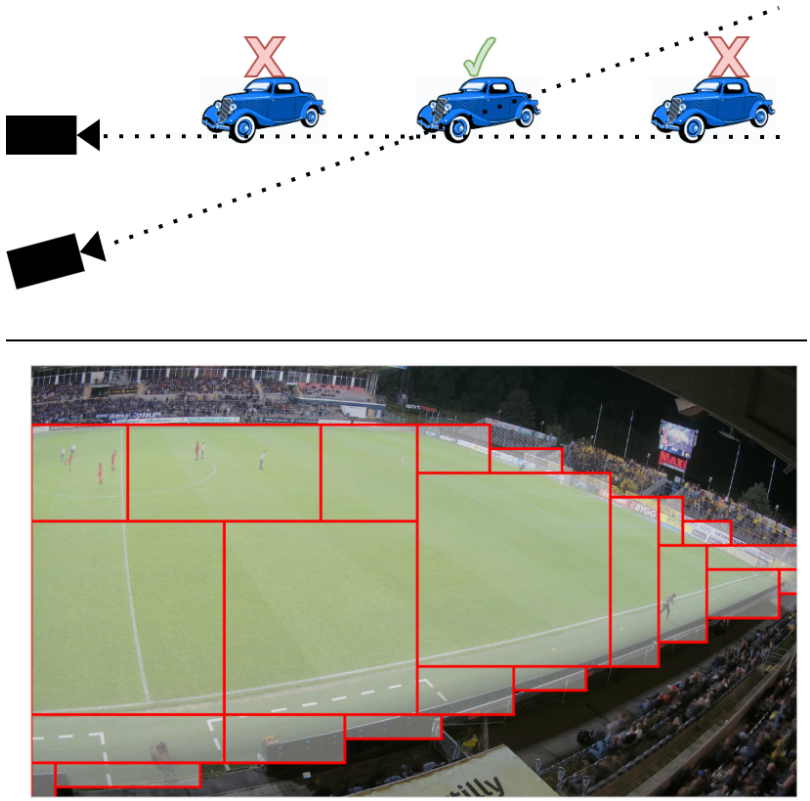
Popular Summary

Chatbots, such as ChatGPT and Le Chat, have in the last few years become well known tools that symbolises the rapid progress of AI research. They have become well known for their usability and are used extensively by both researchers and the public. In parallel with the chatbots, there have been a rapid development in other related fields shuch as computer vision and machine learning, where this thesis teats these subjects.

Two main themes from the thesis are positioning and reconstruction of objects from cameras. By using multiple camera, or one camera taking several images from different positions, you can use a technique called triangulation to find positions. Ordinary cameras we use today are so called projective cameras, where all points along a line in the romm will be shown in the same point in the image. However, it will onlt by the point cloest to the camera that will be seen in the image. Triangulation works through us knowing that a point (or an object) is located along a line from a projective camera, but not where along it. Another camera, from another position and angle, can then with a line of its own cross the first. We can then conclude an actual position. The object and the two cameras form a triangle, hence the name triangulation. An illustration of triangulation can be seen in the top half of Figure 0.1. Us humans can use this technique to some extent with our own eyes, even though greater distance between cameras (or eyes) tend to yield better results. But what happens if we cover en eye, like a pirate? After some time we humans get used to it and function well anyway. This thesis is concerning this very situation where we only have one camera, a so called monocular setup. Then, we lack the ability to use triangulation, however, though addition of machine learning and other solutions this problem can be mitigated.

Article I concerns positioning and reconstruction of cars, seen from another car. This is intended for self driving cars etc where all other road users in the vicinity of hte car needs to be positioned. Except cars, there are other road users, which article II treats, where pedestrians and cyclists are positioned alongside the cars. This can be used in traffic surveillance where you want to detect potentially dangerous situations or accidents.

The subsequent articles treats the topic of positioning of football players on a field. The positioning was made a part of creating a Birds-Eye-View of the field, where it can simplify the detections. The method worked by creating a square grid over the field, where every square was classified as containing a player or not, by the help of a neural network. To improve the performance corrections of the positions were made in the same neural network. The initial work was made in article III with an improvement in article IV. The improvement consisted of Non-Maximum Suppression, a technique that prevents several detections of being made on top of each other. Article V introduced tiling of the input image to the neural network. Through the structure of the neural network used we were able to guarantee the same detections, on a small part of the football field at a time. The tiling is illustrated



Figur 0.1: Triangulation (top half), is possible when camera one, which is a projective camera can see that a car (or a point of it) is somewhere along a line, but not where. Camera two can then with the help of its information, with a line to the car, solve the problem with the help of triangulation. Tiling of a football field (bottom half) does not need to have the same shape or be square. They are done to reduce the calculations needed and memory usage for the image. The image is an example of results from article V.

in the bottom half of Figure 0.1, where each tile is treated by itself. The memory usage could be kept low by analysing one tile at a time. The result from each tile could then be concatenated into detections over the entire football field. Article VI concludes the thesis by putting the system together from articles III, IV and V and also quantising data and weight in the neural network. Quantisation means that instead of using decimal numbers, so called floating point numbers, we only use integers. These are far simpler and quicker to calculate for the network, while also requiring less memory. A drawback is that they can degrade performance. When quantised networks are used they can be used on small, so called embedded devices who lack the big and powerful graphics cards that are usually needed.

The results indicate the possibility of using small and specialised network for positioning where they can be used outside of big server halls or specialised desktop computers.