Introduction to Visual Learning

ECE 285

Xiaolong Wang

xiw012@ucsd.edu

Logistics

- Location: WLH 2005
- Zoom: https://ucsd.zoom.us/my/xiaolonw
- Website: https://xiaolonw.github.io/ece285/sp25/
- Assignments:
  - 4 Homeworks, each 15%
- Final Project:
  - Project proposal, 10%
  - Project report, 30%

Logistics

- TAs:
  - An-Chieh Cheng: a8cheng@ucsd.edu
  - Yandong Ji: yaj011@ucsd.edu

  Additional Helps:

  - Jiarui Xu: jix026@ucsd.edu
  - Yinbo Chen: yic026@ucsd.edu

Logistics

Office Hour, starting next week:

- Monday, 10:00 am - 11:00 am
- Wed, 3:00 pm - 4:00 pm
- (This Wed office hour: Datahub Tutorial: https://ucsd.zoom.us/my/xiaolonw)

- Location: FAH 3rd floor

Logistics

- Canvas (https://canvas.ucsd.edu/courses/64399):
  - Announcements
  - Zoom recordings
  - Slides and assignments

- Piazza:
  - https://piazza.com/class/m8v9d7mwqh23gi
  - Discussions
  - Slides

- GradeScope:
  - https://www.gradescope.com/courses/1012765
  - Entry Code: **42J4YN**
  - Submit assignments

Prerequisite

- know how to use python for programing
- linear algebra

Date	Lecture	Materials	Assignments
Apr 1	Introduction to Visual Learning	Lecture 1	
Apr 3	Image Classification: K-NN and Linear Classifier	Lecture 2	
Apr 8	Multi-Layer Perceptrons and Back-Propagation	Lecture 3	Assignment 1
Apr 10	Convolutional Neural Networks 1	Lecture 4	
Apr 15	Convolutional Neural Networks 2	Lecture 5	
Apr 17	Tutorial on Pytorch	Tutorial Notebook	Assignment 2
Apr 22	Image Segmentation and Visualization	Lecture 7	Assignment 1 due
Apr 24	Object Detection	Lecture 8	
Apr 29	Recurrent Neural Networks	Lecture 9	
May 1	Video Recognition	Lecture 10	Project Proposal Due
May 6	Video Prediction	Lecture 11	Assignment 3
May 8	Self-Attention, Graph Networks, Transformer	Lecture 12	Assignment 2 due
May 13	Vision Transformer	Lecture 13	
May 15	Generative Adversarial Networks	Lecture 14	
May 20	Conditional GAN and Variational Auto-Encoders	Lecture 15	Assignment 4
May 22	Deep 3D Vision	Lecture 16	Assignment 3 due
May 27	Diffusion Models	Lecture 17	
May 29	Multi-Task, Adaptation, Transfer Learning	Lecture 18	
June 3	Self-supervised Learning	Lecture 19	
June 5	Neural Radiance Fields	Lecture 20	Assignment 4 due

Final Project

https://docs.google.com/document/d/1aByplfb_VHFHTaFdZe2TBZXJQx1AR5Zhj3s4S7IGJ_A/edit?usp=sharing

Cannot re-use existing project that is online.

Class Interaction

Please interrupt and ask questions all the time!

Computer Vision with Deep Learning

What is learning?

- The power of learning lies in generalization

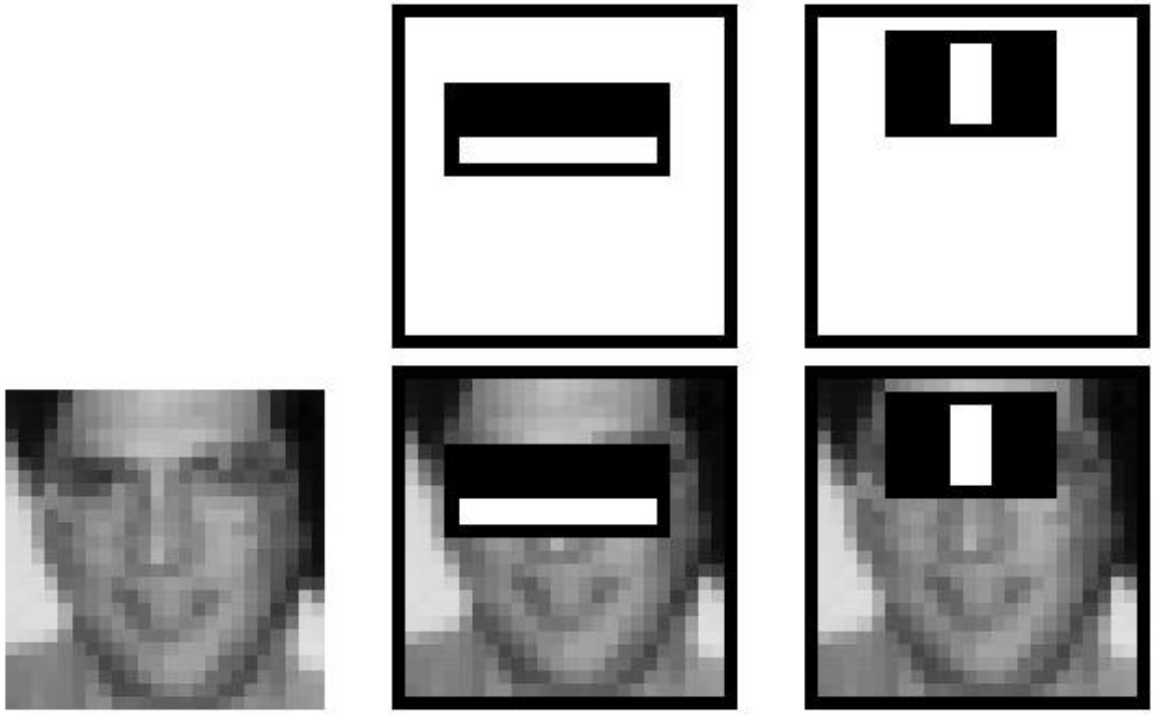


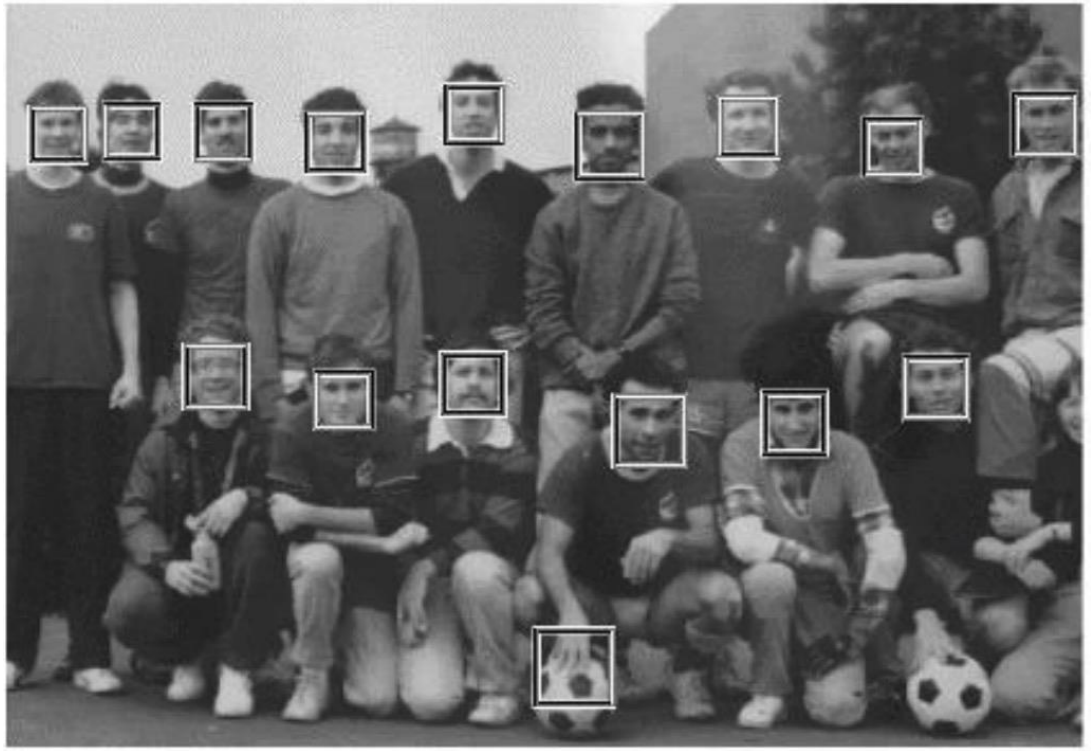

Training Data




Test Data

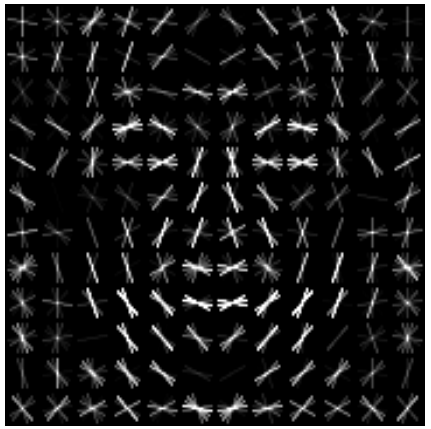
What is Deep about Deep Learning?

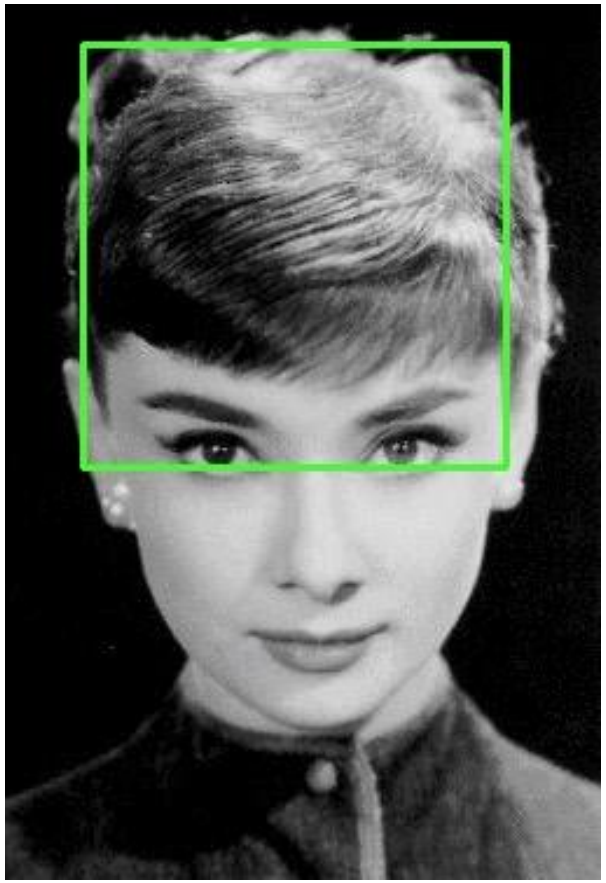


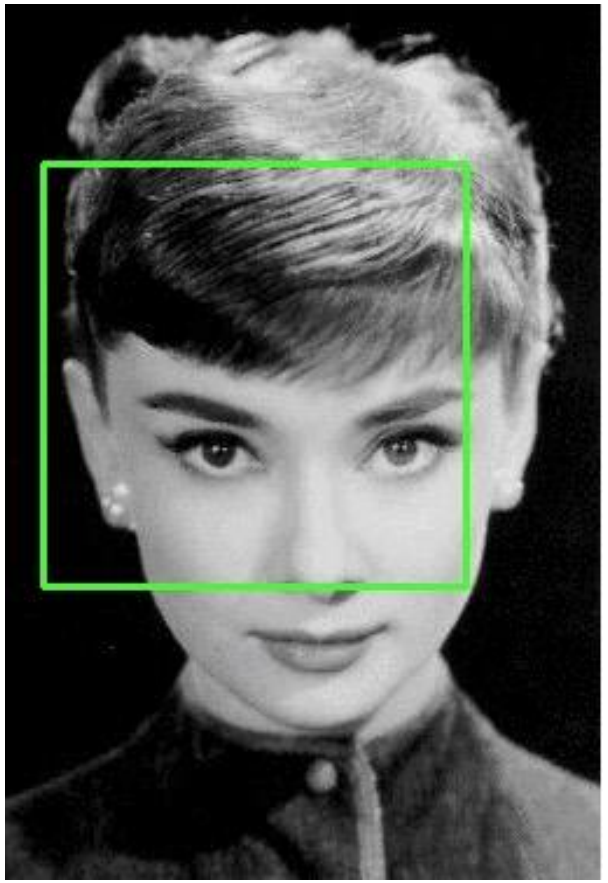


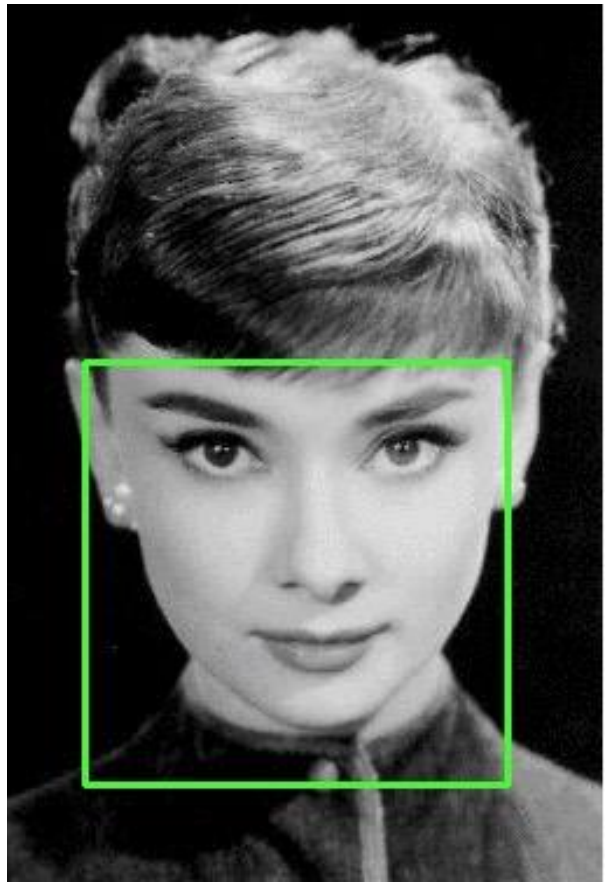
Viola et al. 2001

What is Deep about Deep Learning?







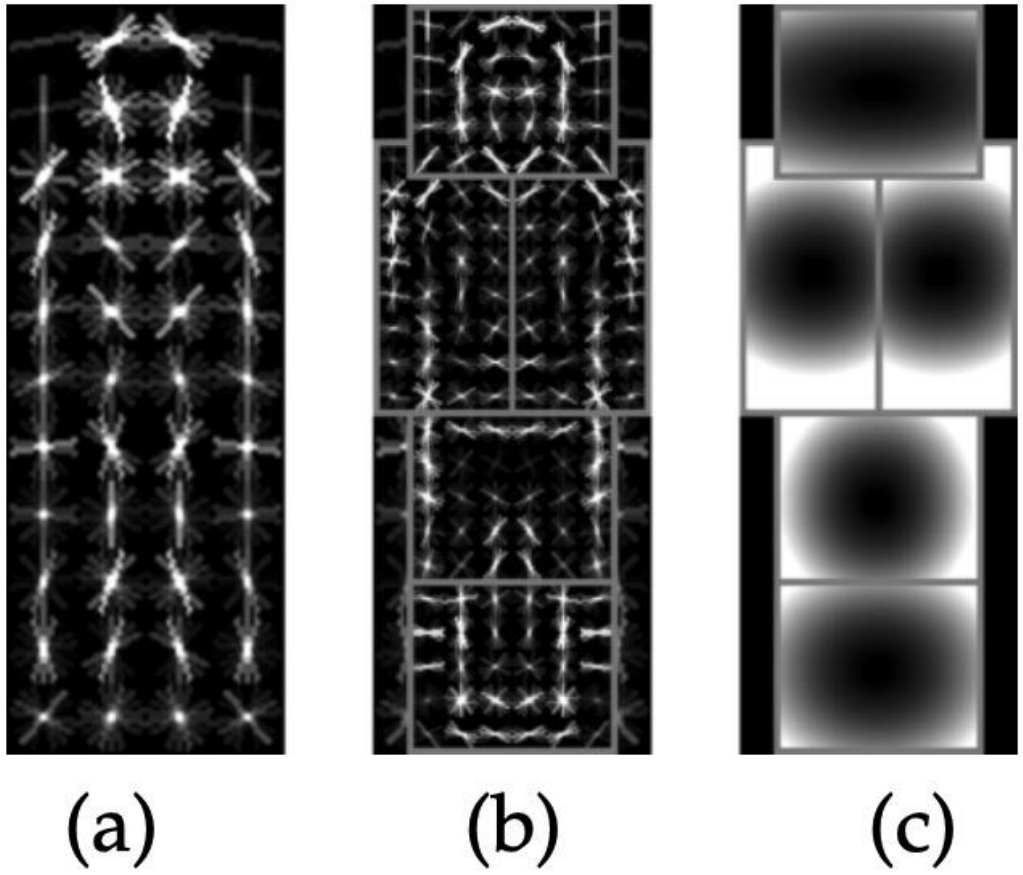


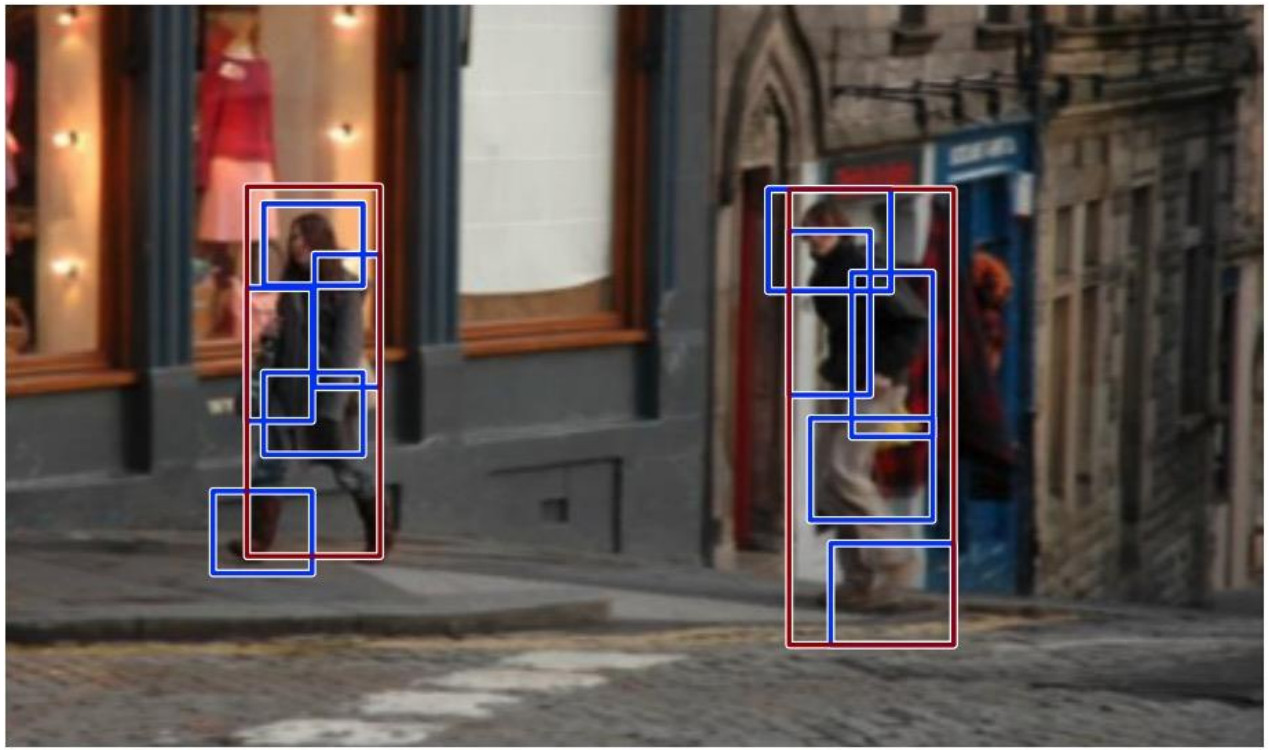
Histograms of Oriented Gradients. Dalal et al. 2005

What is Deep about Deep Learning?

Histograms of Oriented Gradients. Dalal et al. 2005

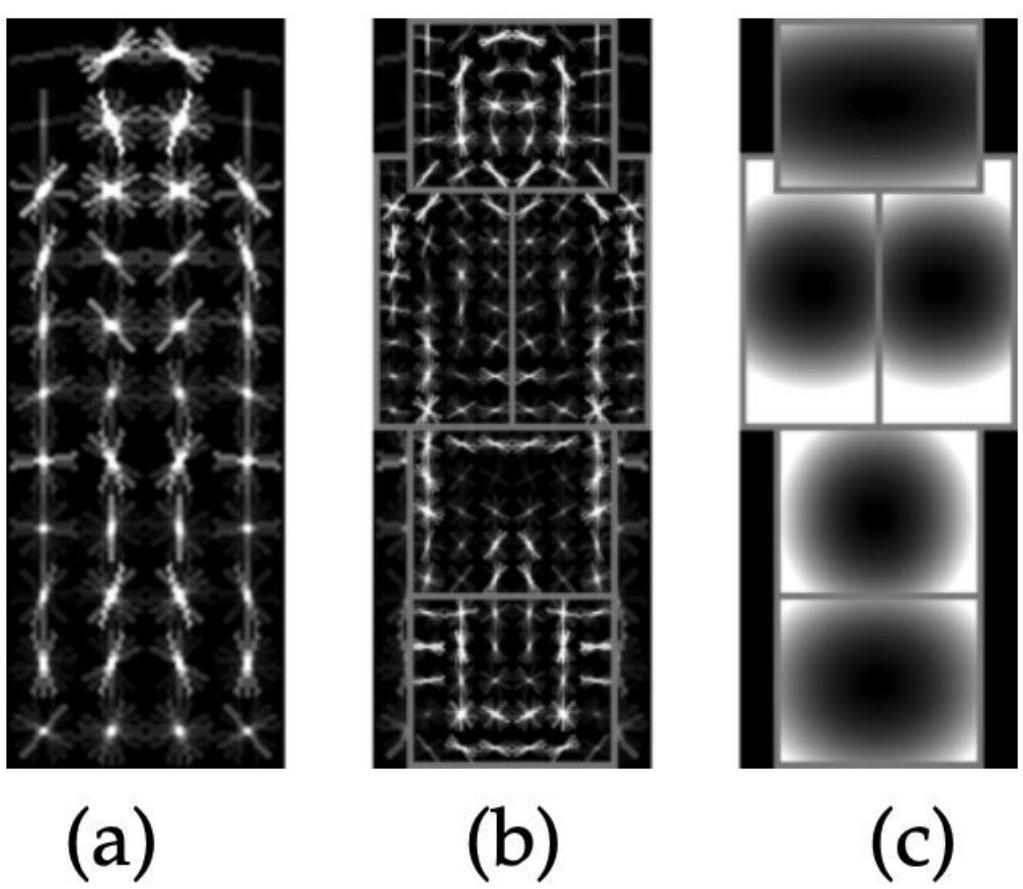
What is Deep about Deep Learning?

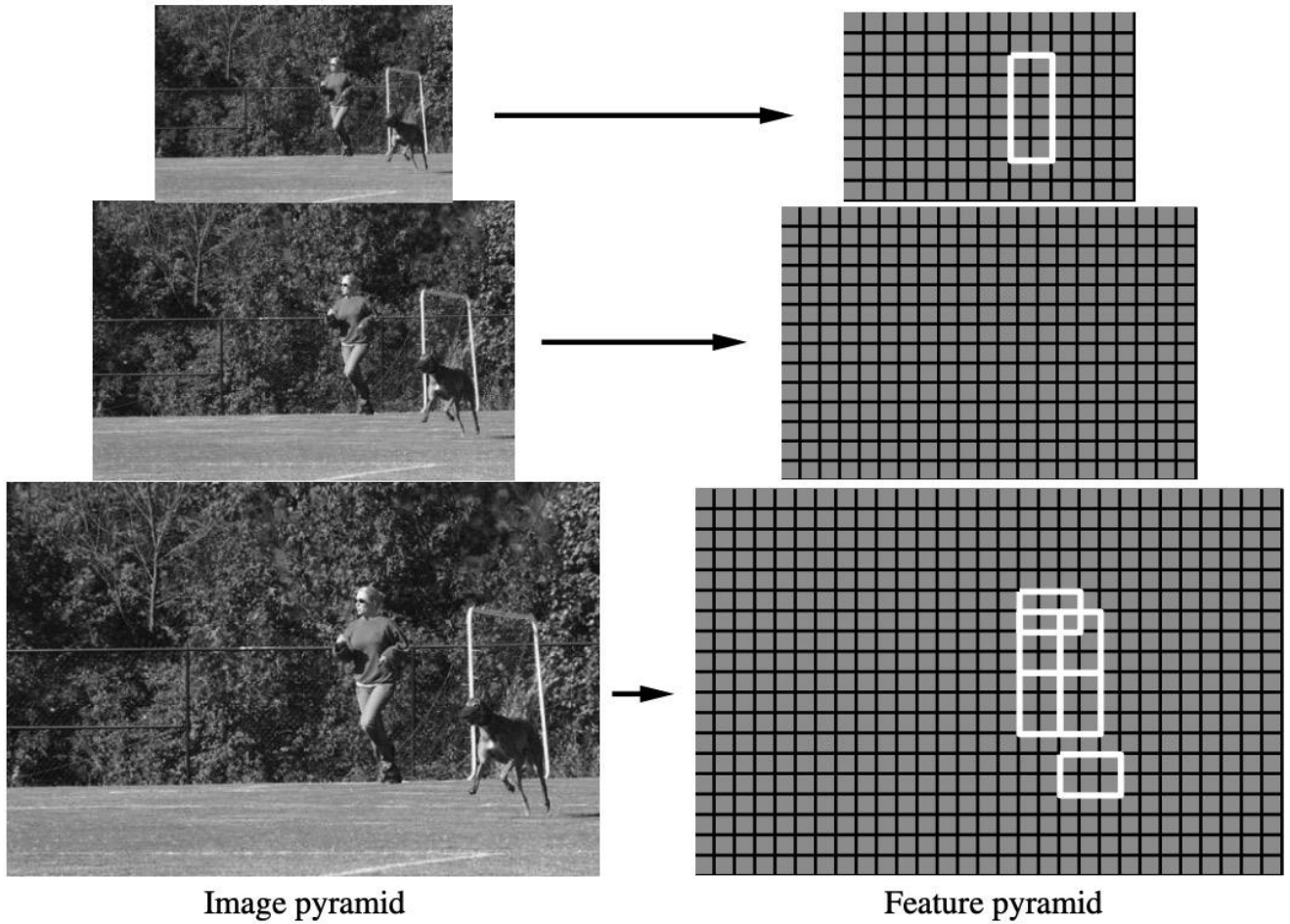




Discriminatively trained Part-based Models. Felzenszwalb et al. 2009

What is Deep about Deep Learning?





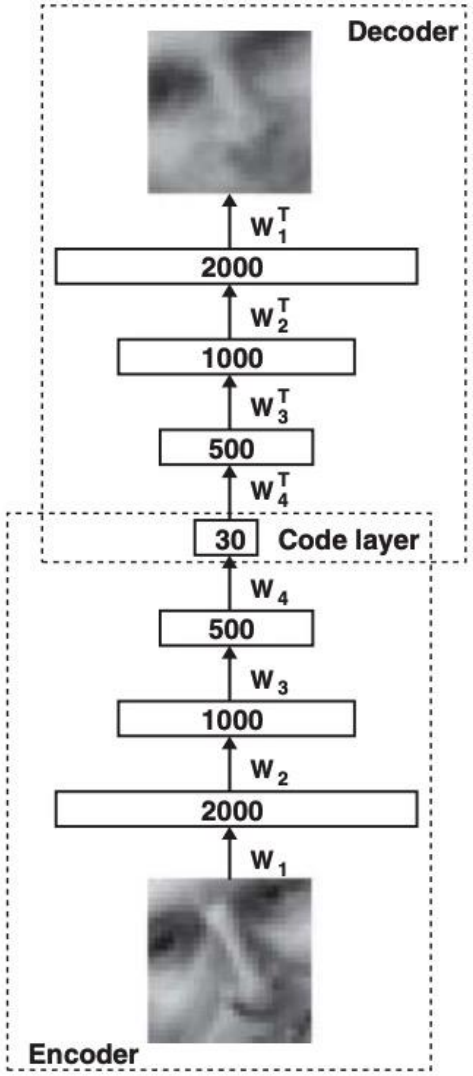
Discriminatively trained Part-based Models. Felzenszwalb et al. 2009

What is Deep about Deep Learning?

- More Layers
  - Previous method: 1-layer classifier (HoG), 2-layer classifier (DPM)
  - Deep Networks: 100, 1000 layers.

- End-to-End Training
  - Previous method: Training each layer of classifier individually.
  - Deep Networks: Training with back-propagation.

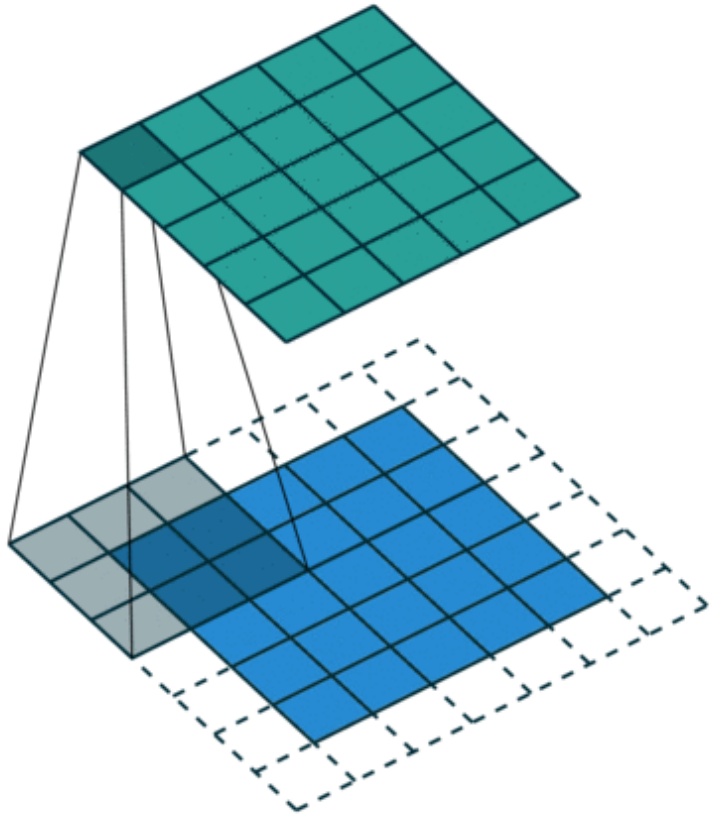
Different Types of Deep Networks

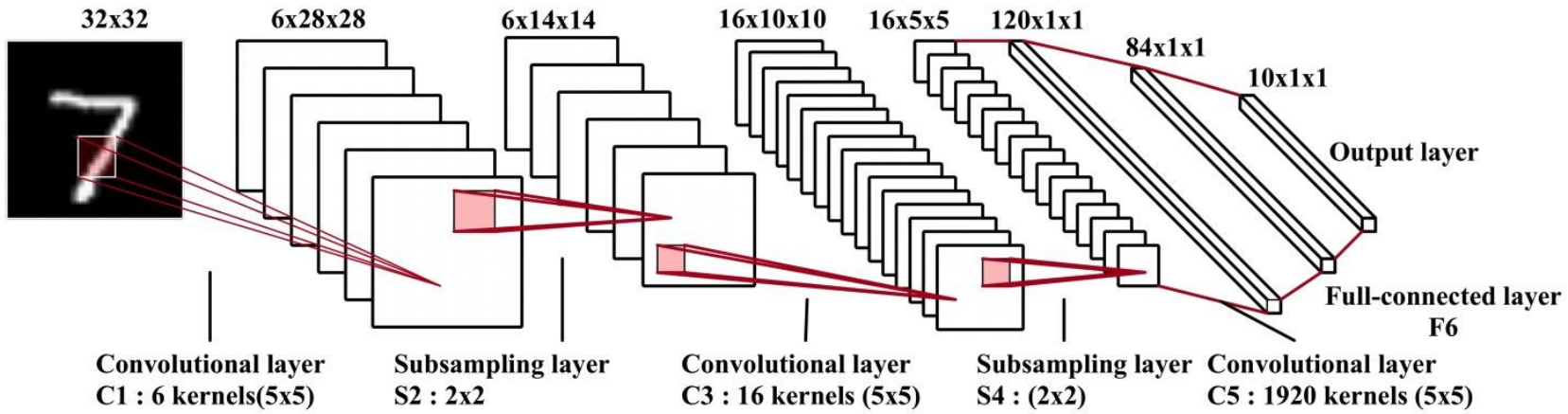


- Multilayer perceptron (MLP)
  - Input image I size : 32 x 32 = 1024
  - First hidden layer h_1 output size: 2000
  - First layer parameters W_1 size: 1024 x 2000
  - $h_1 = I\, W_1$

Hinton et al. 2006

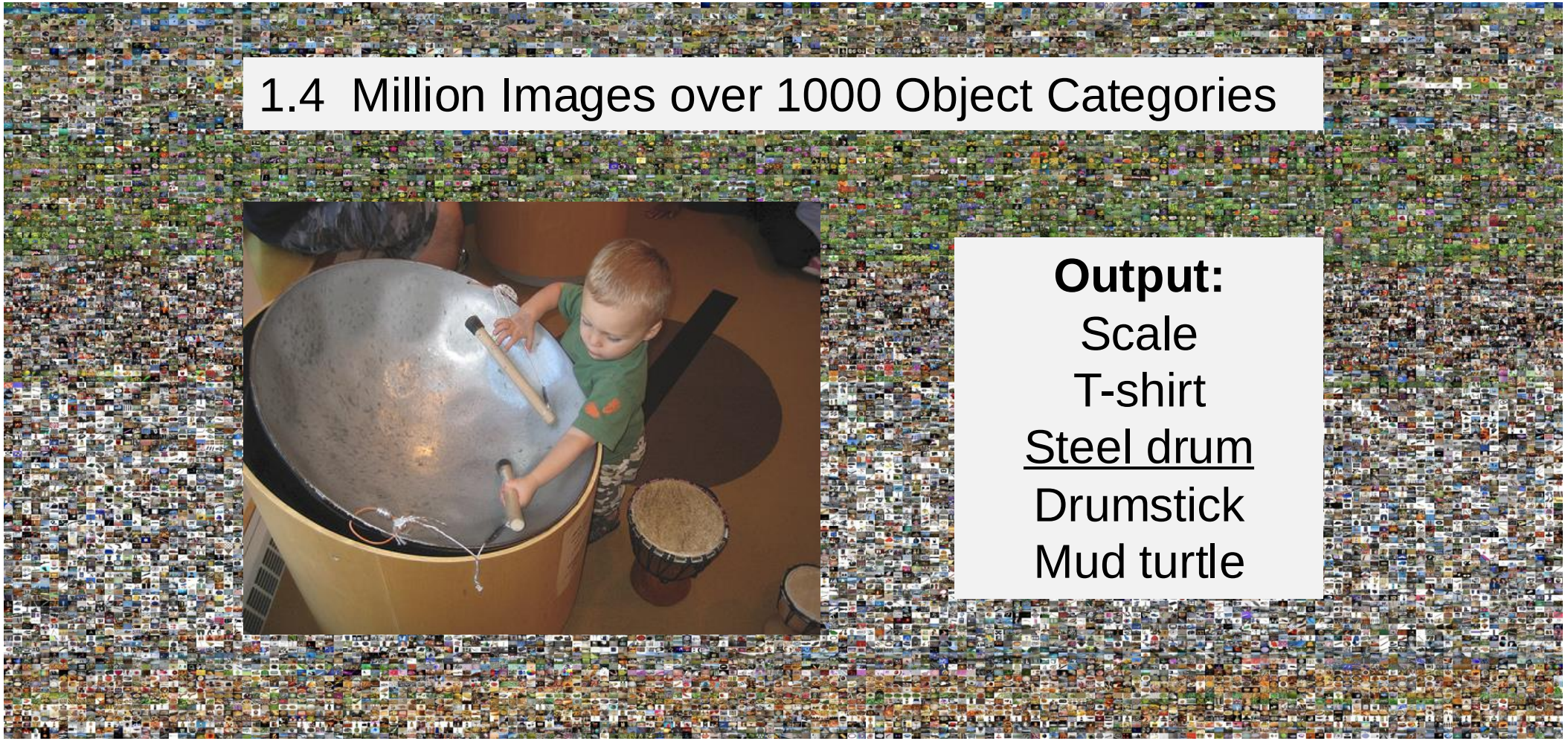
Different Types of Deep Networks





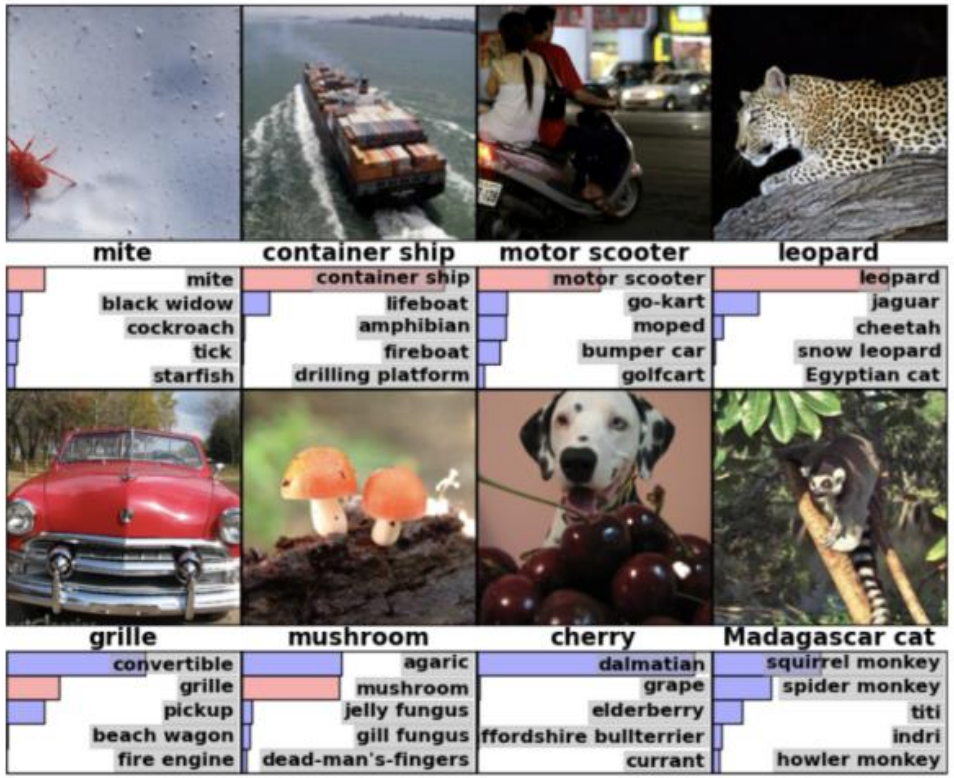
LeCun et al. 1990

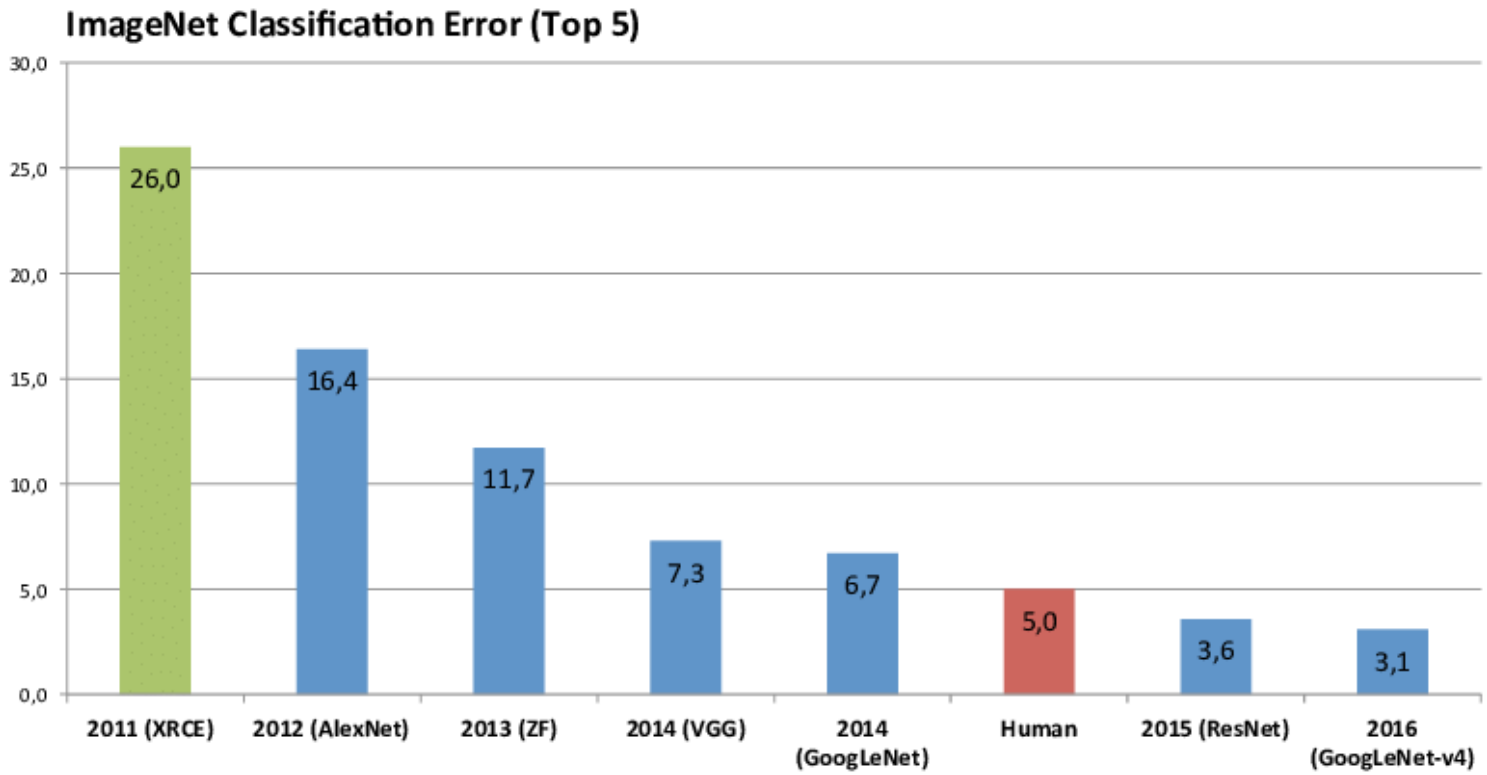
The ImageNet Challenge



Russakovsky et al. 2015

The ImageNet Challenge

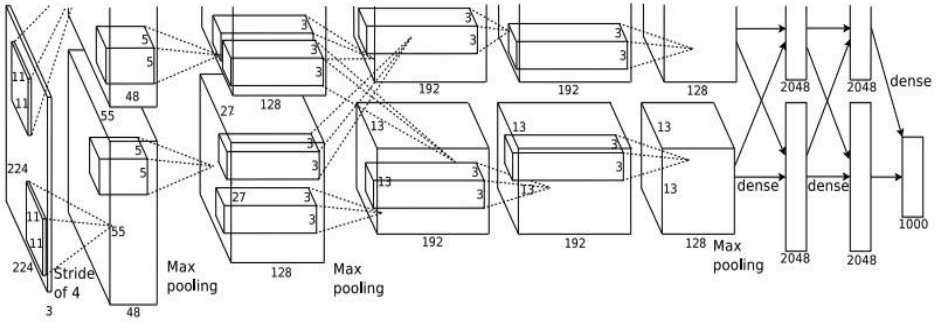


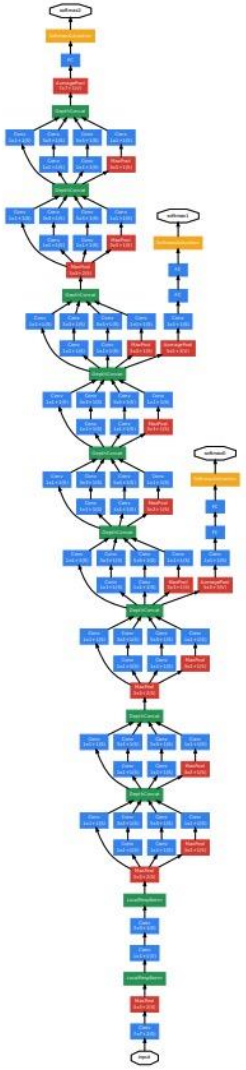


Where does the 5% error human performance come from?

http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/

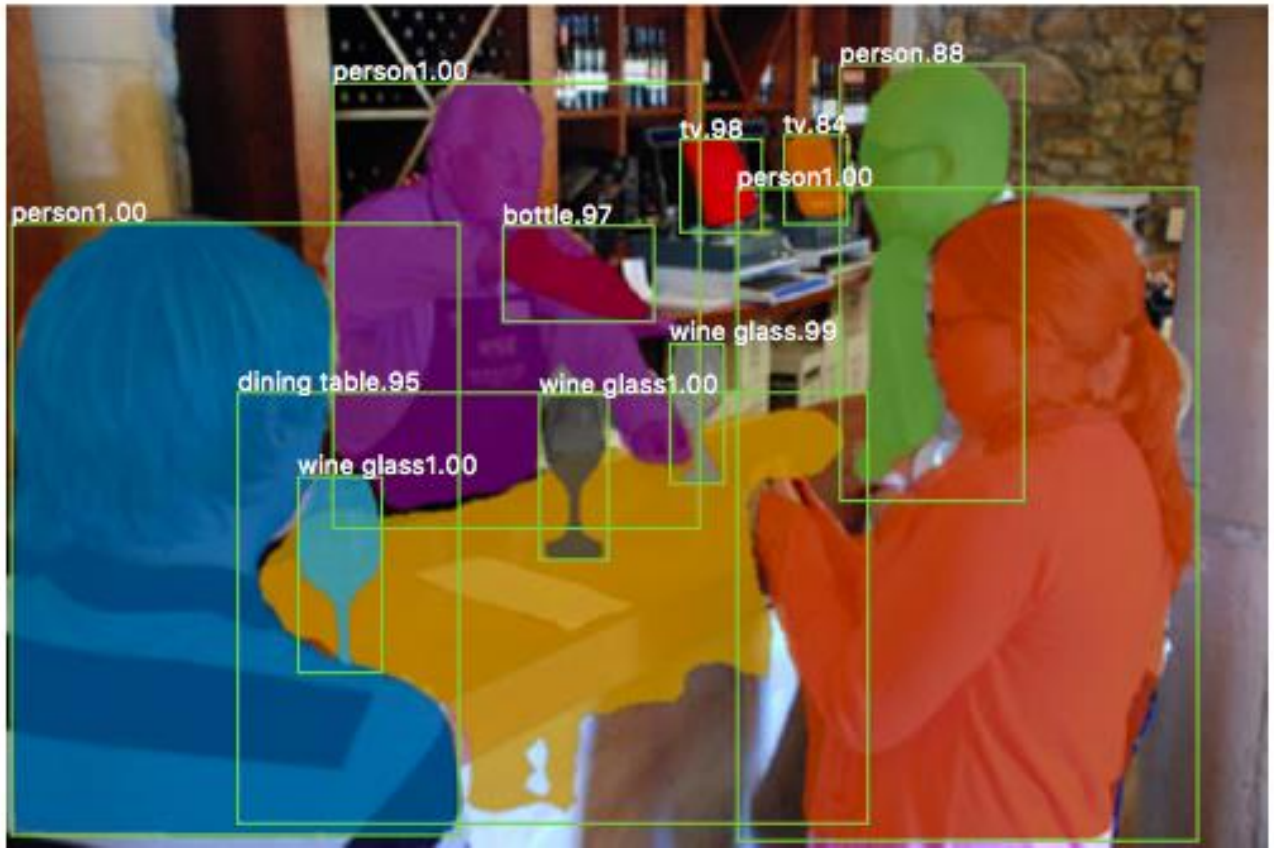
Many-Layer Networks

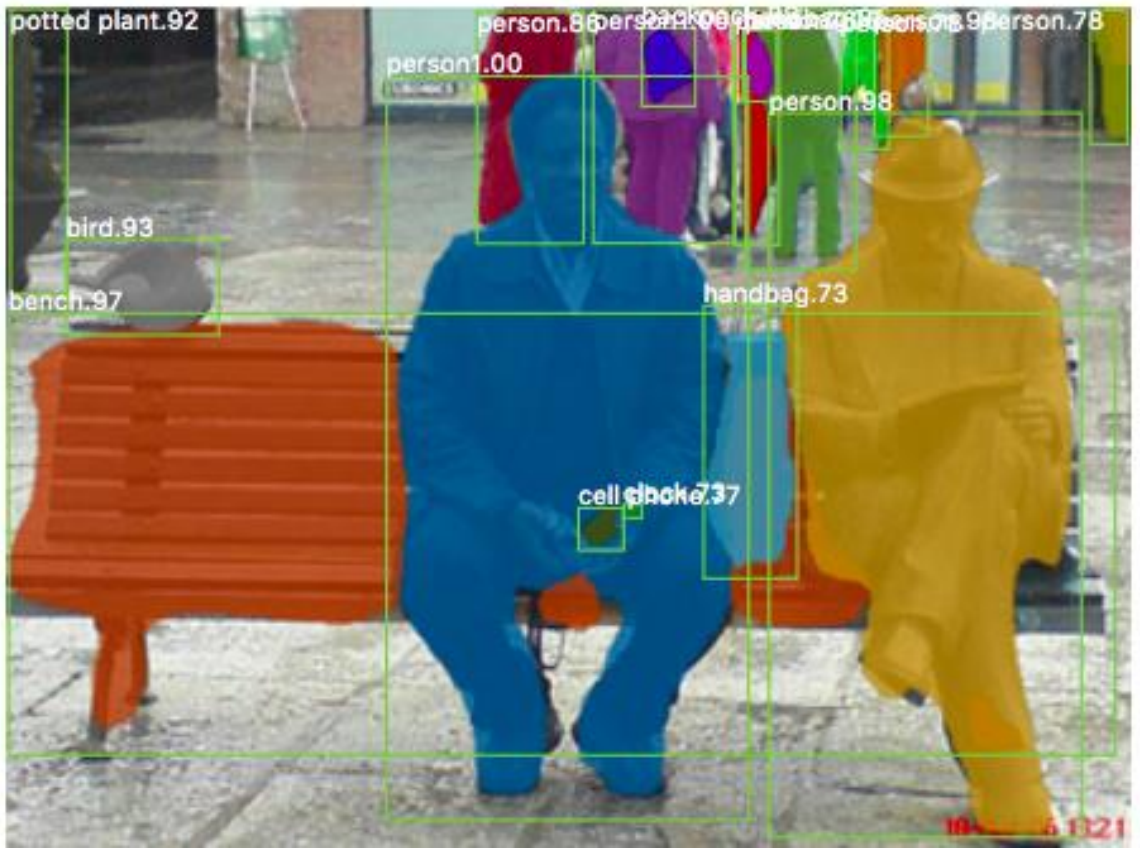




ResNets

Object Detection/Segmentation



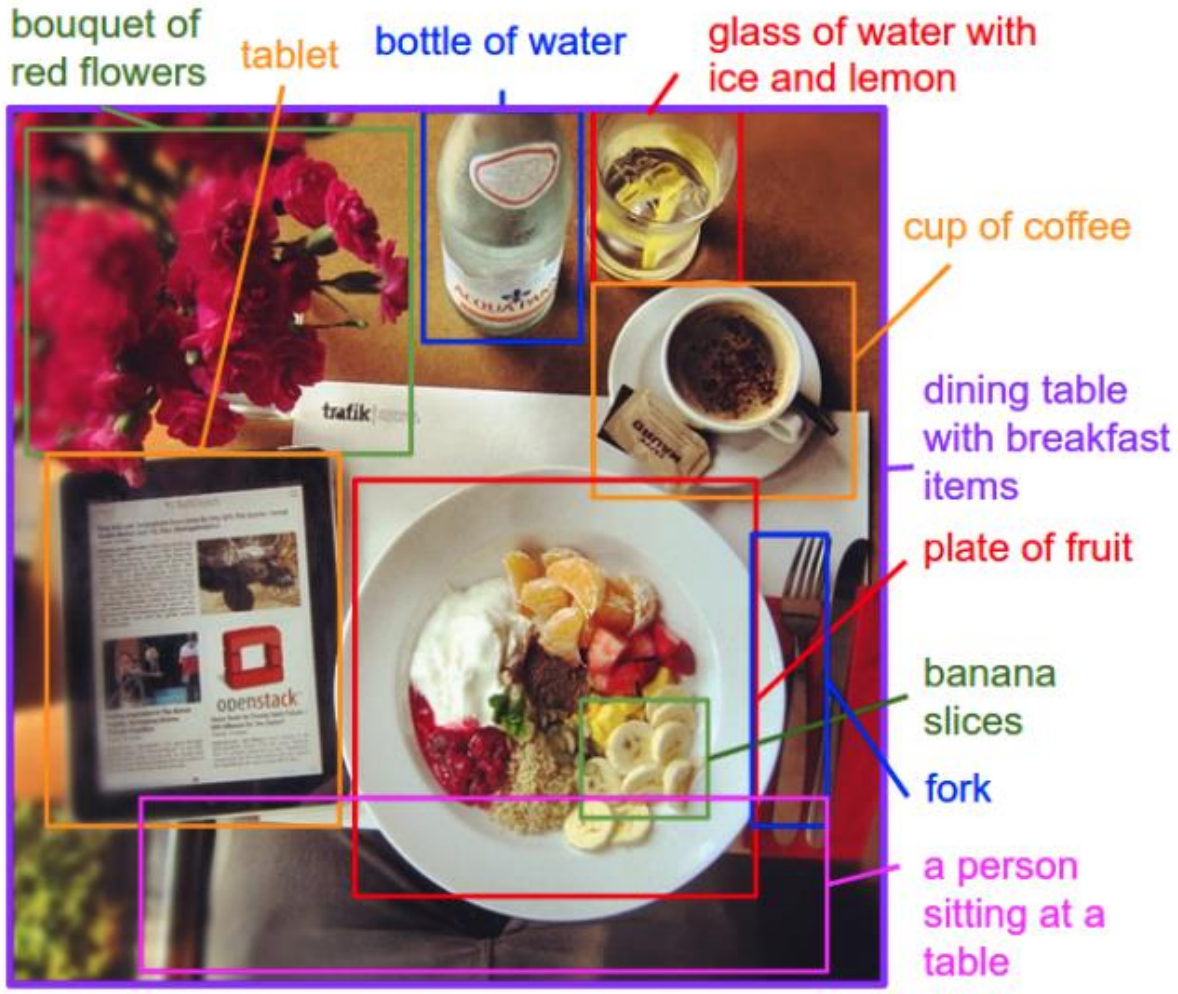


Mask R-CNN. He et al. 2017.

Human Pose Estimation

Mask R-CNN. He et al. 2017.

Image Captioning

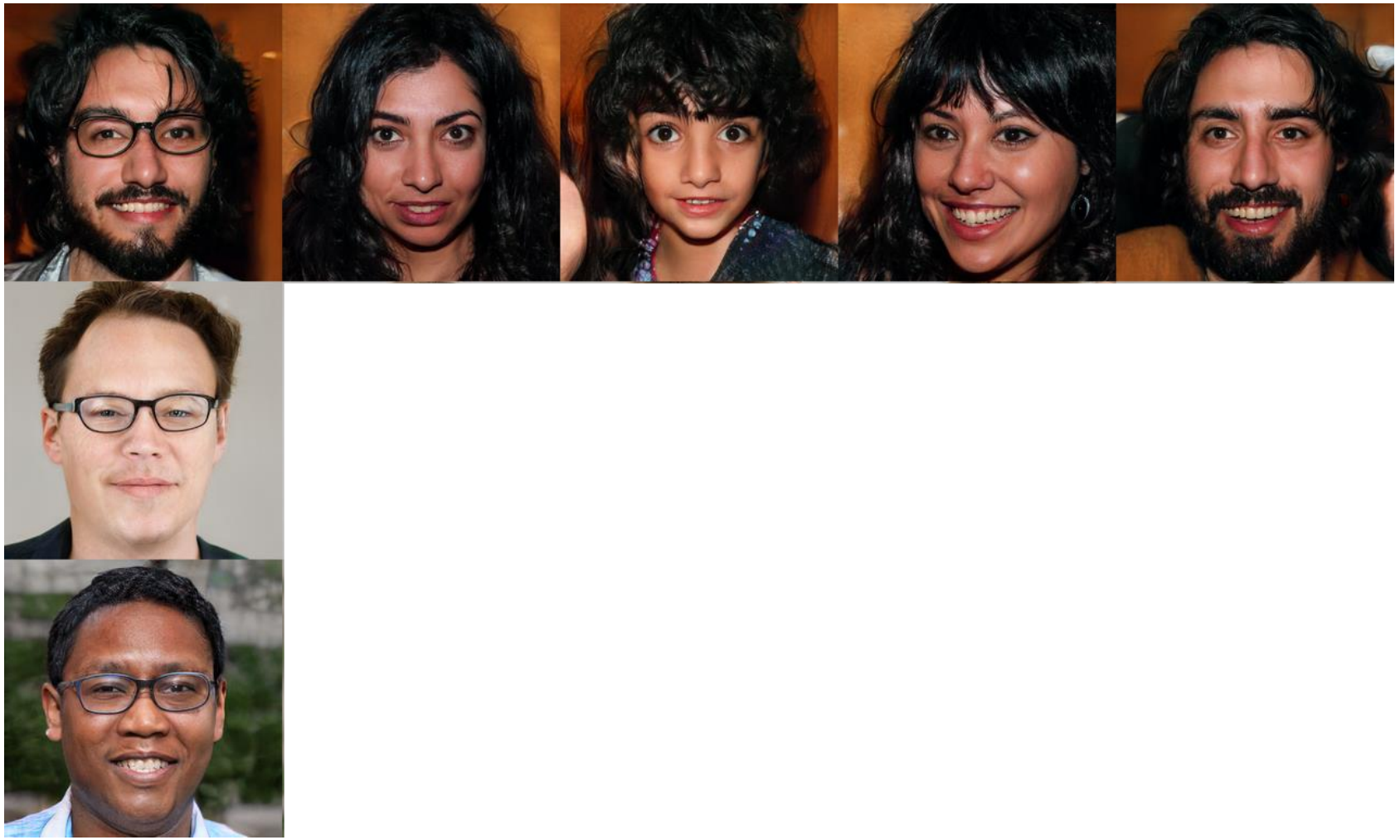


Karpathy et al. 2015.

Image generation

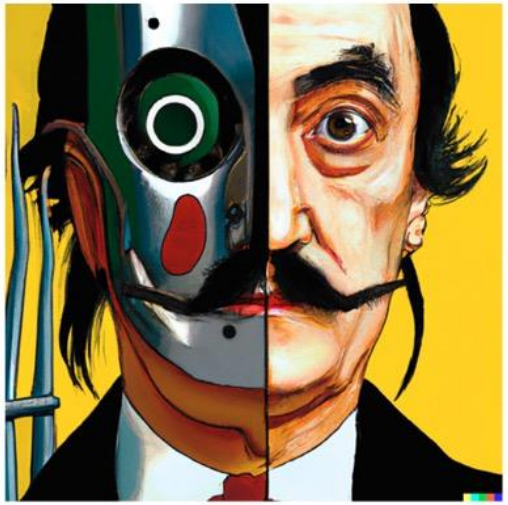
BigGAN. Brock et al. 2019.

Image generation



StyleGAN. Karras et al. 2018.

Image generation with diffusion models



vibrant portrait painting of Salvador Dalí with a robotic half face

a shiba inu wearing a beret and black turtleneck

a close up of a handpalm with leaves growing from it

an espresso machine that makes coffee from human souls, artstation

panda mad scientist mixing sparkling chemicals, artstation

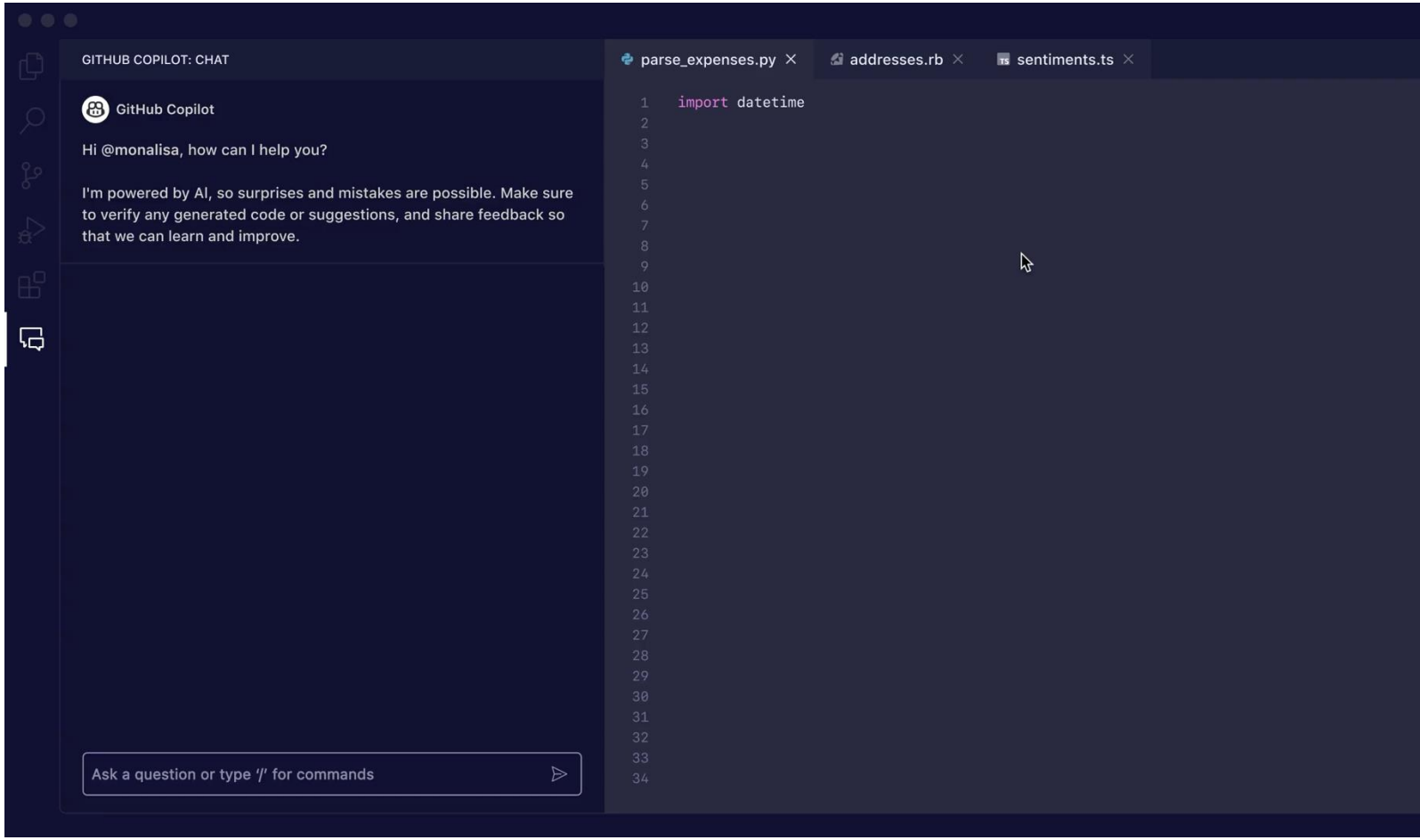
a corgi's head depicted as an explosion of a nebula

Recurrent Neural Networks

```
PANDARUS:
Alas, I think he shall be come approached and the day
When little srain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.
```

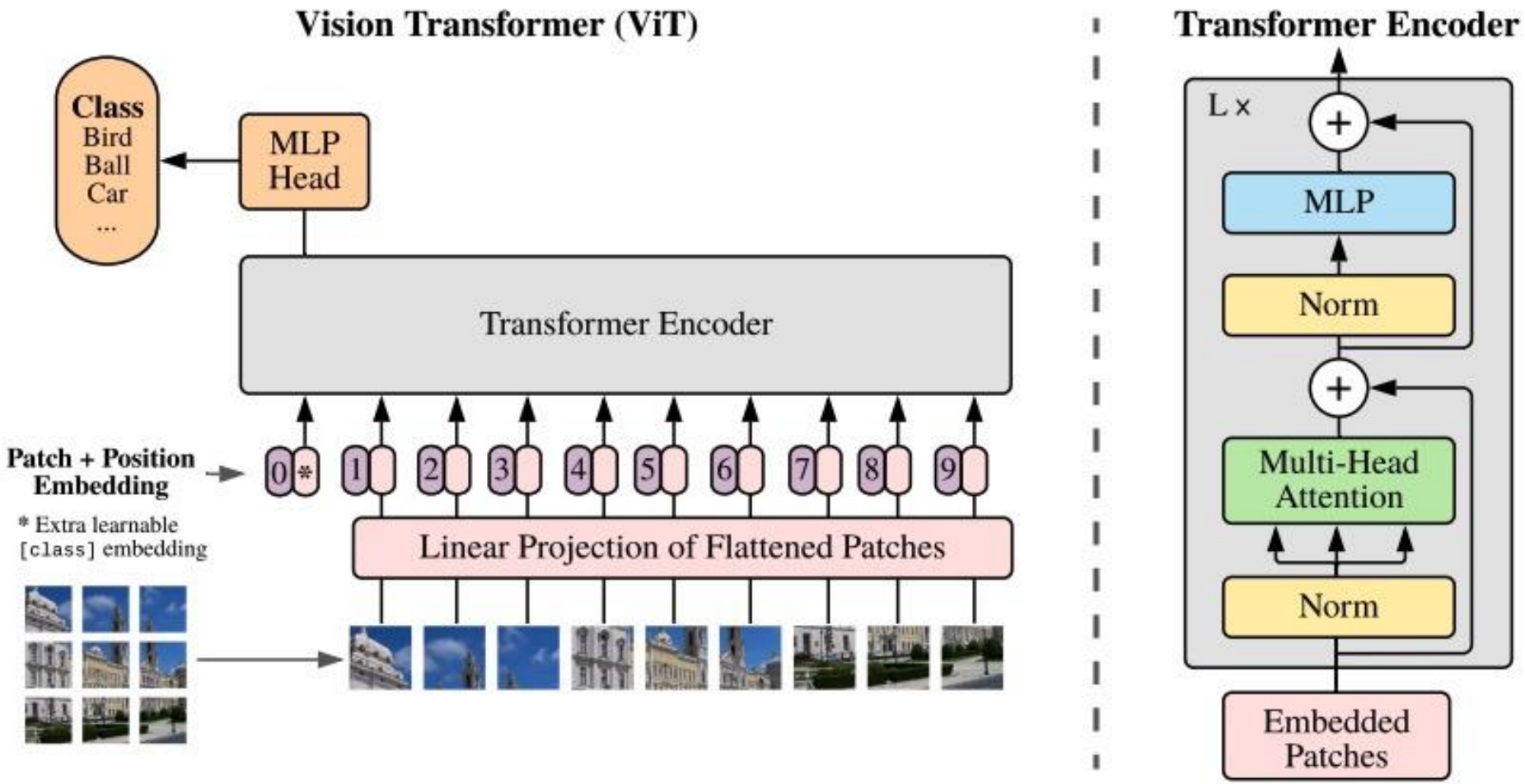
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Language: Transformer, GPT-4



https://github.com/features/copilot

Transformer

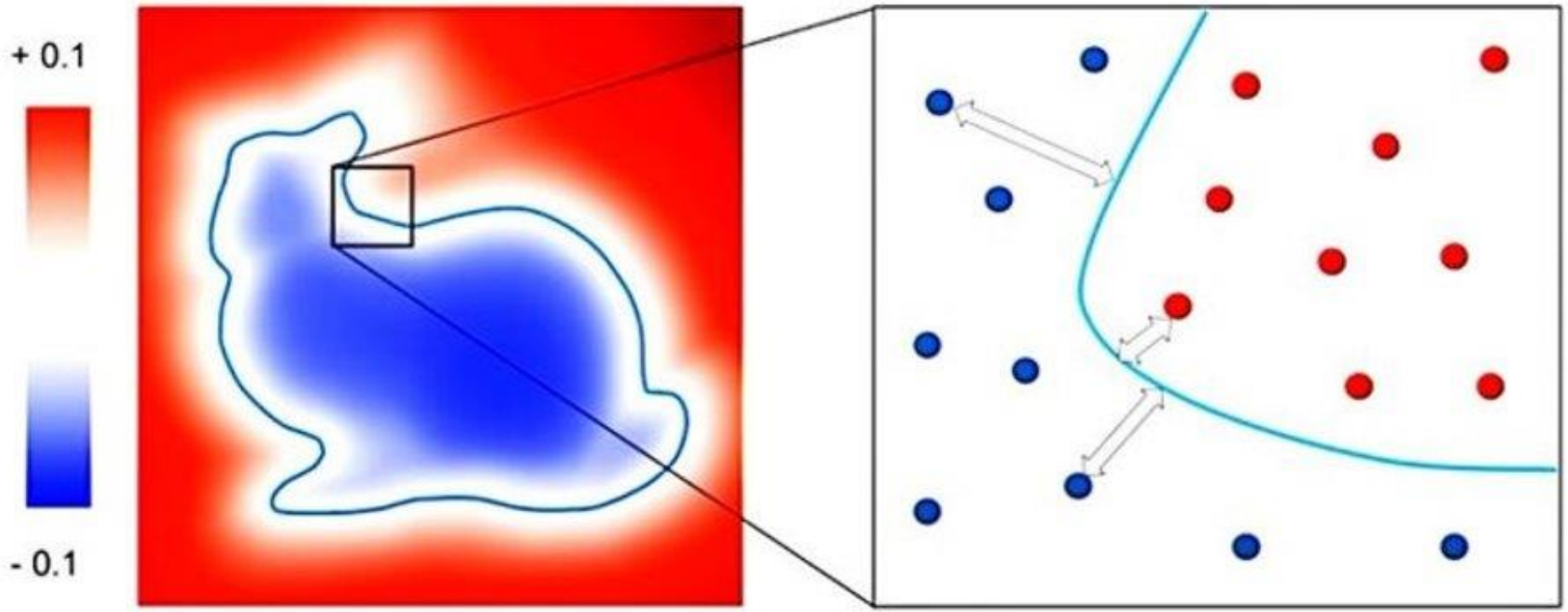


Dosovitskiy et al., 2021

3D Vision: 3D mesh reconstruction from images

Learning Category-Specific Mesh Reconstruction from Image Collections, ECCV, 2018.

3D Vision: Implicit Function

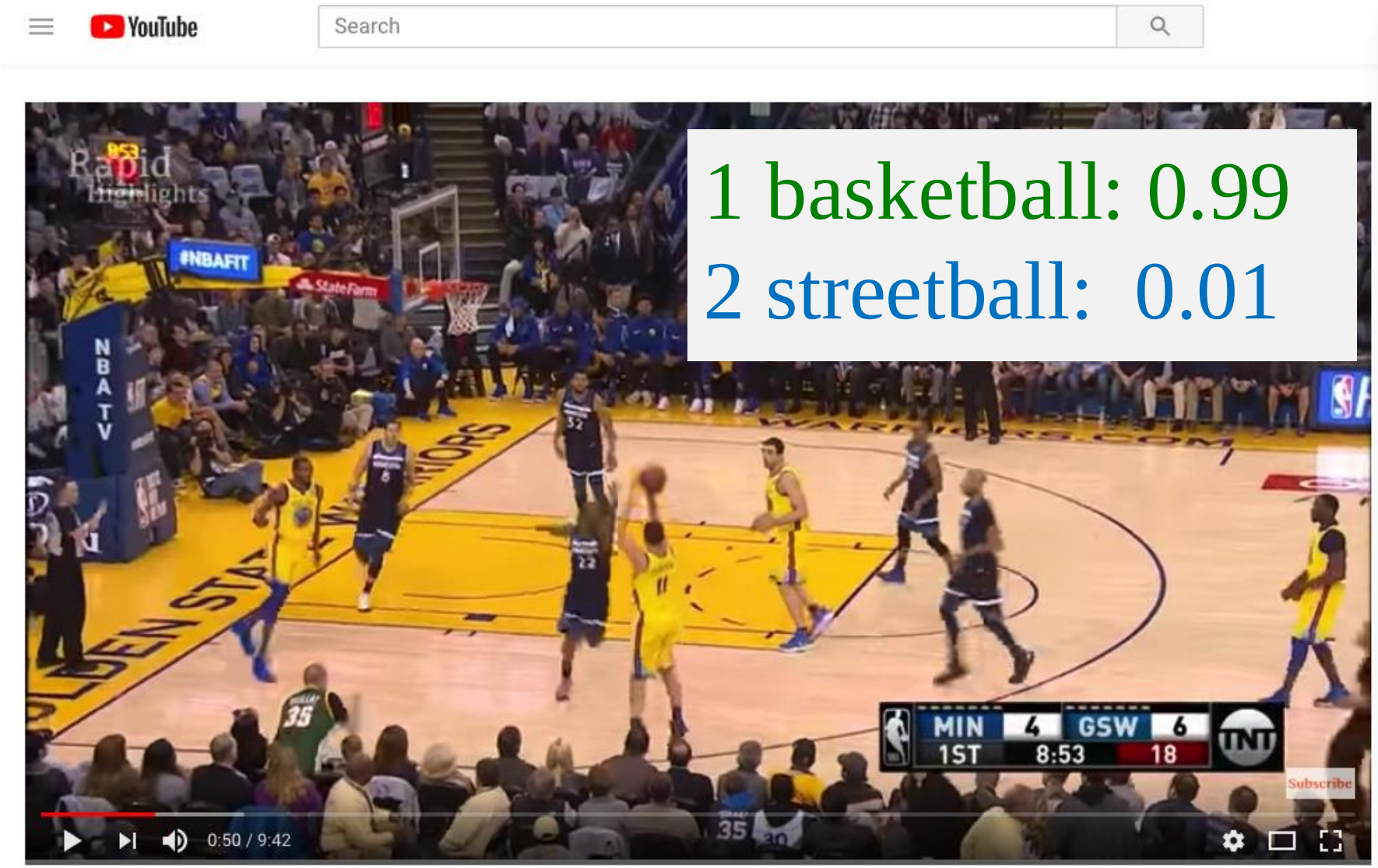


DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019.

3D Vision: Neural Radiance Fields

Mildenhall et al. 2020.

Action Recognition



Video Prediction

Input frames

Ground truth

ℓ_2 result

ℓ_1 result

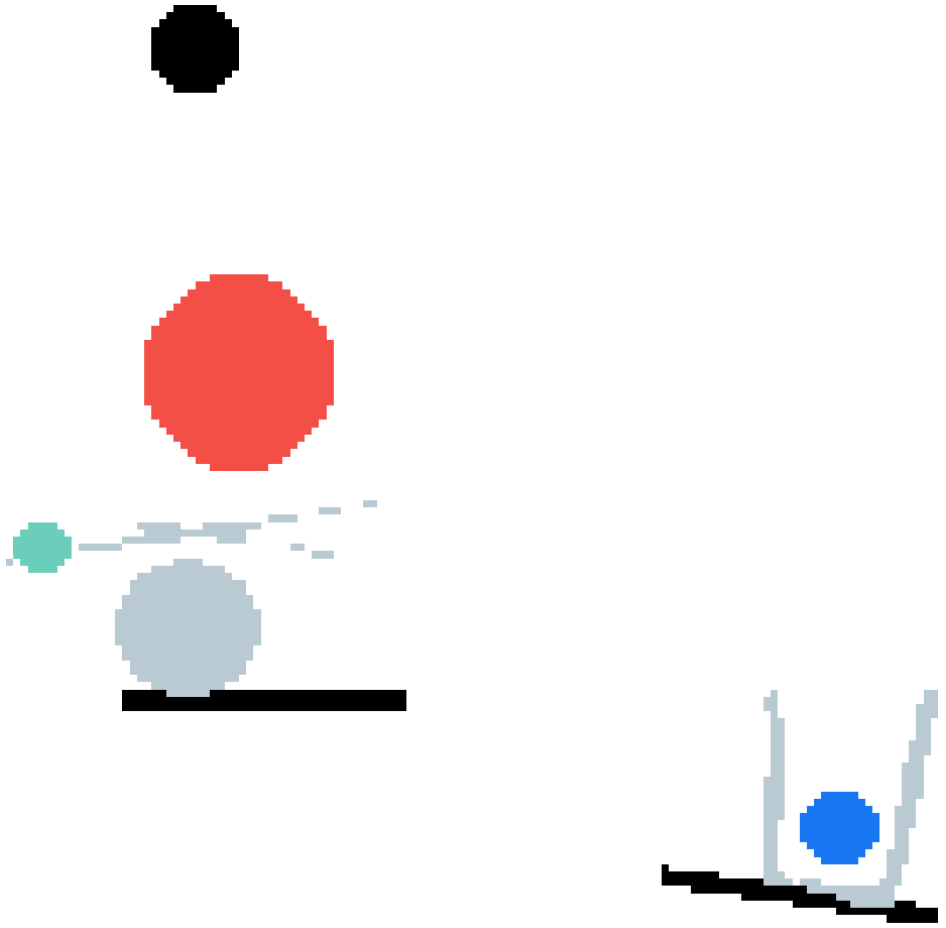
GDL ℓ_1 result

Adversarial result

Adversarial+GDL result

Mathieu et al., 2015

Physical Interaction Prediction



Qi et al., 2021

Statistical learning, Training and Testing

- Training: Learning from the past experience:
  - training dataset
  - demonstrations

- Testing: Generalize to unseen inputs
  - Data that does not exist in training set

Image Classification

input	desired output
	apple
	pear
	tomato
	cow
	dog
	horse

Credit: Svetlana Lazebnik

Image Classification

input **desired output**

apple

pear

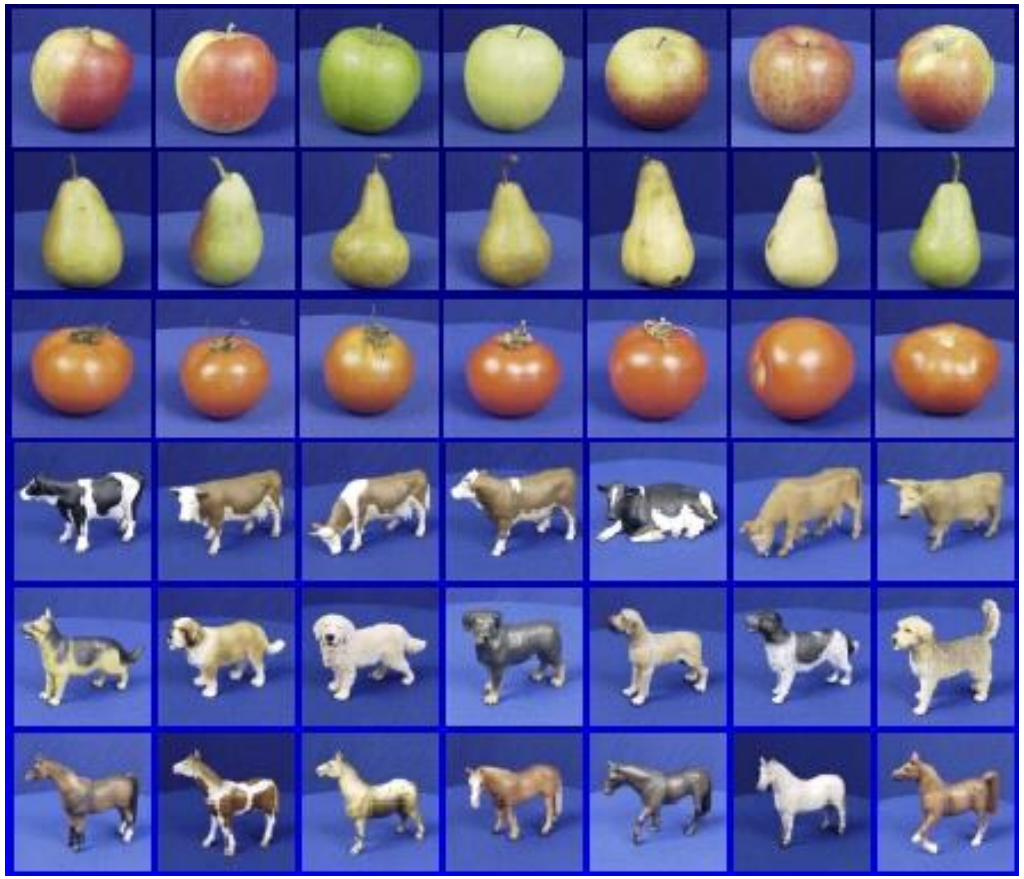
tomato

cow

dog

horse

training data



apple

pear

tomato

cow

dog

horse

Credit: Svetlana Lazebnik

Training time

Training Samples

Training Labels

Features

Training

Learned model

Testing time

Test Sample

Features

Learned model

Prediction

Credit: Svetlana Lazebnik

Supervised Learning

$$y = f(x)$$

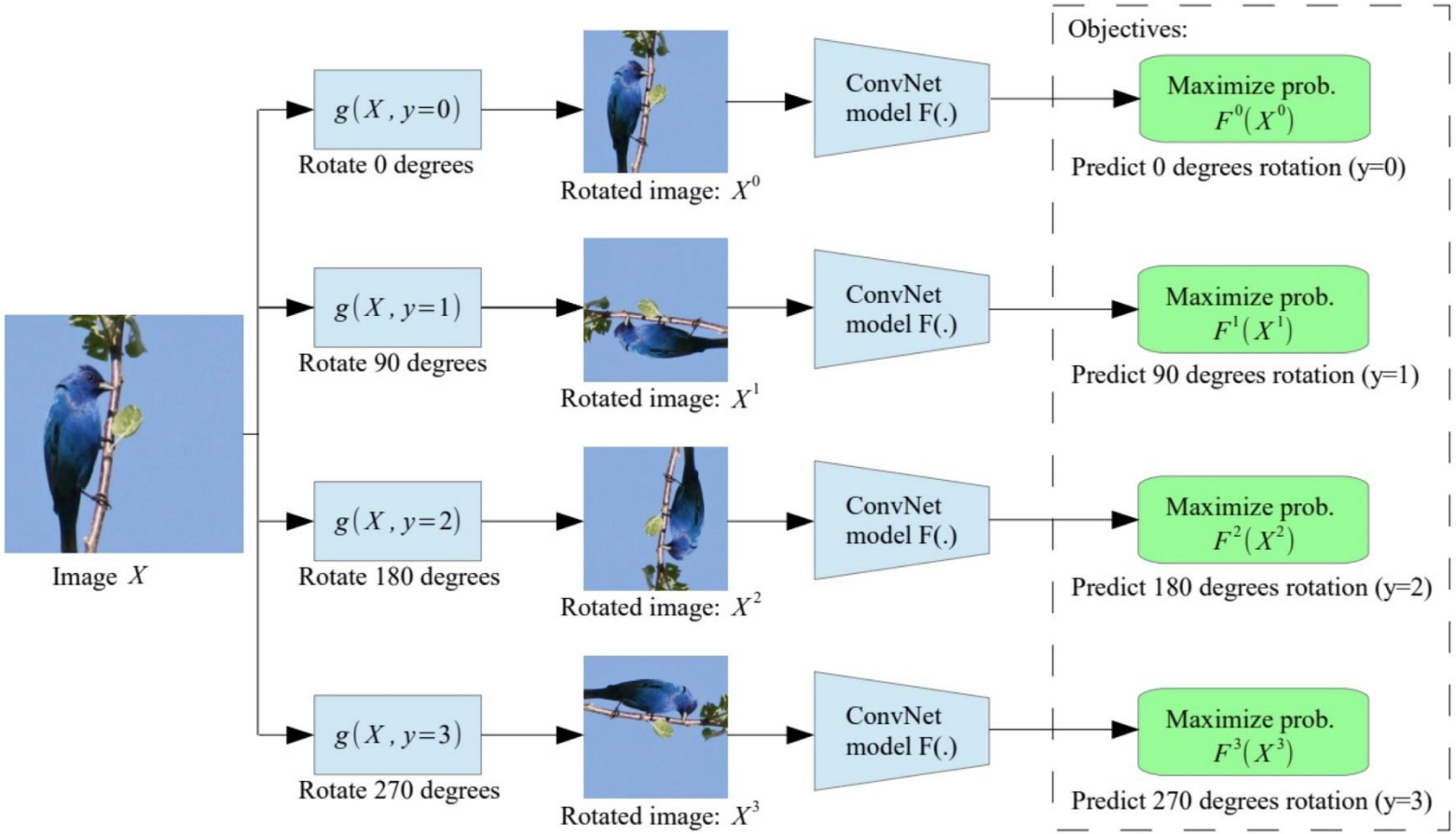
output label — neural network — input image

- **Training** (or **learning**): given a *training set* of labeled examples $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, train a neural network predictor f

- **Testing** (or **inference**): apply neural network f to a new *test example* x and output the predicted value $y = f(x)$

Transfer Learning

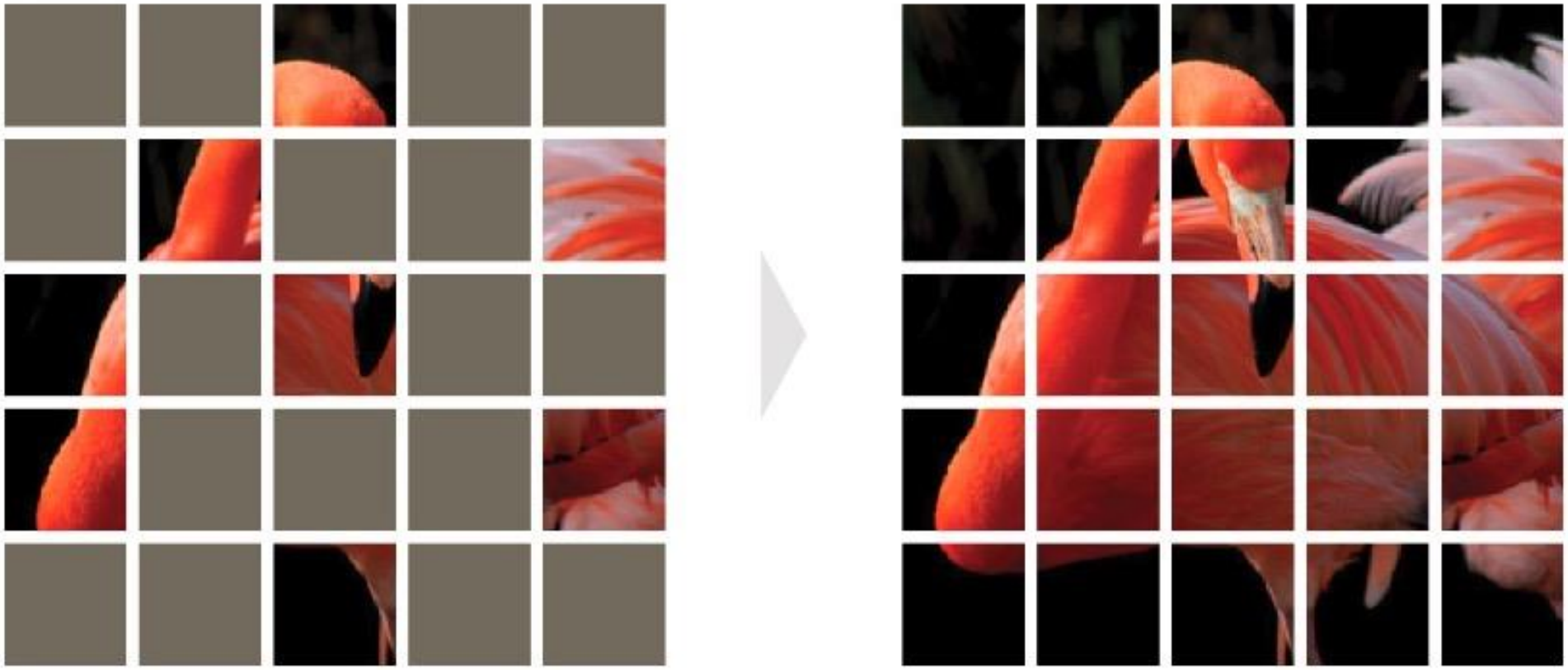
Self-Supervised Learning



[Gidaris et al. 2018]

Self-Supervised Learning with Rotation Prediction

Self-Supervised Learning with MAE



He et al. 2022

This Class

- Computer Vision Research with Deep Learning
- Training and Testing

Coming Tutorial

- There will be a tutorial on how to do/submit assignments This Wednesday, 3:00 - 4:00 pm PST on zoom

- We will use the compute resources in https://datahub.ucsd.edu/