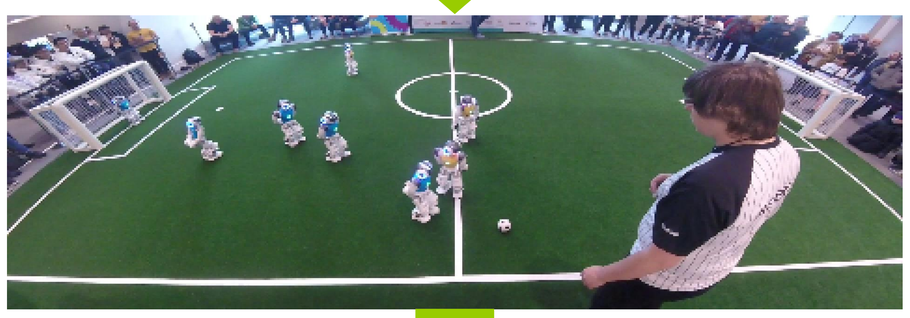
Nao Devils Dortmund[1]
Open Research Challenge

Arne Moos[2] and Dominik Brämer[3]

Idea

Our contribution to this challenge is a real-time SPL robot and ball detection system calculated on a Jetson Nano board using the wide-angle images from a GoPro-like camera.
This allows us to use the recognized image coordinates and convert them into field coordinates for higher level processing.



Hardware

The hardware platform we chose is the NVIDIA Jetson Nano board, which offers a quad-core ARM processor and an Maxwell GPU. It was chosen because this board is already used for a SPL live stream and allows de- and encoding a 4K (H264/HEVC) video stream in real-time allowing to show additional data (e.g., GC data) as an overlay.

Implementation

The machine learning architecture YOLOv3-Tiny (Redmon and Farhadi, 2018) implemented in Tensorflow 2 is used and modified for our object detection task. To achieve real-time performance, the number of filters is reduced and we switched to Depthwise Separable Convolutions.
In addition, the input is cropped (top and bottom) and resized to the input sizes 480x160, 672x224 and 960x320.
Our implementation uses a training/validation split of 70% as well as image augmentation (flip, affine transform, brightness/contrast, …).
For inference, TensorFlow Lite on the CPU is used with varying number of threads.

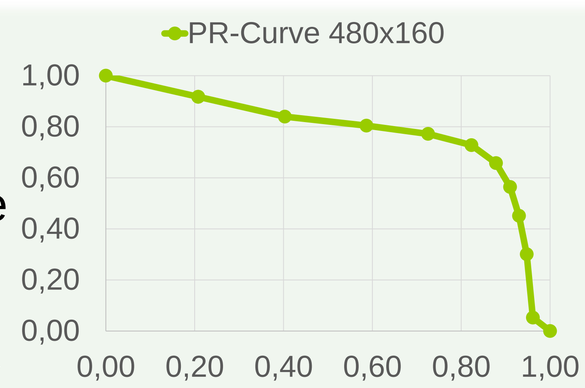
Evaluation and results

Different metrics are used to evaluate the resulting neural networks:

- First, the number of floating point operations (FLOPS) is given along with the number of trainable parameters.
- Then follows the number of frames per second using different thread counts.
- After that the maximum F1-score over all thresholds between 0 and 1 is listed.
- Based on this threshold, the precision and the recall are individually specified.

In addition, the printed precision and recall curve illustrates the influence of different thresholds for the smallest input size.

Outlook and conclusion



At least the first two of the tested resolutions could run faster than 30 FPS.
The output of the network Can and should now be passed to a multi object tracker (SORT, BYTE, ...) to smooth the detection even more.
Because we are using a wide-angle camera the images are distorted. So it can be beneficial to train the net on the undistorted images. But then, in addition to the actual latency of the neural network, there is also the rather computationally intensive conversion.

input size	FLOPS	#parameter	FPS 1T	FPS 2T	FPS 3T	FPS 4T	F1-score$_{50}$	precision$_{50}$	recall$_{50}$
480x160	~85M	168,336	~35	~51	~61	~67	0.773@0.5	0.728	0.823
672x224	~165M	168,336	~18	~26	~31	~34	0.760@0.6	0.690	0.846
960x320	~340M	168,336	~9	~13	~16	~17	0.724@0.5	0.628	0.854

[1]naodevils.de [2]arne.moos@tu-dortmund.de [3]dominik.braemer@tu-dortmund.de

technische universität dortmund

Robotics Research Institute
Section Information Technology