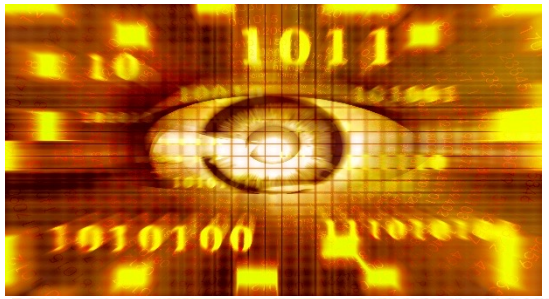
Computer Vision I:

Jingya Wang

Email: wangjingya@shanghaitech.edu.cn

https://faculty.sist.shanghaitech.edu.cn/faculty/wangjingya/



Today’s Class

- Specifics of this course
- What is Computer Vision?
- What is the applications of computer vision?
- Course Contents.

Prerequisites

- **Linear algebra**, basic calculus, and probability
- Experience with image processing or Matlab will help but is not necessary
- **Calculus**
- **Programing**

ST

how to learn computer vision

Here are some steps you can follow to learn computer vision:

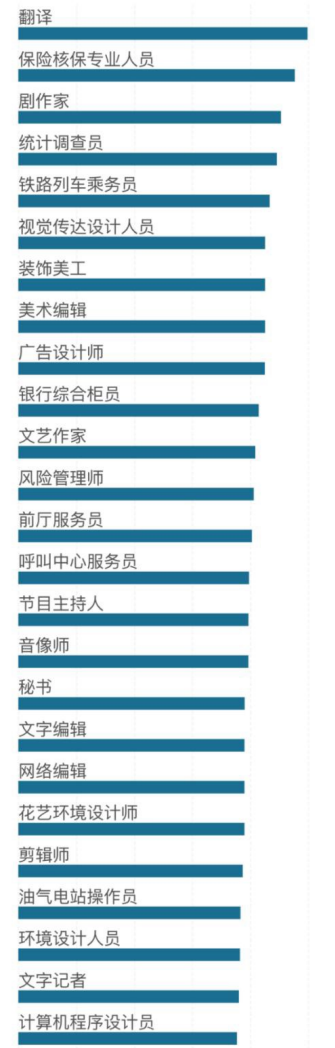
1. Start with the basics: Study linear algebra, calculus, and probability theory, as these are foundational concepts in computer vision.
2. Read introductory textbooks or online resources, such as "Computer Vision: Algorithms and Applications" by Richard Szeliski or the online course "Introduction to Computer Vision" by Georgia Tech.
3. Practice with computer vision software such as OpenCV, scikit-image, or MATLAB.
4. Work on small projects to apply the concepts you've learned, such as image classification, object detection, and image segmentation.
5. Participate in online forums and attend workshops and conferences related to computer vision to stay up-to-date on the latest developments and techniques in the field.
6. Study deep learning and its applications in computer vision, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs).

It's also helpful to have a good understanding of programming and software development, as well as experience with one or more programming languages such as Python, C++, or MATLAB.

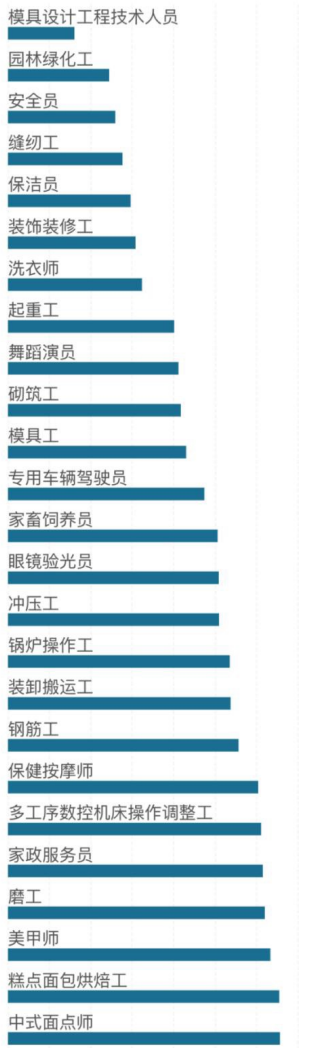
AI替代率最高和最低TOP25

下表列出了AI替代率最高和最低的各25个职业。

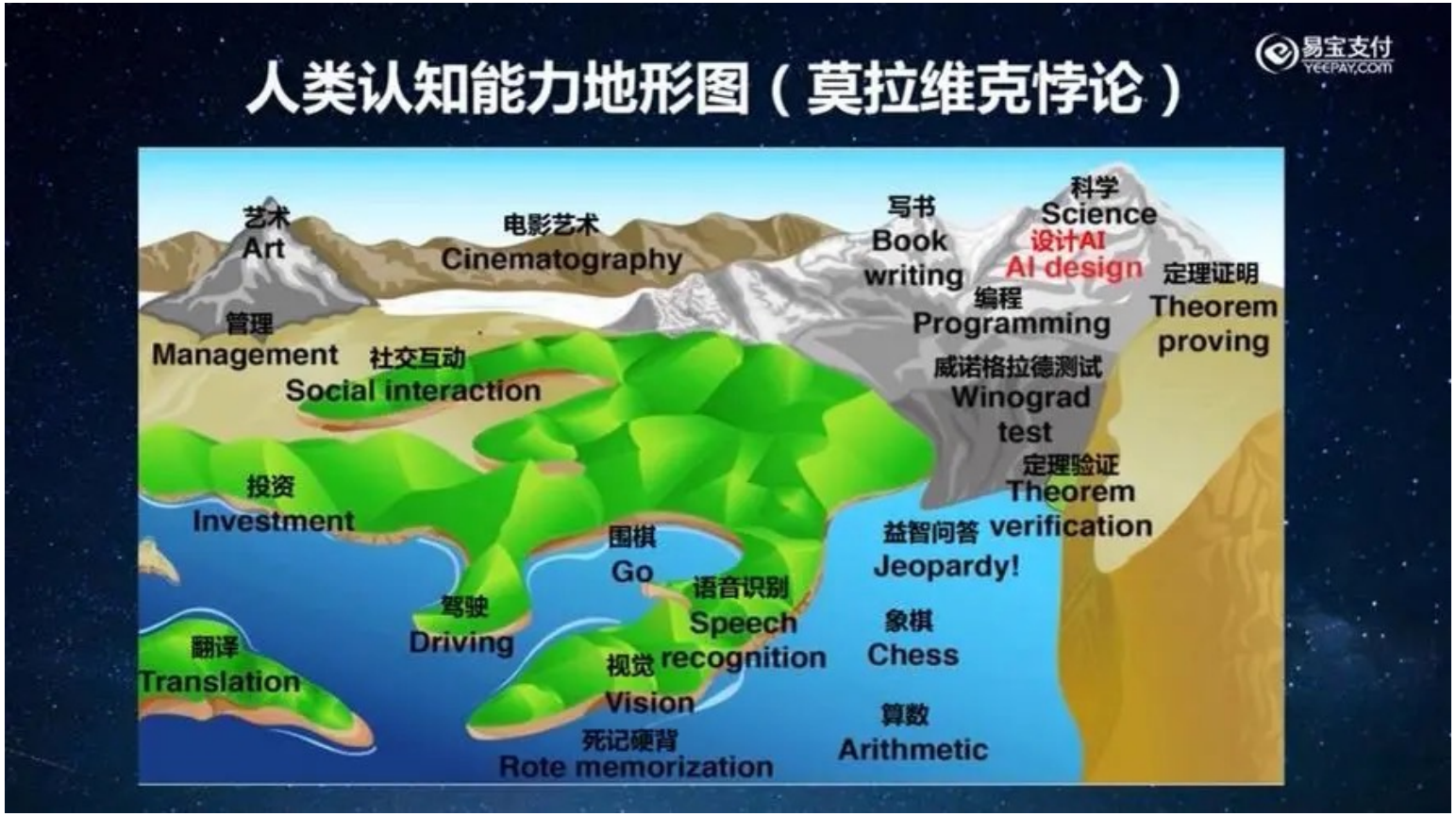
替代率最高的前25名职业



替代率最低的前25名职业



城市数据团



Today’ CV

- **NVIDIA GTC**
  https://www.youtube.com/watch?v=sxHFDKwJGGo
- **Sora**
  https://www.youtube.com/watch?v=XllmgXBQUwA
- **Mobile ALOHA**
  https://www.youtube.com/watch?v=LZ5chzQ9r7s

Why 2023 was the most exciting year in computer vision history (so far)

https://voxel51.com/blog/why-2023-was-the-most-exciting-year-in-computer-vision-history-so-far/

Computer Vision and Nearby Fields

- Computer Graphics: Models to Images
- Comp. Photography: Images to Images
- Computer Vision: Images to Models

- Deep Learning
- Machine Learning

Course Evaluation

- **Assignment (10%)**
  - Assignment 1 release in week 5 (DDL week 8)
  - Assignment 2 release in week 9 (DDL week 12)
- **Small quiz (including course attendance,10%)**
- **Course project (20%)**
  - Maximum 5 people, release in week 7 (DDL week 16)
  - Submit PPT, demo, code and report
- **Final exam(60%)**

Late Policy

- All students have 5 free late days with no penalty.
- You may use up to 3 late days per assignment with no penalty.
- You may not use late days for the final project.
- Once you have exhausted your free late days, we will deduct a late penalty of 10% per additional late day

上海科技大学学生学术诚信规范与管理办法

第三章 学术不端行为的认定

第二十五条 具有以下行为之一者原则上应认定为学术不端：

（一）侵占、抄袭、剽窃他人作业、论文、报告、程序和研究数据等成果；允许或协助他人侵占、抄袭、剽窃自己或他人作业、论文、报告、程序和研究数据等成果；请他人代写或代他人撰写作业、论文、报告、程序等；购买作业、论文、报告、程序、研究数据等作为自己的成果。

（二）代他人签到，冒名顶替他人上课或参加学校要求的其他培养环节；

（三）在考试中采用抄袭或拷贝等不正当手段，窃取他人答案或答题资料等；请他人代替考试、替他人参加考试；偷窃试卷或采用其他各种手段窃取试卷内容者；使用通讯设备或网络手段作弊，组织或参与作弊及其他作弊行为等。

（四）篡改、伪造自己或他人的原始研究数据（包括试验数据、调查数据和软件计算结果等）和研究成果等；隐瞒不利数据用于伪造创新成果和新发现；在接受必要的调查时无法提供原始试验记录；未经严格论证主观臆造学术结论等。

（五）直接引用他人成果、观点、方案、资料、数据等而未标明出处或论文雷同率（包括与直接翻译的成果比对）偏高。

- 抄袭作业(同学作业或者互联网资源)：当次作业记为0分，发现两次及以上交由学院学术委员会处理。
- 考试作弊直接交由学院学术委员会处理。
- Quiz不接受补交（假条须由书院老师认可签字, 成绩不计）

课件发布：blackboard
课程讨论：微信群（助教负责答疑）

My Office Hour ： Tuesday 3-4pm 1c-303b
TA Office Hour : Thursday 8-9 pm 1a-402 or 1a-403

计算机视觉1-2024

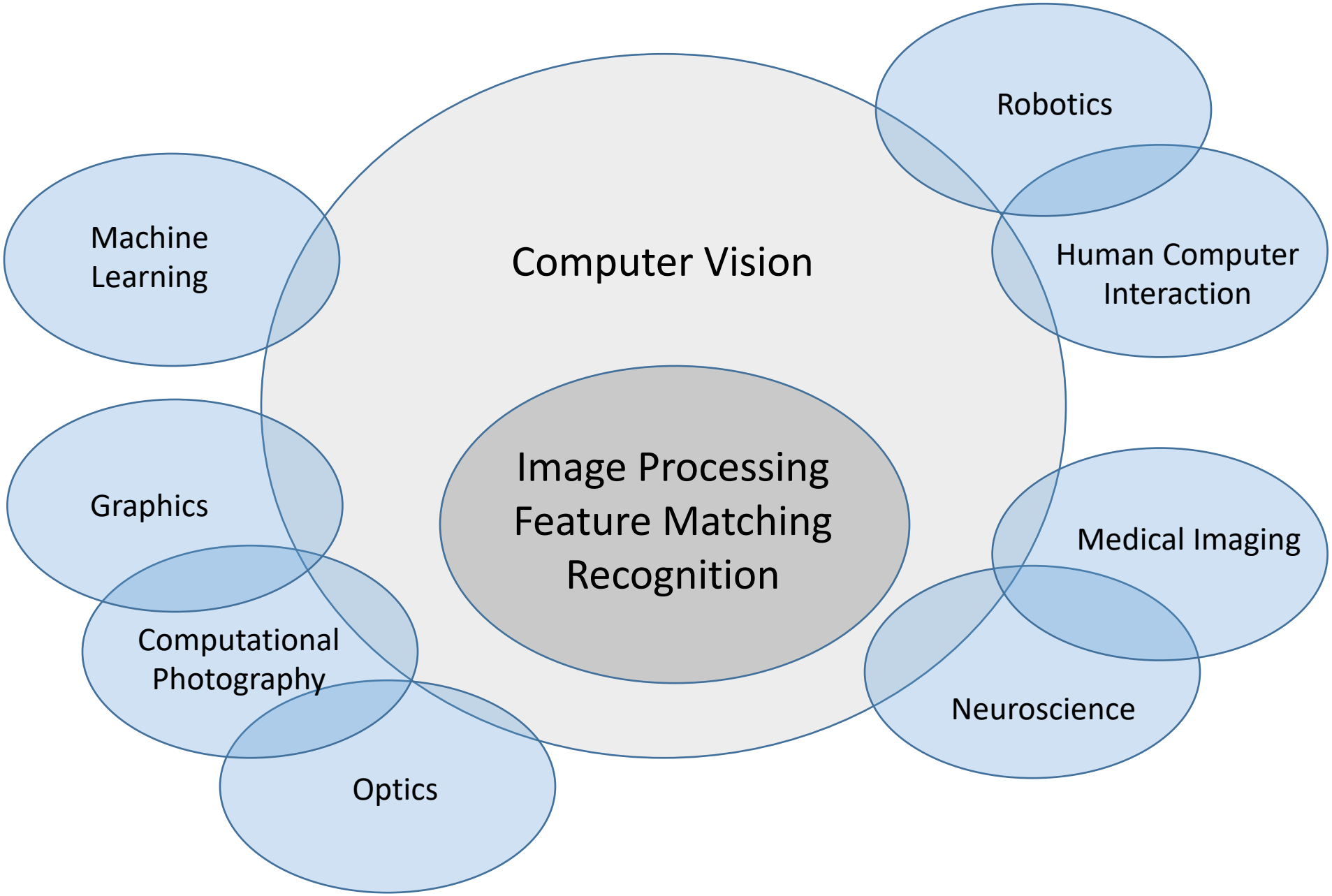
该二维码7天内(3月4日前)有效，重新进入将更新

Online sources：

- Stanford CS231n: Convolutional Neural Networks for Visual Recognition
- Stanford CS231a: Computer Vision: From 3D Reconstruction to Recognition
- Andrew Ng: Deep Learning

Books：

- 1. Richard Szeliski, Computer Vision: Algorithms and Applications, 2010
- 2. Richard Szeliski, Computer Vision: Algorithms and Applications, 2nd ed， 2022
- 3 . [中文版]计算机视觉——算法与应用，艾海舟[译]，清华大学出版社，2012
- 4. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2017


Robotics
Machine Learning
Computer Vision
Human Computer Interaction
Image Processing
Feature Matching
Recognition
Graphics
Medical Imaging
Computational Photography
Neuroscience
Optics

Textbook

Computer Vision: Algorithms and Applications

© 2010 Richard Szeliski, Microsoft Research

http://szeliski.org/Book/

Textbook

Computer Vision: Algorithms and Applications, 2nd ed.

http://szeliski.org/Book/

Course Syllabus (tentative)

- **I. Recognition**
    - Image classification
    - Bag of words based image representation
    - Deep convolutional neural networks

- **II. Image formation and low-level vision**
    - Camera models
    - Light and color
    - Linear filters and edges
    - Feature extraction
    - Optical flow and feature tracking

Course Syllabus (tentative)

- **III. Geometric and 3D vision**
    - Camera calibration
    - Epipolar geometry

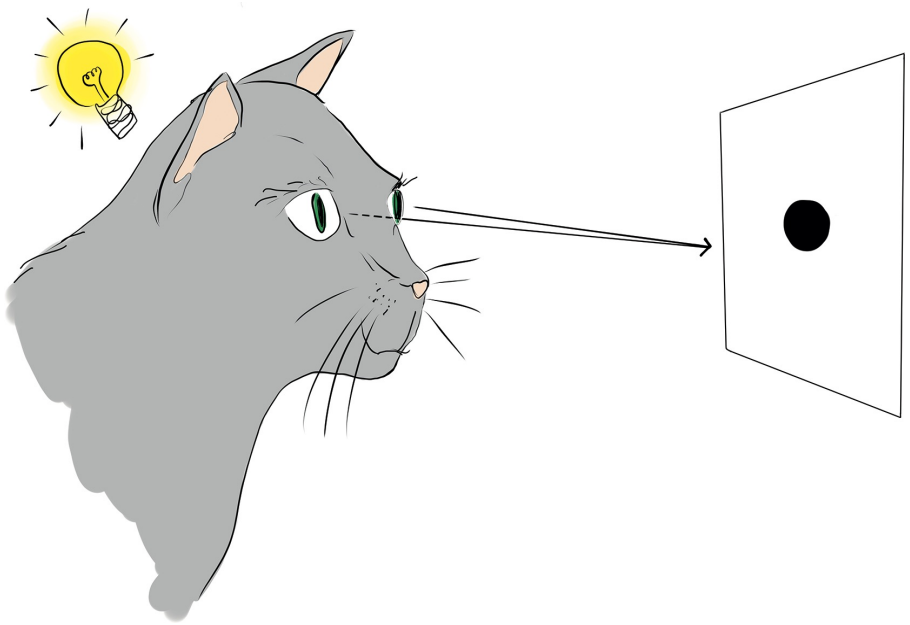
- **IV. Advance CV topics**

TA tutorial

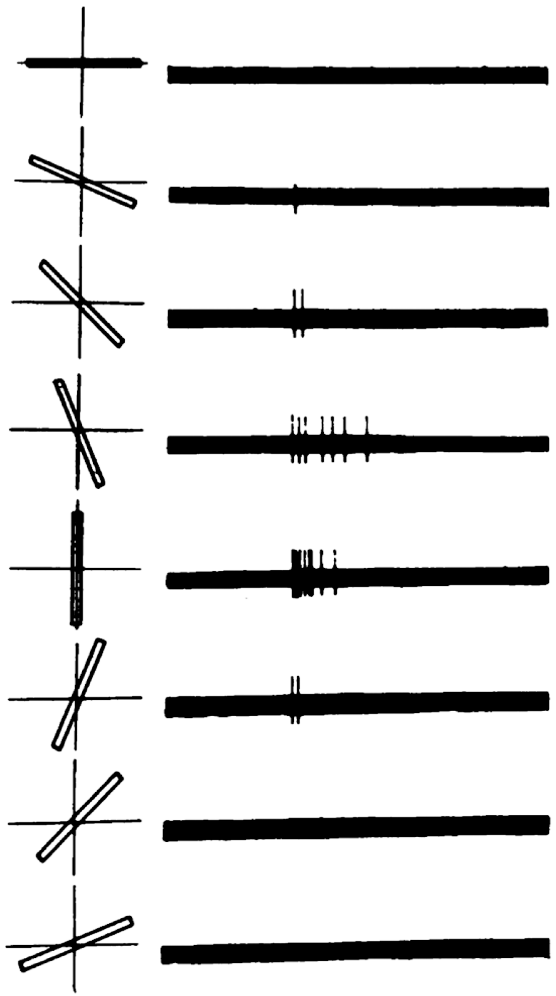
Start from week 4

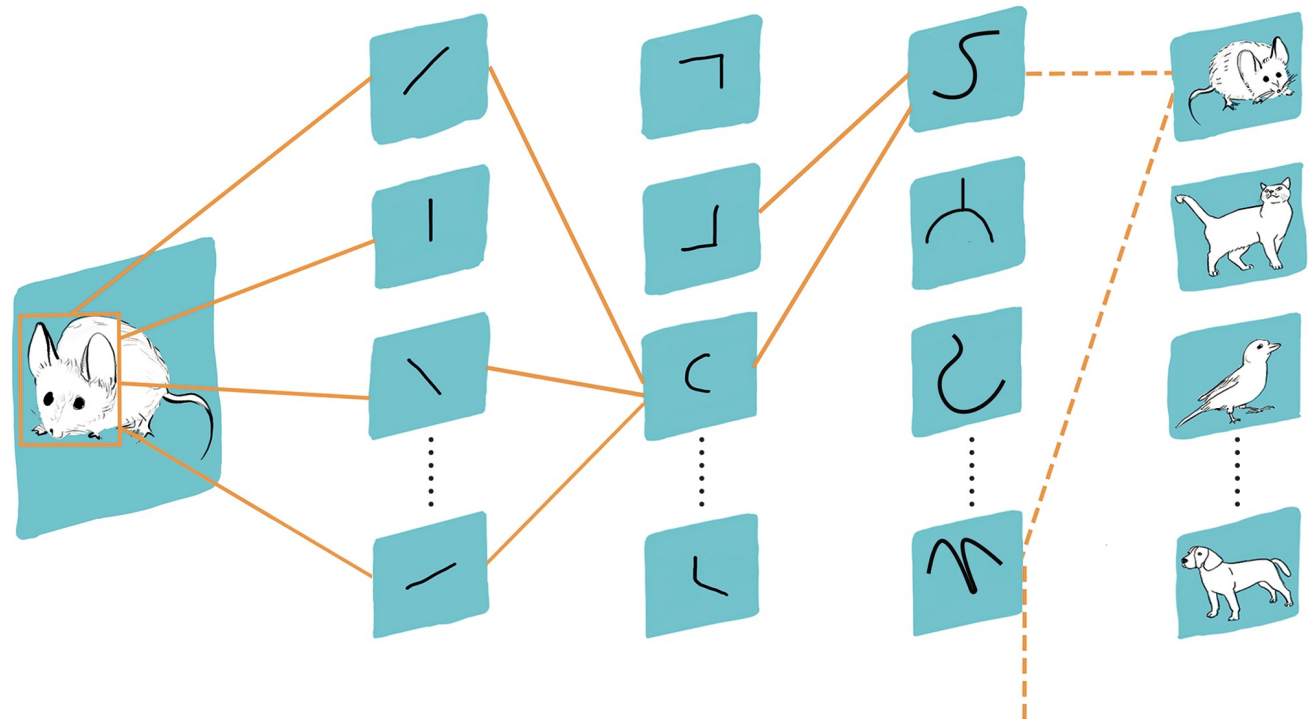
- AI server
- pytorch
- Github
- Advanced CV topics …

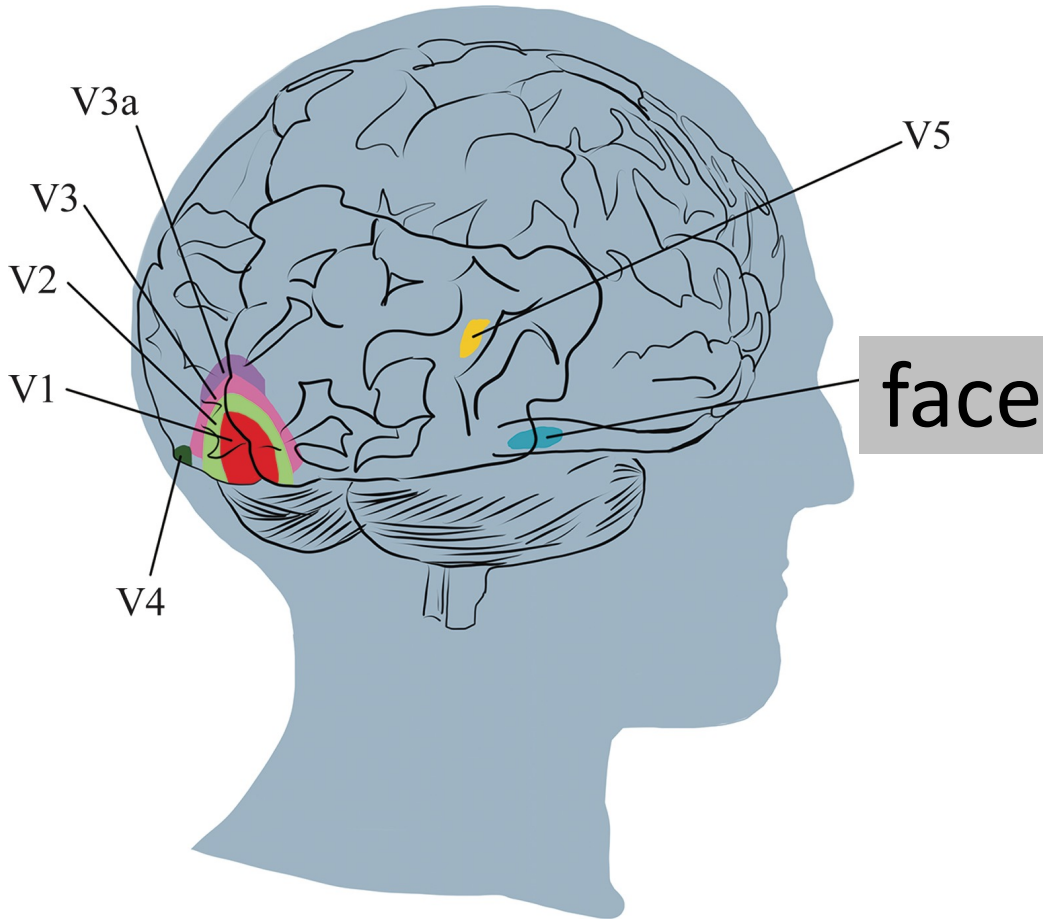
What is Human Vision?



David Hubel， Torsten Wiesel 1959



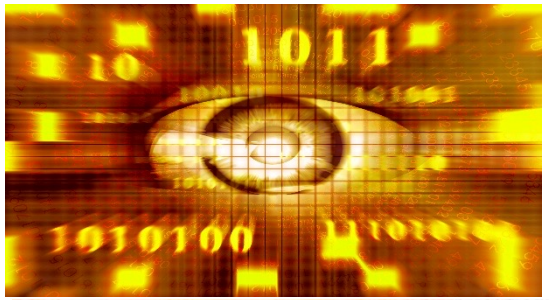



V3a
V3
V2
V1
V4
V5
face

What is Computer Vision?

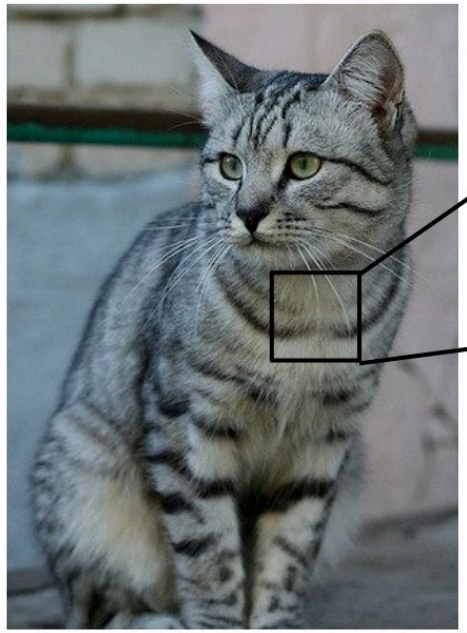
Image Classification: A core task in Computer Vision

cat



What is Computer Vision?

The Problem: Semantic Gap

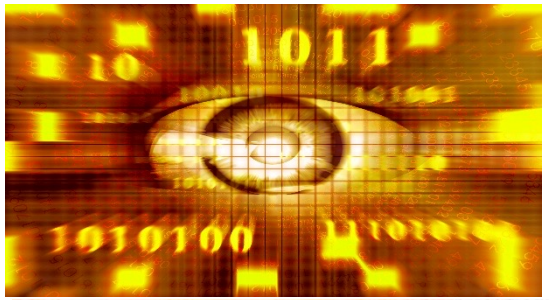



```
[[105 112 108 111 104  99 106  99  96 103 112 119 104  97  93  87]
 [ 91  98 102 106 104  79  98 103  99 105 123 136 110 105  94  85]
 [ 76  85  90 105 128 105  87  96  95  99 115 112 106 103  99  85]
 [ 99  81  81  93 120 131 127 100  95  98 102  99  96  93 101  94]
 [106  91  61  64  69  91  88  85 101 107 109  98  75  84  96  95]
 [114 108  85  55  55  69  64  54  64  87 112 129  98  74  84  91]
 [133 137 147 103  65  81  80  65  52  54  74  84 102  93  85  82]
 [128 137 144 140 109  95  86  70  62  65  63  63  60  73  86 101]
 [125 133 148 137 119 121 117  94  65  79  80  65  54  64  72  98]
 [127 125 131 147 133 127 126 131 111  96  89  75  61  64  72  84]
 [115 114 109 123 150 148 131 118 113 109 100  92  74  65  72  78]
 [ 89  93  90  97 108 147 131 118 113 114 113 109 106  95  77  80]
 [ 63  77  86  81  77  79 102 123 117 115 117 125 125 130 115  87]
 [ 62  65  82  89  78  71  80 101 124 126 119 101 107 114 131 119]
 [ 63  65  75  88  89  71  62  81 120 138 135 105  81  98 110 118]
 [ 87  65  71  87 106  95  69  45  76 130 126 107  92  94 105 112]
 [118  97  82  86 117 123 116  66  41  51  95  93  89  95 102 107]
 [164 146 112  80  82 120 124 104  76  48  45  66  88 101 102 109]
 [157 170 157 120  93  86 114 132 112  97  69  55  70  82  99  94]
 [130 128 134 161 139 100 109 118 121 134 114  87  65  53  69  86]
 [128 112  96 117 150 144 120 115 104 107 102  93  87  81  72  79]
 [123 107  96  86  83 112 153 149 122 109 104  75  80 107 112  99]
 [122 121 102  80  82  86  94 117 145 148 153 102  58  78  92 107]
 [122 164 148 103  71  56  78  83  93 103 119 139 102  61  69  84]]
```

What the computer sees

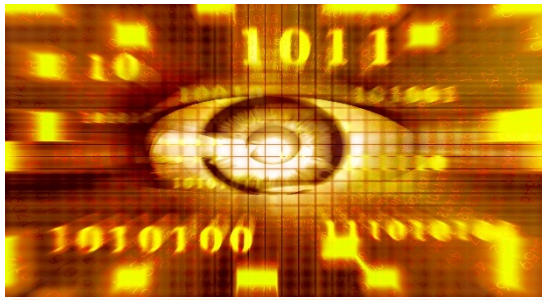
An image is a tensor of integers between [0, 255]:

e.g. 800 x 600 x 3
(3 channels RGB)



What is Computer Vision?

- What are examples of computer vision being used in the world?



Computer Vision

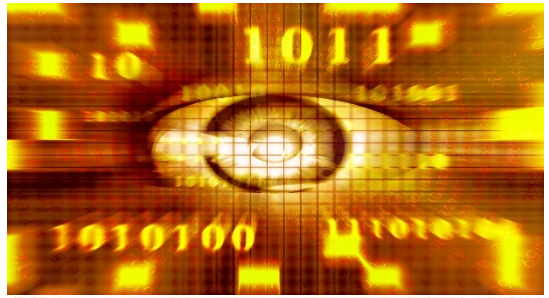
Make computers understand images and video.

What kind of scene?

Where are the cars?

How far is the building?

…

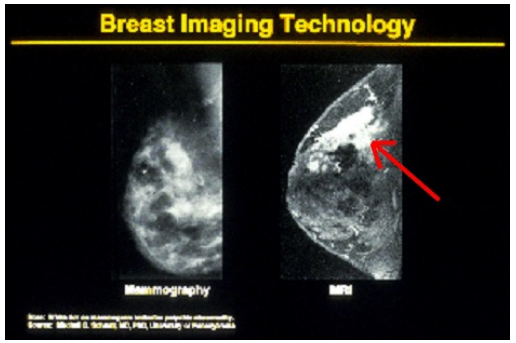


Vision is really hard

- Vision is an amazing feat of natural intelligence
  - Visual cortex occupies about 50% of brain
  - More human brain devoted to vision than anything else

Why computer vision matters

Safety

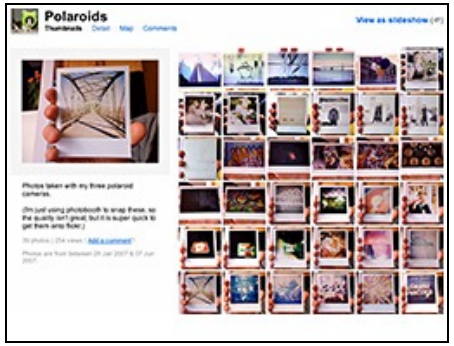


Health

Security

Comfort

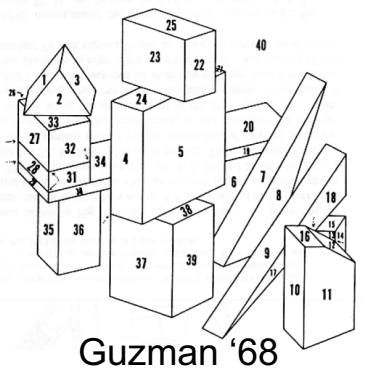
Fun



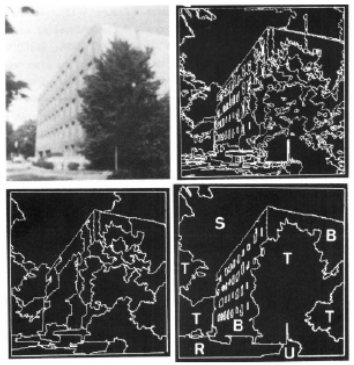
Access

Brief history of computer vision

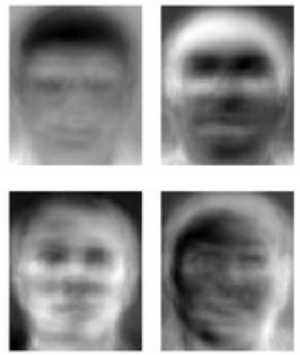
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: deep learning greatly boosts the performance of computer vision.
- 2030's: robot vision uprising?



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

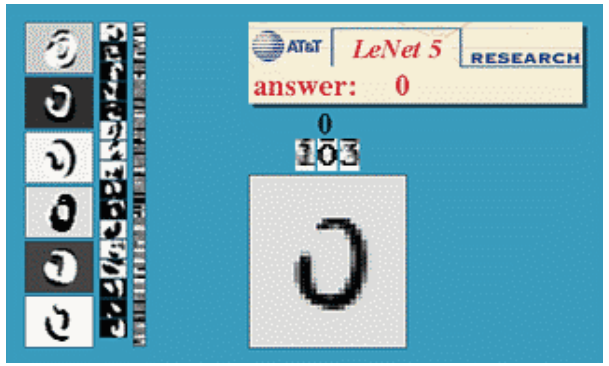
The Applications of Computer Vision

- Optical character recognition
- Face detection
- Image editing
- Medical image analysis
- . . .

Optical character recognition (OCR)

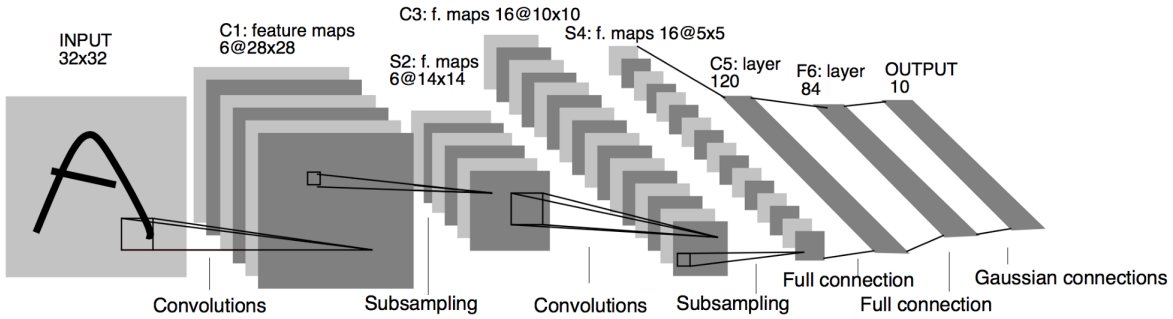
Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

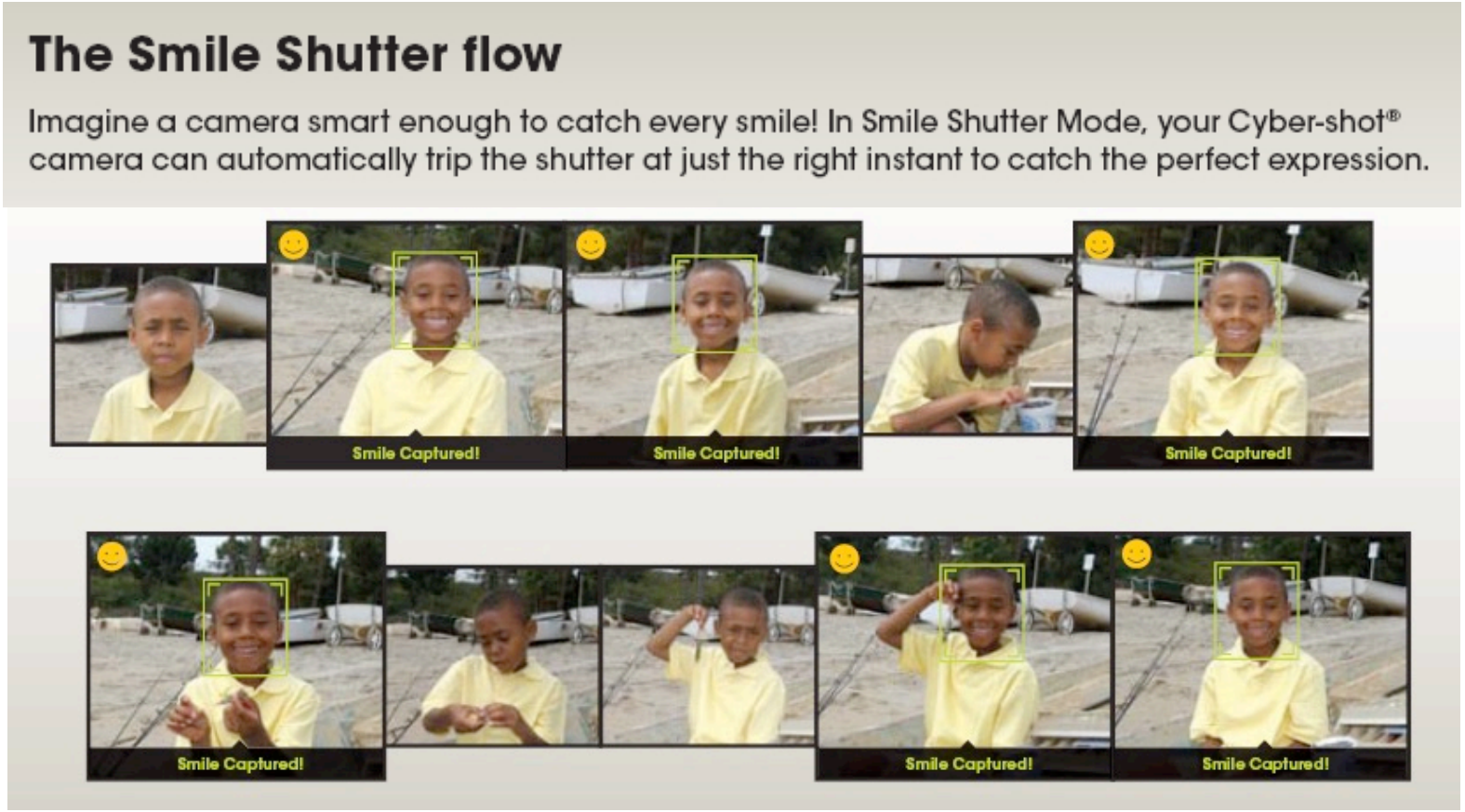


LeNet-5 [1998, paper by LeCun et al.]

Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

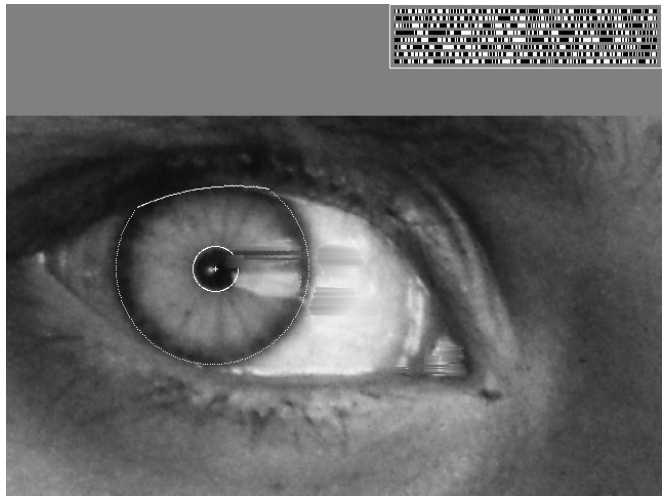
Smile detection

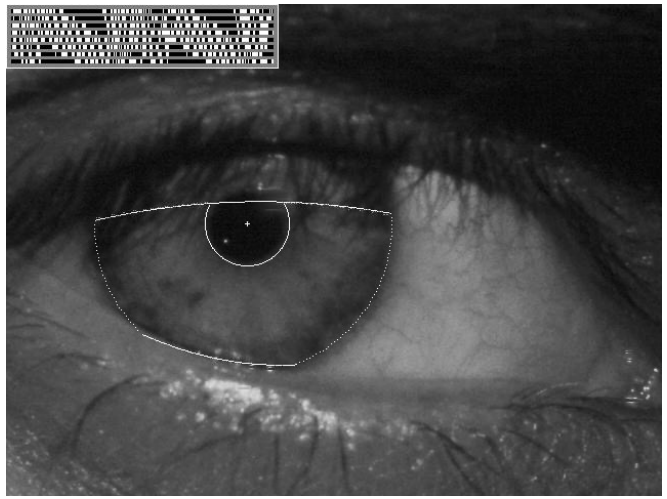


[Sony Cyber-shot® T70 Digital Still Camera]

Vision-based biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story wikipedia





Login without a password…

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

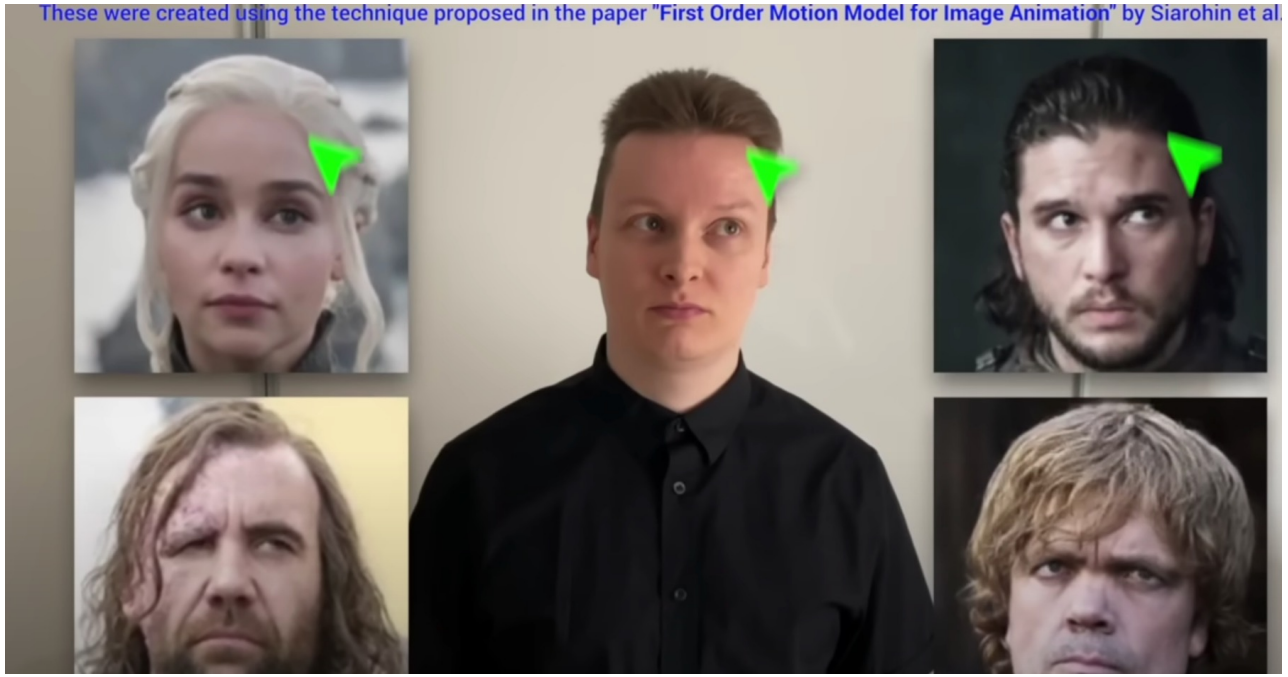
Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

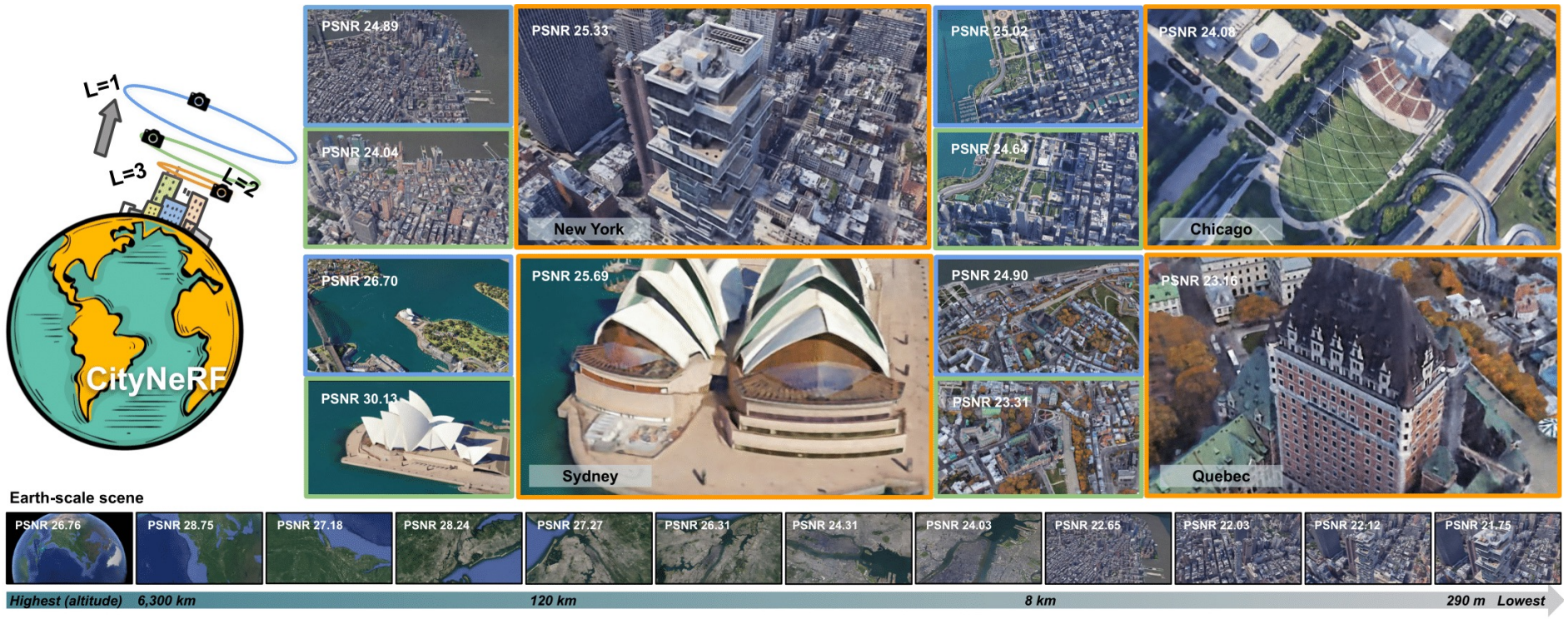
[Google Goggles]()

Face expression transfer

Megapixel DeepFakes
https://www.youtube.com/watch?v=JkUF40kPV4M



City Scale 3D Modeling



https://www.youtube.com/watch?v=TaVOTvN8CpM

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

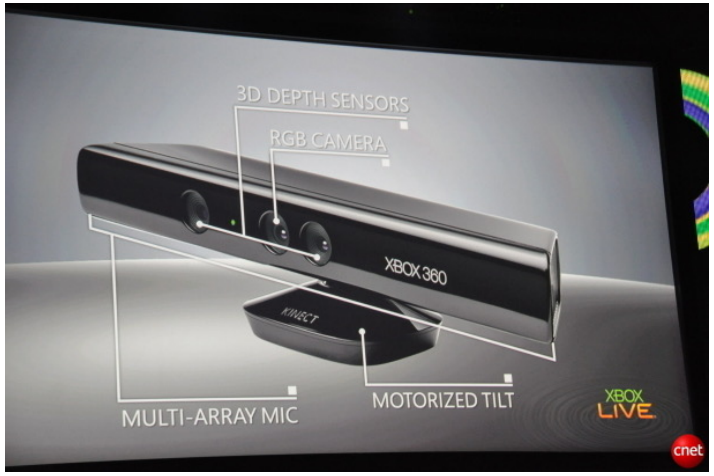
Pirates of the Carribean, Industrial Light and Magic

Sports

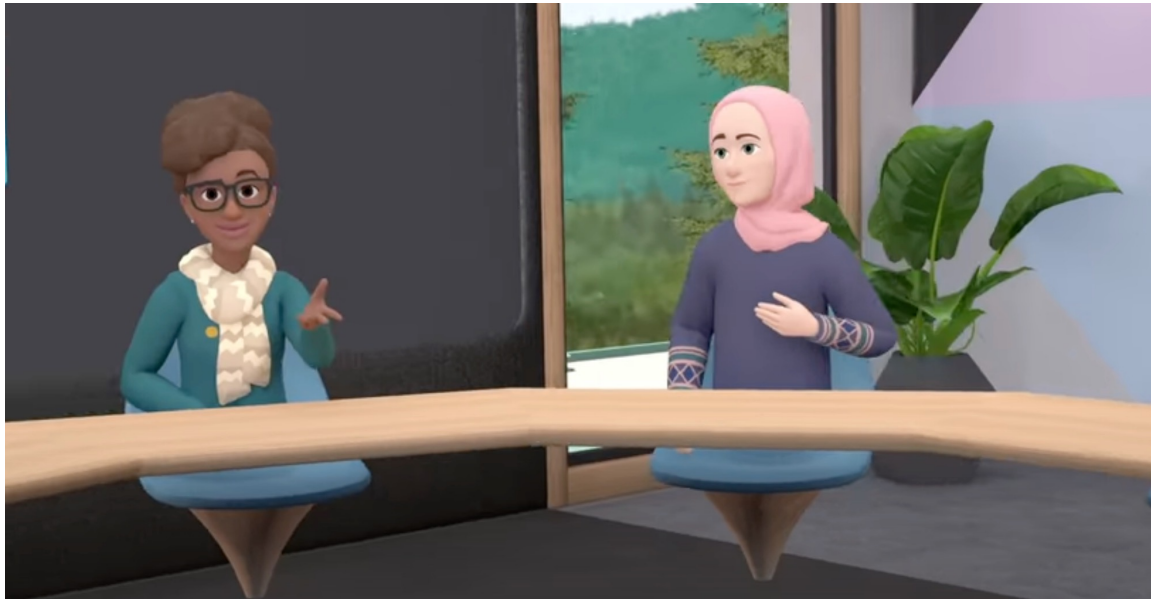
Sportvision first down line
Nice explanation on www.howstuffworks.com

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



Horizon Workrooms

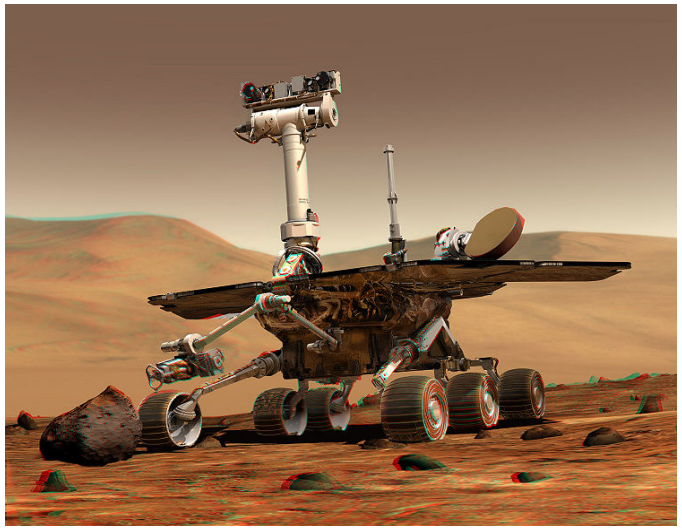


https://www.youtube.com/watch?v=lgj50IxRrKQ
https://www.youtube.com/watch?v=Rncz85tVt5I

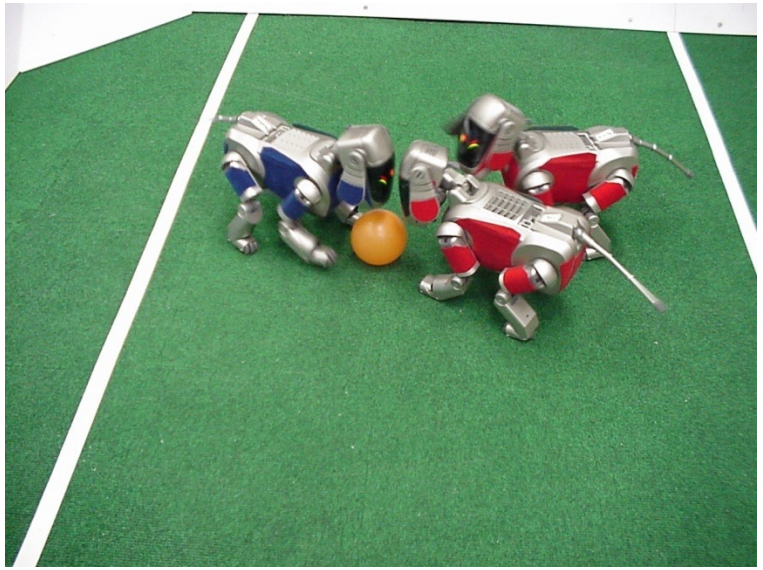
Industrial robots

Vision-guided robots position nut runners on wheels

Mobile robots



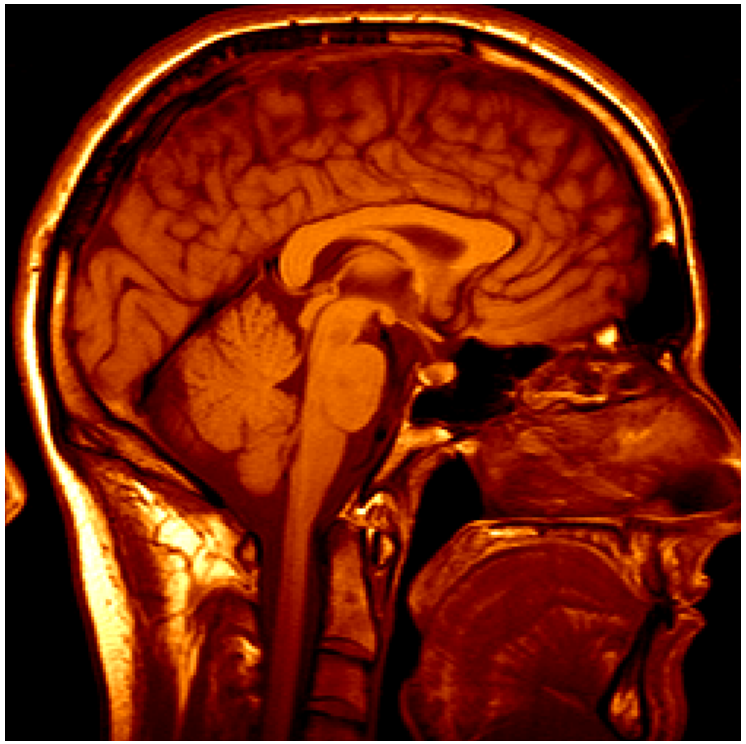
NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover



http://www.robocup.org/

Saxena et al. 2008
STAIR at Stanford

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Case Study: self-driving cars

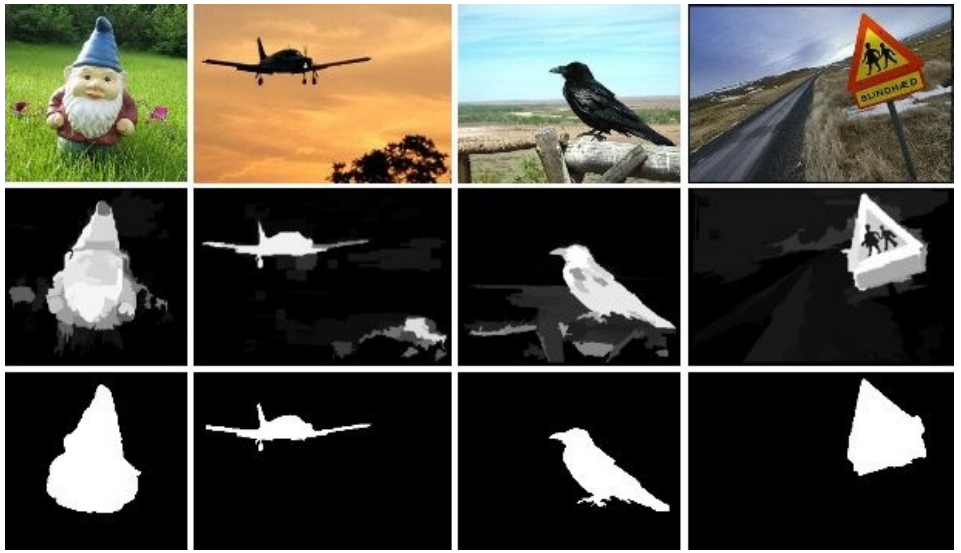
- Google Driverless Car
- Tesla Model S
- Minieye
- https://www.youtube.com/watch?v=PRg5RNU_JLk

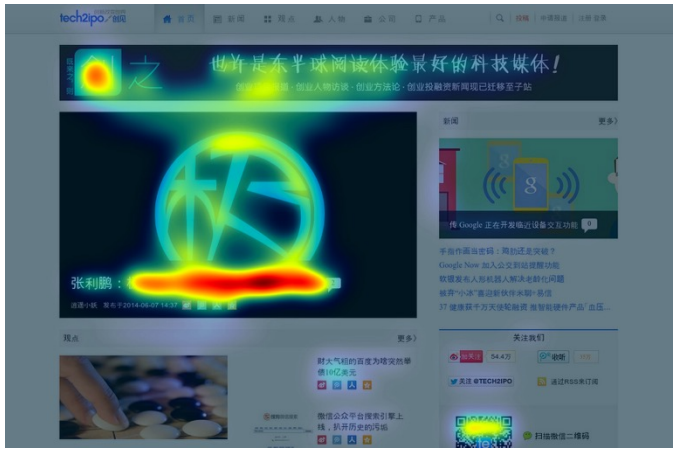
CV subfields:

- **saliency detection**
  - **unsupervised saliency detection/salient object discovery)**
- **segmentation**
  - **(unsupervised image segmentation, semantic segmentation)**
- **object detection**
  - **(face detection/general object detection, supervised methods)**
- **object recognition**
  - **(face recognition/general image classification, (un)supervised methods)**
- **image generation**
  - **Generative adversarial network, style transfer**
- **video processing**
  - **(tracking/event classification(supervised))**

Saliency detection







Application of computer vision

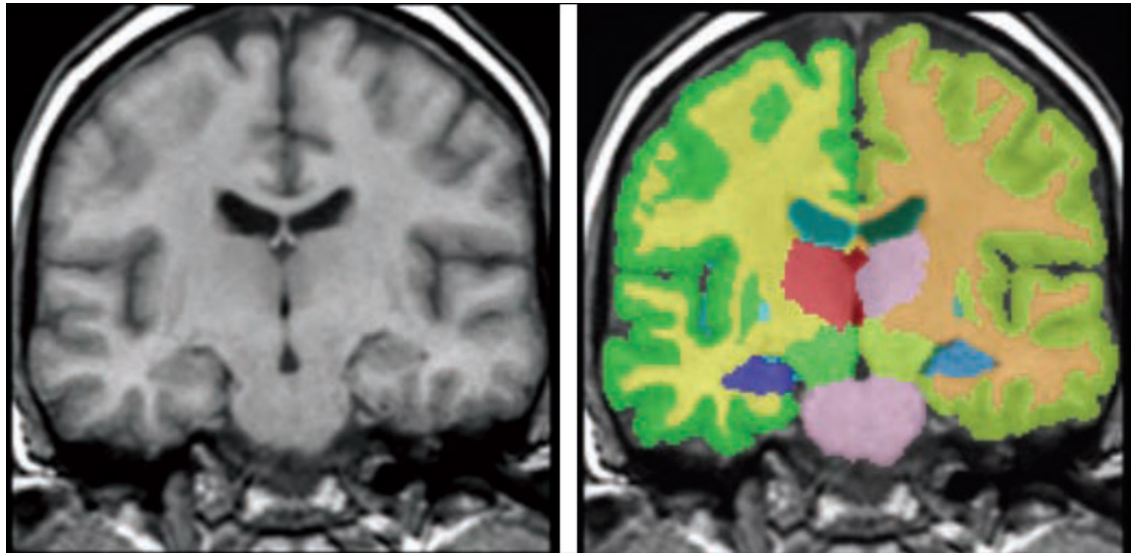
Content based image resizing (Image Retargeting)

https://www.youtube.com/watch?v=6NcIJXTlugc

Website Design

Image/Video Compression

Image Segmentation



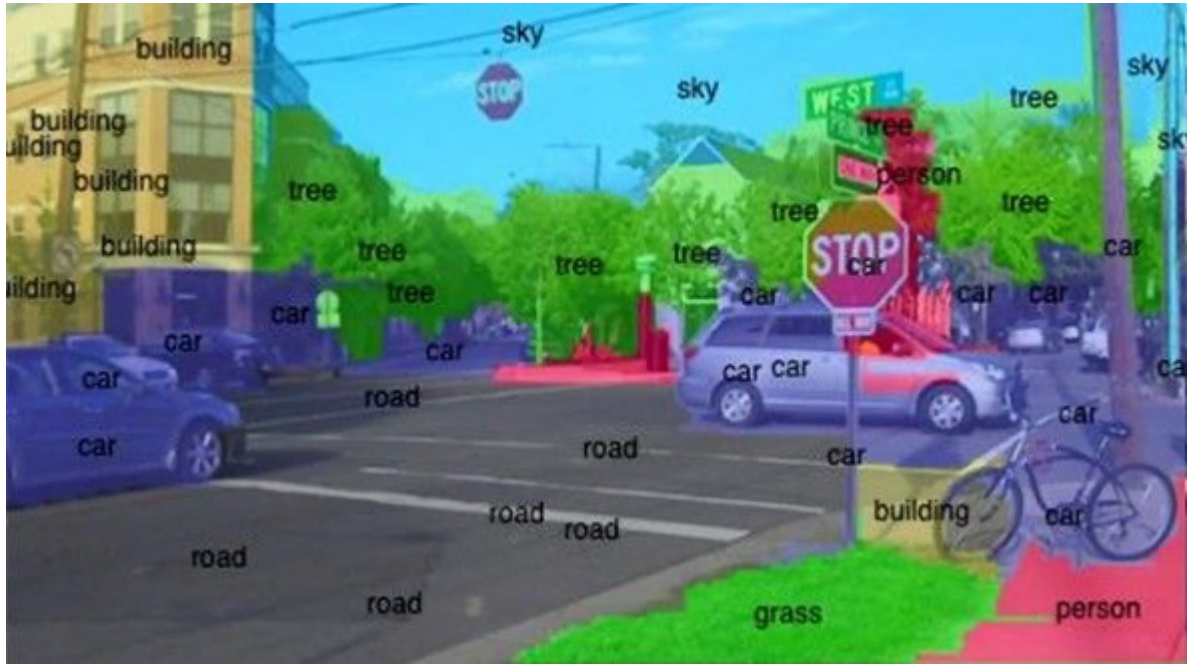
(a) Color Labels (ACA)

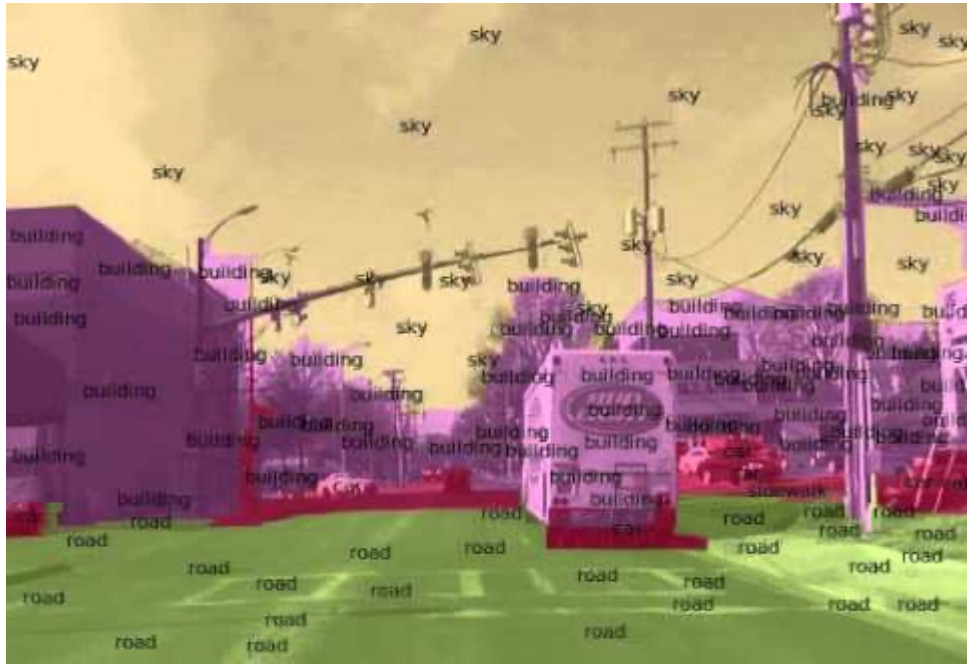
(b) Texture Classes

(c) Crude Segmentation

(d) Final Segmentation

Semantic segmentation





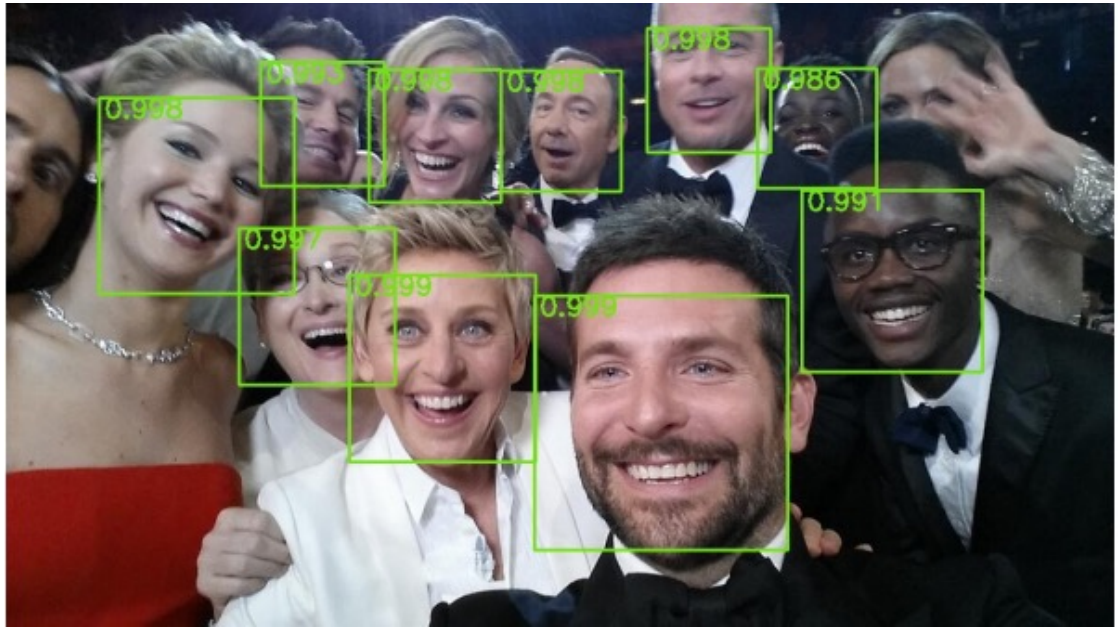
Application

Medical Image Analysis: MRI/CT/PET

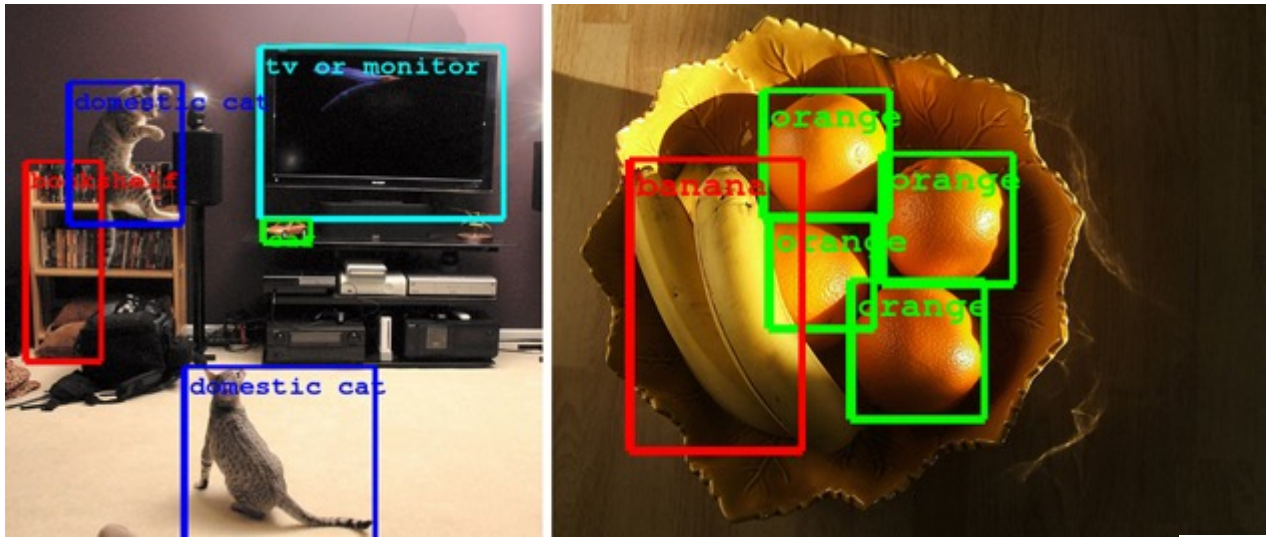
Image retrieval

Finding Images by Sketching

Object Detection



Object recognition



Face recognition

Face identification

Who?

A
B
C
...
...

Multiclass classification

Object Identification

Face verification

Same person or not.

Binary Result

Image generation

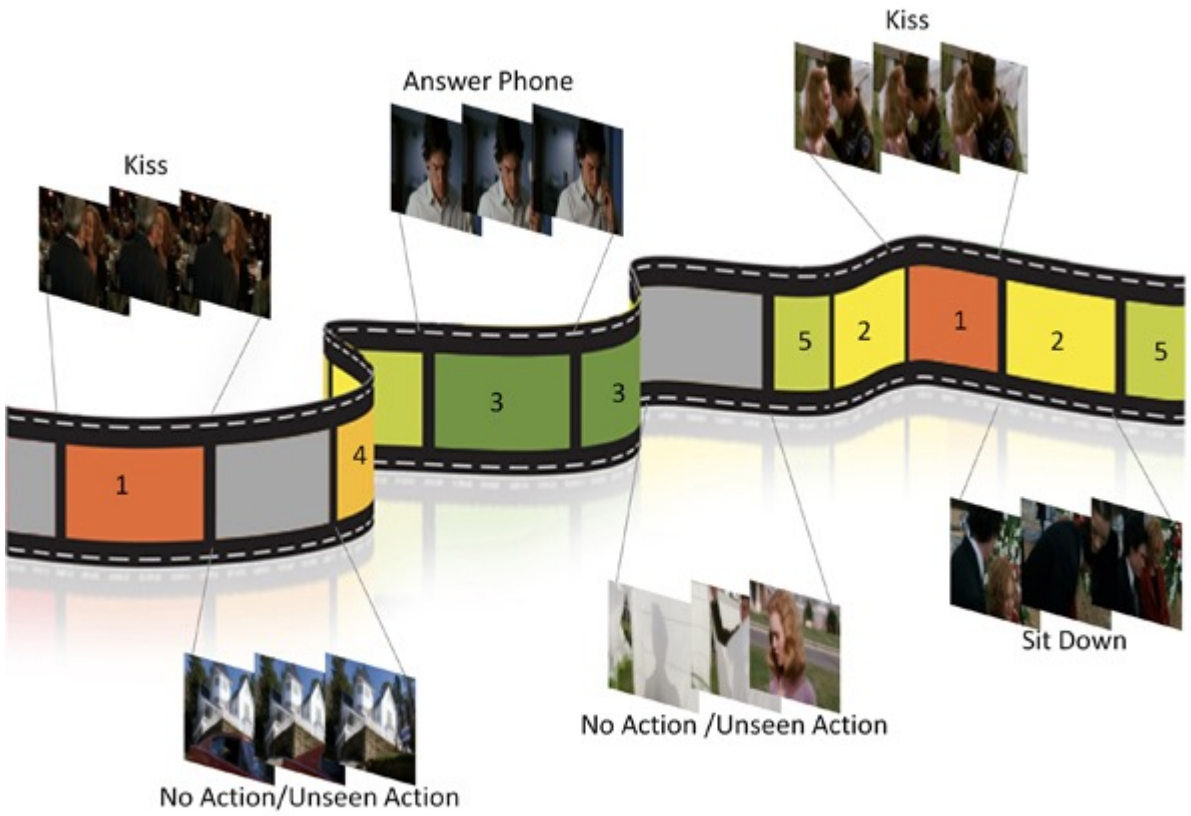
- Face generation
- Style transfer
- Scene generation

https://www.youtube.com/watch?v=OGGjXG562WU

- DALL–E 2

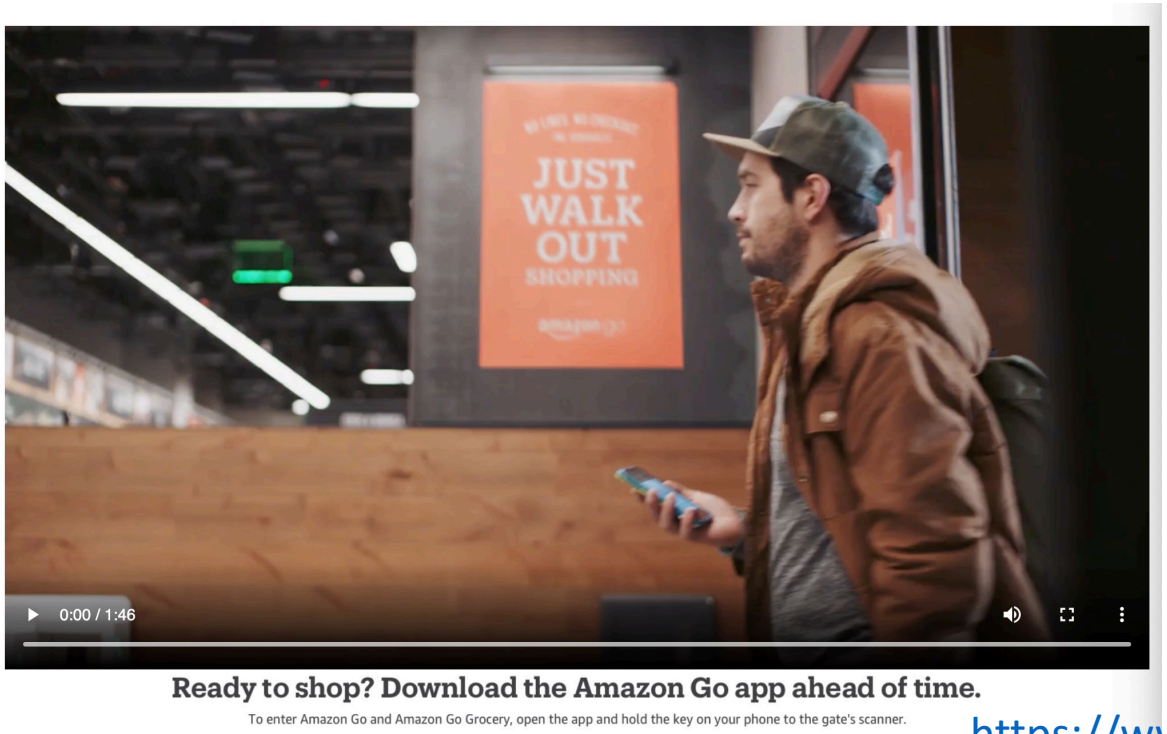
https://www.bilibili.com/video/BV1L3411n7nd/?vd_source=8275a2758ee7d2cac0897efeb47abf25

Video Classification



Case Study: Unmanned supermarket

- **Amazon Go** is a new kind of store featuring the world's most advanced shopping technology. No lines, no checkout – just grab and go! (Stage 3.0)



https://www.youtube.com/watch?v=NrmMk1Myrxc