

图像处理与视觉

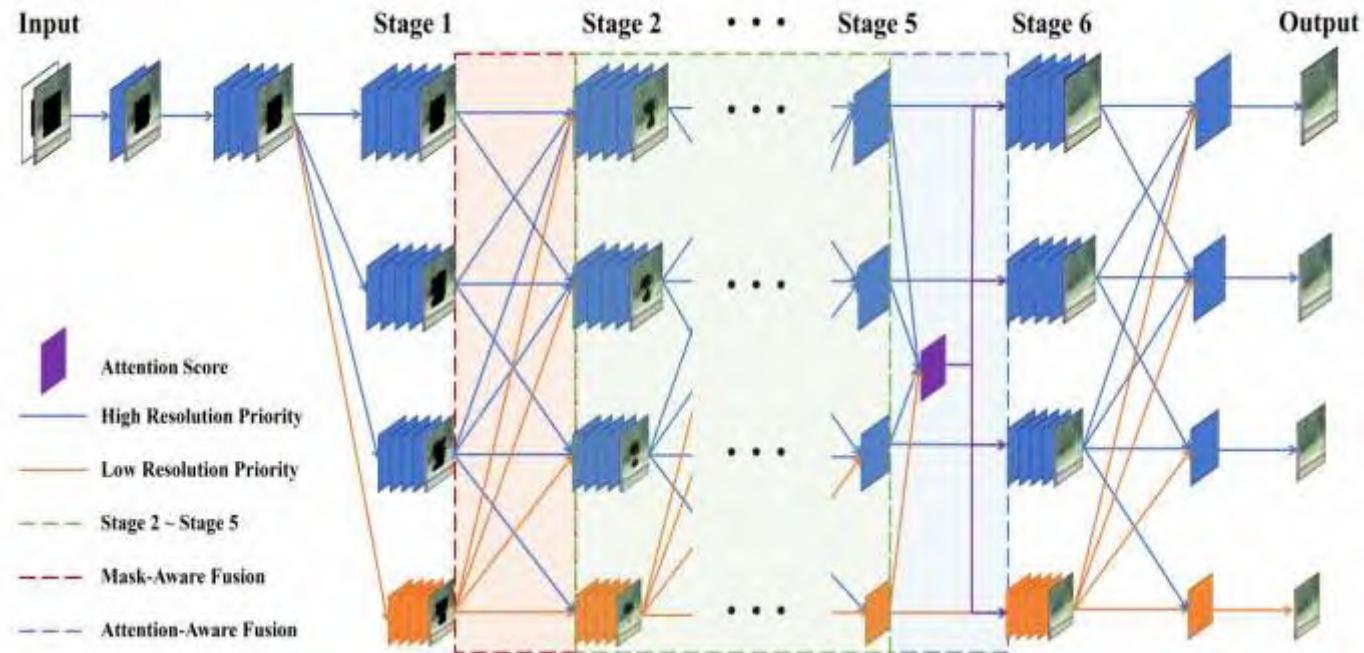
Agenda

- 个人信息介绍 (~2min)
- 课程大纲、作业及评分 (~2min)
- 课程内容介绍/导论 (~15min)
- Q&A

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 电院清源研究院
 - c.sis@sjtu.edu.cn <http://www.qingyuan.sjtu.edu.cn/a/zhang-jian-fu.html>
 - 研究方向：人工智能 | 视觉内容生成 / 深度模型可靠性
- 负责课程
 - John班 大三 专业选修课 《图像处理与视觉》
 - 23秋选课14/17, 评教档次A1, 学科排名81/999
 - 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

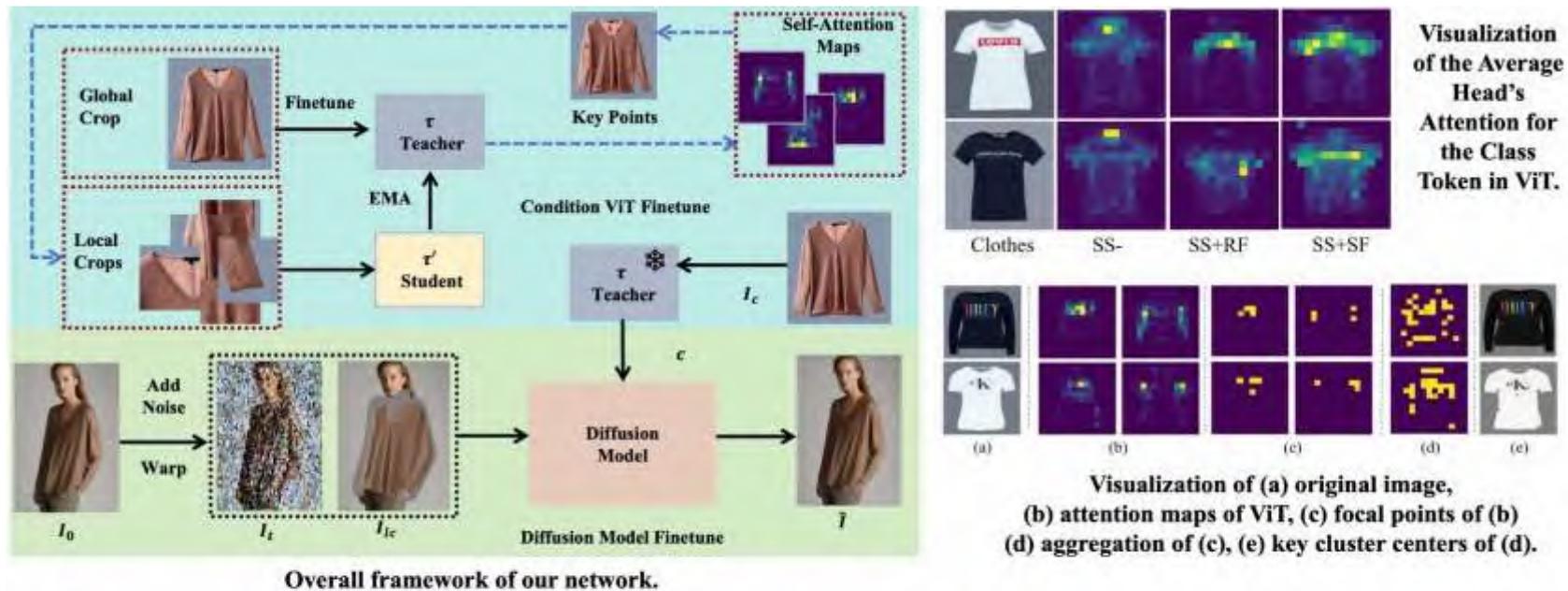


通过修复/编辑区域规划多视角/多尺度融合提升图像/视频修复编辑能力
合作开发 马卡龙玩图，获得Apple APP Store “最佳本土APP”

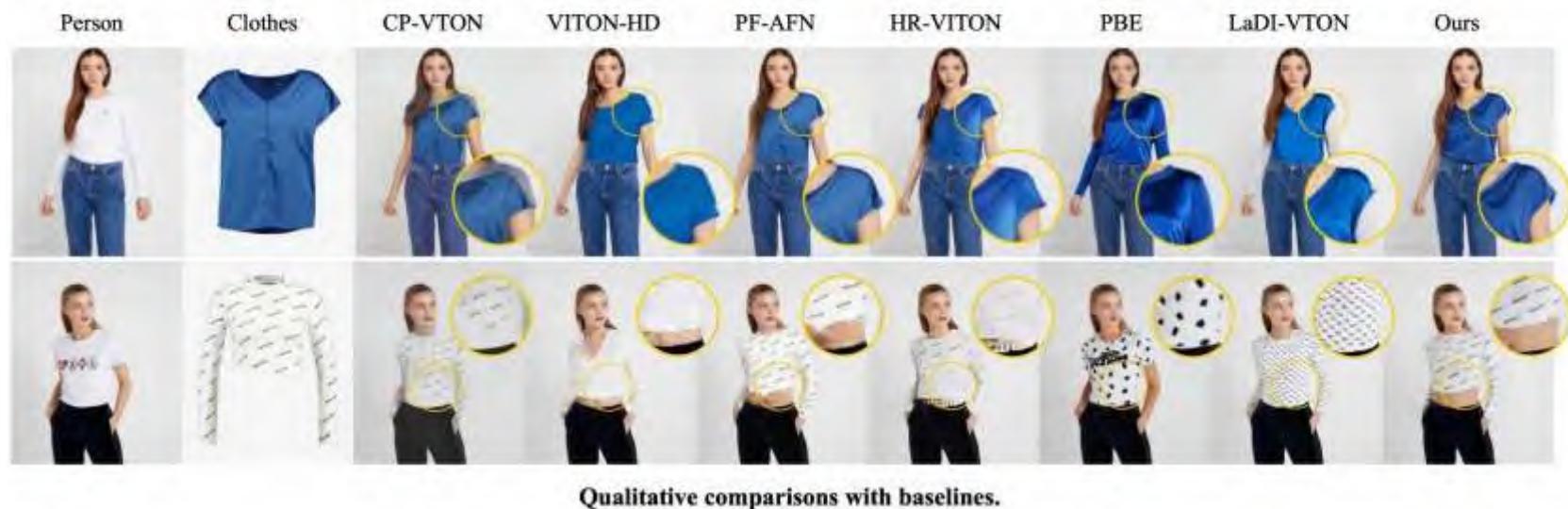
[ACM MM 2019] J Zhang, L Niu, D Yang, L Kang, Y Li, W Zhao, L Zhang. GAIN: Gradient Augmented Inpainting Network for Irregular Holes.
[ICCV 2021] *W Wang, *J Zhang, L Niu, H Ling, X Yang, L Zhang. Parallel Multi-Resolution Fusion Network for Image Inpainting.

[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.

近期工作：虚拟服装试穿



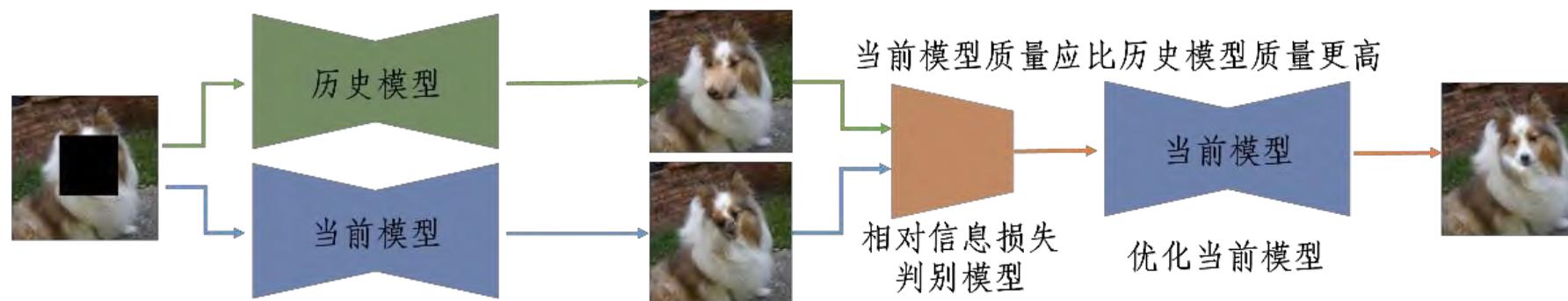
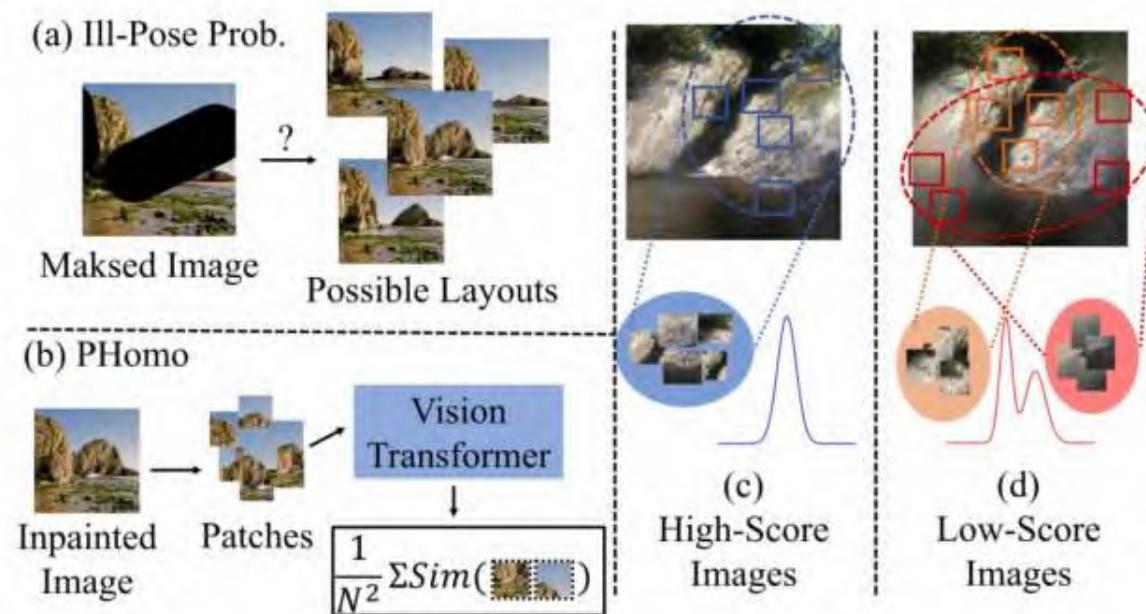
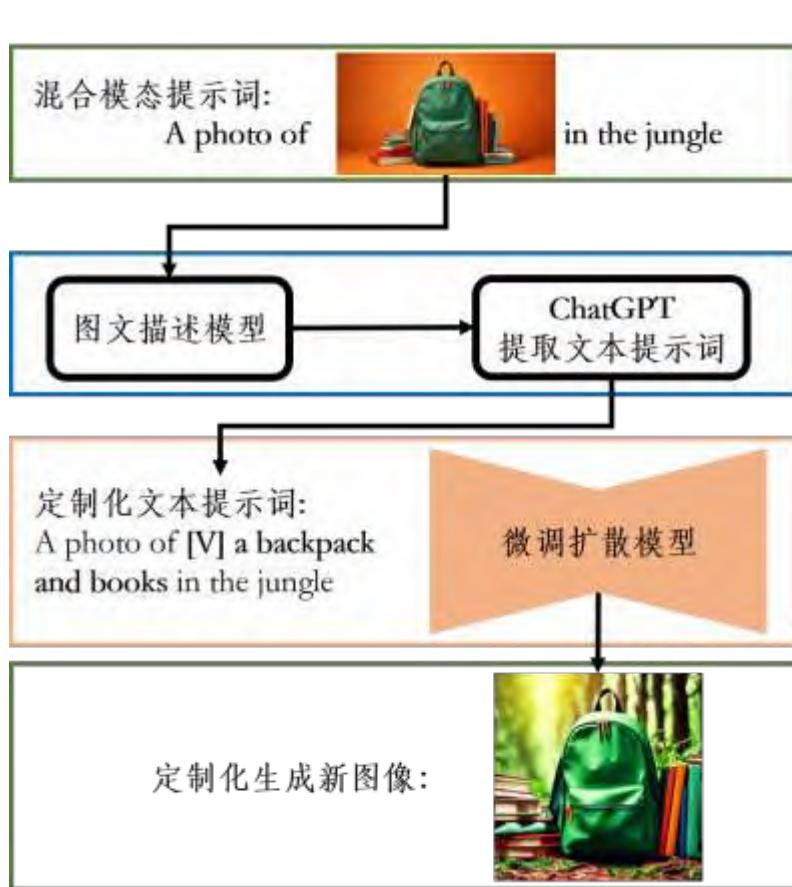
通过服装变形引导扩散模型生成高精度服装试穿图像，利用自监督ViT进一步提升对服装细节的处理。
公开代码在GitHub半年内迅速获得350+ star，为虚拟服装试穿领域目前最有影响力的公开工作。



[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

近期工作：可控生成/生成图像质量评估



自监督评价指标/生成可控性:
 通过评价指标自主学习优化生成模型，自动化的、用户为中心的高质量定制化生成模型。

Arxiv: User-Friendly Customized Generation with Multi-Modal Prompts;
 [ECCV 2024] ComFusion: Personalized Subject Generation in Multiple Specific Scenes From Single Image;
 [CIKM 2024] Assessing Image Inpainting via Re-Inpainting Self-Consistency Evaluation;
 Arxiv: No-Reference Image Inpainting Evaluation Via Patch Homogeneity Assessment

大纲、教材与参考课程

- 1/3传统视觉内容， 2/3深度视觉内容
- [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：图像缝合、目标检测、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%（满足要求即给足分）
 - 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）

课程内容简介

计算机视觉： 建立系统对视觉数据进行处理、感知、推理

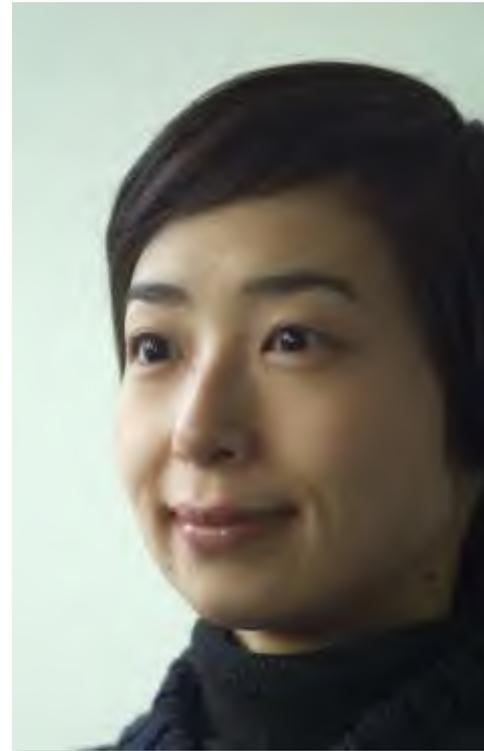
- 图像图形学
 - 我们怎么获得视觉信息
 - 无监督学习，更关注数学建模
- 视觉感知与理解
 - 我们能从视觉信息中获得什么
 - 监督学习，更关注数据驱动
- 视觉内容生成
 - 我们怎么创造新的视觉信息
 - 自监督学习，模型与数据兼备

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"

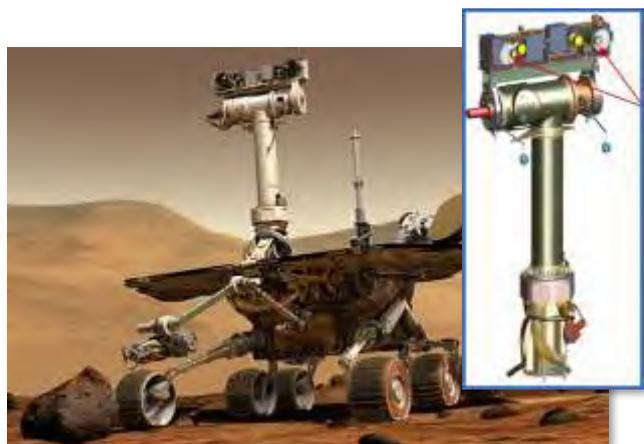
图像图形学的最终目标



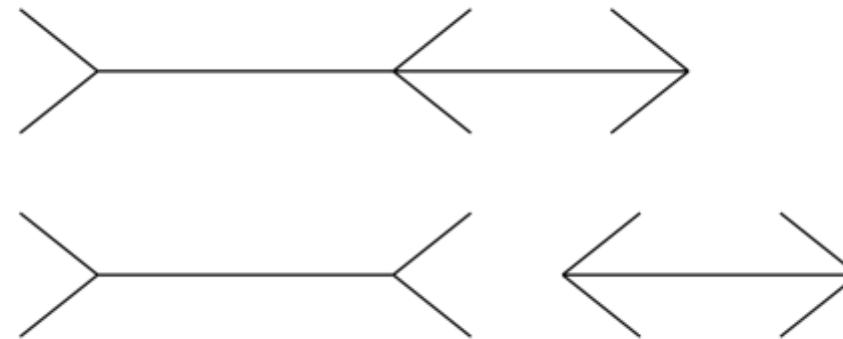
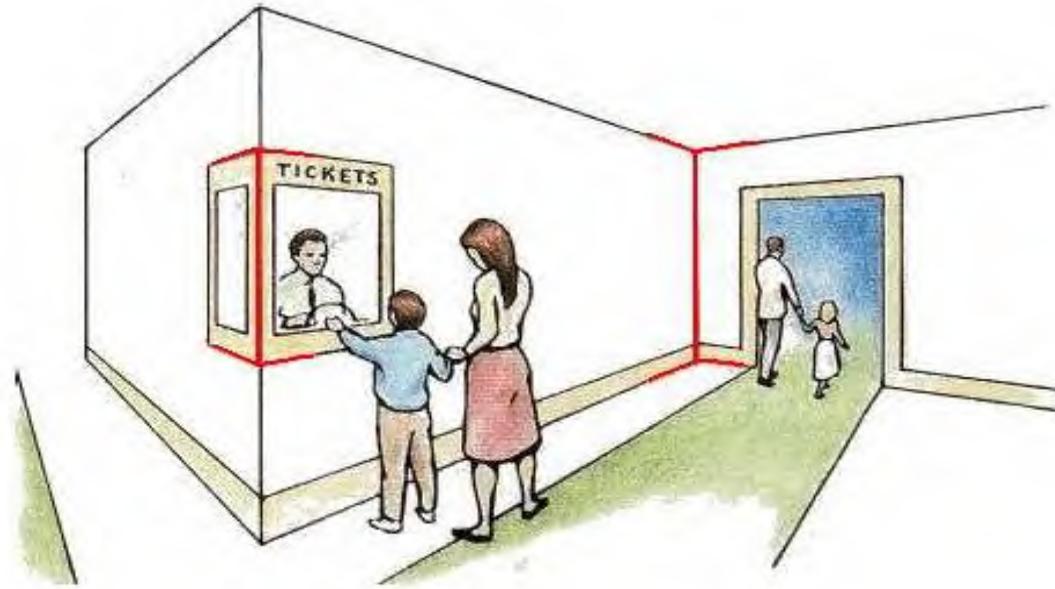
<https://www.bilibili.com/bangumi/play/ep28950?t=717>

《攻壳机动队》

怎样利用视觉信息对世界建模？

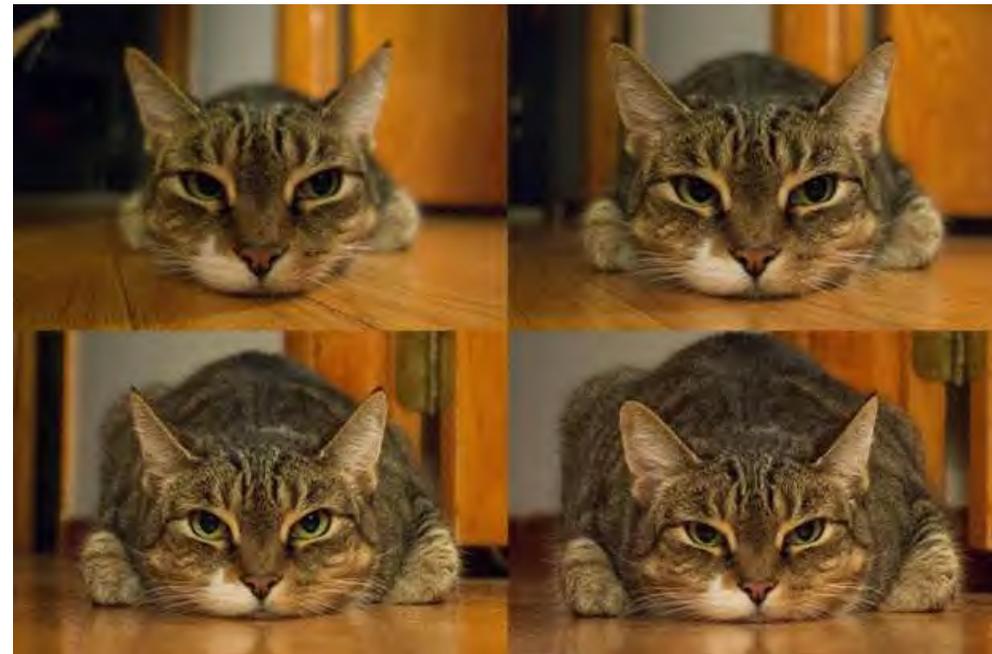


What You See (with Your Eyes) is What You Get?



怎样避免视觉错觉/误差?

提升拍照水平!



广角（短焦）



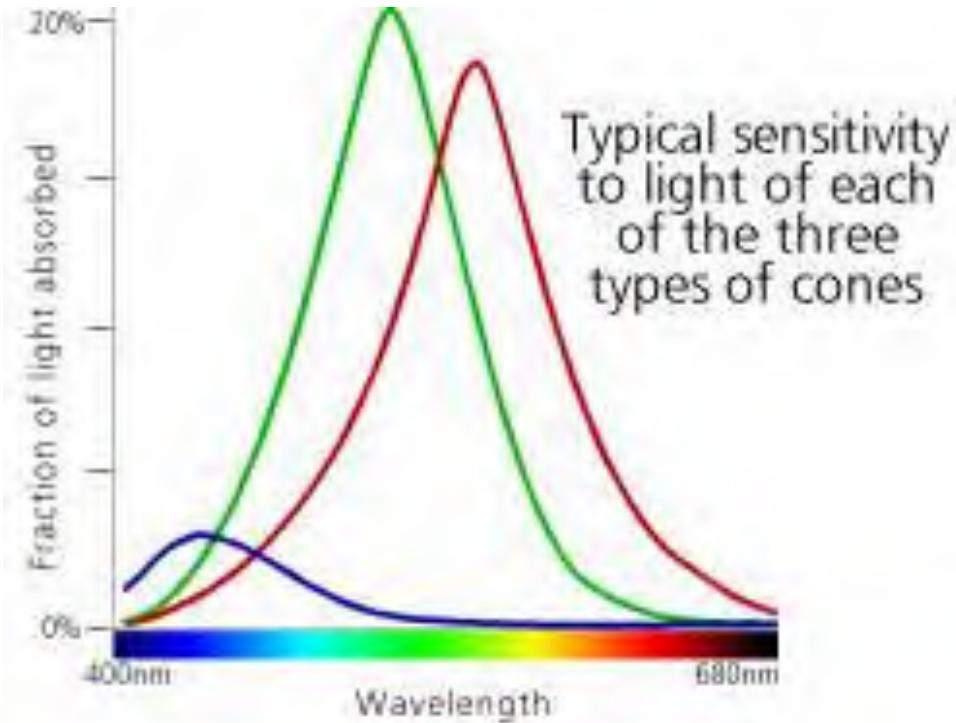
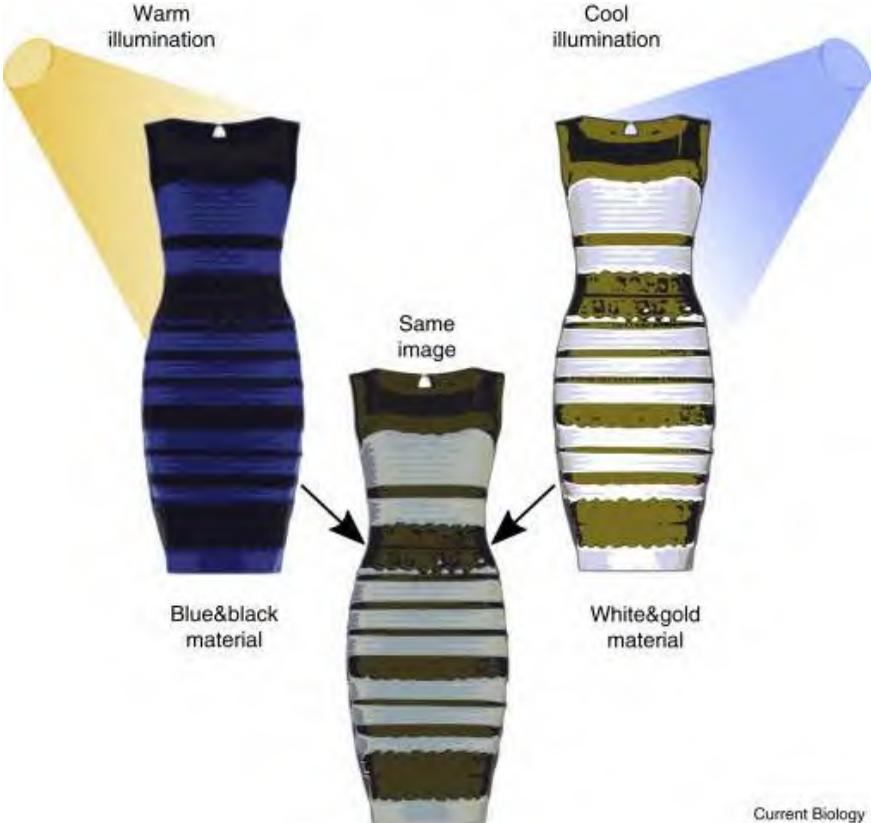
标准镜头



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



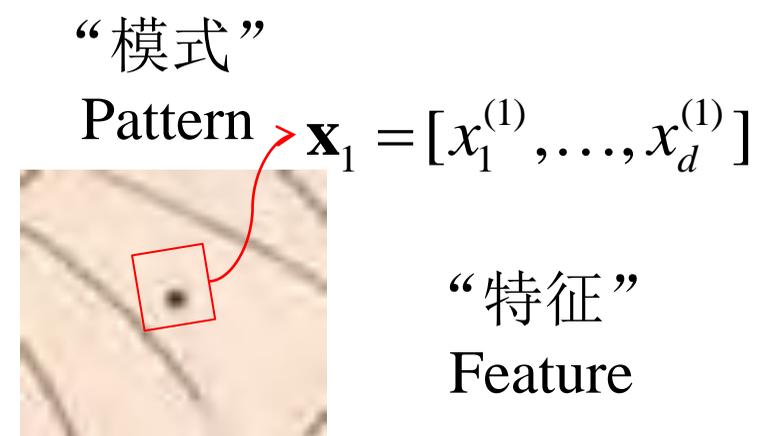
计算机如何感知？

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



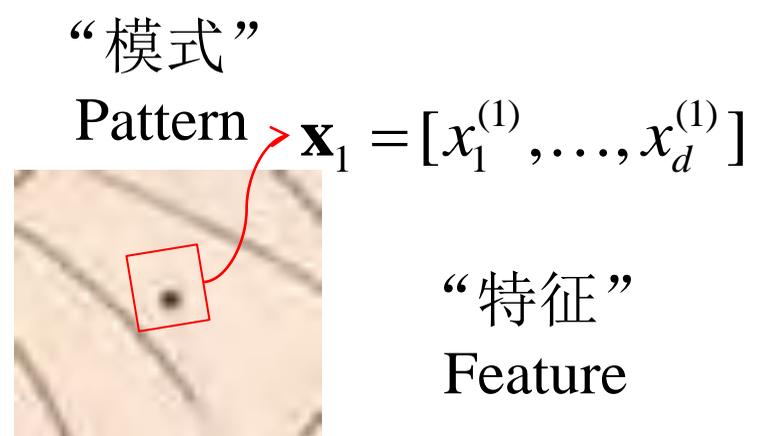
计算机如何感知？

- 让我们以人为例...
 - “音容笑貌”
 - “你**掌心**的**痣**，我总记得**在哪里**”

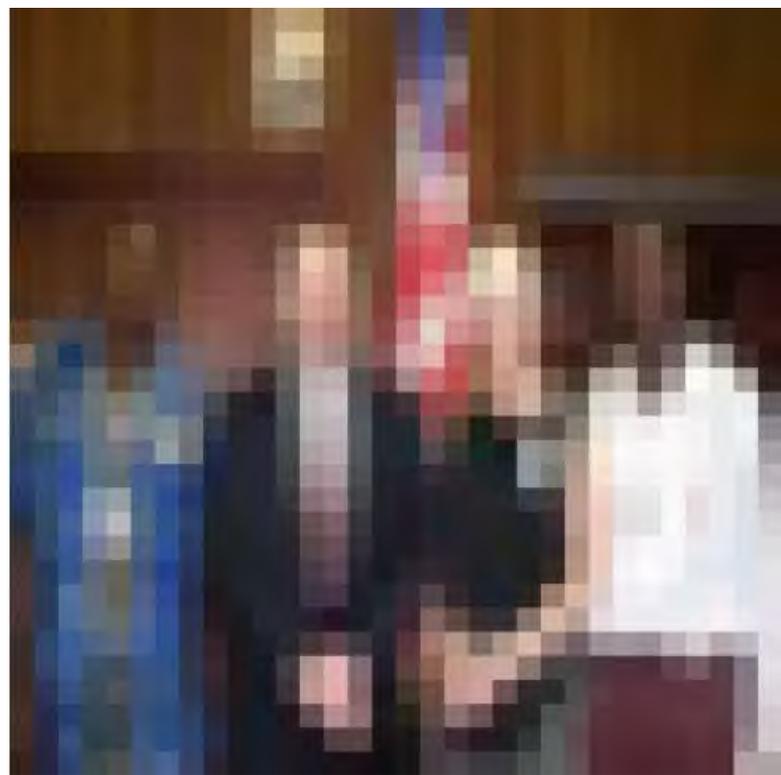


计算机如何感知？

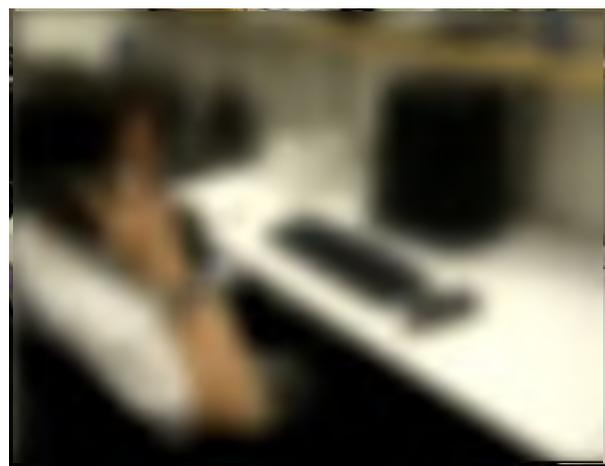
- 让我们以人为例...
 - “音容笑貌”
 - “你掌心的痣，我总记得在哪里”



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、
大量冗余特征

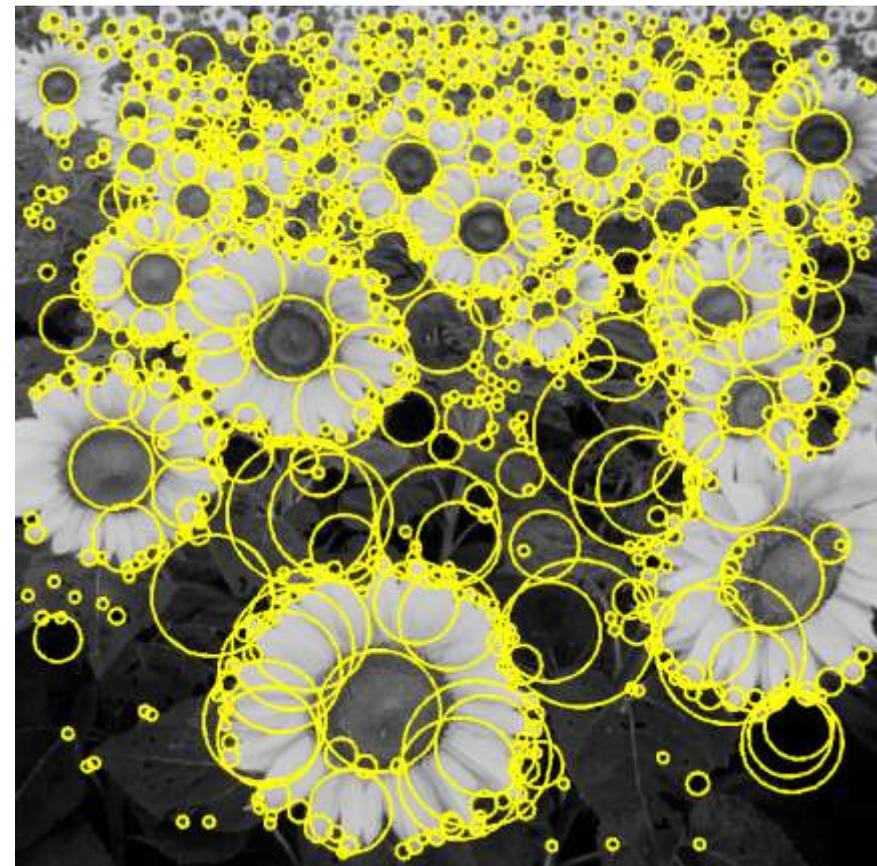
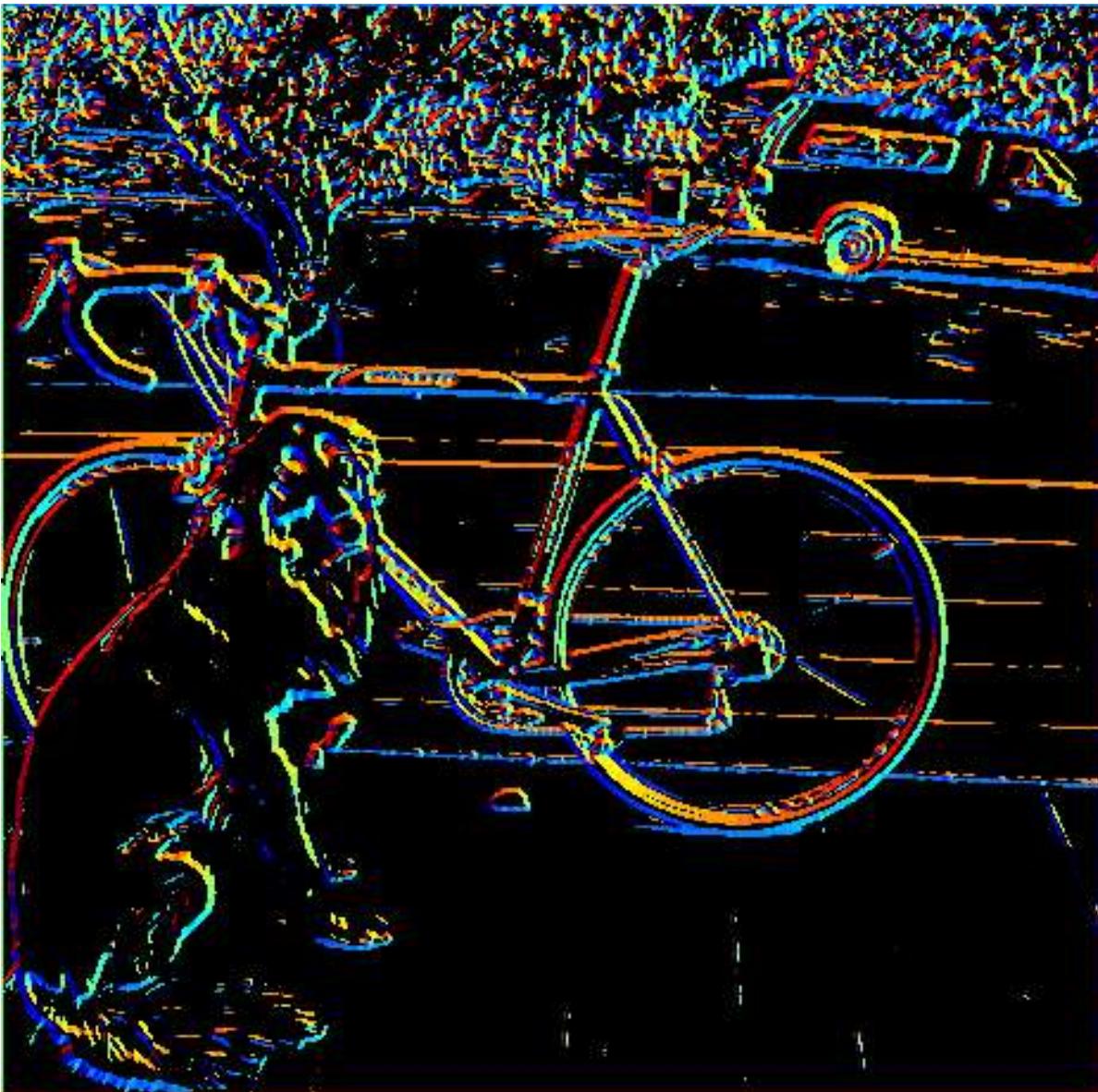


一个房子也许认不出来...

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



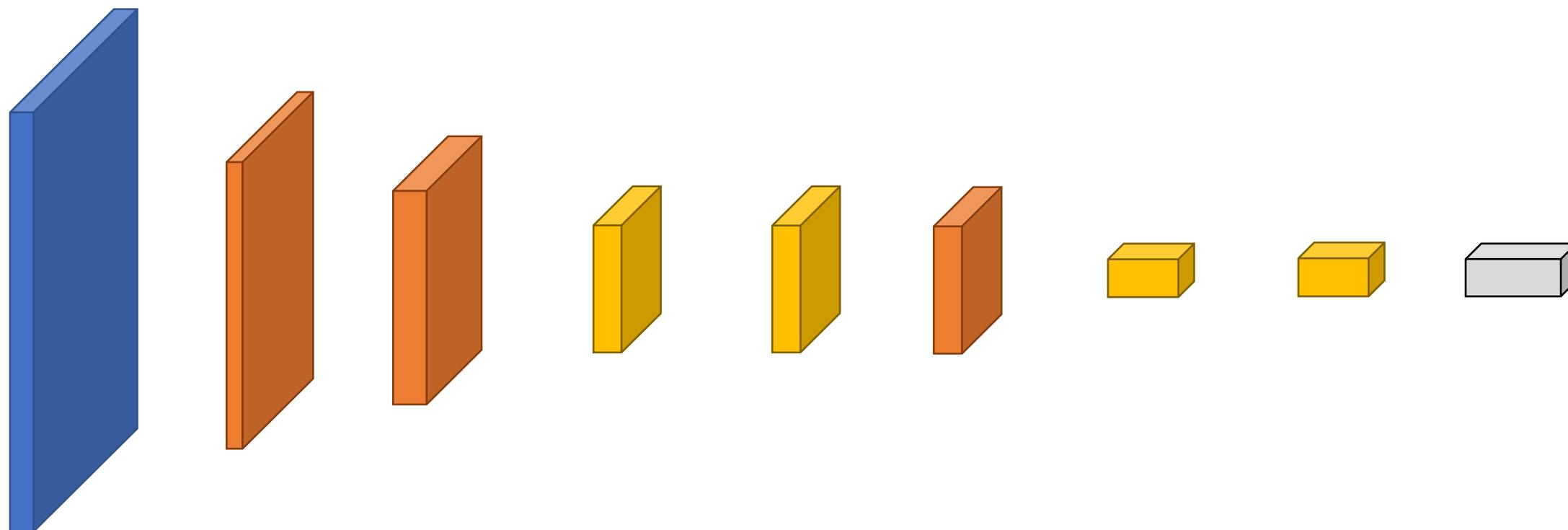
深度模型

更快速的推理判断

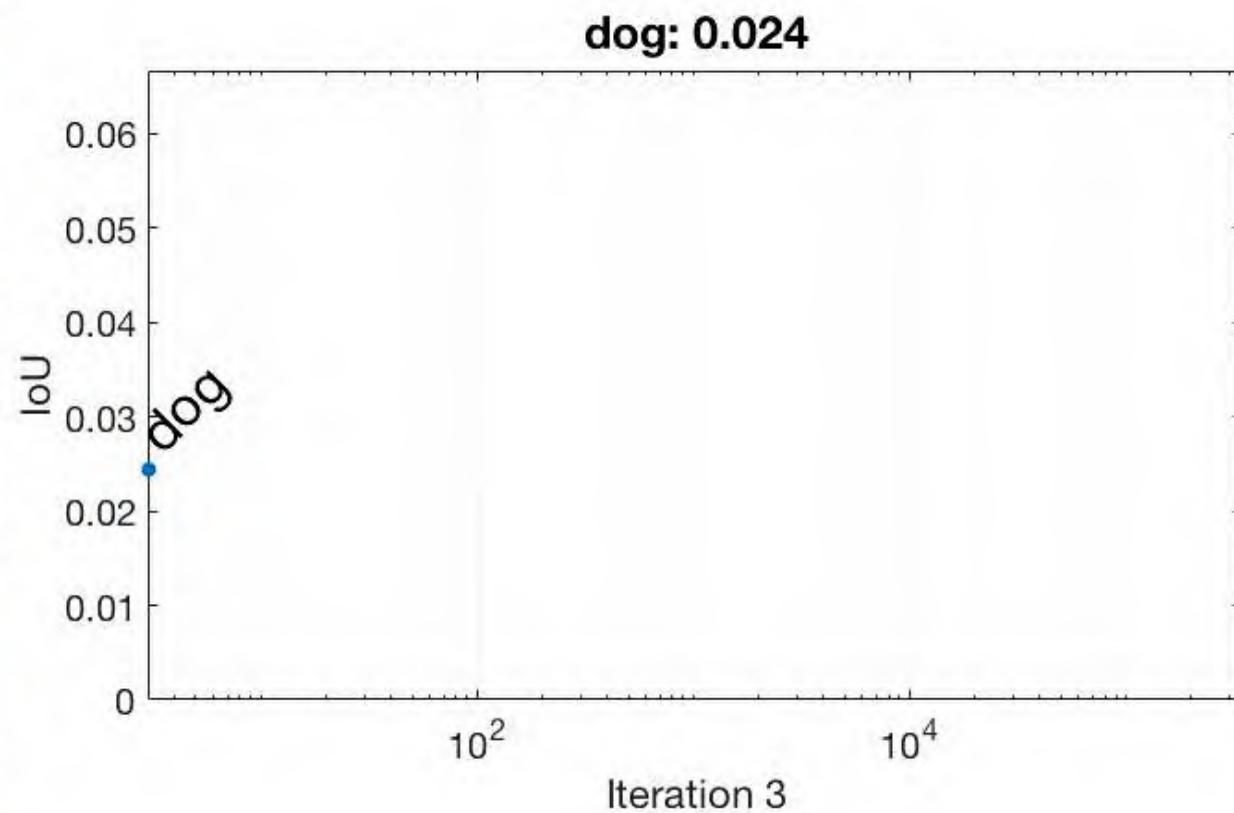
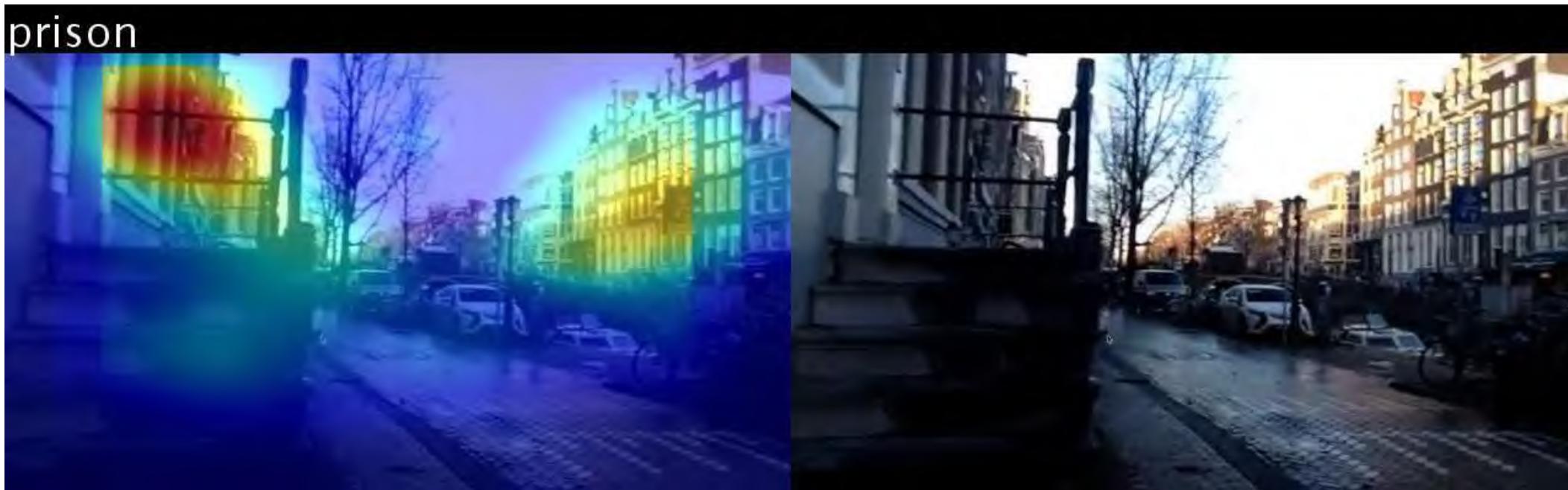
更强的普适性（泛化性）

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

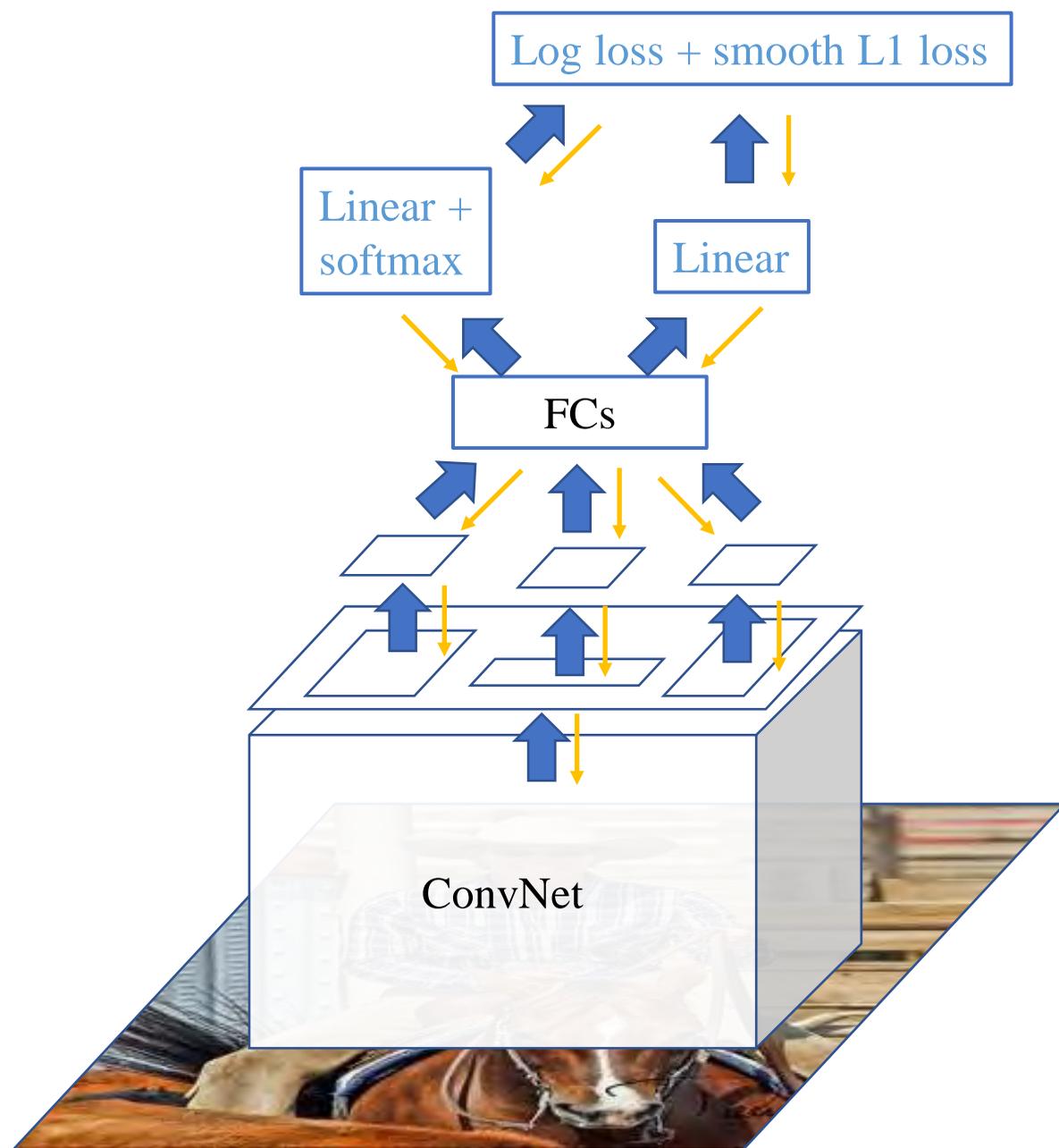


怎样理解/调试视觉深度模型

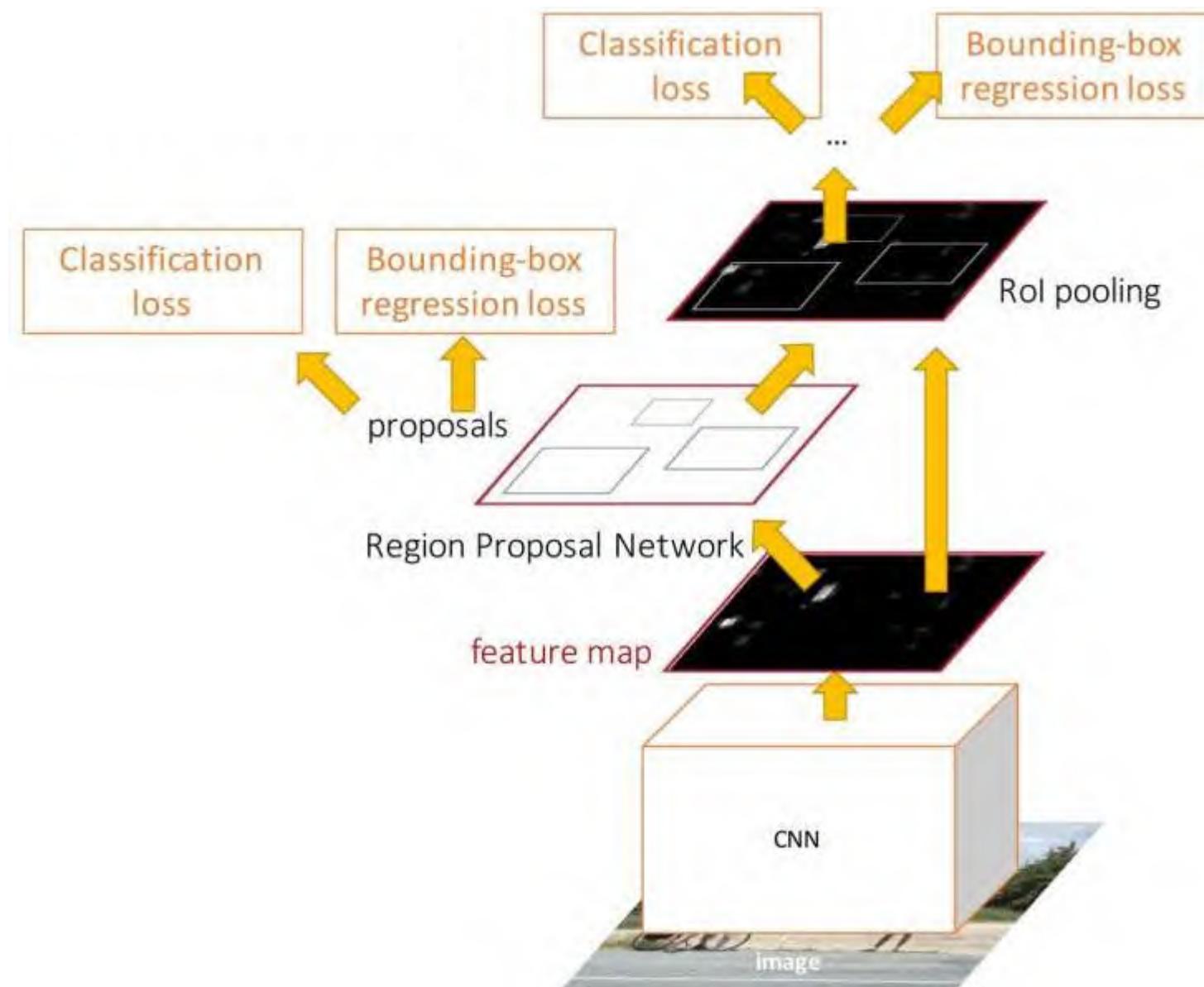


怎样利用深度模型完成视觉任务

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。

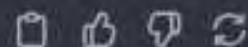
J You



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”)



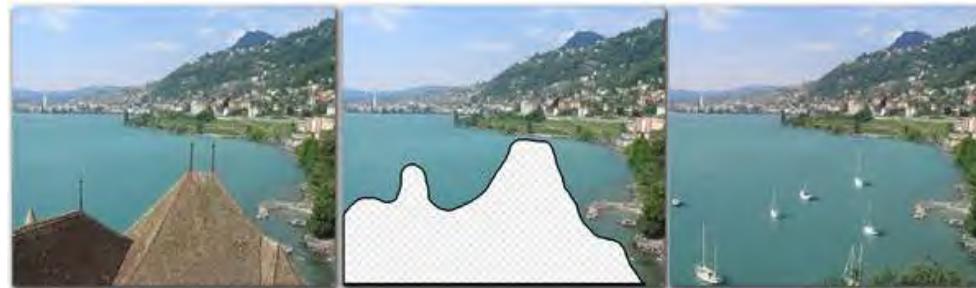
超分辨率 **放大，放大，再放大!**
(source: 2d3)



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”)



Person



Garment



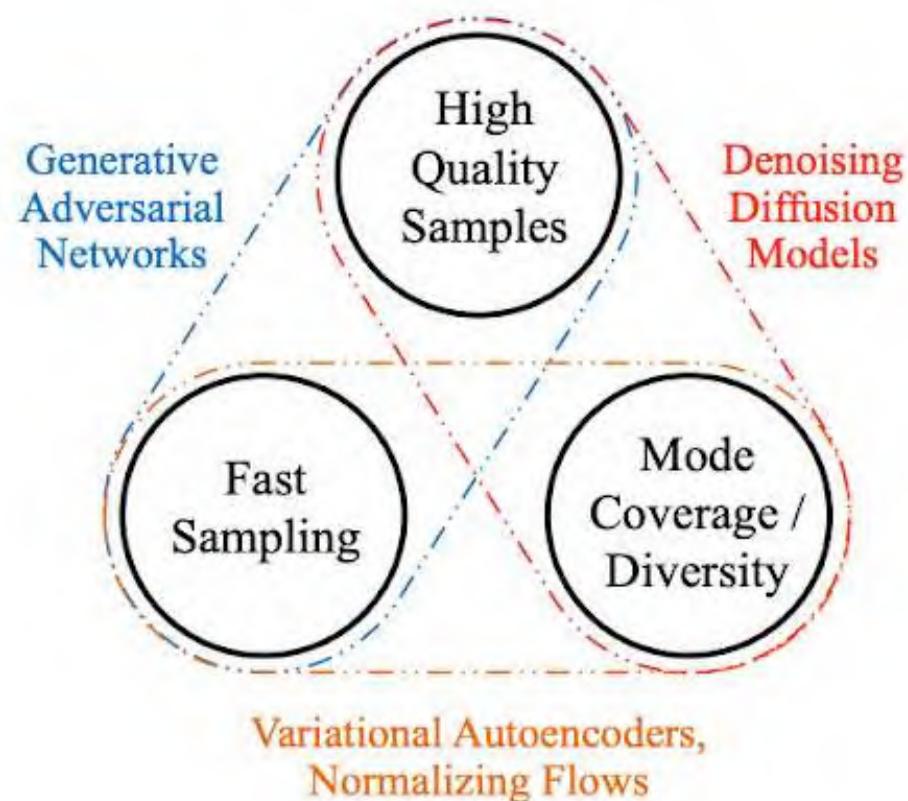
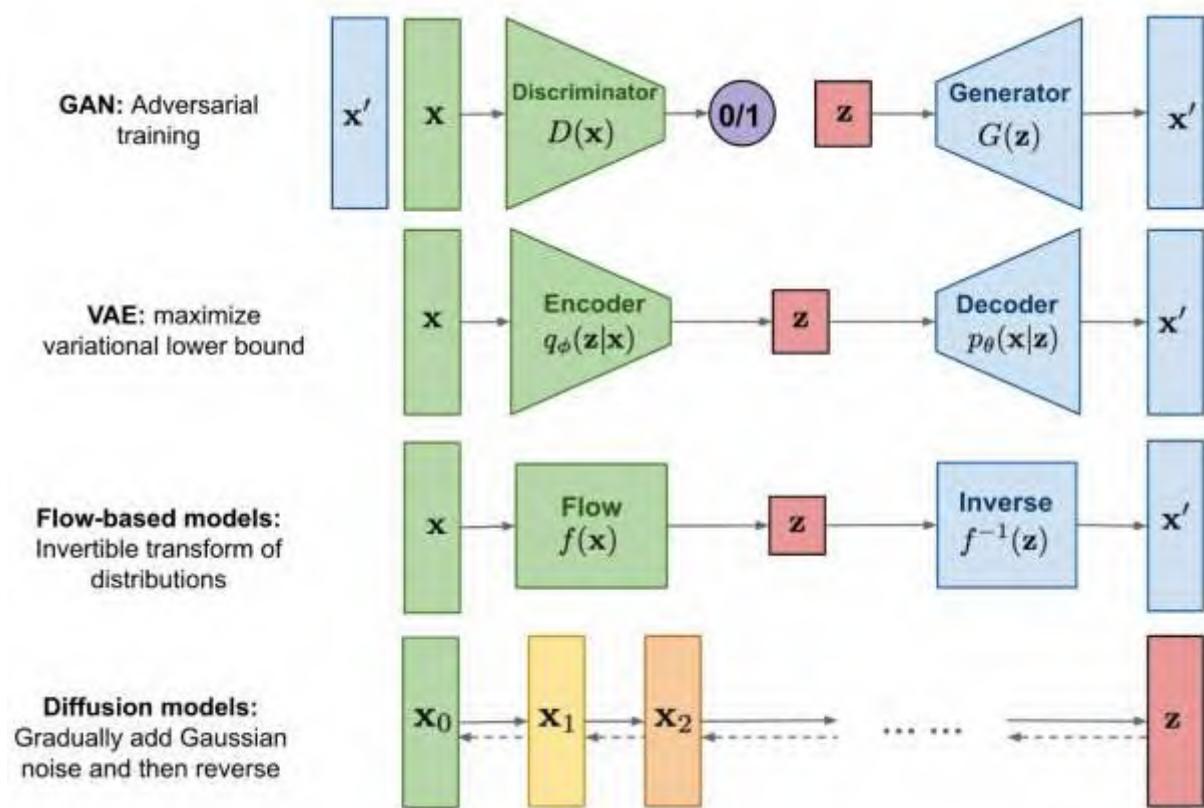
Try-on

Animation

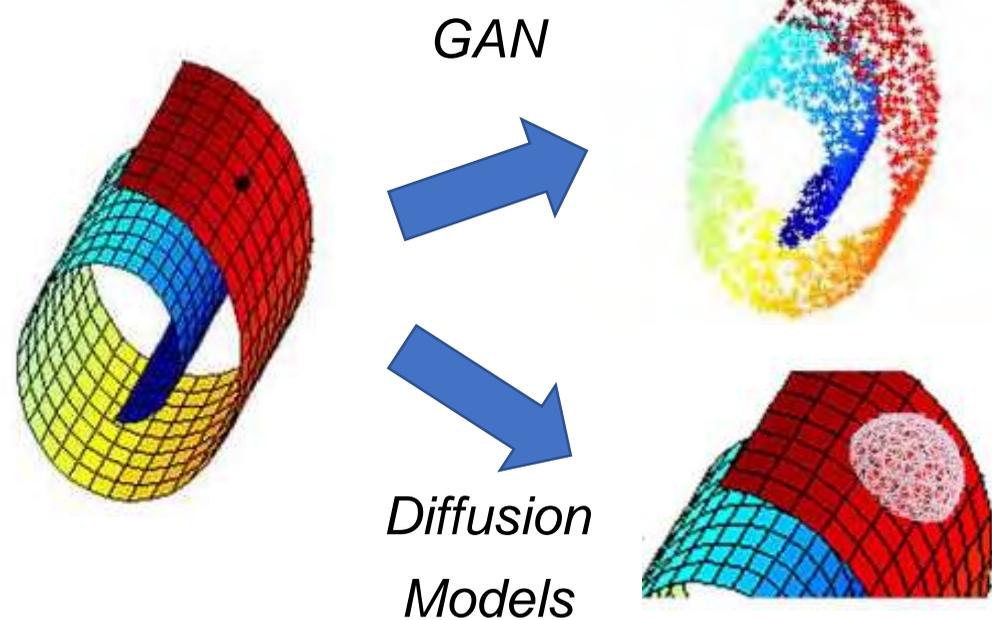
Outfit Anyone + Animate Anyone

层出不穷的生成模型

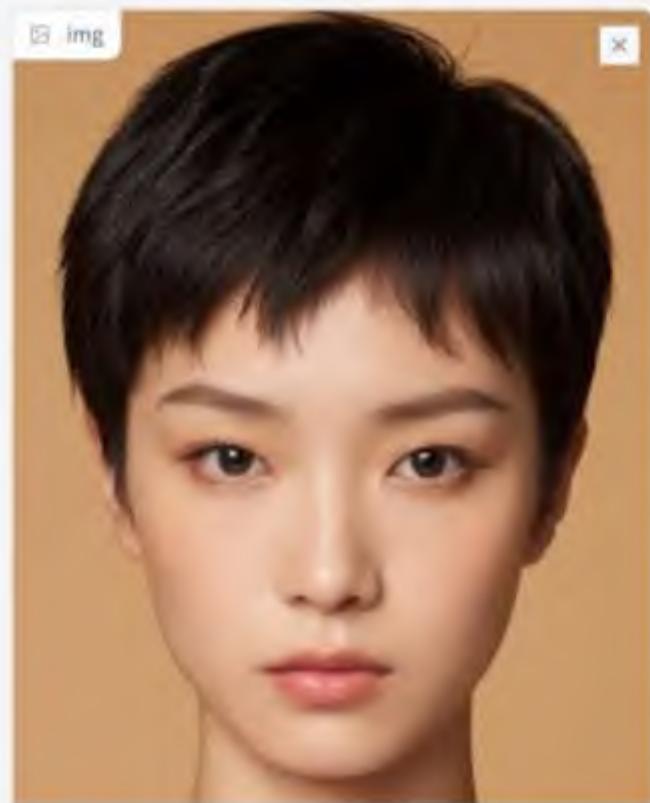
- 这些模型的基本原理是什么？优缺点是什么？



从GAN到Diffusion Models



扩散模型拥有更难以想象的生成能力！



output

Real Image

Real Image	99%
Fake Image	1%



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？



Q&A