Master Seminar:
Practical Applications of Multimedia Retrieval

Hasso Plattner Institute
Dr. Haojin Yang

20.10.2016

Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

- **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:

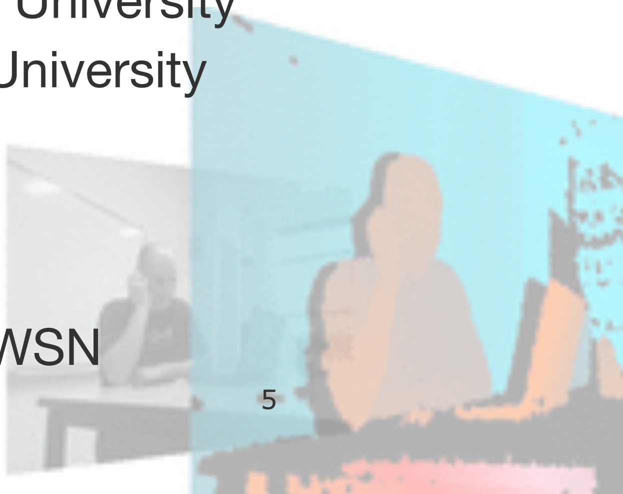
- Document Analysis
- Deep Learning
- Natural Language Processing
- E-Learning

Personal Information

Sheng Luo, M.sc

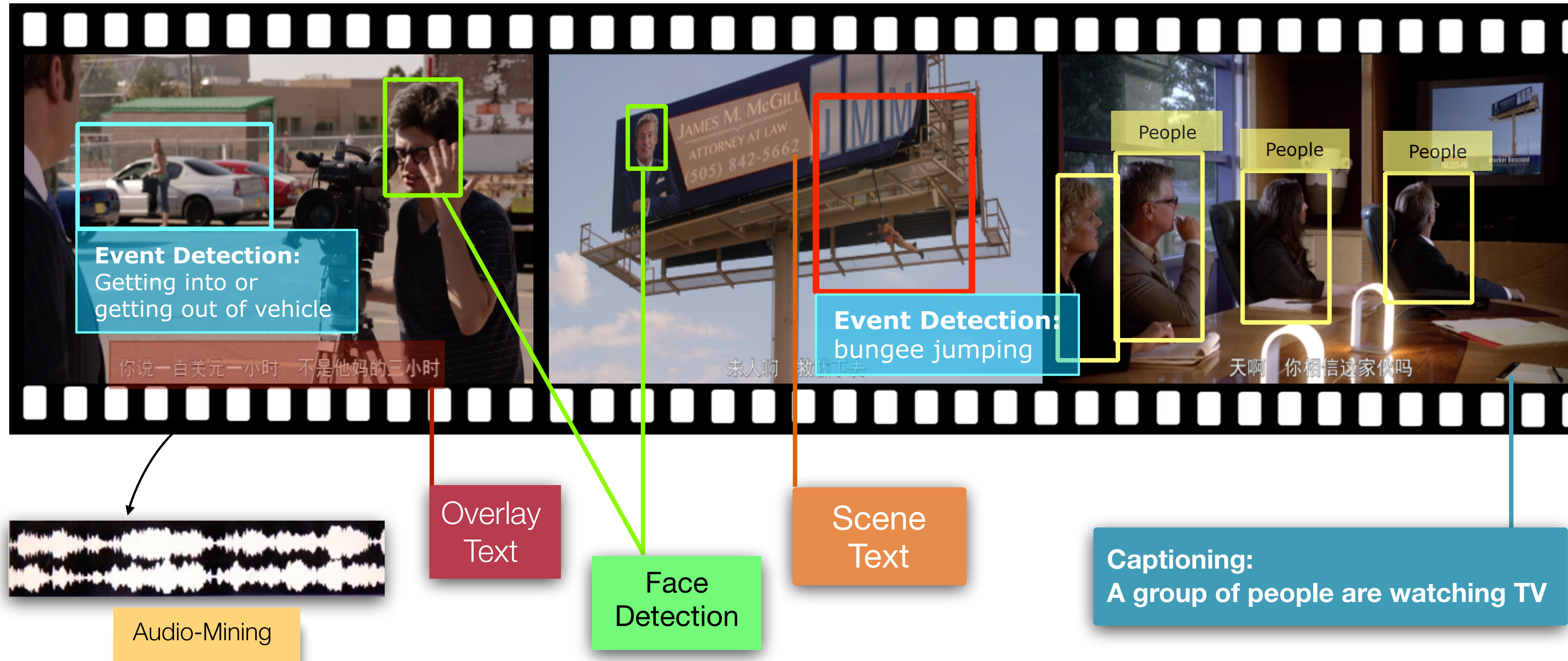
- Research background
  - 2011.09-2014.03 Master of Engineering, Shanghai University
  - 2012.09-2013.09 Master of Engineering, Waseda University
  - 2014.4-now PhD student at HPI

- Research interests
  - Multimedia Retrieval, Deep learning, Robotic and WSN

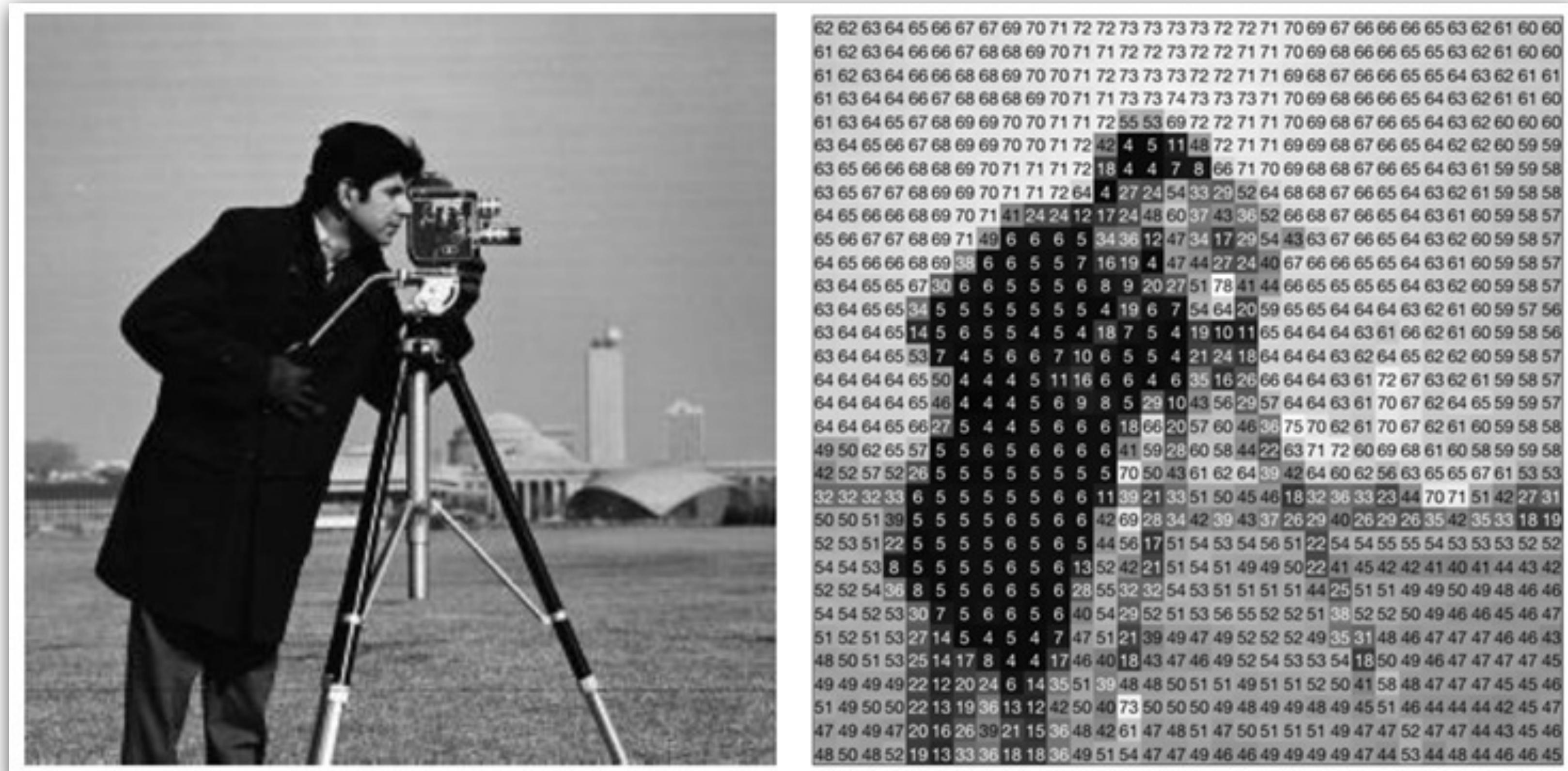


- 400 hours of video uploaded every minute
- Video data —> more than 64% of internet traffics (2014), will be more than 80% in 2019

Automatic Multimedia Analysis



Why Machine Vision So Hard



* *Image copyrights reserved by Facebook AI lab*

Deep Learning for Multimedia Retrieval

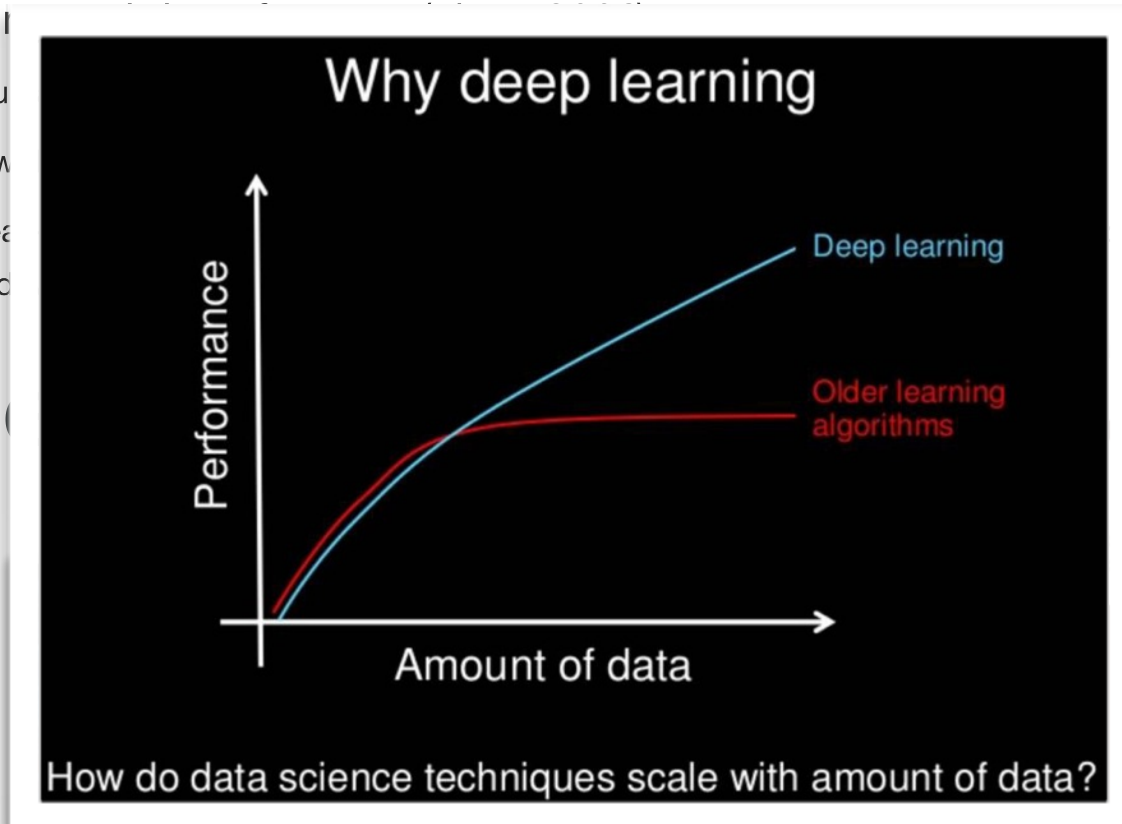
- Deep Learning and deep features (*since 2006*):
  - Simulating human neural network and hierarchically learning features from large scale data
  - Impacting a wide range of multimedia information processing
  - Achieved break-record results in fields like *Speech Recognition*, *Image Classification*,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

Deep learning
as human beings

Deep Learning for Multimedia Retrieval

- Deep Learni
  - Simulating hu
  - Impacting a w
  - Achieved brea
    Detection and

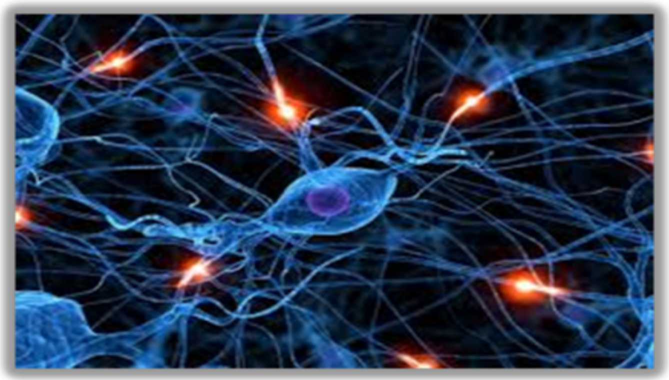
D



Deep Learning for Multimedia Retrieval

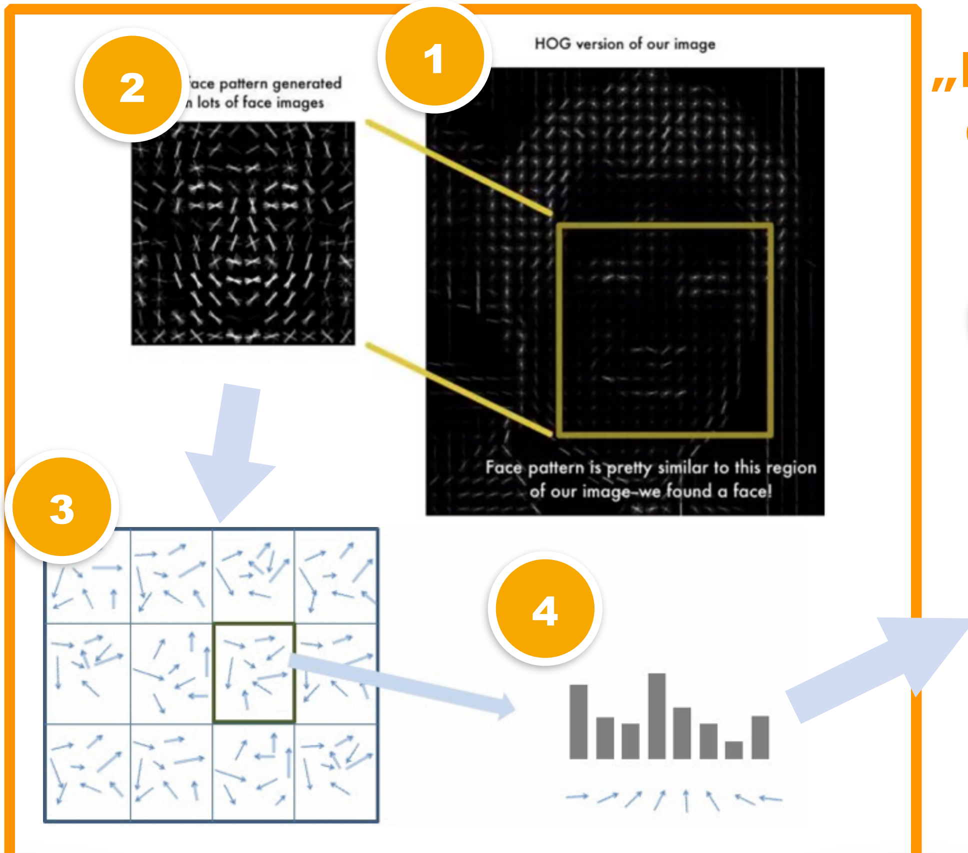
- Deep Learning and deep features (*since 2006*):
  - Simulating human neural network and hierarchically learning features from large scale data
  - Impacting a wide range of multimedia information processing
  - Achieved break-record results in fields like *Speech Recognition*, *Image Classification*,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

Deep learning
as human beings



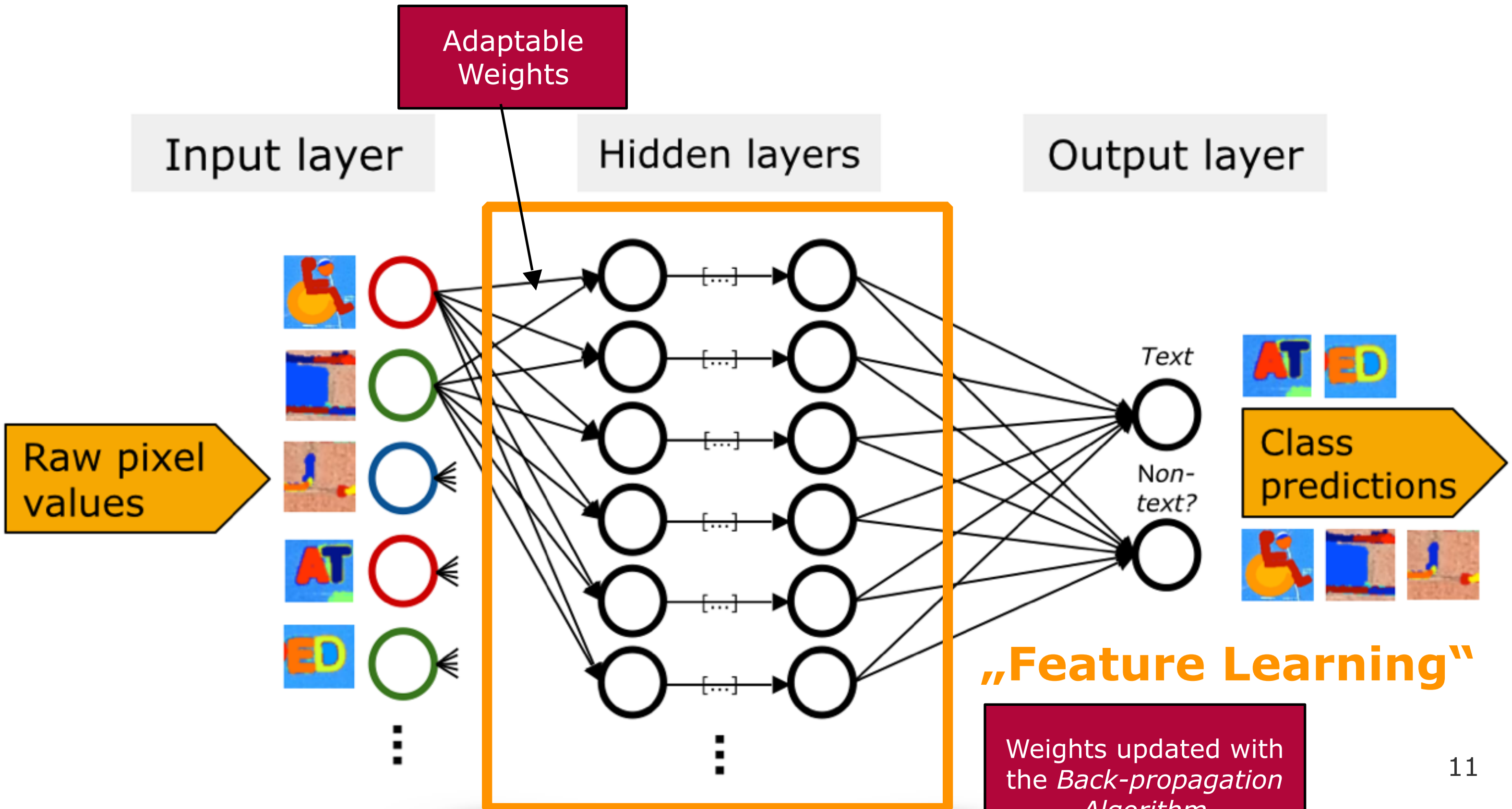
Handcrafted Features Example: HOG

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



„Feature Engineering“
designed by Expert

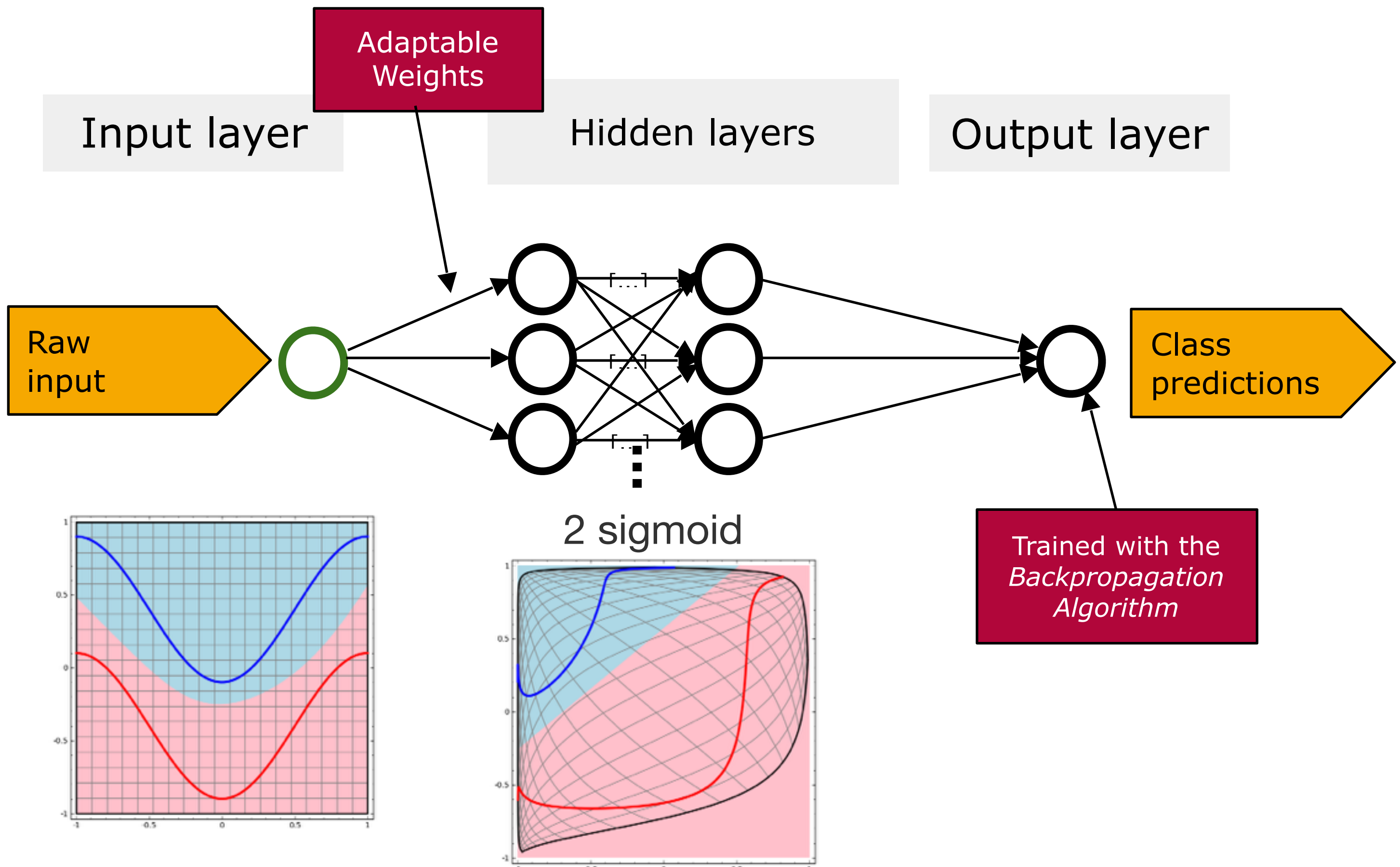
Artificial Neural Networks



„Feature Learning“

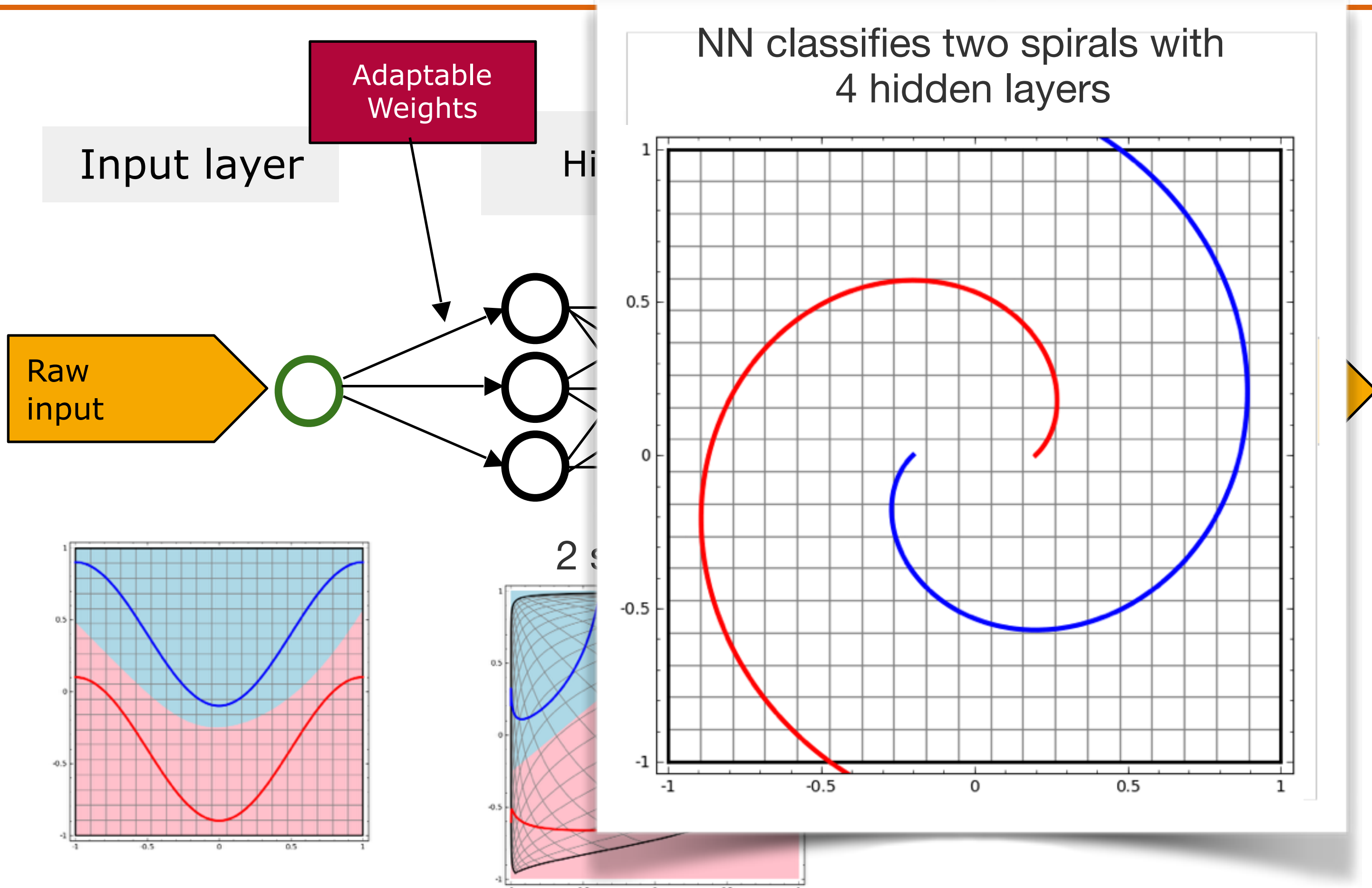
Weights updated with the *Back-propagation Algorithm*

Neural Networks

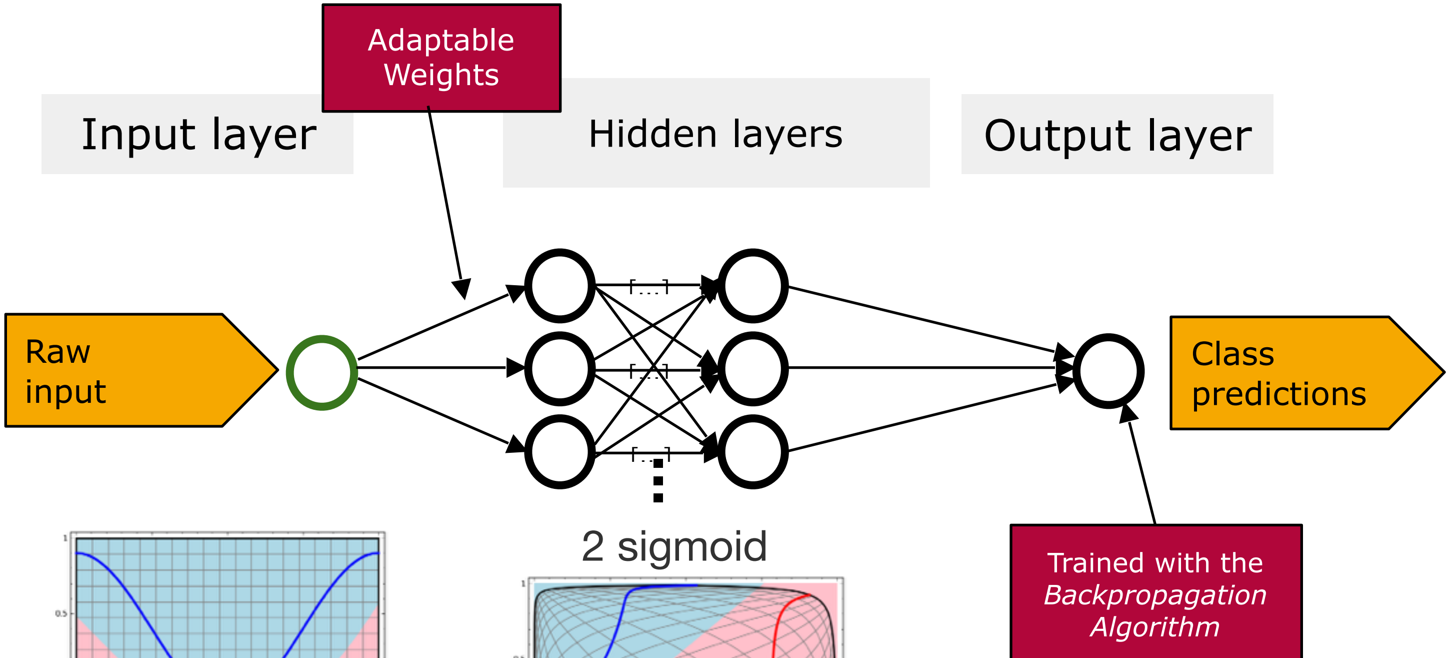


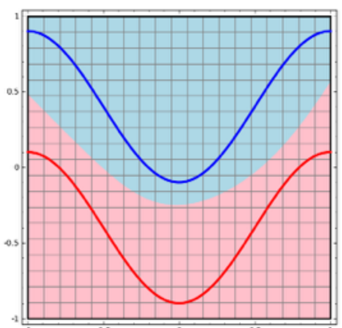
Hasso Plattner Institut

Neural Networks

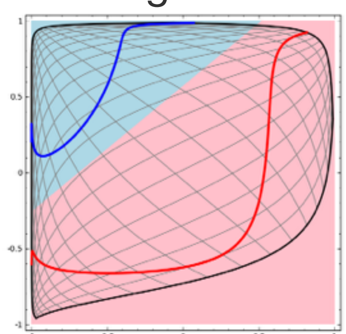


Neural Networks



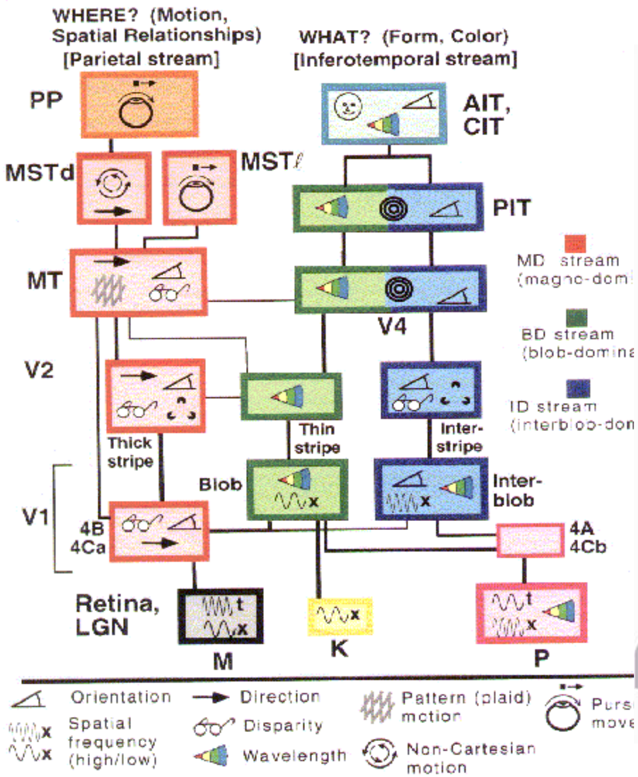


2 sigmoid

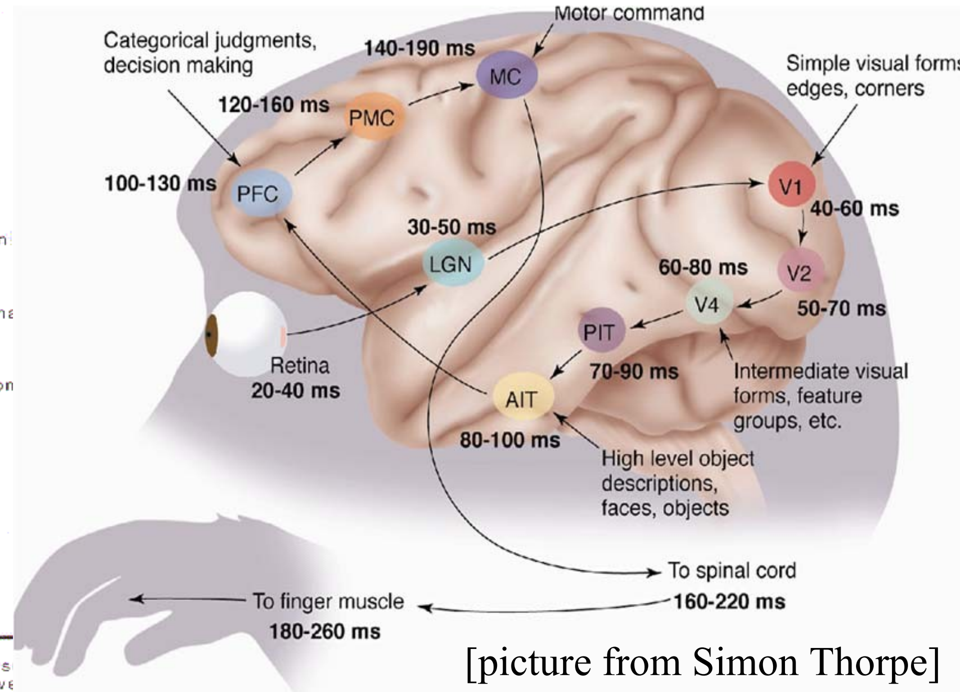


The Mammalian Visual Cortex is Hierarchical

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

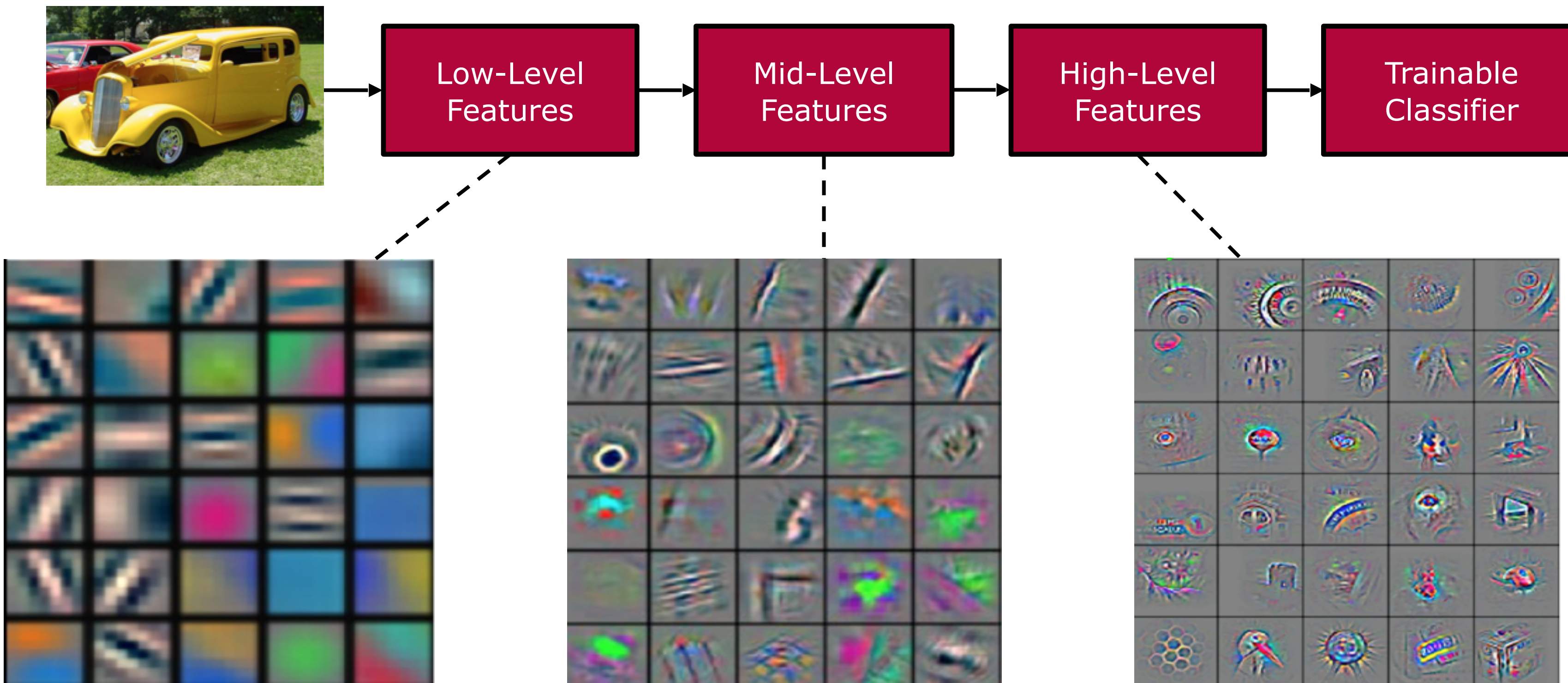


[Gallant & Van Essen]

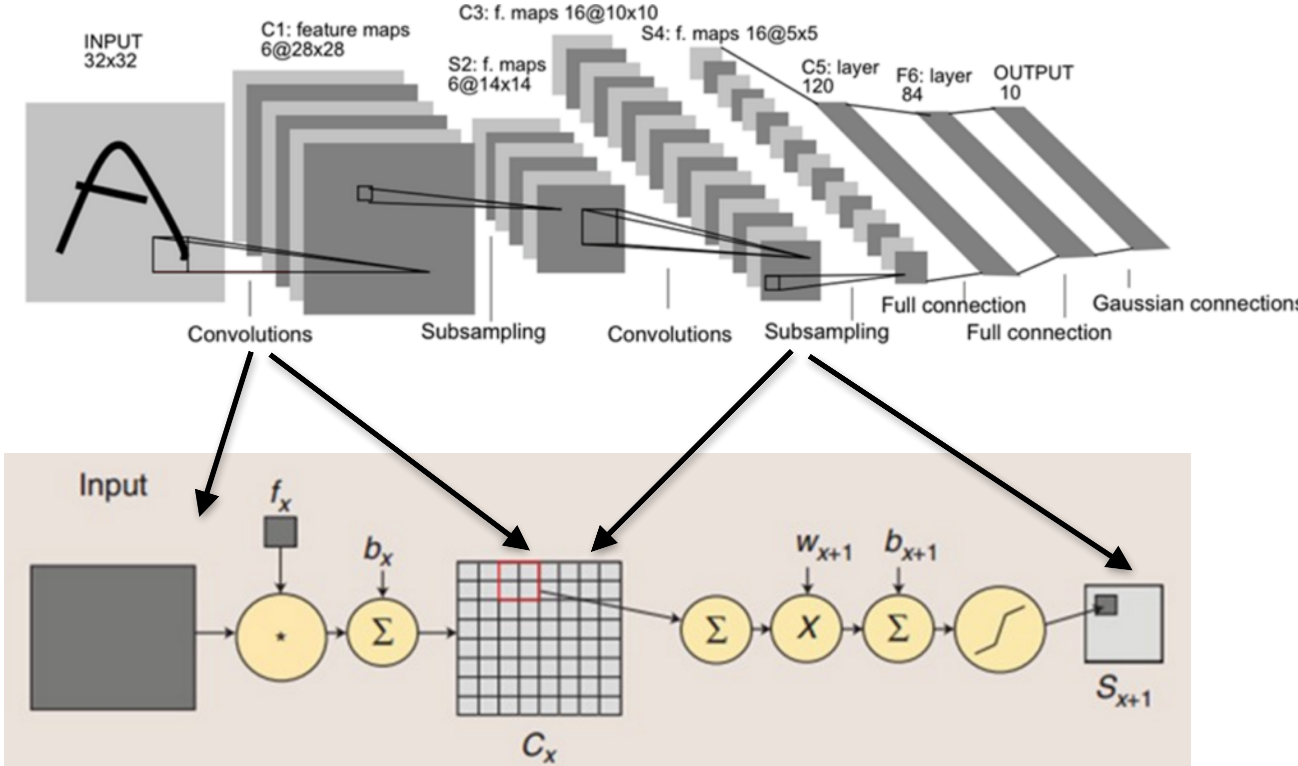


[picture from Simon Thorpe]

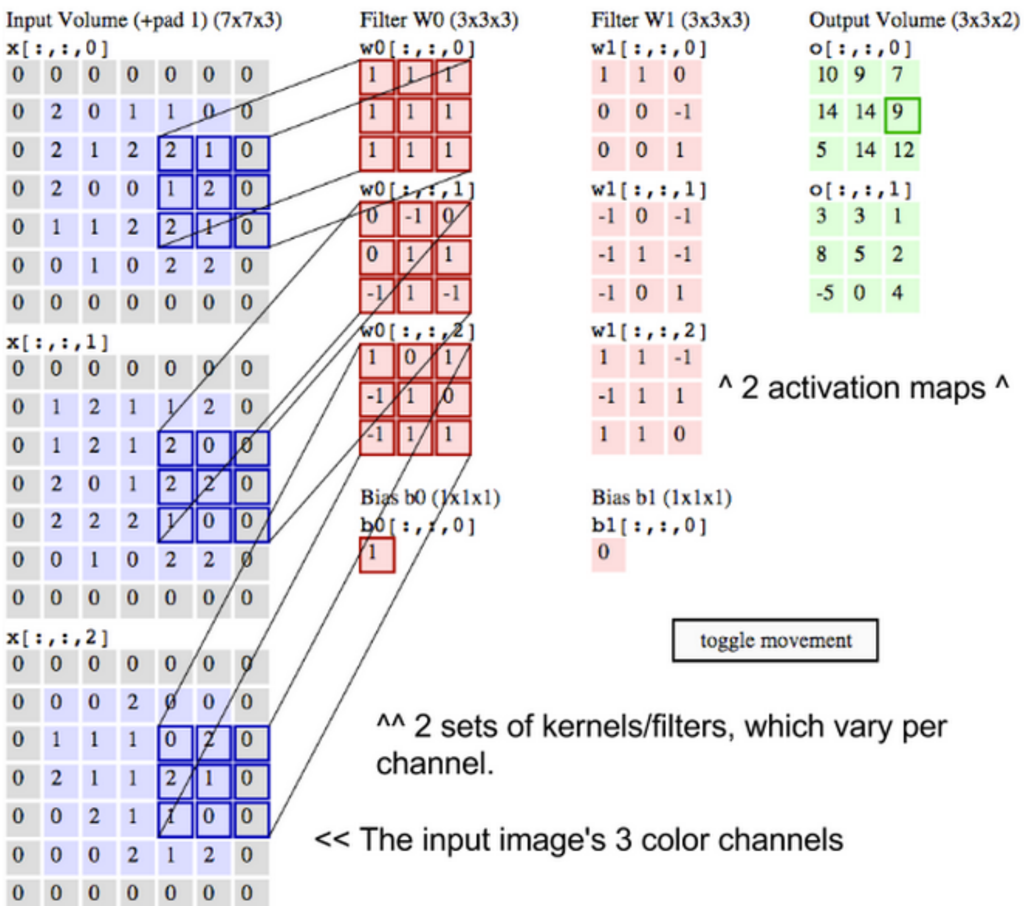
Deep Visual Features

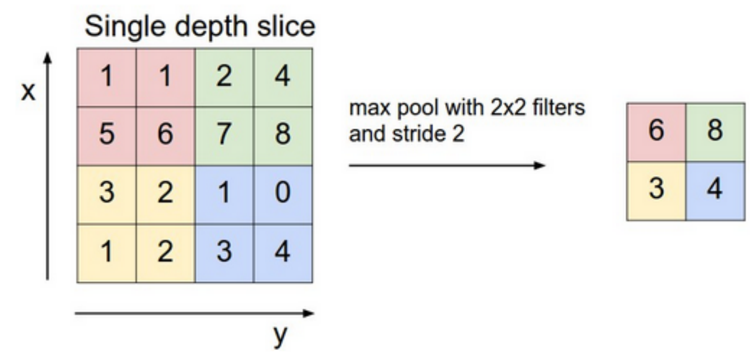


Convolutional Neural Networks



Convolutional Neural Networks





ConvDemo TrainDemo

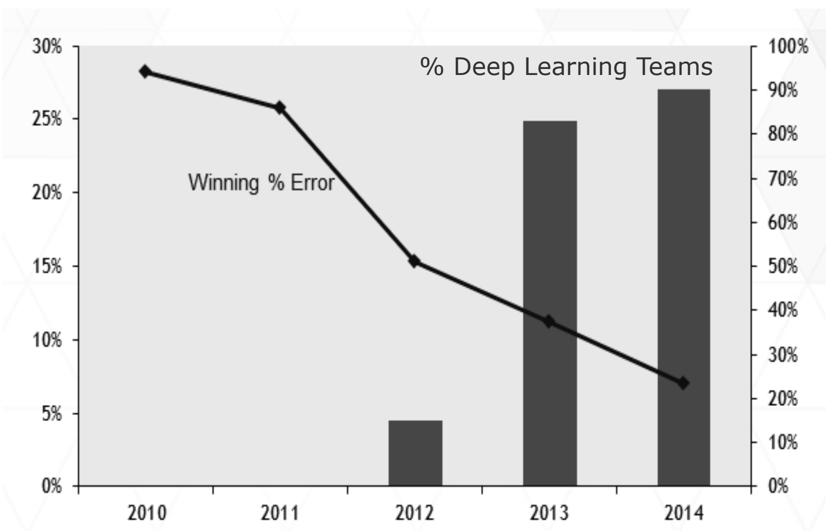
Source from cs231n

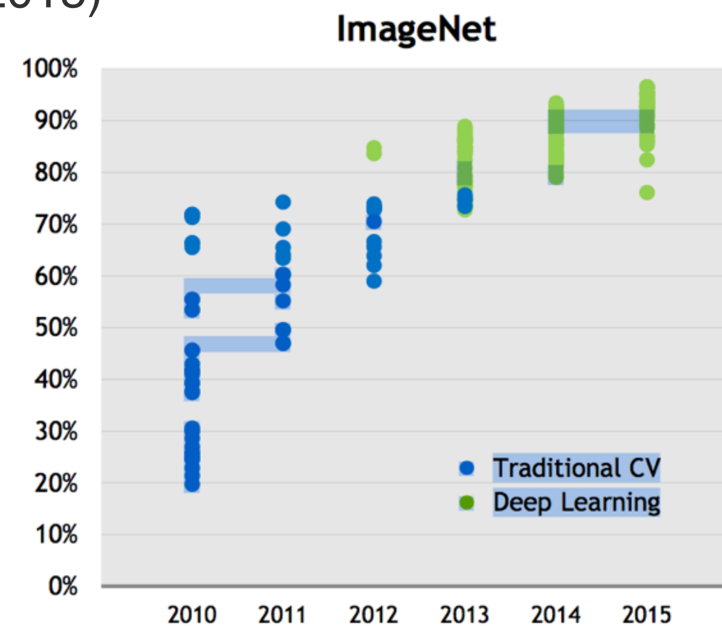
Deep Learning Impact in Research Example

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)

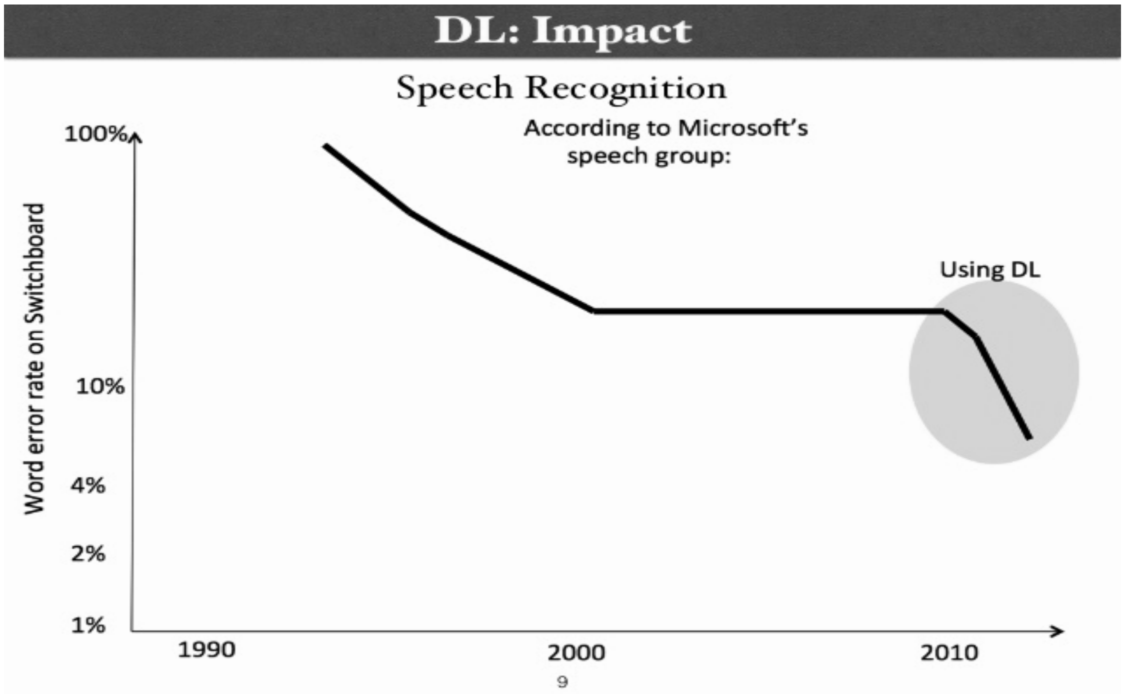




Deep Learning Impact in Research Example

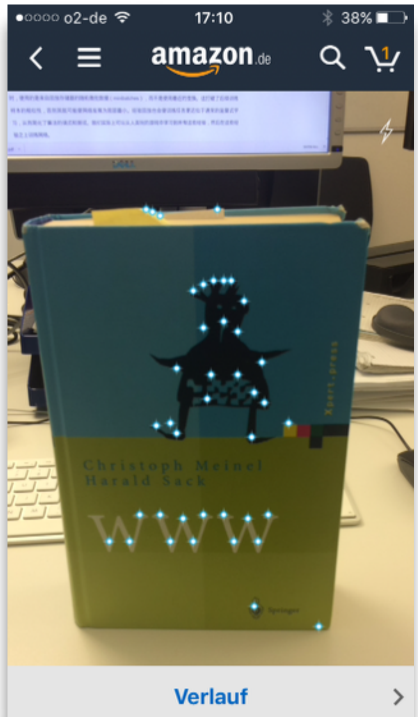
Speech recognition

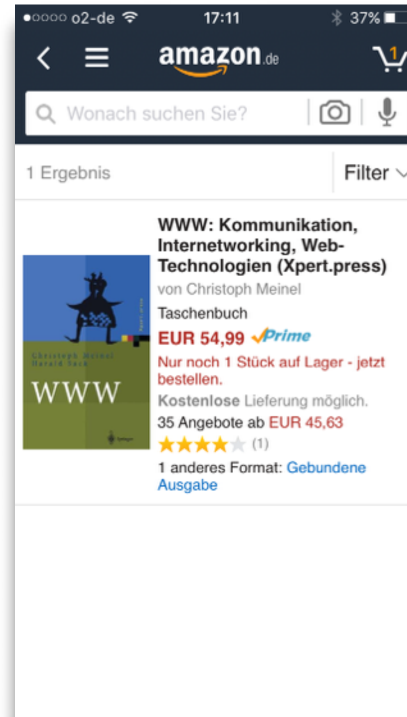
Given an audio file, get word transcription



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

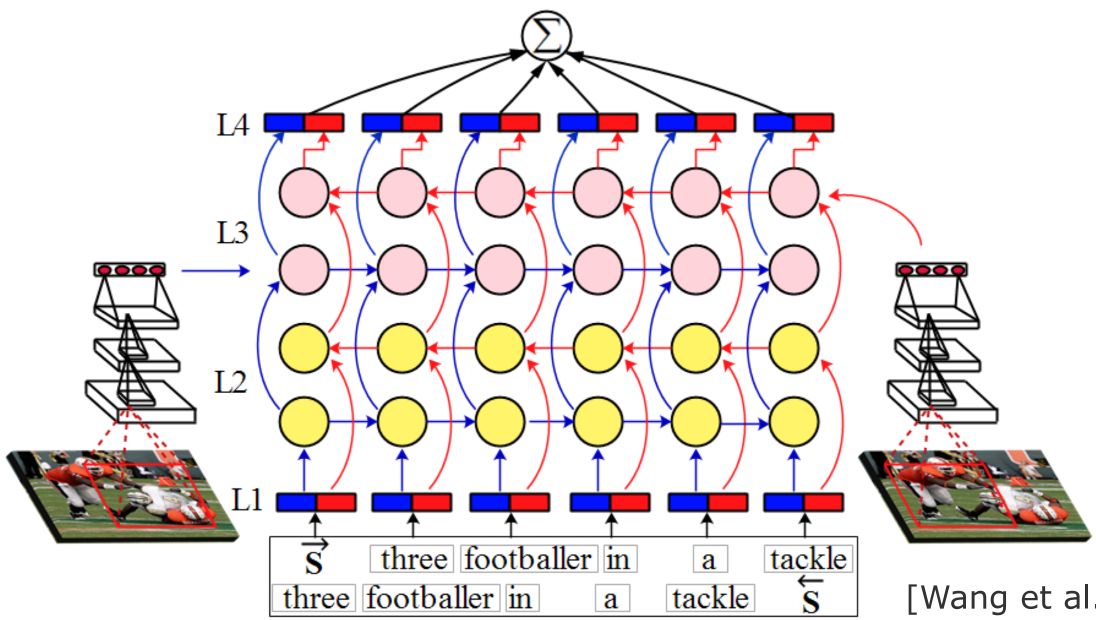
- DL for text processing, NLP
  - *published at INTERSPEECH'16*
- Video classification with CNNs
  - *published at IJCNN'16*
- Lecture video analysis and retrieval (applied in teleTASK and openHPI)
  - *published at IEEE ICALT'16*
- DL for medical image processing
- Audio analysis with DL

Scene Text Recognition

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



[Wang et al. 2016]

Video Classification, Activity Detection

Multiple deep neural networks:

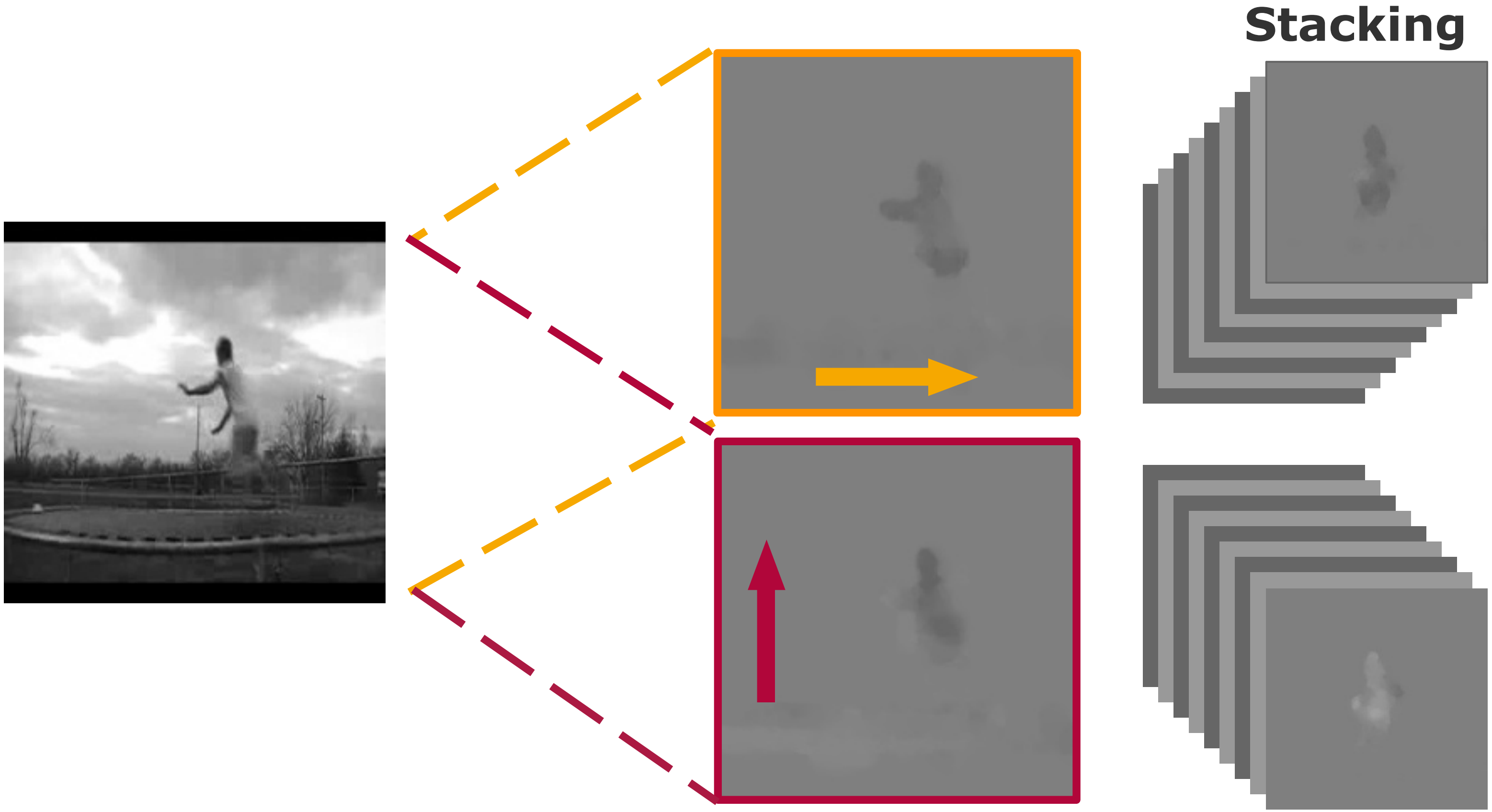
- Spatial: recognizing objects on frames
- Temporal: recognizing motion on multiple frames
- Auditory: acoustical information

Video Classification, Activity Detection

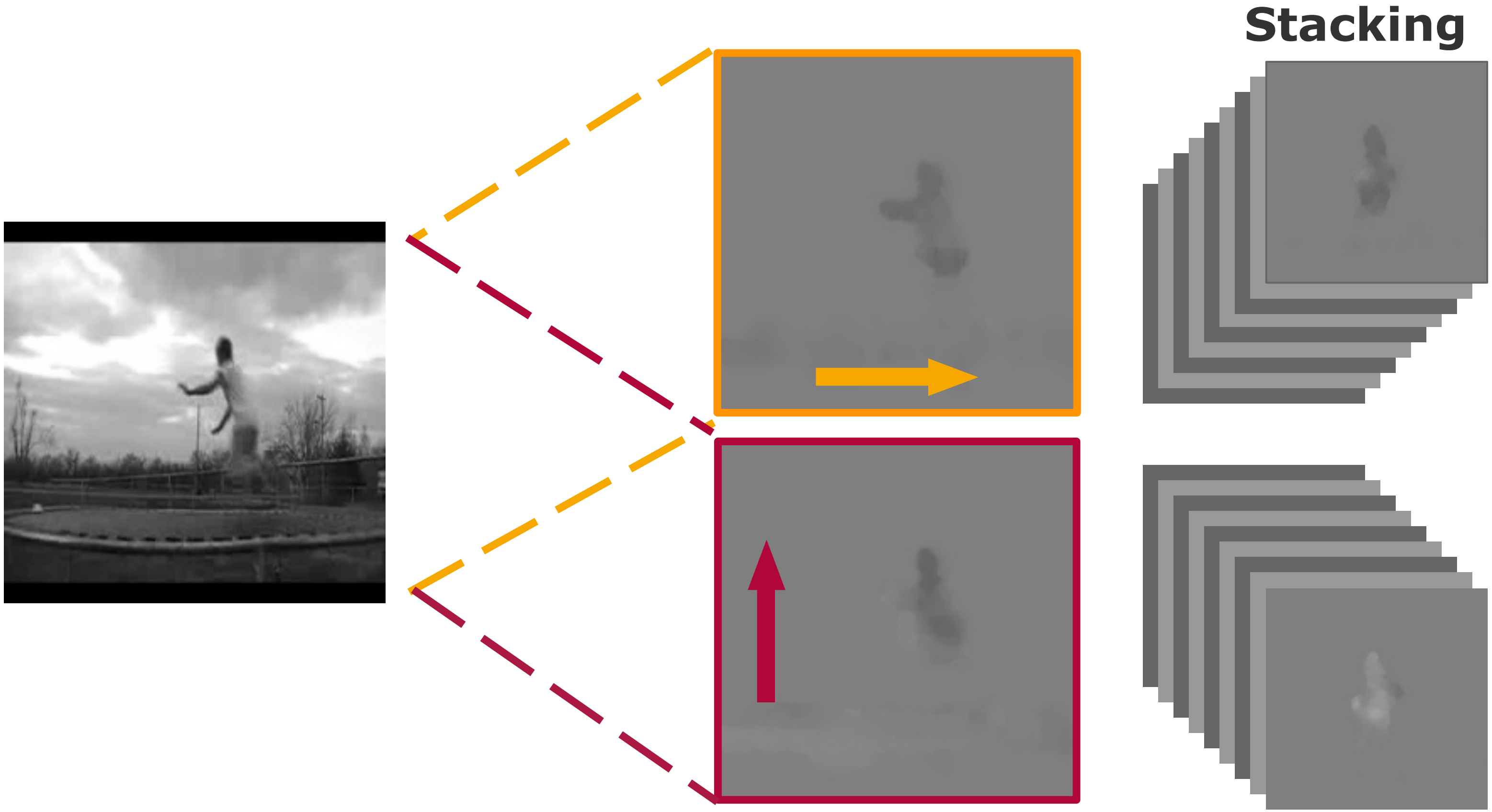
Multiple deep neural networks:

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

Temporal Stream: Dense Optical Flow

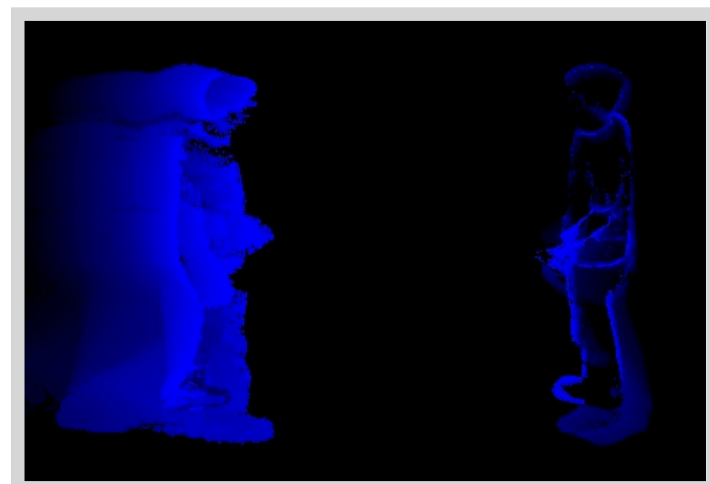


Temporal Stream: Dense Optical Flow



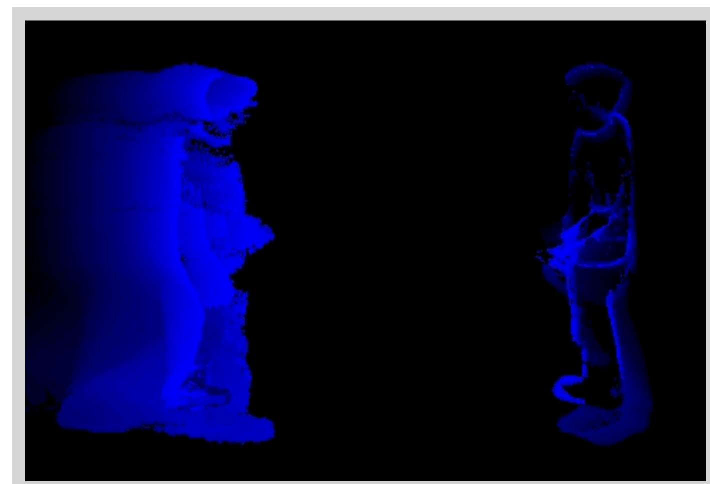
Temporal Stream: Motion History Image

- **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



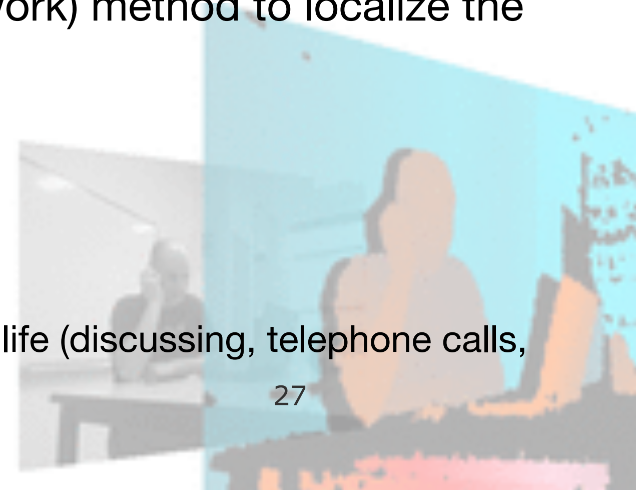
Temporal Stream: Motion History Image

- **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:

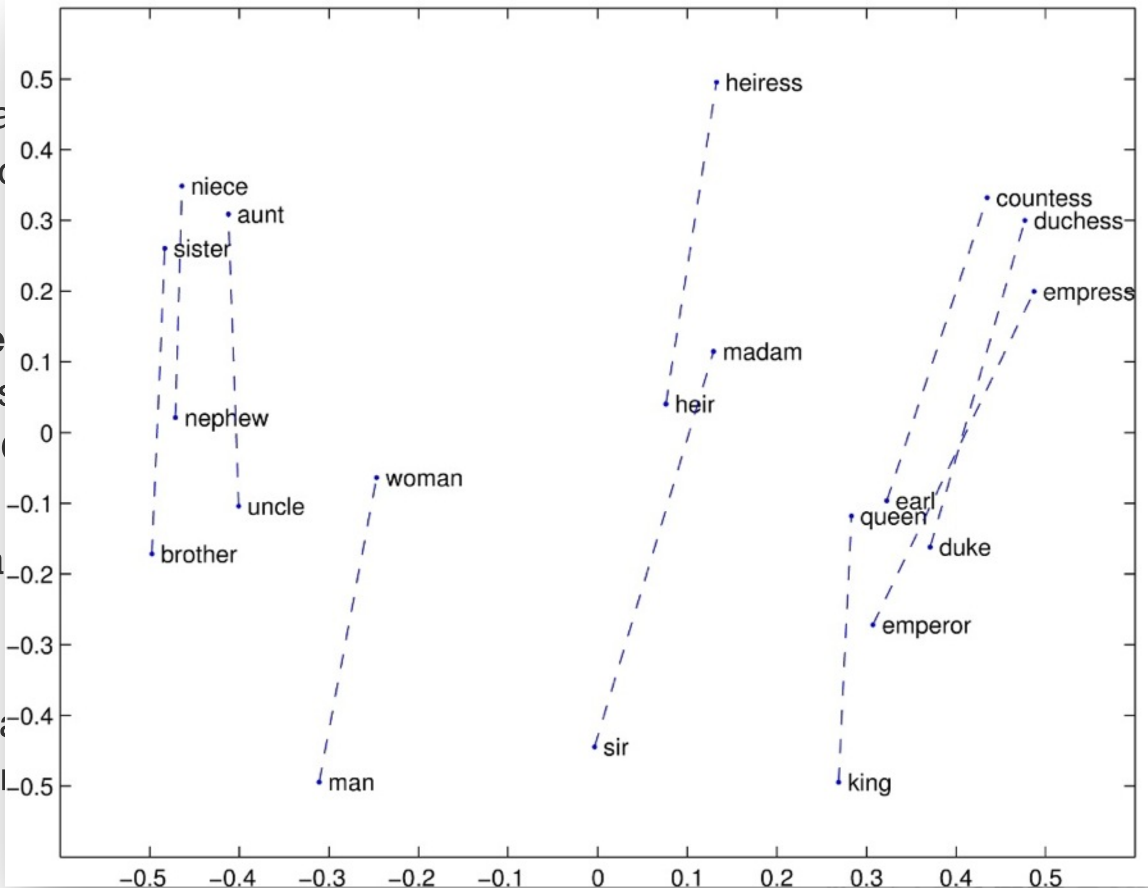
- **Word Vectors** ha
- But the major suc

How:

- Learn the **theore**
- Compare the exis
- Collect as many (
- Test the German
- Apply the Germa

Challenges:

- The amount of tr
- Complicated gra



Topic 2: German Word Vectors and Potential Applications

Why:

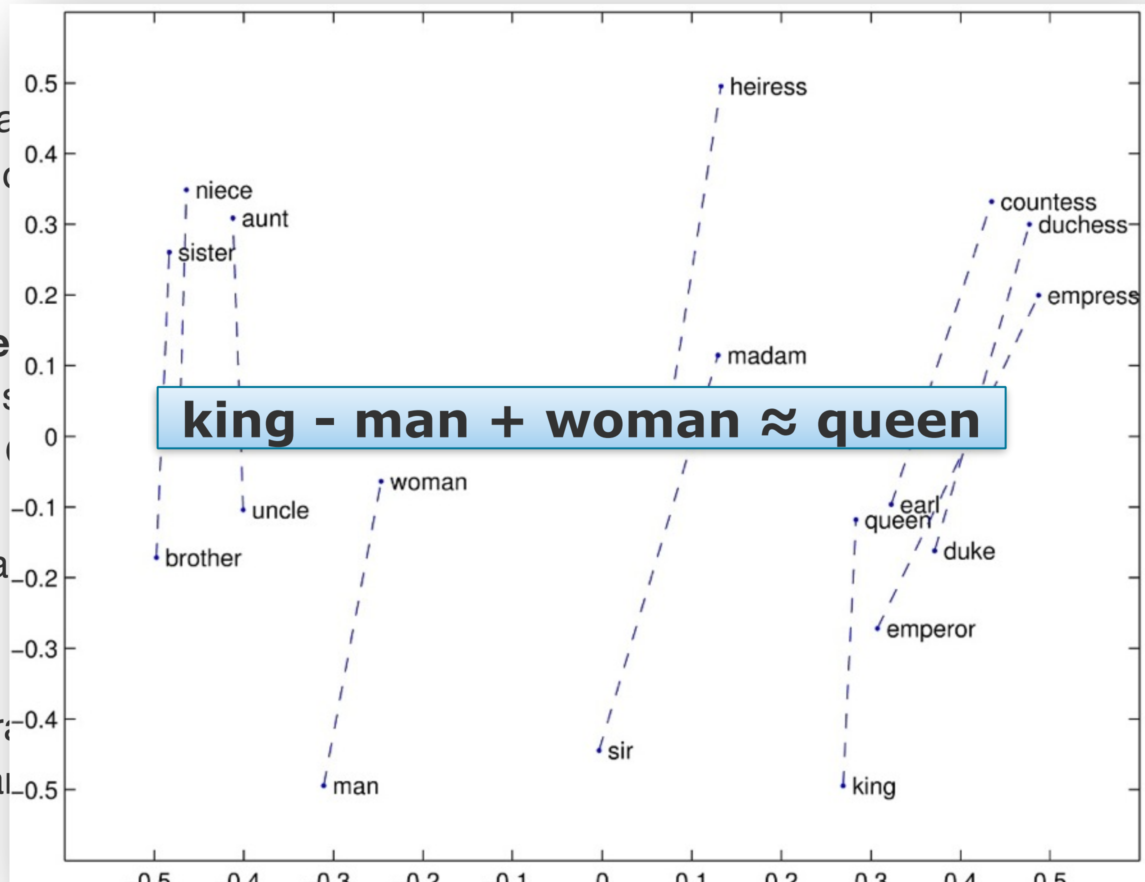
- **Word Vectors** ha
- But the major suc

How:

- Learn the **theore**
- Compare the exis
- Collect as many (
- Test the German
- Apply the Germa

Challenges:

- The amount of tr
- Complicated gra



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**.

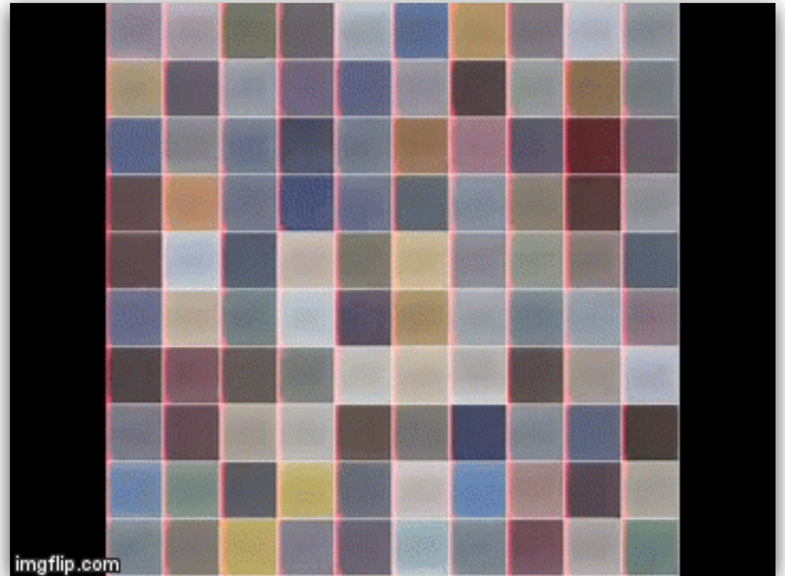
Challenges:

- 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

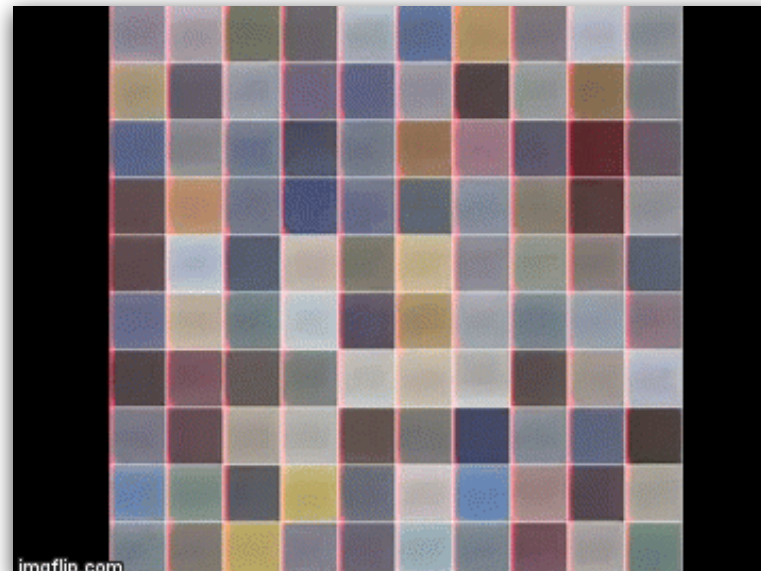
Topic 3: Deep Network For Image Generation

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



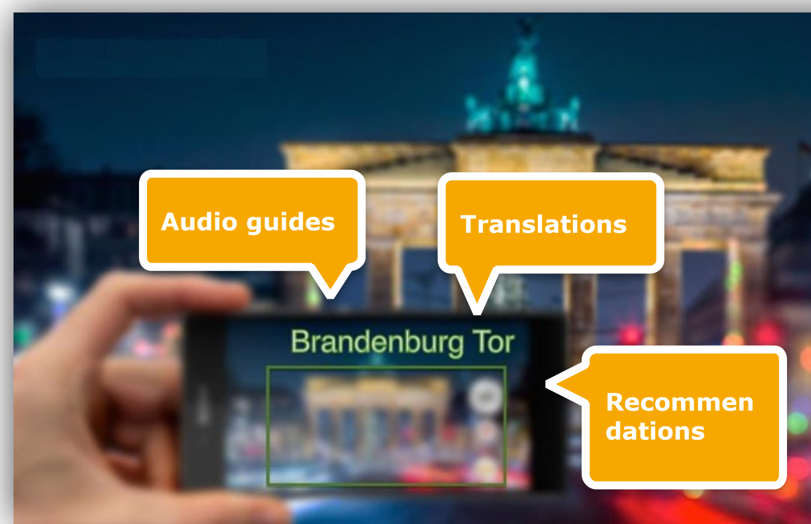
Topic 4: Place Recognizer

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:

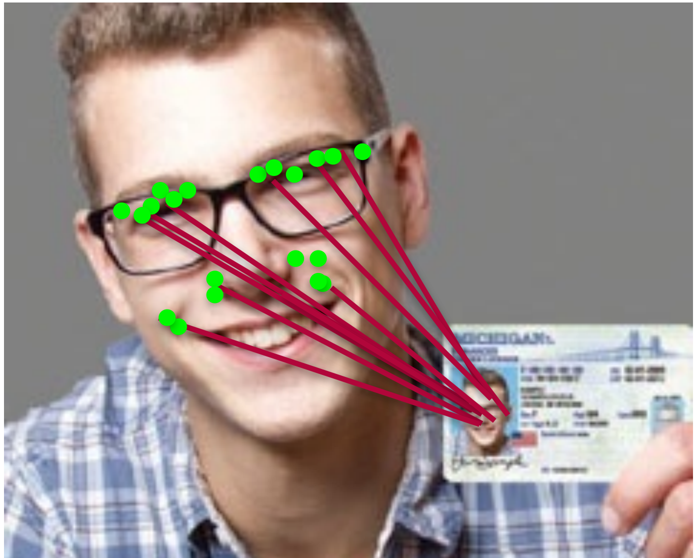
- Learning knowledges of deep learning,
- Apply deep learning technology to mobile application
- Contribute to software design and development



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

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) (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 Lehrveranstaltu

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!