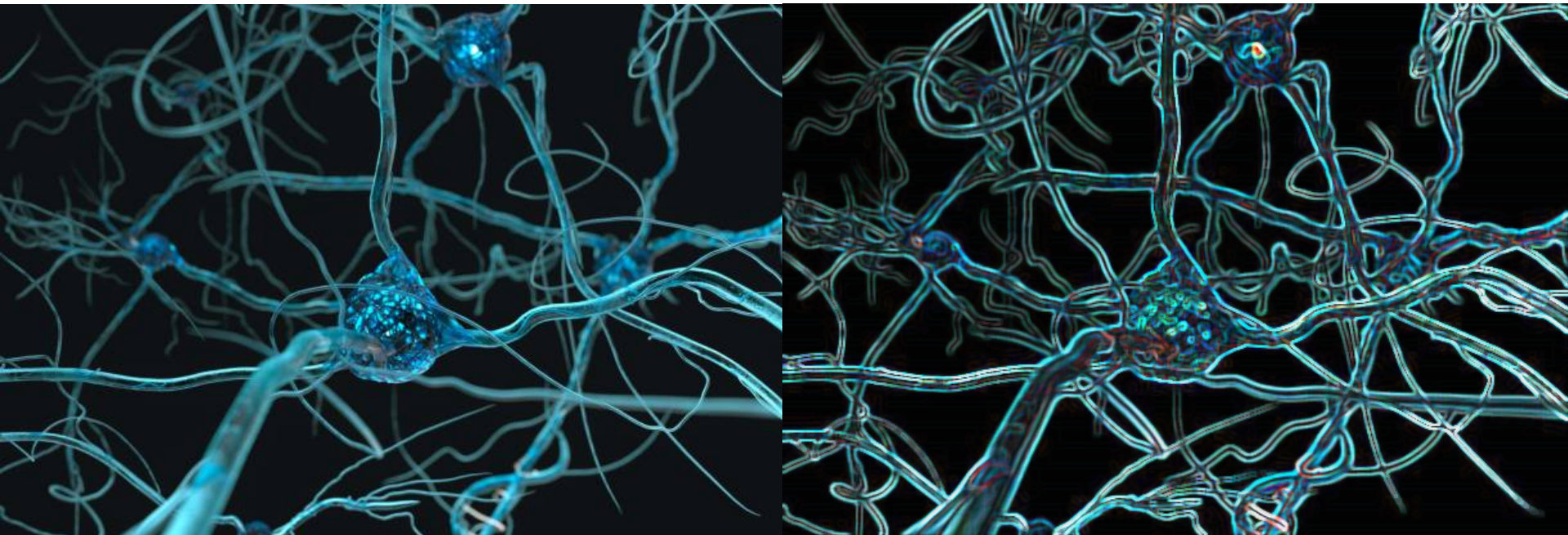


## Deep Neural Networks for Video Coding: Recent Advances and Future Perspectives

Dr. Dan Grois, E-mail: [grois@ieee.org](mailto:grois@ieee.org)



[Gettyimages]

# Agenda

---

**Part I:** Introduction to Deep Neural Networks

**Part II:** Deep Neural Networks for Video Coding

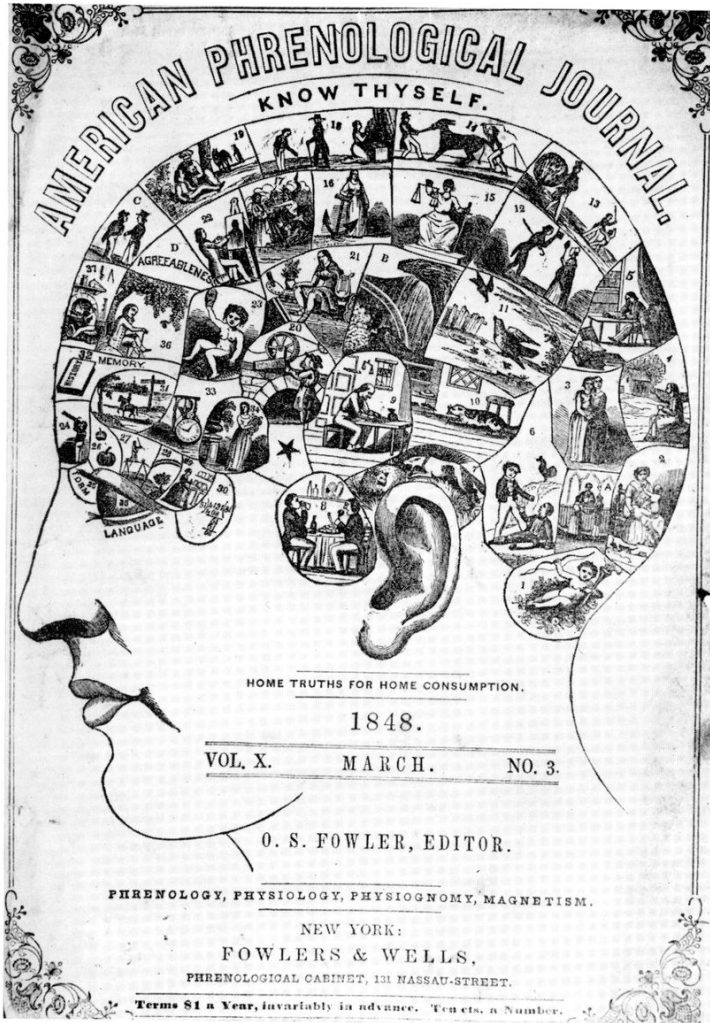
**Part III:** Overview and Future Perspectives

# Part I

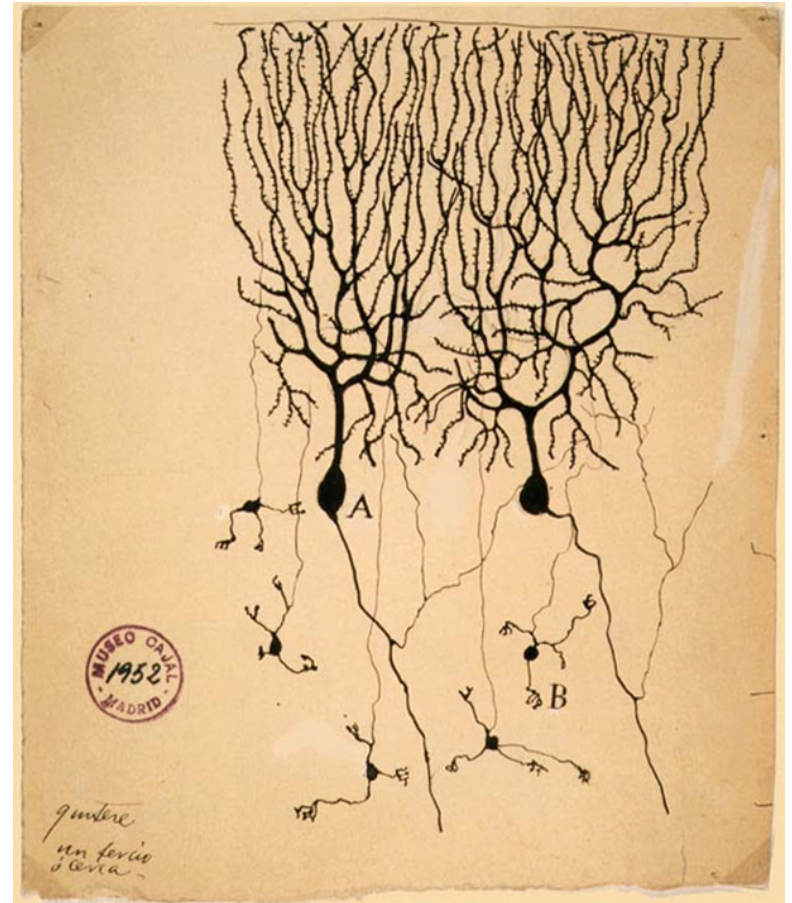
---

## Introduction to Deep Neural Networks (DNNs)

# Biological Neuron



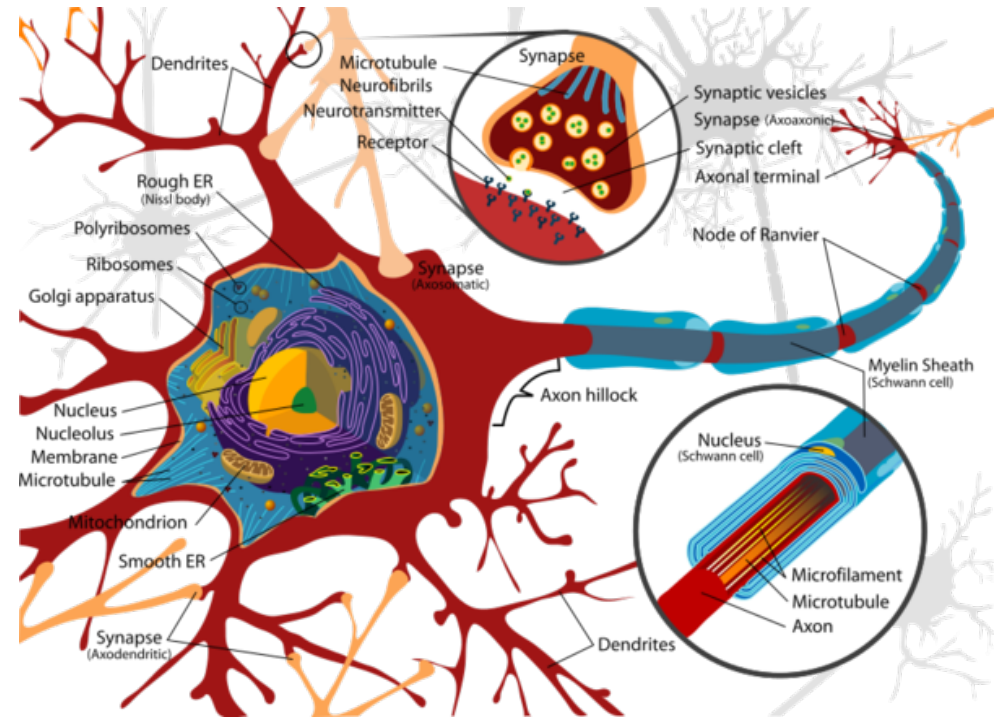
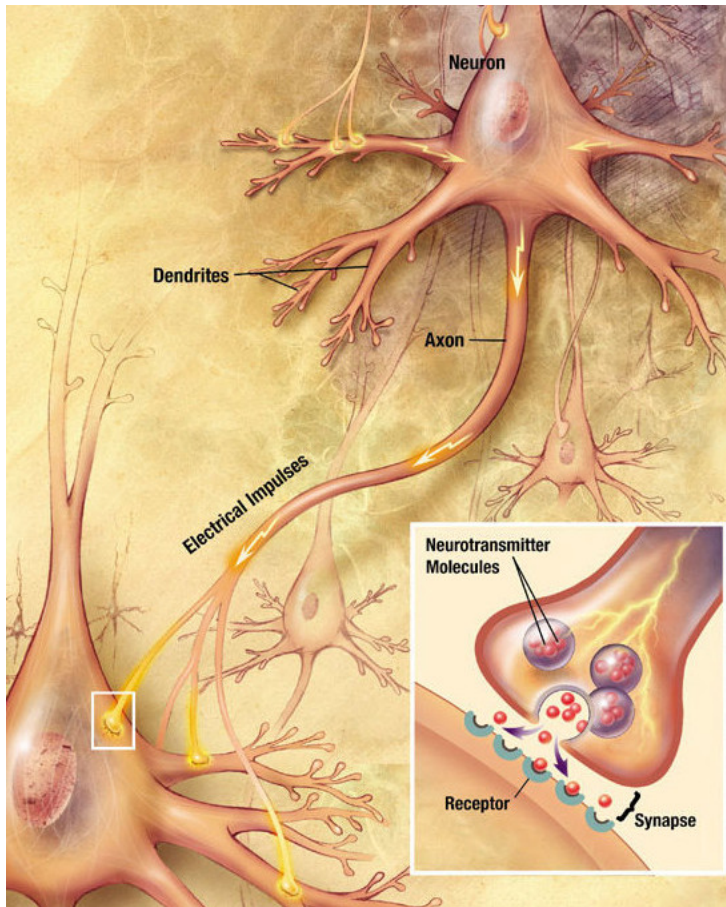
Drawing of neurons in the pigeon cerebellum by Spanish neuroscientist Santiago Ramón y Cajal, 1899.



[<http://en.wikipedia.org>]

[<http://en.wikipedia.org>]

# Biological Neuron (Cont.)



A signal propagating down an axon to the cell body and dendrites of the next cell.

[<http://en.wikipedia.org>]

[LadyOfHats]

# Biological Neuron (Cont.)



# Artificial Neuron: Perceptron

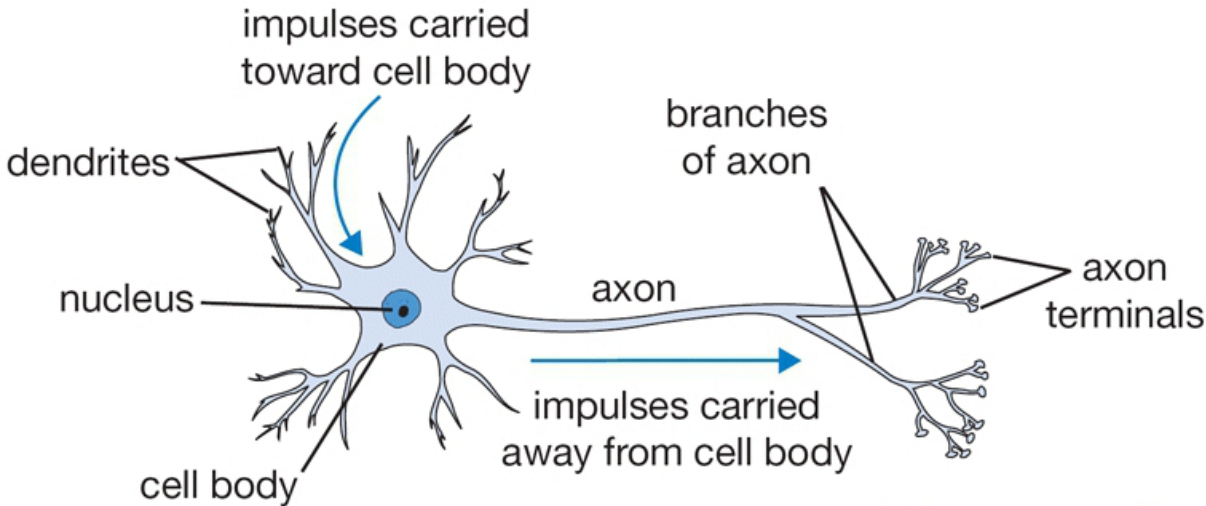
The perceptron's algorithm was invented by Frank Rosenblatt at the Cornell Aeronautical Laboratory, in 1957.



- The **perceptron** is an algorithm for supervised learning of binary classifiers, i.e. it is **a simple binary classification algorithm**.
- A binary classifier is a function which can decide whether or not an input, represented by a vector of numbers, **belongs to some specific class**.
- A perceptron allows to divide input signals into **“Yes” and “No”**.

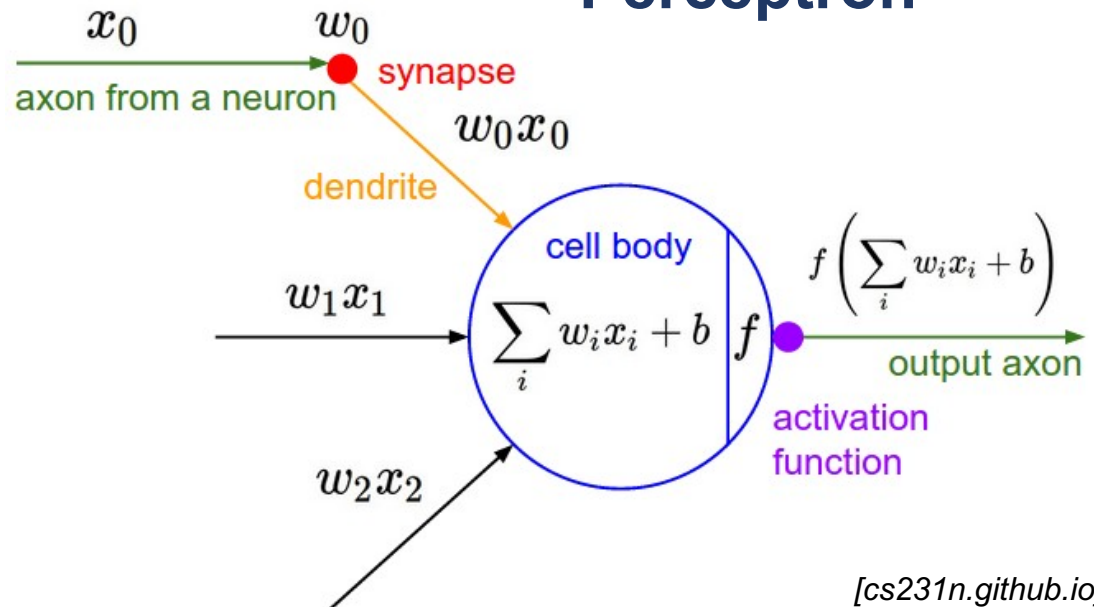
[<http://en.wikipedia.org>]

# Biological vs. Perceptron Structure



**Biological Neuron**

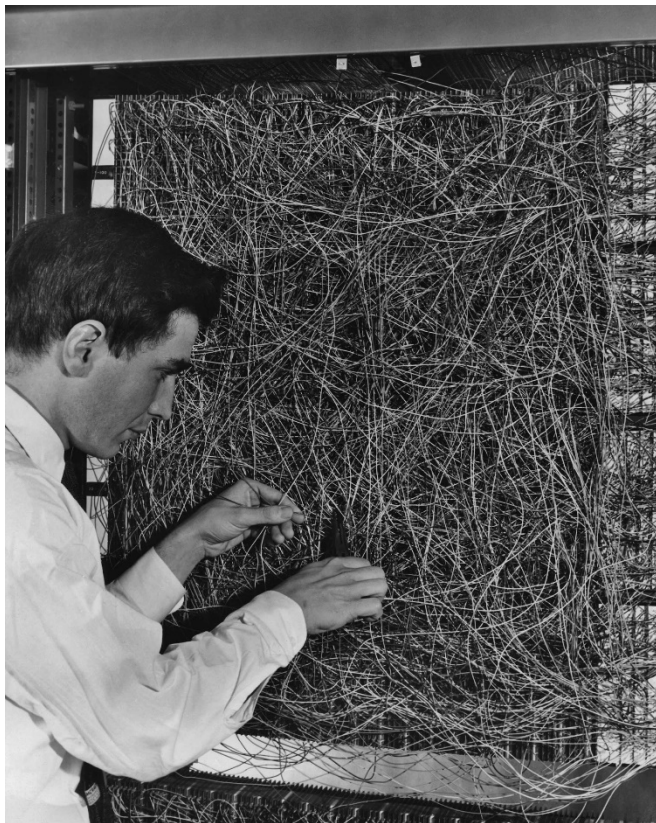
## Artificial Neuron/ Perceptron



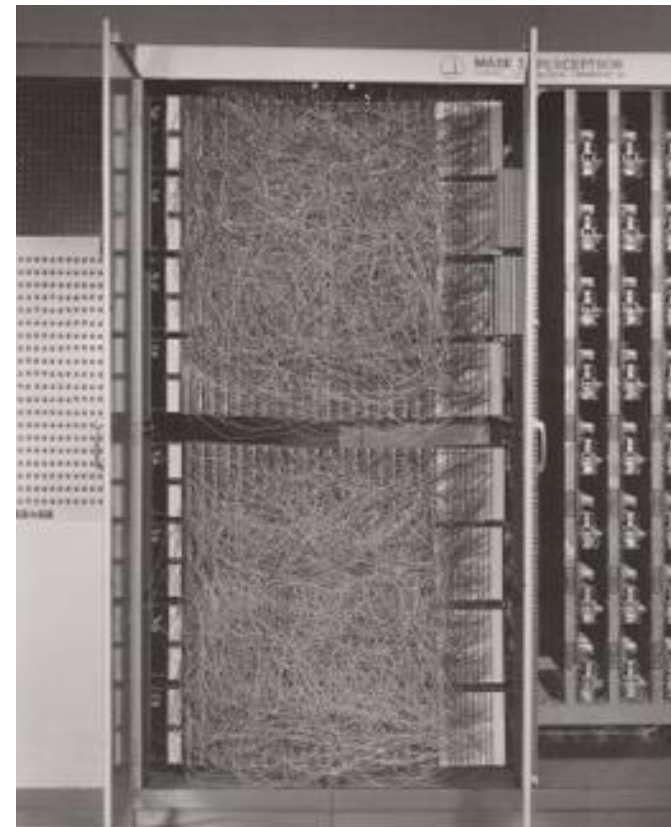
[cs231n.github.io]

# Artificial Neuron: Perceptron (Cont.)

Mark 1 Perceptron Machine Was Designed for **Image Recognition**



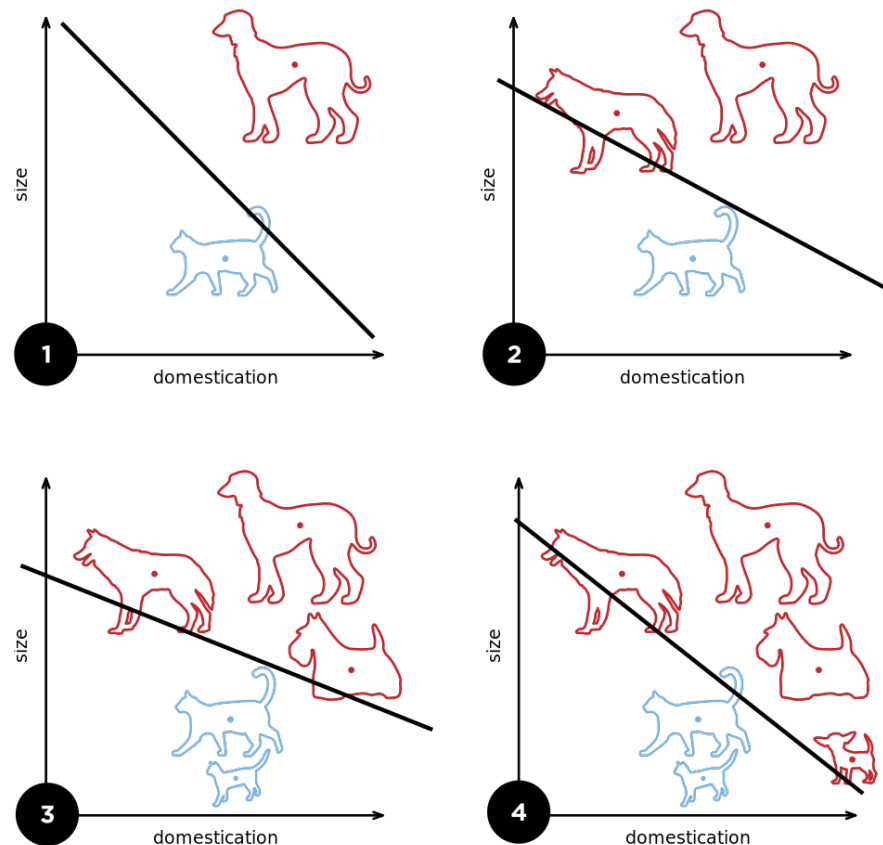
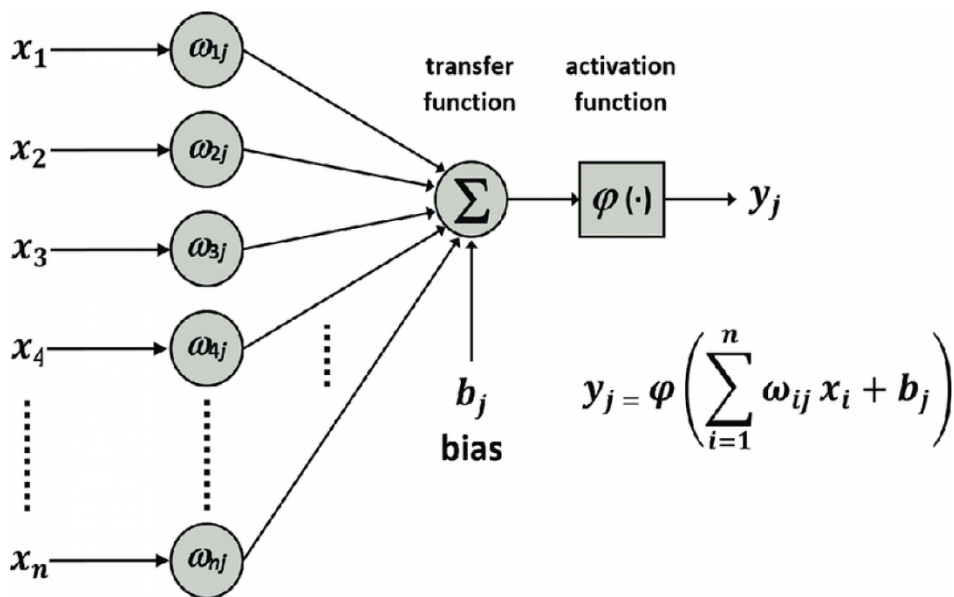
- it had an array of **400 photocells**, randomly connected to the **"neurons"**;
- **weights** were encoded in **potentiometers**;
- **weight updates** during learning were performed by **electric motors**.



Fine-tuning the Perceptron in Buffalo around 1960.

*[Frederic Lewis/Getty Images, <http://en.wikipedia.org>]*

# The Perceptron Updates its Linear Boundary

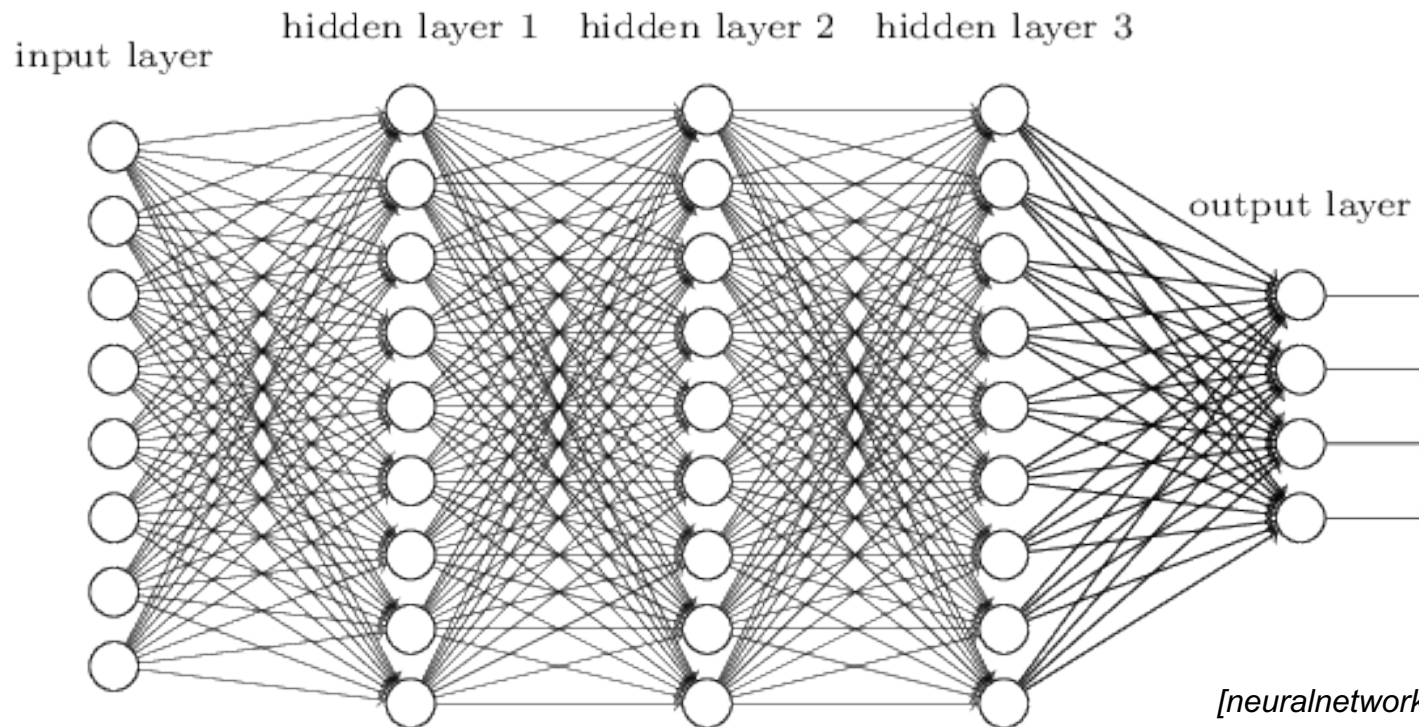


[Álvarez2017]

[<http://en.wikipedia.org>]

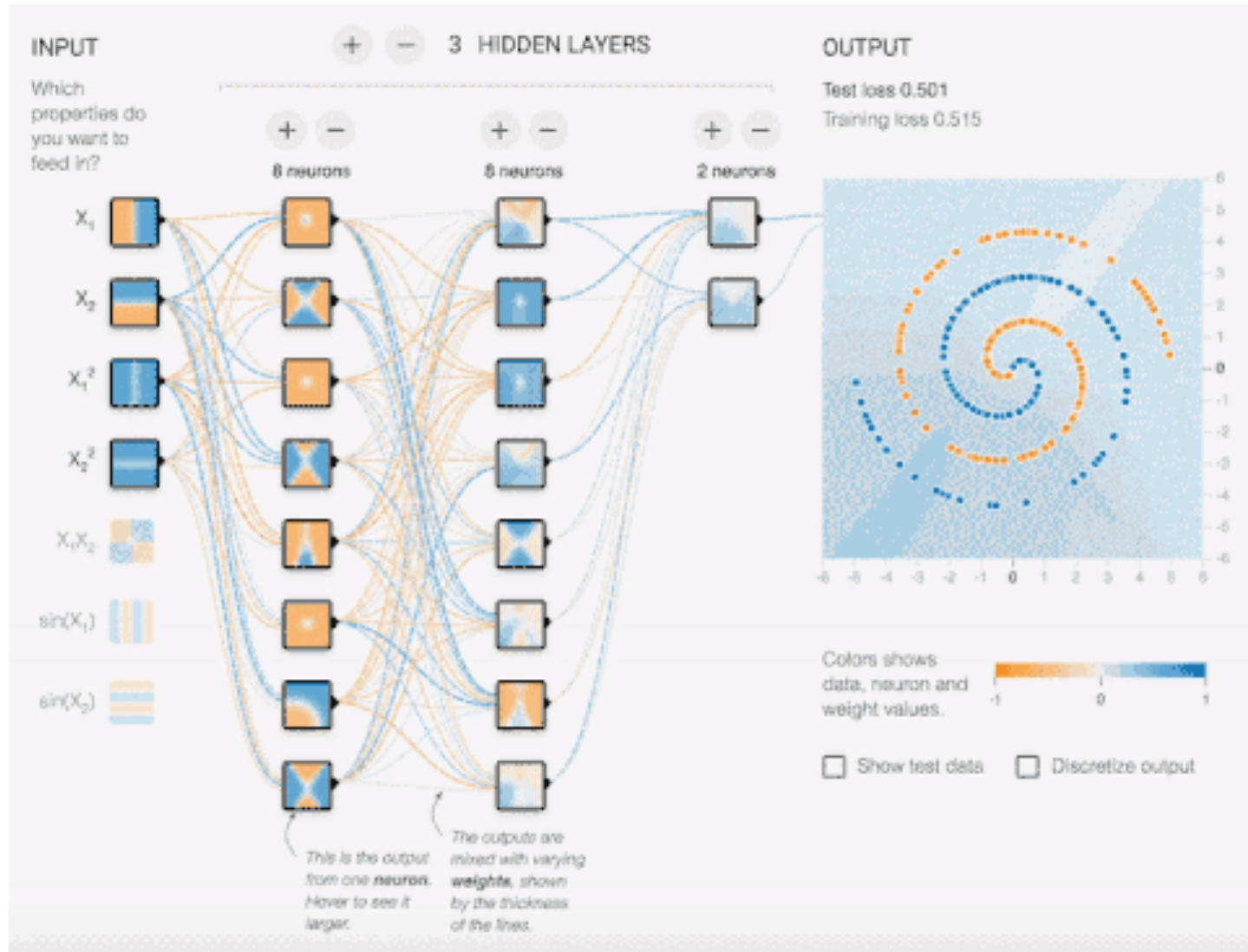
# Multi-Layer Perceptron

- Perceptrons **could not be trained** to recognize many classes of patterns;
- As a result, the neural network research was stuck for many years;
- Finally it was determined that **a feedforward neural network with two or more layers (i.e. a multi-layer perceptron)** has much greater processing power.



[neuralnetworksanddeeplearning.com]

# Sample Process of Neural Network Training



[gifer.com]

# Neural Network Zoo

## A mostly complete chart of Neural Networks

©2019 Fjodor van Veen & Stefan Leijnen [asimovinstitute.org](http://asimovinstitute.org)

- Input Cell
- Backfed Input Cell
- △ Noisy Input Cell
- Hidden Cell
- Probabilistic Hidden Cell
- △ Spiking Hidden Cell
- Capsule Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- △ Gated Memory Cell
- Kernel
- Convolution or Pool

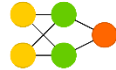
Perceptron (P)



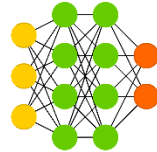
Feed Forward (FF)



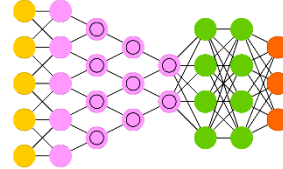
Radial Basis Network (RBF)



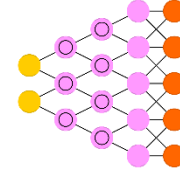
Deep Feed Forward (DFF)



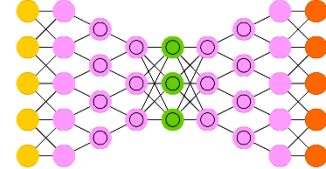
Deep Convolutional Network (DCN)



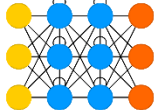
Deconvolutional Network (DN)



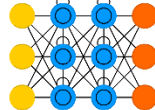
Deep Convolutional Inverse Graphics Network (DCIGN)



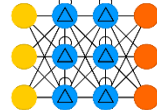
Recurrent Neural Network (RNN)



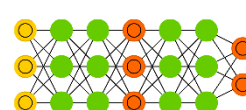
Long / Short Term Memory (LSTM)



Gated Recurrent Unit (GRU)



Generative Adversarial Network (GAN)



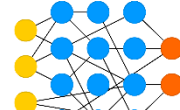
Liquid State Machine (LSM)



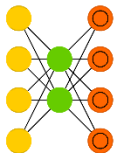
Extreme Learning Machine (ELM)



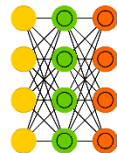
Echo State Network (ESN)



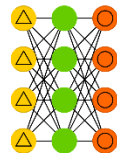
Auto Encoder (AE)



Variational AE (VAE)



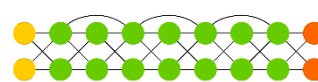
Denosing AE (DAE)



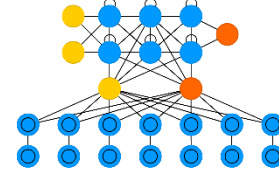
Sparse AE (SAE)



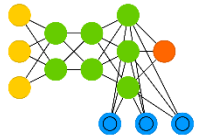
Deep Residual Network (DRN)



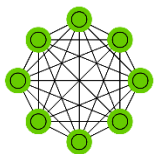
Differentiable Neural Computer (DNC)



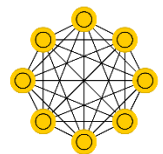
Neural Turing Machine (NTM)



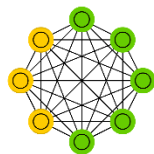
Markov Chain (MC)



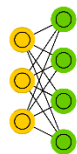
Hopfield Network (HN)



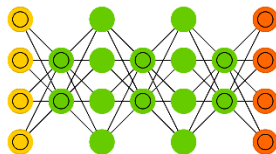
Boltzmann Machine (BM)



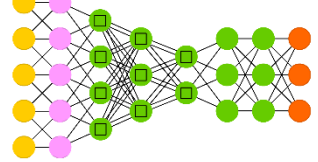
Restricted BM (RBM)



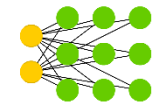
Deep Belief Network (DBN)



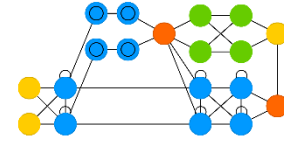
Capsule Network (CN)



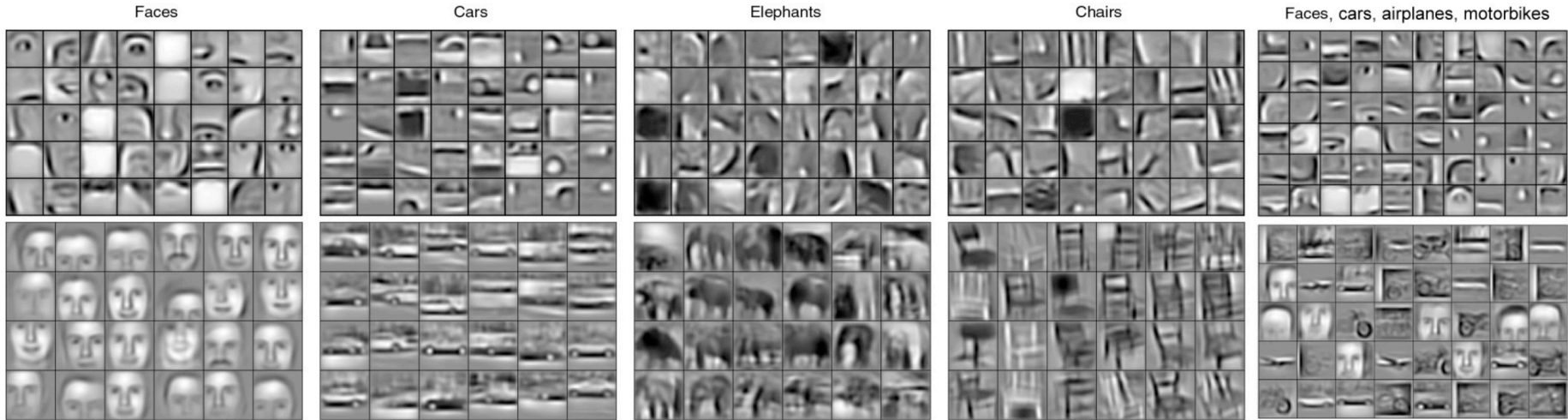
Kohonen Network (KN)



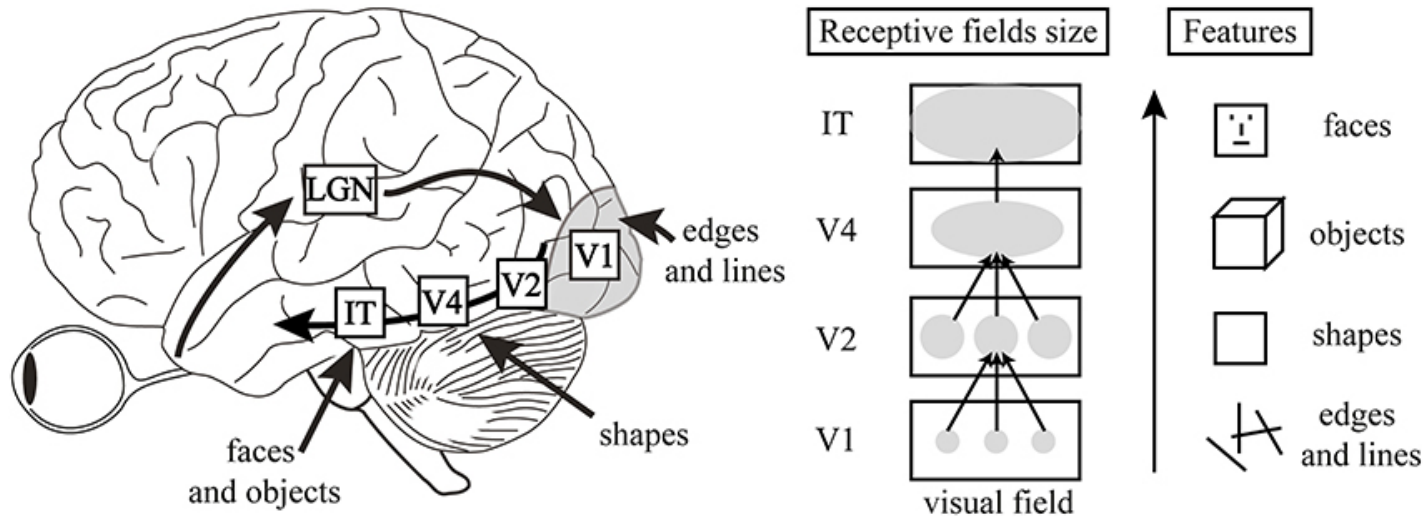
Attention Network (AN)



# Layered Feature Representations

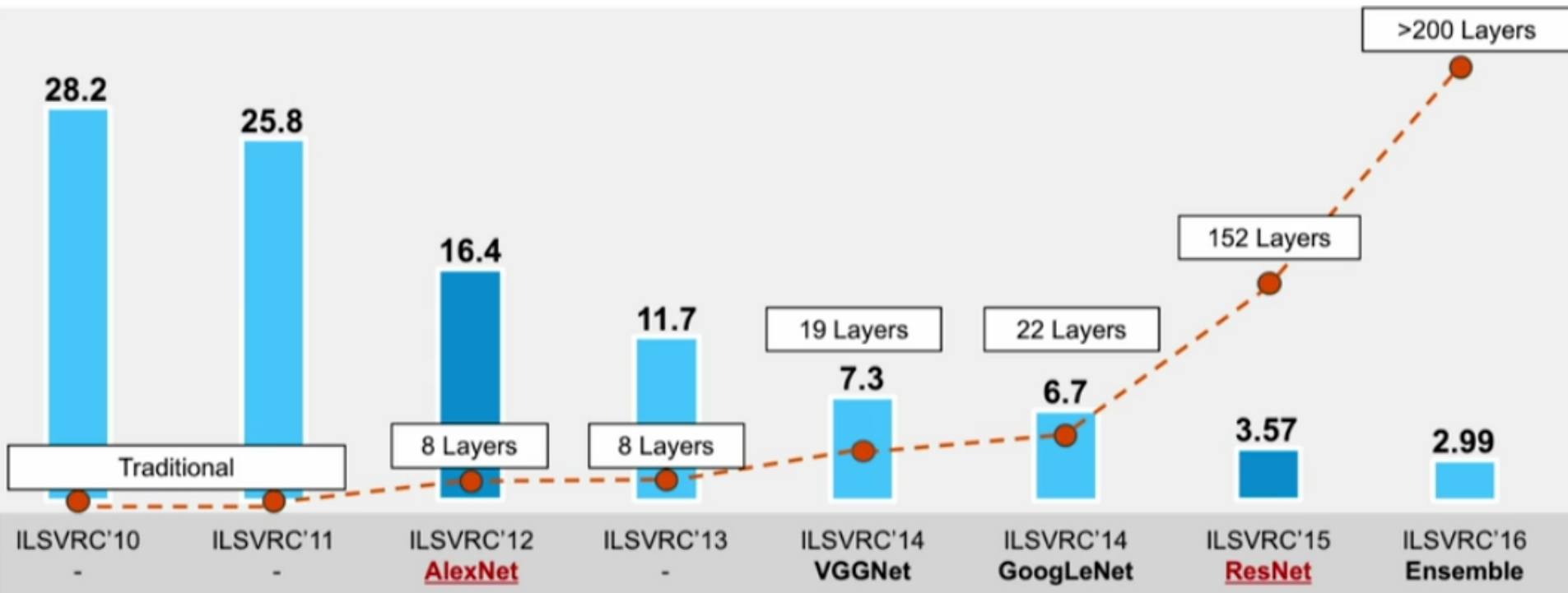


[Lee2011]



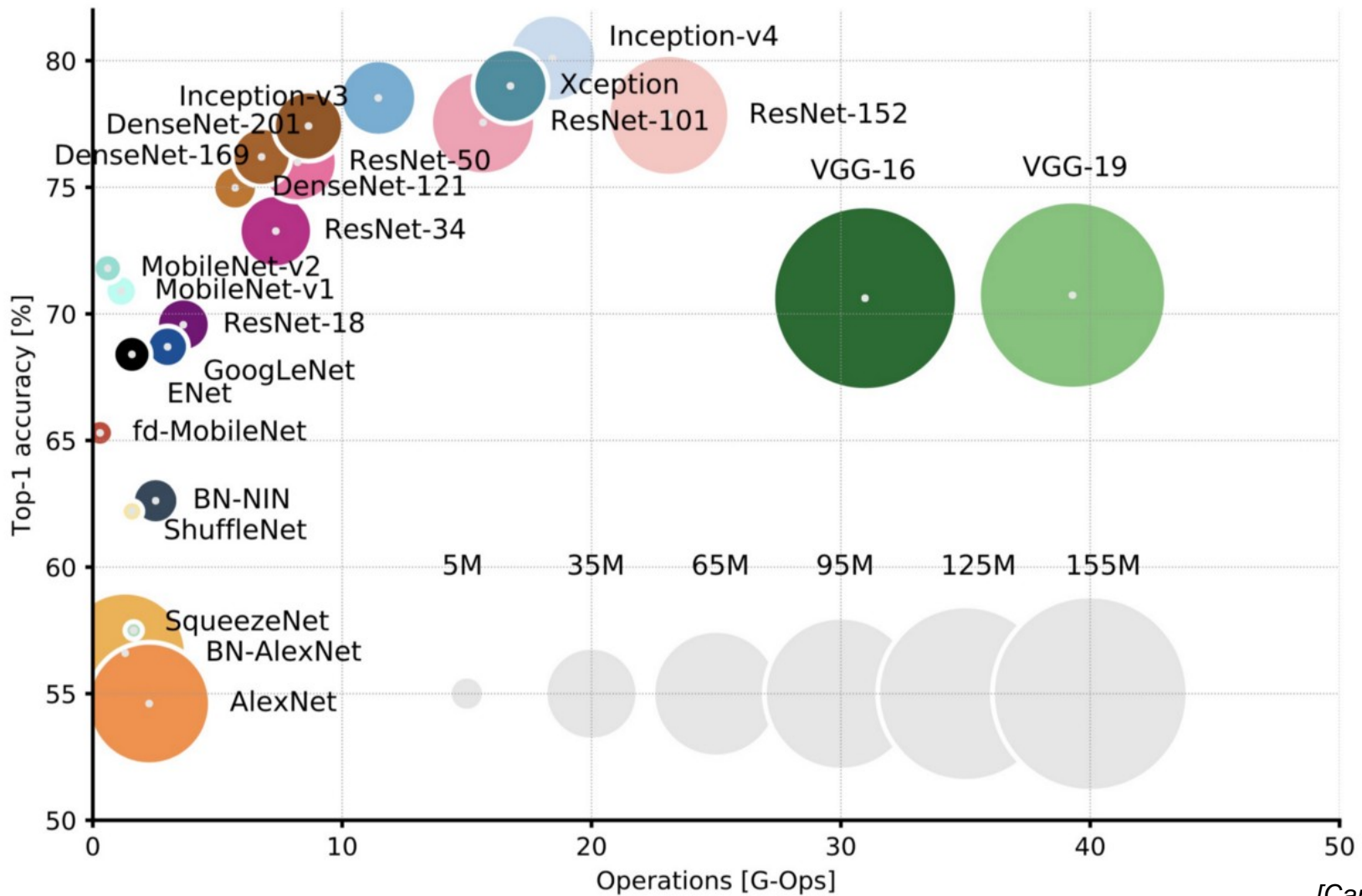
[<https://neurdivness.wordpress.com/2018/05/17/deep-convolutional-neural-networks-as-models-of-the-visual-system-qa/>]

# Error Rate



[SADEK, sqlml.azurewebsites.net]

# Accuracy vs. G-Ops



[Canziani2016]

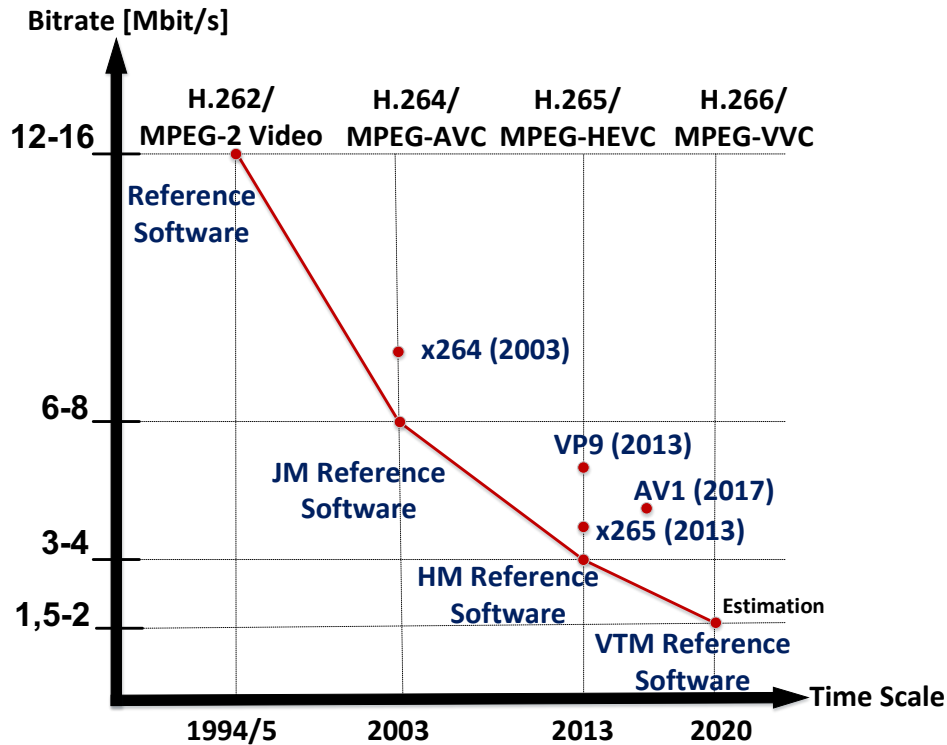
# Part II

---

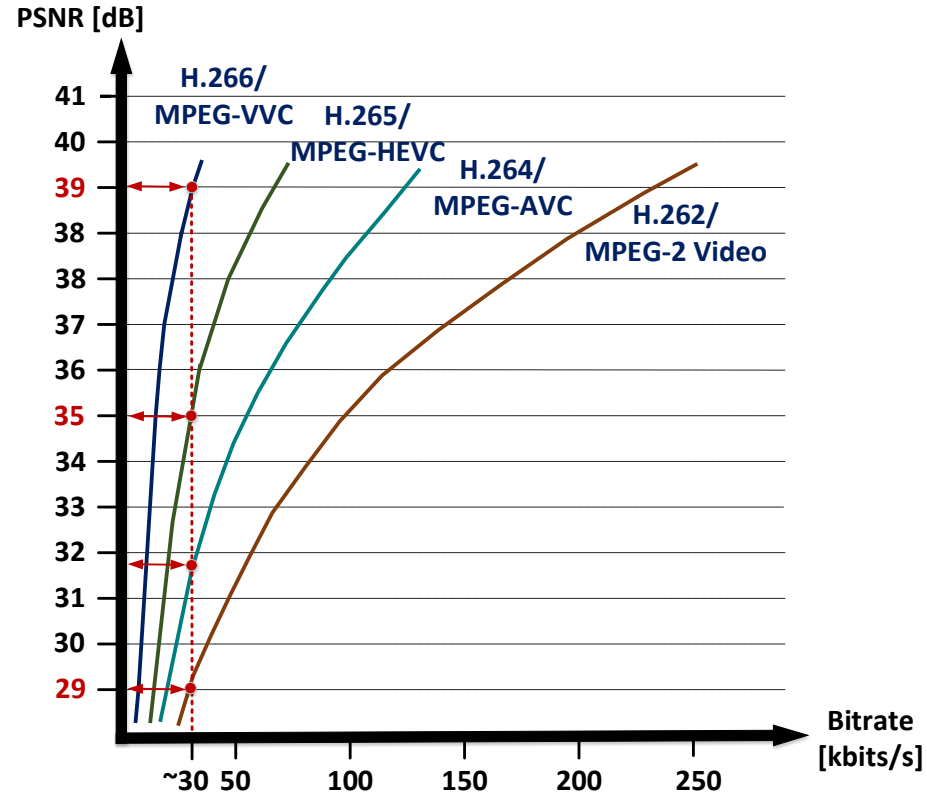
## Deep Neural Networks for Video Coding

# Video Compression

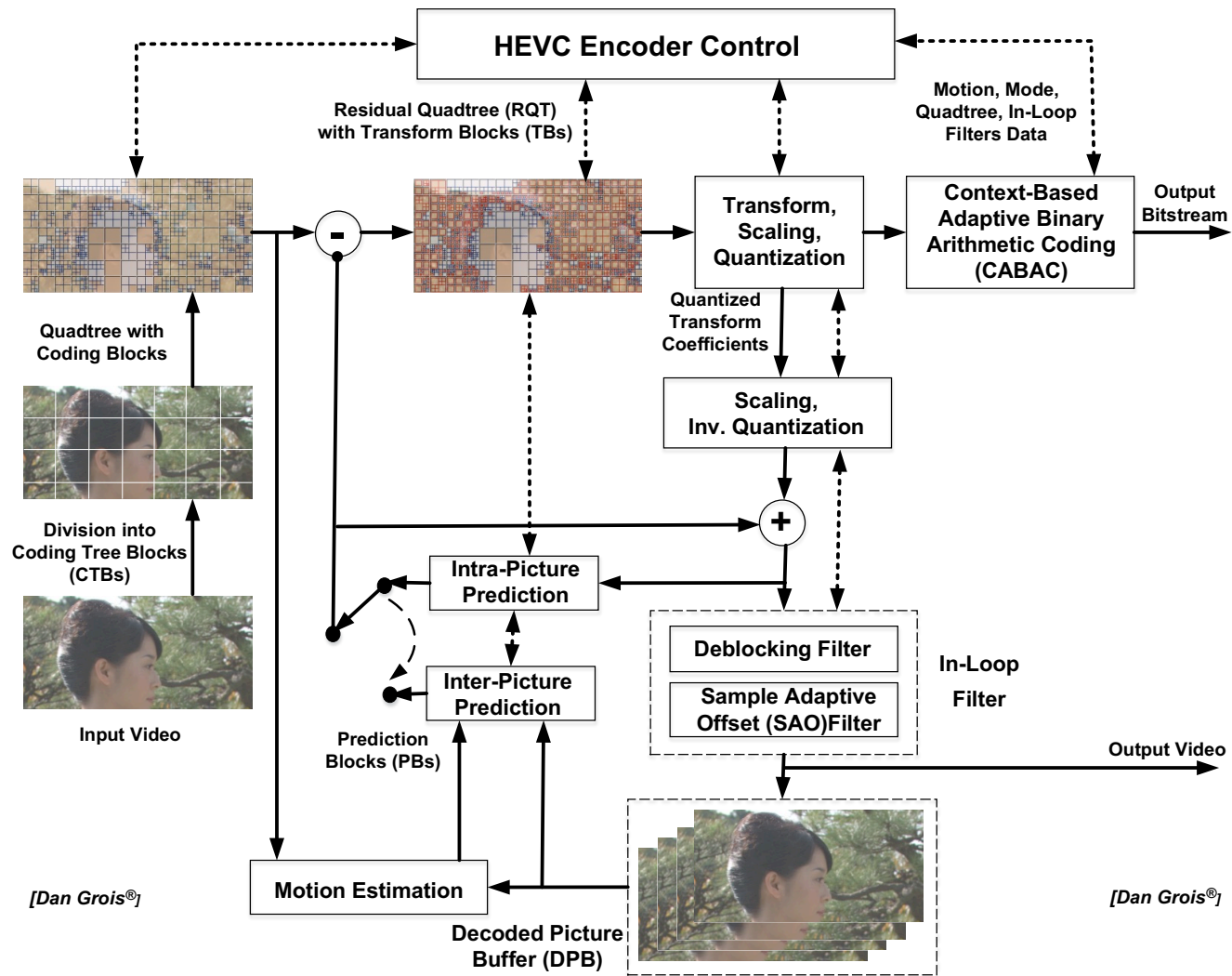
Performance Comparison for 1080p Video Content (SDR, high quality, 24-60 fp/s)



Most Popular MPEG and ITU Video Compression Standards (QCIF, 100 fr.)



# H.265/MPEG-HEVC Coding Loop



# What Can be Improved in Video Coding by DNNs?

- **Intra-Picture Prediction**
- **Inter-Picture Prediction**
- **In-Loop Filtering**
- **Transform**
- **Probability Distribution Prediction**
- **Cross-Component Prediction**
- **Down-Sampling/Up-Sampling**
- **Coding Optimization**



# Intra-Picture Prediction (Cont.)

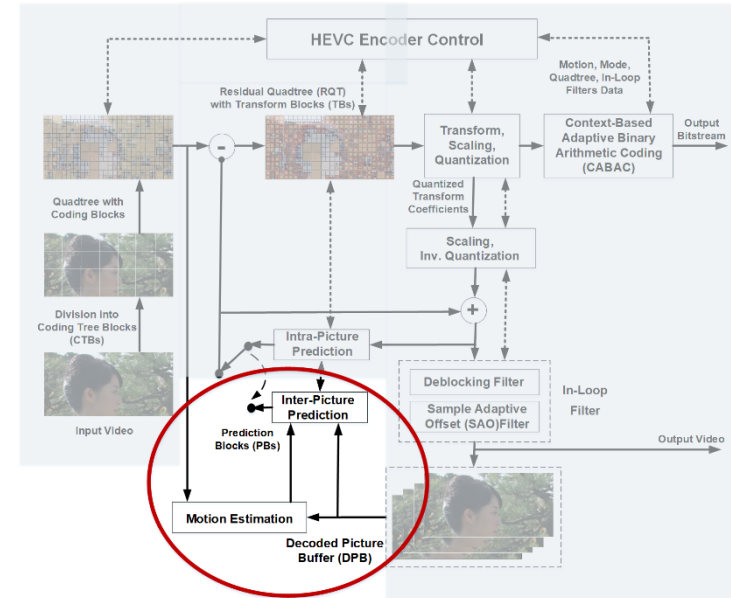
## Improvements by Utilizing DNNs

- The DNN **learns end-to-end mapping** of the reconstructed reference pixels to the current block pixels in order **to handle complex textures**;
- DNN exploits the current block content based on several neighboring blocks – i.e. **utilizing far-located structure data**;
- DNN's outputs the **predicted likelihood of different intra-picture prediction modes**;
- Further, DNN **learns coded prediction residuals** to optimize performance and increase coding efficiency;
- **Currently up to ~11% BD-rate reduction compared to HM**;
- DNN can be further applied at the post-processing stage for artifact reduction in HEVC intra coding.

# Inter-Picture Prediction

## Motivation

- Inter-picture prediction allows to predict **between pictures to remove temporal redundancy**;
- Applying motion estimation (ME) and motion compensation (MC) techniques:
  - ME allows to determine a reference frame block having **the most similar content to a block to be coded**, and
  - MC allows to retrieve such content.



# Inter-Picture Prediction (Cont.)

## Improvements by Utilizing DNNs

- **Extrapolating multiple reference frames** using a trained DNN model;
- **Enhancing bi-prediction** by applying DNN for improving rate-distortion performance;
- **Learning end-to-end mapping** between compensated coding blocks and PUs;
- **Generating interpolation filters** by means of DNN;
- Fractional-pixel Motion Compensation (MC) for improving inter-picture prediction efficiency;
- Providing DNN for **MC luma and chroma components refinement**;
- Learning and **predicting variations between half-/quarter-pixels and integer-position pixels**;
- Providing a single DNN model for generating all sub-pixels;
- **Currently up to ~4% BD-rate reduction compared to HM.**

# In-Loop Filtering

## Motivation

- Quantization process leads to loss of information



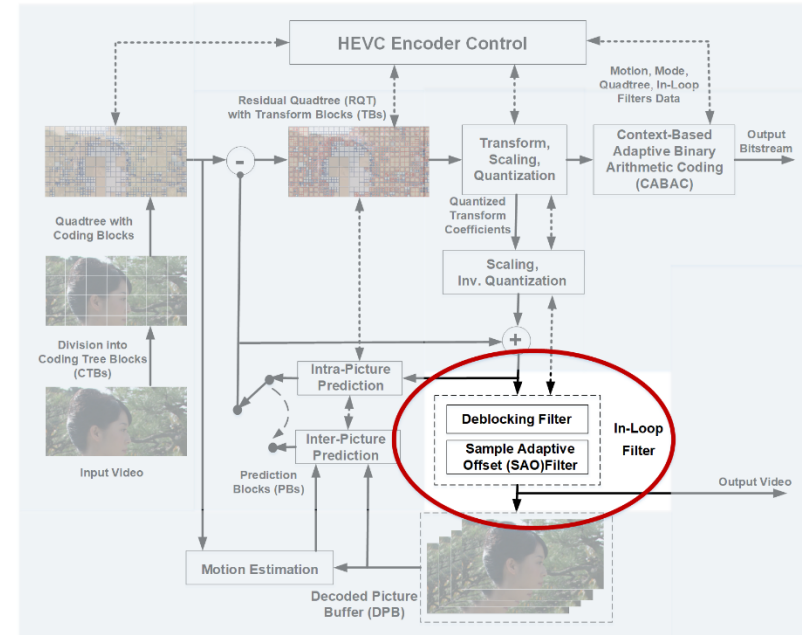
**the reconstructed video has usually lower visual quality than the original one;**

- When this loss is too high



the reconstructed video can have **visible artifacts, such as blockiness, ringing, blurring,** and the like;

- In order to remove such artifacts, in-loop filtering is carried out, which in turn **improves visual quality of the reconstructed pictures.**



# In-Loop Filtering (Cont.)

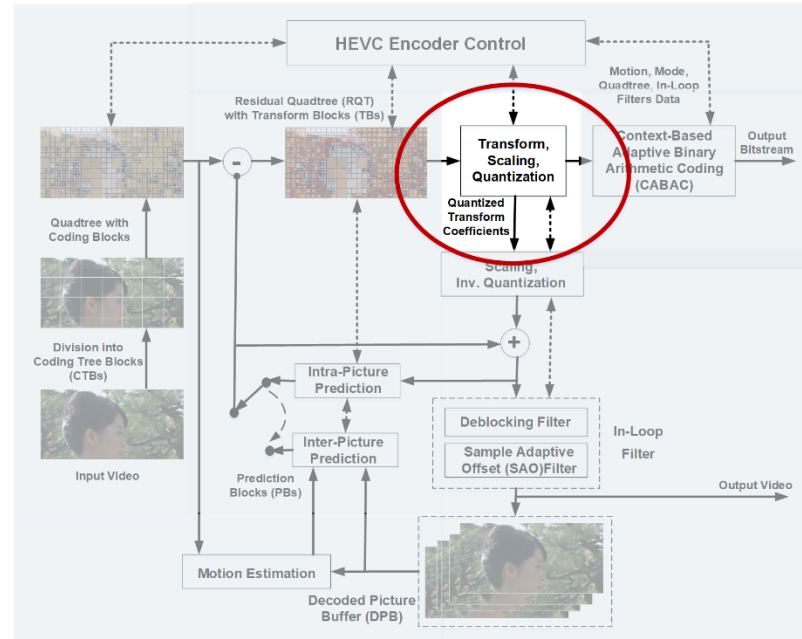
## Improvements by Utilizing DNNs

- **DNN learns distortion features** for intra-picture and inter-picture coding;
- DNN loop filter **adapts to decoded frames with different qualities** and ensure consistency;
- DNN can **replace either a Deblocking filter or Sample Adaptive Offset (SAO) filter, or both;**
- For improving performance, the **QP range can be split into several bands** to train each DNN separately;
- By means of DNNs, **learning non-linear mappings** between distorted/quantized reconstructed frames and corresponding original frames;
- **Currently up to ~10% BD-rate reduction compared to HM;**
- There is a trade-off between the optimal quality enhancement and overall computational complexity.

# Transform

## Motivation

- It is used for decorrelating pixel values by **converting coding block pixel residuals into transform coefficients** in a frequency domain for the further quantization and coding.
- Usually, a **Discrete Cosine Transform (DCT)** or its variation is used.
- In addition, multiple transform can be applied, such as a **primary and secondary transforms**.



# Transform (Cont.)

## Improvements by Utilizing DNNs

- DNN can be used for transform coding – e.g., **to fully replace DCT**;
- Achieving **better compression efficiency than DCT** by applying a DNN-based transform;
- The **transform, quantization, and inverse transform can be jointly trained** to achieve the overall rate-distortion optimization;
- **Currently up to ~9% BD-rate reduction compared to DCT 32x32.**



# Probability Distribution Prediction (Cont.)

## Improvements by Utilizing DNNs

- Estimating probability distribution of:
  - **quantized transform coefficients** (particularly, a DC coefficient of a block from both content and AC coefficients of this block);
  - **intra prediction modes** based on the context.
- **Currently up to ~2% BD-rate reduction compared to HM.**

# Cross-Component Prediction

## Motivation

- There are **dependences between chroma and luma components**:
  - it is desired to efficiently predict chroma components (U, V) from a luma component Y;
- A linear model is usually utilized while its coefficients are estimated from the context by **applying linear regression instead of transmitting them directly**.

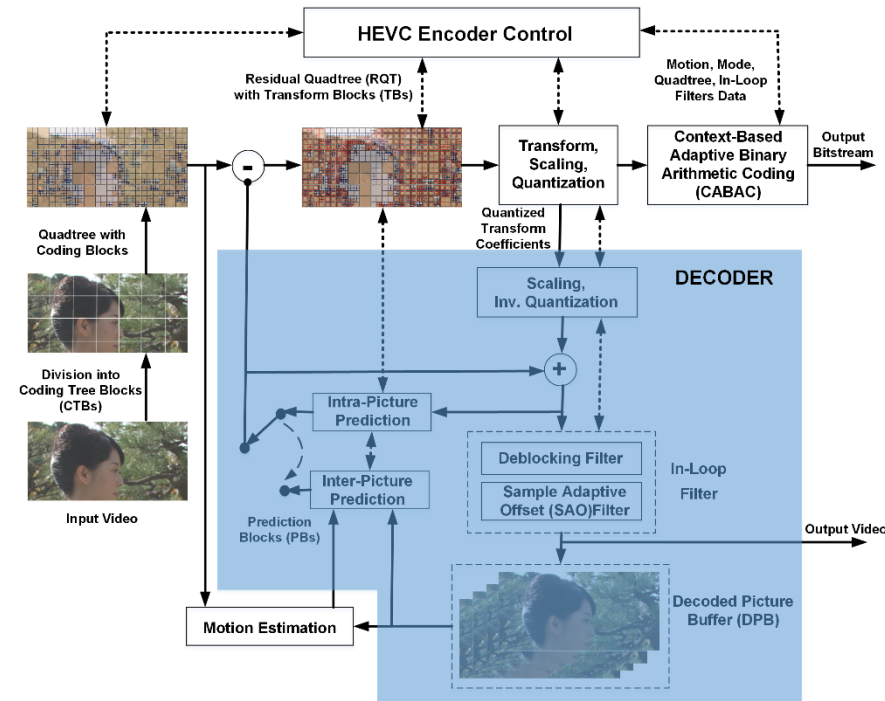
## Improvements by Utilizing DNNs

- The neural network (e.g., CNN, fully connected NN) **extracts features** from the reconstructed luma samples as well as **from neighboring reconstructed luma and chroma samples** for predicting chroma samples in the current block.
- **Currently up to ~1% BD-rate reduction compared to HM.**

# Up-Sampling/Down-Sampling

## Motivation

- Strong demand for **high resolution video**, such as 4K/8K for various applications such as:
  - Home and Digital Cinema;
  - Broadcasting;
  - Gaming;
  - Streaming;
  - VoD;
  - many others...
- However, the transmission bandwidth is limited, so **the resolution can be decreased prior to encoding and increased back after the decoding process** in order to improve the coding efficiency.



# Up-Sampling/Down-Sampling (Cont.)

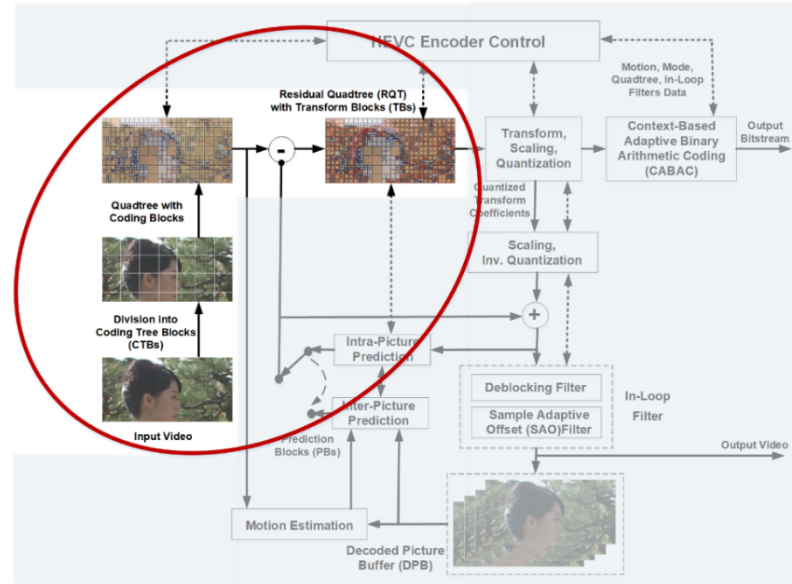
## Improvements by Utilizing DNNs

- DNNs can very efficiently handle **complex image regions**;
- **Down-sampling and up-sampling at different domains**: spatial, temporal, bi-depth **by applying trained convolutional neural networks (DNNs)**:
  - The **DNN learns a difference between the original and degraded block**, which was obtained due to down-sampling – generally, this difference is the **high frequency information** lost during the down-sampling process;
  - Performing spatial up-sampling by means of a DNN during the encoding, and **reconstructing a full-resolution video at the decoder end**;
  - Using different DNNs for **up-sampling luma and chroma components**;
  - Using DNNs for determining either up-sampling filter coefficients, or for **both up-sampling and down-sampling filter coefficients**.
- **Currently up to ~15% BD-rate reduction compared to HM**; however, there is a significant increase in computational complexity.

# Coding Optimization

## Motivation

- The recent and the most popular video codecs, such as H.264/MPEG-AVC, H.265/MPEG-HEVC contain **a very significant number of all possible coding modes**;
- It is desirable to **assign each coding block with an (different) optimal coding mode(s)**.
- For determining optimal coding parameters, usually significant computational resources are used due to:
  - **checking all possible coding modes and coding partitions** based on rate-distortion cost function.



# Coding Optimization (Cont.)

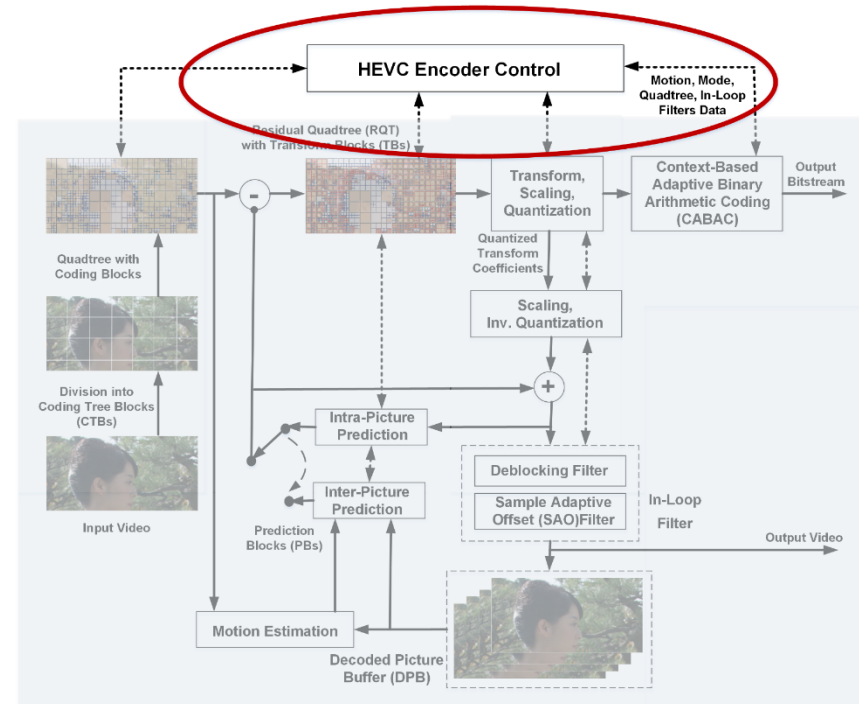
## Improvements by Utilizing DNNs

- **Learning optimal CU/PU partitioning** of the entire CTU;
- Obtaining fast CU/PU mode decisions by **applying early terminated DNN**;
- The partitioning data can be further used to effectively **applying DNN-based quality enhancement process**;
- By using DNN for predicting coding parameters, **computational complexity is significantly decreased** by eliminating the rate-distortion cost calculations;
- **Currently offering up to ~50% complexity reduction with a negligible increase in the BD rate.**

# Coding Optimization (Cont.)

## Motivation

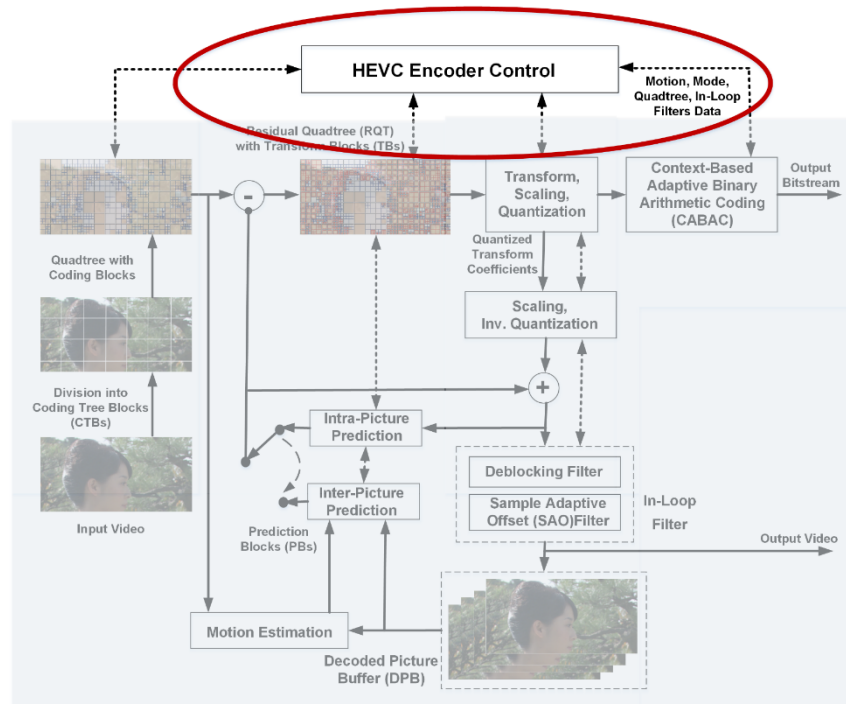
- A transmission **bandwidth is usually limited**;
- Encoder outputs a bitstream in a controlled manner by **utilizing a rate control scheme** in order not to exceed the maximal bit-rate;
- **Frame-level bit allocation** is an important step in the bit-rate control scheme, thereby **allocating bits to each video frame**, further considering inter-frame dependences;
- **Intra-frame rate control** is especially important since:
  - **Intra frames consume a lot of bits**;
  - **Intra-frames quality has a direct impact on coding efficiency** due to the inter-picture prediction.



# Coding Optimization (Cont.)

## Improvements by Utilizing DNNs

- A neural network can be trained to **determine a number of target bits for each GOP/frame/CTU** based on the overall number of bits as well as on the inter-frame coding complexity;



- By using DNNs, **more stable rate control** is provided with a **reduced error on the GOP/frame/CTU-level** with some visual quality improvements;

# Part III

---

## Overview and Future Perspectives of Deep Neural Networks

# Overview and Future Perspectives of Deep Neural Networks

- DNNs video coding parameters are obtained **based on real and practical data**, compared to current hybrid-based codecs based on predefined tools and algorithms, **thereby providing more accurate prediction and larger coding gains**;
- DNN-based video coding **is not limited to pixel values/features of neighbor coding blocks** in the current frame or in several previous frames, compared to state-of-the-art video coding standards;
- DNNs are the best choice for applications requiring **both computer vision/machine analysis and human observation**, since DNNs can be optimized for **both textures and fine details/features**;
- Main reasons why DNN are still not widely deployed:
  - Significant memory requirements;
  - Large computational costs, including significant cost of GPUs.

- For better performance, deeper neural networks with **larger number of parameters and layers** should be considered:
  - Exploring parameters efficiency in terms of **both coding gain and computational complexity, including memory requirements**.
- Due to a variety of video content to be coded (e.g., sport, news, etc.), **utilizing several DNNs to be switched according to specific content** can further significantly improve coding gain.

# What is So Special About These People?



[Karas2019]

# Generating Fake Human Photos

These are not real people

[Karas2018]

# What is Next?

---

## Generating Fully Fake Videos

(i.e. not just changing lips movements)

**When?**

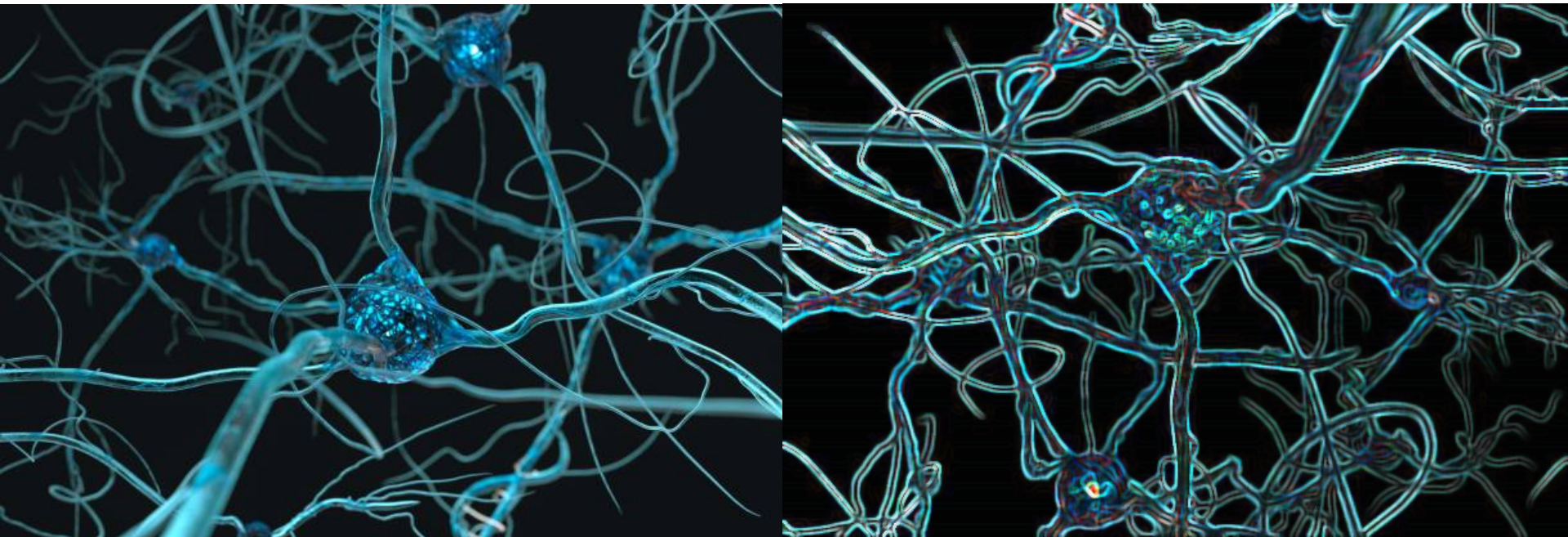
**...in just about 2-3 years from now.**

# References and Further Reading

- Álvarez2017* Daniel Álvarez, Ana Cerezo-Hernández, Graciela López-Muñiz, Tania Álvaro-De Castro, Tomás Ruiz-Albi, Roberto Hornero and Félix del Campo, "Usefulness of Artificial Neural Networks in the Diagnosis and Treatment of Sleep Apnea-Hypopnea Syndrome", IntechOpen, Apr. 2017.
- Canziani2016* Alfredo Canziani, Adam Paszke, Eugenio Culurciello, "An Analysis of Deep Neural Network Models for Practical Applications", CVPR 2017, arXiv:1605.07678.
- Karas2018* Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, " Progressive Growing of GANs for Improved Quality, Stability, and Variation", ICLR 2018, arXiv:1710.10196.
- Karas2019* Tero Karras, Samuli Laine, Timo Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks", CVPR 2019, arXiv:1812.04948.
- Lee2011* Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng, "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks", Commun. ACM 54, 10 (October 2011), 95-103.
- Liu2019* Dong Liu, Yue Li, Jianping Lin, Houqiang Li, Feng Wu, "Deep Learning-Based Video Coding: A Review and A Case Study", arXiv:1904.12462.
- Ma2019* Siwei Ma, Xinfeng Zhang, Chuanmin Jia, Zhenghui Zhao, Shiqi Wang and Shanshe Wang, "Image and Video Compression with Neural Networks: A Review," in IEEE Transactions on Circuits and Systems for Video Technology, Apr. 2019.

# Deep Neural Networks for Video Coding: Recent Advances and Future Perspectives

**Dr. Dan Grois**, [grois@ieee.org](mailto:grois@ieee.org)



[Gettyimages]

07.30.19

Mile High Video – Denver, CO 2019

 COMCAST