position, its inner parameters were adjusted to better predict the object's position for the next image pair. After having trained for multiple hours, sometimes days, the tracker was then evaluated on new, previously not seen image sequences, or videos. This is called the testing phase. The only thing given during the testing phase is the original position, after that, all the future cropping positions are based on predicted positions, requiring the tracker to perform well for the entire sequence.

In summary, the tracker performed very well for several tracking scenarios and had some problems with the more difficult ones. Its size became an issue when applying it to smaller devices but it is possible that the performance can be maintained for smaller trackers as well but requires more extensive testing.