

# ARTIFICIAL INTELLIGENCE

Alexander Schindler  
Scientist  
Information Management  
Center For Digital Safety & Security

AIT Austrian Institute Of Technology GmbH  
[Alexander.schindler@ait.ac.at](mailto:Alexander.schindler@ait.ac.at)



# 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

# FRAGE

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

## BEST AI IN TV AND MOVIES



# BEST AI IN TV AND MOVIES

- **Self-Driving**
- **Listens, understands & answers**
  - Including sarcasm / is sarcastic
  - Sense of humor / tells jokes
- **Sees, hears, „feels“**
- **Plans ahead**
  - Roads, actions, complex, interactions
- **Has human emotions**
  - Is offended, easily bruised, jealous
  - Interprets human interaction
- **Feels social responsibility**
  - start a union for cars
- **Self-awareness**
- **Common knowledge**
- **Recommends**
  - Music, books, social events, ...



# K.I.T.T. ARTIFICIAL INTELLIGENCE IN 2019

- **Self-Driving Cars**
- **Conversational Systems (Chat-Bots)**
  - Listens, understands & answers
    - Including sarcasm / is sarcastic
    - Sense of humor / tells jokes
- **Content Recommendation**
  - Recommends
    - Music, books, social events, ...
- **Sees, hears, „feels“**
- **Common knowledge**
- **General Artificial Intelligence**
  - **Has human emotions**
    - Is offended, easily bruised, jealous
    - Interprets human interaction
  - **Feels social responsibility**
    - start a union for cars
  - **Self-awareness**



# 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)**

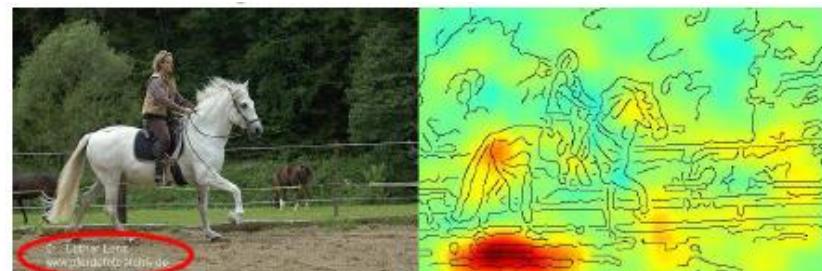
- Horse could solve arithmetic tasks
- In controlled isolated experiment
  - Horse responded to involuntary clues of the trainer and spectators
  - Observer-expectancy effect

- **A System is a 'Horse' (B. Sturm)**

- Focus on production artifacts instead of real semantic 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

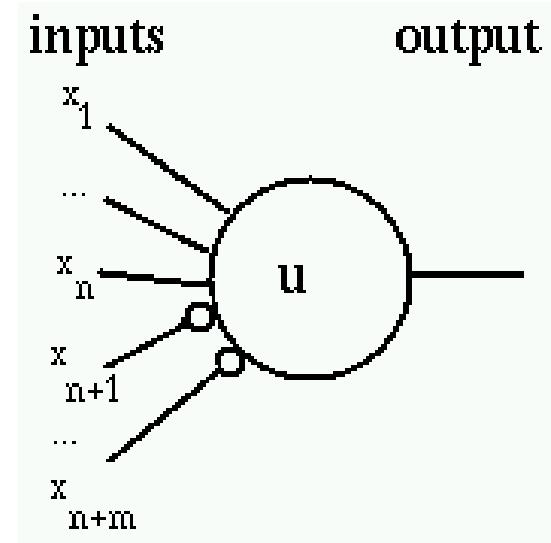
## of Artificial Intelligence

## FRAGE

- Wie alt ist „Künstliche Intelligenz“?
  
- 5 Jahre?
- 10 Jahre?
- 20 Jahre?
- 50 Jahre?
- 80 Jahre?
- 100 Jahre?

# 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)*



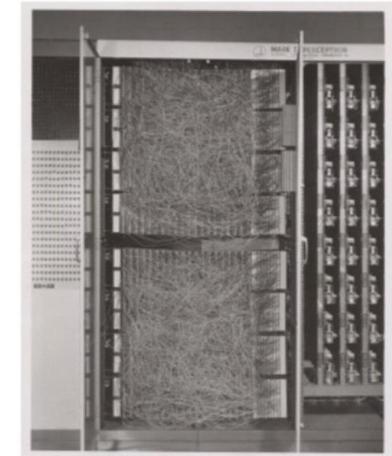
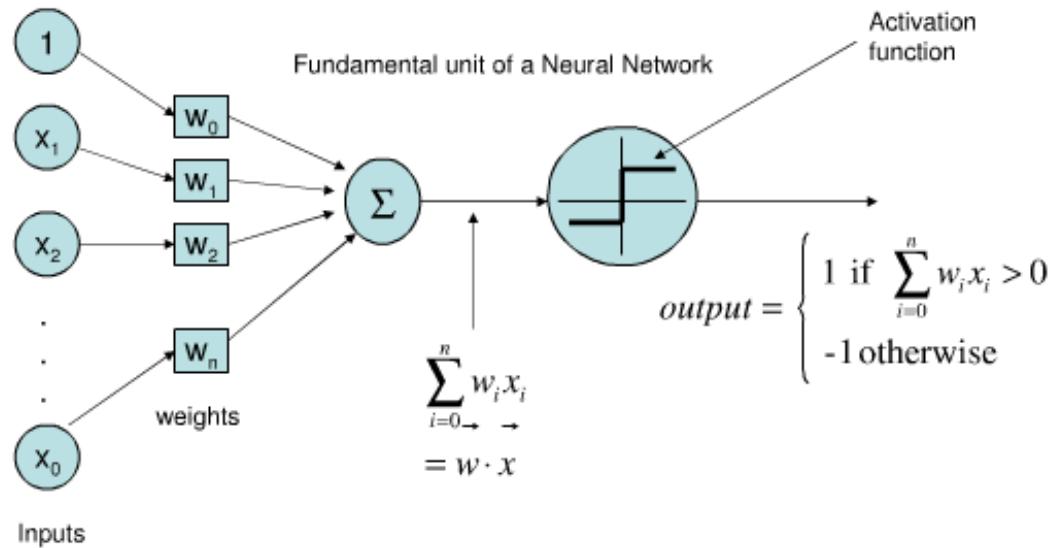
## 1956: The Dartmouth Conference

- **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**



# A BRIEF HISTORY OF AI

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**

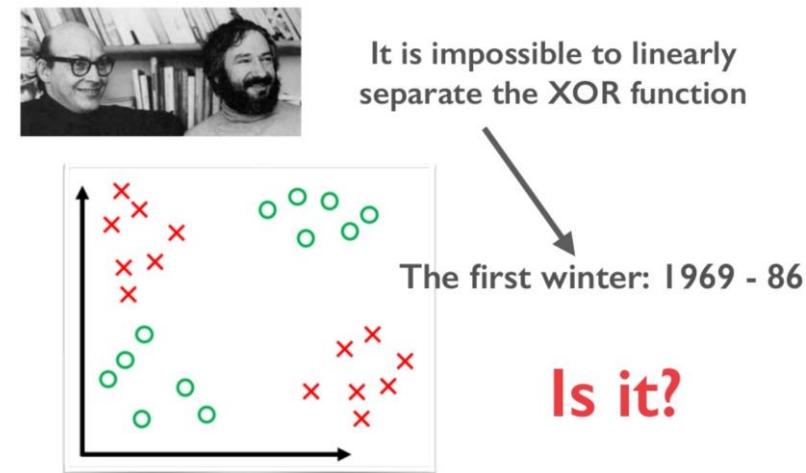
## UPS & DOWNS OF NEURAL NETWORKS

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



## 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

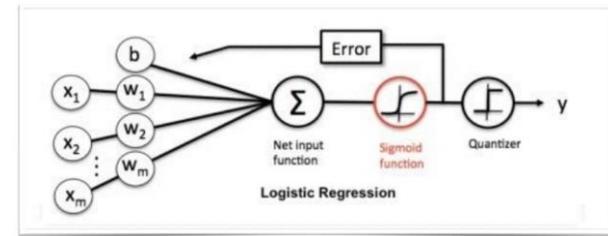
- Birth of expert systems
- Idea is to give AI systems lots of information to start with
  - Rule-based Systems
  - Fuzzy Logic

# 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

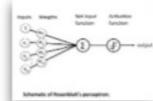
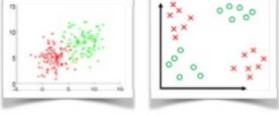
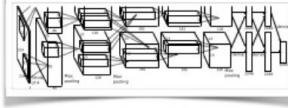


Wikipedia



Wikipedia

# THREE EPOCHS OF NEURAL NETWORKS

	techniques / tricks	hardware	data
1957-69 dawn	perceptron	early mainframes	toy linear, small images, XOR
			
1986-95 golden age	early NNs	workstations	MNIST
			
2006- deep learning	deep NNs	GPU, TPU, Intel Xeon Phi	Imagenet
		  	

# 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

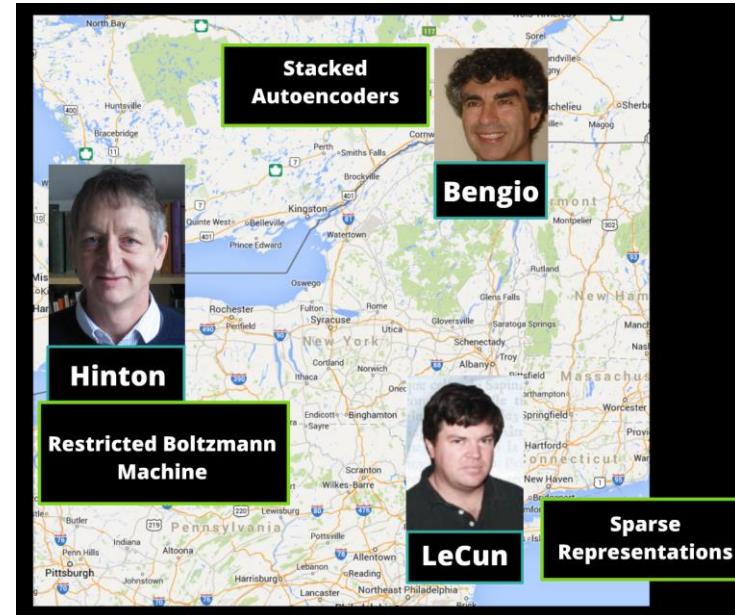
They overcome some problems that caused deep networks not to learn anything at all

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

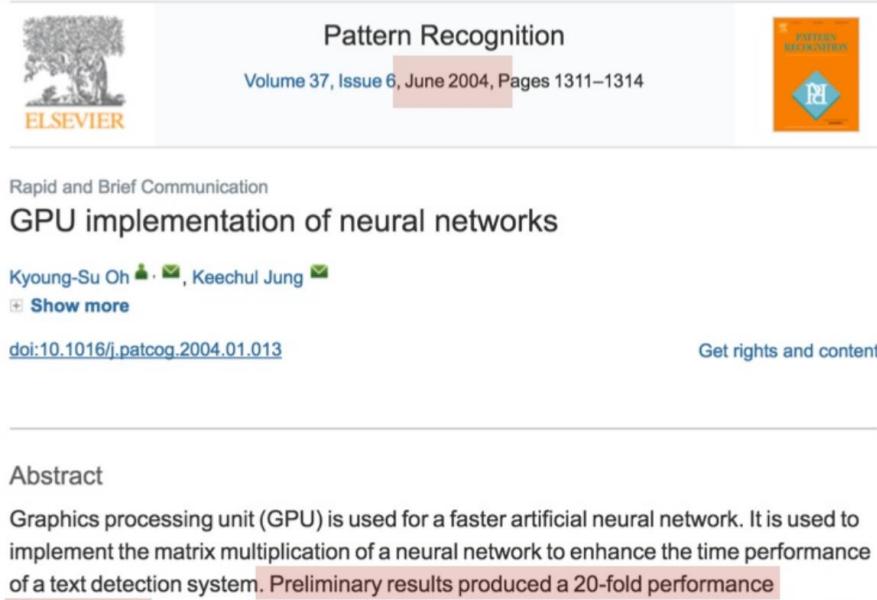
2010s: Important new contributions:

- Simpler initialization (without pre-training)
- Dropout
- Simpler activations: Rectifier Units (ReLUs)
- Batch Normalization

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



# GPUs (2004)



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

ELSEVIER

Rapid and Brief Communication  
**GPU implementation of neural networks**

Kyoung-Su Oh   , Keechul Jung 

[Show more](#)

[doi:10.1016/j.patcog.2004.01.013](https://doi.org/10.1016/j.patcog.2004.01.013)

[Get rights and content](#)

---

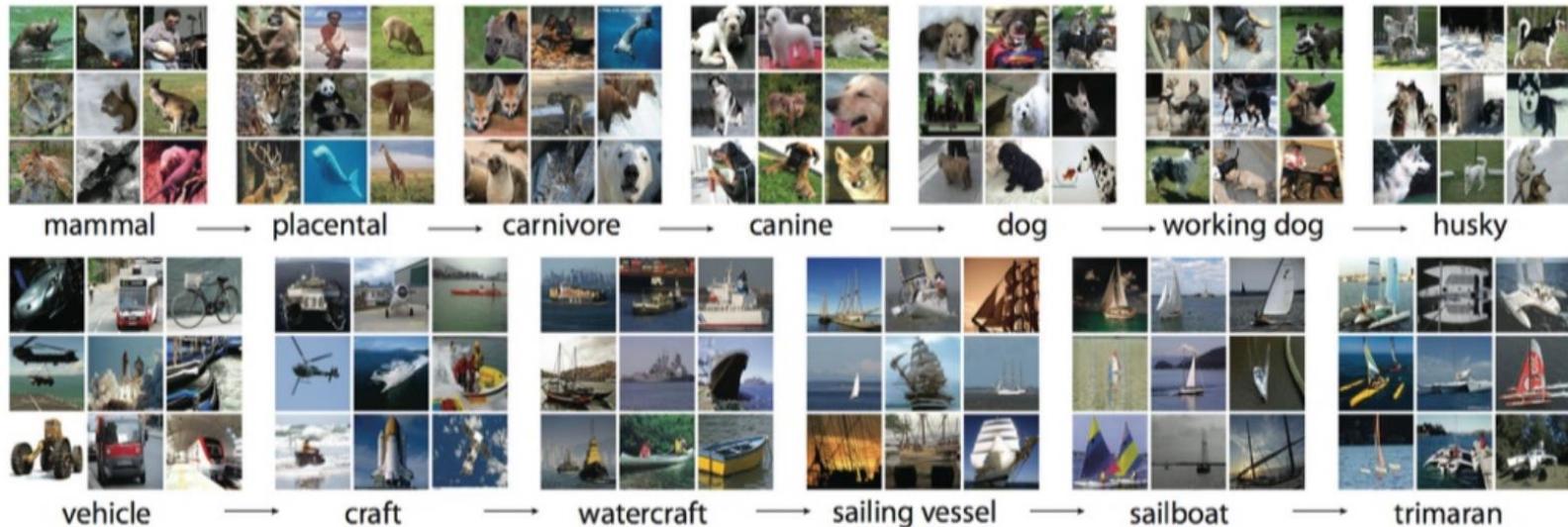
**Abstract**

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



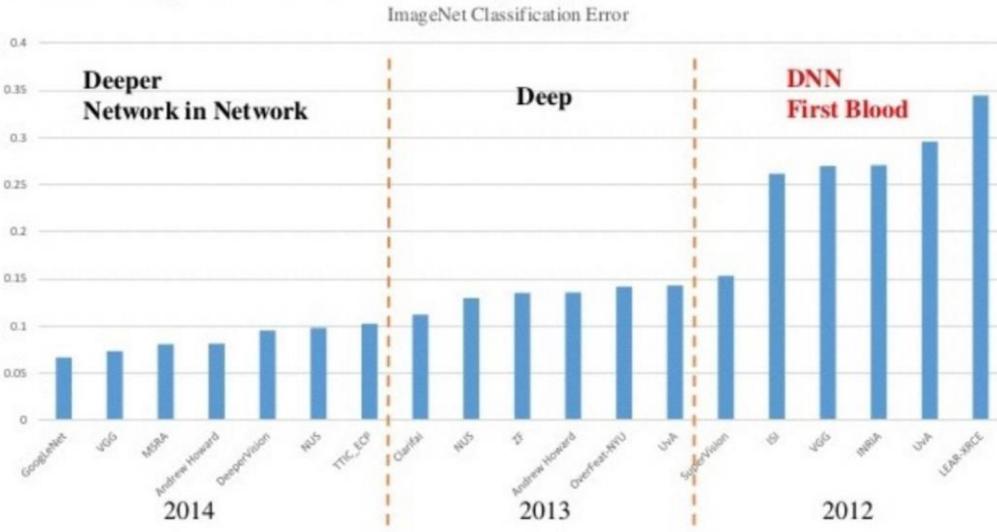
# IMAGENET (2009)

- 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

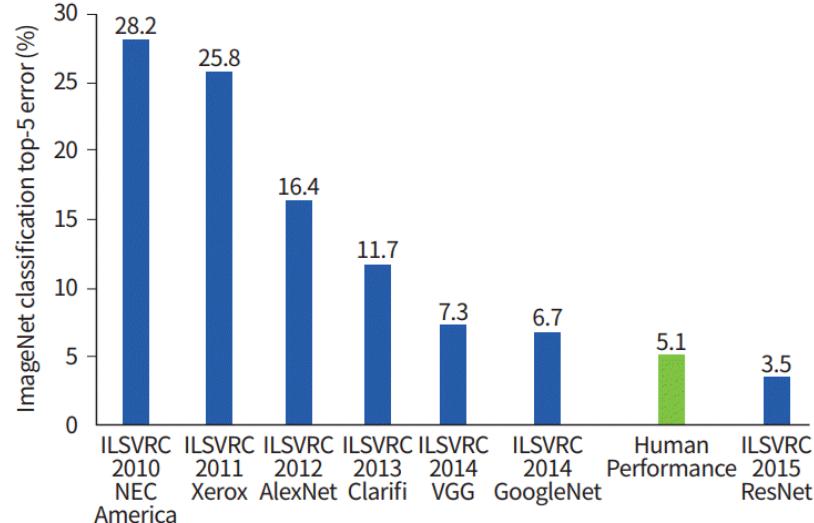


# DROPPING ERROR RATES SINCE THEN

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



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



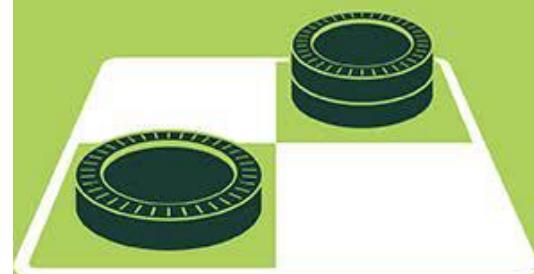
# Introduction to Deep Learning



# THREE EPOCHS OF NEURAL NETWORKS

## ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



50's      1960's      1970's

## MACHINE LEARNING

Machine learning begins to flourish.



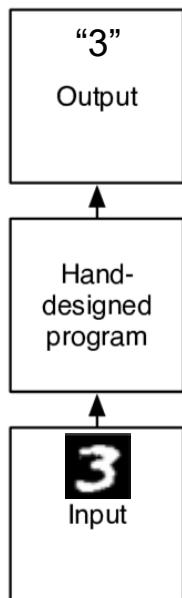
1980's      1990's      2000's      2010's

## DEEP LEARNING

Deep learning breakthroughs drive AI boom.



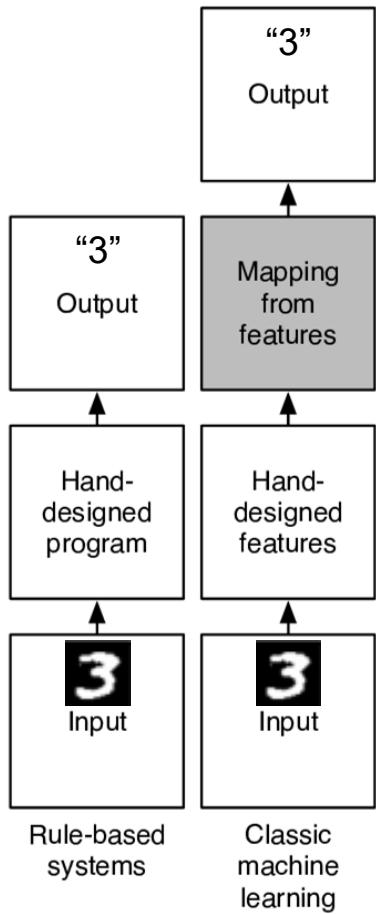
**Rule-based systems:**  
Write algorithm by hand.



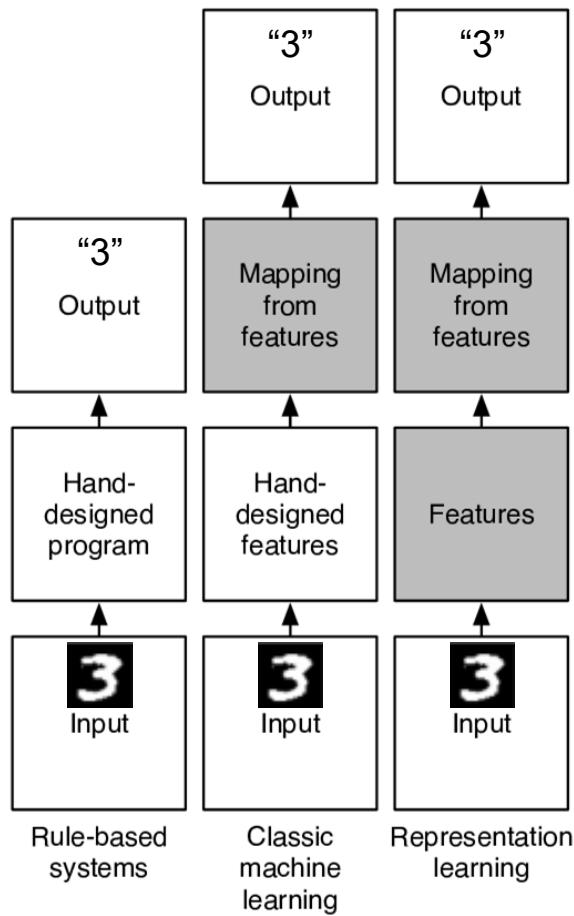
Rule-based  
systems

graphic: Y. Bengio, Deep Learning, MLSS 2015

**Rule-based systems:**  
Write algorithm by hand.



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

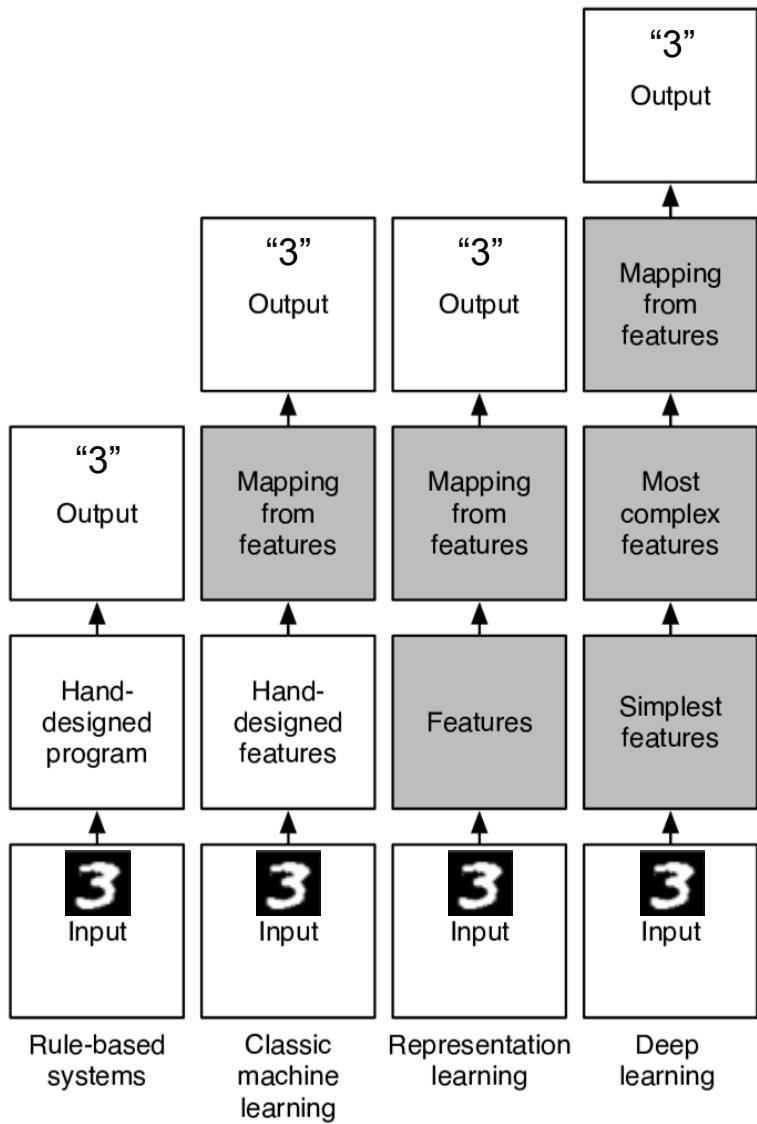


**Rule-based systems:**  
Write algorithm by hand.

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

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

# MACHINE LEARNING PARADIGMS

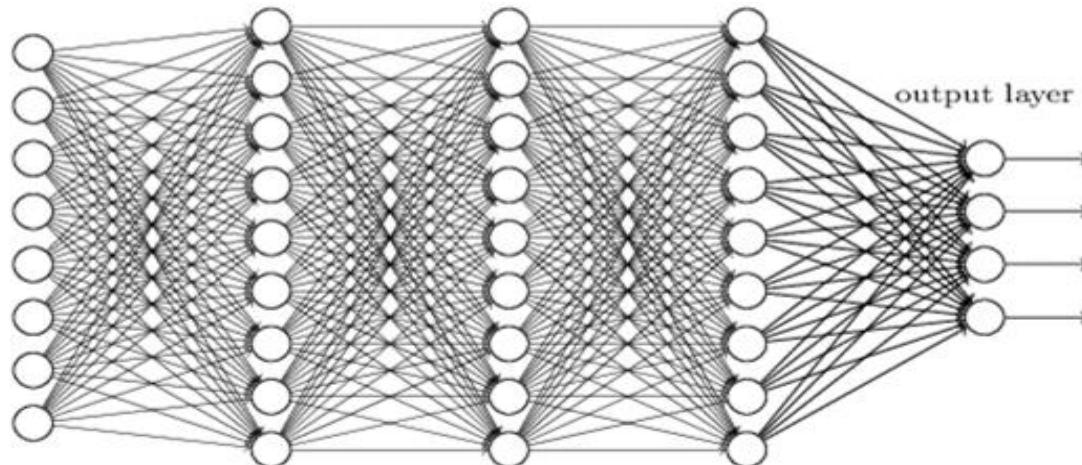


## Deep learning:

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")

# So What is Deep Learning?



# WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

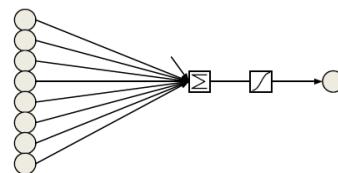
$$y = \sigma(b + w^T x) \quad (\text{equivalent to logistic regression})$$

# WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

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

expression can be visualized as a graph:

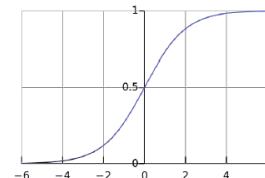


$$\mathbf{x} \qquad b + \mathbf{w}^T \mathbf{x} \qquad y$$

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

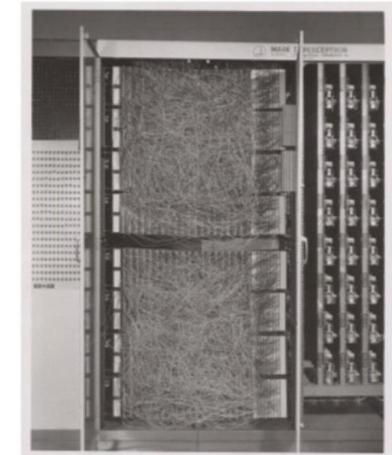
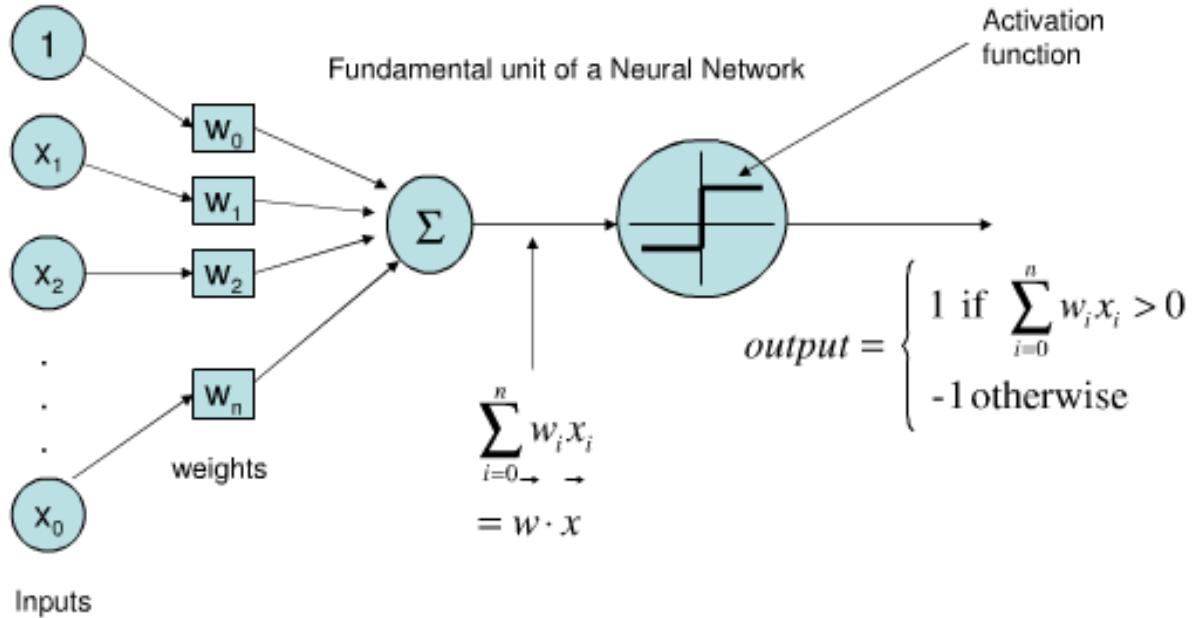
$$b + \mathbf{w}^T \mathbf{x} = b + \sum_i w_i x_i$$

**followed by a nonlinear function.**



## 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

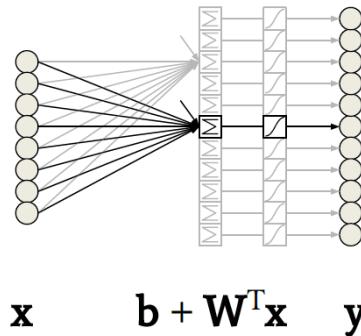
**For the first time a NN could solve simple classification problems merely from training data**

# WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

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

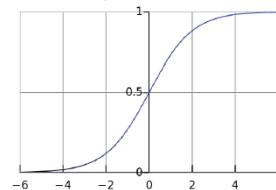
expression can be visualized as a graph:



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

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

**followed by a nonlinear function.**

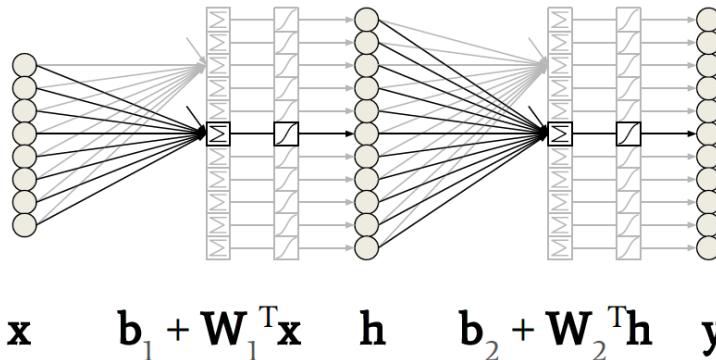


# WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

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

expression can be visualized as a graph:

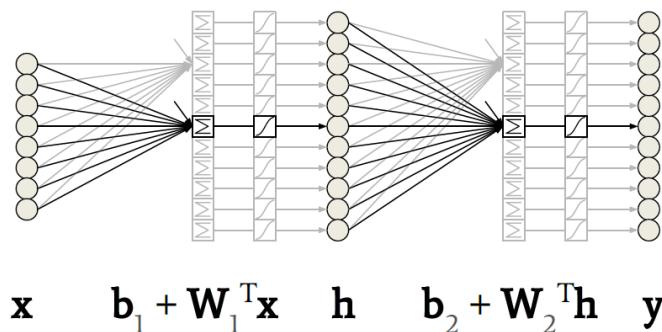


# WHAT ARE ARTIFICIAL NEURAL NETWORKS?

a fancy name for particular mathematical expressions, such as:

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

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!

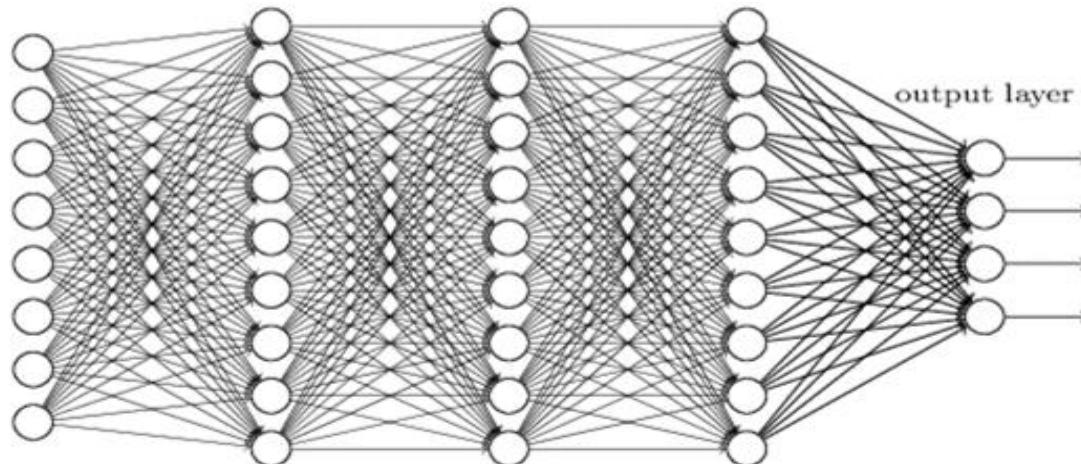
(Cybenko 1998, Hornik 1991)

## 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.

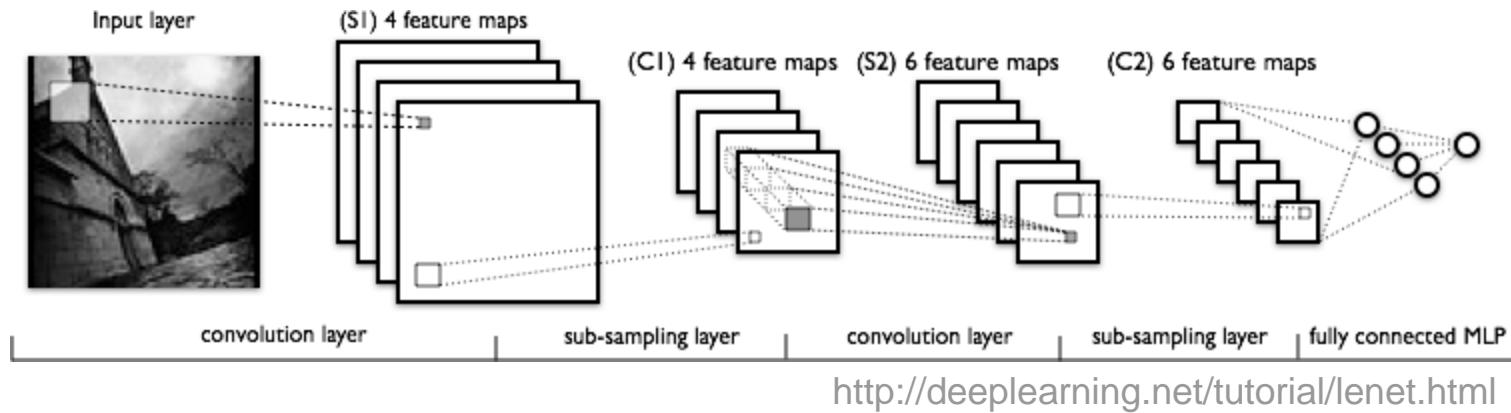
# WHAT MAKES THEM DEEP?

- 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)



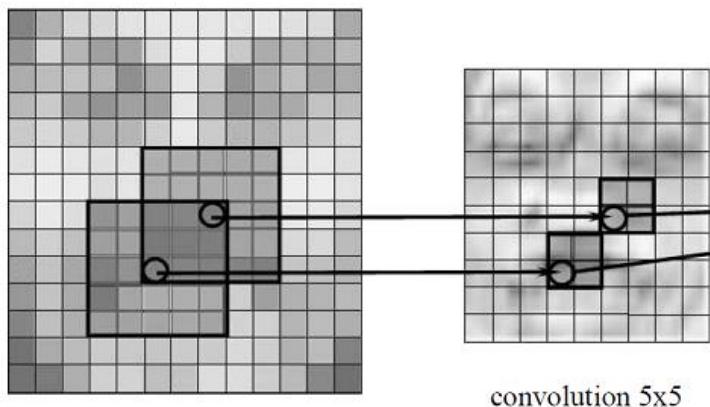
Combines three types of layers:

- **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



Operation	Kernel	Image result
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
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\\_\(image\\_processing\)](https://en.wikipedia.org/wiki/Kernel_(image_processing))

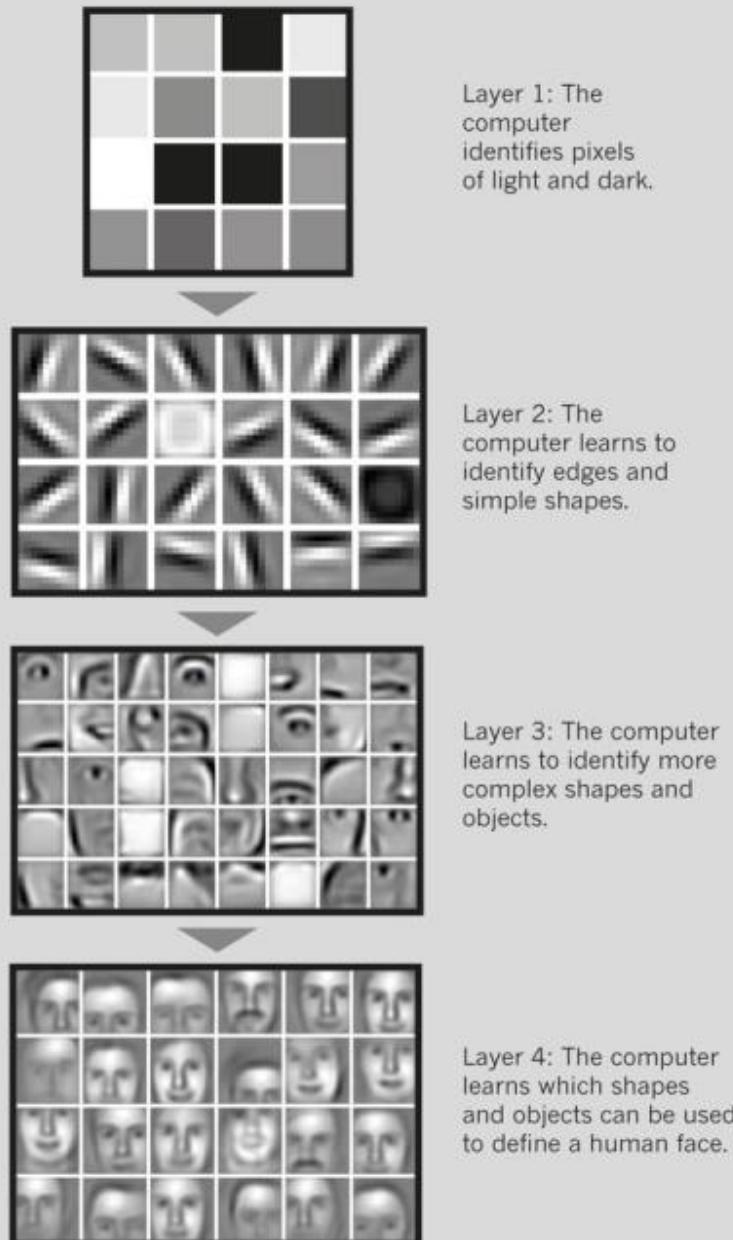
# IMAGE PROCESSING



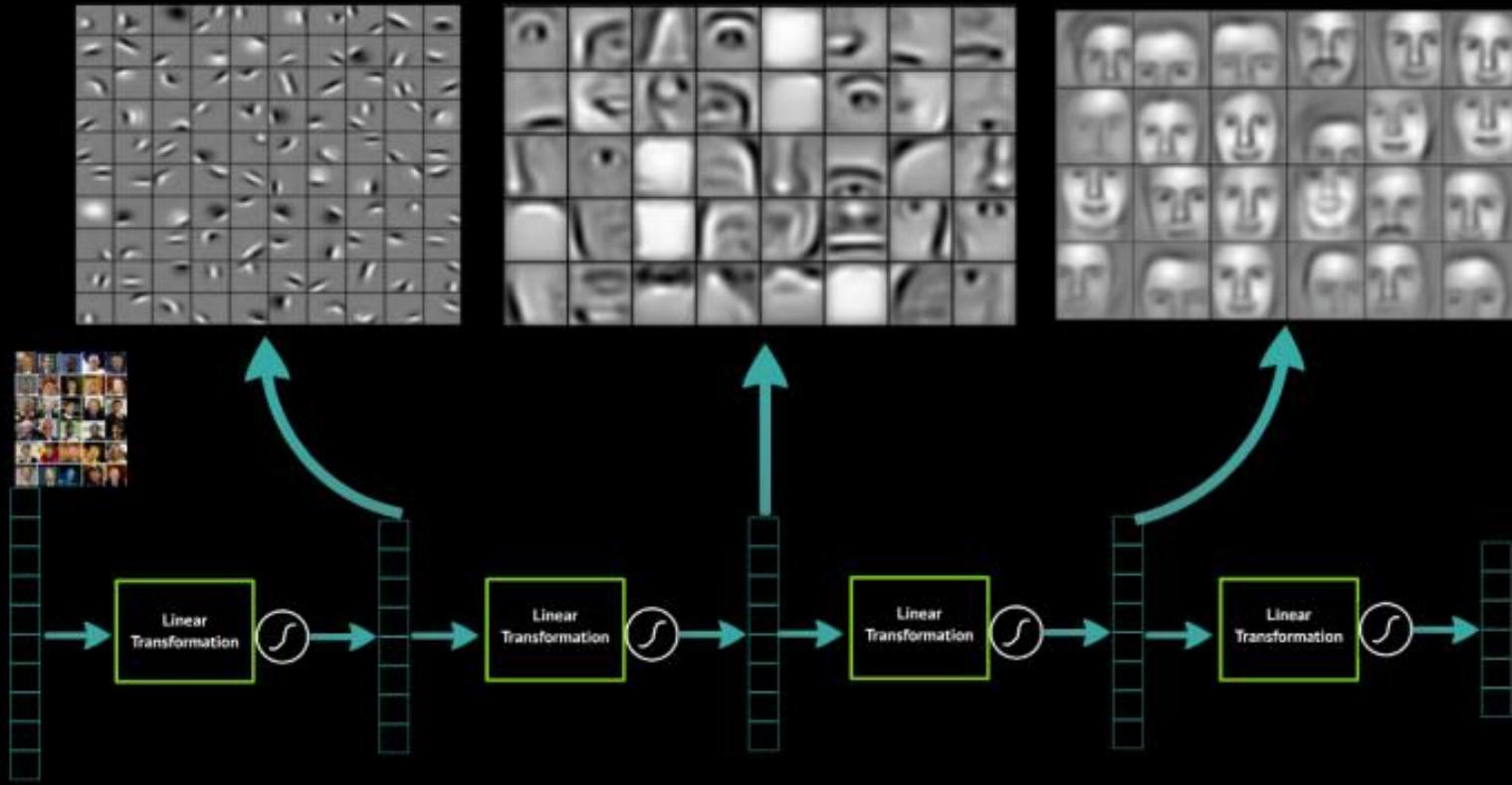
Input

# FACIAL RECOGNITION

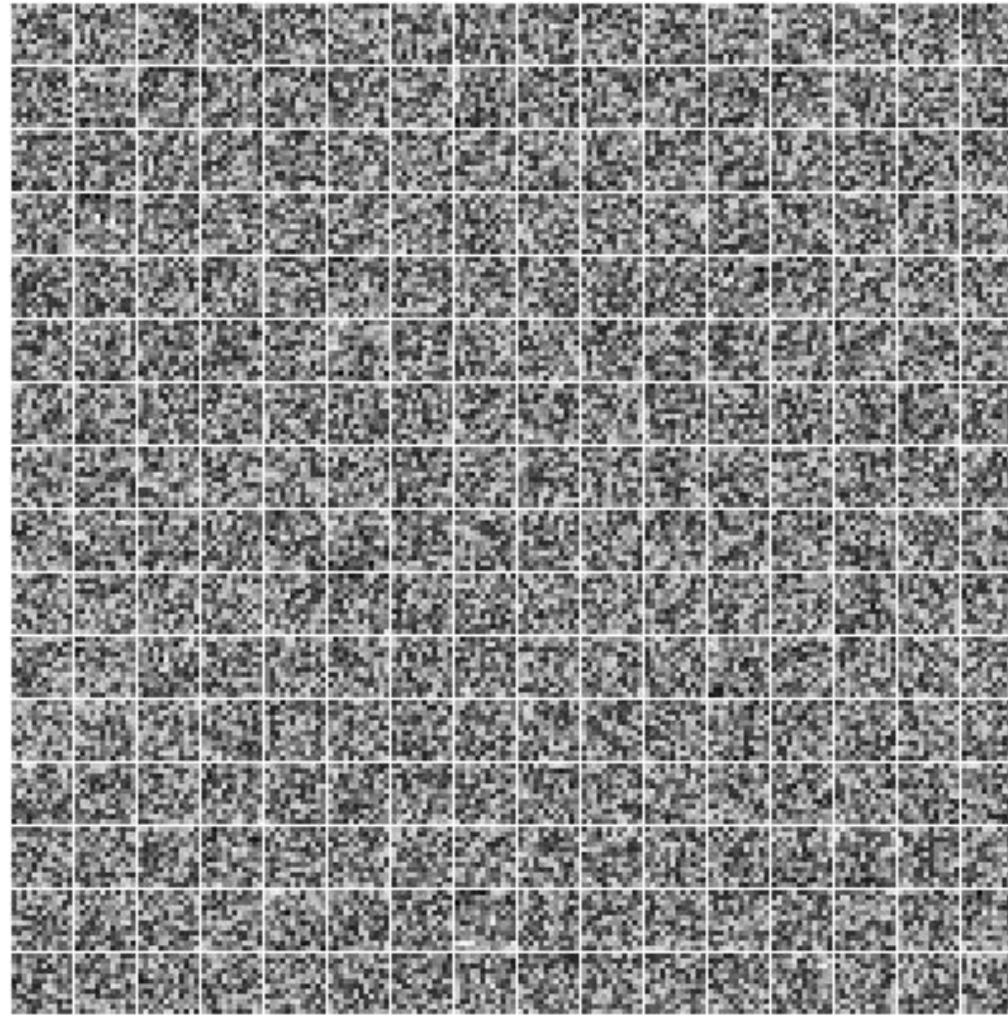
Deep-learning neural networks use layers of increasingly complex rules to categorize complicated shapes such as faces.



## Deep Learning learns layers of features



# IMAGE PROCESSING



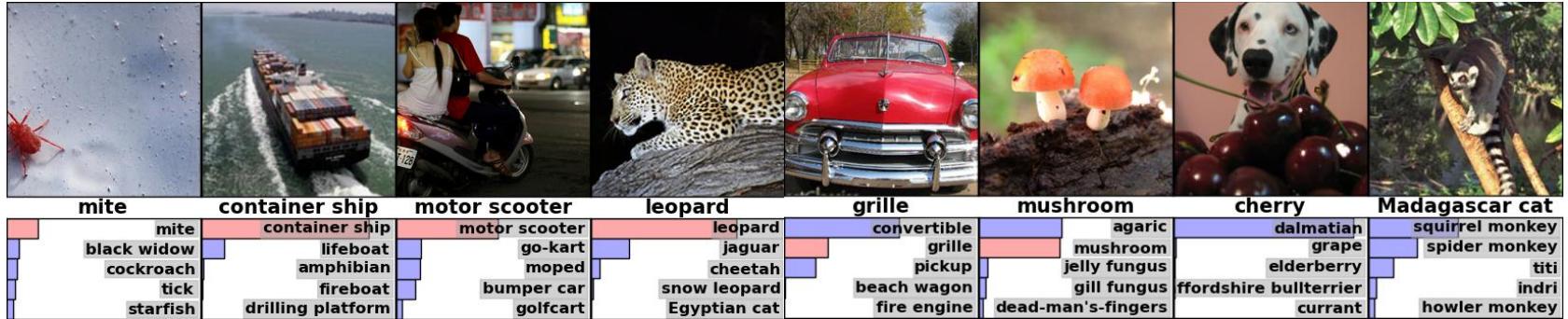
iteration no. 0

# 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

Describes without errors



A person riding a motorcycle on a dirt road.

Describes with minor errors



Two dogs play in the grass.

Somewhat related to the image



A skateboarder does a trick on a ramp.

Unrelated to the image



A dog is jumping to catch a frisbee.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



A little girl in a pink hat is blowing bubbles.



A refrigerator filled with lots of food and drinks.

# DEPTH ESTIMATION

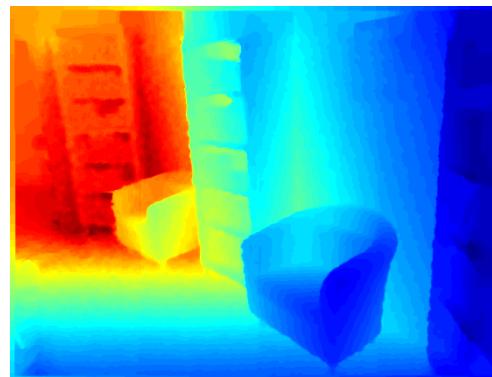
**Input:** Single photograph

**Output:** Its depth map (how far is each pixel from the observer)

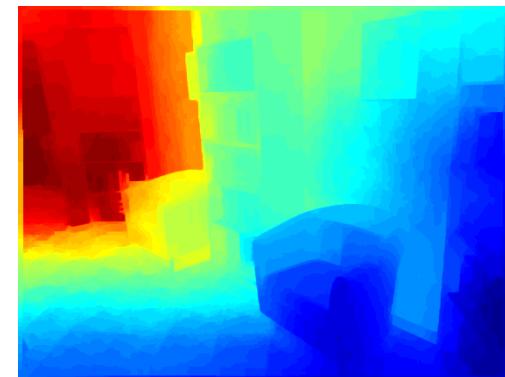
**Method:** fully-convolutional network



input



ground truth

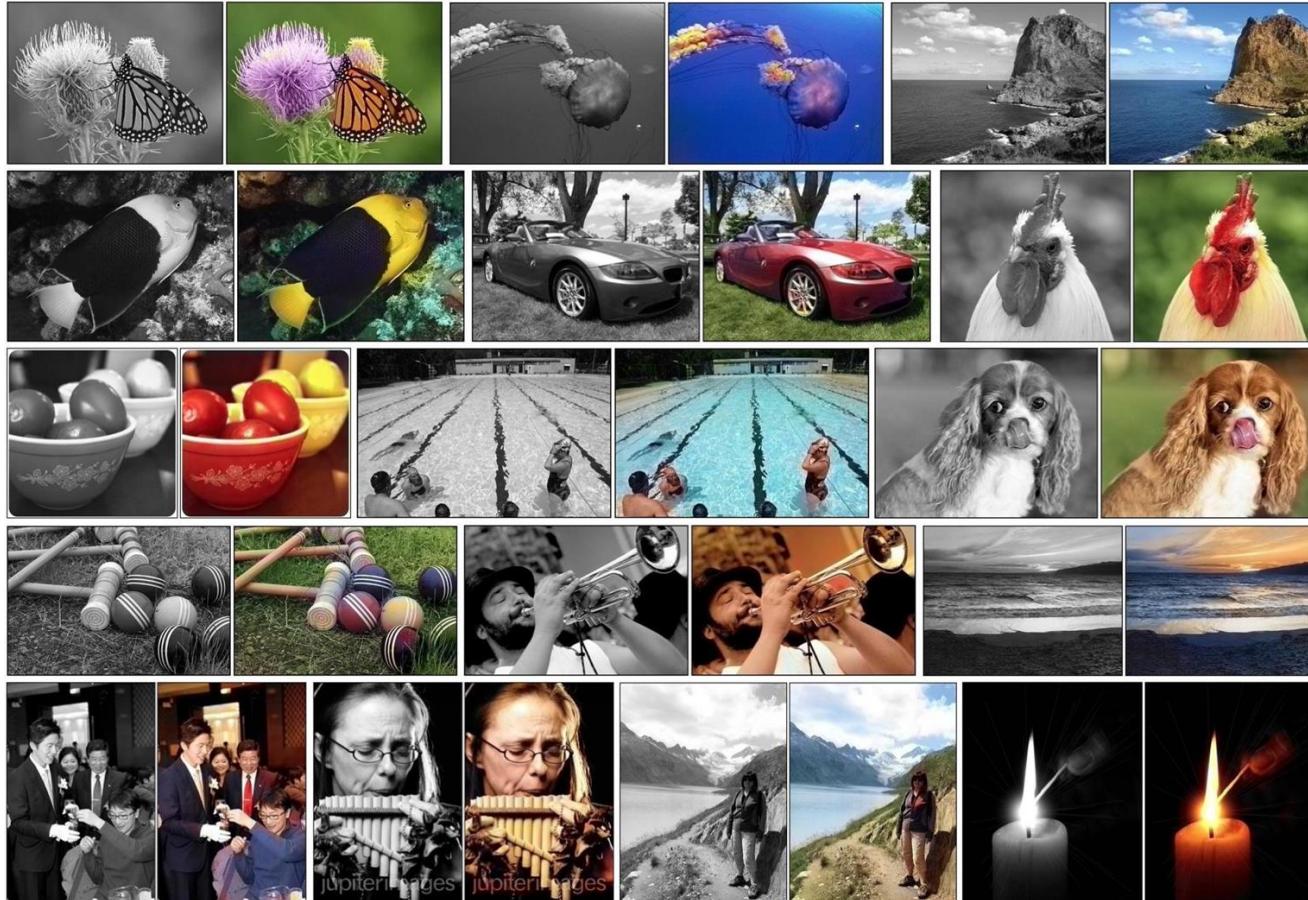


network prediction

# IMAGE COLORIZATION

**Input:** Gray-scale image. **Output:** Colored image.

**Method:** fully-convolutional network

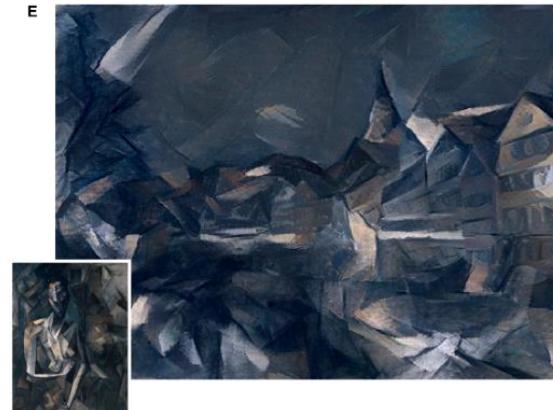


Mar 2016: Colorful Image Colorization, <http://arxiv.org/abs/1603.08511>,  
<http://richzhang.github.io/colorization/>

# 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



# 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](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

# GOOGLE ALPHAGO

**Input:** Go board (stone positions and some helping markers)

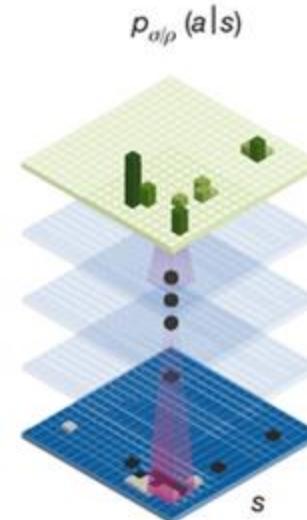
**Output:** Position for next move (policy) or likely winner (value)

**Method:** CNN with supervised and reinforcement learning

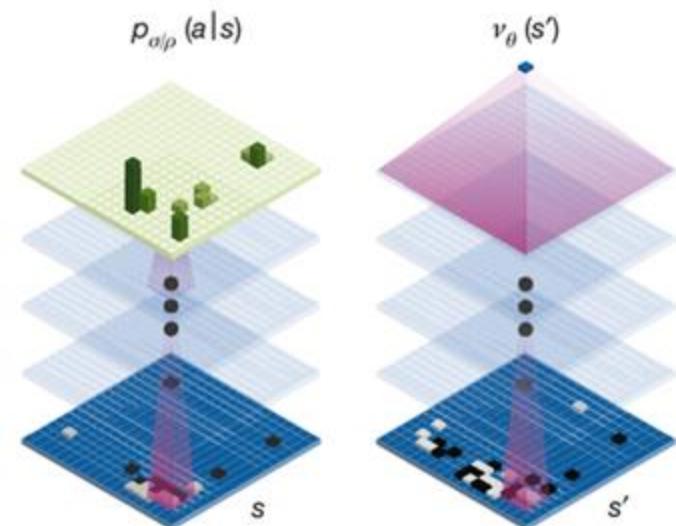
Combined with Monte-Carlo Tree Search for playing



Policy network



Value network



# PROGRESSIVE GROWING OF GANs FOR IMPROVED QUALITY, STABILITY, AND VARIATION

Tero Karras

NVIDIA

{tkarras,

Timo Aila

NVIDIA

,taila,

Samuli Laine

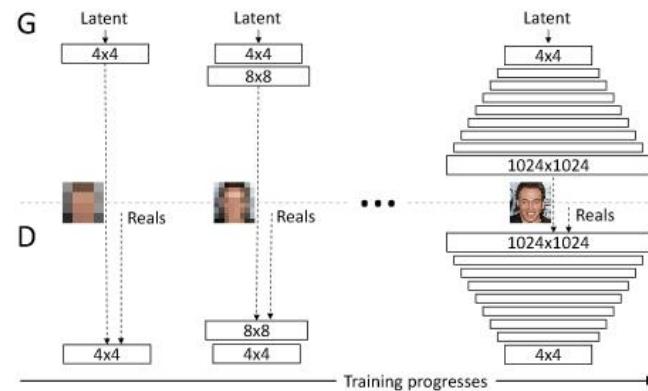
NVIDIA

,slaine,

Jaakko Lehtinen

NVIDIA and Aalto University

jlehtinen}@nvidia.com



# A Style-Based Generator Architecture for Generative Adversarial Networks

Tero Karras  
 NVIDIA

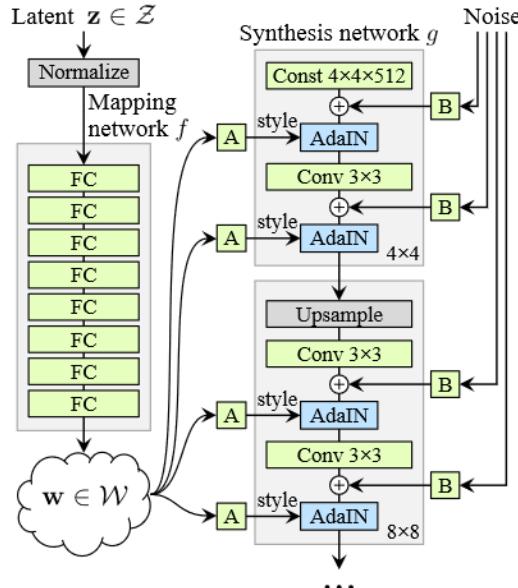
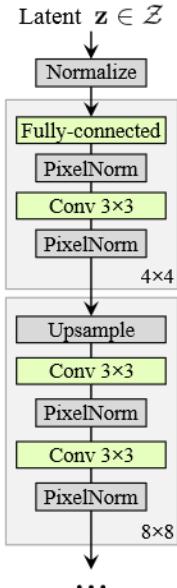
tkarras@nvidia.com

Samuli Laine  
 NVIDIA

slaine@nvidia.com

Timo Aila  
 NVIDIA

taila@nvidia.com



## MODEL GENERATION AI

- Proprietary product
- Full person generation



# FAKE AUDIO

- **No Deep Fake for Audio!!!**
  - Deep Fakes with voice require
    - Traditional audio manipulation
    - Impressionist
- **Approaches to style-transfer in audio**
  - Not promising



Deep Fake VFX - Pity the poor impressionist by Jim Meskimen <https://www.youtube.com/watch?v=Wm3squcz7Aw>

# TACOTRON-2 EXAMPLES

- Synthetic or Real?

“That girl did a video about Star Wars lipstick.”



“She earned a doctorate in sociology at Columbia University.”



“George Washington was the first President of the United States.”

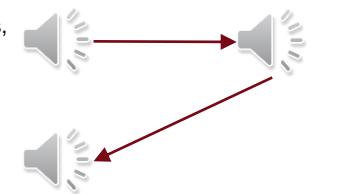


Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... & Saurous, R. A. (2018, April). Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4779-4783). IEEE.

- Stress and intonation
- Questions
- Prosody
  - Intonation, rhythm, tone

## Style / Reference

**Reference text:** Alice was not much surprised at this, she was getting so used to queer things happening.



## Result

**Perturbed text:** Eric was not much surprised at this, he was getting so used to TensorFlow breaking.



## Singing



Skerry-Ryan, R. J., Battenberg, E., Xiao, Y., Wang, Y., Stanton, D., Shor, J., ... & Saurous, R. A. (2018). Towards end-to-end prosody transfer for expressive speech synthesis with tacotron. arXiv preprint arXiv:1803.09047.

# ARTIFICIAL INTELLIGENCE FOR SAFETY AND SECURITY



# AUDIO EVENT DETECTION

Analytic Modules



## Task

Detect and predict audio-events into predefined categories

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

## Use-Case

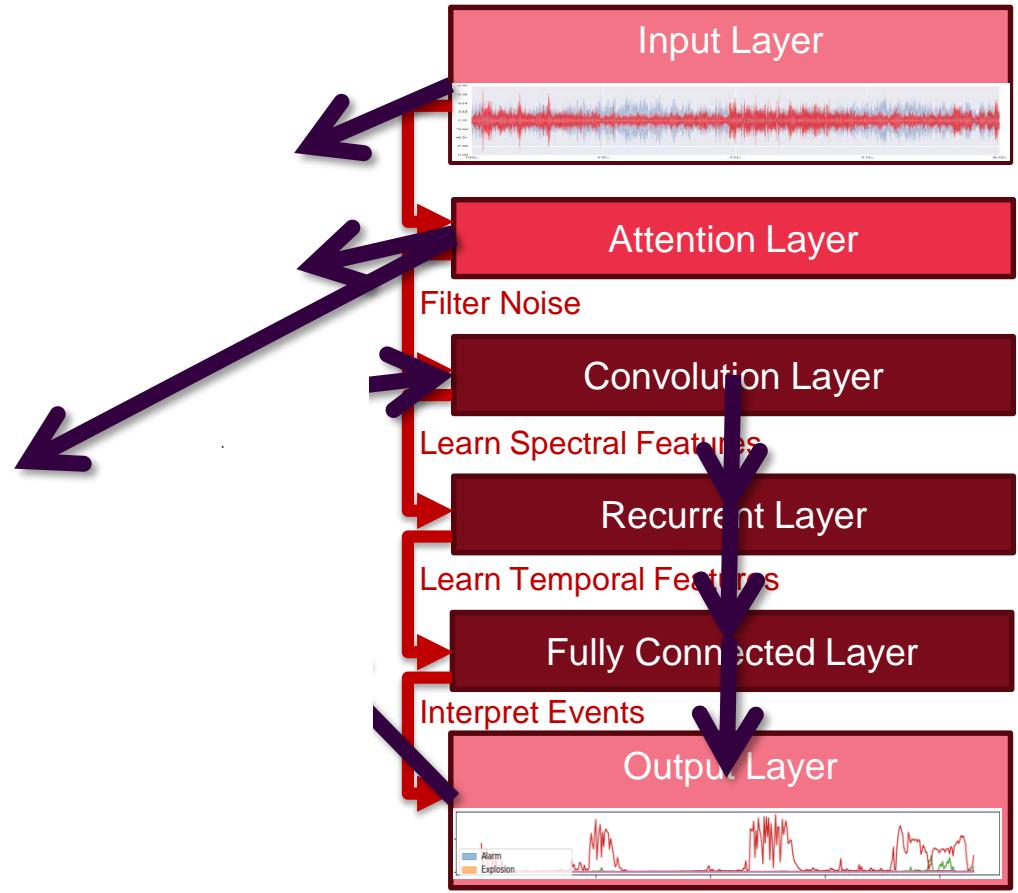
Content filter in mass video-data

Example: attack with firearms

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



# RECURRENT CONVOLUTIONAL NEURAL NETWORKS

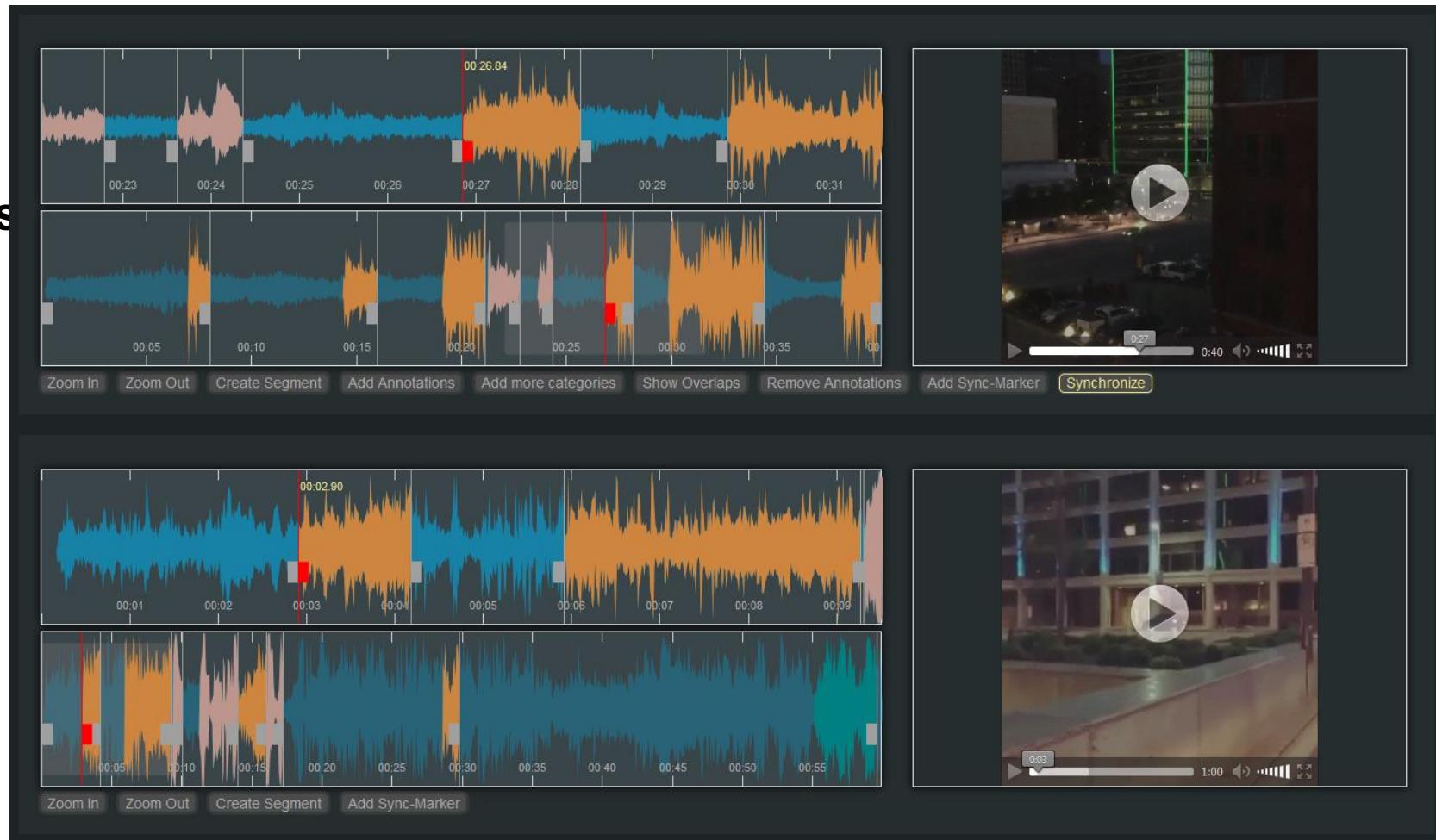


High signal to noise ratio  
Due to attention layer

Smoothing Function  
Temporal Segmentation per Event Category

# AUDIO EVENTS VISUALIZED

Us



# AUDIO SIMILARITY

## Analytic Modules



# AUDIO SIMILARITY SEARCH

## Task

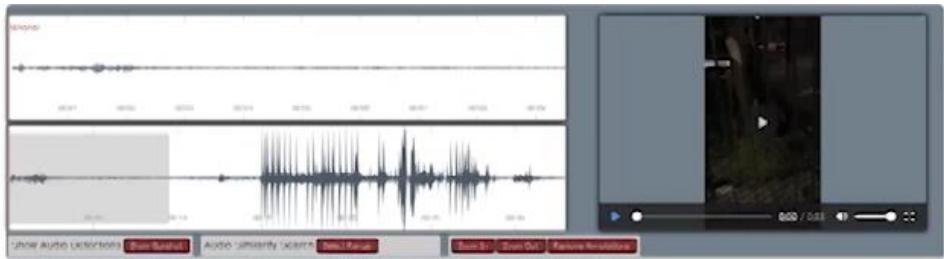
- Searching for video-segments with similar audio-signature
- Sub-Segment video-search

## 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)

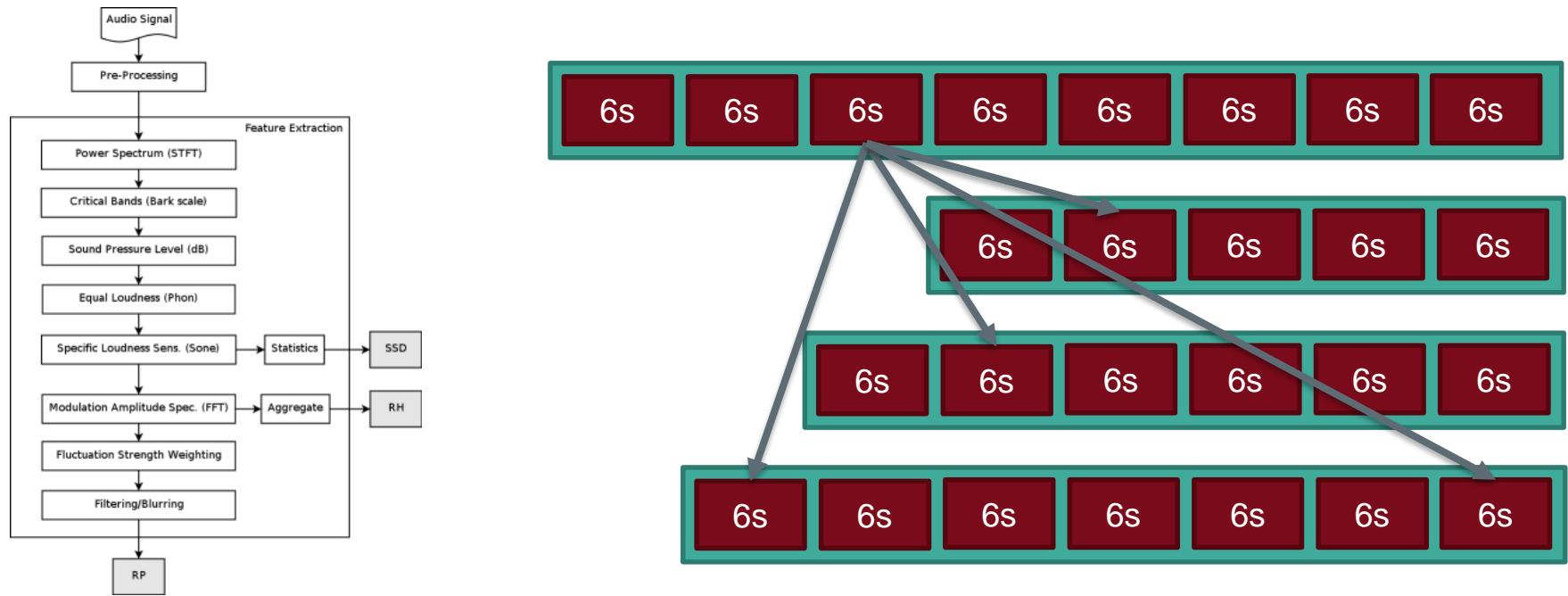
## Solution

- Select range in video
- Retrieve a list of similar sounding video segments
- Sorted by similarity



# 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



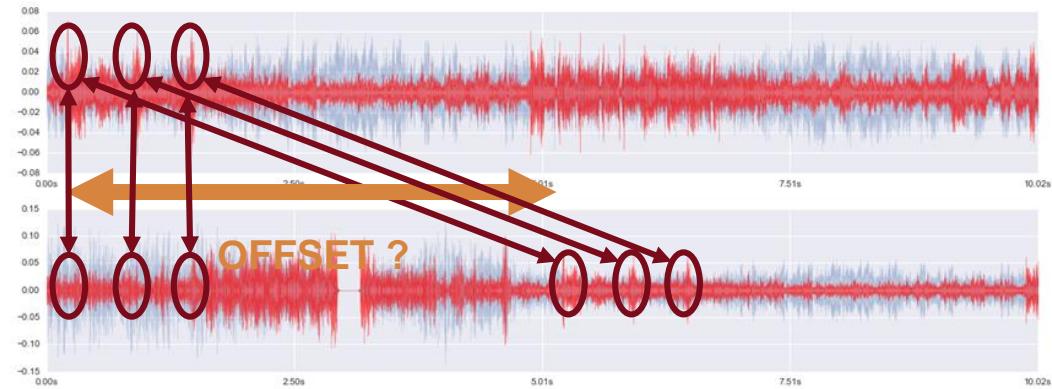
# AUDIO-BASED VIDEO-SYNCHRONIZATION

## ◆ Task

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

## ◆ Technology

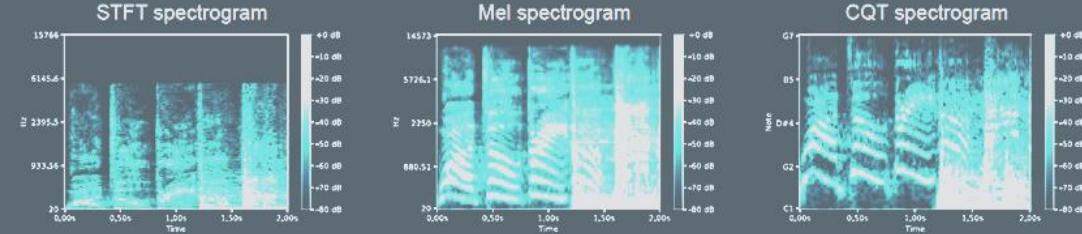
- ◆ Audio-fingerprints (chromaprint)
- ◆ Noise invariant



# SCENE ANALYSIS

## Audio-Visual Scene Understanding

Artificial Intelligence  
Multi-media Analysis



- **Acoustic Scene Classification**

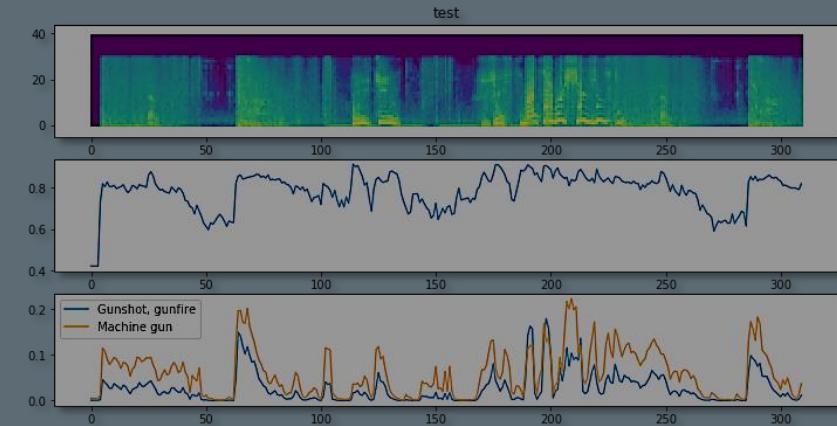
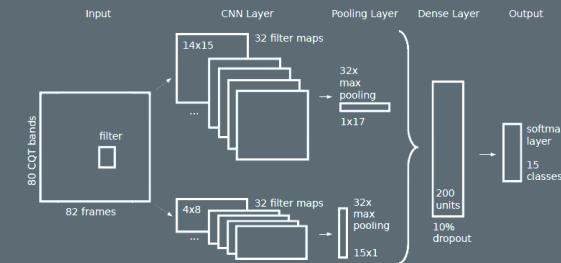
- Identify different acoustic scenes (Bus, Train, Urban Park, Bar)
- Identify different activities (talking, walking, reading, children playing)

- **Audio Event Detection**

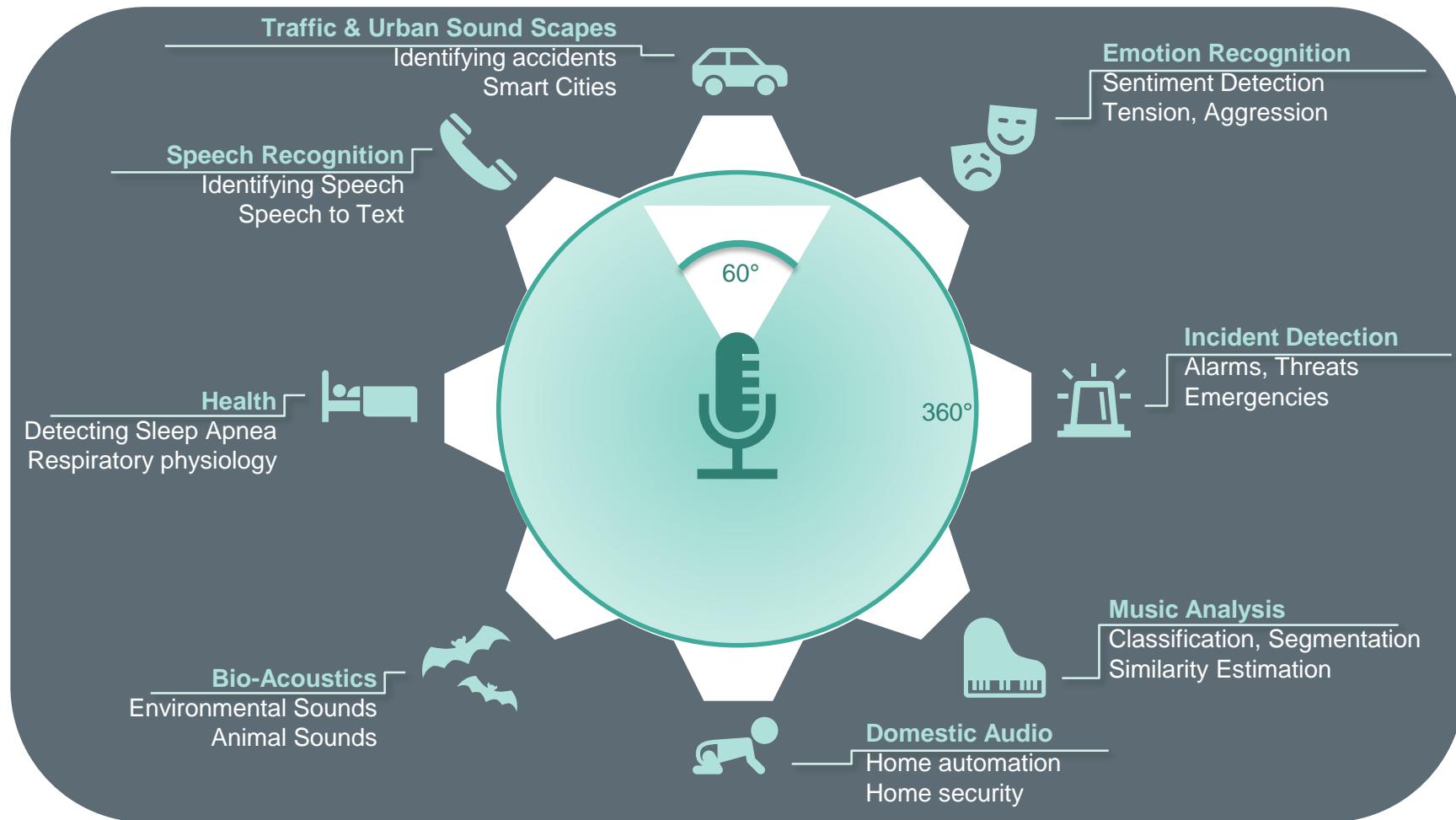
- Detect acoustic events in audio-stream (onsets/length)
- Identify detected Events (Gunshots, explosions, baby cry)

- **Audio-Visual Scene Understanding**

- Combining acoustic with visual information
- Improved interpretability of current scene
- Multi-task learning



# 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

## Image Classification with Deep Neural Networks

Large Scale Deep Learning  
Visual Computing

### • Overview

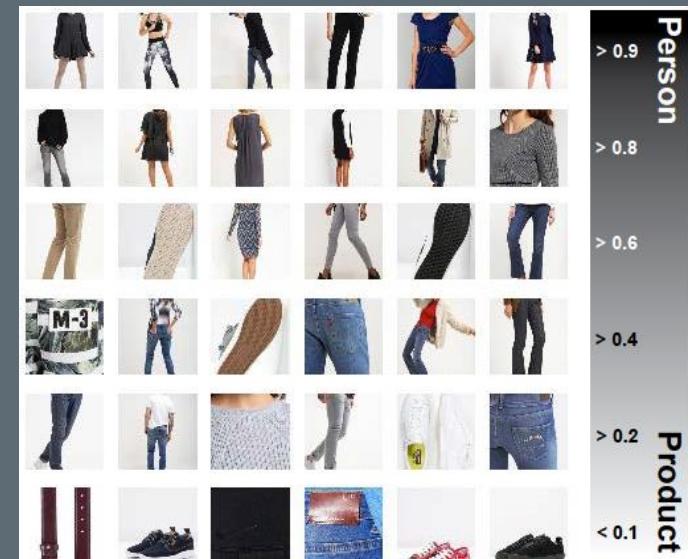
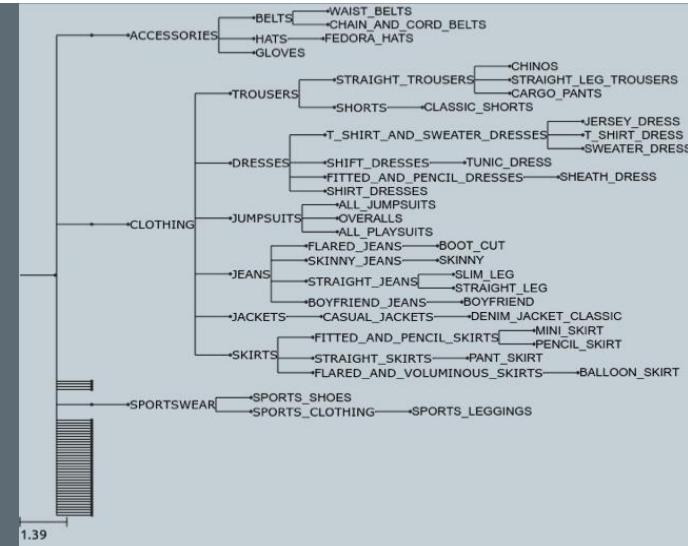
- Empirical study
- Applying deep Convolutional Neural Networks to fashion classification
- Evaluated five CNN architectures
- Custom and pre-trained models
- Evaluated on three tasks
  - Person detection
  - Product classification
  - Gender prediction

### • 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

## 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 ( $1^{\circ}$ E/W/N/S Neighborhood)
  - Random cuts (cut-shuffle-merge)
  - Volume-shifting
  - Noise Overlays



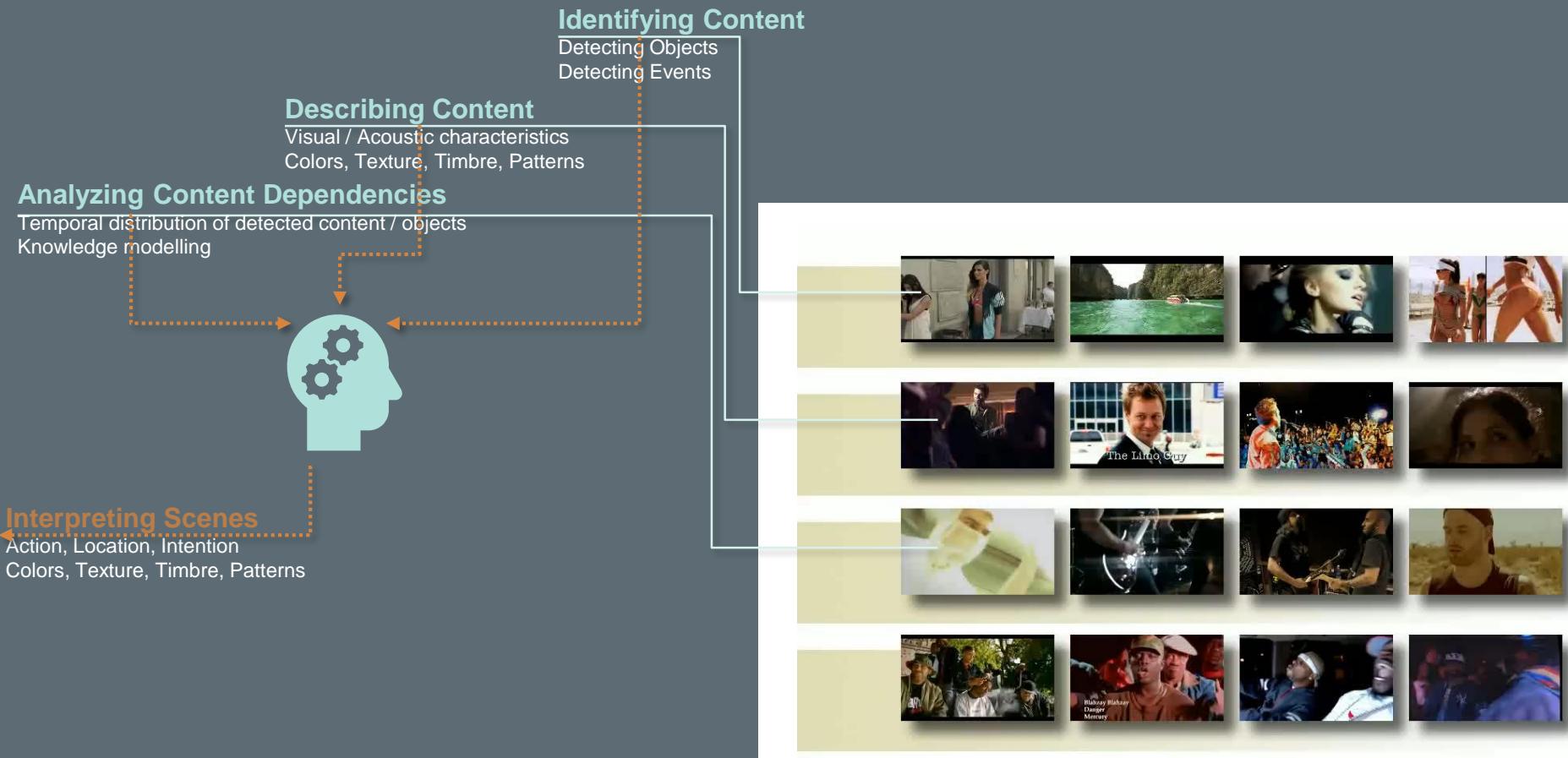
# MULTI- MODAL- ANALYSIS



# MULTI-MEDIA UNDERSTANDING

## Audio-Visual Video Analysis and Classification

Artificial Intelligence  
 Multi-media / Multi-modal



# SENTIMENT DETECTION

## Sentiment Detection

Artificial Intelligence  
 Multi-media / Multi-modal

### Crowd scene / behaviour analysis

Anomaly detection in natural scenes / crowds  
 Threat / escalation estimation

### Audio Analysis

Audio based sentiment detection  
 Arousal, Pleasure, Dominance

### Speaker Sentiment Detection

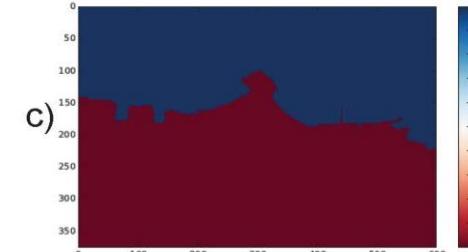
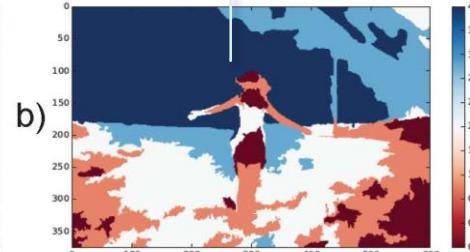
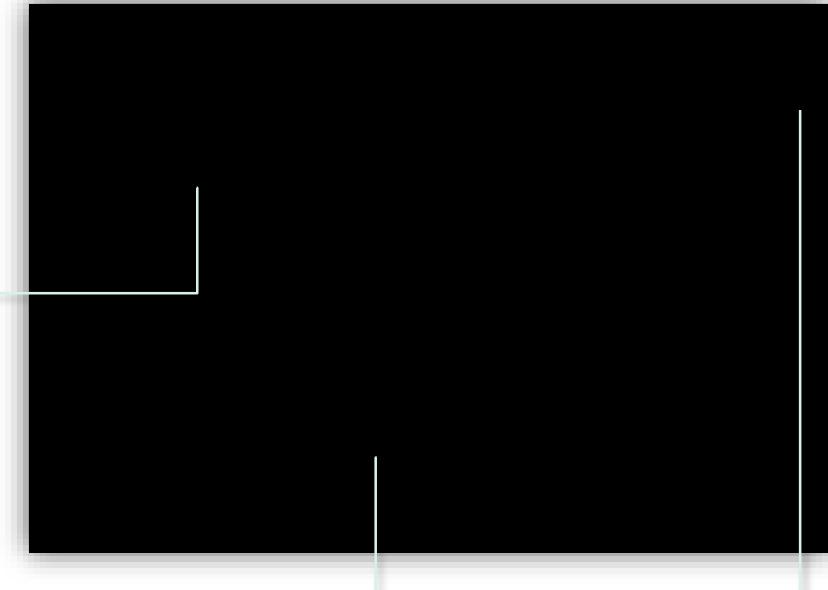
Sentiment of speaker  
 Tension, excitement, affection

### Affective Contrasts

Cold / Warm  
 Light / Dark

### Color Statistics

Calculating color distributions  
 Deriving higher level features

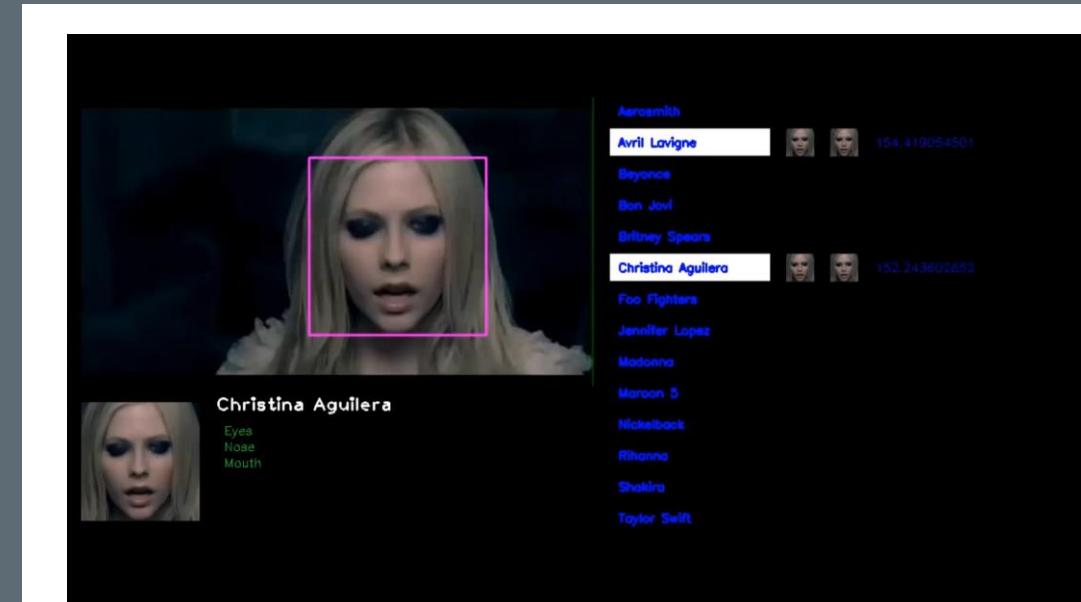


# SPEAKER / SINGER IDENTIFICATION

## Audio-Visual Signer Identification

Artificial Intelligence  
Multi-media / Multi-modal

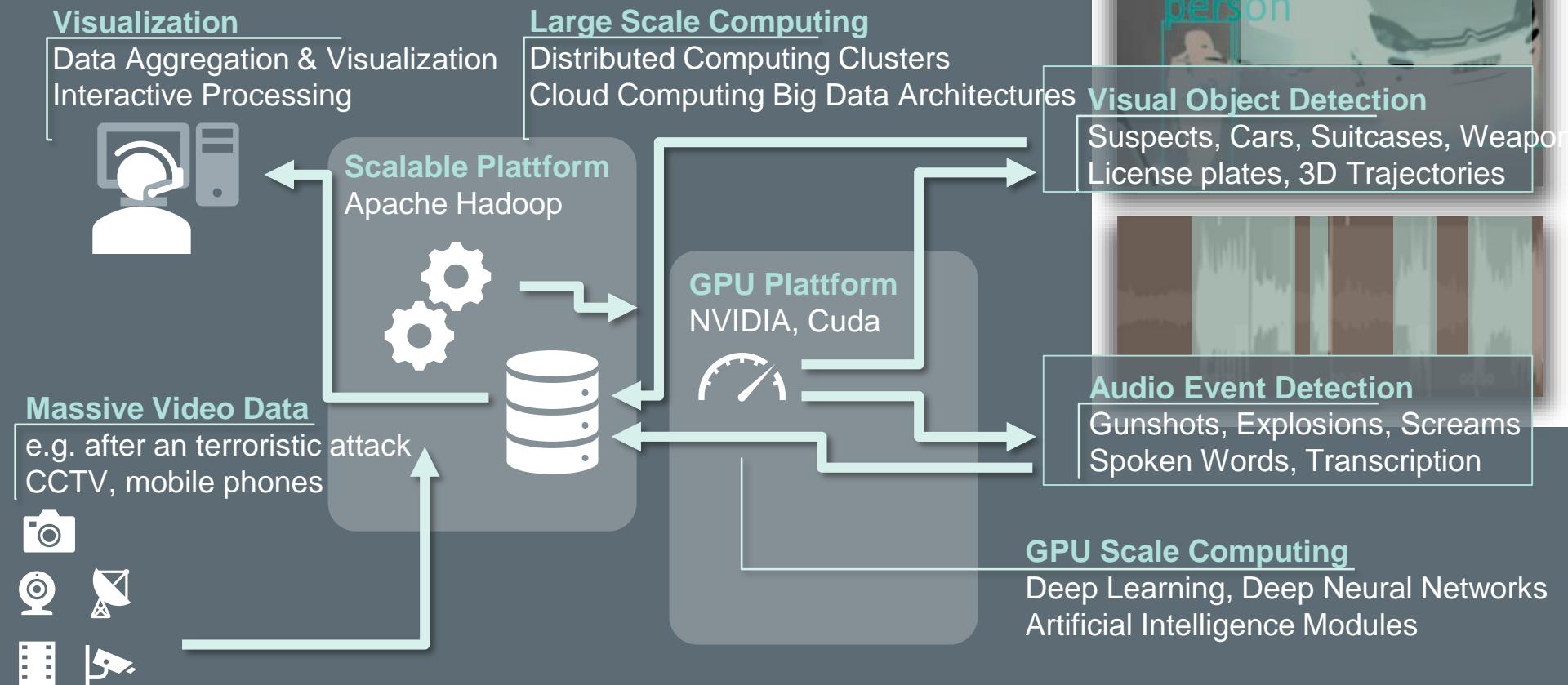
- **Identify Speaker / Singer in Audio/Video Sequence**
  - Based on segmentation (identified vocalized segments)
  - Supervised identification (models trained on declared persons)
  - Un-supervised identification (relative similarity estimation)
- **Audio-Identification**
  - Based on acoustic models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

## Example: Forensic Analytics in Massive Video Content

Artificial Intelligence  
Scalable Computing



# FURTHER RESOURCES

## Code & Tutorials

# Vienna Deep Learning Meetup



**Vienna's largest monthly event on Deep Learning & AI**

## The Organizers:



Thomas Lidy  
Musimap



Alex Schindler  
AIT & TU Wien



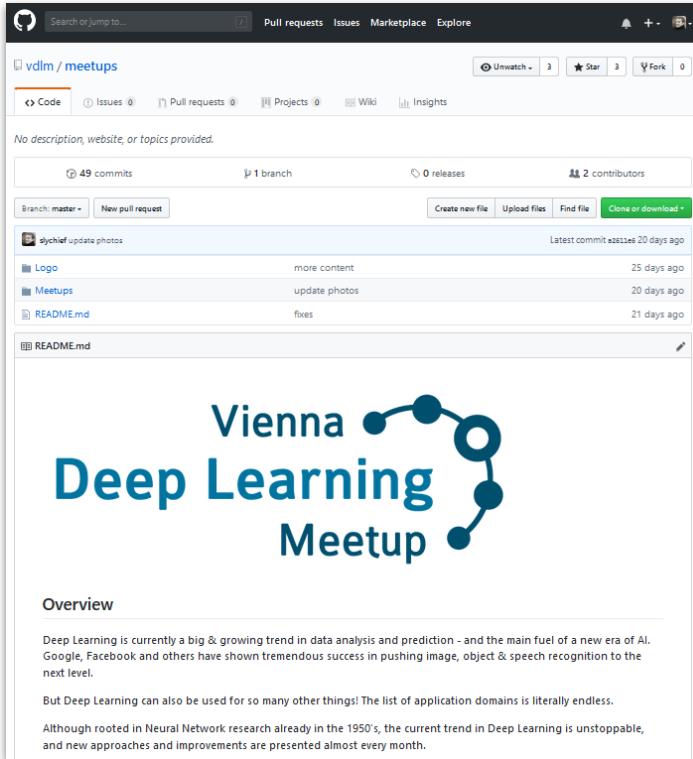
René Donner  
contextflow



Jan Schlüter  
OFAI & UTLN

**[www.meetup.com/Vienna-Deep-Learning-Meetup](https://www.meetup.com/Vienna-Deep-Learning-Meetup)**

## VDLM ON GITHUB



The screenshot shows the GitHub repository page for `vdlm/meetups`. The repository has 49 commits, 1 branch, 0 releases, and 2 contributors. The latest commit was made 20 days ago. The repository contains files like `Logo`, `Meetups`, and `README.md`. Below the repository details, there is a preview of the `README.md` file which contains the following content:

**Vienna Deep Learning Meetup**

**Overview**

Deep Learning is currently a big & growing trend in data analysis and prediction - and the main fuel of a new era of AI. Google, Facebook and others have shown tremendous success in pushing image, object & speech recognition to the next level.

But Deep Learning can also be used for so many other things! The list of application domains is literally endless.

Although rooted in Neural Network research already in the 1950's, the current trend in Deep Learning is unstoppable, and new approaches and improvements are presented almost every month.

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

### Talks

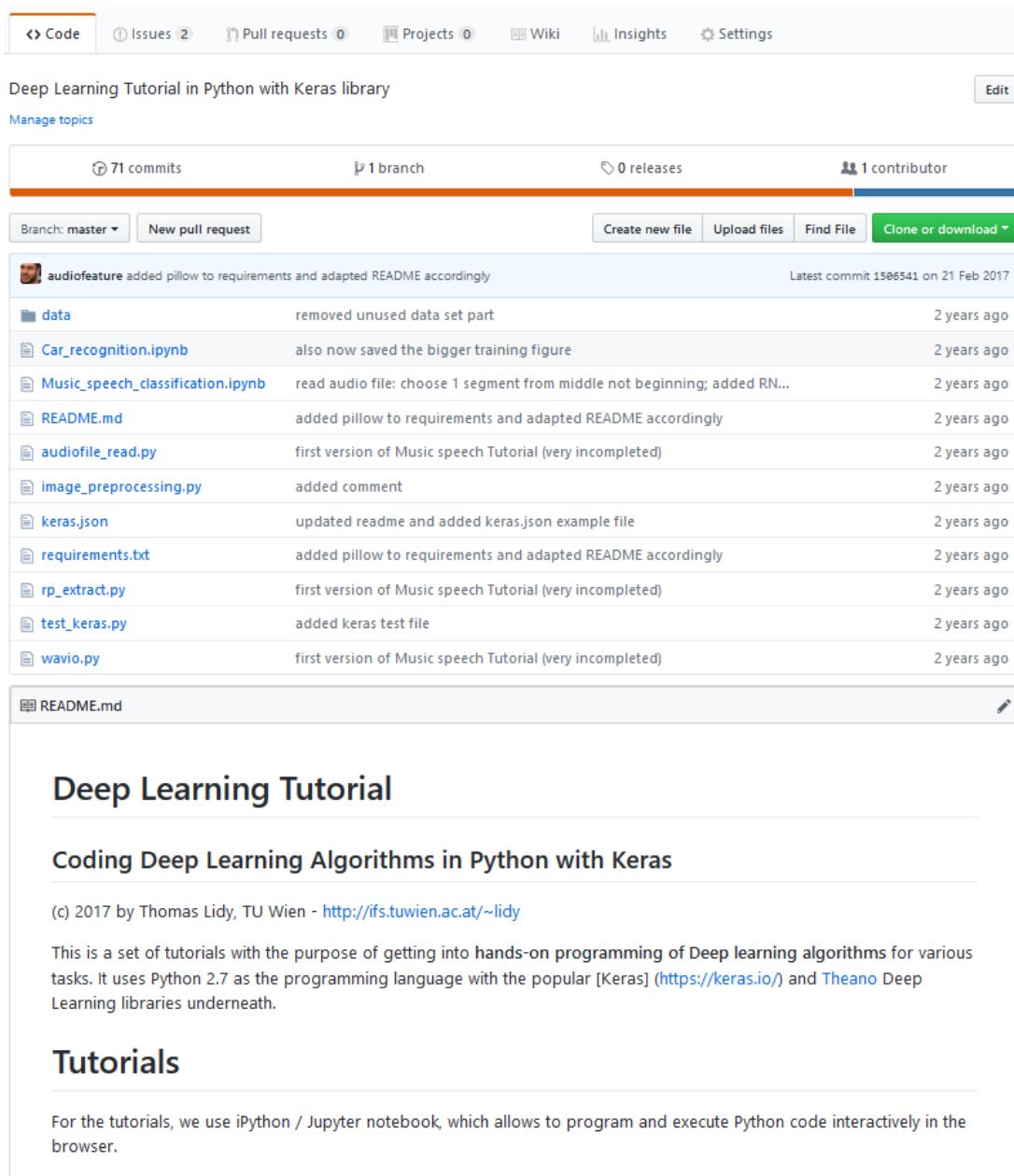
Date	MU#	Speaker	Topic	Slides
2016-04-07	1	Thomas Lidy	An overview presentation of Deep Learning	<a href="#">pdf</a>
2016-04-07	1	Jan Schlüter	History, Approaches, Applications	<a href="#">pdf</a>
2016-05-09	2	Alex Champandard	Neural Networks for Image Synthesis	
2016-05-09	2	Gregor Mitscha-Baude	Recurrent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

[github.com/vdlm/meetups](https://github.com/vdlm/meetups)

# DEEP LEARNING TUTORIAL

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

[https://github.com/tuwien-musicir/DL\\_Tutorial](https://github.com/tuwien-musicir/DL_Tutorial)



Deep Learning Tutorial in Python with Keras library

Manage topics

71 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find File Clone or download

File / Commit	Description	Date
audiofeature added pillow to requirements and adapted README accordingly	Latest commit 1506541 on 21 Feb 2017	
data removed unused data set part	2 years ago	
Car_recognition.ipynb also now saved the bigger training figure	2 years ago	
Music_speech_classification.ipynb read audio file: choose 1 segment from middle not beginning; added RN...	2 years ago	
README.md added pillow to requirements and adapted README accordingly	2 years ago	
audiofile_read.py first version of Music speech Tutorial (very incompletely)	2 years ago	
image_preprocessing.py added comment	2 years ago	
keras.json updated readme and added keras.json example file	2 years ago	
requirements.txt added pillow to requirements and adapted README accordingly	2 years ago	
rp_extract.py first version of Music speech Tutorial (very incompletely)	2 years ago	
test_keras.py added keras test file	2 years ago	
wavio.py first version of Music speech Tutorial (very incompletely)	2 years ago	

README.md

## Deep Learning Tutorial

### Coding Deep Learning Algorithms in Python with Keras

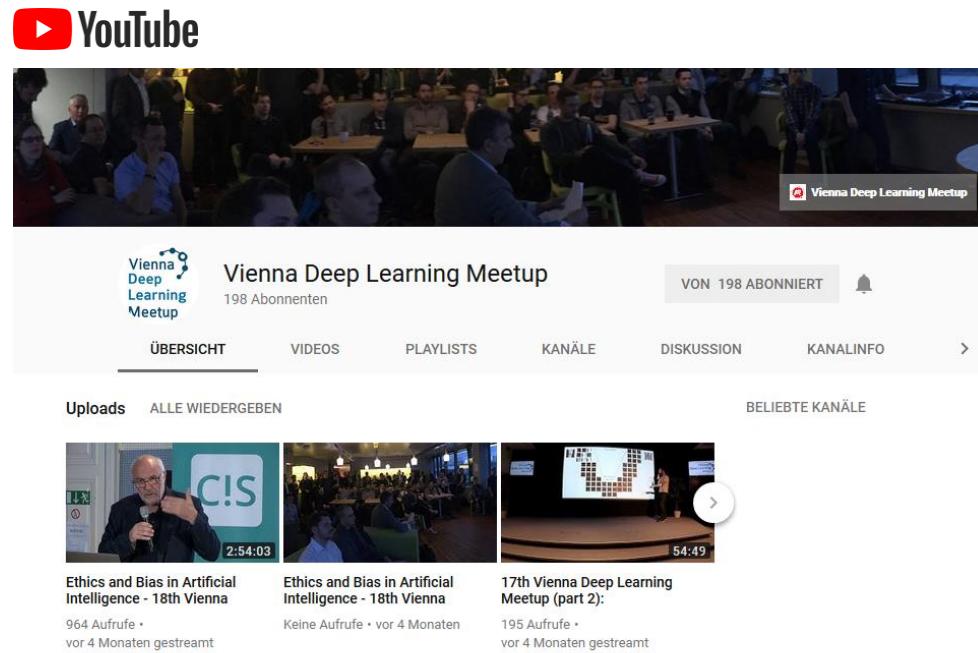
(c) 2017 by Thomas Lidy, TU Wien - <http://ifs.tuwien.ac.at/~lidy>

This is a set of tutorials with the purpose of getting into hands-on programming of Deep learning algorithms for various tasks. It uses Python 2.7 as the programming language with the popular [Keras] (<https://keras.io/>) and **Theano** Deep Learning libraries underneath.

## Tutorials

For the tutorials, we use iPython / Jupyter notebook, which allows to program and execute Python code interactively in the browser.

## VDLM YOUTUBE CHANNEL



The screenshot shows the YouTube channel page for 'Vienna Deep Learning Meetup'. The channel has 198 subscribers. The main video thumbnail shows a group of people at a meetup. Below the video, there are three video thumbnails: 'Ethics and Bias in Artificial Intelligence - 18th Vienna' (964 views), 'Ethics and Bias in Artificial Intelligence - 18th Vienna' (Keine Aufrufe), and '17th Vienna Deep Learning Meetup (part 2)' (195 views). The channel navigation bar includes ÜBERSICHT, VIDEOS, PLAYLISTS, KANÄLE, DISKUSSION, and KANALINFO.

[www.youtube.com/ViennaDeepLearningMeetup](https://www.youtube.com/ViennaDeepLearningMeetup)

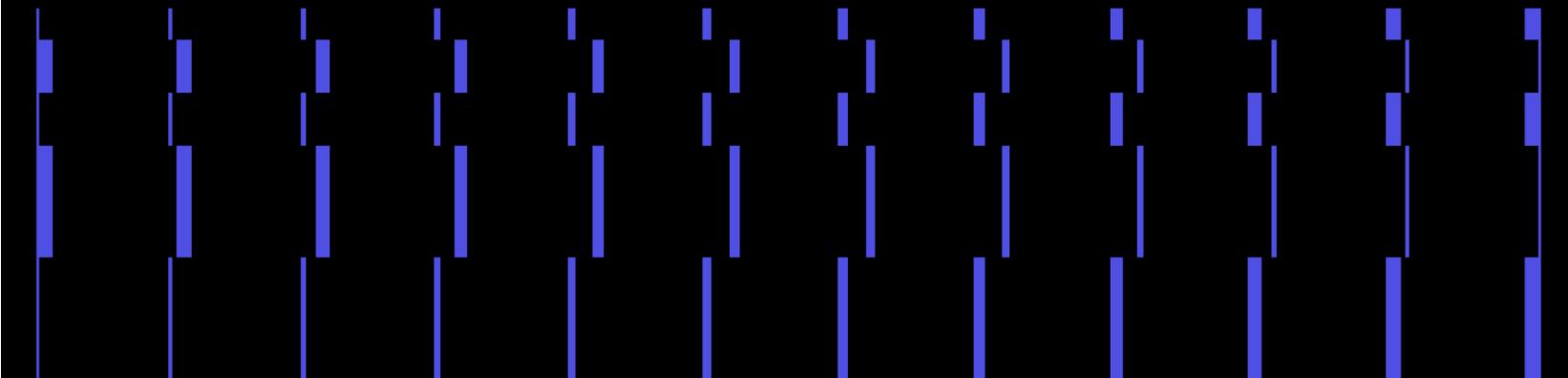
# INTERNATIONAL DIGITAL SECURITY FORUM

INTERNATIONAL  
DIGITAL  
SECURITY  
FORUM  
VIENNA

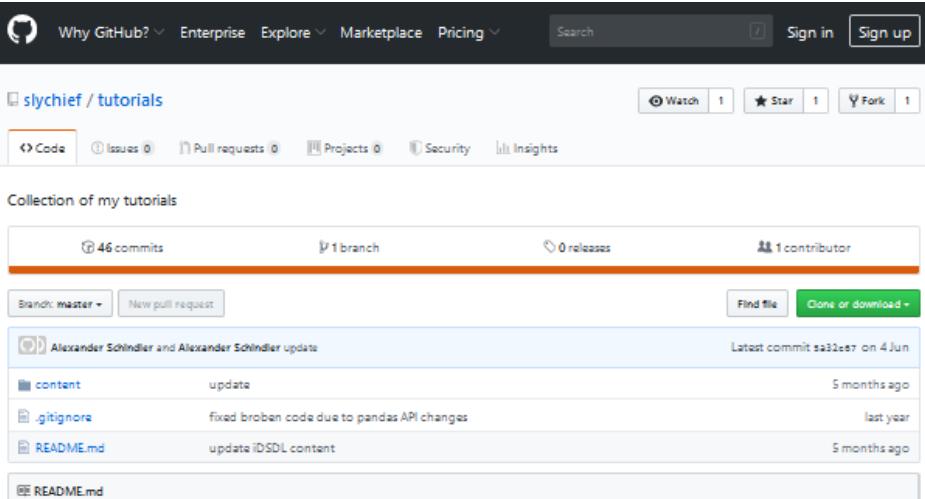
2.-4. December 2020

[www.idsfo.io](http://www.idsfo.io)

Organised by



# SLIDES OF THIS TALK



Why GitHub? Enterprise Explore Marketplace Pricing

Search Sign in Sign up

slychief / tutorials

Code Issues Pull requests Projects Security Insights

46 commits 1 branch 0 releases 1 contributor

Alexander Schindler and Alexander Schindler update

Latest commit sa32ce7 on 4 Jun

.content update 5 months ago

.gitignore fixed broken code due to pandas API changes last year

README.md update iDSDL content 5 months ago

README.md

## Presentations, Tutorials and Teaching materials

(c) 2019 by Alexander Schindler



Alexander Schindler, AIT Austrian Institute of Technology

Alexander Schindler researches audio-visual aspects of music information. He is currently employed as a researcher with the Vienna University of Technology and as scientist at the Austrian Institute of Technology (AIT). He has specialized on applying deep learning methods to analyze music related information in visual media such as music videos or album-art images, as well as sound event detection and audio similarity estimations. His research interests include information retrieval, specifically audio and video retrieval, image processing and machine learning with a focus on deep neural networks.

Website: <http://ifs.tuwien.ac.at/~schindler>

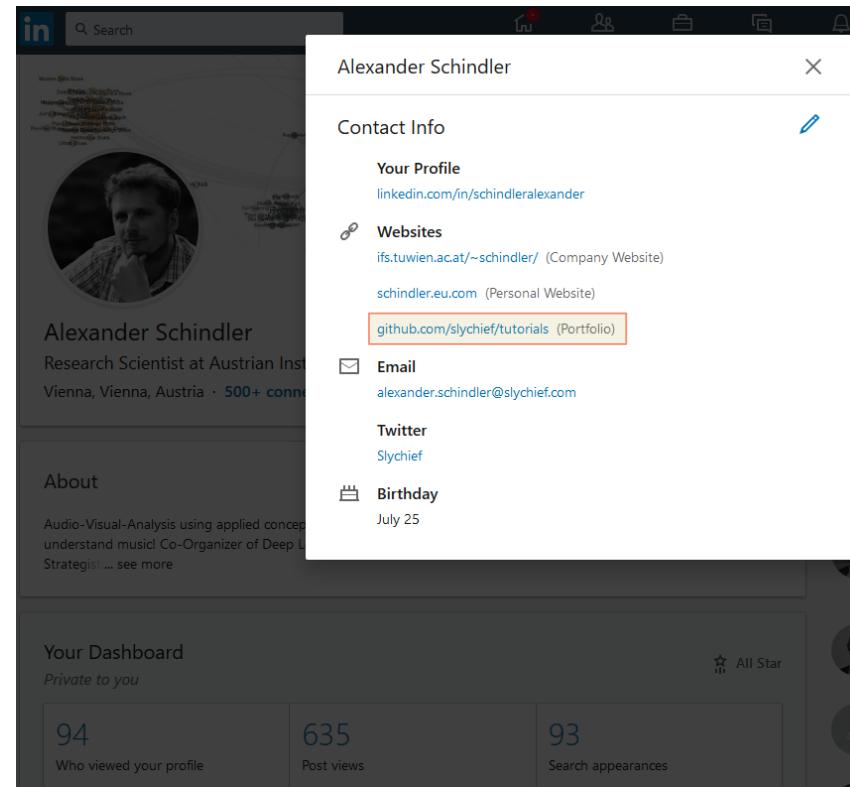
Twitter: <https://twitter.com/Slychief>

LinkedIn: <https://www.linkedin.com/in/schindleralexander>

## Overview

Date	Event	Description	Link
2019-05-24	Innovationslehrgang Data Science und Deep Learning (iDSDL)	Deep Learning Application Examples	<a href="#">more</a>

- Further Questions?
- Connect via LinkedIn!



Alexander Schindler

Contact Info

Your Profile [linkedin.com/in/schindleralexander](https://linkedin.com/in/schindleralexander)

Websites [ifs.tuwien.ac.at/~schindler/](https://ifs.tuwien.ac.at/~schindler/) (Company Website) [schindler.eu.com](http://schindler.eu.com) (Personal Website) [github.com/slychief/tutorials](https://github.com/slychief/tutorials) (Portfolio)

Email [alexander.schindler@slychief.com](mailto:alexander.schindler@slychief.com)

Twitter Slychief

Birthday July 25

Your Dashboard Private to you All Star

94 Who viewed your profile 635 Post views 93 Search appearances

# THANK YOU!



AUSTRIAN INSTITUTE  
OF TECHNOLOGY

## ALEXANDER SCHINDLER

Scientist

Information Management

Center for Digital Safety & Security

AIT Austrian Institute of Technology GmbH  
Donau-City-Straße 1 | 1220 Wien  
T +43 50550-2902 | M +43 664 8251454 | F +43 50550-2813  
[alexander.schindler@ait.ac.at](mailto:alexander.schindler@ait.ac.at) | [www.ait.ac.at](http://www.ait.ac.at)