Pattern Recognition: Statistics to Deep Networks

Anil K. Jain

https://www.cse.msu.edu/~jain

Michigan State University

Beijing Academy of AI (BAAI) Annual Conference, June 21-23, 2020

Outline

- Beginning of AI
- Alphabet soup: AI, PR, NN, DM, DS, ML, DNN,…
- Statistics to deep networks
- Face recognition
- Privacy concerns
- Next decade of AI

Artificial Intelligence (AI)

- …. making a machine behave in ways that would be called intelligent if a human were so behaving.

McCarthy, Minsky, Rochester & Shannon, 1956

- Turing test (1951) , “imitation game”, tests if a computer can successfully pretend to be a human in a dialogue via screen & keyboard. *Dictionary.com*

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955, AI Magazine, Vol. 27(4), 2006

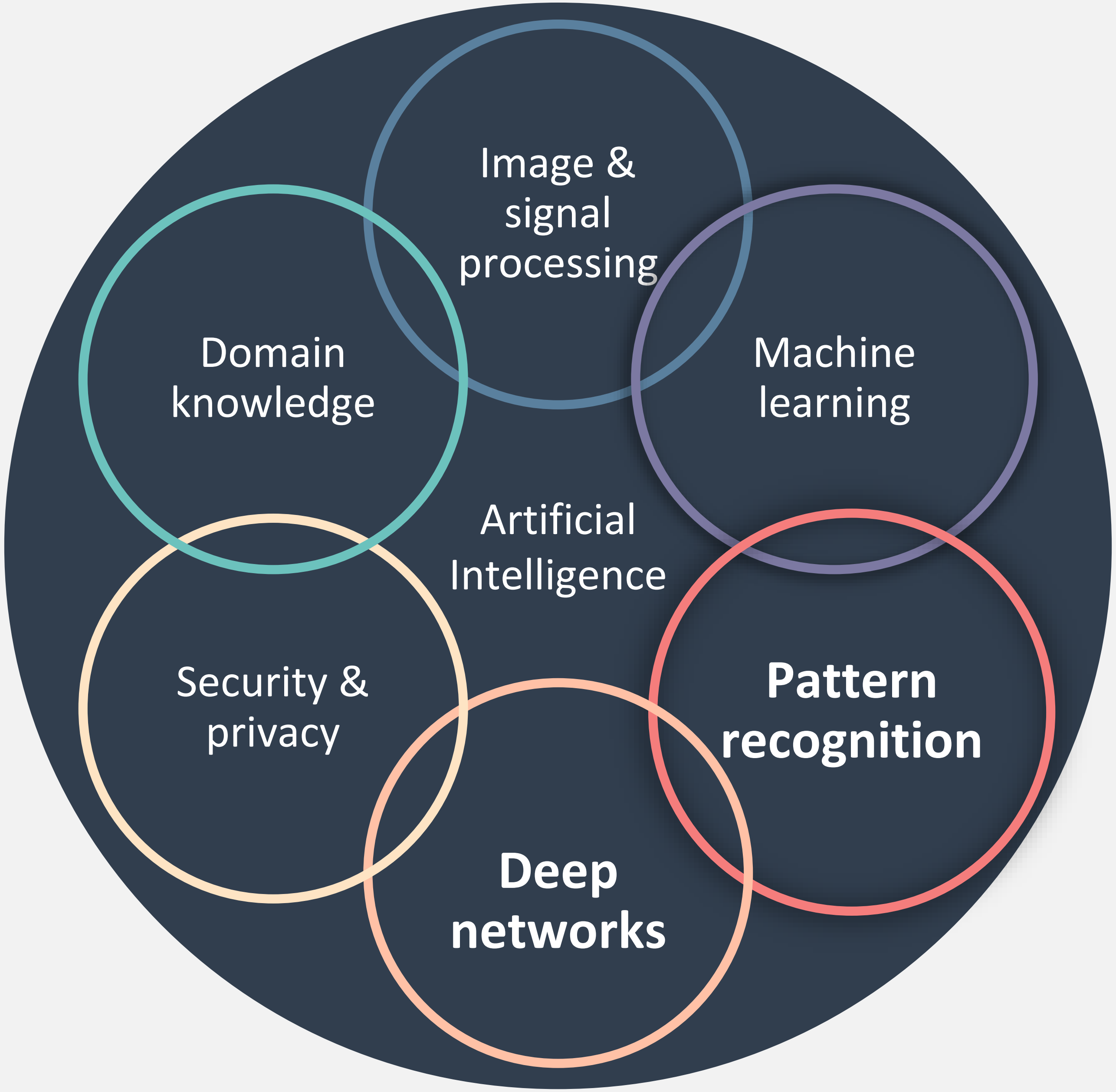
Pattern Recognition

By pattern recognition we mean the extraction of the significant features from a background of irrelevant detail. … it is the kind of thing that brains seem to do very well….that computing machines do not do very well yet. **O. G. Selfridge, 1955**

AI: General-purpose intelligence; P.R.: Domain-specific intelligence

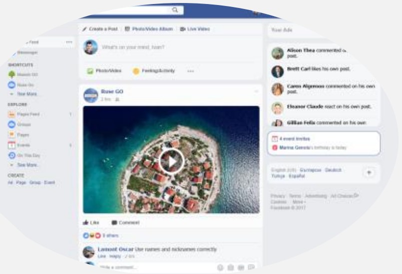
Selfridge, "Pattern recognition and modern computers." In Proceedings of the Western Joint Computer Conf, pp. 91-93. March 1-3, 1955.

Artificial Intelligence: Many Facets



Most-Influential Technologies

2006 — Facebook's News Feed



2007 — Netflix Streaming

2010 — Facebook's Instagram

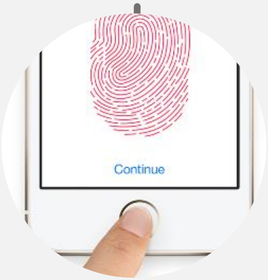
2011 — Uber

2012 — Tesla's Model S

2012 — Apple iPad

2013 — Ring's Doorbell

2013 — Apple TouchID



2014 — Amazon Alexa

2015 — Apple Watch

2017 — Apple FaceID

https://www.washingtonpost.com/technology/2019/12/26/we-picked-most-influential-technologies-decade-it-isnt-all-bad/

AI Hype

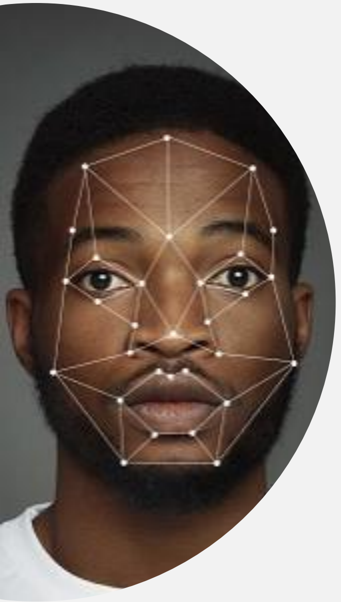
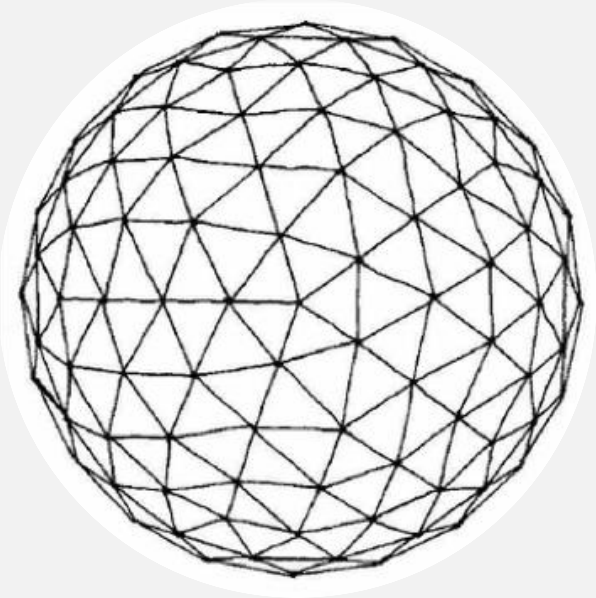
“Adjusts settings, based on the load, to provide the most optimized washing cycle.”

“We overestimated the arrival of autonomous vehicles.” - Ford CEO Jim Hackett

Hype surrounding AI has peaked & troughed over the years as the abilities of the technology get overestimated and then re-evaluated. . *bbc.com/news/technology-51064369*

http://www.lgnewsroom.com/2019/09/lg-washing-machines-with-artificial-intelligence-and-direct-drive-motor-roll-out-region-wide/

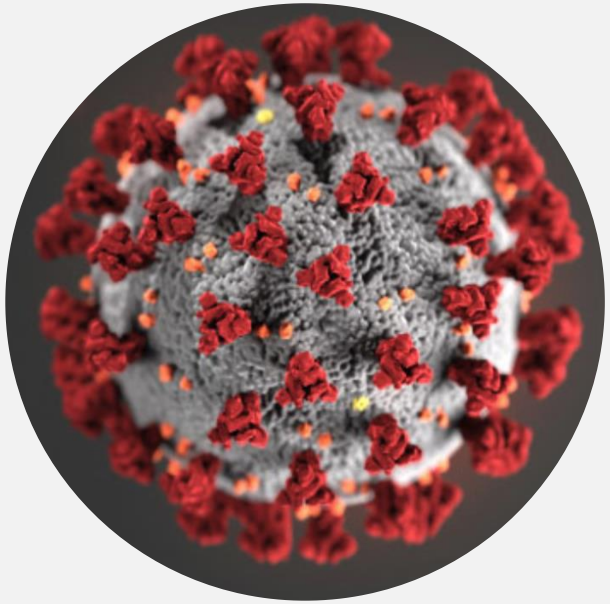
https://emerj.com/ai-adoption-timelines/self-driving-car-timeline-themselves-top-11-automakers/

What is a Pattern?

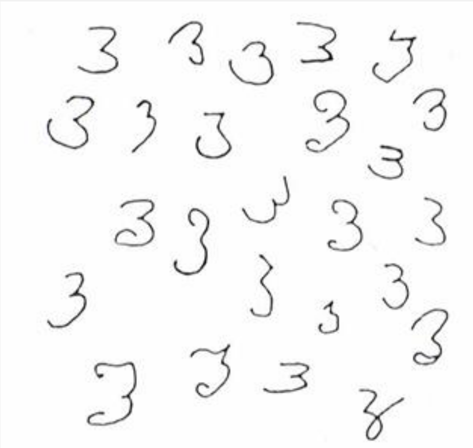
A pattern is the opposite of a chaos; it is an entity vaguely defined, that could be given a name.

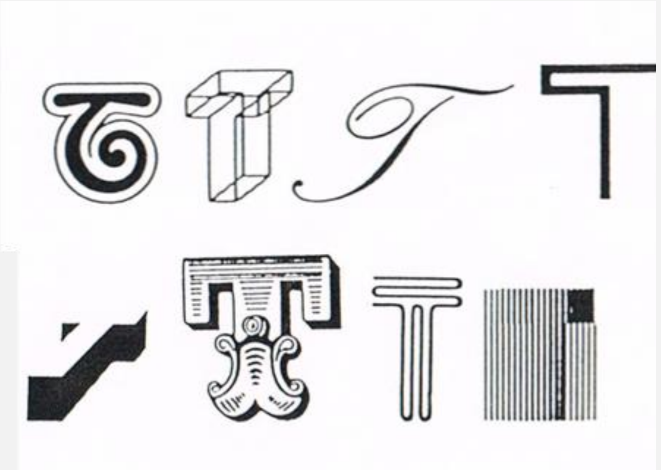
S. Watanabe, 1985




Pattern Class

- Collection of similar, not necessarily identical, patterns
- Class is defined by a model or examples
- How to define similarity, fundamental to intelligent systems





American Elm

Ginkgo

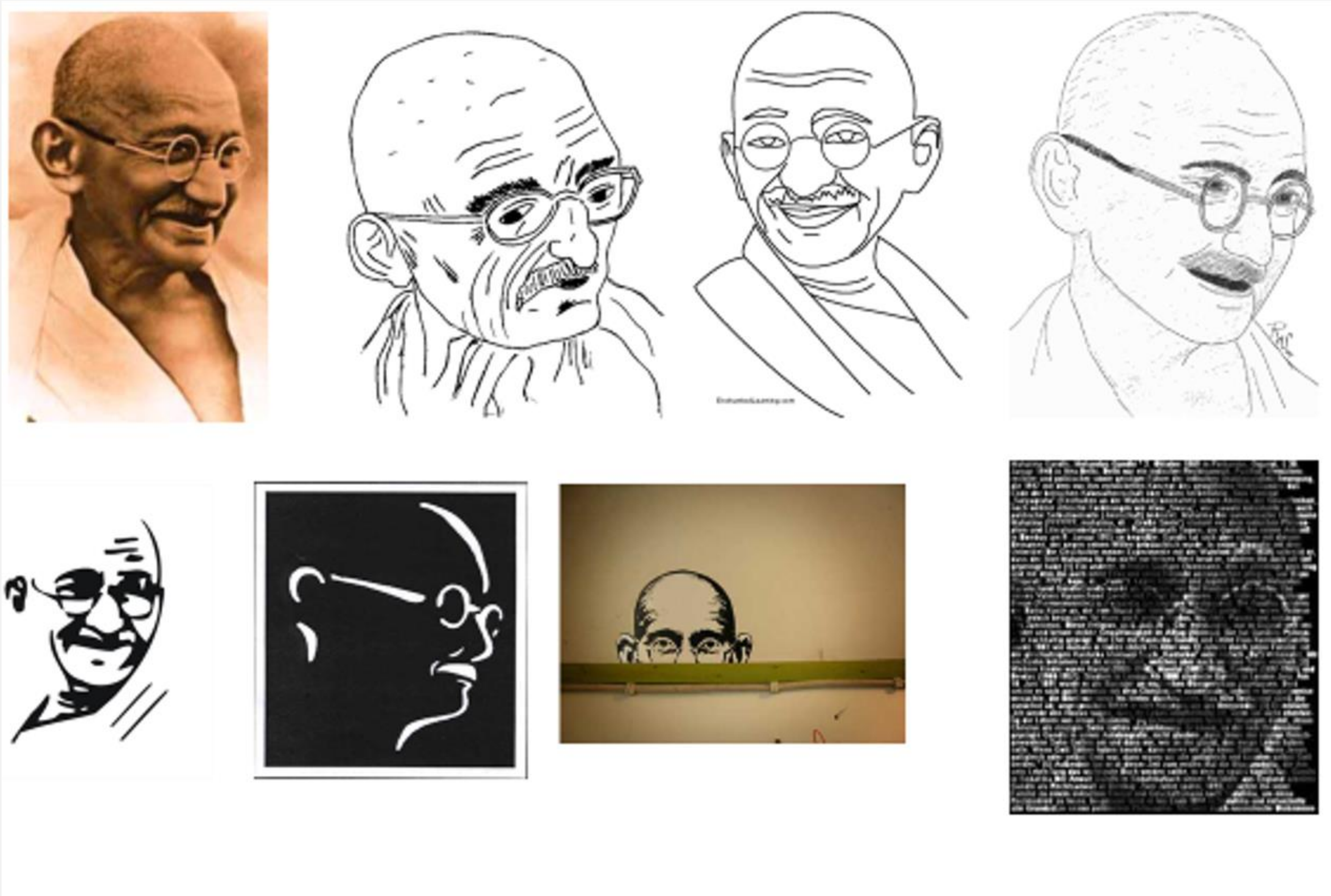
Wallow

Spruce

Larch

Birch

Palm



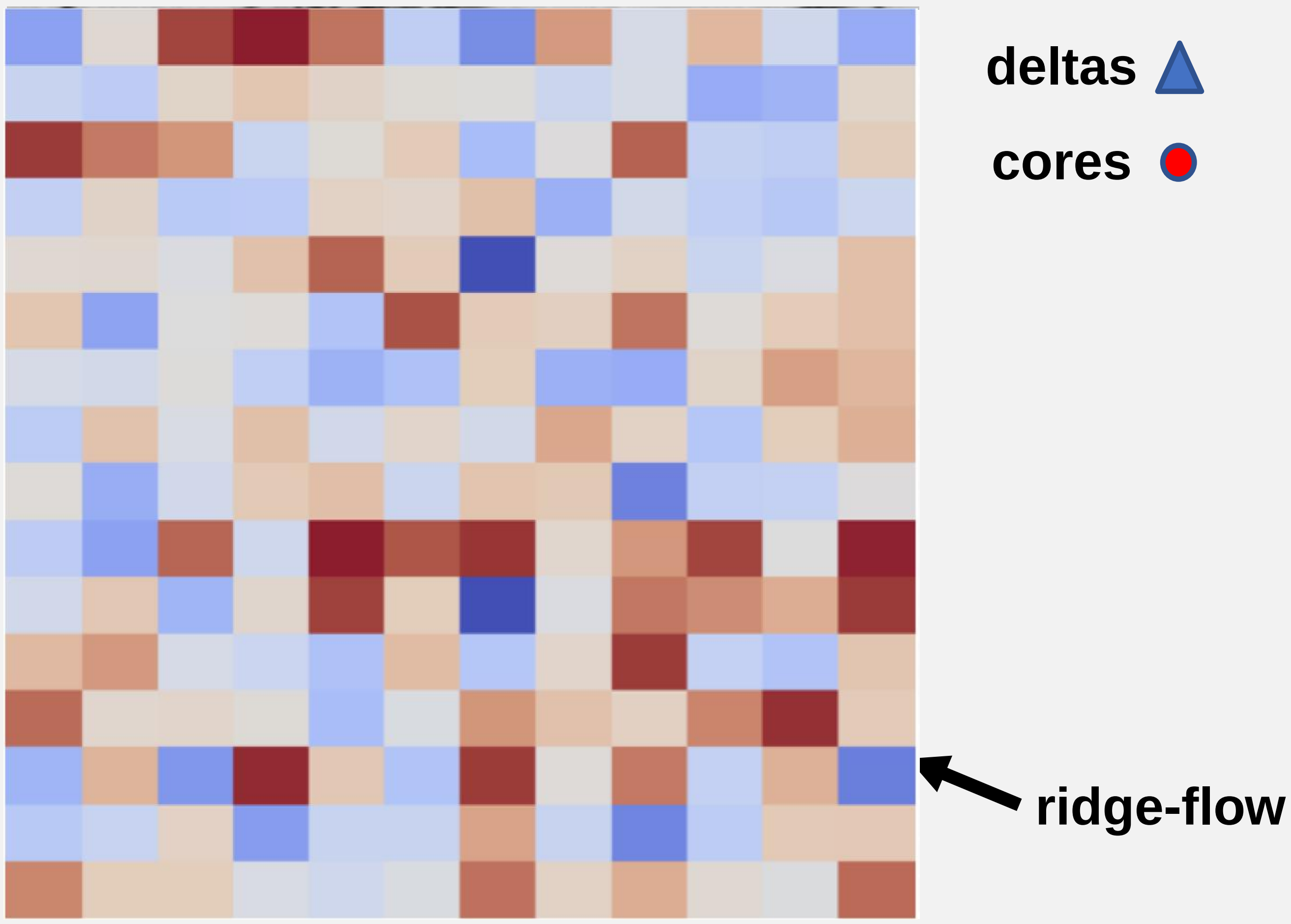
Intra-class Variability

Inter-Class Similarity

Learn a compact & discriminative representation for pattern classes

www.cbsnews.com/8301-503543_162-57508537-503543/chinese-mom-shaves-numbers-on-quadruplets-heads

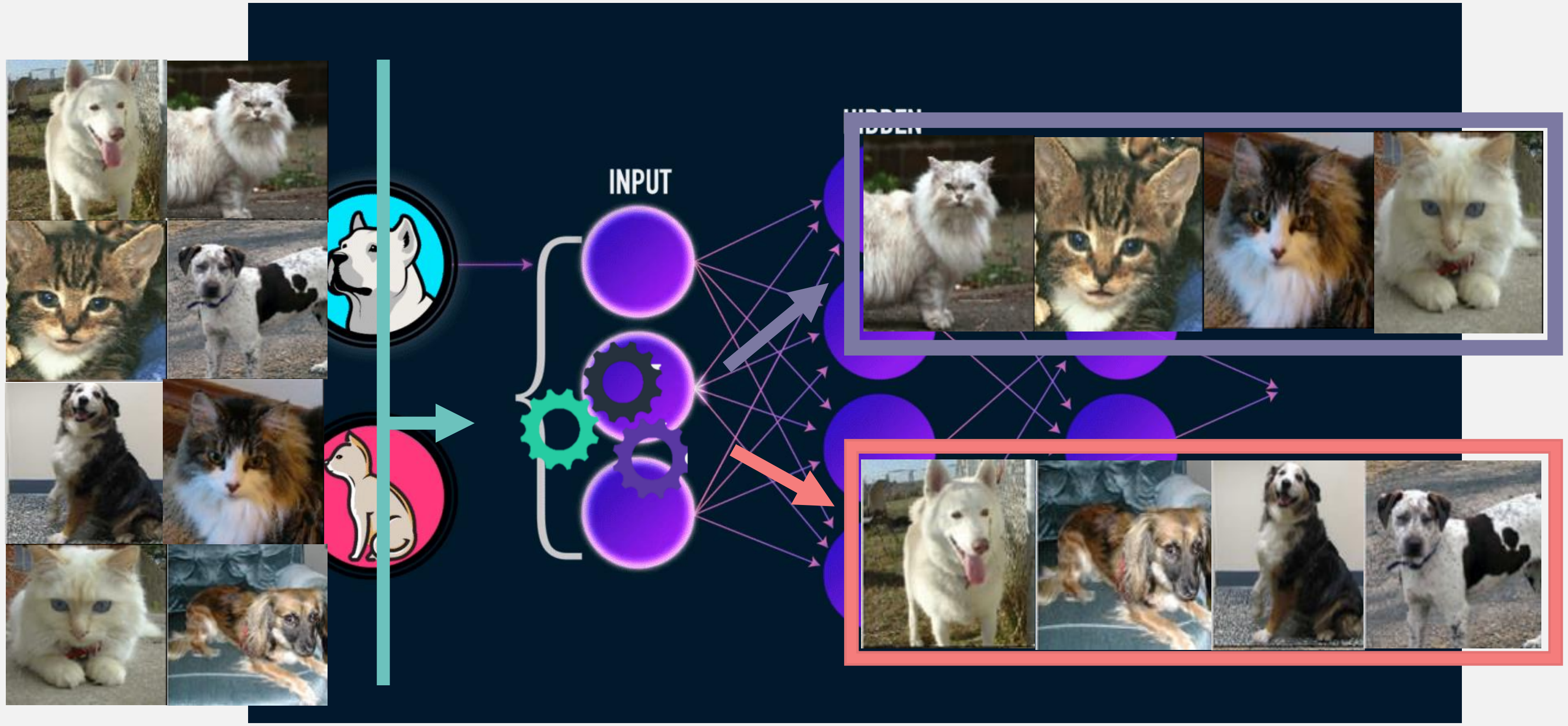
Representation, Matching and Similarity



Fusion of multiple representations can boost recognition performance

Recognition (Learning)

Assign patterns to known classes (classification) or group them to define classes (clustering)



Clustering (Unsupervised learning)

Model-driven Approach: Linear Discriminant (1936)

Fisher (1890-1962)

Input: Features $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_\mathbf{n})$
Labeled data (by pattern class) for 2 classes
Statistical model: $\mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\mathrm{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$

Output: Class label of the input

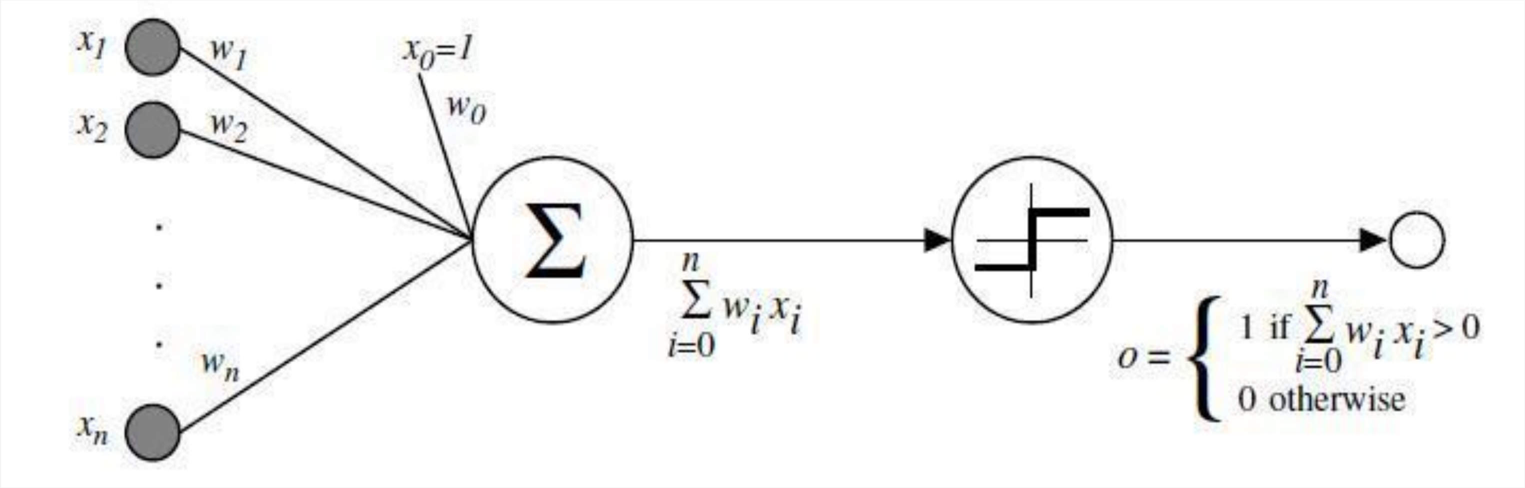
Learning: Estimate model parameters $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$

R.A. Fisher, The Use of Multiple Measurements in Taxonomic Problems, Annals of Eugenics, 1936

Data-Driven Approach: Perceptron (1958)

First biologically motivated network that learns to classify patterns

Rosenblatt (1928-1971)



Input: Features $(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$
Labeled data

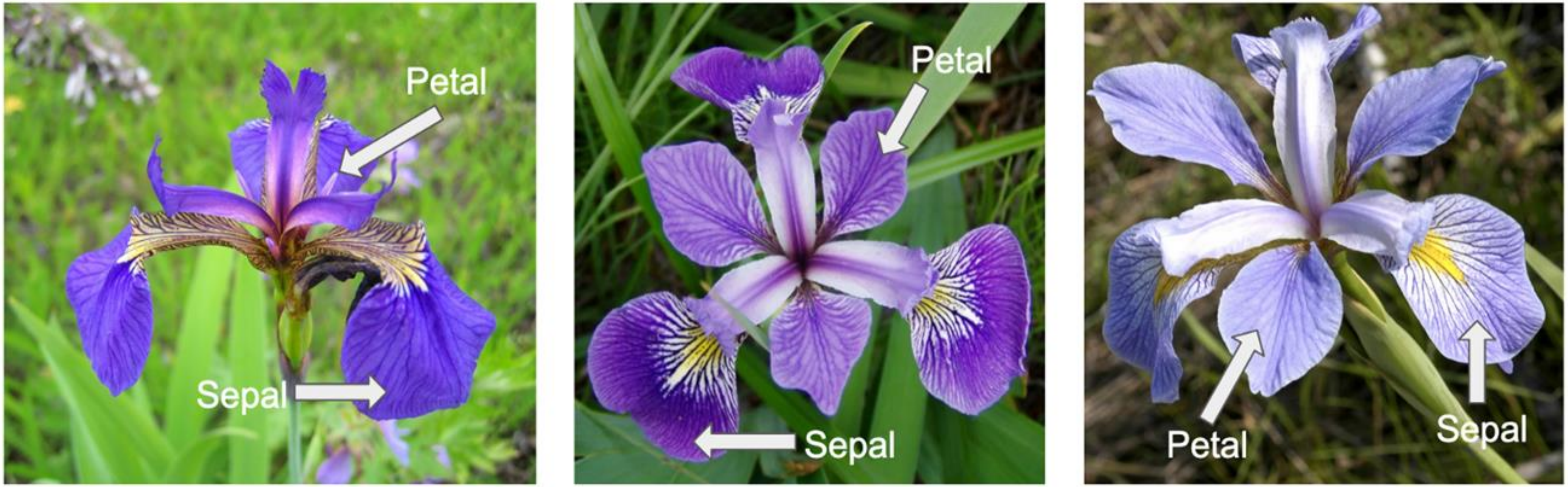
Output: Class label of the input

Learning: Network weights (w_0, w_1,….,w_n)

F. Rosenblatt. The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory, 1957

Fisher’s Iris Data

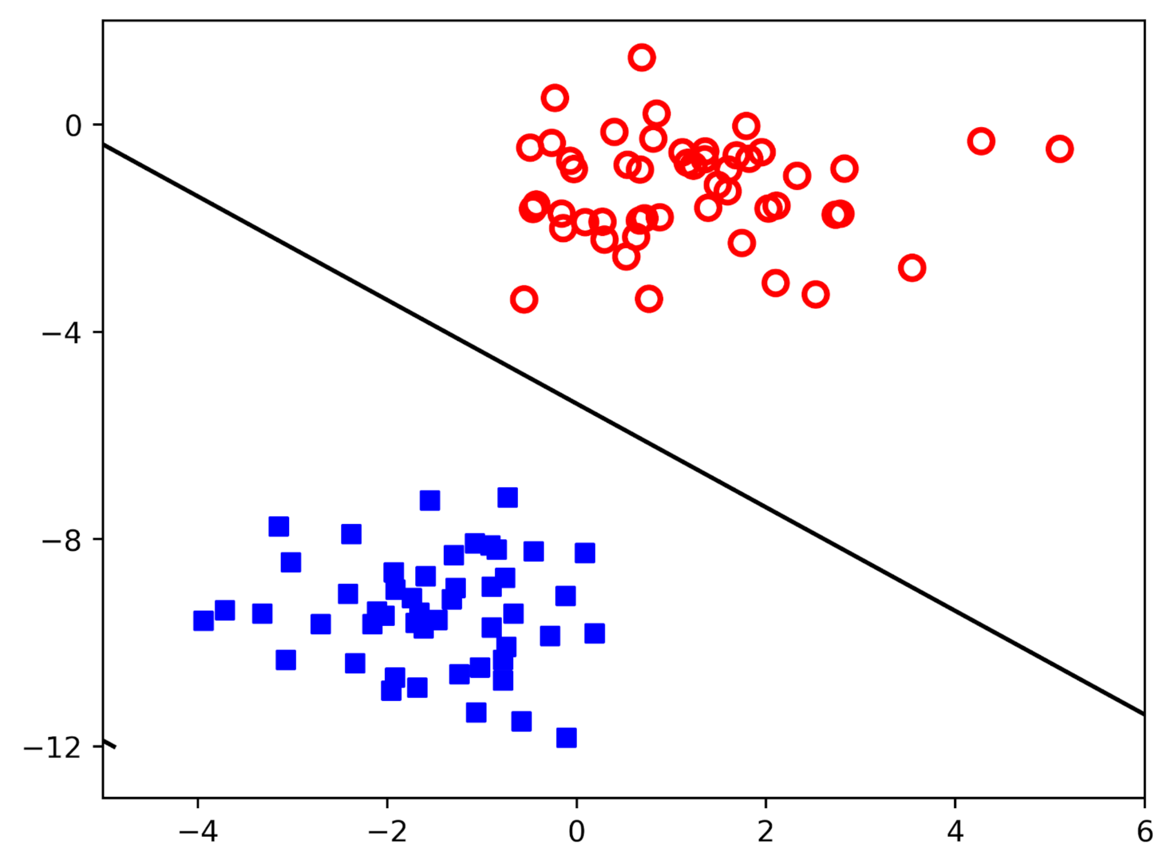
Iris setosa *Iris versicolor* *Iris virginica*



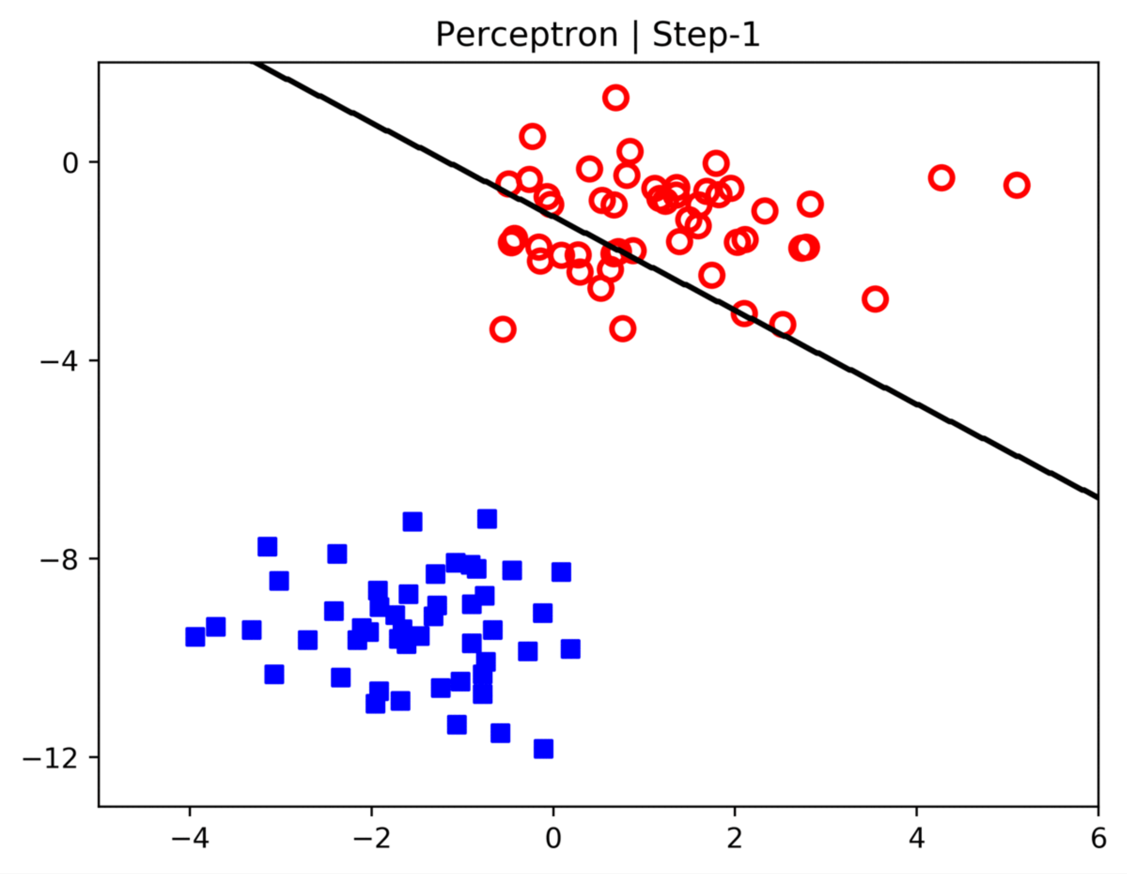
4 features (sepal length and width, petal length and width), 50 samples/class

https://archive.ics.uci.edu/ml/datasets/iris

Linearly Separable Data



Estimate (μ_1, μ_2, Σ)



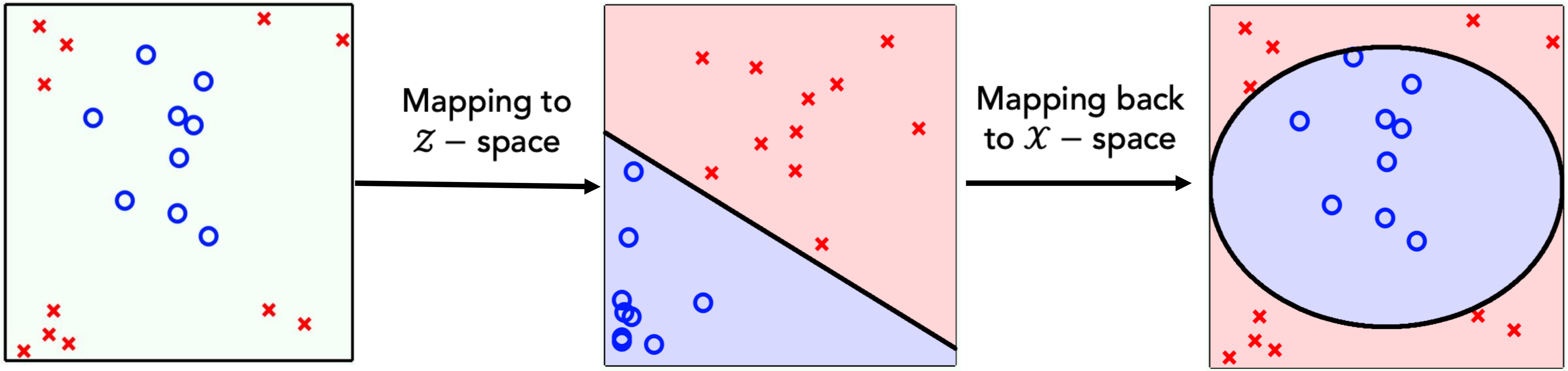
Perceptron Learning Algorithm

Linear Discriminant and Perceptron do not work for non-linearly separable data

Linear to Quadratic Classifiers and SVM

Statistical model: $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$

Nonlinear kernel: Transform data to linearly separable space



Original feature space $\mathcal{X}$

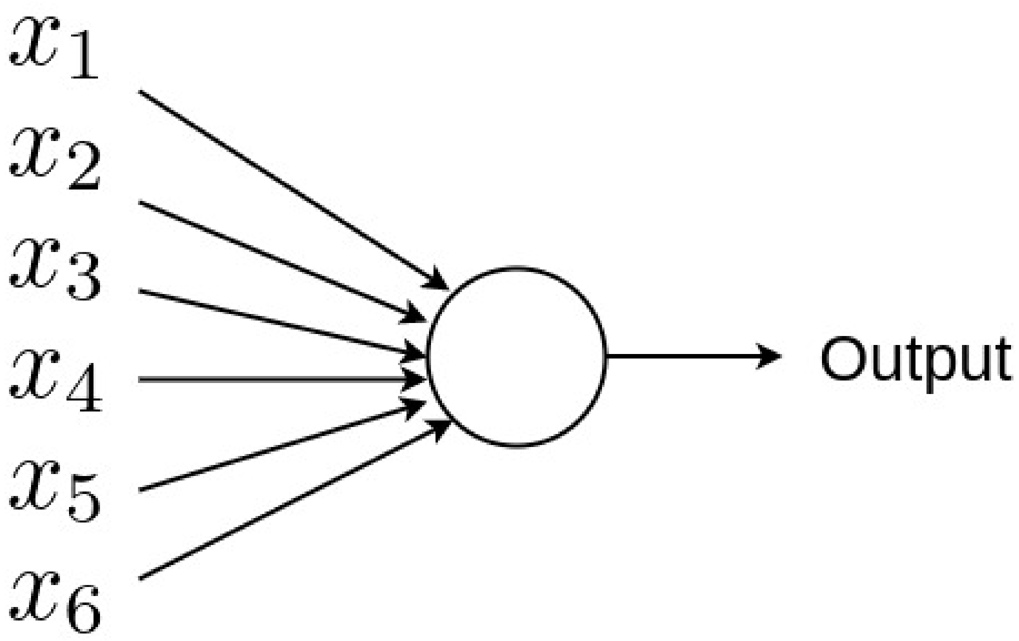
Decision boundary after non-linear transformation $z - \phi(\mathrm{x})$

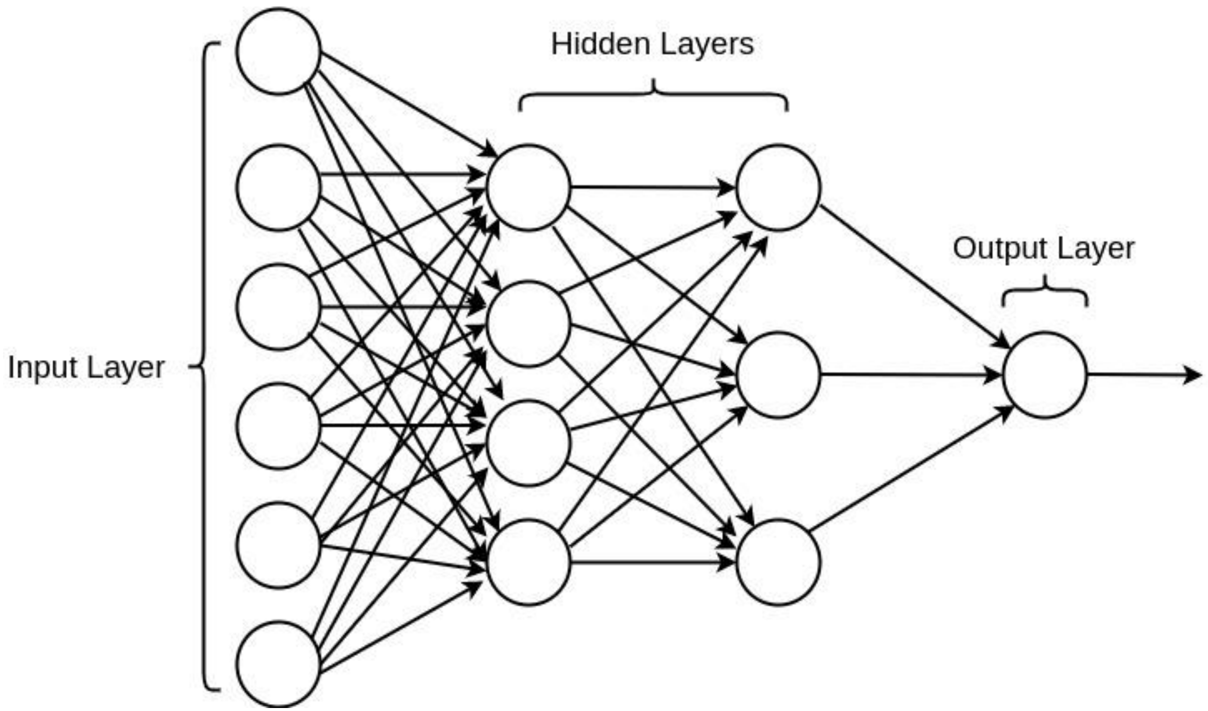
Non-linear decision boundary in the original space

Abu-Mostafa, Magdon-Ismail, Lin, "Learning from Data", AML Book, 2012

T. W. Anderson, "Classification into Multivariate Normal Distributions with Unequal Covariance Matrices, JASA, 1960

Perceptron to Multi-layer Neural Networks





Perceptron

(7 parameters to learn)

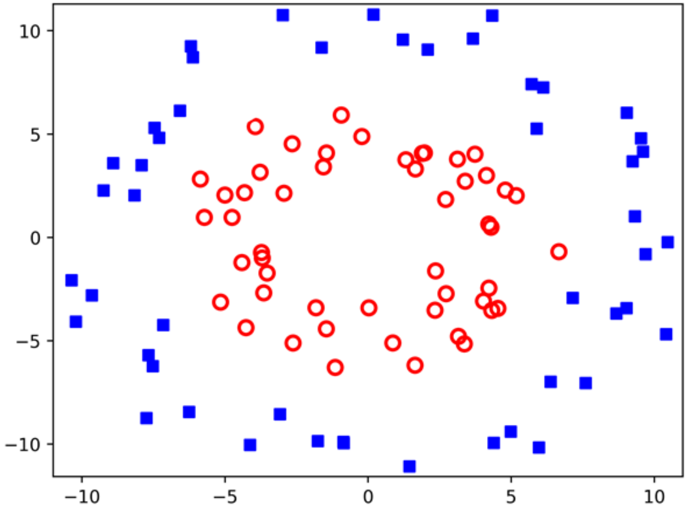
Rosenblatt's Perceptron learning algorithms

2-Hidden layer neural network

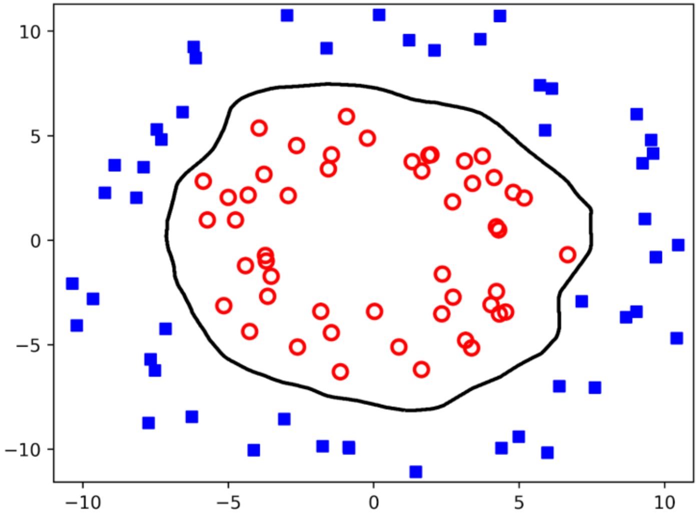
(47 parameters to learn)

Backpropagation learning algorithm:
Werbos, 1974; Rumelhart, Hinton & Williams, 1986

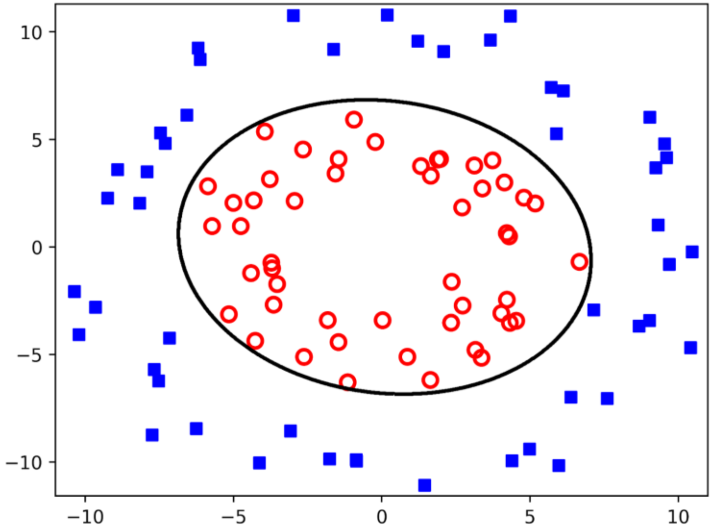
Non-Linearly Separable Data



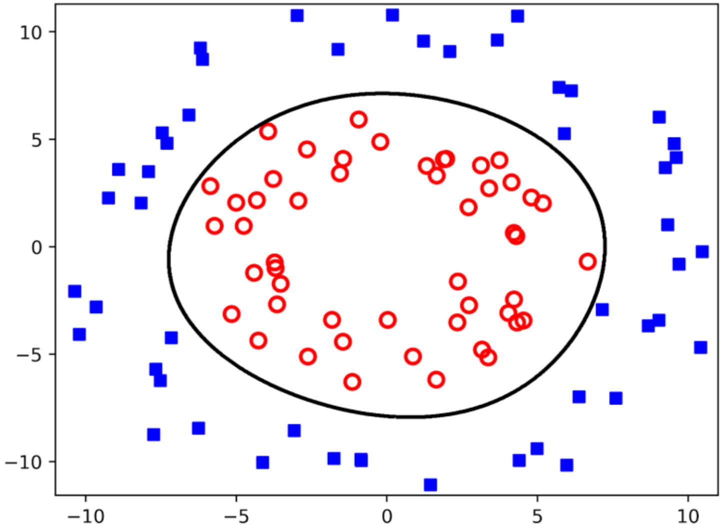
Input Data



2-Hidden Layer Network



Quadratic Classifier

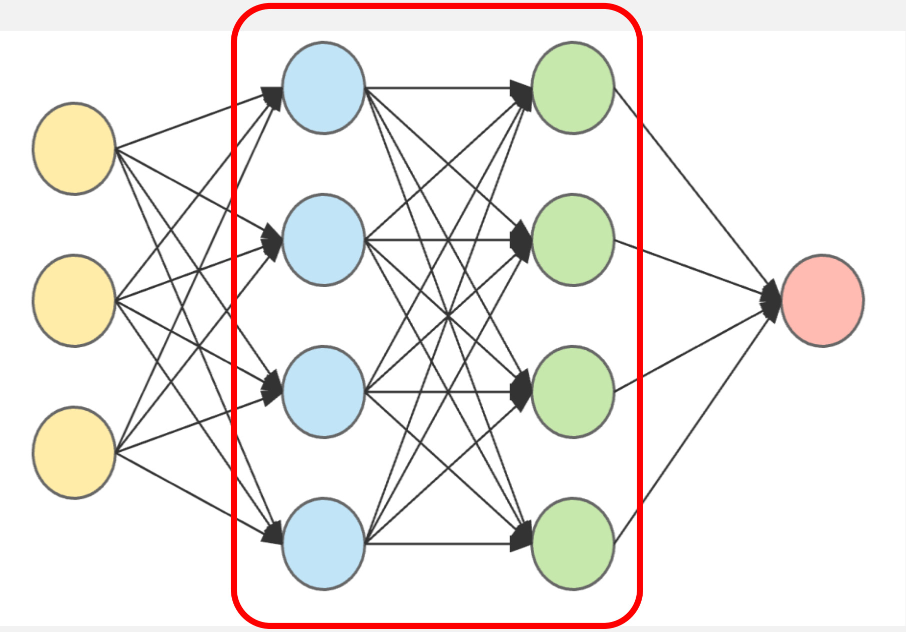


SVM

Deep Networks

End-to-end approach to jointly learn features and predictor

Data → **Hand-crafted Features** → Learning Algorithm → Prediction

Data →  → Prediction

Learned Features

Why are Deep Networks So Popular?

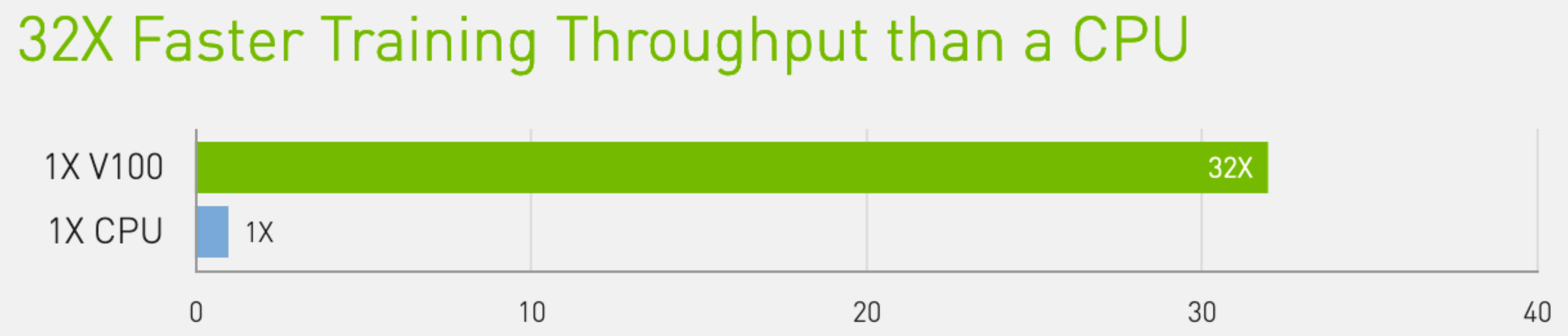
Large-scale annotated datasets

- **ImageNet: 14M *images* from 22K *classes* collected from the web**

http://www.image-net.org

Why are Deep Networks So Popular?

Faster Computation



NVIDIA Tesla V100

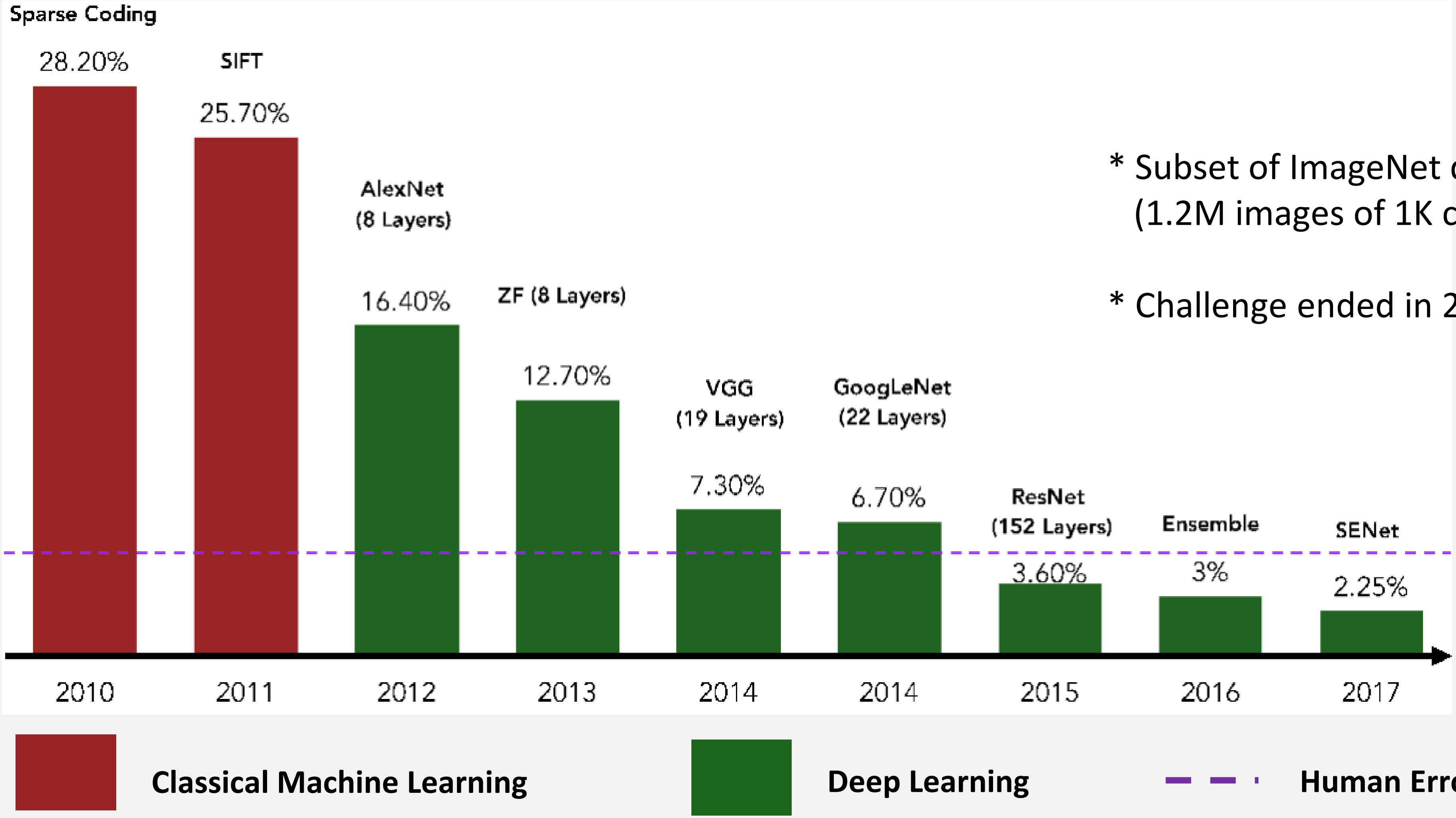
RAM: **32-64 GB**
Tensor Performance: **100 TFLOPS**
Memory Bandwidth: **900GB/s**
Cost: **$10,664**

https://www.nvidia.com/en-us/data-center/v100

23

Why are Deep Networks So Popular?

Top-5 Classification Error Rates (%) on ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)*



* Subset of ImageNet dataset (1.2M images of 1K categories)

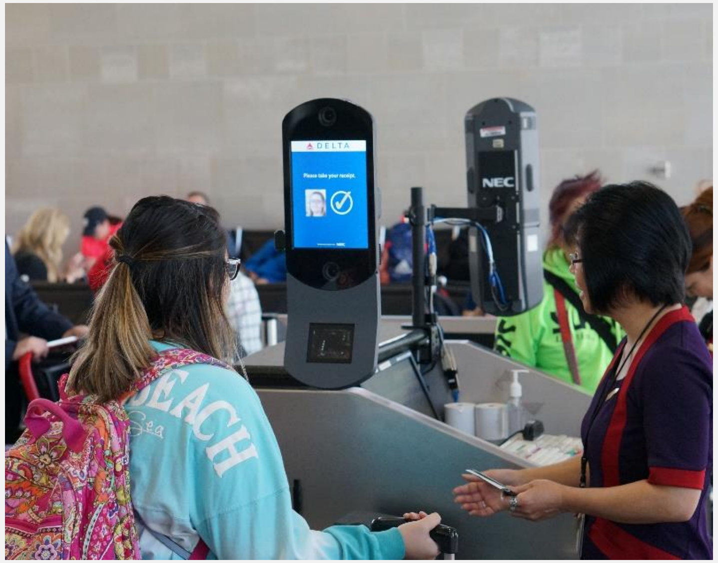
* Challenge ended in 2017

http://www.image-net.org/challenges/LSVRC

Automated Face Recognition

Networked CCTV cameras

Entry into the Unites States



Exit from the United States

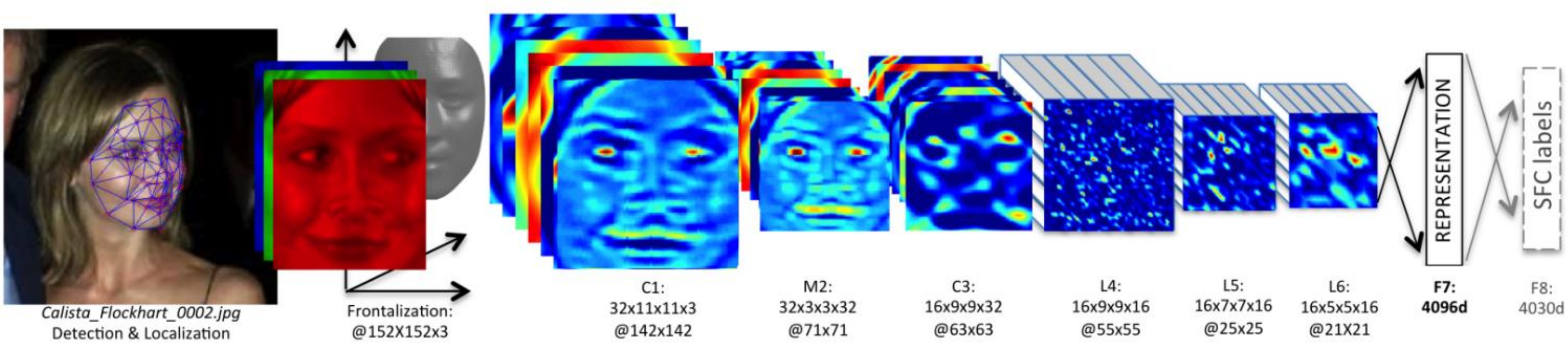
Face Search

Probe

Gallery

Find a person of interest

DeepFace



Multiple layers of neurons stacked together and connected to a small area in previous layer (120M parameters)

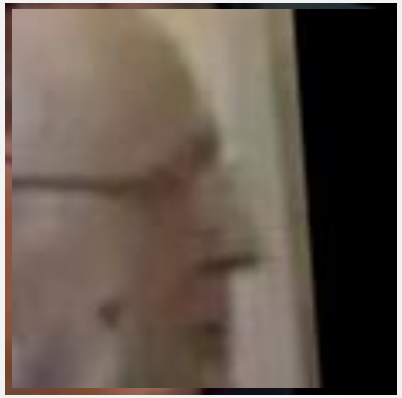
Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. "Deepface: Closing the gap to human-level performance in face verification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1701-1708. 2014.

State-of-the-Art: Authentication

LFW (2009) TAR = 99.2% @ FAR = 0.1%

NIST IJB-S (2018) TAR = 4.86% @ FAR = 0.1%

State-of–the-Art: Search



Probe

Top-5 Retrievals

Rank: 50

Results on IJB-C using ArcFace* (rank-1 search accuracy = 94.5%)

J. Deng, J. Guo, N. Xue, & S. Zafeiriou. "Arcface: Additive angular margin loss for deep face recognition." In CVPR 2019.

Interpretability

What kind of faces does the network see?

- reconstructing the potential appearance from deep face features

High Quality

Medium Quality

Poor Quality

Visualizing CosFace* features via a decoder trained on MS-Celeb-1M (5.8M images of 85K subjects)

*CosFace: H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu. "Cosface: Large margin cosine loss for deep face recognition." In CVPR, 2018.

Y. Shi and A. K. Jain, "Probabilistic Face Embeddings", ICCV 2019.

Fairness: Demographic Bias

- **At most 1% difference in accuracies between race and gender classes**

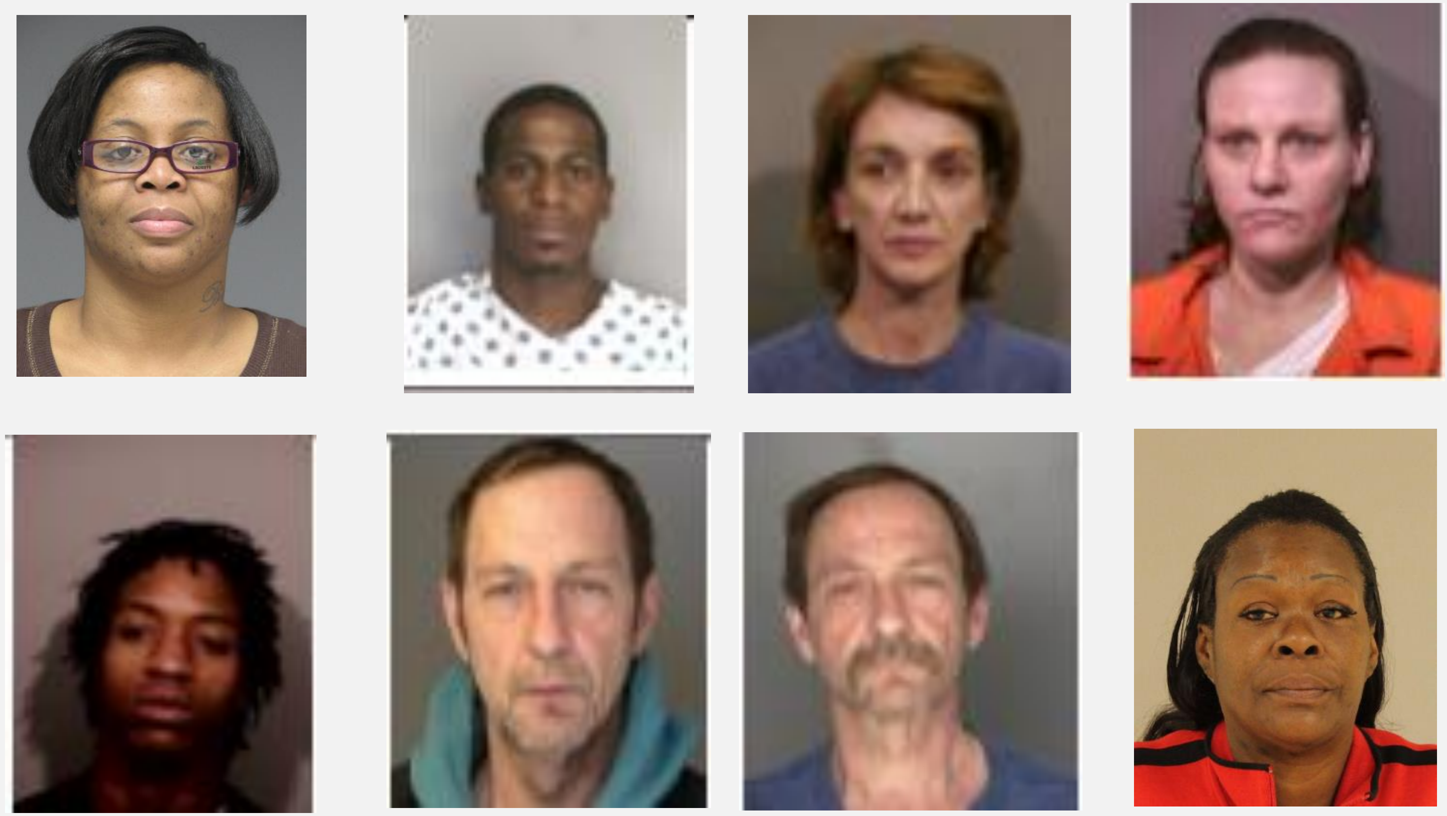
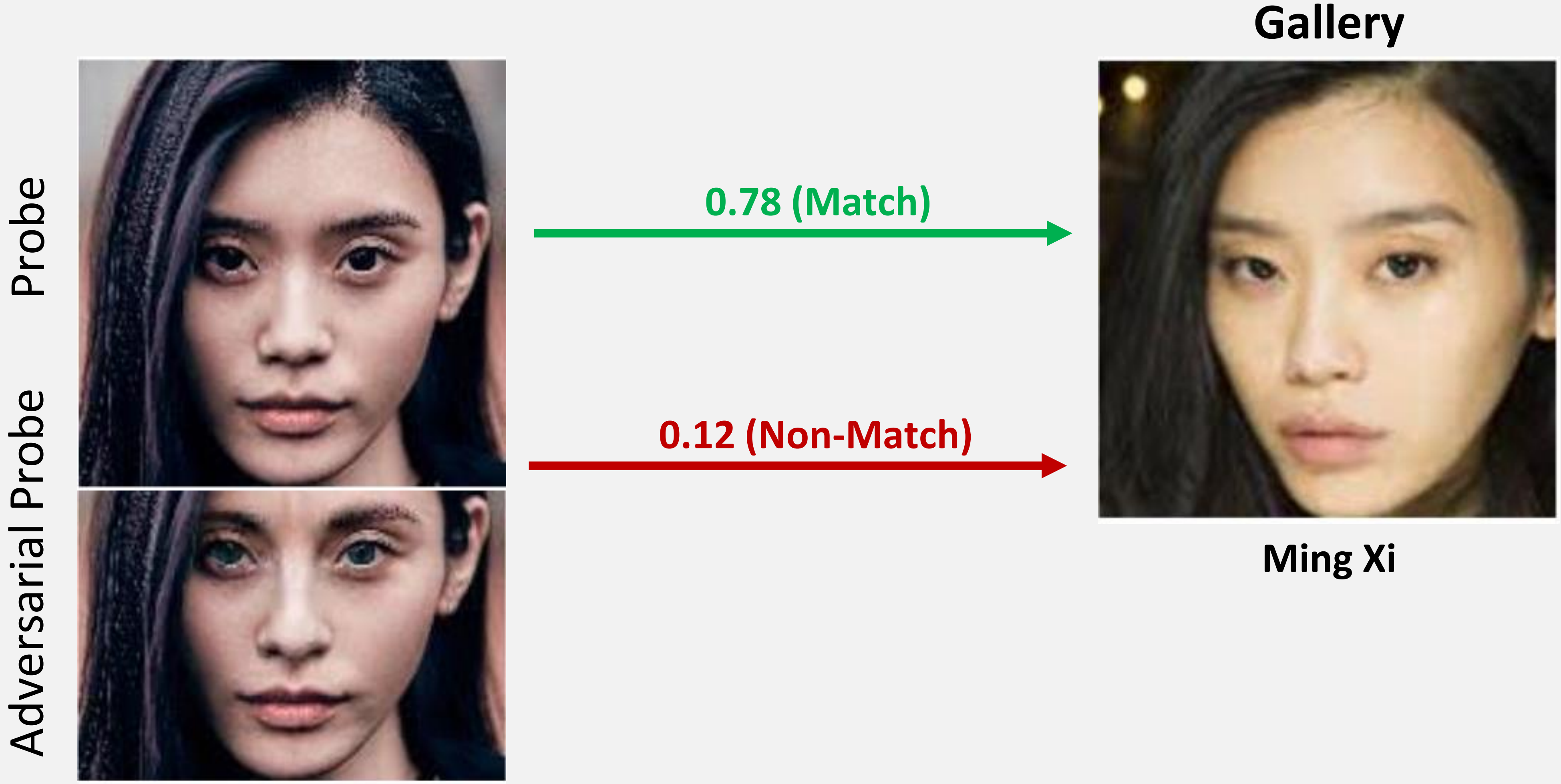


Figure 64: "For the mugshot images, error tradeoff characteristics for white females, black females, black males and white males.", NIST.gov Face Recognition Vendor Test (FRVT) 1:1 Ongoing, Nov. 11, 2019

Digital Image Manipulation



D. Deb, J. Zhang, and A. K. Jain, "AdvFaces: Adversarial Face Synthesis", arXiv:1908.05008, 2019.

Security vs. Privacy

Summary

- Many of our daily tasks involve recognizing patterns: faces, vehicles, pedestrians, voice, trees, buildings,…
- Two approaches: Model-based & data-driven (deep networks)
- Training a recognition algorithm needs large labeled data
- DNs are now popular: (i) no modelling, (ii) access to large data
- DNs provide state-of-the-art: object, face & speech recognition
- DNs are “brittle” and cannot explain their actions
- Another AI Winter? (1974–1980; 1987–1993)

Next Decade of AI

- **Access to labeled data:** Utilize synthetic & unlabeled data
- **Domain knowledge:** Combine top-down & bottom-up
- **Network capacity:** How many pattern classes can it separate?
- **Adversarial attacks:** Brittle to robust networks
- **Explainability:** How does a network make a decision?
- **User privacy:** Safeguard users' private data
- **Global good:** Design AI to improve lives of extremely poor (~1bn)