End-to-end object tracking in high-resolution videos

Lasse Regin Nielsen s123815@student.dtu.dk
Nikolaj Vestbjerg Christensen s123631@student.dtu.dk

October 24, 2016

1 Introduction

You are watching your favourite team play a soccer match. The camera is zooming in around the action packed area of the ball, which intensifies the experience. This scenario is in reality handled by a lots of camera men following the ball with their cameras. This is a costly process and prawn to human errors. What if it was possible to have the camera zoom in automatically on the area of action with perfection and no delay. This is essentially what this project can lead to among other things as we try to make a deep learning network that can learn to track where the ball is in the videoframes.

[1] [2]

2 Datasets

The dataset used in this project is provided by `Sportcaster` and contains a number of video samples from Ultra HD recordings of different football matches. Each frame has an associated data dictionary, containing information about the position of the ball in the frame - if the ball is visible.

I.e. our dataset contains frames of dimension 3840×2160 with, and with out a ball and it's corresponding position coordinates.

3 Methods

4 Results

References

[1] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", 2015. [Online]. Available: `http://arxiv.org/abs/1409.4842`.

[2] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking", C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds., pp. 809–817, 2013. [Online]. Available: `http://papers.nips.cc/paper/5192-learning-a-deep-compact-image-representation-for-visual-tracking.pdf`.