ARTIFICIAL INTELLIGENCE & SECURITY

Alexander Schindler
Scientist
Information Management
Center For Digital Safety & Security

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

ALEXANDER SCHINDLER

- **Scientist @ AIT**
  - Applied Artificial Intelligence at 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)

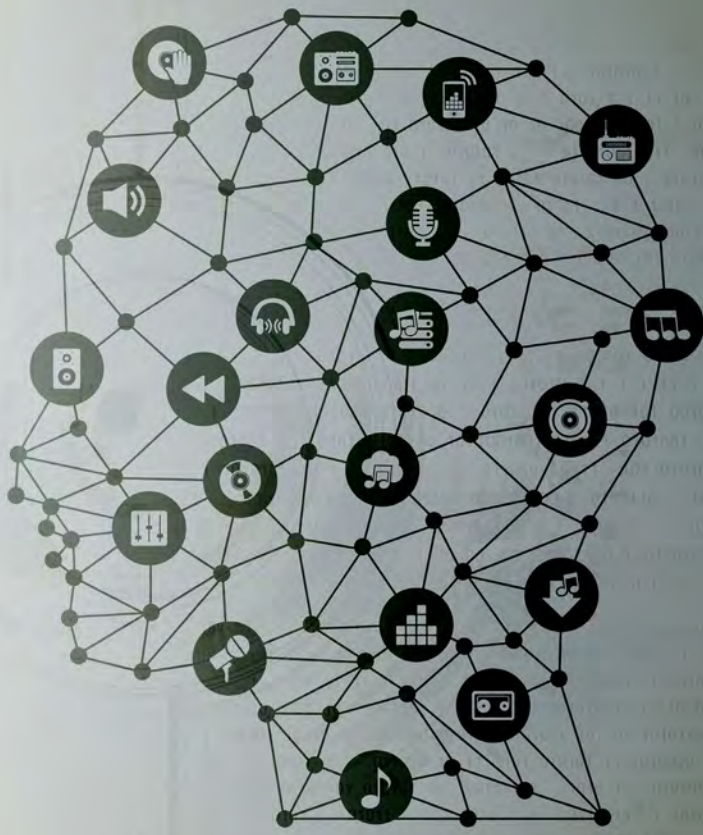
OUTLINE

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

AUSTRIAN INSTITUTE OF TECHNOLOGY

Introduction to

Artificial Intelligence



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

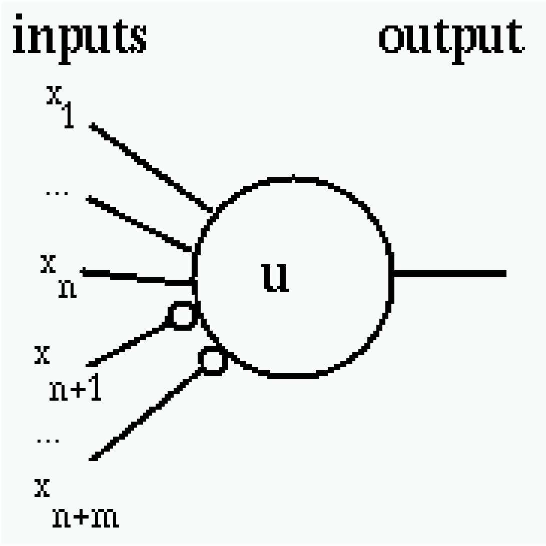
HISTORY

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
  - *Stochastic neural analog reinforcement calculator (SNARC)*



1956: **The Dartmouth Conference**

- **two-month workshop** for researchers interested in
  - neural networks
  - the study of intelligence

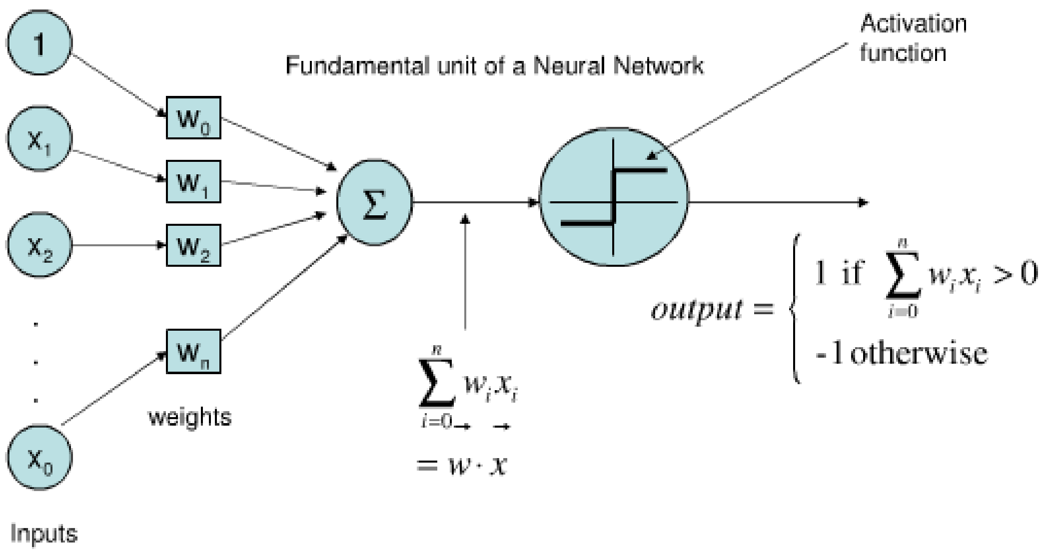
- Organizers
  - John McCarthy (Stanford)
  - Marvin Minsky (MIT)
  - Herbert Simon (CMU)
  - Allen Newell (CMU)
  - Arthur Samuel (IBM)

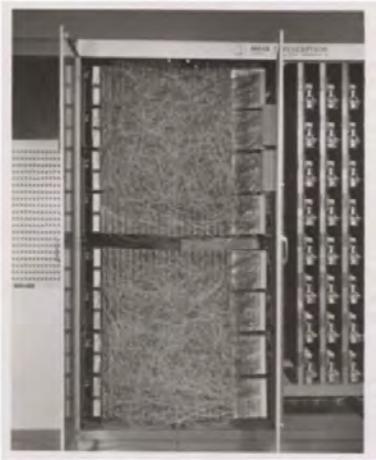
Wikipedia

- **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

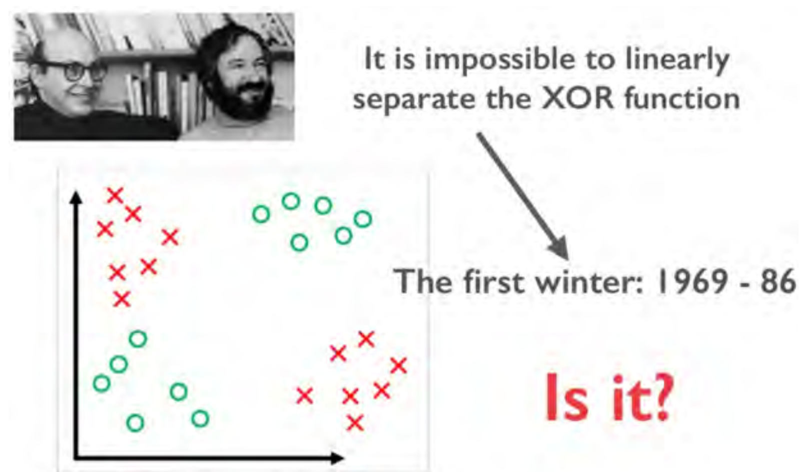
- 1952-1962: Checkers player (by Arthur Samuel)
- 1957: Newell and Simon: "within ten years a digital computer will be the world's chess champion"
- 1967: MacHack achieved class-C rating in tournament chess

UPS & DOWNS OF NEURAL NETWORKS

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



UPS & DOWNS OF NEURAL NETWORKS

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**

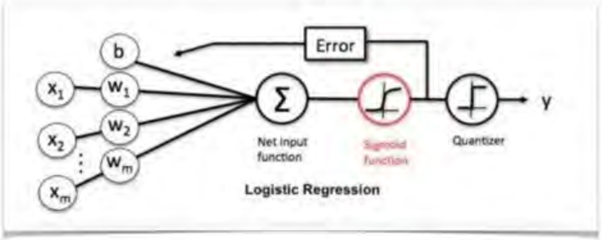
- Birth of expert systems
- Idea is to give AI systems lots of information to start with
  - Rule-based Systems
  - Fuzzy Logic

UPS & DOWNS OF NEURAL NETWORKS

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

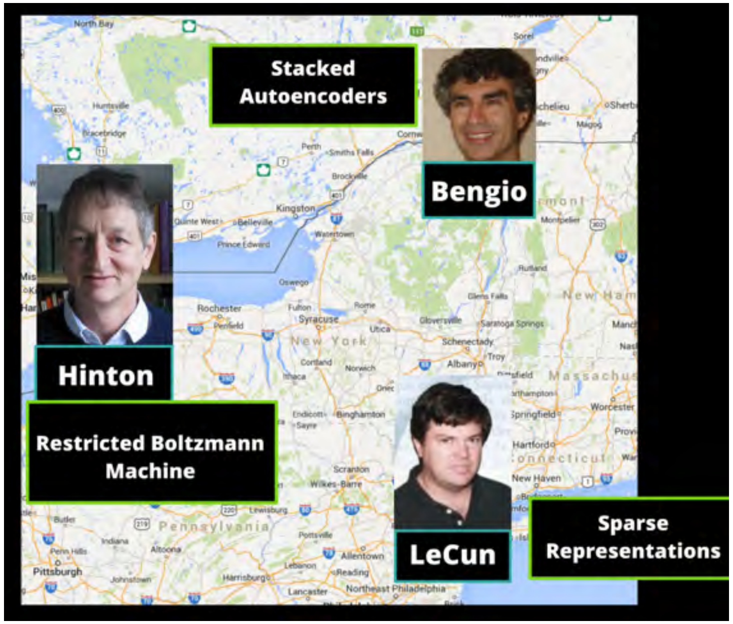
Wikipedia

Wikipedia

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

AUSTRIAN INSTITUTE OF TECHNOLOGY

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

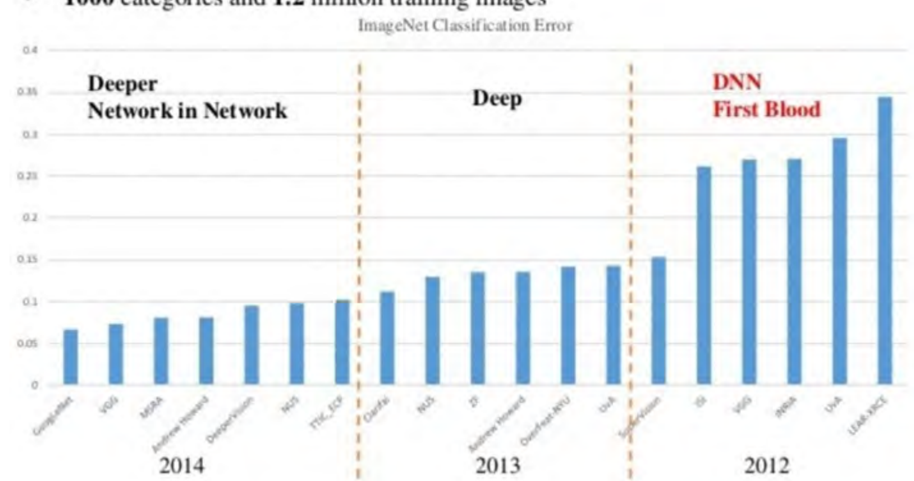
2012: AlexNet (first deep net) pushed state-of the art object recognition in images by far (16.4% error). Most people in object recognition now use deep nets.

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

2015: ResNets half object recognition again to 3.6%.

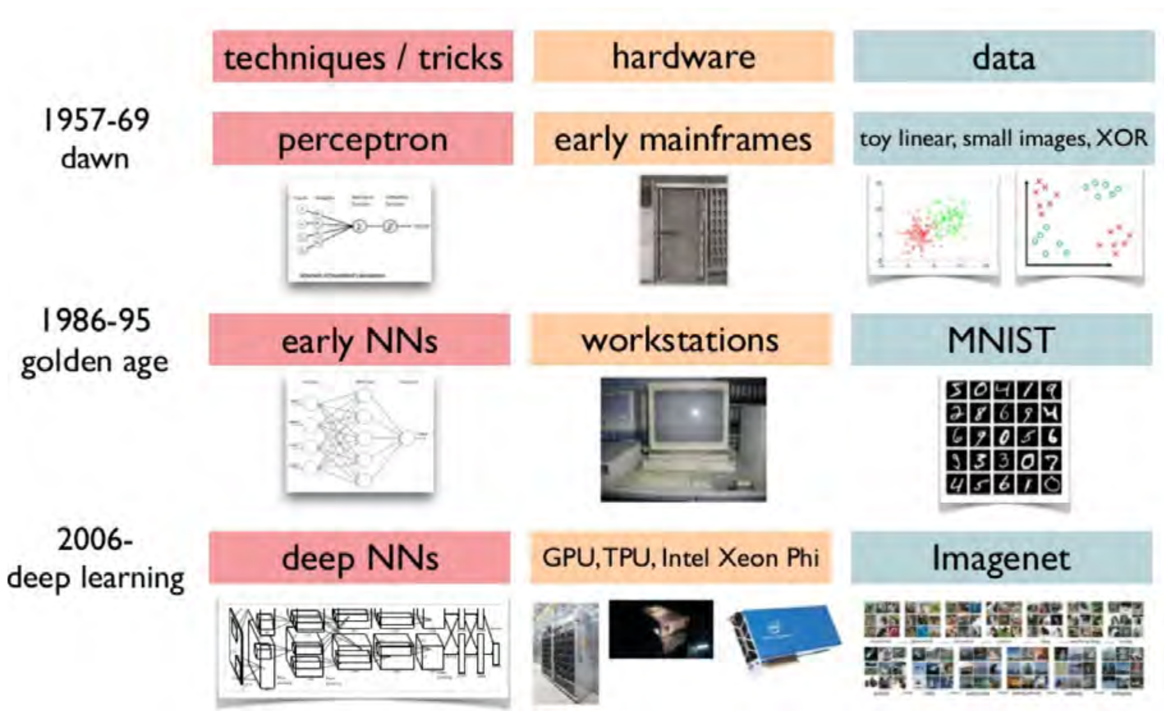
2016: Google DeepMind's AlphaGo beats Go champion 4:1

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



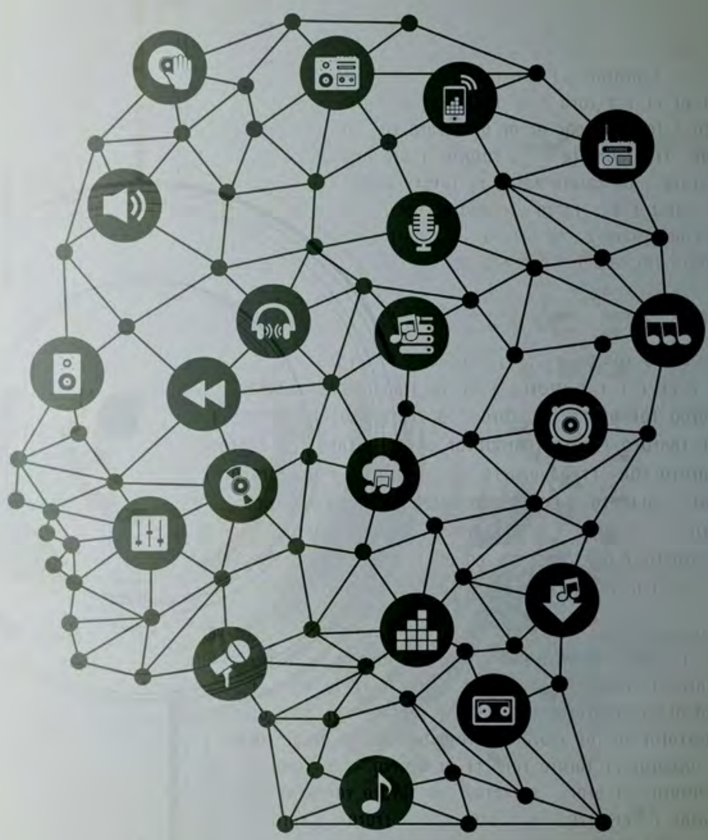
Li Fei-Fei: ImageNet Large Scale Visual Recognition Challenge, 2014 http://image-net.org/

THREE EPOCHS OF NEURAL NETWORKS

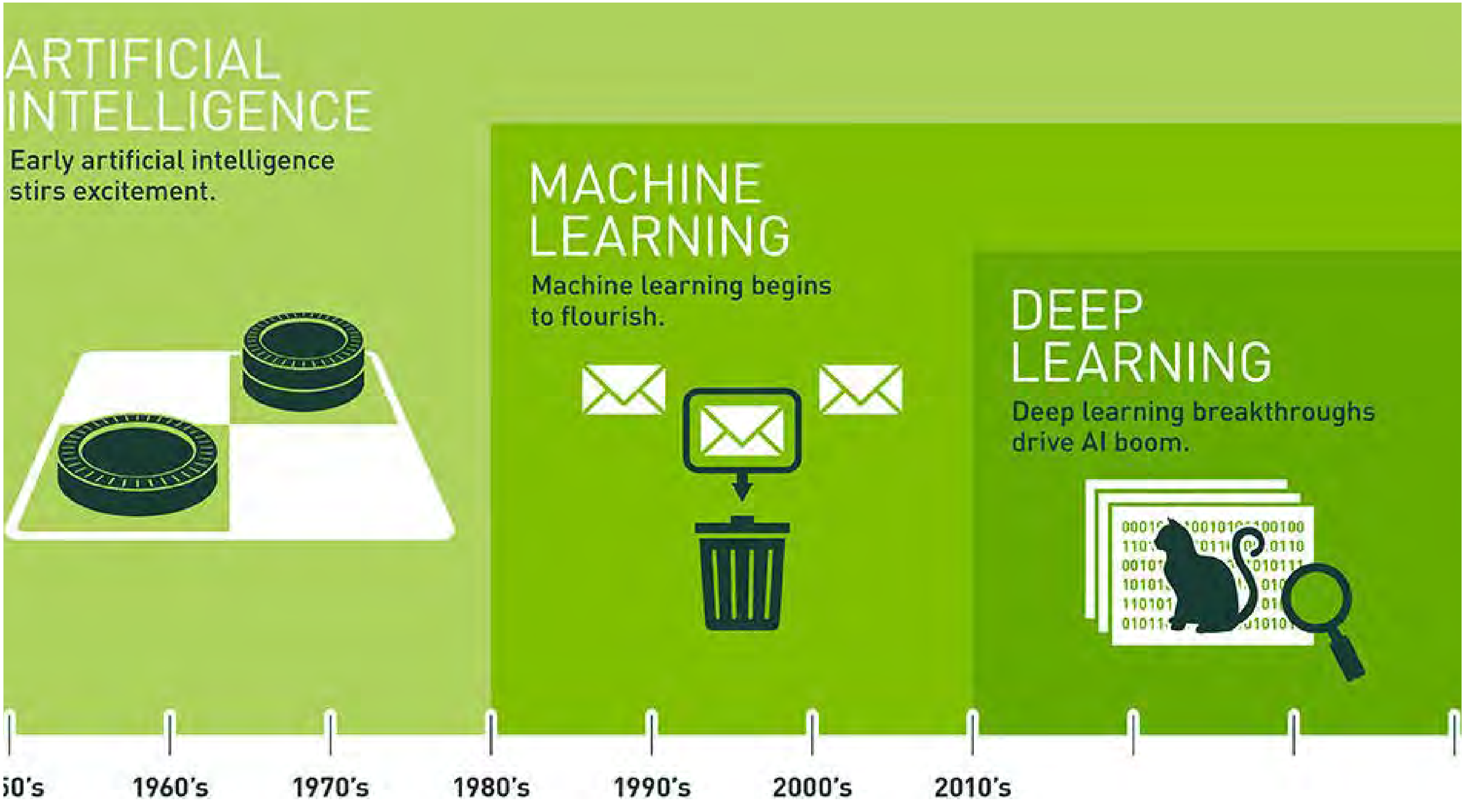


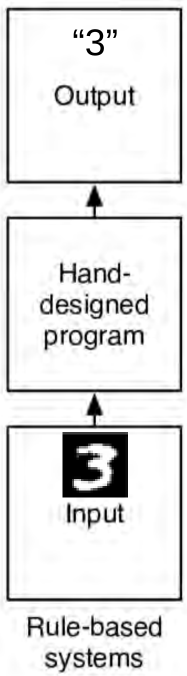
Introduction to

Deep Learning



THREE EPOCHS OF NEURAL NETWORKS



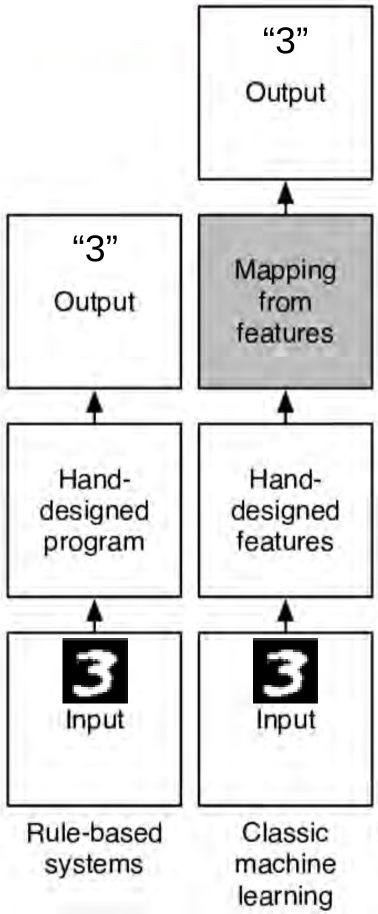


Rule-based systems:

Write algorithm by hand.

graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS

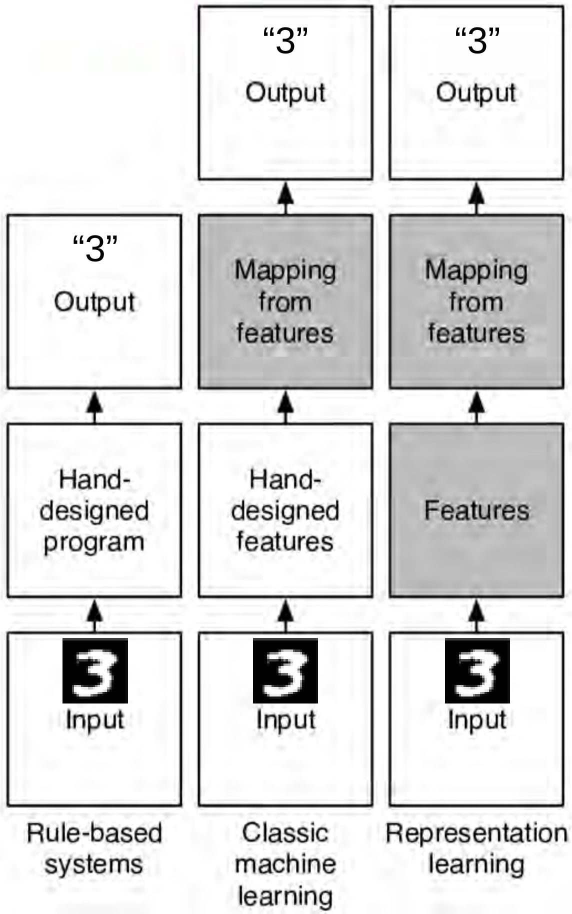


Rule-based systems:
Write algorithm by hand.

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

graphic: Y. Bengio, Deep Learning, MLSS 2015

MACHINE LEARNING PARADIGMS



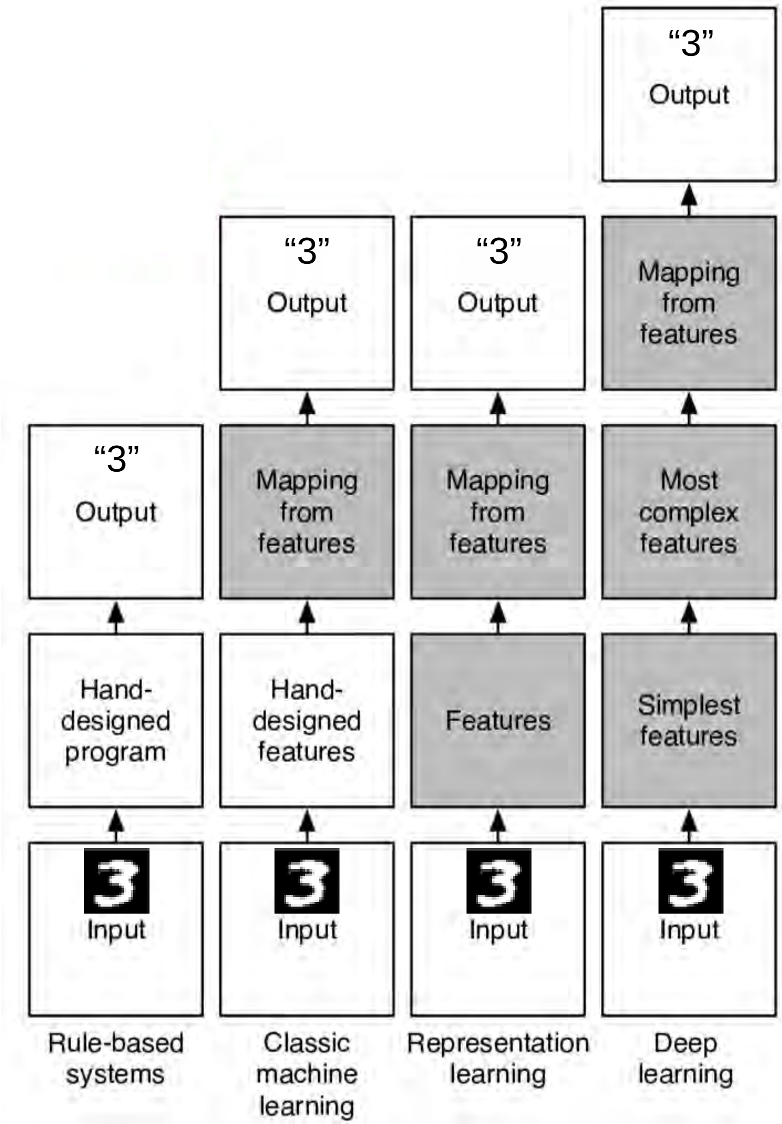
Rule-based systems:
Write algorithm by hand.

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

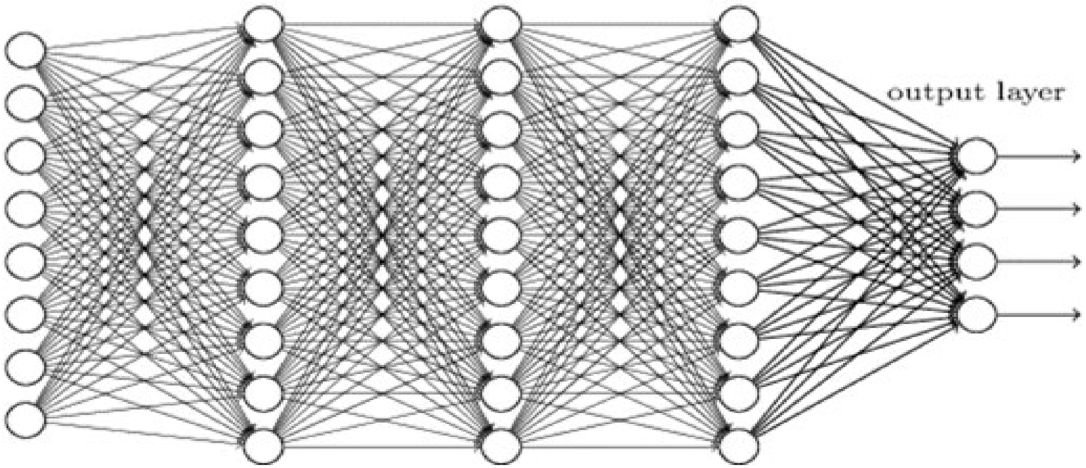


Deep learning:
Learn a stack of many simpler functions to map input to output.

- 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:

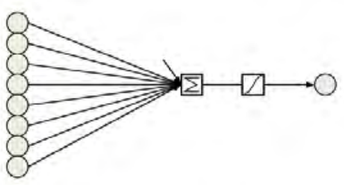
$$y = \sigma(b + \mathbf{w}^T\mathbf{x}) \qquad \text{(equivalent to logistic regression)}$$

WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

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

expression can be visualized as a graph:

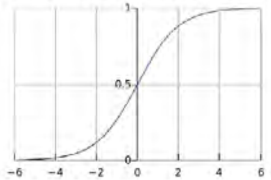


$\mathbf{x}$ $\qquad b + \mathbf{w}^T\mathbf{x}$ $\qquad y$

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

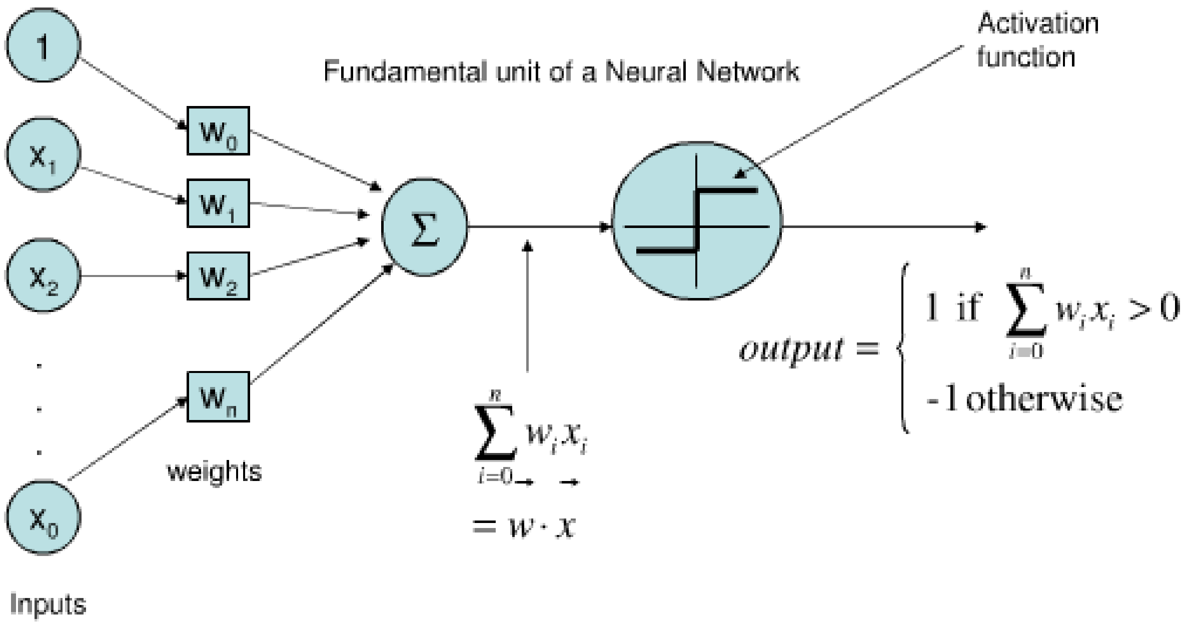
$$b + \mathbf{w}^T\mathbf{x} = b + \Sigma_i w_i x_i$$

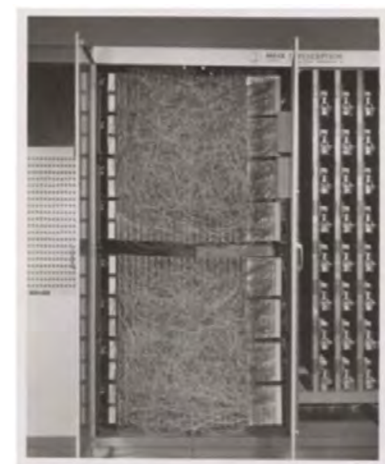
followed by a nonlinear function.



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

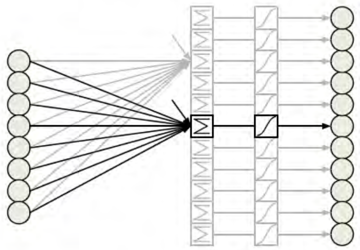
For the first time a NN could solve simple classification problems merely from training data

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:

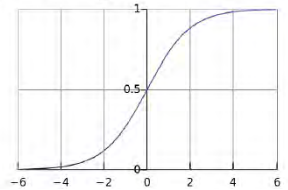


$\mathbf{x}$ $\qquad$ $\mathbf{b} + \mathbf{W}^T\mathbf{x}$ $\qquad$ $\mathbf{y}$

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.

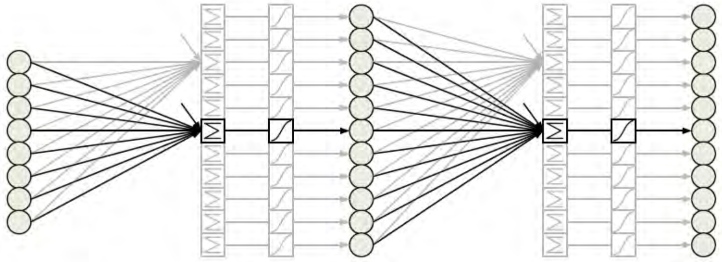


WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

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

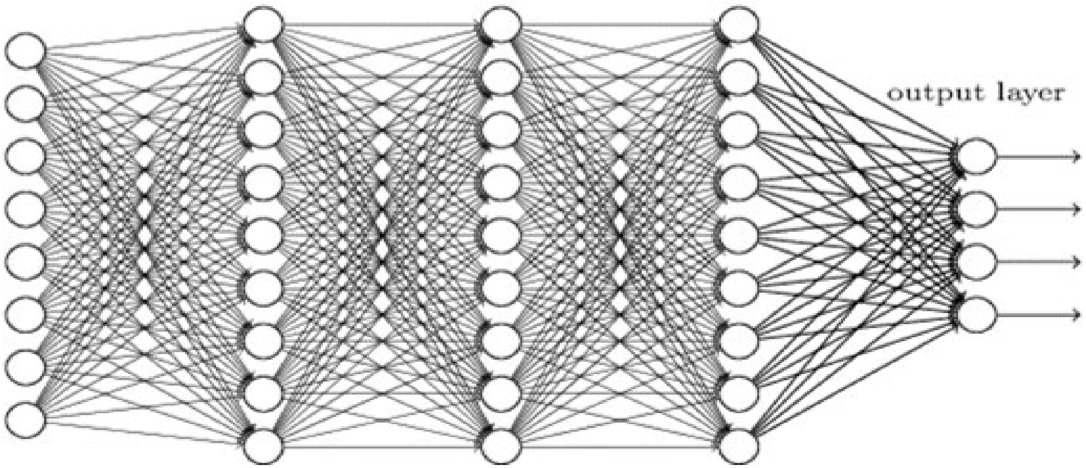
expression can be visualized as a graph:



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

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)



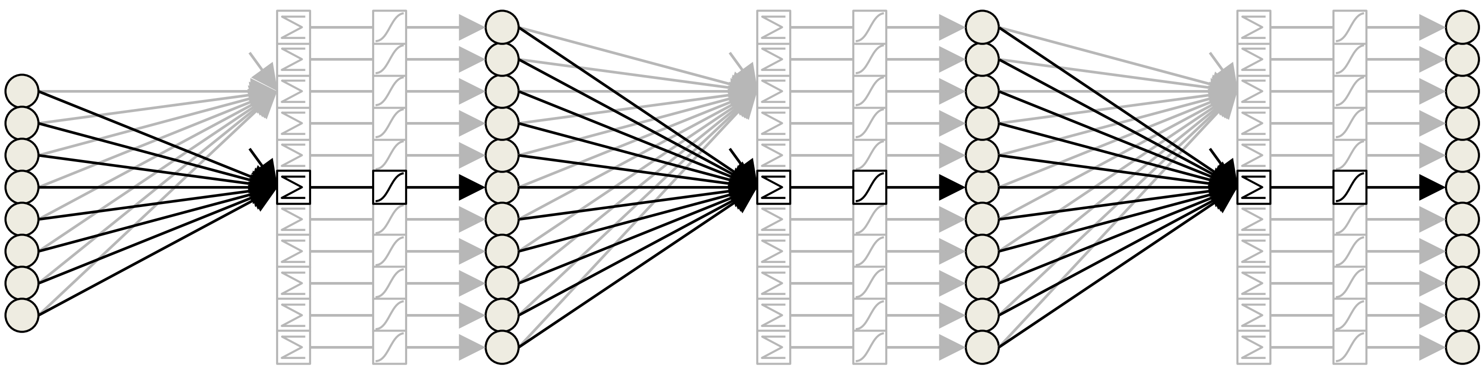
More hidden layers = deep

… MATHEMATICALLY:

mathematical expressions, such as:

$$\mathbf{y} = \sigma(\mathbf{b}_3 + \mathbf{W}_3^T\sigma(\mathbf{b}_2 + \mathbf{W}_2^T\sigma(\mathbf{b}_1 + \mathbf{W}_1^T\mathbf{x})))$$

expression can be visualized as a graph:



$\mathbf{x}$ $\quad$ $\mathbf{b}_1 + \mathbf{W}_1^T\mathbf{x}$ $\quad$ $\mathbf{h}_1$ $\quad$ $\mathbf{b}_2 + \mathbf{W}_2^T\mathbf{h}_1$ $\quad$ $\mathbf{h}_2$ $\quad$ $\mathbf{b}_3 + \mathbf{W}_3^T\mathbf{h}_2$ $\quad$ $\mathbf{y}$

that's why deep NNs need so much computational power!

MODERN DNN ARCHITECURES

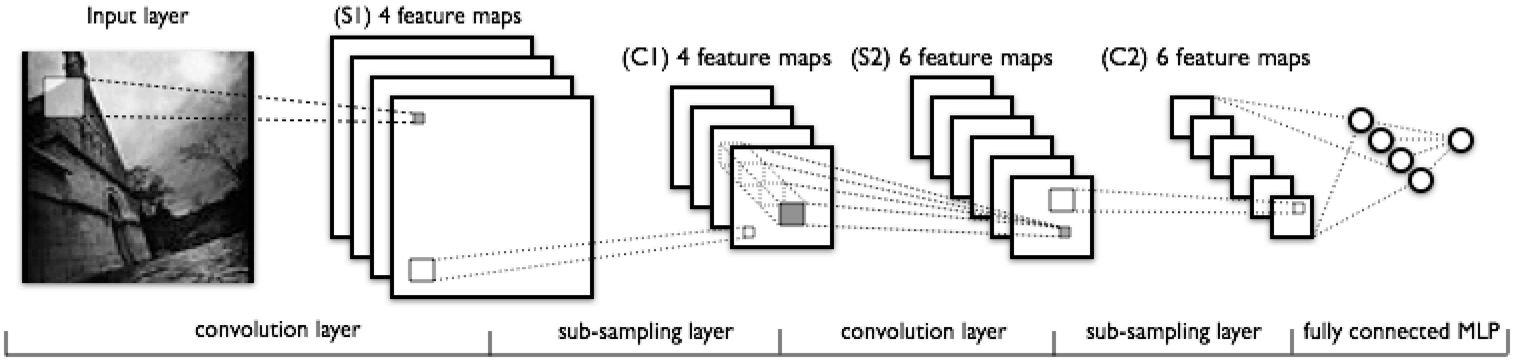
There are two prevalent DNN architectures which are mostly used today:

- **Convolutional Neural Networks** (CNN or ConvNets)
  for data where spatial vicinity is an important concept, e.g. in images, audio spectrograms, MRI (volume) data

- **Recurrent Neural Networks** (RNN, LSTM, GRU)
  for sequences of data, e.g. time series, text (translation, etc.)
  speech, audio, etc.

(or a combination of the two, e.g. for speech translation, video, image captioning ...)

CONVOLUTIONAL NEURAL NETWORK (CNN)



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

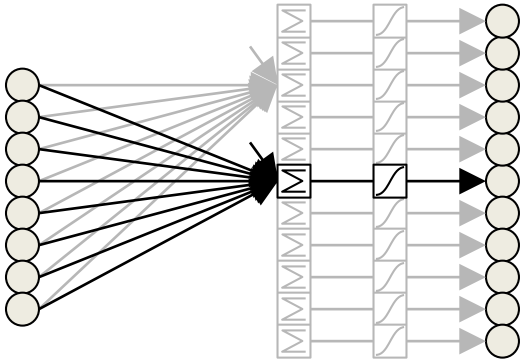
Combines three types of layers:

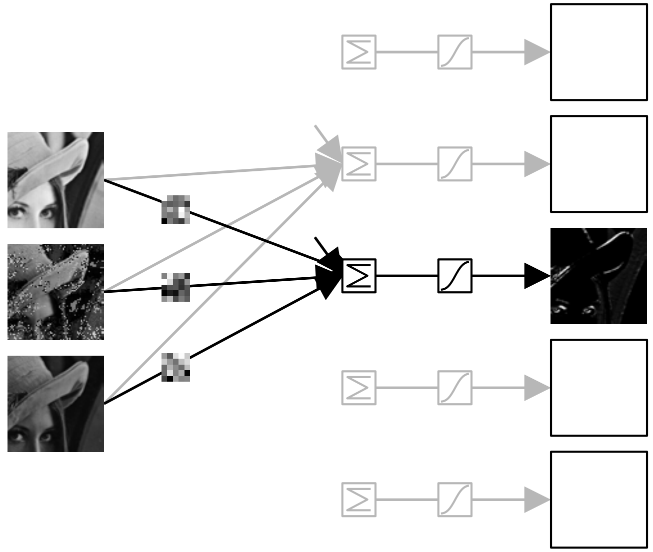
- **Convolutional layer:** performs 2D convolution of 2D input with multiple learned 2D kernels
- **Subsampling layer:** replaces 2D patches by their maximum ("max-pooling") or average
- **Fully-connected layer:** computes weighted sums of its input with multiple sets of learned coefficients

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

Fully-connected layer:
Each **input** is a **scalar** value,
each **weight** is a **scalar** value,
each output is the sum of inputs **multiplied** by weights.

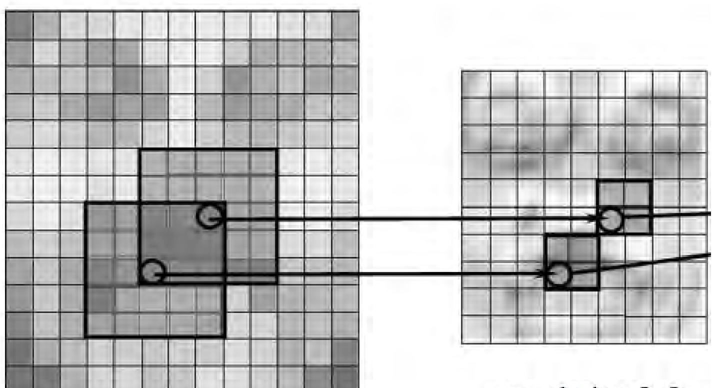
Convolutional layer:
Each **input** is a **tensor** (e.g., 2D),
each **weight** is a **tensor**,
each output is the sum of inputs **convolved** by weights.



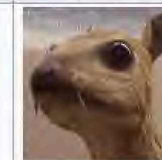
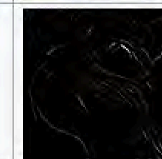
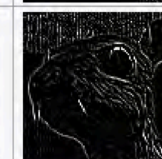
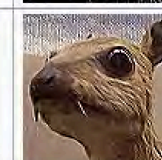
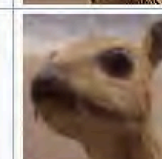


MOTIVATION FOR CONVOLUTIONS

- 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

POOLING STEP

Second very important aspect of a CNN:
(also called subsampling or downsampling)

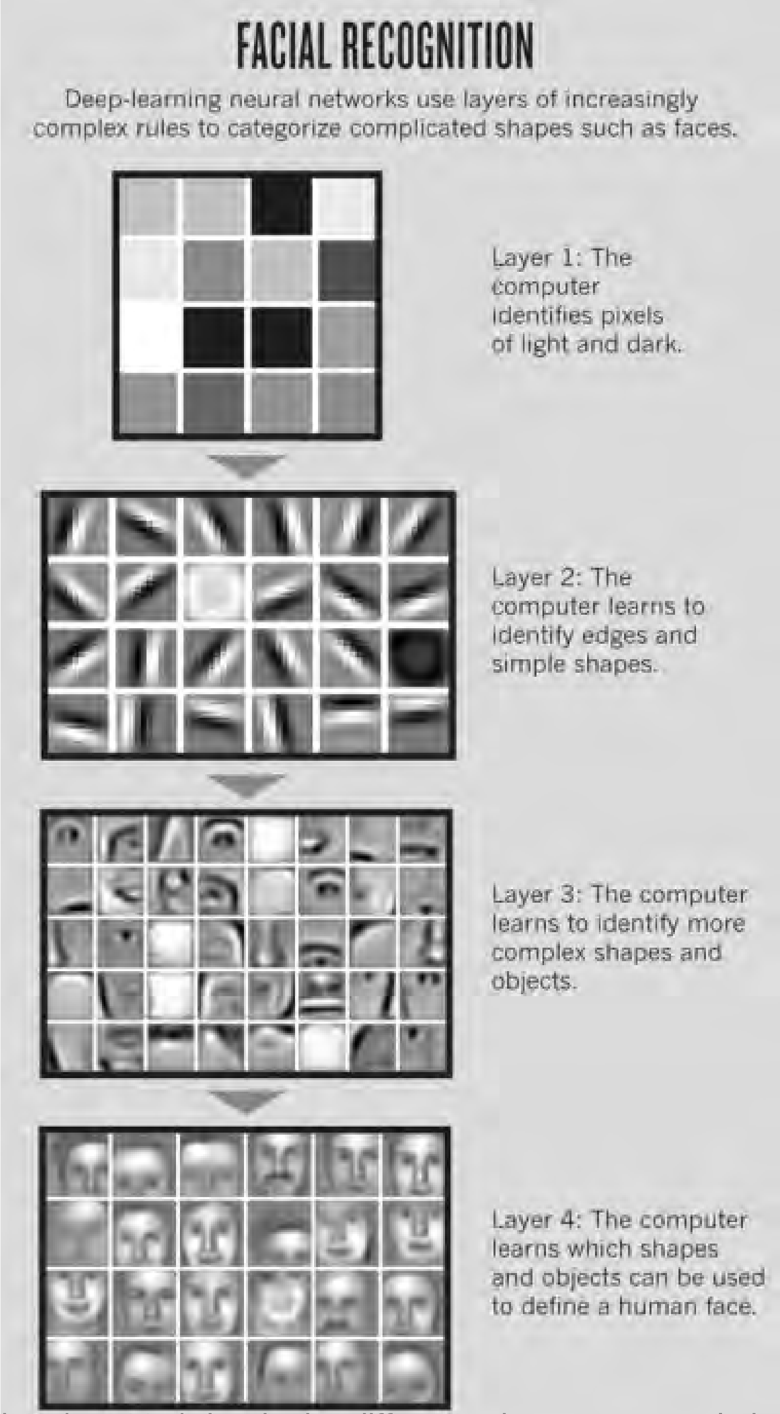
A **pooling layer** <u>reduces</u> the size of feature maps (i.e. output of a CNN layer and thus the input to the next layer)

1	**3**	**5**	-3
-2	**3**	2	-1
1	2	-2	**2**
-1	**3**	-2	1

→

3	5
3	2

Max pooling: take the max. activation across small regions (e.g. 2x2, as in the example above)



AUSTRIAN INSTITUTE OF TECHNOLOGY

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

FACE RECOGNITION WITH CNNS

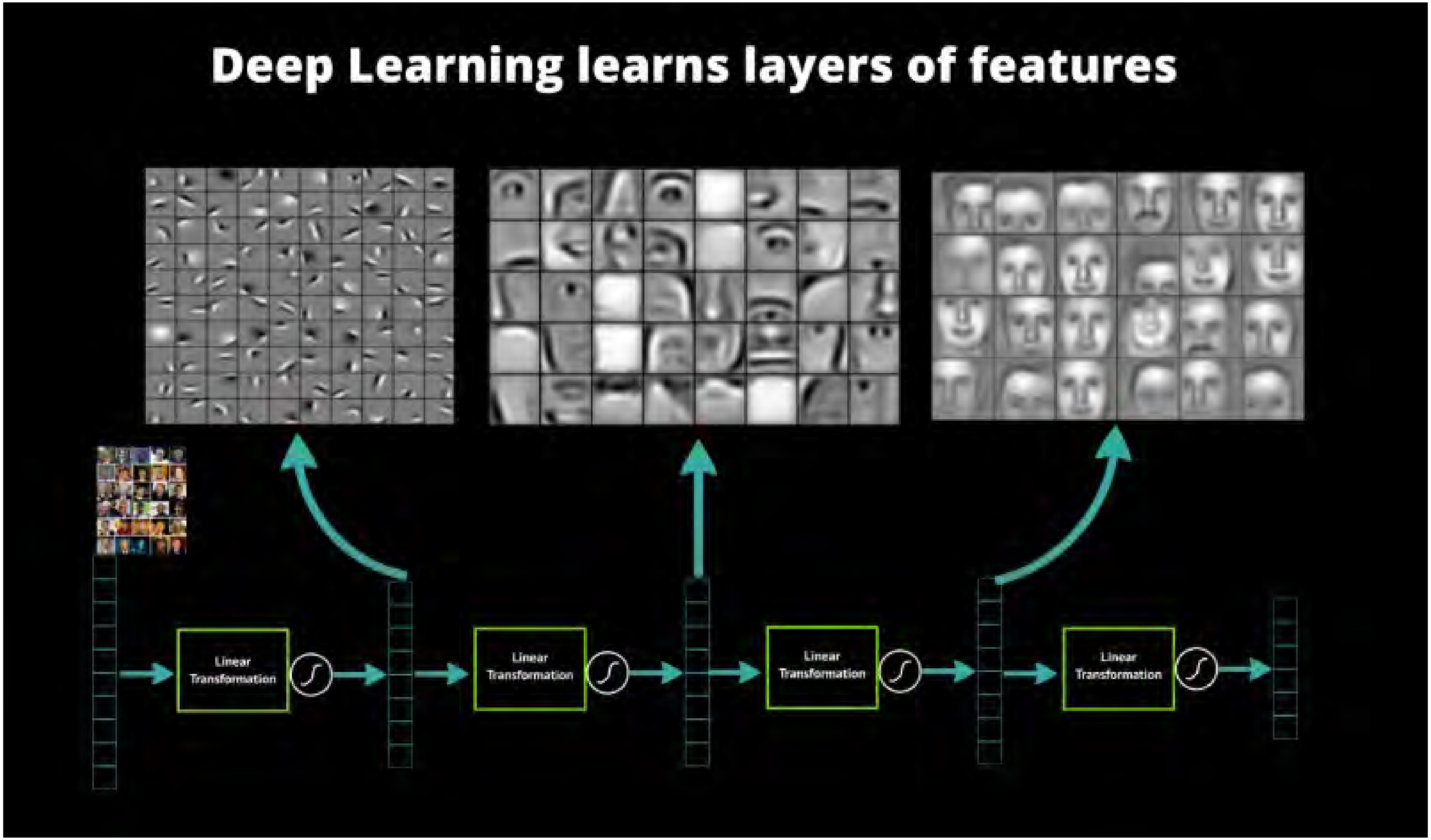
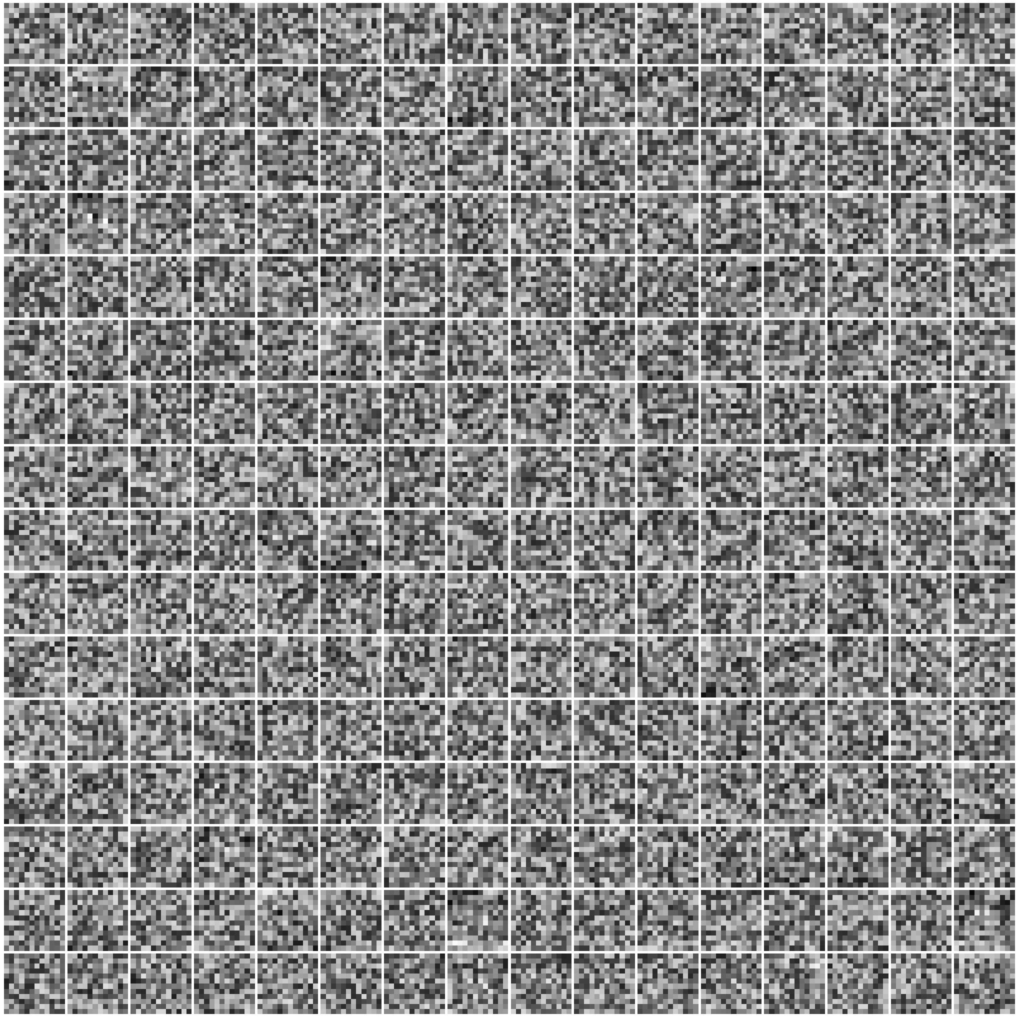


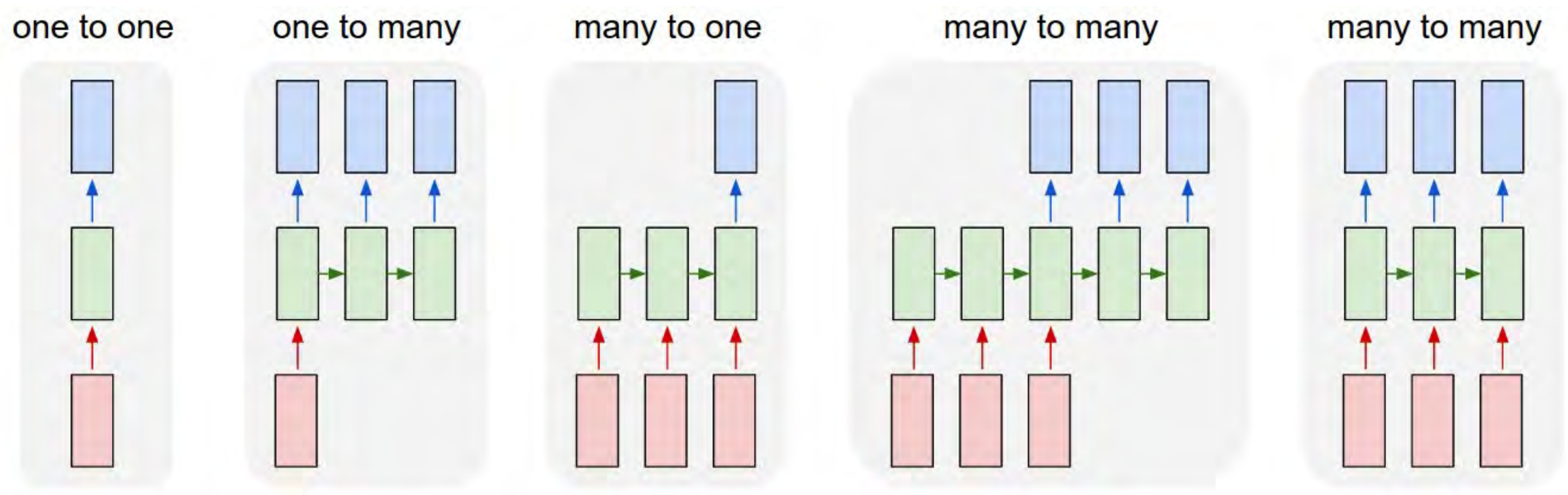
IMAGE PROCESSING



iteration no 0

RECURRENT NEURAL NETWORKS

Process **sequences** of data in various ways:



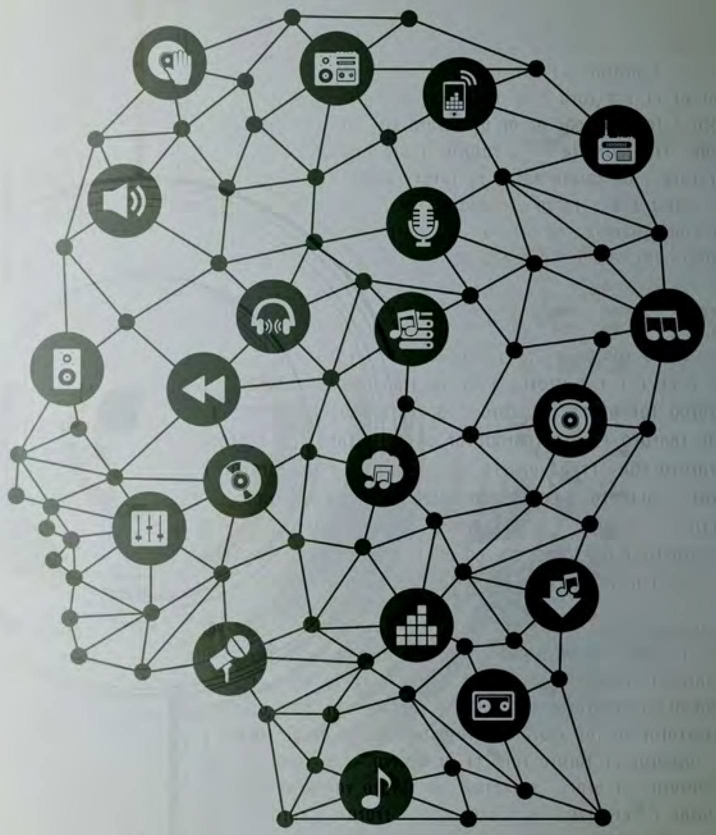
2) Image description (output: multiple words or sentences)
3) Sentiment analysis (input: text, output: mood category)
4) Translation (English -> French)
5) Synced input / output (e.g. frame-wise video categorization)

Image: http://karpathy.github.io/2015/05/21/rnn-effectiveness/

AUSTRIAN INSTITUTE OF TECHNOLOGY

Artificial Intelligence

& Security



FORENSIC INVESTIGATIONS OF TERRORISTIC ATTACKS

- **Context**
  - Forensic Investigation
  - Investigating video data after a terroristic attack

- **Objectives**
  - Spot suspects
  - Follow hints by civilian witnesses
  - Collect and secure evidence
  - Prevent immediate or subsequent attacks

FORENSIC INVESTIGATIONS
OF
TERRORISTIC ATTACKS

- **Obstacles**
  - Great increase in the number of public and private cameras
  - Massively increasing volume of video data to be analysed
    - Boston Marathon Bombing 5.000h
    - Toulouse and Montauban:10.000h (35TB)
  - Time pressure
    - Timely content evaluation of video mass data is of considerable importance

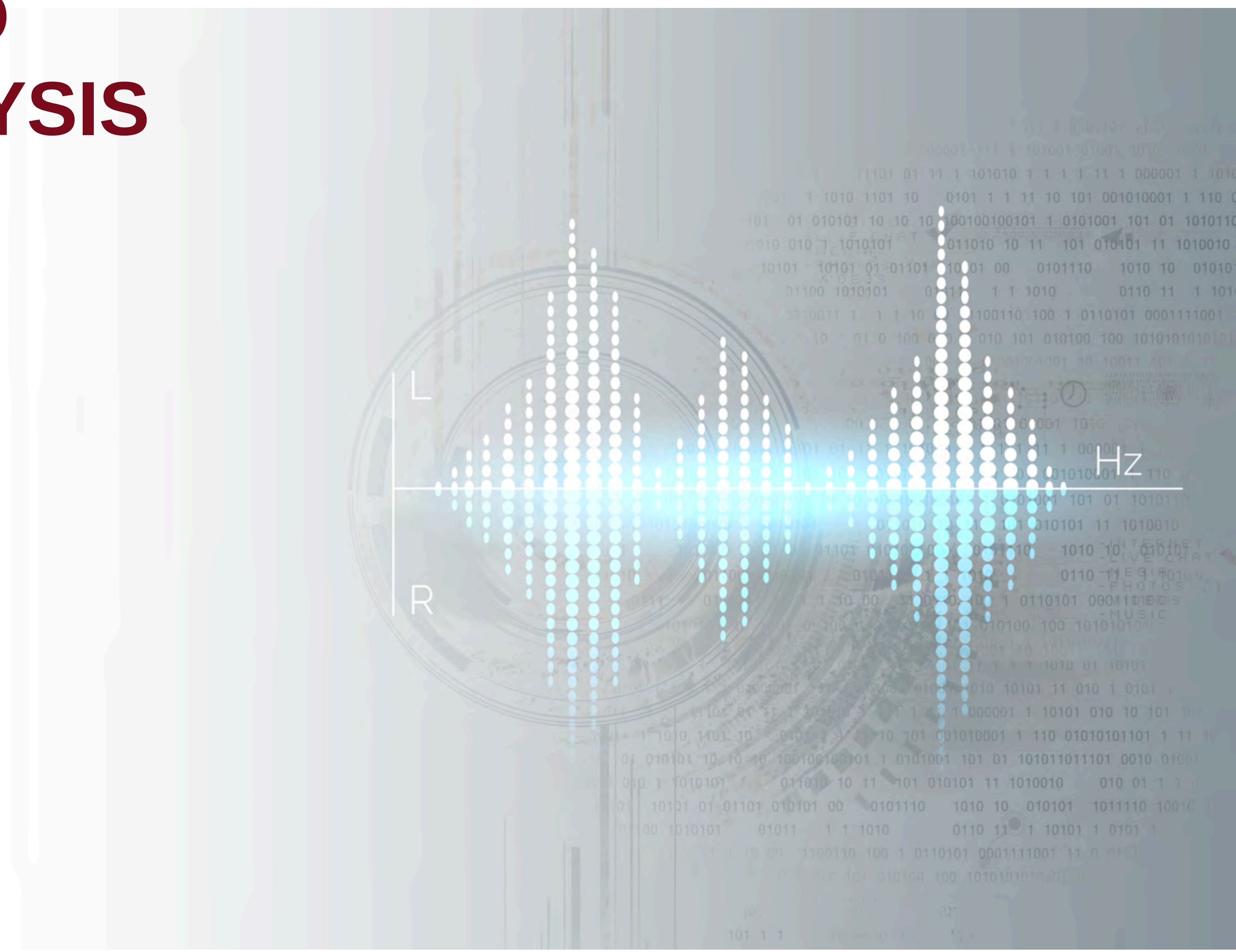
FORENSIC INVESTIGATIONS OF TERRORISTIC ATTACKS

- **Initial Situation (before project)**
  - manual viewing/processing of the video material
  - Personnel-intensive: time span from several hundred to several thousand hours

  - **Technical, supporting tools necessary**

- **Projects Goals and Outcomes**
  - 2 Projects
    - FLORIDA (Bi-Lateral funding Austria/Germany) => intial research
    - VICTORIA (H2020) => TRL 6 - 10
  - Large-scale computing platform
  - Analytical modules

AUDIO ANALYSIS



AUDIO EVENT DETECTION

Analytic Modules

Task

Detect and predict audio-events into predefined categories

- Gunshots, explosions, emergency vehicles, scream, speech, Alarm

Use-Case

Content filter in mass video-data

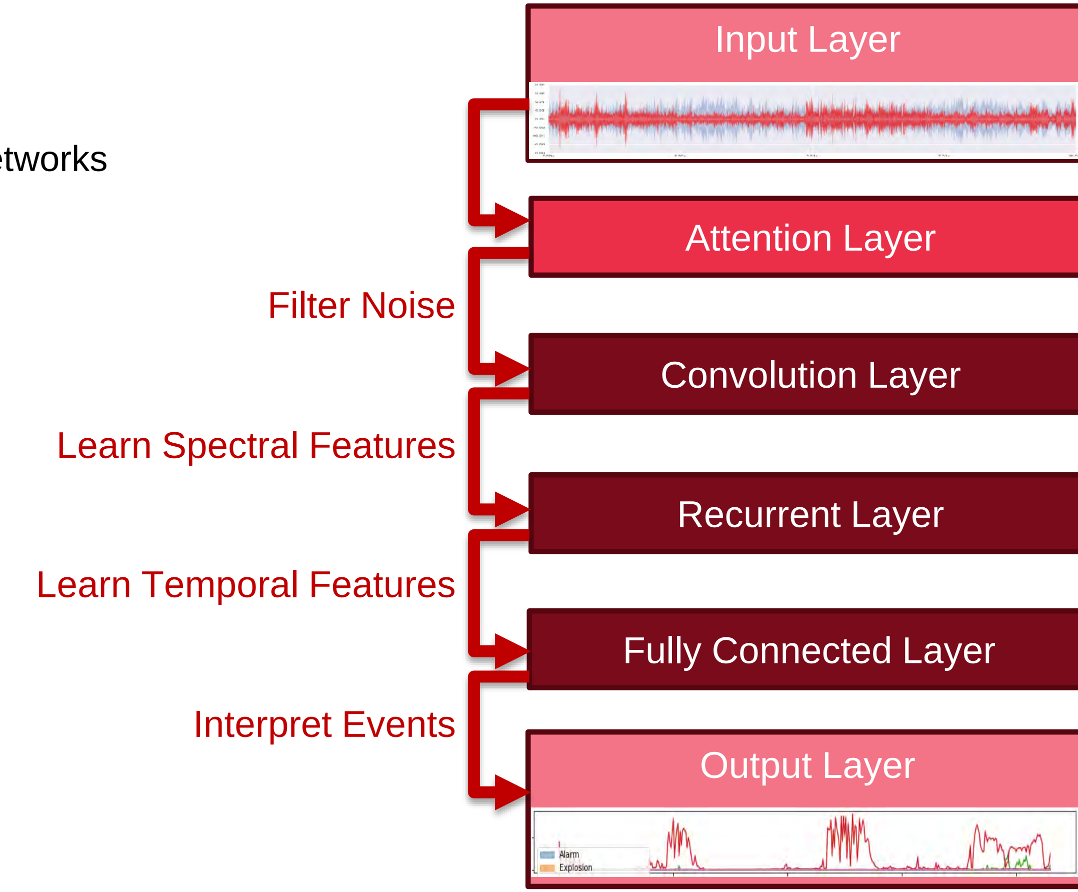
Example: attack with firearms

- => initiate search by filtering all videos which contain *Gunshots* (sorted by confidence)

AUDIO EVENT DETECTION

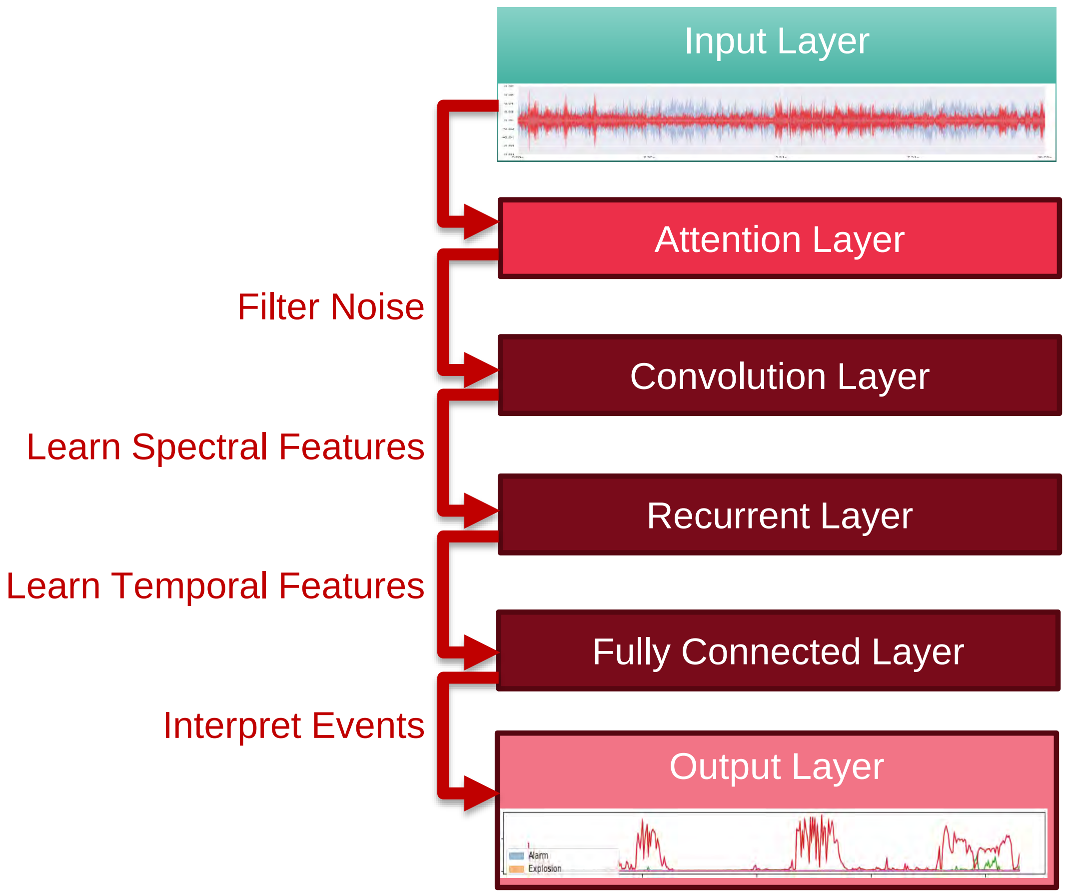
Applied Technology

Recurrent Convolutional Neual Networks
With Attention Layer



AUDIO EVENT DETECTION

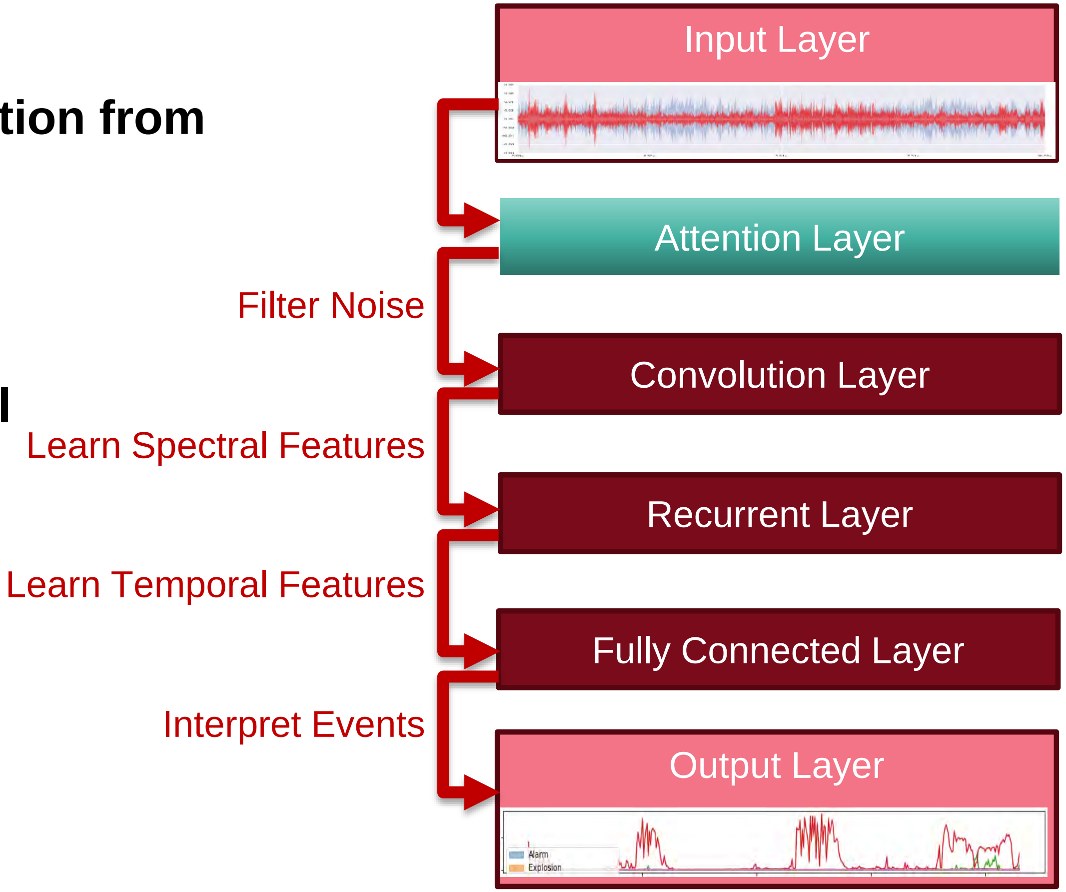
1. **Input representation**
   Common: Mel-Spectrograms



AUDIO EVENT DETECTION

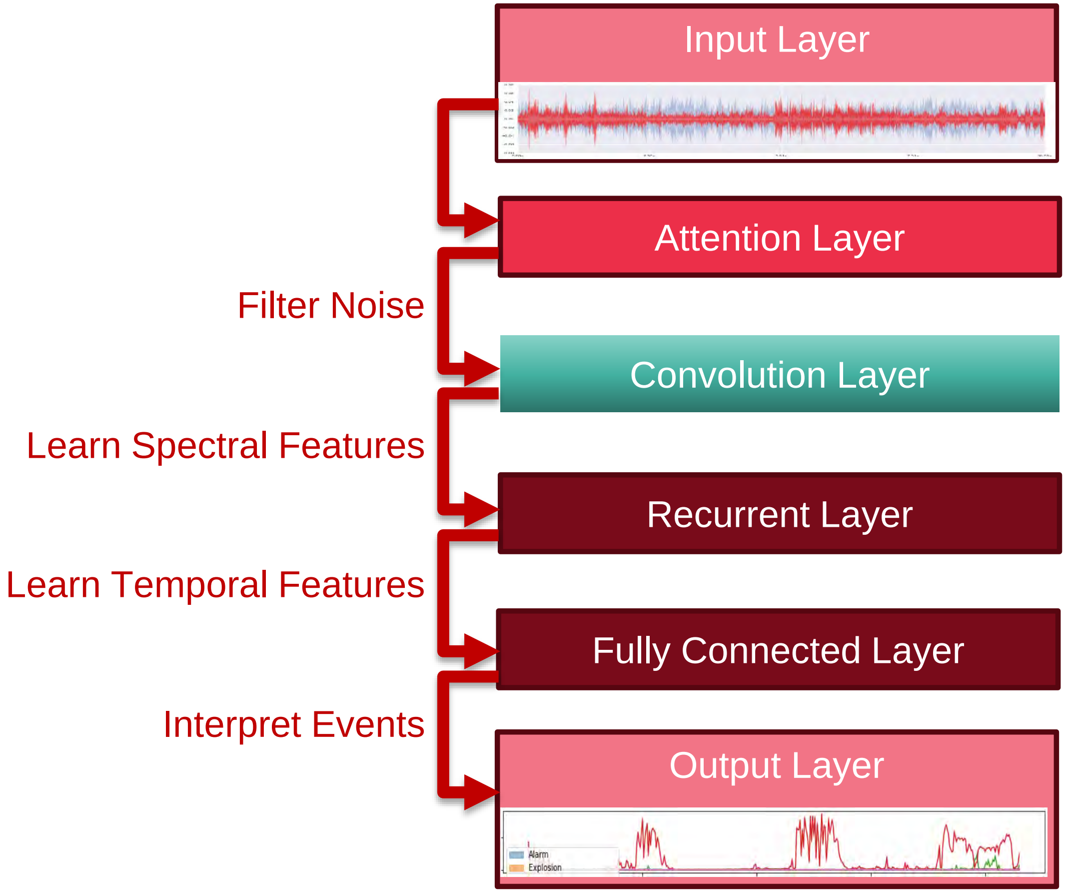
1. Input representation
   - Common: Mel-Spectrograms

- **Attention Layer**
  - **Filter non-relevant information from Input**
  - **Help to learn faster**
  - **Better convergence**
  - **Better generalization**
  - **Smoother prediction signal**



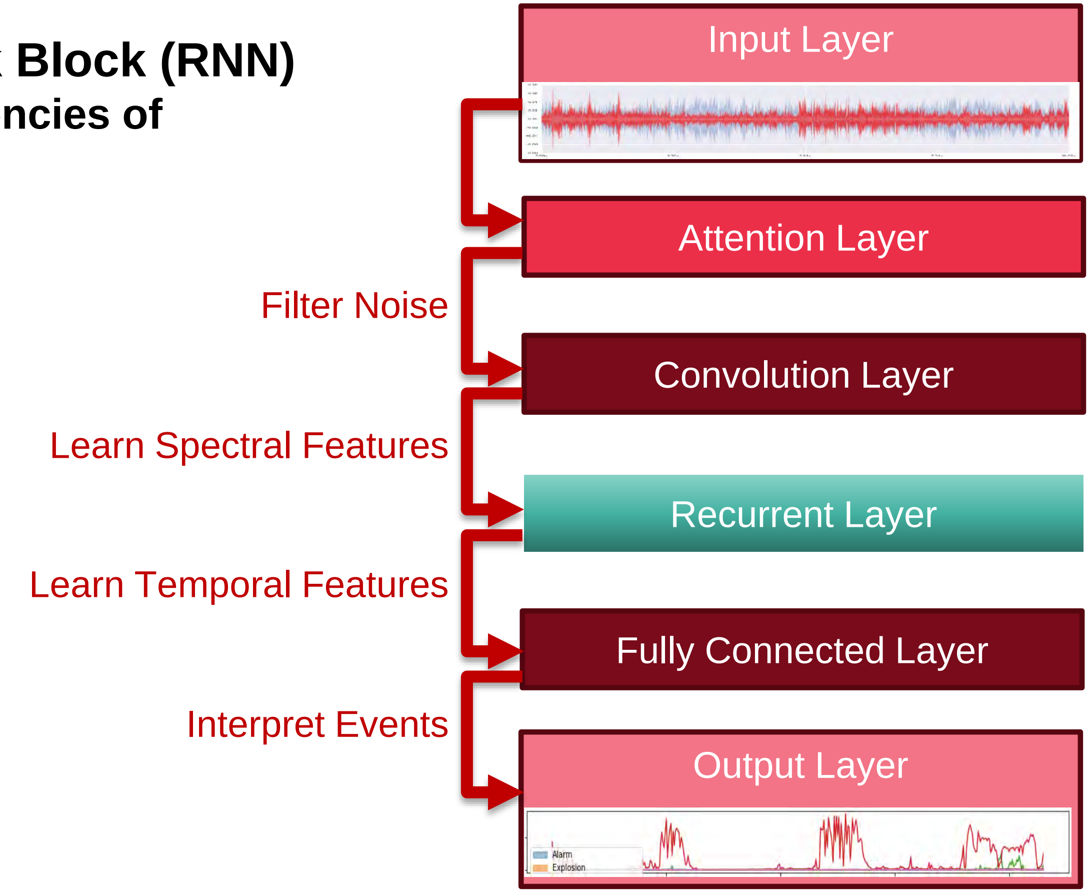
AUDIO EVENT DETECTION

1. Input representation
   - Common: Mel-Spectrograms
2. **Convolutional Neural Network Block (CNN)**
   - **Learn audio embeddings**



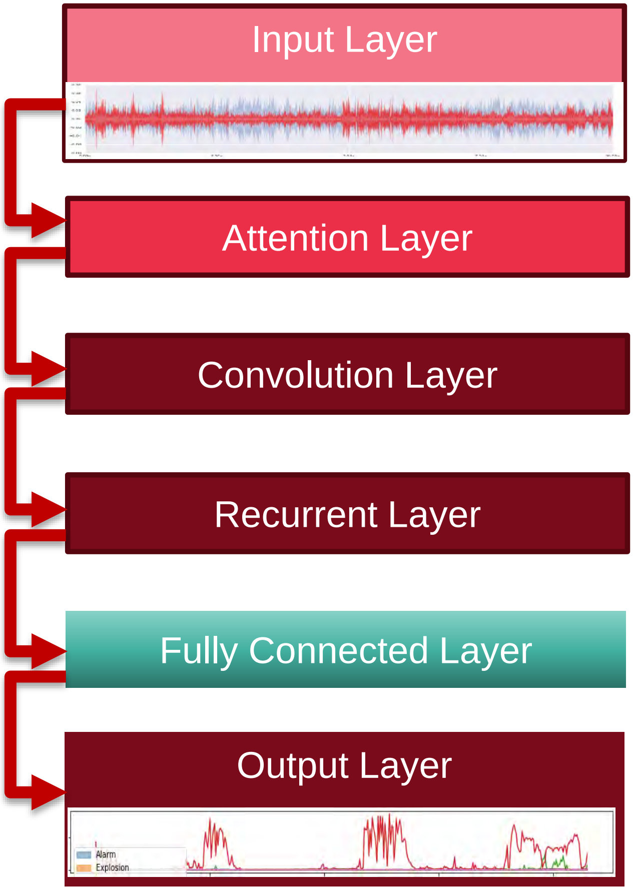
AUDIO EVENT DETECTION

1. Input representation
   - Common: Mel-Spectrograms
2. Convolutional Neural Network Block (CNN)
   - Learn audio embeddings
3. **Recurrent Neural Network Block (RNN)**
   - **Learn Temporal dependencies of embeddings**



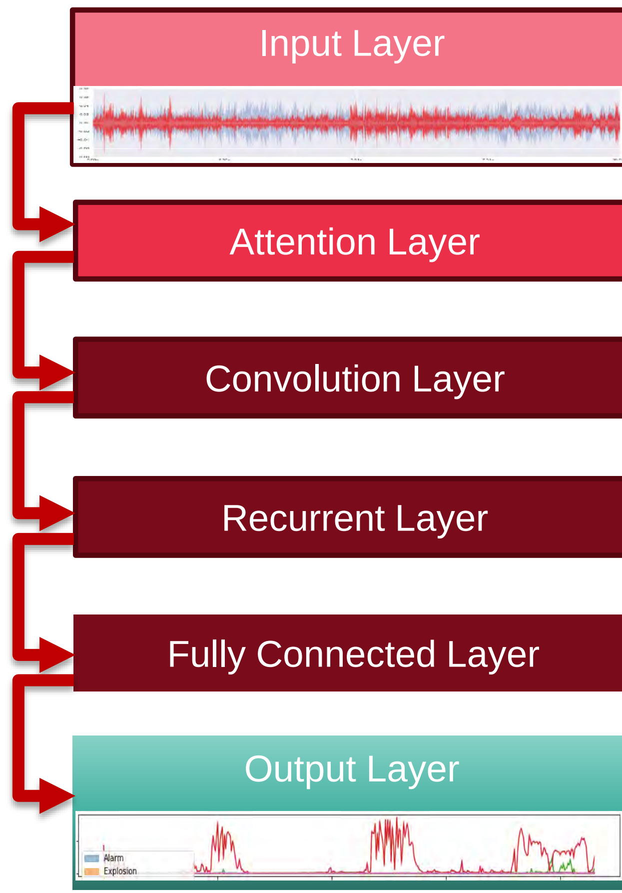
AUDIO EVENT DETECTION

1. Input representation
   Common: Mel-Spectrograms
2. Convolutional Neural Network Block (CNN)
   Learn audio embeddings
3. Recurrent Neural Network Block (RNN)
   Learn Temporal dependencies of embeddings
4. **Array of Fully Connected Layers**
   One Layer per temporal dimension (Time-Distributed)
   Dimensionality of Layer = Number of classes



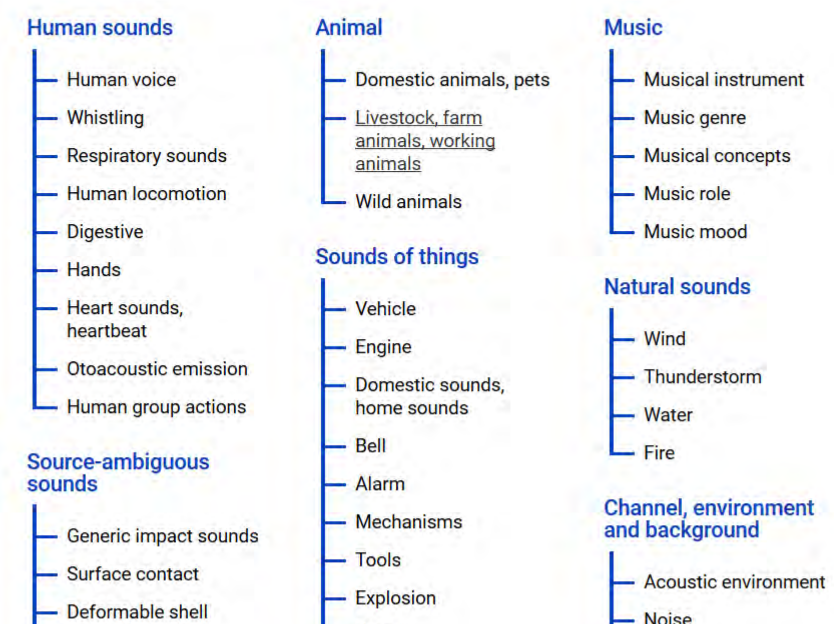
AUDIO EVENT DETECTION

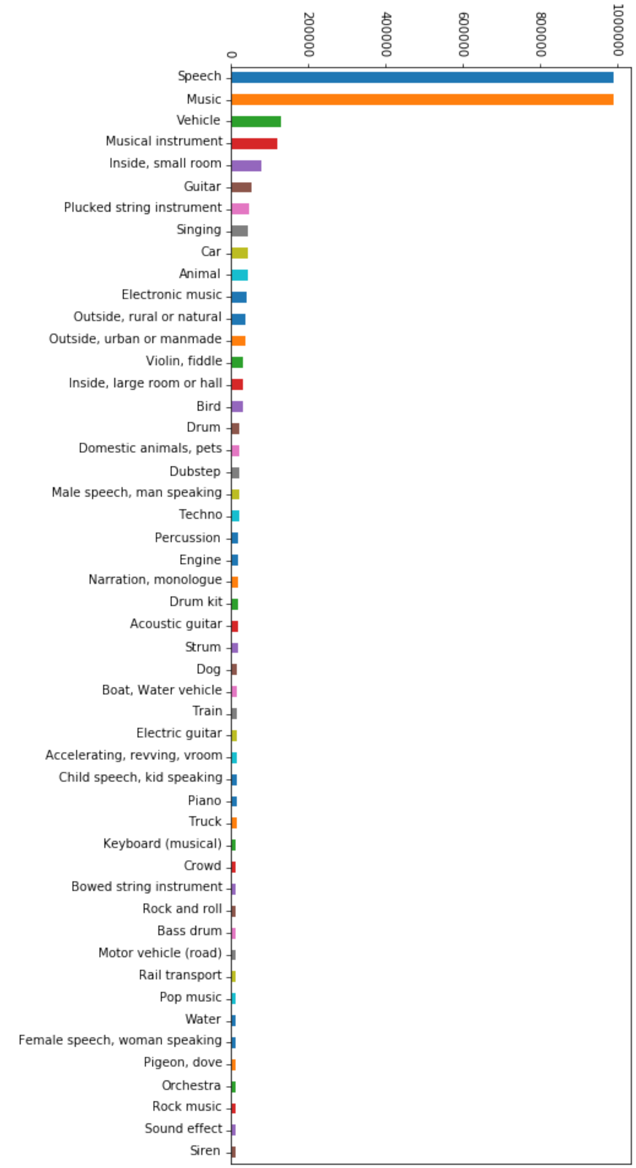
1. Input representation
   Common: Mel-Spectrograms
2. Convolutional Neural Network Block (CNN)
   Learn audio embeddings
3. Recurrent Neural Network Block (RNN)
   Learn Temporal dependencies of embeddings
4. Array of Fully Connected Layers
   One Layer per temporal dimension (Time-Distributed)
   Dimensionality of Layer = Number of classes
5. **Outputs**
   Strong Labels – Training & Inference
      Output of Time-Distributed Fully Connected Layers
   Weak Labels - Training
      Output Layer aggregation (e.g. avg, max)
      Multi label prediction



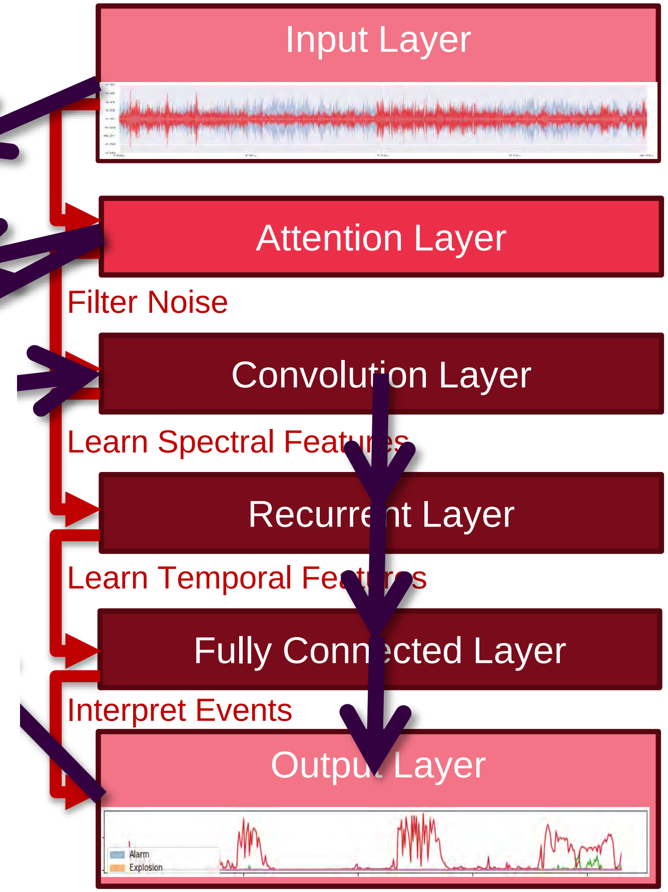
Google Audio Set

- 2M Videos
- 632 audio events
- annotaded according acoustic categories
- Weakly labelled (10s)
- Currently largest source of data





RECURRENT CONVOLUTIONAL NEURAL NETWORKS



High signal to noise ratio
Due to attention layer

Smoothing Function
Temporal Segmentation per Event Category

AUDIO EVENTS VISUALIZED

Us

AUSTRIAN INSTITUTE OF TECHNOLOGY

AUDIO SIMILARITY

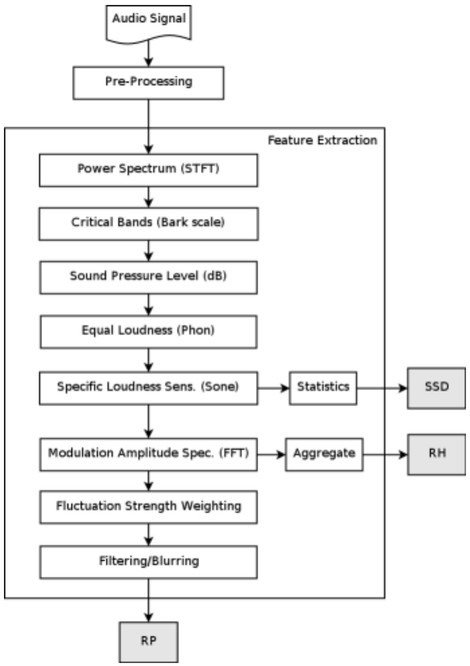
Analytic Modules

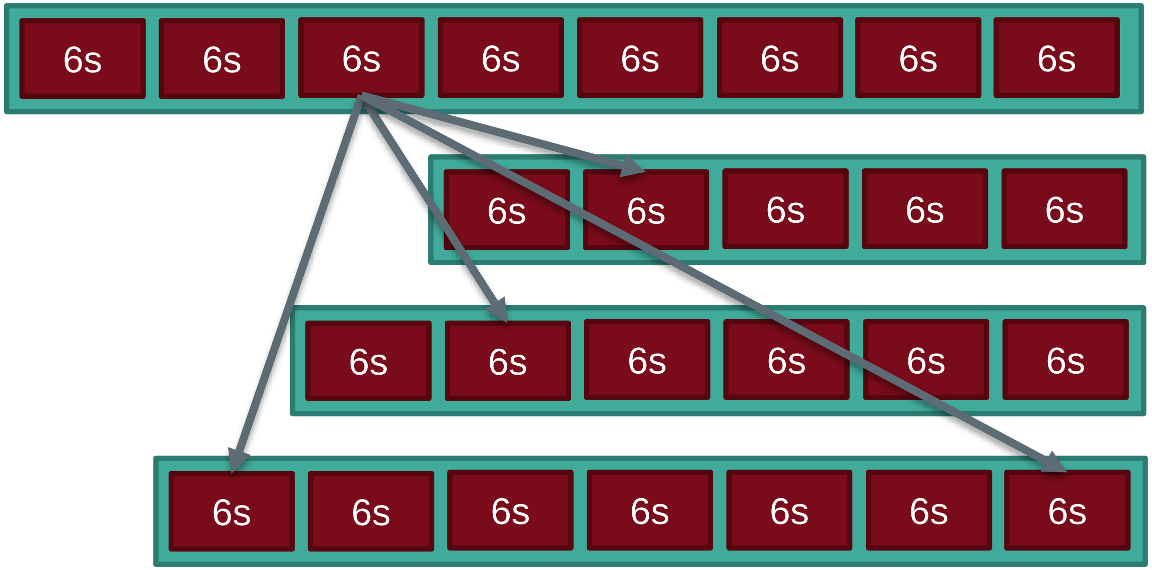
AUDIO SIMILARITY SEARCH

- **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





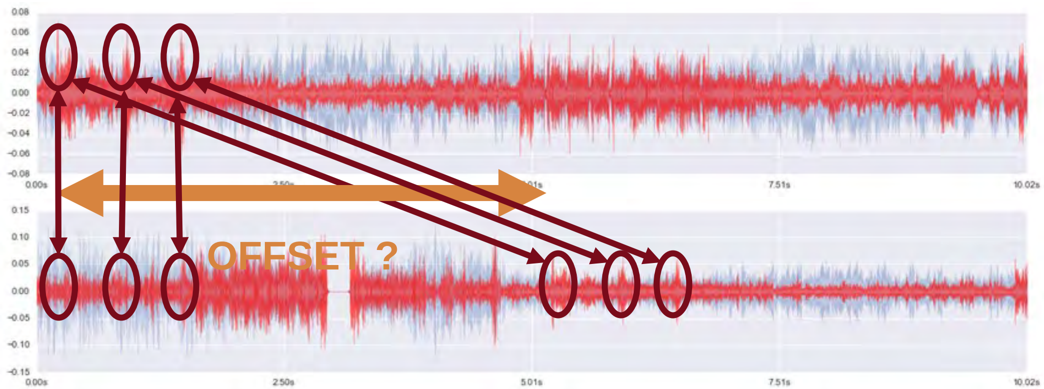
AUDIO SIMILARITY – INSTANT LOCALIZATION (WITHOUT GPS)

AUDIO-BASED VIDEO-SYNCHRONIZATION

Analytic Module

AUDIO-BASED VIDEO-SYNCHRONIZATION

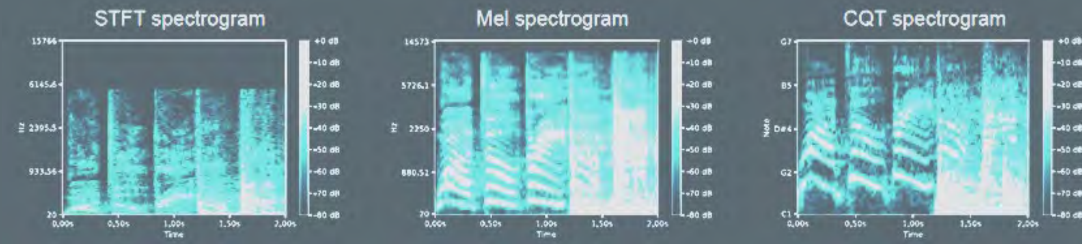
- **Task**
  - Synchronize various video files with unreliable time metadata
  - Use audio-signature to relatively align video files
- **Technology**
  - Audio-fingerprints (chromaprint)
  - Noise invariant



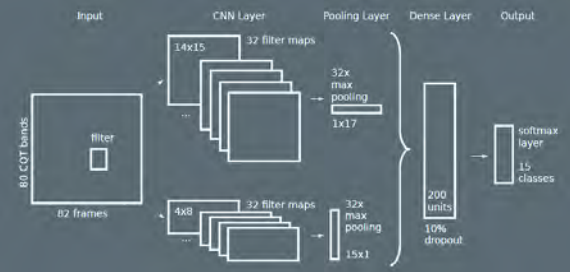
SCENE ANALYSIS

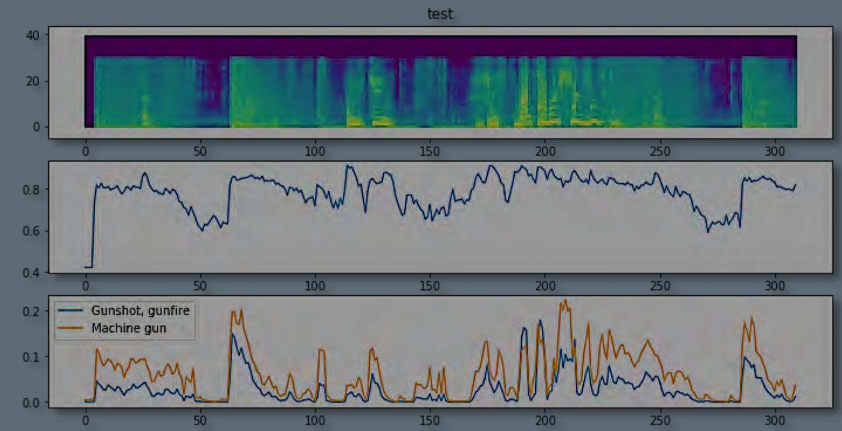
Audio-Visual Scene Understanding

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





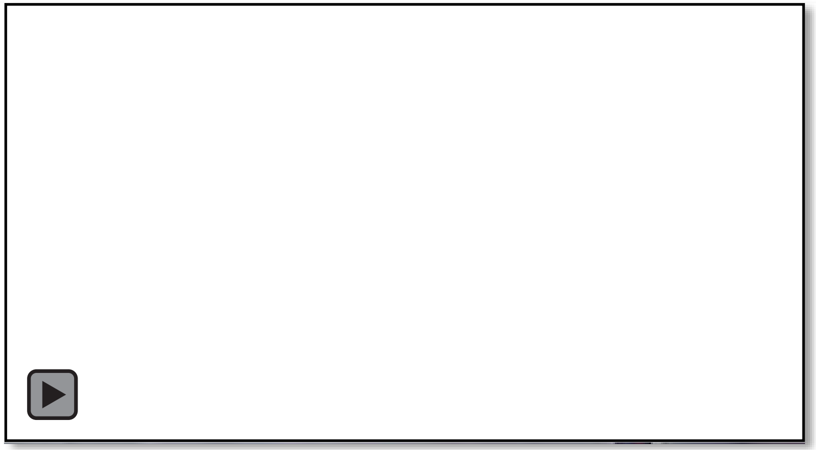
AUSTRIAN INSTITUTE OF TECHNOLOGY

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

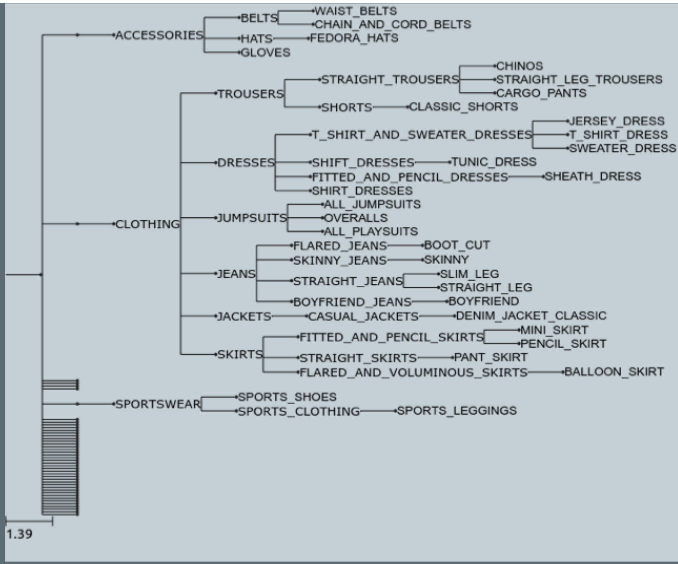
FASHION IMAGE CLASSIFICATION

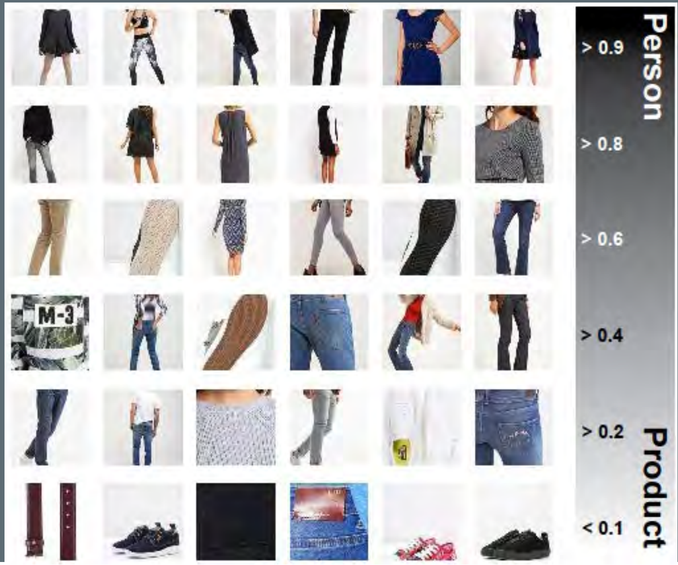
AUSTRIAN INSTITUTE OF TECHNOLOGY

Image Classification with Deep Neural Networks

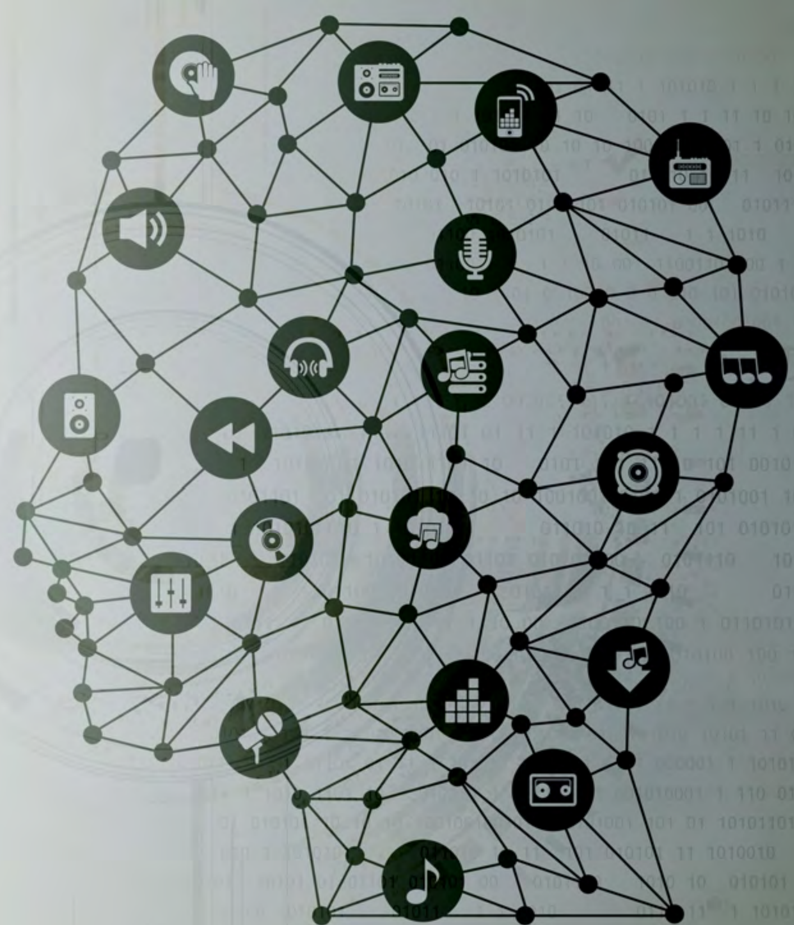
Large Scale Deep Learning
Visual Computing

- **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





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

AUSTRIAN INSTITUTE OF TECHNOLOGY

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

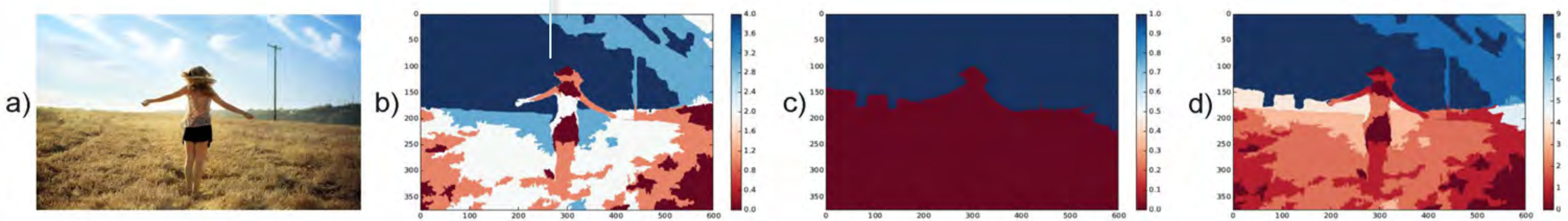
Cold / Warm
Light / Dark

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

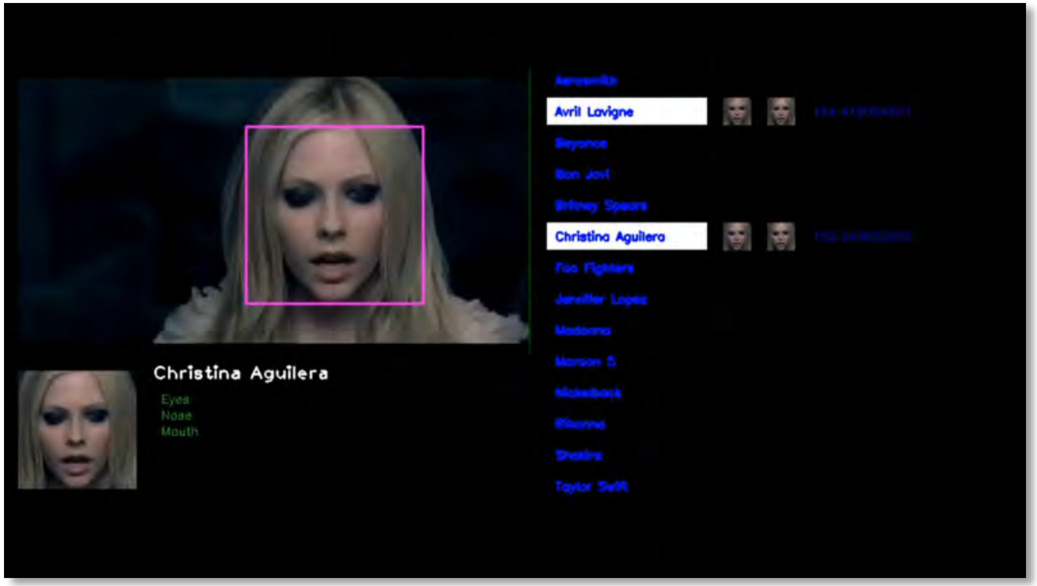
Audio-Visual Signer Identification

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



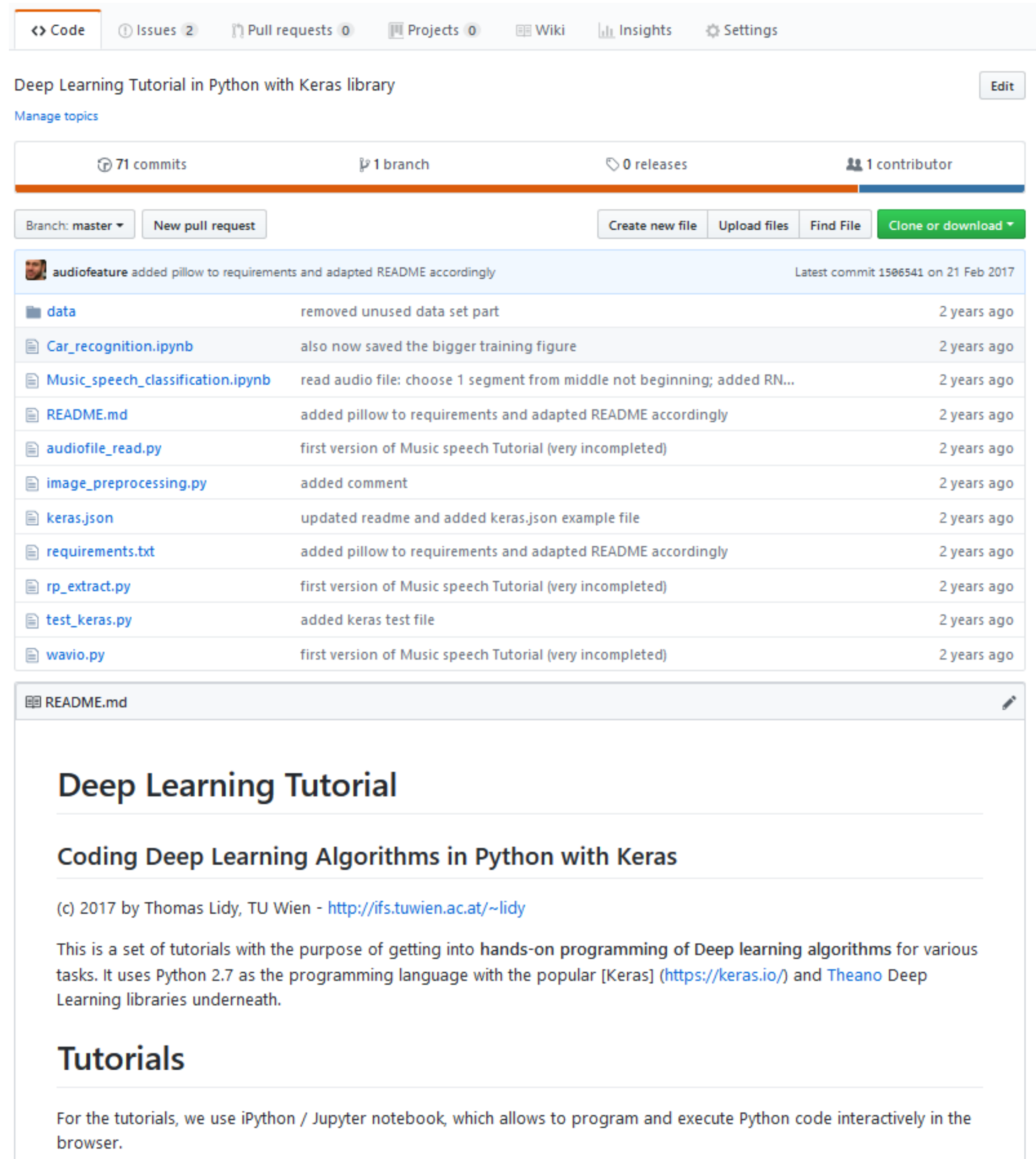
FURTHER RESOURCES

Code & Tutorials

DEEP LEARNING TUTORIAL

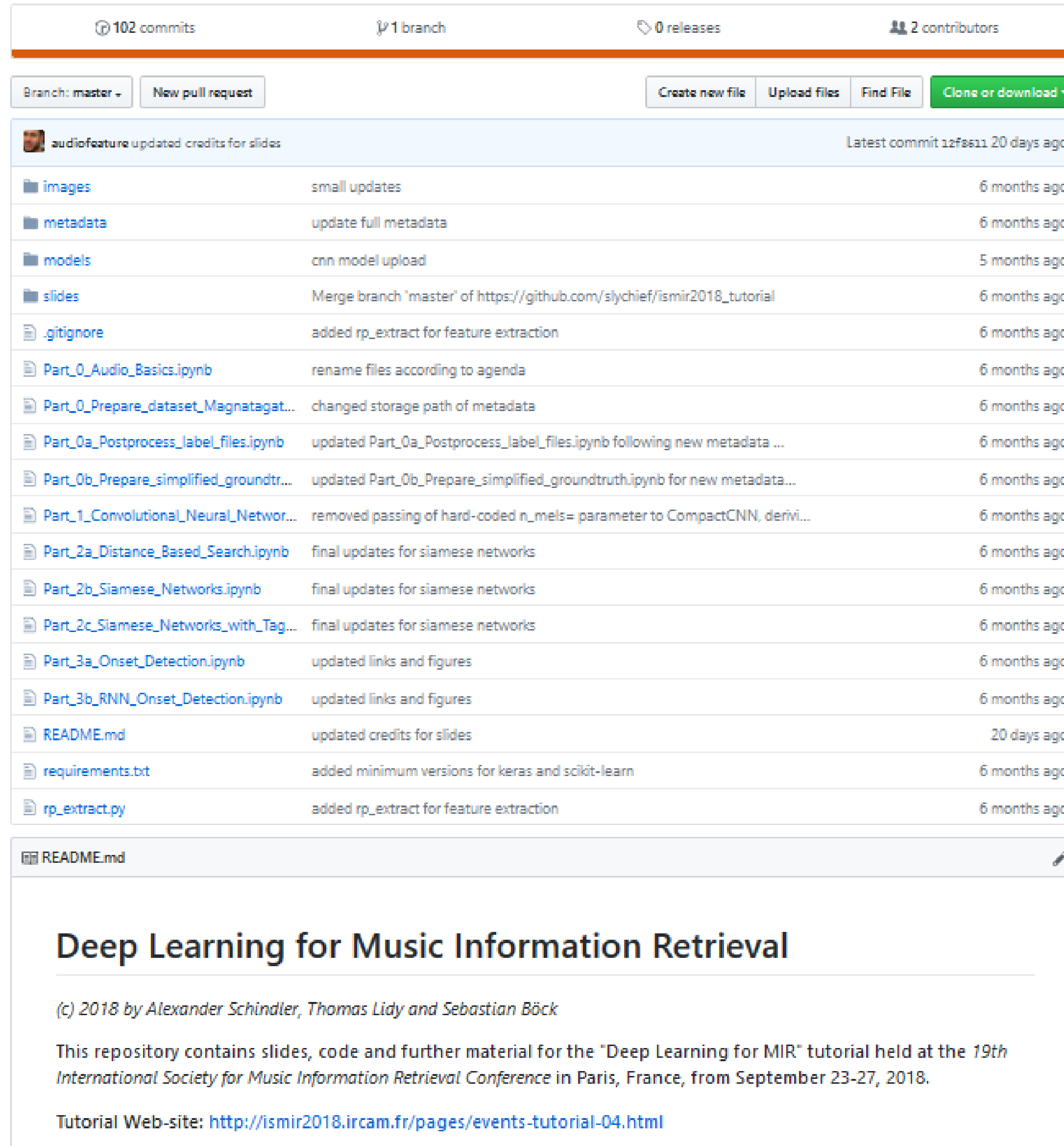
- Beginners Tutorial
  - Simple Image Processing
  - Simple Audio Processing
  - Python
  - Keras
  - Tensorflow

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



DEEP LEARNING TUTORIAL FOR MIR

- Advanced Tutorial
  - Extensive Tutorial
  - Audio Analysis
    - Classification
    - Similarity
    - Onsets Detection
  - Advanced Concepts
    - CNNs
    - LSTMs
    - Siamese Networks
  - Code & Data provided
  - Presented @ main MIR Conference ISMIR 2018



https://github.com/slychief/ismir2018_tutorial

AUSTRIAN INSTITUTE OF TECHNOLOGY

Vienna Deep Learning Meetup

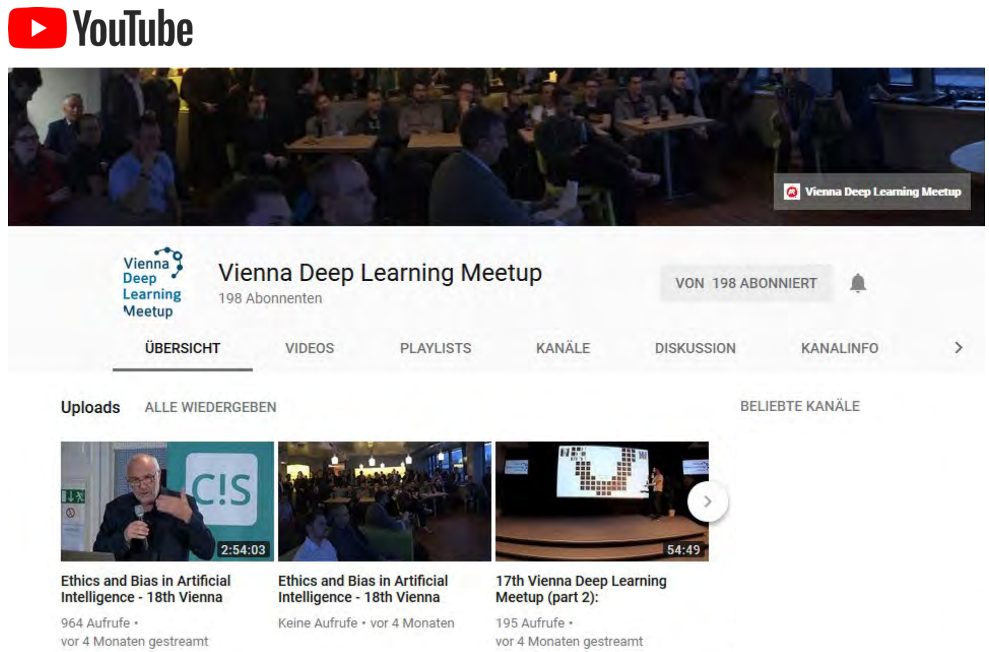
Vienna's largest monthly event on Deep Learning & AI

The Organizers:

			
Thomas Lidy Musimap	Alex Schindler AIT & TU Wien	René Donner contextflow	Jan Schlüter OFAI & UTLN

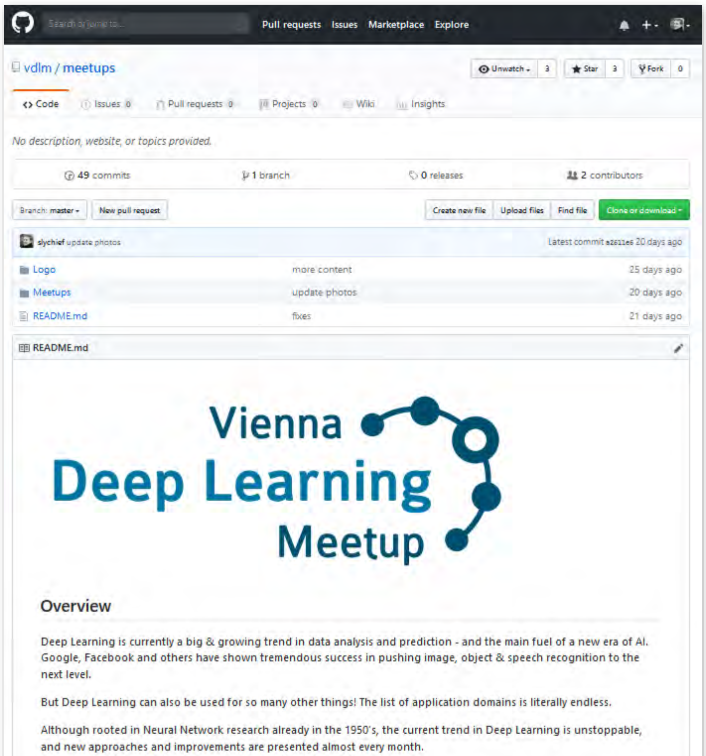
www.meetup.com/Vienna-Deep-Learning-Meetup

VDLM YOUTUBE CHANNEL



www.youtube.com/ViennaDeepLearningMeetup

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

THANK YOU!

QUESTIONS?