Tracking of professional swimmers on video clips taken during competitions

T. Vincent, R. Carmigniani, V. Lepetit - ENPC

Context

In sport, collecting data on the performances is a key for improving the results of any sportsman. The problem with swimming is that no solution have been found to record accurate data. Especially during official events where devices are prohibited on the swimmers and in the water...

Thus, we have shaped a new way of getting the position of swimmers using the video clips that are taken by the French Swimming Federation during competitions. The new way uses deep learning technics to track the position of swimmers. To estimate the position of swimmers we have asked the neural network to focus on the head of the swimmer.

The only assumptions that were made are that the camera has to be on the side of the pool, perpendicular to the swimming way and that the camera has to be fixed during the entire race. No assumptions were made on the type of swimming.

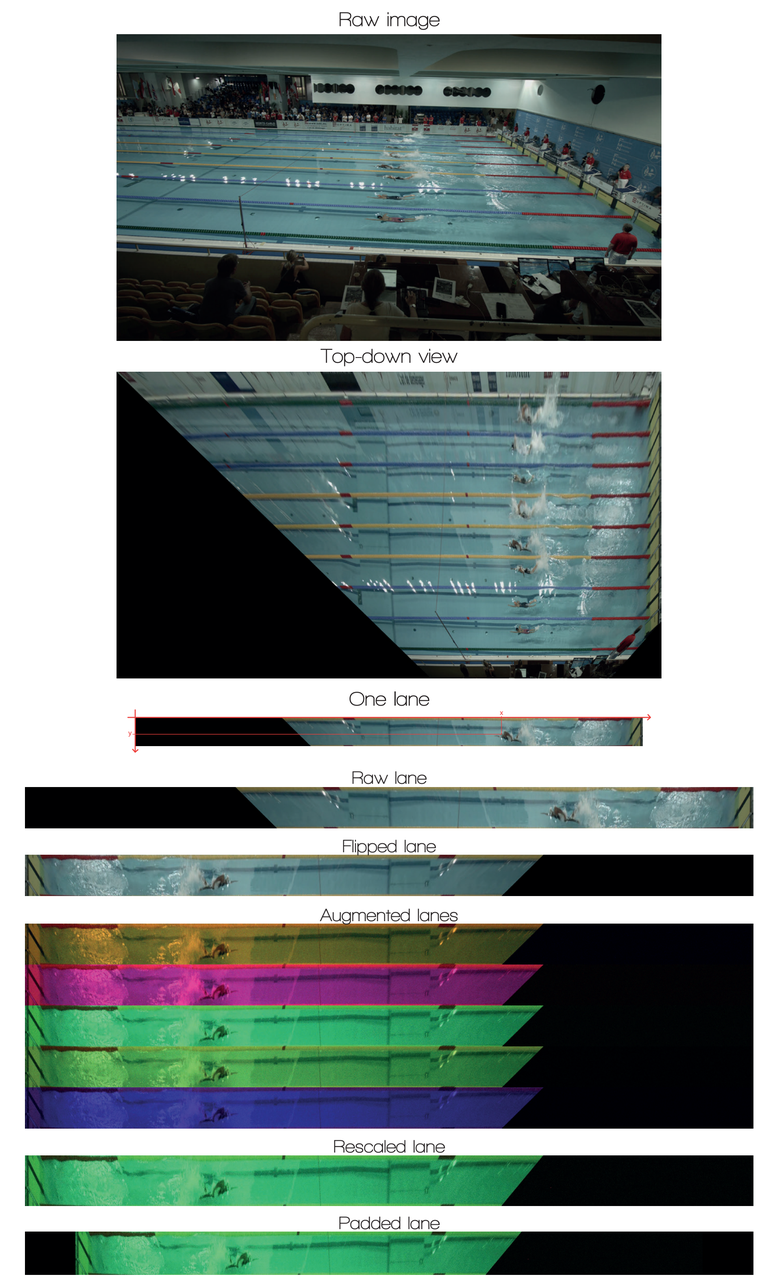
Overview

The first step of the project was to create a database since no one has delivered a free database linked with the problem we are looking at. Our goal was to craft a database that could be used by other researchers who want to work on this problem. In total, 16800 lanes where labelled by hand. On average it tooks three seconds per frame to point at the head of the swimmers. This database is now freely accessible on the github of the project.

Then, each frame of the video passes through a pre-processing step, which is describe on the right side, before the neural network estimates the position of the head of the swimmer. The idea of this step is to make sure that the swimmer is always swimming toward the right side with the same size. This is not always the case since videos can be made at different moment of the race with different zooms. Moreover, the colors of the lanes are changed to make sure that the network will manage to find the head of the swimmer in more complex situations.

Two neural networks are used to get the position of the head of the swimmers. The details are explained underneath.

Pre processing



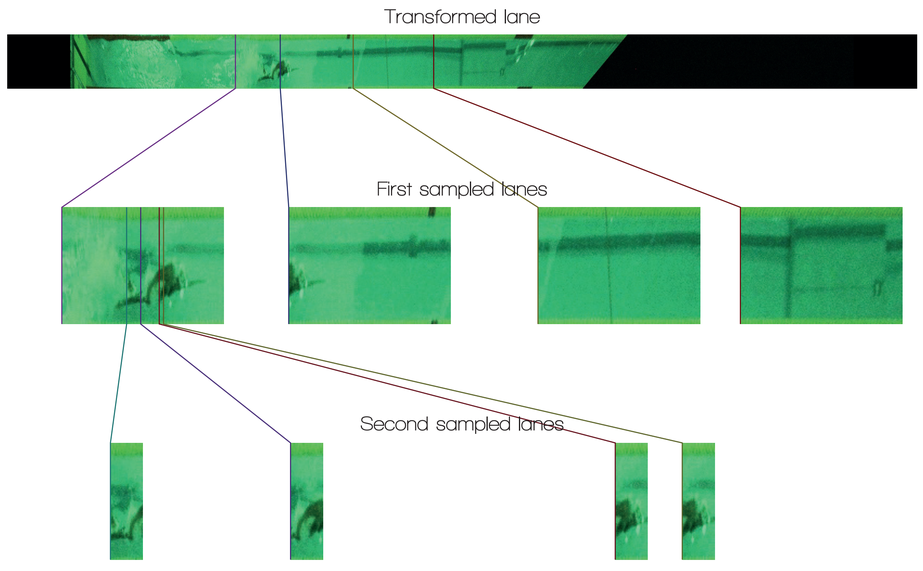
After spliting the video into frames, the pre-processing step computes the top-down view of the pool before dividing the image into lanes. It is at this stage that the data is labelled. Then, the lane is flipped if the swimmer is not going to the right and the color is changed. Finally, the scale is moved to a common one and the image is padded to reach a common size for all the lanes.

Neural Networks

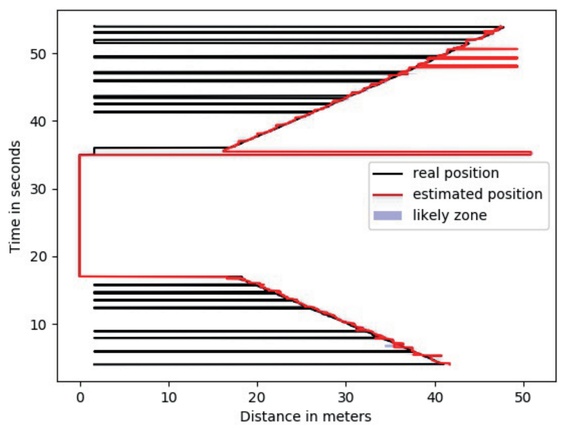
Once the pre-processing step is finished, a first batch of random samples is taken. Each of these samples is passed through a first neural network that has to determine if the head of the swimmer is inside the frame. The loss of the network is simply the binary crossentropy loss. Three convolutional layers and two fully connected layers are applied. We achieved to reach 98.2 % of accuracy on the validation set for this problem.

After that, a second neural network with the same shape as the previous network is applied to the samples that where generated previously. This time, the network outputs the probability that the head is inside the frame and also the the column where the head is. Thus the loss is different: it is the sum of the crossentropy loss on the probabilities and the mean square error. A trade-off between those two losses have been tuned. The accuracy has raised up to 92 % on the validation set for the classification problem.

During the evaluation, instead of taking random samples, samples are taken uniformly over the lane with some recovery in order to be sure that the head is not cut between two samples.



Results



The graph shows in black the position of a swimmer's head estimated by a human, horizontal lines represents the seconds where the human did not succeed to point at the head of the swimmer. The red line is the estimation of the network. Thus, the network perforoms better on this validation video since it manages to find the position of the swimmer even when the human cannot see it.

Difficulties

The first difficulty was the fact that this problem has not been studied a lot so the state of art was not developped a lot. Then, crafting a database that can be widely used was a great challenged. Finally, we decided that our algorithm should be able to deal with all kind of video. We had to implement an algorithm that can cope with different quality, occlusions, difference perspectives and zooms.

References and Contact

A. Alfalou A.Verney P.Hellard D.Benarab,T.Napoléon. Optimized swimmer tracking system based on a novel multirelated targets approach. ISEN Brest, Département Vision, Brest Cedex 2, France, 2016.

This work was supported by the NePTUNE project funded by the French National Agency of Research (ANR) [ANR-19-STHP-0004NePTUNE]

Git : https://github.com/remingtonCarmi/TrackingSwimmingENPC
LinkedIn : https://www.linkedin.com/in/th%C3%A9o-vincent-ba573a180/
Mail : theo.vincent@eleves.enpc.fr