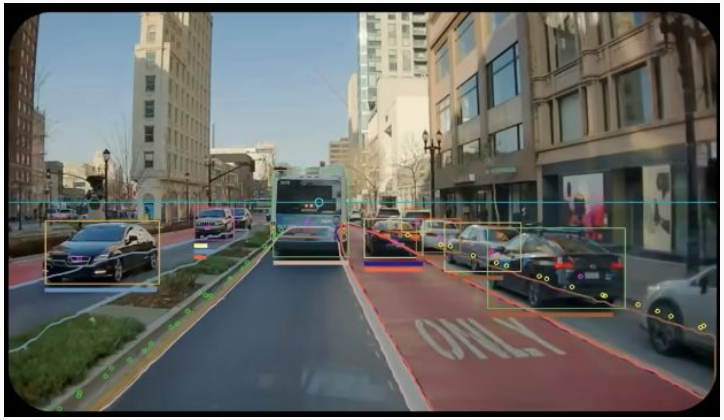
Intro & Topics Overview

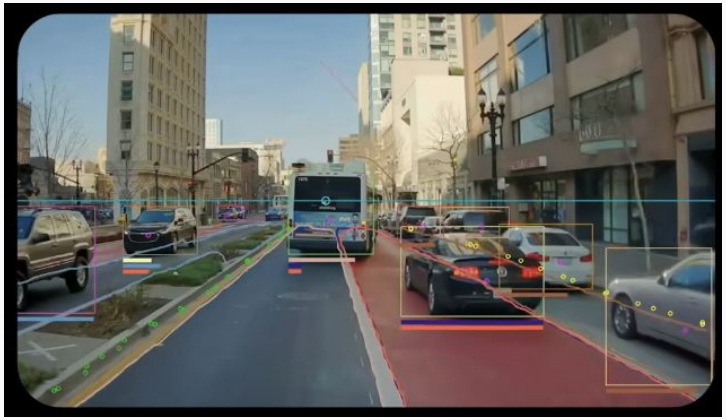
CSCI 611, Summer 2025, Week 1 Session 1

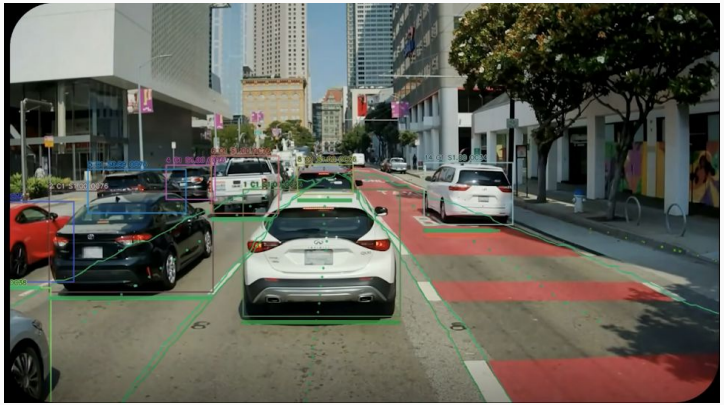
About Me

- PhD Computer Science, Wayne State University
- Hewlett Packard Laboratories, researcher on computer vision, video coding, streaming and content distribution; extensively published.
- Vuclip, Streaming Media Startup, from Series A to acquization, Software Architect → VP Engineering → Chief Technology Officer
- HaydenAI, Vision-base AI Startup, from beginning to post Series B (2024), co-founder and Chief Technology Officer → Chief Science Officer

Past Work



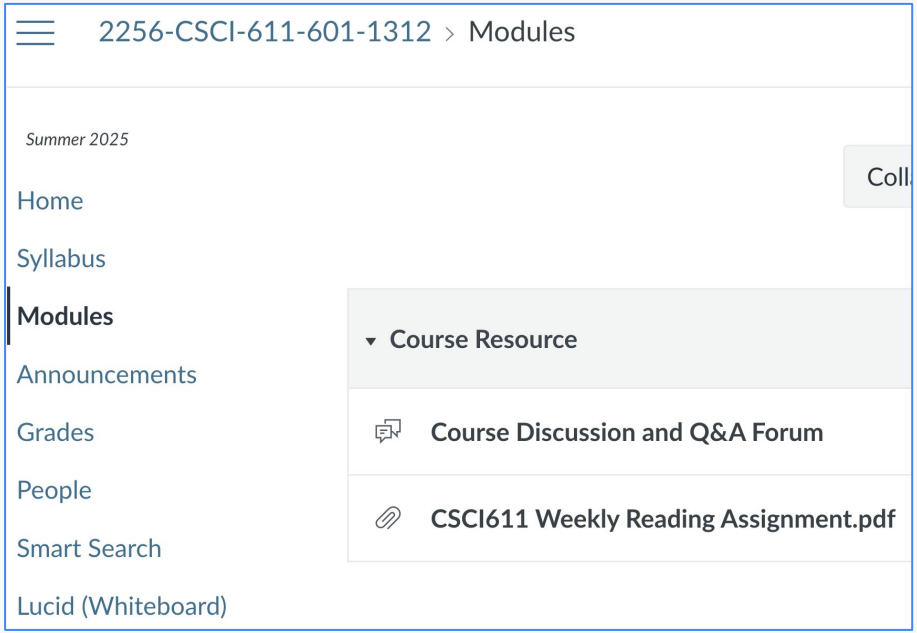




Your Turn [share in Canvas/Discussion]

- Survey
  - Python Experience?
  - C++ Experience ?
  - Jupyter Notebook or Google Co-Lab?
  - Github Experience?
- More to share?
  - Your Name, number of years
  - Your Expectation
  - Your Favorite Topics in Machine Learning
  - Questions?
  - ...

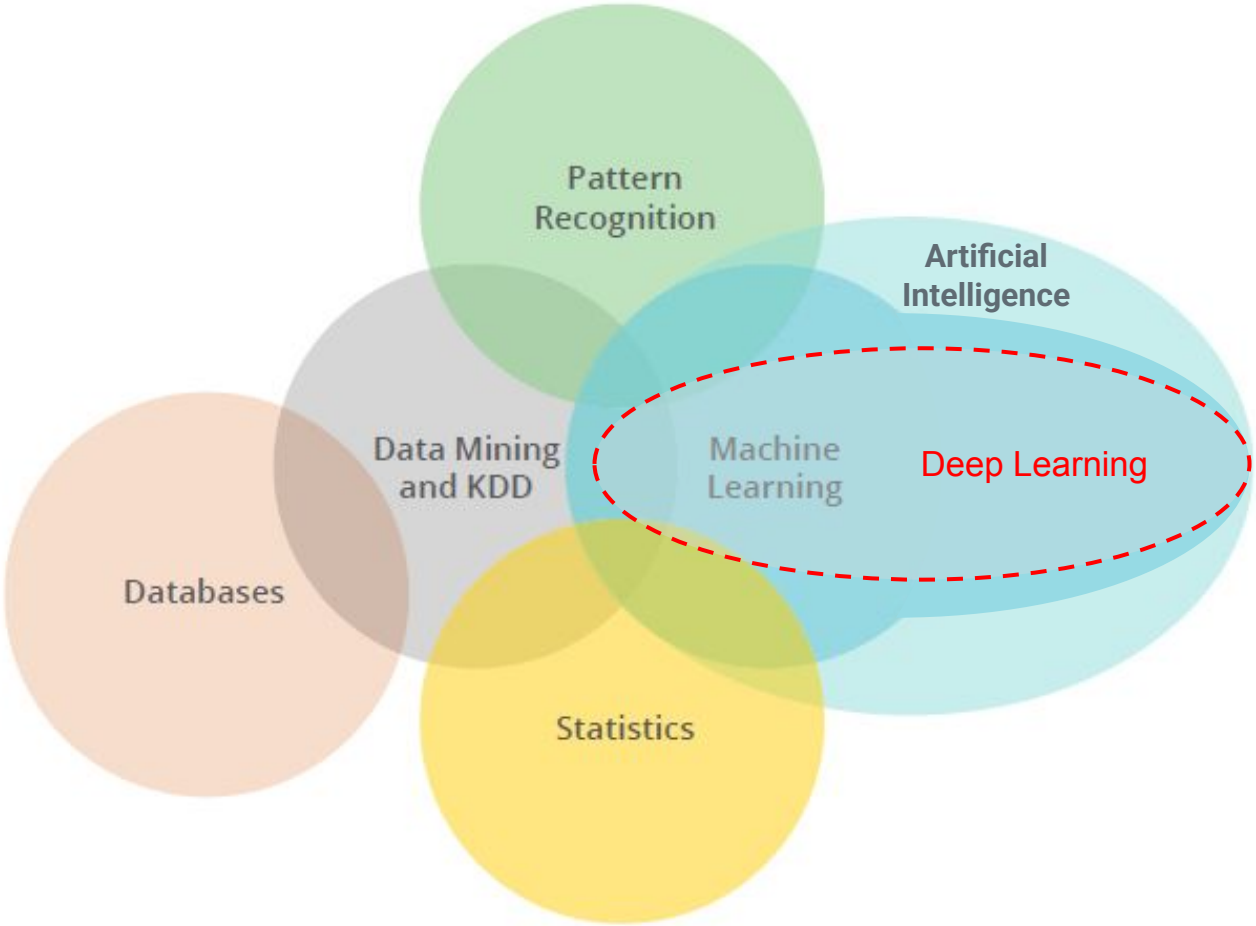
Discussion Forum & Syllabus



- Syllabus is available on Canvas: https://canvas.csuchico.edu/courses/42089/assignments/syllabus

Data → Intelligence

structure feature semantic



original graph from: https://www.altexsoft.com/whitepapers/machine-learning-bridging-between-business-and-data-science/

611 Topics

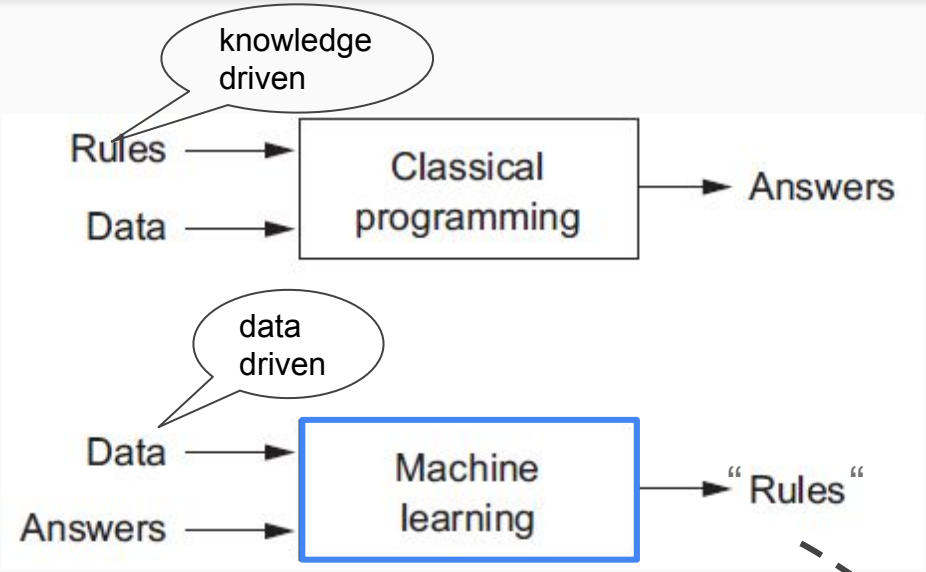
Machine Learning → Deep Learning

- Categories of ML
  - Statistical
  - Supervised (Neural Network based)
  - Unsupervised
  - Reinforcement
- Neural Nets Based ML
  - Multilayer Perceptron
- Neural Nets Based DL
  - Convolutional Neural Nets
  - Residual Neural Nets
  - Deep Neural Nets
  - Edge Neural Nets

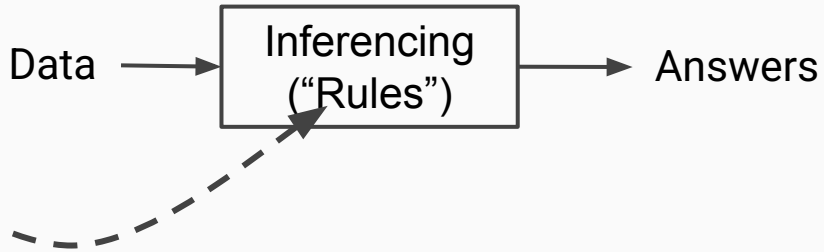
DL → Generative AI → GPT → AGI/ASI

- **Feature** → Semantics
  - Style Transfer
  - GAN → Auto Encoder
- **Memory**: Recurrent Neural Nets
  - RNN → LSTM
  - LSTM → BiLSTM → BART
- **Context**: Transformer
  - BERT → Transformer
  - Vision Transformer
- **Large Scope Context**: GPT
  - Large Language Model
  - LLM → LCM → WM

Classic AI → Machine Learning

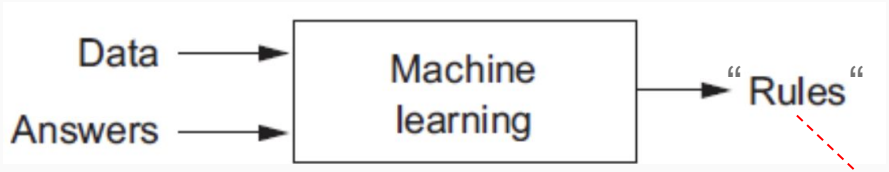


Rules =?= Knowledge → Intelligence



Machine Learning Full Circle

Learning Flow:



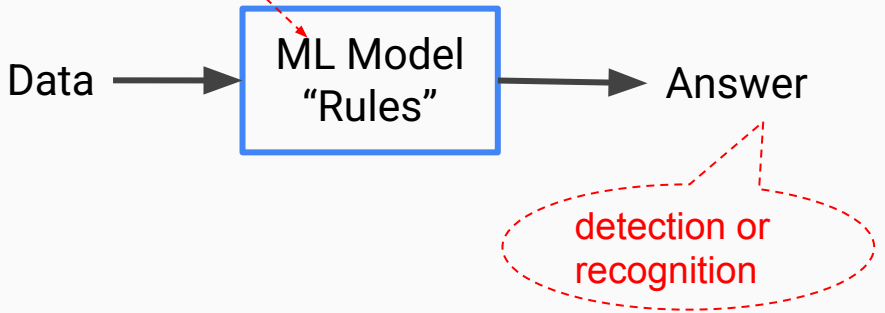
Linear model parameters (y = **m**x + **b**)

Decision tree structure

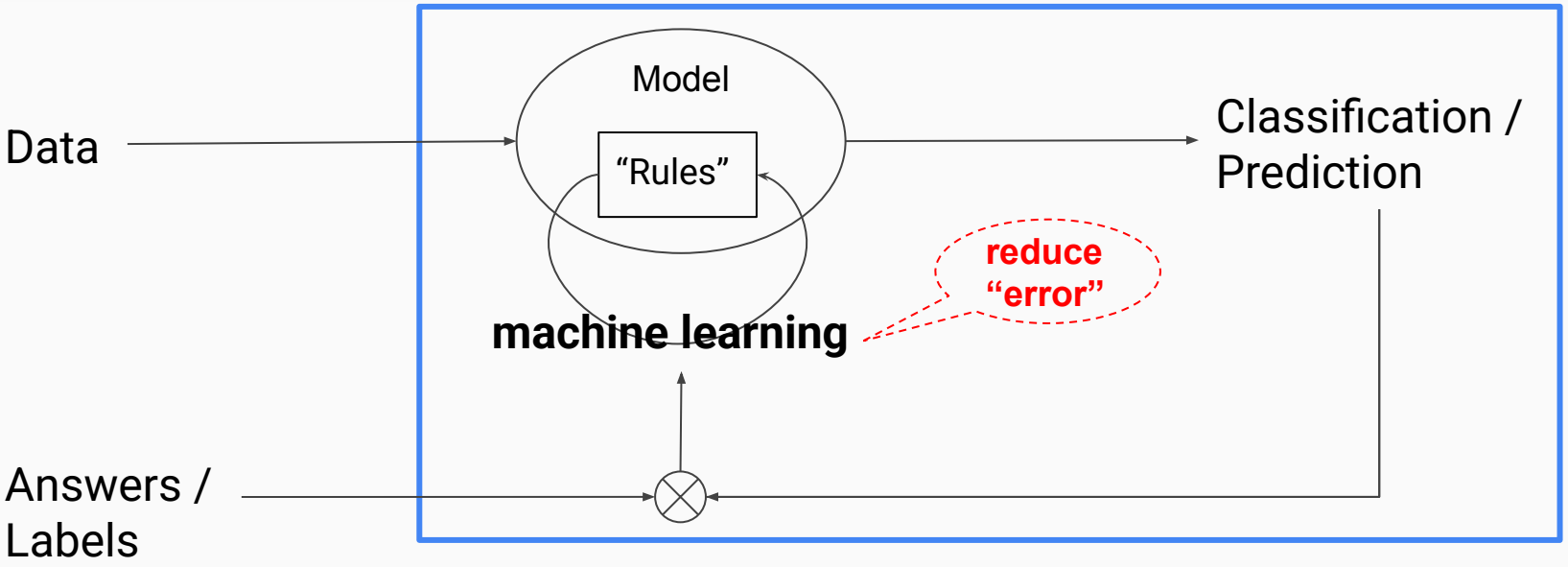
Weights in Neural Nets

Learned
Knowledge
Intelligence

Inferencing Flow:

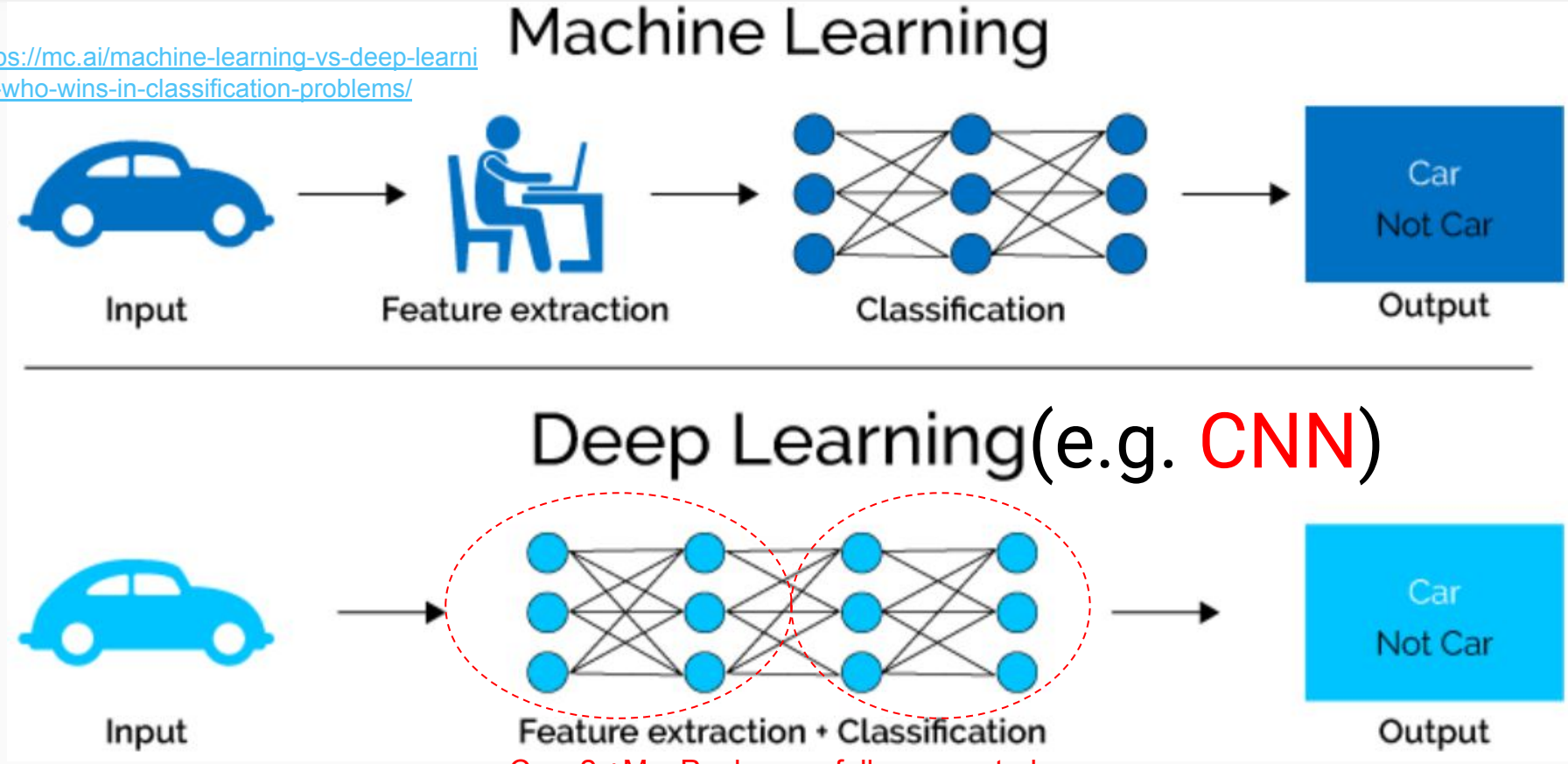


Principle of Machine Learning: Reducing Error → Training



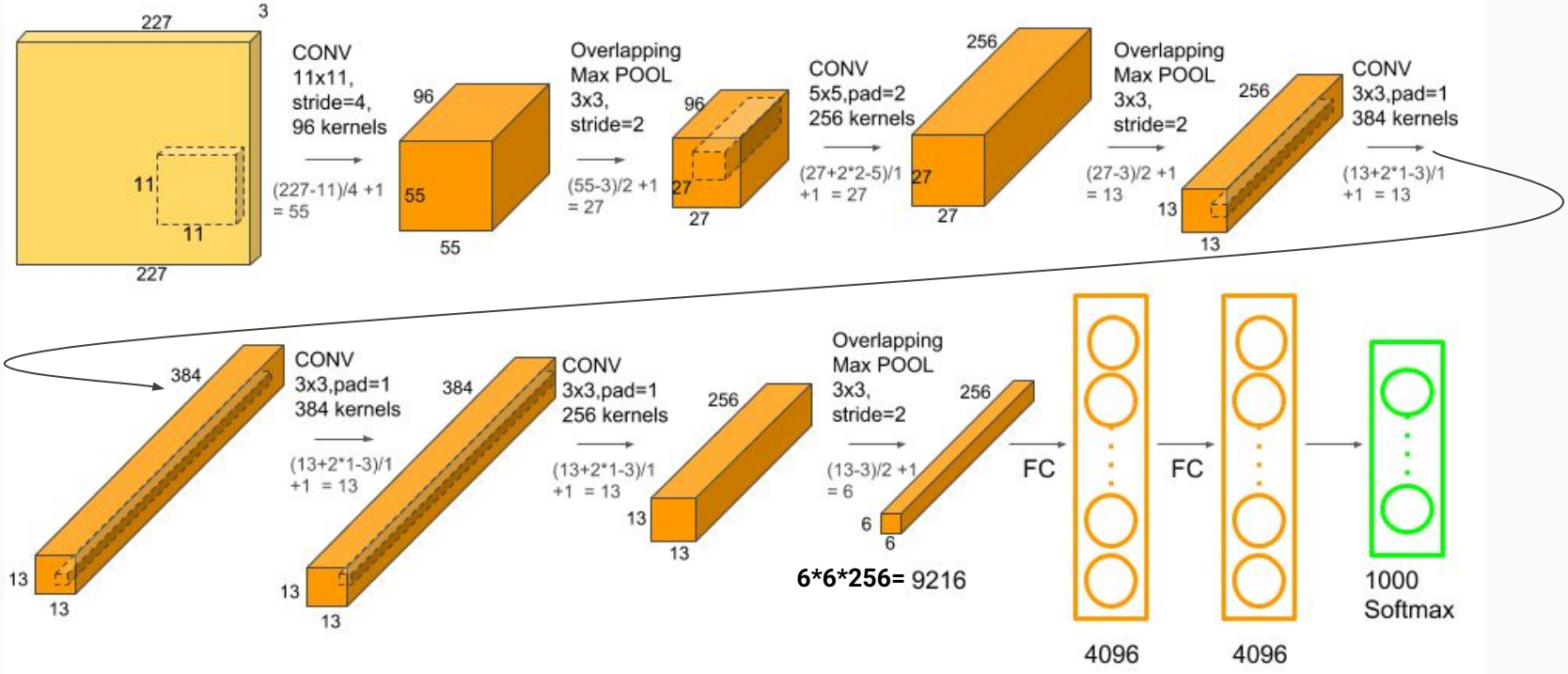
Machine Learning → Deep Learning

https://mc.ai/machine-learning-vs-deep-learning-who-wins-in-classification-problems/

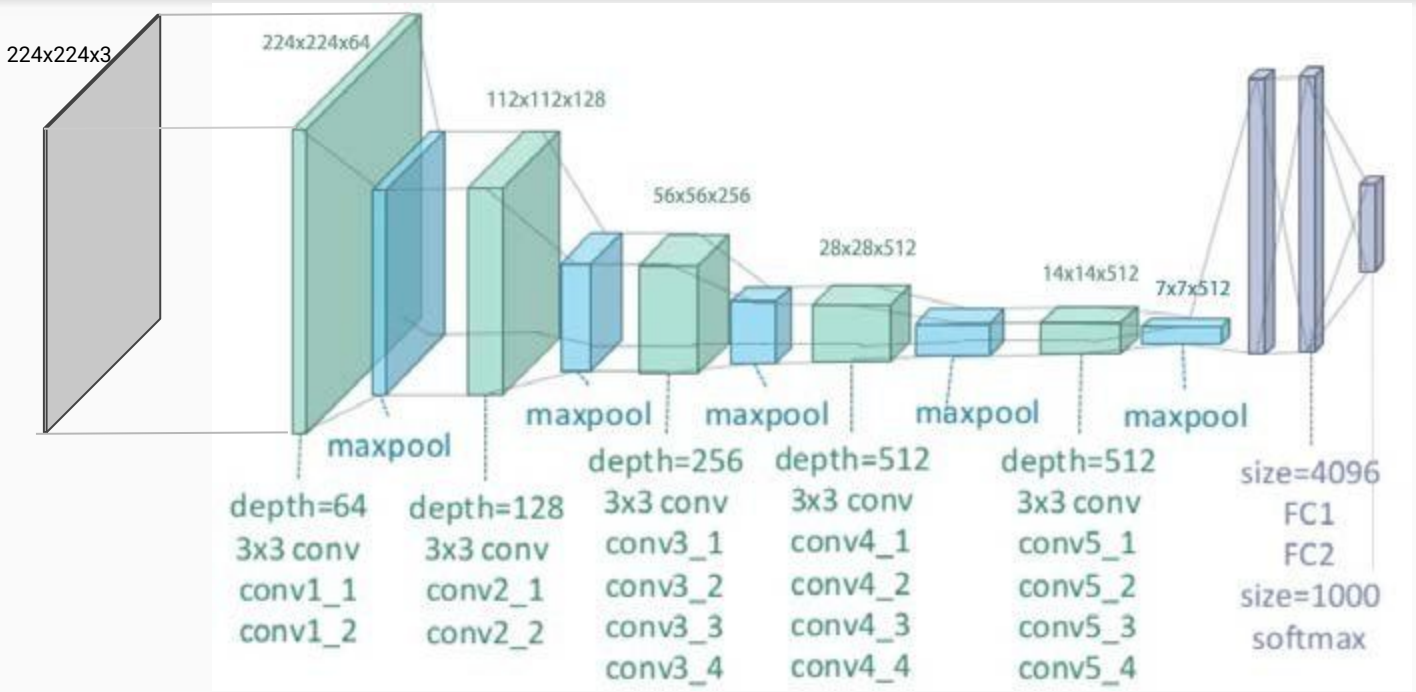


AlexNet

https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/



VGG19

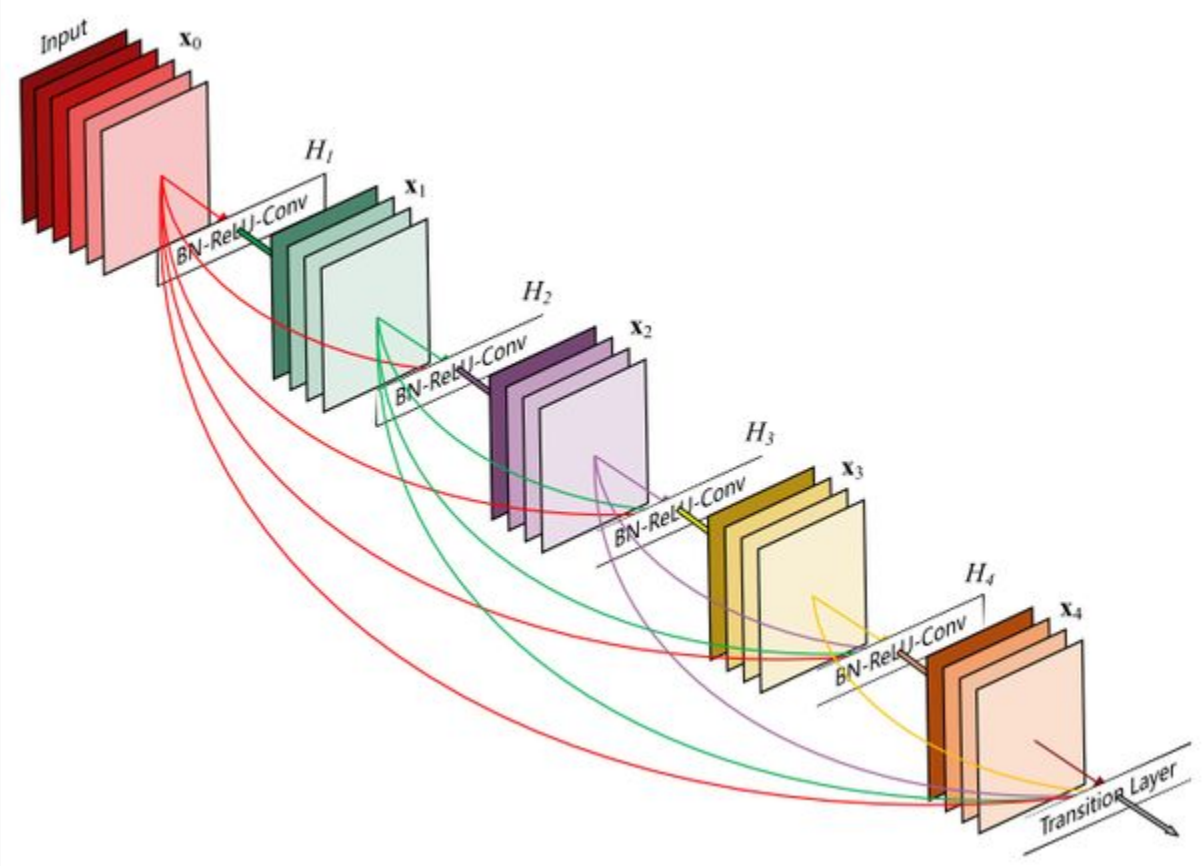


- 16 conv layers (conv1_1 to conv5_4) + 3 fully connected layers = 19 layers
- all small 3x3 kernel

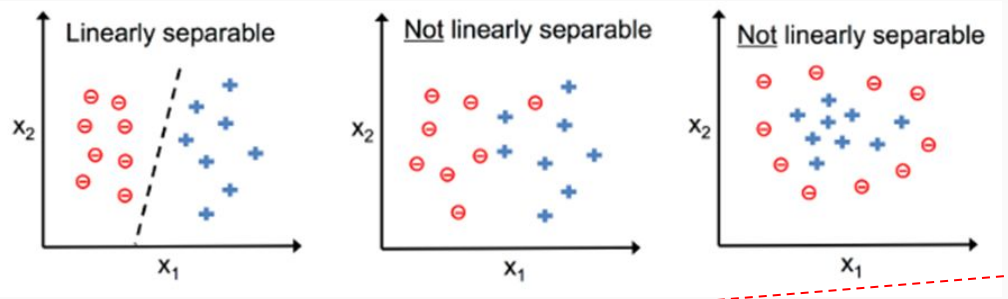
ResNet Extension

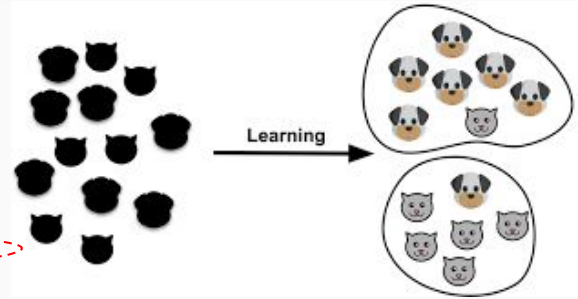
DenseNet: connects all layers directly with each other. the input of each layer consists of the feature maps of all earlier layer, and its output is passed to each subsequent layer. ⇒ Encourage feature Reuse

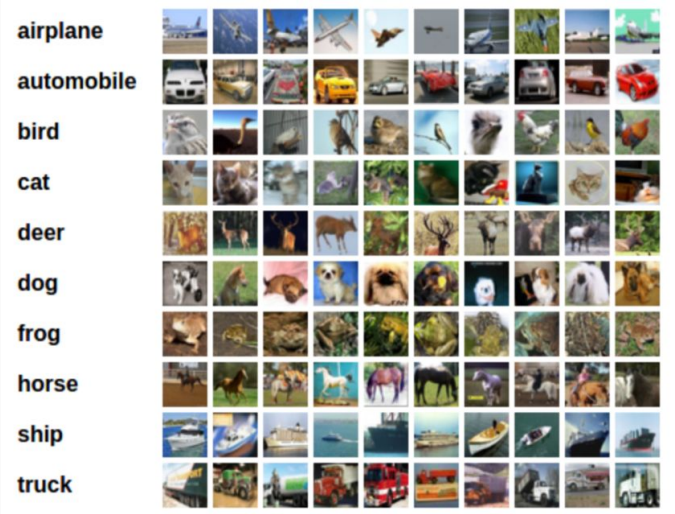
ResNet can be considered as an ensemble of many smaller networks



ML Problems and Applications (Basics)



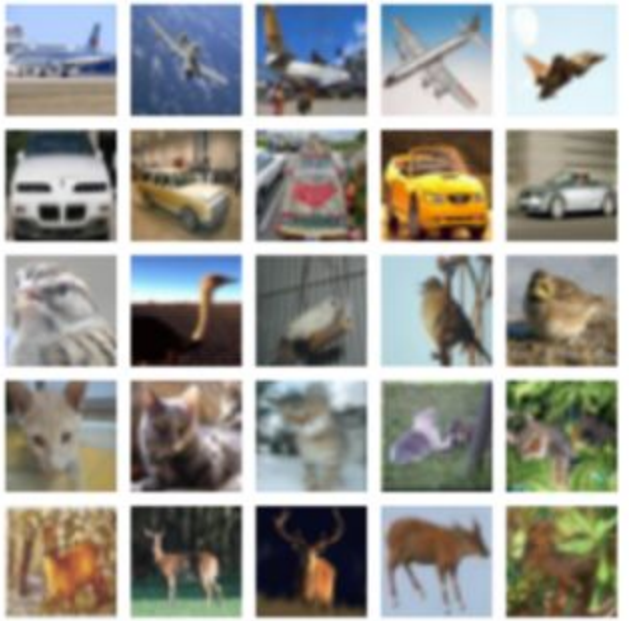


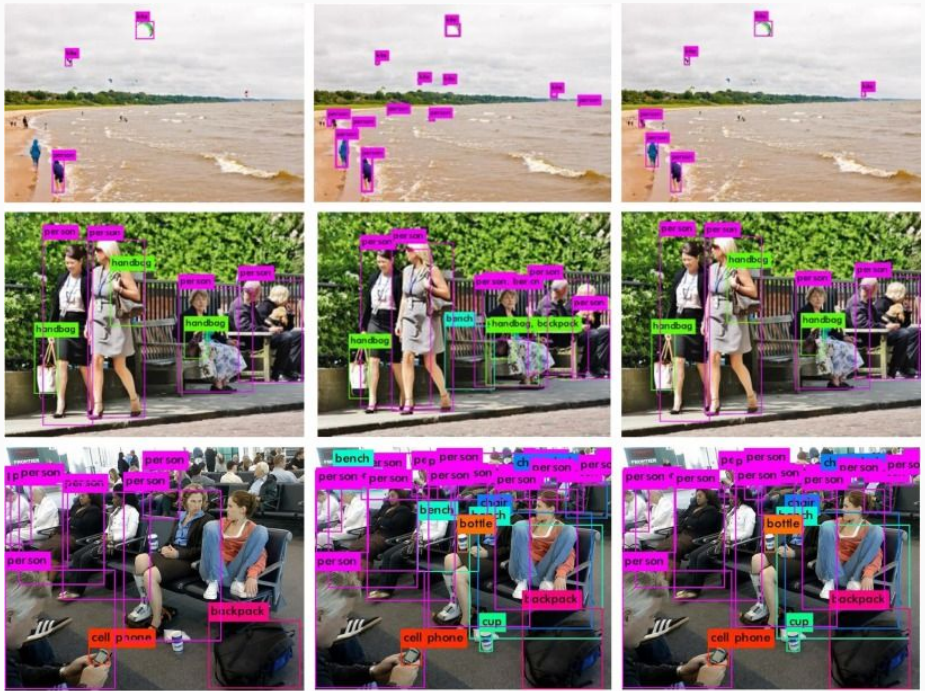


Imagenet (https://www.image-net.org/download.php)

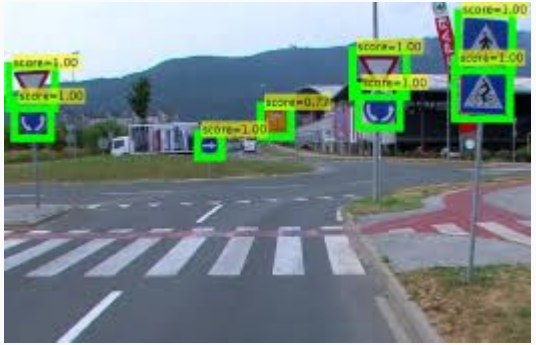
The most highly-used subset of ImageNet is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012-2017 image classification and localization dataset. This dataset spans **1000** object classes and contains **1,281,167** training images, 50,000 validation images and 100,000 test images. This subset is available on Kaggle.

More Than Object Classification?



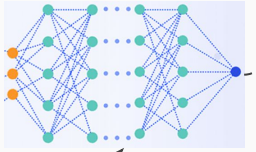


Object + Location ⇒ Detect Small Object

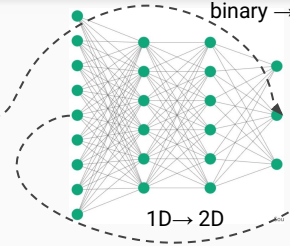


My prior work ⇒ pptx
Further research opportunities

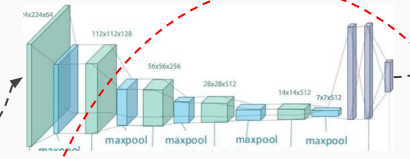
Coverage So Far



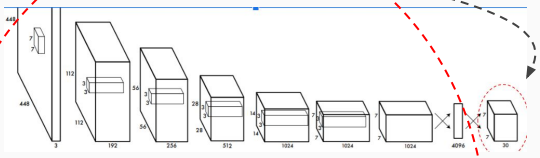
single → multilayer
linear → nonlinear bound



binary → multiclass
1D→ 2D



single object→ multiple objects /w attributes per picture



Perceptron

activation / loss
gradient descent
back propagation

multiclass ⇒

MLP

dense/fully connected
sigmoid→ softmax
normalization
dropout

image ⇒

CNN

conv for AutoFE
maxpooling for multiscale
overfitting

⇩ AlexNet

vanishing gradient

⇩ VGG19

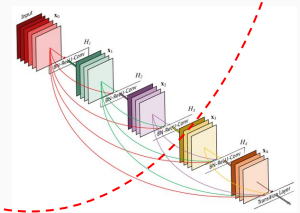
residual connection

⇩ ResNet

feature reuse
batch normalization

deeper learning
higher accuracy

▼ DenseNet



multiple objects
+ location ⇒

YOLO

multiple, overlapped with location
speed (1x1 conv)
depthwise conv
real time → video

⇩

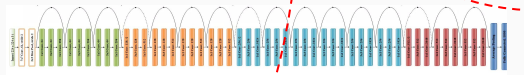
SSD

use multiple feature maps

⇩

higher speed
higher perf.
lower power

▼ MobileNet (EdgeML)
EfficientNet



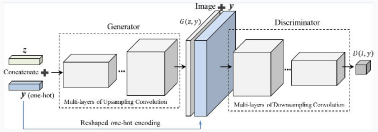
style transfer ⇐

StyleGAN

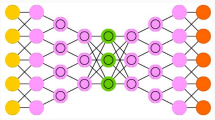
generator
discriminator
content derivation

⇩

GAN



⇩



Auto-Encoder

Variational AE

feature extraction → embedding
compression ← dimension reduction

Before Generative AI

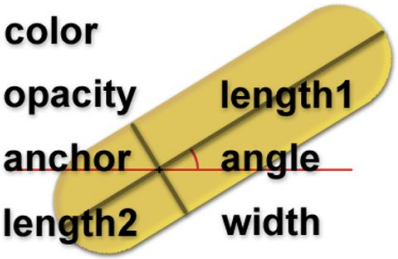
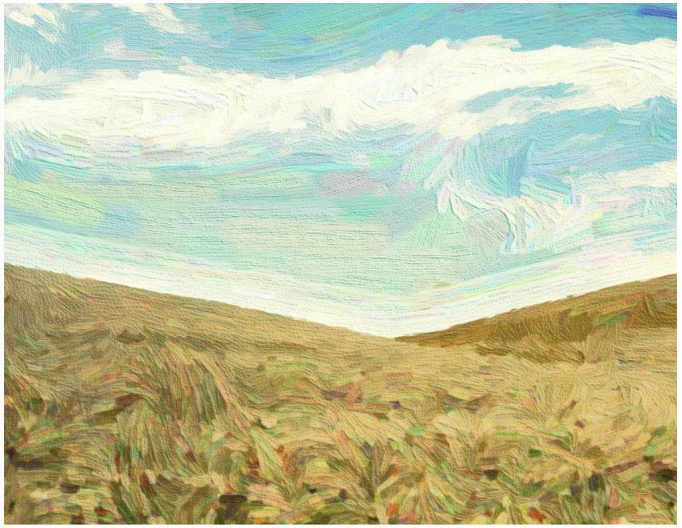


Figure 2: Each Brush stroke is an object with the following properties: Anchor point in image coordinates, angle of orientation in degrees, width in pixels, lengths in both directions from the anchor point in pixels, color (R, G, B) for the current and past several frames, and opacity. In addition the brush stroke knows if it is "New" and/or if it is "Strong." Anchor, angle, width, and lengths are kept in floating point coordinates.

https://faculty.cc.gatech.edu/~hays/papers/IVBPA_Final.pdf



Extracted Feature by CNN

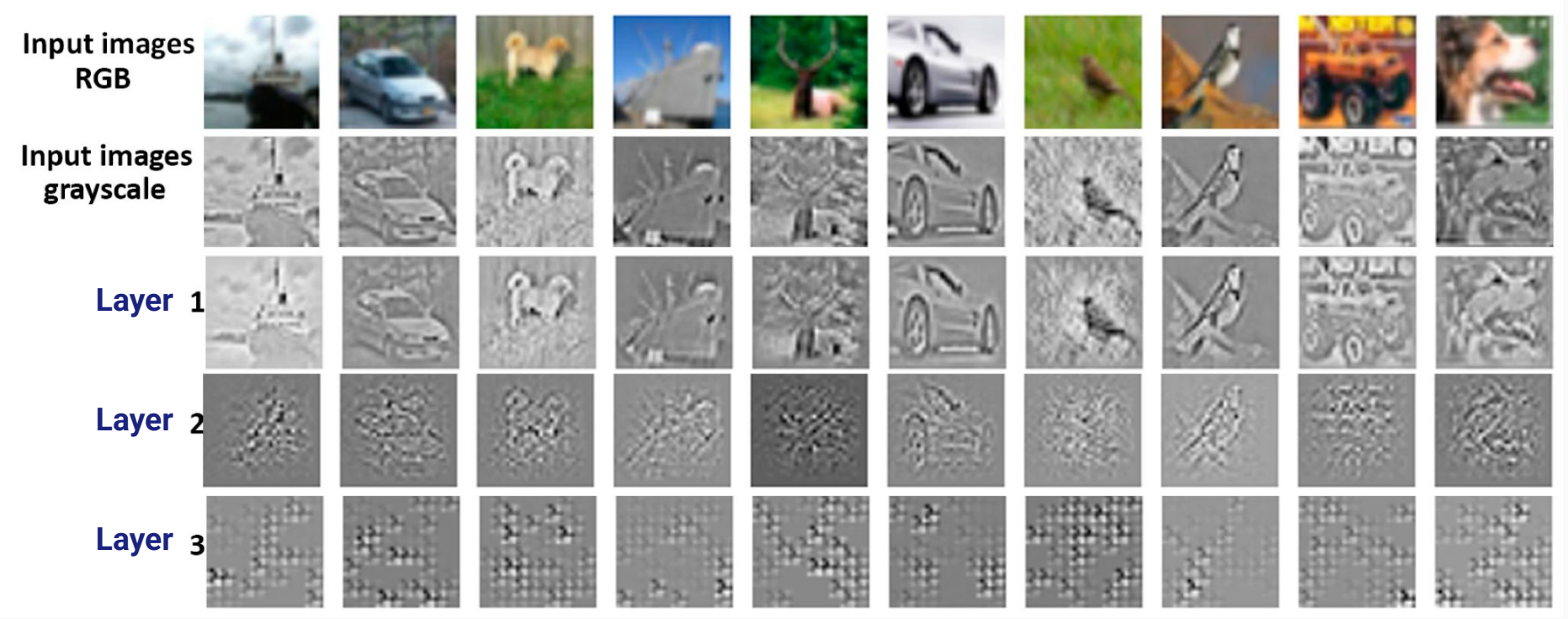
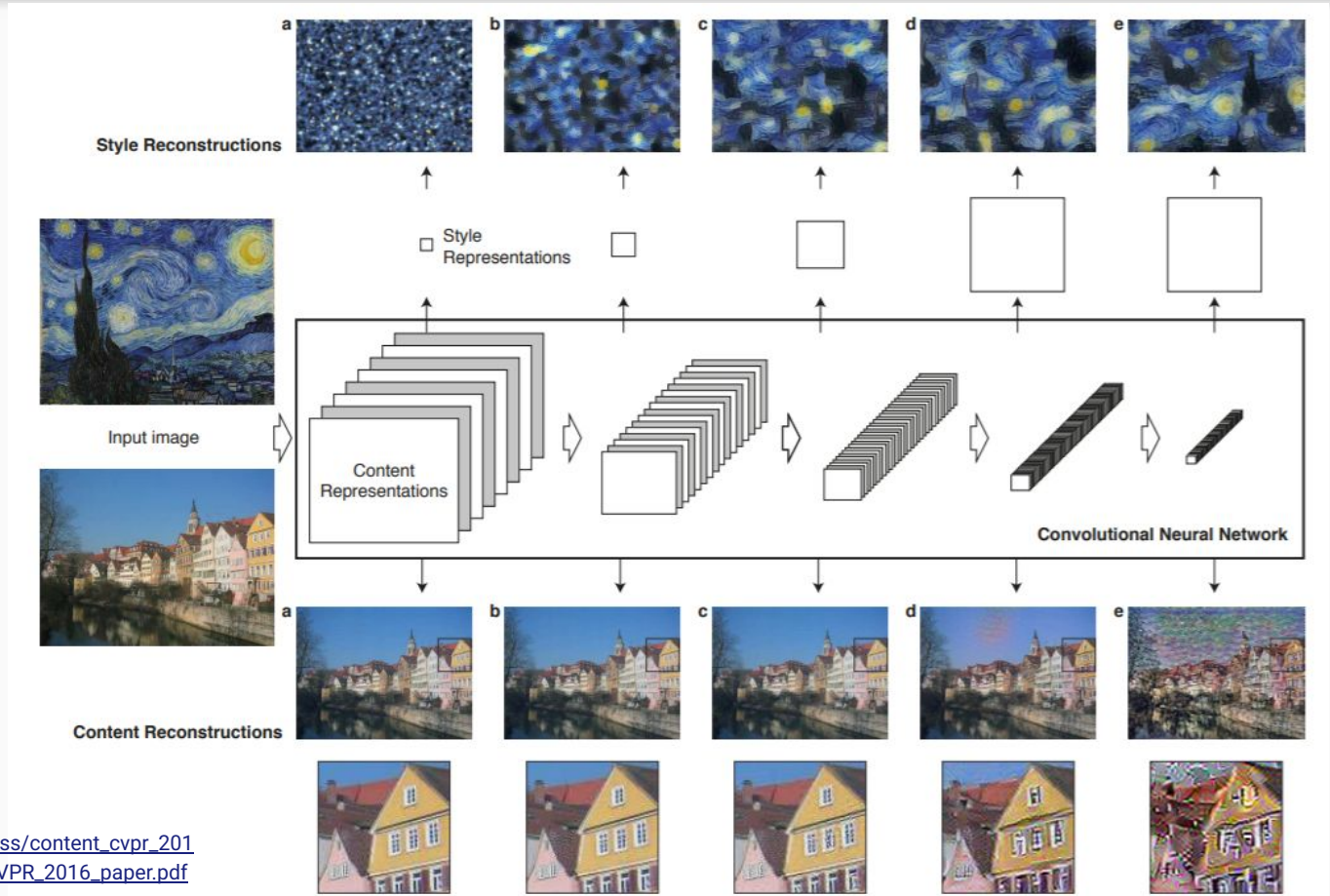


Image Style Transfer [Gatys et al.]

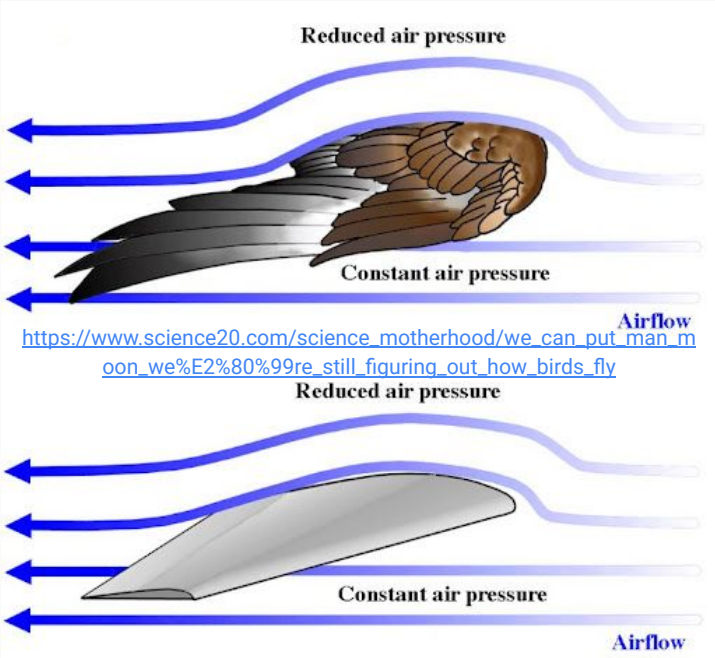


https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Gatys_Image_Style_Transfer_CVPR_2016_paper.pdf

Given Start Image, Generate the world

Foundational World Model – water

Bionic Modeling of Bird → Physics of Flying



https://www.science20.com/science_motherhood/we_can_put_man_moon_we%E2%80%99re_still_figuring_out_how_birds_fly

vs. Interpretability of Neural Nets

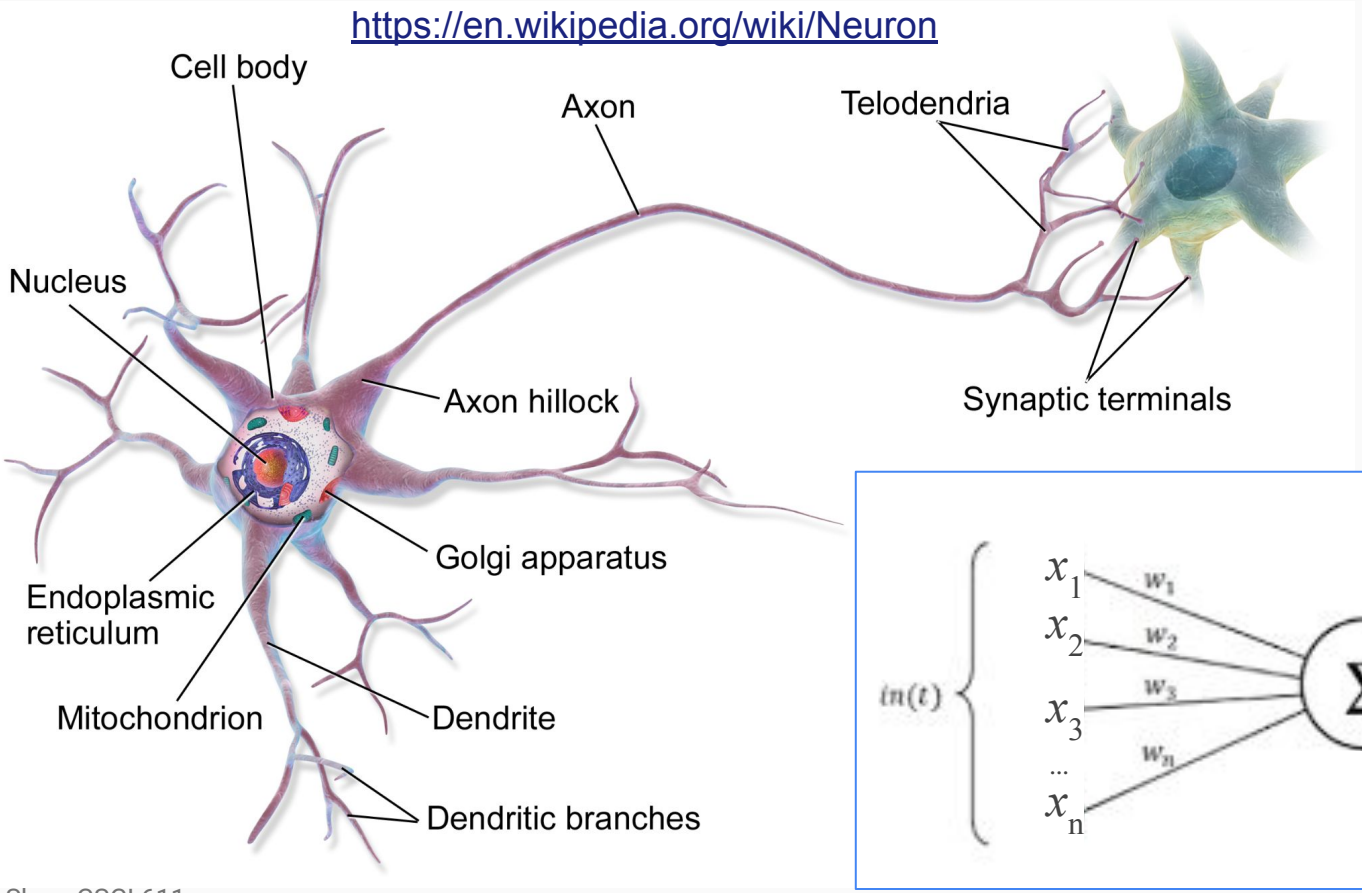
Peter Norvig: we already reach AGI. 1903 White brother build first airplane, only years later it's for commercial. Airplane nowadays are just bigger faster and more comfortable.

Same argument on AI→AGI?

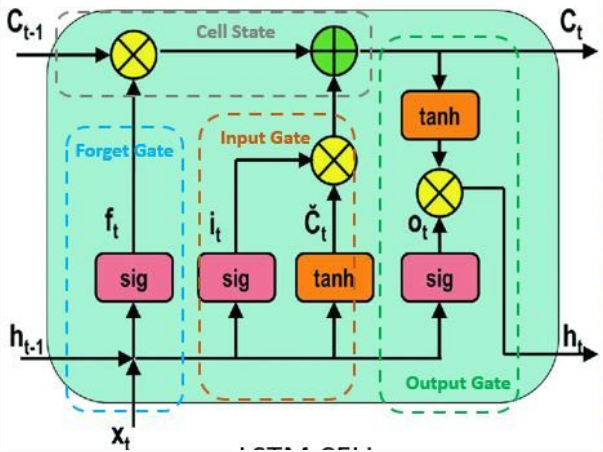
- Stuart Russell: No
- But is there a same progression …

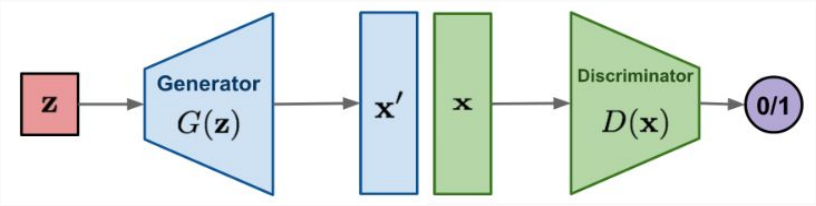
Bionic Modeling of Neuron → Neural Network

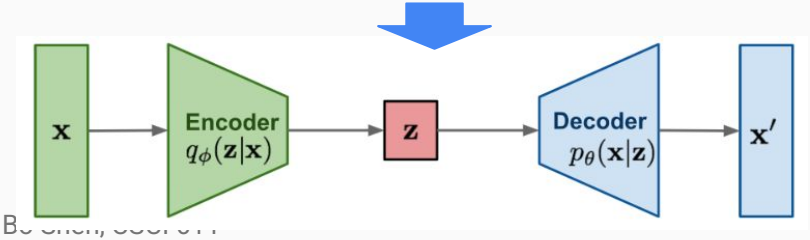
https://en.wikipedia.org/wiki/Neuron

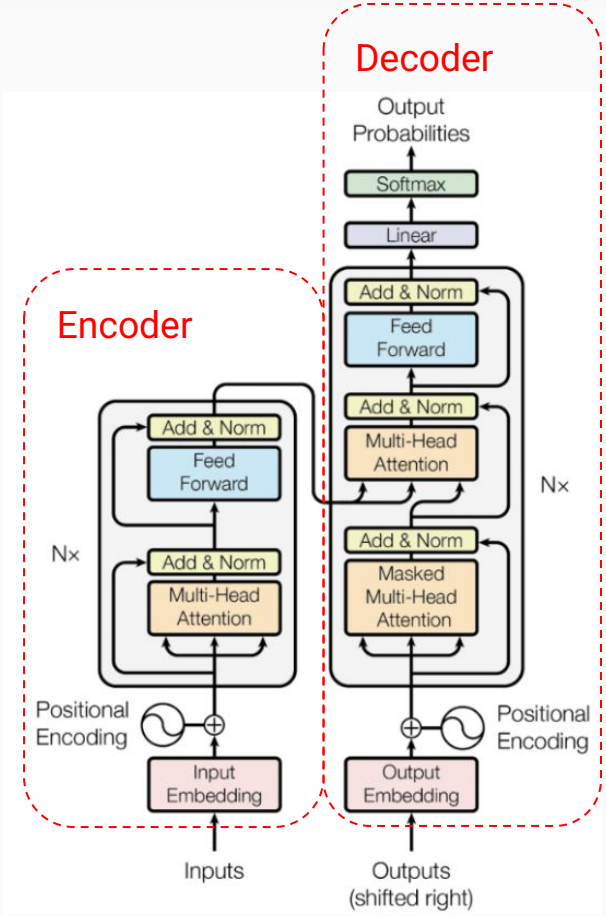


Bionic ⇒ Semantic

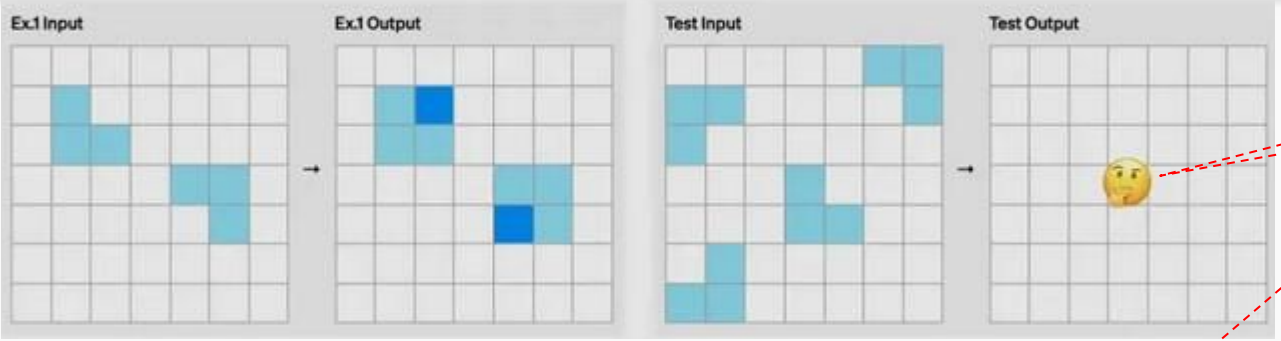




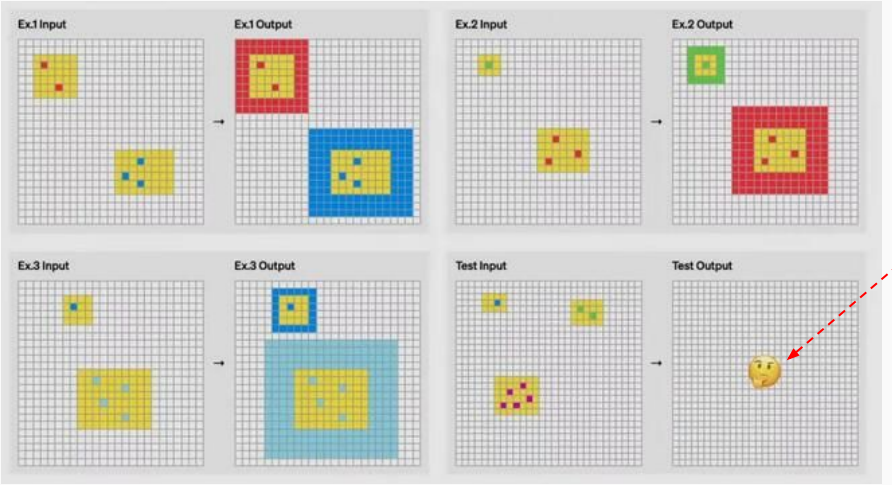




ARC-AGI



What's the rule?

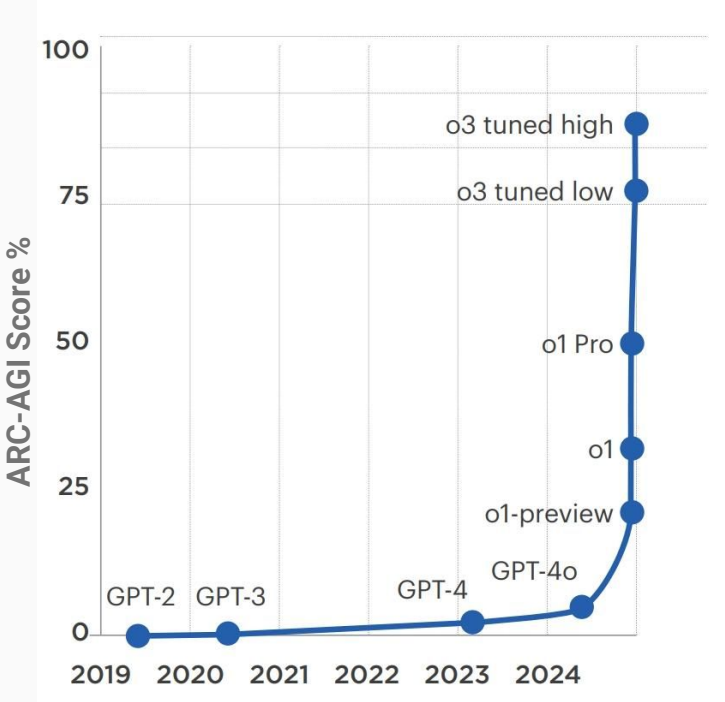


- ARC is composed of a set of visual puzzles that require understanding of basic concepts such as objects, boundaries and spatial relationships
- ARC-AGI tasks are a series of three to five input and output tasks followed by a final task with only the input listed.
- Each task tests the utilization of a specific learned skill based on a minimal number of cognitive priors.

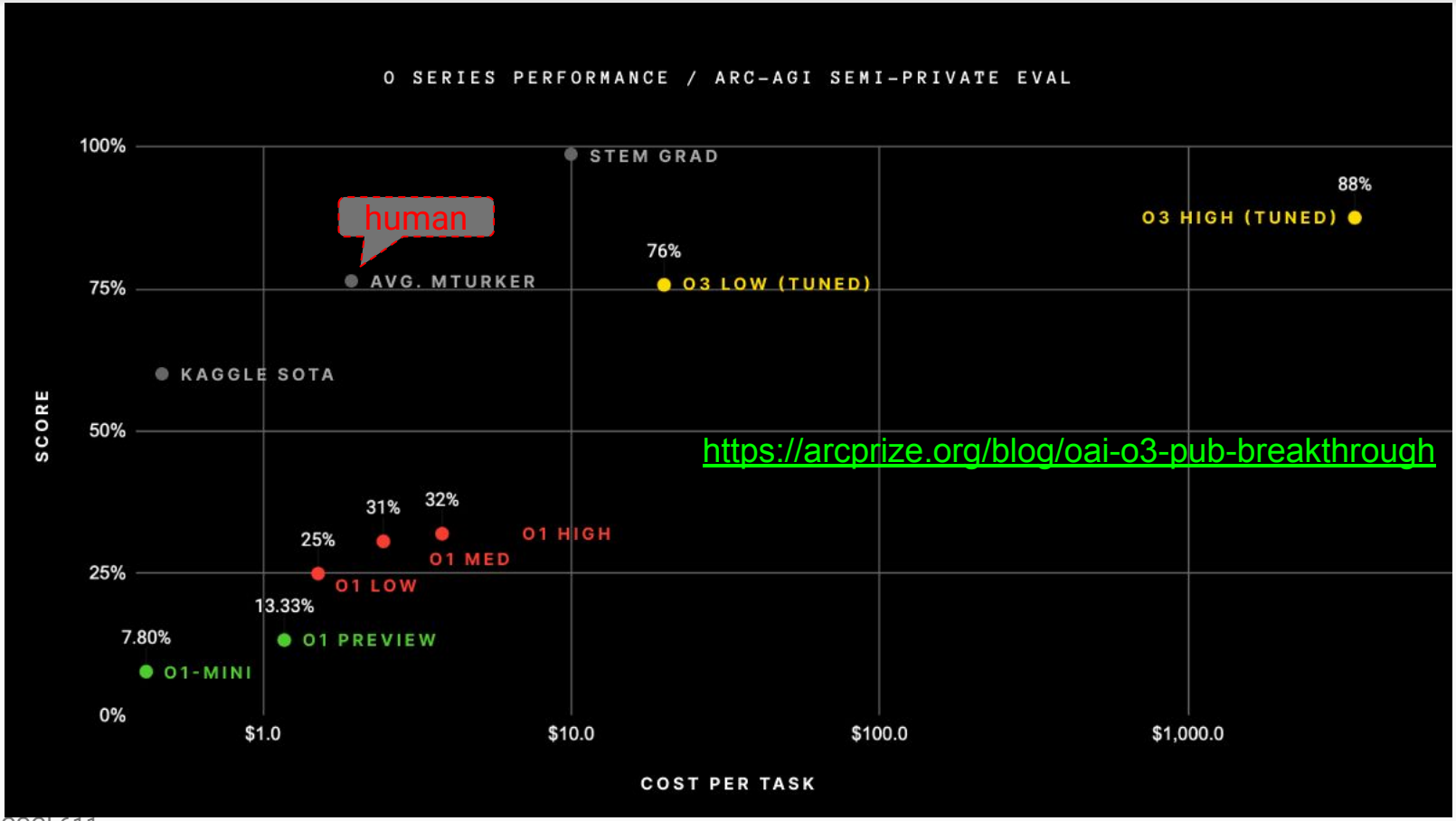
ARC-AGI

The ARC-AGI benchmark is based on the Abstract Reasoning Corpus, which tests an AI system's ability to adapt to novel tasks and demonstrate fluid intelligence.

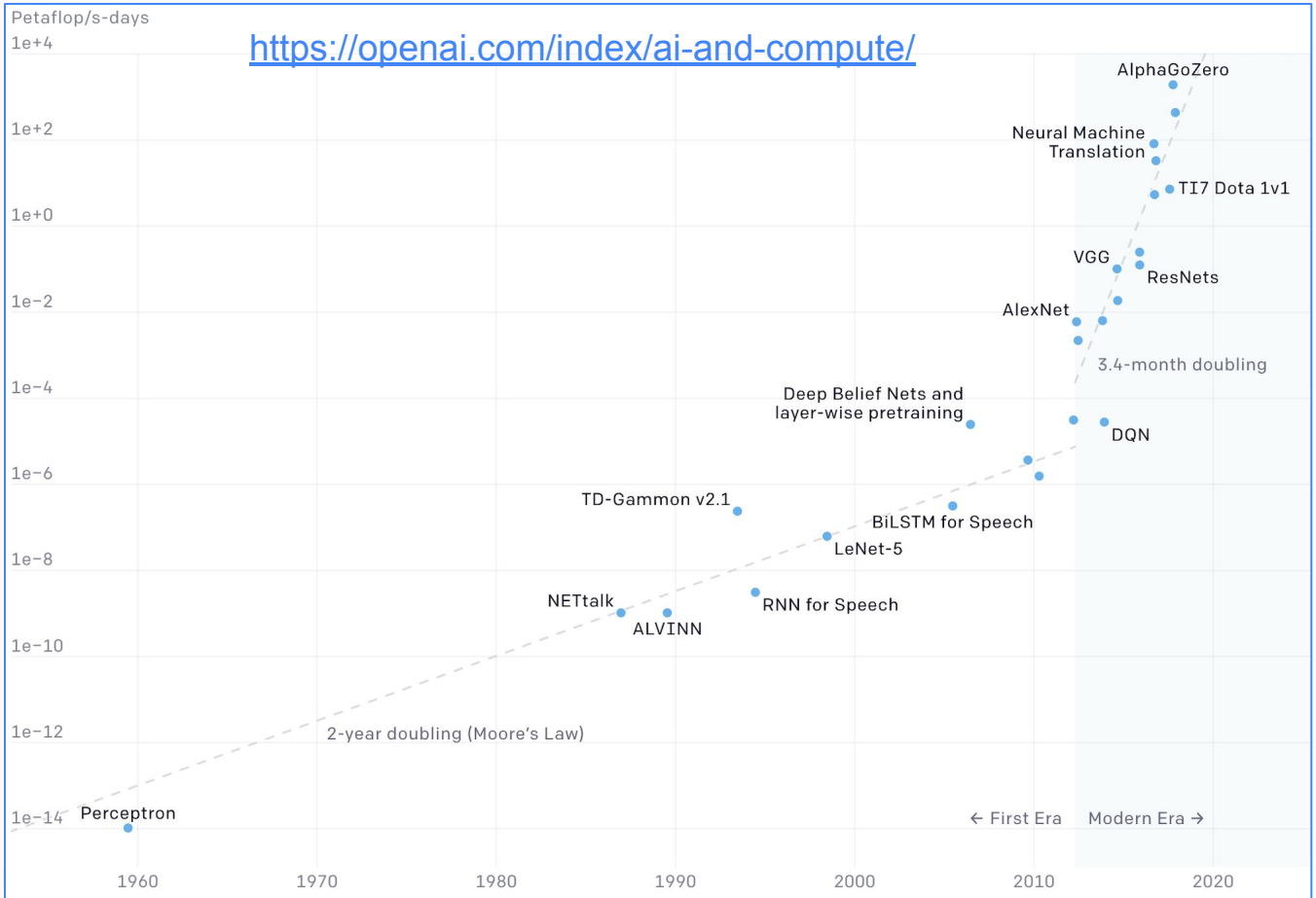
- ARC is composed of a set of visual puzzles that require understanding of basic concepts such as objects, boundaries and spatial relationships
- ARC-AGI tasks are a series of three to five input and output tasks followed by a final task with only the input listed.
- Each task tests the utilization of a specific learned skill based on a minimal number of cognitive priors.



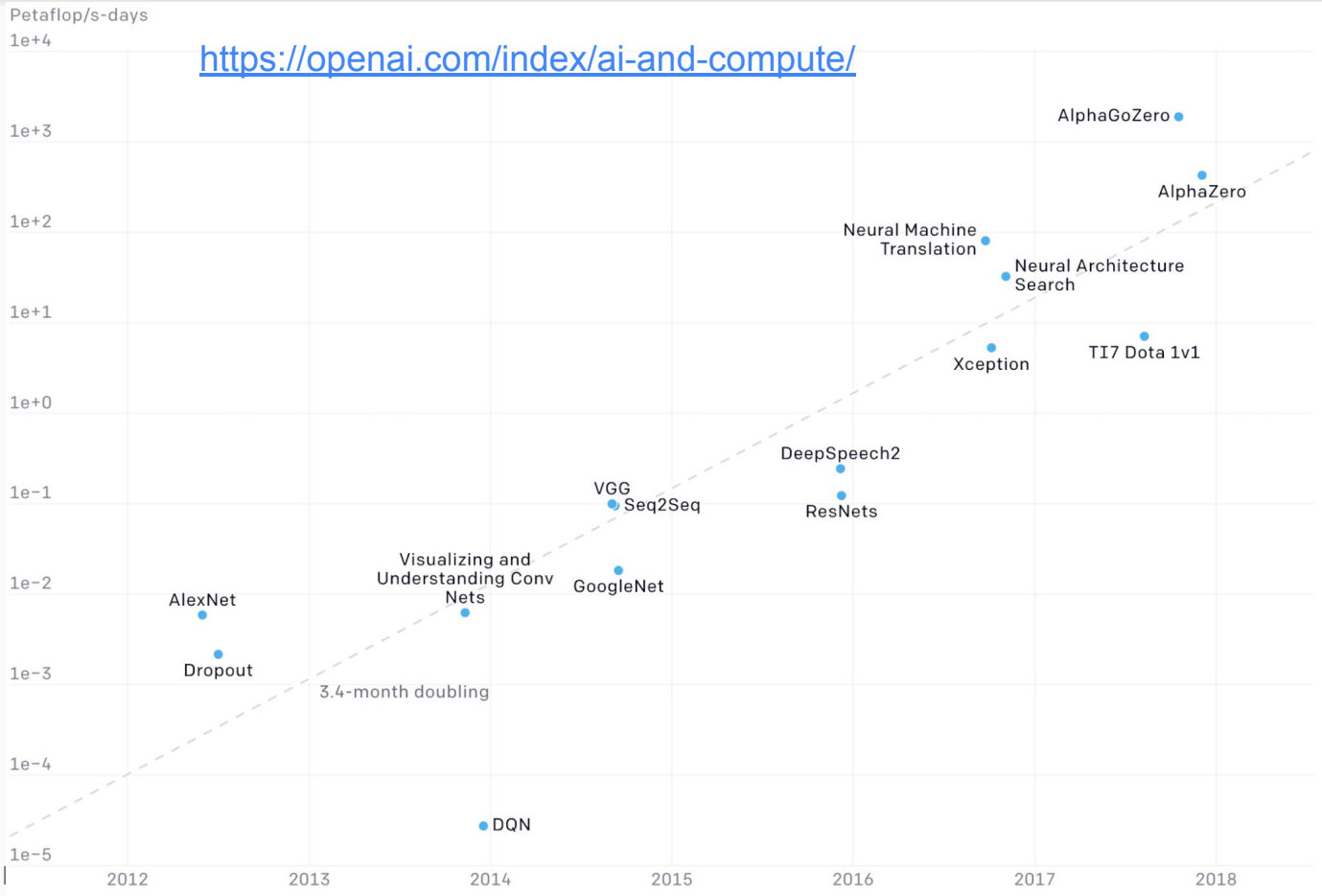
Score vs Cost



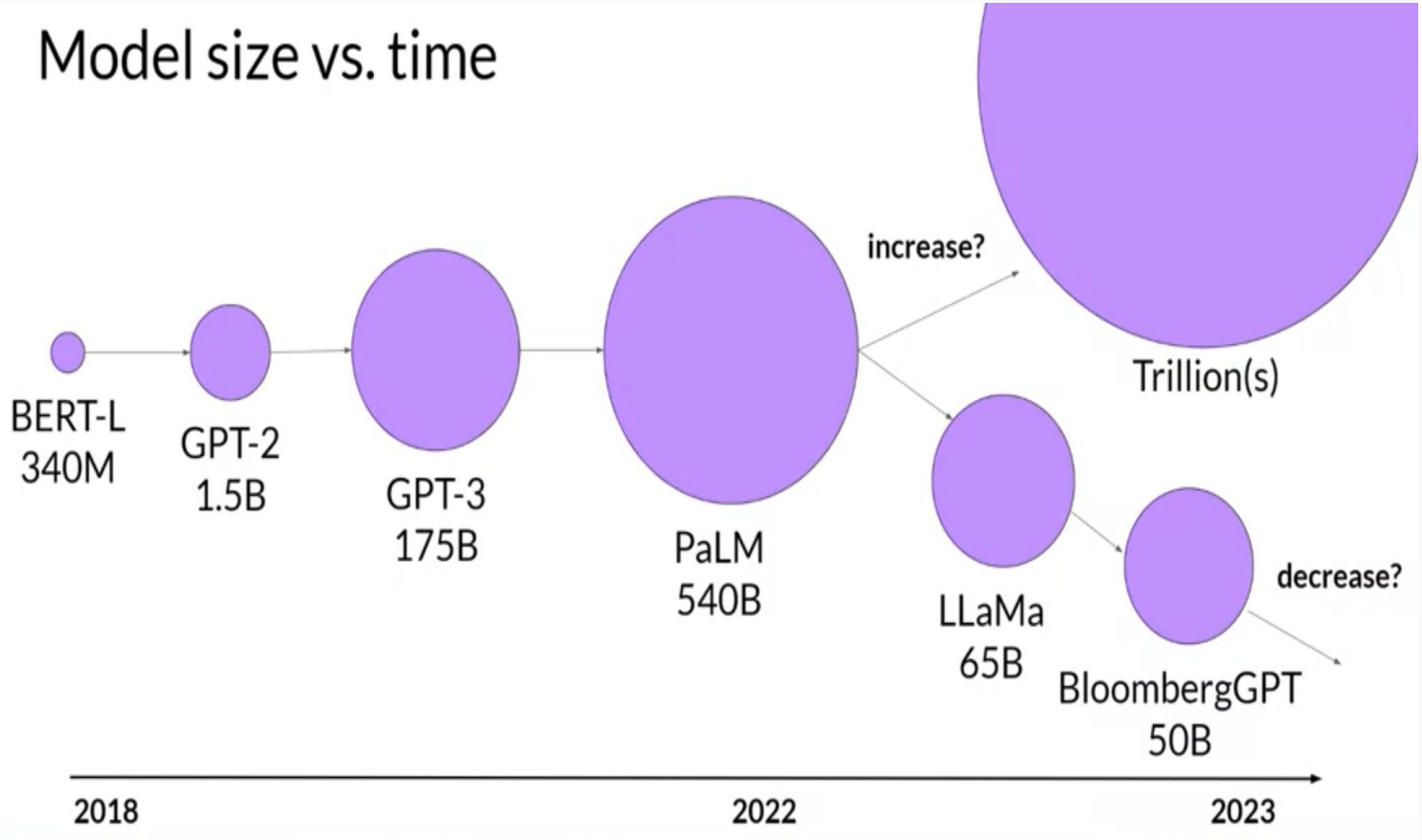
Two Eras of Compute Usage in Training AI Systems



Modern Era


Petaflop/s-days
1e+4
1e+3
1e+2
1e+1
1e+0
1e−1
1e−2
1e−3
1e−4
1e−5
2012
2013
2014
2015
2016
2017
2018
https://openai.com/index/ai-and-compute/
AlphaGoZero
AlphaZero
Neural Machine Translation
Neural Architecture Search
TI7 Dota 1v1
Xception
DeepSpeech2
ResNets
VGG
Seq2Seq
Visualizing and Understanding Conv Nets
GoogleNet
AlexNet
Dropout
3.4-month doubling
DQN

Model Size vs Training Time/Cost



GPT-4
1.76T

DeepSeekV3
671B

Exhaustion of Human-Generated Data

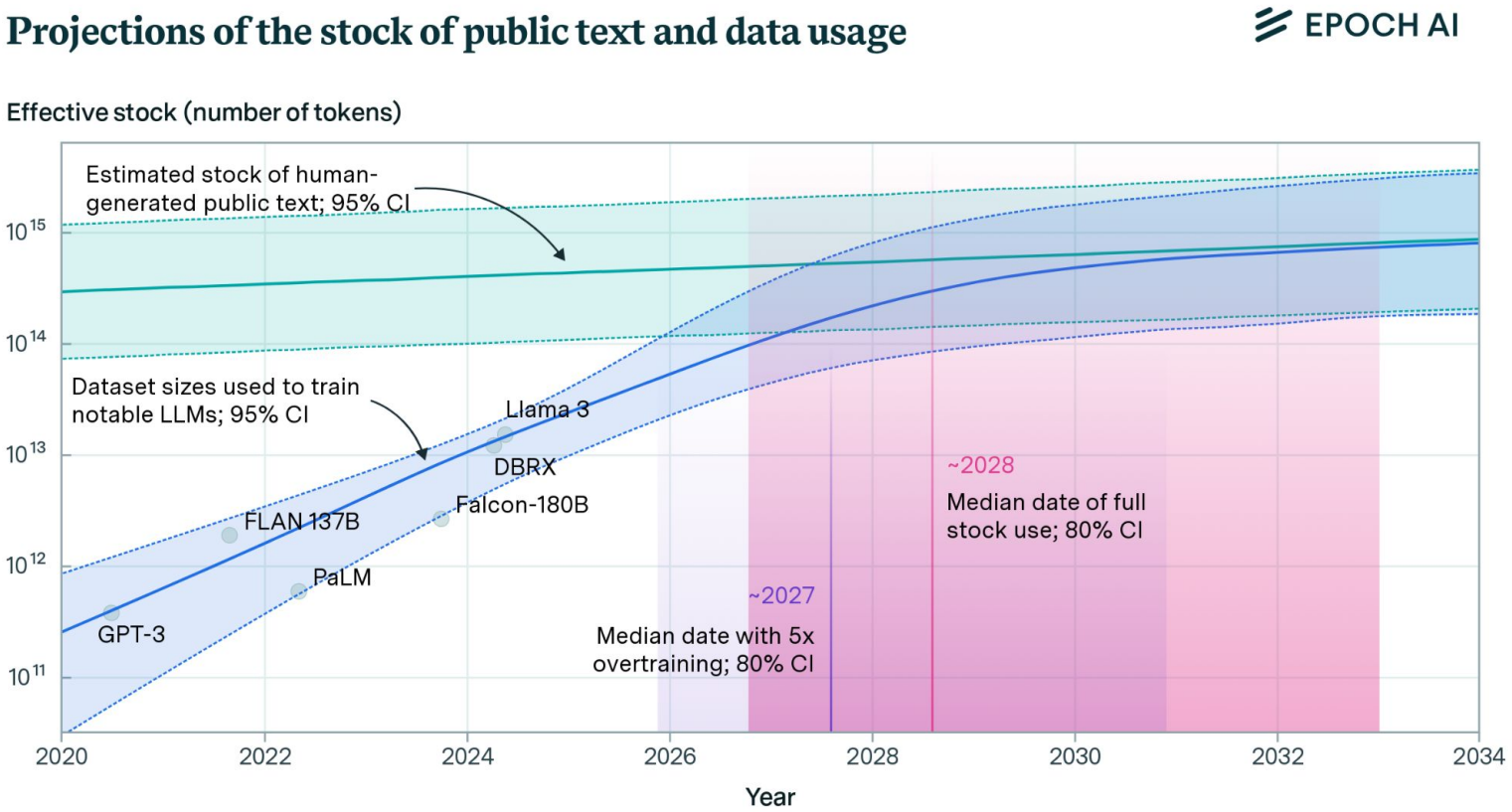
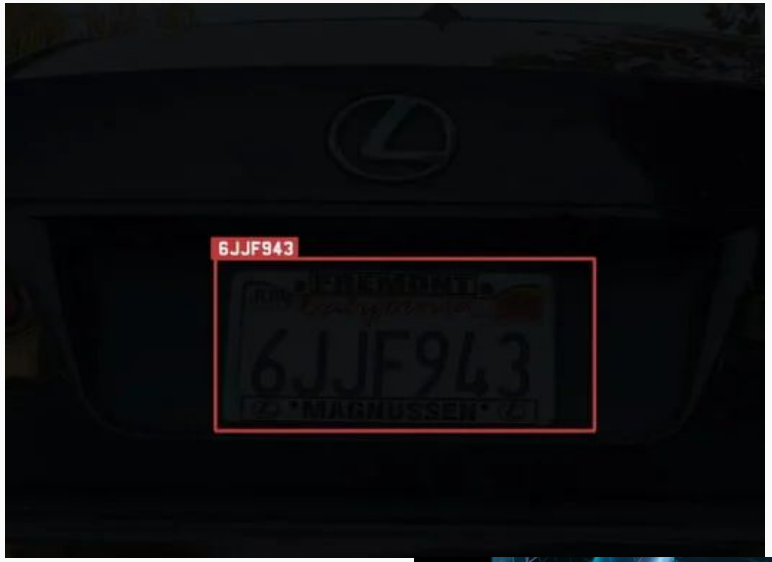
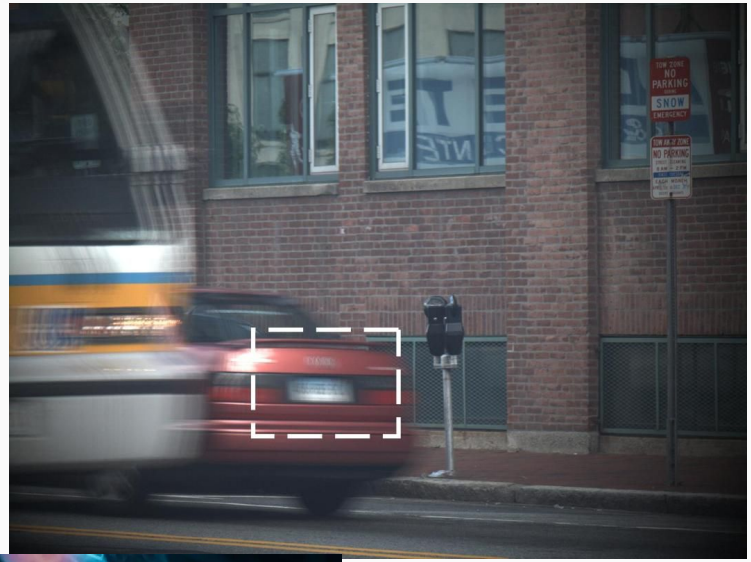


Figure 2: Projection of effective stock of human-generated public text and dataset sizes used to train notable LLMs. Individual dots represent dataset sizes of specific notable models. The dataset size projection is a mixture of an extrapolation of historical trends and a compute-based projection that assumes models are trained compute-optimally.

https://epoch.ai/blog/will-we-run-out-of-data-limits-of-llm-scaling-based-on-human-generated-data

AGI → ASI





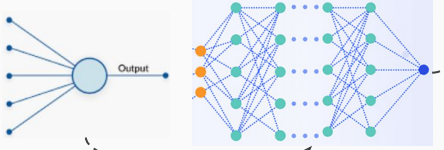
Super Sensor ⇒ Super Intelligence

Super Algorithm → ASI

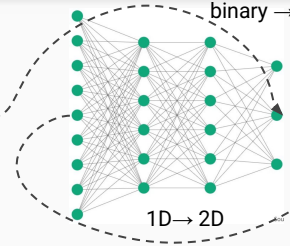
Super Algorithm
https://www.cs.jhu.edu/~misha/ReadingSeminar/Papers/Malzbender01.pdf

Super Sensor + Super Algorithm
⇒ ASI

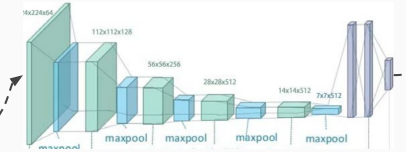
First Half



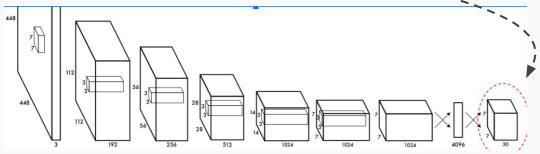
single → multilayer
linear → nonlinear bound



binary → multiclass



single object→ multiple objects /w attributes per picture



Perceptron

activation / loss
gradient descent
back propagation

multiclass ⇒

MLP

dense/fully connected
sigmoid→ softmax
normalization
dropout

image ⇒

CNN

conv for AutoFE
max pooling for multiscale
overfitting

multiple objects
+ location ⇒

YOLO

multiple, overlapped with location
speed (1x1 conv)
depthwise conv
real time → video

AlexNet

vanishing gradient

SSD

use multiple feature maps

VGG19

residual connection

higher speed
higher perf.
lower power

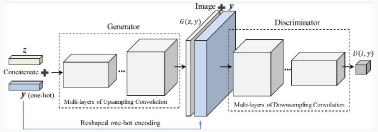
MobileNet (EdgeML)

Efficient Net

style transfer ⇐

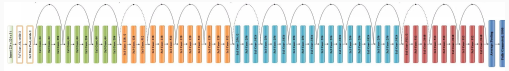
StyleGAN

generator
discriminator
content derivation



GAN

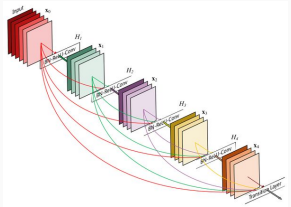
ResNet

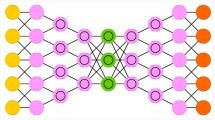


feature reuse
batch normalization

deeper learning
higher accuracy

DenseNet



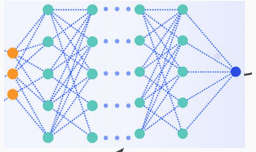


Auto-Encoder

Variational AE

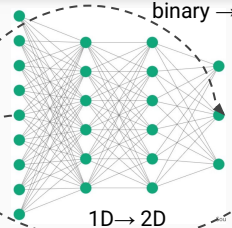
feature extraction → embedding
compression ← dimension reduction

Second Half

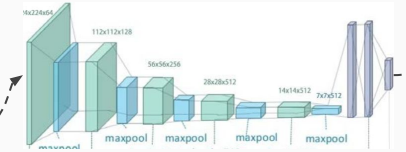


single → multilayer
linear → nonlinear bound

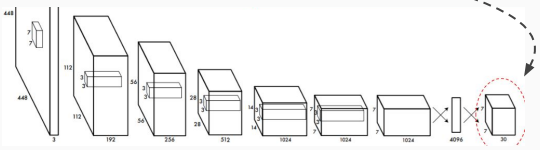
binary → multiclass



1D→ 2D



single object→ multiple objects /w attributes per picture



Perceptron

activation / loss
gradient descent
back propagation

multiclass ⇒

MLP

dense/fully connected
sigmoid→ softmax
normalization
dropout

image ⇒

CNN

conv for AutoFE
max pooling for multiscale
overfitting

multiple objects
+ location ⇒

YOLO

multiple, overlapped with location
speed (1x1 conv)
depthwise conv
real time → video

AlexNet

vanishing gradient

SSD

use multiple feature maps

VGG19

residual connection

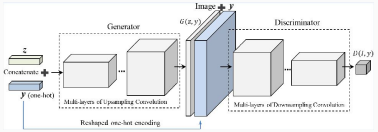
higher speed
higher perf.
lower power

MobileNet (EdgeML)
Efficient Net

style transfer ⇐

StyleGAN

generator
discriminator
content derivation



GAN

deeper learning
higher accuracy

ResNet

feature reuse
batch normalization

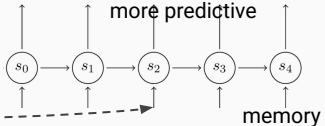
sequential signal ⇒

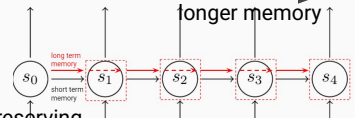
RNN ⇒ LSTM/GRU

vanishing gradient

more predictive

longer memory

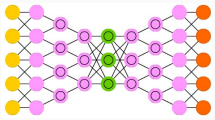




DenseNet

style from other → style from past

memory preserving
one hot → embedding
word2vec



Auto-Encoder
Variational AE

feature extraction → embedding
compression ← dimension reduction

attention← recurrent

transformer encoder

GPT ⇐ ViT ⇐ BERT ⇐ Transformer ⇐ BRNN/BiLSTM

Q&A