ARTIFICIAL INTELLIGENCE

Alexander Schindler
Scientist
Information Management
Center For Digital Safety & Security

AIT Austrian Institute Of Technology Gmbh
Alexander.schindler@ait.ac.at

AUSTRIAN INSTITUTE OF TECHNOLOGY

DR. ALEXANDER SCHINDLER

- **Scientist @ AIT**
  - Multi-Modal Machine Learning
  - Center for Safety and Security
- **Project-Assistant - TU-Wien (part-time)**
  - Research
  - Teaching
    - Data Science
    - Information Retrieval
    - Intelligent Audio and Music Analysis
- **Research Interests**
  - Audio / Music Analysis
  - Audio-Visual Analysis
  - Machine Learning / Deep Learning
  - Artificial Intelligence
- **Event Organization**
  - Vienna Deep Learning Meetup
  - AI-Summit 2017
  - Ethics & Bias in AI 2018
  - WeAreDevelopers AI Congress 2018 (Partner)
  - Tutorials on Deep Learning (ML-Prague 2018, ISMIR 2018)
  - Int. Workshop on Music Speech and Mind (SMM2019, Vienna)

OUTLINE

- Introduction to Artificial Intelligence
- Introduction to Deep Learning
- Artificial Intelligence for Security

Introduction to

Artificial Intelligence

FRAGE

- Was ist eine Künstliche Intelligenz?
- Was stellen Sie sich unter einer Künstlichen Intelligenz vor?
- Wer kennt eine Künstliche Intelligenz?

DEFINITION

of Artificial Intelligence

WHAT IS AI?

- Academic Discipline / Computer Science

- *"Artificial Intelligence (AI) is the part of computer science concerned with* ***designing intelligent computer systems****, that is, systems that exhibit* ***characteristics we associate with intelligence in human behavior****"* (Barr & Feigenbaum, 1981)
  - Understanding language
  - Learning
  - Reasoning
  - solving problems

- ***Scientific Goal:*** To determine which ideas about knowledge representation, learning, rule systems, search, and so on, explain various sorts of real intelligence.

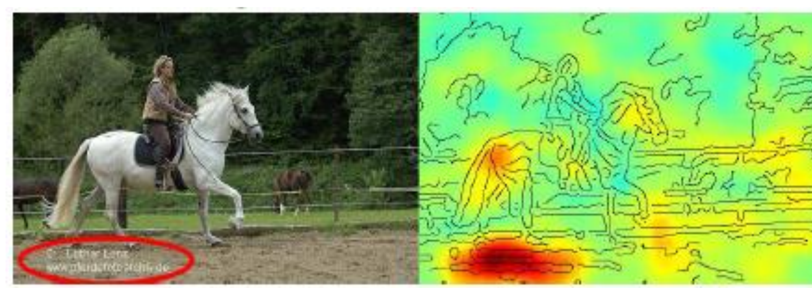
- ***Engineering Goal:*** To solve real world problems using AI techniques such as knowledge representation, learning, rule systems, search, and so on.

INTELLIGENT COMPUTATIONS ASSOCIATED WITH HUMAN INTELLIGENCE

- **Clever Hans (1907)**
  - Horse could solve arithmetic tasks
  - In controlled isolated experiment
    - Horse responded to involuntary clues of the trainer and spectators
    - Observer-expectancy effect

- **A System is a ‚Horse' (B. Sturm)**
  - Focus on production artifacts instead of real sematic concepts
  - Cannot be associated with Human Intelligence

Wikipedia



Sebastian Lapuschkin et al., *Analyzing Classifiers: Fisher Vectors and Deep Neural Networks*

WHAT IS AI?

- **Roots / Foundations of AI**
  - Philosophy (e.g. Descartes, Leibnitz)
  - Logic / Mathematics (e.g. Gödel)
  - Computation (e.g. Turing, von Neumann)
  - Psychology / Cognitive Science (knowledge representation)
  - Biology / Neuroscience (Connectionism, Neural Networks)
  - Evolution (Genetic Programming)

- **Sub-fields of AI**
  - Neural Networks / Machine Learning
  - Evolutionary Computation
  - Computer Vision
  - Robotics
  - Expert Systems
  - Speech Processing
  - Natural Language Processing
  - Planning

WHAT IS AI?

- **AI-Effect**
  - AI successfully solves a problem
  - the problem is no longer a part of AI
  - Examples:
    - Digit Recognition
    - Optical Character Recognition (OCR)

  - "AI is whatever hasn't been done yet.“ (D. Hofstadter, 1980)

- **AI Hype Today**
  - Inverse AI-Effect
  - Companies claim to use AI
    - linear regression
    - Rule-based Systems

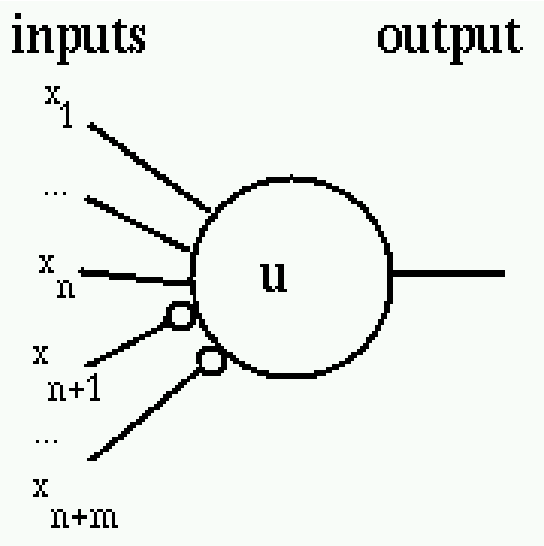
HISTORY

of Artificial Intelligence

A BRIEF HISTORY OF AI

- 1943: McCulloch and Pitts propose a model of artificial neurons
    - *“A Logical Calculus of the Ideas Immanent in Nervous Activity”*

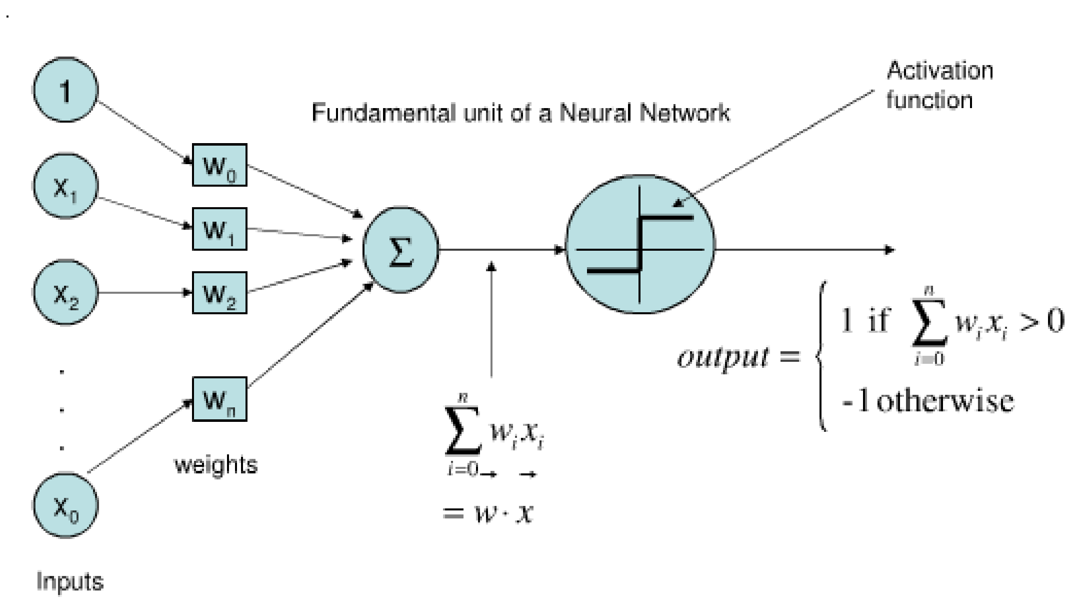
- 1956: Minsky and Edmonds build first neural network computer, the *Stochastic neural analog reinforcement calculator (SNARC)*

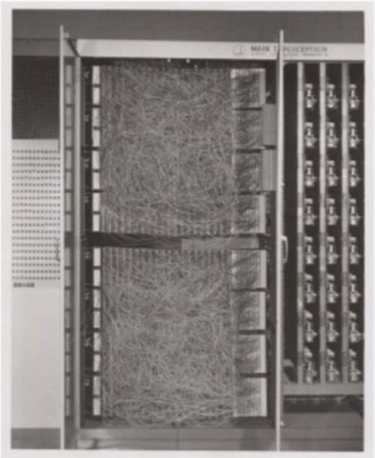


1956: **The Dartmouth Conference**

- **two-month workshop** for researchers interested in
  - neural networks
  - the study of intelligence
- Organizers
  - John McCarthy (Stanford)
  - Marvin Minsky (MIT)
  - Herbert Simon (CMU)
  - Allen Newell (CMU)
  - Arthur Samuel (IBM)
- **Agreement** to adopt a **new name** for this field of study:
  - **Artificial Intelligence**

1958: **Perceptron** by Frank Rosenblatt





Weights were encoded in potentiometers, and weight updates during learning were performed by electric motors.

Linear binary classifier using a step function

For the first time a NN could solve simple classification problems merely from training data

UPS & DOWNS OF NEURAL NETWORKS

1952 - 1969: Golden years of AI (funded by DARPA):

- Solve algebra
  - 1956 Logic Theorist
  - 1961 SAINT
- Reasoning
- Semantic nets

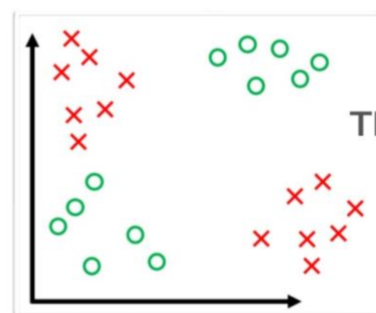
- 1952-1962: Checkers player (by Arthur Samuel)
- 1957: Newell and Simon: "within ten years a digital computer will be the world's chess champion"
- 1967: MacHack achieved class-C rating in tournament chess

1969: *Perceptrons: An Introduction to Computational Geometry*
by Marvin Minsky and *Seymour Papert*

- shown that XOR problem cannot be solved by Perceptron
  - But: they argued for locally connected neurons
    - Easier to implement in the 1960s
  - Fully connected three-layered Perceptrons can model XOR function

- Book was popular and is often cited as a show-stopper for AI

It is impossible to linearly separate the XOR function



The first winter: 1969 - 86

Is it?

1970s: **First AI-Winter**

- Book by Minsky and Papert was not the only problem
- AI problems appear to be too big and complex
- Computers are very slow, very expensive, and have very little memory (compared to today)
  - neural networks were tiny and could not achieve (the expected) high performance on real problems
  - Datasets were small

- Pessimism in the AI community
- followed by pessimism in the press
- followed by a severe cutback in funding
- followed by the “end” of serious research on Neural Networks
- First AI-Winter

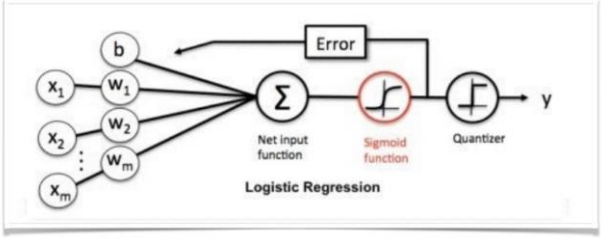
1969 - 1979: **Knowledge-based Systems**

- Birth of expert systems
- Idea is to give AI systems lots of information to start with
  - Rule-based Systems
  - Fuzzy Logic

1980s: Governments (starting in Japan) and industry provide AI with billions of dollars. **Boom of “expert systems”.**

1986: **Backpropagation** had been invented in the 1970s, but only 1986 it became popular through a famous paper by David Rumelhart, Geoffrey Hinton, and Ronald Williams. It showed that also complex functions became solvable through NNs by using multiple layers.

Late 1980s: Investors - despite actual progress in research - became disillusioned and withdrew funding again.



SECOND AI-WINTER

1991: Hornik proved 1 hidden layer network can model any continuous function (universal approximation theorem)

1991/92 Vanishing Gradient: problem in multi-layer networks where training in front layers is slow due to backpropagation diminishing the gradient updates through the layers. Identified by Hochreiter & Schmidhuber who also proposed solutions.

1990s - mid 2000s:
Due to lack of computational power, interest in NNs decreased again and other Machine Learning models, such as Bayesian models, Decision Trees and Support Vector Machines became popular.

UPS & DOWNS OF NEURAL NETWORKS

1996: Deep Blue (IBM) beats world chess champion

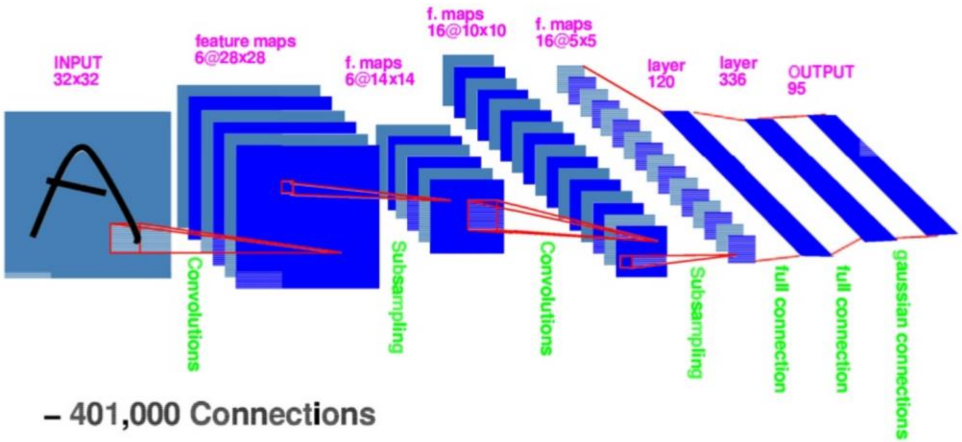
2005: Stanford robot won the DARPA Grand Challenge by driving autonomously for 131 miles along an unrehearsed desert

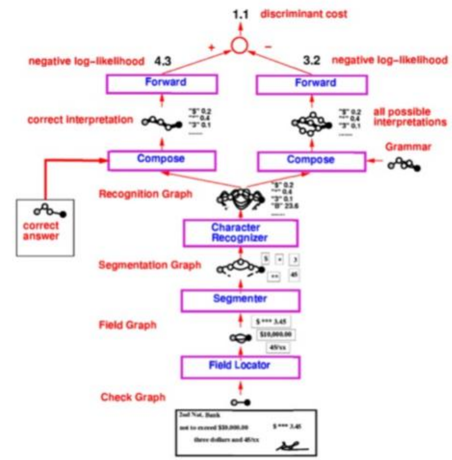
2011: Watson (IBM), defeated the two greatest Jeopardy! champions

Wikipedia

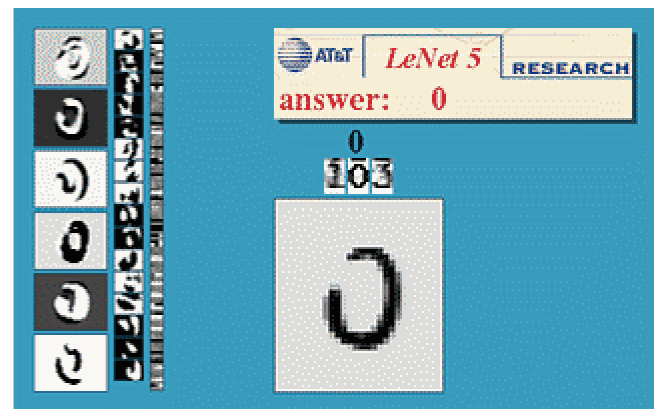
Wikipedia

AT&T CHECK READER (LECUN, BENGIO, 1996)



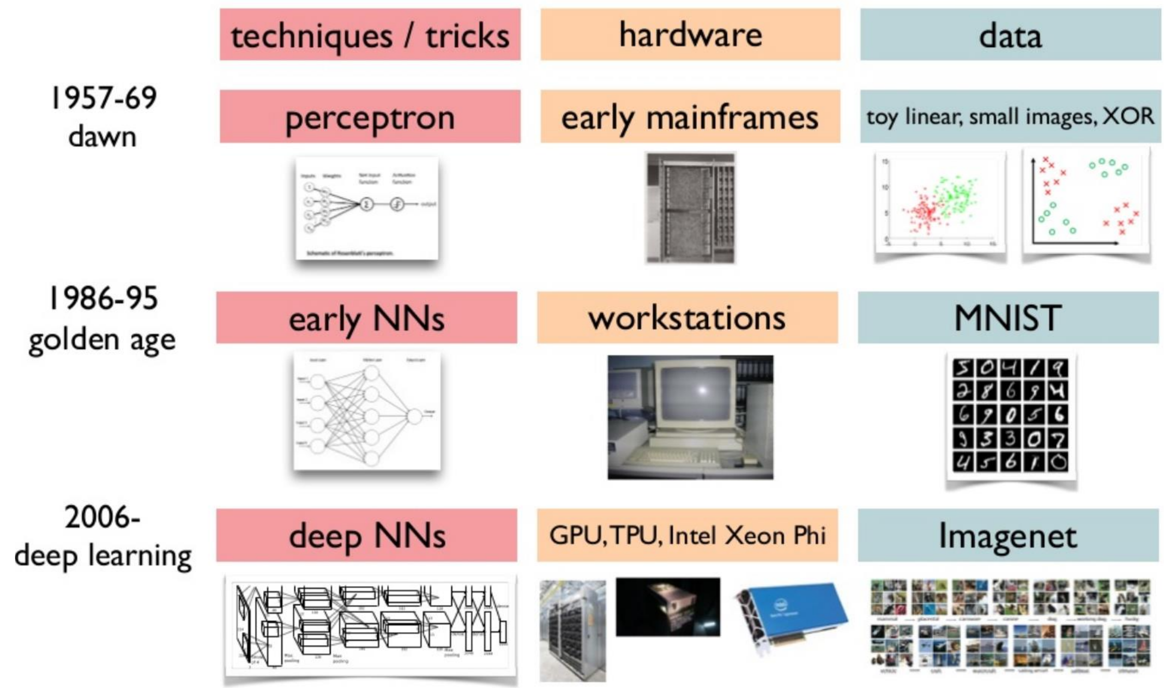


- 401,000 Connections
- 100,000 free parameters
- Trained with 500,000 character samples
  (Full ASCII set, machine printed and handwritten)



http://yann.lecun.com/exdb/lenet/

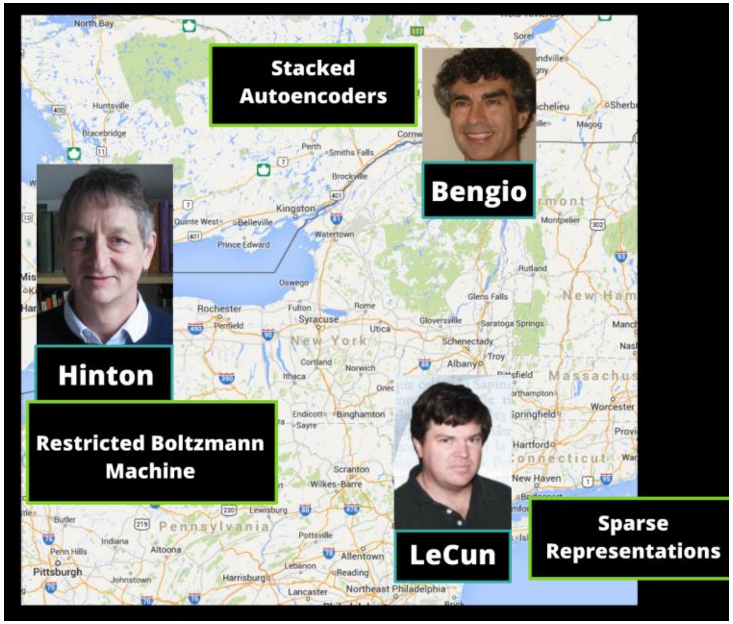
THREE EPOCHS OF NEURAL NETWORKS



RESURRECTION OF DEEP LEARNING IN THE 2000S

2000s: Hinton, Bengio and LeCun ("The fathers of the age of deep learning") join forces in a project

They overcome some problems that caused deep networks not to learn anything at all



2006: Breakthrough with Layer-wise pre-training by unsupervised learning (using RBMs)

2010s: Important new contributions:
- Simpler initialization (without pre-training)
- Dropout
- Simplier activations: Rectifier Units (ReLUs)
- Batch Normalization

→ not a re-invention of NNs but paved the way for very deep NNs

https://www.datarobot.com/blog/a-primer-on-deep-learning/

RECENT SUCCESSES OF DEEP LEARNING

2012/13: MNIST Number Recognition Error rate brought down to 0.21%

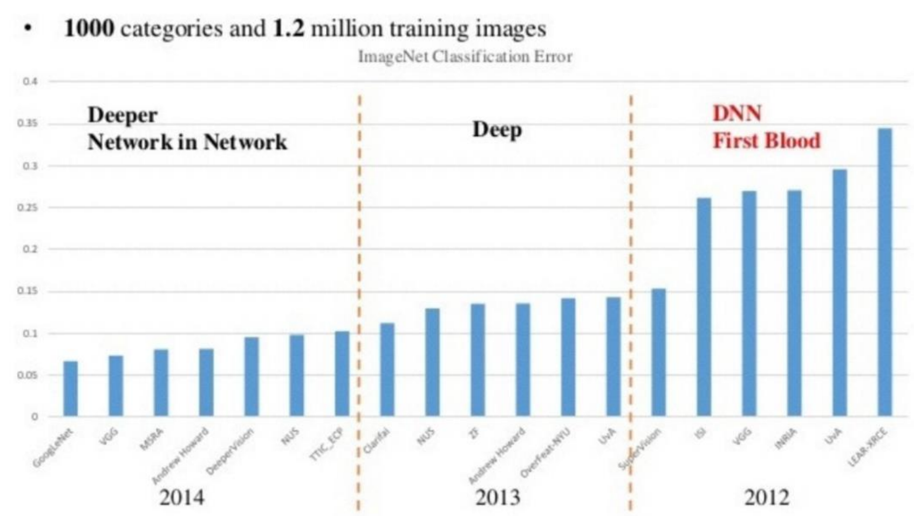
2012: Google's X Lab's DNN learned the concept of cats and humans from Youtube

2012: AlexNet (first deep net) pushed state-of the art object recognition in images by far (16.4% error). Most people in object recognition now use deep nets.

2014: GoogLeNet again brings the error on object recognition down to half (6%)

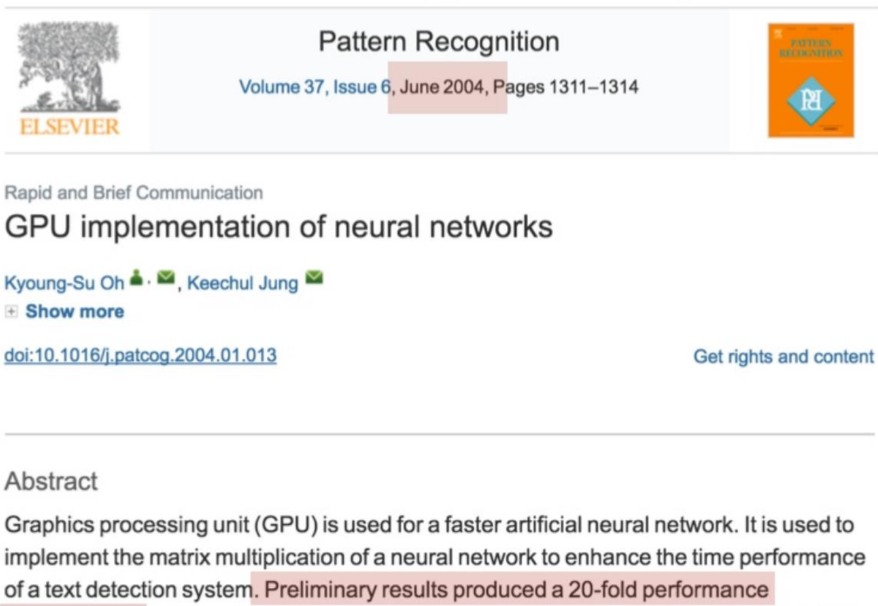
2015: ResNets half object recognition again to 3.6%.

2016: Google DeepMind's AlphaGo beats Go champion 4:1



Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 http://image-net.org/

GPUs (2004)


ELSEVIER

Pattern Recognition

Volume 37, Issue 6, June 2004, Pages 1311–1314

Rapid and Brief Communication

GPU implementation of neural networks

Kyoung-Su Oh, Keechul Jung

Show more

doi:10.1016/j.patcog.2004.01.013

Get rights and content

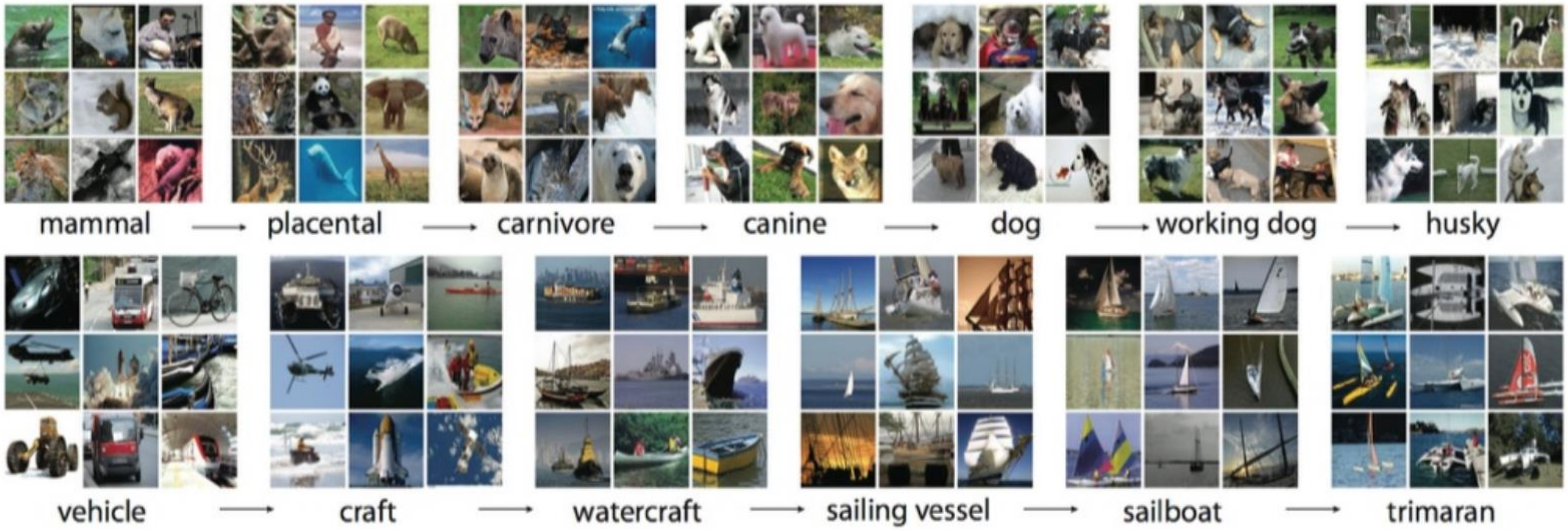
Abstract

Graphics processing unit (GPU) is used for a faster artificial neural network. It is used to implement the matrix multiplication of a neural network to enhance the time performance of a text detection system. Preliminary results produced a 20-fold performance

Author unknown CC BY-SA-NC

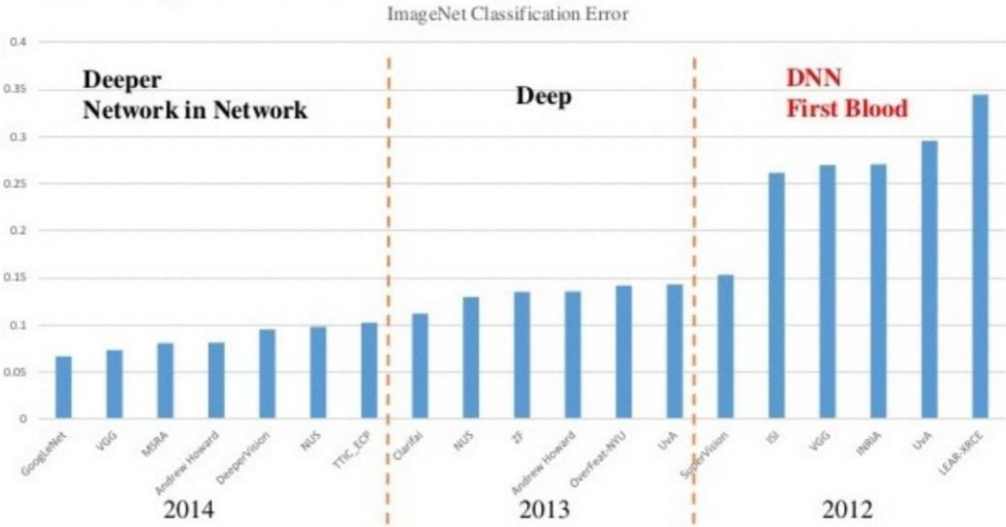
IMAGENET (2009)

- Large Database for Visual Object Recognition
- More than 20.000 categories
- Aligned to WordNet
- Large Scale Visual Recognition Challenge (ILSVRC)
- Significant impact on Deep Learning

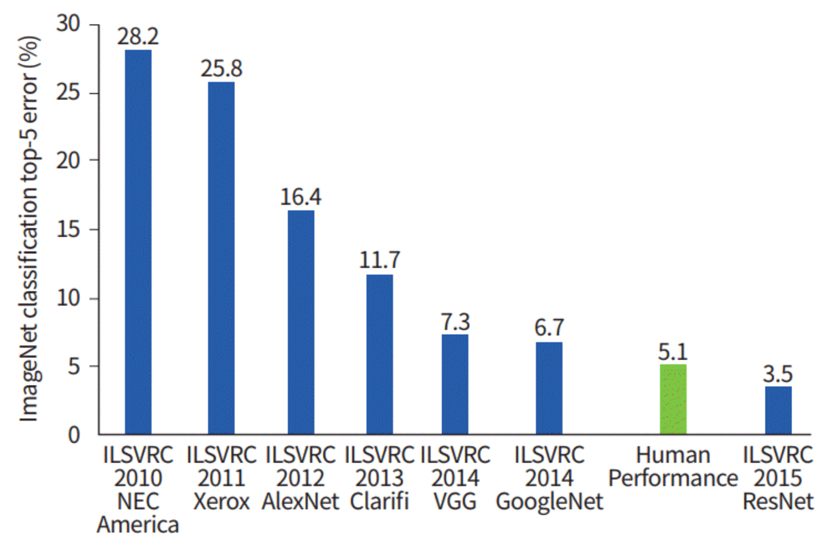


DROPPING ERROR RATES SINCE THEN

- **1000** categories and **1.2** million training images



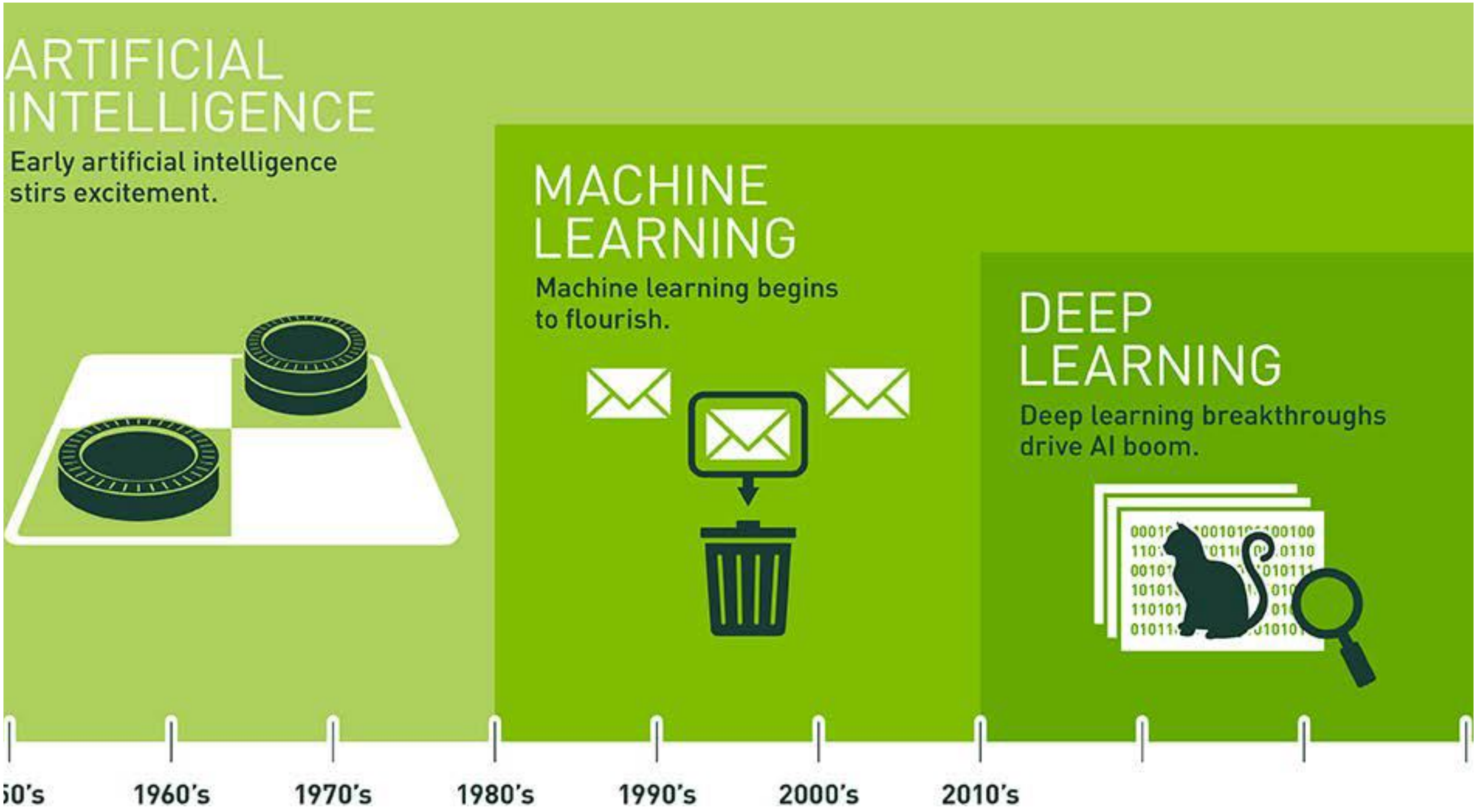
Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 http://image-net.org/



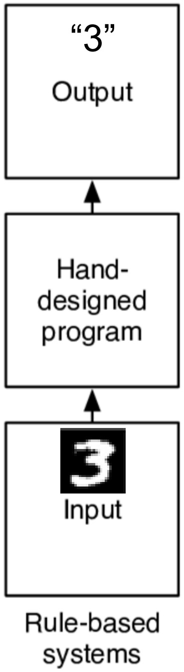
Introduction to

Deep Learning

THREE EPOCHS OF NEURAL NETWORKS

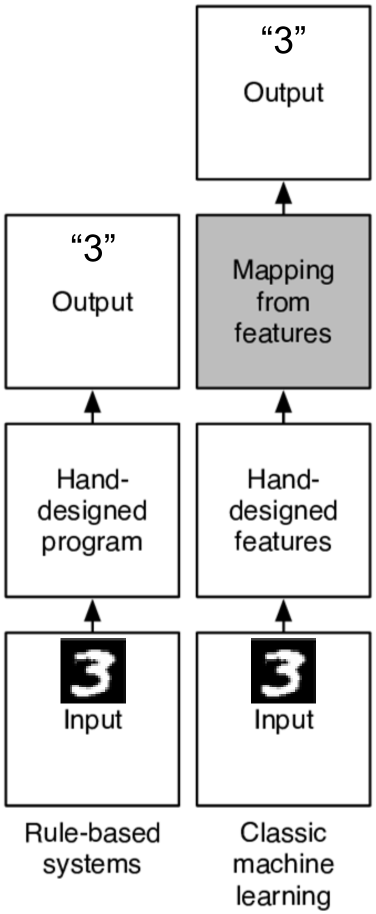


Rule-based systems:
Write algorithm by hand.



graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS

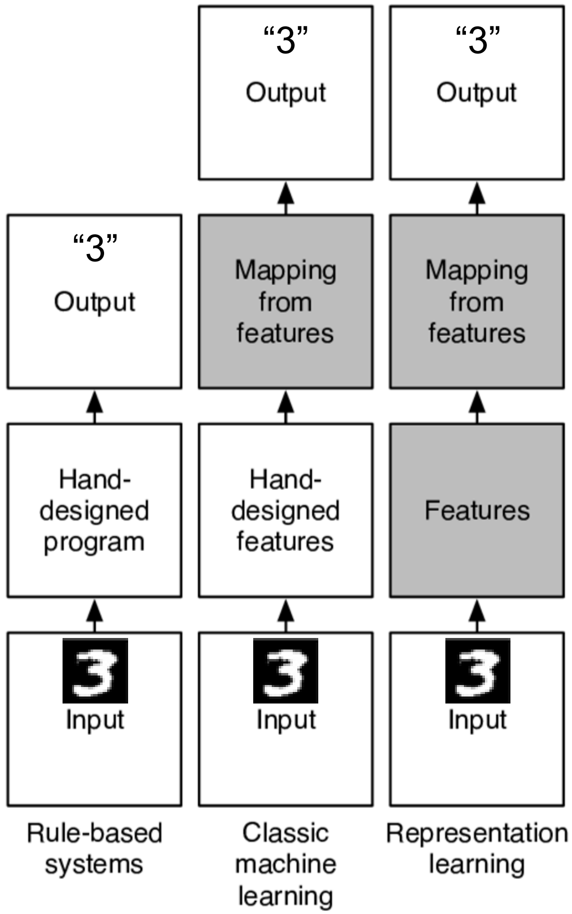


Rule-based systems:
Write algorithm by hand.

Classic machine learning:
Write feature extractor by hand, train classifier on top.

graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS



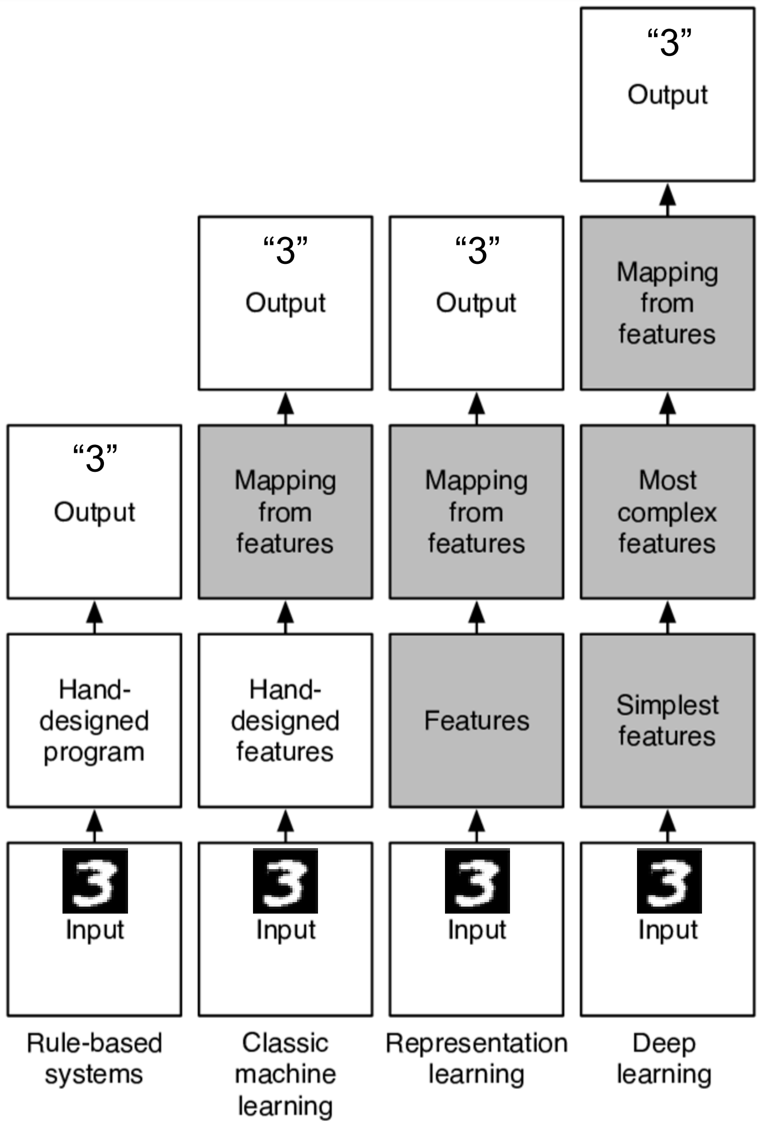
Rule-based systems:
Write algorithm by hand.

Classic machine learning:
Write feature extractor by hand, train classifier on top.

Representation learning:
Learn feature extractor (often unsupervised), train classifier on top.

graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS

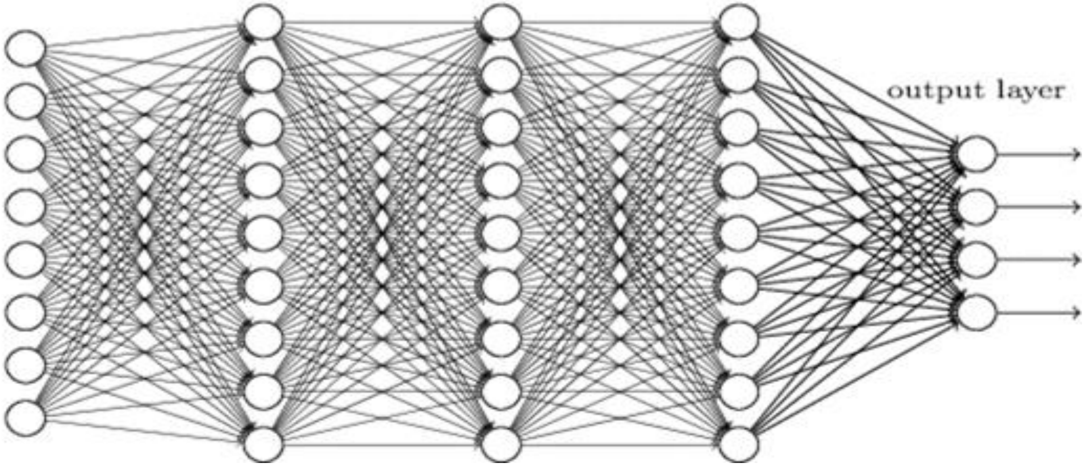


Deep learning:
Learn a stack of many simpler functions to map input to output.

- Often, that stack is a ***neural network***.
- Often, it is trained on raw input: optimize features & classification together, ***minimize hand-crafting***. ("end-to-end learning")

graphic: Y. Bengio, Deep Learning, MLSS 2015

So What is Deep Learning?



WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

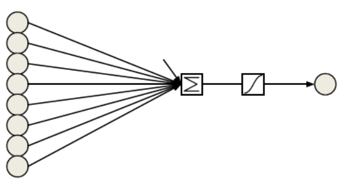
$y = \sigma(b + \mathbf{w}^T \mathbf{x})$ (equivalent to logistic regression)

WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$y = \sigma(b + \mathbf{w}^T\mathbf{x})$ (equivalent to logistic regression)

expression can be visualized as a graph:

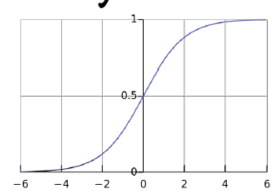


$\mathbf{x}$ $b + \mathbf{w}^T\mathbf{x}$ y

Output value is computed as a **weighted sum of its inputs,**

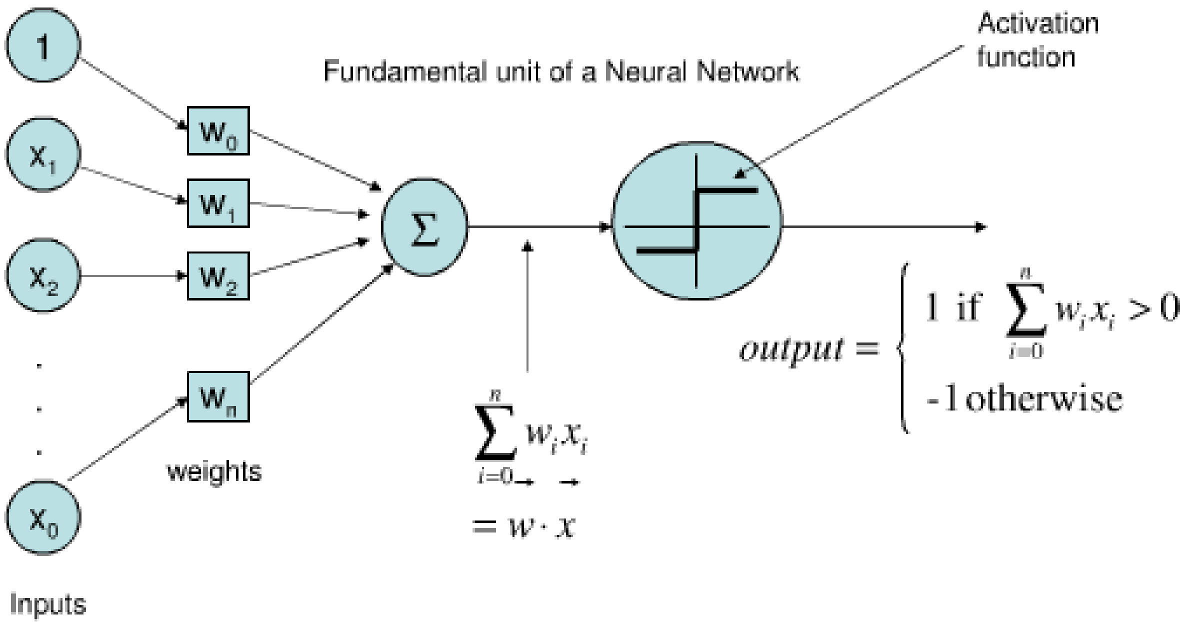
$$b + \mathbf{w}^T\mathbf{x} = b + \Sigma_i w_i x_i$$

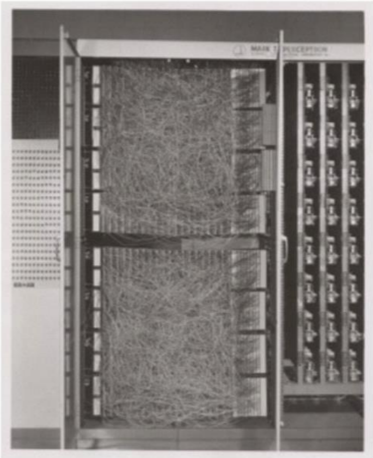
followed by a nonlinear function.



Origins of Neural Networks:
The Perceptron

1958 by Frank Rosenblatt





Weights were encoded in potentiometers, and weight updates during learning were performed by electric motors.

Linear binary classifier using a step function

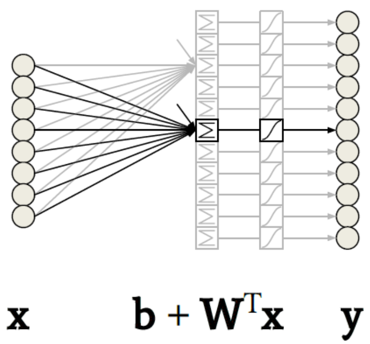
For the first time a NN could solve simple classification problems merely from training data

WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$$\mathbf{y} = \sigma(\mathbf{b} + \mathbf{W}^T\mathbf{x}) \qquad \text{(multiple logistic regressions)}$$

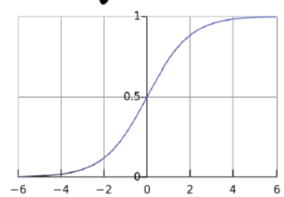
expression can be visualized as a graph:



Output values are computed as **weighted sums of their inputs,**

$$\mathbf{b} + \mathbf{W}^T\mathbf{x} = b_j + \Sigma_i w_{ij} x_i$$

followed by a nonlinear function.

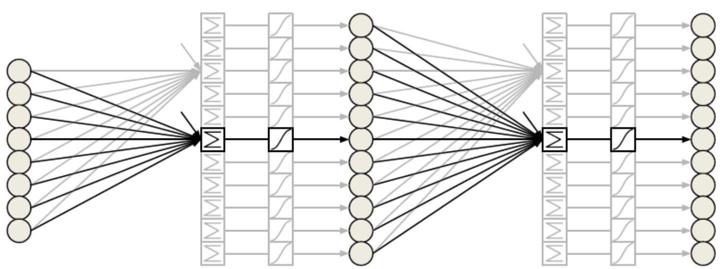


WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$$\mathbf{y} = \sigma(\mathbf{b}_2 + \mathbf{W}_2{}^{T}\sigma(\mathbf{b}_1 + \mathbf{W}_1{}^{T}\mathbf{x})) \qquad \text{(stacked logistic regressions)}$$

expression can be visualized as a graph:



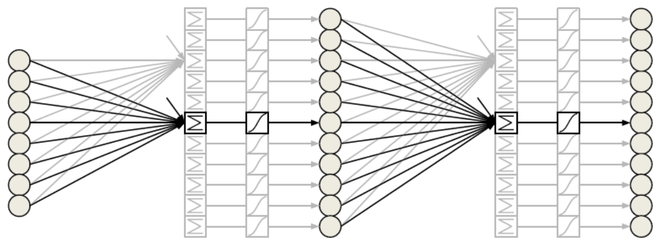
$\mathbf{x} \qquad \mathbf{b}_1 + \mathbf{W}_1{}^{T}\mathbf{x} \qquad \mathbf{h} \qquad \mathbf{b}_2 + \mathbf{W}_2{}^{T}\mathbf{h} \qquad \mathbf{y}$

WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

$$\mathbf{y} = \sigma(\mathbf{b}_2 + \mathbf{W}_2^T \sigma(\mathbf{b}_1 + \mathbf{W}_1^T \mathbf{x})) \qquad \text{(stacked logistic regressions)}$$

expression can be visualized as a graph:



Universal Approximation Theorem:
This can model any continuous function from $\mathbb{R}^n$ to $\mathbb{R}^m$ arbitrarily well (if $\mathbf{h}$ is made large enough).

MATHEMATICAL REASONS FOR GOING "DEEP"

A neural network with a single hidden layer of enough units can approximate any continuous function arbitrarily well. In other words, it can solve whatever problem you're interested in!
(Cybenko 1998, Hornik 1991)

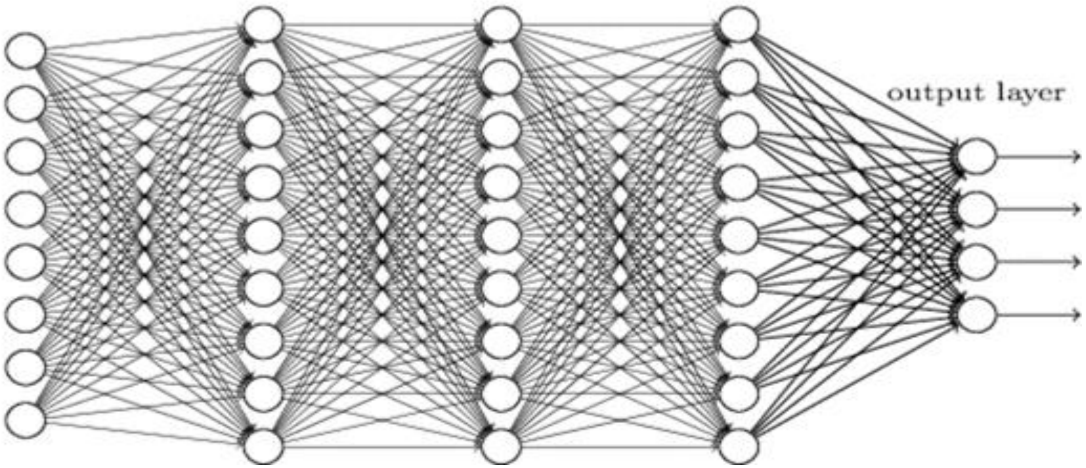
But:

- "Enough units" can be a very large number. There are functions representable with a small, but deep network that would require exponentially many units with a single layer.
  (e.g., Hastad et al. 1986, Bengio & Delalleau 2011)

- The proof only says that a shallow network *exists*, it does not say how to find it. Evidence indicates that it is easier to train a deep network to perform well than a shallow one.

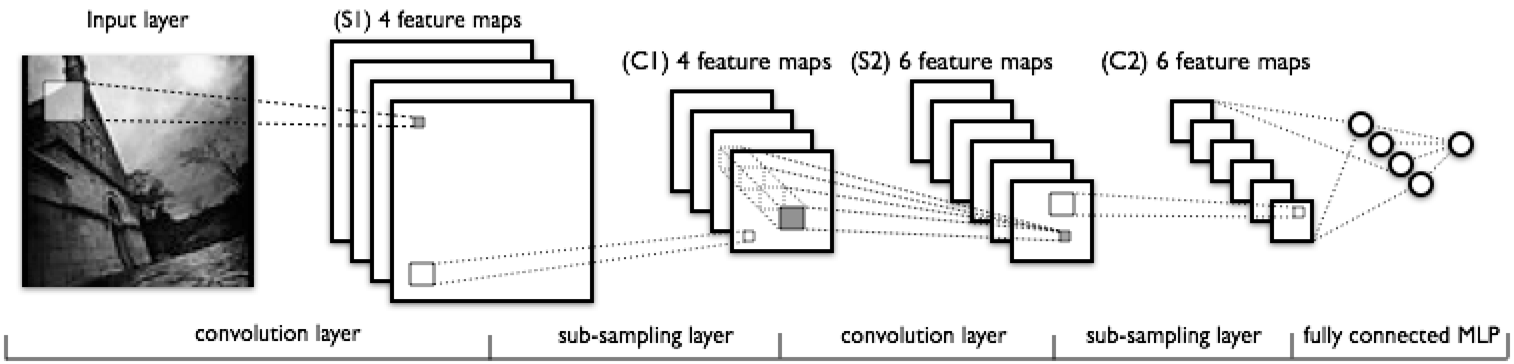
WHAT MAKES THEM DEEP?

- Neural Networks can learn any arbitrary function
- That‘s what makes them so powerful
- The more layers they have, the more expressive they can be
- However, mathematically the problem gets more difficult to solve
- And computationally it becomes extreme (or unfeasible)
  (that‘s why GPUs are needed)



More hidden layers = deep

CONVOLUTIONAL NEURAL NETWORK (CNN)



http://deeplearning.net/tutorial/lenet.html

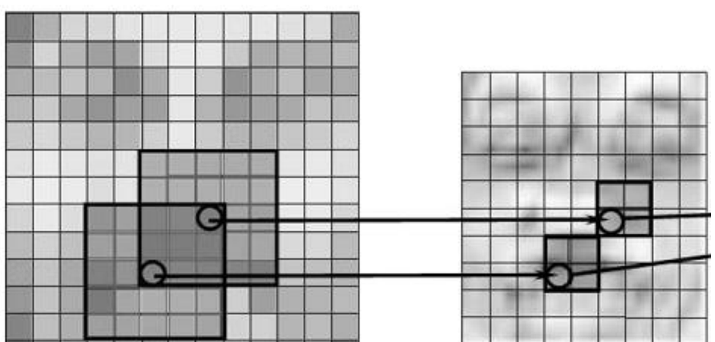
Combines three types of layers:

- **Convolutional layer:** performs 2D convolution of 2D input with multiple learned 2D kernels
- **Subsampling layer:** replaces 2D patches by their maximum (“max-pooling”) or average
- **Fully-connected layer:** computes weighted sums of its input with multiple sets of learned coefficients

Applies a nonlinear function after each linear operation (without, a deep network would be linear despite its depth).

MOTIVATION FOR CONVOLUTIONS

- Apply local filter kernels
- These kernels are the neurons that are learned

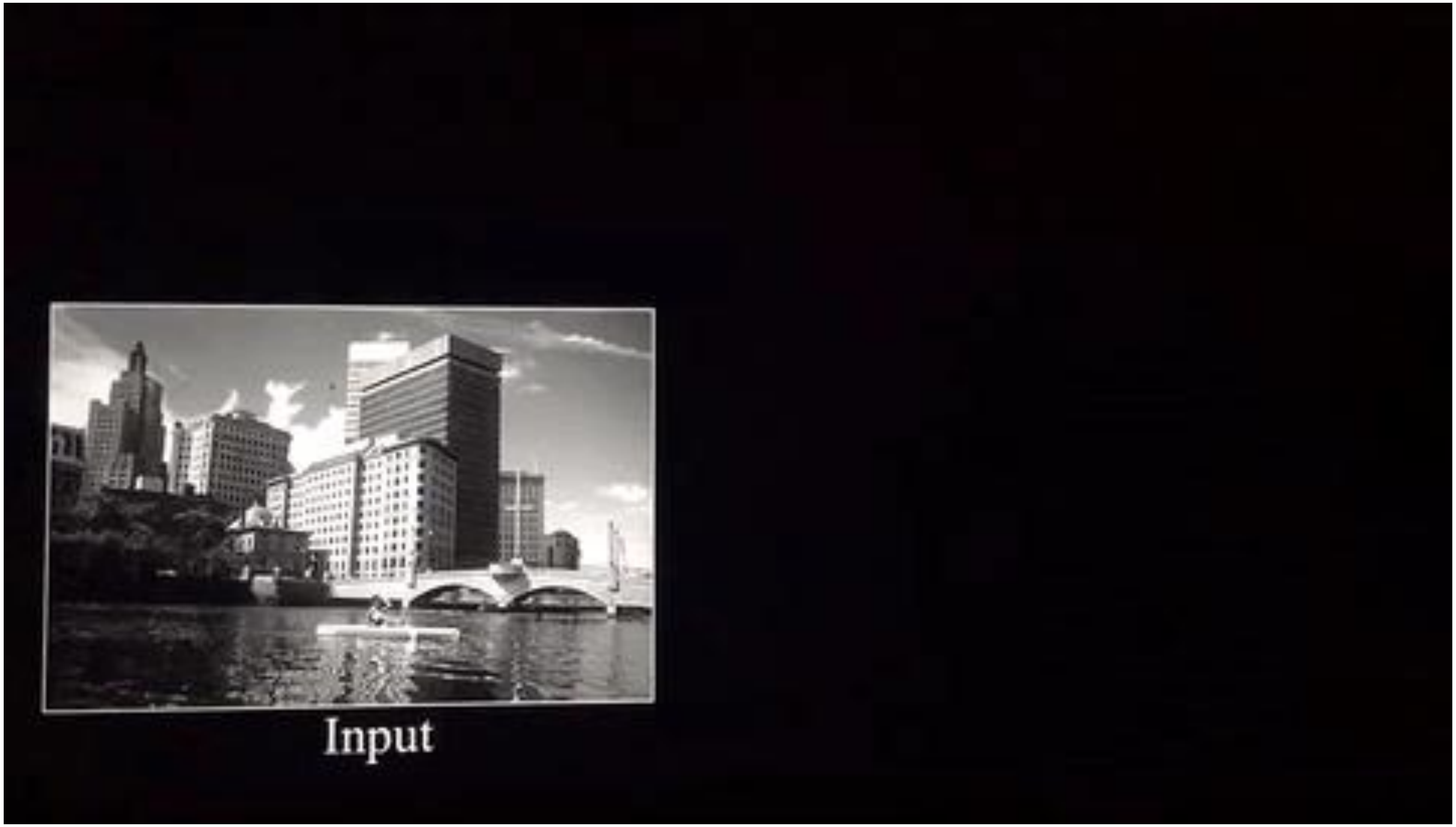


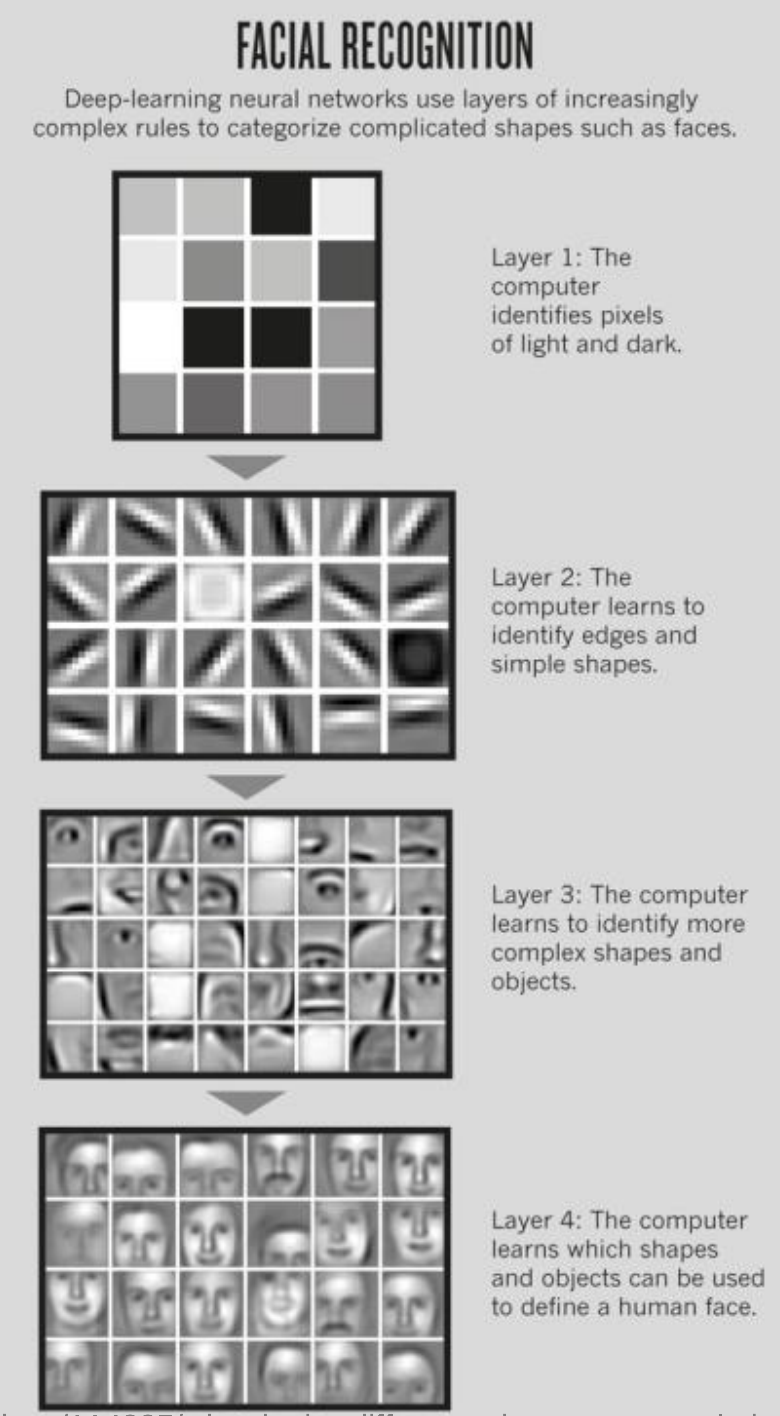
convolution 5x5

Operation	Kernel	Image result
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	

Images: http://sanghyukchun.github.io/75/
https://en.wikipedia.org/wiki/Kernel_(image_processing)

IMAGE PROCESSING





http://stats.stackexchange.com/questions/114385/what-is-the-difference-between-convolutional-neural-networks-restricted-boltzma

FACE RECOGNITION WITH CNNS

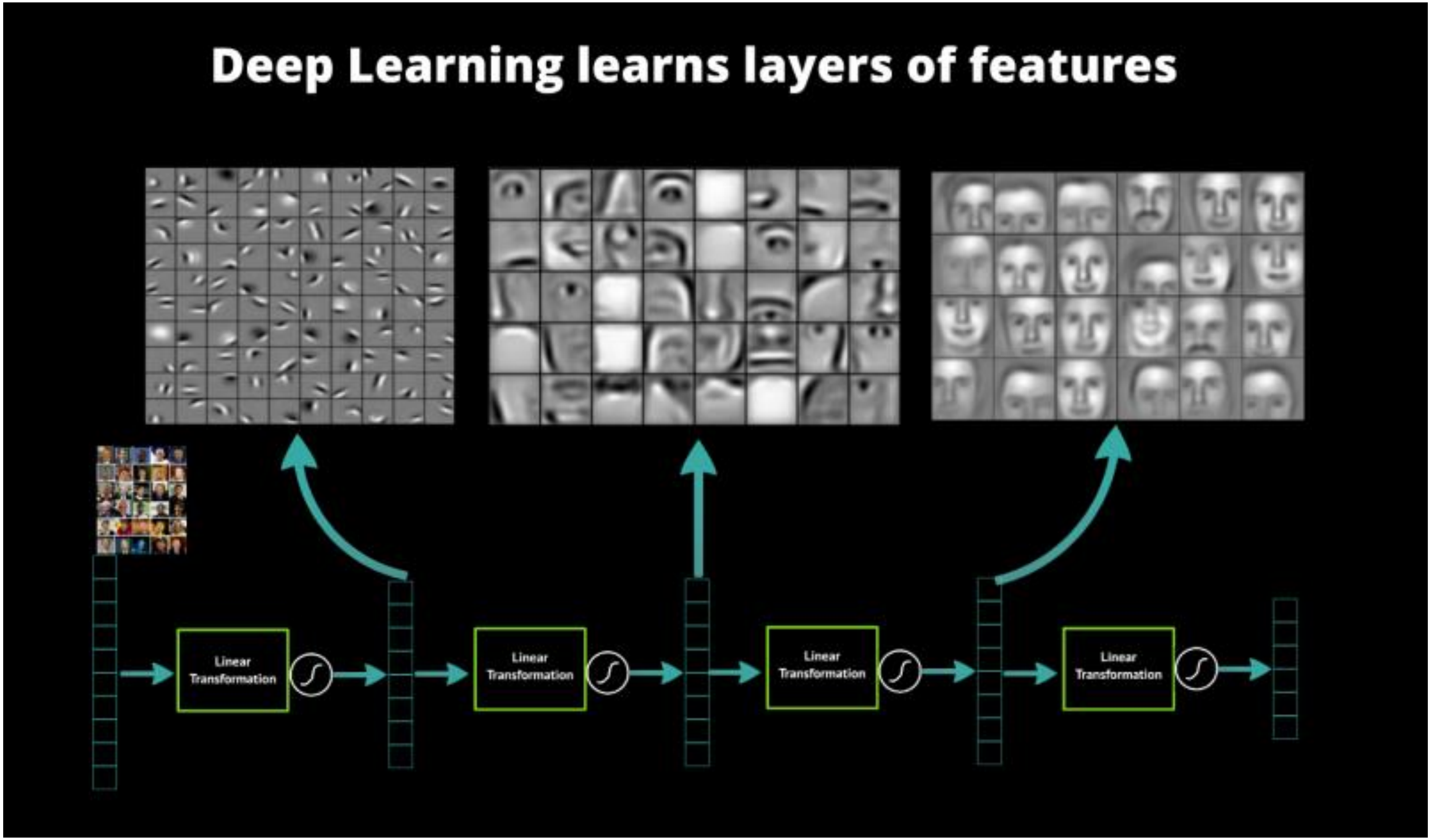
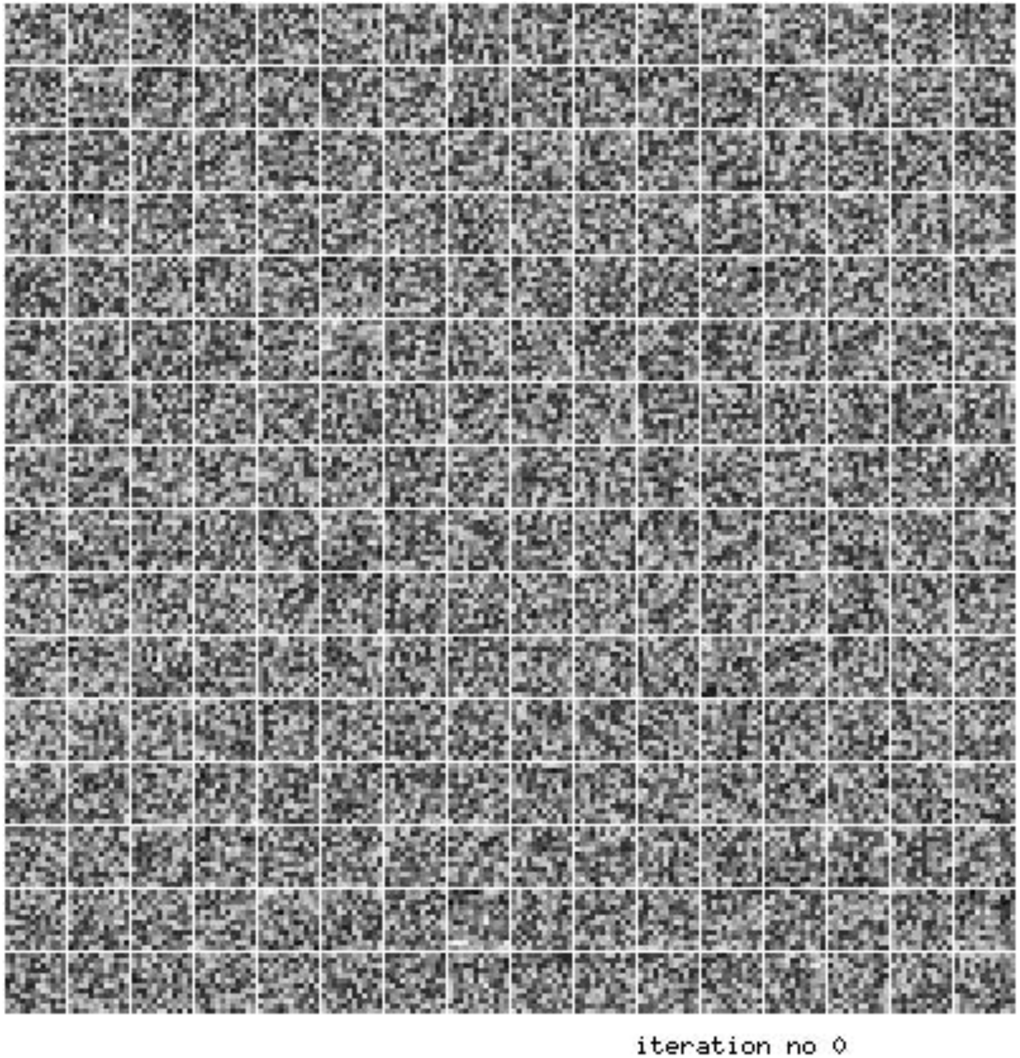


IMAGE PROCESSING

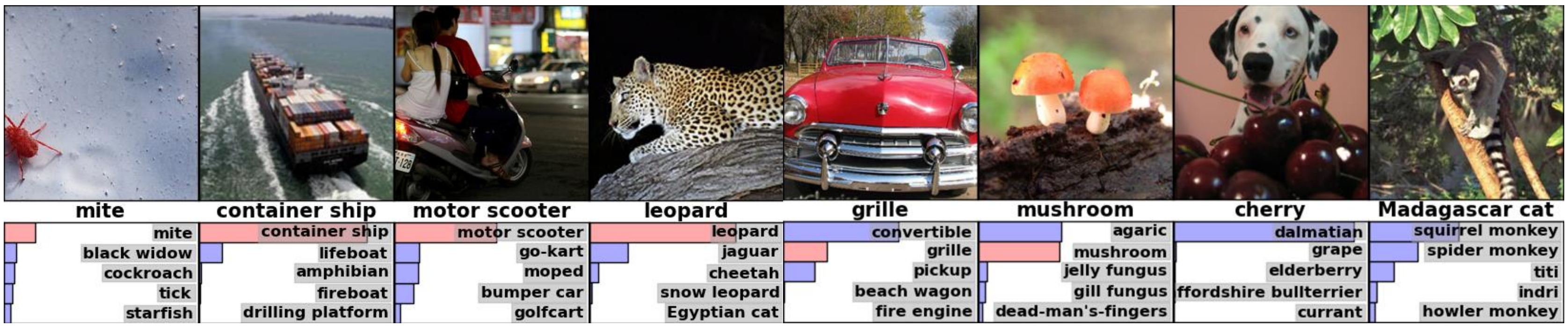


What is it good for?

Application Examples

OBJECT CLASSIFICATION

IMAGENET Large Scale Visual Recognition Challenge:
1.2 million training images of 1000 classes



- 2013: Clarifai: 11.7% top-5 error, or 11.2% with extra data (most contestants used deep nets now).
- 2014: GoogLeNet: 6.7%. Andrej Karpathy (a human): 5.1%.
- 2015: ResNets: 3.6%.

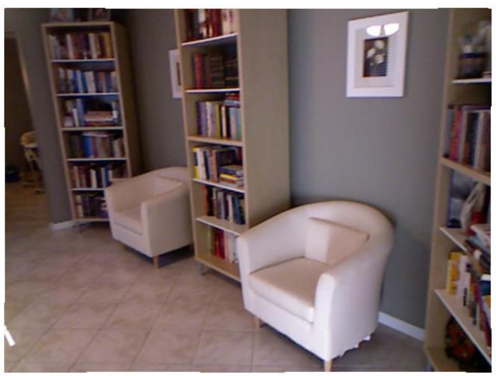
IMAGE CAPTIONING

Input: Photograph. **Output:** Textual description.
Method: CNN to analyze image, RNN to output text

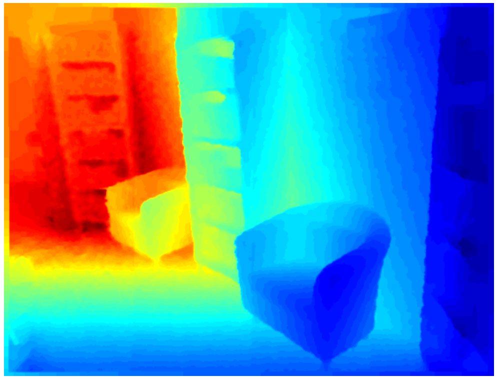
Nov 2014: Show and Tell: A Neural Image Caption Generator, http://arxiv.org/abs/1411.4555

DEPTH ESTIMATION

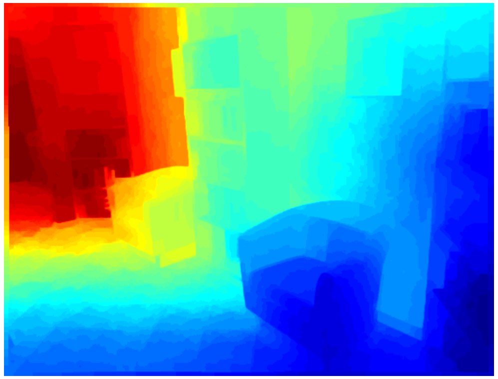
Input: Single photograph
Output: Its depth map (how far is each pixel from the observer)
Method: fully-convolutional network



input



ground truth

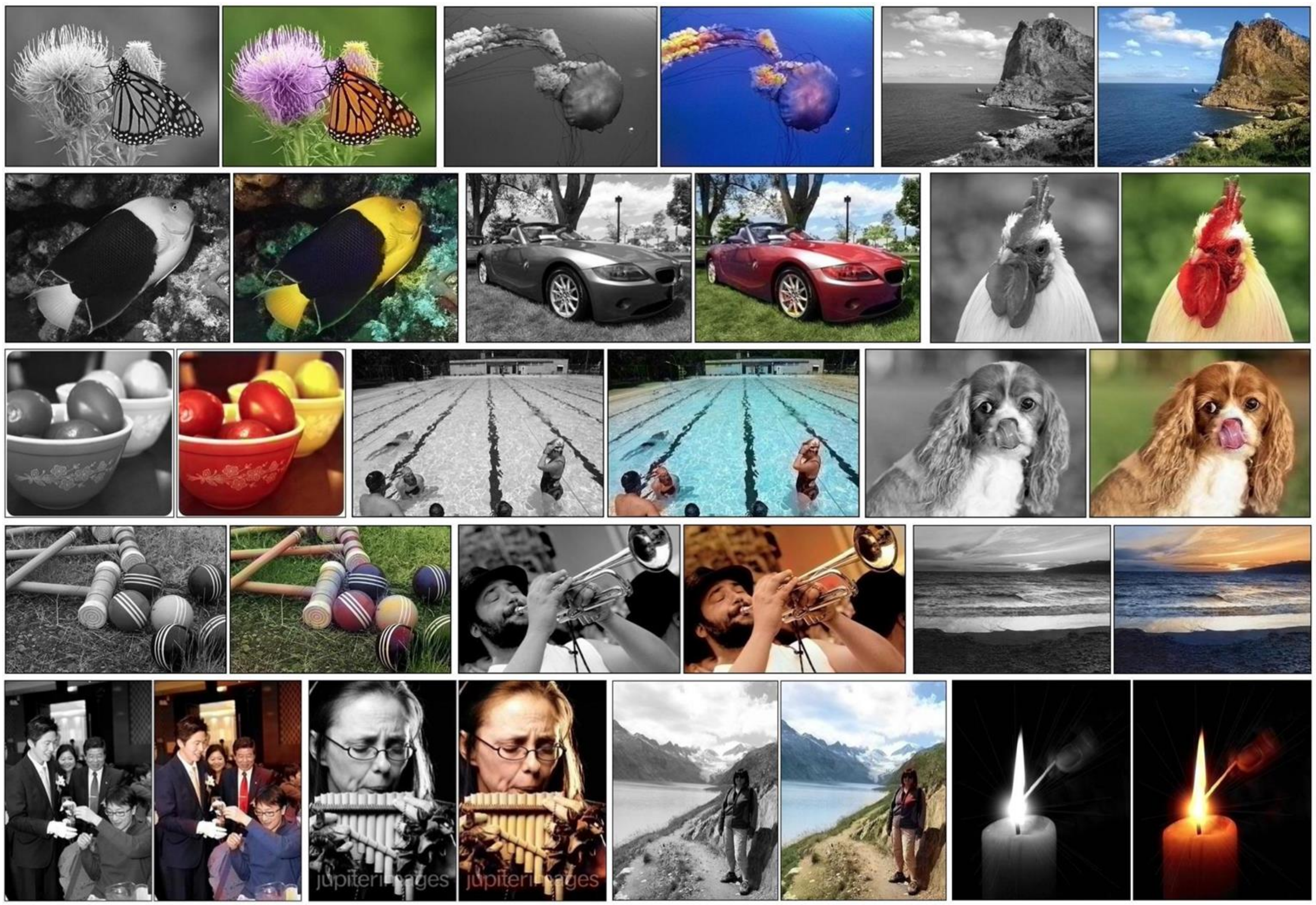


network prediction

Feb 2015: Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields, http://arxiv.org/abs/1502.07411

IMAGE COLORIZATION

Input: Gray-scale image. **Output:** Colored image.
Method: fully-convolutional network



Mar 2016: Colorful Image Colorization, http://arxiv.org/abs/1603.08511,
http://richzhang.github.io/colorization/

NEURAL STYLE TRANSFER

Starting point: ConvNet trained for object classification
Use network in reverse: Find image that matches high-level representation of photograph and low-level repr. of painting

Aug 2015: A Neural Algorithm of Artistic Style, http://arxiv.org/abs/1508.06576

IMAGE GENERATION

Generative Adversarial Networks: One network transforms random noise into images, a second one tries to distinguish generated from real images. Both are trained at the same time.

Photographs of bed rooms that do not actually exist

Nov 2015: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, http://arxiv.org/abs/1511.06434, https://github.com/Newmu/dcgan_code

IMAGE GENERATION

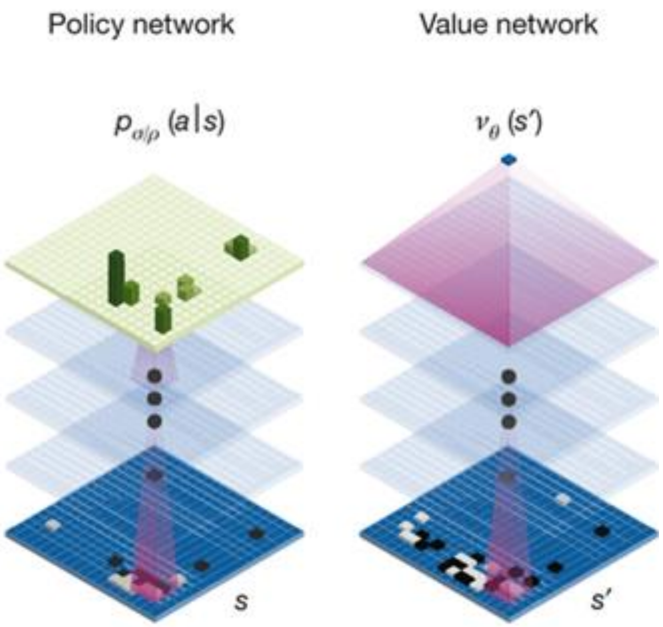
Generative Adversarial Networks: One network transforms random noise into images, a second one tries to distinguish generated from real images. Both are trained at the same time.

Generated album covers

Nov 2015: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, http://arxiv.org/abs/1511.06434, https://github.com/Newmu/dcgan_code

GOOGLE ALPHAGO

Input: Go board (stone positions and some helping markers)
Output: Position for next move (policy) or likely winner (value)
Method: CNN with supervised and reinforcement learning
Combined with Monte-Carlo Tree Search for playing



Nature 526, Jan 2016: Mastering the game of Go with deep neural networks and tree search

QUESTIONS?

AUSTRIAN INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE FOR SAFETY AND SECURITY

AUSTRIAN INSTITUTE OF TECHNOLOGY

AUDIO EVENT DETECTION

Analytic Modules

AUDIO EVENT DETECTION

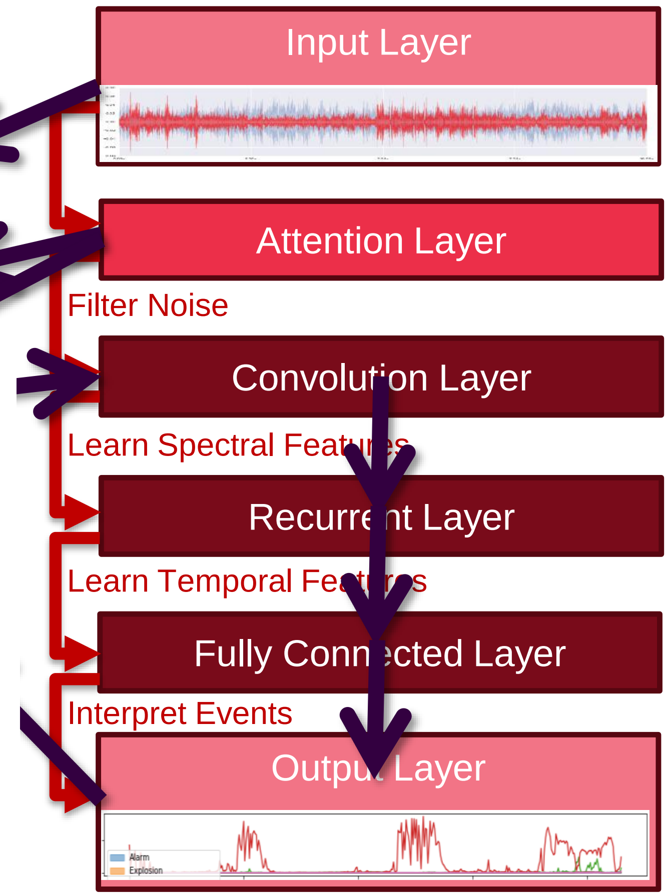
Task

- Detect and predict audio-events into predefined categories
  - Gunshots, explosions, emergency vehicles, scream, speech, Alarm

Use-Case

- Content filter in mass video-data
- Example: attack with firearms
  - => initiate search by filtering all videos which contain *Gunshots* (sorted by confidence)

RECURRENT CONVOLUTIONAL NEURAL NETWORKS



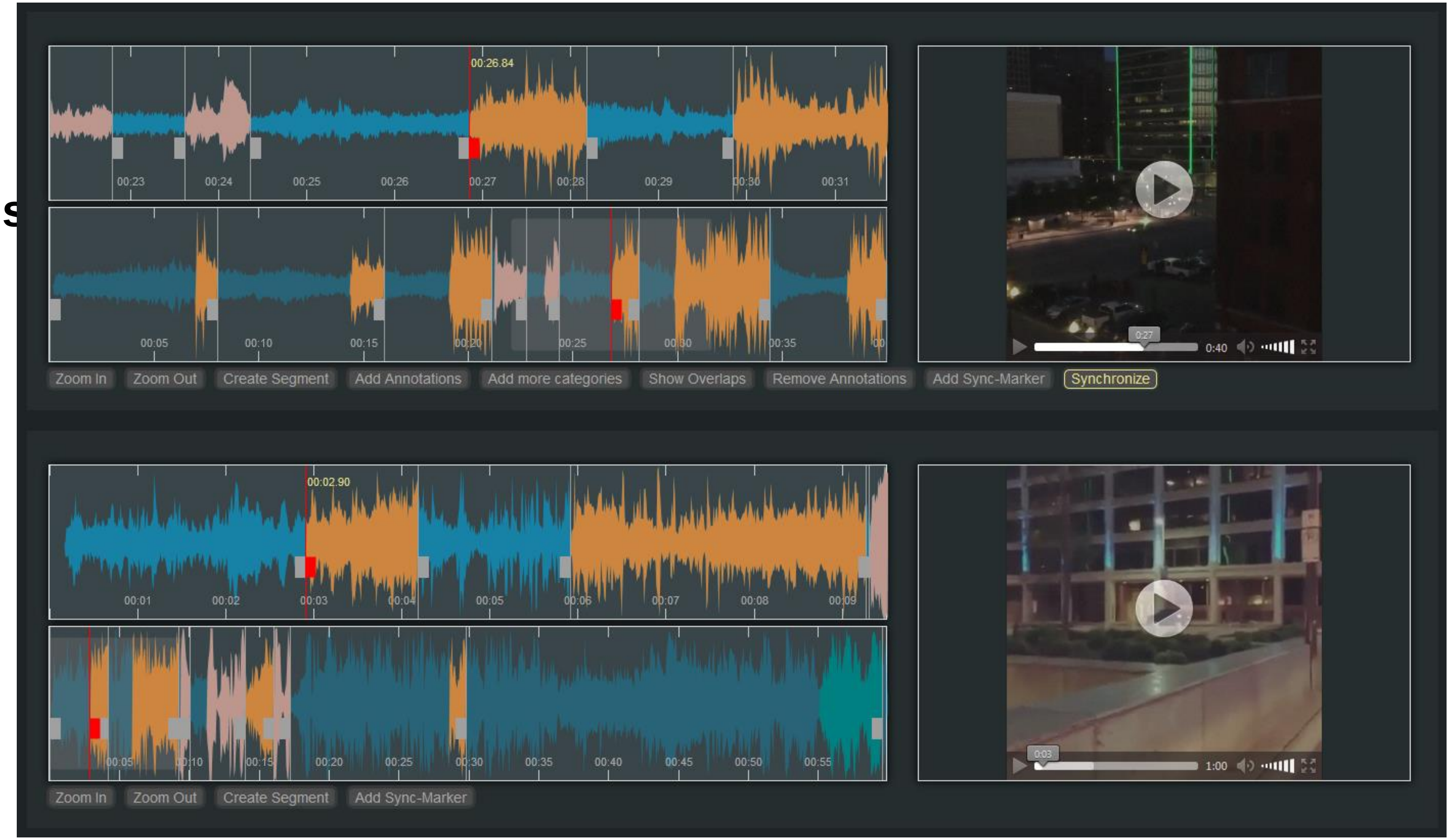
High signal to noise ratio
Due to attention layer

Smoothing Function
Temporal Segmentation per Event Category

AUDIO EVENTS VISUALIZED

AUSTRIAN INSTITUTE OF TECHNOLOGY

Us



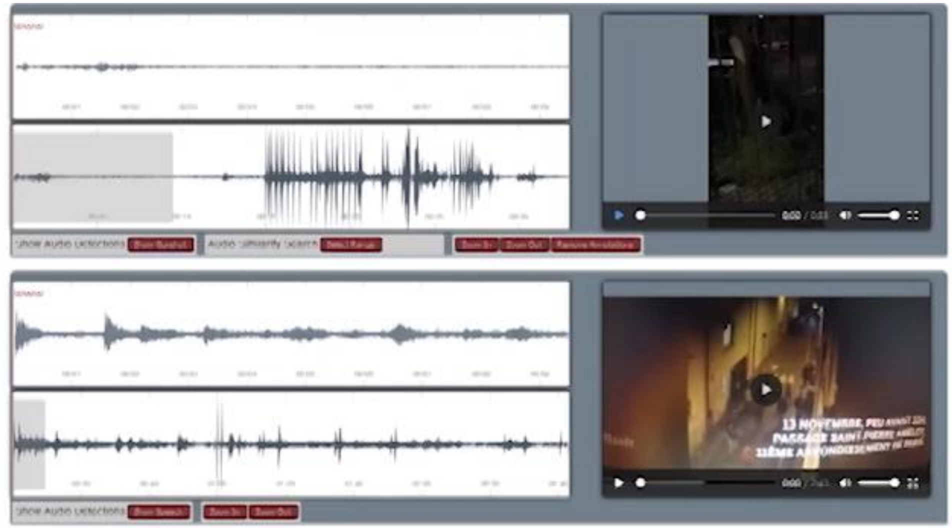
AUSTRIAN INSTITUTE OF TECHNOLOGY

AUDIO SIMILARITY

Analytic Modules

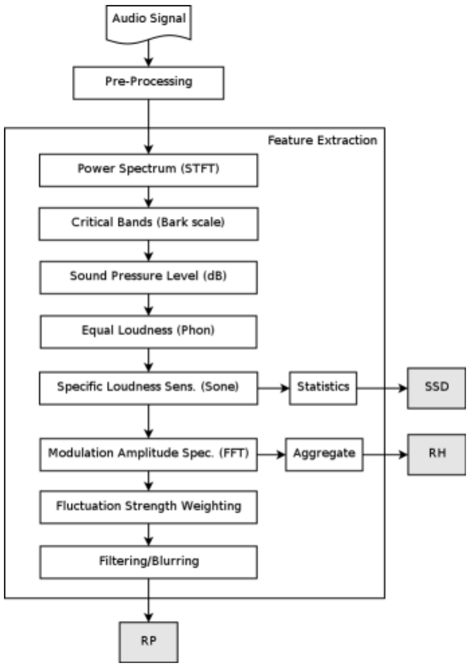
AUDIO SIMILARITY SEARCH

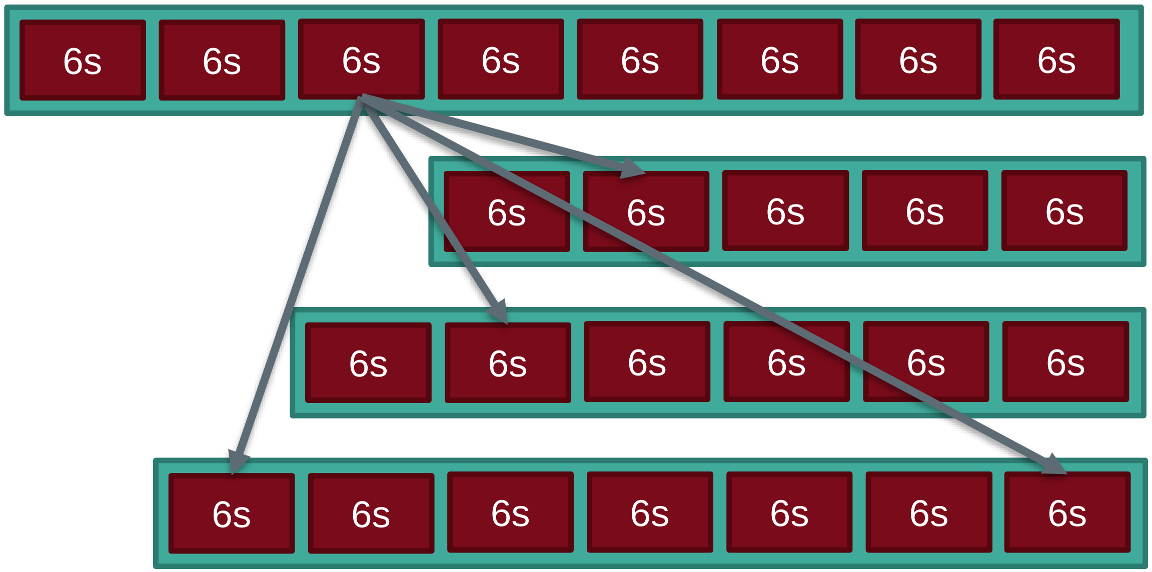
- **Task**
  - Searching for video-segments with similar audio-signature
  - Sub-Segment video-search
- **Use-Case**
  - Suspect could not be identified in one video
  - Select segment and search for others using audio-signature
  - Instant localization (videos close to audio source)
- **Solution**
  - Select range in video
  - Retrieve a list of similar sounding video segments
  - Sorted by simlarity



AUDIO SIMILARITY SEARCH

- **Audio features extracted for each 6s segment**
  - Rhythm Patterns (repetitiveness in audio)
  - Statistical Spectrum Descriptors
- Nearest Neighbor search using late fusion in a normalized feature space





AUDIO SIMILARITY – INSTANT LOCALIZATION (WITHOUT GPS)

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

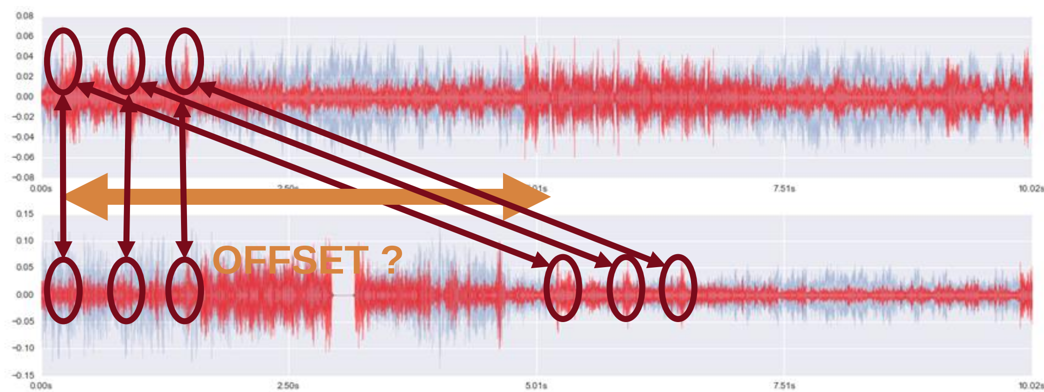
AUDIO-BASED VIDEO-SYNCHRONIZATION

Analytic Module

AUDIO-BASED VIDEO-SYNCHRONIZATION

- **Task**
  - Synchronize various video files with unreliable time metadata
  - Use audio-signature to relatively align video files

- **Technology**
  - Audio-fingerprints (chromaprint)
  - Noise invariant

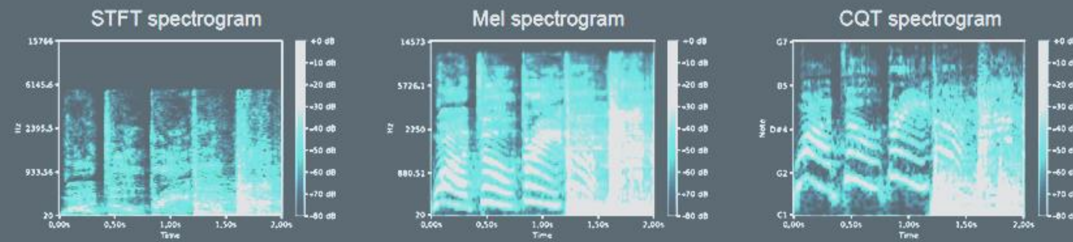


SCENE ANALYSIS

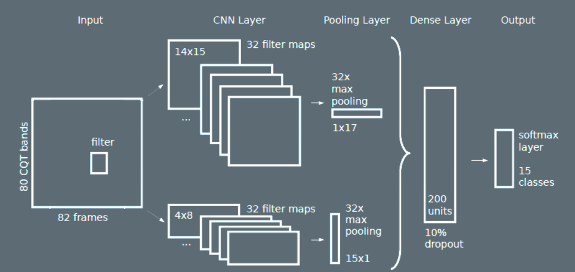
AUSTRIAN INSTITUTE OF TECHNOLOGY

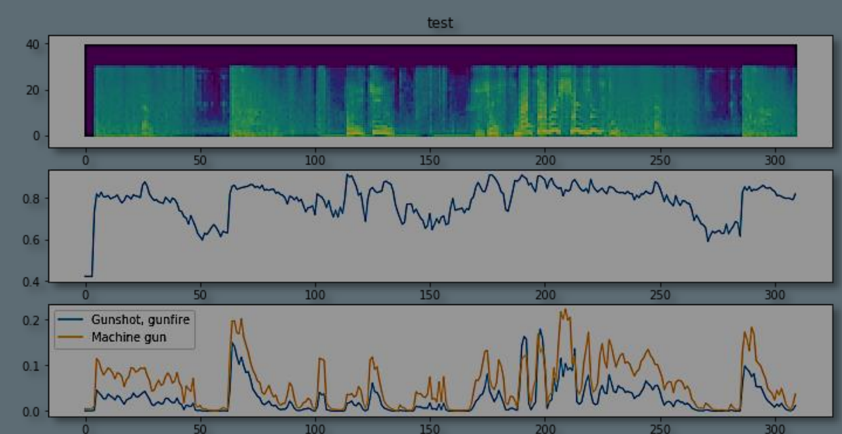
Audio-Visual Scene Understanding

Artificial Intelligence
Multi-media Analysis



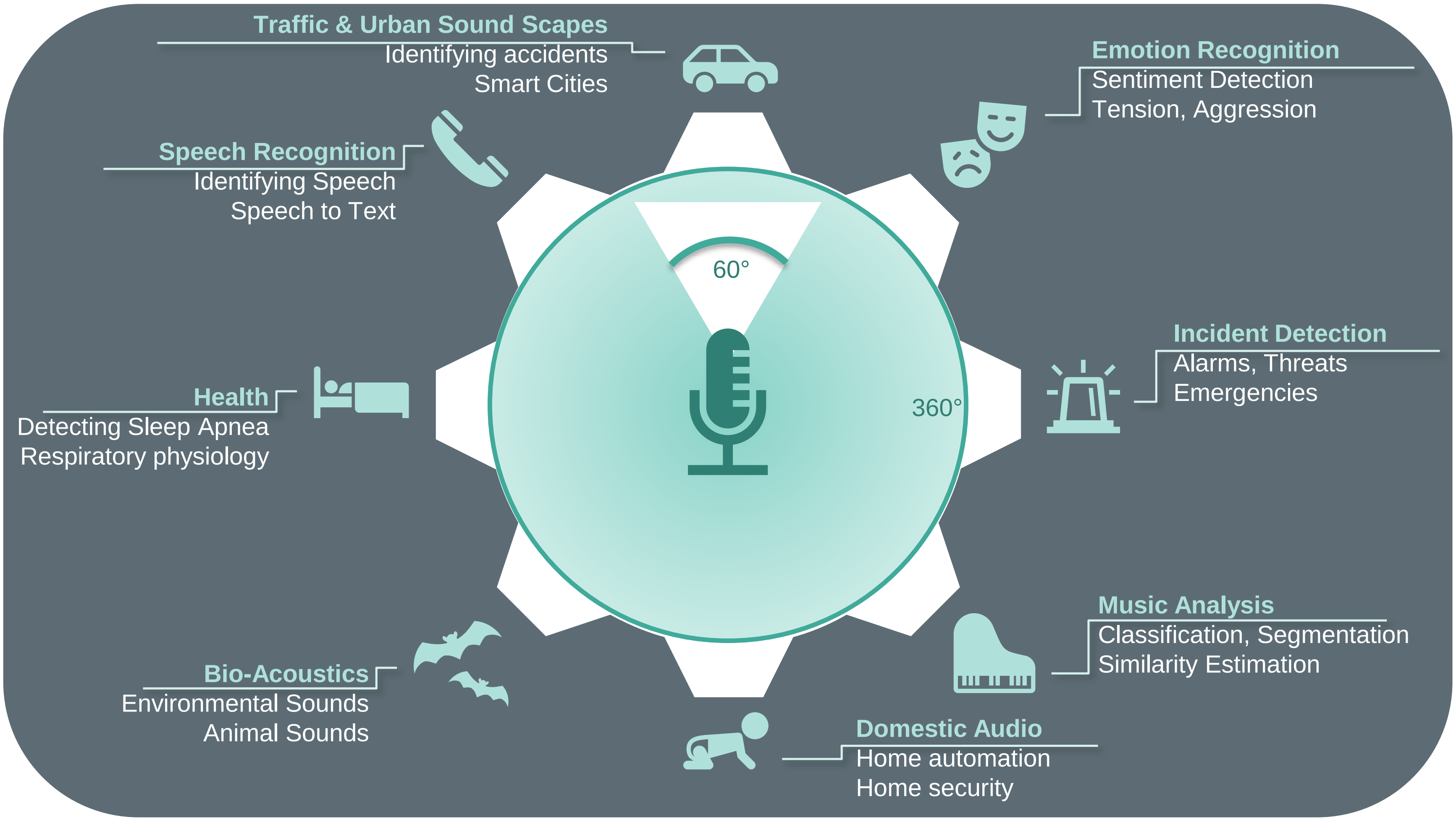
- **Acoustic Scene Classification**
  - Identify different acoustic scenes (Bus, Train, Urban Park, Bar)
  - Identify different activites (talking, walking, reading, children playing)
- **Audio Event Detection**
  - Detect acoustic events in audio-stream (onsets/length)
  - Identify detected Events (Gunshots, explosions, baby cry)
- **Audio-Visual Scene Understanding**
  - Combining acoustic with visual information
  - Improved interpretability of current scene
  - Multi-task learning





AUSTRIAN INSTITUTE OF TECHNOLOGY

AUDIO ANALYSIS



VISUAL COMPUTING

OBJECT DETECTION AND TRACKING

Detecting Persons and Objects in Videos

FLORIDA and VICTORIA Project
Scalable Computing

- **Overview**
  - Tracking Objects and Persons in Videos
  - Video-Frame Segmentation / Bounding Boxes
- **Technology**
  - Regional Convolutional Neural Networks (R-CNN)
  - YOLO
  - Connected Vision
- **Application**
  - Applied in two security sensitive projects
  - Post attack video forensich

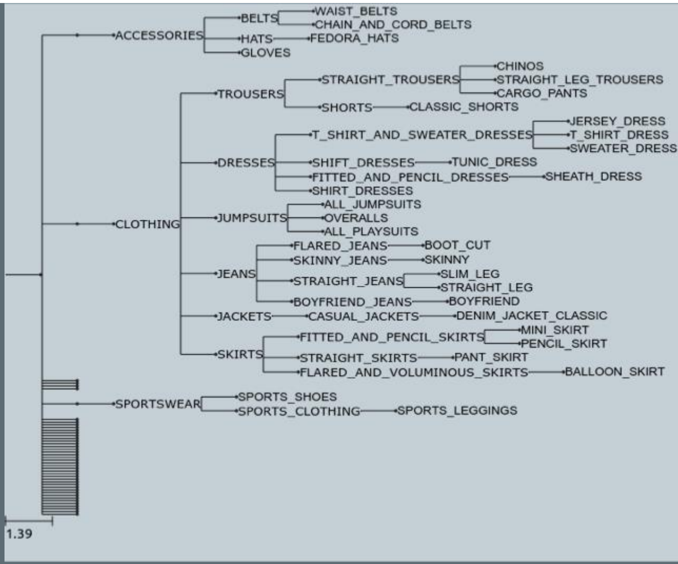
FASHION IMAGE CLASSIFICATION

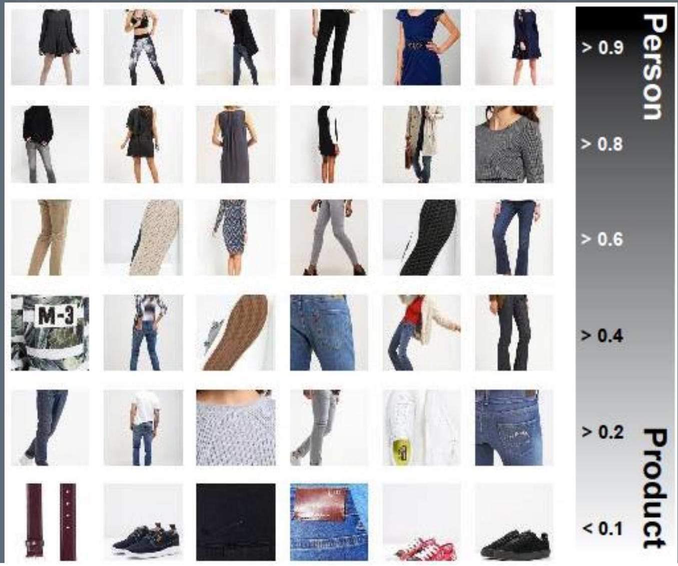
AUSTRIAN INSTITUTE OF TECHNOLOGY

Image Classification with Deep Neural Networks

Large Scale Deep Learning
Visual Computing

- **Overview**
  - Empirical study
  - Applying deep Convolutional Neural Networks to fashion classification
  - Evaluated five CNN architectures
  - Custom and pre-trained models
  - Evaluated on three tasks
    - Person detection
    - Product classification
    - Gender prediction
- **Dataset**
  - 234.884 Product Images
  - 39.474 Products
  - 7.833 Person labels
- **Conclusions**
  - Despite large dataset and reduced number of classes
  - Pretrained models outperform from-scratch training
    - Product classification – 79.1%
    - Gender prediction – 88.0%
  - Preprocessing of ground-truth required





BIRD SONG IDENTIFICATION

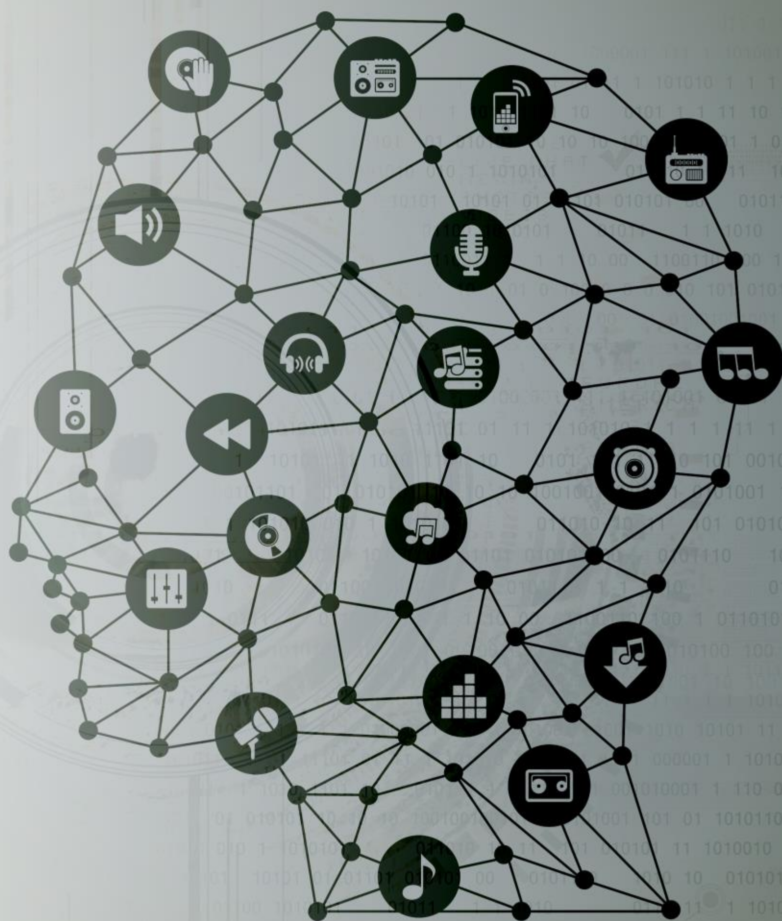
AUSTRIAN INSTITUTE OF TECHNOLOGY

Bio-Acoustics

Artificial Intelligence
Ecology, Environmental Protection

- **Large Scale Classification Problem**
  - Identify 1500 different bird species
  - Large variations in audio quantity and quality per species
- **Multi-Modal Neural Network Approaches**
  - Combining Audio-Information with Geo-Information
  - Normalizing temporal information by region (e.g. Dusk/Dawn)
- **Sophisticated Data-Augmentation**
  - Time-stretching (faster/slower)
  - Pitch-shifting (higher/lower)
  - Mixing recordings from same species
  - Mixing surrounding recordings (1°E/W/N/S Neighborhood)
  - Random cuts (cut-shuffle-merge)
  - Volume-shifting
  - Noise Overlays

MULTI-MODAL-ANALYSIS



MULTI-MEDIA UNDERSTANDING

Audio-Visual Video Analysis and Classification

Artificial Intelligence
Multi-media / Multi-modal

Identifying Content

Detecting Objects
Detecting Events

Describing Content

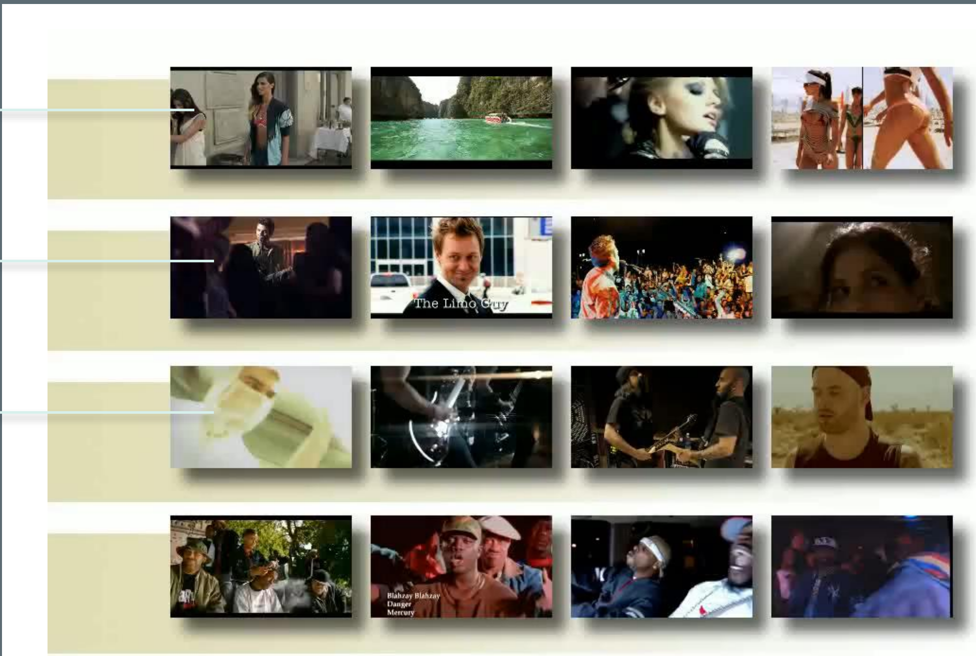
Visual / Acoustic characteristics
Colors, Texture, Timbre, Patterns

Analyzing Content Dependencies

Temporal distribution of detected content / objects
Knowledge modelling

Interpreting Scenes

Action, Location, Intention
Colors, Texture, Timbre, Patterns



SENTIMENT DETECTION

Sentiment Detection

Artificial Intelligence
Multi-media / Multi-modal

Crowd scene / behaviour analysis

Anomality detection in natural scenes / crowds
Threat / escalation estimation

Audio Analysis

Audio based sentiment detection
Arousal, Pleasure, Dominance

Speaker Sentiment Detection

Sentiment of speaker
Tension, excitement, affection

Affective Contrasts

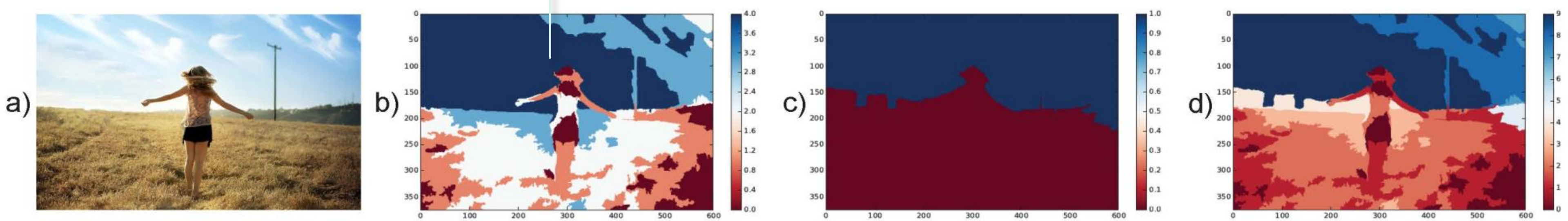
Cold / Warm
Light / Dark

Affective Interpretation

Deduced from Color Information
Arousal, Pleasure, Dominance

Color Statistics

Calculating color distributions
Deriving higher level features



SPEAKER / SINGER IDENTIFICATION

AUSTRIAN INSTITUTE OF TECHNOLOGY

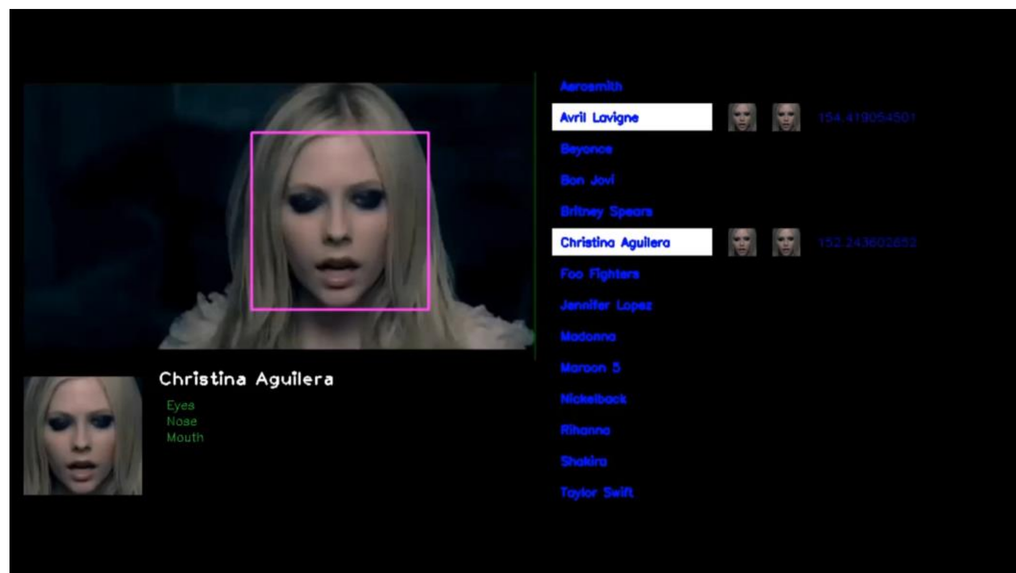
Audio-Visual Signer Identification

Artificial Intelligence
Multi-media / Multi-modal

- **Identify Speaker / Singer in Audio/Video Sequence**
  - Based on segmentation (identified vocalized segments)
  - Supervised identification (models trained on declared persons)
  - Un-supervised identification (relative similarity estimation)

- **Audio-Identification**
  - Based on acoustic models
  - Distinguis between speakers
  - Transcription of spoken words

- **Audio-Visual Identification**
  - Based on acoustic and visual models
  - More accurate
  - Visual segmentation / boxes
  - Visual identification / tracking



AI SOLUTION EXAMPLE

AUSTRIAN INSTITUTE OF TECHNOLOGY

Example: Forensic Analytics in Massive Video Content

Artificial Intelligence
Scalable Computing

Visualization
Data Aggregation & Visualization
Interactive Processing

Large Scale Computing
Distributed Computing Clusters
Cloud Computing Big Data Architectures

Visual Object Detection
Suspects, Cars, Suitcases, Weapor
License plates, 3D Trajectories

Scalable Plattform
Apache Hadoop

GPU Plattform
NVIDIA, Cuda

Audio Event Detection
Gunshots, Explosions, Screams
Spoken Words, Transcription

Massive Video Data
e.g. after an terroristic attack
CCTV, mobile phones

GPU Scale Computing
Deep Learning, Deep Neural Networks
Artificial Intelligence Modules

FURTHER RESOURCES

Code & Tutorials

AUSTRIAN INSTITUTE OF TECHNOLOGY

Vienna Deep Learning Meetup

Vienna’s largest monthly event on Deep Learning & AI

The Organizers:

Thomas Lidy
Musimap

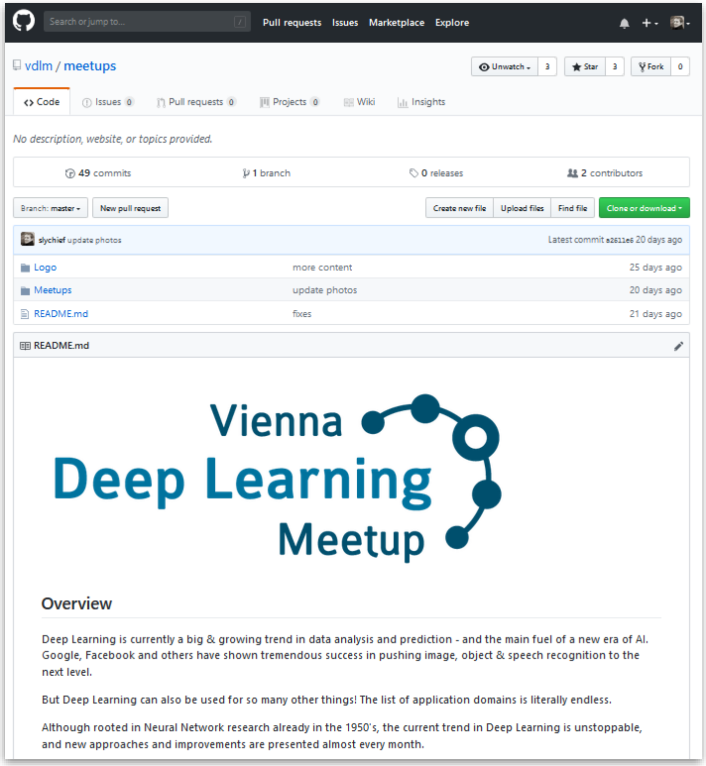
Alex Schindler
AIT & TU Wien

René Donner
contextflow

Jan Schlüter
OFAI & UTLN

www.meetup.com/Vienna-Deep-Learning-Meetup

VDLM ON GITHUB



- ➔ Talks
- ➔ Slides
- ➔ Videos
- ➔ Wiki with beginner's resources

Talks

Date	MU#	Speaker	Topic	Slides
2016-04-07	1	Thomas Lidy	An overview presentation of Deep Learning	pdf
2016-04-07	1	Jan Schlüter	History, Approaches, Applications	pdf
2016-05-09	2	Alex Champandard	Neural Networks for Image Synthesis	
2016-05-09	2	Gregor Mitscha-Baude	Recurrent Neural Networks	pdf
2016-06-06	3	Jan Schlüter	Open-source Deep Learning with Theano and Lasagne	pdf
2016-09-22	5	Josef Puchinger	Deep Learning & The Future of Automation	
2016-09-22	5	Christoph Körner	Going Deeper with GoogLeNet and CaffeJS	pdf

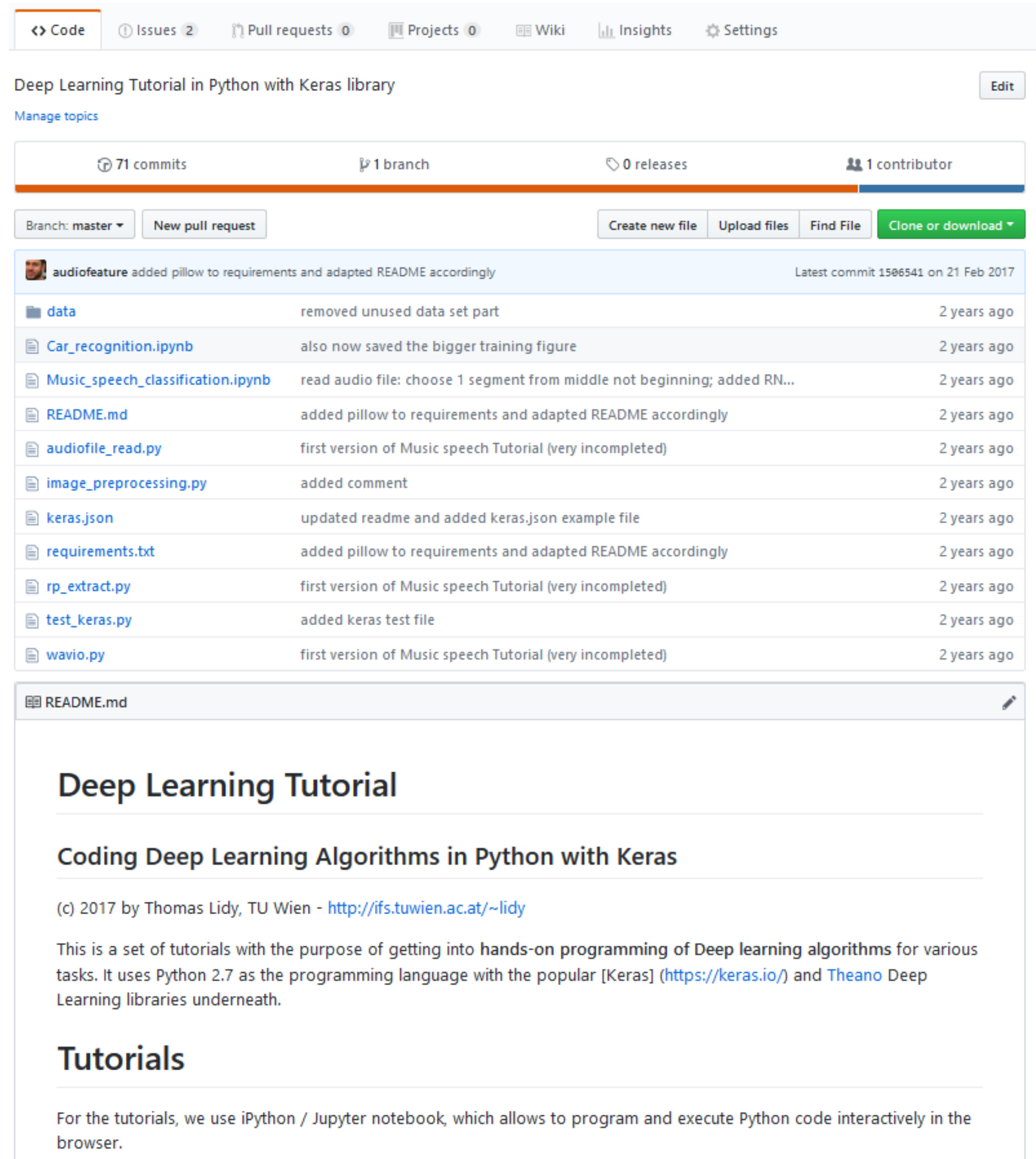
github.com/vdlm/meetups

AUSTRIAN INSTITUTE OF TECHNOLOGY

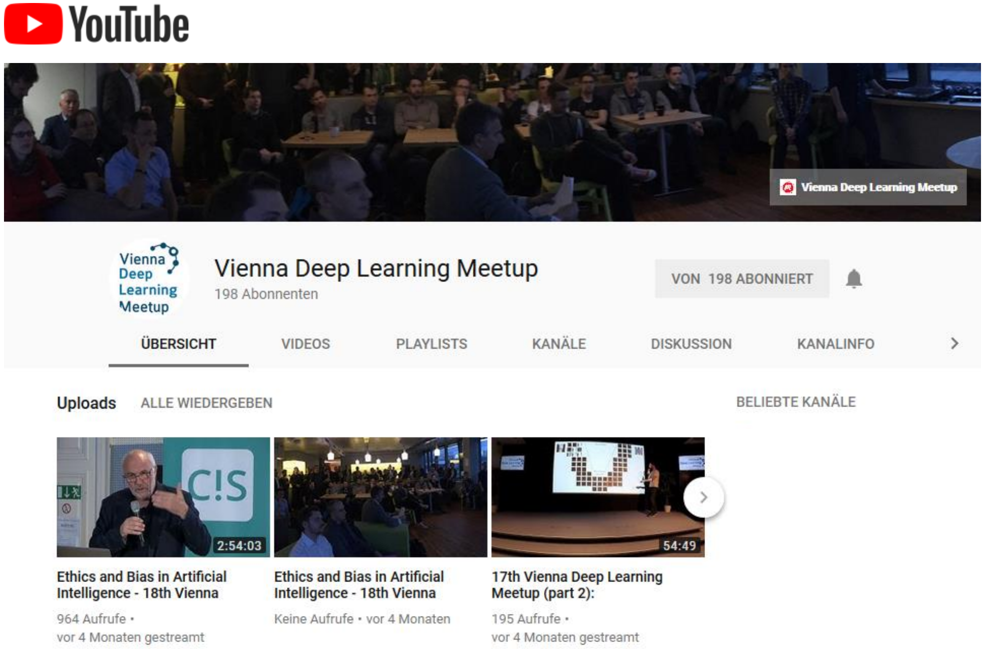
DEEP LEARNING TUTORIAL

- Beginners Tutorial
  - Simple Image Processing
  - Simple Audio Processing
  - Python
  - Keras
  - Tensorflow

https://github.com/tuwien-musicir/DL_Tutorial

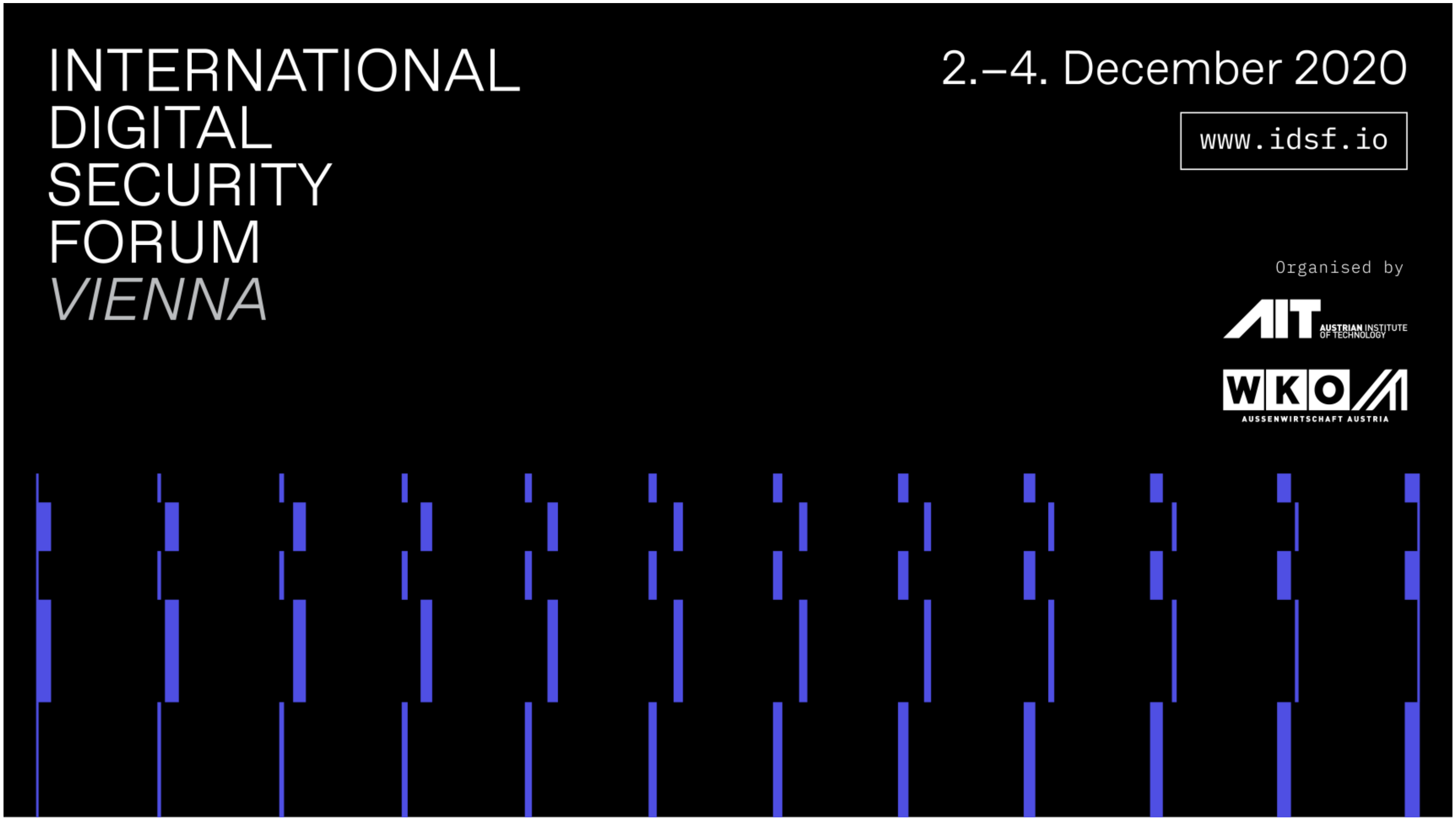


VDLM YOUTUBE CHANNEL

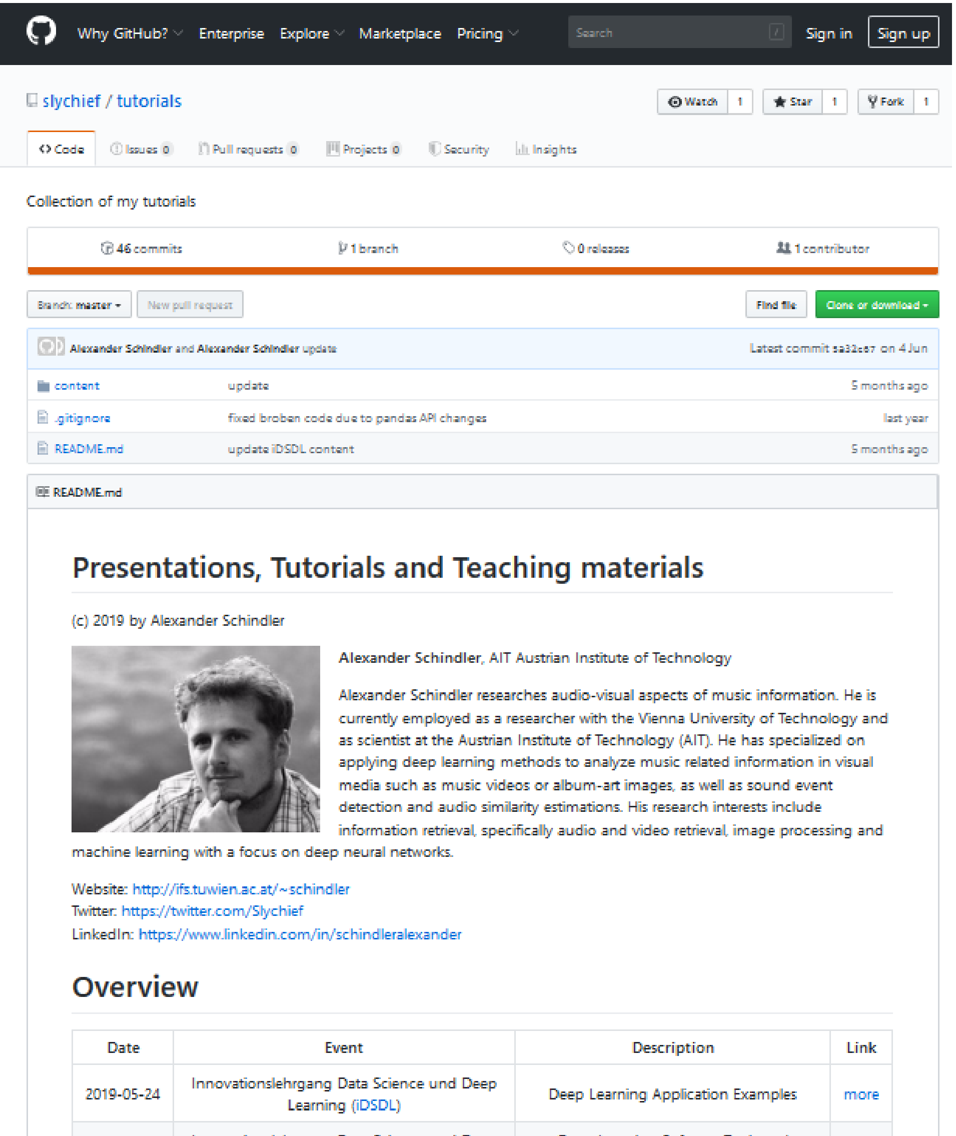


www.youtube.com/ViennaDeepLearningMeetup

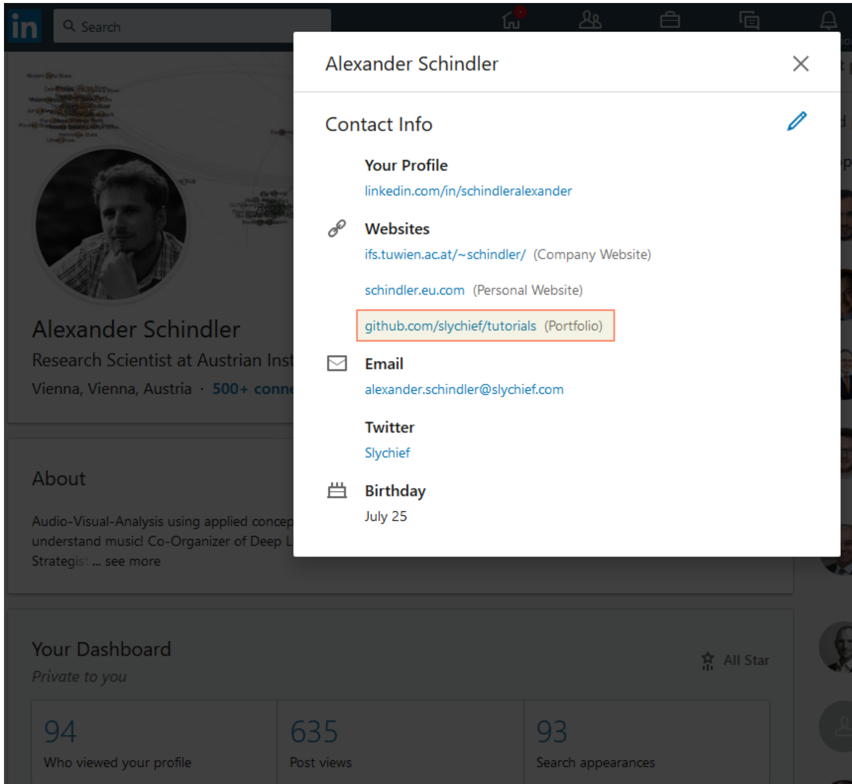
INTERNATIONAL DIGITAL SECURITY FORUM



SLIDES OF THIS TALK



- Further Questions?
- Connect via LinkedIn!



AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

THANK YOU!