



Master Seminar:
Practical Applications of Multimedia Retrieval
Hasso Plattner Institute
Dr. Haojin Yang
20.10.2016



- **Dr. Haojin Yang**

- Dipl.-Ing study at TU-Ilmenau (2002-2007)
- Software engineer (2008-2010)
- PhD student, internet technology and system, at HPI (2010-2013)
- Senior researcher, chair of Internet technologies and systems
- Research interest: multimedia analysis, computer vision, machine learning/deep learning, information retrieval etc.
- Web: <http://hpi.de/meinel/lehrstuhl/team-fotos/postdocs/haojin-yang.html>



Personal Information

Christian Bartz, M.sc



■ Research background

- 2010~2013 Bachelor Degree (Hasso-Plattner-Institute)
- 2013~2016 Master Degree (Hasso-Plattner-Institute)
- 2016~ PhD Student at Hasso-Plattner-Institute

■ Research interests

- Computer vision, deep learning, text recognition
data generation



Personal Information

Xiaoyin Che, M.sc

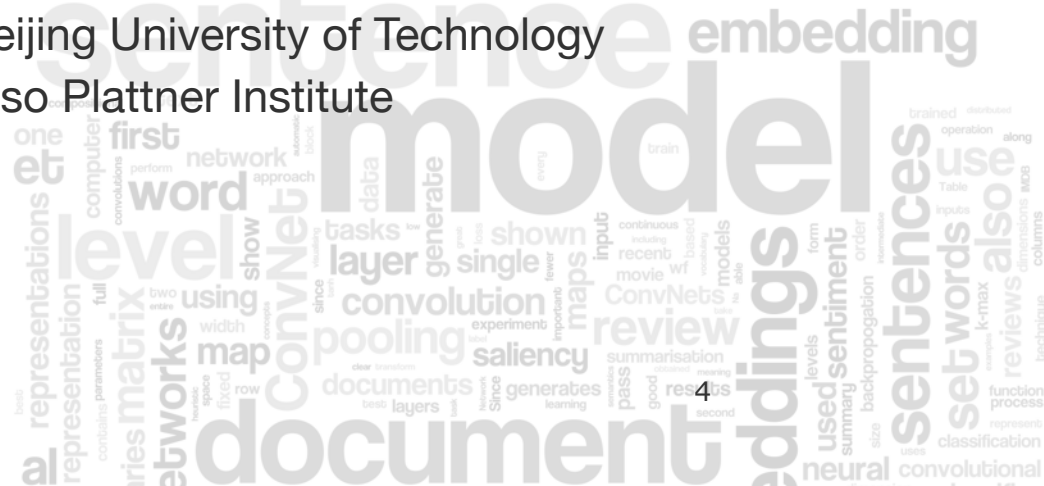


Education:

- 2005~2009 Bachelor Degree in Beijing University of Technology
- 2009~2012 Master Degree in Beijing University of Technology
- 2012~ PhD Student in Hasso Plattner Institute

Research Topics:

- Document Analysis
- Deep Learning
- Natural Language Processing
- E-Learning

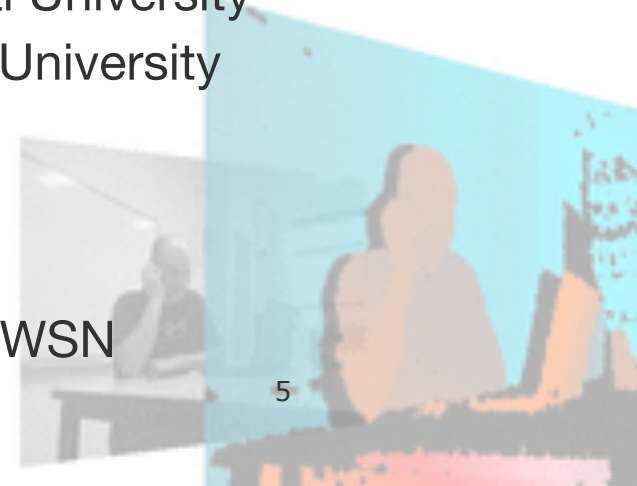


Personal Information

Sheng Luo, M.sc



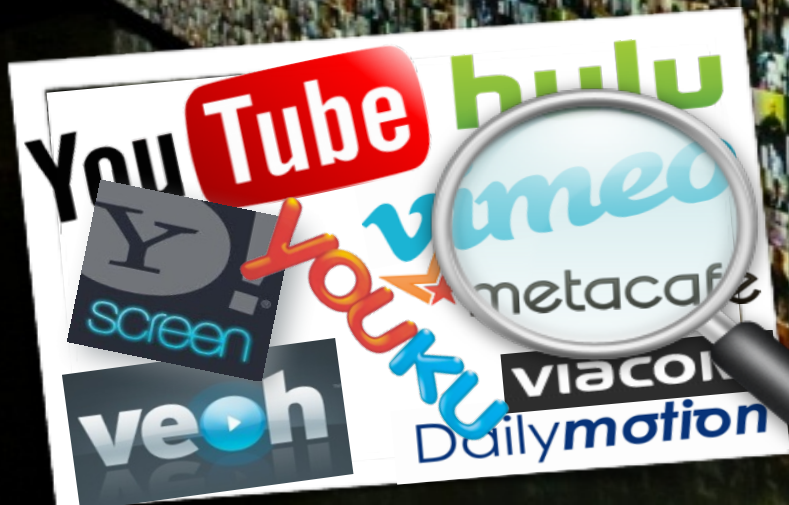
- Research background
 - 2011.09-2014.03 Master of Engineering, Shanghai University
 - 2012.09-2013.09 Master of Engineering, Waseda University
 - 2014.4-now PhD student at HPI
- Research interests
 - Multimedia Retrieval, Deep learning, Robotic and WSN



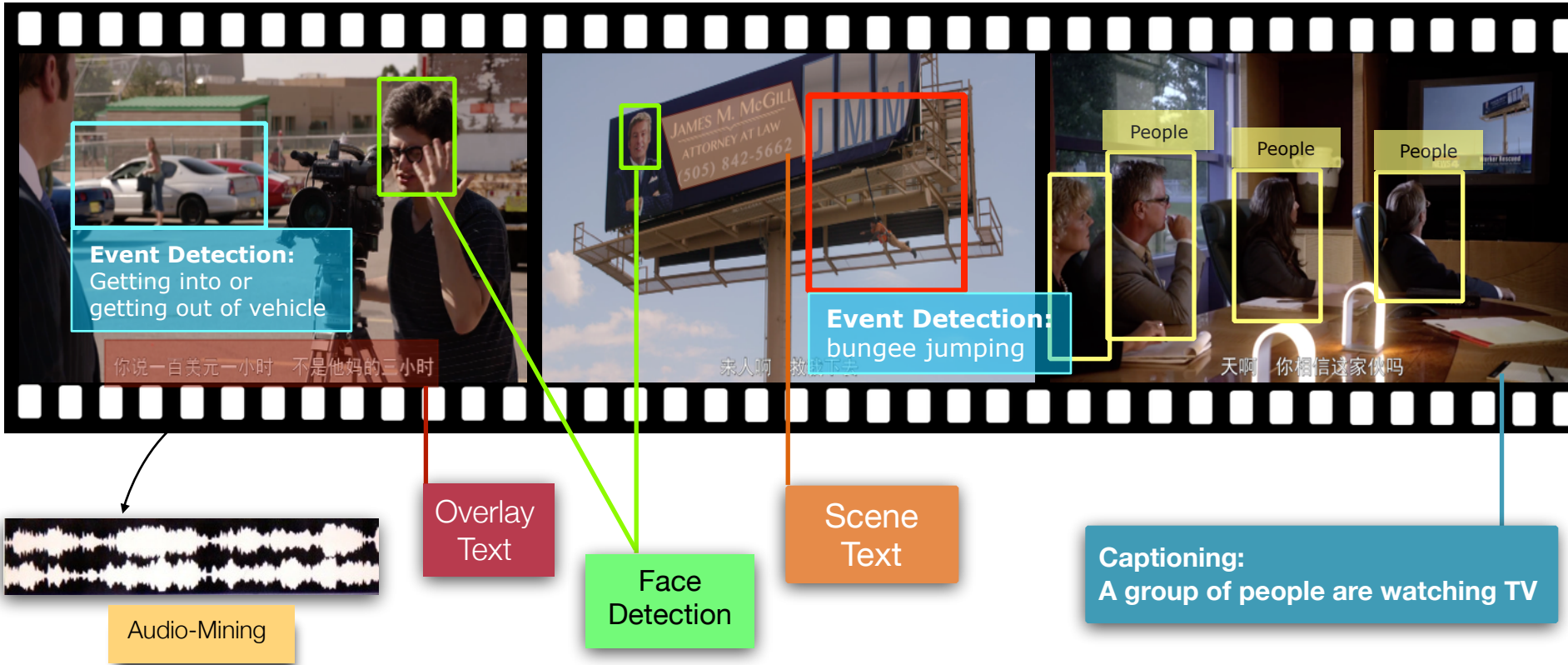


You Tube

- 400 hours of video uploaded every minute
- Video data —> more than 64% of internet traffics (2014), will be more than 80% in 2019



Automatic Multimedia Analysis



Why Machine Vision So Hard



62	62	63	64	65	66	67	67	69	70	71	72	72	73	73	73	73	72	72	71	70	69	67	66	66	66	65	63	62	61	60	60	
61	62	63	64	66	66	67	68	68	69	70	71	71	72	73	73	72	72	71	71	70	69	68	66	66	65	65	63	62	61	60	60	
61	62	63	64	66	66	68	68	69	70	70	71	72	73	73	73	72	72	71	71	69	68	67	66	66	65	65	64	63	62	61	61	
61	63	64	64	66	67	68	68	69	70	71	71	73	73	74	73	73	73	71	70	69	68	66	66	65	64	63	62	61	61	60		
61	63	64	65	67	68	69	69	70	70	71	71	72	55	53	69	72	72	71	71	70	69	68	67	66	65	64	63	62	60	60	60	
63	64	65	66	67	68	69	69	70	70	71	72	42	4	5	11	48	72	71	71	69	69	68	67	66	65	64	62	62	60	59	59	
63	65	66	66	68	68	69	70	71	71	71	72	18	4	4	7	8	66	71	70	69	68	68	67	66	65	64	63	61	59	59	58	
63	65	67	67	68	69	69	70	71	71	72	64	4	27	24	54	33	29	52	64	68	67	66	65	64	63	62	61	59	58	58		
64	65	66	66	68	69	70	71	41	24	24	12	17	24	48	60	37	43	38	52	66	68	67	66	65	64	63	61	60	59	58	57	
65	66	67	67	68	69	71	49	6	6	6	5	34	36	12	47	34	17	29	54	43	63	67	66	65	64	63	62	60	59	58	57	
64	65	66	66	68	69	38	6	6	5	5	7	16	19	4	47	44	27	24	40	67	66	66	65	65	64	63	61	60	59	58	57	
63	64	65	65	67	30	6	6	5	5	5	6	8	9	20	27	51	78	41	44	66	65	65	65	64	63	62	60	59	58	57		
63	64	65	65	34	5	5	5	5	5	5	4	19	6	7	54	64	20	59	65	65	64	64	64	63	62	61	60	59	57	56		
63	64	64	65	14	5	6	5	5	4	5	4	18	7	5	4	19	10	11	65	64	64	64	63	61	66	62	61	60	59	58	56	
63	64	64	65	37	4	5	6	6	7	10	6	5	5	4	21	24	18	64	64	64	63	62	64	65	62	62	60	59	58	57		
64	64	64	64	65	50	4	4	4	5	11	16	6	6	4	6	35	16	26	66	64	64	63	61	72	67	63	62	61	59	58	57	
64	64	64	64	65	46	4	4	4	5	6	9	8	5	28	10	43	56	29	57	64	64	63	61	70	67	62	64	65	59	59	57	
64	64	64	65	66	27	5	4	4	5	6	6	6	18	66	20	57	60	46	36	75	70	62	61	70	67	62	61	60	59	58	58	
49	50	62	65	57	5	5	6	5	6	6	6	41	59	28	60	58	44	22	63	71	72	60	69	68	61	60	58	59	59	58		
42	52	57	52	26	5	5	5	5	5	5	5	70	50	43	61	62	64	39	42	64	60	62	56	63	65	65	67	61	53	53		
32	32	32	33	6	5	5	5	5	5	6	6	11	39	21	33	51	50	45	46	18	32	36	33	23	44	70	71	51	42	27	31	
50	50	51	39	5	5	5	6	5	6	5	6	42	69	28	34	42	39	43	37	26	29	40	26	29	26	35	42	35	33	18	19	
52	53	51	22	5	5	5	6	5	6	5	6	5	44	56	17	51	54	53	54	56	51	22	54	54	55	55	54	53	53	53	52	52
54	54	53	8	5	5	5	6	5	6	6	13	52	42	21	51	54	51	49	49	50	22	41	45	42	42	41	40	41	44	43	42	
52	52	54	36	8	5	5	6	6	5	6	28	55	32	32	54	53	51	51	51	51	44	25	51	51	49	49	50	49	48	46	46	
54	54	52	53	30	7	5	6	6	5	6	40	54	29	52	51	53	56	55	52	52	51	38	52	52	50	49	46	46	45	46	47	
51	52	51	53	27	14	5	4	5	4	7	47	51	21	39	49	47	49	52	52	49	35	31	48	46	47	47	47	46	46	43		
48	50	51	53	25	14	17	8	4	4	17	46	40	18	43	47	46	49	52	54	53	53	54	18	50	49	46	47	47	47	45	45	
49	49	49	49	22	12	20	24	6	14	35	51	39	48	48	50	51	51	49	51	51	52	50	41	58	48	47	47	47	45	45	46	
51	49	50	50	22	13	19	36	13	12	42	50	40	73	50	50	50	49	48	49	48	49	45	51	46	44	44	44	42	45	47		
47	49	49	47	20	16	26	39	21	15	36	48	42	61	47	48	51	47	50	51	51	51	49	47	47	52	47	47	44	43	45	46	
48	50	48	52	19	13	33	36	18	18	36	49	51	54	47	47	49	46	46	49	49	49	47	44	53	44	48	44	46	46	45		

Deep Learning for Multimedia Retrieval

- Deep Learning and deep features (*since 2006*):
 - Simulating human neural network and hierarchically learning features from large scale data
 - Impacting a wide range of multimedia information processing
 - Achieved break-record results in fields like *Speech Recognition, Image Classification, Object Detection* and *Nature Language Processing* etc.

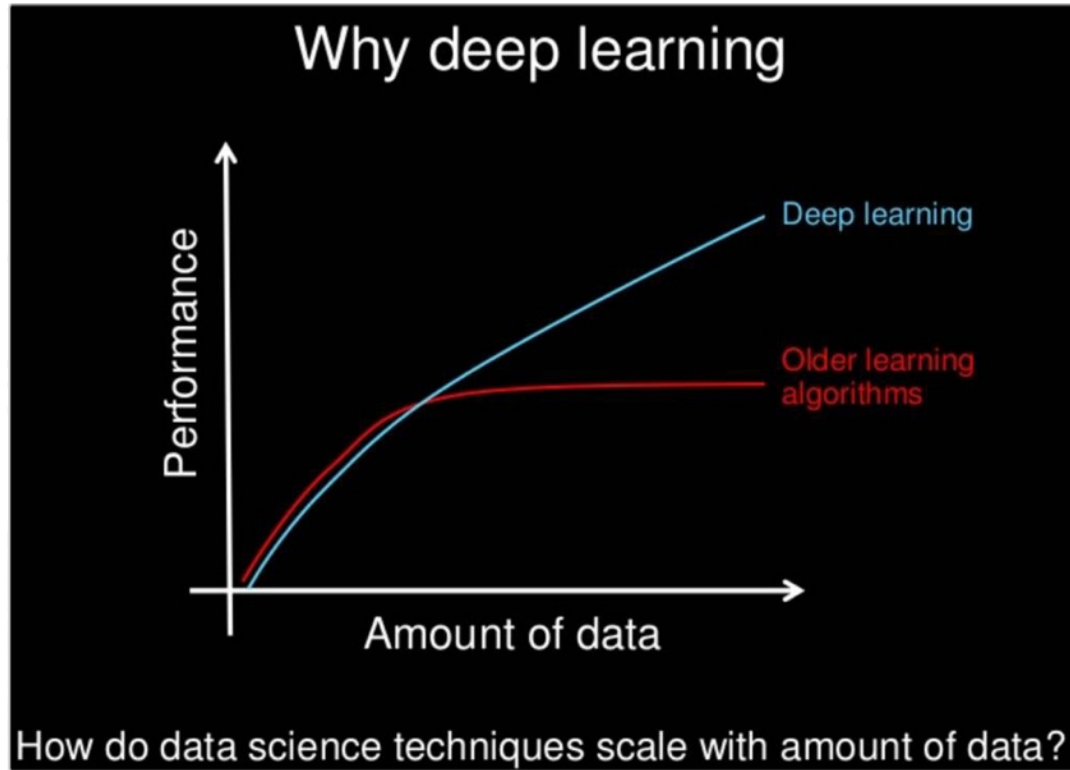
Deep learning as human beings



Deep Learning for Multimedia Retrieval

- Deep Learning
 - Simulating human perception
 - Impacting a wide range of applications
 - Achieved breakthroughs in many domains
 - Detection and*

D



Deep Learning for Multimedia Retrieval

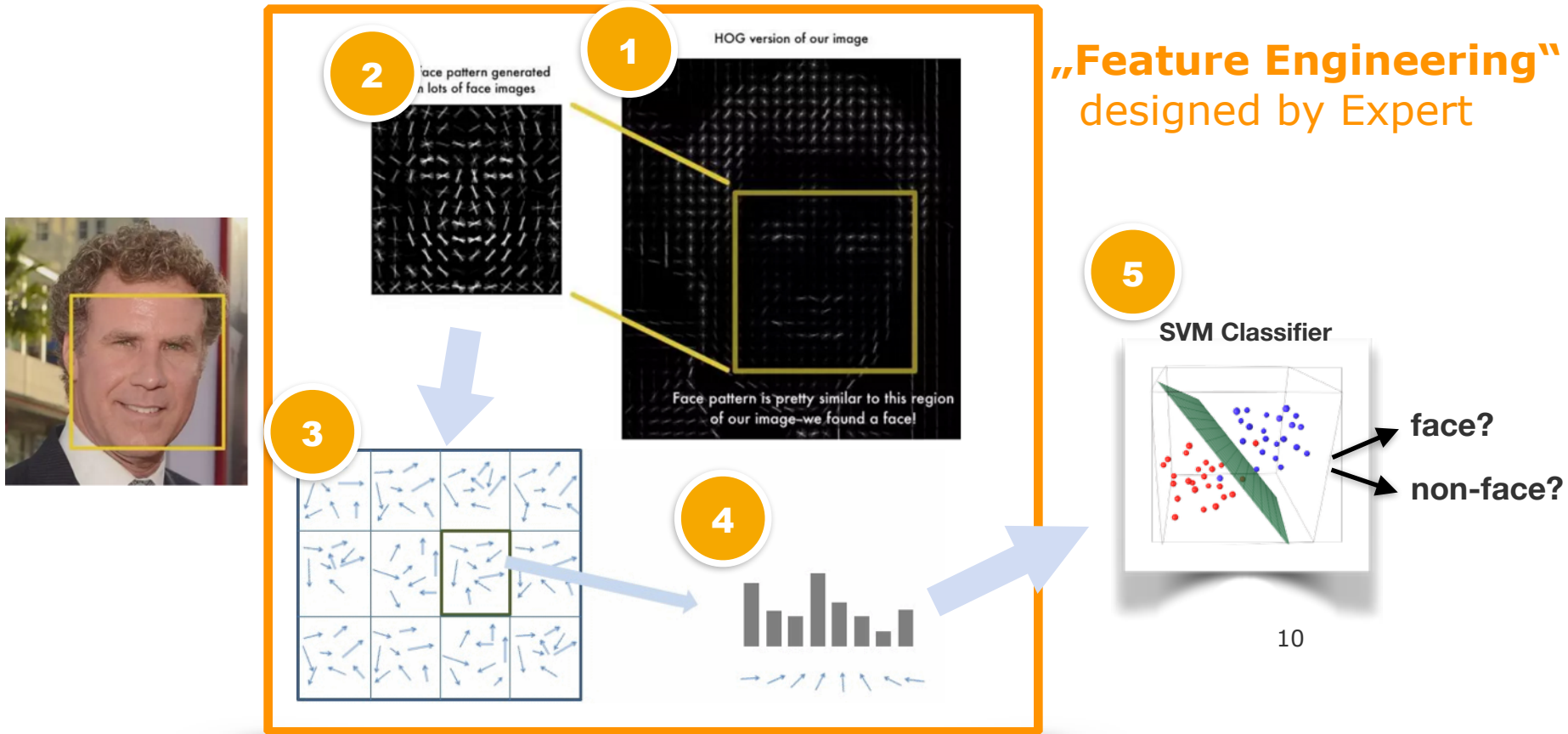
- Deep Learning and deep features (*since 2006*):
 - Simulating human neural network and hierarchically learning features from large scale data
 - Impacting a wide range of multimedia information processing
 - Achieved break-record results in fields like *Speech Recognition, Image Classification, Object Detection* and *Nature Language Processing* etc.

Deep learning as human beings



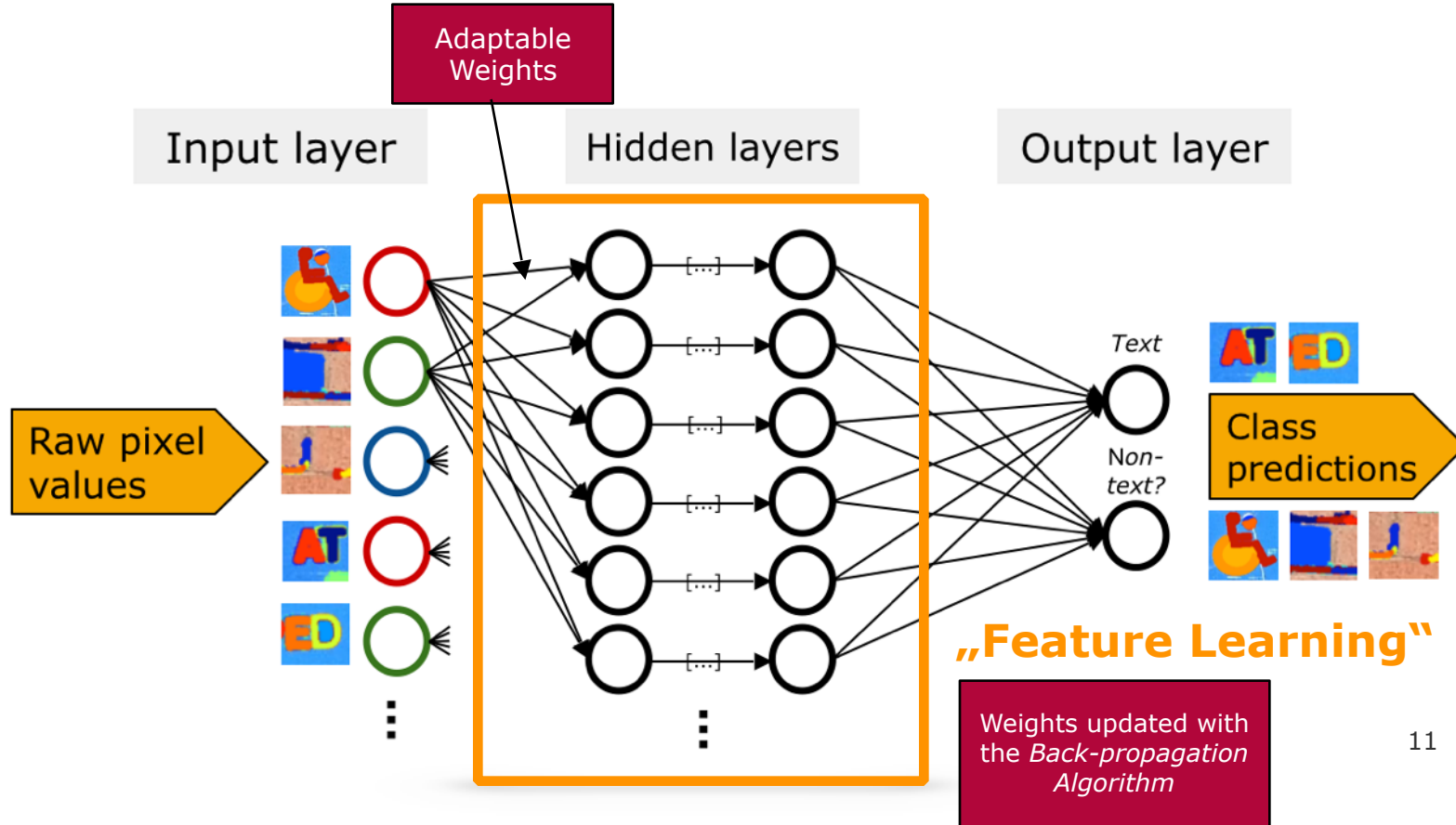
Handcrafted Features Example: HOG

- HOG (Histogram of Oriented Gradients) feature for face detection

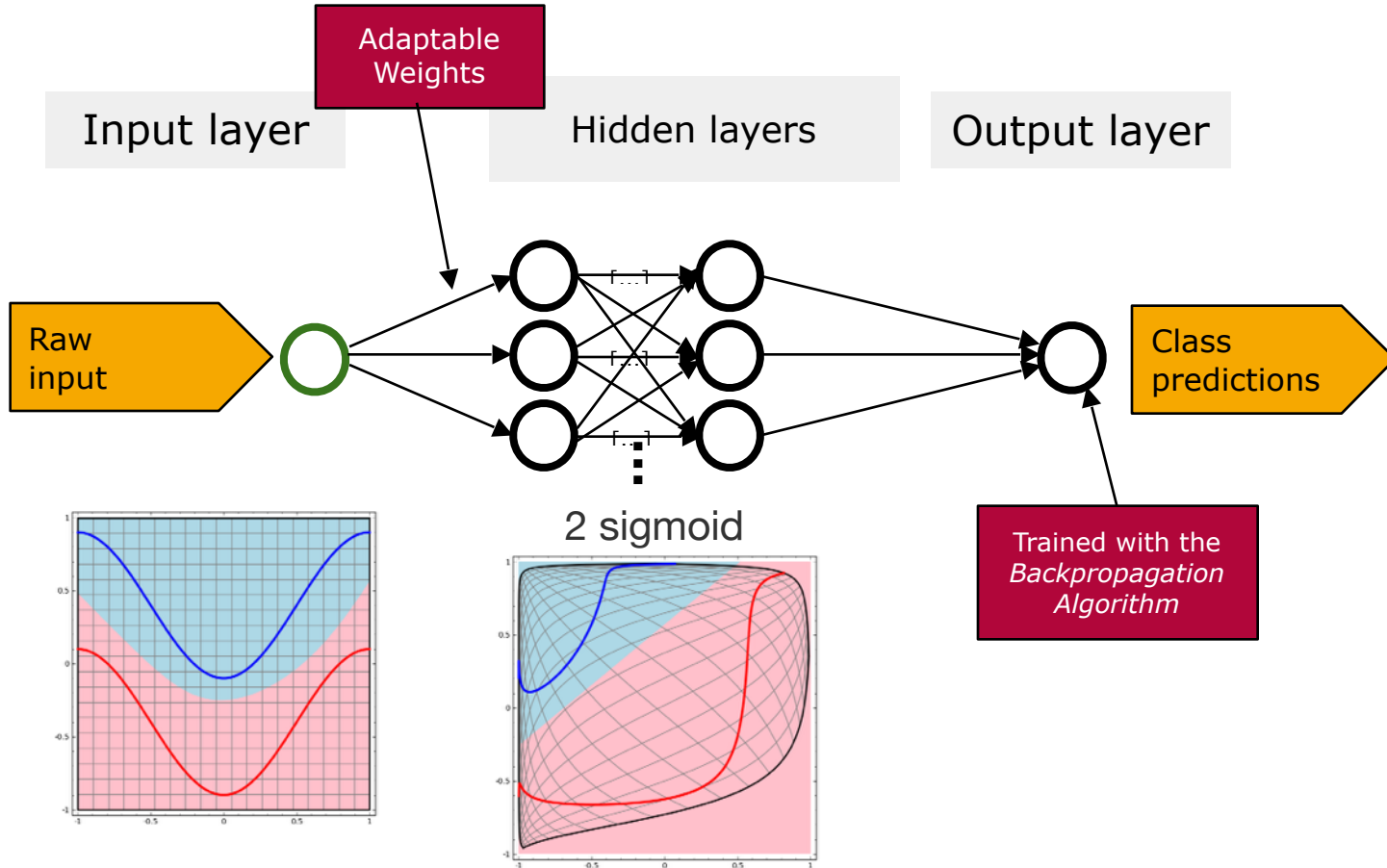


„Feature Engineering“
designed by Expert

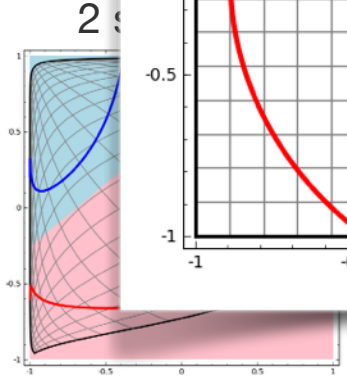
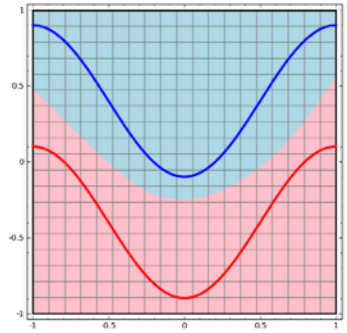
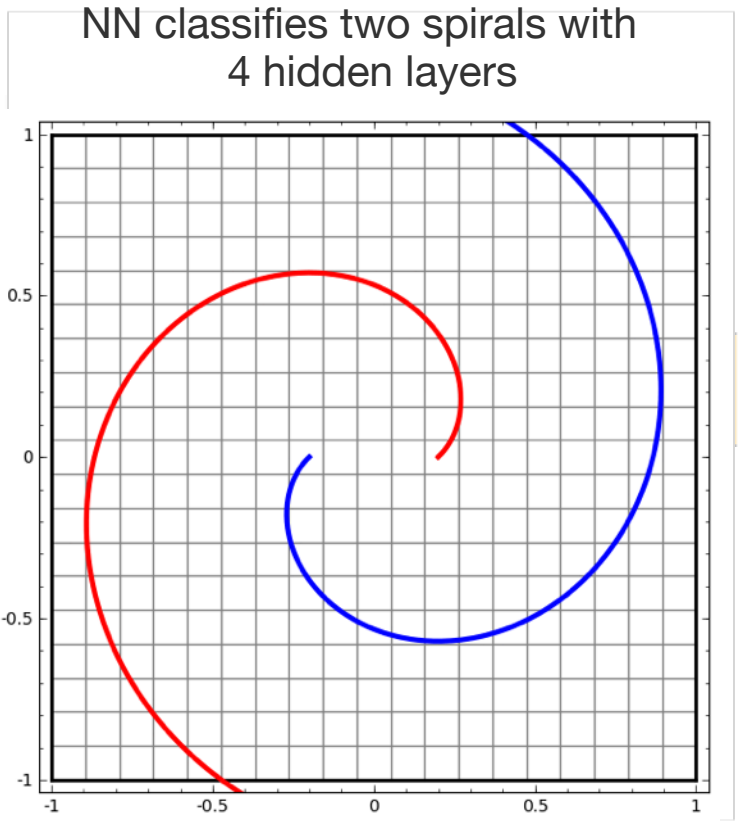
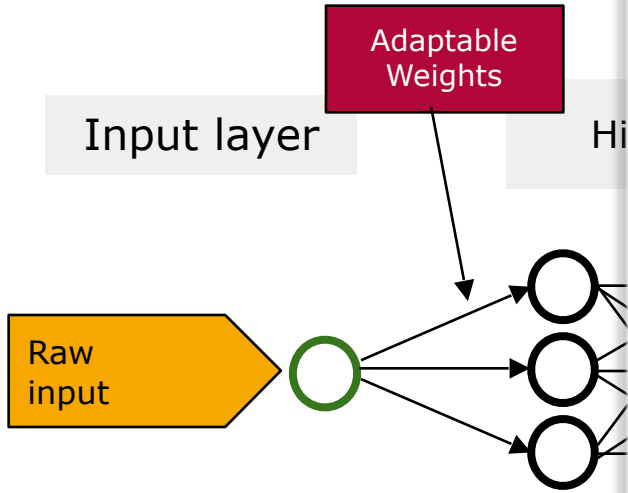
Artificial Neural Networks



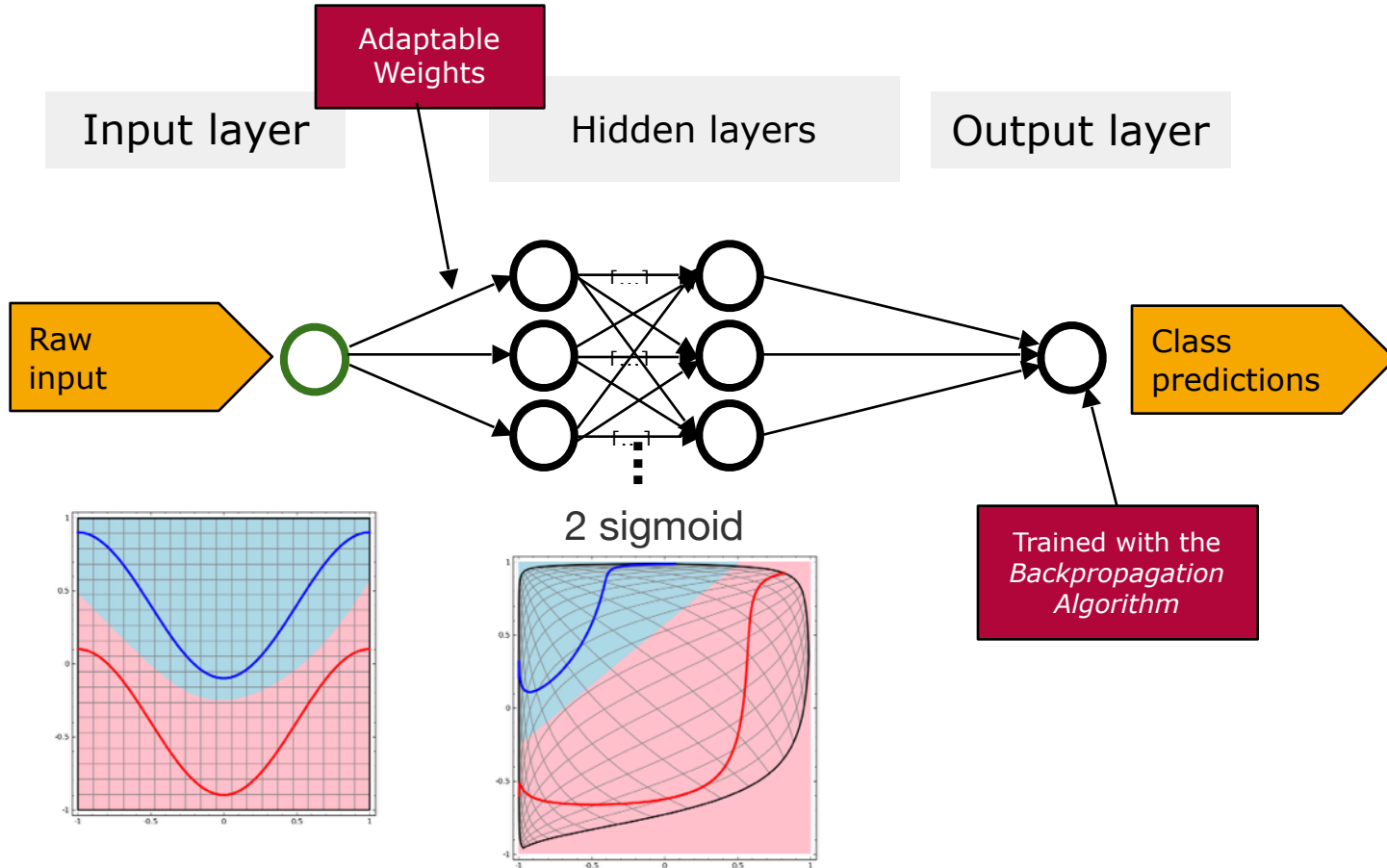
Neural Networks



Neural Networks

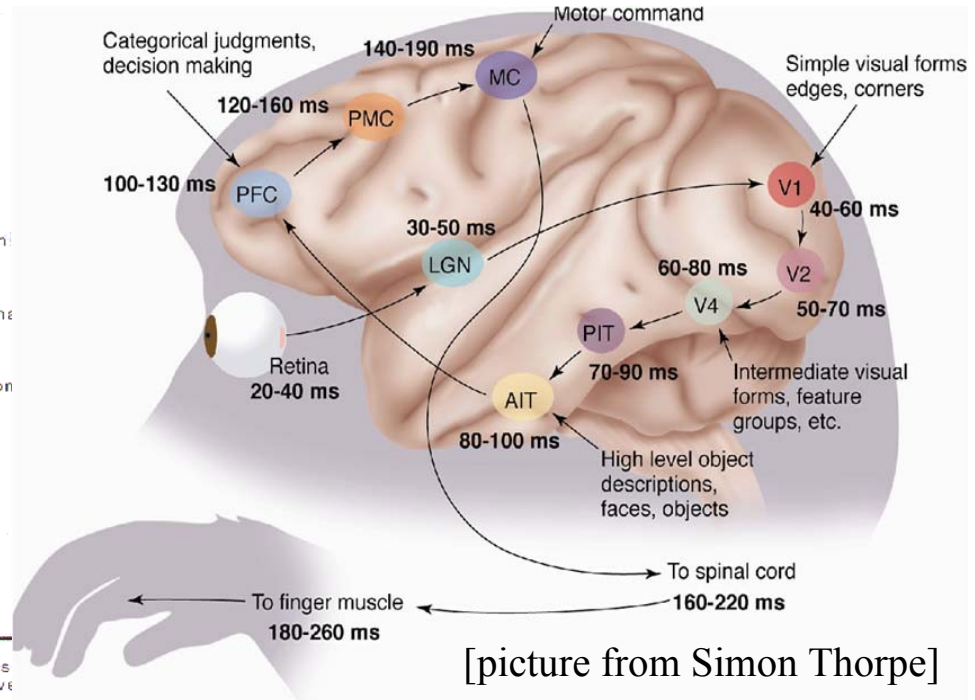
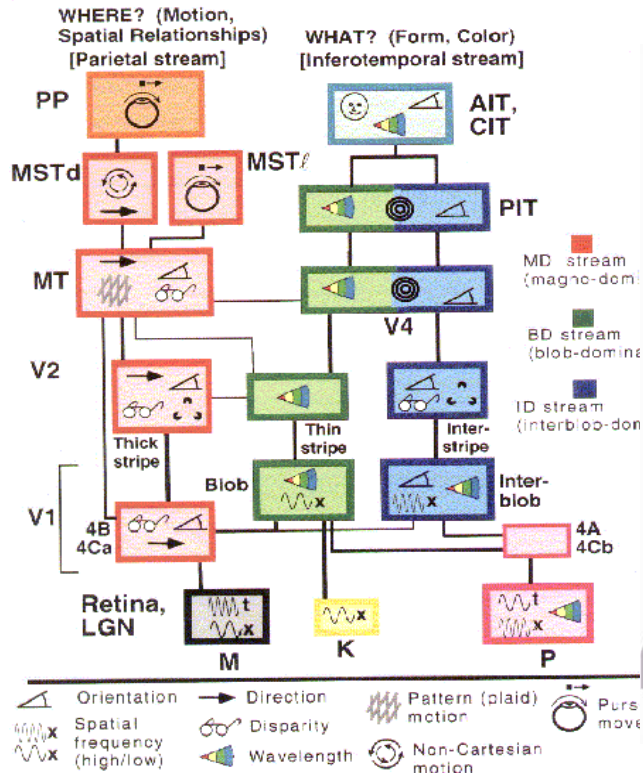


Neural Networks



The Mammalian Visual Cortex is Hierarchical

- The ventral (recognition) pathway in the visual cortex has multiple stages
 - Retina - LGN - V1 - V2 - V4 - PIT - AIT
 - Lots of **intermediate representations**



[picture from Simon Thorpe]

[Gallant & Van Essen]

Deep Visual Features

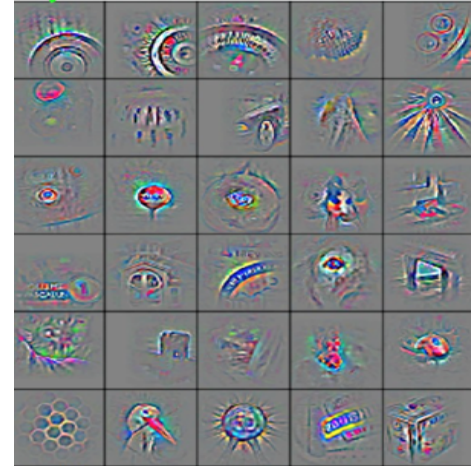
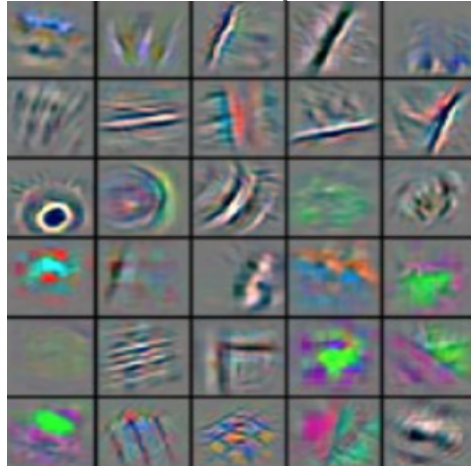


Low-Level Features

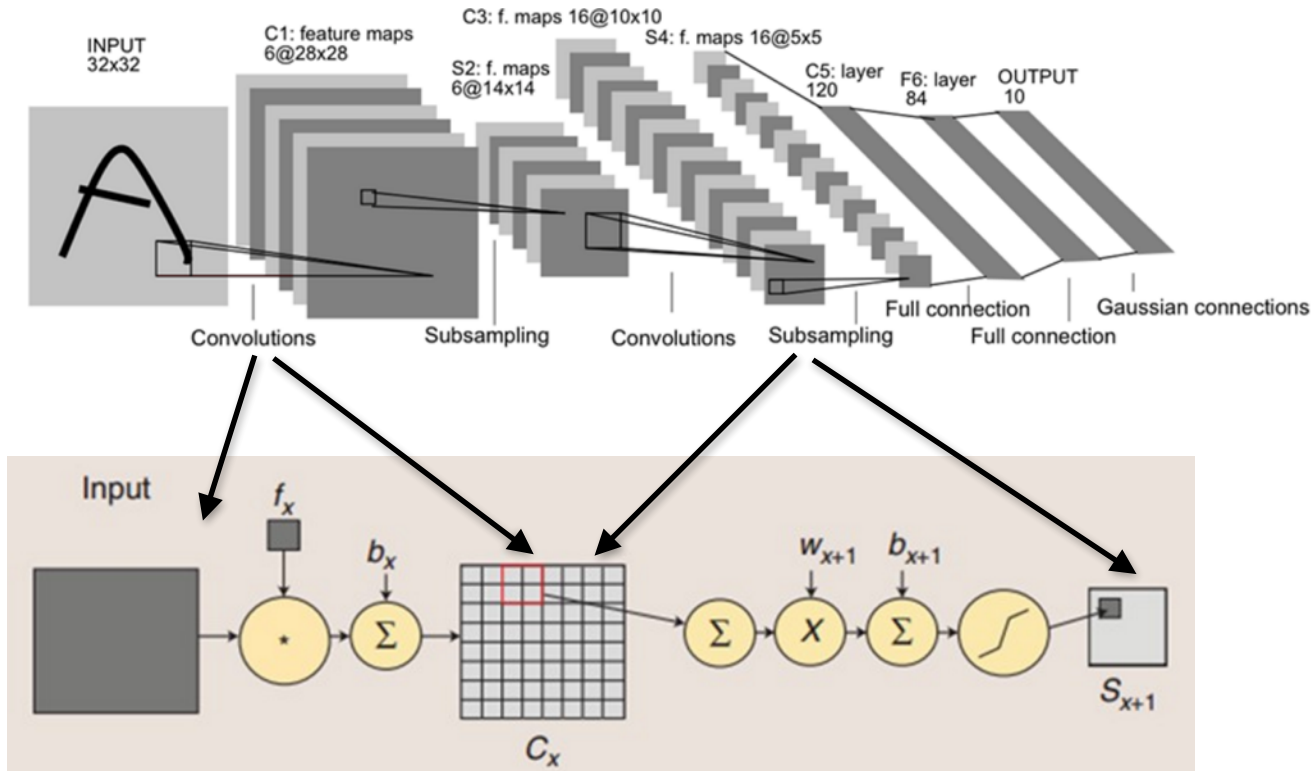
Mid-Level Features

High-Level Features

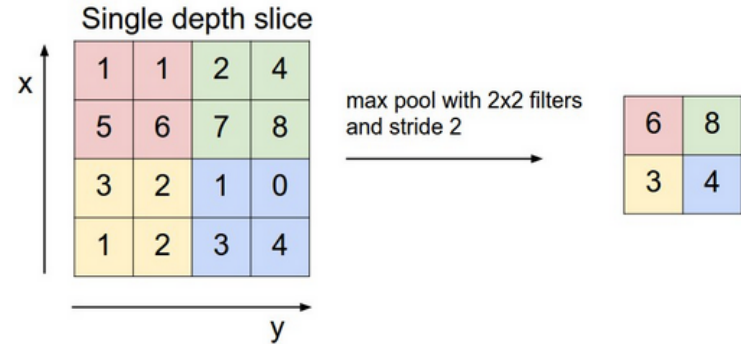
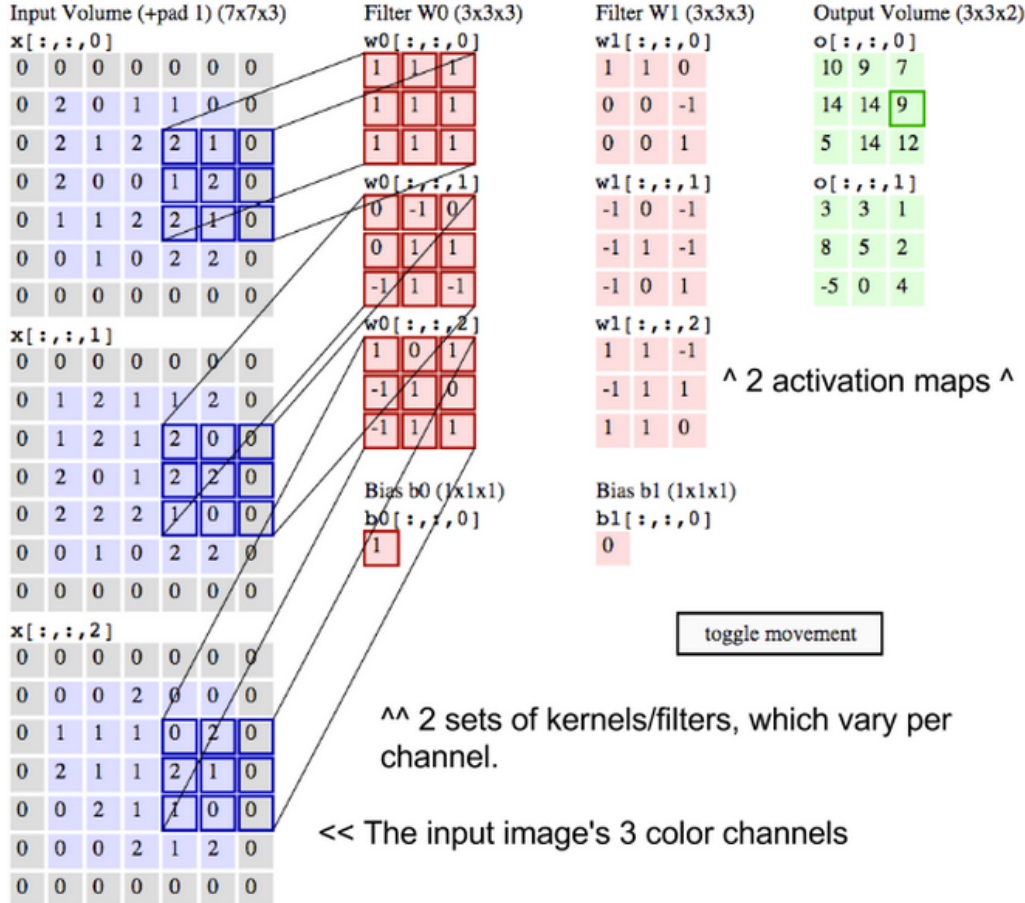
Trainable Classifier



Convolutional Neural Networks



Convolutional Neural Networks



[ConvDemo](#) [TrainDemo](#)

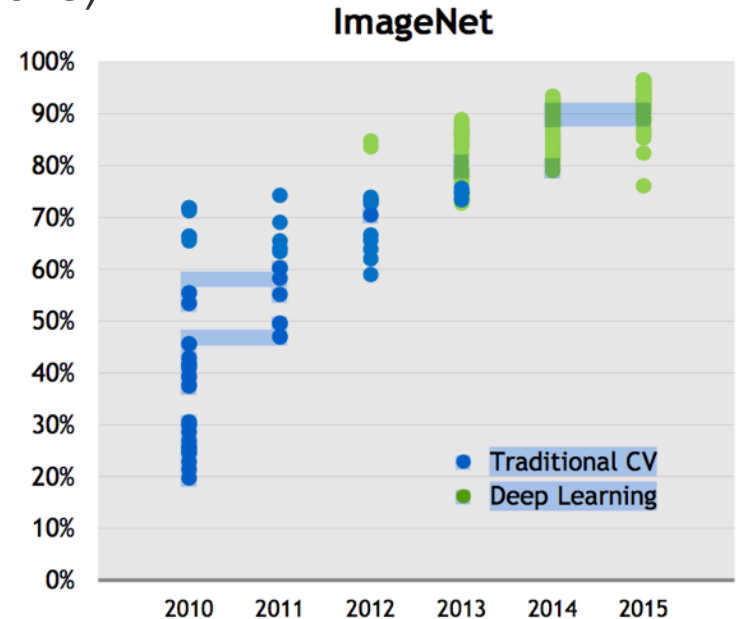
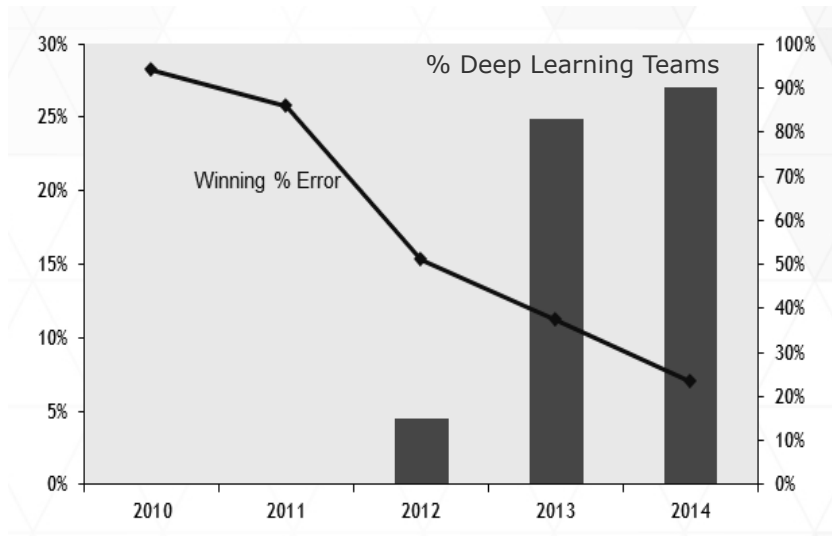
Source from [cs231n](#)

Deep Learning Impact in Research Example

Image classification (ImageNet Challenge)

Given an image, classify what is depicted

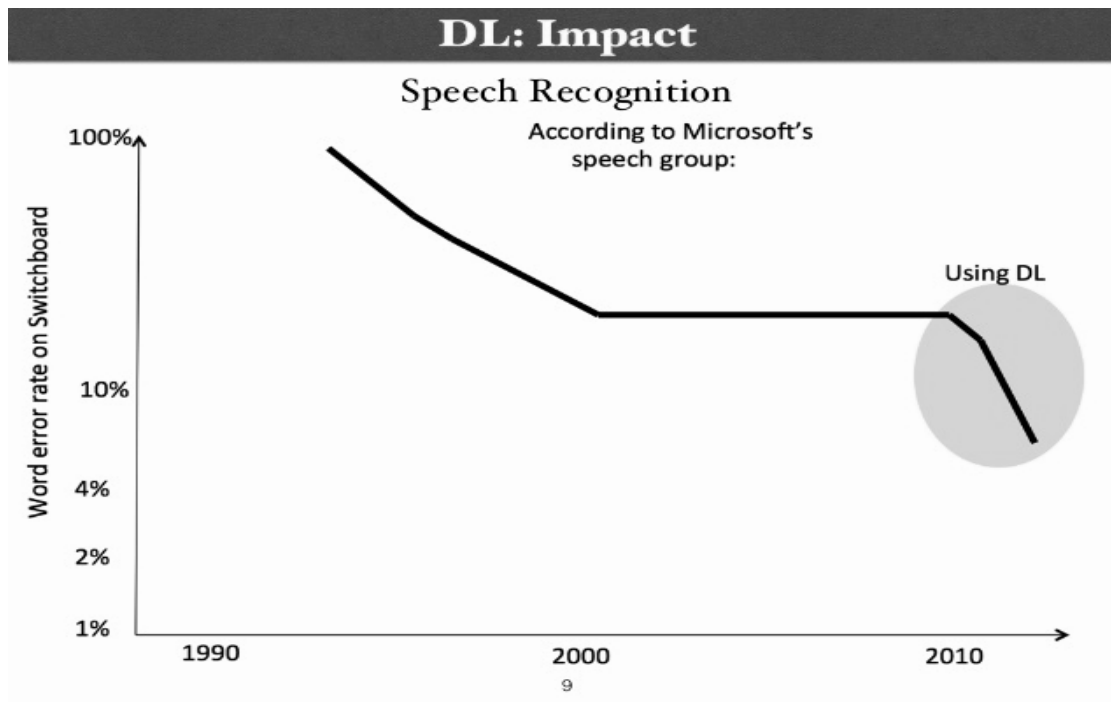
Recent winners: 8-layer AlexNet (2012), 22-layer GoogleNet (Google 2014), 152-layer ResNet (Microsoft 2015)



Deep Learning Impact in Research Example

Speech recognition

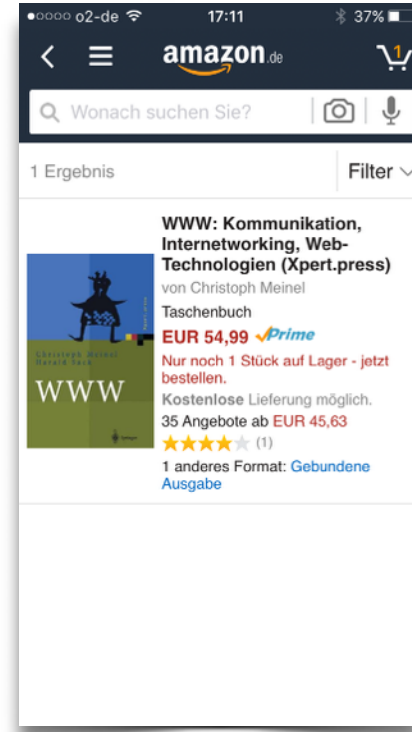
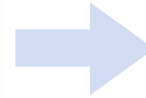
Given an audio file, get word transcription



Machine Vision Applications



Recognize books, barcodes etc.



Current Research Topics

- SceneTextReg: a real-time video text detection and recognition framework using deep CNN and RNN
 - *demonstrated at ACM ICMR'15, IEEE ICASSP'16, ACM Multimedia'16*
- Neural visual translator: Image/video captioning
 - *published at ACM Multimedia'16*
- Human action recognition, event detection in video
 - *published at ICONIP'16*
- Deep semantic retrieval for multimodal data
 - *published at MTAP Journal 2016*
- DL for metrics learning
 - *published at ISVC'16*

Current Research Topics

- DL for text processing, NLP
 - *published at INTERSPEECH'16*
- Video classification with CNNs
 - *published at IJCNN'16*
- Lecture video analysis and retrieval (applied in teleTASK and openHPI)
 - *published at IEEE ICALT'16*
- DL for medical image processing
- Audio analysis with DL

Scene Text Recognition



H O R N B Y
HORNBY

I F
IF

B R O W
BROW

N 2 5
N°25

5 2
52

S H I N I N G
SHINING

0 0 0
0.00

T E A I V I
TEAMI

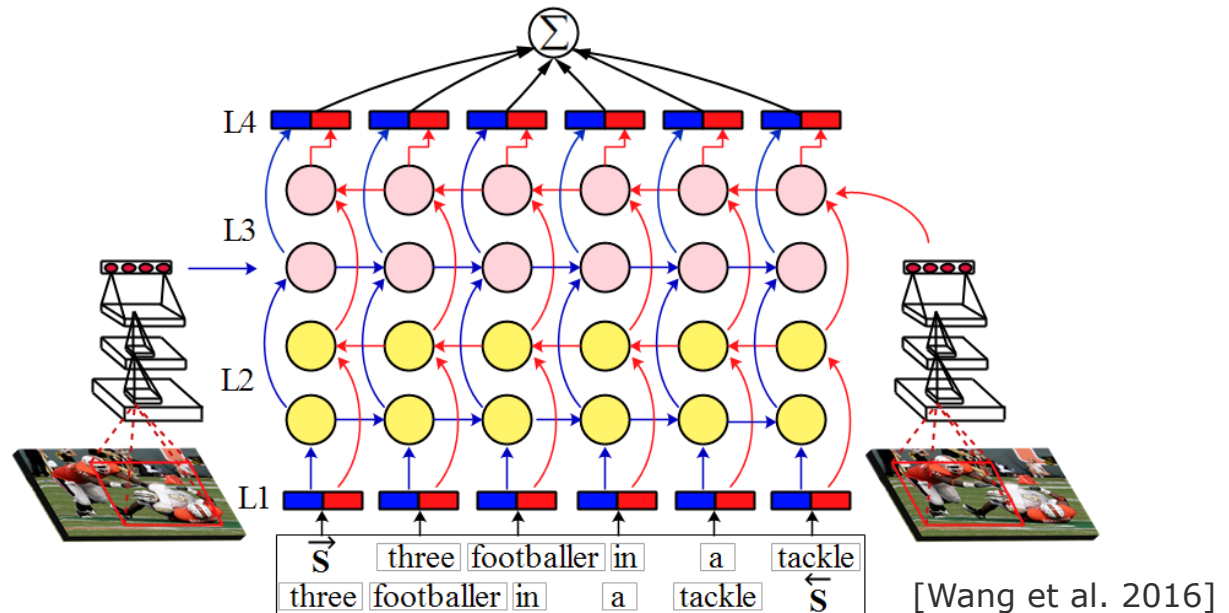
PRIORY
PRIORY

F I R S T
FIRST

Neural Visual Translator

Image/Video Captioning

- Image representation from deep CNN model
- Image to sentence via Bi-directional LSTM (Long short-term memory)
- Achieved state-of-the-art



Video Classification, Activity Detection

Multiple deep neural networks:

- Spatial: recognizing objects on frames
- Temporal: recognizing motion on multiple frames
- Auditory: acoustical information



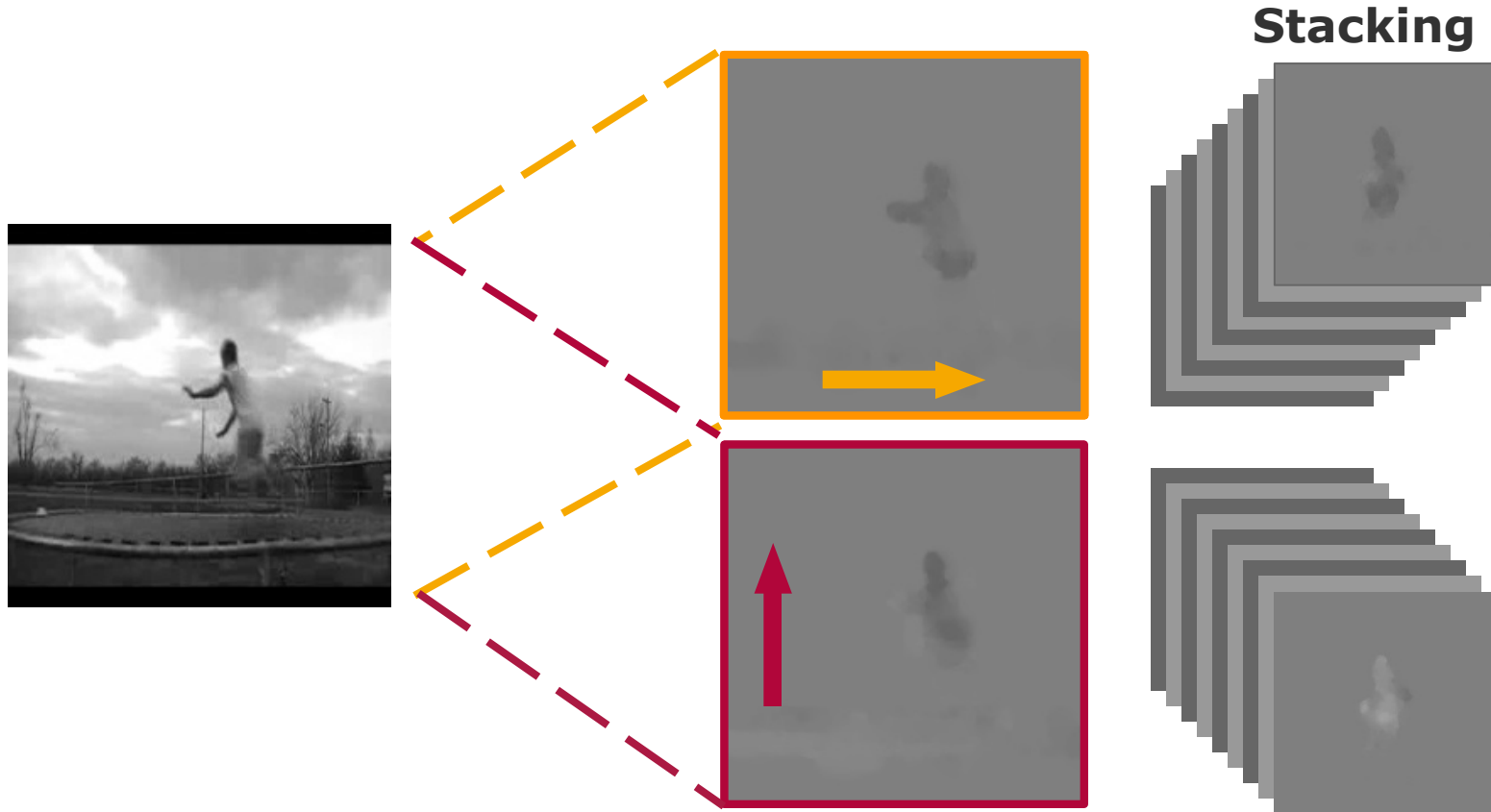
Video Classification, Activity Detection

Multiple deep neural networks:

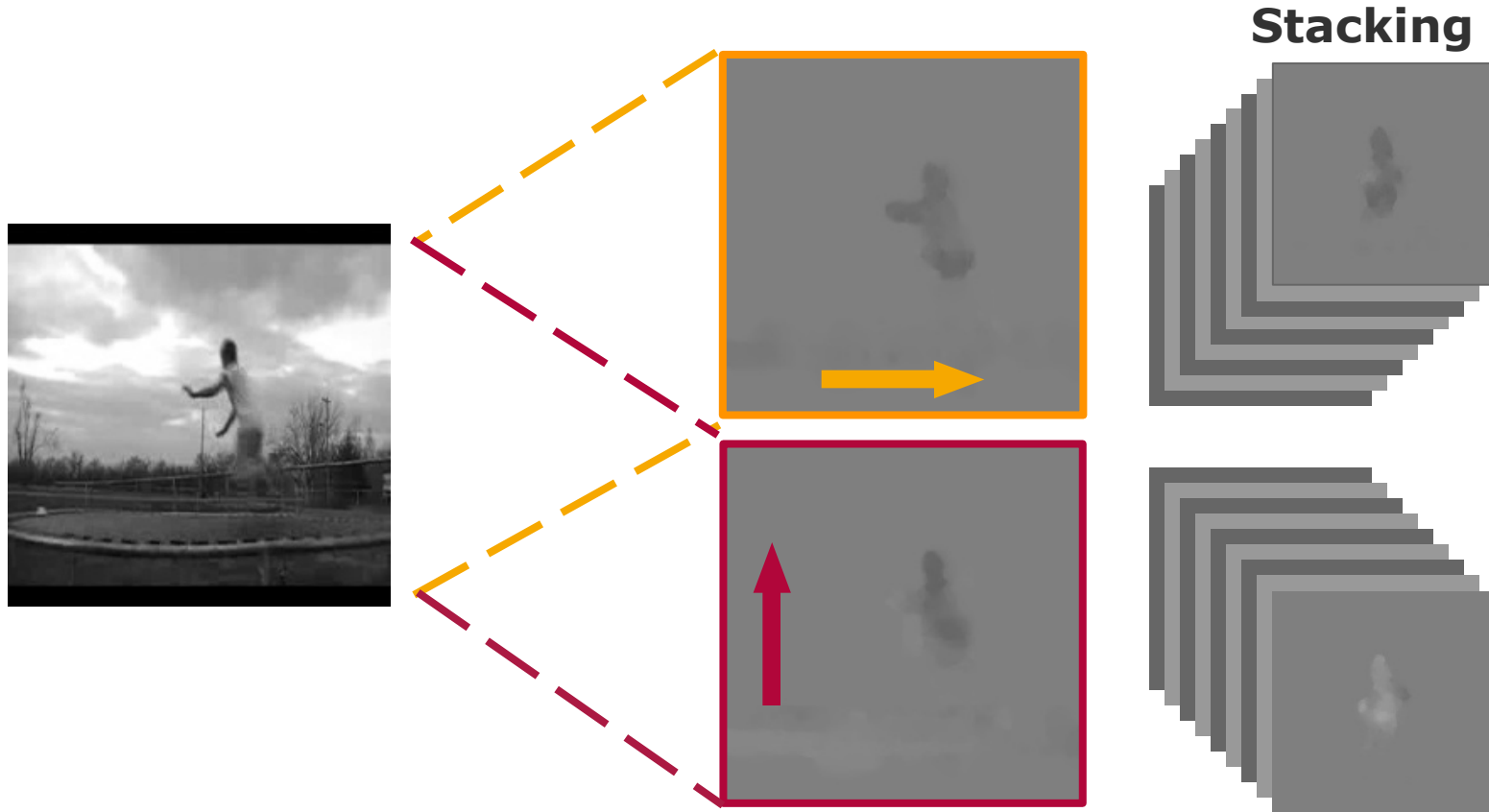
- Spatial: recognizing objects on frames
- Temporal: recognizing motion on multiple frames
- Auditory: acoustical information



Temporal Stream: Dense Optical Flow



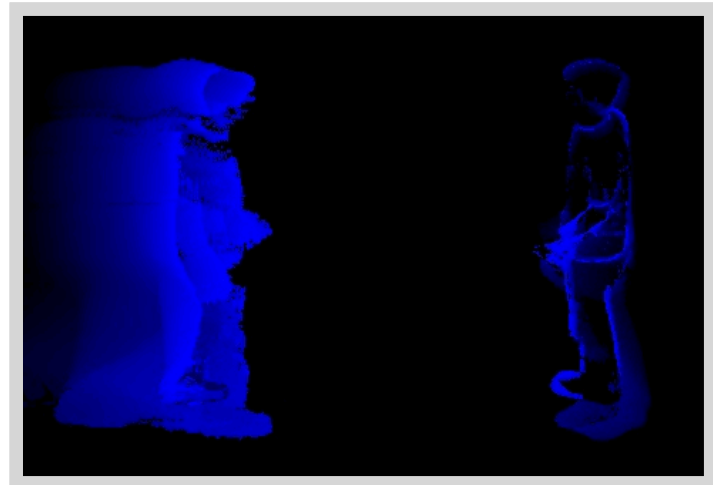
Temporal Stream: Dense Optical Flow



Temporal Stream: Motion History Image

- **Advantages:**

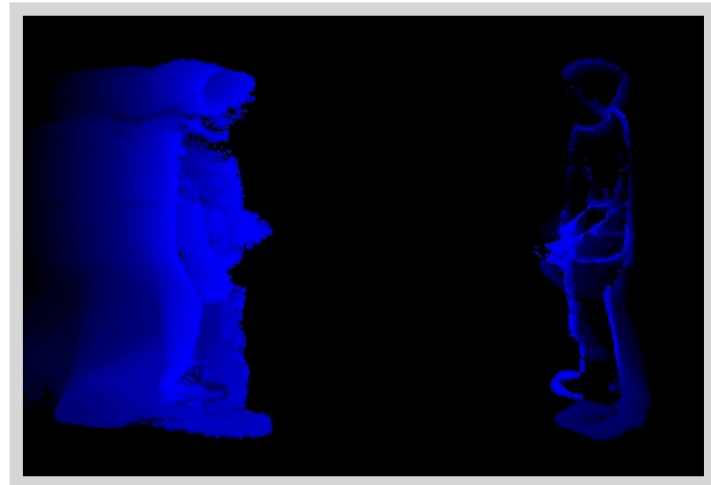
- insensible to the background noise
- representing motion changes in a single image → simplifies the training and prediction process
- low computation cost → real-time application



Temporal Stream: Motion History Image

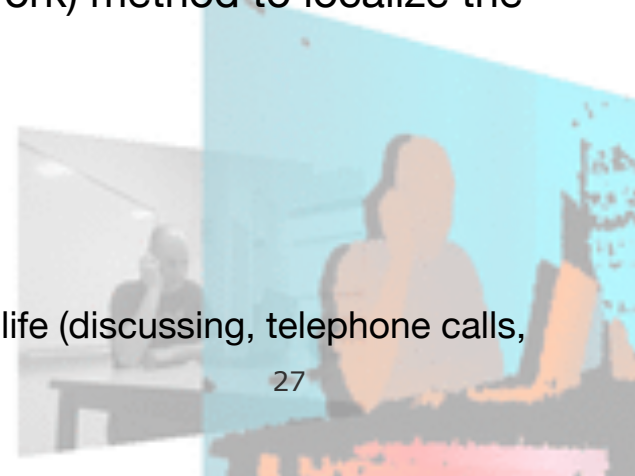
- **Advantages:**

- insensible to the background noise
- representing motion changes in a single image → simplifies the training and prediction process
- low computation cost → real-time application



Topic 1: Indoor Human Activities Recognition

- Core question:
 - How to localize the activity in a static video frame
 - How to capture it in temporal video stream
- Potential solution: two-stream neural networks
 - Faster RCNN (Region based Convolutional Neural Network) method to localize the potential activities in static frames
 - Optical flow or MHI to express motion changes
- Datasets
 - LIRIS dataset (gray/rgb/depth videos), various activities from daily life (discussing, telephone calls, giving an item etc.)



Topic 2: German Word Vectors and Potential Applications

Why:

- **Word Vectors** have been proven to be successful in many NLP apps.
- But the major successes are achieved in English, **not German**.

How:

- Learn the **theoretical background** of Word Vectors.
- Compare the existing WV generation tools and choose the most suitable one.
- Collect as many **German textual data** as possible for the training.
- Test the German word vectors obtained with some **measurements**.
- Apply the German word vectors into **potential applications**.

Challenges:

- The amount of training data (only Wiki dataset is not enough).
- Complicated grammar system, especially the verbs.



Topic 2: German Word Vectors and Potential Applications

Why:

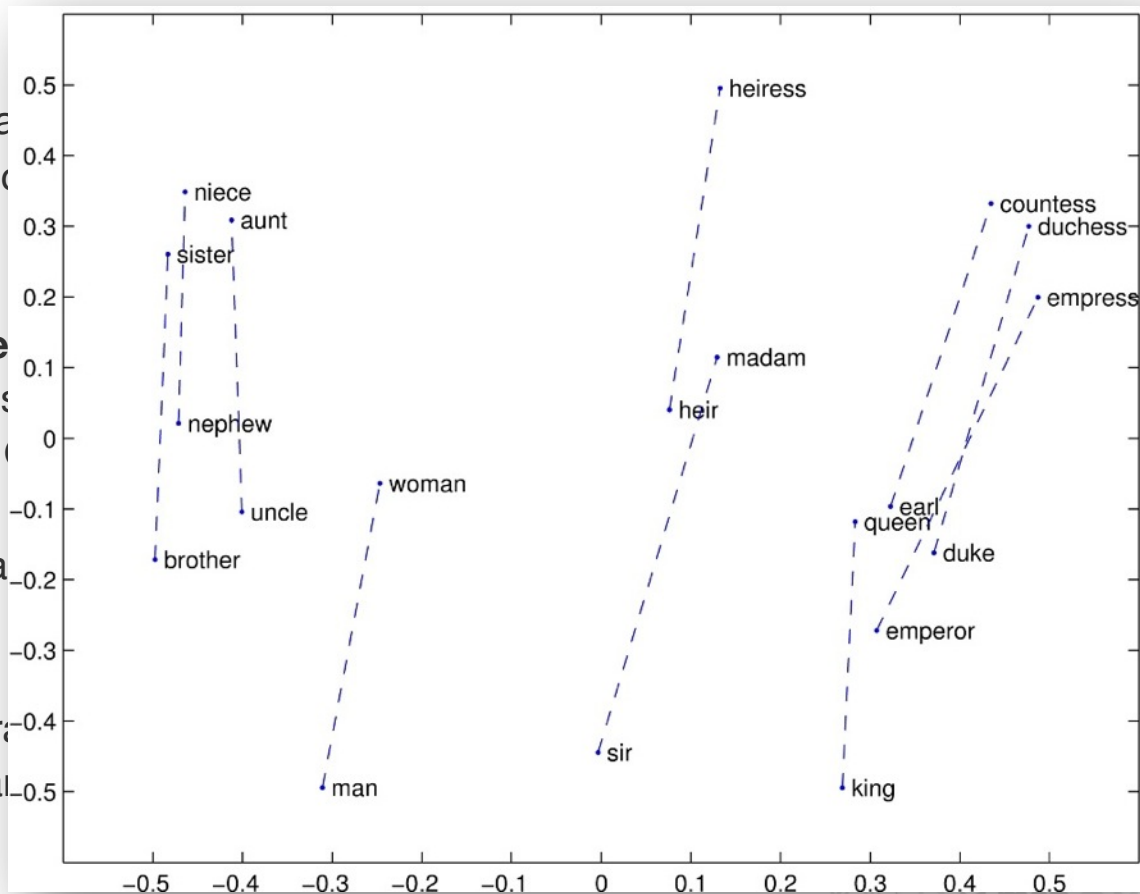
- Word Vectors have many applications
- But the major success is in NLP

How:

- Learn the **theoretical** model
- Compare the existing models
- Collect as many data as possible
- Test the German word vectors
- Apply the German word vectors

Challenges:

- The amount of training data is limited
- Complicated grammar



Topic 2: German Word Vectors and Potential Applications

Why:

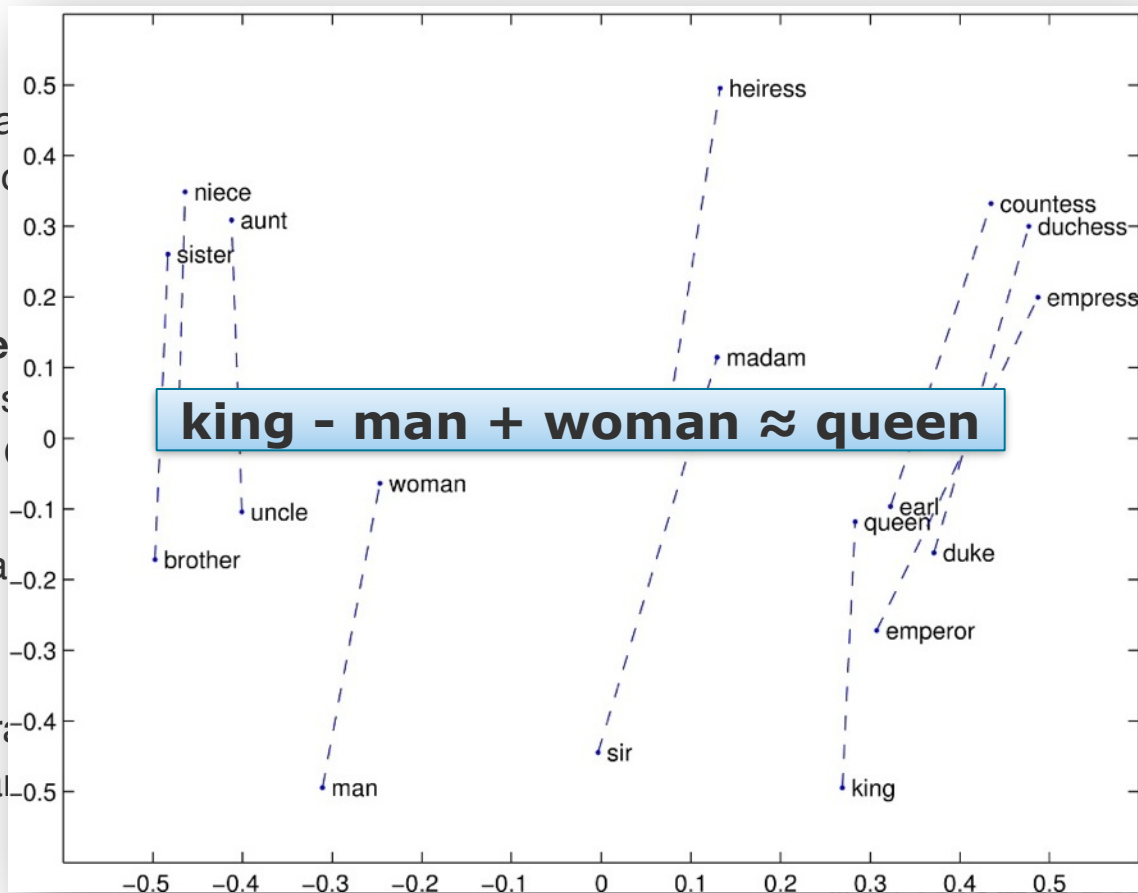
- Word Vectors have many applications
- But the major success is in machine translation

How:

- Learn the **theoretical** word relationships
- Compare the existing word vectors
- Collect as many word pairs as possible
- Test the German word vectors
- Apply the German word vectors

Challenges:

- The amount of training data is limited
- Complicated grammatical structures



Topic 2: German Word Vectors and Potential Applications

Why:

- **Word Vectors** have been proven to be successful in many NLP apps.
- But the major successes are achieved in English, **not German**.

How:

- Learn the **theoretical background** of Word Vectors.
- Compare the existing WV generation tools and choose the most suitable one.
- Collect as many **German textual data** as possible for the training.
- Test the German word vectors obtained with some **measurements**.
- Apply the German word vectors into **potential applications**.

Challenges:

- The amount of training data (only Wiki dataset is not enough).
- Complicated grammar system, especially the verbs.



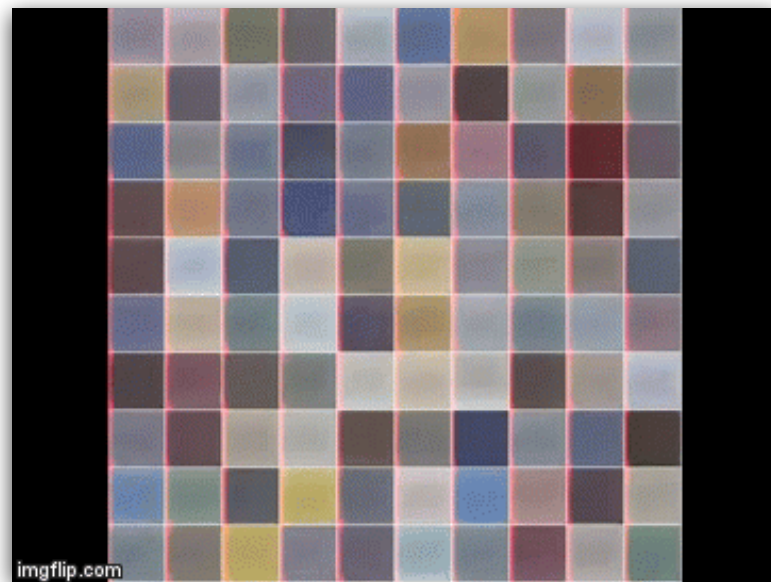
Topic 3: Deep Network For Image Generation

Motivation:

- Deep learning systems need huge amounts of training data
- Getting training data for deep learning is difficult

Possible Solution:

- System that automatically generates training images
- Such a system could be based on:
 - Attention modeling
 - Recurrent Neural Networks



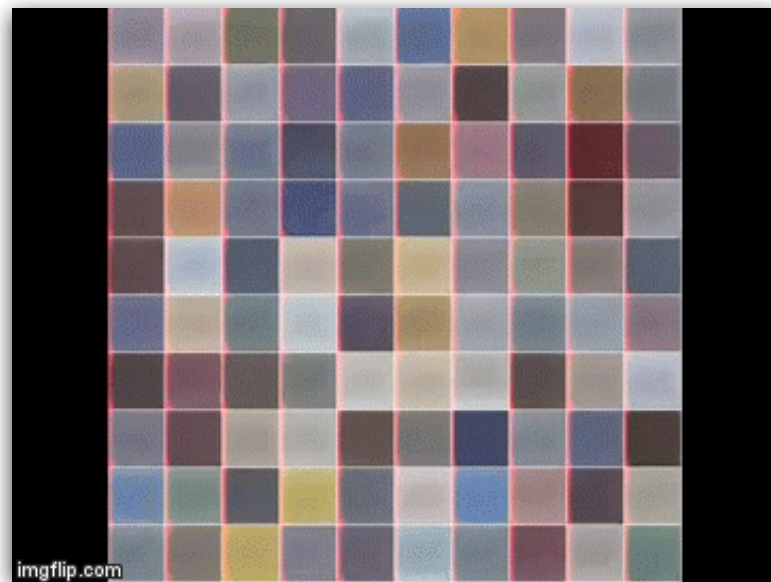
Topic 3: Deep Network For Image Generation

Motivation:

- Deep learning systems need huge amounts of training data
- Getting training data for deep learning is difficult

Possible Solution:

- System that automatically generates training images
- Such a system could be based on:
 - Attention modeling
 - Recurrent Neural Networks



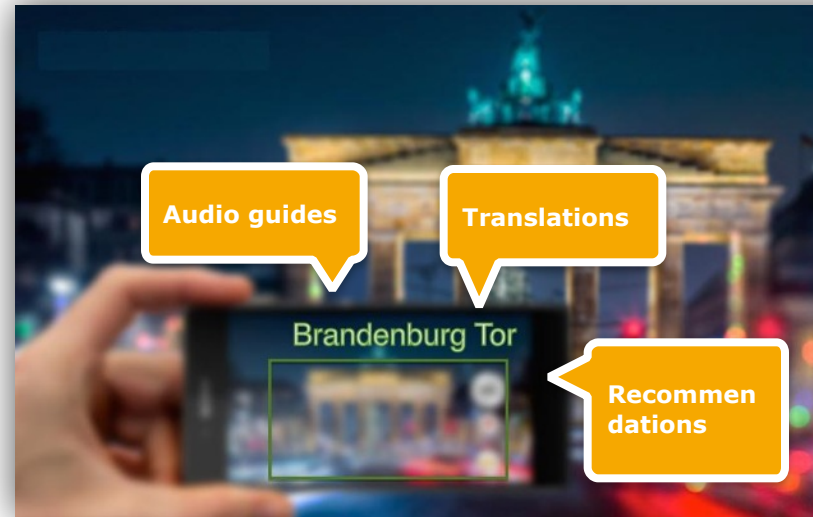
Topic 4: Place Recognizer

Idea:

- Build an App with machine vision feature, e.g. place recognition in real-time using android phone
- Training images and information retrieval from Google maps and flickr
- Apply deep model to extract visual feature for place recognition
- Recommendations and useful features, z.B. audio guides, translations...
- More idea from you...

Your participation:

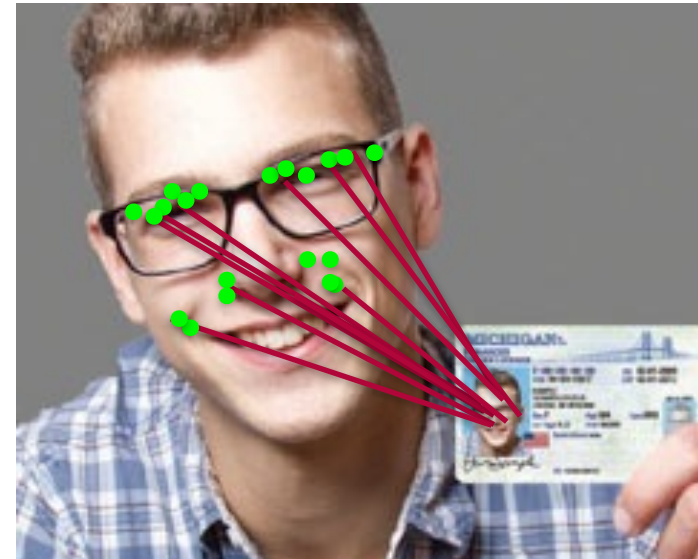
- Learning knowledges of deep learning,
- Apply deep learning technology to mobile application
- Contribute to software design and development



Topic 5: Deep Face Representation

Face representation with CNN

- Workflow:
 - Face detection -> **frontal face alignment** -> facial representation -> classification
 - Deep face model learning -> robust face representation ✓
 - Demo app for face identification
 - e.g. **Android app (unlock screen?)**
- Datasets
 - CASIA-WebFace dataset (train): 10k subjects, 490k images
 - LFW dataset (test): 5.7k subjects, 13k images
- Difficulties:
 - Lighting effect, blur problem
 - Multi-scale
 - Geometrical distortion

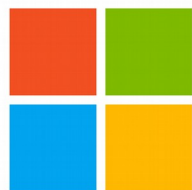


Tools and Hardware

- Caffe: deep Learning framework by Berkeley vision lab
- Chainer: a flexible framework of neural networks
- Google's TensorFlow
- CNNdroid: open source library for GPU-accelerated execution of trained deep convolutional neural networks on Android
- Chair's GPU Server



ENCNP



Leistungserfassung

- The final evaluation will be based on:
 - Initial implementation / idea presentation, 10% (Anfang Dezember)
 - Final presentation, 20% (09.02.2017)
 - Report/Documentation, 12-18 pages (single column), 30% (bis Ende Februar)
 - Implementation, 40% (bis Ende Februar)
 - Participation in the seminar (bonus points)
- Wahl der Themen **bis 27.10.16**: anmelden on Doodle (verlinkt im HPI website der Lehrveranstaltung)

Leistungserfassung

- The final evaluation will be based on:
 - Initial implementation / idea presentation, 10% (Anfang Dezember)
 - Final presentation, 20% (09.02.2017)
 - Report/Documentation, 12-18 pages (single column) 20% (bis Ende Februar)
 - Implementation, 40% (bis Ende Februar)
 - Participation in the seminar (bonus points)
- Wahl der Themen **bis 27.10.16**: anmelden (verlinkt im HPI website der Lehrveranstaltung)



Leistungserfassung

- The final evaluation will be based on:
 - Initial implementation / idea presentation, 10% (Anfang Dezember)
 - Final presentation, 20% (09.02.2017)
 - Report/Documentation, 12-18 pages (single column), 30% (bis Ende Februar)
 - Implementation, 40% (bis Ende Februar)
 - Participation in the seminar (bonus points)
- Wahl der Themen **bis 27.10.16**: anmelden on Doodle (verlinkt im HPI website der Lehrveranstaltung)

Ansprechpartner

Dr. Haojin Yang

Senior Researcher

Office: H-1.22

Phone: +49 (0)331-5509-511

Email: haojin.yang@hpi.de

Xiaoyin Che, M.sc

PhD Student

Office: H-1.22

Email: xiaoyin.che@hpi.de

Sheng Luo, M.sc

PhD Student

Office: H-1.21

Email: sheng.luo@hpi.de

Christian Bartz, M.sc

PhD Student

Office: H-1.11

Email: chrisitan.bartz@hpi.de

Thank you for your ATTENTION!