Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MIN Faculty
Department of Informatics

Using FCNNs for Goalpost Detection in the RoboCup Humanoid Soccer Domain

Jonas Hagge

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Technical Aspects of Multimodal Systems

4. June 2019

Outline

Motivation

- goalposts important for soccer
  - localization
  - shooting to goal
  - avoiding goalkeeper
- deep learning
- FCNNs have been shown to work for ball detection
- rules restricting other sensors

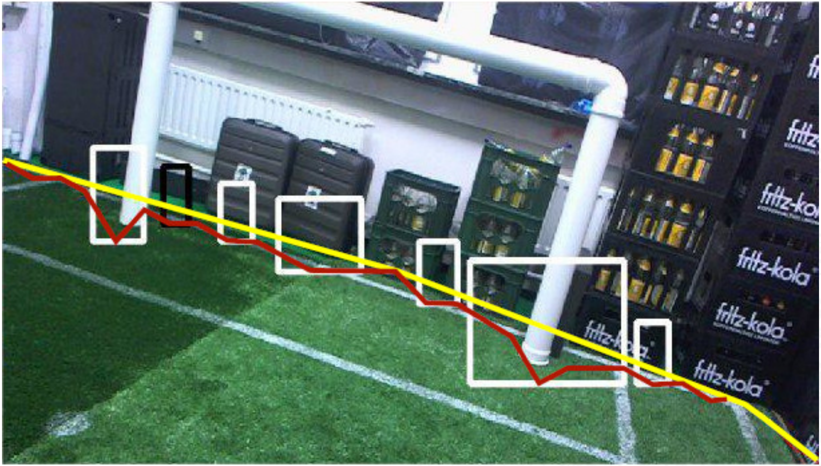
Conventional Approach

- currently used approach works based on field boundary
- obstacles block way of field boundary
- mostly white obstacle is classified as goalpost

Example of Conventional Approach

The conventional approach relying on the detection of the field boundary to find dents in it and then color detection to find goalposts.

Conventional Approach - False Positives



The conventional approach sometimes detects far too many false positives.

Conventional Approach - False Negatives

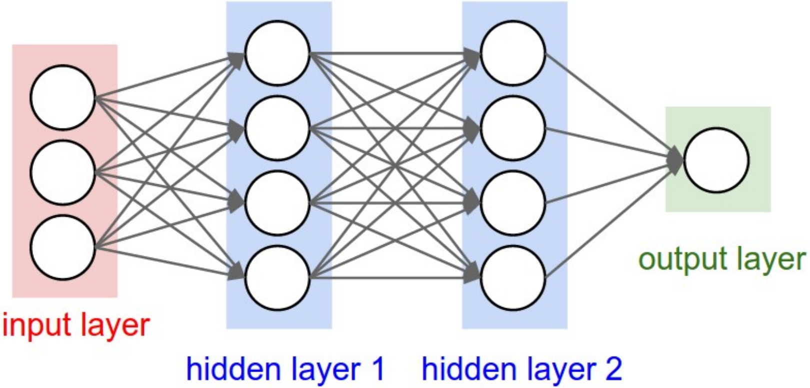
The conventional approach also has the problem of detecting too many false negatives.

Computing Platform

- Nvidia Jetson TX2 [1]
  - 256-core NVIDIA Pascal^{TM} GPU
  - 1.3 TOPs
  - Dual core and Quad core CPU
  - 8 GB Ram
- NVIDIA TITAN X Pascal [2]
  - used for training
  - 3584 NVIDIA CUDA Cores
  - 11 TFLOPS
  - 12 GB of GPU memory
- Logitech C910 camera [3]
  - 1080p
  - USB 2.0
  - our team:
    - static color balance
    - dynamic exposure

[1] https://developer.nvidia.com/embedded/develop/hardware
[2] https://www.nvidia.com/en-us/geforce/products/10series/titan-x-pascal/
[3] https://support.logitech.com/en_us/product/hd-pro-webcam-c910/specs

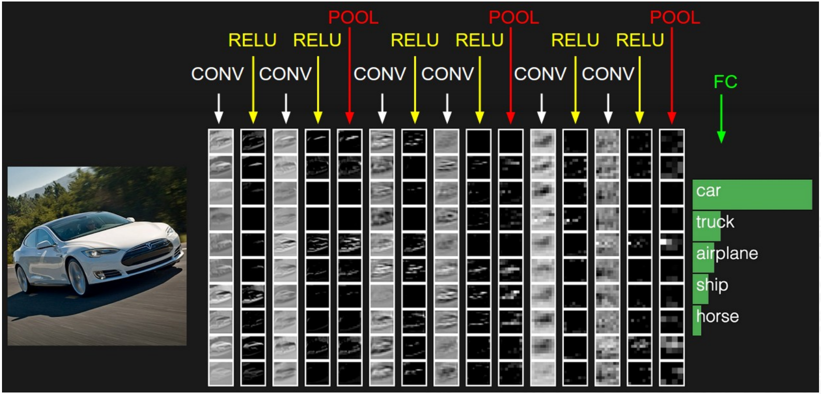
Convolutional Neural Networks



[4] Image of a conventional neural network

[4] http://cs231n.github.io/convolutional-networks/

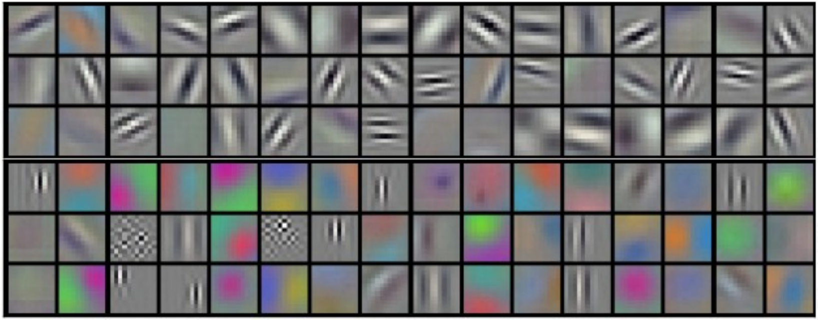
Convolutional Neural Networks



[5] Example of the layers a convolutional neural network

[5] http://cs231n.github.io/convolutional-networks/

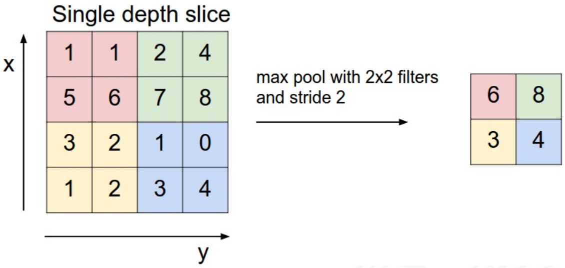
Convolutional Neural Networks - Filter Example



[KSH12] An example of what filter can look like. This example is from the AlexNet, which was trained for detection of significantly more classes.

- A filter is a matrix multiplication of learned weights and pixel values.

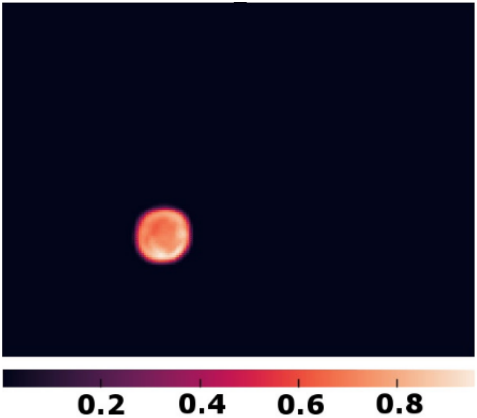
Convolutional Neural Networks - Max Pooling



[6] A max pool layer is used to down sample the image.

[6] http://cs231n.github.io/convolutional-networks/

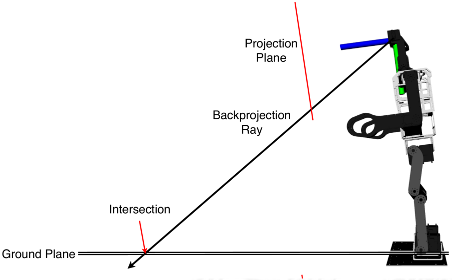
Convolutional Neural Networks - Example Output Image



[SBB18] An example image of what the FCNN output could look like for a ball. Black (no activation) where no object was detected and brighter the more activation there is per pixel.

Transformer

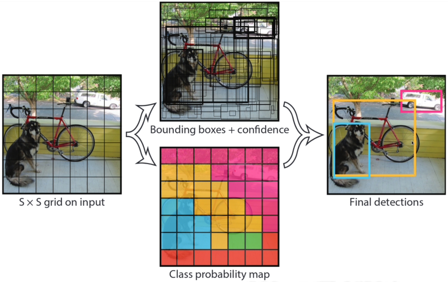
- Transformer is needed to get coordinates from image space to Cartesian space



[Gü19] The transformer works by knowing the motor positions and the position of the ground and inferring where an object lies by calculating at which angle an object would be where in the image space.

YOLO

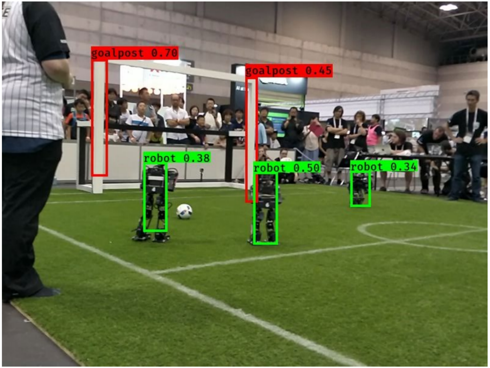
- used by multiple teams in RoboCup Soccer domain



[RDGF16] The YOLO approach works by splitting the image in to a grid, simultaneously detecting possible bounding boxes and calculating class probabilities and finally combining them to get the final detection

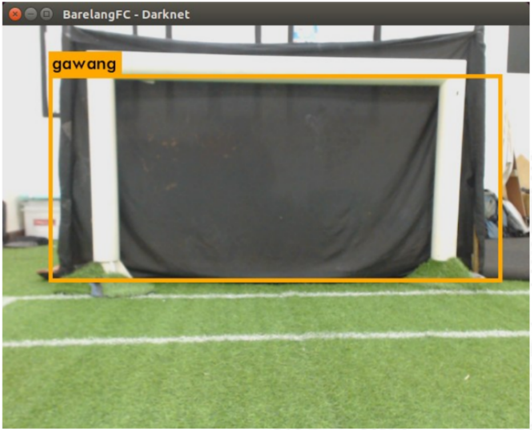
YOLO in the RoboCup Domain

- bounding box approach is less accurate
- less accurate transformations is bad for localization



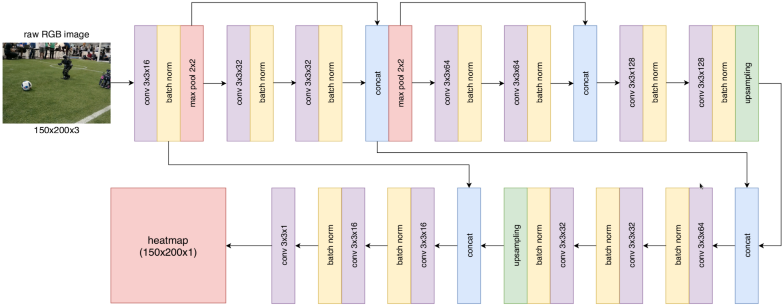
[Pie19] The bounding boxes in this example are accurate, but still less precise than what an FCNN could achieve with being pixel precise

YOLO in the RoboCup Domain



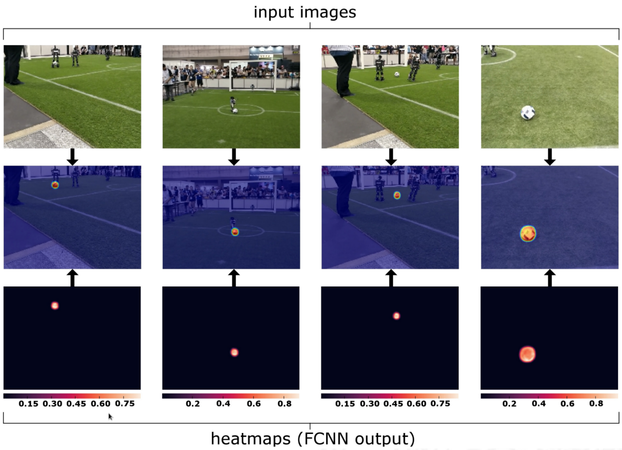
[RAS+17] Team Barelangs approach tries to detect the whole goal. This shows the inaccurate bounding boxes generated by the YOLO architecture. This would significantly harm the precision of the localization

Ball Architecture



[SBB18] The architecture currently used by the Hamburg Bit-Bots robots for ball detection which was developed by Speck et al. The approach presented here builds upon this architecture to also find goalposts.

Ball Architecture Example Image



[SBB18] Example output generated by the neural network from Speck et al. to detect the ball.

Data Set

- human labeled bounding boxes or polygons
- custom export formats possible → YAML

Ball labeled by a human[7]

[7] https://imagetagger.bit-bots.de/

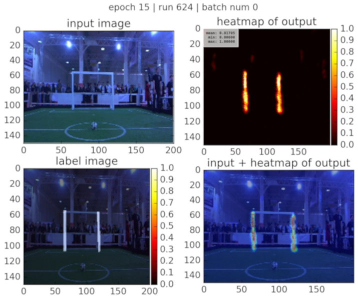
Data Set

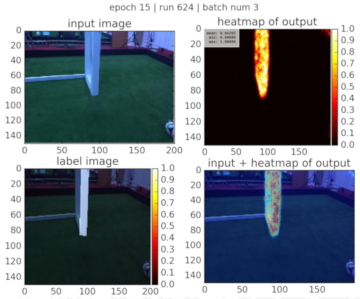
- thousands of labels created for the purpose of this bachelor thesis

Goalpost labeled by a human[8]

[8] https://imagetagger.bit-bots.de/

Single Channel - full goalpost

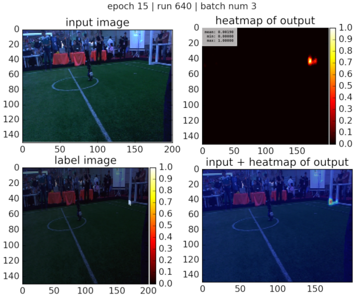


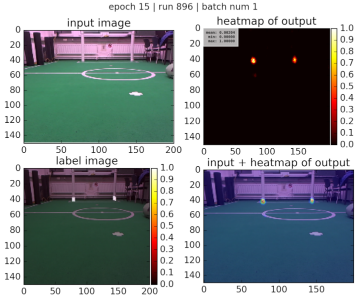


Single Channel - Bottom Part

- bottom part of label
  - depending on post processing beneficial for transforming into cartesian space
  - detection is not good enough to detect e.g. distance from height
  - other parts of goalpost irrelevant
- runtime difference would be high

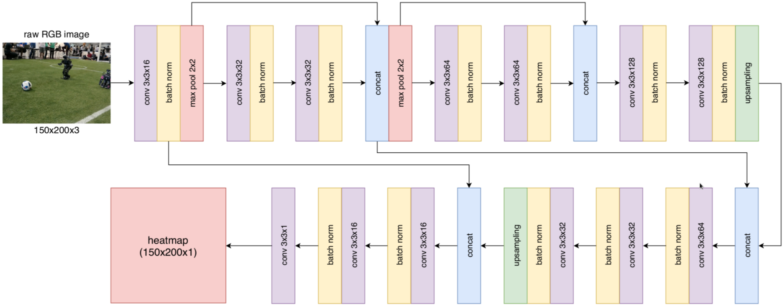
Single Channel - Bottom Part





Approach of using only the bottom part is able to detect the goalpost.

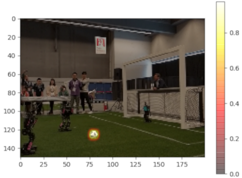
Architecture for two Channels

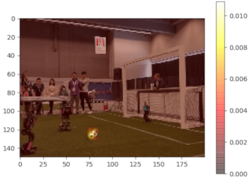


[SBB18] Architecture difference is just another filter in the last layer and thus also 2 heatmaps as a result. (Image of Speck et al. architecture for reference)

Two Channel

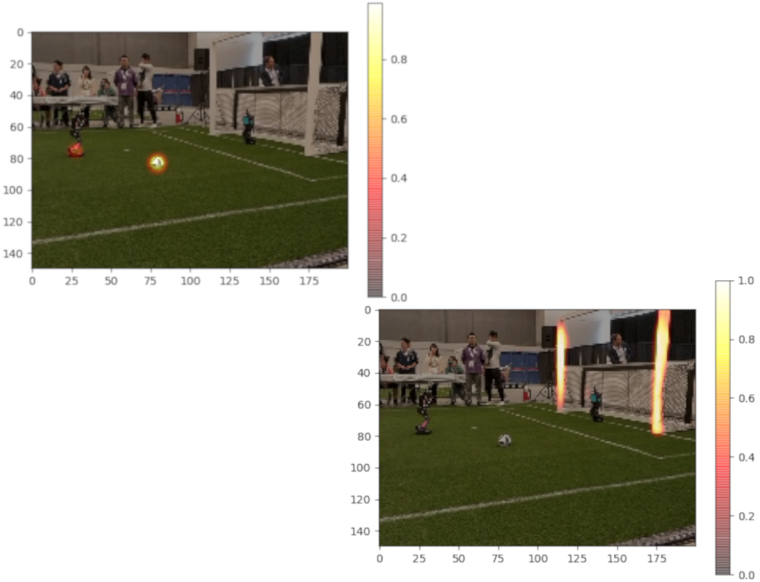
- first convolutional layers can use the same filters $\rightarrow$ less runtime needed
- splitting using bottom part of label





The top left image shows the activation for the ball. The bottom right image the activation for the goalpost. The ball is detected, while the goalpost layer has almost no activation.

Two Channel with full Goalpost Label



Ball Detection Example

Goalpost Detection Example

Evaluation

- bottom part of goalpost single layer:
  - IOU of 0.4523
- two channel with full goalpost:
  - IOU for ball: 0.6578
  - IOU for goalpost: 0.4032

Live Demo Video

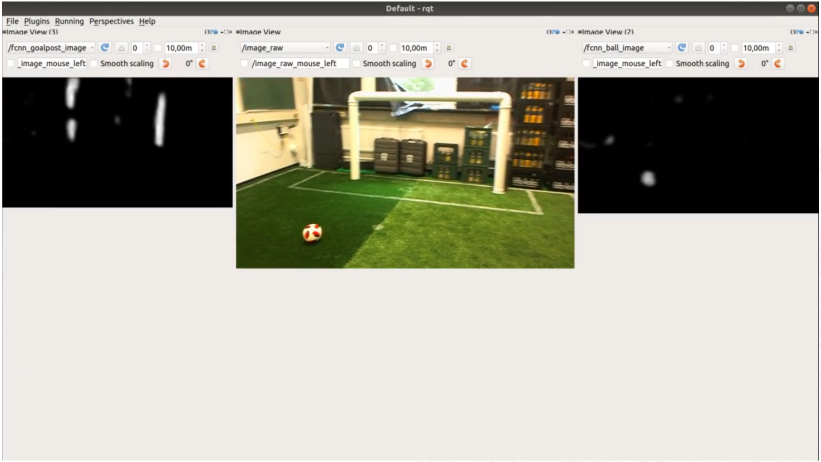


Image from the output of the neural network next to the input image.

Conclusion

- the architecture works for detection of two classes
- results worse than for single class detection, but useful due to run time trade off
- could be improved with more data
- bottom part of label approach necessary for field of view of robot
- Future Work:
  - detection of more than two classes (e.g. Field markings and Robots)

[Gü19] Jasper Güldenstein.
Comparison of measurement systems for kinematic calibration of a humanoid robot.
Bachelor Thesis at the University of Hamburg, 2019.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton.
Imagenet classification with deep convolutional neural networks.
In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[Pie19] Marius Pierenkemper.
Object detection with yolo by means of transfer learning in the robocup context.
Bachelor Thesis at the University of Hamburg, 2019.

[RAS+17] Eko Rudiawan, Riska Analia, P Daniel Sutopo, Hendawan Soebakti, et al.
The deep learning development for real-time ball and goal detection of barelang-fc.

In *2017 International Electronics Symposium on Engineering Technology and Applications (IES-ETA)*, pages 146–151. IEEE, 2017.

[RDGF16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi.
You only look once: Unified, real-time object detection.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[SBB18] Daniel Speck, Marc Bestmann, and Pablo Barros.
Towards real-time ball localization using cnns.
Robot World Cup XXII. Springer, 2018.