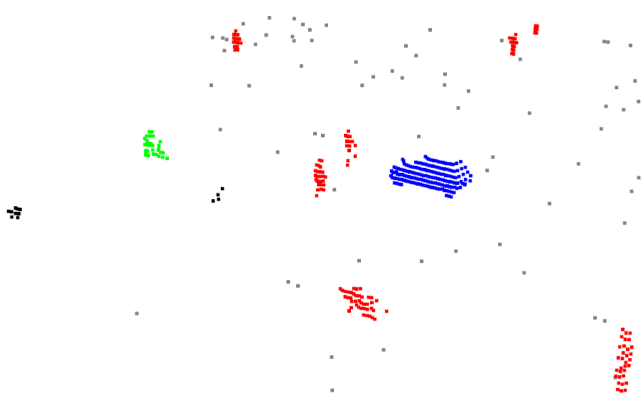
Tracking and Classifying Objects With a Lidar Sensor

By ROBIN BERNSTÅHLE & HJALMAR LIND

The year is 2122. The world is subjected to a rule by sentient fascist robots. Humans are subjugated to the robots' violent means and sophisticated technology. The rulers see you, track you, wherever you are. There is no way to hide.

Is this the future we are heading towards? While popular Hollywood films tend to suggest so every time Artificial Intelligence, machine learning and similar subjects are depicted, the truth is somewhat more nuanced and perhaps... Boring. AI-based methods for solving problems often show great potential towards solving their specific tasks, but then also nothing more nor less. We may be far from the singularity, but perhaps one step closer to a more sophisticated tracking system using LiDAR sensors. We have in our project tackled the very specific task of examining whether deep learning-based methods of tracking and classifying objects appearing in an environment monitored by a LiDAR sensor could perform better than some more classically used methods. Let's break this down.

So what is a LiDAR sensor? Much like a traditional radar, it uses non-visible light to interact with its environment. In a nutshell, it fires infrared lasers which bounce onto its surroundings, before returning to the sensor. The measured travel time of the laser beam provides an easy way to calculate the distance to nearby objects, since we know how fast the beam is travelling. As such, a LiDAR can perceive depth and yield a 3D-image of its surroundings. This 3D image is represented by a point cloud, a collection of 3D points, as illustrated in the image below. In this image we see pedestrians (red), a cyclist (green), a car (blue) and some dogs (black), together with some noise.



Once we are at this stage, i.e. having measured the surroundings and identified a couple of objects, the next step is to *track* and *classify* these objects. The tracking means to essentially answer the question: *"Over time, what objects are there that enter and leave the scene, and where do they go?"*. Once we have answered that question, we move on to classification, and ask *"What are the objects? Humans? Cars? Animals?"*. We have tried to tackle both of these problems by developing deep learning-based models. Deep learning models, or deep neural networks as they are also called, are based on a collection of neurons which loosely model the biological brain. These networks are created for specific tasks, and are then trained to perform these tasks by showing them thousands of examples. The training posed one of the main issues in this area: there is not much training data easily available. We handled this by simulating LiDAR data in Carla, an open source addon to Unreal Engine. With this addon we could successfully simulate pedestrians, cyclists and cars interacting in a realistic urban environment.

The first model trained was the tracker, in which recorded tracks from Carla was used as training data. The neural network structure in this model was based on previous research in a similar field. Parallel to this model we also created a classical baseline model for comparison, using algorithms proven to work well in tracking. The deep-learning based tracker seemed to work well at solving certain steps in the tracking process, but failed in other regards. We ended up creating a hybrid model, in which some steps of the tracking process were solved by classical algorithms, while others were solved by deep learning-based methods. A final comparison showed that while this hybrid model had potential, and in most cases performed on par with the baseline, the baseline model was still superior.

For the classification problem we created a model based on a network known as PointNet. This network was trained to identify pedestrians, cyclists and cars in Carla. It became very good at identifying objects in the simulated data, but the question was how well does this translate to real world success? Fairly well it turns out, unless people do unexpected movements not encountered during the simulation. Another path we explored was creating a model which could detect and identify objects at the same time. Here we could see some great potential, but further work needed to be done for the results to be satisfactory.

All in all, there were some good results to be extracted from both the tracking and classification parts of the projects. Deep learning-based methods have certainly proven to be the best approach to the classification problem, while in tracking it shows potential for future research and applications. In the end, any learning-based methods are only as good as they are taught, and only on their specific tasks. Seeing these technologies applied in for example surveillance settings should make us feel safer. At least for now.