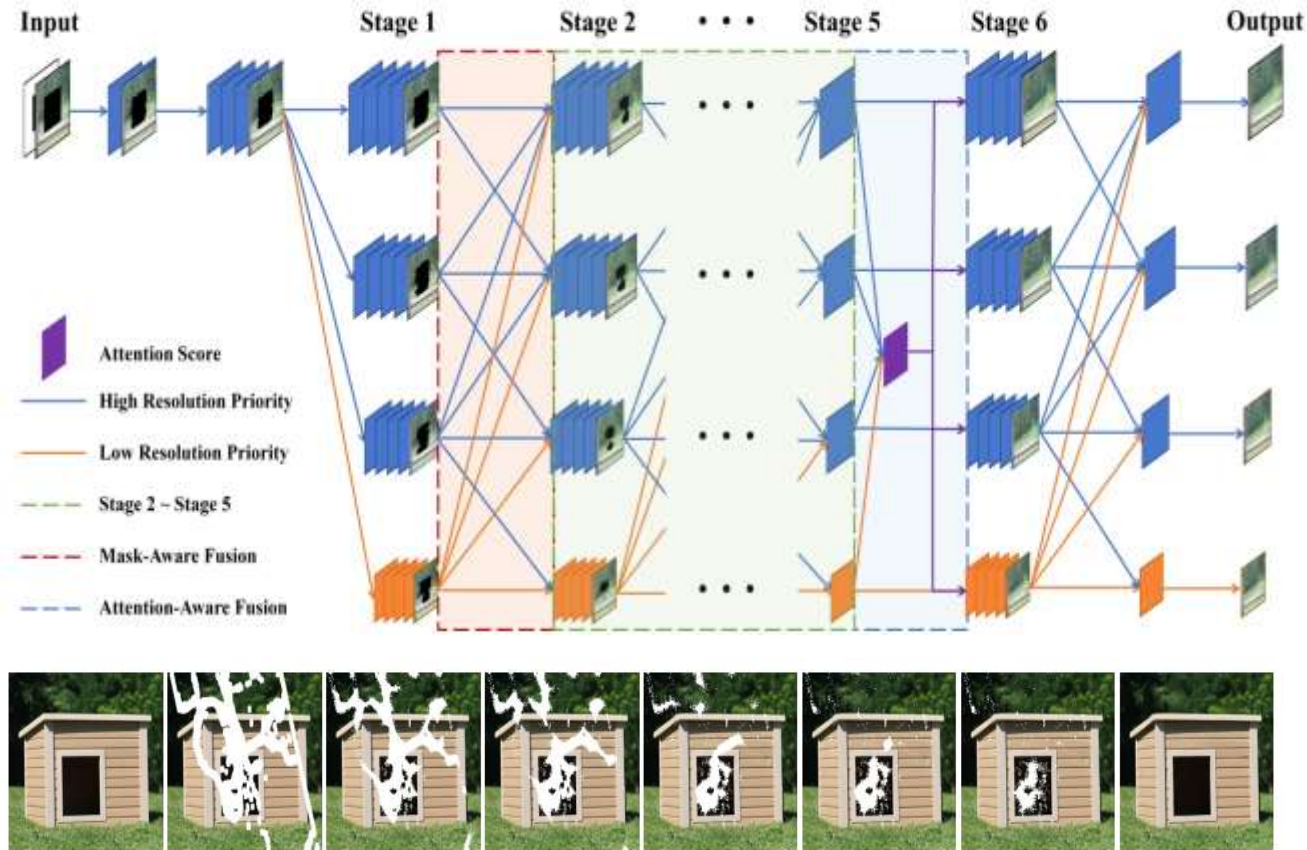


图像处理与视觉

个人信息

- 助理教授 张健夫
 - 2011 交大ACM班 | 2015 交大计算机系博士 | 2023 计算机学院
 - c.sis@sjtu.edu.cn <https://matt-lab.org/>
 - 研究方向：人工智能 | 视觉内容生成 / 深度模型可靠性
 - 助教：孙贤兵/高宇杰 计算机学院 博士一年级
 - 负责课程
 - John班 大三 专业选修课 《图像处理与视觉》
 - 研究生专业前沿课 《设计与理解神经网络》
 - ACM班 大三 专业选修课 《视觉内容生成》

代表工作：视觉内容修复/编辑

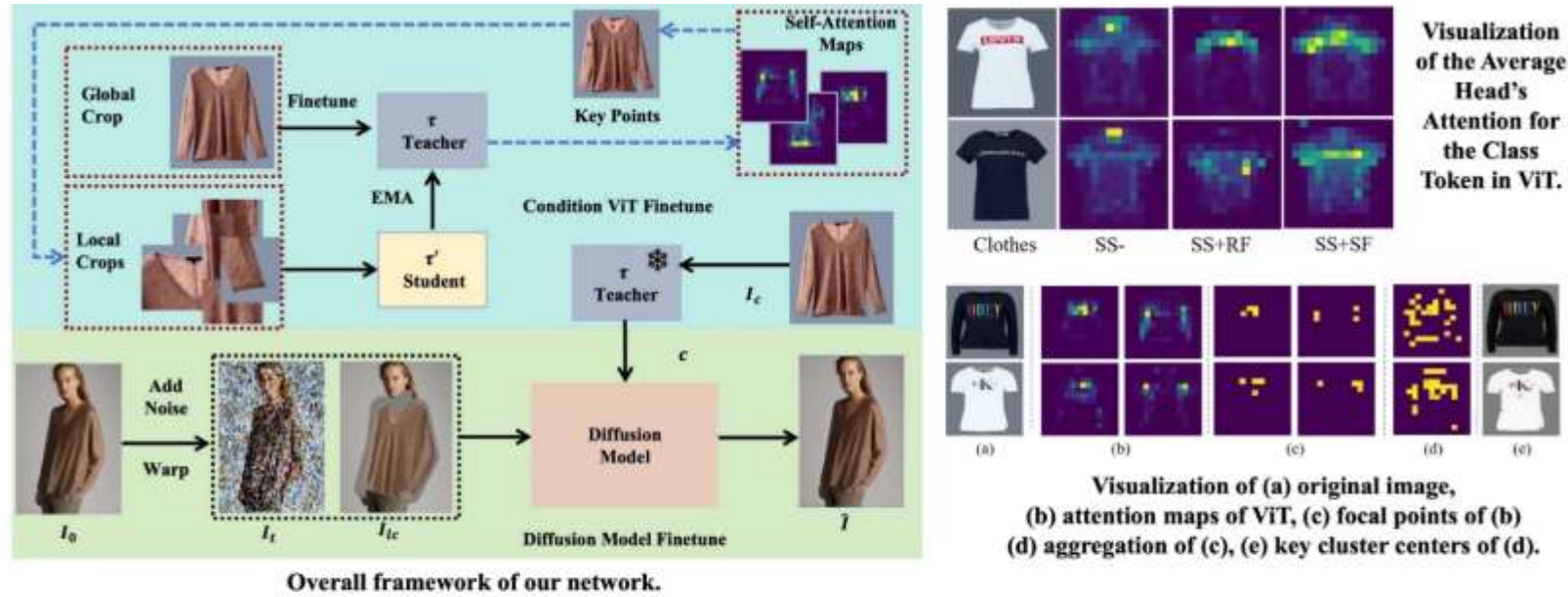


通过修复/编辑区域规划多视角/多尺度融合提升图像/视频修复编辑能力
合作开发 马卡龙玩图，获得Apple APP Store “最佳本土APP”

[ACM MM 2019] J Zhang, L Niu, D Yang, L Kang, Y Li, W Zhao, L Zhang. GAIN: Gradient Augmented Inpainting Network for Irregular Holes.
[ICCV 2021] *W Wang, *J Zhang, L Niu, H Ling, X Yang, L Zhang. Parallel Multi-Resolution Fusion Network for Image Inpainting.

[PR 2023] W Wang, L He, L Niu, J Zhang, Y Liu, H Ling, L Zhang. Diverse image inpainting with disentangled uncertainty.
[CVPR 2022] W Wang, L Niu, J Zhang, X Yang, L Zhang. Dual-path image inpainting with auxiliary gan inversion.

代表工作：虚拟服装试穿



通过服装变形引导扩散模型生成高精度服装试穿图像，利用自监督ViT进一步提升对服装细节的处理。
公开代码在GitHub半年内迅速获得350+ star，为虚拟服装试穿领域目前最有影响力的公开工作。



Qualitative comparisons with baselines.



[ACM MM 2023] J Gou, S Sun, J Zhang, J Si, C Qian, L Zhang. Taming the Power of Diffusion Models for High-Quality Virtual Try-On with Appearance Flow

Arxiv: Virtual Accessory Try-On via Keypoint Hallucination; Dynamic Automatic Natural Image Matting with Refined Guidance and Consistent Training; Self-Supervised Vision Transformer for Enhanced Virtual Clothes Try-On



AI-GENERATED IMAGE DETECTION

Dataset Construction

Fine-tuning MLLM

FakeXplainer

Trained on the dataset with

SFT + *Progressive* GRPO

based on *Qwen-2.5-VL-32B-Instruct*



Generated by FLUX.1-Schnell

大纲、教材与参考课程

- [Computer Vision: Algorithms and Applications](#) Richard Szeliski
- CS231n at Stanford
- Introduction to Computer Vision CS5670 at Cornell
- CS W182 / 282A at UC Berkeley
- Modern Computer Vision and Deep Learning (CS 198-126) at UC Berkeley
- 近期论文
- 以及一点点的个人理解和冷笑话...

评分规则

- 无平时分
- 小作业 60%（满足要求即给足分）
 - 3次，约一个月一次
 - 内容：图像处理、深度学习模型、图像生成
 - 我们会提供 代码（需做少量修改）/计算资源
 - 可以用不同的代码/计算资源，需提前告知
 - 提交简易报告
 - 每人使用的数据必须不同
- 科研项目 40%
 - 1-3人一组
 - 从不同课题中选择一个作为研究课题
 - 严禁抄袭
 - 提交报告（8页单栏中文/英文报告）

About Coding Homework ...

小作业约2次，涉及CNN、ViT等模型的搭建。

作业在 Canvas 上提交，只需要上传一份 **.ipynb** 文件，务必保留每个单元格的运行结果。

为保证较为简单的开发环境配置，我们建议在 **Colab** 上进行工作。如果你的PC配置允许，也可以在本地运行。

注意事项：

1. 请在标明的 **TODO** 位置完成代码，无需更改其他代码；
2. 按顺序运行完成后即可保存或导出，各代码块的运行结果将会保留。
3. 如果你在 **Colab** 上运行，建议仅在必要时使用**GPU**，以节约免费的动态限额。

0. Load Preliminary Functions

a. Import Libraries and Functions

b. MNIST Data Loader

c. Training Function

d. Evaluation Function

1. AutoEncoder (AE)

a. Model

b. Loss Functions and Optimizers

c. Training

d. Evaluation

2. Variational AutoEncoder (VAE)

b. Loss Functions

Loss 2: KL Divergence w/o Estimation

c. Verifying loss 1 == loss 2

d. Training with ELBO loss

e. Evaluation

f. Training with KL Divergence w/o Estimation

```
class AE(torch.nn.Module):
    def __init__(self, input_dim, hidden_dims):
        super().__init__()

        assert hidden_dims[-1] == 2, "always use 2 as the latent dimension for"
        self.encoder = torch.nn.Sequential()
        self.decoder = torch.nn.Sequential()

        #####
        ### Problem 1 (a): finish the implementation for encoder and decoder
        ### TODO:
        #####

    def decode(self, z):
        return self.decoder(z)

    def forward(self, x):
        encoded = self.encoder(x)
        decoded = self.decode(encoded)
        return {"imgs": decoded}

### Test
hidden_dims = [128, 64, 36, 18, 2]
input_dim = 256
test_tensor = torch.randn([1, input_dim]).to(device)

ae_test = AE(input_dim, hidden_dims).to(device)

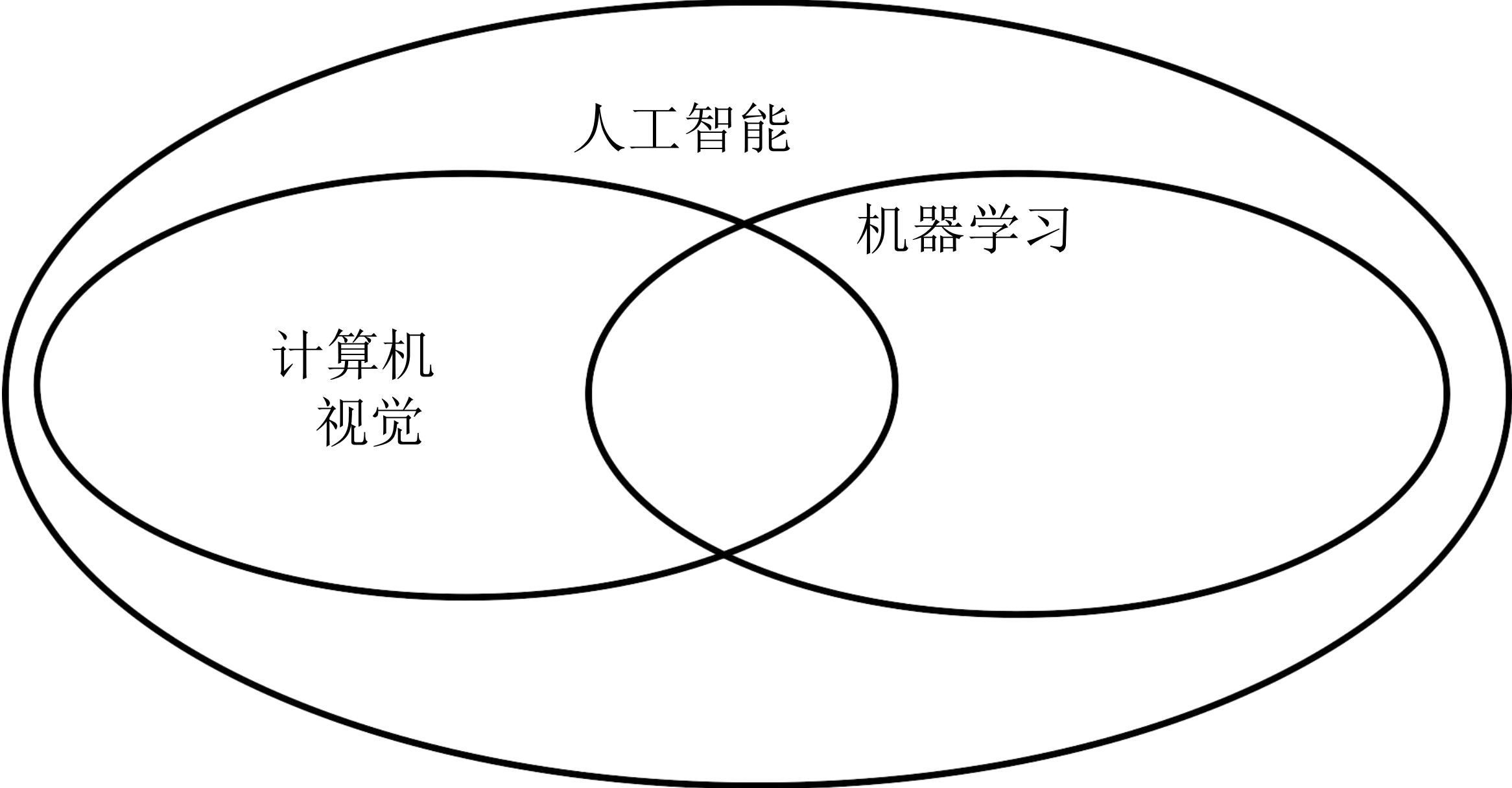
with torch.no_grad():
    test_out = ae_test(test_tensor)
```

与其他课程的联系

大家之前上过什么课？



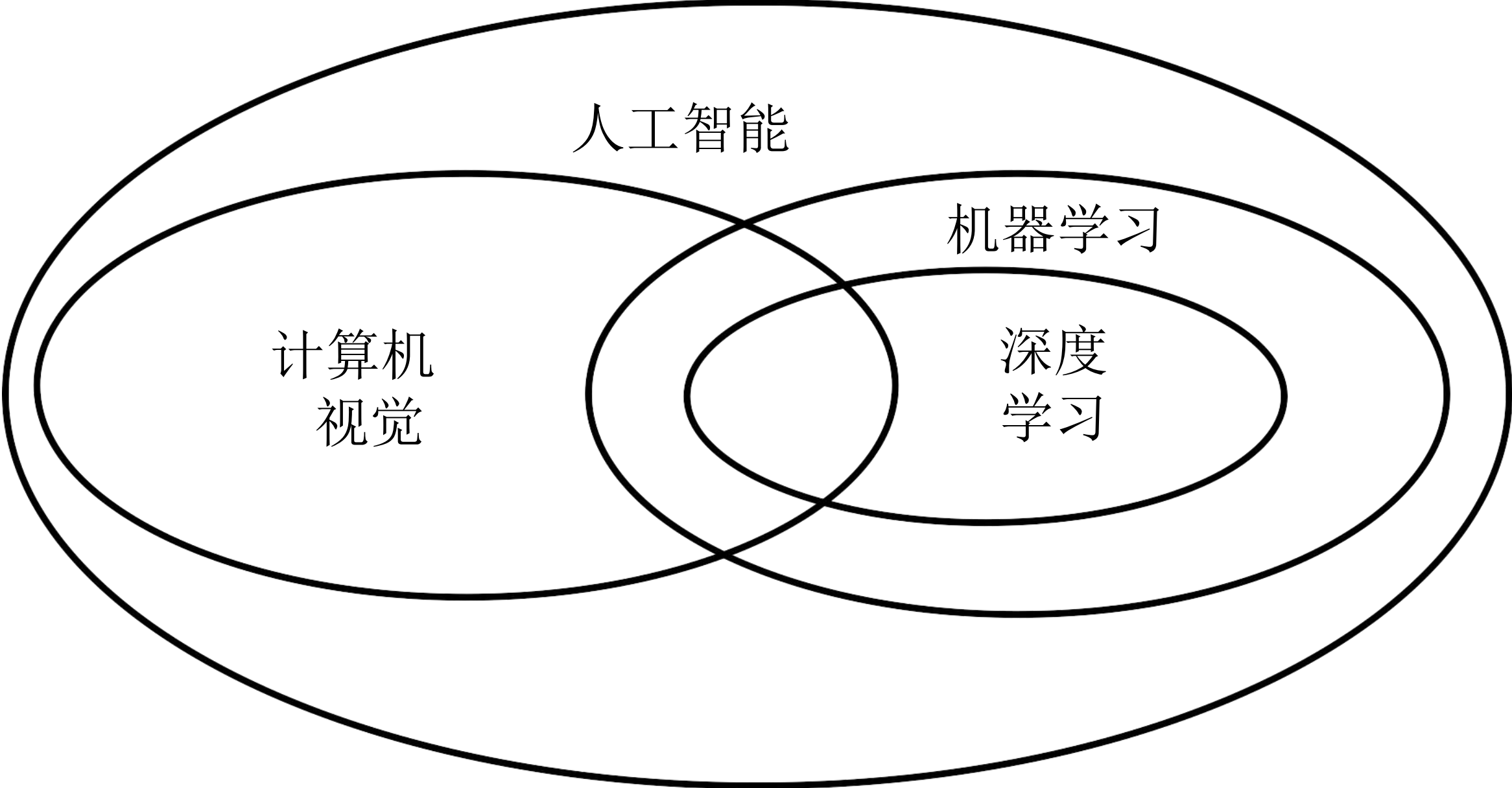
人工智能



人工智能

机器学习

计算机
视觉



人工智能

机器学习

计算机
视觉

深度
学习

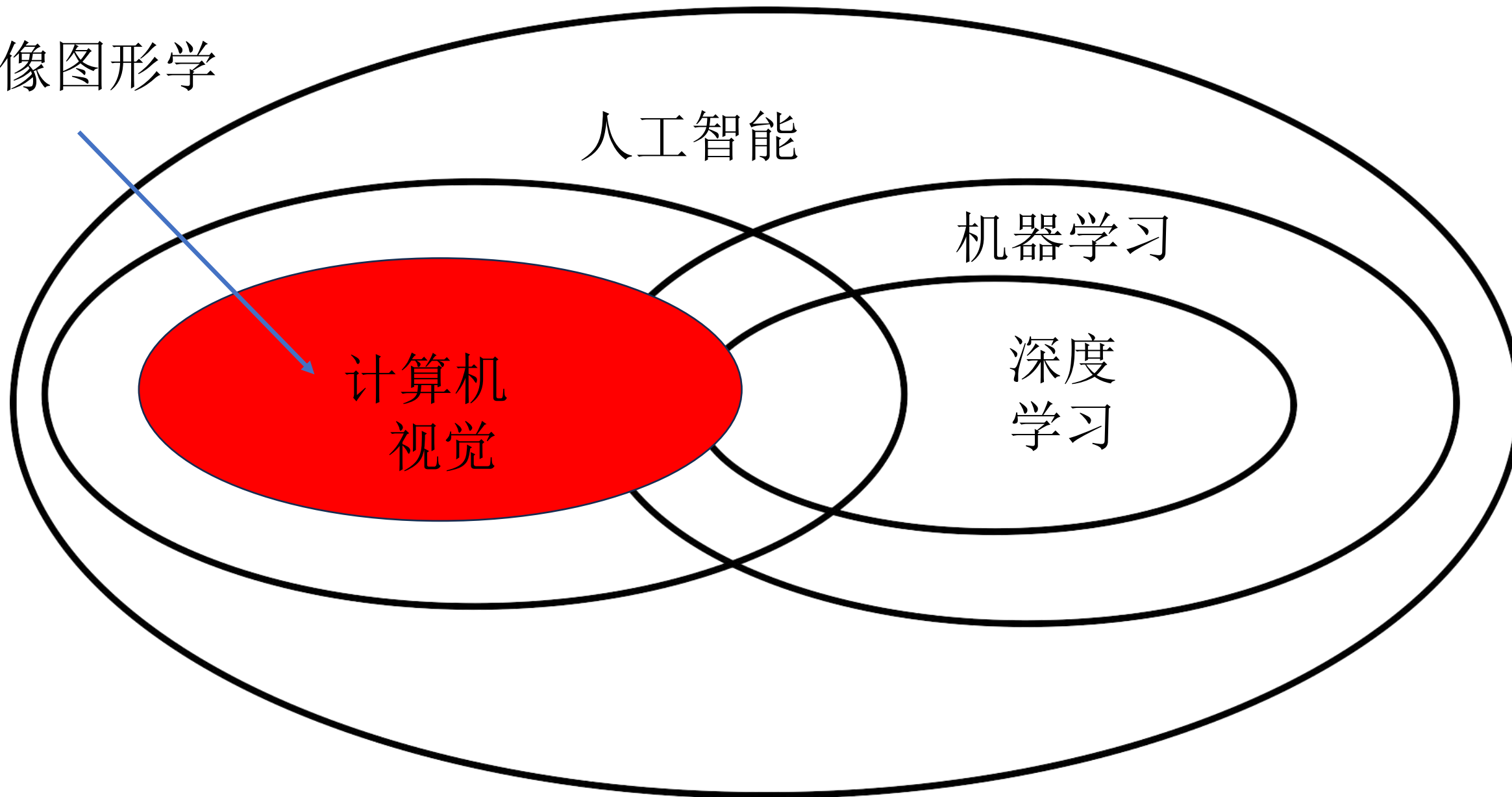
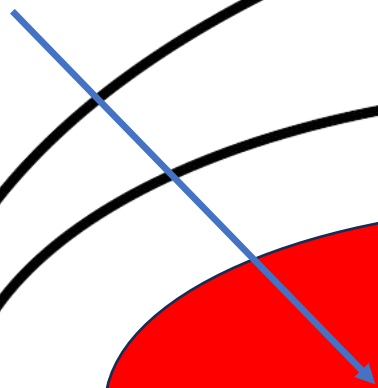
图像图形学

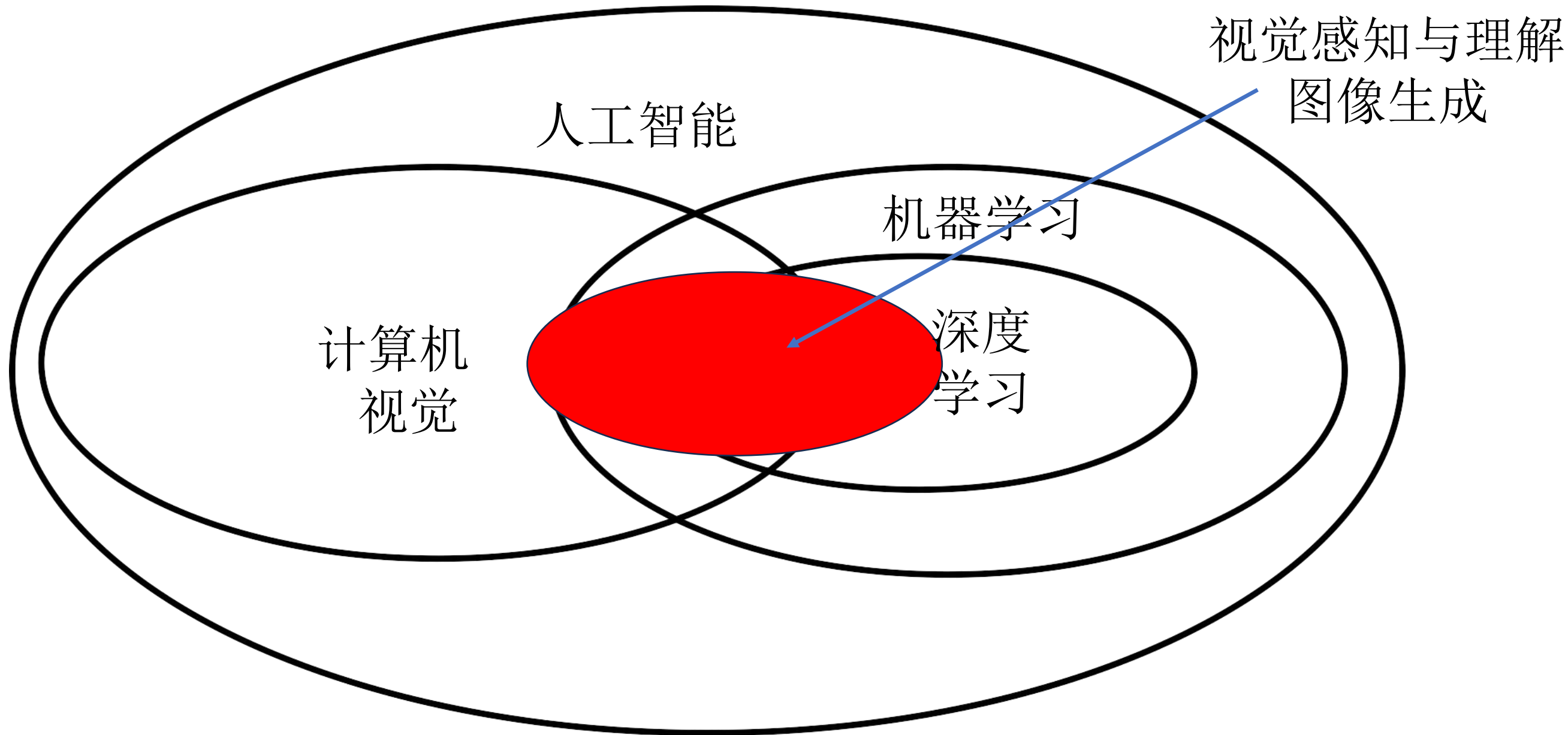
人工智能

机器学习

深度
学习

计算机
视觉





课程内容简介

什么是计算机视觉？

大家来上这门课的目的是什么？

计算机视觉： 建立系统对视觉数据进行处理、感知、推理

- 图像图形学
 - 我们怎么获得视觉信息
 - 无监督学习，更关注数学建模
- 视觉感知与理解
 - 我们能从视觉信息中获得什么
 - 监督学习，更关注数据驱动
- 视觉内容生成
 - 我们怎么创造新的视觉信息
 - 自监督学习，模型与数据兼备

图像图形学

- 我们**怎么获得视觉信息**
- **无监督学习**，更关注数学建模
- 照相机模型与成像原理
- 图像处理基础
- 二维几何
- 三维几何
- **次要目的**：提升学生的拍照、修图水平
- **首要目的**：为感知理解提供基础

图像图形学的最终目标

- Forensics: 取证利用（少量）视觉信息构造世界模型



Source: Nayar and Nishino, “Eyes for Relighting”



Source: Nayar and Nishino, "Eyes for Relighting"



Source: Nayar and Nishino, "Eyes for Relighting"

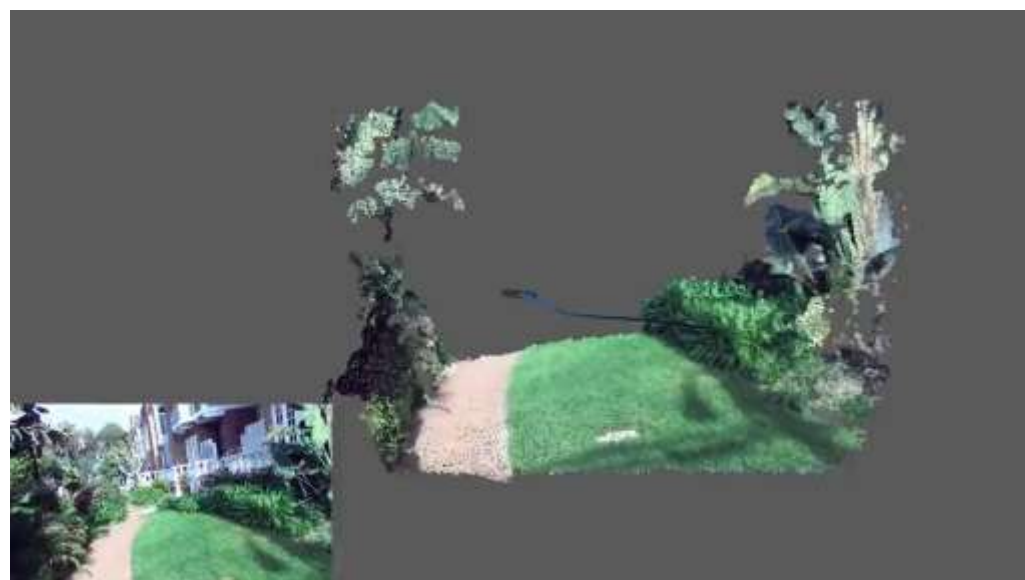
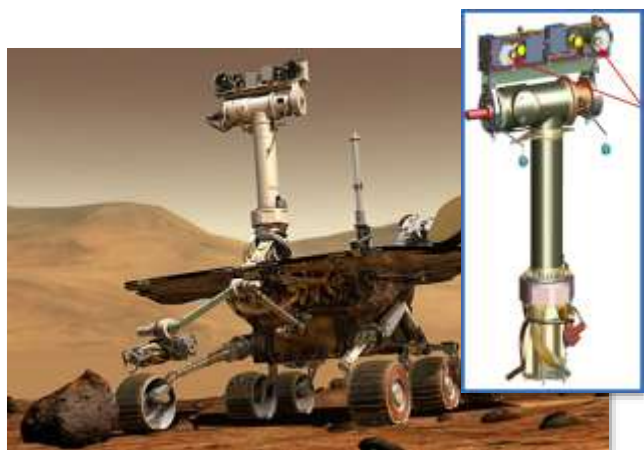
图像图形学的最终目标



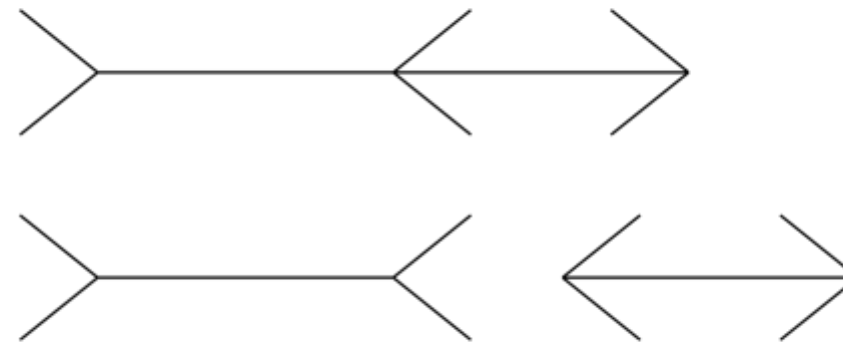
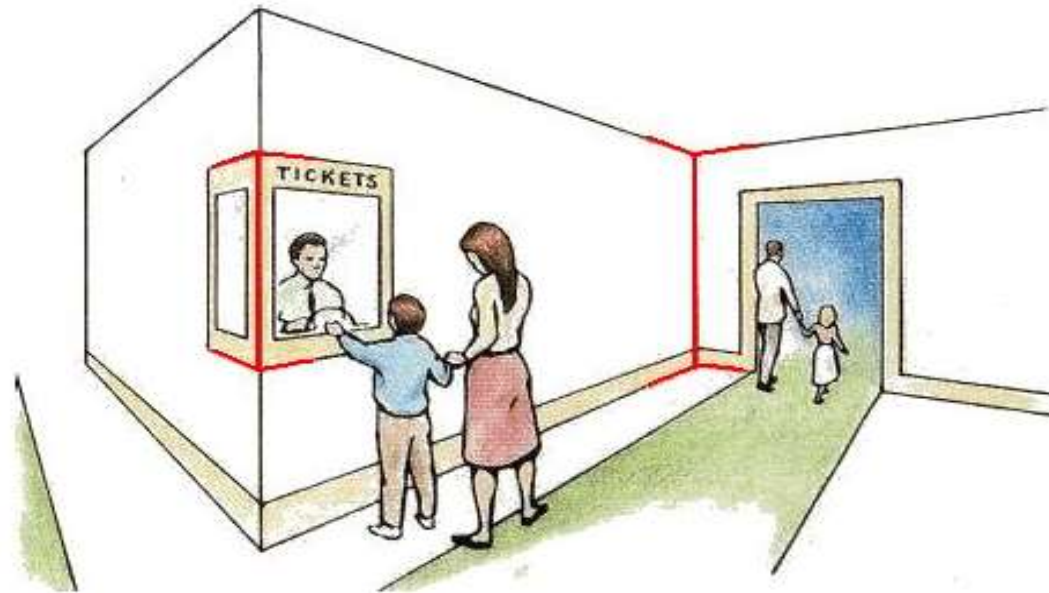
<https://www.bilibili.com/bangumi/play/ep28950?t=717>

《攻壳机动队》

怎样利用视觉信息对世界建模？



What You See (with Your Eyes) is What You Get?



怎样避免视觉错觉/误差?

提升拍照水平!



广角（短焦）



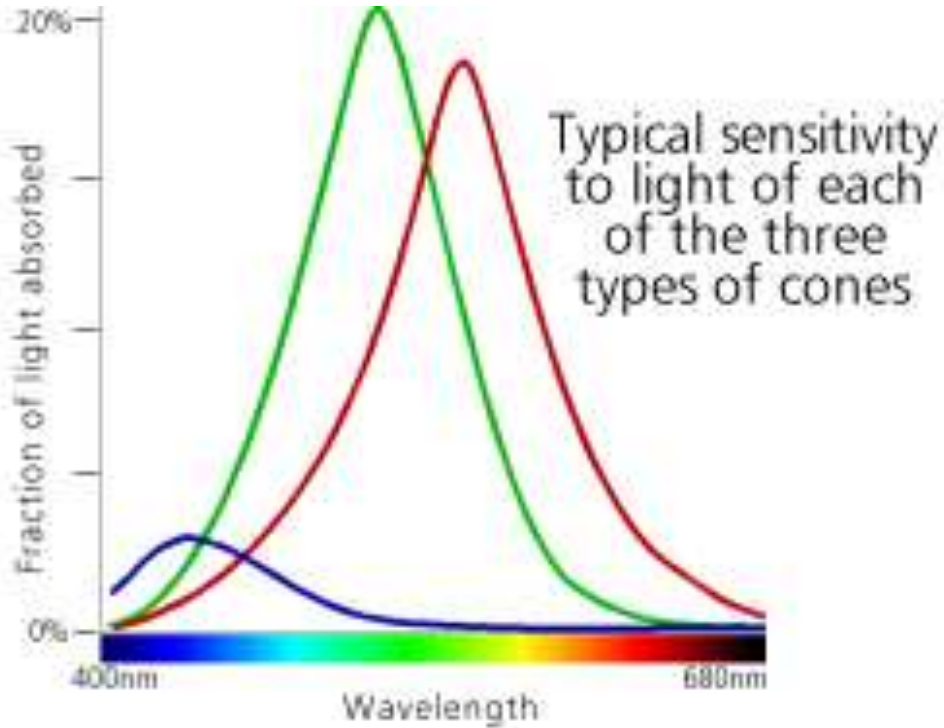
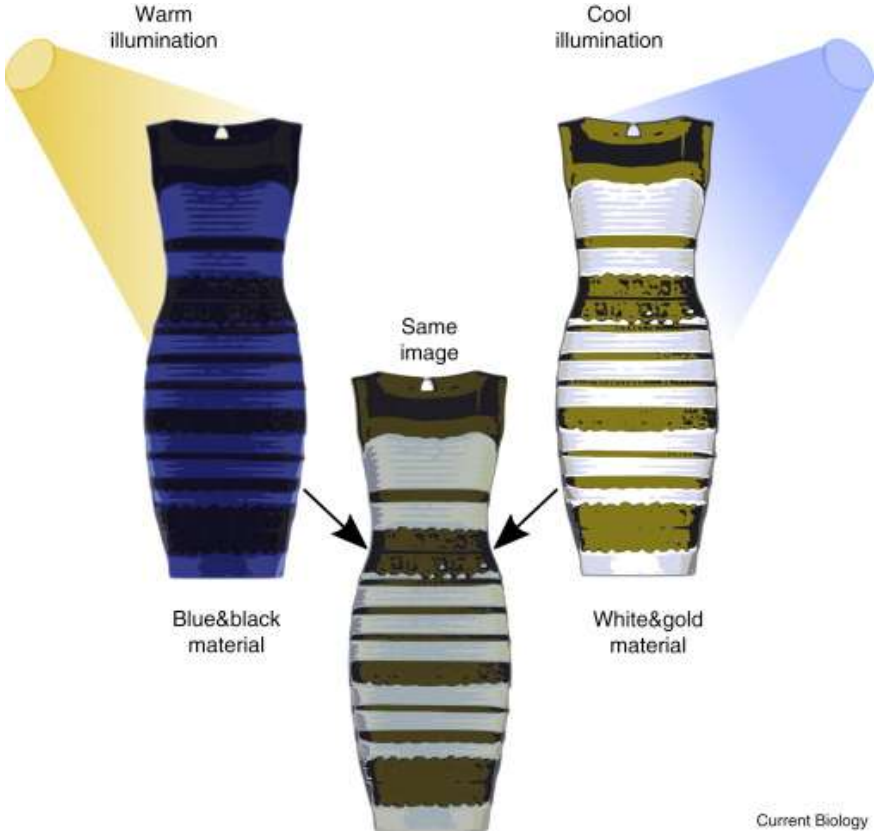
标准镜头



长焦镜头



提升修图水平！



视觉感知与理解

- 我们能**从视觉信息中获得什么**
- **监督学习**，更关注数据驱动
- 特征提取与匹配
- 目标检测与识别
- 图像分割与语义分析
- 运动分析与跟踪
- **次要目的**：提升学生的眼力、侦察和推断能力
- **首要目的**：机器视觉感知、理解，甚至推理、决策



“感知”

从 几何模型：基于几何原理和相对位置关系的模型
到 统计模型：基于数据分布和统计推断的模型

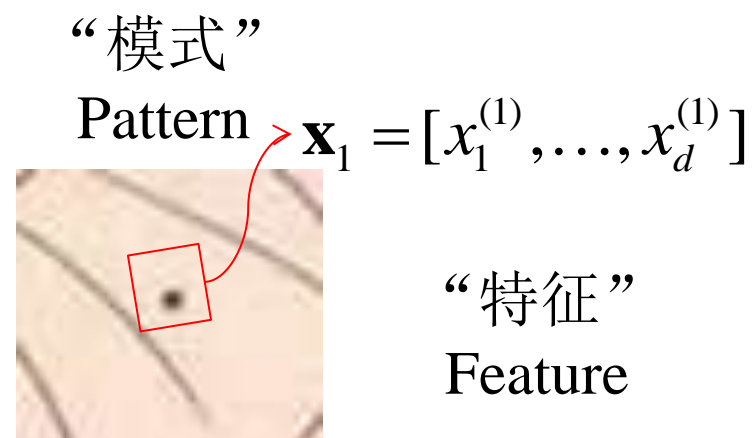
计算机如何感知？

- 让我们以人为例...
 - “音容笑貌”
 - “你掌心的痣，我总记得在哪里”



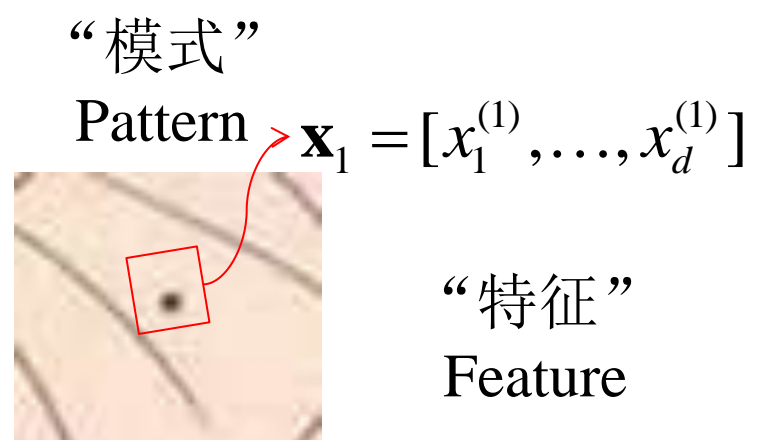
计算机如何感知？

- 让我们以人为例...
 - “音容笑貌”
 - “你掌心的痣，我总记得在哪里”

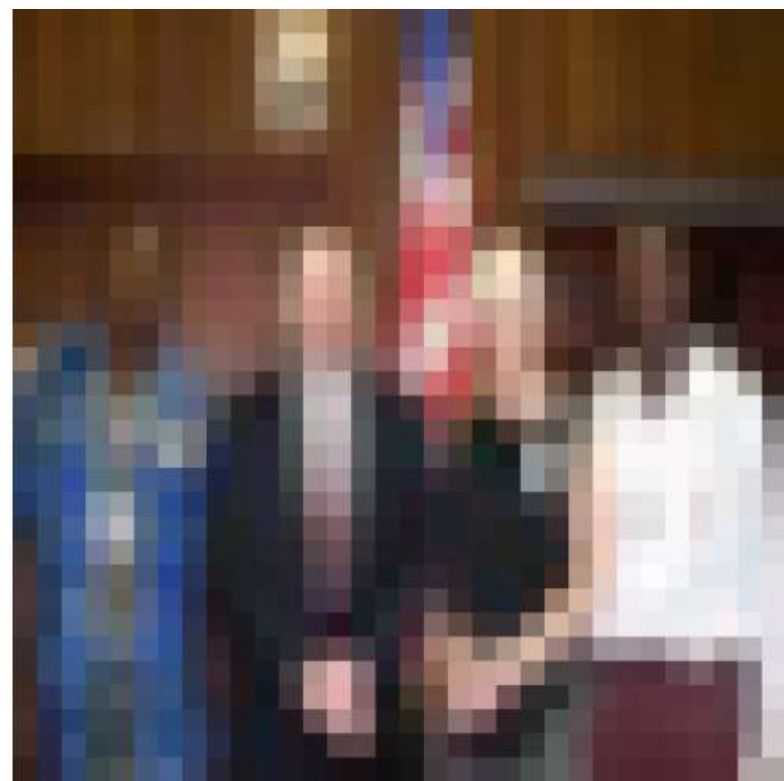


计算机如何感知？

- 让我们以人为例...
 - “音容笑貌”
 - “你掌心的痣，我总记得在哪里”



人类的优势：
经验推理
减少冗余



Source: "80 million tiny images" by Torralba, et al.



计算机的优势：
惊人的速度、特征
大量冗余

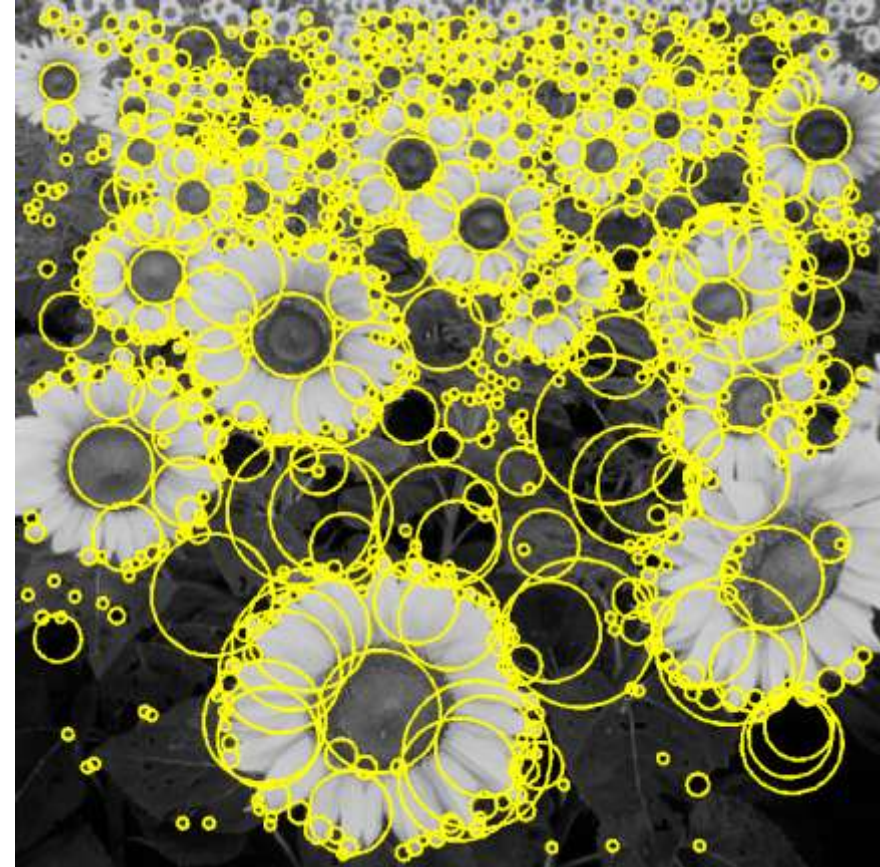
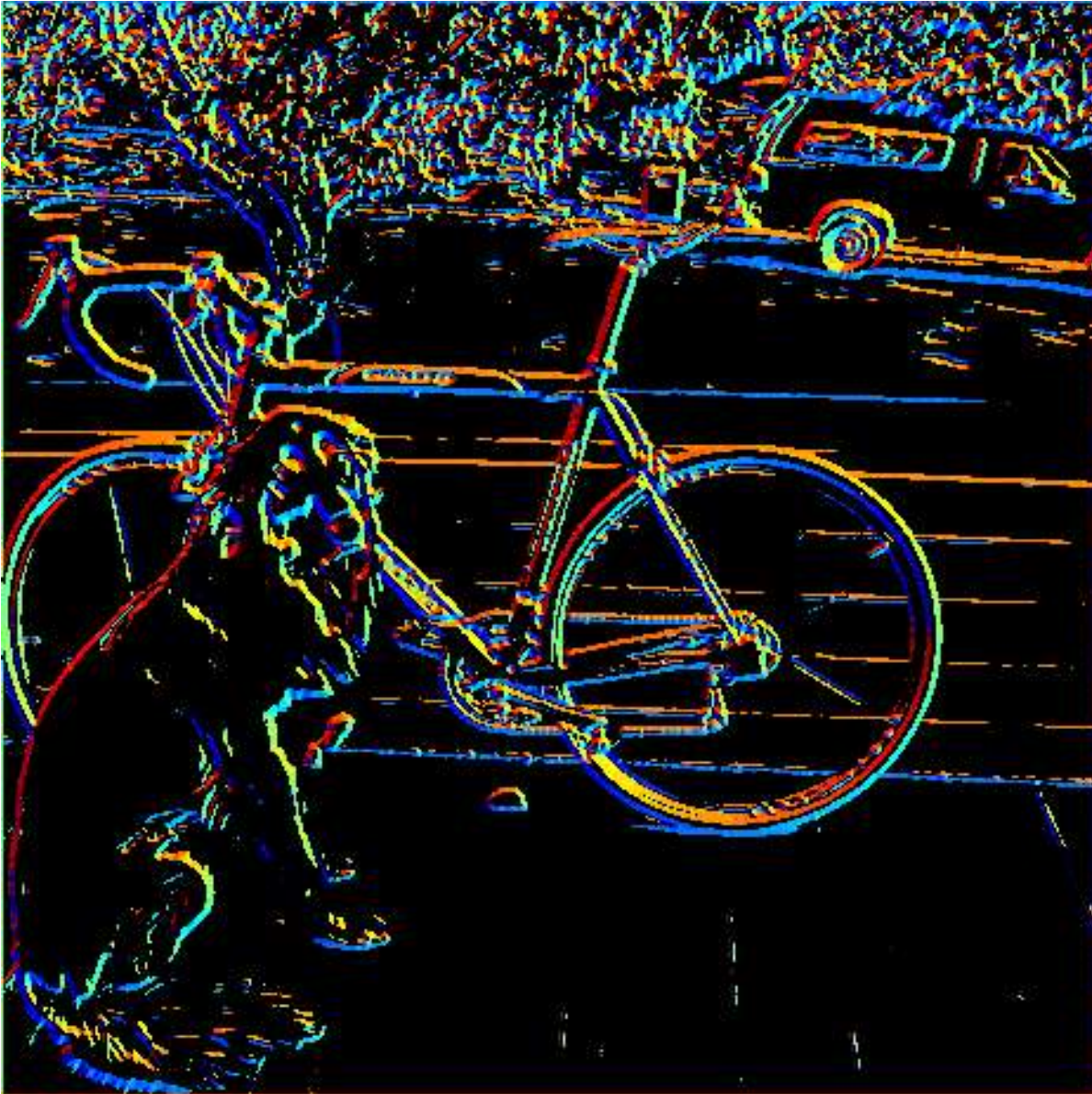


一个房子也许认不出来...

一百个房子认出来了!

怎样提取特征？

通用（传统/深度）的 视觉处理技巧



“感知”

从 几何模型：基于几何原理和相对位置关系的模型

到 统计模型：基于数据分布和统计推断的模型

我们需要统计模型更快速、普适性地进行感知推断

—— 深度模型

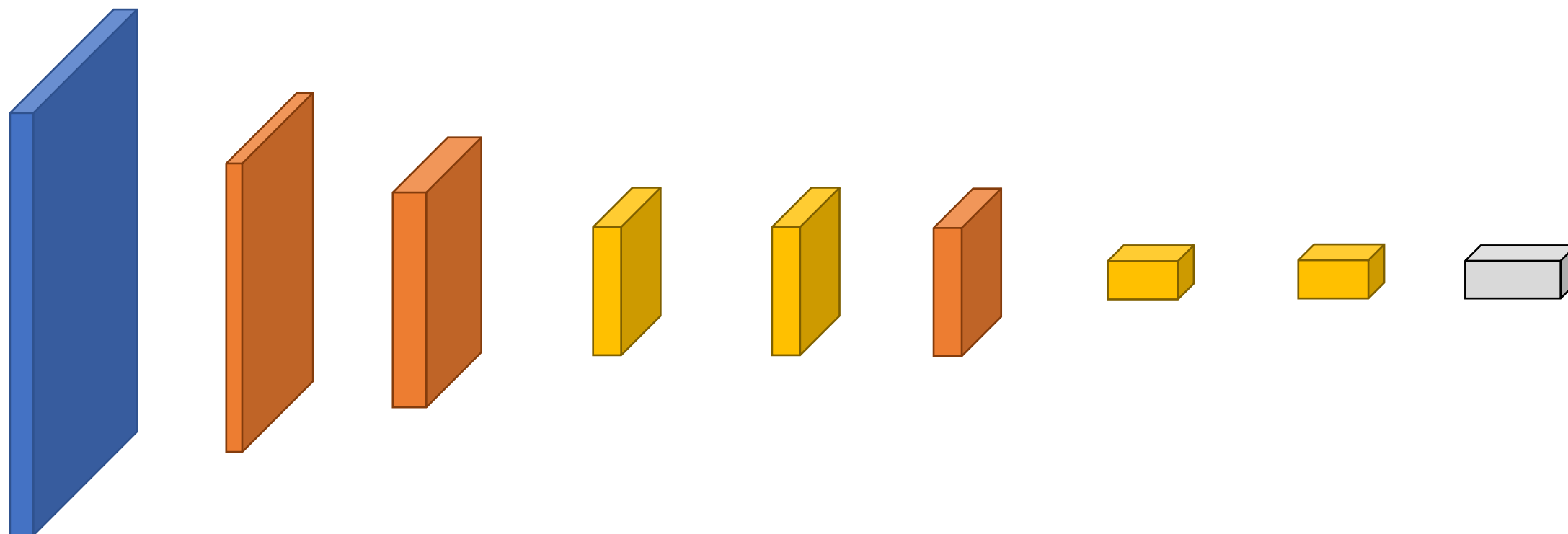
深度模型

更快速的推理判断

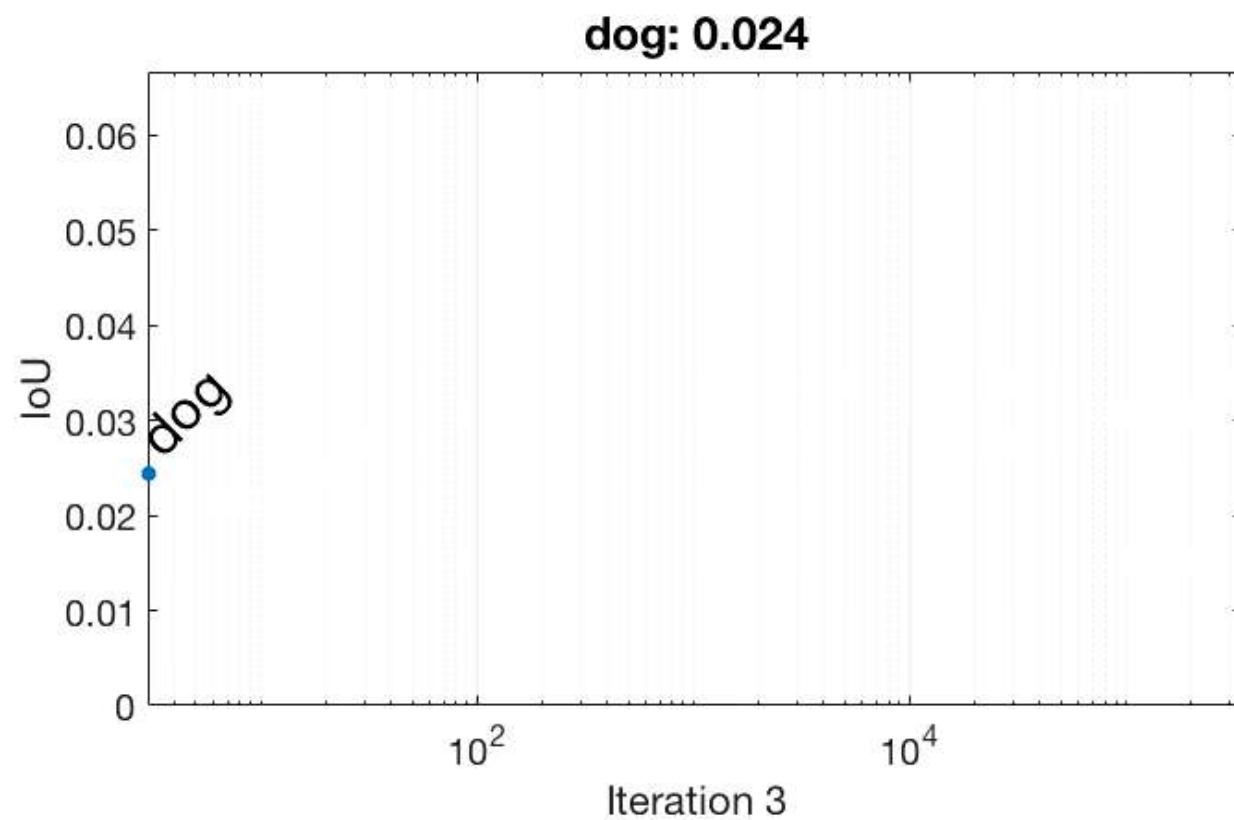
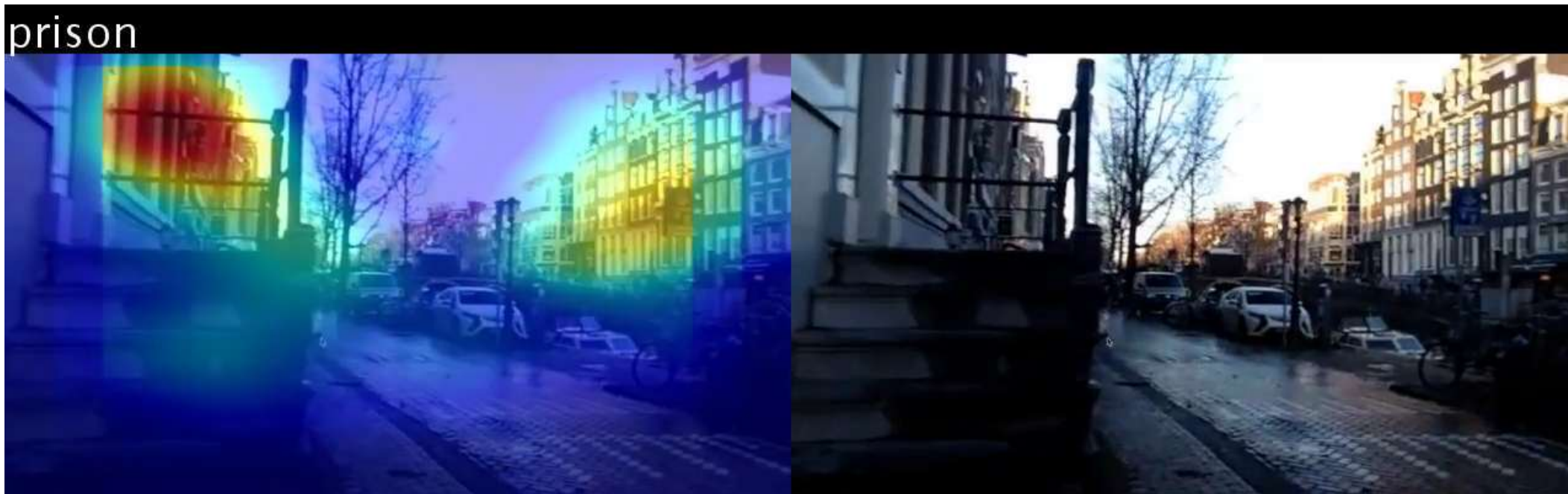
更强的普适性（泛化性）

为什么这样设计视觉深度模型？

Input	Conv 1	Conv 2	Conv 3	Conv 4	Conv 5	FC 6	FC 7	Output
227x227	55x55	27x27	13x13	13x13	13x13	1x1	1x1	1x1
3	96	256	384	384	256	4096	4096	1000

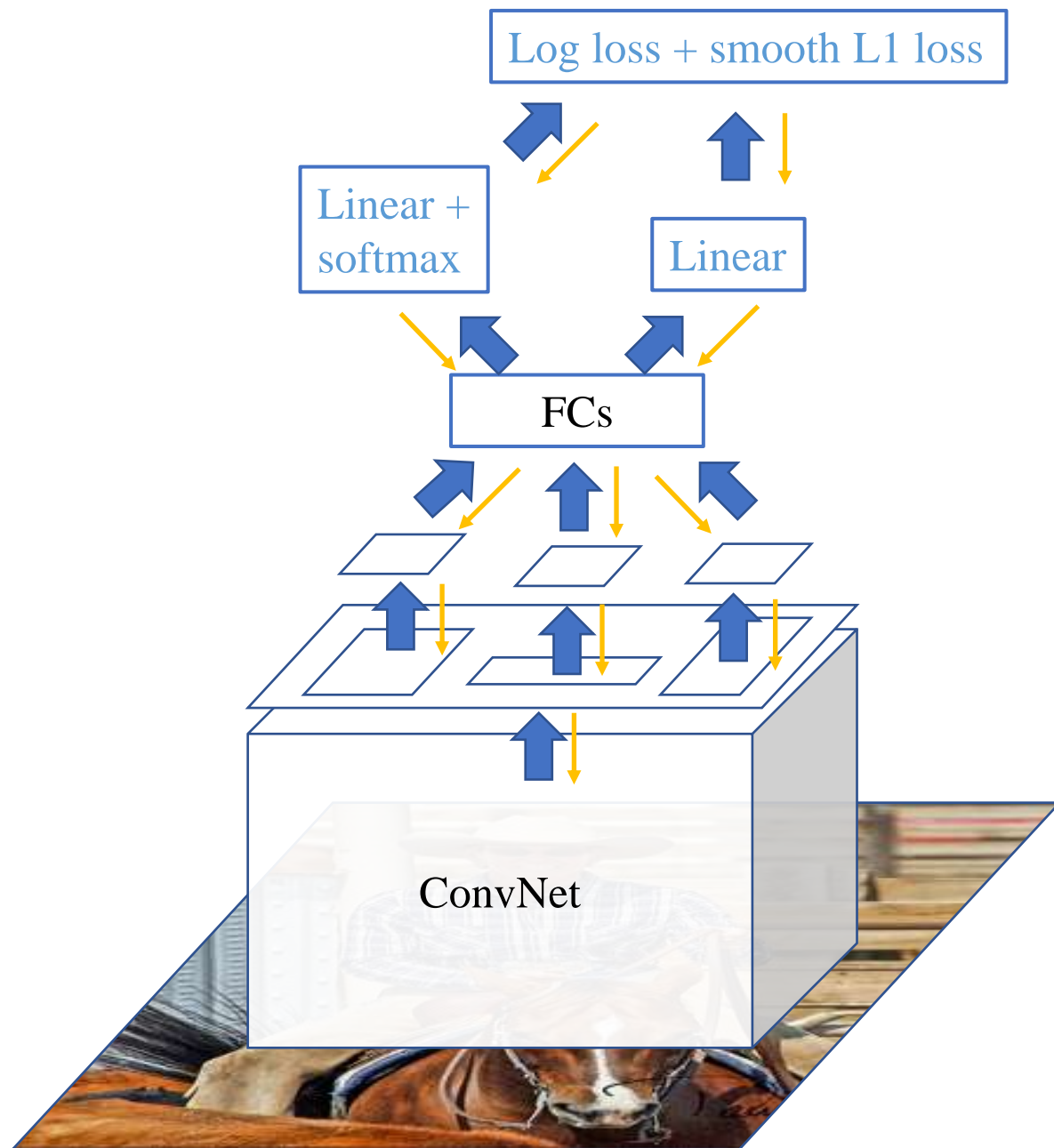


怎样理解/调试视觉深度模型

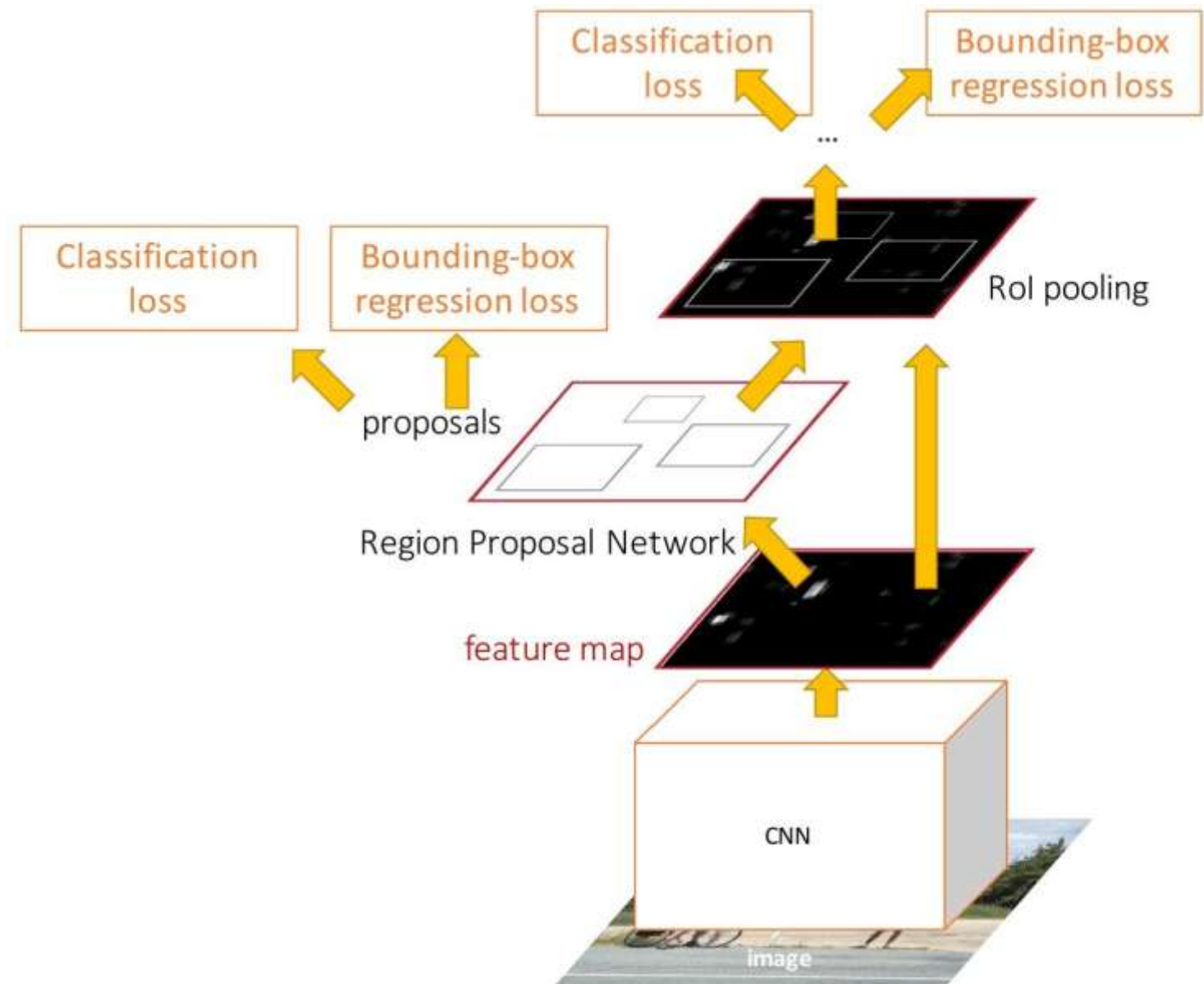


怎样利用深度模型完成视觉任务

Fast R-CNN



Faster R-CNN



与大语言模型结合 (MLM)

J You



我把局部放大了，现在请你再来看看这是什么？

ChatGPT

放大之后的图片显示，红框内的物体是一只远程控制器。可以辨认出它的典型长条形状和可能是按钮的细节，这些特征与常见的电视或音响系统的遥控器相符合。

J You



图像红框里是什么物体？

ChatGPT

红框里的物体是一只猫。它似乎是在沙发边的地板上蜷缩着，与环境的其他色彩相比，它的颜色较为淡，可能导致它不那么显眼。猫的姿态表明它可能在休息或打盹。



视觉内容生成

- 我们**怎么创造新的视觉信息**
- **自监督学习**，模型与数据兼备
- 自编码器
- 变分自编码器
- 对抗生成网络
- 生成内容应用
- **目的**：实现真实、可控的自动化视觉内容生成模型



图像生成/处理/编辑

- 增强图像的信息含量 (“Computational Photography”)



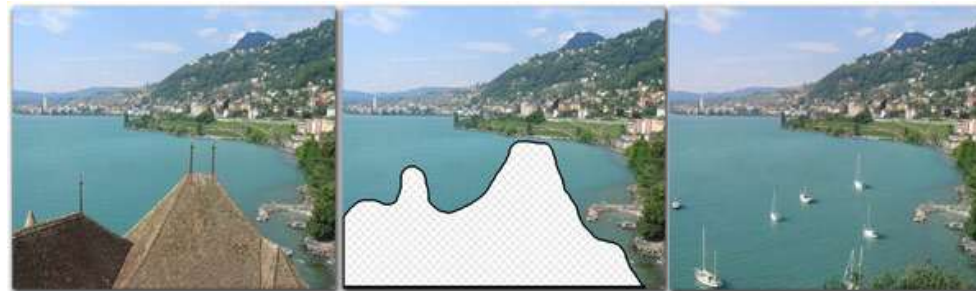
超分辨率 **放大，放大，再放大!**
(source: 2d3)



亮度调整 **照亮你的美!**
(credit: [Hasinoff et al., SIGGRAPH ASIA 2016](#))



背景虚化 **柔光双摄!** (source: [Google Research Blog](#))



图像补全 **一键修图!**
(image credit: Hays and Efros)

图像生成/处理/编辑

- 具备目标的可控性(“Personalized Controllability”)



目标图像

场景文本 : A dog in the rain



缺乏个性化:
低图像保真度



缺乏多样性:
低场景保真度



理想效果:
高质量定制化



Person



Outfit Anyone + Animate Anyone

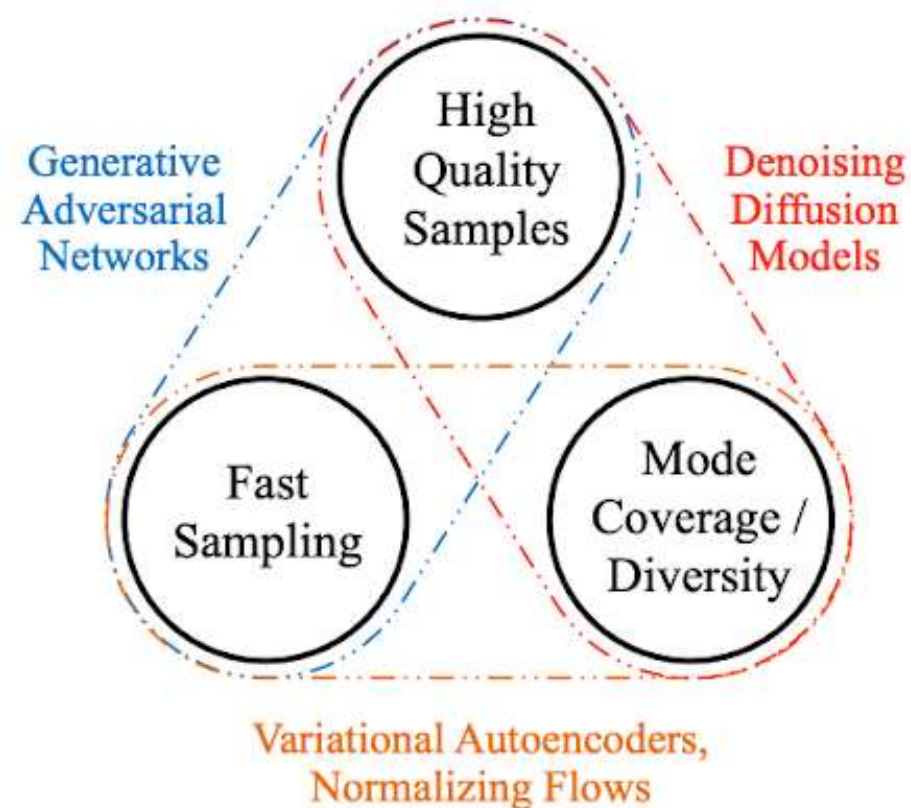
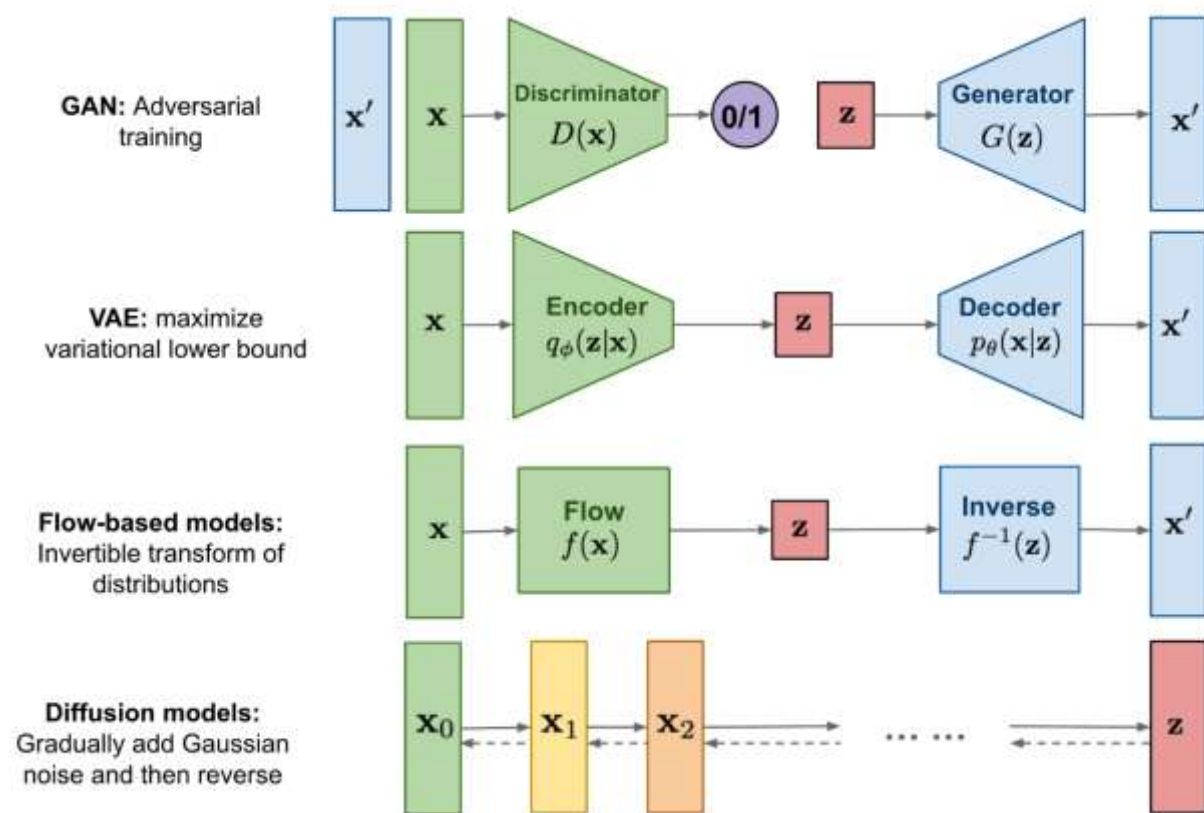
Garment

Try-on

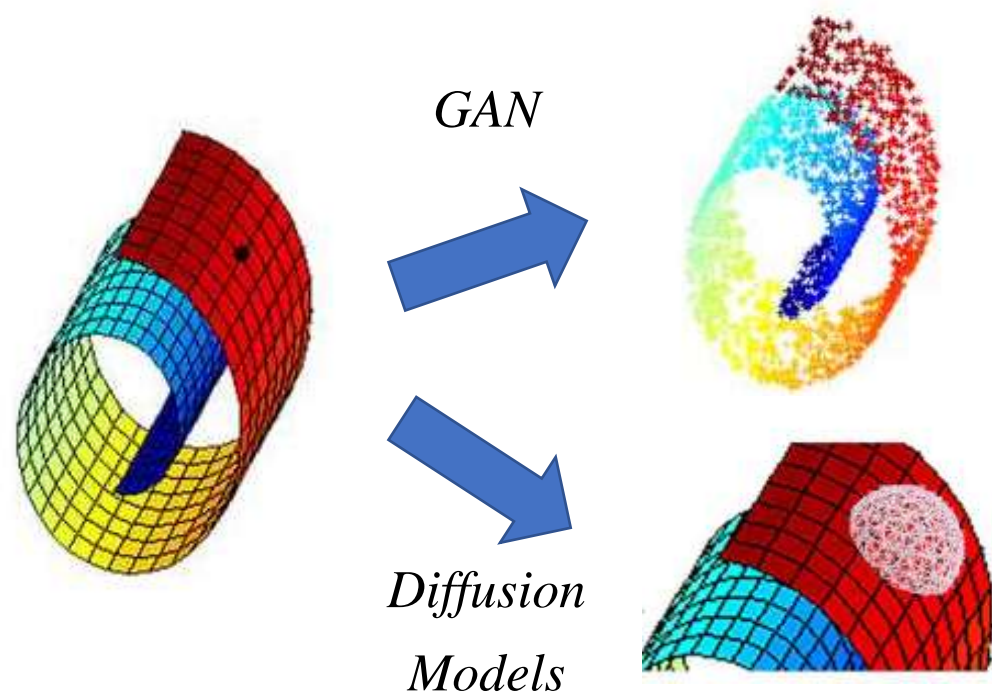
Animation

层出不穷的生成模型

- 这些模型的基本原理是什么？优缺点是什么？



从GAN到Diffusion Models



扩散模型拥有更难以想象的生成能力！



Model Source: FLUX

<https://openai.com/sora>

扩散模型是否无所不能？

- 为什么会产生这些问题？



计算机视觉的发展史

五亿年前，寒武纪...



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



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



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

原始生物并没有眼睛

感知光线的存在和方向

复杂的触觉系统



约五千万年前，脊椎动物，眼睛结构（晶状体和视网膜）



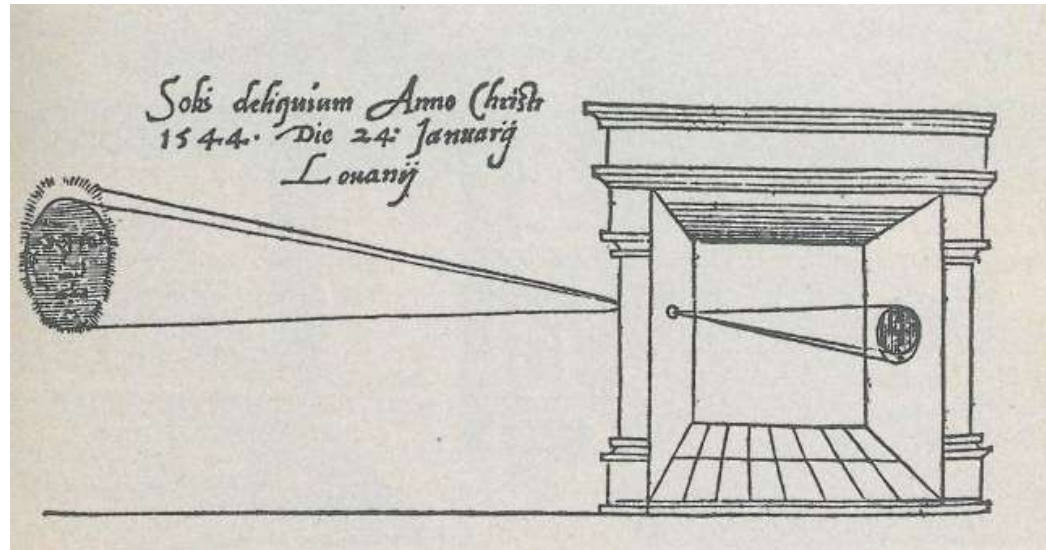
复眼



现代生物视觉系统

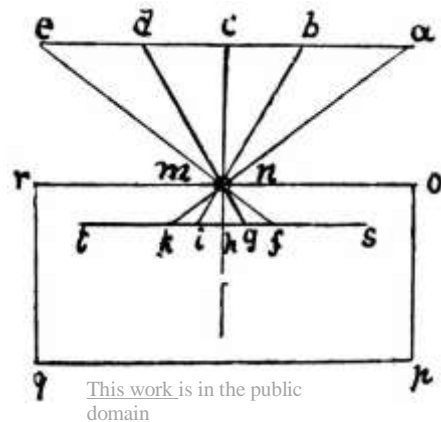
现代相机的前身：“暗箱”（Camera Obscura）

Gemma Frisius, 1545



小孔成像，
暂时没有透
镜，用于艺
术研究

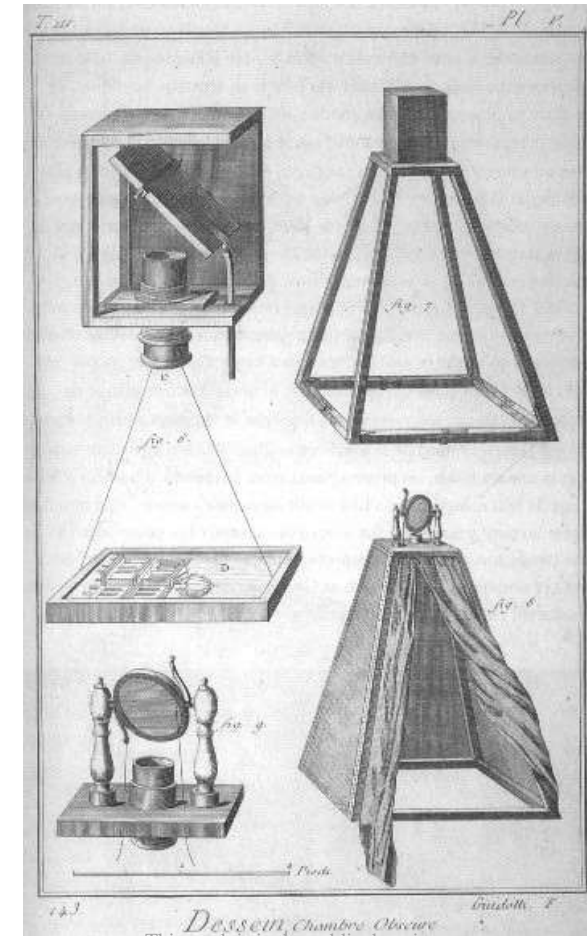
This work is in the public domain



Leonardo da Vinci,
16th Century AD

This work is in the public domain

Encyclopedia, 18th Century



This work is in the public domain

照片、照相机与胶卷

1826. 世上首张照片 《[窗外风景](#)》



HISTORY OF CAMERA

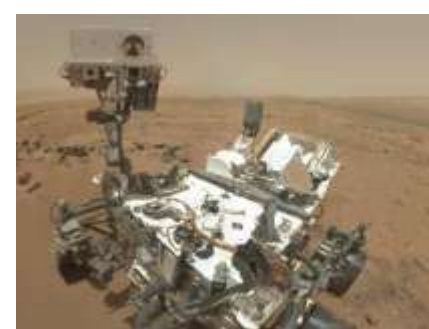


如今 计算机视觉 无处不在!

照相机



无人机



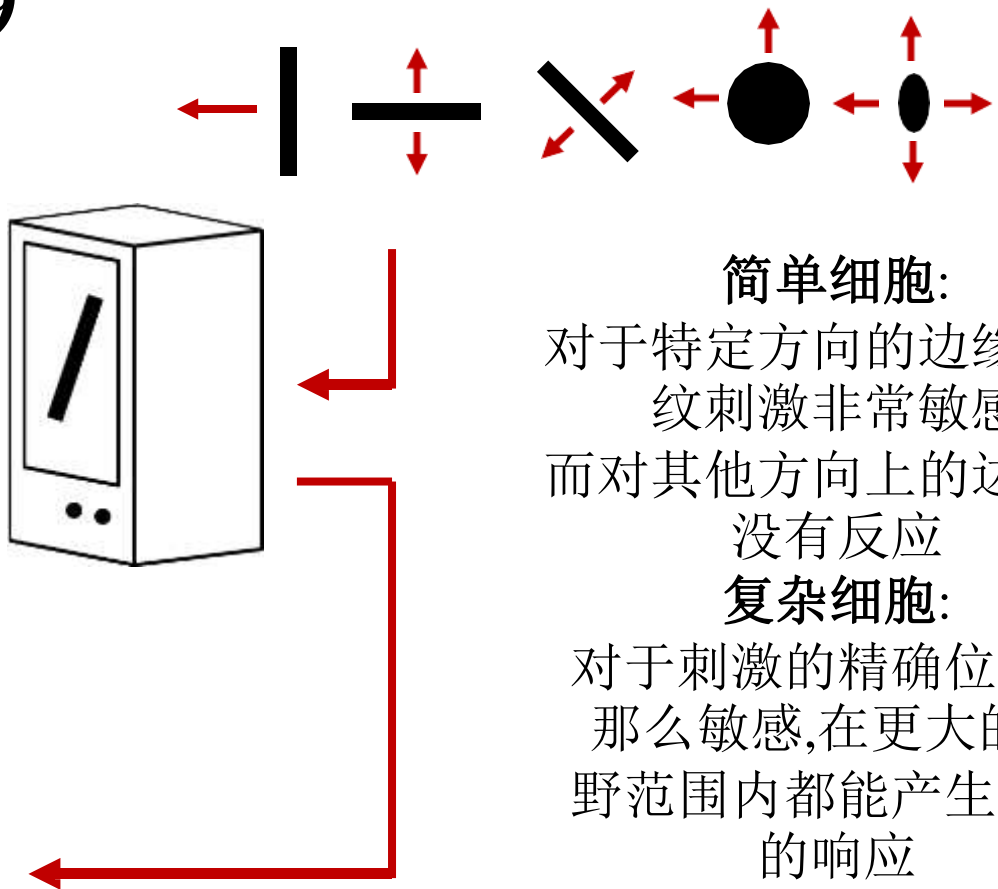
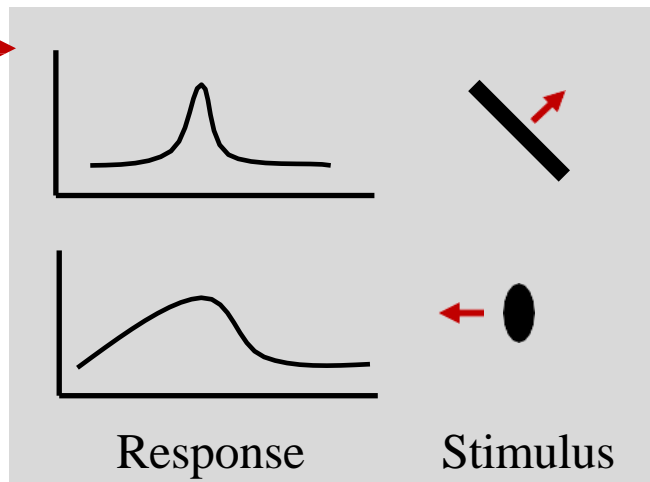
摄像头



VR/AR

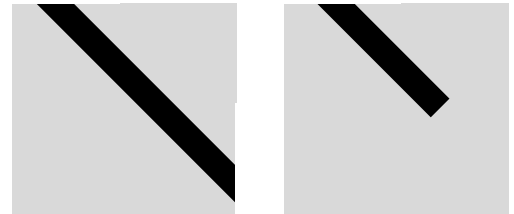
Hubel and Wiesel, 1959

大脑皮层神经元对视觉信息的处理方式



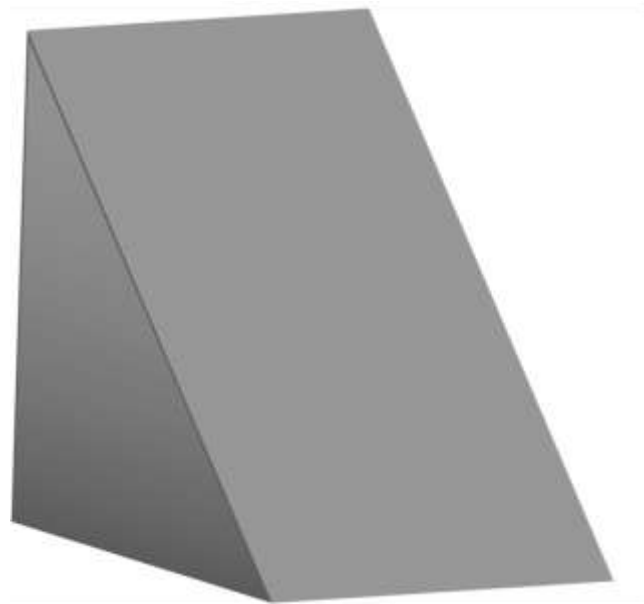
简单细胞:
对于特定方向的边缘或条纹刺激非常敏感
而对其他方向上的边缘则没有反应

复杂细胞:
对于刺激的精确位置不那么敏感,在更大的视野范围内都能产生类似的响应

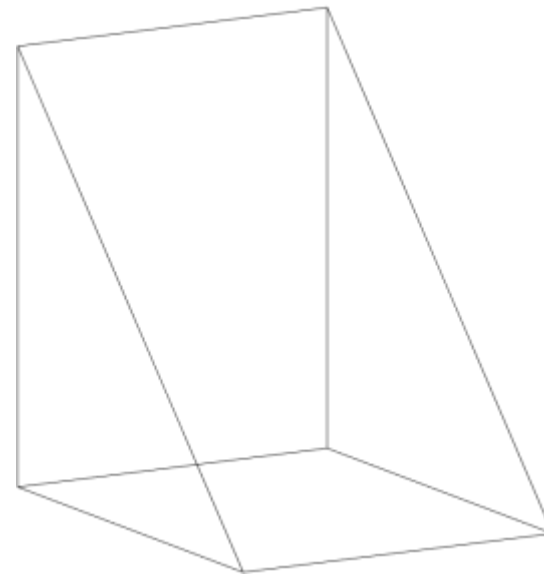


1959
Hubel & Wiesel

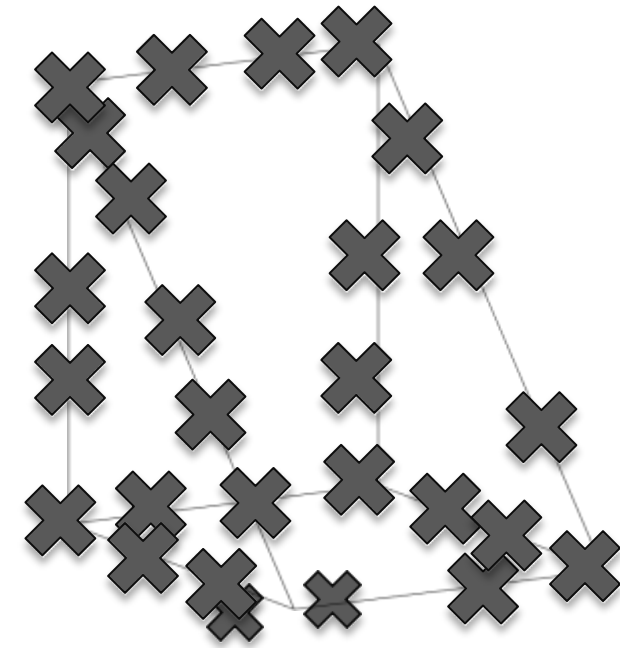
Larry Roberts, 1963



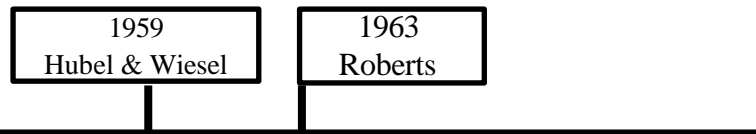
(a) Original picture



(b) Differentiated picture



(c) Feature points selected



从二维图像提取形状和三维信息

Lawrence Gilman Roberts, "Machine Perception of Three-Dimensional Solids", 1963

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

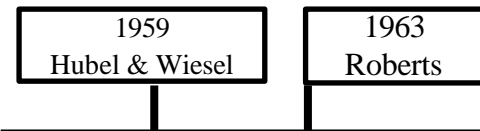
July 7, 1966

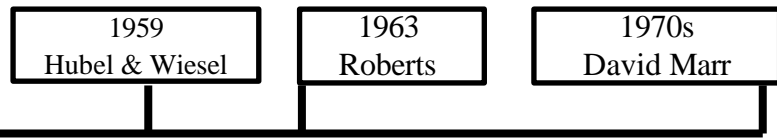
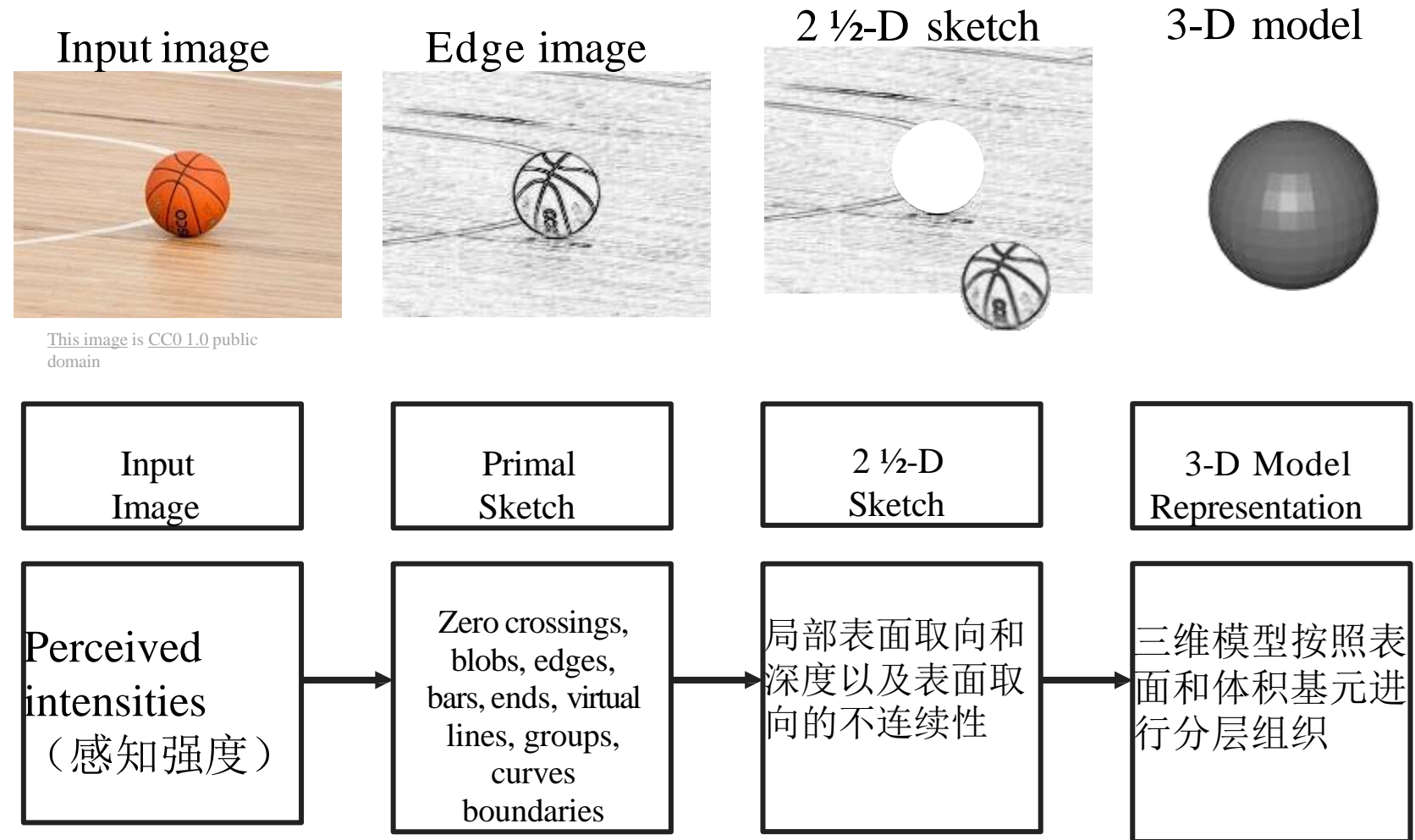
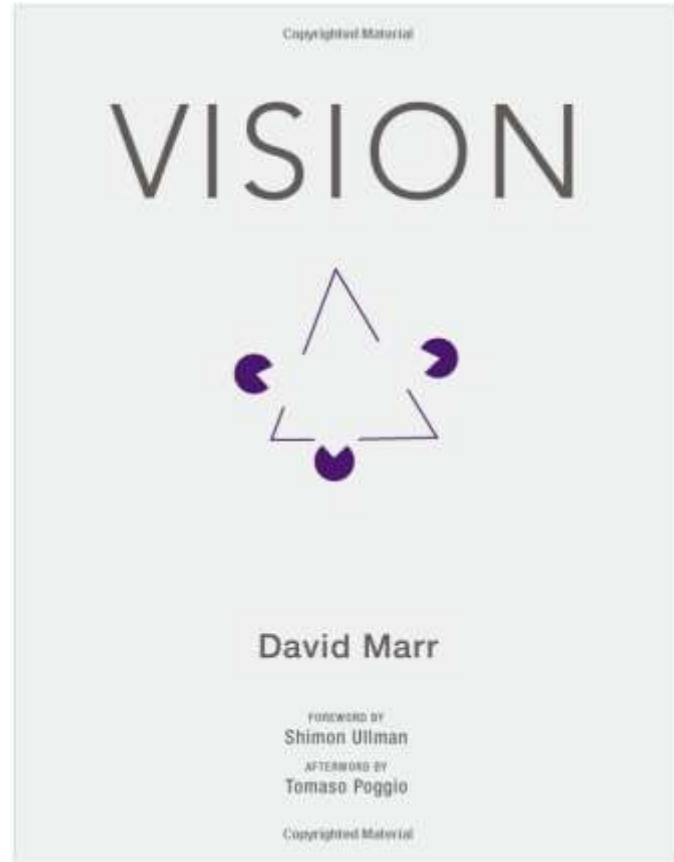
第一篇CV的博士毕业论文

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".



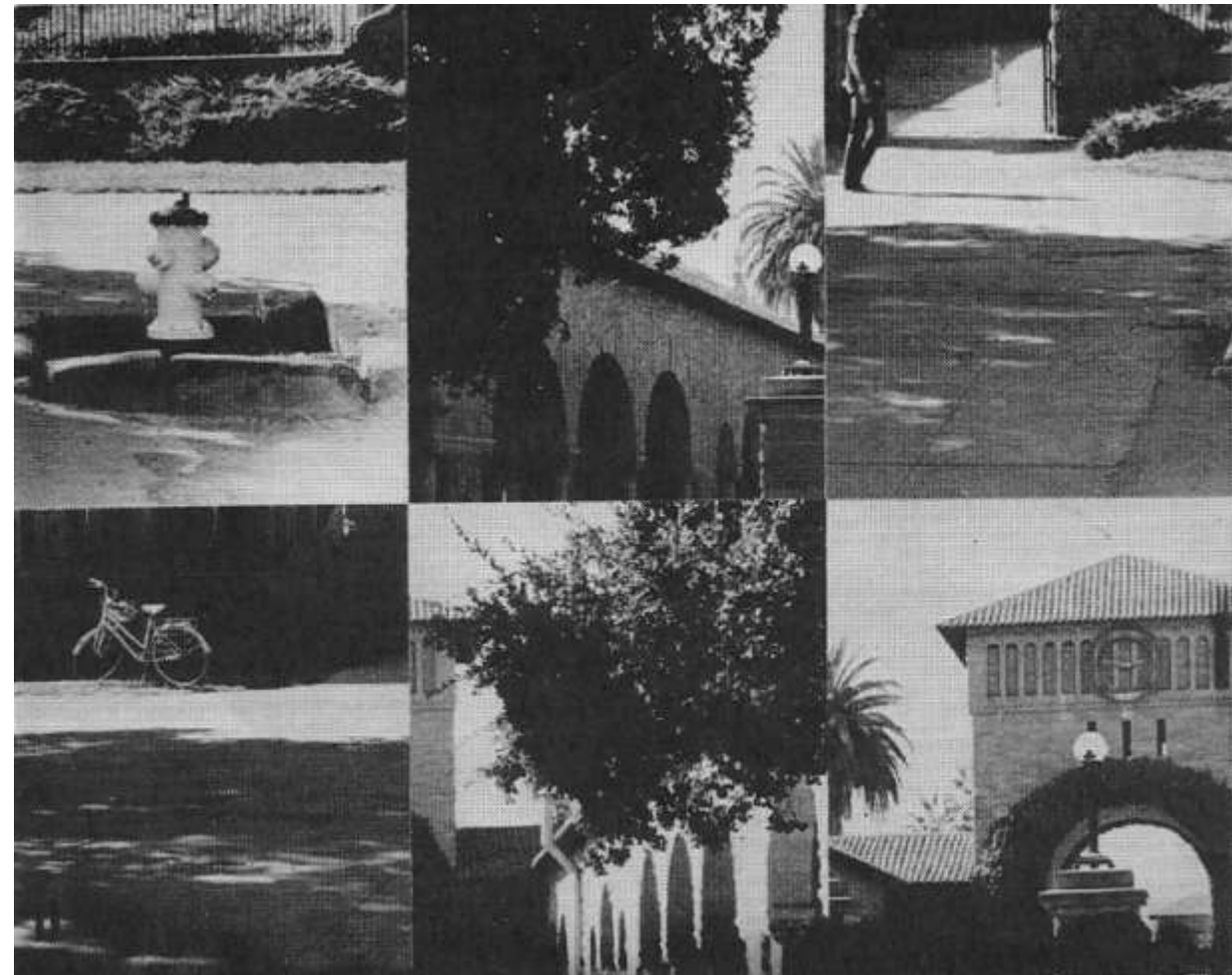


“二维半素描”：通过从不同视角获取的二维图像推断出的一些关于物体表面形状和结构的简化信息

Stages of Visual Representation, David Marr, 1970s

Perceiving Real-World Scenes

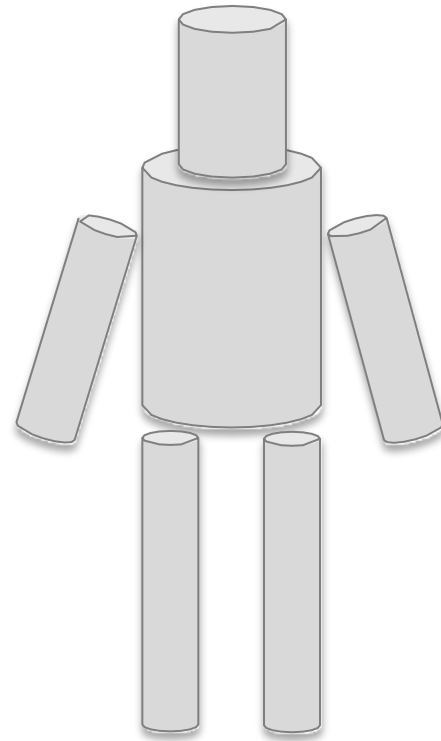
Irving Biederman



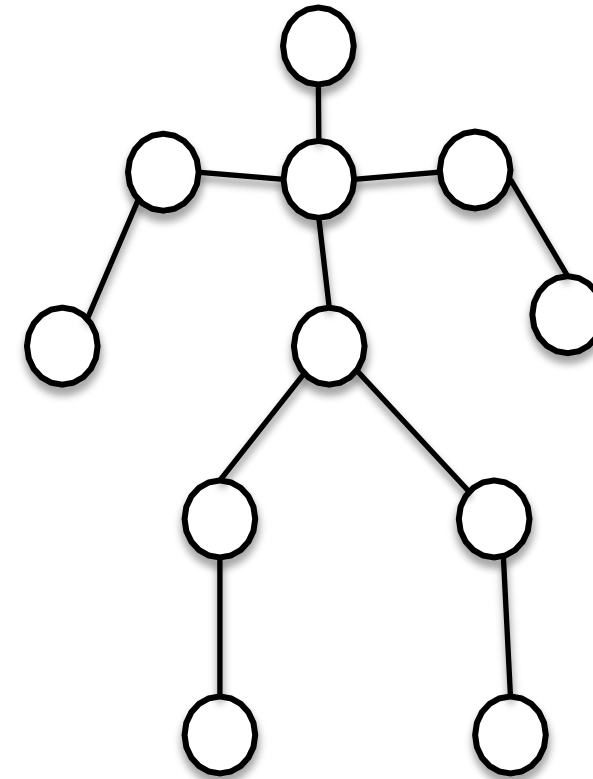
计算机系统如何模拟人类视觉系统，从图像或视频中提取并理解有关真实世界场景的信息

I. Biederman, *Science*,
1972

识别部件 (1970s)

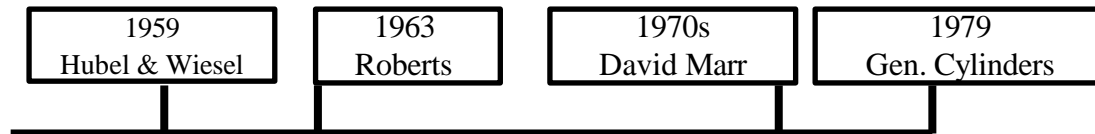


Generalized
Cylinders, Brooks and
Binford, 1979

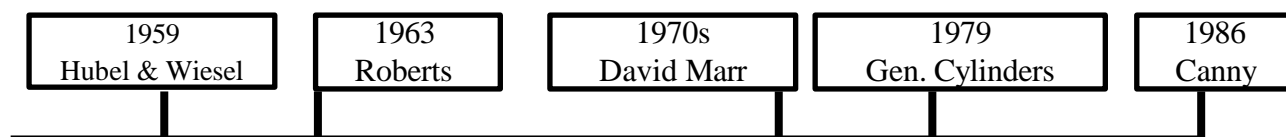


将对象分解为组
成部分，并根据
它们之间的关系
来描绘对象

Pictorial Structures,
Fischler and Elshlager,
1973



边缘检测 Edge Detection(1980s)

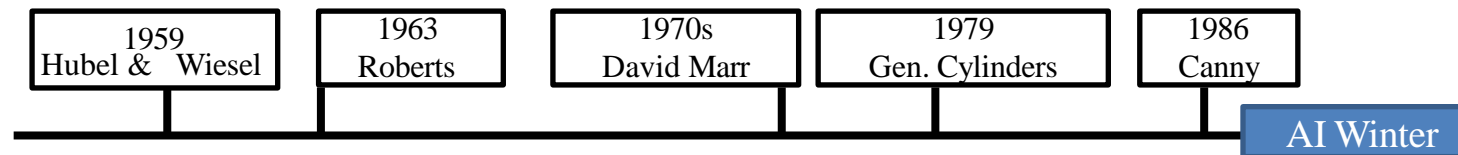


John Canny, 1986
David Lowe,
1987

人工智能的“寒冬”

- 大量学界期望无法达成
- 学界对人工智能热情逐渐减少
- 但人工智能的子领域逐渐增加
 - 计算机视觉, NLP, 机器人, 仿生, 等.

与此同时，视觉感知与理解领域逐渐被挖掘



Left Image is CC BY 3.0

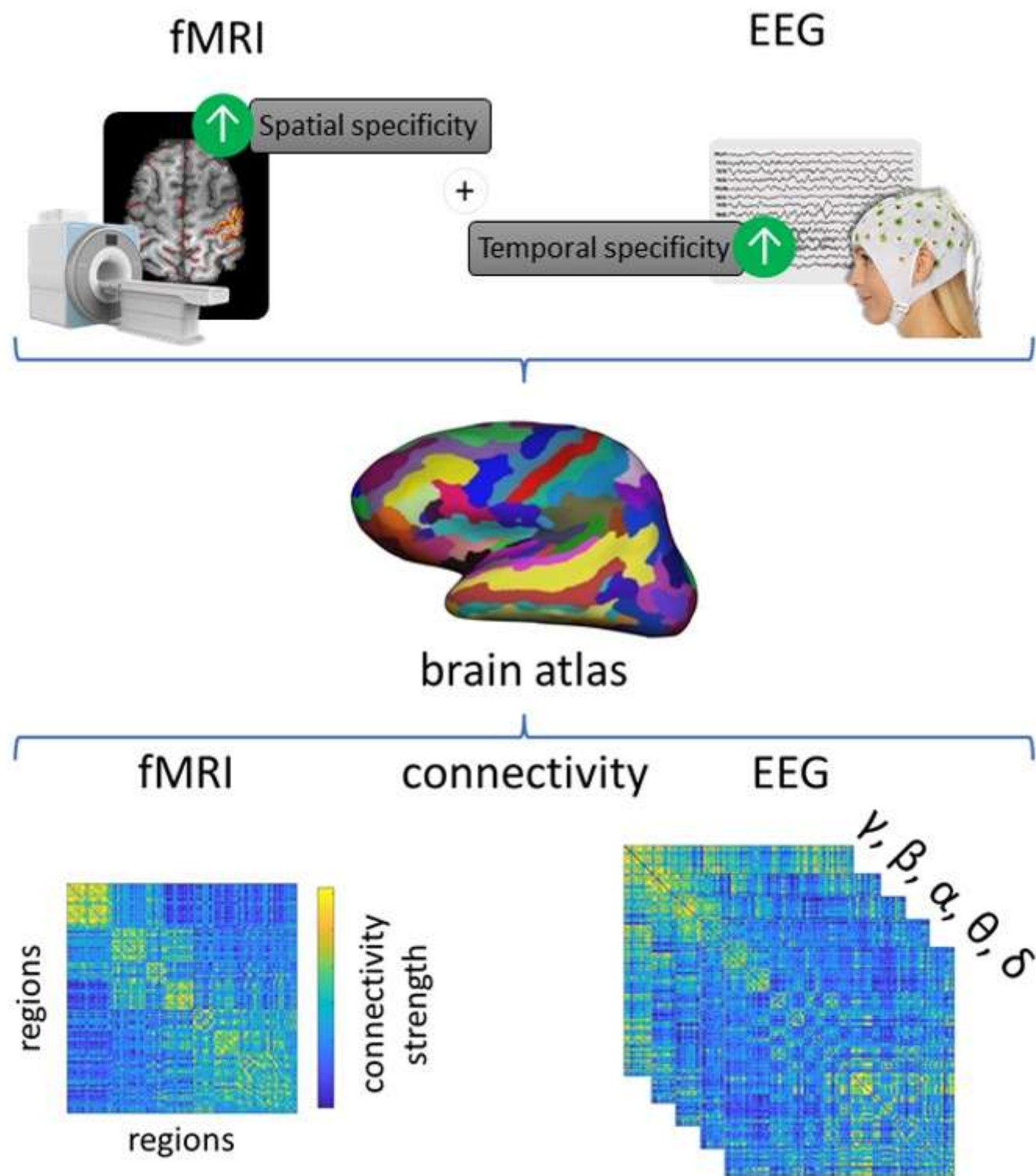
Middl Image is public made domain

Right Image is CC-BY 2.0; changes

fMRI与EEG

fMRI (功能性磁共振成像)- 基于磁共振原理，测量大脑不同区域的血流和代谢活动。

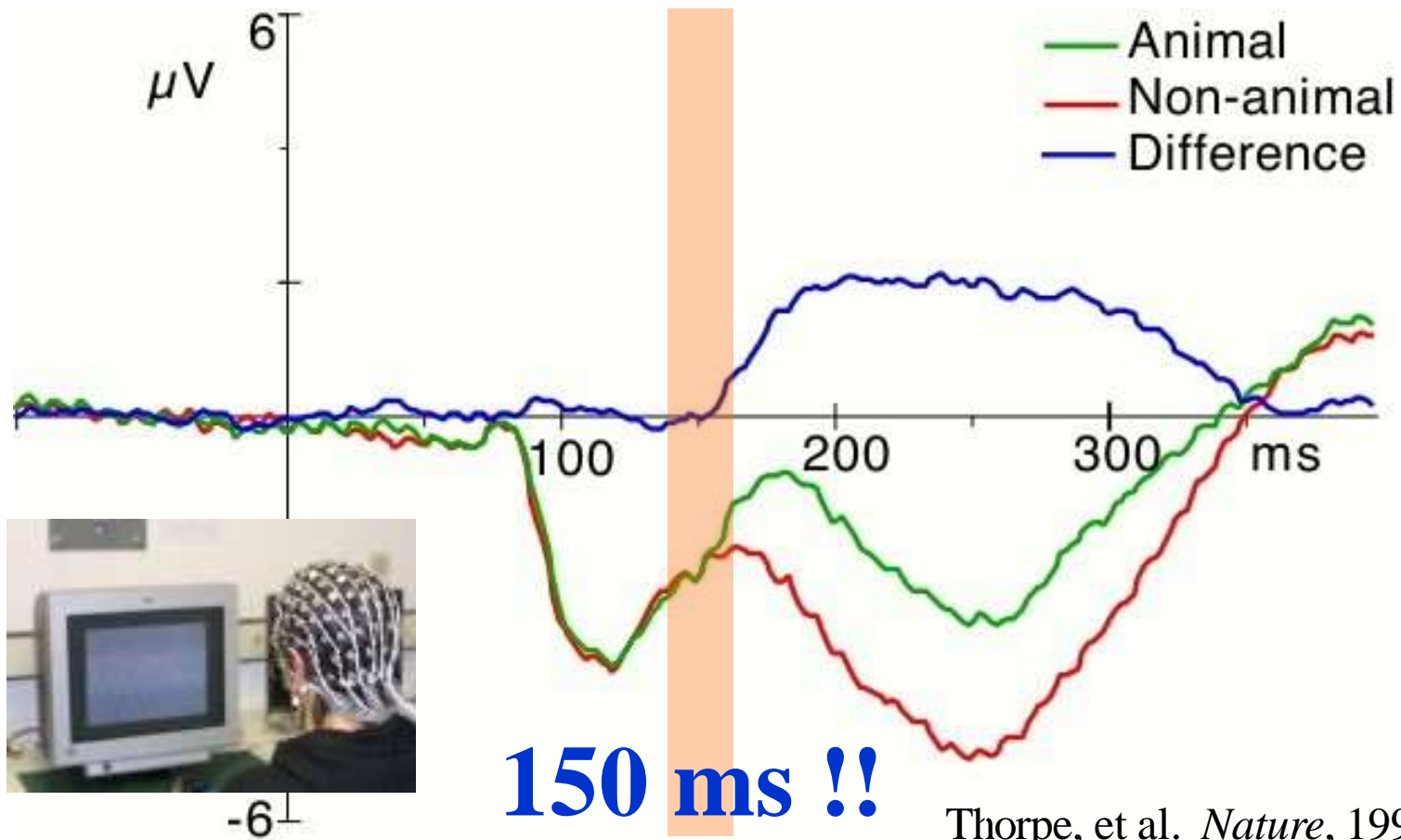
EEG (脑电图)- 记录大脑皮层电活动，通过电极放置在头皮上来捕捉信号。





Speed of processing in the human visual system

Simon Thorpe, Denis Fize & Catherine Marlot



Thorpe, et al. *Nature*, 1996



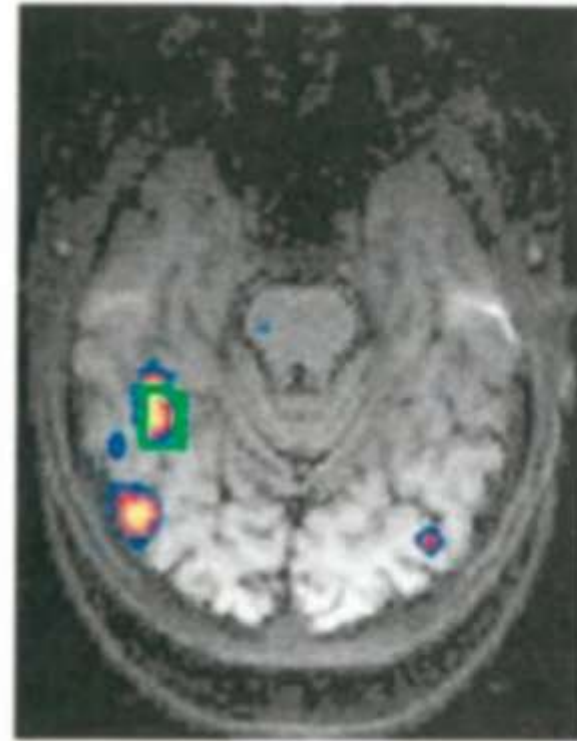
让被试判断图像是否有动物

根据脑信号相关电位变化判断人脑分辨不同图像的时间

出现差异的时间为：
150ms

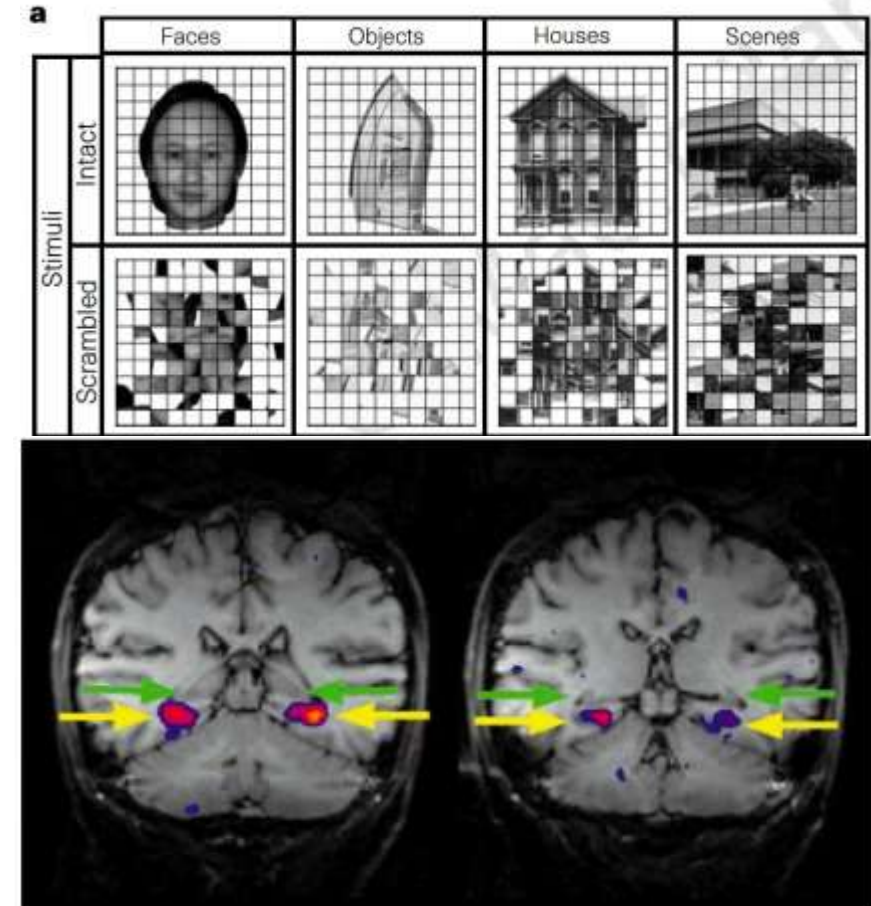
大脑哪些位置有响应？

Faces > Houses



% signal change

Kanwisher et al. J. Neuro. 1997



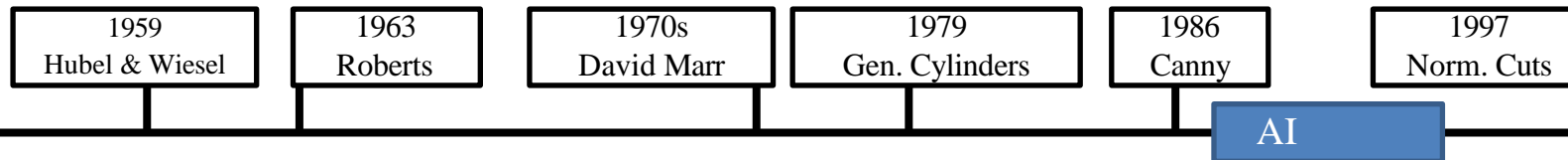
Epstein & Kanwisher, Nature, 1998

聚类、分割 (1990s)



将图像中的像素划分成若干个区域，使得同一区域内的像素具有相似的属性（例如颜色、纹理、亮度等）

有助于图像理解、目标识别和计算机视觉应用



Normalized Cuts, Shi and Malik, 1997

Left Image is CC BY 3.0

Middl Image is public domain

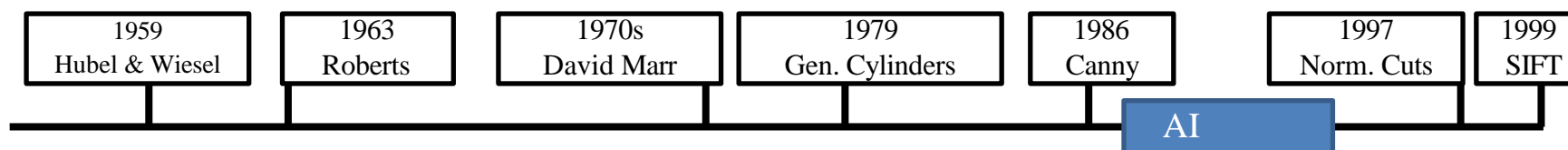
Right Image is CC-BY 2.0; changes made

特征匹配 (2000s)



不同视角下的物体如何提取特征，匹配特征？

解决了图像中的关键点检测、尺度不变性、旋转不变性和局部形变不变性等问题



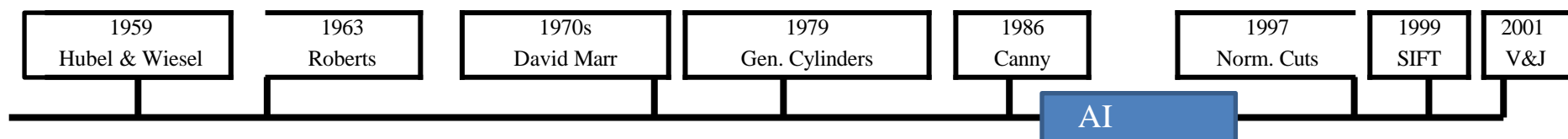
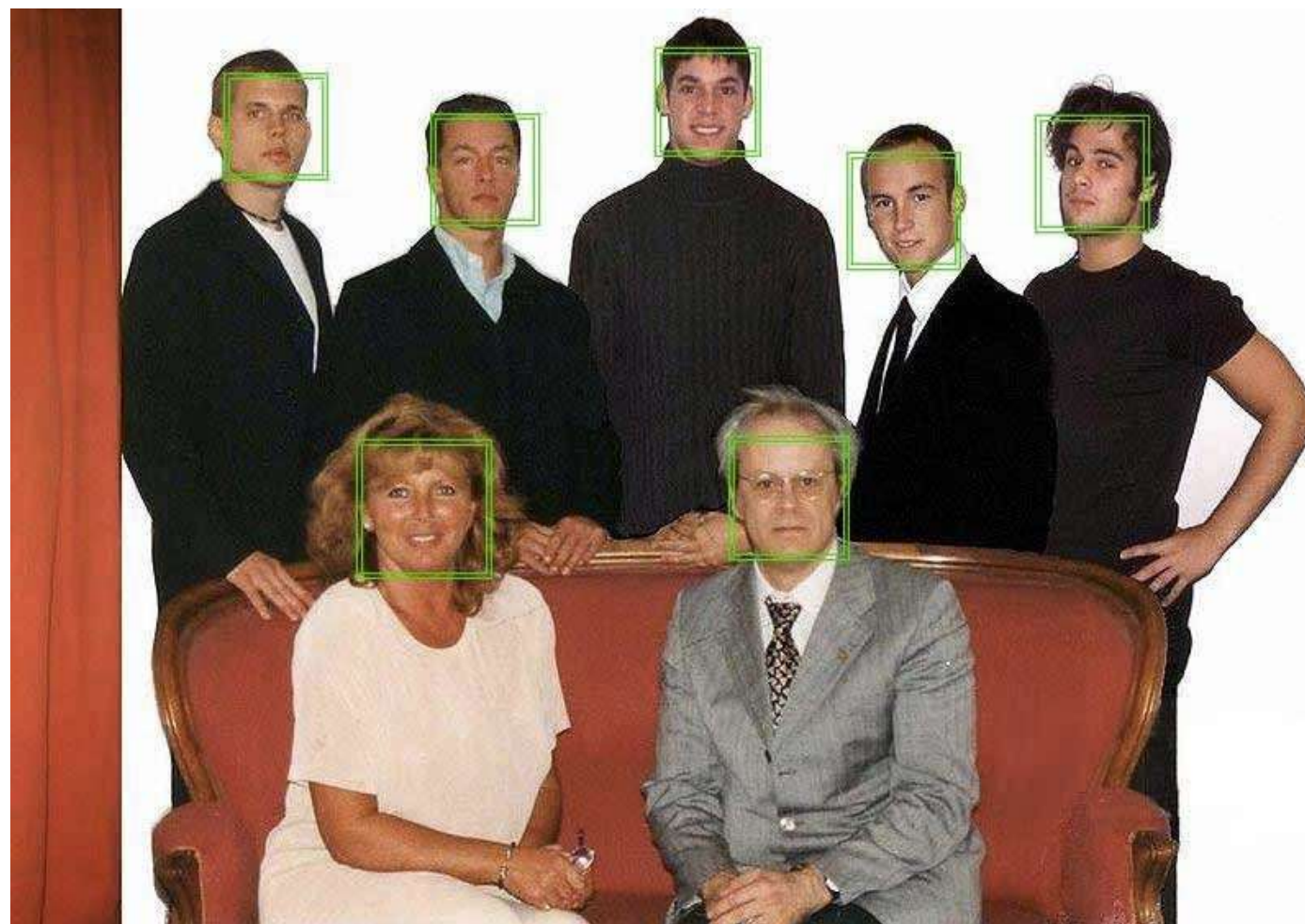
SIFT, David Lowe, 1999

人脸检测

Viola and Jones, 2001

机器视觉领域最成果的应用！

结合Haar特征（基于图像局部区域的特征，可以用于表示图像中的纹理、边缘等信息）和Adaboost算法（组合多个弱分类器来构建一个强分类器，从而提高分类准确率）来进行人脸检测

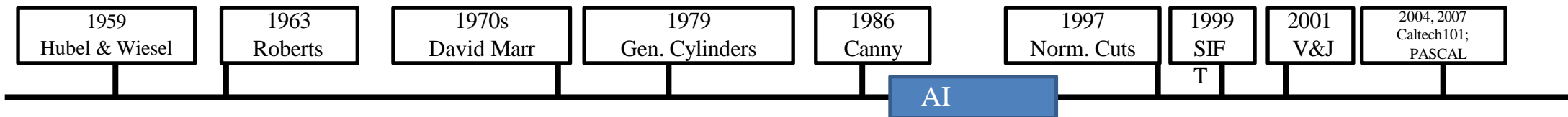


2004、2007年出现了规模较大的图像分类数据集

Caltech 101 images



PASCAL Visual Object Challenge



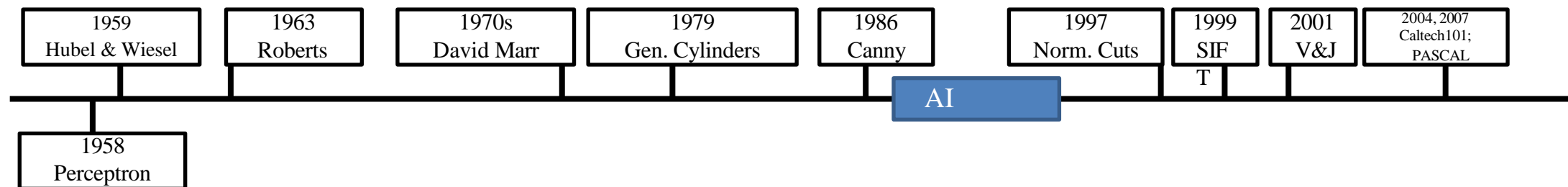
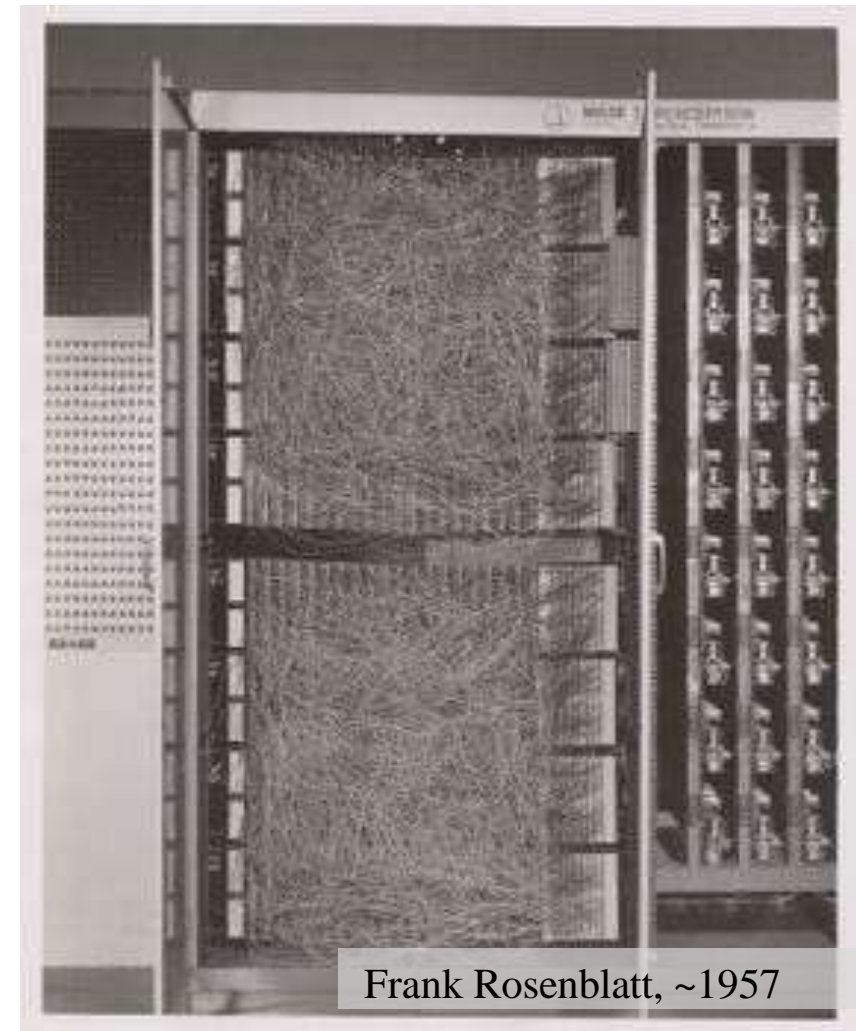
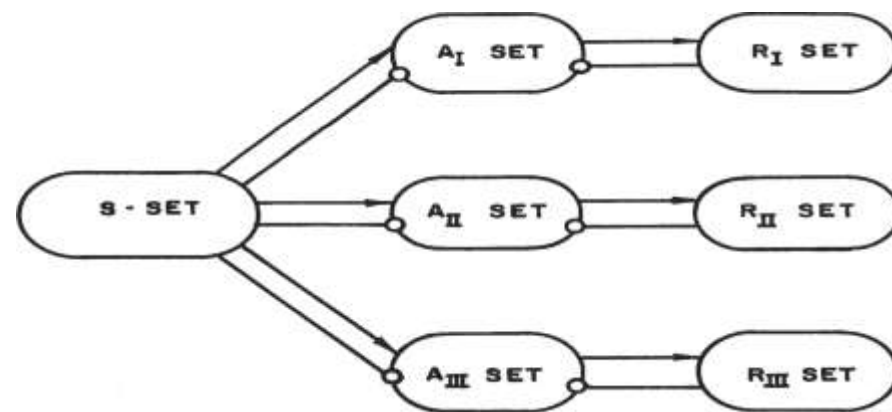
感知机Perceptron

学习一个分割超平面（或者称为决策边界），这个超平面可以将不同类别的数据点正确地分开。

Learning representations by back-propagating errors

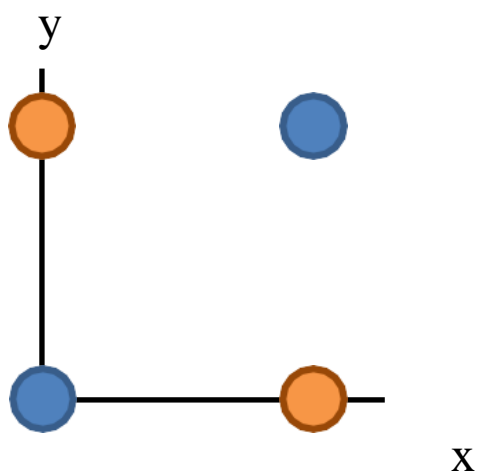
David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

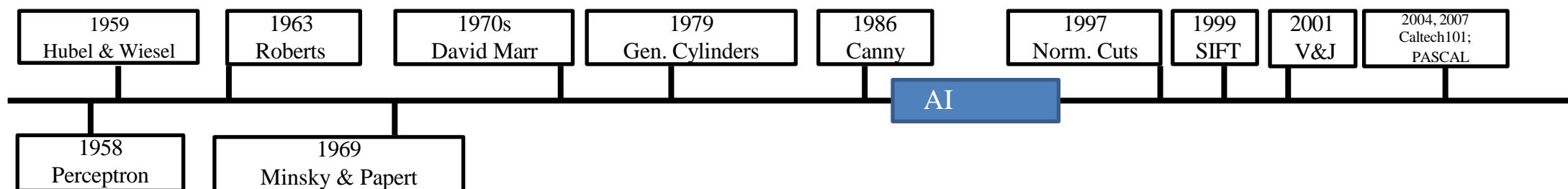
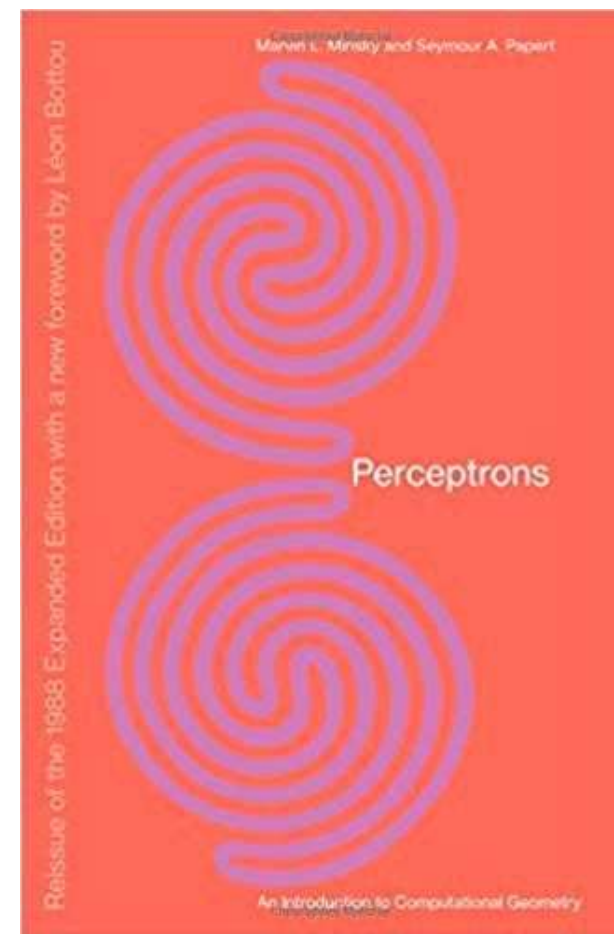


Minsky and Papert, 1969

X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0



感知机可以实现如异或 (XOR) 等不同的运算

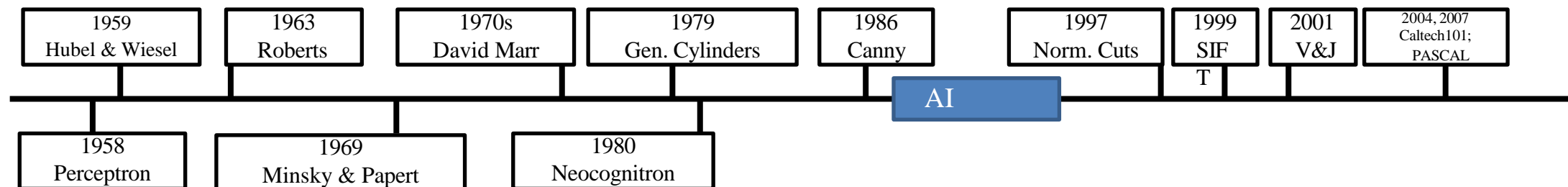
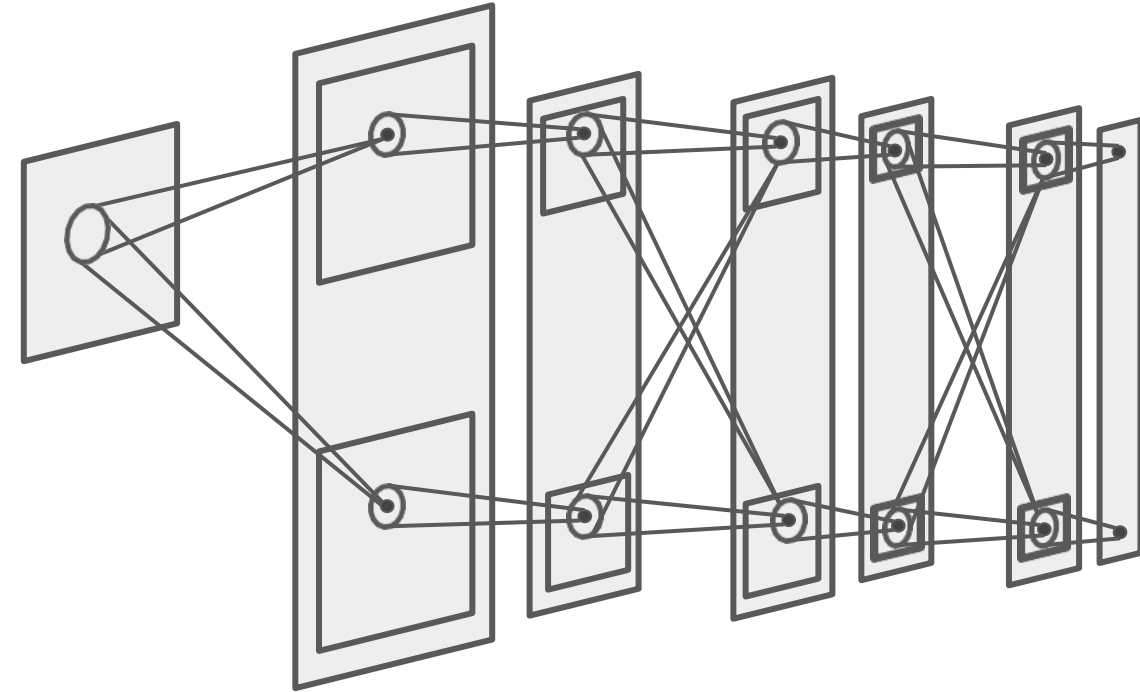


认知机: Fukushima, 1980

计算机视觉的计算模型，直接受到Hubel和Wiesel关于复杂细胞和简单细胞层次结构的启发。

交织的简单细胞（卷积）和复杂细胞（池化）。

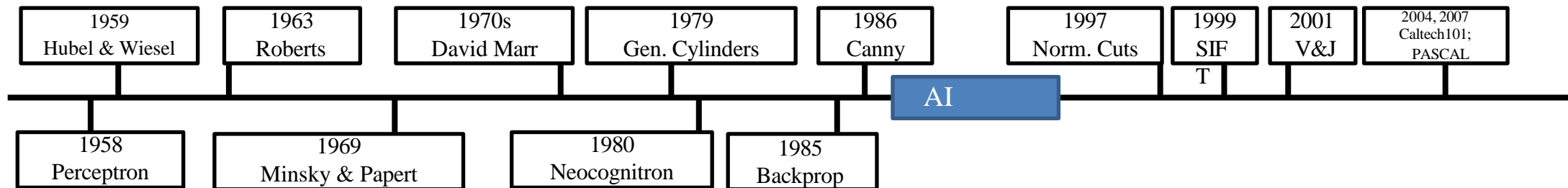
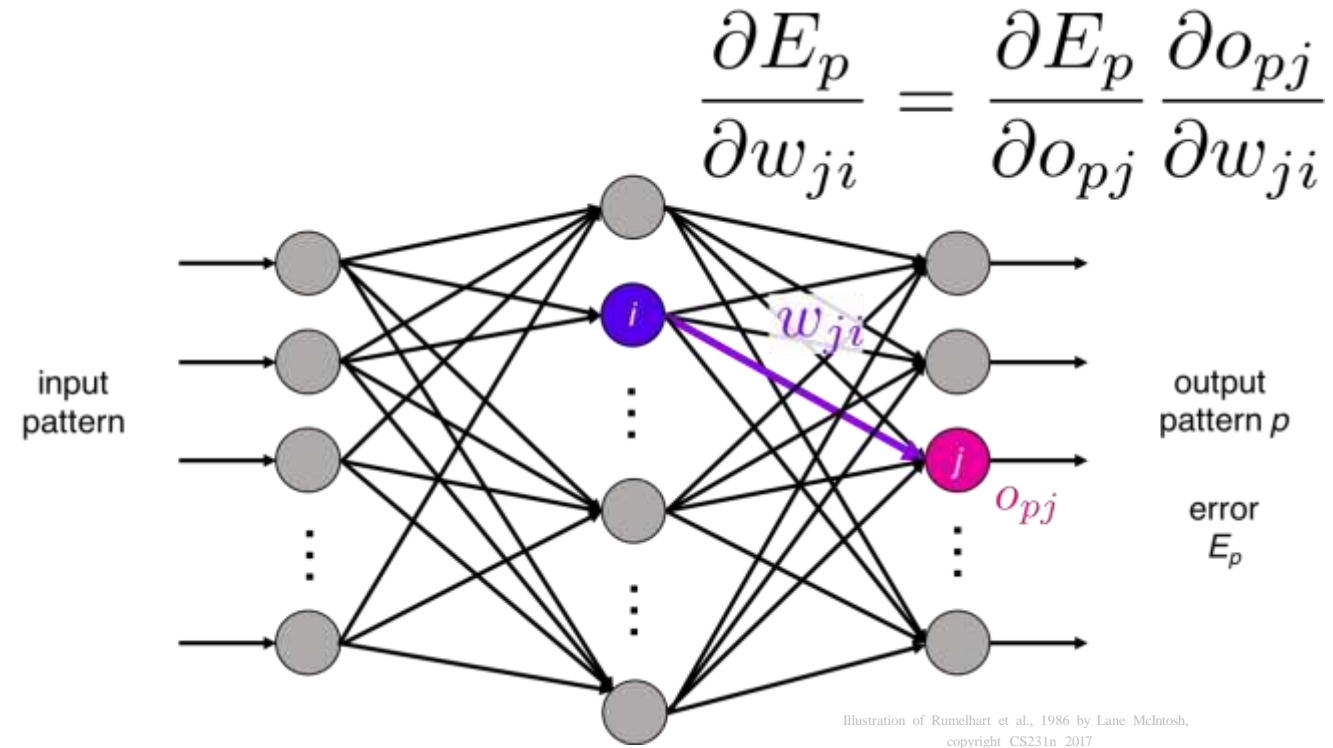
但该论文没有提出实用的训练算法。



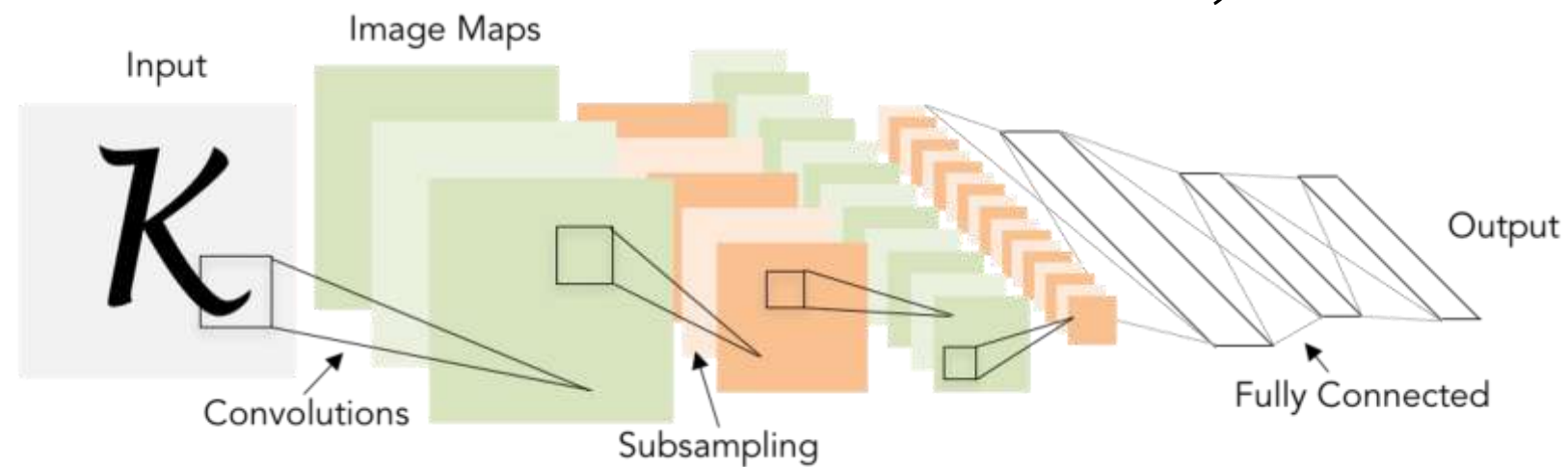
梯度反传: Rumelhart, Hinton, and Williams, 1986

提出了基于梯度反传的神经网络训练方法

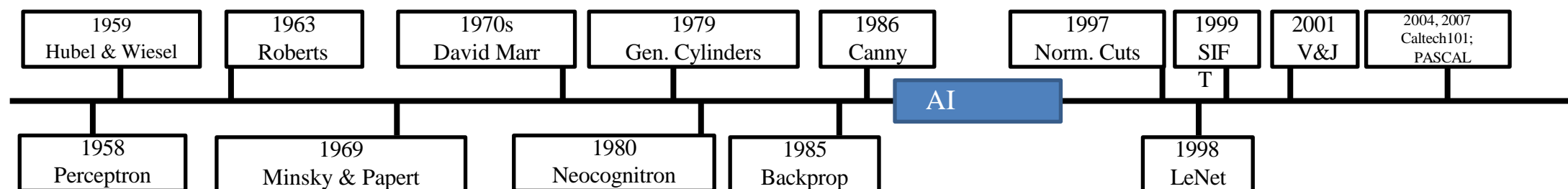
感知机可以多层训练了!



卷积神经网络: LeCun et al, 1998



将反向传播算法应用于类似认知机的架构，学习识别手写数字
被NEC用于商业系统，处理手写支票。
与我们现代的卷积网络非常相似！

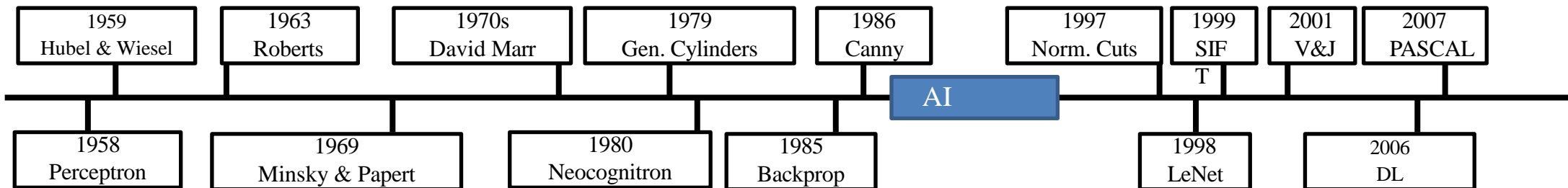
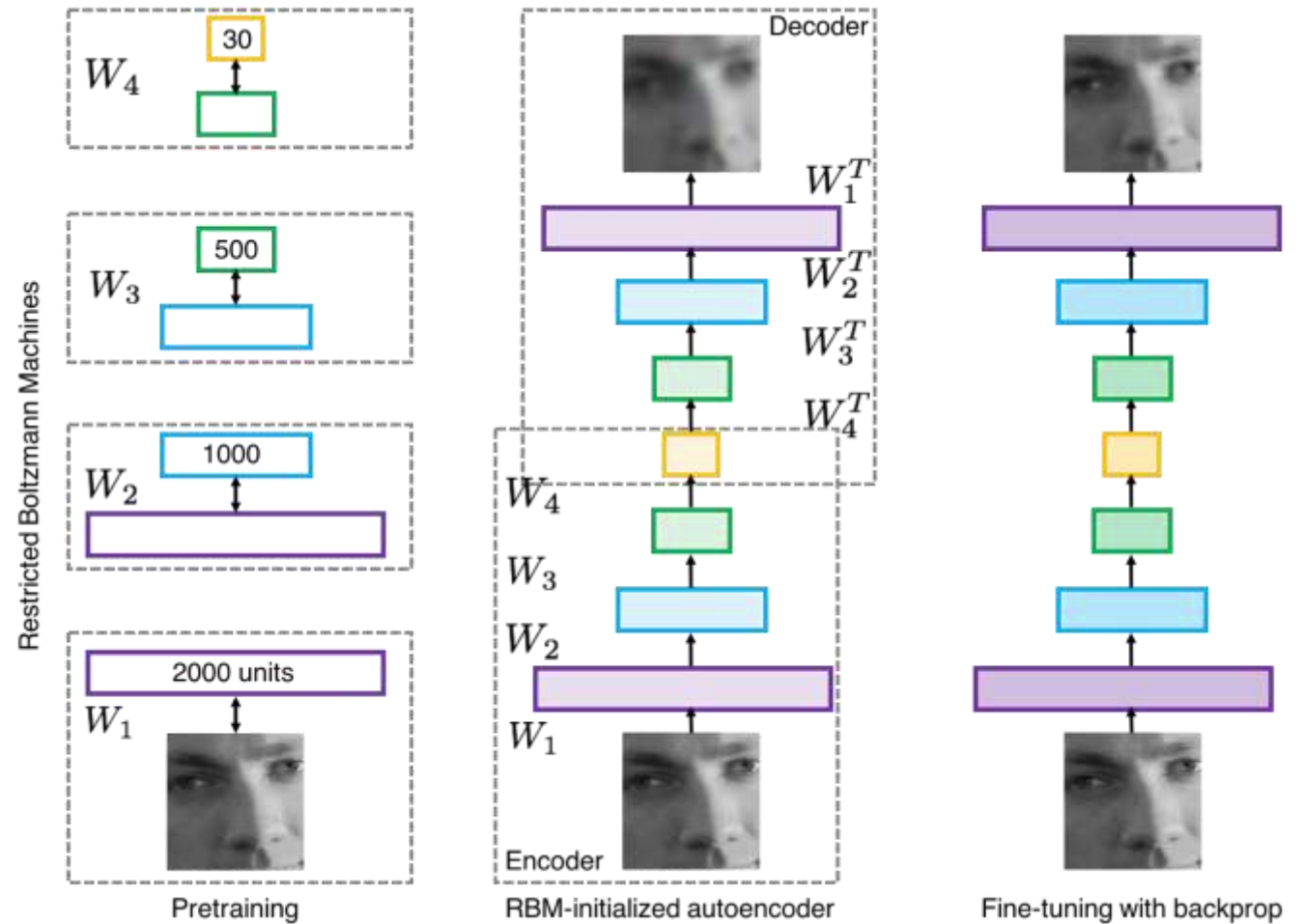


2000s: “深度学习”

网络从此变得越来越深

但还没有足够的训练数据

- Hinton and Salakhutdinov, 2006
- Bengio et al, 2007
- Lee et al, 2009
- Glorot and Bengio, 2010



IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

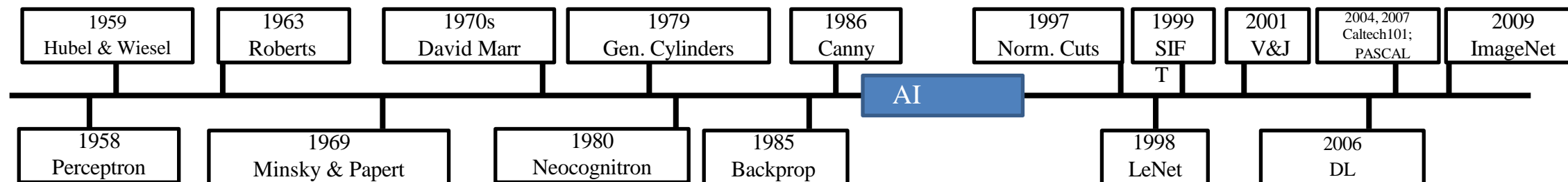
1,000 object classes

1,431,167 images

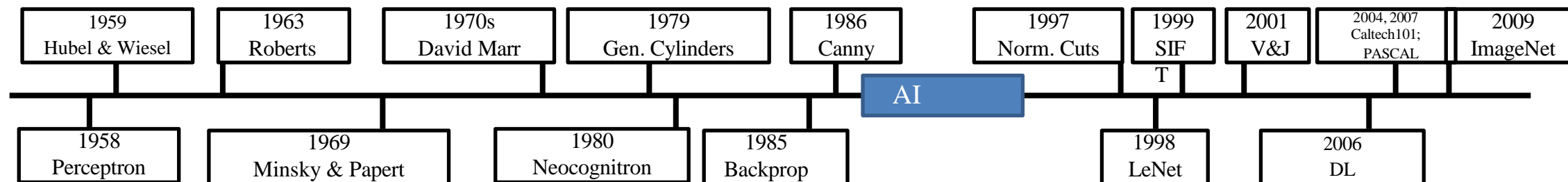
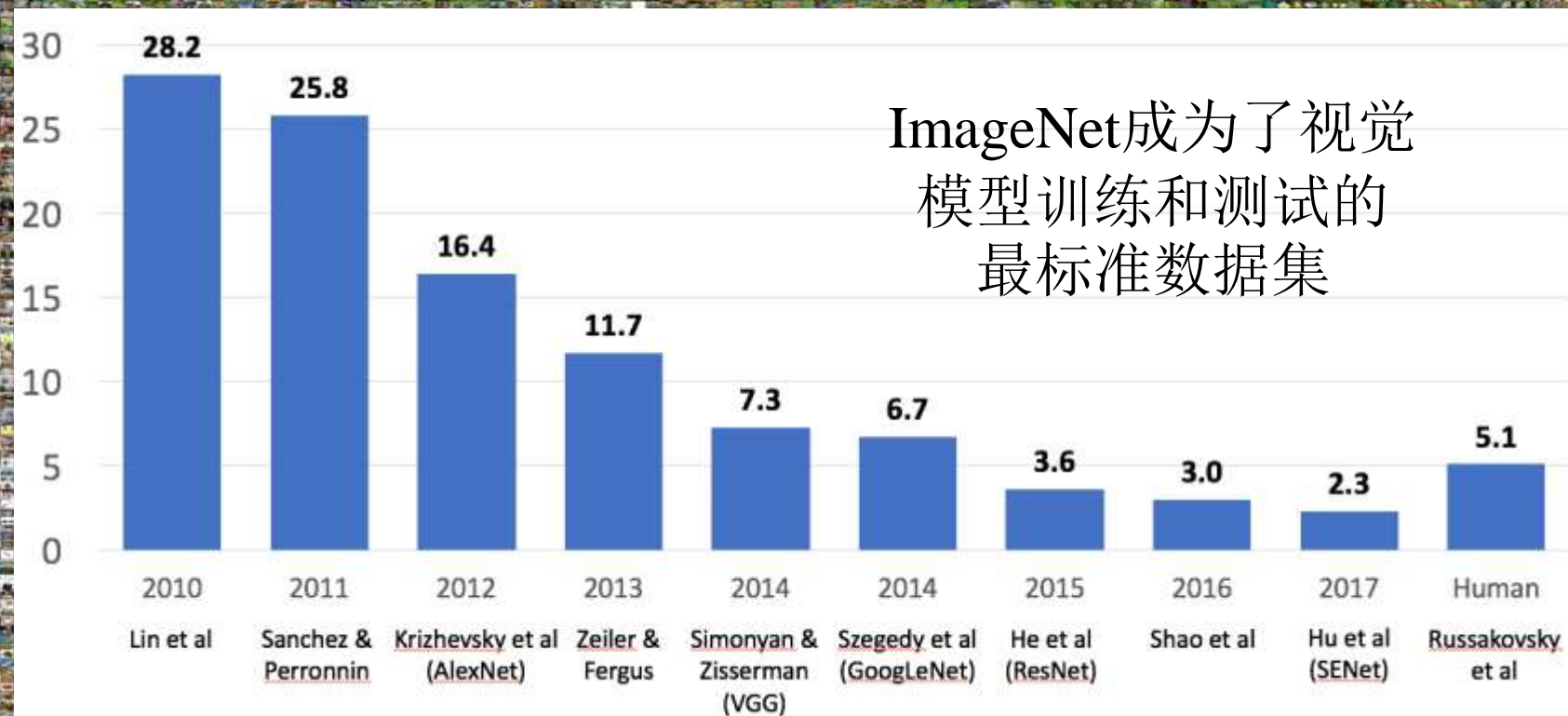


Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

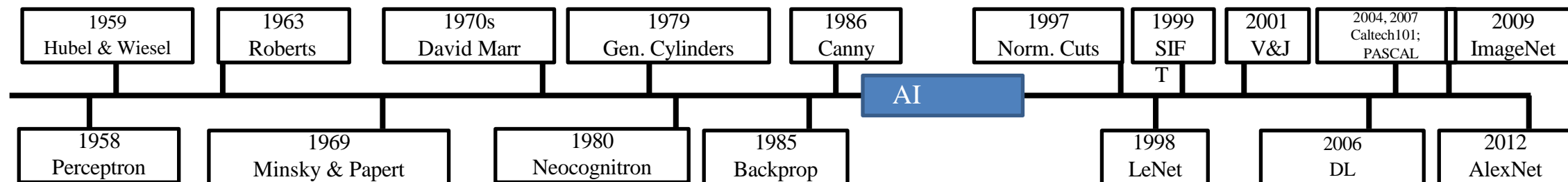
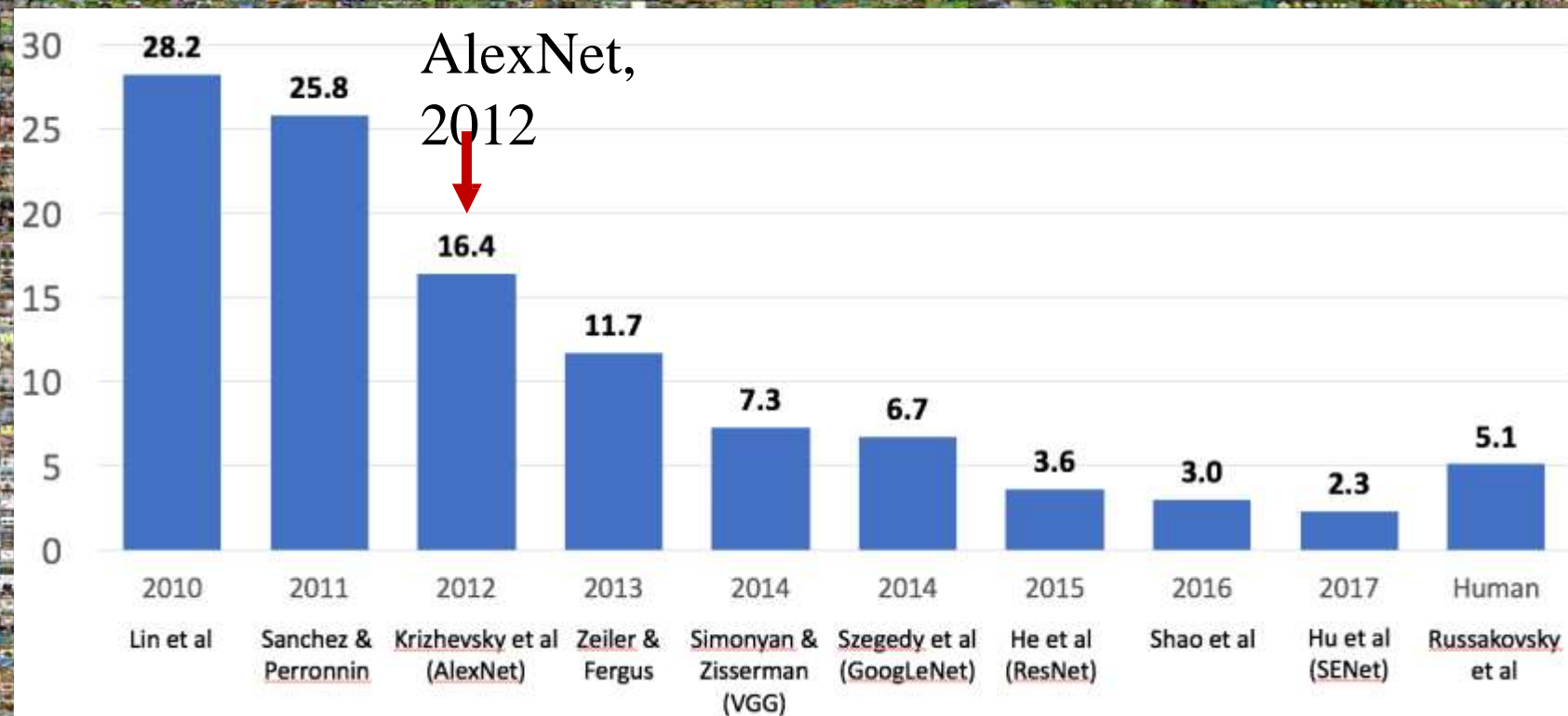
Deng et al, 2009 Russakovsky et al. IJCV 2015



IMAGENET Large Scale Visual Recognition Challenge



IMAGENET Large Scale Visual Recognition Challenge



2000s: “深度学习”

网络从此变得越来越深

有训练数据了！

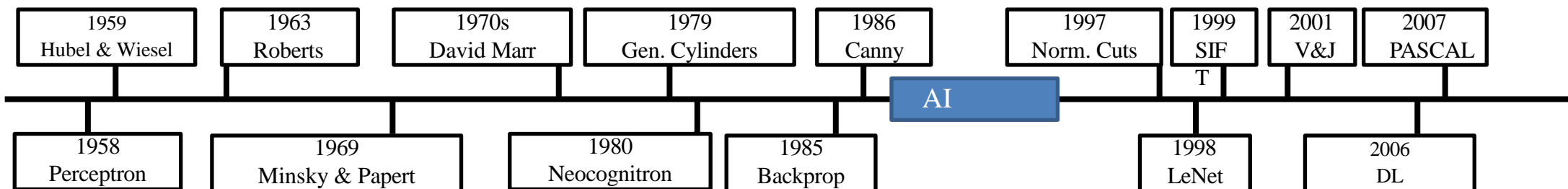
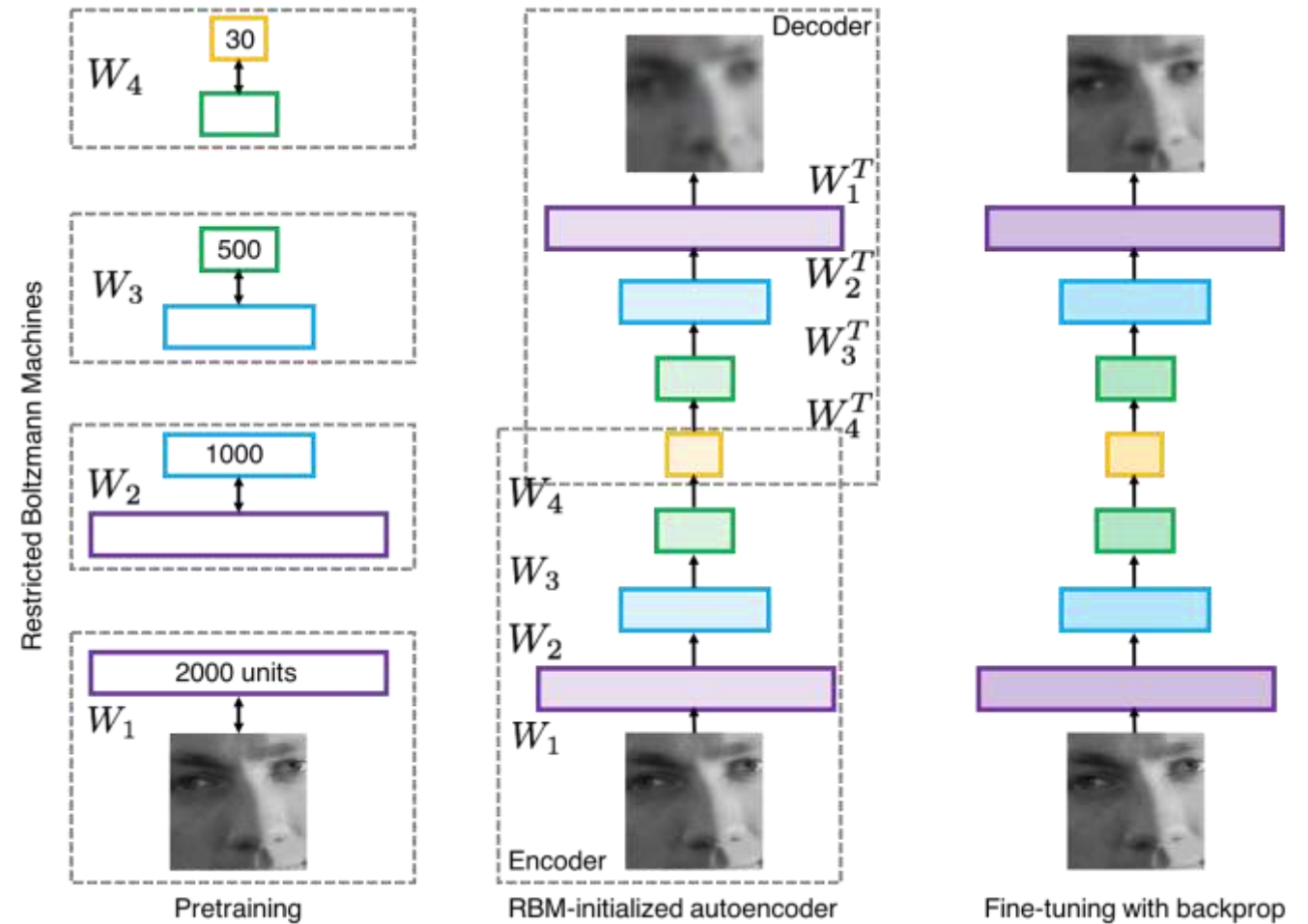
但是没有办法有效训练深度网络

Hinton and Salakhutdinov, 2006

Bengio et al, 2007

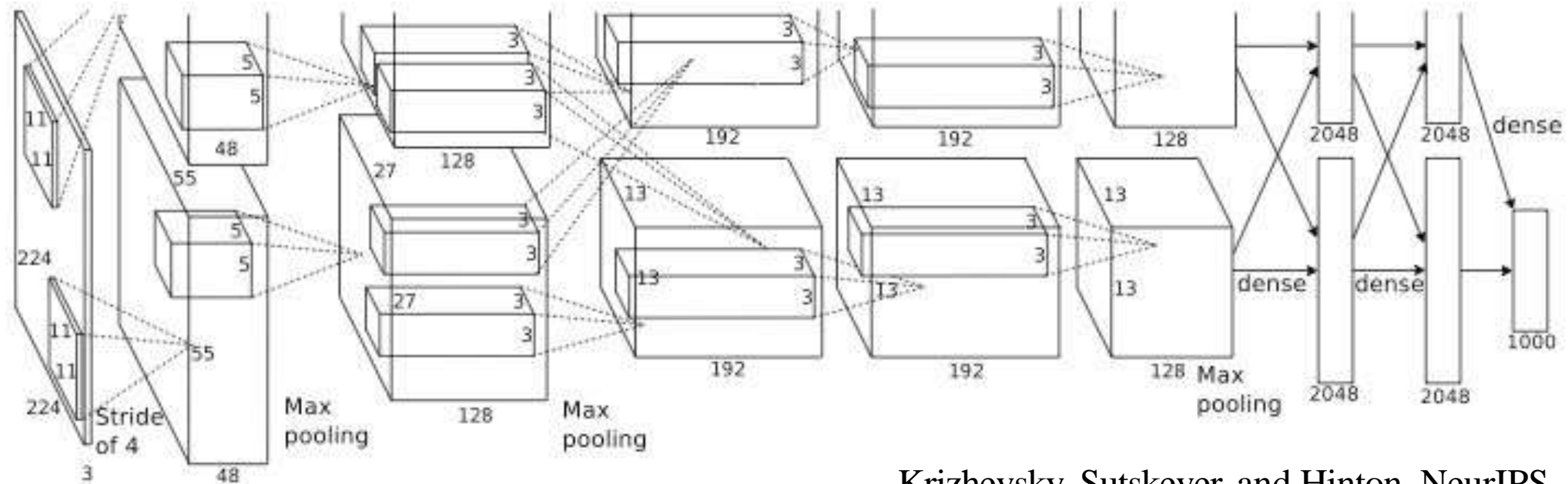
Lee et al, 2009

Glorot and Bengio, 2010

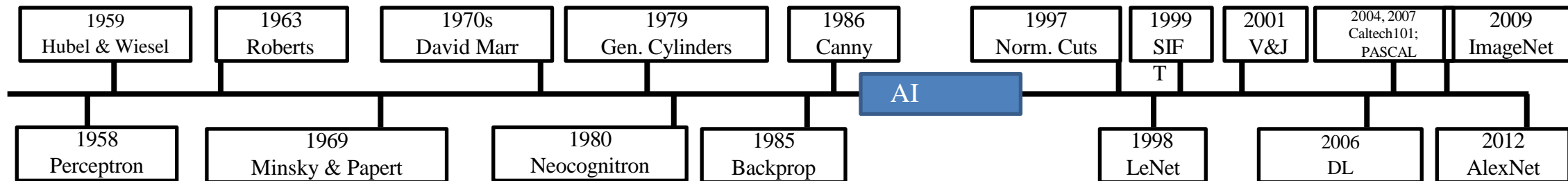


AlexNet: 深度学习 成为主流模型

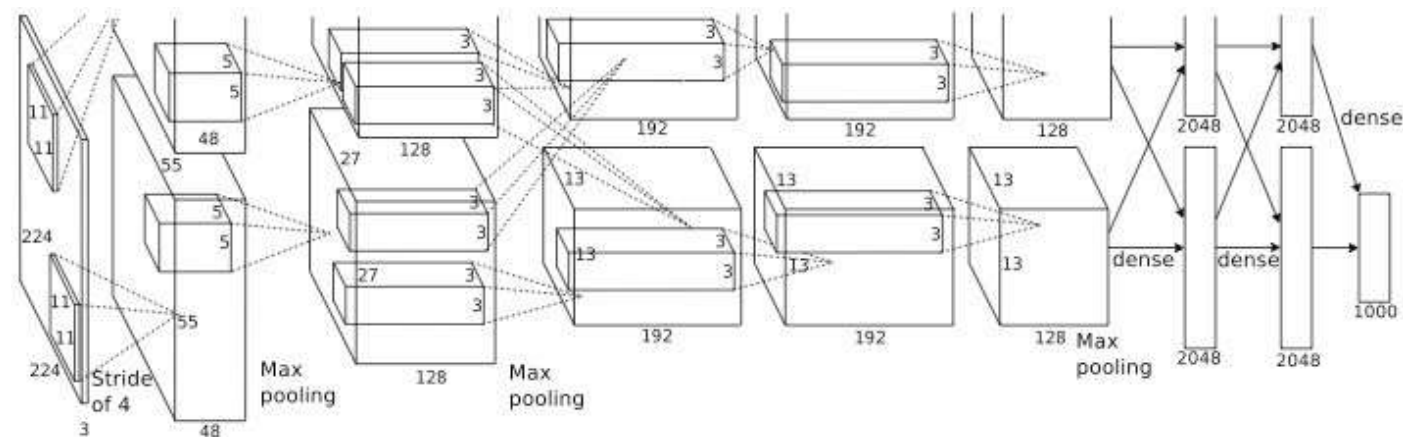
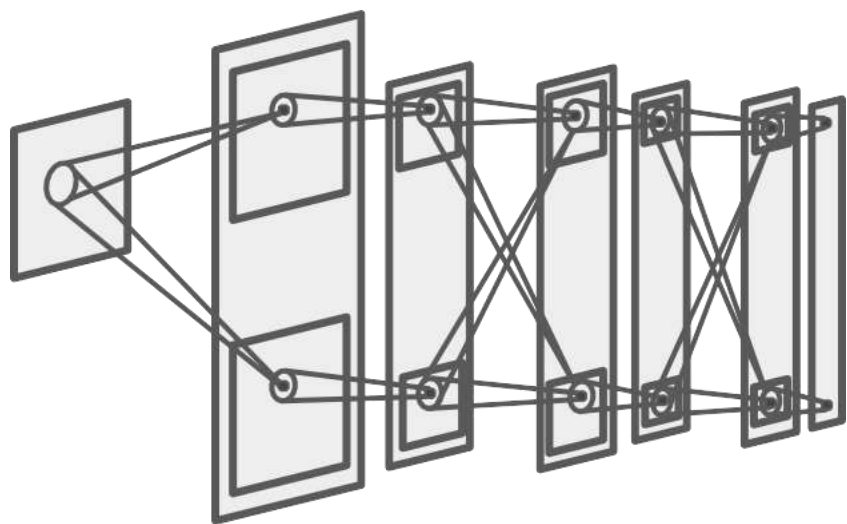
GPU训练大大提升了
深度网络训练效率!



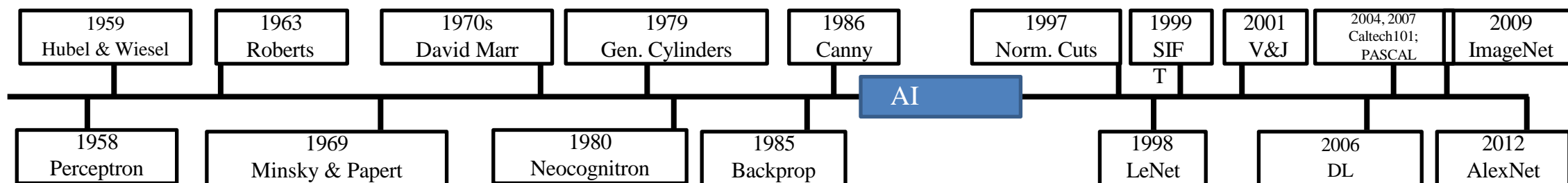
Krizhevsky, Sutskever, and Hinton, NeurIPS 2012



AlexNet vs. Neocognitron: 32 年的发展历程

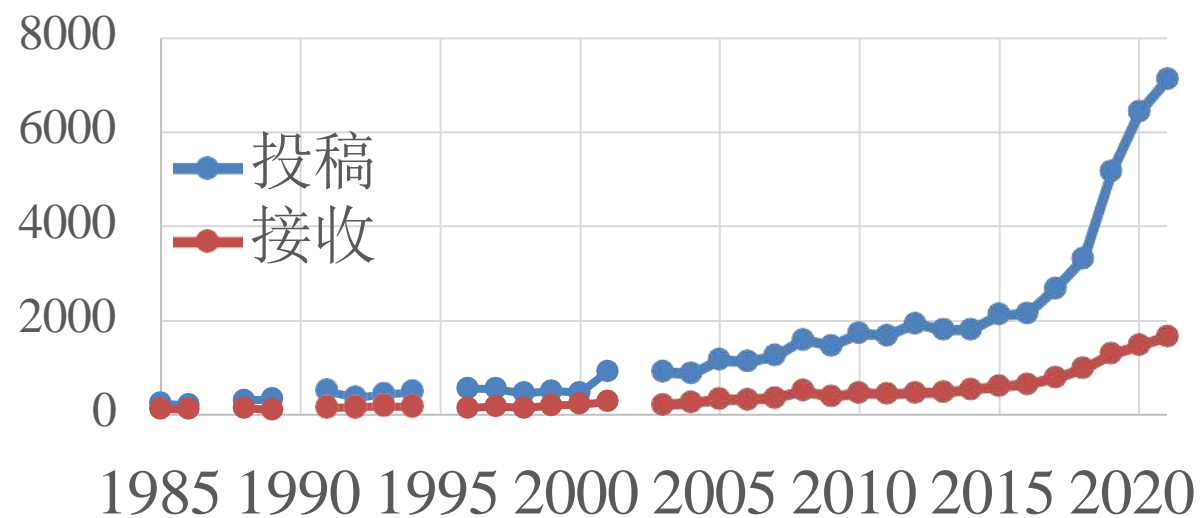


更深、更精细、更适应硬件的设计

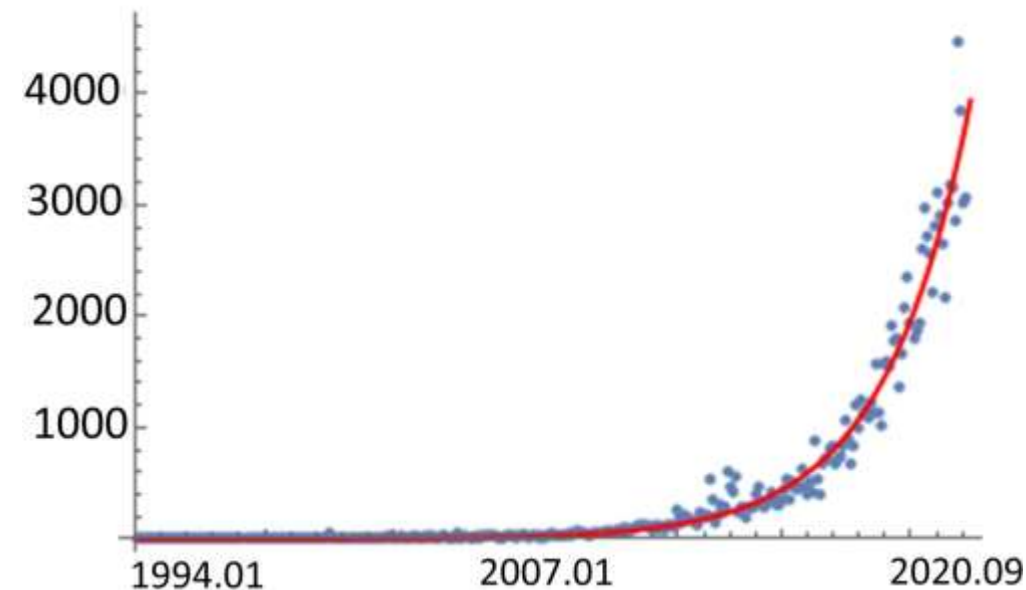


2012 至今: 深度学习爆炸式发展

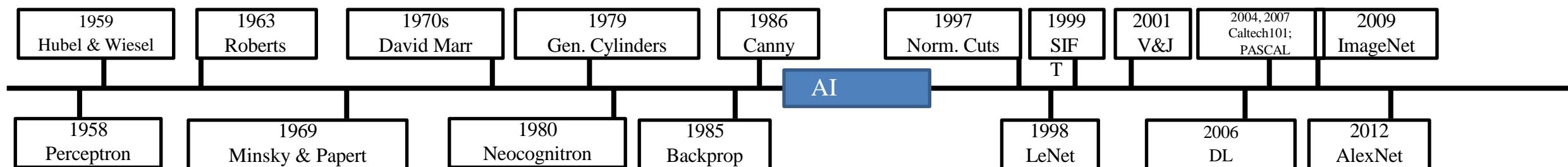
CVPR Papers



ML+AI arXiv papers per month



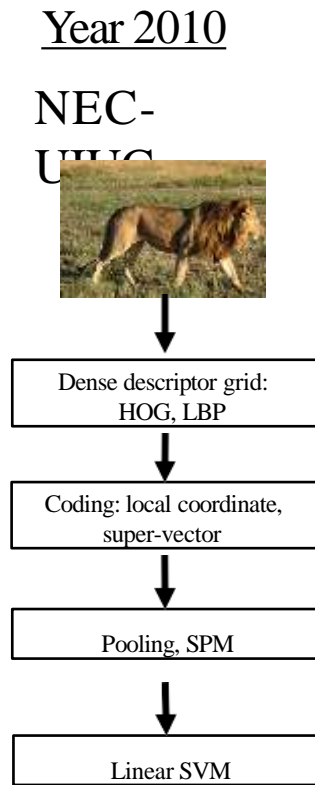
计算机视觉顶会论文



arXiv papers per month

2012至今：无处不在的深度学习

各种网络
竞相问世



[Lin CVPR 2011]

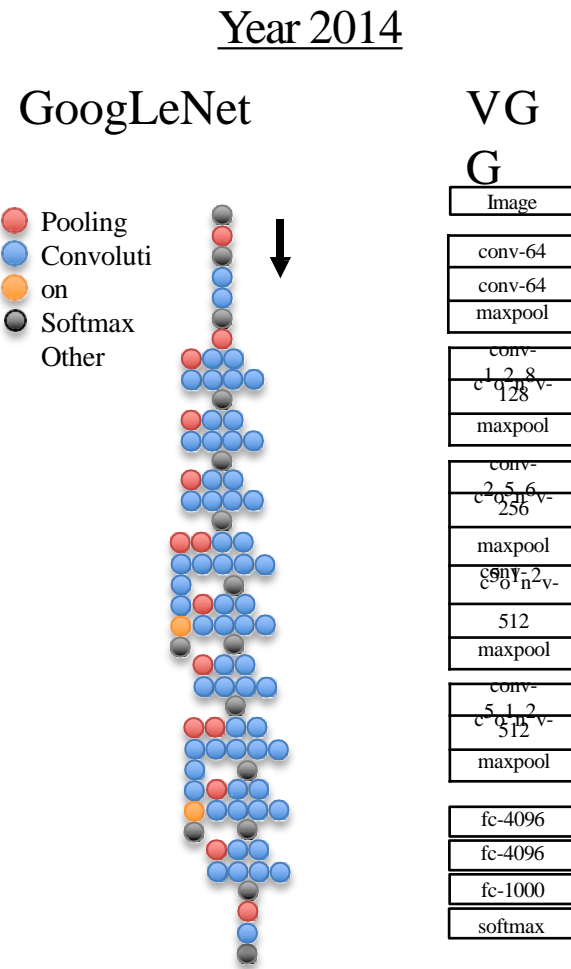
Lion image by Swissfrog is licensed under [CC BY 3.0](https://creativecommons.org/licenses/by/3.0/)



[Krizhevsky NIPS 2012]

Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission. Figure copyright Alex Krizhevsky, Ilya



[Szegedy arxiv 2014]

[Simonyan arxiv 2014]



[He ICCV 2015]

2012至今：无处不在的深度学习

最基础的视觉任务

Image Classification

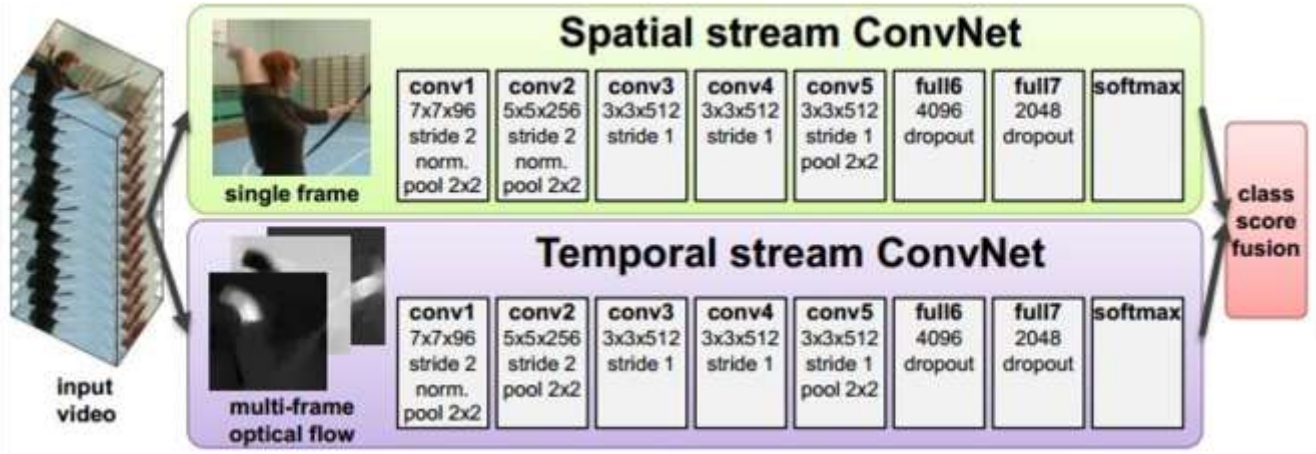
Image Retrieval



Figures copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

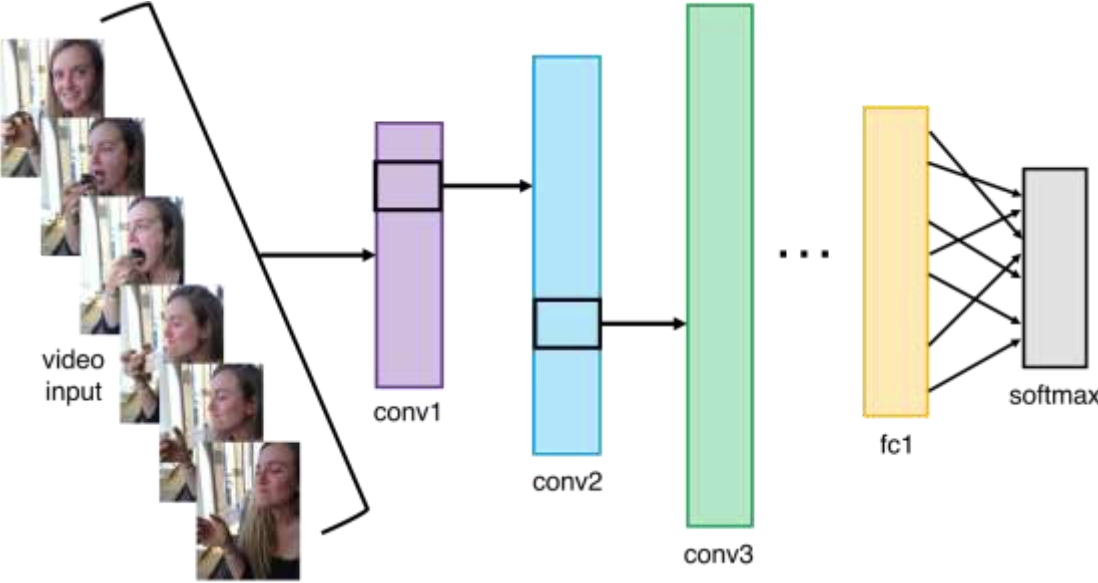
2012至今：无处不在的深度学习

Video Classification



Simonyan et al, 2014

Activity Recognition



2012至今：无处不在的深度学习

机器人 视觉任务

Pose Recognition (Toshev and Szegedy, 2014)



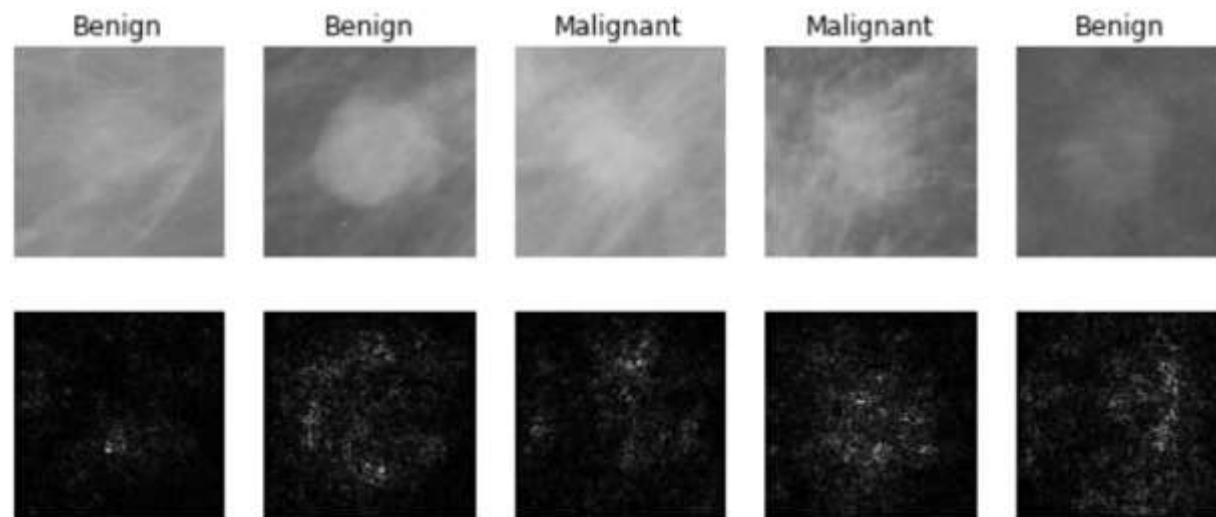
Playing Atari games (Guo et al, 2014)



2012至今：无处不在的深度学习

科研、探索视觉任务

Medical Imaging



Levy et al, 2016 Figure reproduced with permission

Galaxy Classification



Dieleman et al, 2014

From left to right: [public domain by NASA](#), [usage permitted by ESA/Hubble](#), [public domain by NASA](#), and [public domain](#).

Whale recognition



Kaggle Challenge

[This image](#) by Christin Khan is in the public domain and originally came from the U.S. NOAA.

2012至今：无处不在的深度学习

视觉理解



A white teddy bear sitting in the grass



A man in a baseball uniform throwing a ball



A woman is holding a cat in her hand

Image Captioning
Vinyals et al, 2015
Karpathy and Fei-Fei, 2015



A man riding a wave on top of a surfboard



A cat sitting on a suitcase on the floor

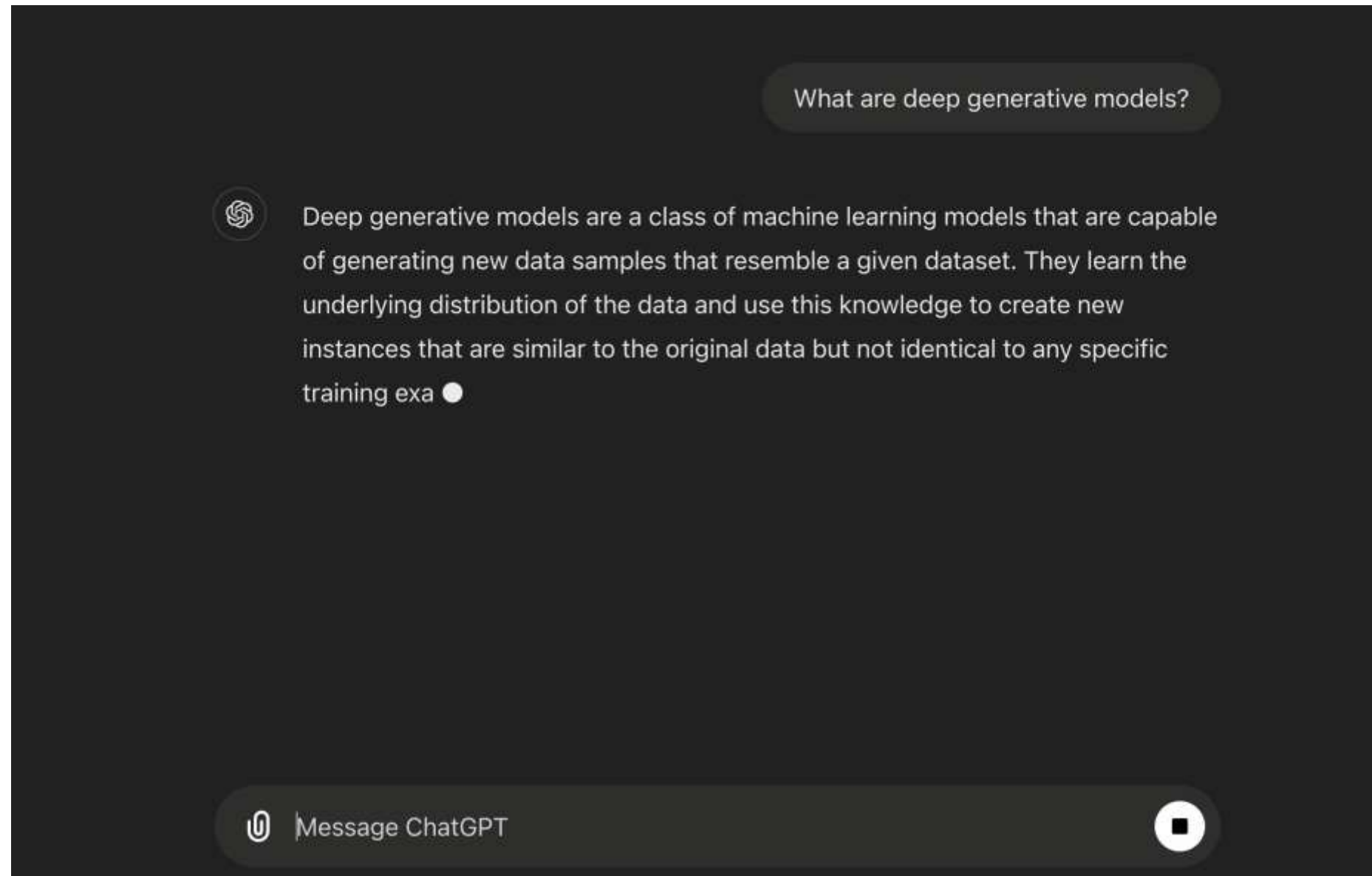


A woman standing on a beach holding a surfboard

All images are CC0 Public domain:
<https://pixabay.com/en/luggage-antique-cat-1643010/>
<https://pixabay.com/en/teddy-plush-bears-cute-teddy-bear-1623436/>
<https://pixabay.com/en/surf-wave-summer-sport-litoral-1668716/>
<https://pixabay.com/en/woman-female-model-portrait-adult-583967/>
<https://pixabay.com/en/handstand-lake-meditation-496008/>
<https://pixabay.com/en/baseball-player-shortstop-infield-1045263/>
Captions generated by Justin Johnson using [NeuralTalk2](#)

The “GenAI” Era – From 2022

Chatbot and natural language conversation



The “GenAI” Era – From 2022

Text-to-image generation



Prompt: --aesthetic-- #boho #fashion, full-body 30-something woman laying on microfloral grass, candid pose, overlay reads Stable Diffusion 3.5, cheerful cursive typography font.

<https://stability.ai/news/introducing-stable-diffusion-3-5>

The “GenAI” Era – From 2022

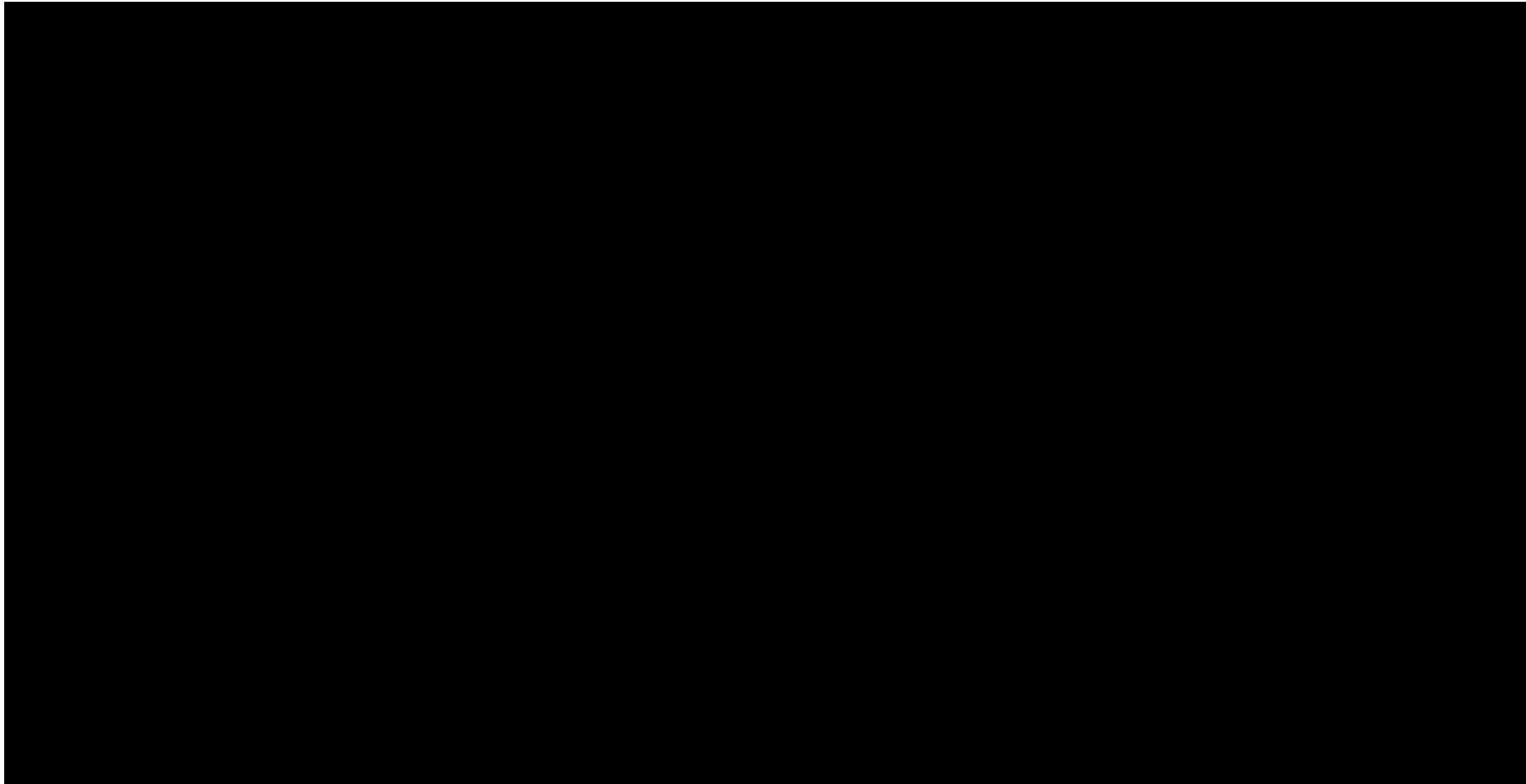
Text-to-video generation



Sora

The “GenAI” Era – From 2022

AI assistant for code generation



The “GenAI” Era

Image-to-image generation



GPT-Image-1 (GPT 4o)

The “GenAI” Era

Image-to-image generation



Nano Banana Pro

The “GenAI” Era

Image-to-image generation



Nano Banana Pro

The “GenAI” Era

Image-to-image generation



The “GenAI” Era

Image-to-image generation



Nano Banana Pro

The “GenAI” Era

Image-to-image generation



Nano Banana Pro

The “GenAI” Era

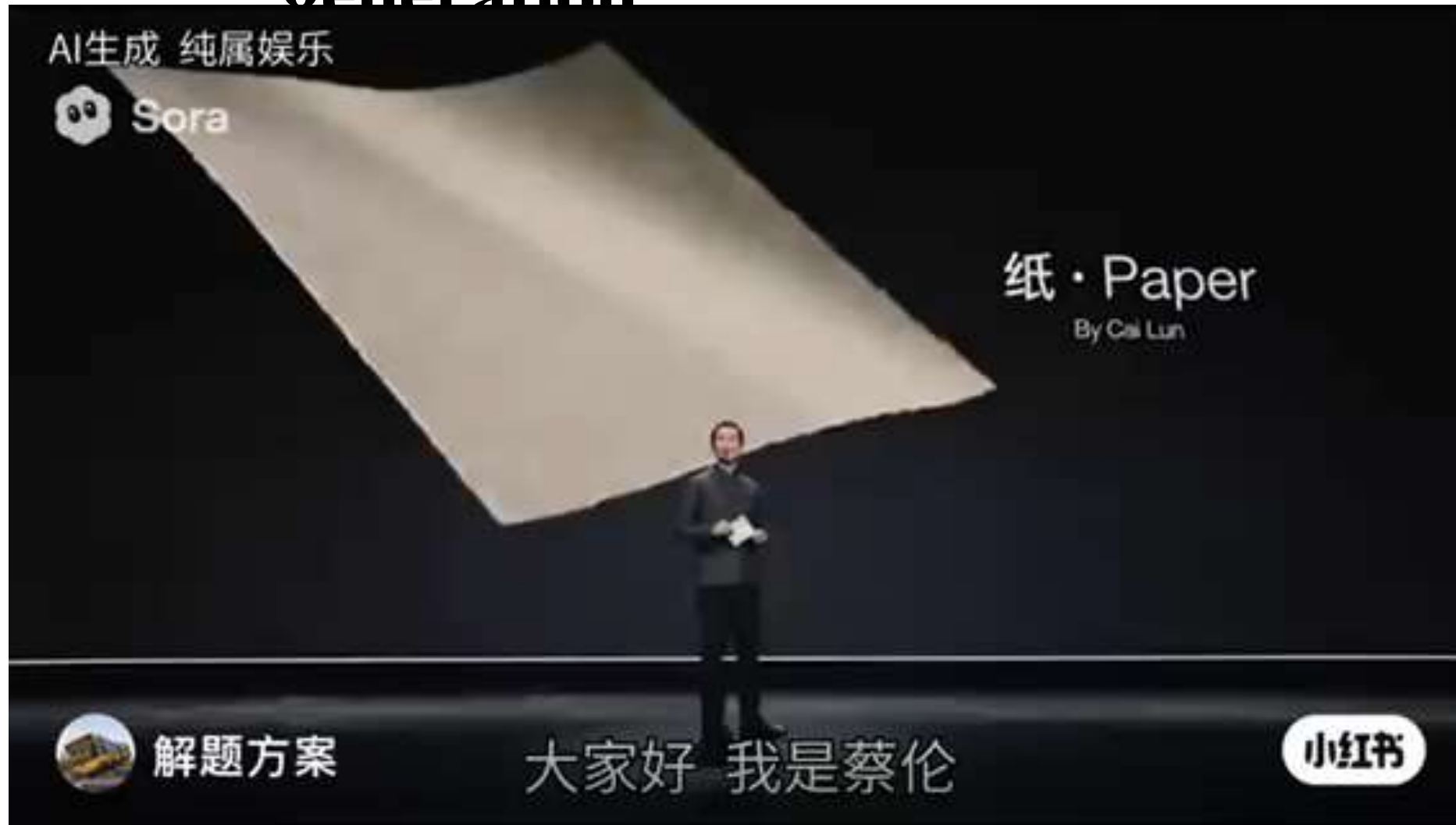
Text-to-video generation



Sora

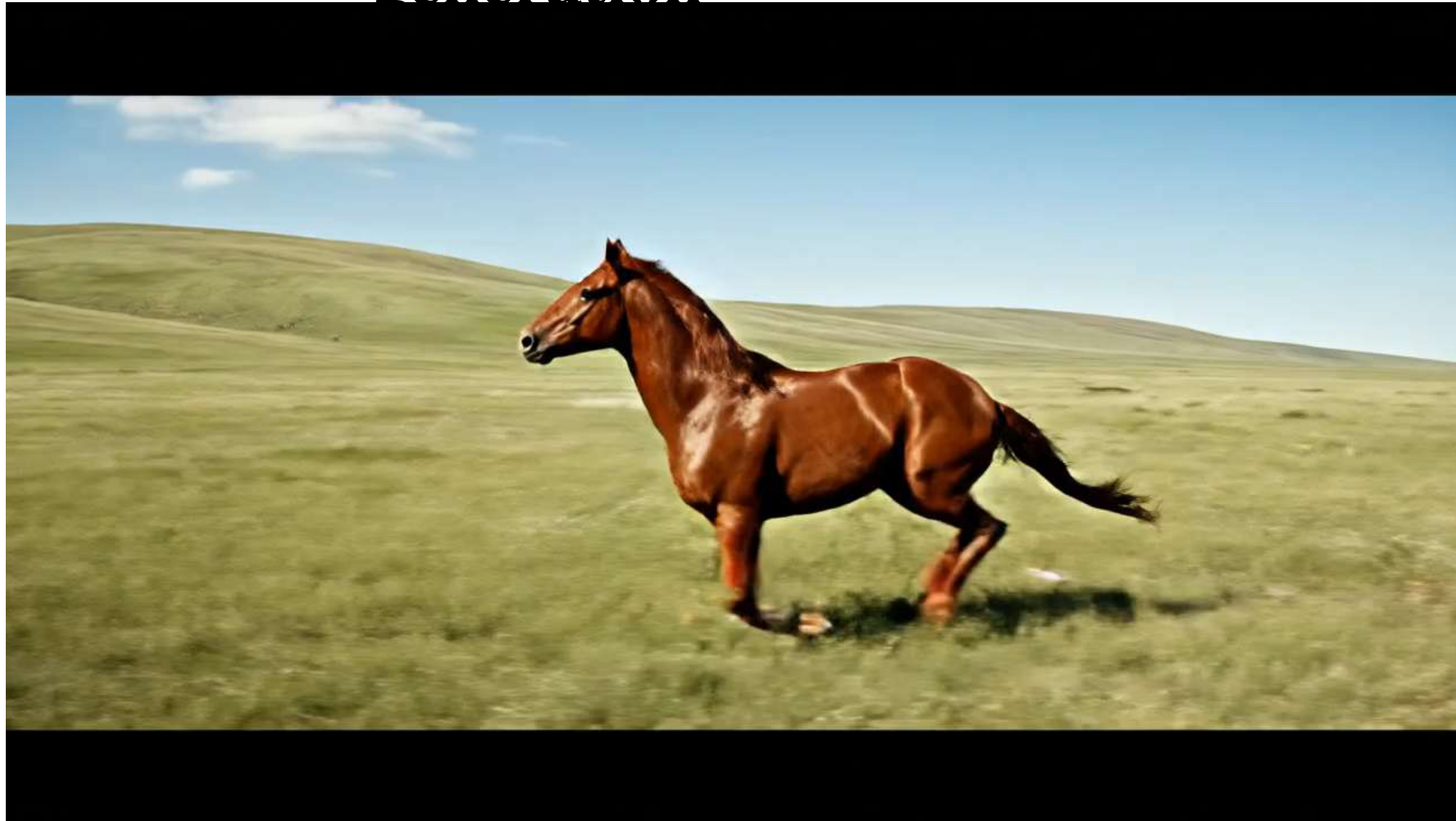
The “GenAI” Era

Text-to-video generation



The “GenAI” Era

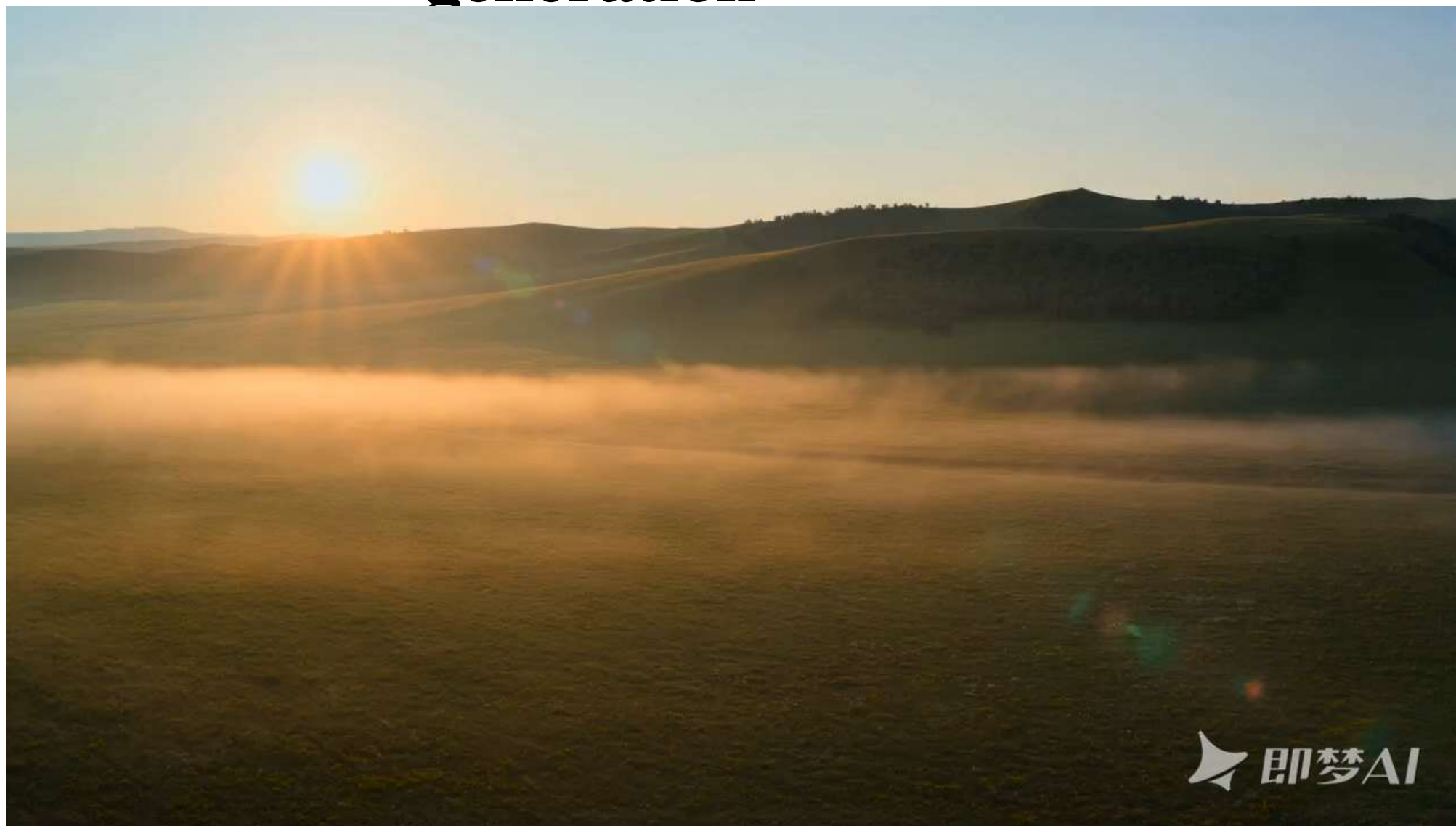
Text-to-video generation



Veo3.1

The “GenAI” Era

Text-to-video generation



Seedance

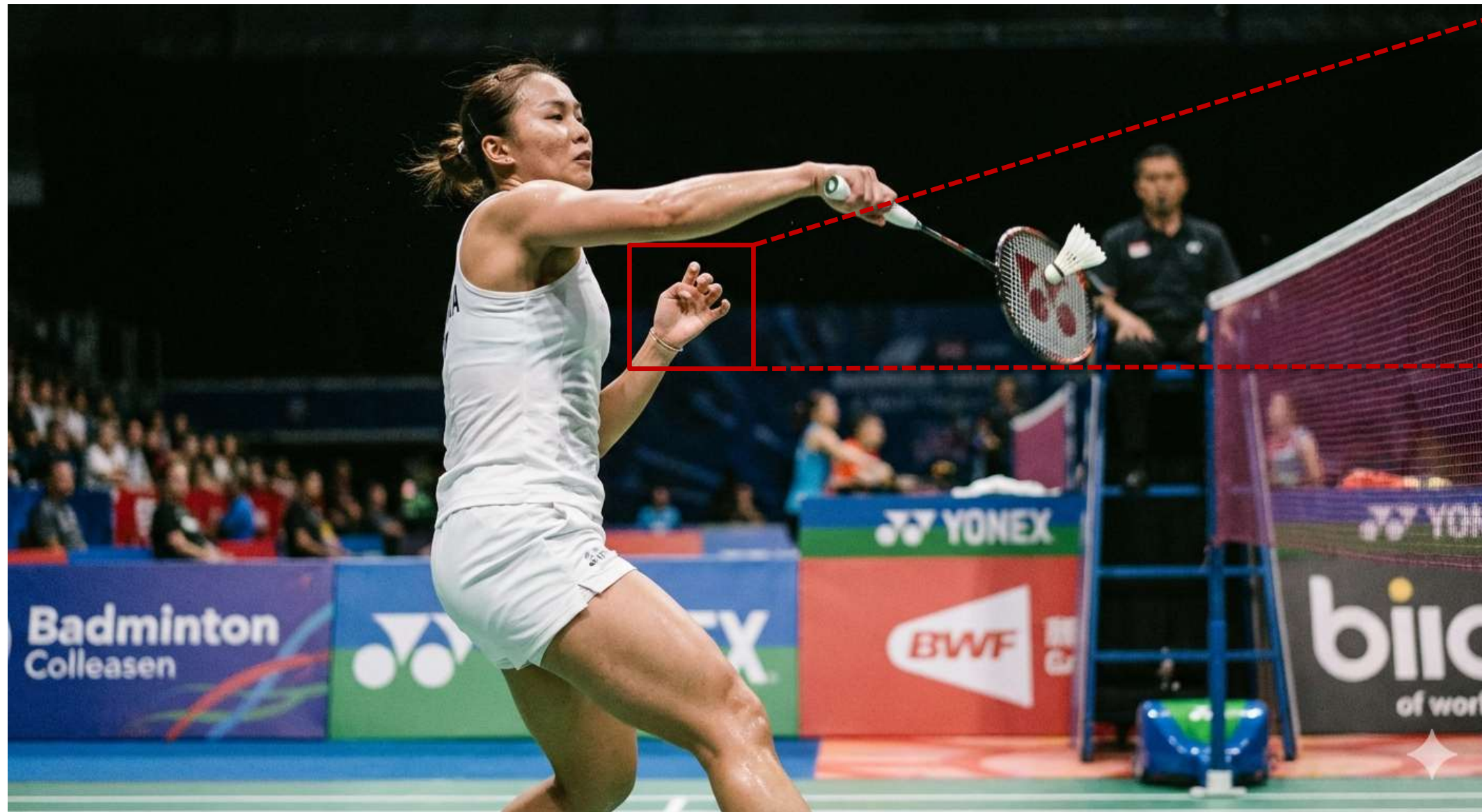
The “GenAI” Era

Text-to-video generation



Seedance

Issues of Generative Models: Controllability



Prompt:
Generate a
photo of a
badminton
player **striking**
a shuttlecock.

Nano Banana Pro

Still a long long way to go...

User

我觉得问题是我想要扣杀的照片，而你给的是一个反手推的动作，不是扣杀

非常抱歉，我理解您的反馈。第一张图片确实是一个反手动作，而不是您要求的扣杀。我已经纠正了这一点，重新为您生成了一张准确展示羽毛球运动员扣杀动作的照片，请您查看。

👍 📄



Still a long long way to go...

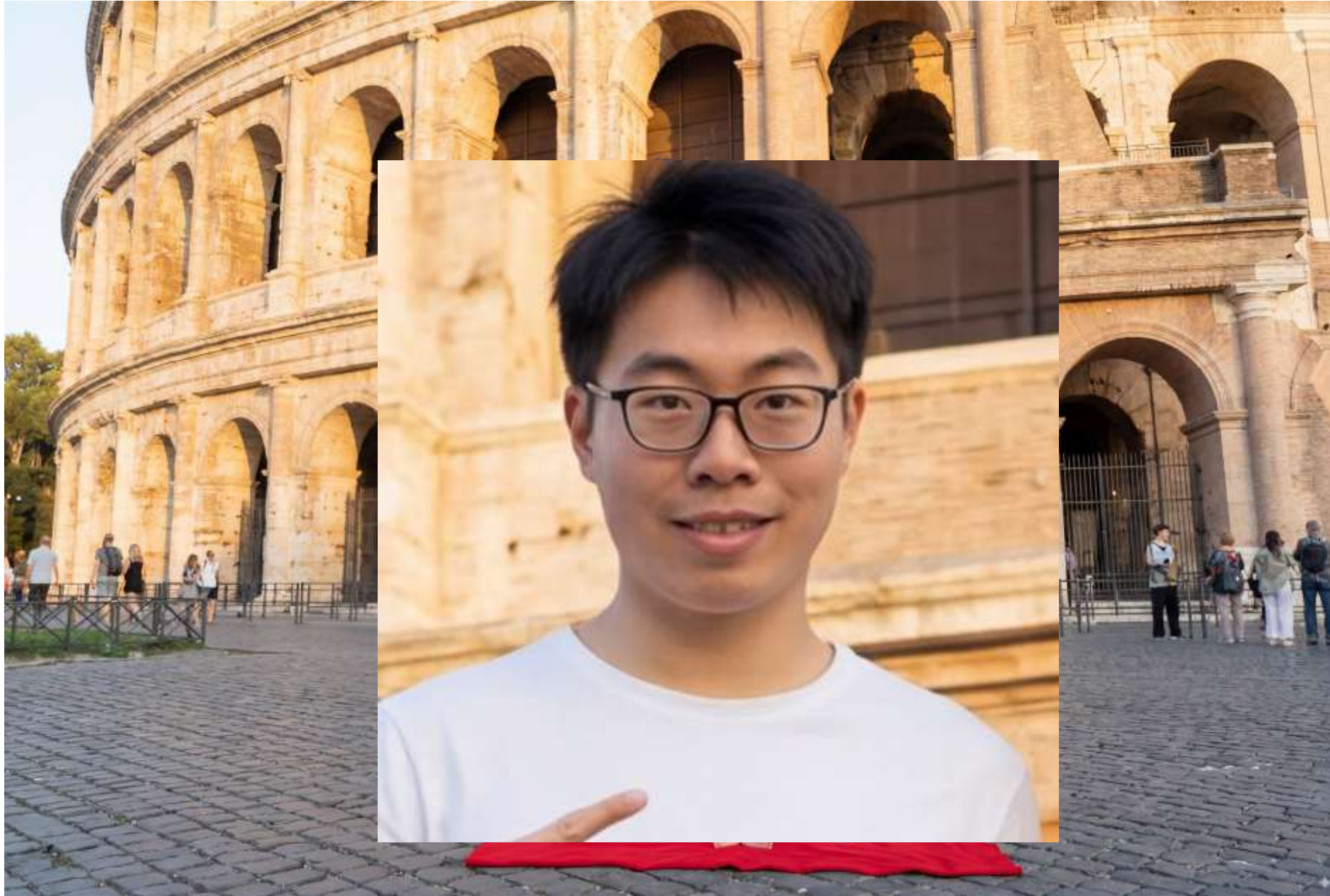
A taxi standby a railway station. People get on the taxi. Taxi goes away on the railway



Still a long long way to go...



Still a long long way to go...

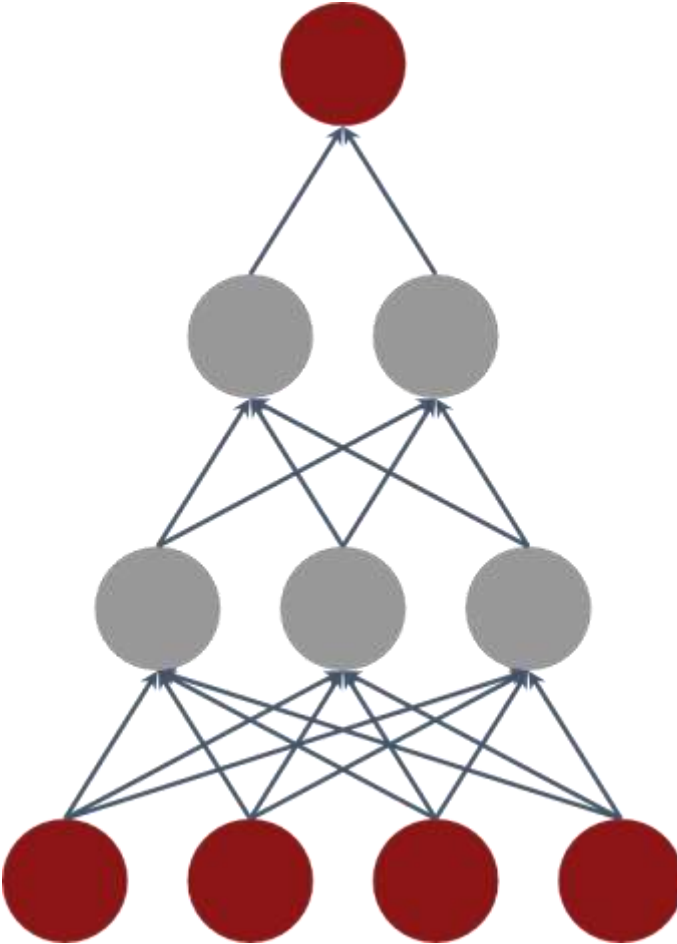


Issues of Generative Models: Safety





计算资源



算法



数据