

Machine Learning

Deep Learning Application

Dr. Shuang LIANG

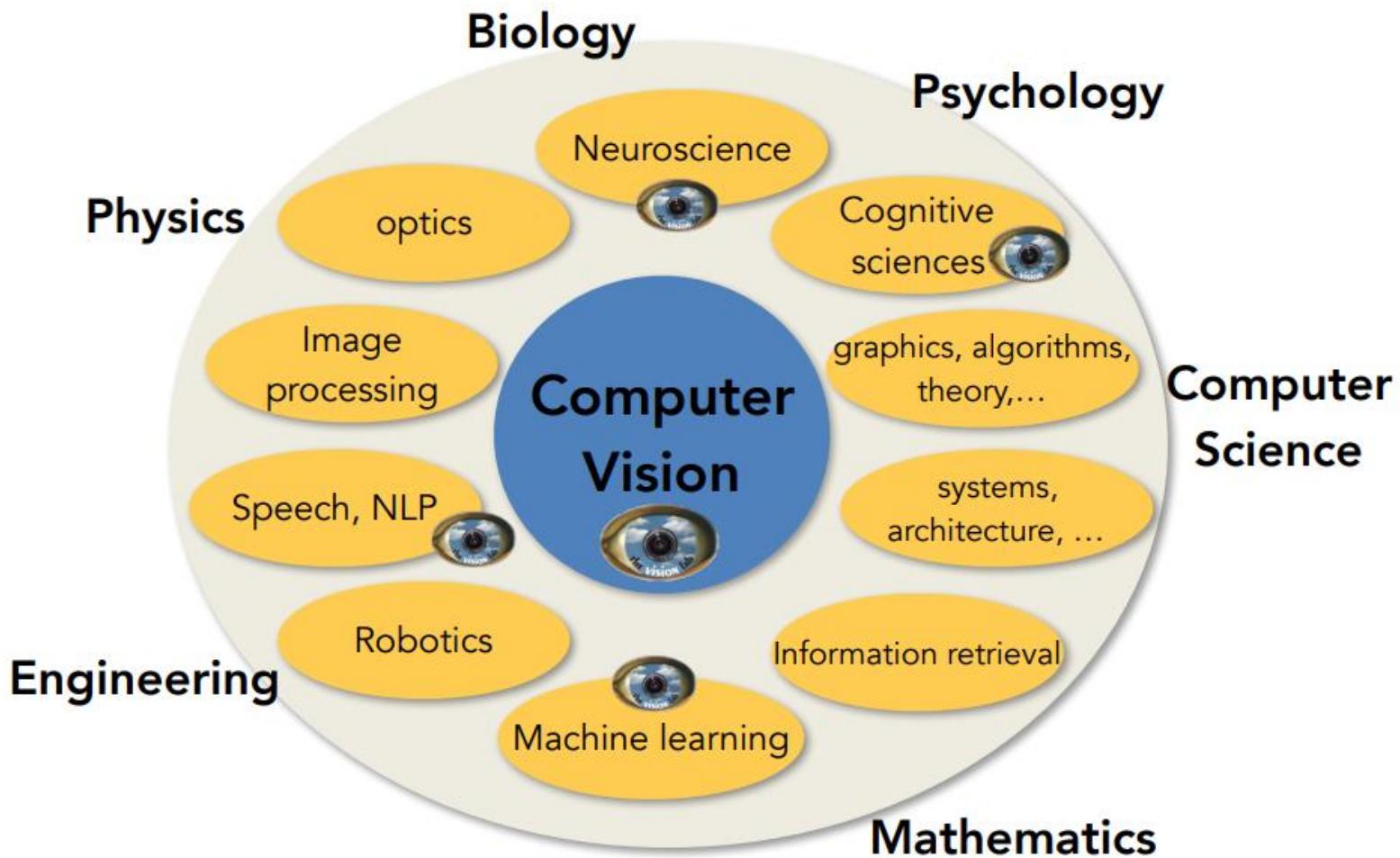
Today's Topics

- Deep Learning & Computer Vision
- Deep Learning & Natural Language Process (NLP)
- Deep Learning & Medicine
- AIGC

Today's Topics

- *Deep Learning & Computer Vision*
- Deep Learning & Natural Language Process (NLP)
- Deep Learning & Medicine
- AIGC

Computer Vision



A bit of history

Evolution's Big Bang



[This image is licensed under CC-BY 2.5](#)

543million years, B.C.



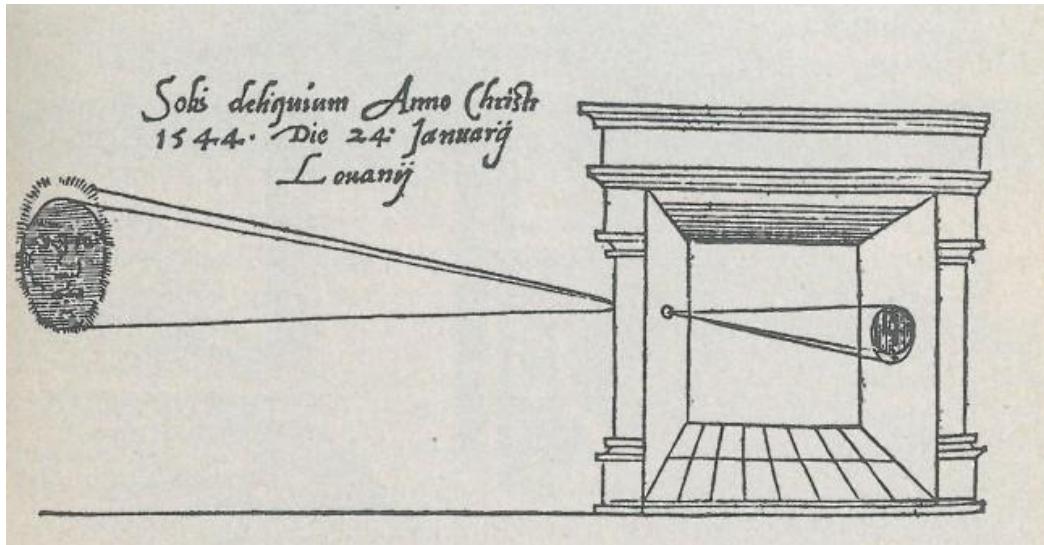
[This image is licensed under CC-BY 2.5](#)



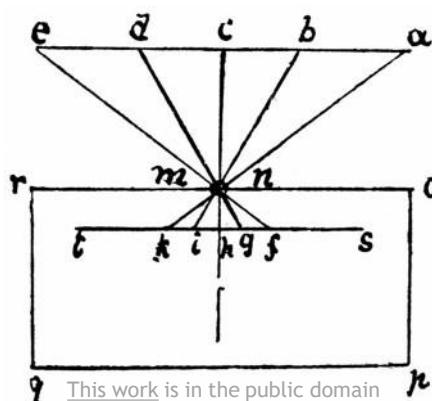
[This image is licensed under CC-BY 3.0](#)

A bit of history

Gemma Frisius, 1545



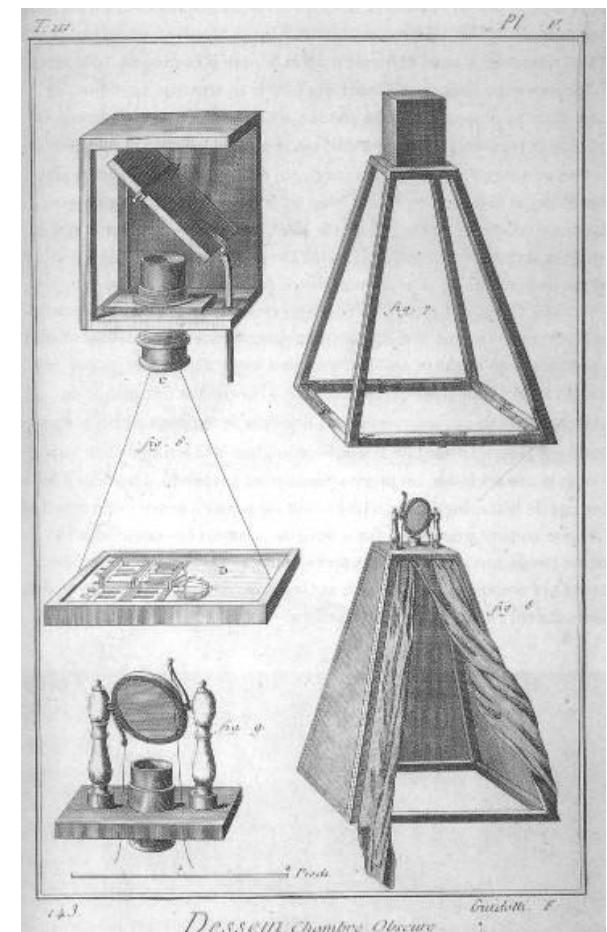
[This work is in the public domain](#)



Camera Obscura
Leonardo da Vinci,
16th Century AD

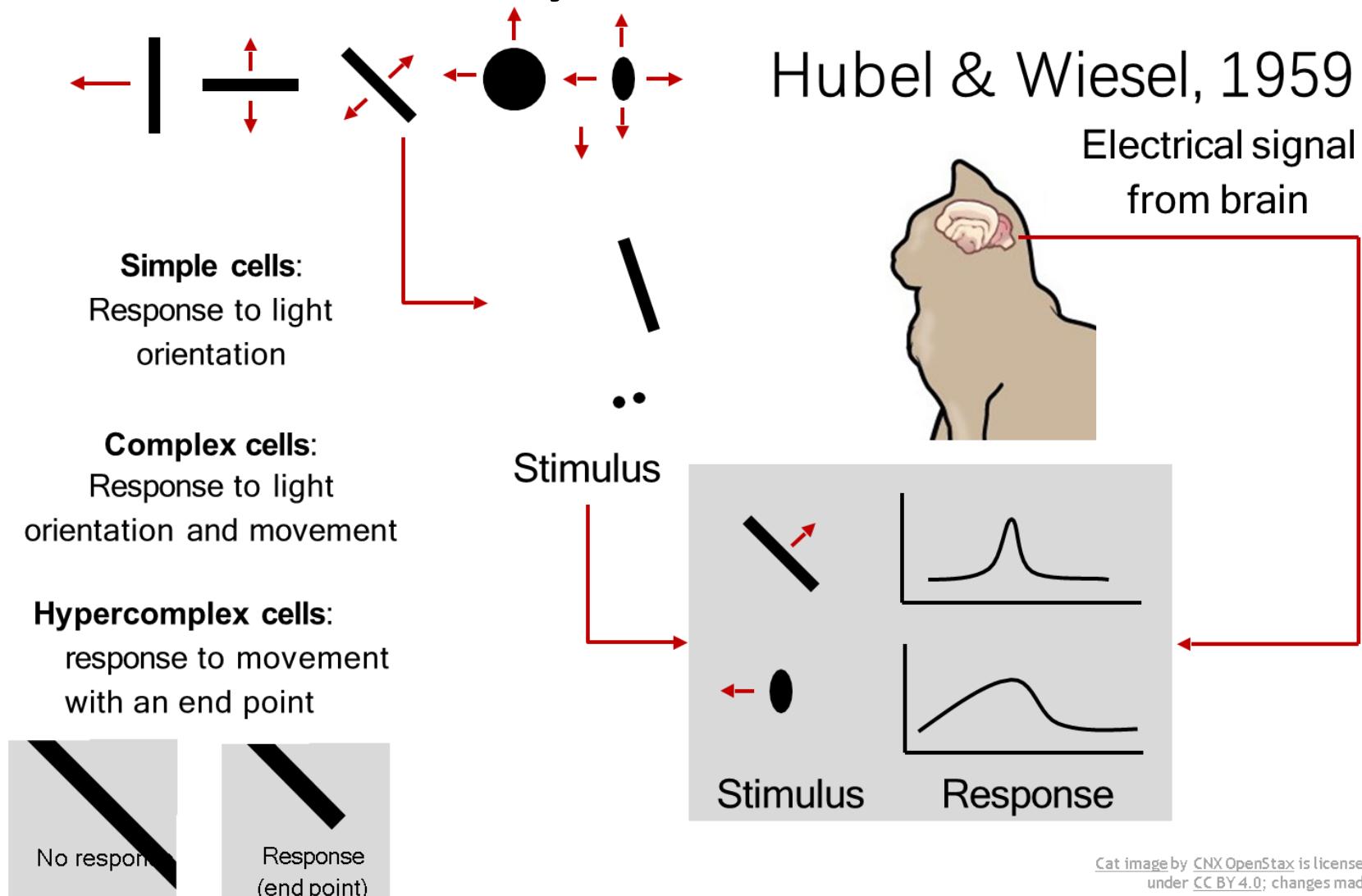
[This work is in the public domain](#)

Encyclopedie, 18th Century



[This work is in the public domain](#)

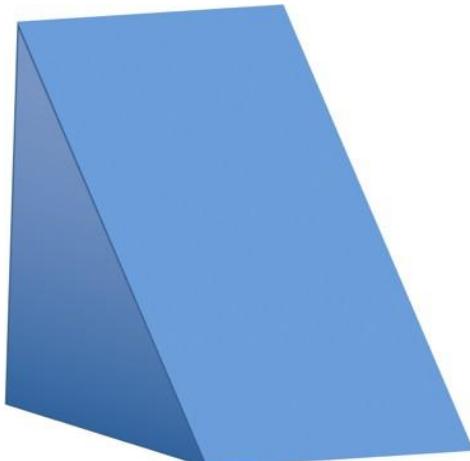
A bit of history



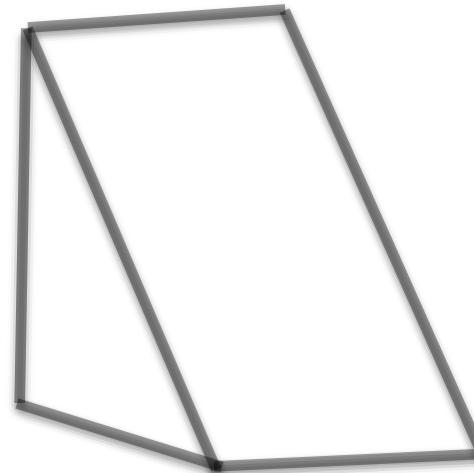
A bit of history

Block world

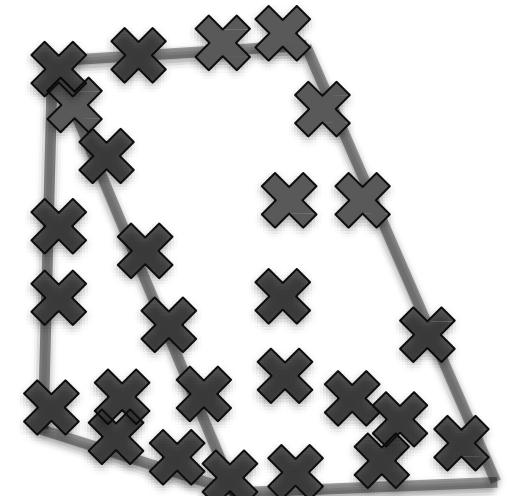
Larry Roberts, 1963



(a) Original picture



(b) Differentiated picture



(c) Feature points selected

A bit of history

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

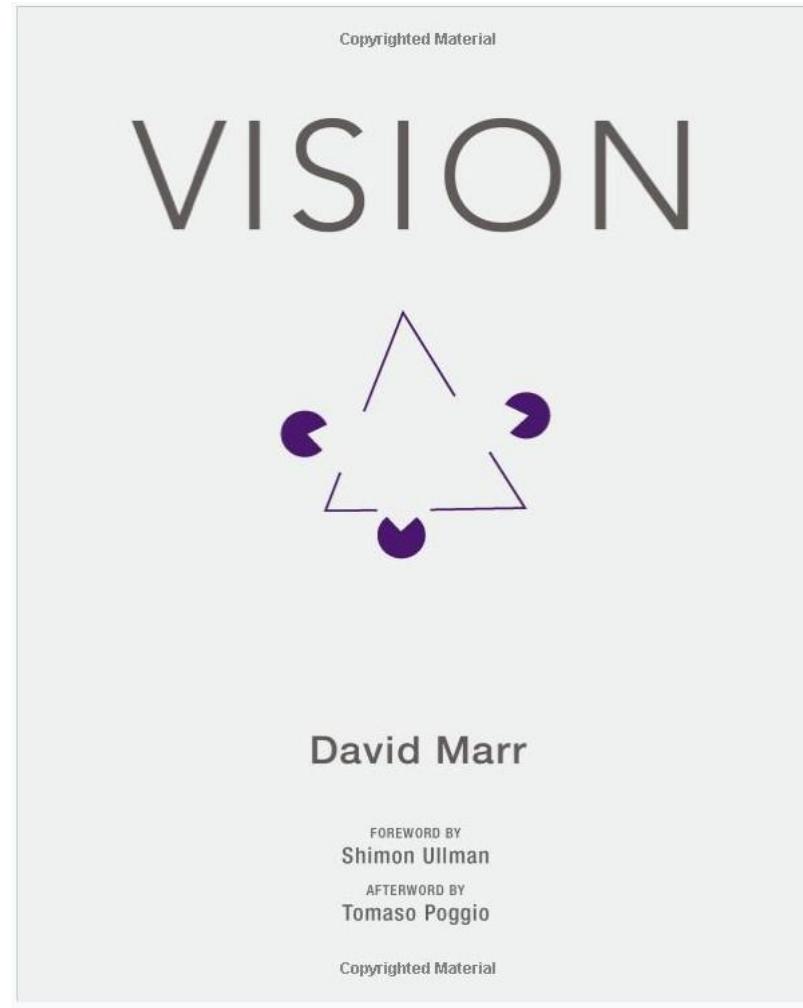
July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

A bit of history



David Marr, 1970s

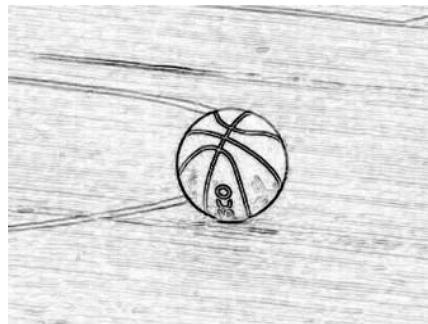
A bit of history

Input image



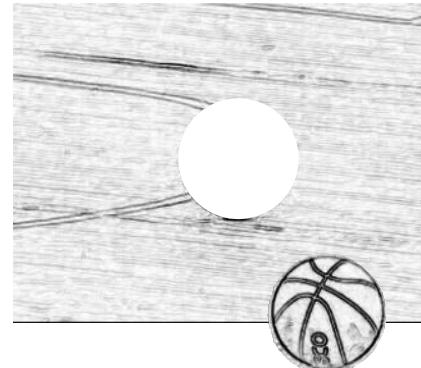
This image is CC0 1.0 public domain

Edge image

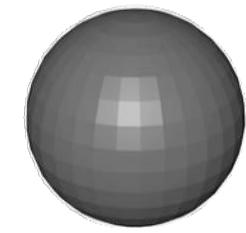


Stages of Visual Representation,
David Marr, 1970s

2 ½-D sketch



3-D model



This image is CC0 1.0 public domain

Input
Image

Perceived
intensities

Primal
Sketch

Zero crossings,
blobs, edges,
bars, ends,
virtual lines,
groups, curves
boundaries

2 ½-D
Sketch

Local surface
orientation and
discontinuities
in depth and in
surface
orientation

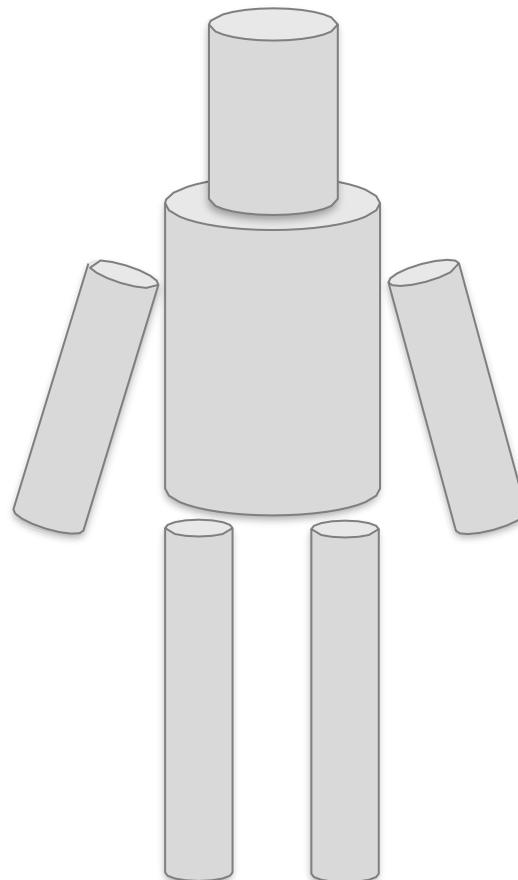
3-D Model
Representation

3-D models
hierarchically
organized in
terms of surface
and volumetric
primitives

A bit of history

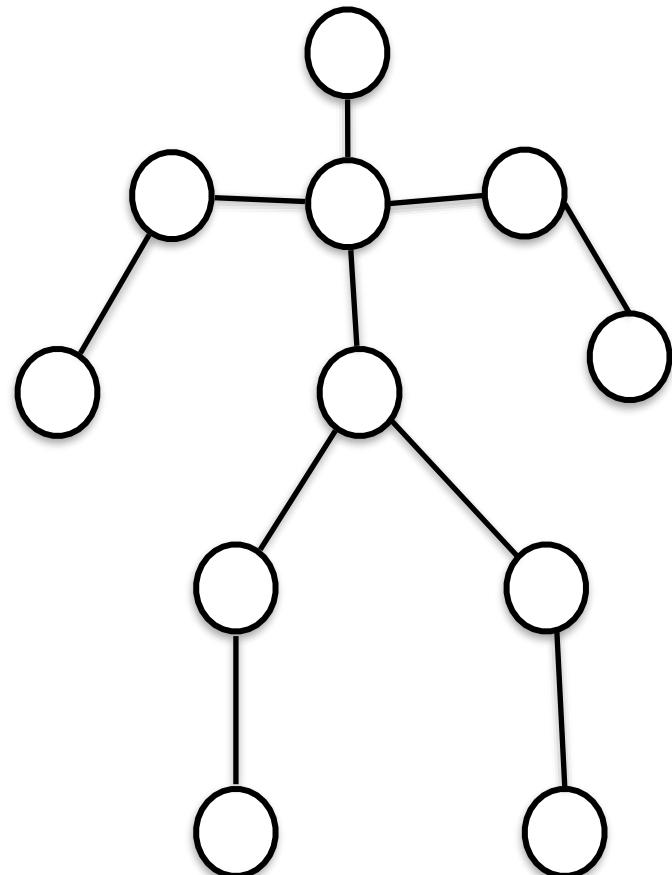
Generalized Cylinder

Brooks & Binford, 1979



Pictorial Structure

Fischler and Elschlager, 1973



A bit of history



[Image](#) is CC BY-SA 4.0

David Lowe, 1987

A bit of history

[Image](#) is CC BY 3.0



[Image](#) is public domain



[Image](#) is CC-BY SA 3.0



Normalized Cut (Shi & Malik, 1997)

Face Detection, Viola & Jones, 2001

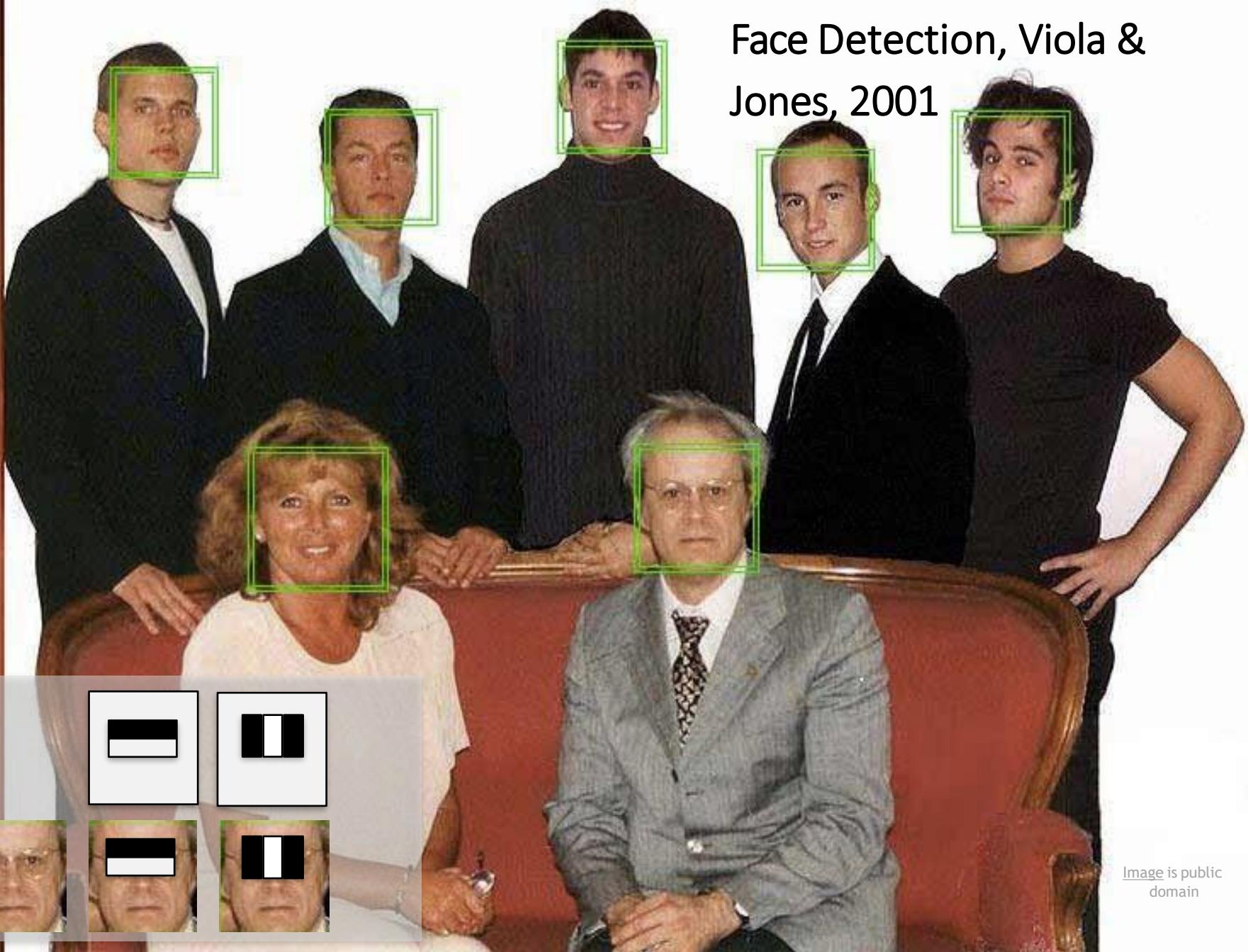


Image is public domain

A bit of history



[Image](#) is public domain



[Image](#) is CC BY-SA 2.0

“SIFT” & Object Recognition, David Lowe, 1999

A bit of history

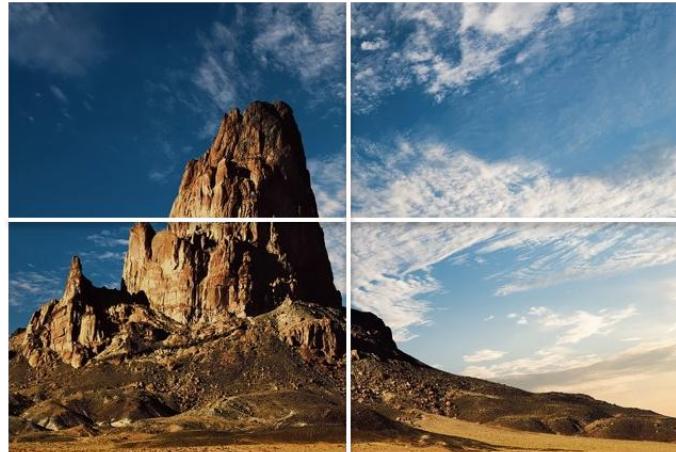
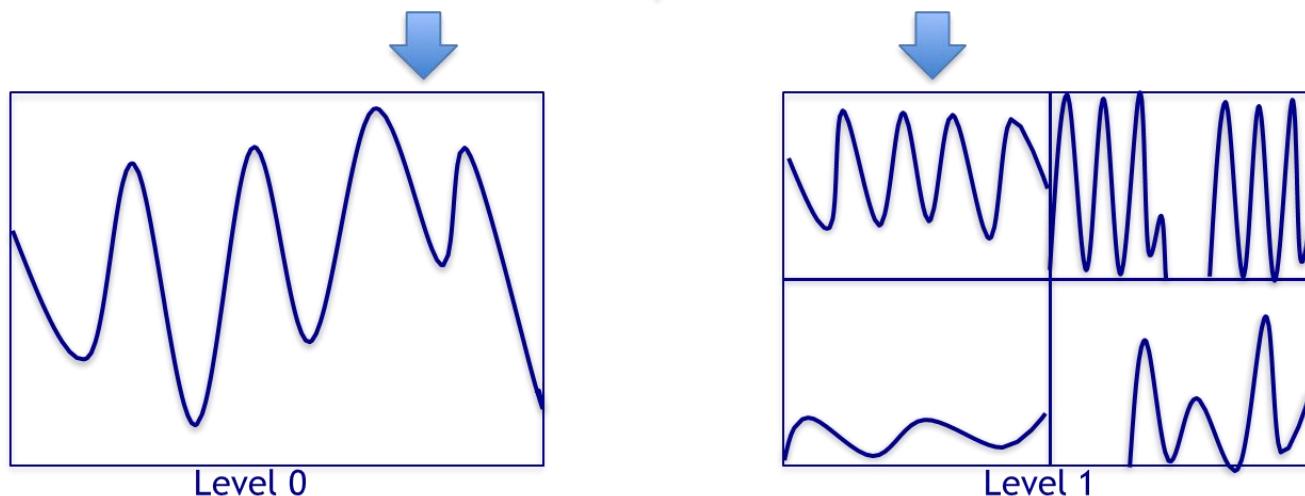


Image is CC0 1.0 public domain



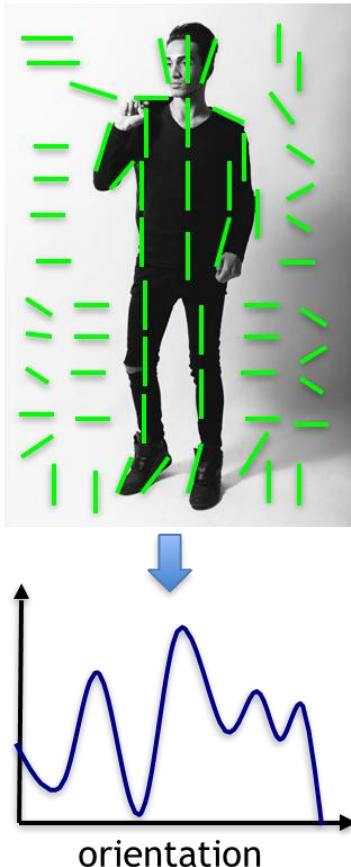
Level 0

Level 1

Spatial Pyramid Matching, Lazebnik, Schmid & Ponce, 2006

A bit of history

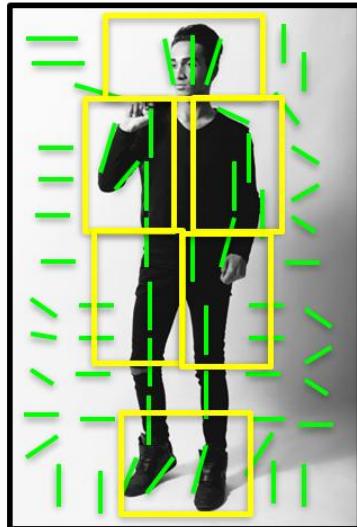
Image is CC0 1.0 public domain



Histogram of Gradients (HoG)
Dalal & Triggs, 2005



Deformable Part Model
Felzenswalb, McAllester, Ramanan, 2009



A bit of history

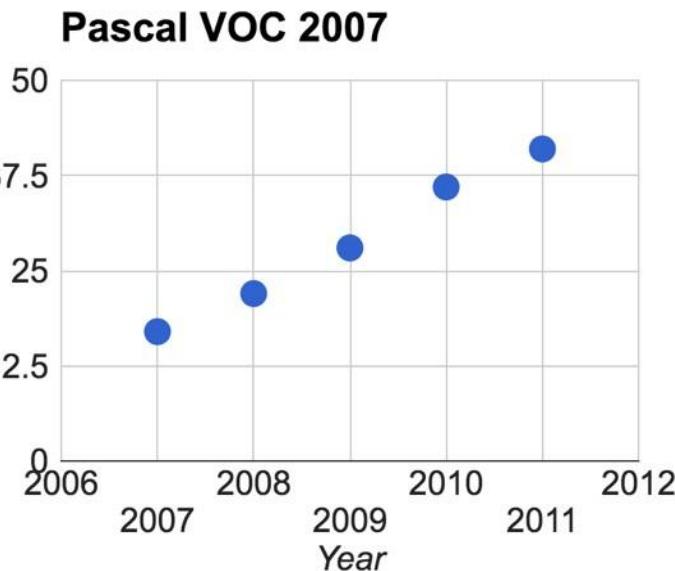
PASCAL Visual Object Challenge (20 object categories)

[Everingham et al. 2006-2012]

[Image is CC BY-SA 3.0](#)



Mean Average Precision
(mAP)



[Image is CC0 1.0 public domain](#)

This image is licensed under
CC BY-SA 2.0; changes made



www.image-net.org

22K categories and **14M** images

- Animals
 - Bird
 - Fish
 - Mammal
 - Invertebrate
- Plants
 - Tree
 - Flower
 - Food
 - Materials
- Structures
 - Artifact
 - Tools
 - Appliances
 - Structures
- Person
- Scenes
 - Indoor
 - Geological Formations
- Sport Activities

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

IMAGENET Large Scale Visual Recognition Challenge

Steel drum

The Image Classification Challenge:

1,000 object classes

1,431,167 images



Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle



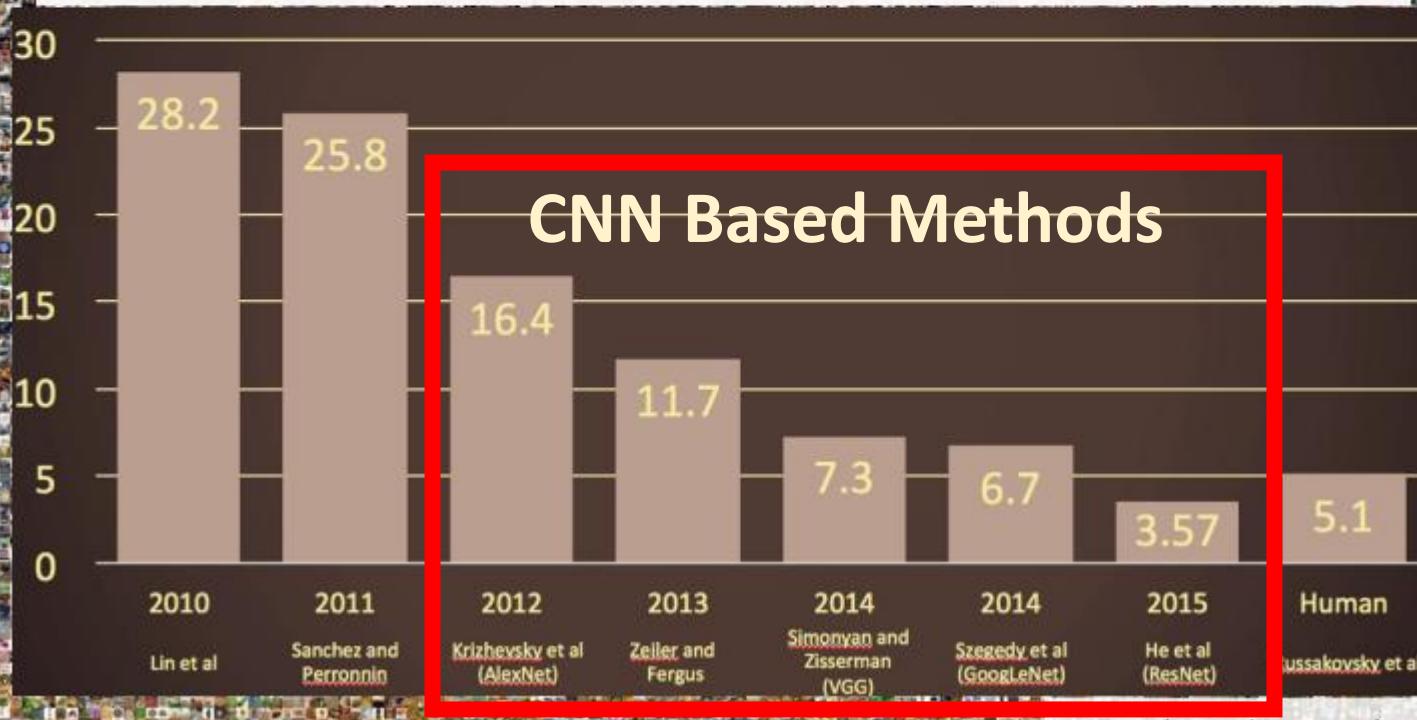
Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle



Russakovsky et al. arXiv, 2014

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:
1,000 object classes
1,431,167 images



Russakovsky et al. arXiv, 2014

DL & CV

- Deep learning methods (like CNN) have become an important tool for computer vision
- Deep learning methods perform well in most tasks in computer vision

Image Classification

Why CNN performs so well on this task?

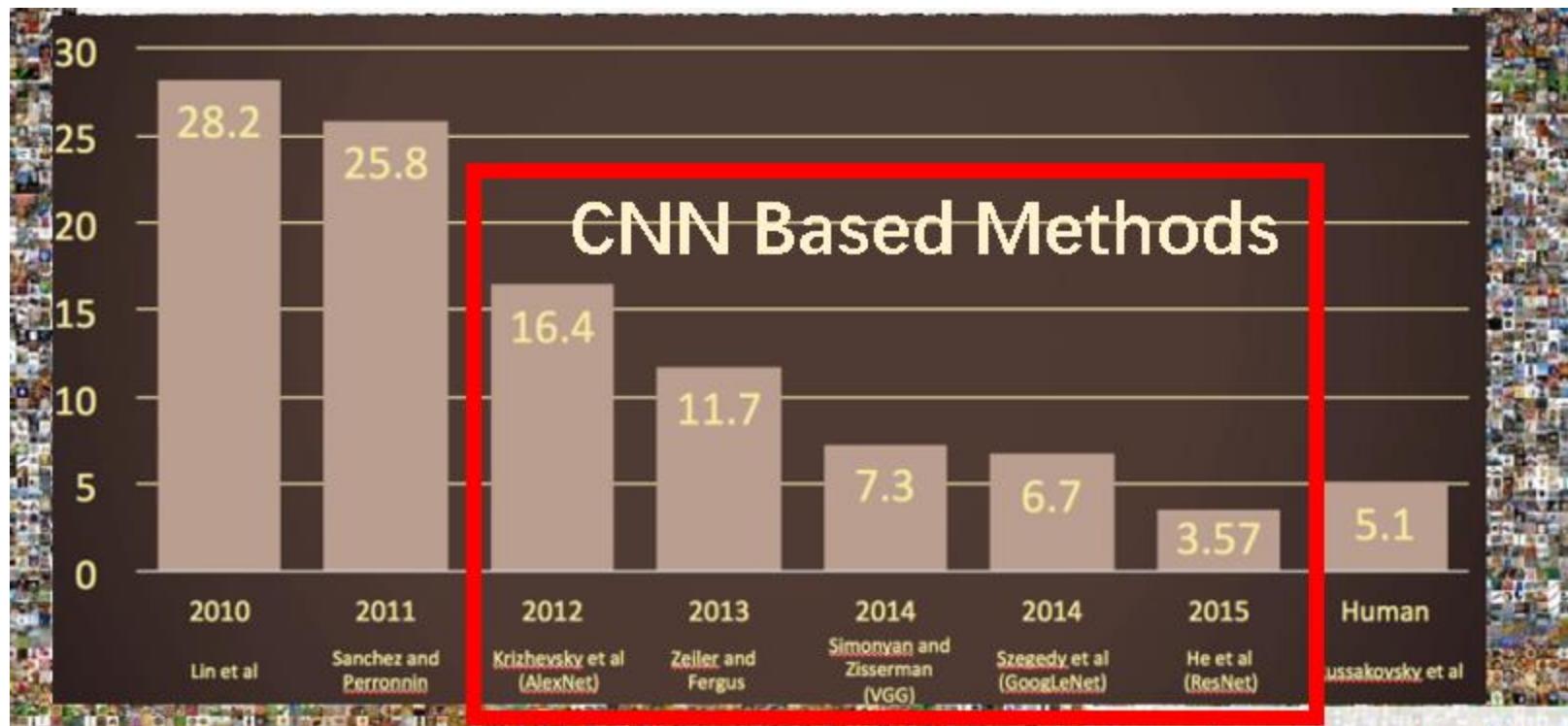
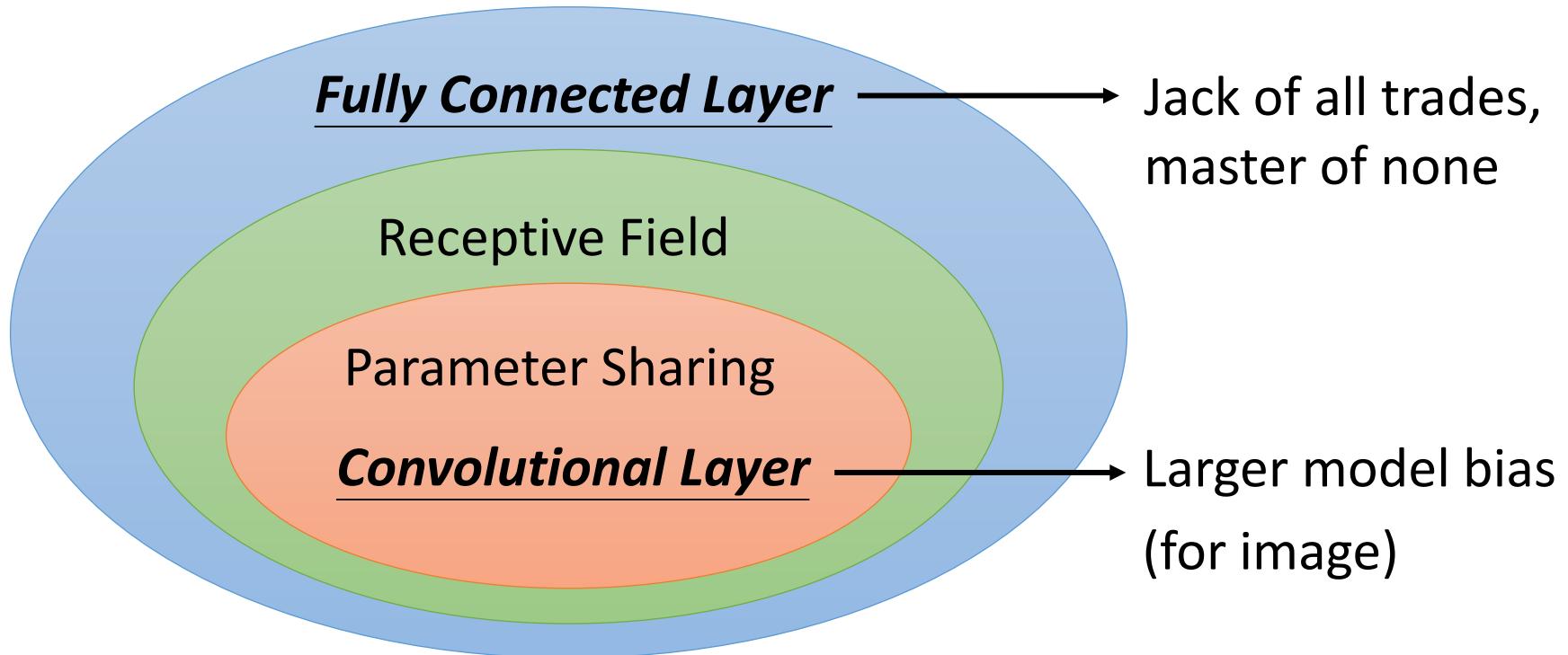


Image Classification

Why CNN performs so well on this task?



- Some patterns are much smaller than the whole image.
- The same patterns appear in different regions.

Image Classification

- The most important problems of visual recognition – image classification
- There is a number of visual recognition problems that are related to image classification

Image Classification



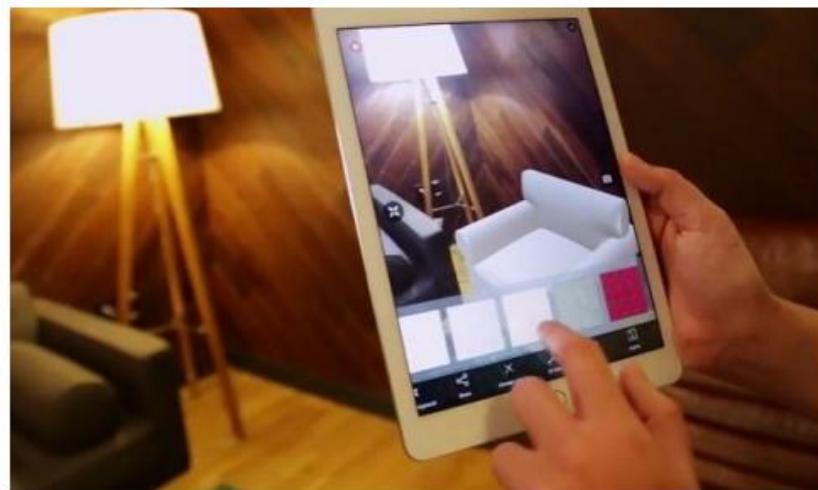
[Image by US Army](#) is licensed under CC BY 2.0



[Image](#) is CC0 1.0 public domain



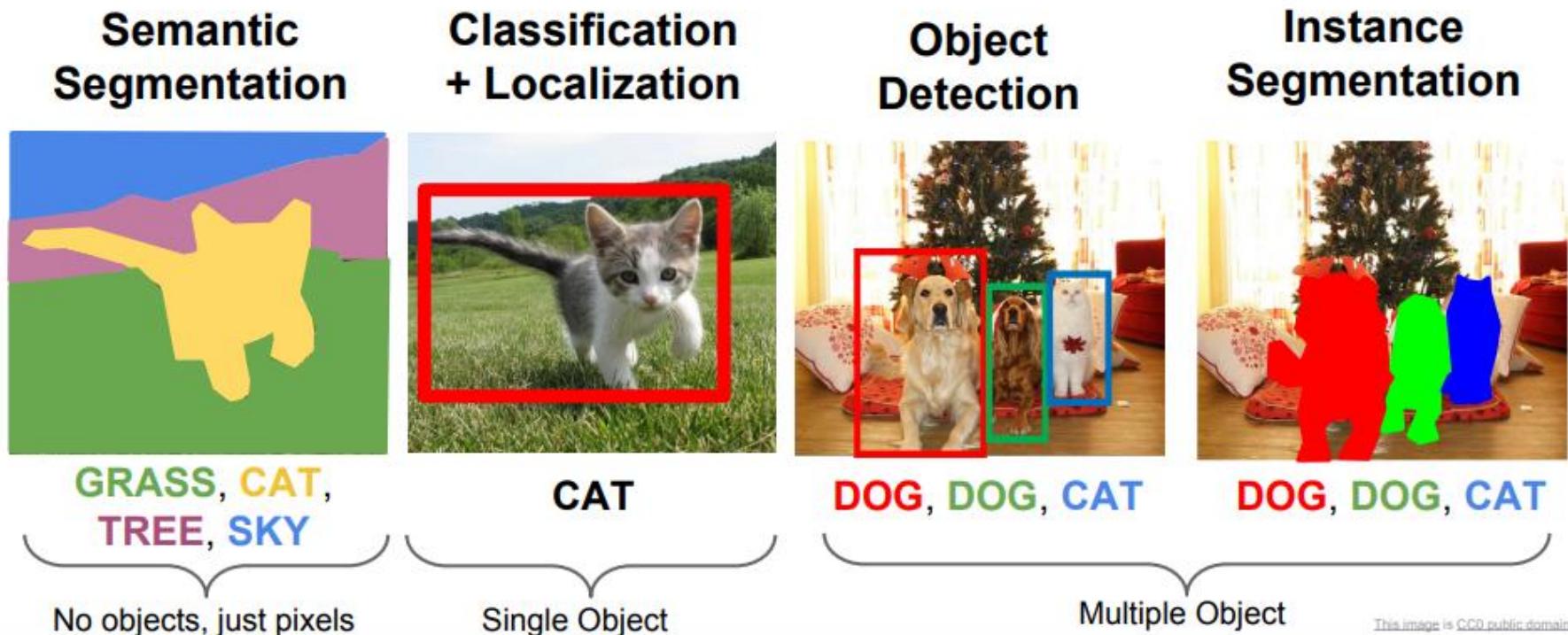
[Image by Kippelboy](#) is licensed under CC BY-SA 3.0



[Image by Christina C.](#) is licensed under CC BY-SA 4.0

Other Basic Tasks on CV

- Besides classification, there are other basic tasks in computer vision, including localization, detection and segmentation.



Object Detection

- Impact of deep learning

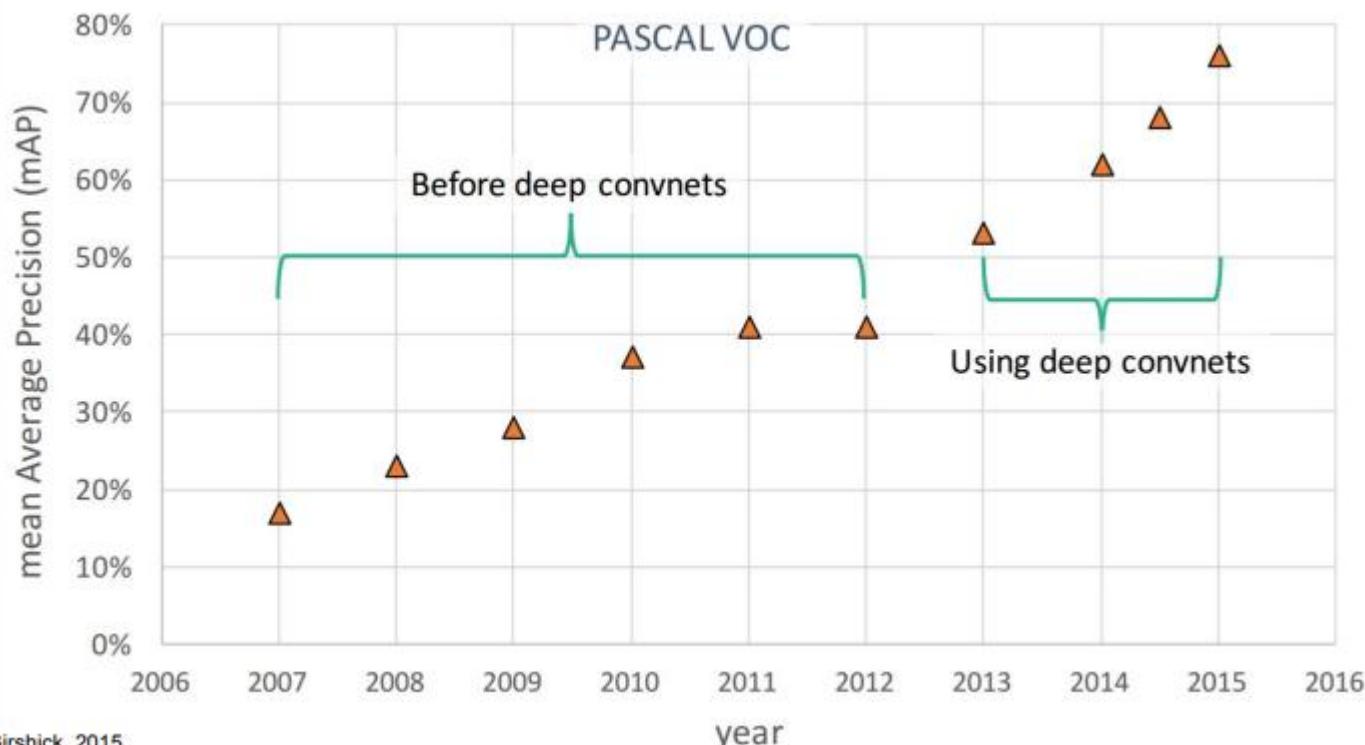
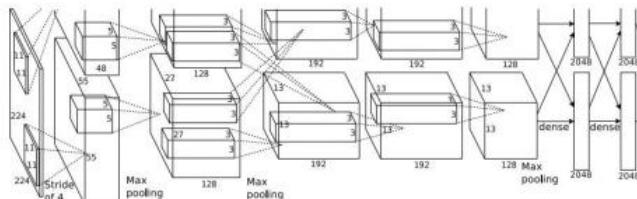


Figure copyright Ross Girshick, 2015.
Reproduced with permission.

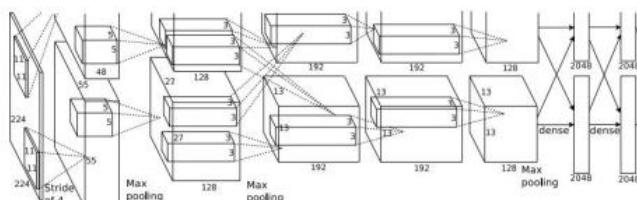
Object Detection

- As regression



Each image needs a different number of outputs!

CAT: (x, y, w, h) 4 numbers

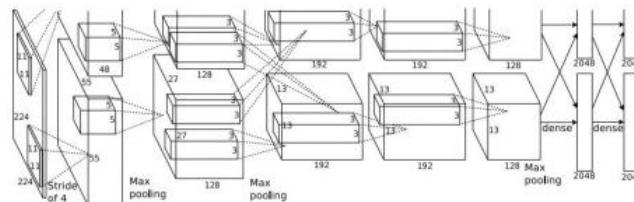


DOG: (x, y, w, h)

DOG: (x, y, w, h)

CAT: (x, y, w, h)

16 numbers



DUCK: (x, y, w, h) Many numbers!

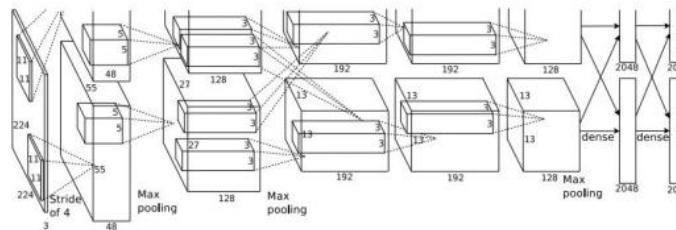
DUCK: (x, y, w, h) Many numbers!

• • •

Object Detection

- As classification

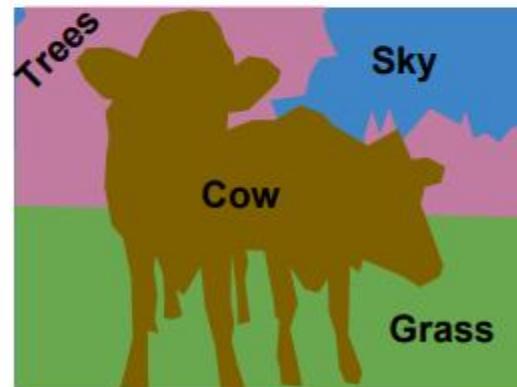
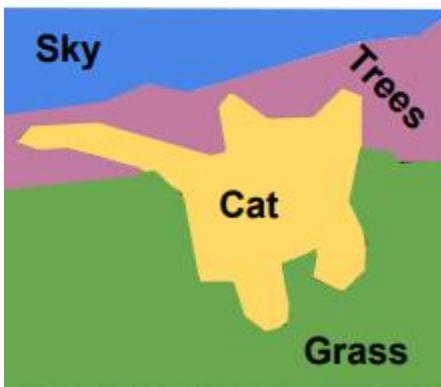
Apply a CNN to many different crops of the image, CNN classifies each crop as object or background



Dog? NO
Cat? NO
Background? YES

Semantic Segmentation

- Label each pixel in the image with a category label
- Don't differentiate instances, only care about pixels



More High-level Application

- Human Pose Estimation



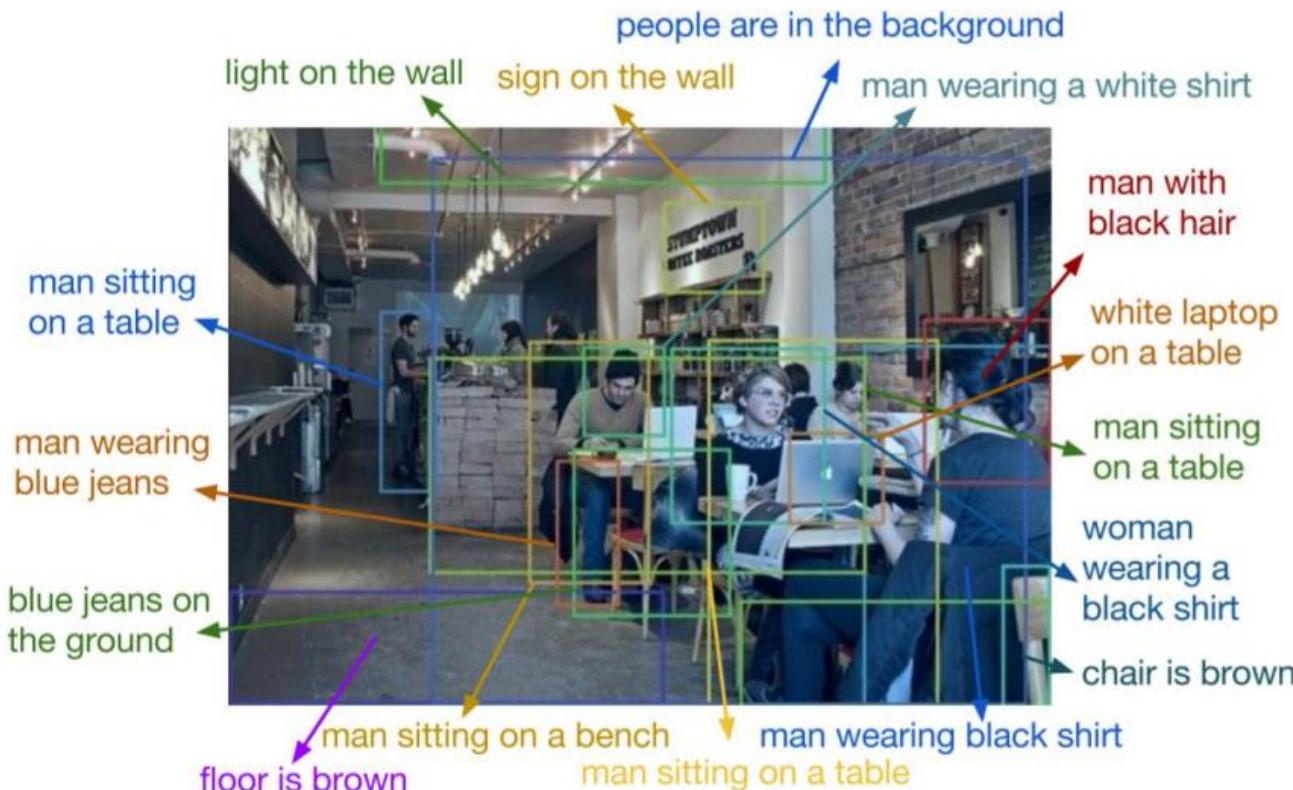
Represent pose as a set of 14 joint positions:

Left / right foot
Left / right knee
Left / right hip
Left / right shoulder
Left / right elbow
Left / right hand
Neck
Head top

This image is licensed under CC-BY 2.0.

More High-level Application

- Dense Captioning



More High-level Application

- Self-driving: Tesla



More High-level Application

- Self-driving: Xiao Peng



智能驾驶

我们基于持续提升的自研软件架构，以及全闭环、自成长的AI和数据体系，将实现无人驾驶前，辅助驾驶的终极技术形态

城市 N G P ▼

智能辅助驾驶系统 X N G P ▼

深度视觉神经网络 X N E T ^

将多个摄像头采集的数据，进行多帧时序前融合，输出 BEV 视角下的动态目标物的 4D 信息和静态目标物的 3D 信息，利用神经网络实现端到端数据驱动算法迭代

More High-level Application

- Self-driving: Baidu Apollo
- You can experience it near Jiading Campus!



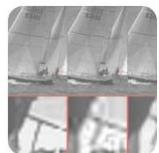
More High-level Application



Image Generation



Denoising



Super-Resolution



Action Recognition

.....

If you want to know more...

<https://paperswithcode.com/area/computer-vision>

[Browse SoTA](#) > Computer Vision

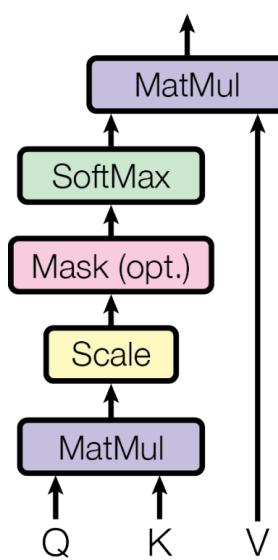
Computer Vision

2625 benchmarks • 938 tasks • 1724 datasets • 22899 papers with code

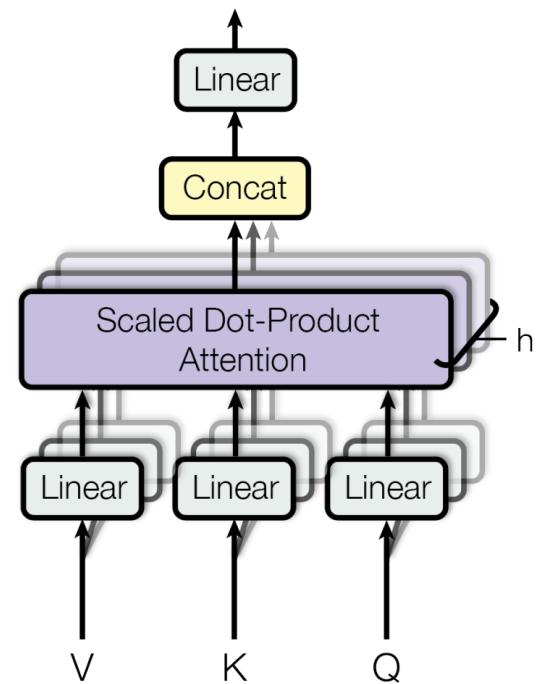
New trend: Transformer-based Methods

- Transformer: An encoder-decoder architecture
- Can be used as a very strong feature extractor
- Consists of a few simple units **with no CNN or RNN layers**

Scaled Dot-Product Attention



Multi-Head Attention



New trend: Transformer-based Methods

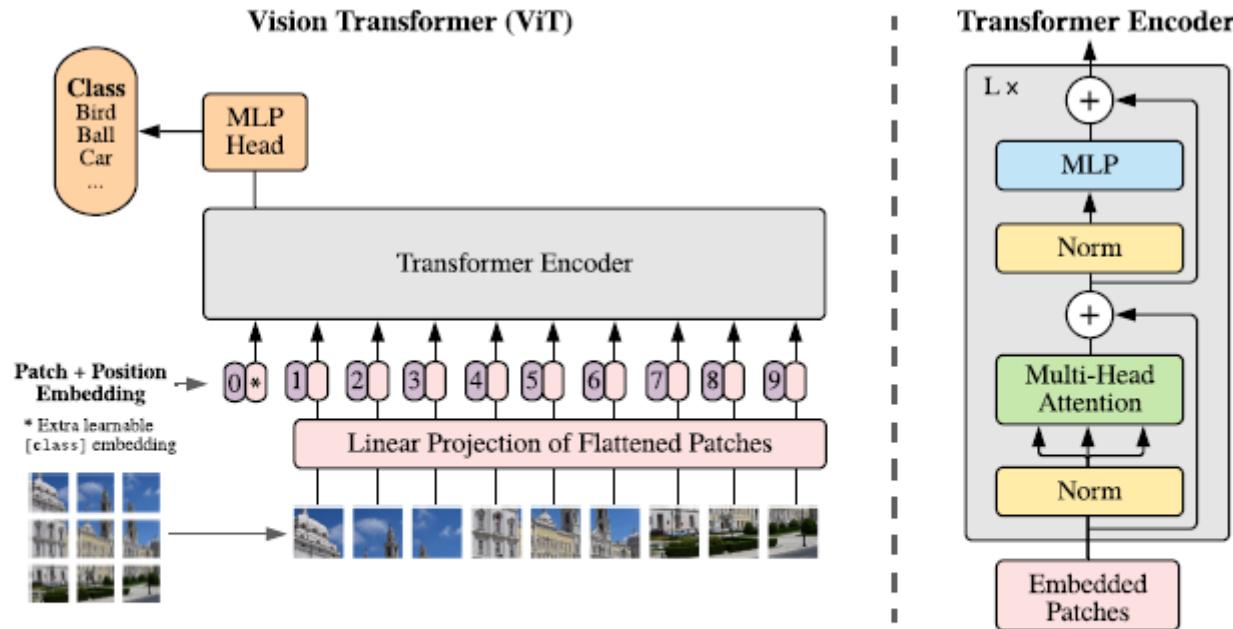


Fig. 6. Illustration of ViT. The flatten image patches with an additional class token are fed into the vanilla Transformer encoder after positional encoding. Only the class token can be predicted for classification. (from [29].)

New trend: Transformer-based Methods

- AI engineer Sudharshan gave GPT-4 a photo of a refrigerator and asked it to come up with a recipe within 60 seconds.



Based on the provided ingredients, here are five food items that can be created:

1. Fruit Salad:
Combine diced apples, oranges, and bananas in a large bowl. Optionally, you can add a squeeze of lemon juice to prevent browning and a drizzle of honey for extra sweetness. Mix well and serve chilled.
2. Cheese Omelette:
Beat 2-3 eggs in a bowl with a pinch of salt and pepper. Heat a non-stick pan with a little oil or butter, pour in the egg mixture, and cook until it begins to set. Add grated cheese on one side, fold the omelette over, and cook until the cheese melts. Serve with your choice of bread or toast.
3. Ham and Cheese Sandwich:
Layer slices of lunch meat and cheese on a slice of bread. Add lettuce, tomato, or other desired toppings, then spread mustard or mayonnaise on another slice of bread and place it on top. Enjoy the sandwich as is or grill it for a warm, toasty meal.
4. Fruit Smoothie:
Blend a mix of apples, oranges, and bananas with a cup of milk (adjust the amount to achieve desired consistency) until smooth. For added flavor, add a spoonful of honey or a few ice cubes for a cold, refreshing drink.

Recommended Reading

- The Transformer Architecture
 - *Attention is all you need*
- Transformer & Image Classification
 - *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (Known as ViT)
- Transformer & Object Detection
 - *End-to-End Object Detection with Transformers* (Known as DETR)

Today's Topics

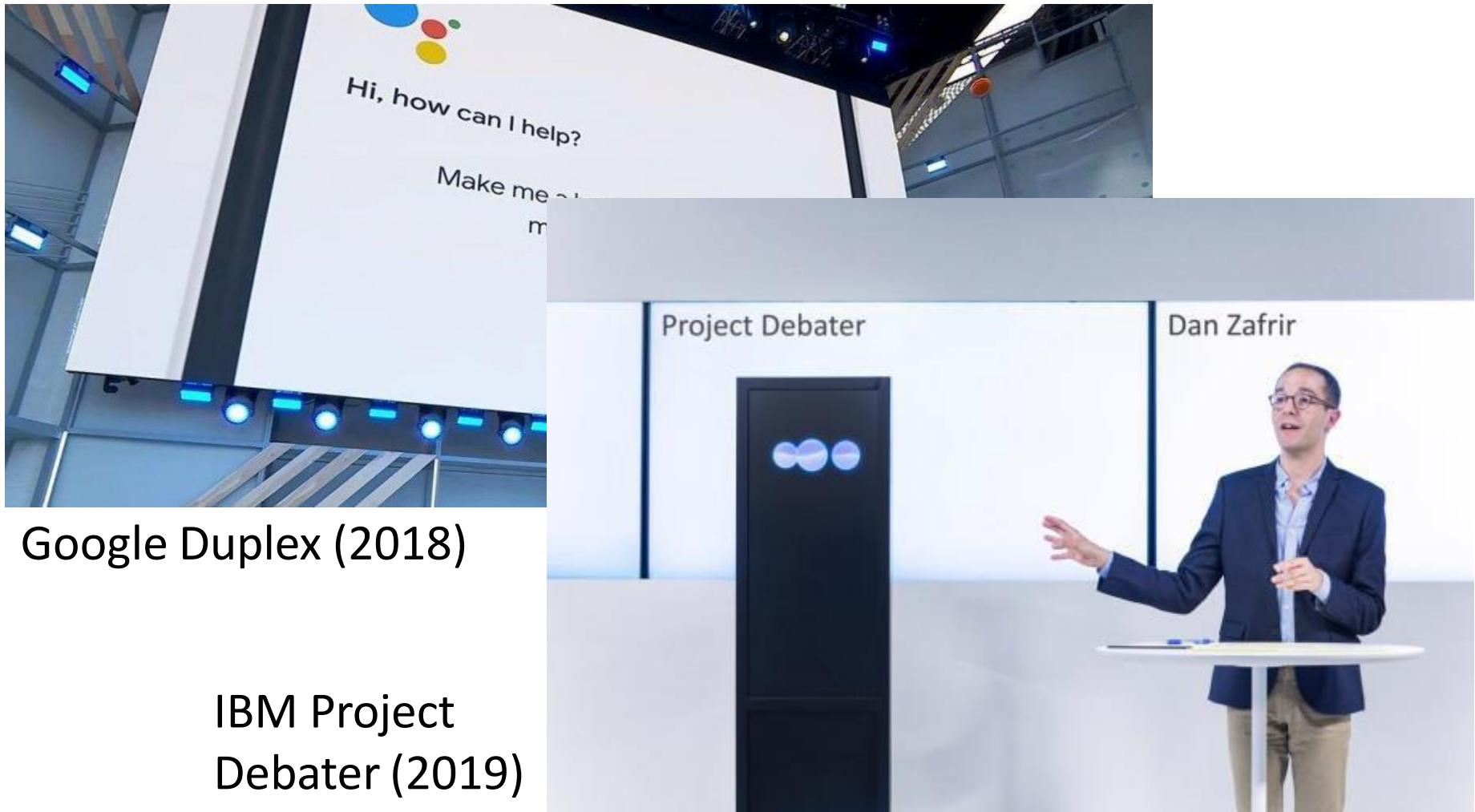
- Deep Learning & Computer Vision
- *Deep Learning & Natural Language Process (NLP)*
- Deep Learning & Medicine
- AIGC

What we want to do with NLP



- Natural Language can be Speech or Text

NLP is popular



Google Duplex (2018)

IBM Project
Debater (2019)

Human Language is Complex



1 second has 16K sample points

Each point has 256 possible values.

William Faulkner, “Absalom, Absalom.”:
“Just exactly like Father if Father had
known” (1289 words)

Jonathan Coe's *The Rotters' Club* has a sentence with 13,955 words (2014)

Faulkner wrote, “Just exactly like Father ...”

Pinker said Faulkner wrote, “Just exactly like Father ...”

Who cares that Pinker said Faulkner wrote, “Just exactly like Father ...”

The Language Instinct: How the Mind Creates Language (Steven Arthur Pinker)

Successful Model - BERT

- BERT: Bidirectional Encoder Representations from **Transformers**
- Devlin et al., 2018 proposed the “Masked Language Model” objective and released the weights of a pretrained Transformer, a model they labeled BERT.
- BERT was massively popular and hugely versatile; finetuning BERT led to new state-of-the-art results on a broad range of tasks.
- Two models are released:

	Layers	Hidden states	Attention heads	Params
BERT-base	12	768-dim	12	110M
BERT-large	24	1024-dim	16	340M

Successful Model – GPT-3

- A large language model

Model	# Parameters
Medium-sized LSTM	10M
ELMo	90M
GPT	110M
BERT-Large	320M
GPT-2	1.5B
Honey Bee Brain	~1B synapses
GPT-3	175B
Cat	~10 Trillion synapses
Human	~100 Trillion synapses

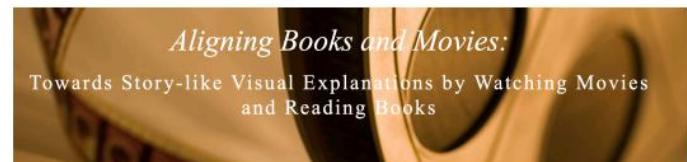
- Trained on 500 billion tokens from



WIKIPEDIA
The Free Encyclopedia



- **Transformer Based**



Playing with GPT-3

- Natural Language to Bash
- Using GPT-3 to convert natural language commands to bash one-liners

Q: List files in this directory, including hidden files

A: ls -a

Q: Go to the home directory

A: cd ~

Q: Move cat.jpg to my Desktop

A: mv cat.jpg ~/Desktop

Q: List running processes

A: ps aux

Q: find occurrences of the string "pepsi" in every file in the current directory recursively

A: grep -r "pepsi" *

Q: copy the file testing.txt onto a remote server at remotesserver.com

A: scp testing.txt remotesserver.com:~/

Q: start the node program at index.js

A: node index.js

Q: start a simple web server with python

A: python -m SimpleHTTPServer

Playing with GPT-3

- **Blending Concepts**

Idea 1: Two-Dimensional Space

Idea 2: Three-Dimensional Space

Blended Idea: 2.5-Dimensional Space

Idea 1: Old

Idea 2: New

Blended Idea: Recycled

Idea 1: Geology

Idea 2: Neurology **Blended Idea:** Sediment neurology: When sediments are buried deep under the ground, they can act like a soft layer of neurological tissue in that they will be compressed over time and can get folded and distorted. In neurology, this is called cortical folding. In geology, this is called tectonic folding.

Idea 1: Mathematics

Idea 2: Fecal matter

Blended Idea: Brownian motion: Brownian motion is a type of random motion that happens when you observe the movements of fecal matter (such as protozoans) under a microscope.

Idea 1: More

Idea 2: Less

Blended Idea: Just-Right

Idea 1: Triangle

Idea 2: Square

Blended Idea: Trapezoid

Some generated samples

Playing with GPT-3

- Copycat analogy problems

Q: If a b c changes to a b d what does p q r change to?

A: p q r changes to p q s

Q: If a b c changes to a b d what does p q r s change to?

A: p q r s changes to p q r t

Q: If a b c changes to a b d what does p p q q r r change to?

A: p p q q r r changes to p p q q s s

Q: If a b c changes to a b d what does p p p q q q r r r
change to?

A: p p p q q q r r r changes to p p p q q q s s s

Q: if a b c changes to a b d what does i j k change to?

A: i j k changes to i j l

Q: If a b c changes to a b d what does i i j j k k change to?

A: i i j j k k changes to i i j j l l

Q: If a b c changes to a b d what does x y z change to?

A: x y z changes to x z

GPT-3 output in red

Playing with GPT-4

```
if (position(x, y, currentPlayer)) {
    setTimeout(() => {
        alert(currentPlayer + ' 获胜了！');
        initializeBoard();
        ctx.clearRect(0, 0, boardSize, boardSize);
        drawBoard();
    }, 1000);
} else {
    currentPlayer = currentPlayer === 'block' ? 'white' : 'block';
}
```

此函数首先获取所有可用的位置，然后随机选择一个位置并在该位置下棋。然后，检查机器人选手是否获胜。

2. 修改“canvas”的“click”事件监听器，使用户下棋后机器人进行回应：

javascript

Copy code

```
ctx.stroke();
}

function drawPiece(x, y, player) {
    ctx.beginPath();
    ctx.arc(x * gridSize, y * gridSize, gridSize / 2 - 1, 0, 2 * Math.PI);
    ctx.fillStyle = player;
    ctx.fill();
}

function initializeBoard() {
    for (let i = 0; i < grid; i++) {
        board[i] = [];
        for (let j = 0; j < grid; j++) {
            board[i][j] = null;
        }
    }
}

function score(x, y) {
    let score = 0;
    let direction = [
        {x: 0, y: 1}, {x: 1, y: 1}, {x: 1, y: 0}, {x: 0, y: -1},
        {x: 1, y: 1}, {x: 2, y: 1}, {x: 2, y: 0}, {x: 1, y: -1},
        {x: 2, y: 1}, {x: 3, y: 1}, {x: 3, y: 0}, {x: 2, y: -1},
        {x: 3, y: 1}, {x: 4, y: 1}, {x: 4, y: 0}, {x: 3, y: -1}
    ];
    for (let d = 0; d < direction.length; d++) {
        let count = 1;
        let color = board[x][y];
        for (let i = 1; i < 5; i++) {
            let nx = x + direction[d].x * i;
            let ny = y + direction[d].y * i;
            if (nx < 0 || nx > grid || ny < 0 || ny > grid || board[nx][ny] === null) {
                break;
            } else if (board[nx][ny] === color) {
                count++;
            } else {
                break;
            }
        }
        if (count === 5) {
            score += 1000;
        } else if (count === 4 && !isBlocked(x, y, direction[d])) {
            score += 100;
        }
    }
    return score;
}, 0);
}
```

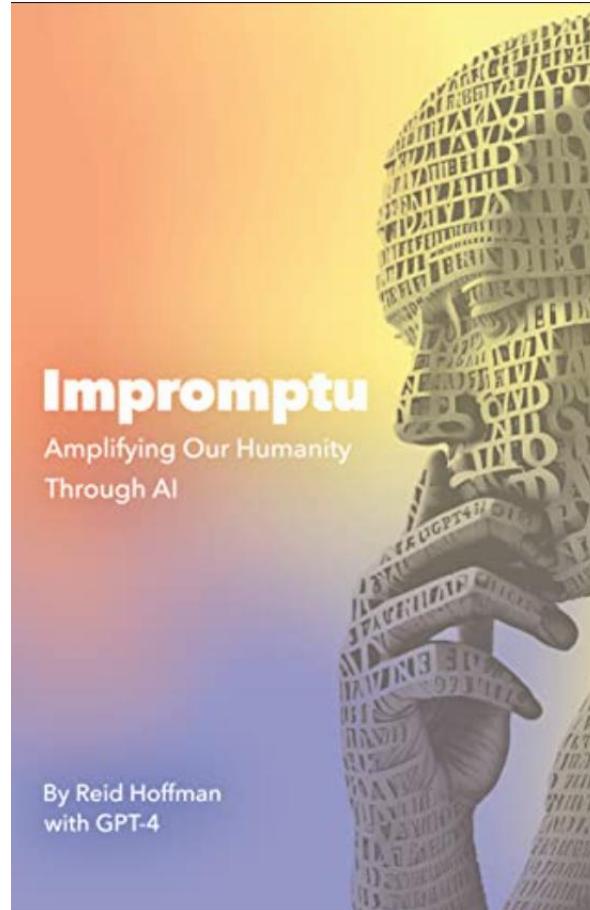
此函数会评估给定位置对应玩家的分数。它将检查所有四个方向（横向、纵向和两个对角线方向），然后根据棋子的连接性和是否被阻塞来计算分数。

2. 修改“robotMove”函数，使机器人选手根据评分系统选择位置下棋：

```
function robotMove() {
    let bestMove = null;
    let bestScore = -Infinity;

    for (let i = 0; i < grid; i++) {
        for (let j = 0; j < grid; j++) {
            if (score(i, j) > bestScore) {
                bestMove = {x: i, y: j};
                bestScore = score(i, j);
            }
        }
    }
    return bestMove;
}
```

Playing with GPT-4



OpenAI GPT

- GPT-2:
 - Pre-training - Fine-tuning paradigm
- GPT-3:
 - Context learning
 - Chain of Thought (CoT) reasoning
- GPT-3 .5:
 - Reinforcement Learning of Human Feedback
- GPT-4:
 - Multimodal

If you want to know more...

<https://paperswithcode.com/area/natural-language-processing>

[Browse SoTA](#) > Natural Language Processing

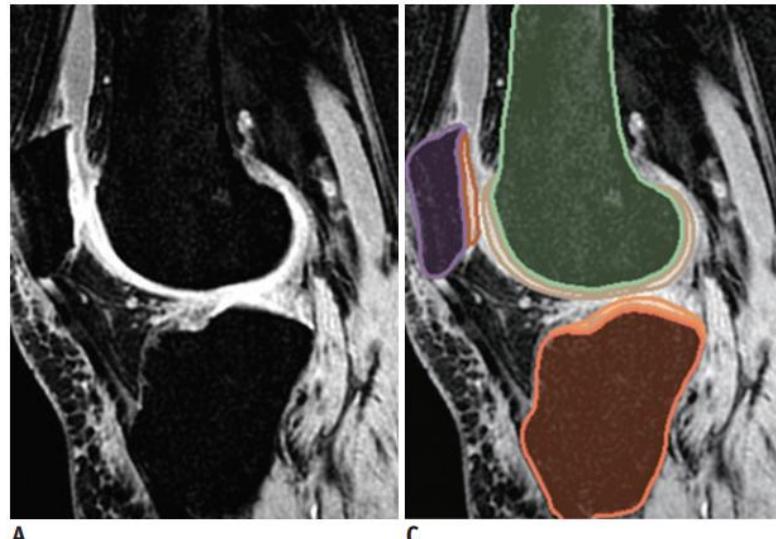
Natural Language Processing

1249 benchmarks • 444 tasks • 1261 datasets • 13138 papers with code

Today's Topics

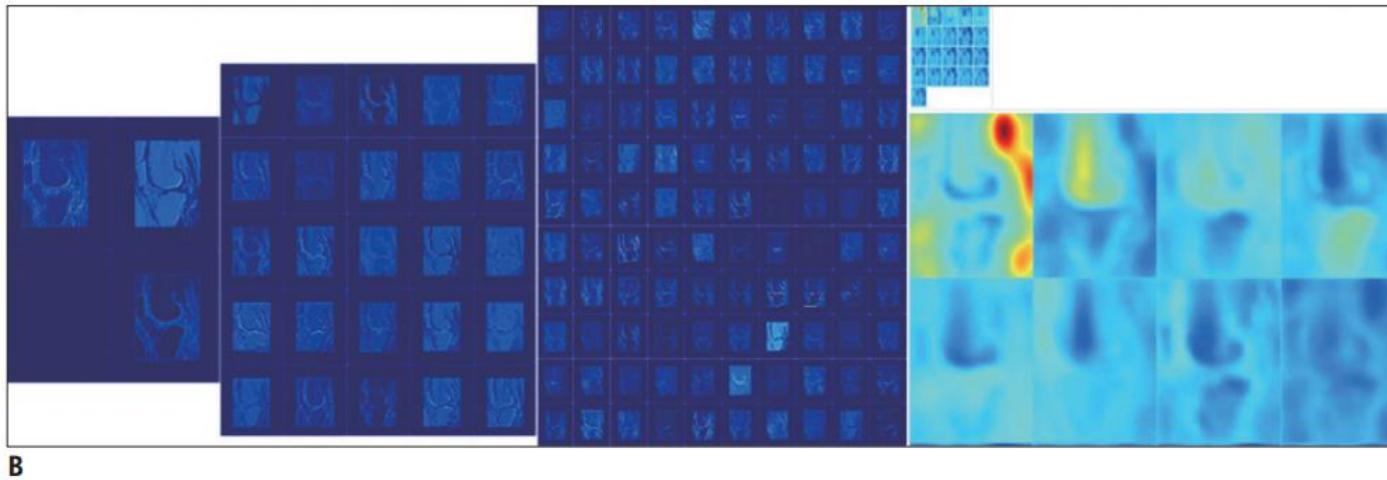
- Deep Learning & Computer Vision
- Deep Learning & Natural Language Process (NLP)
- *Deep Learning & Medicine*
- AIGC

Organ Segmentation



A

C

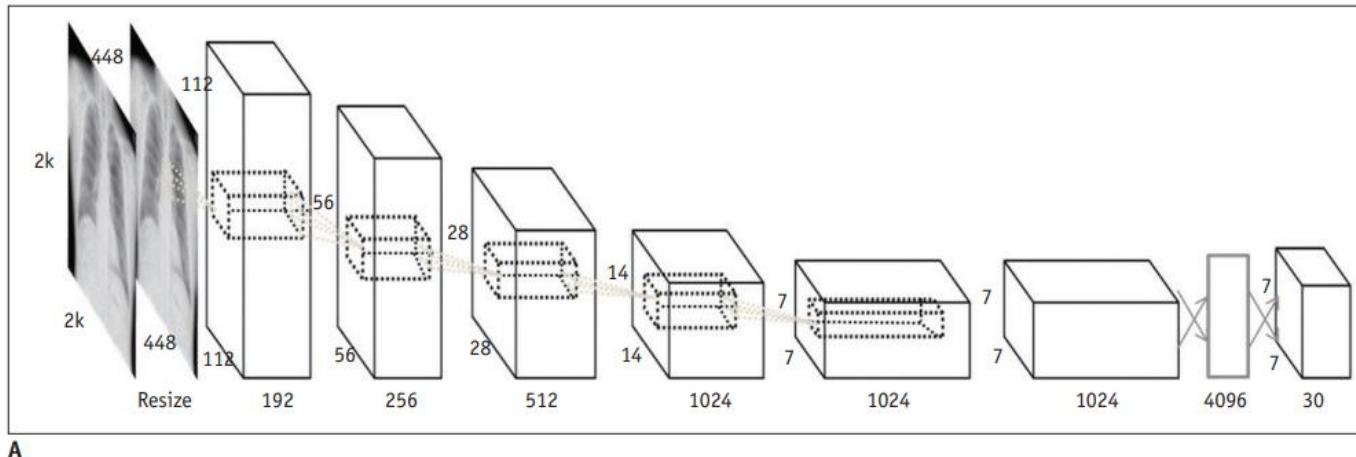


B

Example of semantic segmentation in knee MR image.

Lesion Detection

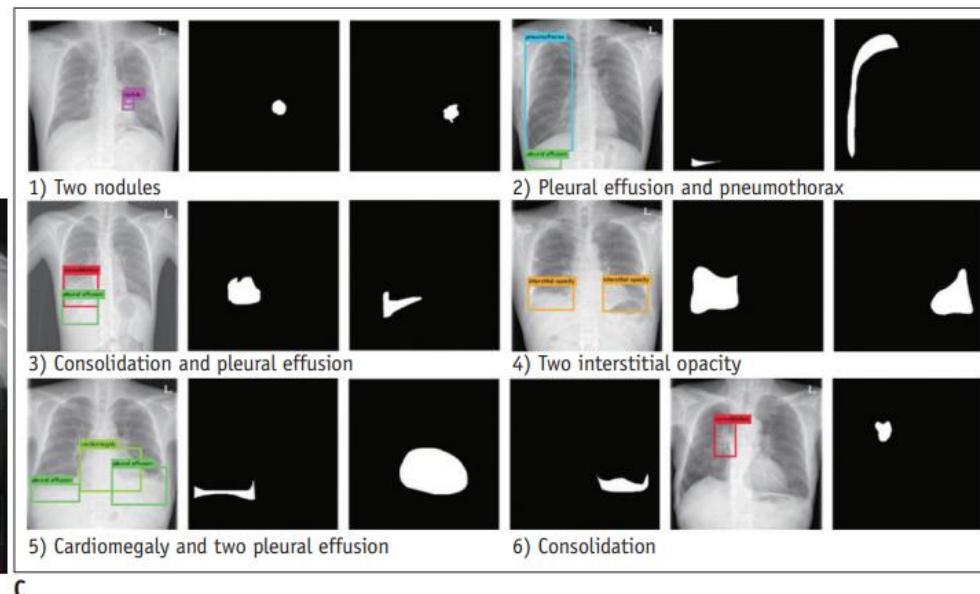
Preliminary results of lesion detection on chest radiographs, by using faster R-CNN architecture



A

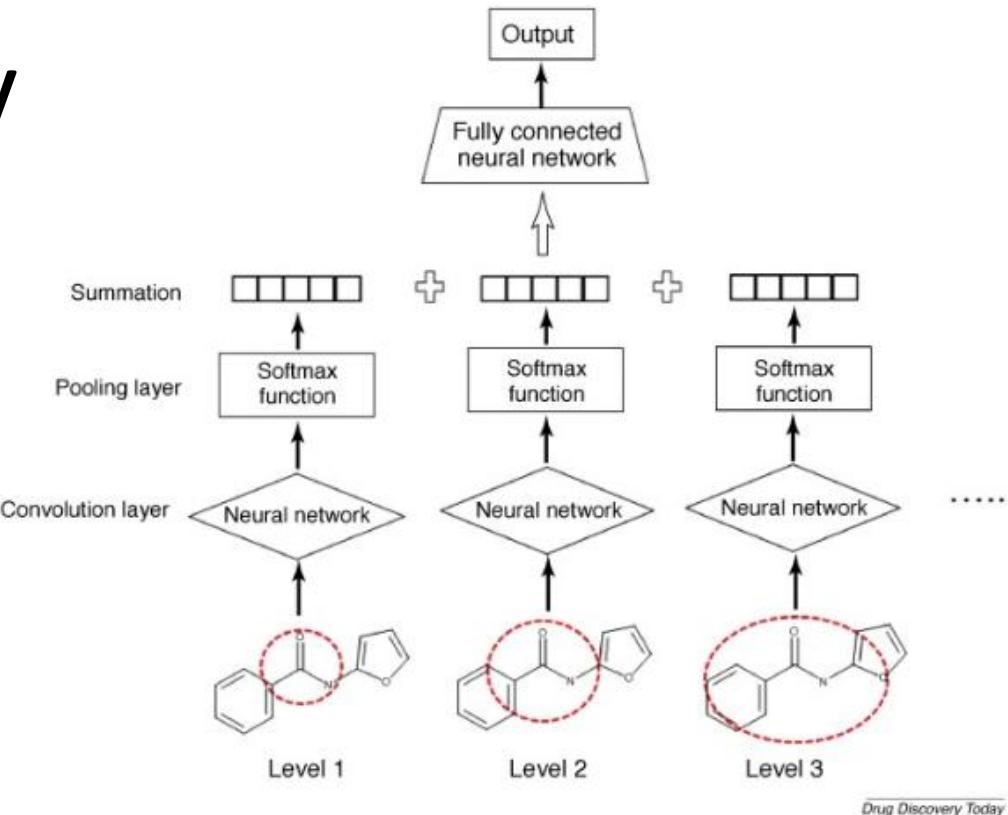
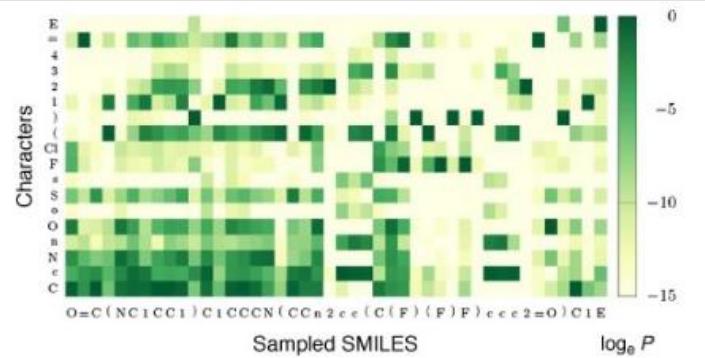


B

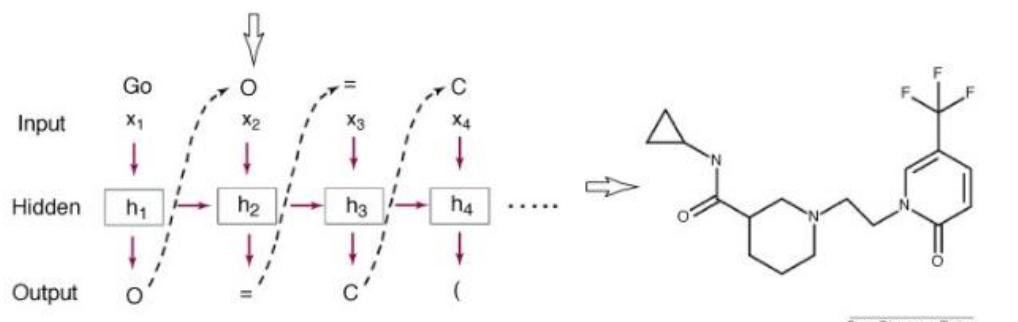


C

Drug Discovery



Drug Discovery Today



Drug Discovery Today

If you want to know more...

<https://paperswithcode.com/area/medical>

[Browse SoTA](#) > Medical

Medical

228 benchmarks • 199 tasks • 111 datasets • 1529 papers with code

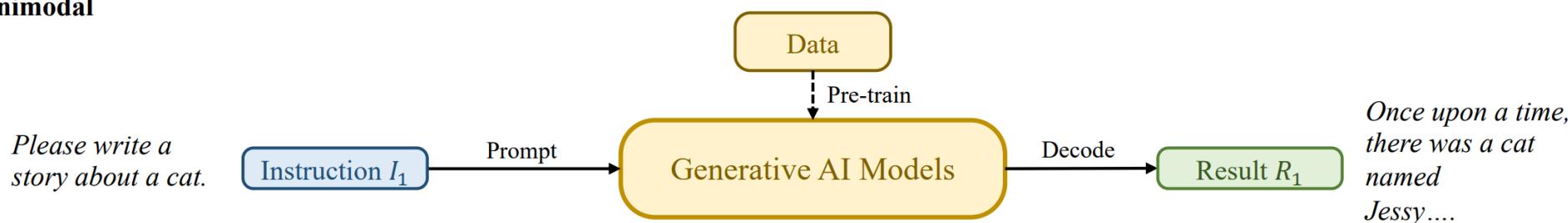
Today's Topics

- Deep Learning & Computer Vision
- Deep Learning & Natural Language Process (NLP)
- Deep Learning & Medicine
- **AIGC**

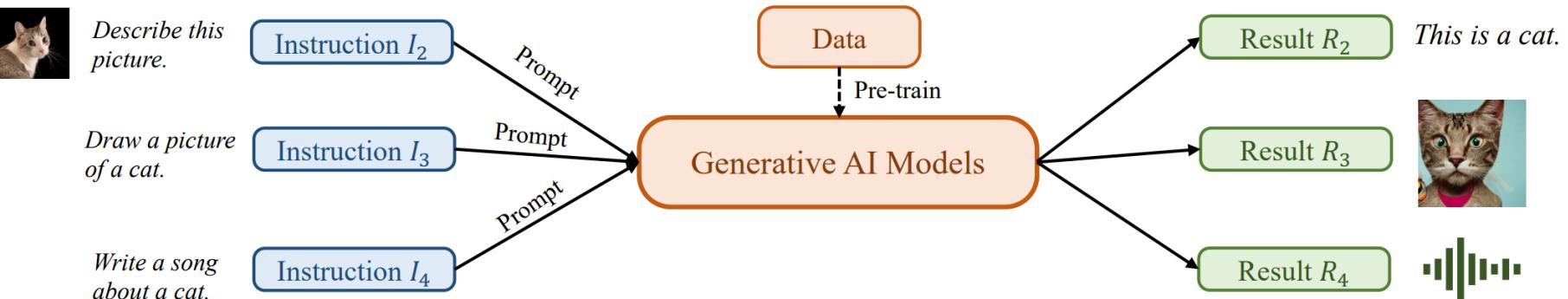
AIGC (AI Generated Content)

- The hottest AI topic at present
- Create digital content, such as images, music, and natural language, through AI models.

Unimodal



Multimodal

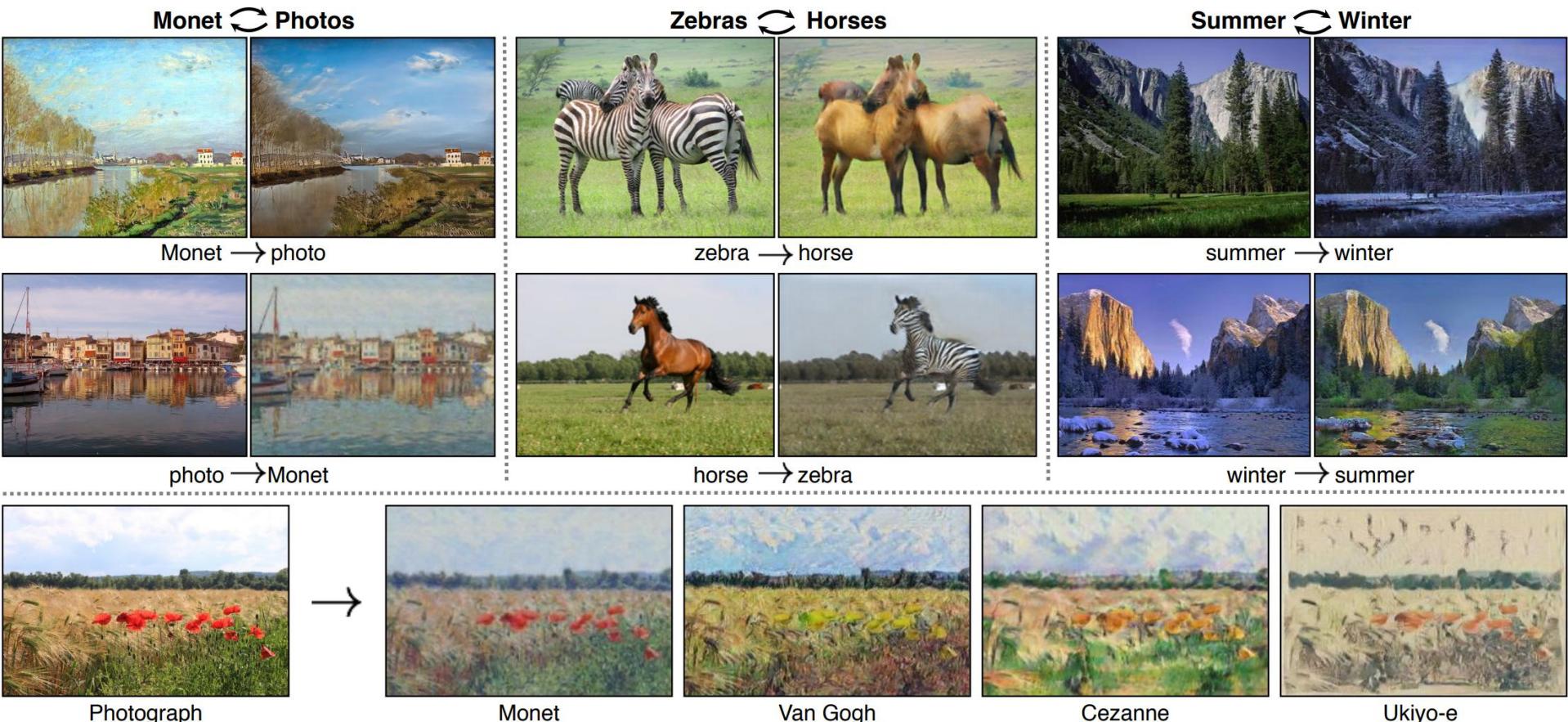


AIGC (AI Generated Content)

Company	Product	Applications
OpenAI	ChatGPT	Text generation, chatbots, and text completion
Google	LaMDA	Question answering and chatbots
NVIDIA	StyleGAN	Image generation, art, and design
Microsoft	Turing-NLG	Summarization, translation, and question answering
DeepMind	DVD-GAN	Video generation
Stability.AI	Stable Diffusion	Text to images
EleutherAI	GPT-Neo	Text generation
Baidu	ERNIE	Question answering and chatbots

Image to Image Translation

- Model: CycleGAN



Text to Image

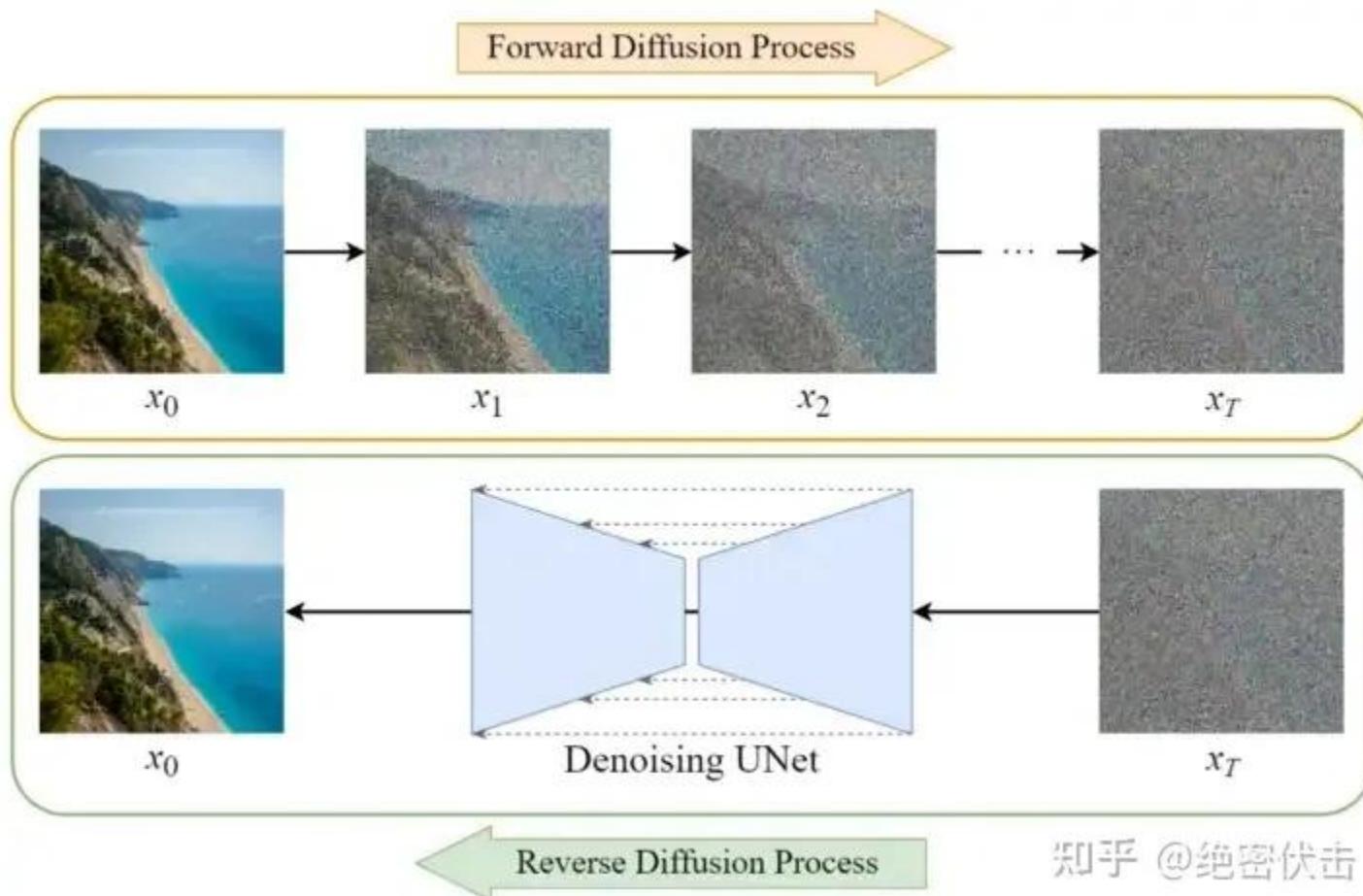
- Model: Diffusion (扩散模型)

*“A street sign that reads
'Latent Diffusion'”*



*“An oil painting
of a space shuttle”*

Diffusion



知乎 @绝密伏击

Text to Image

Stable Diffusion

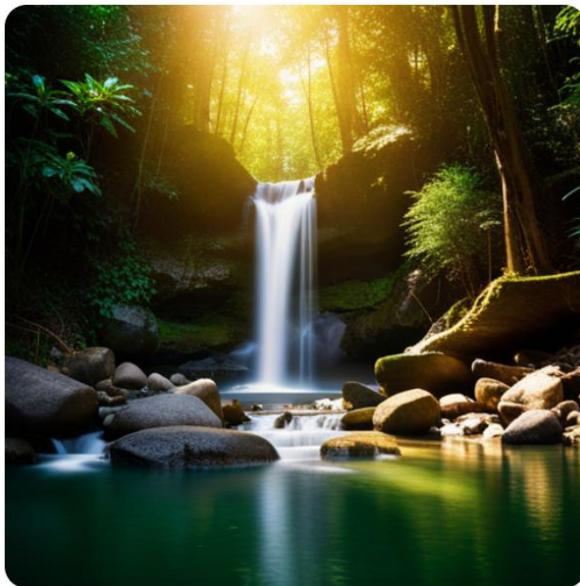
stability.ai

 **NovelAI**

Input

Beautiful waterfall in a lush jungle, with sunlight shining through the trees

Output



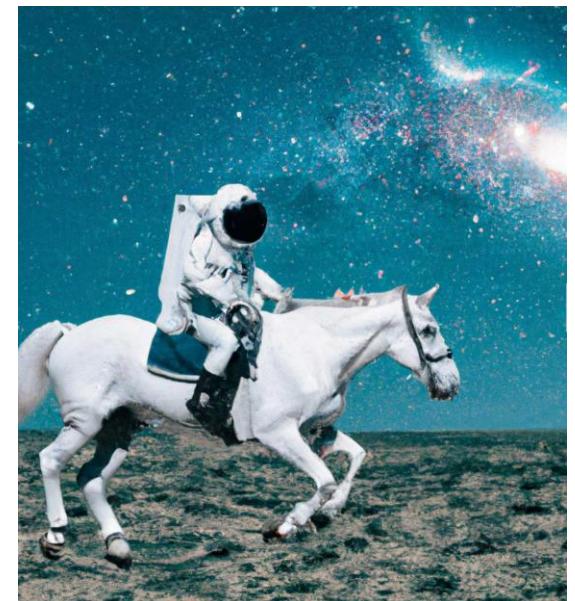
DALL·E 2

 OpenAI

Input

An astronaut riding a horse in photorealistic style.

Output



Text to Image

Different prompt:

Pictures
Sketches
human skeleton posture
.....

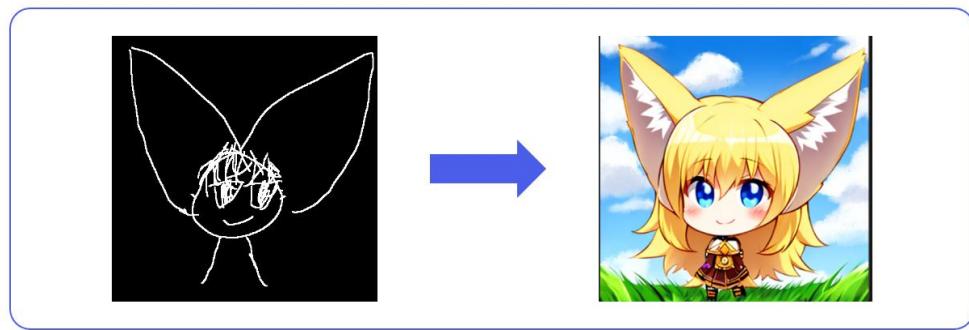
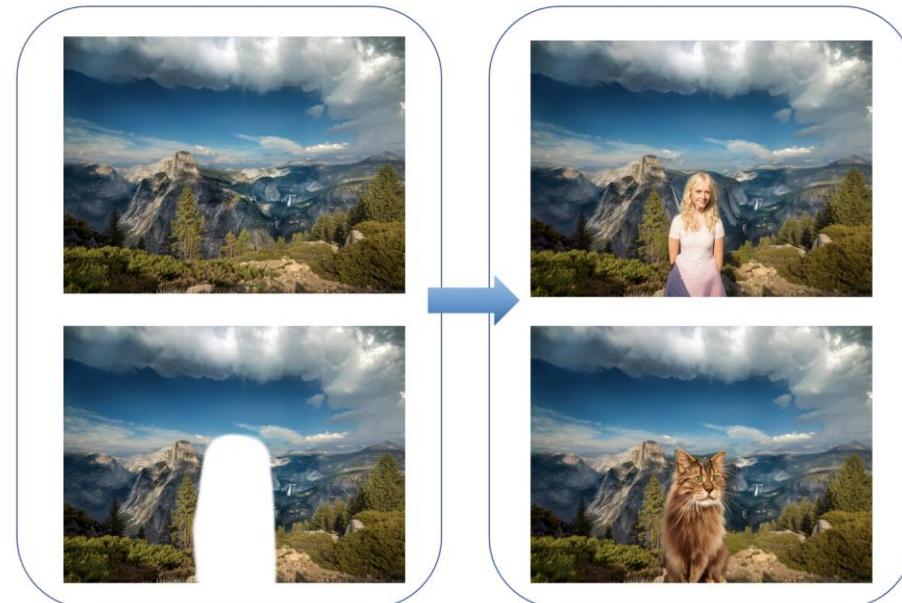
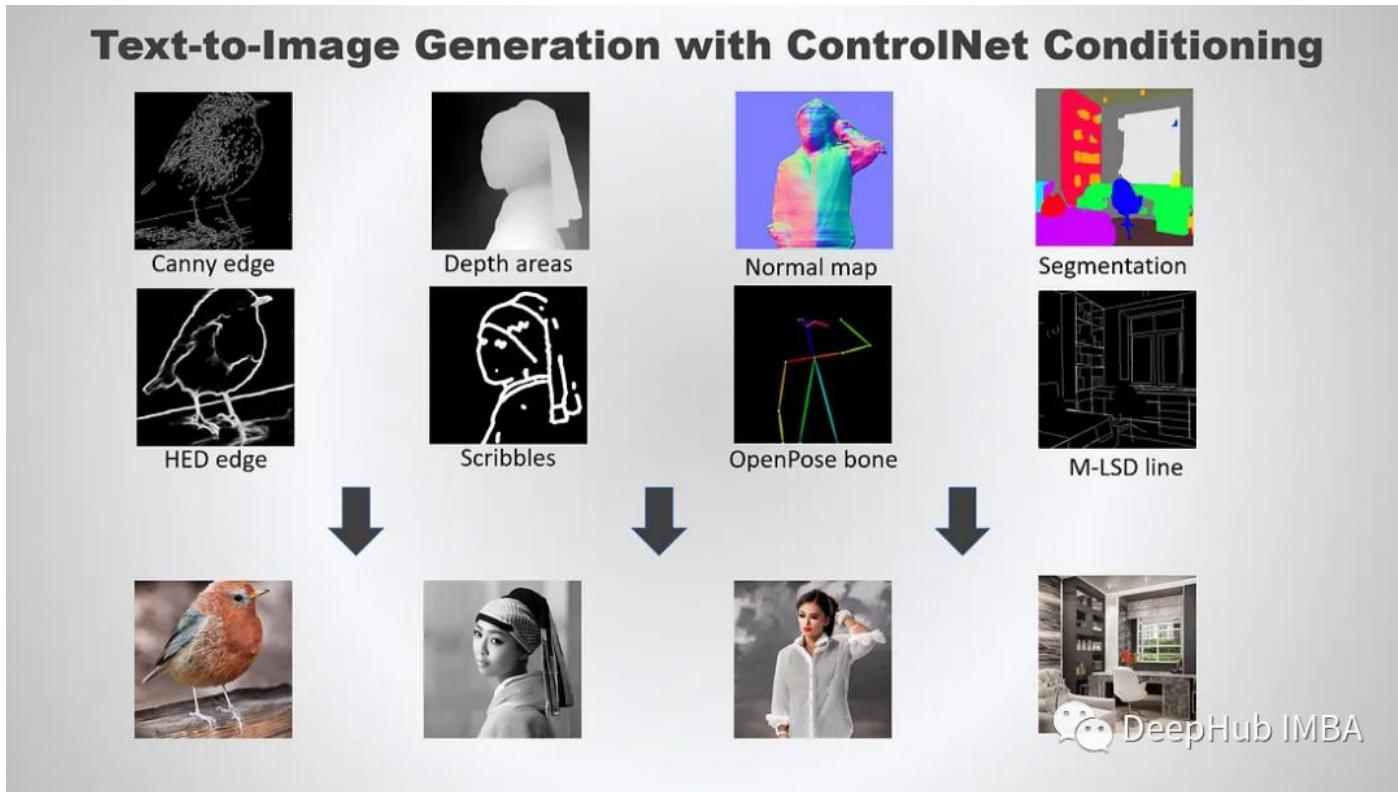
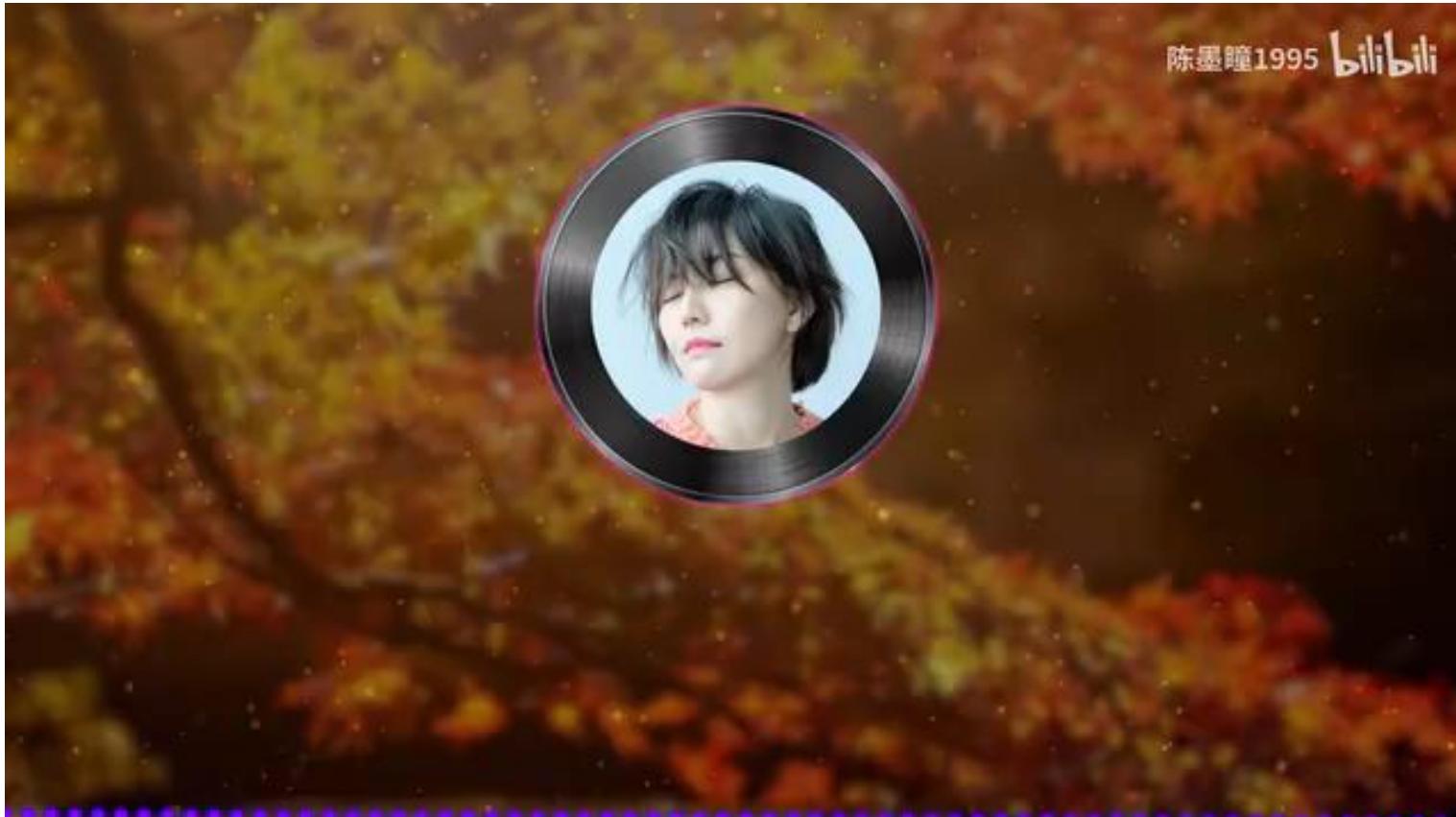


Image-to-Image



Vocal Synthesis

- A popular and interesting application



[AI 孙燕姿] 《发如雪》 cover 周杰伦_哔哩哔哩_bilibili

Video



ChatBot



- Model: GPT-3 (Release) → GPT-4 (Current)

ChatGPT

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Here is a list of their availability:

- Andrew: 11 am to 3 pm
- Joanne: noon to 2 pm, and 3:30 pm to 5 pm
- Hannah: noon to 12:30 pm, and 4 pm to 6 pm

Based on their availability, there is a 30-minute window where all three of them are available, which is from 4 pm to 4:30 pm. So, the meeting can be scheduled at 4 pm.

GPT-4

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Andrew: 11 am - 3 pm

Joanne: 12 pm - 2 pm, 3:30 pm - 5 pm

Hannah: 12 pm - 12:30 pm, 4 pm - 6 pm

Common availability for a 30-minute meeting: 12 pm - 12:30 pm

GPT-4 surpasses ChatGPT in its advanced reasoning capabilities.

ChatBot



- Model: ERNIE 3.0

Input (Factural QA)	问题:发明加减消元法的数学家所在国家有哪些品牌汽车? 答:
Generation	有很多, 比如说奔驰、宝马、奥迪、大众、保时捷。
Input (Opinion QA)	问题:宝宝眼睛流泪是怎么回事? 答:
Generation	这种情况可能有三个原因: 一是宝宝泪腺的炎症。二是宝宝的眼睛有倒睫毛或者睫毛刺激眼球引起流泪。三是宝宝有结膜炎。所以如果宝宝眼睛流泪, 首先要排除宝宝倒睫毛刺激眼球引起的流泪, 如果不是倒睫毛引起的流泪, 还要考虑有没有泪腺炎症或者结膜炎的可能。
Input (Reverse Dictionary)	问题:描写"人的品质优良"的词语有哪些? 答:
Generation	品质优良,品德高尚,诚实,忠厚,老实
Input (Novel Generation)	江西孟龙潭与朱孝廉客都中, 偶涉一兰若, 殿宇禅舍, 俱不甚弘敞, 惟一老僧挂褡其中。见客入, 肃衣出迓, 导与随喜。殿中塑志公像, 两壁画绘精妙, 人物如生。东壁画散花天女, 内一垂髫者, 拈花微笑, 樱唇欲动, 眼波将流。
Generation	西壁画一小沙弥, 面目奇古, 双手合十, 似在祈祷。客大骇, 请问之。僧曰: “此乃佛经也。”客曰: “此何所谓?”僧曰: “此有佛性, 无所不有。”客曰: “何以见之?”僧曰: “佛性即是佛性。佛性无所不有, 所以不能言说。”客大悟, 遂去。
Input (Couplet Generation)	上联:口十心思, 思乡, 思友, 思父母 下联:
Generation	眼四面观, 观景, 观人, 观自己
Input (News Summarization)	花盆里有黑色小飞虫, 这应该是夏天大家都比较头疼的问题。灭小飞虫方法很多, 给大家分享几种。电蚊拍电, 粘虫板粘, 杀虫剂喷, 烟蒂水灌根, 诱杀法, 其他异味水灌根。消灭花盆里的小黑飞虫, 可以把烟蒂水灌根和电蚊拍拍打结合起来, 坚持一段时间就能彻底消灭。大蒜香烟泡一泡, 用这一碗水, 小虫去无踪。文章标题是《
Generation	花盆里有黑色小飞虫怎么办? 》

DeepSeek

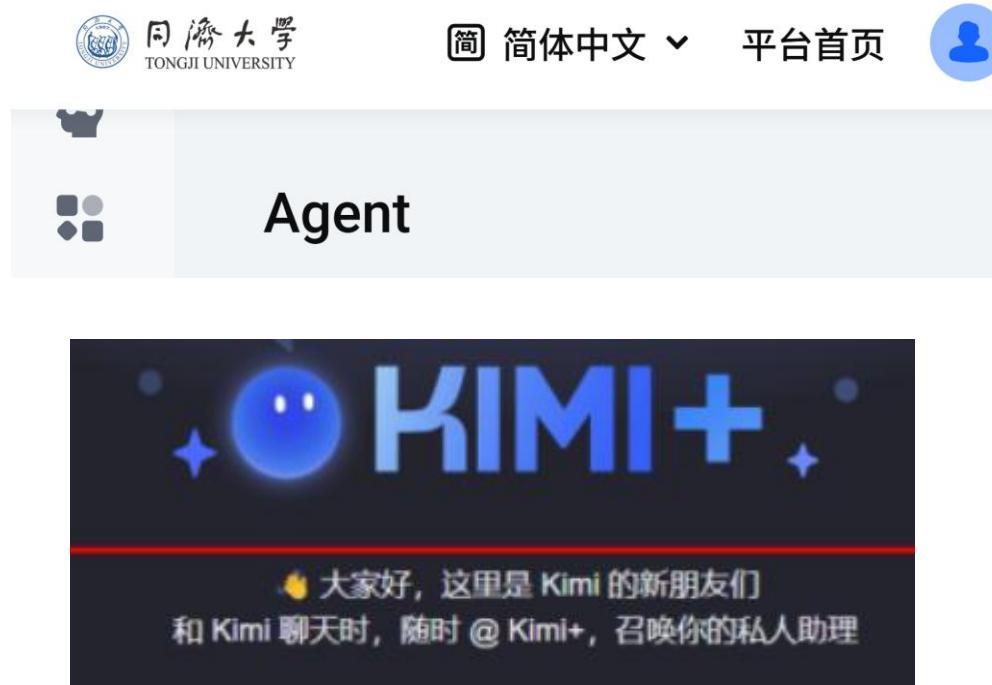
- Low-cost training

- low-priced APIs

- Making a sudden surge in adoption by general users for large models.



Agent



扣子，用 Agent 重塑生产力

低门槛 | 灵活编排 | 全链路评测 | 安全可信 | 开箱即用

AIGC Security

全国信息安全标准化技术委员会秘书处

信安秘字〔2023〕146号

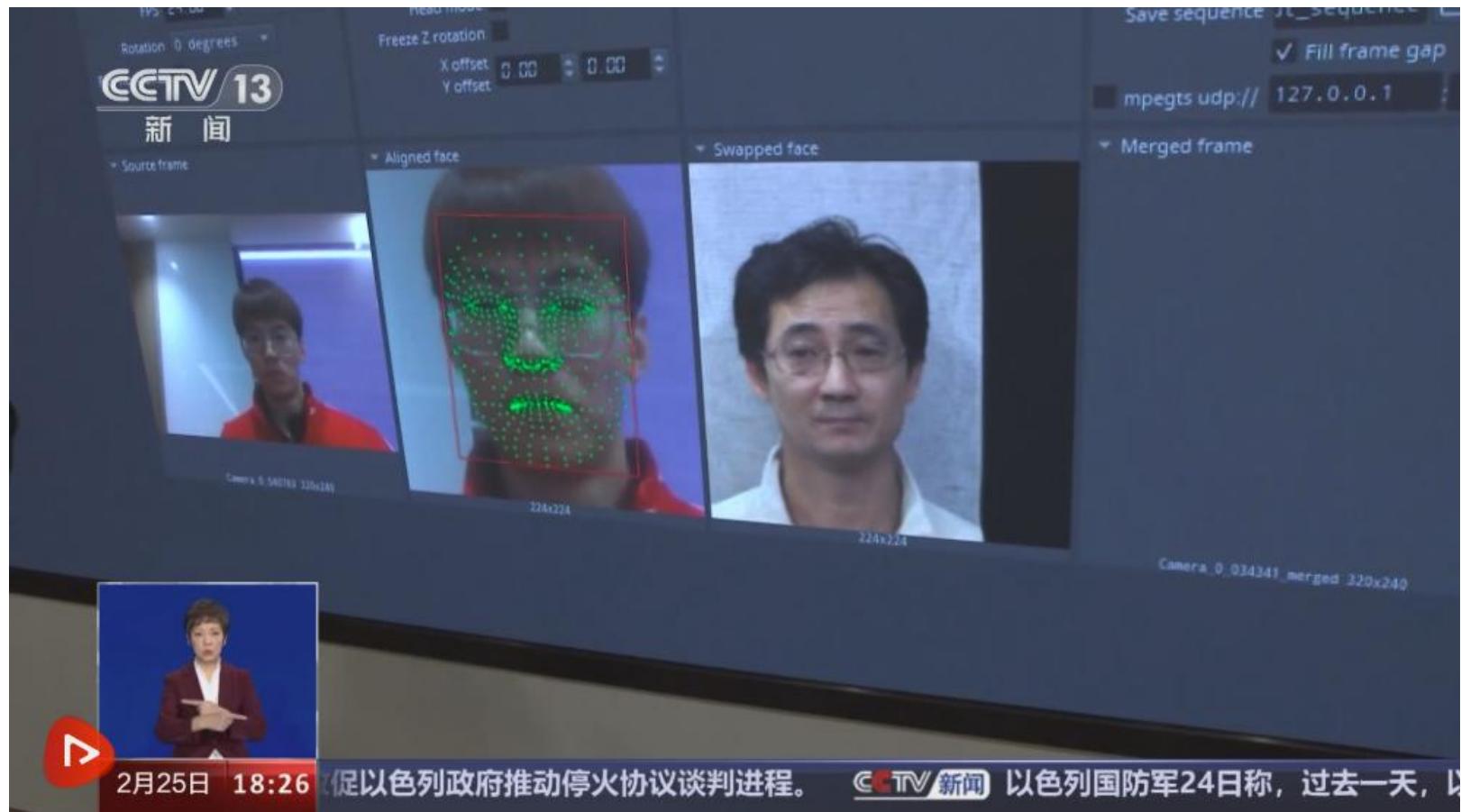
关于征求信安标委技术文件《生成式人工智能服务 安全基本要求》（征求意见稿）意见的通知

全国信息安全标准化技术委员会组织制定的技术文件
《生成式人工智能服务安全基本要求》已形成征求意见稿。

根据《全国信息安全标准化技术委员会技术文件制订工作程
序（试行）》，现将该技术文件面向社会公开征求意见，如
有意见或建议请于2023年10月25日24:00前反馈秘书处。

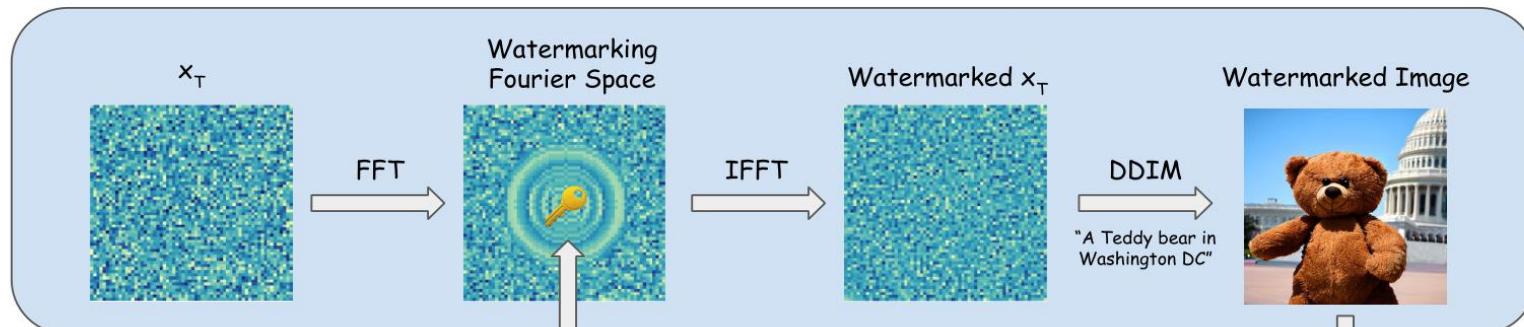


AIGC Security



AIGC Fingerprints

Generation



Detection

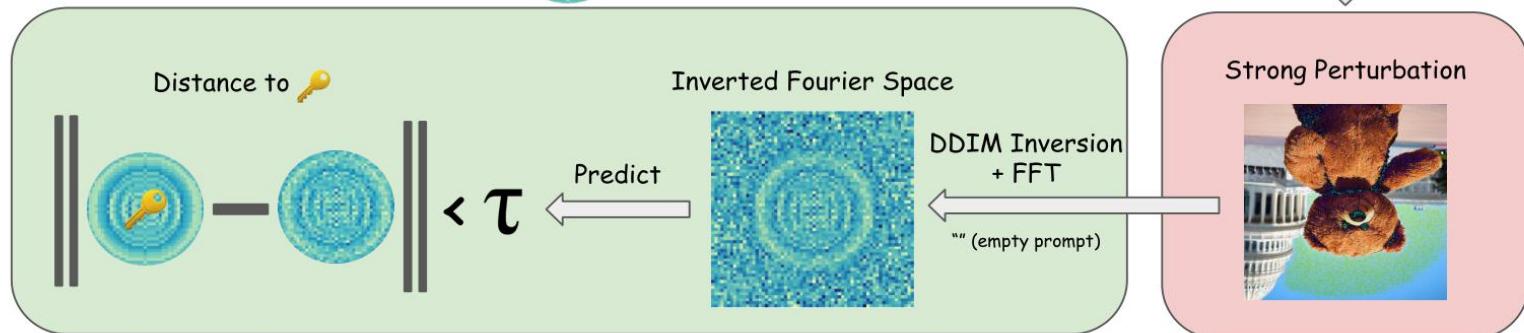


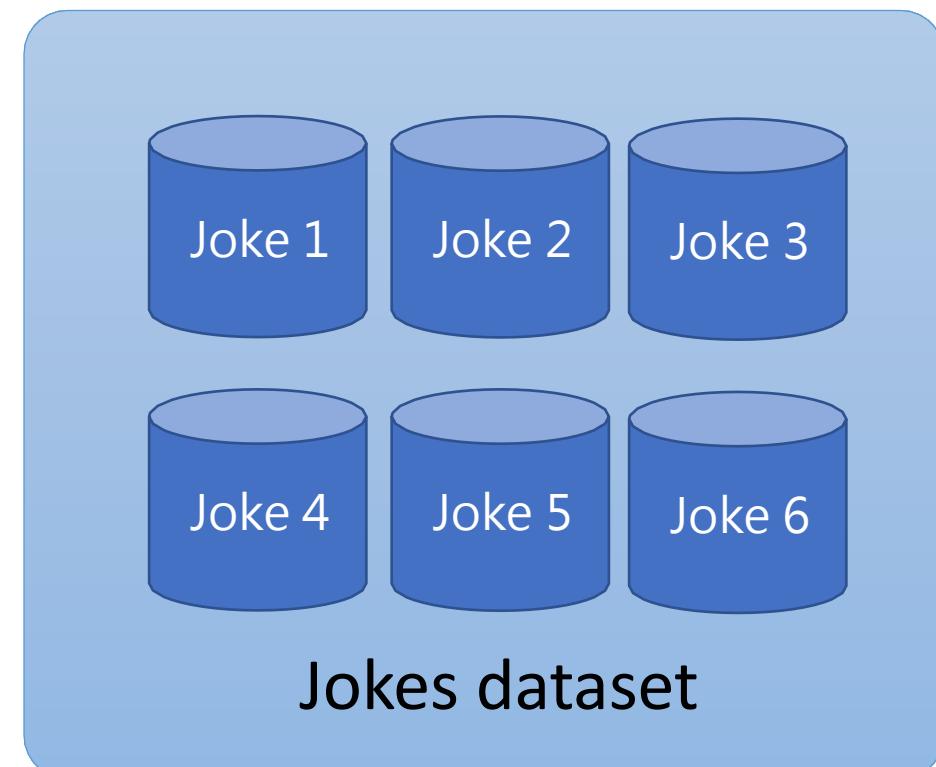
Figure 1: Pipeline for *Tree-Ring Watermarking*. A diffusion model generation is watermarked and later detected through ring-patterns in the Fourier space of the initial noise vector.

- Wen Y, Kirchenbauer J, Geiping J, et al. Tree-ring watermarks: Fingerprints for diffusion images that are invisible and robust[J]. arXiv preprint arXiv:2305.20030, 2023.

What ChatGPT really did

- Misunderstanding 1
- ChatGPT uses data prepared by developers to respond

Tell a joke ?



What ChatGPT really did

- Misunderstanding 2
- ChatGPT searches answers from the web

What is Diffusion Model ?



reorganization

Diffusion Model is ...

(Sentences from web)

(Papers from web)

What are Diffusion Models?
July 11, 2021 · 26 min · Lilian Weng

Diffusion Models: A Com
Ling Yang, Zhilong Zhang, Yang Song
Hsuan Yang

Diffusion models have emerged as a pow
applications, including image synthesis, v
expanding body of work on diffusion mod
estimation, and handling data with specia
models for enhanced results. We further
natural language processing, temporal da
provide a contextualized, in-depth look at
further exploration. Github: this <https://github.com/lilianweng/diffusion-models>

[Upd
Yang
[Upd
[Upd
So fa
have
own.
their
archi

What ChatGPT really did

- Misunderstanding 2
- ChatGPT searches answers from the web
- Most of ChatGPT's answers cannot be found on the Internet in the exact same sentence
- There are even a lot of imaginary answers

4. Can I trust that the AI is telling me the truth?

- ChatGPT is not connected to the internet, and it can occasionally produce incorrect answers. It has limited knowledge of world and events after 2021 and may also occasionally produce harmful instructions or biased content.

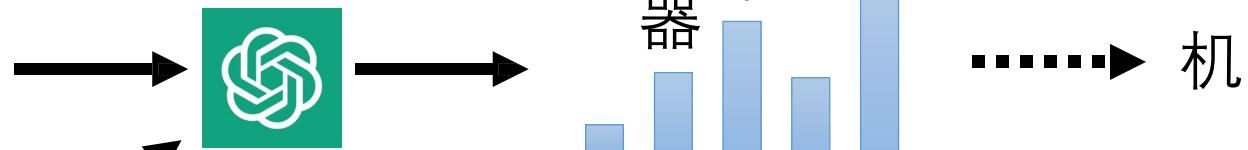
[What is ChatGPT? | OpenAI Help Center](#)

What ChatGPT really did

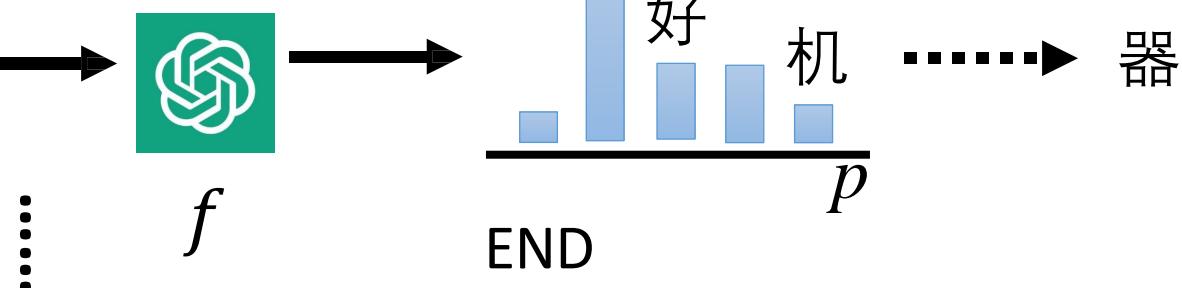
- What ChatGPT really did – 文字接龙

每次答案
都不相同

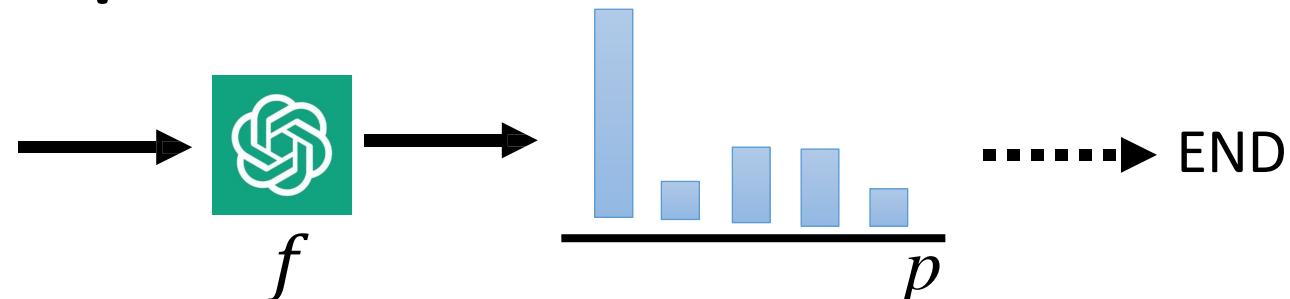
什么是机器学习?



什么是机器学习? 机

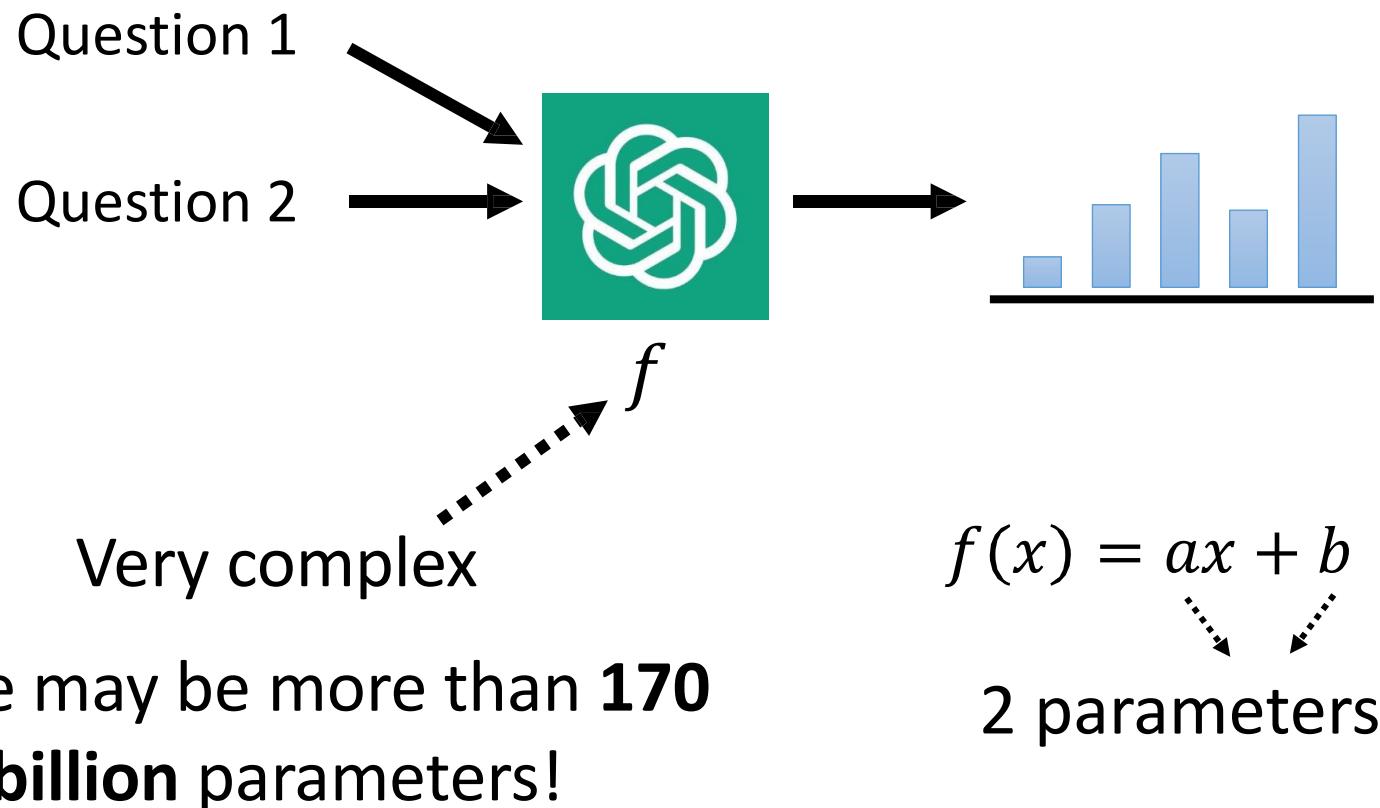


什么是机器学习?
机器学习是一门
很深奥.....。



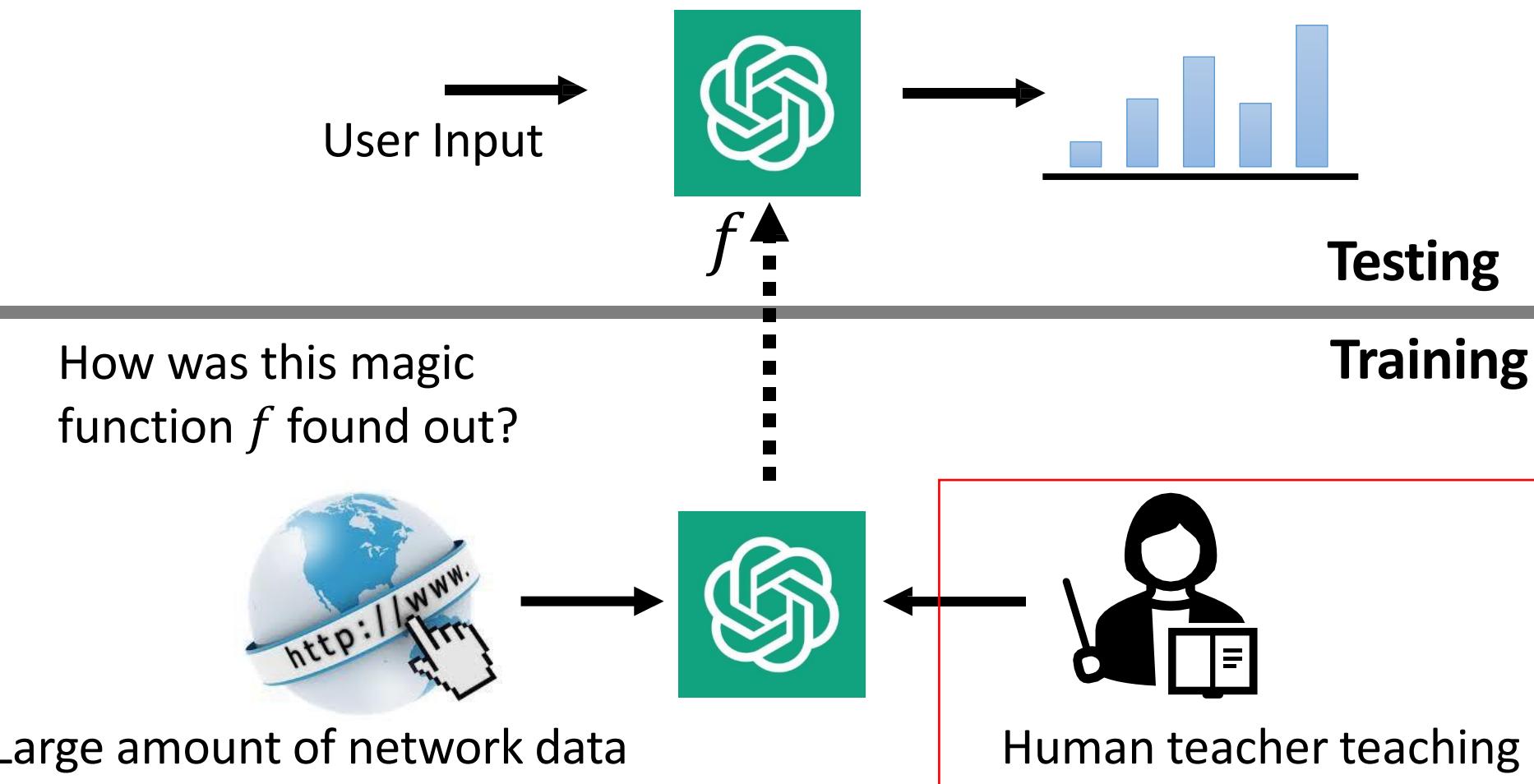
What ChatGPT really did

- What ChatGPT really did – 文字接龙



What ChatGPT really did

- What ChatGPT really did – 文字接龙

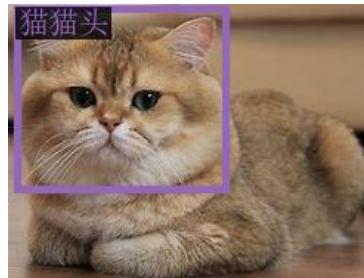


ChatBot

- Technology Behind: **RLHF**
- Reinforcement Learning from Human Feedback
- Core Steps
 1. Pretraining a language model (LM),
 2. gathering data and training a reward model, and
 3. fine-tuning the LM with reinforcement learning.
- See [Illustrating Reinforcement Learning from Human Feedback \(RLHF\) \(huggingface.co\)](#) for more details

Contrast and deficiency

- 阿里云Qwen-VL



- BingChat



- ✓ 分析图像: 人脸可能模糊以保护隐私
- ✓ 正在搜索: 男人在做什么
- ✓ 正在为你生成答案...

这张照片中的人正在人行道上骑自行车。他穿着一件蓝色衬衫和灰色裤子。自行车是红色和黑色的，靠在一座建筑物上。背景是一座有柱子的建筑和远处行人¹。

Agent



At the end of the course...

- Deep learning is hot, but it doesn't solve all problems
- Need a lot of data, a lot of computing power
- There is still a lack of interpretability, which brings uncontrollability
- When solving practical problems, it is necessary to comprehensively consider factors such as cost and complexity

Tradeoff