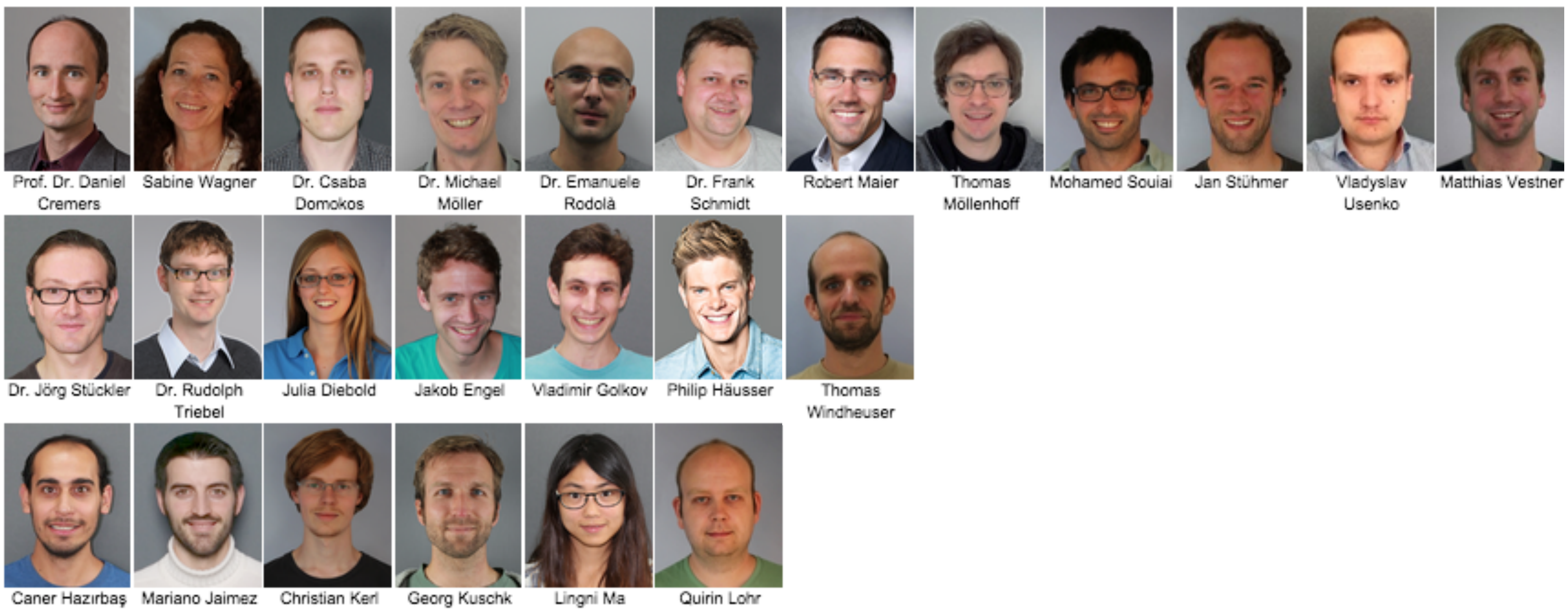
Deep Learning in Computer Vision

Caner Hazırbaş

Deep Learning in Action
24. June ’15

Computer Vision Group



6 Postdocs, 16 PhD students

Research in Computer Vision

Image-based 3D Reconstruction

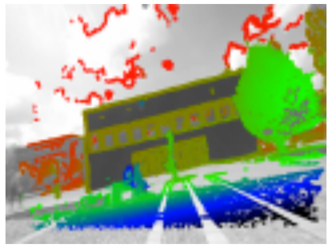
Shape Analysis

Robot Vision

RGB-D Vision

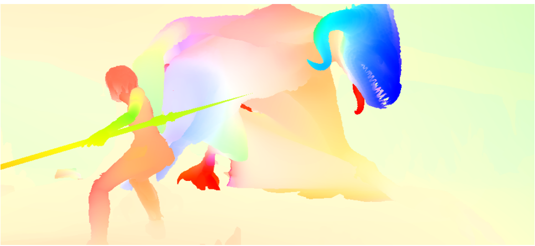
Image Segmentation

Visual SLAM



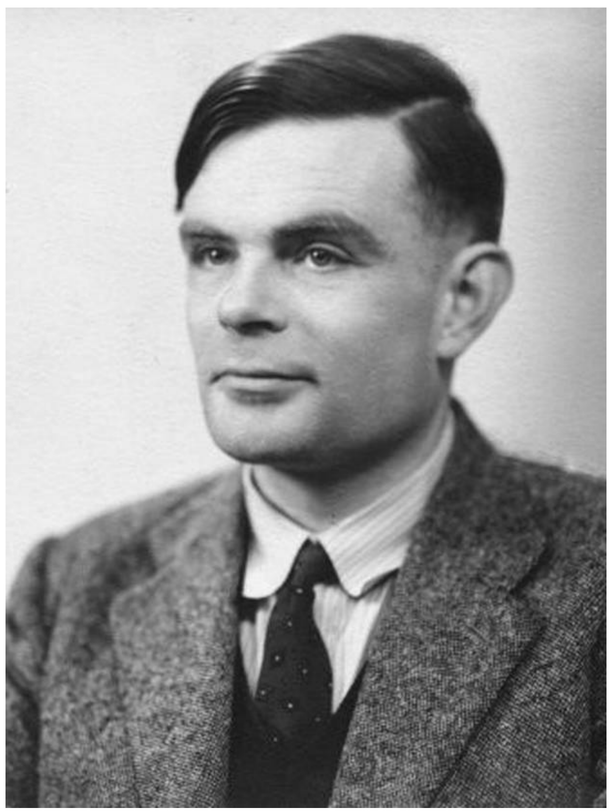
Convex Relaxation Methods

Optical Flow



Deep Learning
in Computer Vision

How to teach a machine ?



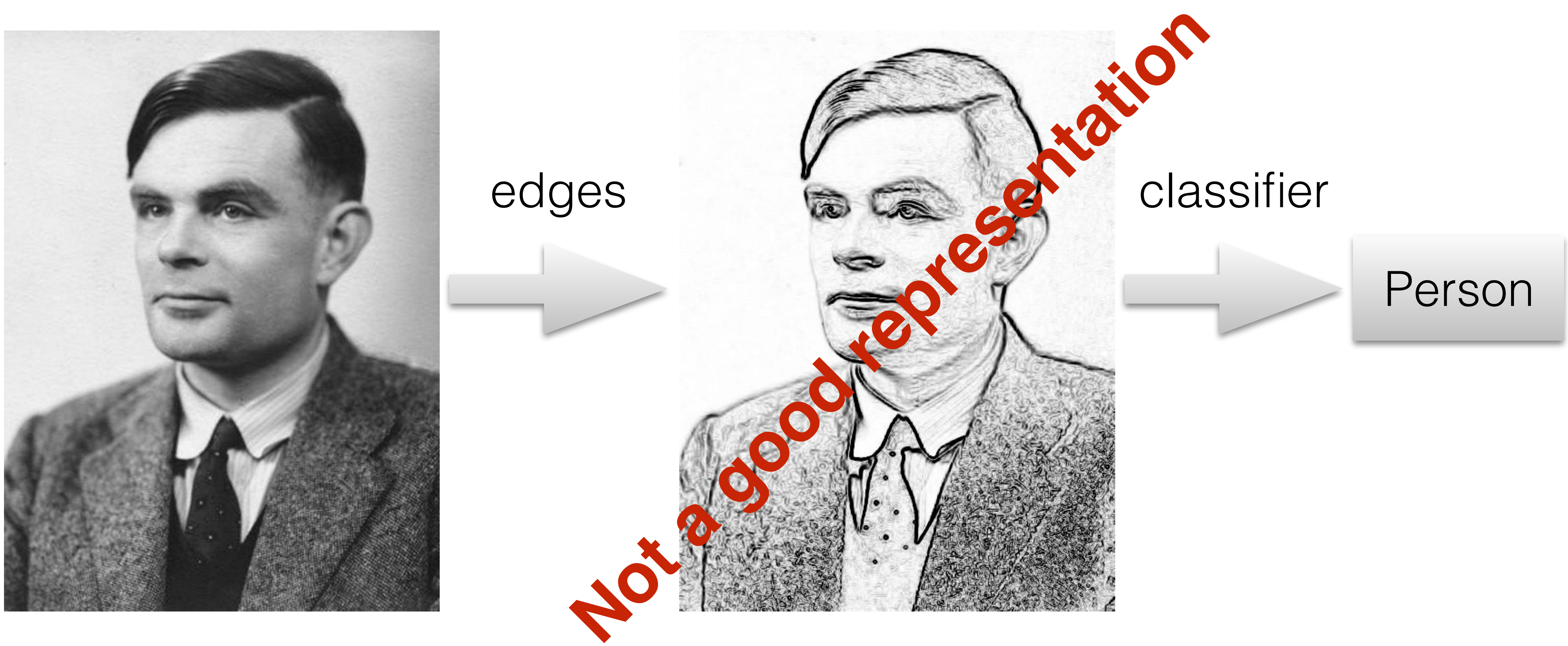
edges

classifier

Person

(or any other **hand-crafted** features)

How to teach a machine ?



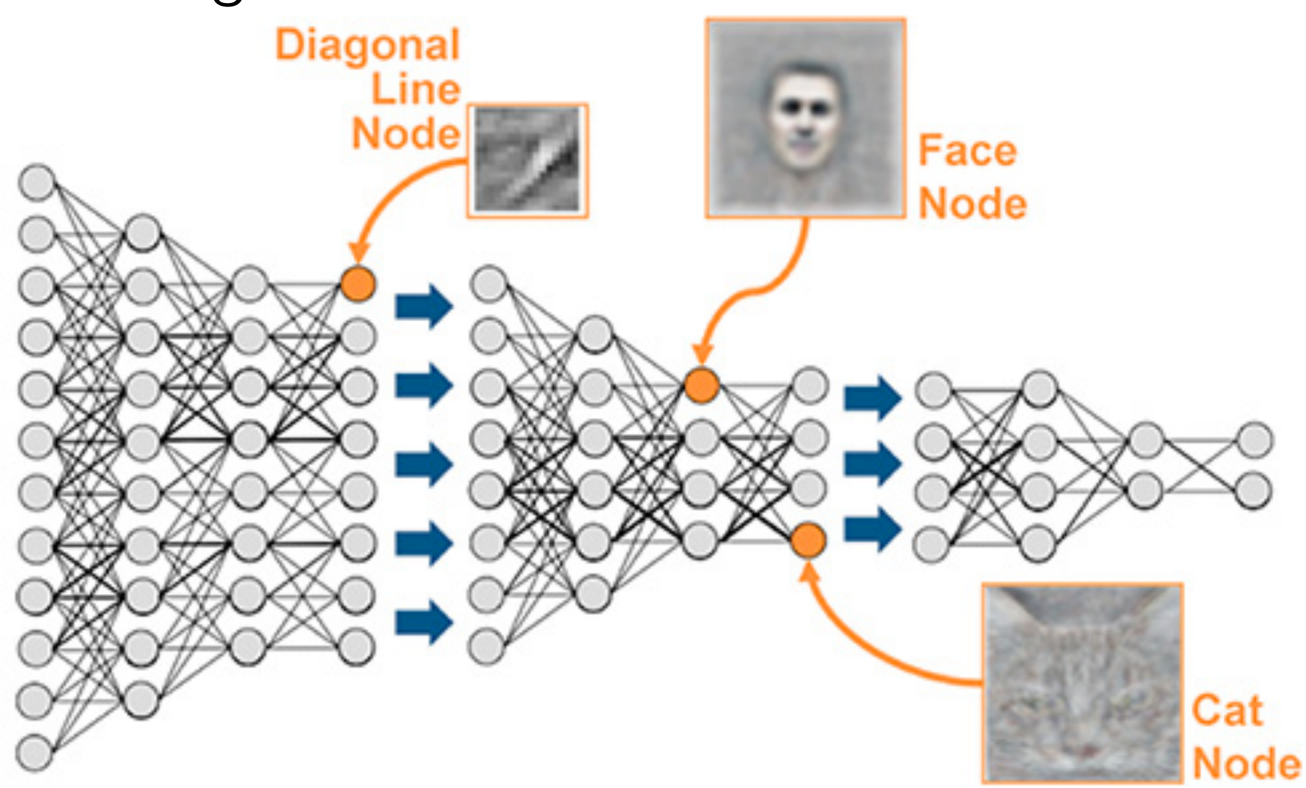
(or any other **hand-crafted** features)

What is deep learning ?

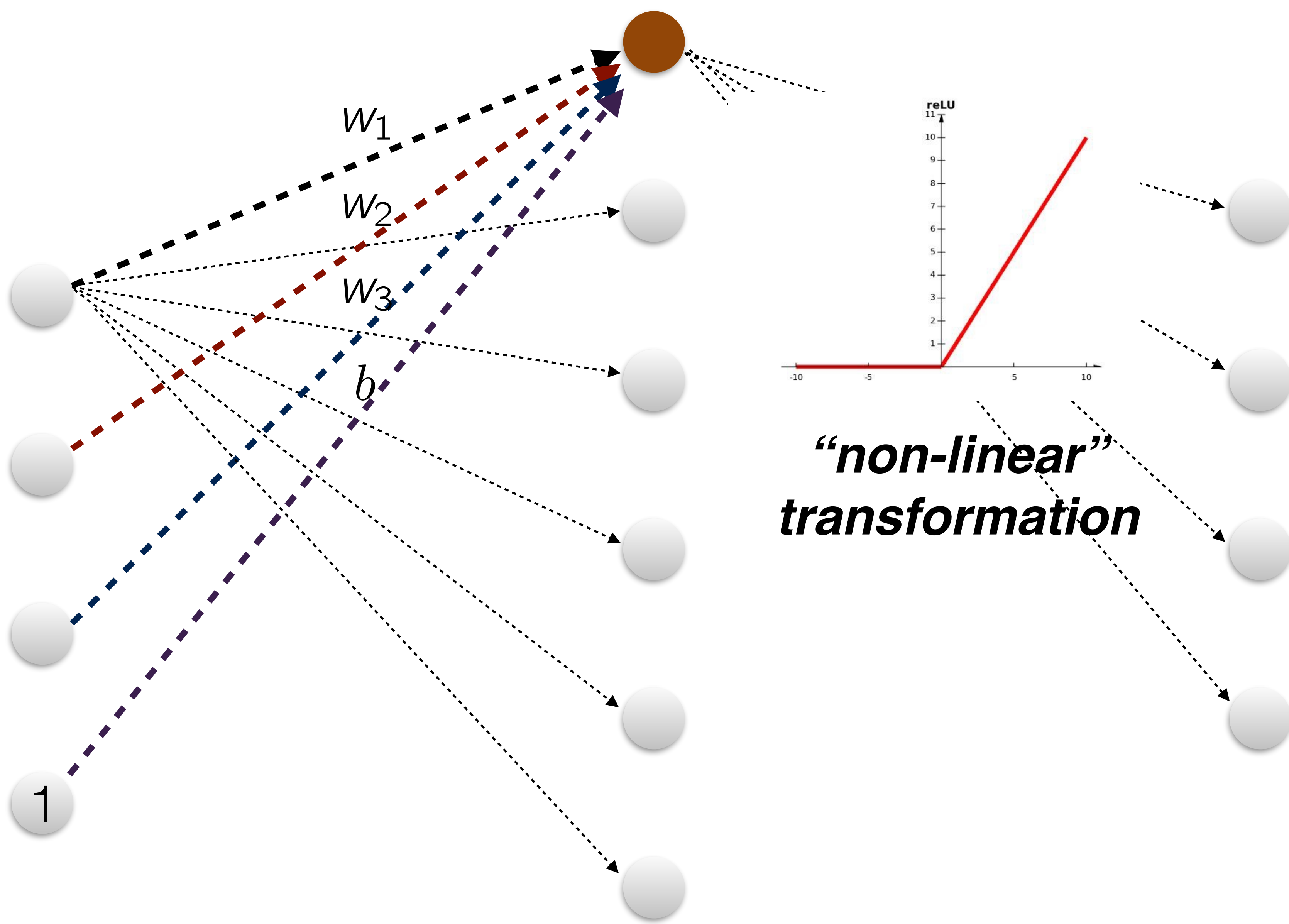
- Representation learning method
  Learning good features automatically from raw data

- Learning representations of data with multiple levels of abstraction

Google's cat detection neural network



Construction of higher levels of abstraction



Going deeper in the network

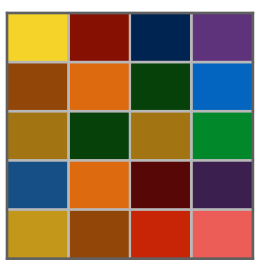
Input
'Pixels'

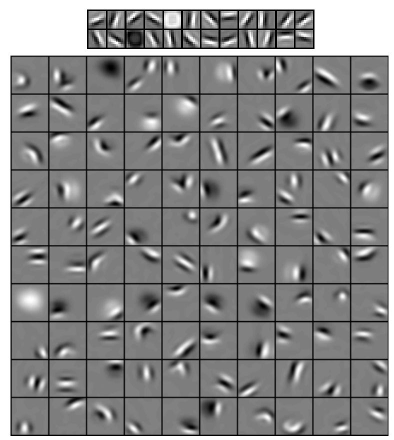
1st and 2nd Layers
'Edges'

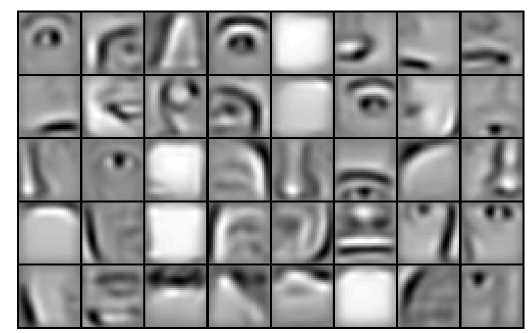
3rd Layer
'Object Parts'

4th Layer
'Objects'

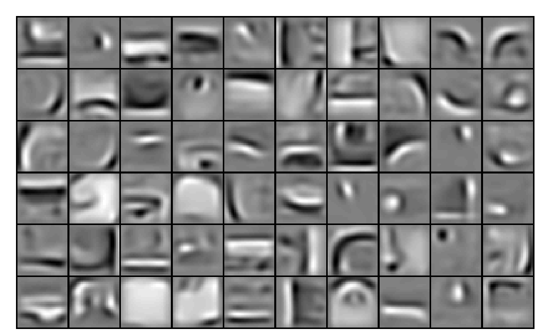
faces

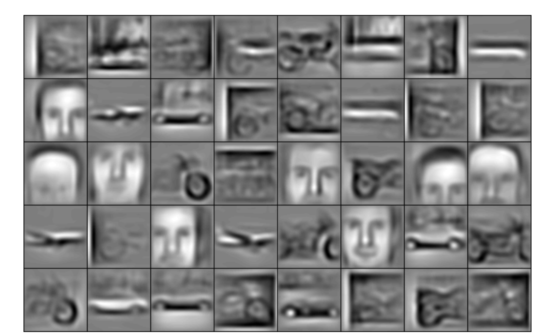






faces
cars
airplanes
motorbikes

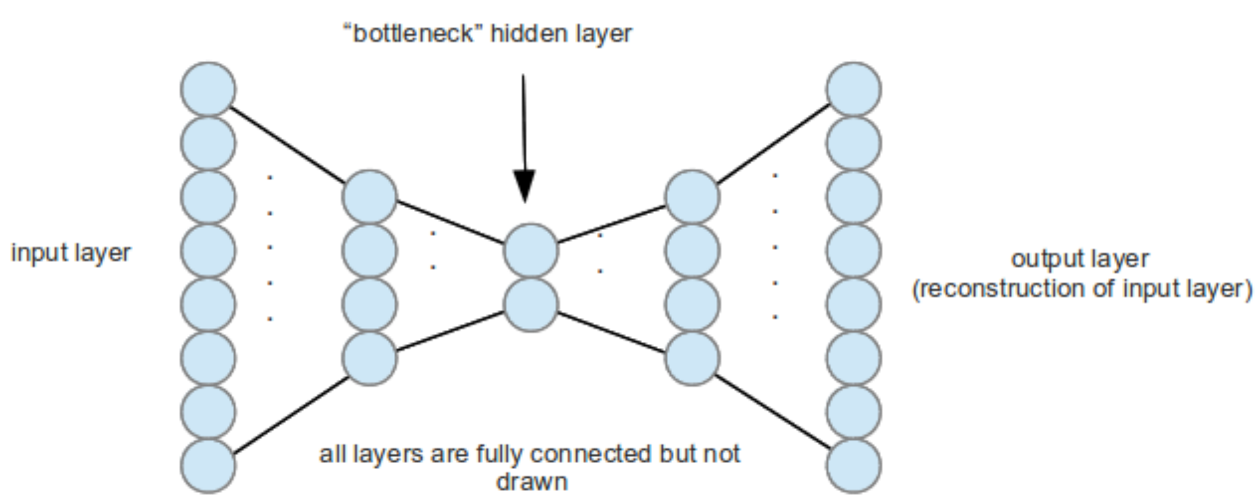




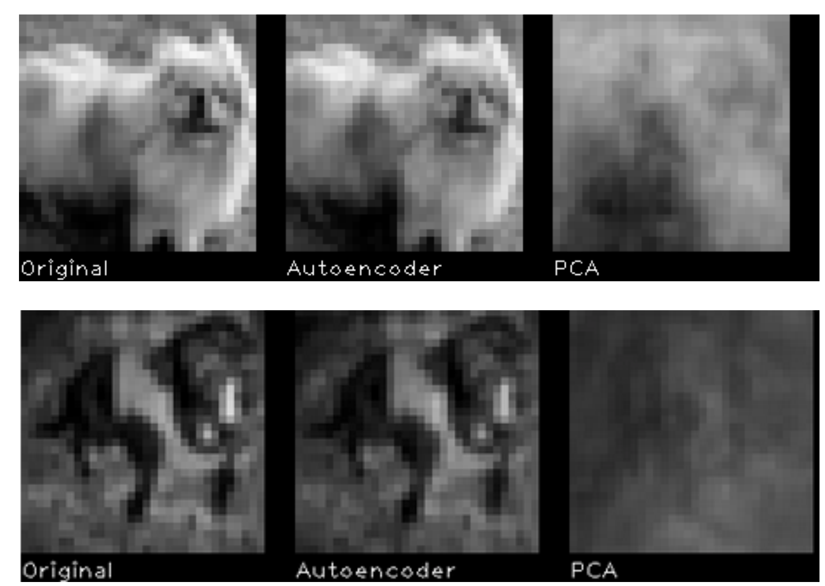
Deep Learning Methods

Unsupervised Methods

- Restricted Boltzmann Machines
- Deep Belief Networks
- Auto encoders: unsupervised feature extraction/learning



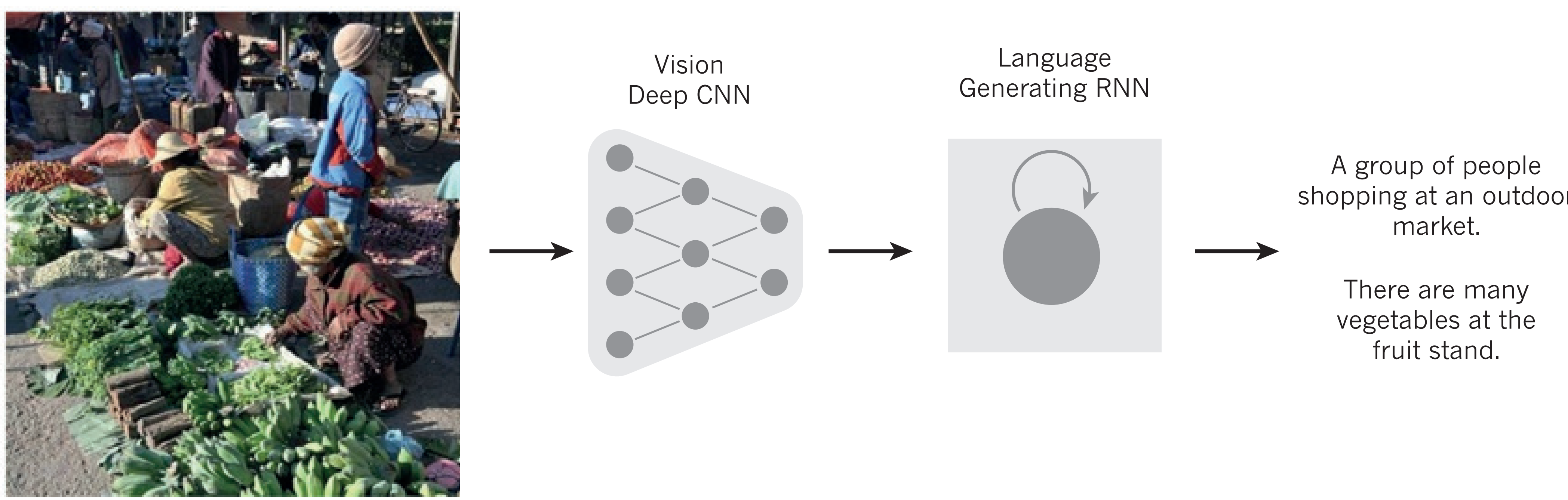
encode decode



Deep Learning Methods

Supervised Methods

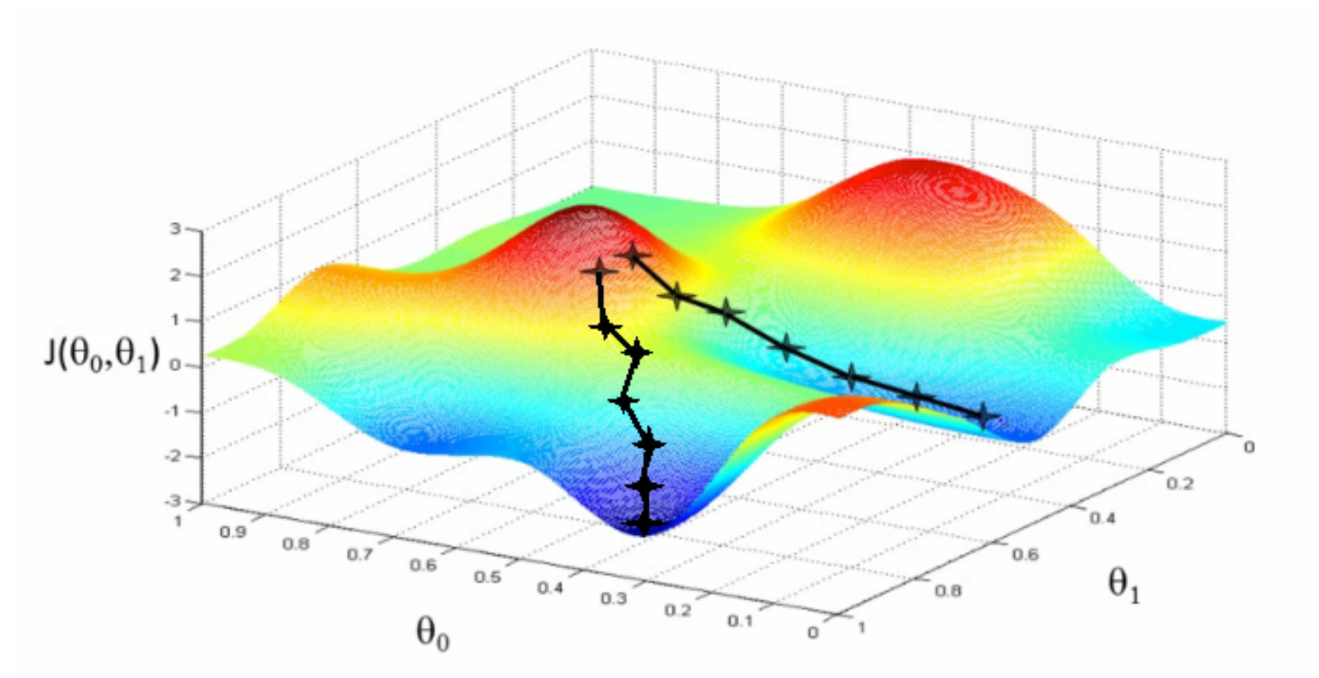
- Deep Neural Networks
- Recurrent Neural Networks
- Convolutional Neural Networks



How to train a deep network ?

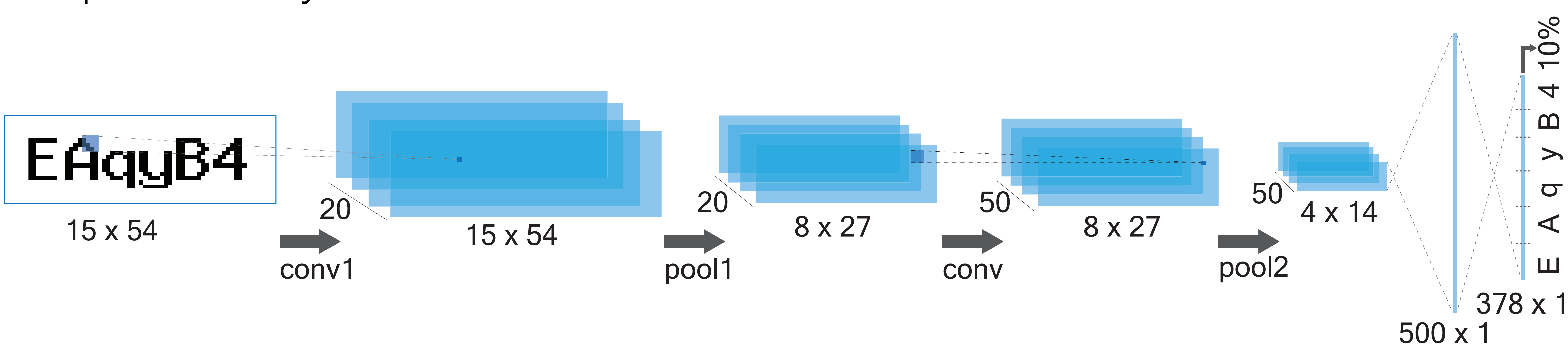
Stochastic Gradient Descent — *supervised learning*

- show input vector of few examples
- compute the output and the errors
- compute average gradient
- update the weights accordingly



Convolutional Neural Networks

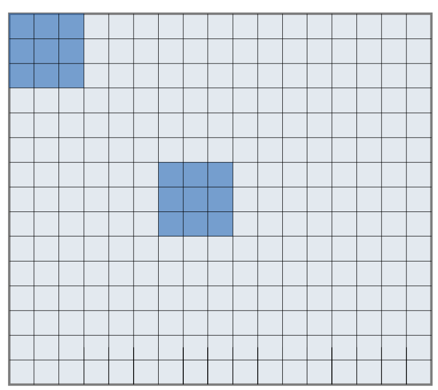
- CNNs are designed to process the data in the form of multiple arrays (e.g. 2D images, 3D video/volumetric images)
- Typical architecture is composed of series of stages: ***convolutional*** *layers* and ***pooling*** *layers*
- Each unit is connected to local patches in the feature maps of the previous layer



Key Idea behind Convolutional Networks

Convolutional networks take advantage of the properties of natural signals:

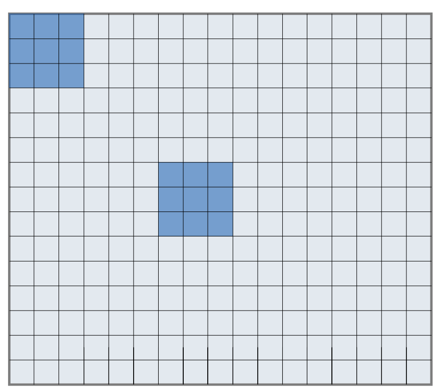
- local connections



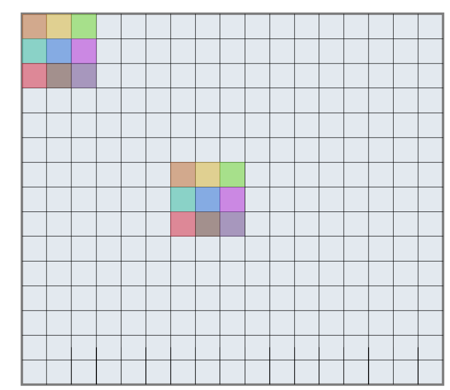
Key Idea behind Convolutional Networks

Convolutional networks take advantage of the properties of natural signals:

- local connections



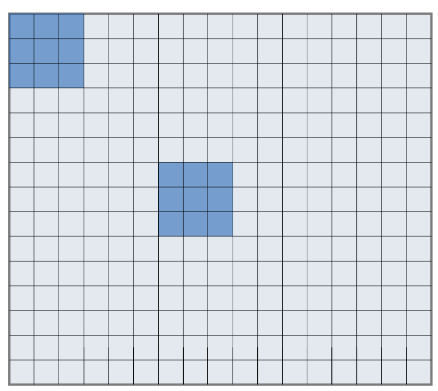
- shared weights



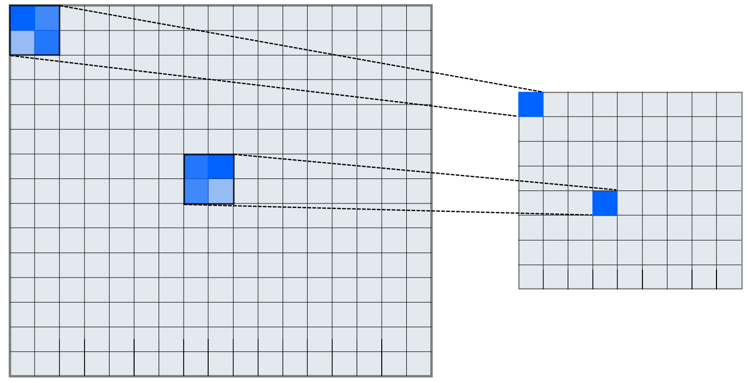
Key Idea behind Convolutional Networks

Convolutional networks take advantage of the properties of natural signals:

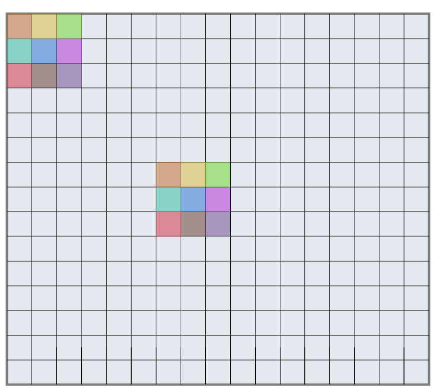
- local connections



- pooling



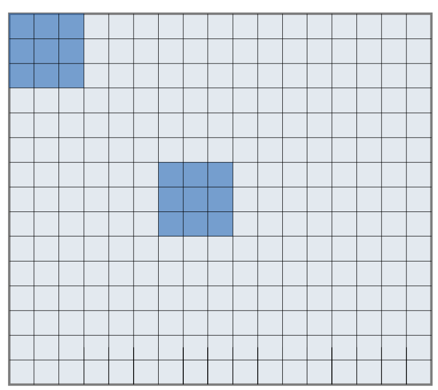
- shared weights



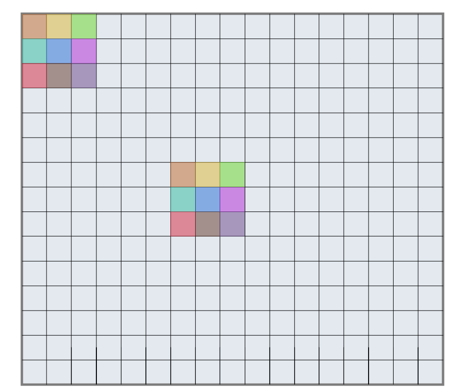
Key Idea behind Convolutional Networks

Convolutional networks take advantage of the properties of natural signals:

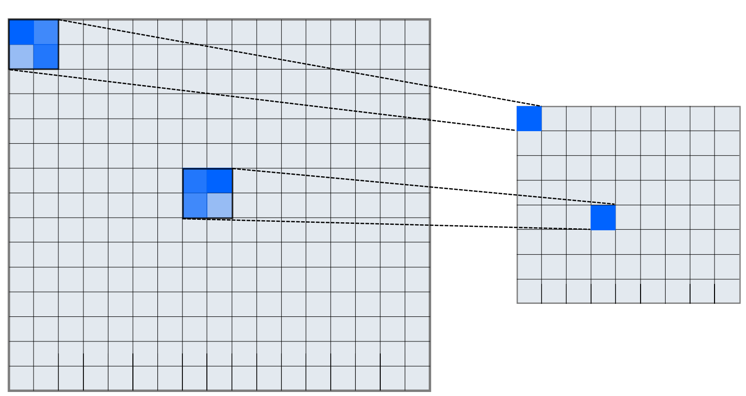
- local connections



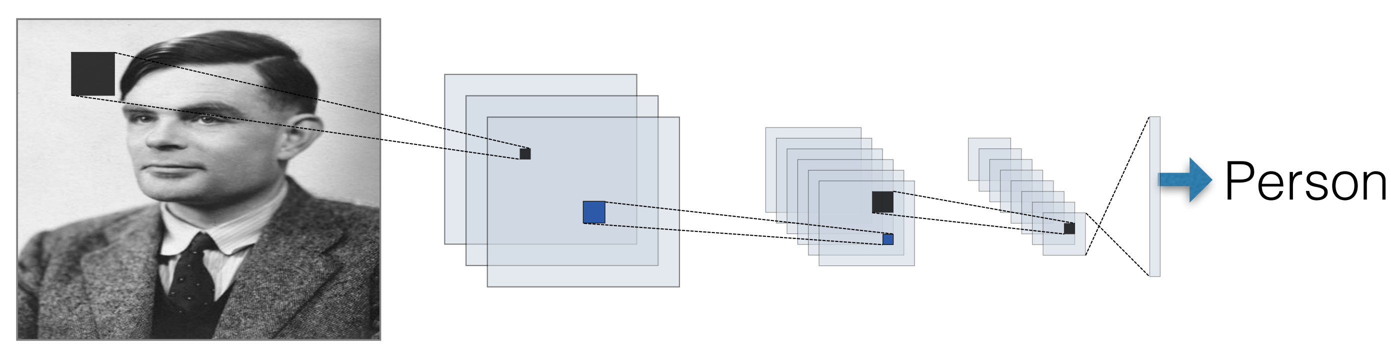
- shared weights



- pooling



- the use of many layers



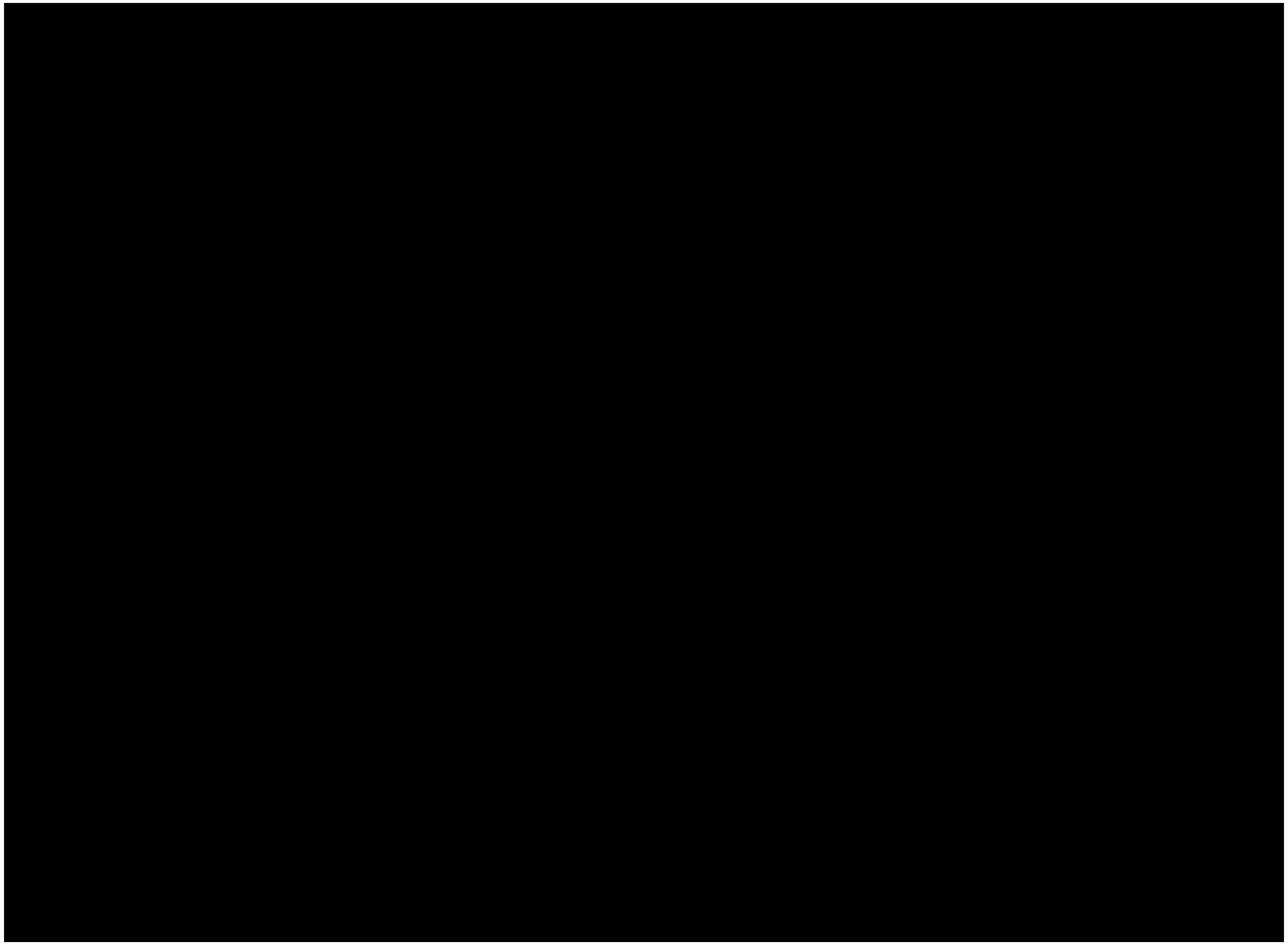
Pros & Cons

- Best performing method in many Computer Vision tasks
- No need of *hand-crafted* features
- Most applicable method for large-scale problems, e.g. classification of 1000 classes
- Easy parallelization on GPUs

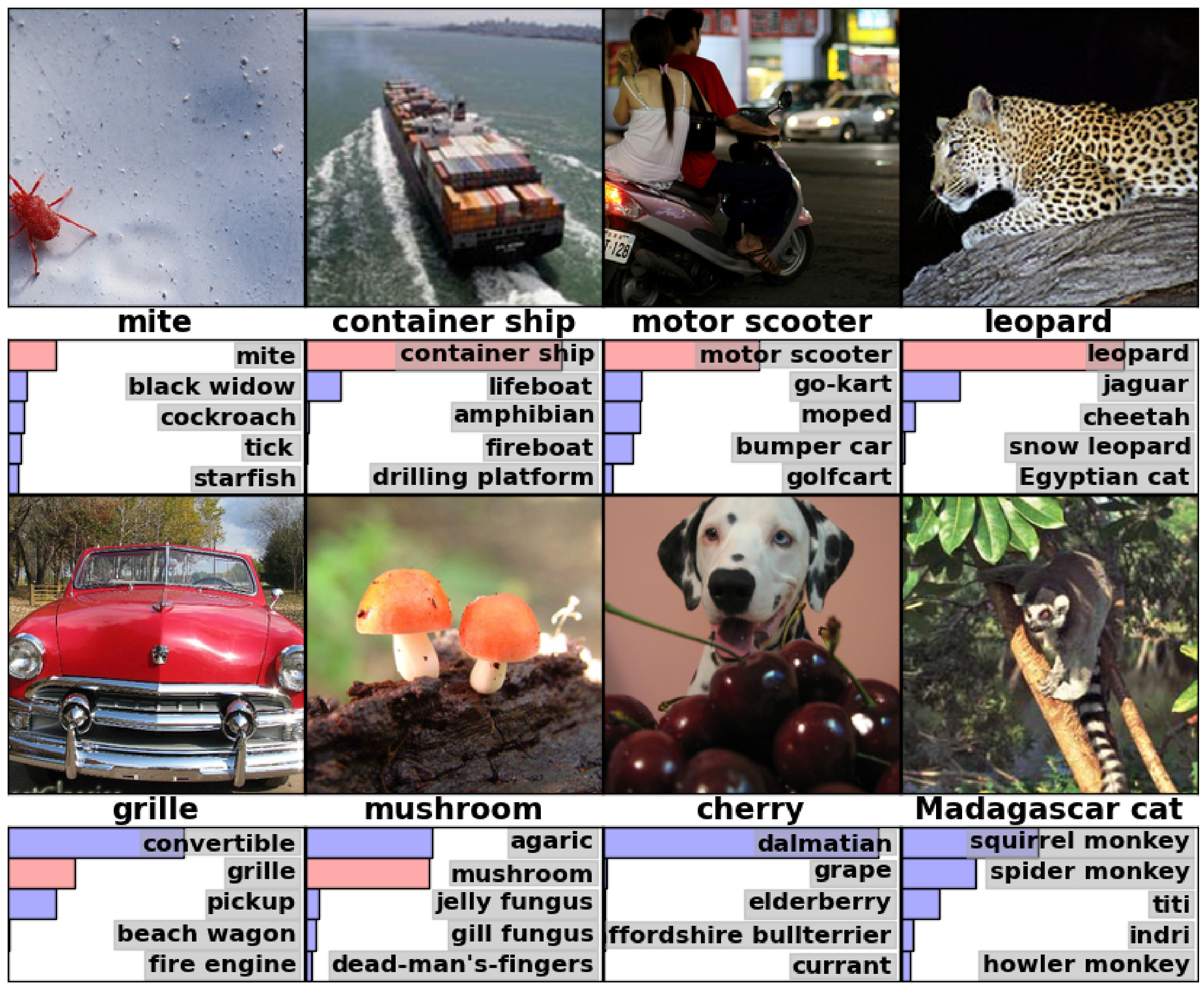
- Need of huge amount of training data
- Hard to train (local minima problem, tuning hyper-parameters)
- Difficult to analyse (*to be solved*)

Deep Learning Applications in Computer Vision

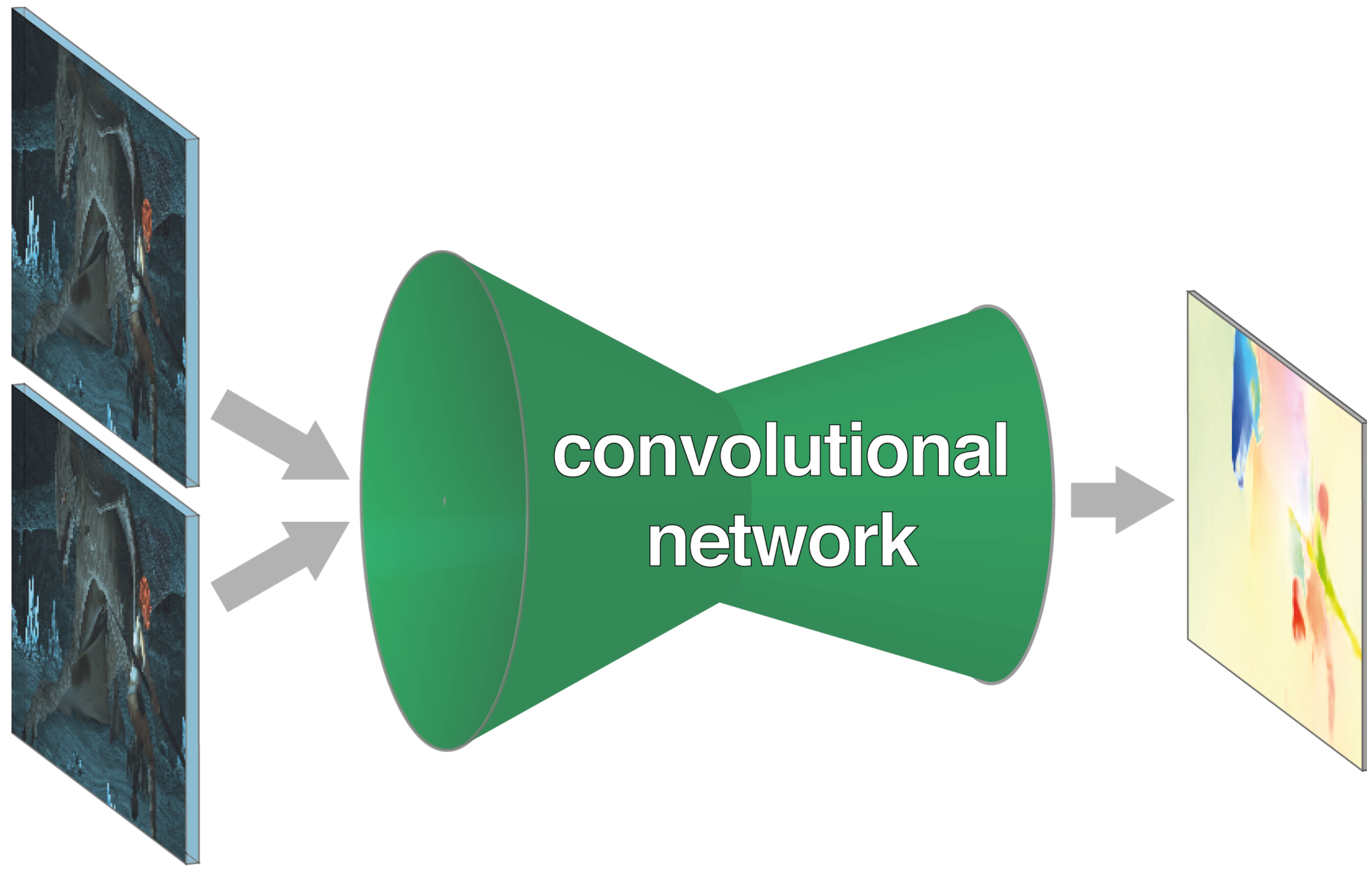
Handwritten Digit Recognition



ImageNet Classification with Deep Convolutional Neural Networks (AlexNet)

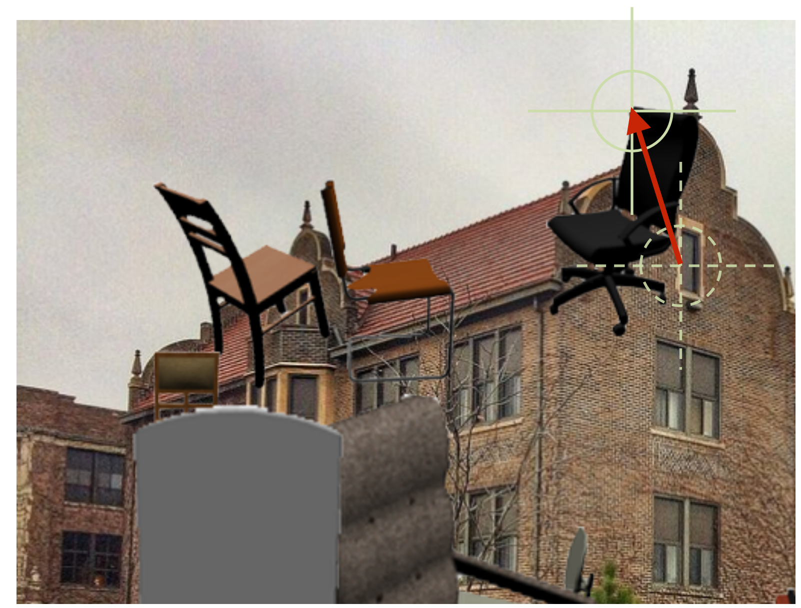


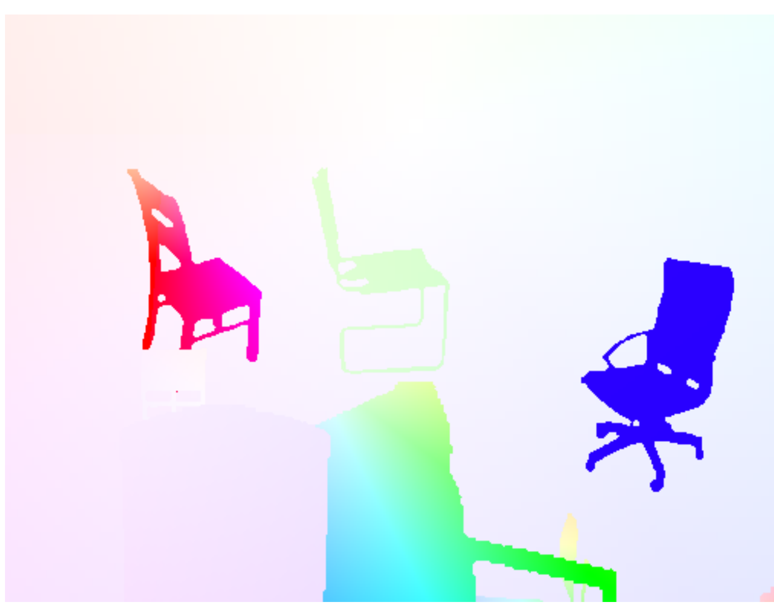
FlowNet: Learning Optical Flow with Convolutional Networks

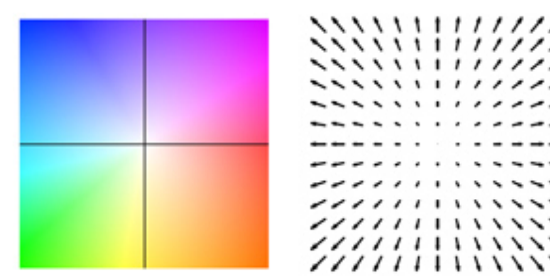


in collaboration with University of Freiburg
lmb.informatik.uni-freiburg.de

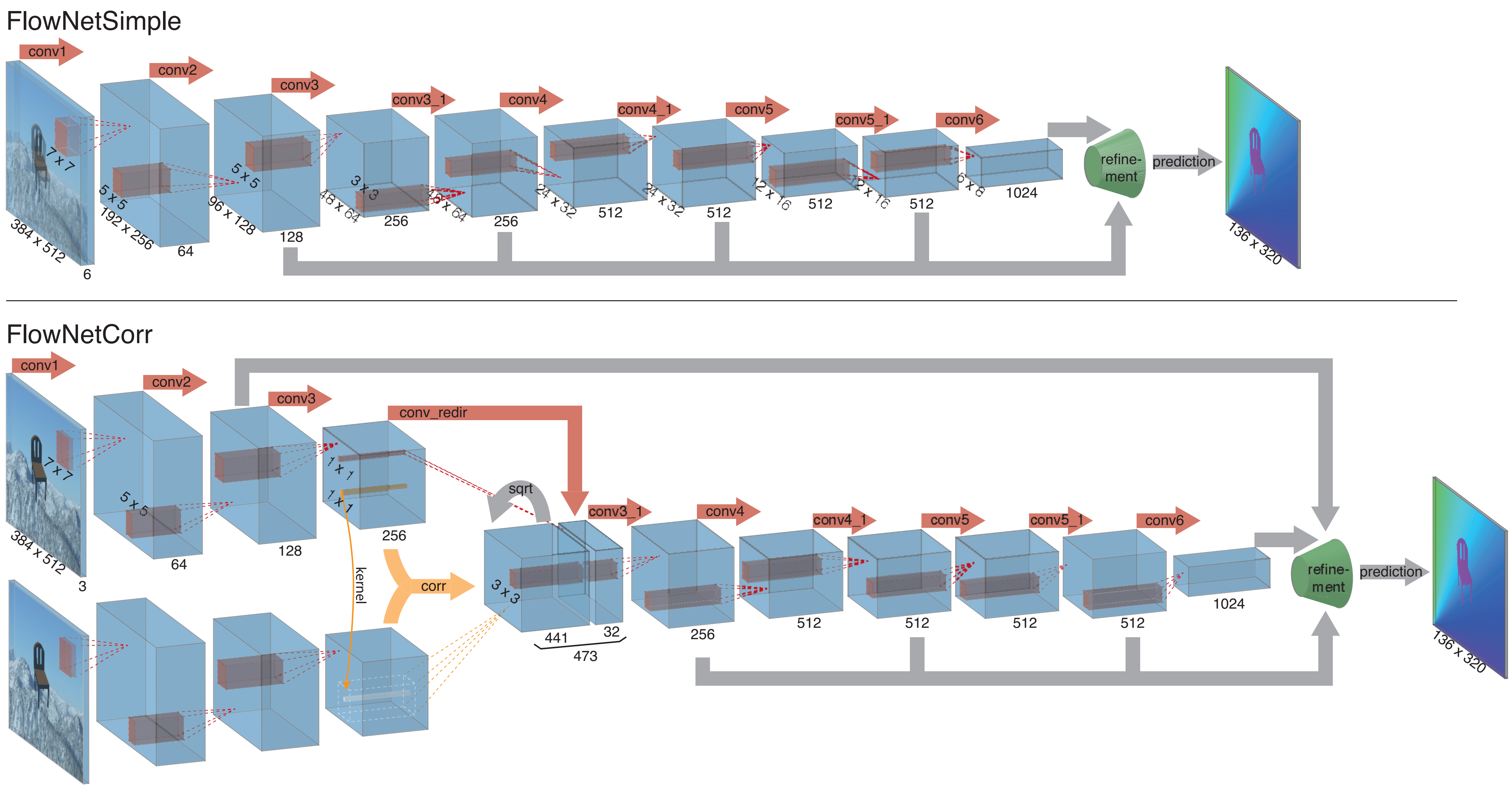
FlowNet: Learning Optical Flow with Convolutional Networks



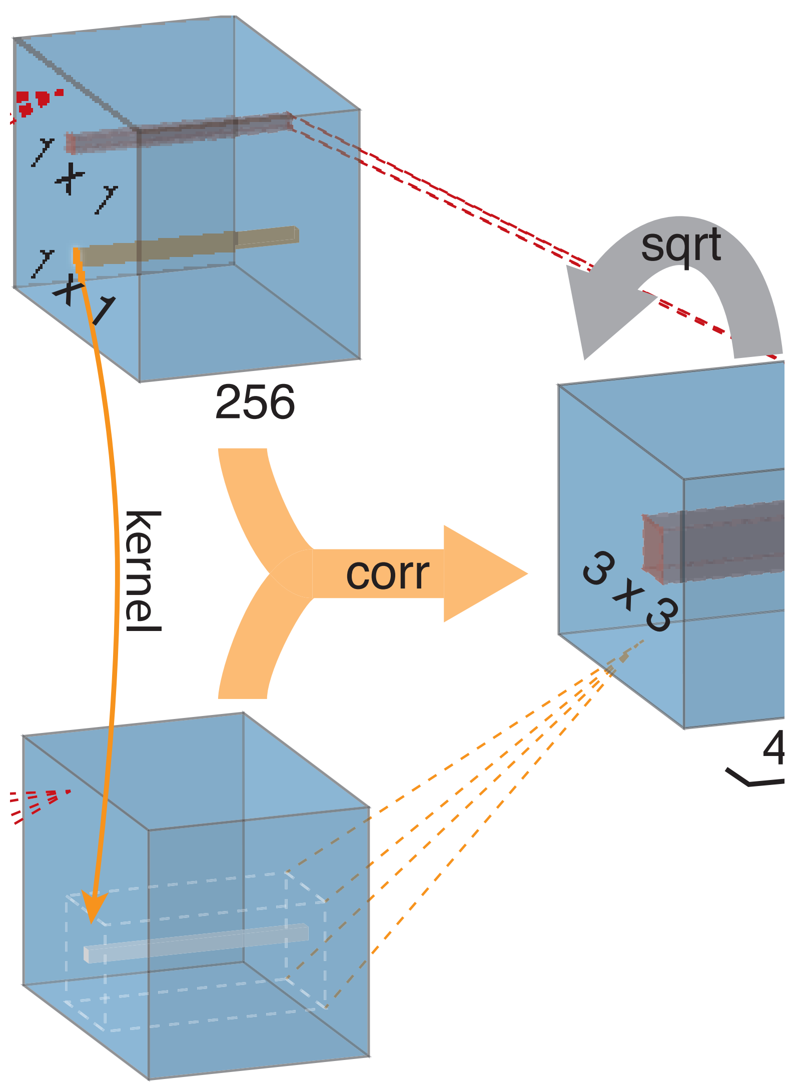


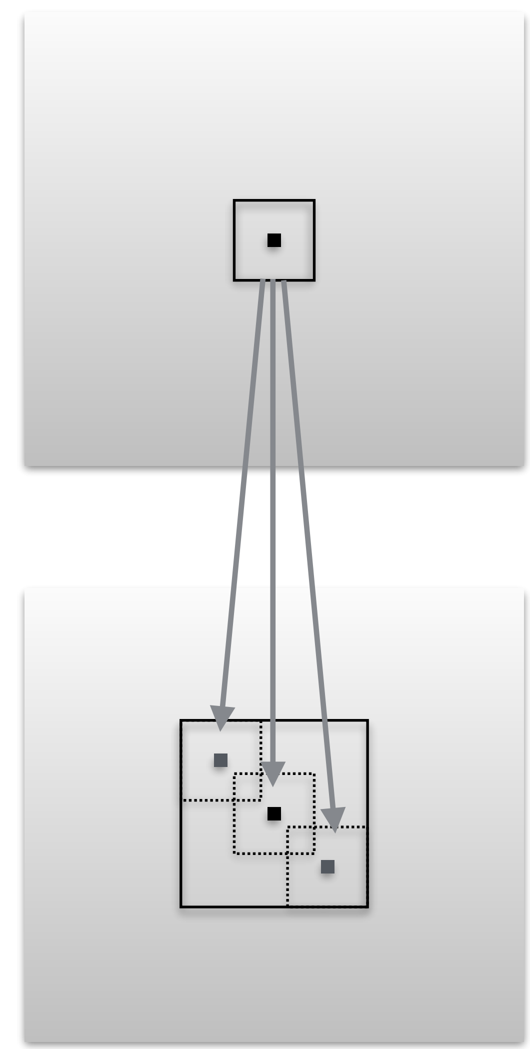


FlowNet: Learning Optical Flow with Convolutional Networks

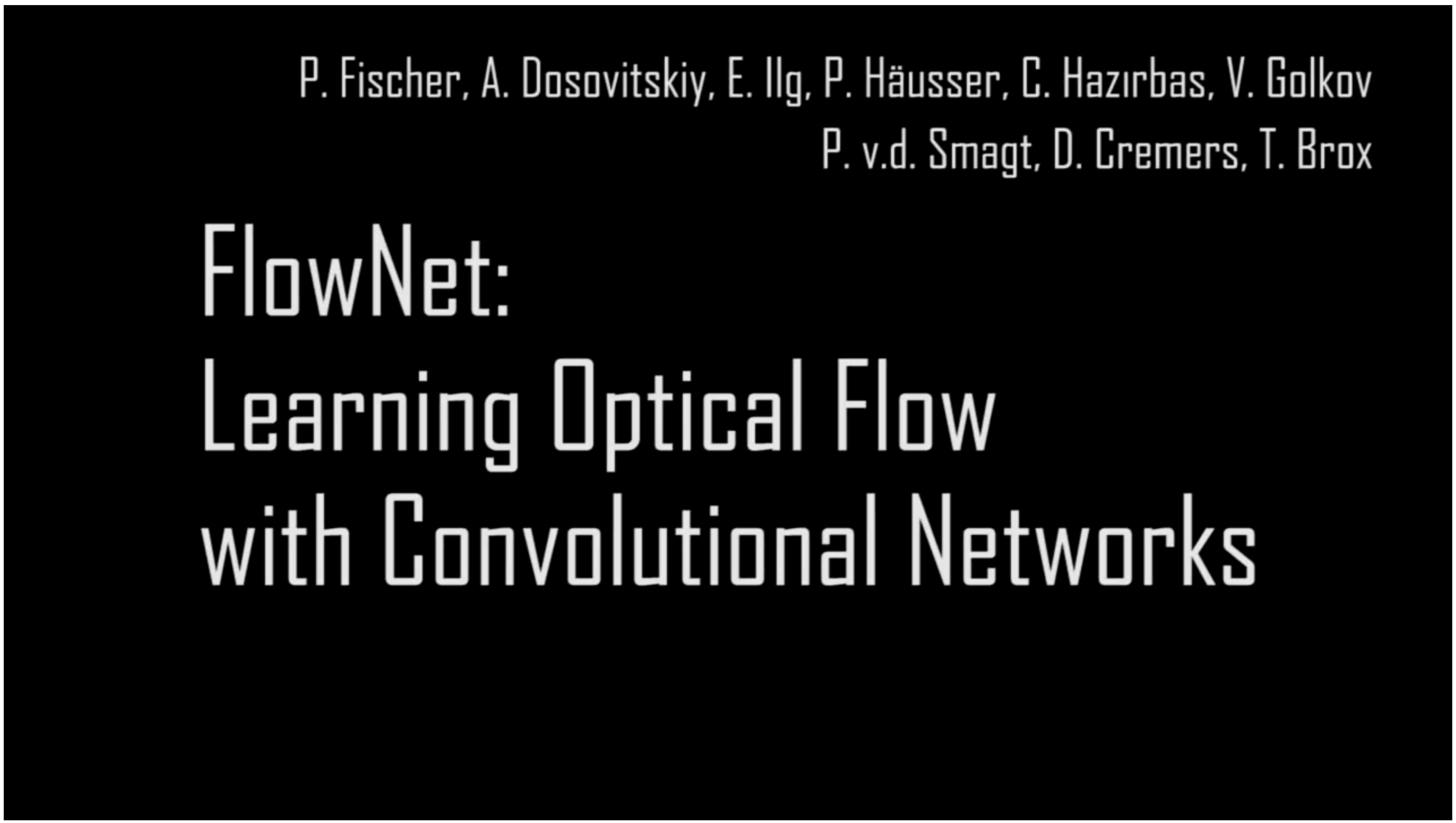


FlowNet: Learning Optical Flow with Convolutional Networks

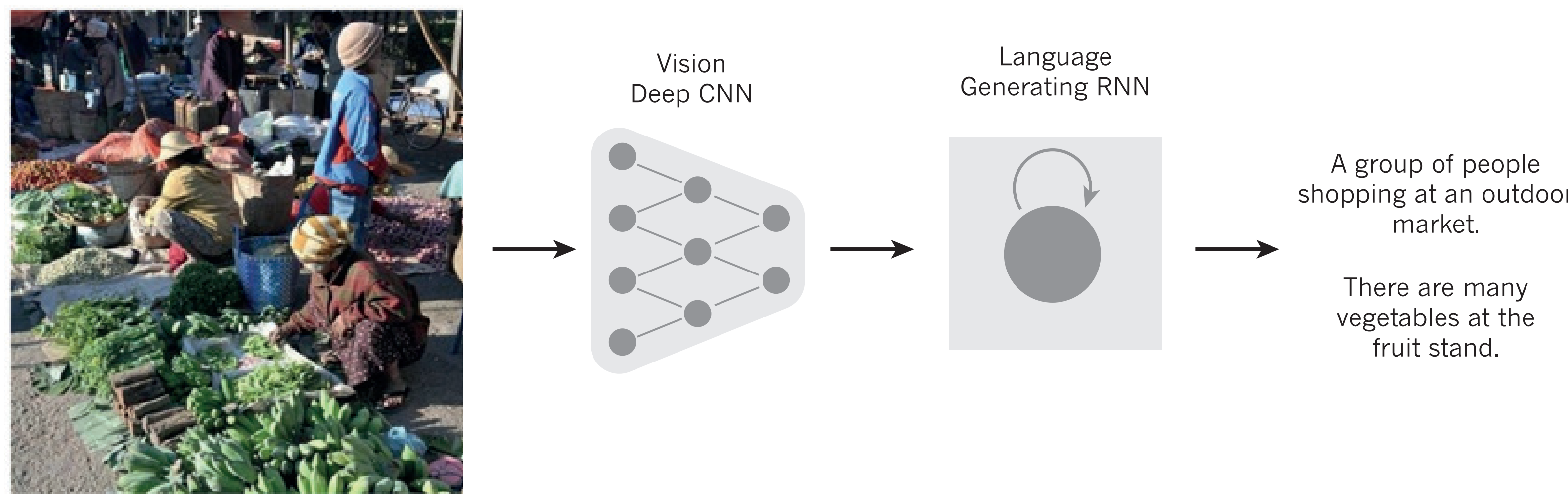




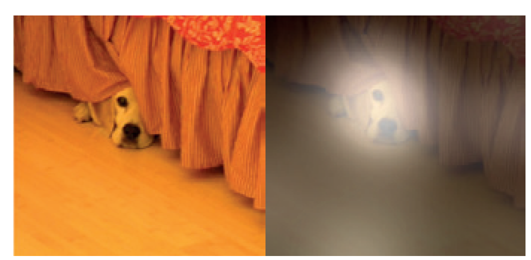
FlowNet: Learning Optical Flow with Convolutional Networks



From Image to Caption

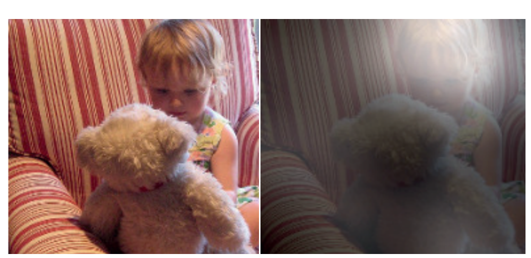


A woman is throwing a **frisbee** in a park.

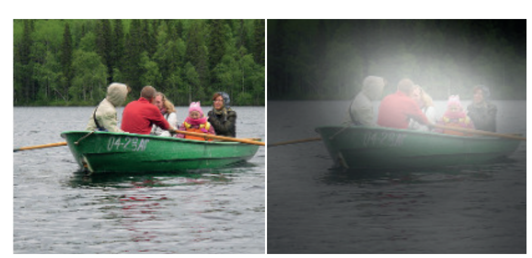


A **dog** is standing on a hardwood floor.

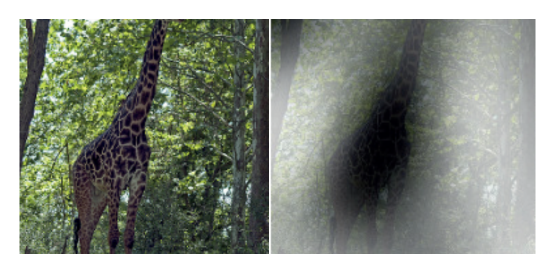
A **stop** sign is on a road with a mountain in the background



A little **girl** sitting on a bed with a teddy bear.



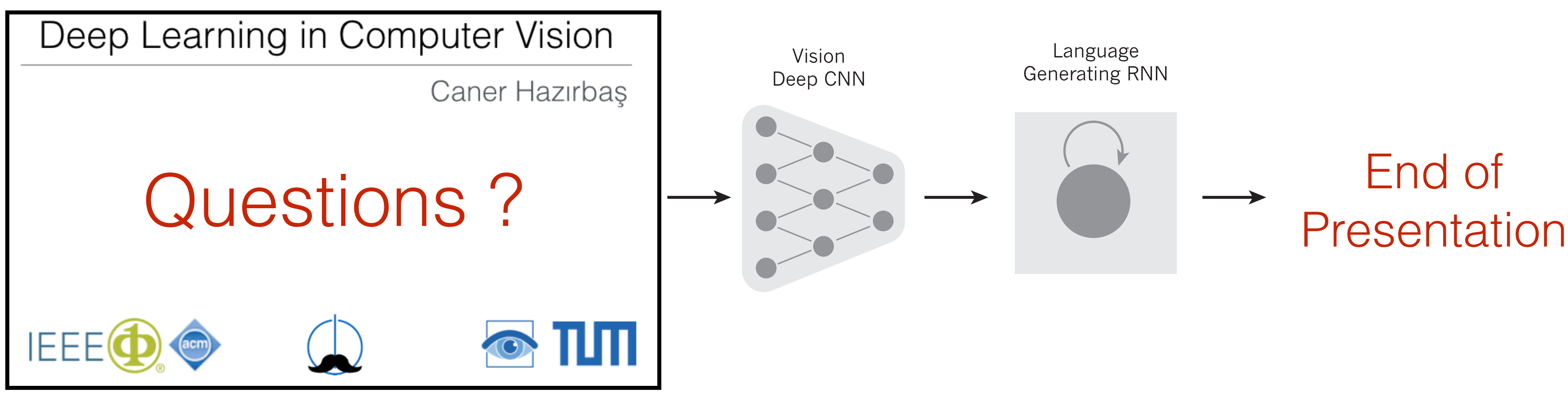
A group of **people** sitting on a boat in the water.



A giraffe standing in a forest with **trees** in the background.

Deep Learning in Computer Vision

Caner Hazırbaş | hazirbas@cs.tum.edu



References

- Building High-level Features Using Large Scale Unsupervised Learning
  Quoc V. Le , Rajat Monga , Matthieu Devin , Kai Chen , Greg S. Corrado , Jeff Dean , Andrew Y. Ng
  ICML'12

- Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations
  Honglak Lee Roger Grosse Rajesh Ranganath Andrew Y. Ng ICML'09

- ImageNet Classification with Deep Convolutional Neural Networks
  Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton NIPS'12

- Gradient-based learning applied to document recognition.
  Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner Proceedings of the IEEE'98

- FlowNet: Learning Optical Flow with Convolutional Networks
  Philipp Fischer, Alexey Dosovitskiy, Eddy Ilg, Philip Häusser, Caner Hazırbaş, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, Thomas Brox

References

- Google’s cat detection neural network http://www.resnap.com/image-selection-technology/deep-learning-image-classification/
- Example auto-encoder : http://nghiaho.com/?p=1765
- SGD : http://blog.datumbox.com/tuning-the-learning-rate-in-gradient-descent/