Lab: CudaVision – Learning Vision Systems on Graphics Cards (MA-INF 4308)

Introduction

1.6.2016

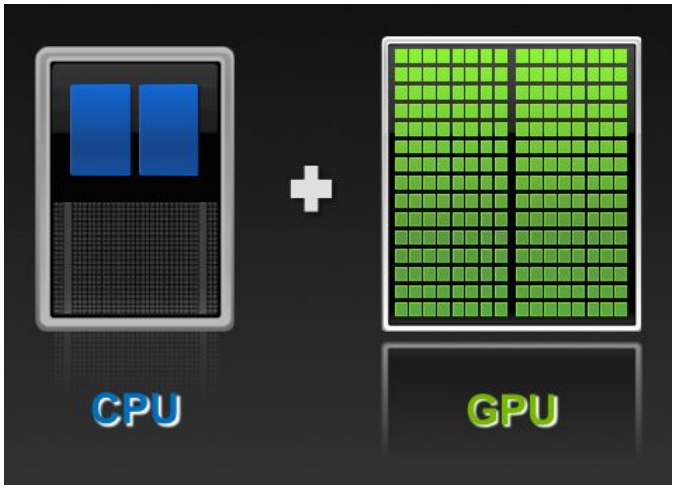
Prof. Sven Behnke
Dr. Seongyong Koo

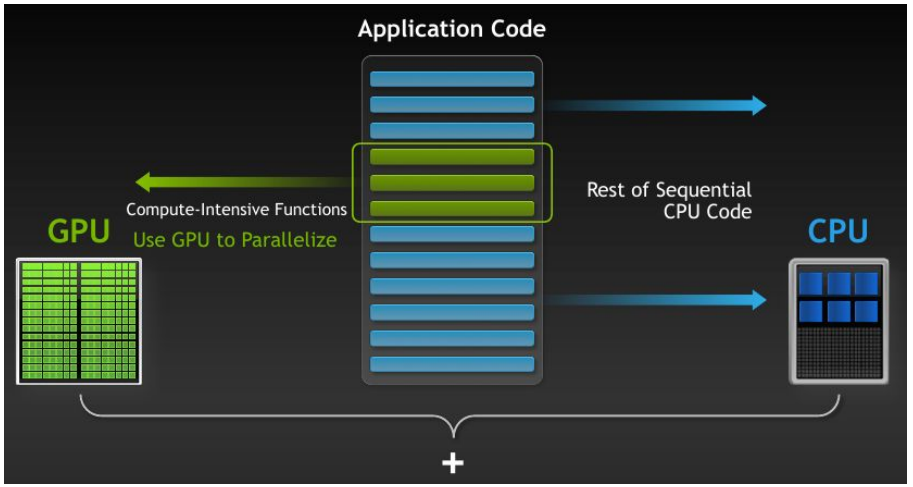
universitätbonn ais

Background:
Image processing on GPUs

Algorithms for the analysis of images mostly work independently on different regions of an image. These algorithms are therefore inherently **parallel** and can greatly profit from **parallel hardware (GPUs)**.

Due to the availability of general purpose programming interfaces like **CUDA**, the immense speed of graphics cards can be put to work for a multitude of parallel tasks.





Background:
ImageNet and Deep learning

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images (currently, an average of over five hundred images per node)

[http://www.image-net.org/]

[https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures]

Background: ImageNet and Deep learning

Large Scale Visual Recognition Challenge (ILSVRC) evaluates algorithms for object detection and image classification at large scale with ImageNet.

- A object detection challenge on fully labeled data for 200 categories of objects
- An image classification plus object localization challenge with 1000 categories.

GT: horse cart
1: horse cart
2: minibus
3: oxcart
4: stretcher
5: half track

GT: birdhouse
1: birdhouse
2: sliding door
3: window screen
4: mailbox
5: pot

GT: forklift
1: forklift
2: garbage truck
3: tow truck
4: trailer truck
5: go-kart

GT: coucal
1: coucal
2: indigo bunting
3: lorikeet
4: walking stick
5: custard apple

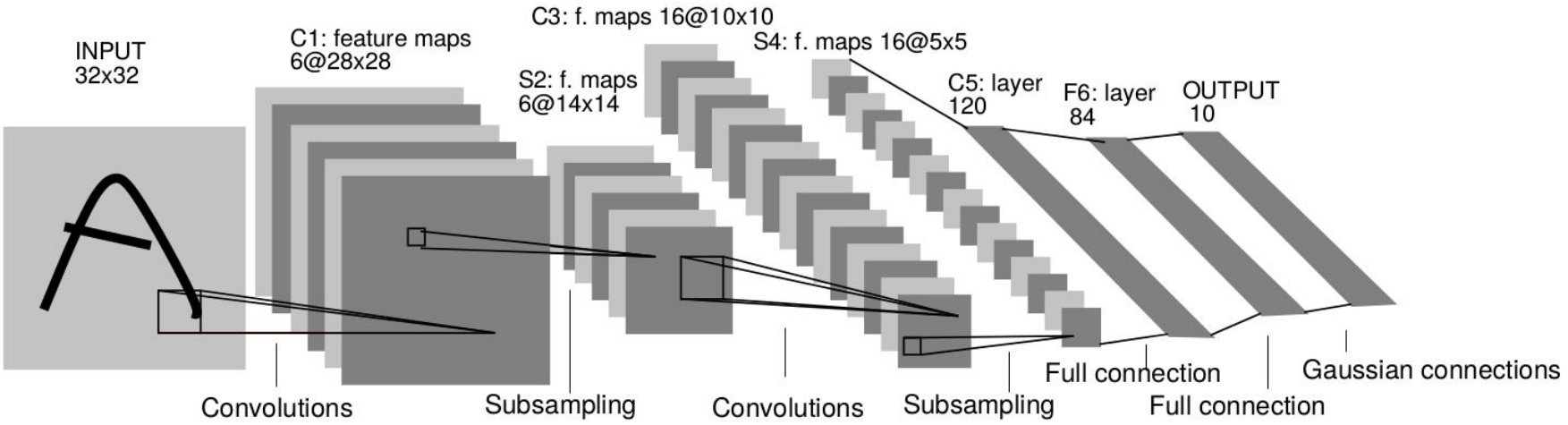
GT: komondor
1: komondor
2: patio
3: llama
4: mobile home
5: Old English sheepdog

GT: yellow lady's slipper
1: yellow lady's slipper
2: slug
3: hen-of-the-woods
4: stinkhorn
5: coral fungus

[He et al. 2015]
Delving Deep into Rectifiers:Surpassing Human-Level Performance on ImageNet Classification

Background:
ImageNet and Deep learning

From 2012, **Deep Convolutional Neural Network** significantly improved upon the best performance in the multiple image databases including ImageNet.



Architecture of LeNet-5, a Convolutional Neural Network for digits recognition

[LeCun et al. 1998]
Gradient-based learning applied to document recognition

In this lab,

We learn how to implement deep learning algorithms from the area of visual pattern recognition using **TensorFlow** library with python.

1. (1-3 June) TensorFlow basic operations, MNIST/custom dataset
2. (6-10 June) Linear regression / Logistic regression
3. (13-17 June) Multi-layer Perceptron (MLP)
4. (20-24 June) Convolutional Neural Network (CNN)

5. (27 June - 1 July) Fine-tuning a pre-trained model (VGG)
6. (4-8 July) Denoising Convolutional Autoencoder
7. (11-15 July) Recurrent Neural Network (LSTM model)
8. (18-22 July) Sequence to sequence model (seq2seq)

Final assignment with real-world dataset
(In lectuer-free time, 1st August - 30th September)

[https://www.tensorflow.org/]

In this lab,

Weekly assignment

- Issued on every Monday
- Each topic with reading material and a simple python code
- Follow the example code first and write your own code for more complicated assignment
- Officially 2 hours a week in Room I.28
- Anytime accessible

CudaLab

- 4 Cuda (cuda8 - cuda11)
  - GTX680, GTX980, GTX980, GTX780
- 3 Bigcuda (bigcuda1, bigcuda3, bigcuda4)
  - Tesla K20c, GTX titan, GTX titan X
- Students can login with Informatik ID

In this lab,

Grading

- 30% weekly assignments
- 70% report for the final assignment
  (7-9 pages, technical report format)

Official lab time

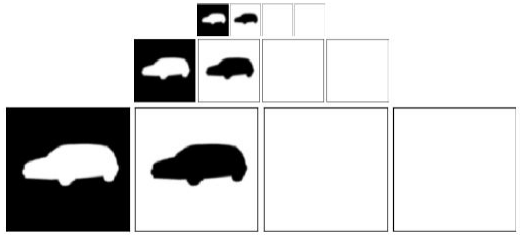
- From 6th June to 22nd July

	Mon	Tue	Wed	Thr	Fri
9-11					
13-15	O				
15-17					

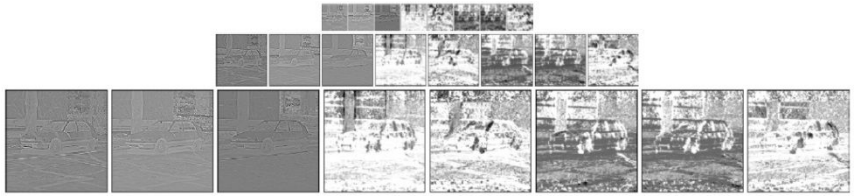
✦ Two weeks (4th-15th July) will be called off

Deep learning research in AIS,
Object-class Segmentation

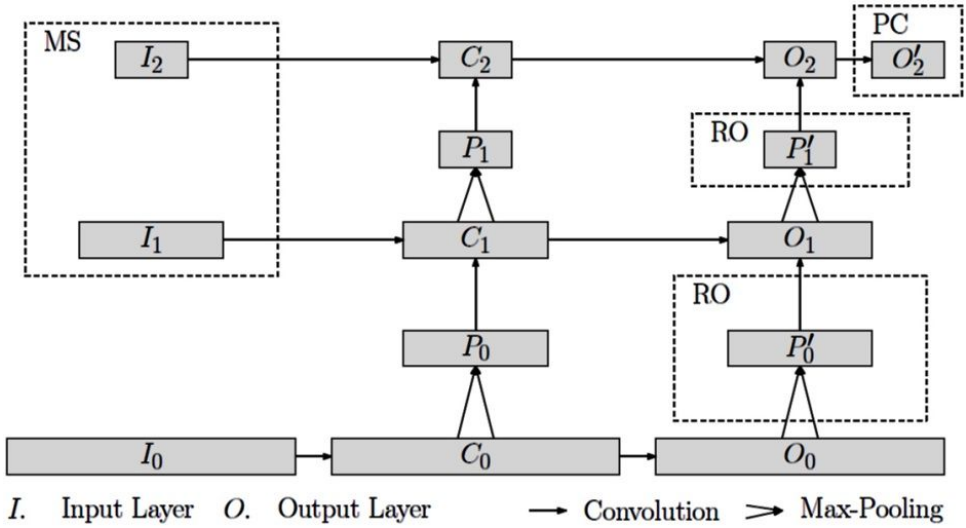
Class annotation per pixel



Multi-scale input channels



[Schulz, Behnke 2012]



Evaluated on MSRC-9/21
and INRIA Graz-02 data sets

Input Output Truth

Input Output Truth

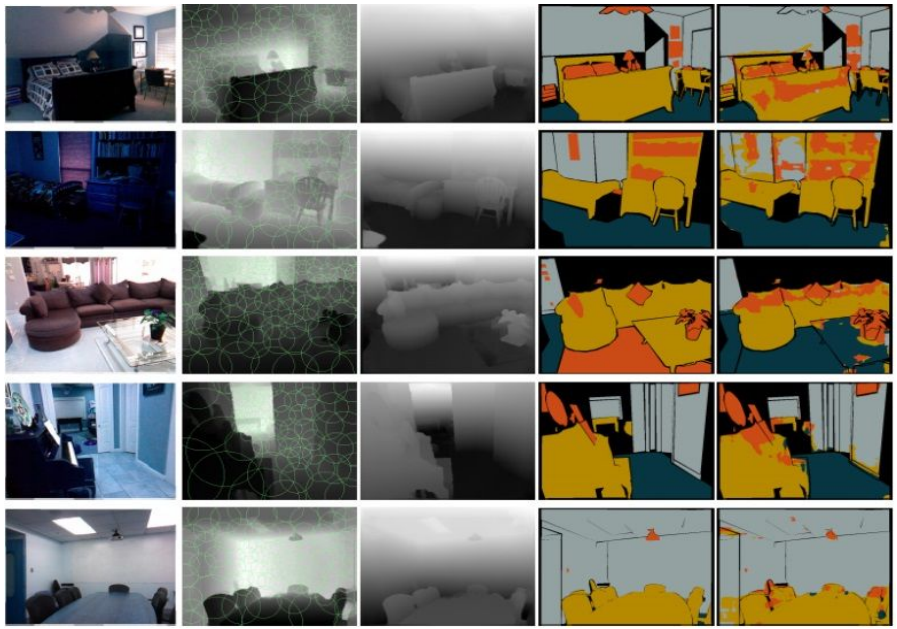
Object Detection in Images

- Bounding box annotation
- Structured loss that directly maximizes overlap of the prediction with ground truth bounding boxes
- Evaluated on two of the Pascal VOC 2007 classes

[Schulz, Behnke, ICANN 2014]

RGB-D Object-Class Segmentation

- Kinect-like sensors provide dense depth
- Scale input according to depth, compute pixel height



RGB Depth Height Truth Output

NYU Depth V2

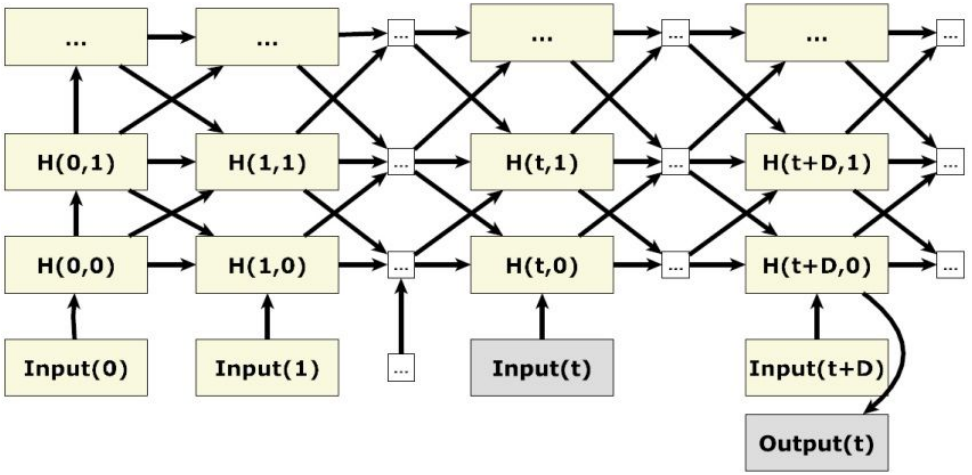
Method	floor	struct	furnit	prop	Class Avg.	Pixel Acc.
CW	84.6	70.3	58.7	52.9	66.6	65.4
CW+DN	87.7	70.8	57.0	53.6	67.3	65.5
CW+H	78.4	74.5	55.6	62.7	67.8	66.5
CW+DN+H	93.7	72.5	61.7	55.5	70.9	70.5
CW+DN+H+SP	91.8	74.1	59.4	63.4	72.2	71.9
CW+DN+H+CRF	93.5	80.2	66.4	54.9	**73.7**	**73.4**
Müller et al.[8]	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
Couprie 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

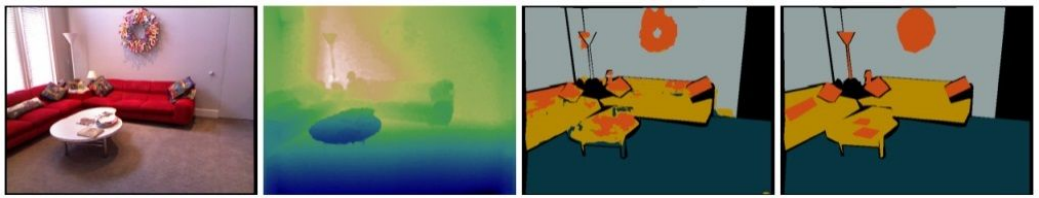
CW is covering windows, H is height above ground, DN is depth normalized patch sizes. SP is averaged within superpixels and SVM-reweighted. CRF is a conditional random field over superpixels [8]. Structure class numbers are optimized for class accuracy.

[Schulz, Höft, Behnke, ESANN 2015]

Neural Abstraction Pyramid for RGB-D Video Object-class Segmentation

- NYU Depth V2 contains RGB-D video sequences
- Recursive computation is efficient for temporal integration





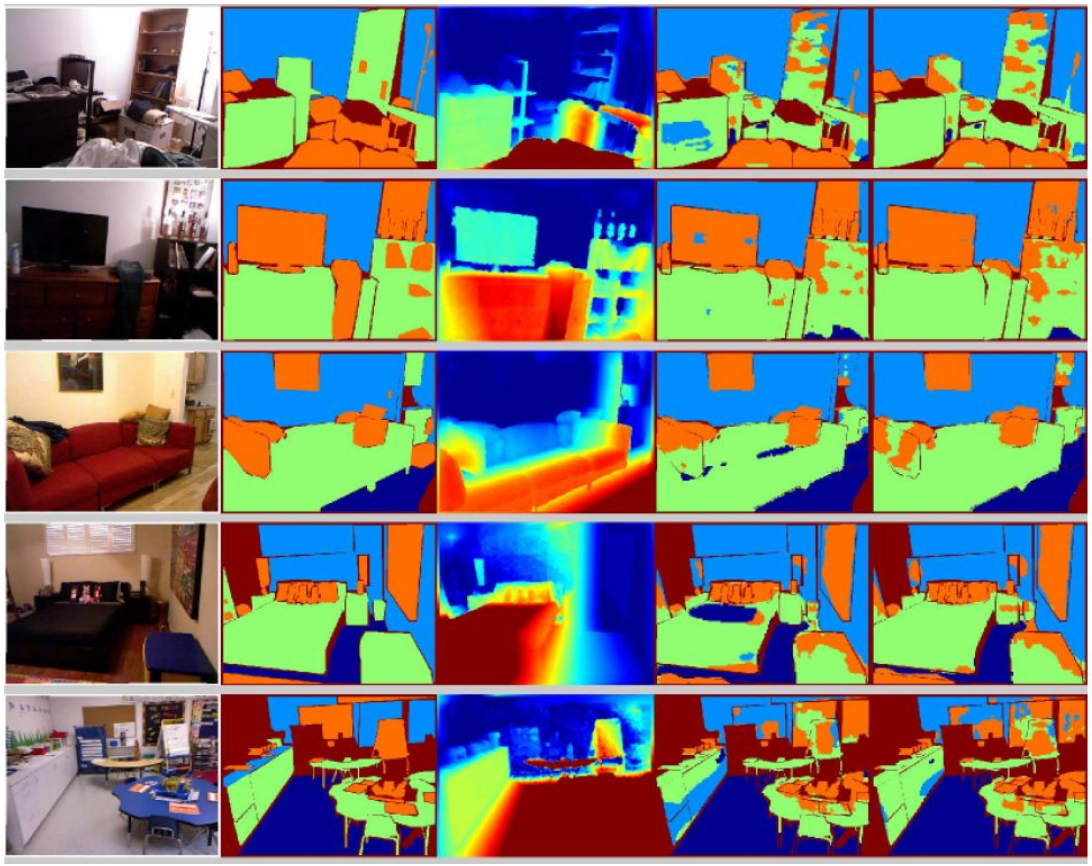
RGB Depth Output Truth

Method	Class Accuracies (%) ground	struct	furnit	prop	Average (%) Class	Pixel
Höft *et al.* [19]	77.9	65.4	55.9	49.9	62.0	61.1
Unidirectional + MS	73.4	66.8	**60.3**	49.2	62.4	63.1
Schulz *et al.* [20] (no height)	87.7	70.8	57.0	53.6	67.3	65.5
Unidirectional + SW	**90.0**	**76.3**	52.1	**61.2**	**69.9**	**67.5**

[Pavel, Schulz, Behnke, IJCNN 2015]

Geometric and Semantic Features for RGB-D Object-class Segmentation

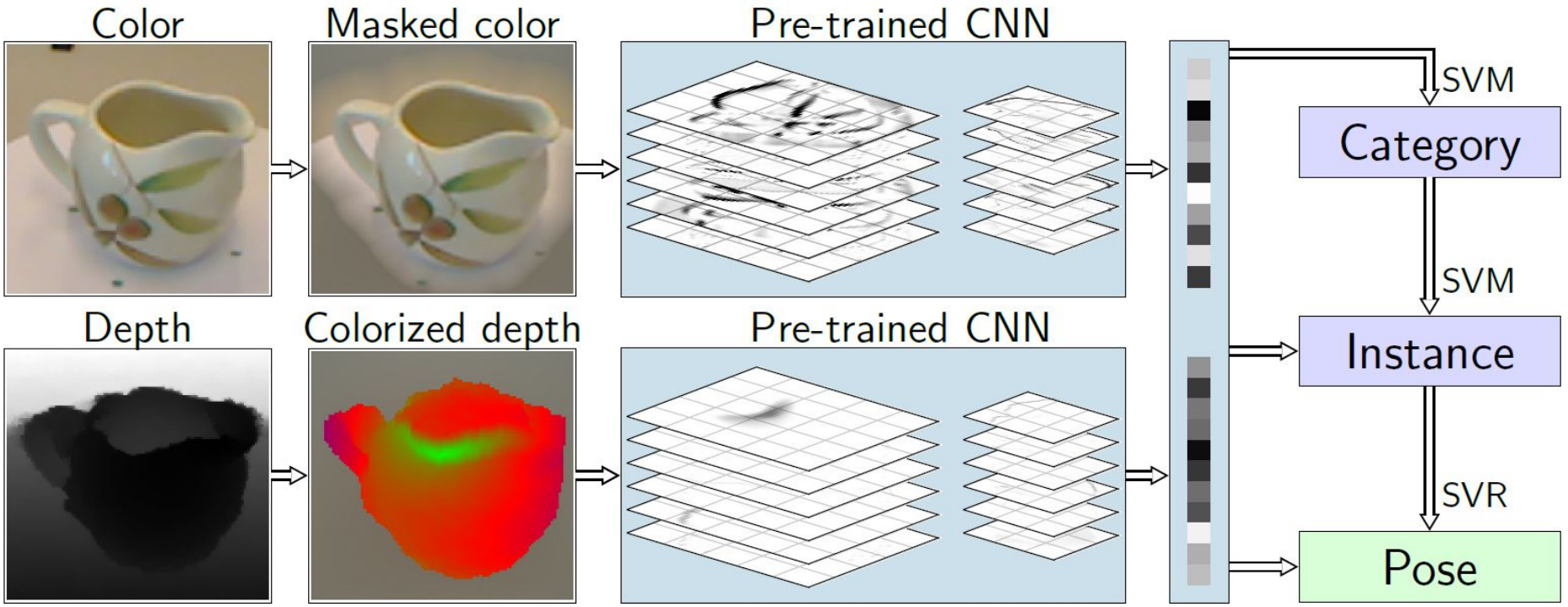
- New **geometric** feature: distance from wall
- **Semantic** features pretrained from ImageNet
- Both help significantly



[Husain et al. ICRA2016, RA-L]

RGB-D Object Recognition and Pose Estimation

- Use pretrained features from ImageNet



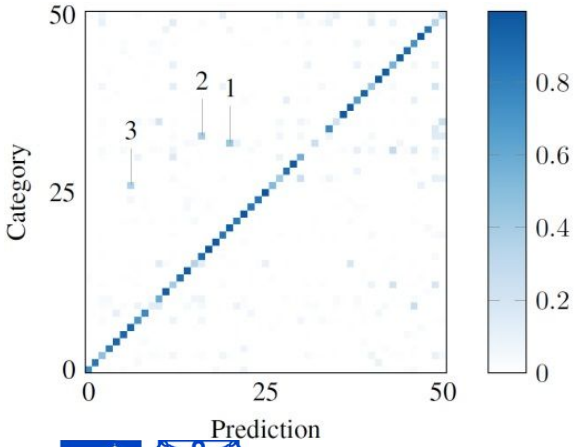
[Schwarz, Schulz, Behnke, ICRA2015]

Recognition Accuracy

- Improved both category and instance recognition

Method	Category Accuracy (%)		Instance Accuracy (%)	
	RGB	RGB-D	RGB	RGB-D
Lai *et al.* [1]	74.3 ± 3.3	81.9 ± 2.8	59.3	73.9
Bo *et al.* [2]	82.4 ± 3.1	87.5 ± 2.9	**92.1**	92.8
PHOW[3]	80.2 ± 1.8	—	62.8	—
Ours	$\mathbf{83.1 \pm 2.0}$	88.3 ± 1.5	92.0	**94.1**
Ours	$\mathbf{83.1 \pm 2.0}$	$\mathbf{89.4 \pm 1.3}$	92.0	**94.1**

- Confusion matrix



1: pitcher / coffe mug

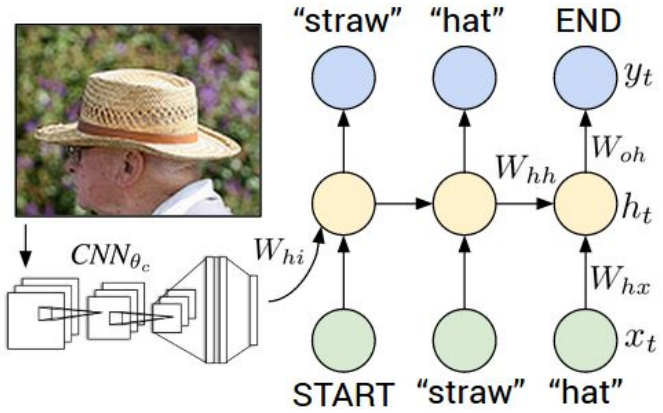
2: peach / sponge

[Schwarz, Schulz, Behnke, ICRA2015]

Deep learning research from other labs,
Generating Image Captions

- Multimodal recurrent neural network generative model

[Karpathy, Fei-Fei 2015]

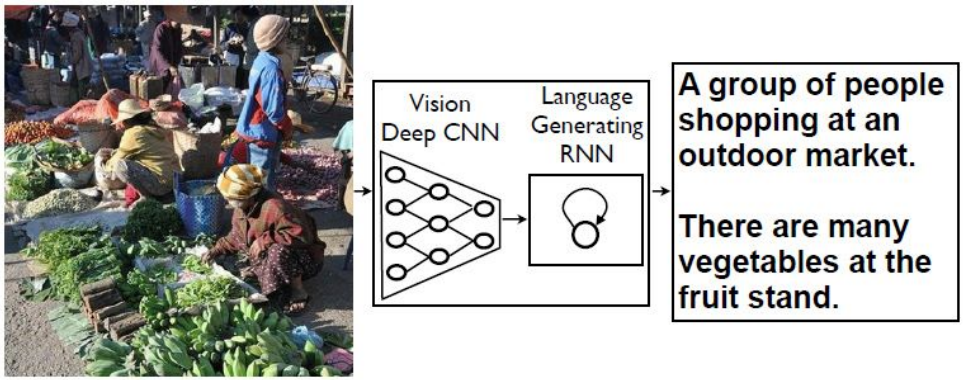


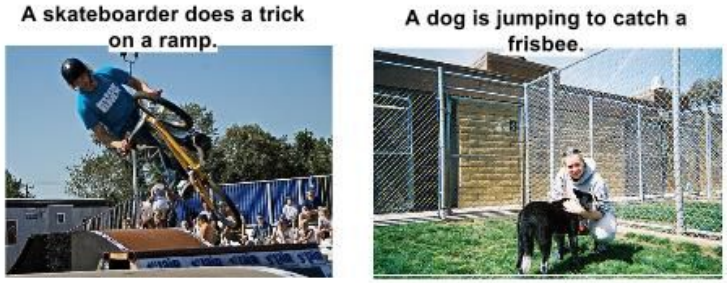
man in black shirt is playing guitar.

construction worker in orange safety vest is working on road.

two young girls are playing with lego toy.

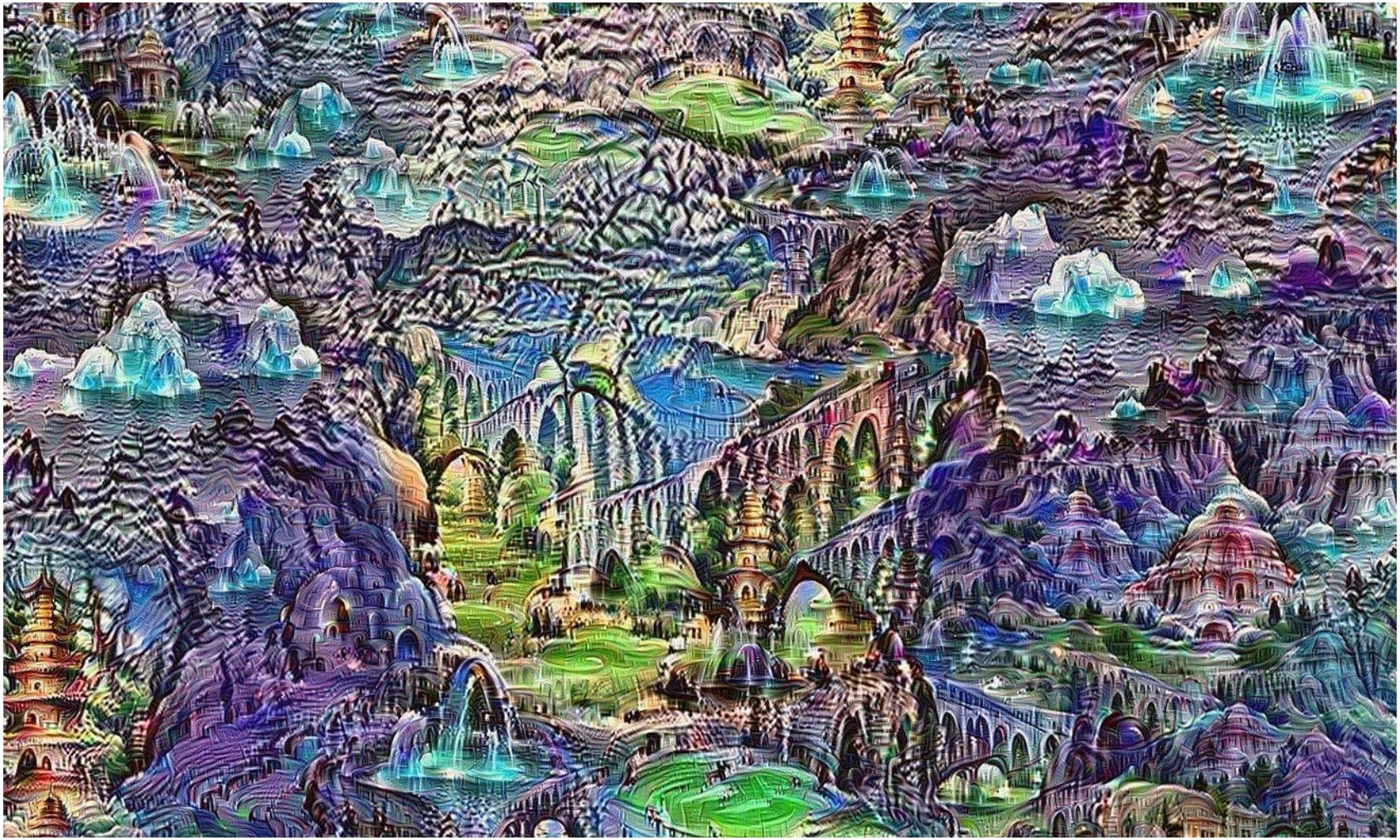
Generating Image Captions





[Vinyals et al. 2015]

Dreaming Deep Networks



[Mordvintsev et al 2015]

Painting Style Transfer

Original

Turner

van Gogh

Munch

[Gatys et al. 2015]