



# Lecture 1. **Introduction to Pattern Recognition and Computer Vision (PRCV)**

**Computer Vision**

Guanbin Li,

School of Computer Science and Engineering, Sun Yat-Sen University

# 扫码签到

---



# Instructor Introduction

---



- Associate Professor
- PhD from HKU
- Research Interest: Computer Vision (Visual Understanding, Machine Learning, Cross-modal visual analysis, Generative AI)  
<http://guanbinli.com>, <https://cse.sysu.edu.cn/content/2687>
- Our Lab: <http://www.sysu-hcp.net/>
- Email: [liguanbin@mail.sysu.edu.cn](mailto:liguanbin@mail.sysu.edu.cn)
- Office: 东校区南基础实验楼D302

TAs:



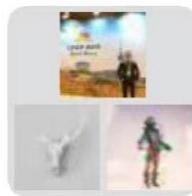
张嘉诚  
硕士二年级  
研究方向：半监督物体检测



吴振华  
硕士二年级  
研究方向：医学影像三维重建

# Join in....

---



群聊: 20 级-PRCV 课程



# Prerequisites

---

- Proficiency in Python (NumPy)
  - Linear Algebra
  - Probability and Statistics
  - Calculus
- 

## Textbook Reference

- Computer Vision: A Modern Approach by Forsyth & Ponce
  - Computer Vision: Theory and Application - Rick Szeliski 2010
- 

## Course Reference

[http://vision.stanford.edu/teaching/cs131\\_fall2021/index.html](http://vision.stanford.edu/teaching/cs131_fall2021/index.html)

<http://cs231n.stanford.edu/>

# Doing Research in PRCV

---

- Proficiency in Python (NumPy)
  - Linear Algebra
  - Calculus
  - Probability and Statistics
- 

Top Conference and Journal: CVPR/ICCV/ECCV/TPAMI/TIP

Arxiv Computer Vision: <https://arxiv.org/list/cs.CV/recent>

---

Deep learning based CV:

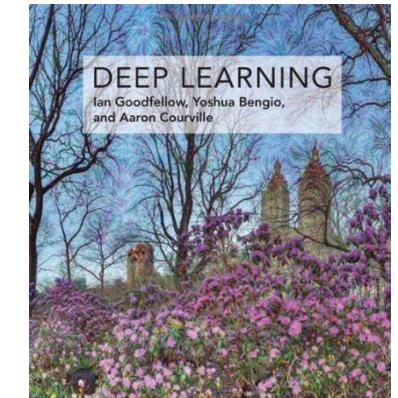
<http://cs231n.stanford.edu/>

Open source framework: Pytorch, tensorflow, caffe...

Learning material:

[https://pan.baidu.com/s/1CbNv7OtAhEKFDmEmUY0f\\_g](https://pan.baidu.com/s/1CbNv7OtAhEKFDmEmUY0f_g)

提取码: 4wr5



# Schedule and Syllabus

---

- L1. Introduction to PRCV
- | L14. Image Classification with Linear Classifiers
- L2. Pixels and Filters
- | L15. Regularization and Optimization
- L3. Filters and convolutions
- | L16. Neural Networks and BP
- L4. Images and transformations
- | L17. Image Classification with CNNs
- L5. Edge detection
- | L18. CNN Architectures
- L6. Features and fitting
- | L19. Training Neural Networks
- L7. Feature descriptors
- | L20. Visualization and Understanding
- L8. Resizing
- | L21. Object Detection and Image Segmentation
- L9. Clustering and Segmentation
- | L22. Recurrent Neural Networks
- L10. Dimensionality Reduction
- | L23. Attention and Transformers
- L11. Face Identification
- | L24. Video Understanding
- L12. Visual Bag of Words
- | L25. Generative Models
- L13. Detecting objects by parts
- | L26. Self-supervised Learning

# Grading

---

- 平时成绩: 40%
  - 期末大作业: 60%
- 
- 平时成绩: 课堂签到 10%; 3次平时实验小作业 30% (锻炼实践能力)
  - 期末大作业 60%: 从26个主题中选择其中一个深入理解, 实验, 并撰写相关技术的小型综述、算法研究或算法应用 (三选一) 论文

# What is PRCV, ChatGPT says...

Li

请简要描述一下计算机视觉是什么？



计算机视觉是一门涉及计算机科学、数学和工程学的跨学科领域，其目标是让计算机像人一样能够解析和理解图像和视频数据。具体而言，计算机视觉致力于开发算法和技术，使得计算机可以对图像或视频数据进行处理、识别、分类、跟踪和重建等任务，以实现人机交互、智能感知和自主决策等功能。



计算机视觉涉及多个方面，如图像处理、模式识别、机器学习、人工智能等，其应用范围广泛，包括自动驾驶、人脸识别、安防监控、医疗影像分析、虚拟现实等领域。随着深度学习等新技术的不断发展，计算机视觉领域也在不断推陈出新，为我们带来了更多的可能性和应用场景。

# What is PRCV, ChatGPT says...

LI

请简要描述一下模式识别是什么？



模式识别是计算机科学、数学和统计学等多学科交叉领域的一部分，其目的是研究如何使用计算机和数学方法来对数据进行分类、聚类、分割、识别、检测等任务，以找到数据中的规律、模式和结构。模式识别在许多领域都有应用，如计算机视觉、自然语言处理、生物医学、金融风控等。



模式识别通常涉及以下步骤：

1. 数据预处理：对数据进行清洗、过滤、转换等操作，以便后续的处理和分析。
2. 特征提取：从数据中提取有意义的特征，用来描述数据的属性和特性。这是模式识别中非常重要的一步。
3. 特征选择：对提取出的特征进行选择和降维，以提高算法的效率和准确度。
4. 模型训练：使用已经标注好的数据，利用机器学习等方法训练出模型，以实现对新数据的预测和分类。
5. 模型评估：对训练好的模型进行评估和测试，以确保其在新数据上的泛化能力和准确度。

在模式识别中，常用的技术包括支持向量机、决策树、神经网络、随机森林等。模式识别的应用非常广泛，例如人脸识别、语音识别、手写字识别、金融欺诈检测等领域。

# What is PRCV, ChatGPT says...

LI

模式识别中的模式指的是什么



在模式识别领域中，“模式”通常指代数据中的重复、规律性或特定的形式。例如，一个手写数字“5”就是一个模式，一个人脸的轮廓也可以看做一个模式，一段语音信号的频谱特征也可以被视为一个模式。这些模式可以用来描述和区分不同的数据，如手写数字的分类、人脸的识别等。

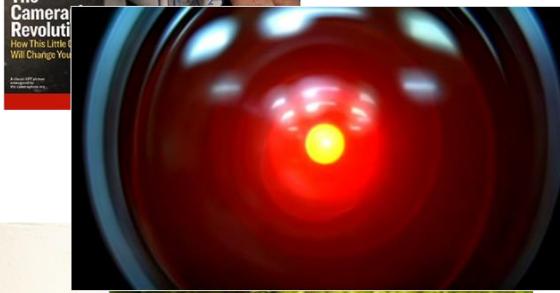


在模式识别中，模式通常是由许多特征组成的，如数字“5”可以由其笔画的轮廓线和关键点等特征来描述。因此，在模式识别中，特征提取和选择也是非常重要的步骤，以便正确地描述和区分不同的模式。

总之，模式识别中的“模式”可以指数据中的各种重复、规律性或特定形式，而模式识别的目的则是从数据中提取这些模式，并对其进行分类、聚类、识别等任务。

# Computer vision is everywhere

Computer Vision: to realize human visual functions (the perception, recognition and understanding of three-dimensional scenes in the objective world)

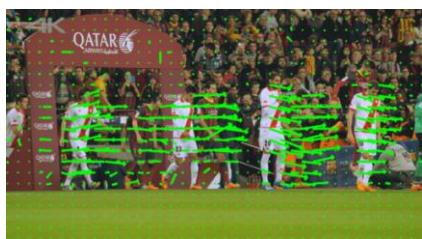


# Computer vision is everywhere

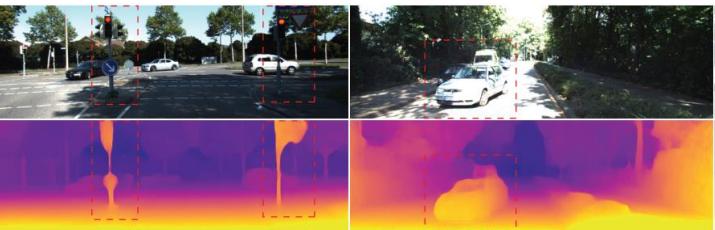
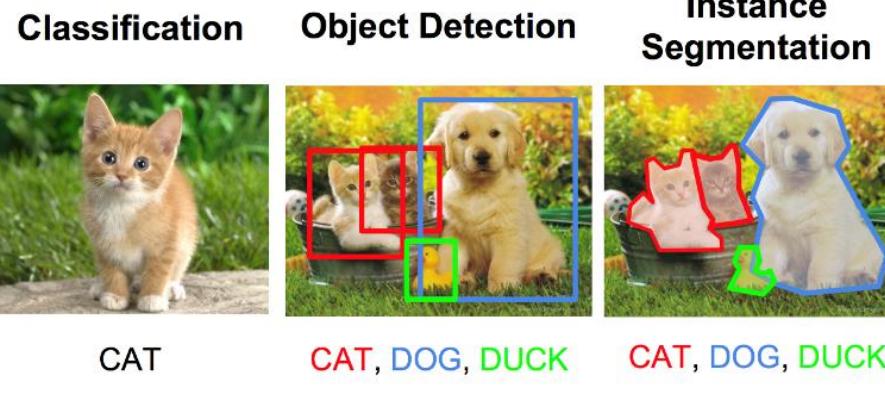
## Low-level Vision



## Mid-level Vision



## High-level Vision



# Origins of computer vision

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
PROJECT MAC

Artificial Intelligence Group  
Vision Memo. No. 100.

July 7, 1966

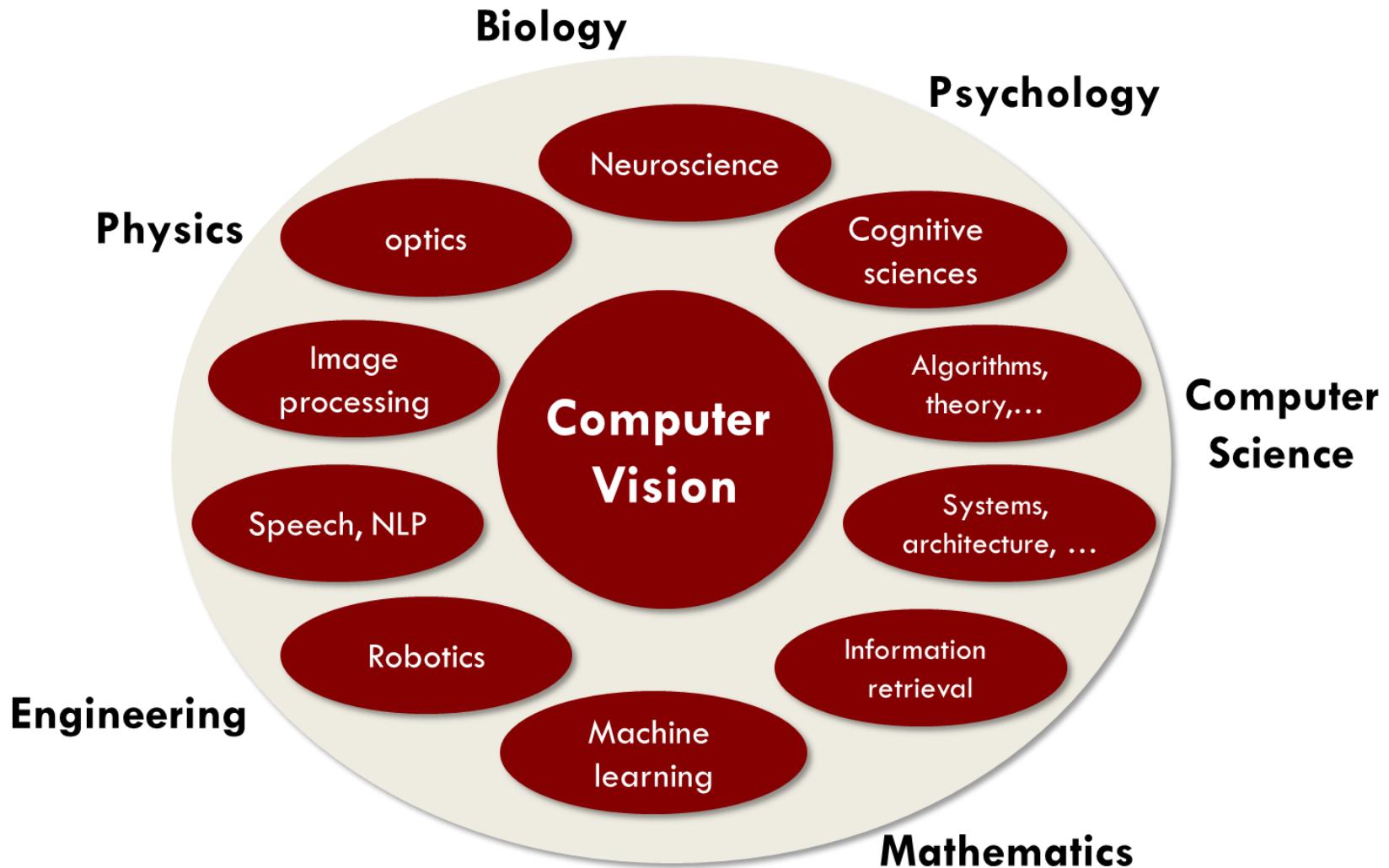
## THE SUMMER VISION PROJECT

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".

an MIT undergraduate summer project

# Introduction to computer vision



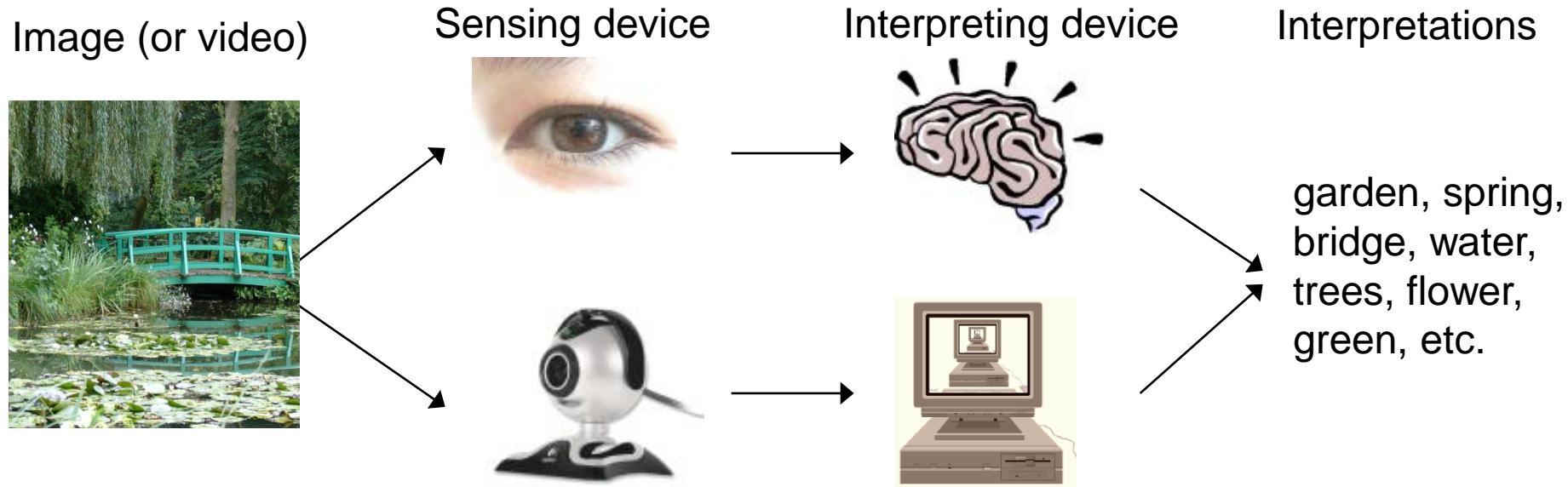
# Today's agenda

---

- Introduction to computer vision
- Course overview

# What is (computer) vision?

视觉是指人类和动物通过眼睛感知和理解外部世界的过程。在视觉过程中，眼睛接收到光线并将其转化为电信号，这些信号然后通过神经系统传递到大脑中的视觉中枢，被加工和解释成有意义的视觉信息。



# Every image tells a story

---



- Vision: “To see is to know what is where by looking”
- Computer Vision: To design a computer model of vision.
- Computer Vision: Part of artificial intelligence (AI)
- Turing Test of Computer Vision: To answer any question regarding an image, what can be answered by a human.



# The goal of computer vision

To bridge the gap between pixels and “meaning”

Enable computers to have human-like visual perception, allowing them to extract information, analyze, and understand it from image and video data

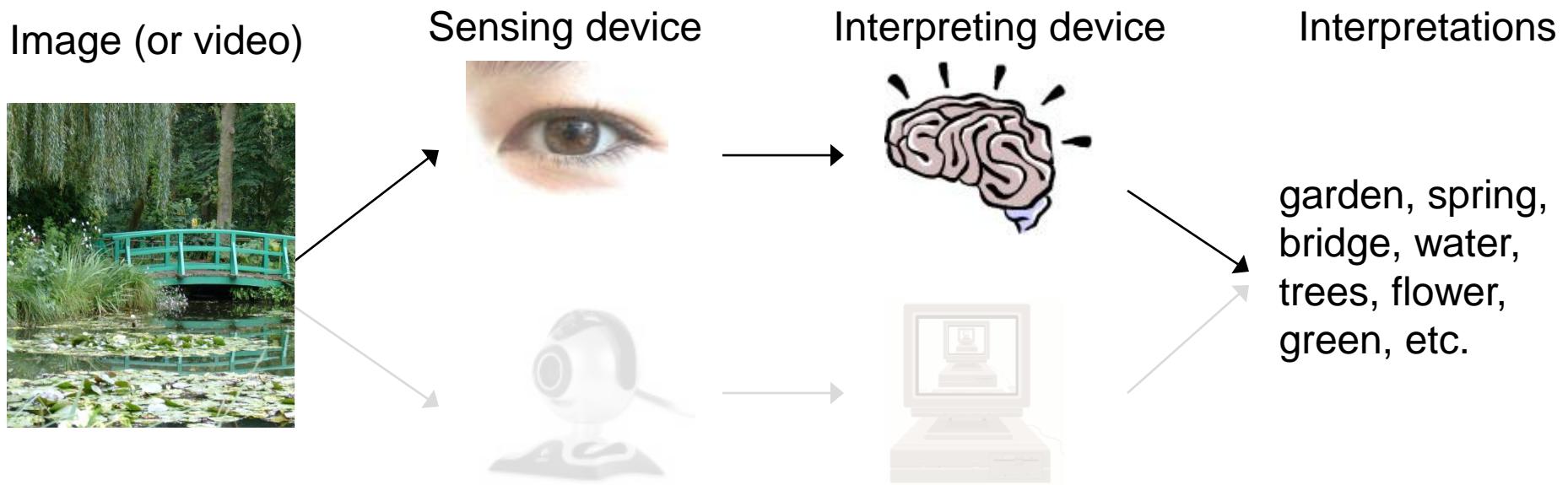


What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

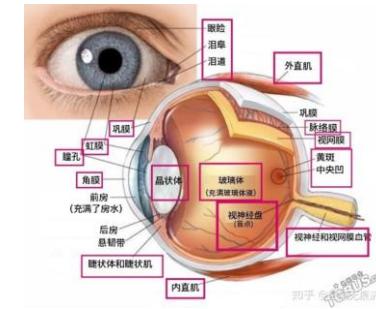
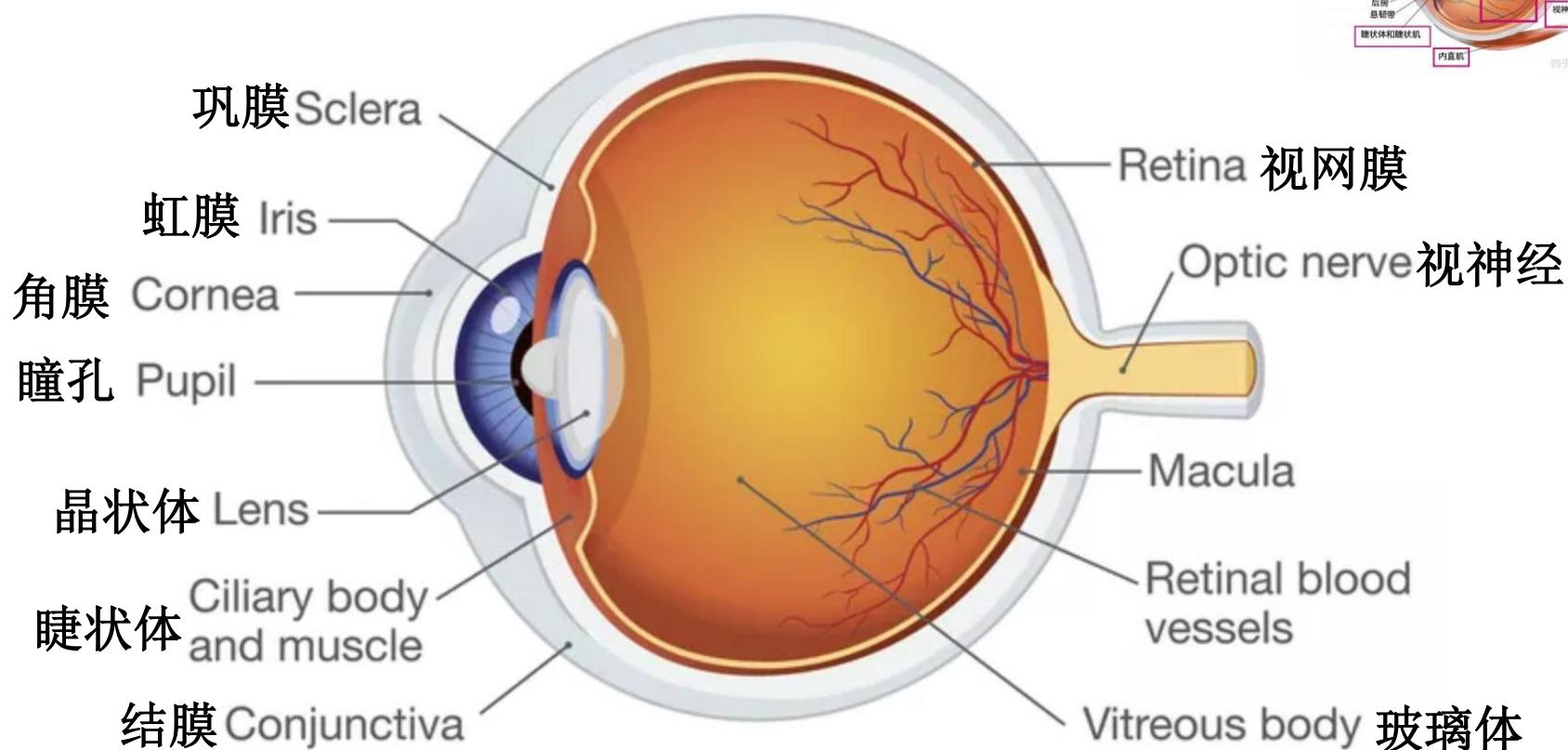
What a computer sees

# What is (computer) vision?

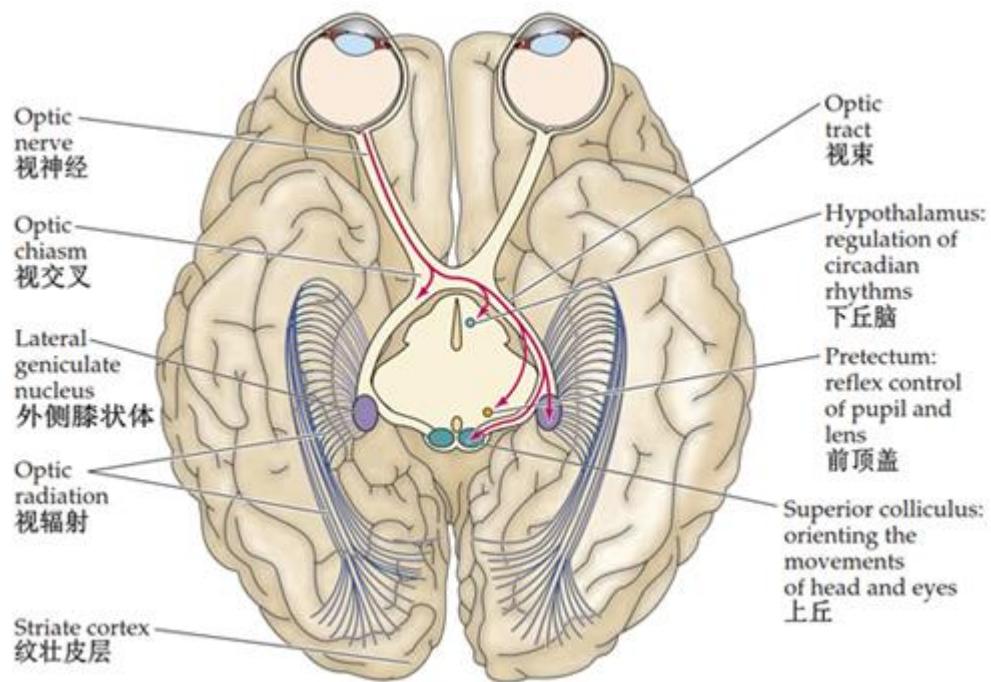
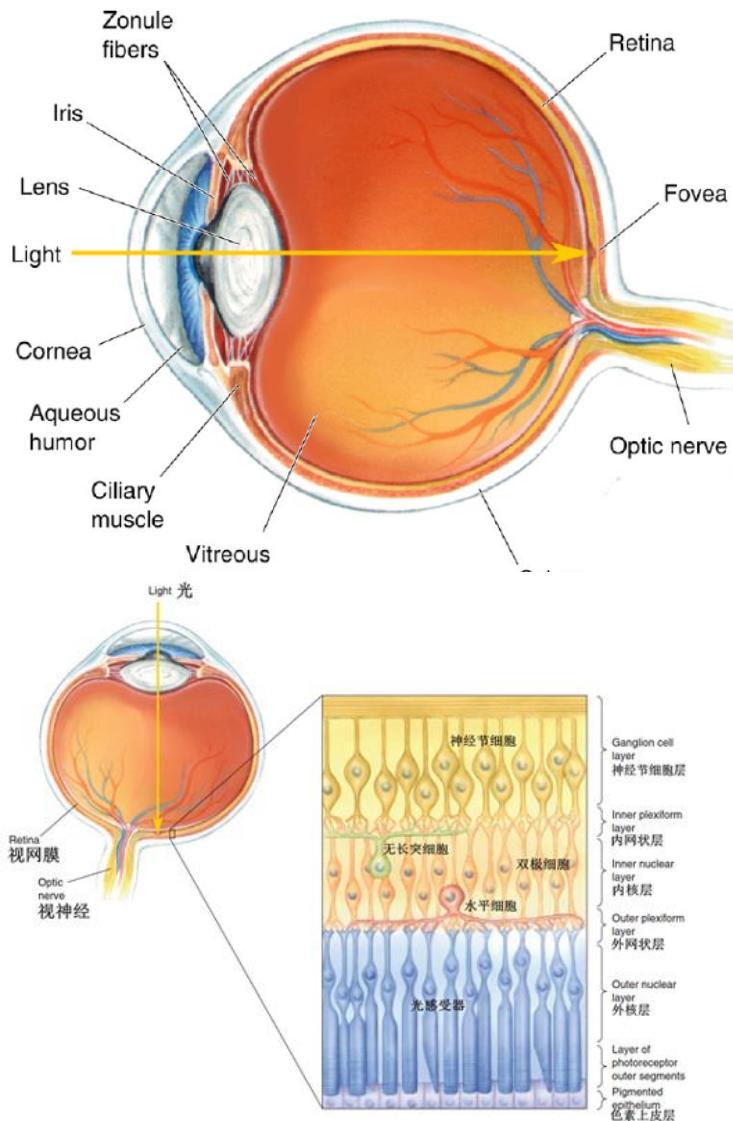


# Human Eye Anatomy

## Human Eye Anatomy



# Visual system



视网膜神经节细胞投射的主要脑区和结构

外膝体：视觉信息的中转站

视皮层：初级视觉皮层 (V1)、纹外皮层 (V2-V5)

腹侧通路（处理物体形状、颜色）：V1, V2, V4

背侧通路（处理运动信息）：V1, V2, V5

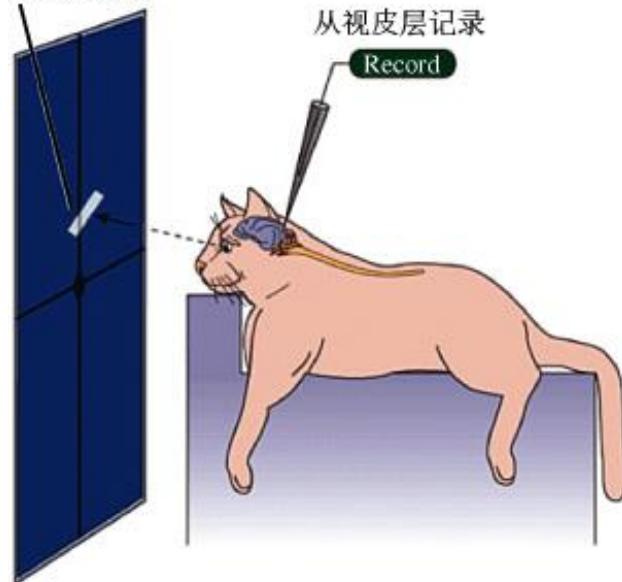
# 1981: Nobel Prize in medicine

## 方向选择性细胞 (Orientation Selective Cell)

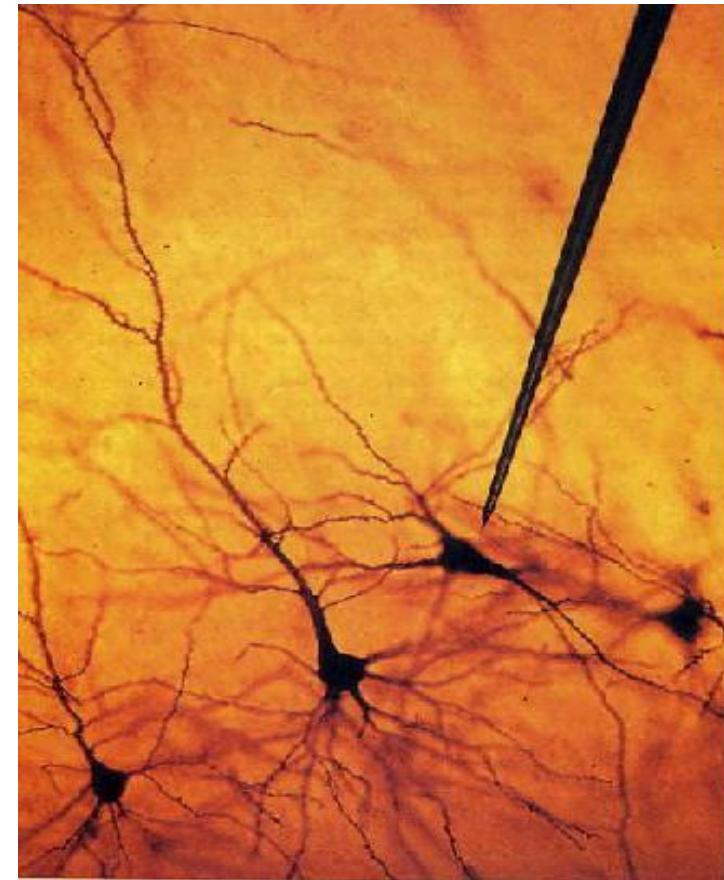
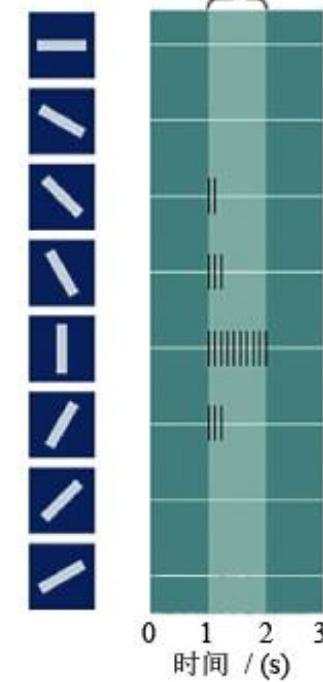
视觉皮层的神经元不是简单地对所有视觉刺激做出相同的响应，而是对于不同类型的刺激有选择性的响应，这为后来的视觉神经科学和计算机视觉领域的研究奠定了基础。

A 实验装置

投射到屏幕上的光棒刺激



B 刺激方位 刺激反应



1981年共享了诺贝尔生理学或医学奖，以表彰他们在“视觉系统信息加工”的重要贡献

# Human vision is superbly efficient

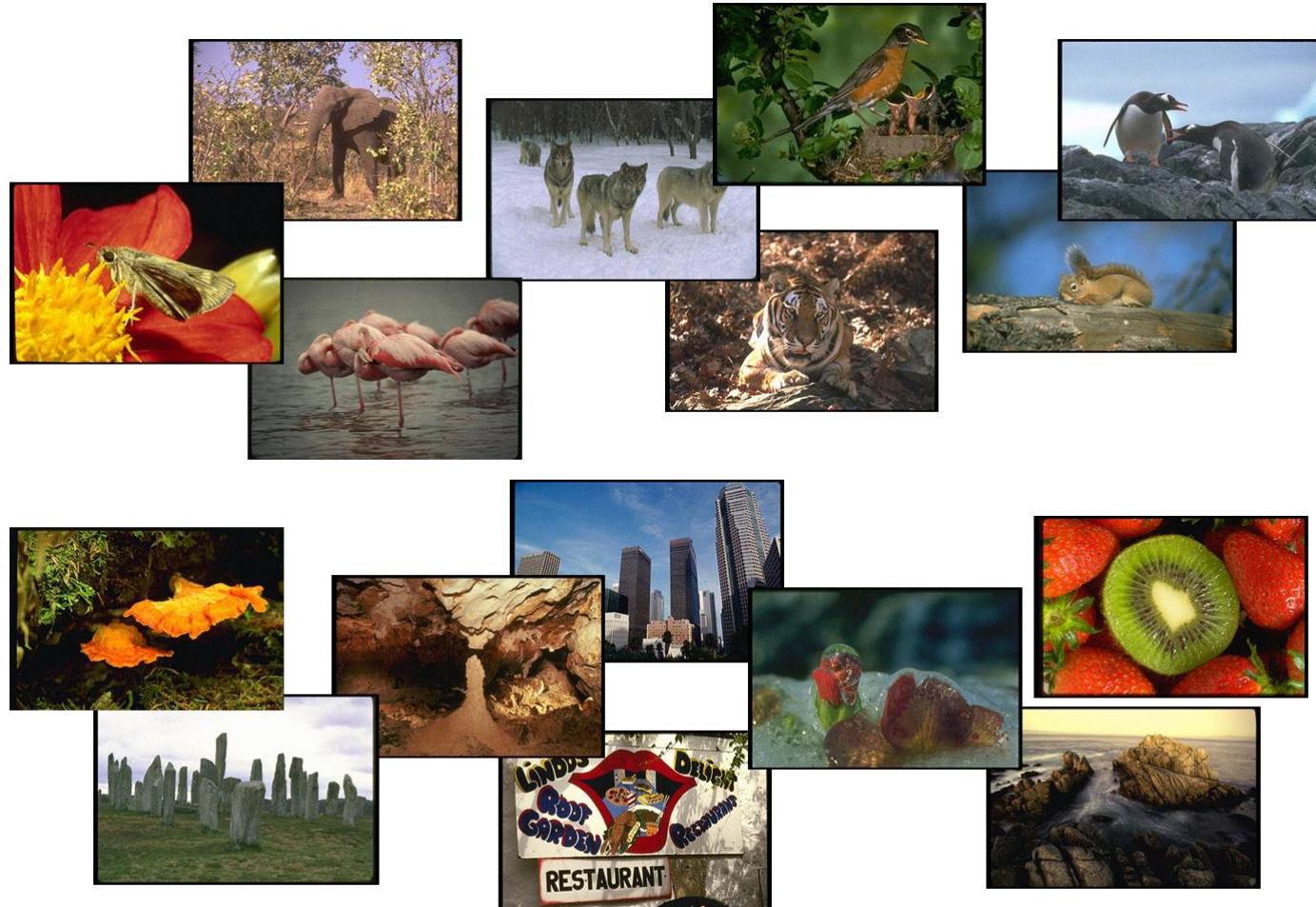
100 ms per frame



Potter, Biederman,  
etc. 1970s

Capable of perceiving a vast range of visual information, including color, depth, motion, and texture, with remarkable speed and accuracy. Highly adaptable and can adjust to different lighting conditions, focus on different distances, and quickly identify patterns and objects.

# Efficiency of human vision

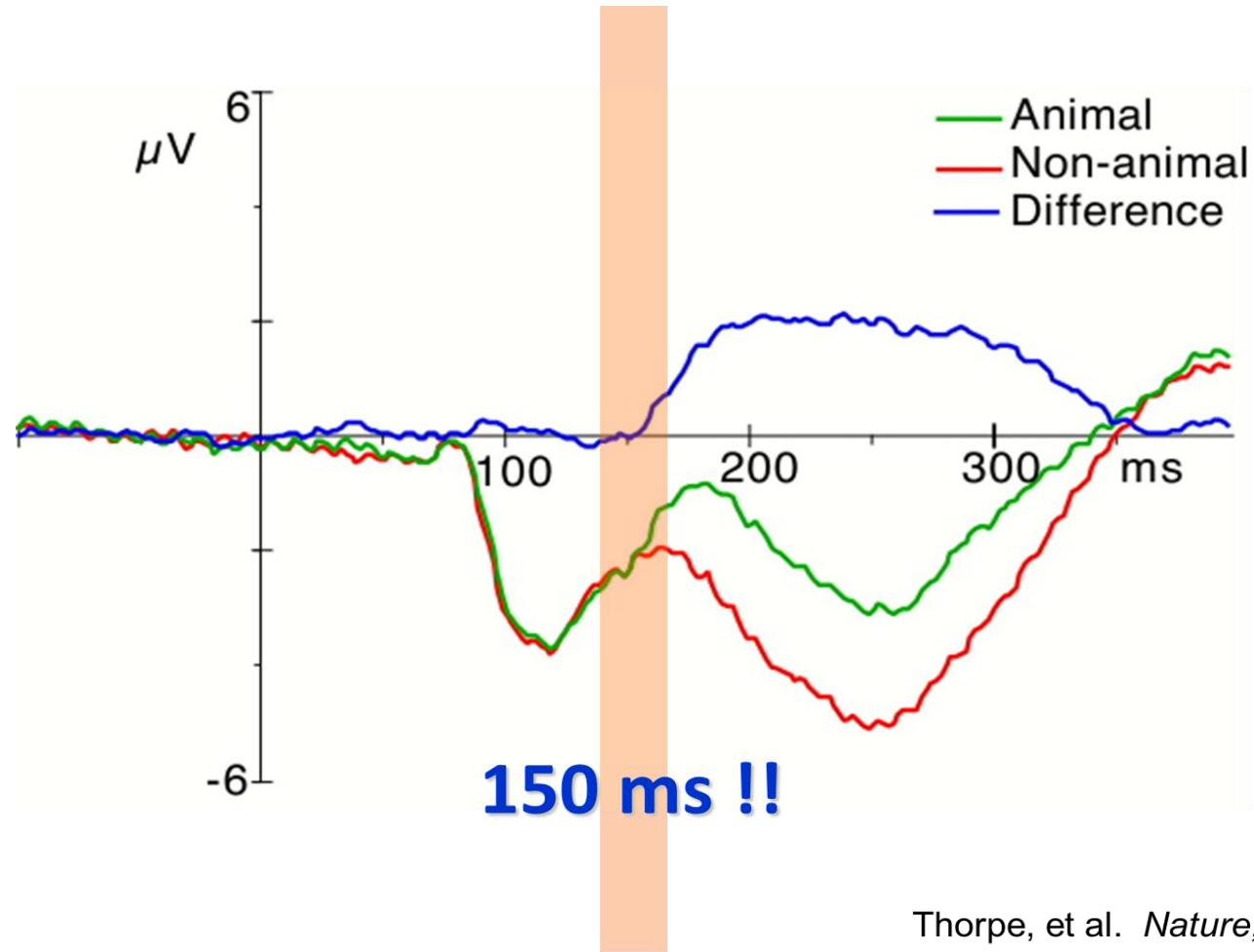


Thorpe, et al. *Nature*, 1996

27ms flash, Very complex pictures – multiple animals, city, textures, etc

# Efficiency of human vision

脑内神经元的活动通过离子传导到大脑的皮层，固定在头上的导电电极感应到微弱的电信号，差分放大、滤波、数模转换等一系列手段将电信号转成脑电波的原始数据。



Thorpe, et al. *Nature*, 1996

# Change Blindness (变化盲视)



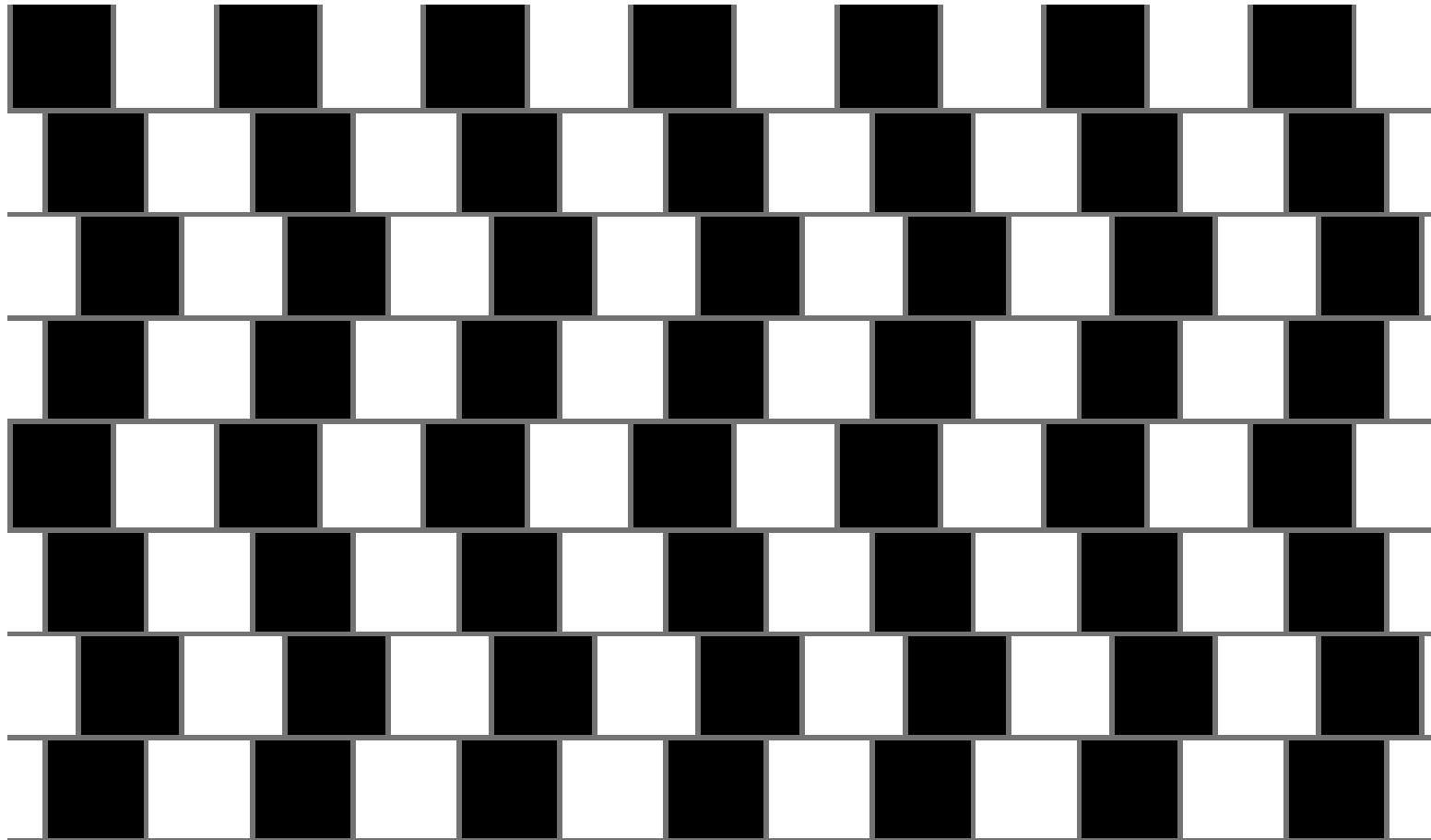
# Change Blindness

---



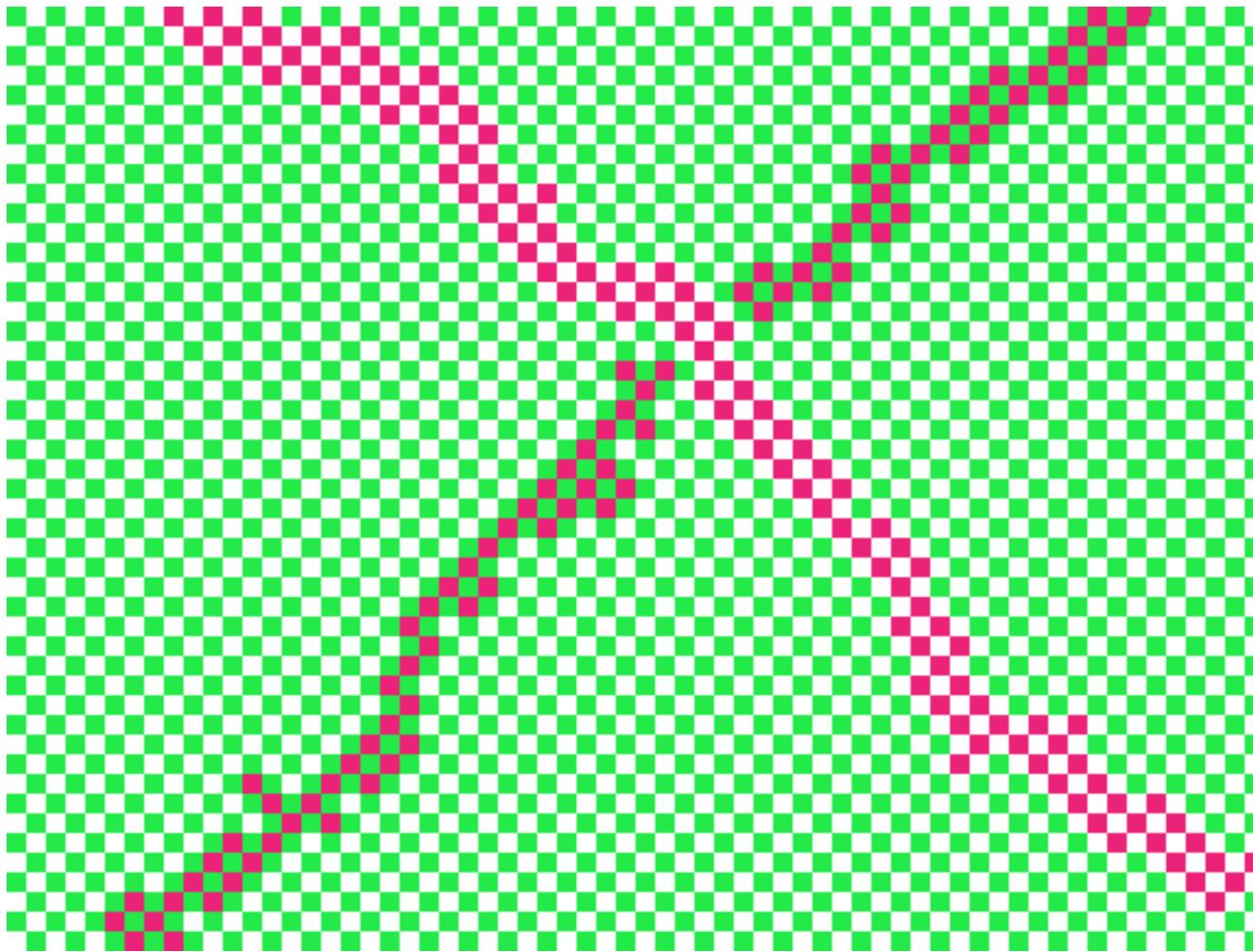
# Vision illusion

光学幻觉、运动幻觉、色彩幻觉、大小幻觉等。



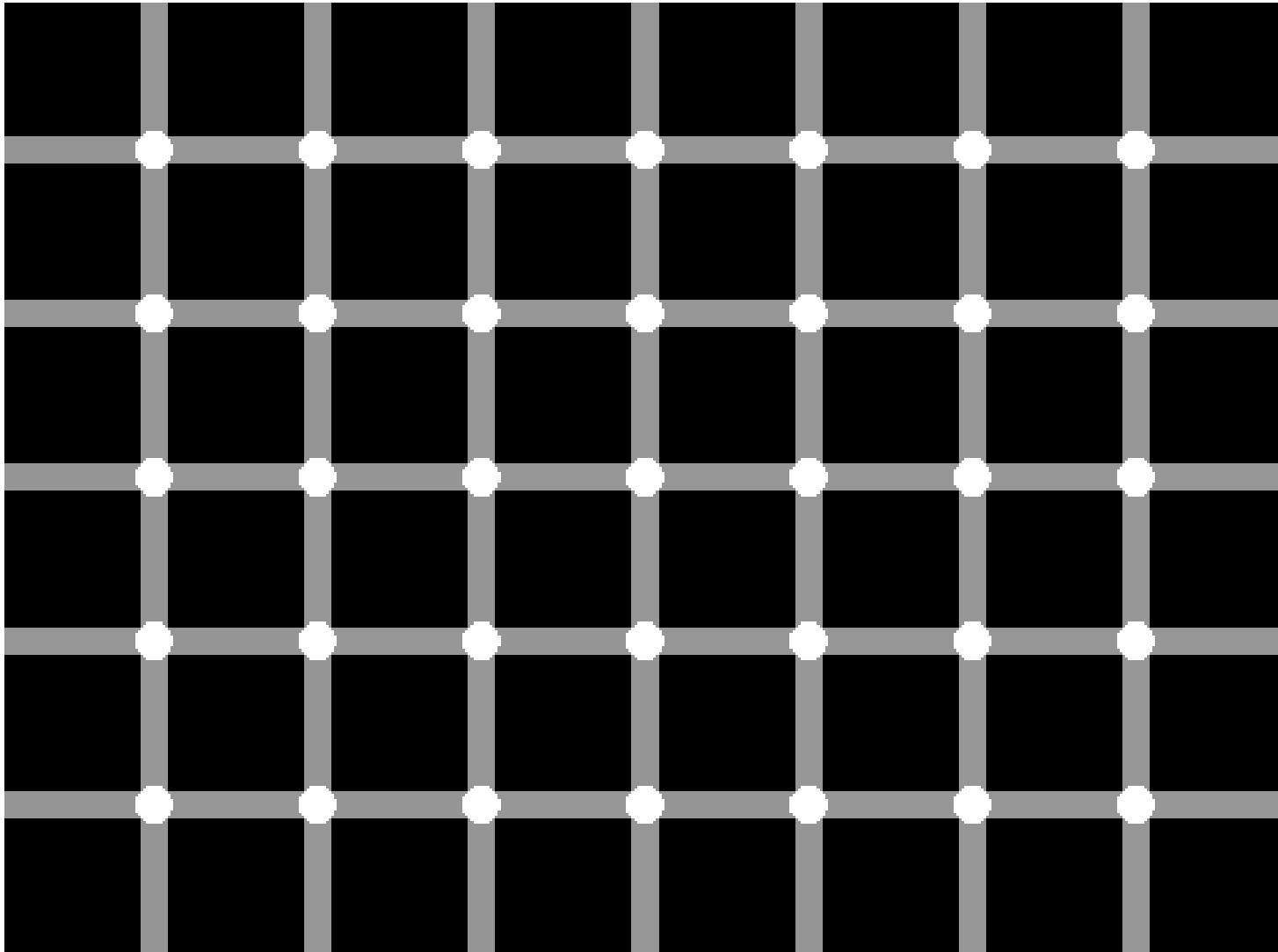
# Vision illusion

---



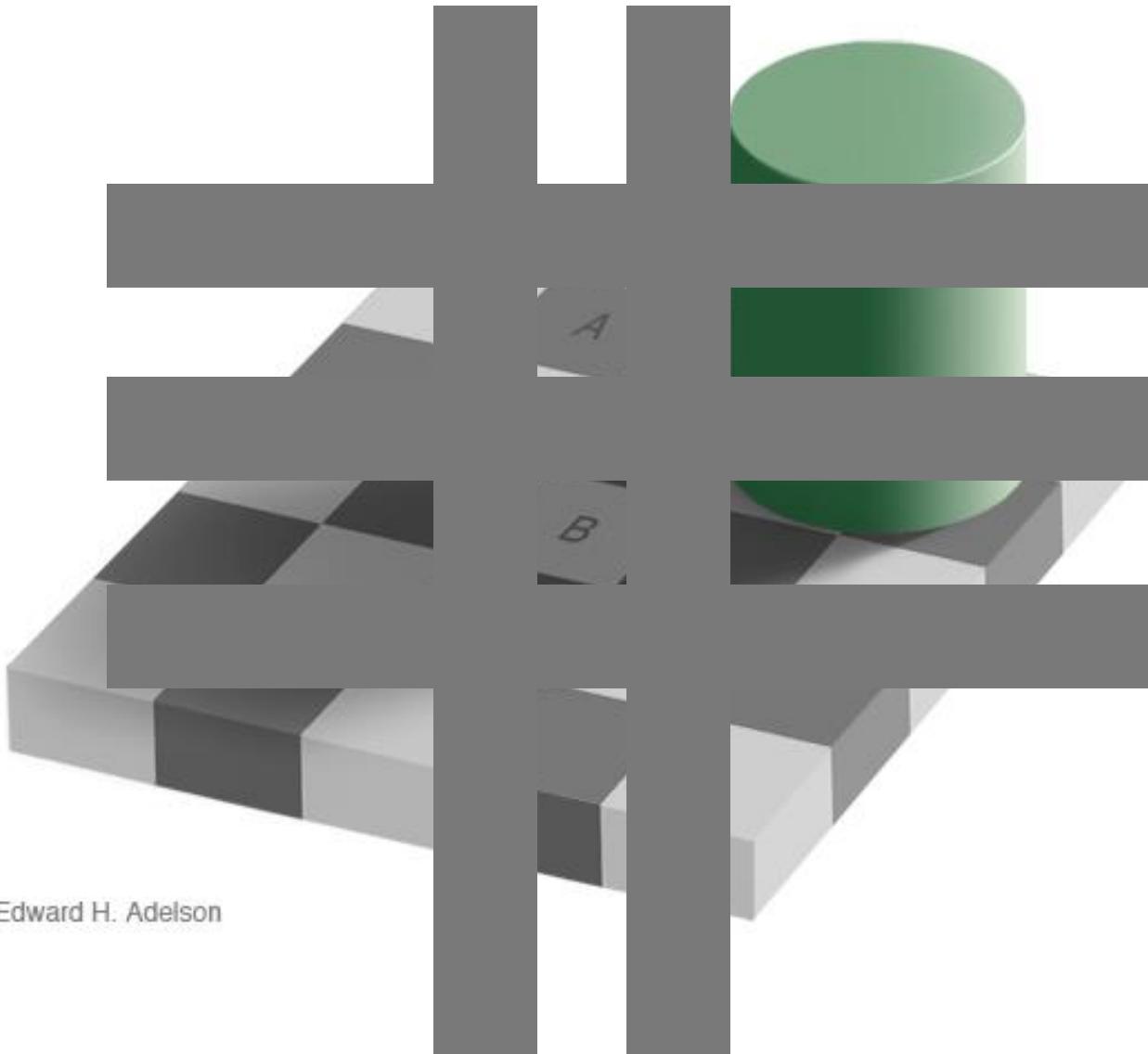
# Vision illusion

---



# Vision illusion

---



Edward H. Adelson

# Vision illusion

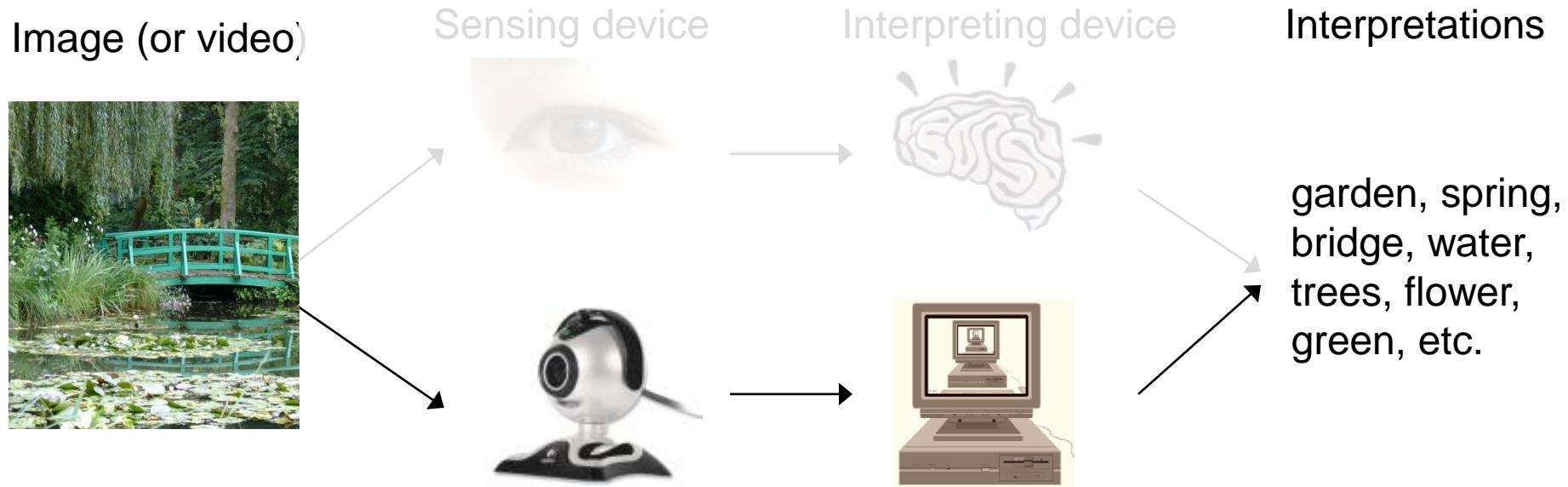
---



大脑受到周围区域的刺激，尝试去理解和解释图像

格式塔效应是指我们的大脑会根据周围环境和上下文来解释和理解我们的感知信息，而不是只看到孤立的元素。

# What is (computer) vision?



# Paintings in 1838

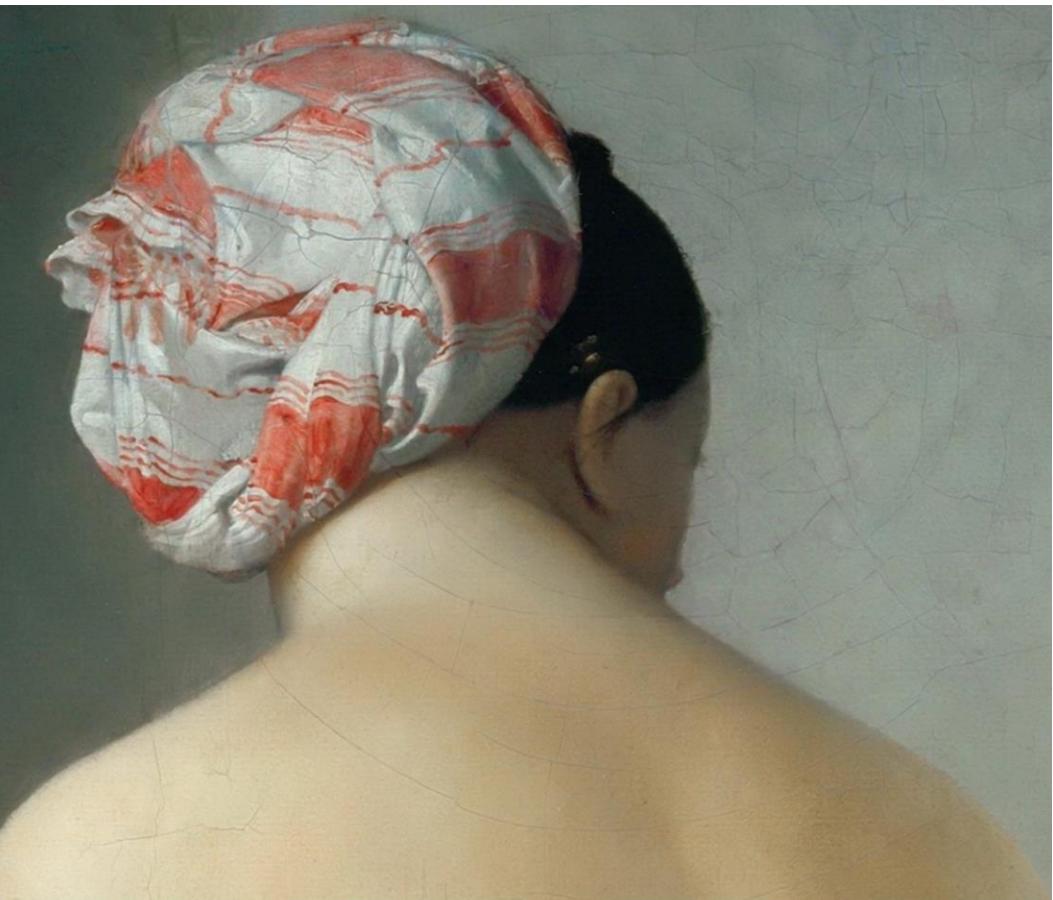
---



1812: Jacques-Louis-David  
The Emperor Napoleon at his Study at the Tuilleries

# Paintings in 1838

---



1808: Ingres, *La grande baigneuse*

# The birth of photography

公认的世界上第一幅照片是法国人尼埃普斯于1826年拍摄。

涂有沥青的白蜡板+8小时曝光（日光蚀刻法）



法国发明家尼埃普斯（窗外的风景）

# Demonstration of Daguerreotype

---



“From today, painting is dead”  
— painter Paul Delaroche  
at a demonstration of the Daguerreotype,  
1839

# Demonstration of Daguerreotype

达盖尔银版照相法是利用镀有碘化银的钢板在暗箱里进行曝光，然后以水银蒸汽进行显影，再以普通食盐定影，得到的实际上是一个金属负像，但十分清晰而且可以永久保存。



1838: Boulevard du Temple, Daguerre

# Daguerreotype (达盖尔银版照相法)

本视频由日本摄影师再现了达盖尔摄影术的过程

# The goal of computer vision

---



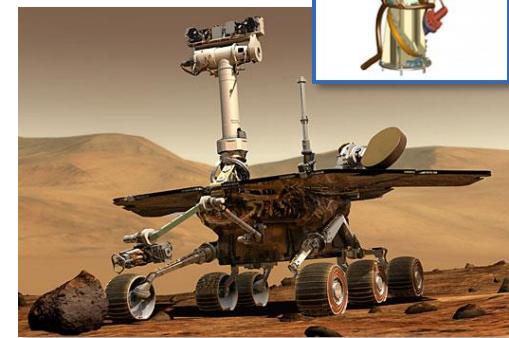
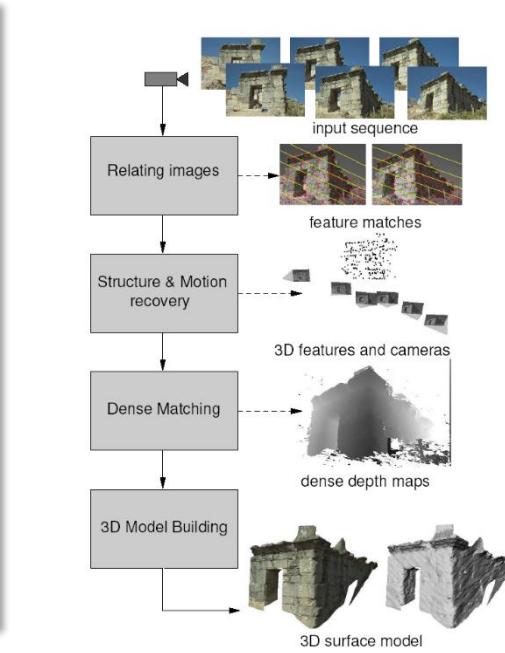
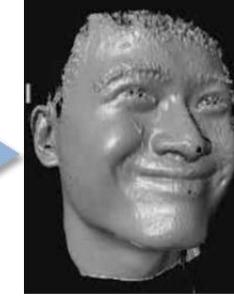
# Information extraction from image

---

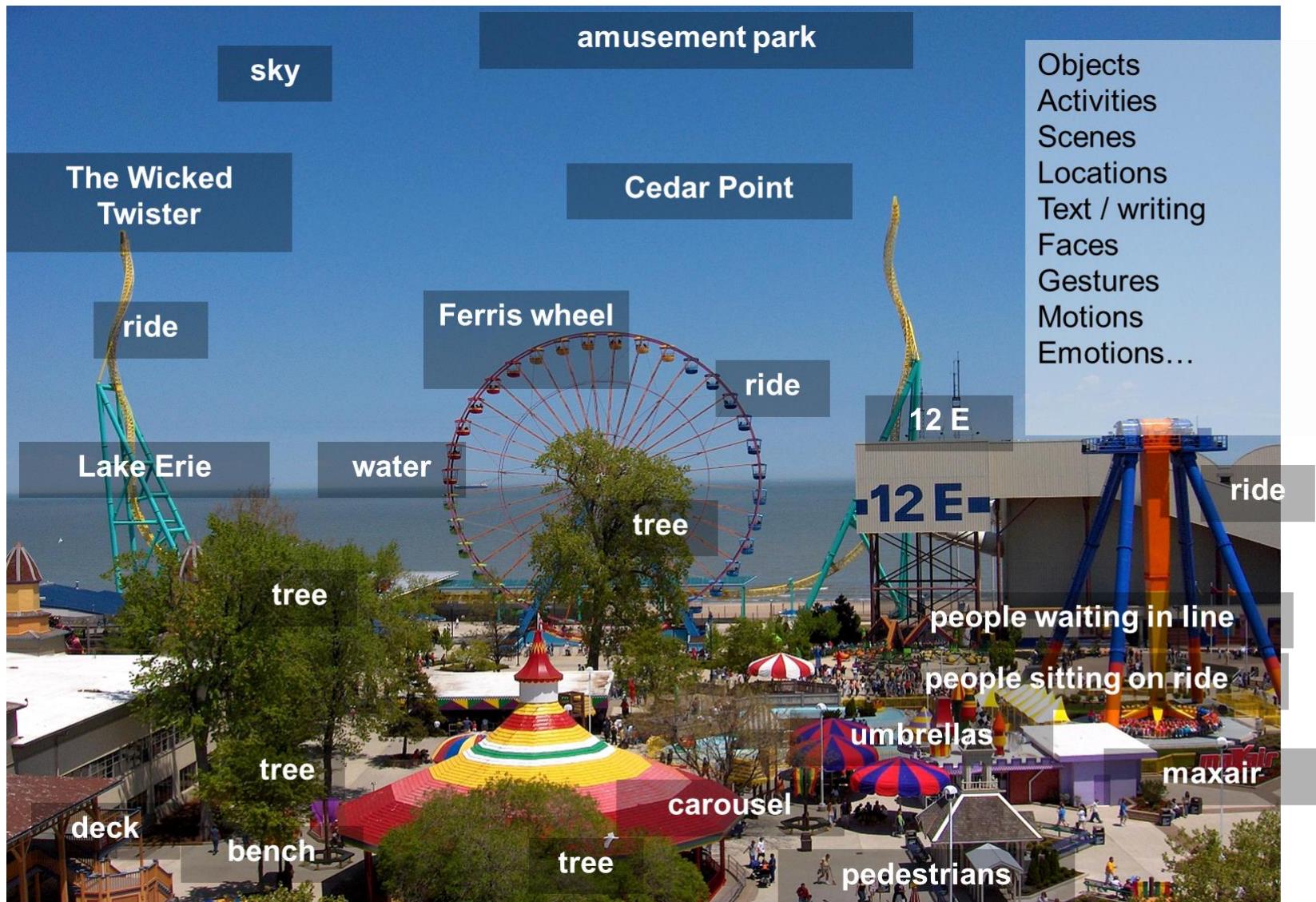
- Metric 3D information
- Semantic information

# Vision as measurement device

Compute the 3D shape of the world



# Vision as source of semantic information

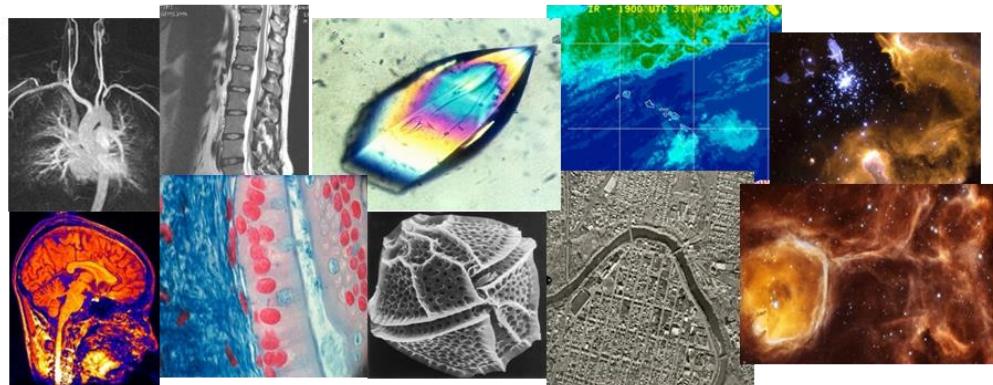


# Why study computer vision?

- Vision is useful: Images and video are everywhere!



Surveillance and security



Medical and scientific images

# **Why study computer vision?**

---

**80% of all web traffic is images and videos**

**Majority of the internet is dark matter without computer vision**

# Application of computer vision

- Improve photos (“Computational Photography”)



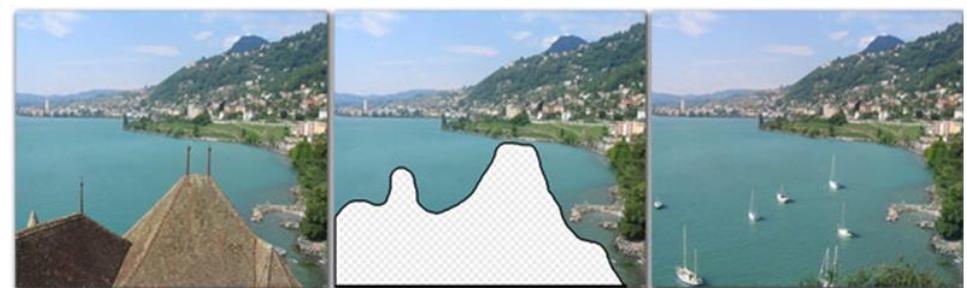
Super-resolution (source: 2d3)



Low-light photography  
(credit: [Hasinoff et al., SIGGRAPH ASIA 2016](#))



Depth of field on cell phone camera  
(source: [Google Research Blog](#))



Inpainting / image completion  
(image credit: Hays and Efros)

# Application of computer vision

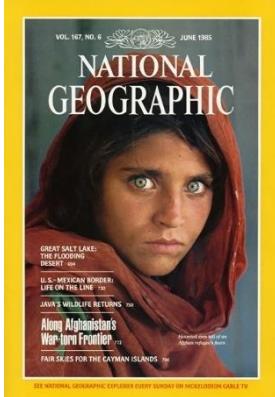
## Face detection



图片来源：视觉中国 www.vcg.com

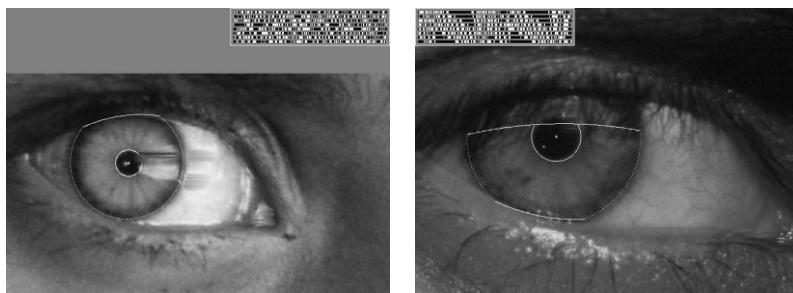
# Application of computer vision

## Biometrics



Fingerprint scanners on many new smartphones and other devices

### How the Afghan Girl was Identified by Her Iris Patterns



虹膜包含有很多相互交错的斑点、细丝  
、冠状、条纹、隐窝等的细节特征



Face unlock on Apple iPhone X  
See also <http://www.sensiblevision.com/>

# Application of computer vision

Vision in supermarkets (无人零售)



# Application of computer vision

## Automotive safety

The collage consists of several panels:

- A top-left panel shows a car from above with three camera types labeled: "rear looking camera" (top left), "forward looking camera" (top right), and "side looking camera" (bottom). Above the car, the text "Our Vision. Your Safety." is displayed. Navigation arrows for "manufacturer products" and "consumer products" are at the top.
- A top-right panel titled "News" contains two news items:
  - "Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System"
  - "Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end" with a link to "all news".
- A bottom-left panel features a "EyeQ Vision on a Chip" image and a "Vision Applications" section showing a pedestrian detection scenario with a person walking across a crosswalk.
- A large bottom-right panel depicts a futuristic car interior with a dashboard screen displaying a 3D augmented reality view of the road ahead, showing other vehicles and a pedestrian.

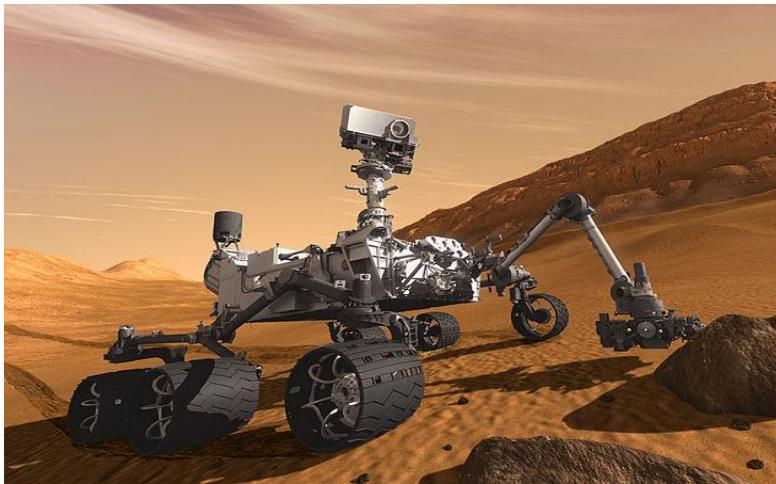
# Application of computer vision

## Intelligent medical image analysis



# Application of computer vision

## Robotics



NASA's Mars Curiosity Rover  
[https://en.wikipedia.org/wiki/Curiosity\\_\(rover\)](https://en.wikipedia.org/wiki/Curiosity_(rover))



Amazon Picking Challenge  
<http://www.robocup2016.org/en/events/amazon-picking-challenge/>



Amazon Prime Air

# Application of computer vision

## Virtual Reality



# Application of computer vision

## Augmented Reality



# 什么是人工智能？

LI

简述什么是人工智能



人工智能（Artificial Intelligence，简称AI）是一门研究如何使计算机模拟人类智能的学科。



它涉及模拟、理解和实现人类智能的各个方面，包括感知、推理、学习、语言处理、决策等。通过利用计算机和机器学习等技术，人工智能可以使计算机像人一样完成复杂任务，或在某些任务上超越人类的表现。

LI

什么是强人工智能和弱人工智能



强人工智能和弱人工智能是两种不同的人工智能分类。



强人工智能是指一种拥有与人类智能相似的能力的人工智能系统。强人工智能具备理解、推理、决策等人类智能的各种能力，能够在不同领域进行高效且自主的思考，其智能水平可与人类媲美或超过人类。

弱人工智能是指只能完成单一或有限领域任务的人工智能系统，其智能水平只能针对特定任务进行优化，且无法具备超越人类的综合性智能。

目前，大部分的人工智能系统都是弱人工智能，能够完成特定领域的任务，如图像识别、语音识别、自然语言处理等。而强人工智能仍处于研究和开发阶段，目前还没有完全实现。

# 什么是人工智能？

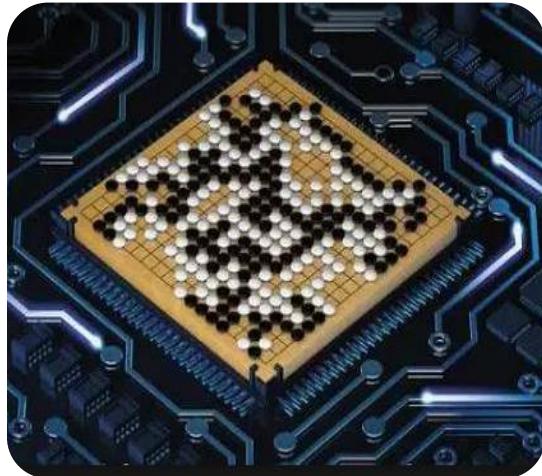
1. 由人创造出来的智能机器（或者系统）：智能驾驶、智能监控、智能影像分析
2. 类比人类智慧（常识、推理、情感、主观能动性...）



# 人工智能的分类

## 弱人工智能

- 不能真正推理和解决问题
- 看起来像人而已



## 强人工智能

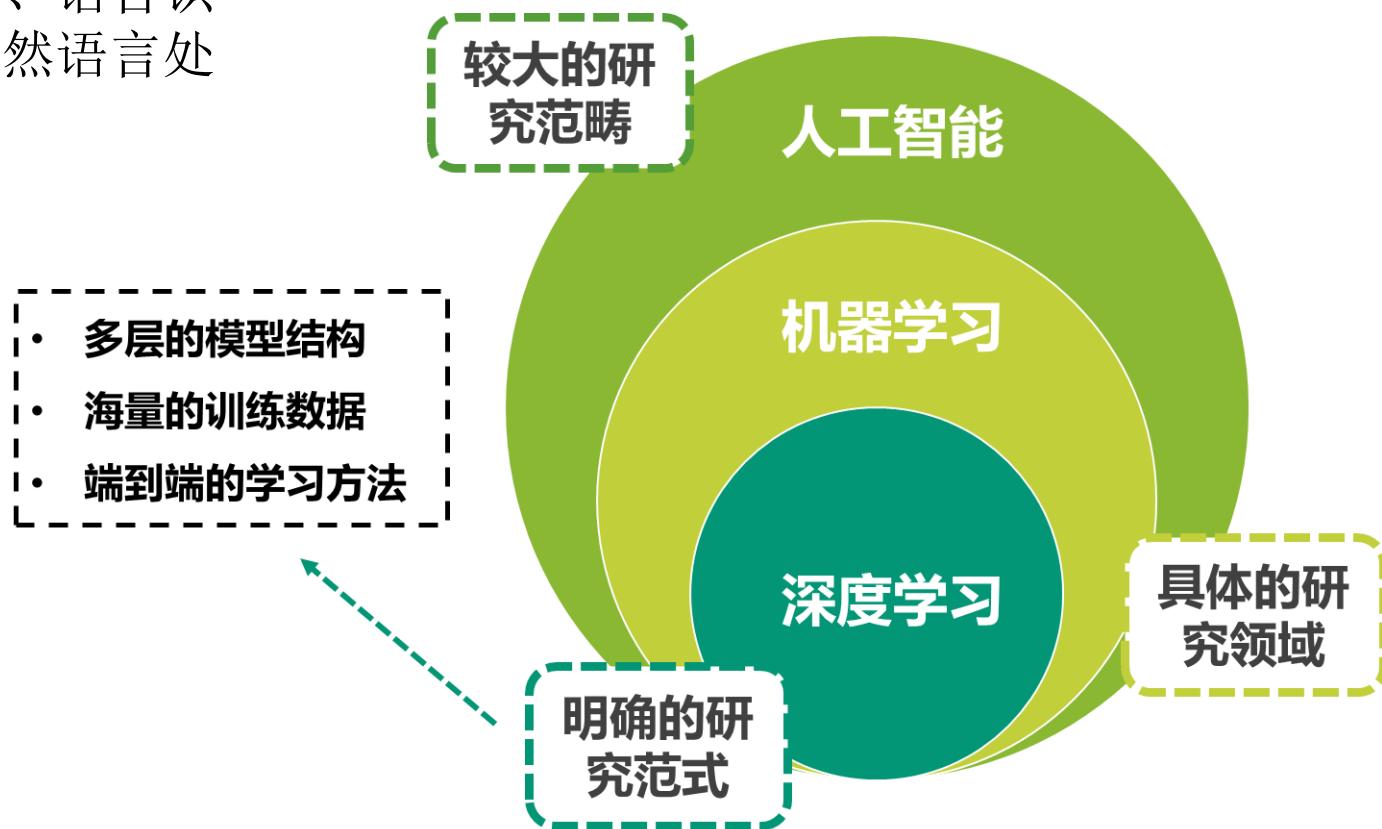
- 能够真正推理和解决问题
- 有自我意识的智能体



# 三个熟悉的概念

人工智能是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

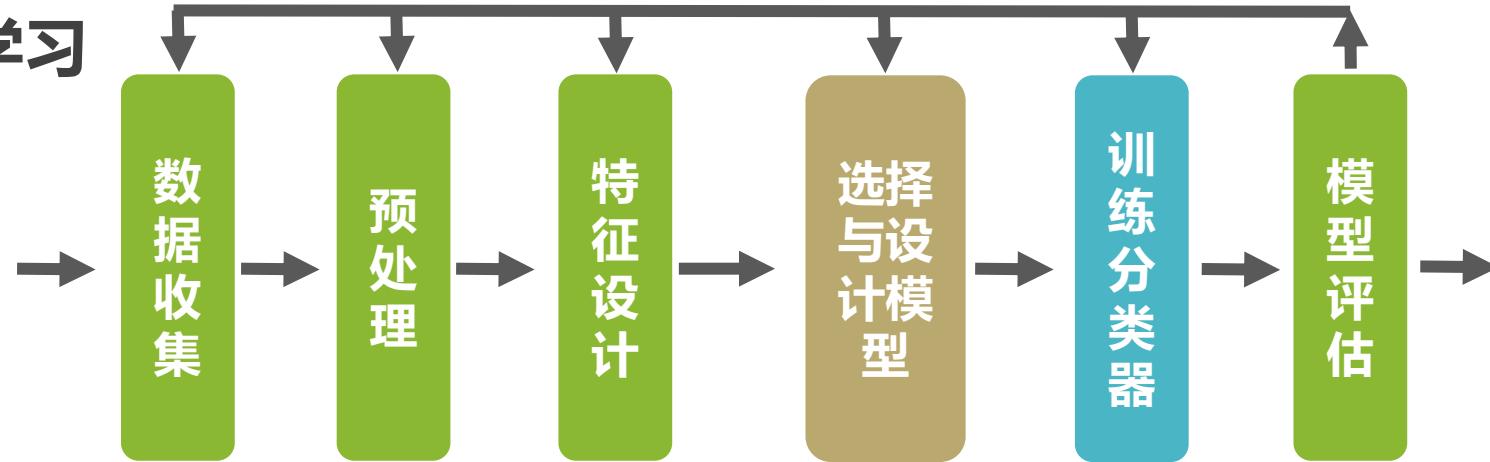
包括例如：机器人、语言识别、图像识别、自然语言处理和专家系统等



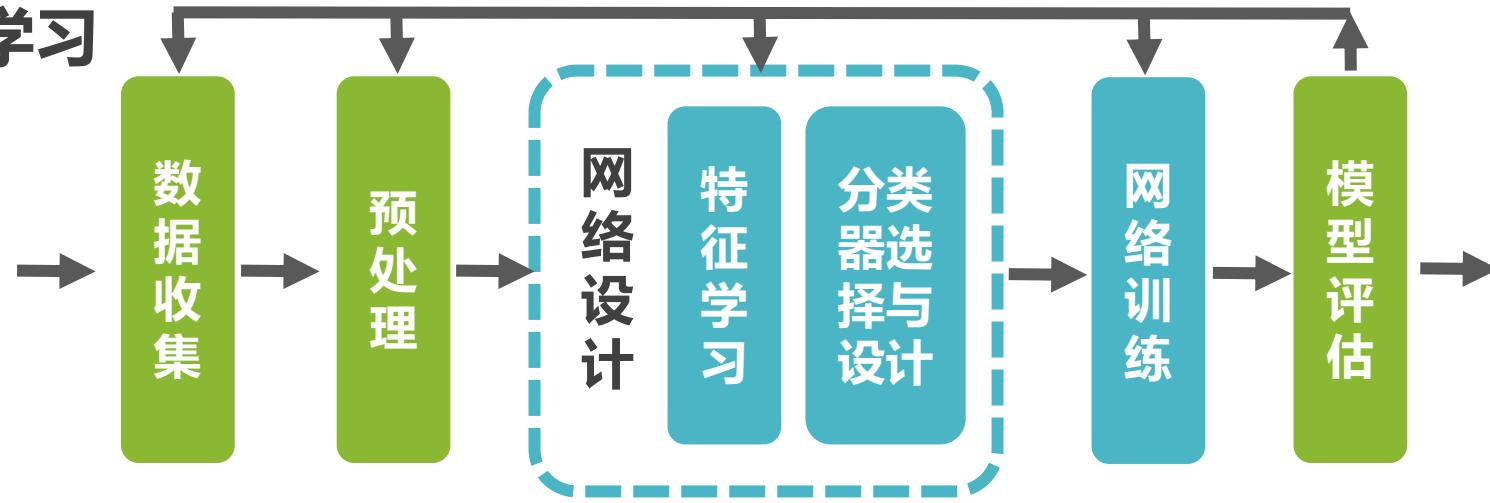
2018/5/4

# 深度学习的基本流程

机器学习



深度学习



# 人工智能的应用领域



自然语言处理



计算机视觉



智能控制

# 人工智能的应用领域

目标：像人类一样去理解世界



视觉



语音



语言



知识



搜索

计算视觉

人脸识别

视频检测

**医学影像**

智能语音

声纹识别

语音识别

文本分析

语言理解

拼写检查

自动翻译

知识图谱

学术知识

智能推荐

知识探索

智能检索

图片检索

新闻检索

视频检索

网页检索

# 深度学习推动了计算机视觉的变革

## Face Verification



## Image Classification



深度神经网络

## Robotics



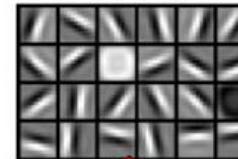
## Automated Driving



object models



object parts  
(combination  
of edges)



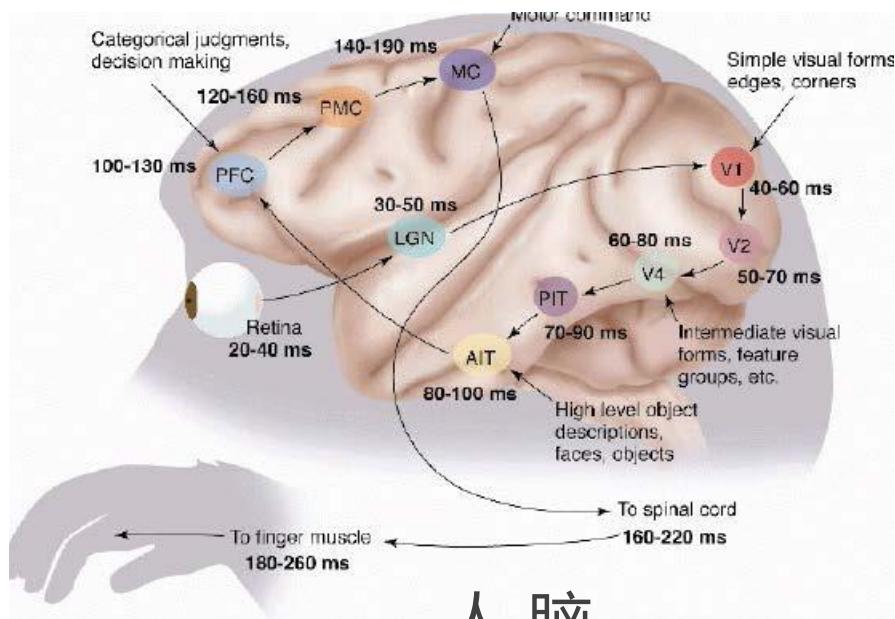
edges



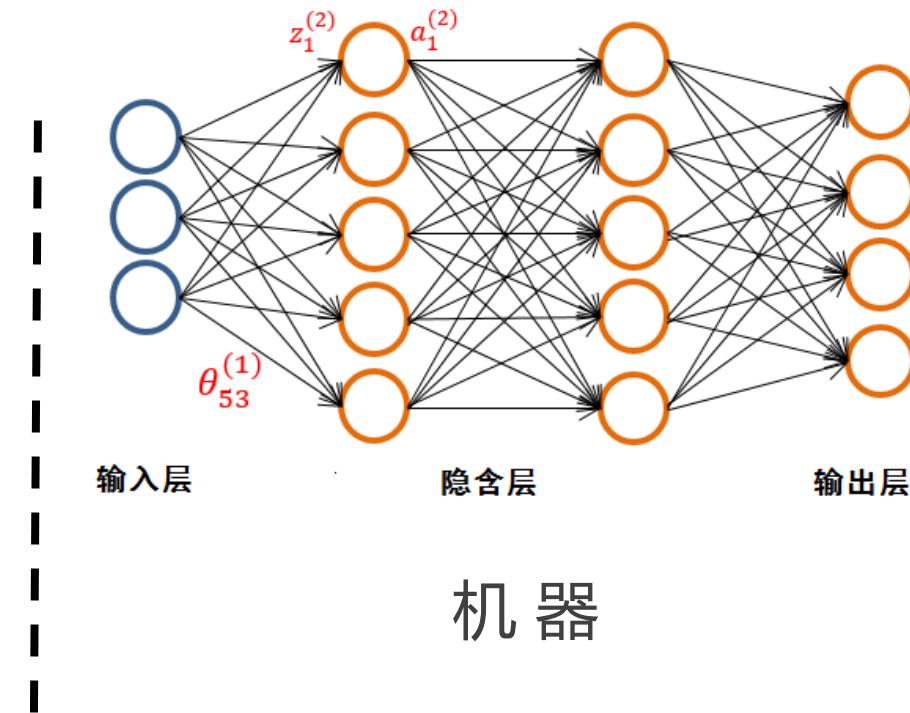
pixels

端到端的特征学习

# 深度学习的前身（上世纪八十年代）



人 脑



机 器

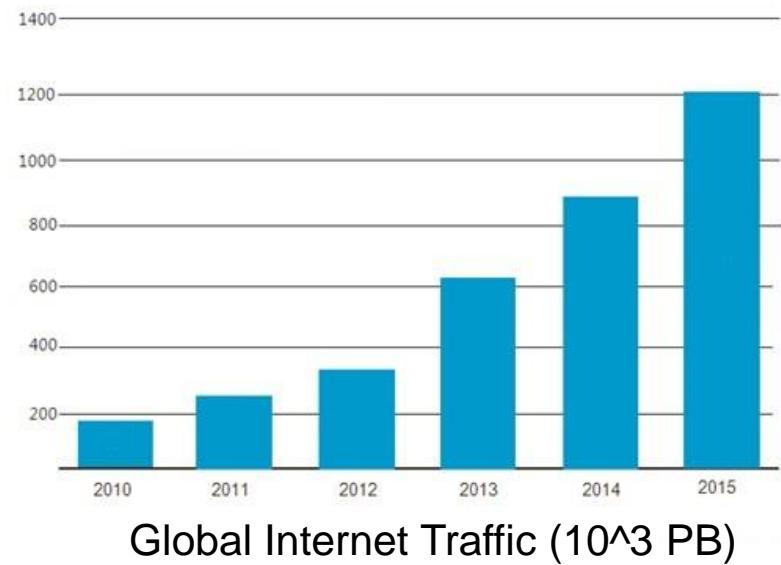
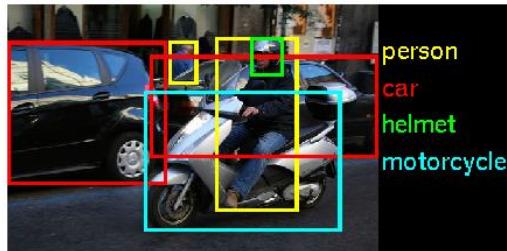
- 早期神经网络通过**模拟大脑认知的机理**解决各种机器学习问题
- 模型有大量的参数 → 较高的计算复杂度
- 需要大量的数据 → 否则会导致模型的过拟合
- 与其他的机器学习方法比较 → 没有明显的优势

# 新时代的契机（二十一世纪）

## 数据的爆炸式增长

**IMAGENET** Large Scale Visual  
Recognition Challenge (ILSVRC) 2012-2015

200 categories	516,849 images	DET
1000 categories	1,431,167 images	CLS-LOC



## 计算资源的极大丰富

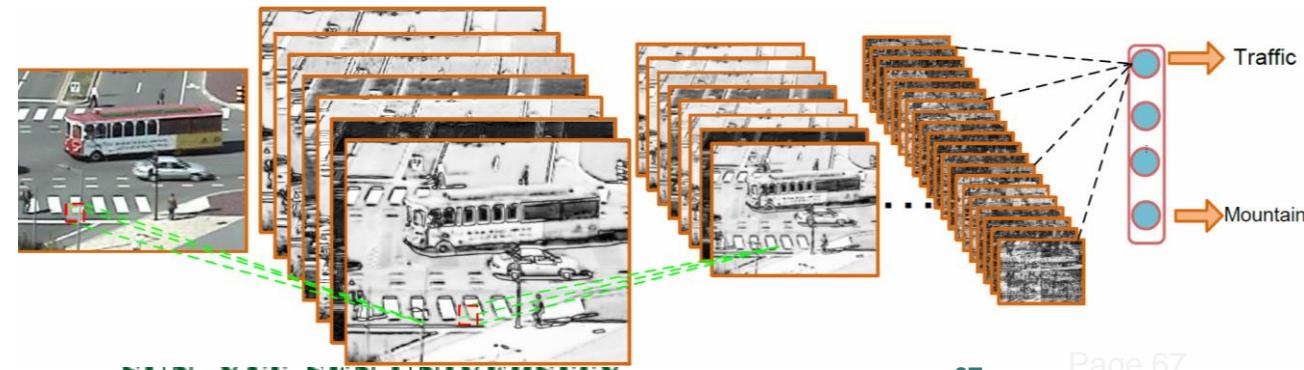


# 深度学习的大规模应用（2012年以来）



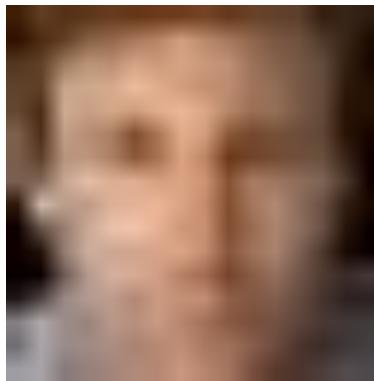
## 卷积神经网络(CNN)

- 判别式模型
- 大规模数据需求



# CV Research in HCP Lab

## Low level vision (Face hallucination/super-resolution)



Low Resolution Face Image

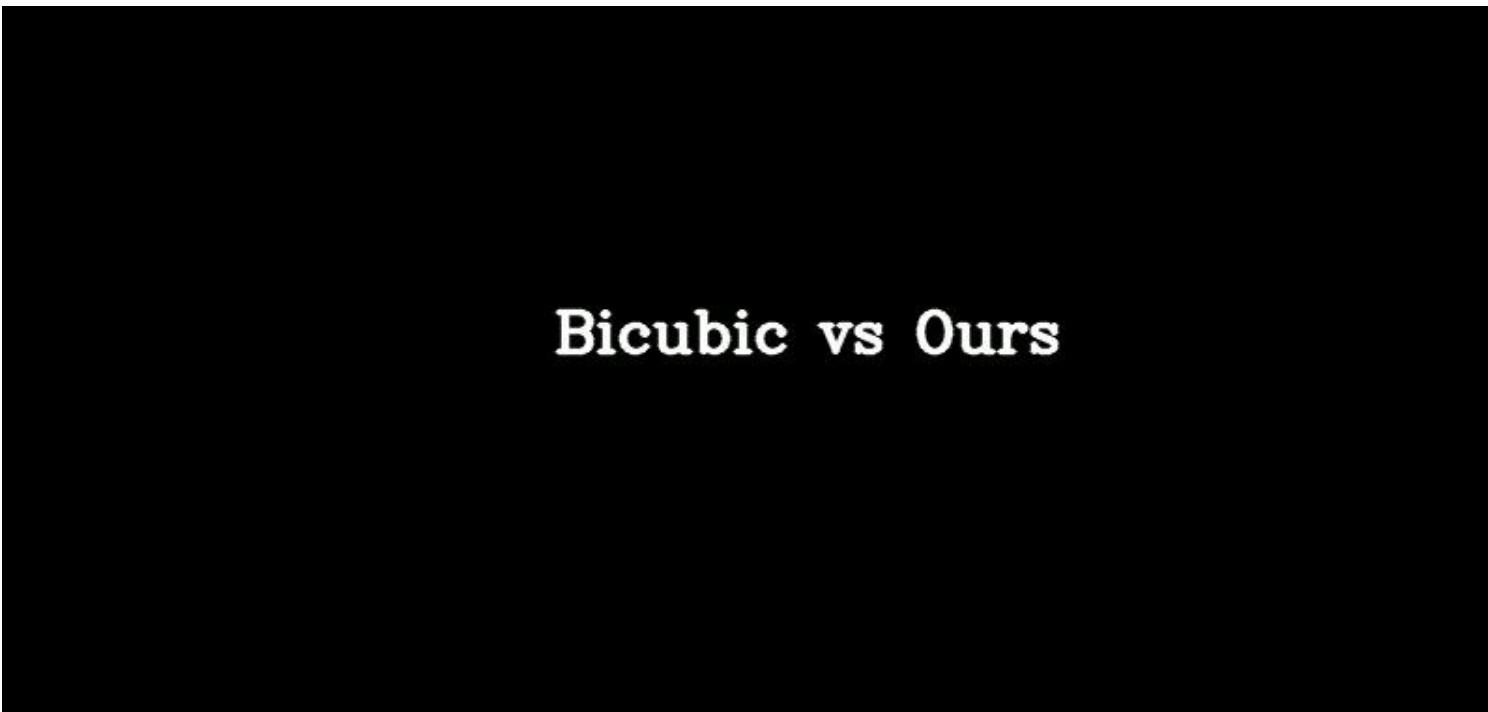


High Resolution Face Image

# CV Research in HCP Lab

---

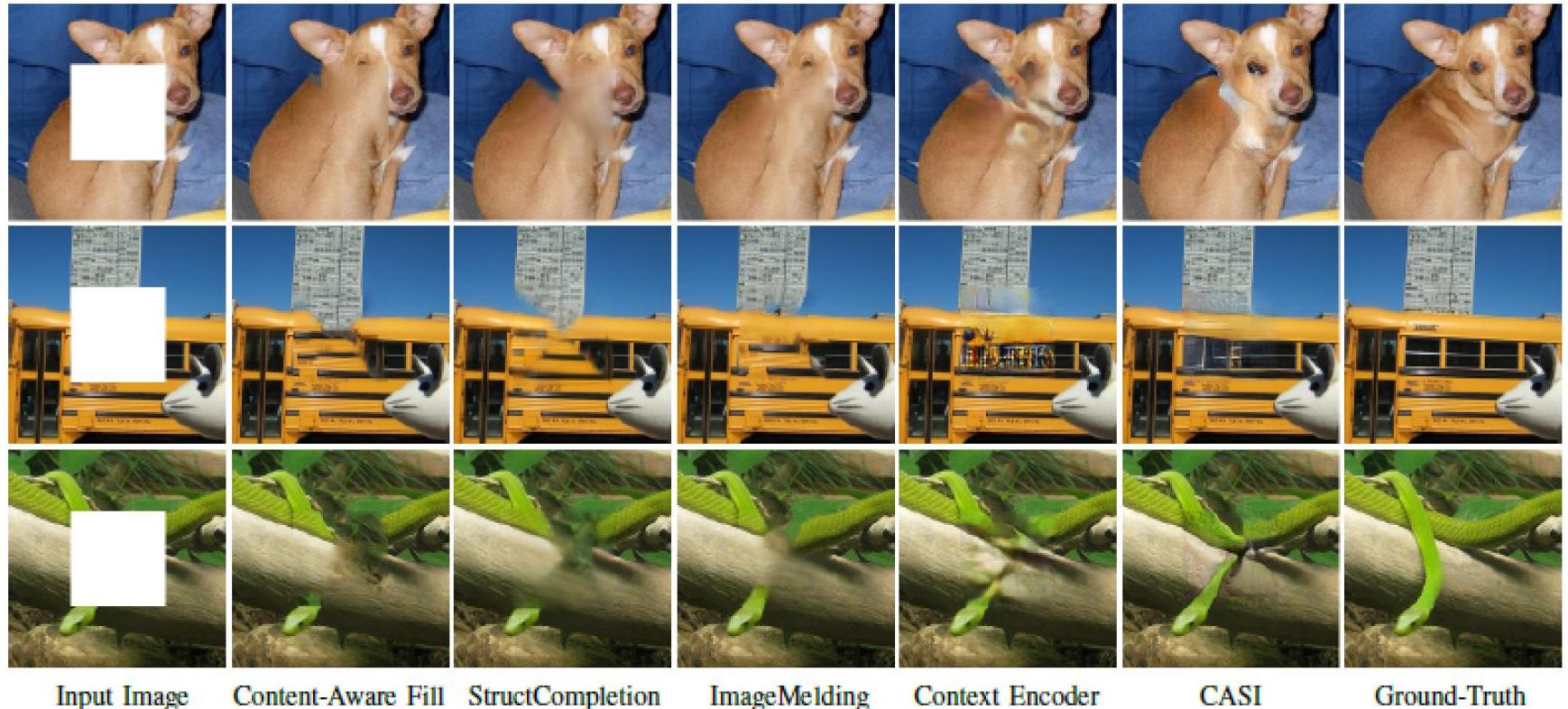
## Low level vision (Face hallucination/super-resolution)



**Bicubic vs Ours**

# CV Research in HCP Lab

## Low level vision (Image Inpainting)



# CV Research in HCP Lab

---

## Low level vision (De-raining)



# CV Research in HCP Lab

## Middle level vision (Visual Saliency)



应用需求推动研究发展



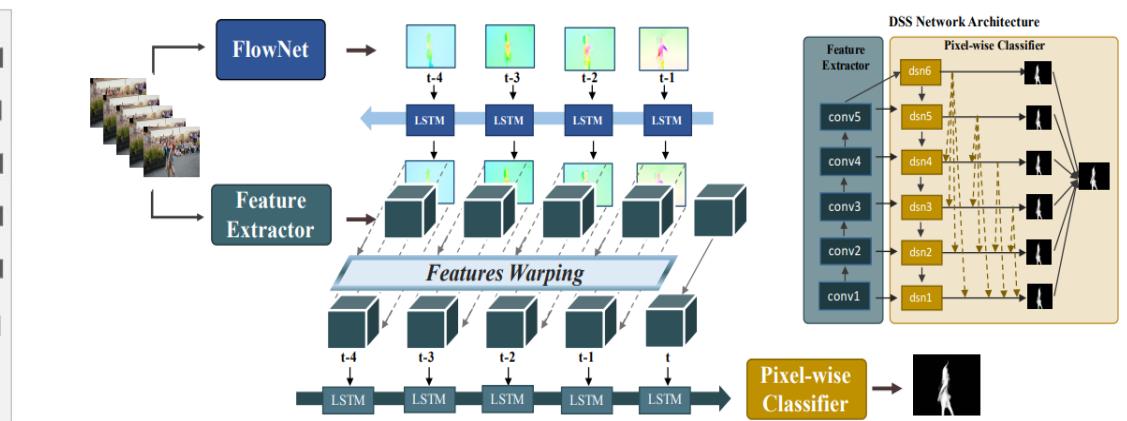
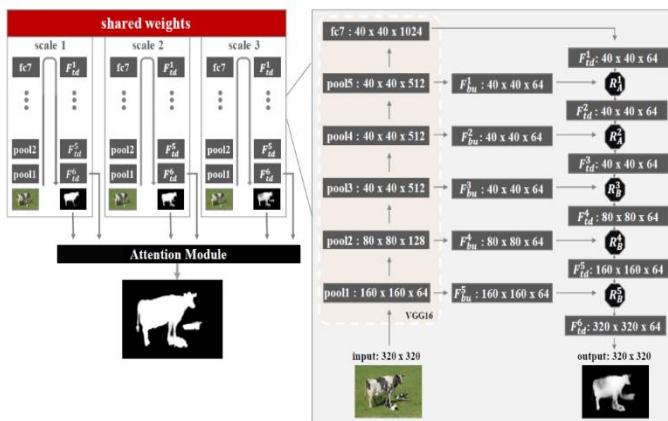
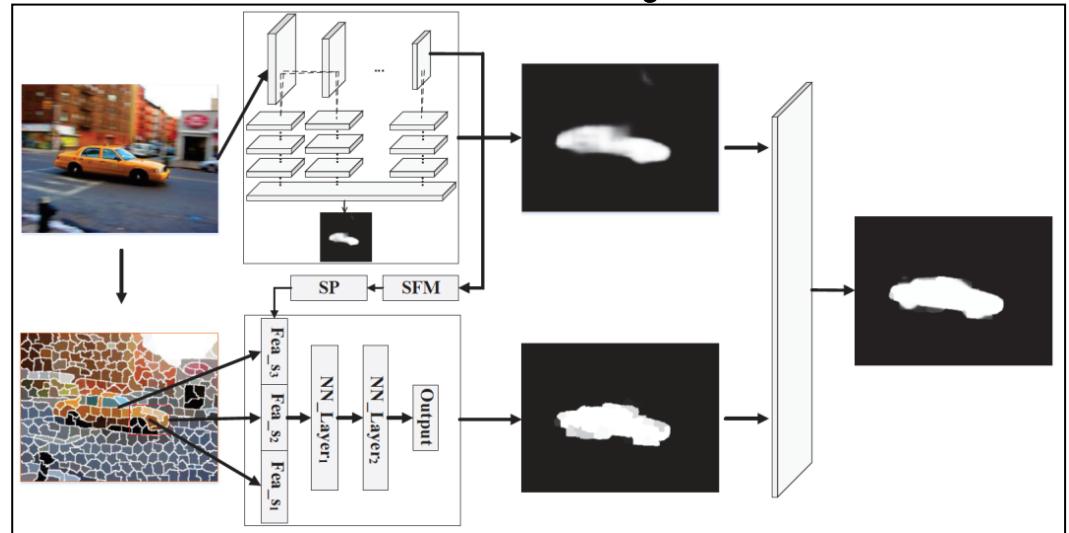
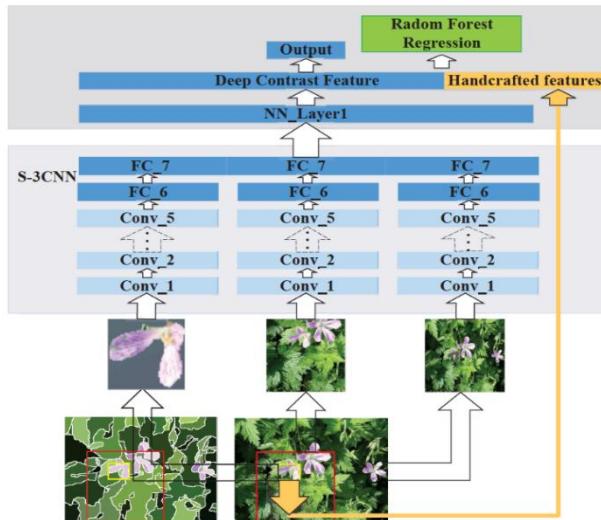
图片拼贴

内容感知的图像缩放

背景虚化

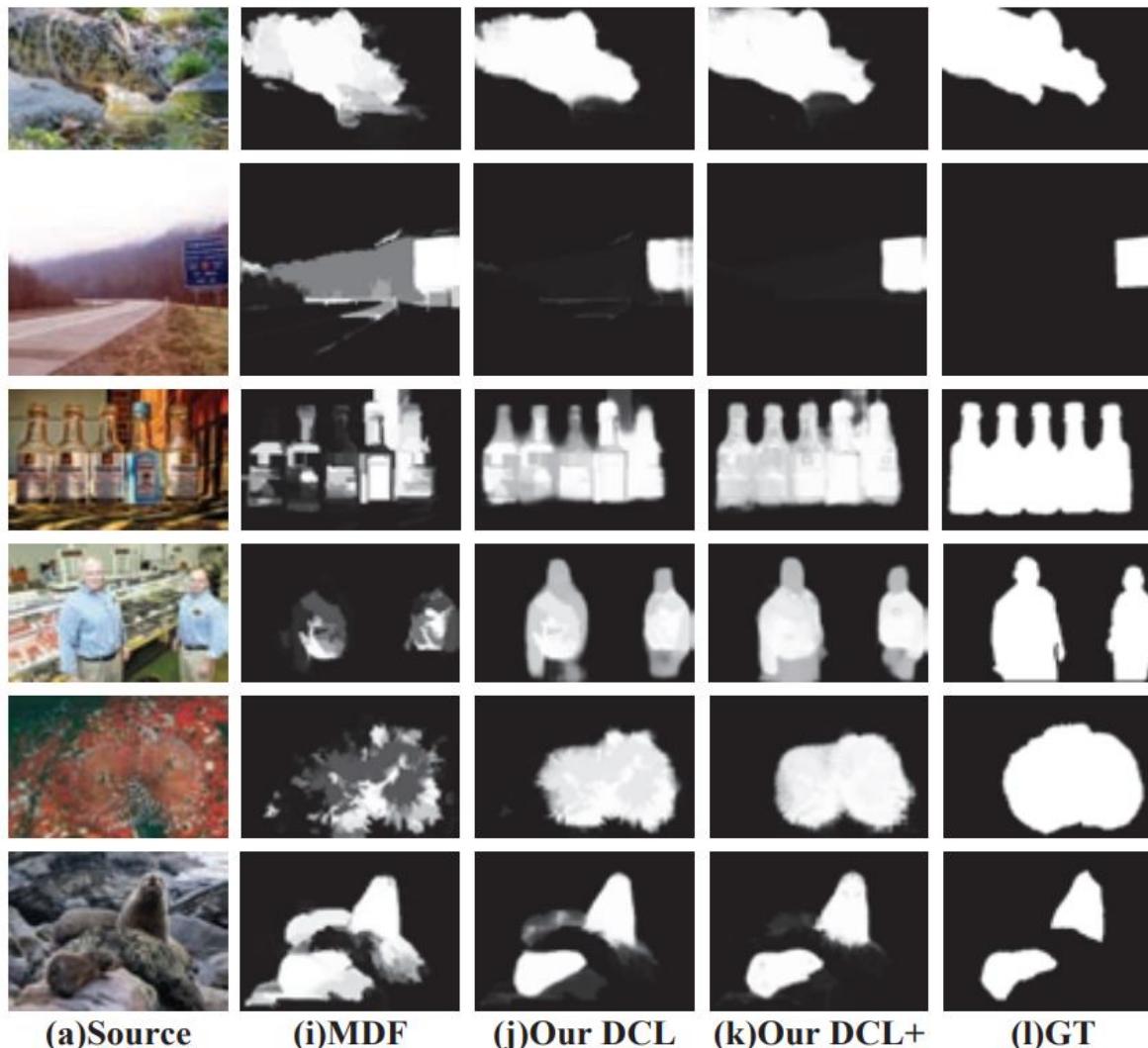
# CV Research in HCP Lab

## Middle level vision (Visual Saliency)



# CV Research in HCP Lab

## Salient Object Detection



# CV Research in HCP Lab

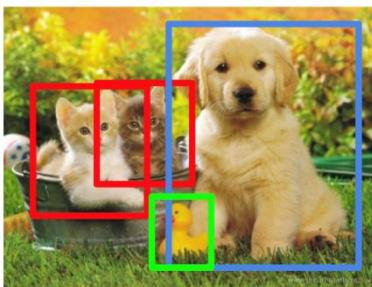
## High level vision (Image Understanding)

Classification



CAT

Object Detection



CAT, DOG, DUCK

Instance Segmentation

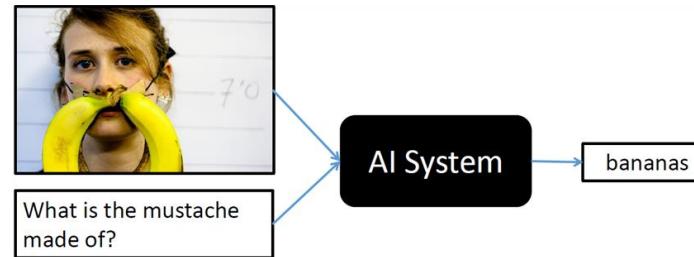


CAT, DOG, DUCK

man in green to right



## 高层视觉语义理解

A woman is throwing a frisbee in a park.

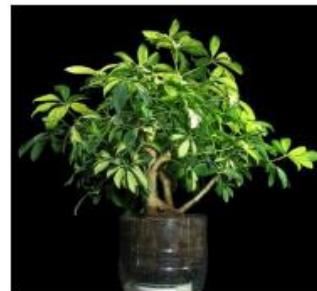
Image/Video Captioning

Visual Question Answering

Visual Grounding

# CV Research in HCP Lab

## Multi-label Image Recognition



Single-label images from ImageNet

horse&  
person



dog&  
person



Multi-label images from Pascal VOC

# CV Research in HCP Lab

## Semantic Segmentation for Human parsing



# CV Research in HCP Lab

---

virtual try-on



# CV Research in HCP Lab

## Semantic Segmentation for Medical Image Analysis



# CV Research in HCP Lab

视觉生成（虚拟试发）



# CV Research in HCP Lab

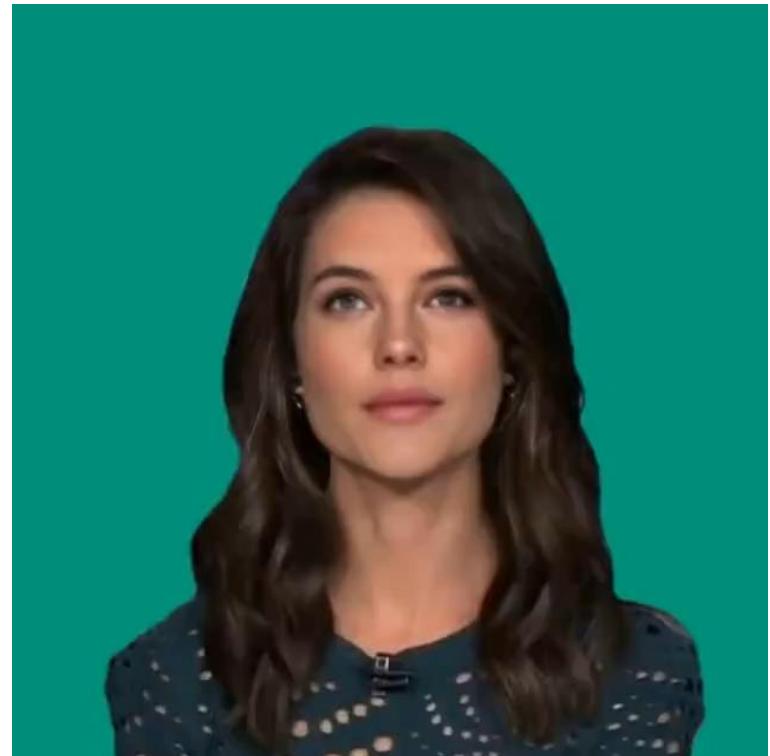
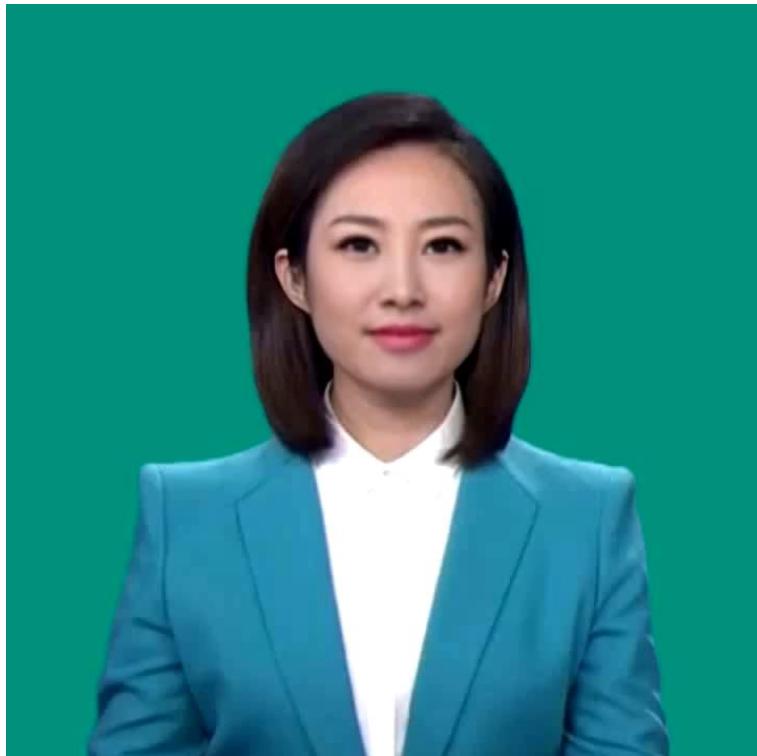
视觉生成（虚拟试发）



# CV Research in HCP Lab

---

Talking Video Generation



# CV Research in HCP Lab

---

Talking Video Generation

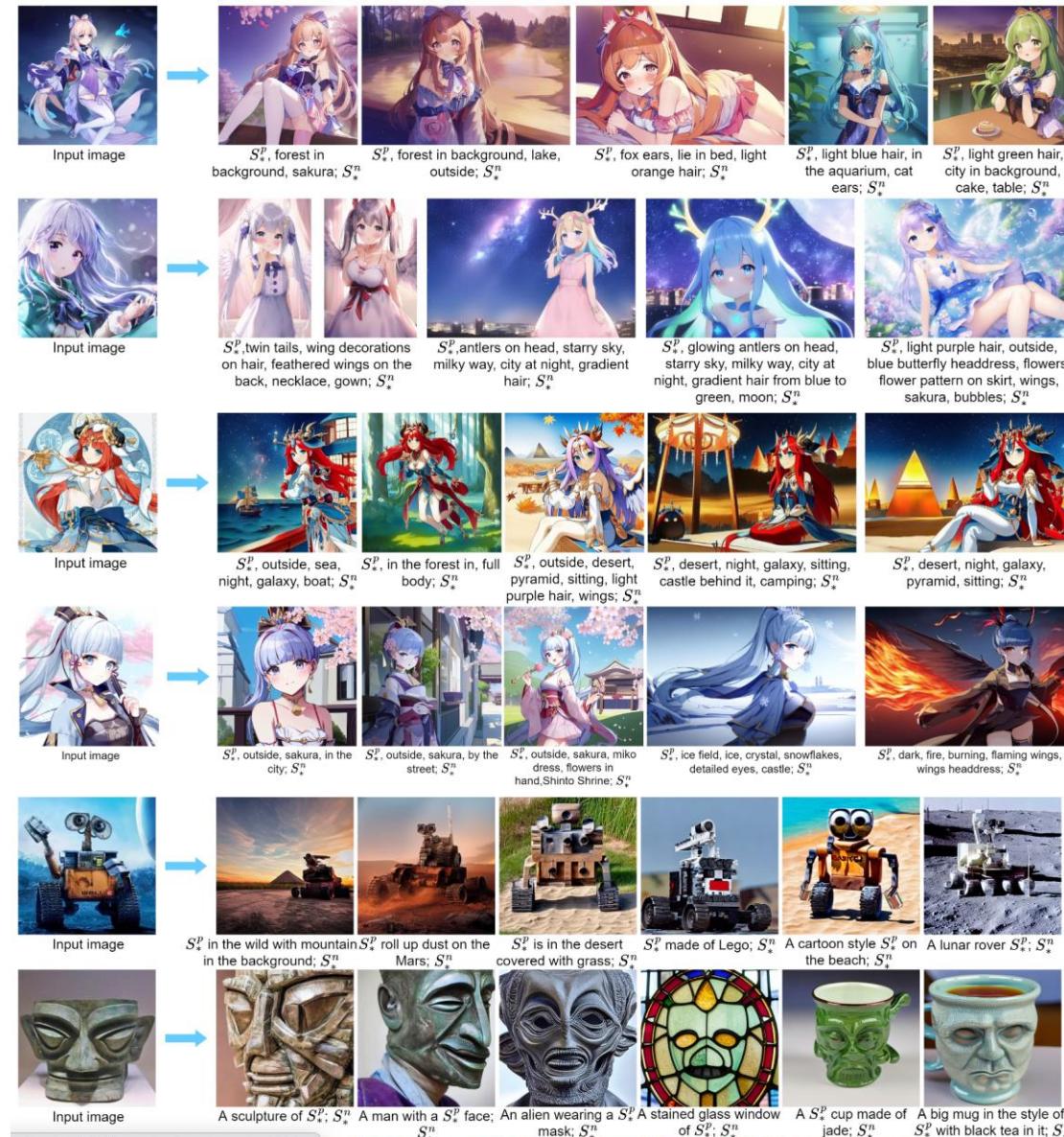


# CV Research in HCP Lab

---



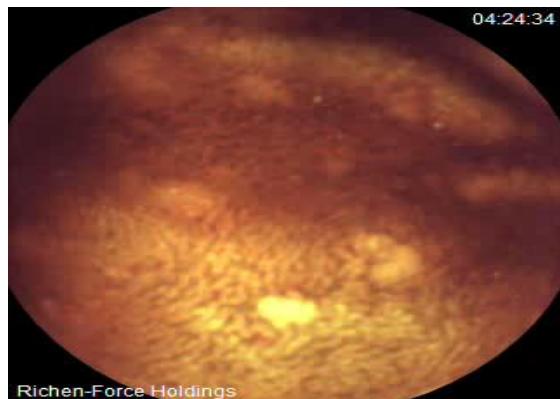
# Controllable Text-To-Image Generation



# CV Research in HCP Lab

# Semantic Segmentation for Medical Image Analysis

# X光小儿肺部分割 (MICCAI2019)



胶囊内镜

# 鲁棒医学影像分割 (AAAI2019)



直肠超声

# 3D胰腺分割 (MICCAI2019)

# 胃肠息肉分割 (MICCAI2020)

合作研究机构：  
中山附六、珠江医院、  
广州妇幼、深睿医疗

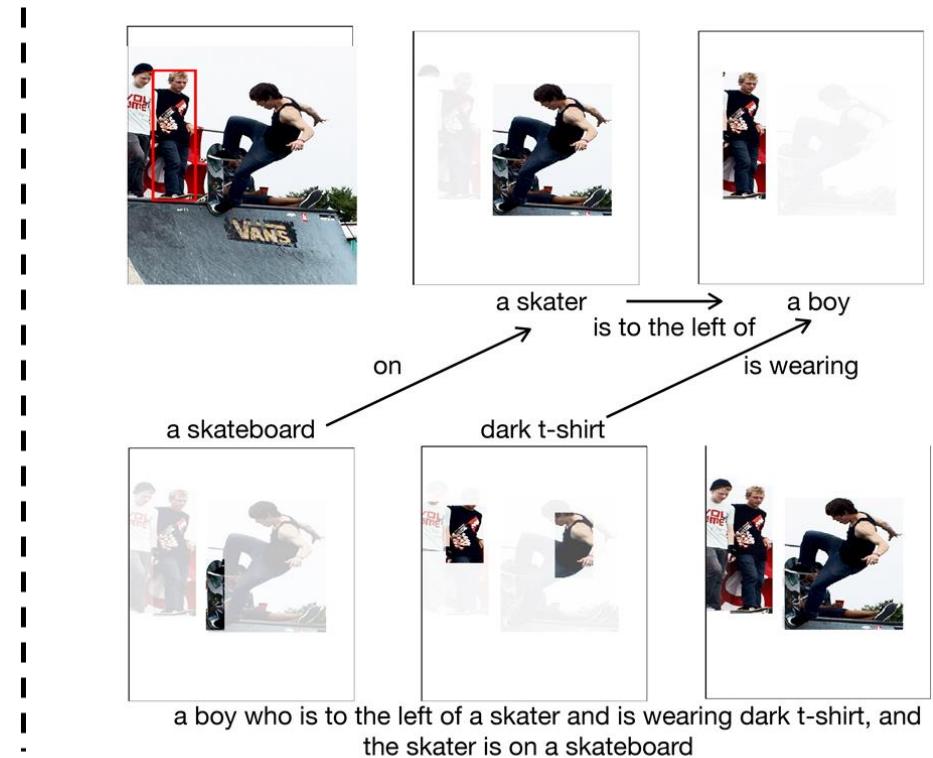
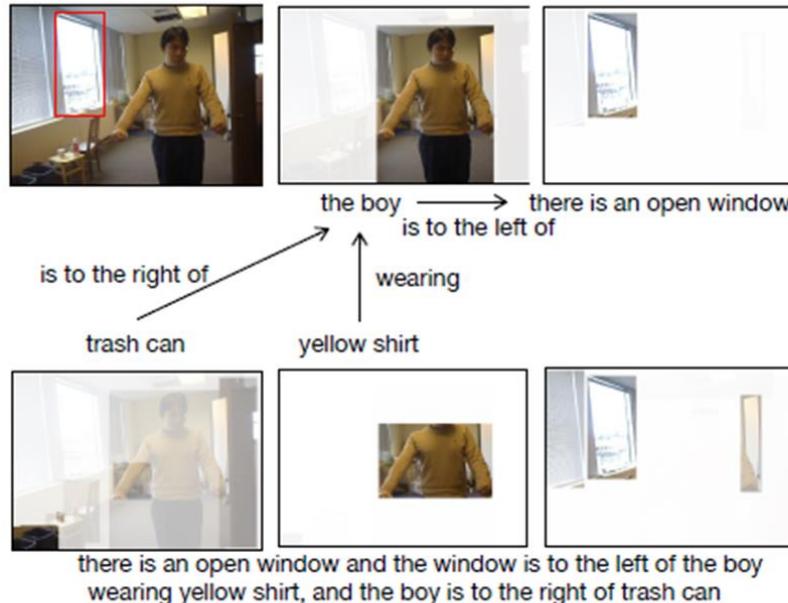
# CV Research in HCP Lab

## Visual Captioning



# CV Research in HCP Lab

## Visual Grounding



# More CV Research Topics...

---

- My home page: <http://guanbinli.com/pb-list.html>
- HCP Lab: <https://www.sysu-hcp.net/projects/cv/index.html>

If you are excited about computer vision research, feel free to contact me privately. Welcome to join the following research topics.

1. Face swap/Talking Head Video
2. Vision and Language / Prompt tuning
3. Data Efficient Visual Understanding
4. Diffusion model
5. Medical Image Processing

# Questions?