

Lab CudaVision

Learning Vision Systems on Graphics Cards (MA-INF 4308)

Introduction Session

21.04.2021

PROF. SVEN BEHNKE, ANGEL VILLAR-CORRALES

Contact: villar@ais.uni-bonn.de

About Me

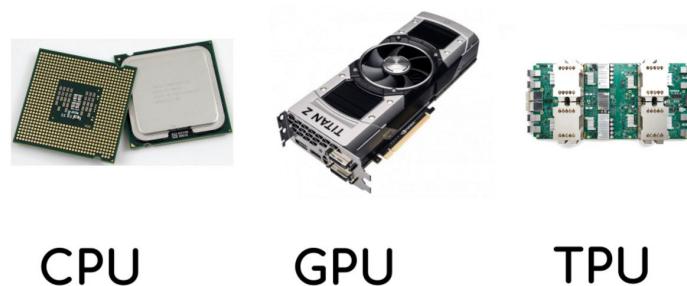
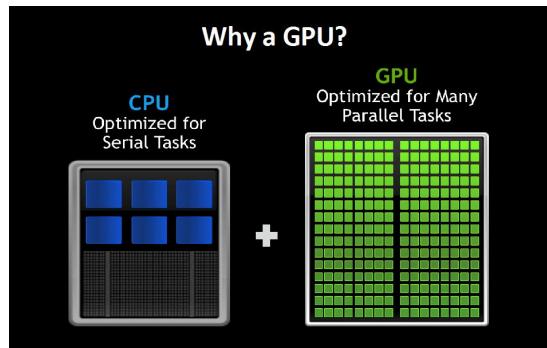
- PhD Candidate at AIS since 02.2021
- Before:
 - M. Sc. at University of Erlangen-Nuremberg
 - B. Sc. at University of Vigo (Spain)
- Research interests:
 - Deep Learning
 - Computer Vision
 - Music Information Retrieval

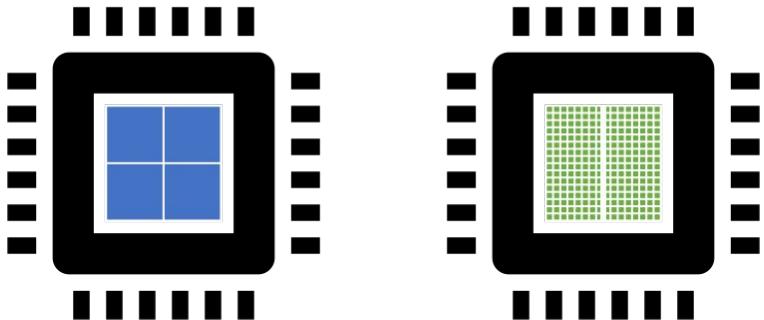


Motivation

Why Image Processing on GPUs?

- Image processing and analysis algorithms are inherently parallel:
 - Convolutions
 - Filtering
- Advancements on parallel computing devices: GPUs or TPUs
- Availability of programming interfaces: CUDA, OpenCL, ...



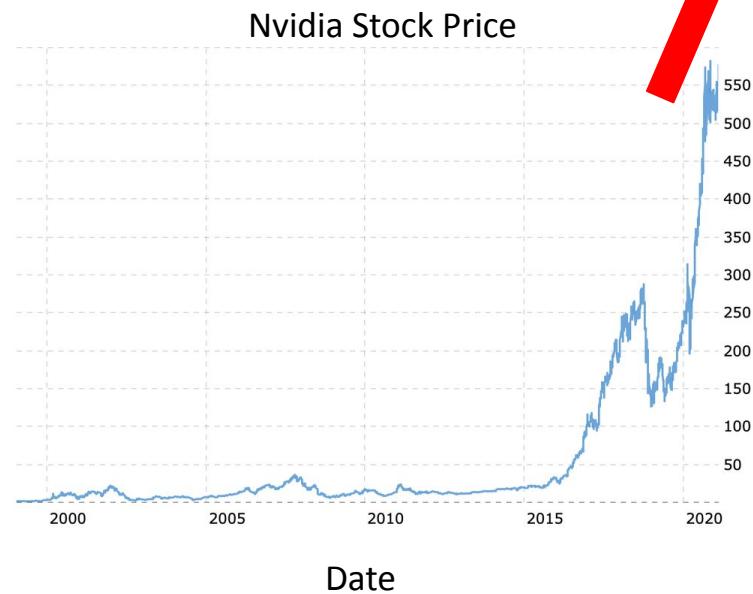
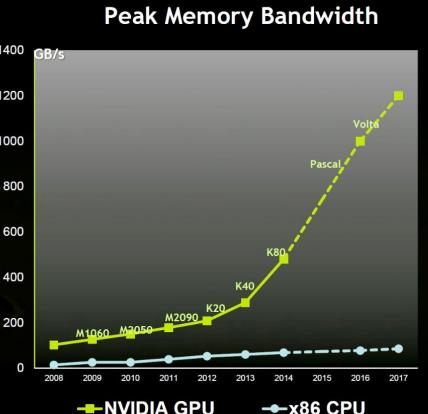
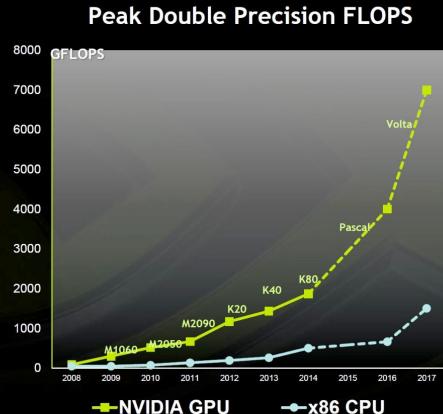


CPU	GPU
Central Processing Unit	Graphics Processing Unit
4-8 Cores	100s or 1000s of Cores
Low Latency	High Throughput
Good for Serial Processing	Good for Parallel Processing
Quickly Process Tasks That Require Interactivity	Breaks Jobs Into Separate Tasks To Process Simultaneously
Traditional Programming Are Written For CPU Sequential Execution	Requires Additional Software To Convert CPU Functions to GPU Functions for Parallel Execution

<https://towardsdatascience.com/parallel-computing-upgrade-your-data-science-with-a-gpu-bba1cc007c24>

CPU vs. GPU Performance

GPU Motivation (I): Performance Trends



7

ImageNet Challenge

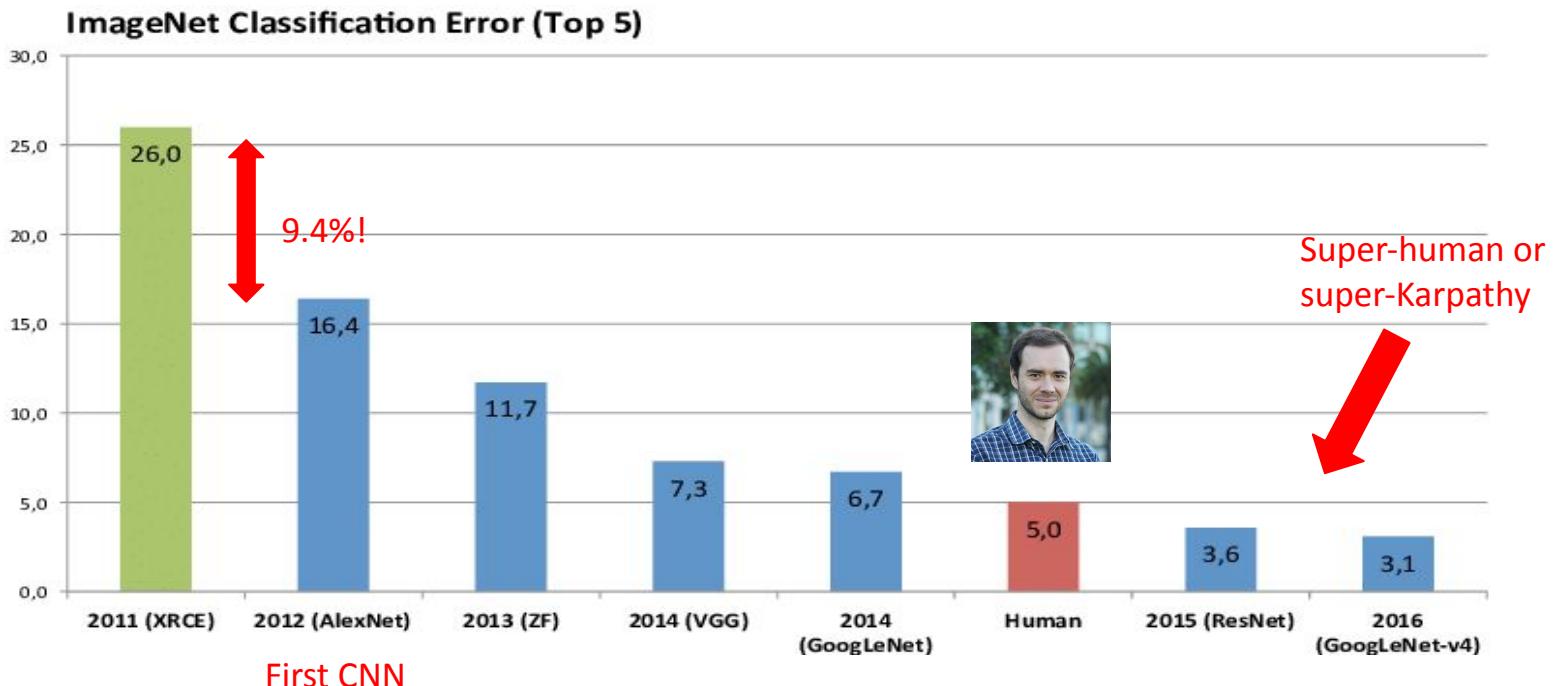
- \approx 14 million images labelled into \approx 1000 classes
- Natural images from the Internet
- Fostered research on computer vision and machine learning
- **2012:** Deep learning breakthrough by Krizhevsky et al. [1]

IMAGENET



[1]Krizhevsky, A. et al.. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*, 2012

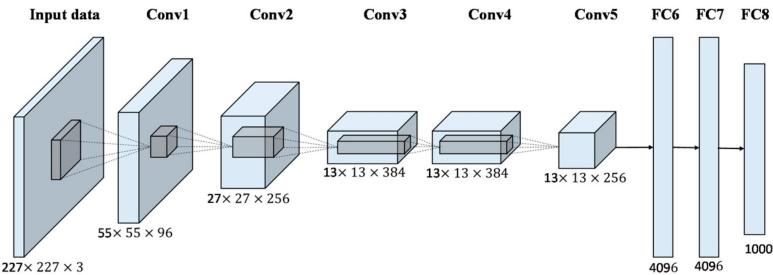
ImageNet SOTA



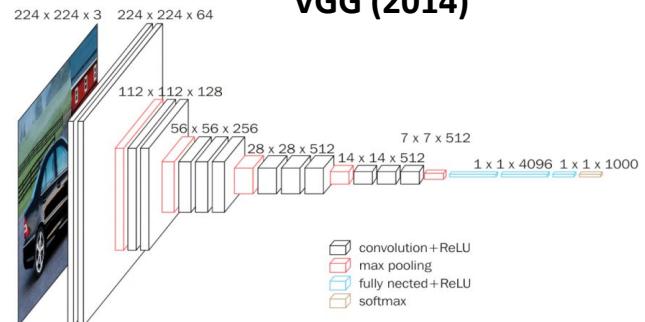
<https://devopedia.org/imagenet>

Architectures

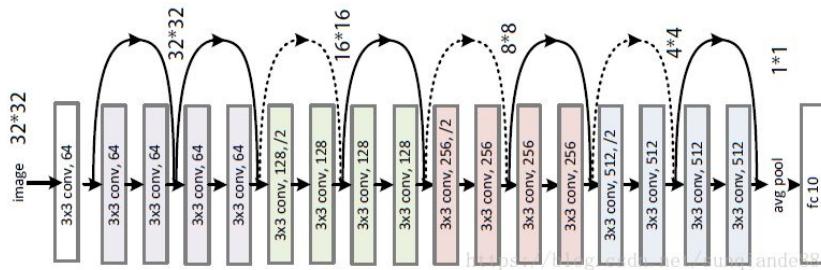
AlexNet (2012)



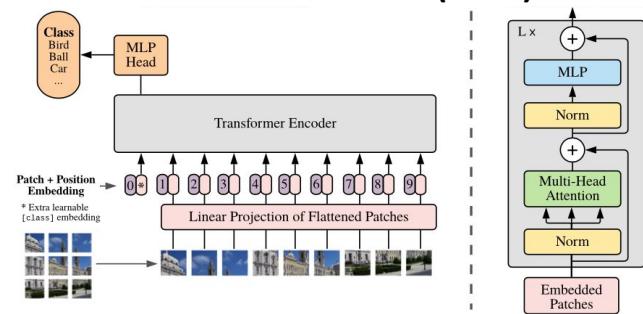
VGG (2014)



ResNet (2015)

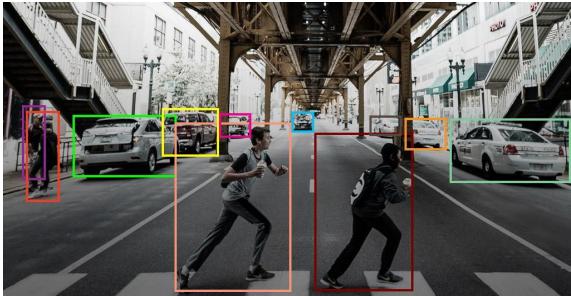


Transformers (2020)

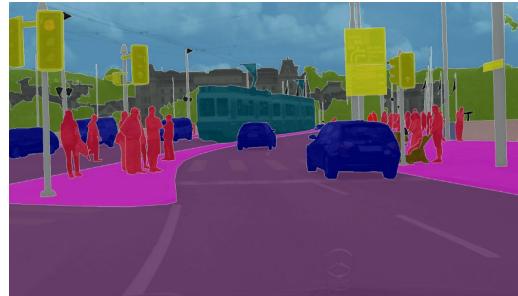


Applications in Vision

Object Detection



Semantic Segmentation



Human Pose Estimation



Image Synthesis



Deep Learning Users

Google



amazon

SIEMENS



TESLA



NETFLIX



BostonDynamics

In this Lab...

In this Lab...

- Implementing deep learning algorithms for visual pattern recognition
 - Python programming language
 - PyTorch framework
 - Deep learning basics
- 8 Lab sessions (30%) in 14 weeks
- Final project (70%) in lecture free period
 - Code/Results
 - Technical Report (6-10 pages)

Organization

- Meeting time: 2 hours a week, online meeting
 - Discuss solution to previous assignment
 - Review some theory
 - Run sample code
 - Provide next assignment
 - Questions
- Room 0.057 accessible during working hours
- Accessible GPUs (Informatik ID):
 - 6 GPU servers (cuda7 - cuda12) with GTX680 / 780 / 980
 - 4 Bigcuda (bigcuda1, 3, 4, 5) with GTX-Titan / GTX-Titan X / Tesla K20c
 - Free online resources (Google Colab/ Kaggle kernels)



Assignments

- Each session covers one topic
- Take-home assignment
 - Similar to what we do during the session
 - Due one day before follow up session
- Assignments & project can be done in pairs
 - Highly recommended!
- Send me a mail with the name of your partner or tell me if you need one

Topics Covered

1. Python & PyTorch basics
2. Autograd, Linear/Logistic regression
3. Fully Connected Neural Networks (FC or MLP)
4. Convolutional Neural Networks (CNNs)
5. Popular architectures and transfer learning
6. Autoencoders (AEs), Denoising and Variational AEs
7. Recurrent Neural Networks (RNNs & LSTM)
8. Generative Adversarial Networks (GAN)
9. Extra credit assignment (Optional)

Registration

- Register in BASIS before 30.04
- Please fill this form before 28.04 (select the time slots that work for you)

<https://forms.gle/7ABPPLDTHwWsmtBa8>

	Monday	Tuesday	Wednesday	Thursday	Friday
9-11					
11-13					
13-15					
15-17					

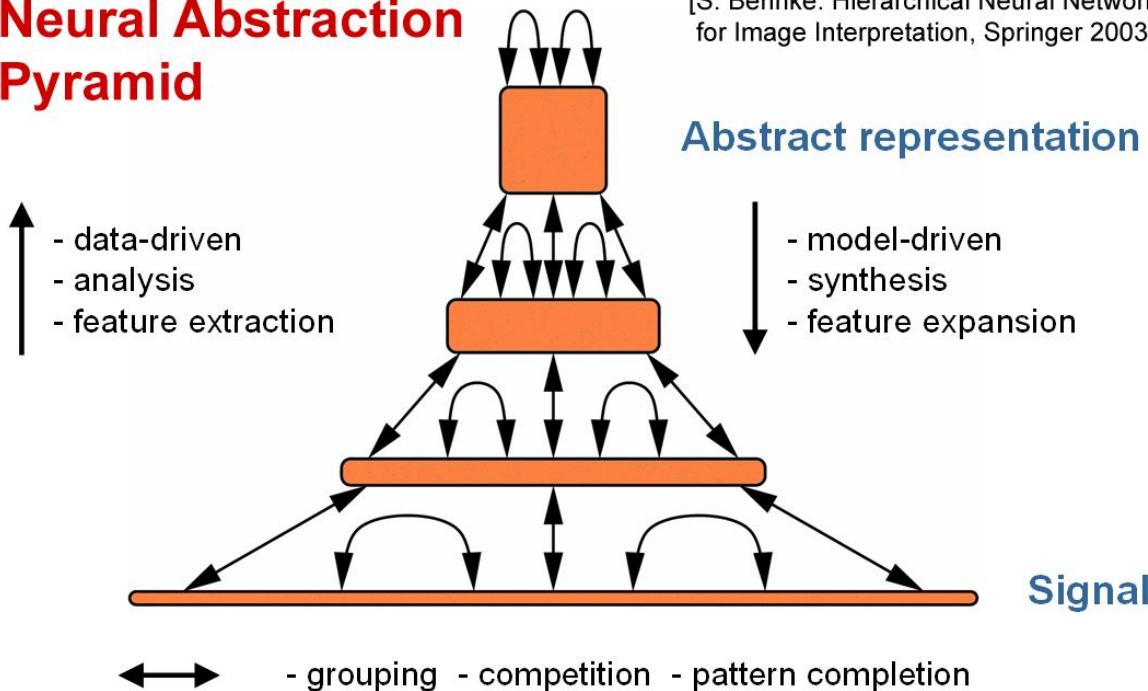
Some of our research...

Neural Abstraction Pyramid

Neural Abstraction Pyramid

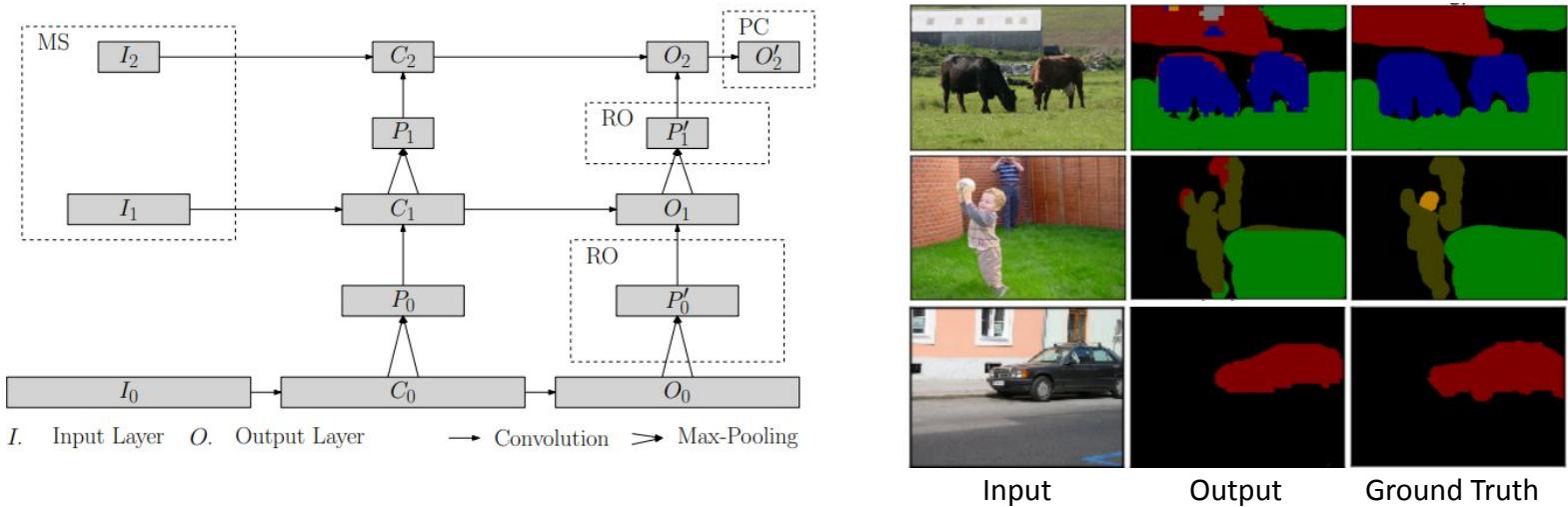
[S. Behnke: Hierarchical Neural Networks for Image Interpretation, Springer 2003]

Abstract representation



Object-class Segmentation

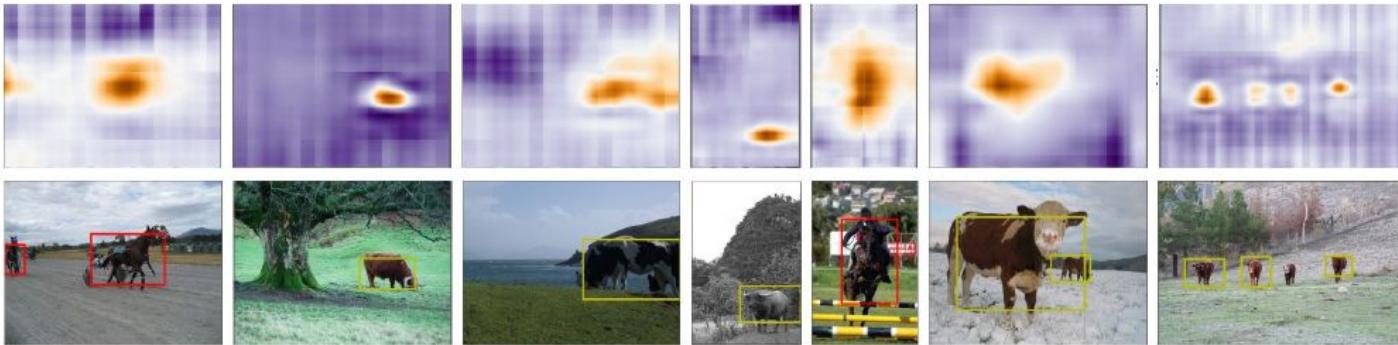
- Multi-scale CNN for RGB-pixel segmentation



[Schulz and Behnke. ESANN 2012]

Object Detection

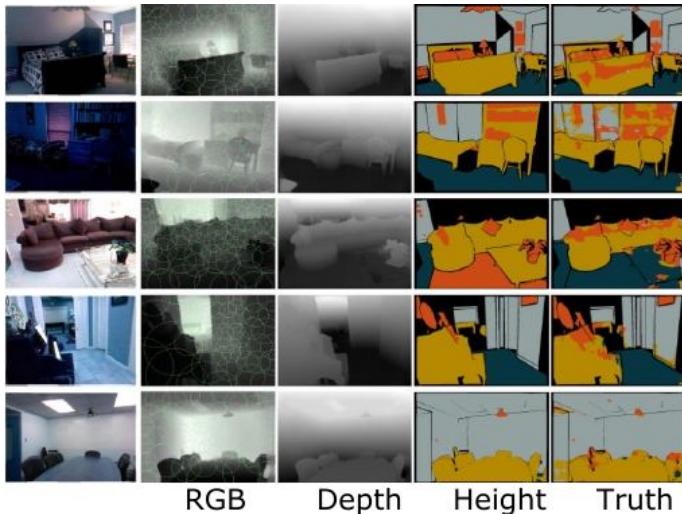
- Bounding box object detection
- Novel structured loss function to directly maximize overlap of predicted and ground truth bounding boxes
- Evaluation on two difficult classes from Pascal VOC dataset



[Schulz and Behnke. ICANN 2014]

RGB-D Object Segmentation

- Use of kinect-like sensors to obtain depth values



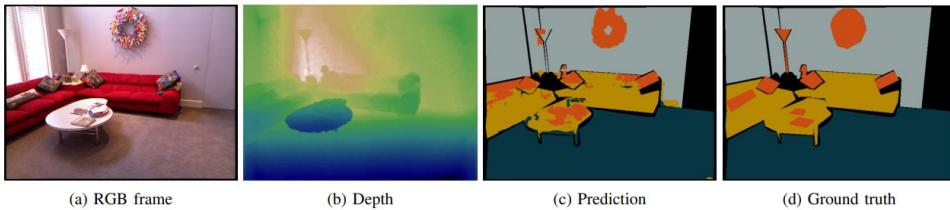
Method	floor	struct	furnit	prop	Class Avg.	Pixel Acc.
CW	84.6	70.3	58.7	52.9	66.6	65.4
CW+DN	87.7	70.8	57.0	53.6	67.3	65.5
CW+H	78.4	74.5	55.6	62.7	67.8	66.5
CW+DN+H	93.7	72.5	61.7	55.5	70.9	70.5
CW+DN+H+SP	91.8	74.1	59.4	63.4	72.2	71.9
CW+DN+H+CRF	93.5	80.2	66.4	54.9	73.7	73.4
Müller et al.[8]	94.9	78.9	71.1	42.7	71.9	72.3
Random Forest [8]	90.8	81.6	67.9	19.9	65.1	68.3
Couprise et al.[9]	87.3	86.1	45.3	35.5	63.6	64.5
Höft et al.[10]	77.9	65.4	55.9	49.9	62.3	62.0
Silberman [12]	68	59	70	42	59.7	58.6

CW is covering windows, H is height above ground, DN is depth normalized patch sizes. SP averaged within superpixels and SVM-reweighted. CRF is a conditional random field over superpixels [8]. Structure class numbers are optimized for class accuracy.

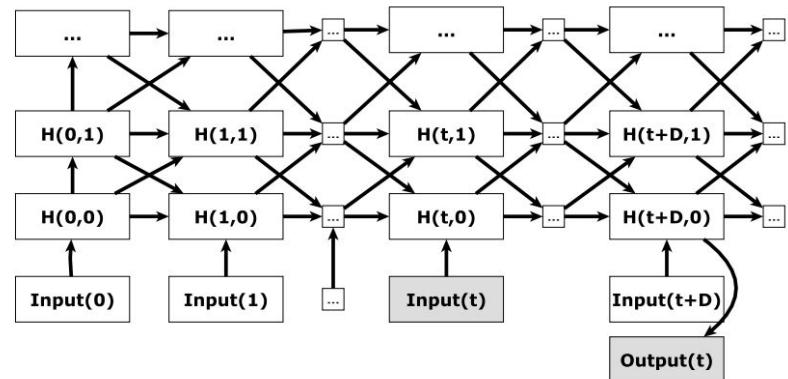
[Schulz, Höft and Behnke. ESANN 2015]

Object Segmentation from RGB-D Video

- Video processing with multi-scale Convolutional RNNs
- Iterative refinement through different time steps



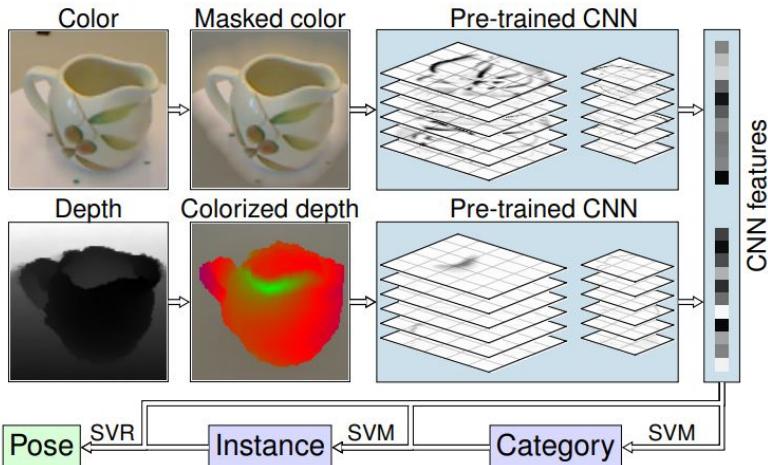
Method	Class Accuracies (%)				Average (%)	
	ground	struct	furnit	prop	Class	Pixel
Unidirectional + SW	90.0	76.3	52.1	61.2	69.9	67.5
Schulz <i>et al.</i> [20]	93.6	80.2	66.4	54.9	73.7	73.4
Müller and Behnke [22]	94.9	78.9	79.7	55.1	71.9	72.3
Stückler <i>et al.</i> [21]	90.8	81.6	67.9	19.9	65.0	68.3
Couprise <i>et al.</i> [23]	87.3	86.1	45.3	35.5	63.5	64.5
Höft <i>et al.</i> [19]	77.9	65.4	55.9	49.9	61.1	62.0
Silberman <i>et al.</i> [17]	68	59	70	42	59.6	58.6



[Pavel, Schulz, and Behnke. IJCNN 2015,
Neural Networks 2017]

Computer Vision with Pretrained Features

- Object recognition and pose estimation
- Pretrained features from ImageNet
- Improved classification and estimation performance



Evaluation on the Washington RGB-D Objects dataset

Method	Category Accuracy (%)		Instance Accuracy (%)	
	RGB	RGB-D	RGB	RGB-D
Lai <i>et al.</i> [12]	74.3 ± 3.3	81.9 ± 2.8	59.3	73.9
Bo <i>et al.</i> [14]	82.4 ± 3.1	87.5 ± 2.9	92.1	92.8
PHOW[18]	80.2 ± 1.8	—	62.8	—
Ours	83.1 ± 2.0	89.4 ± 1.3	92.0	94.1

[Schwarz, Schulz and Behnke. ICRA 2015]

Amazon Bin-Picking Challenge

- Picking a large variety of objects
- Placing them on a shelf or packing boxes
- NimbRo team came in 2nd
- Computer vision challenge



[Schwarz et al. ICRA 2017]

Object Capture and Scene Synthesys

- Capture data with a turn table
- Rendered realistic scenes



Object Detection and Segmentation

- RefineNet architecture [1]
- Trained on rendered data



[1] Lin et al. CVPR 2016

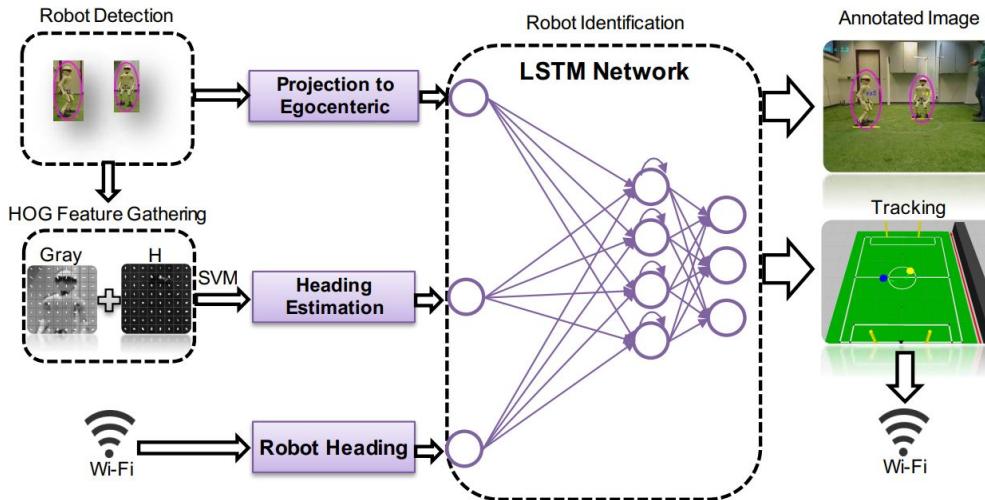
Soccer Robots

- NimbRo participates in humanoid soccer robot competitions (RoboCup)
- Challenging perception scenario



Robot Tracking and Identification

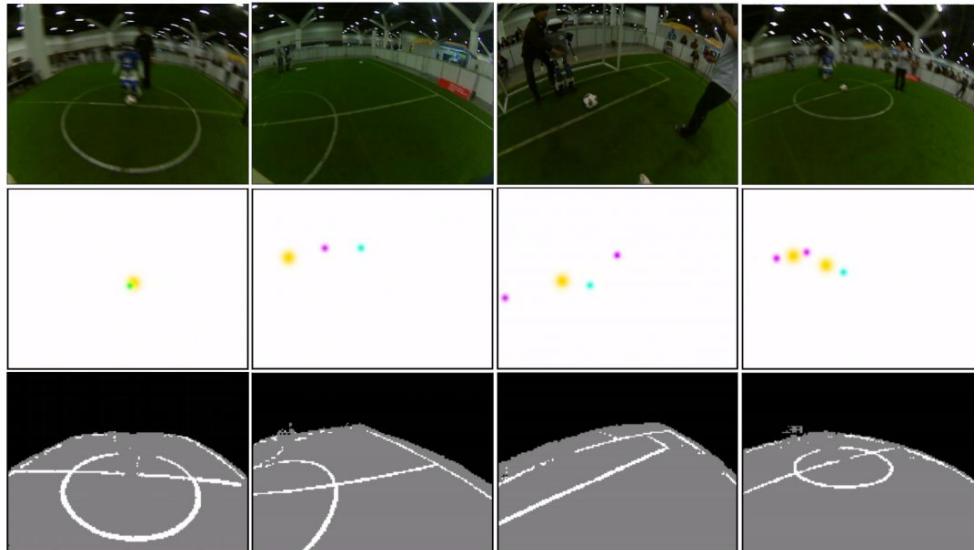
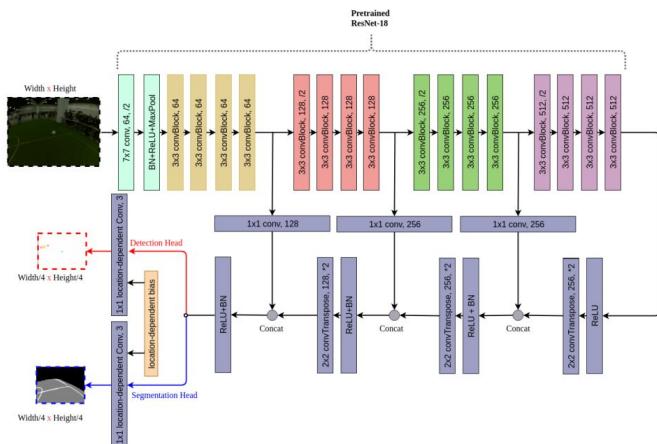
- Real time robot tracking
- Scalable and able to handle occlusions



[Farazi and Behnke. IROS 2017]

Scene Understanding

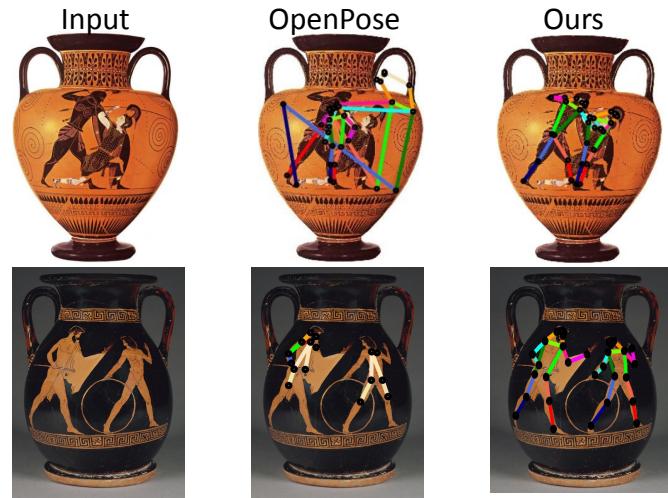
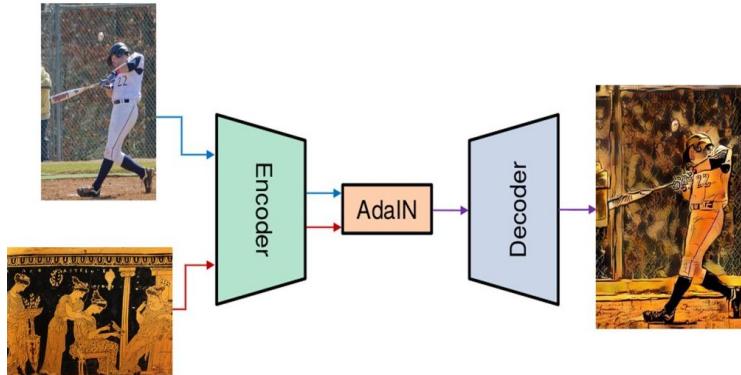
- End-to-end convolutional model
- Robot and ball detection
- Soccer field segmentation



[Rodriguez et al. RoboCup 2019]

Style Transfer Learning

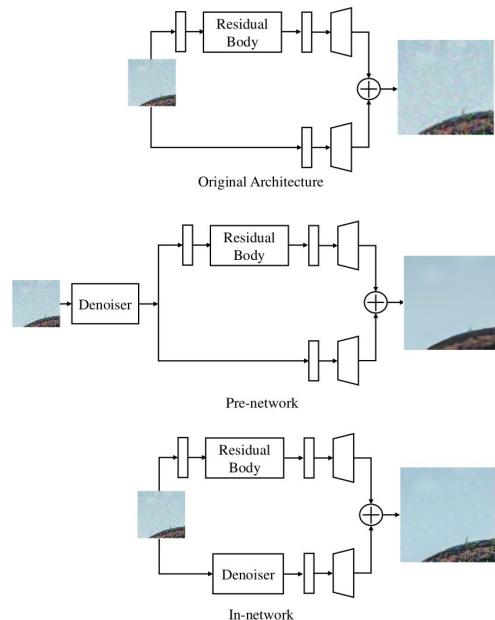
- Person detection and human pose estimation through style transfer learning
- Two stage pipeline to learn in challenging domains with few annotations



[Villar-Corrales, Mahdu, Kosti, et al. 2020]

Denoising and Super-Resolution

- Architectural designs for joined denoising and super-resolution



[Villar-Corrales, Schirrmacher and Riess. ICASSP 2021]

And much more...

UAV Perception



3D Deep Learning

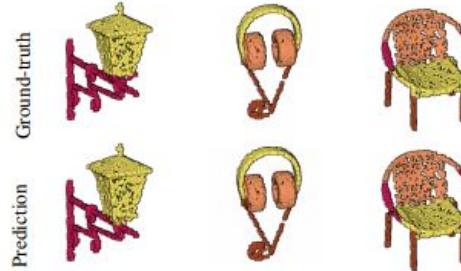
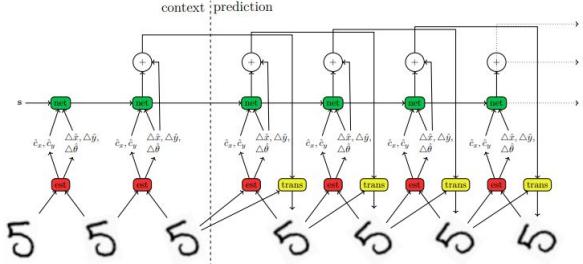


Fig. 8: ShapeNet [39] results of our method.

Signal Processing and Deep Learning



Scene Synthesis



Once Again...

Registration

- Register in BASIS before 30.04
- Please fill this form before 28.04 (select the time slots that work for you)

<https://forms.gle/7ABPPLDTHwWsmtBa8>

	Monday	Tuesday	Wednesday	Thursday	Friday
9-11					
11-13					
13-15					
15-17					

Questions?

