# Computer Vision

CS308

Feng Zheng

SUSTech CS Vision Intelligence and Perception

Week 1


南方科技大学

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

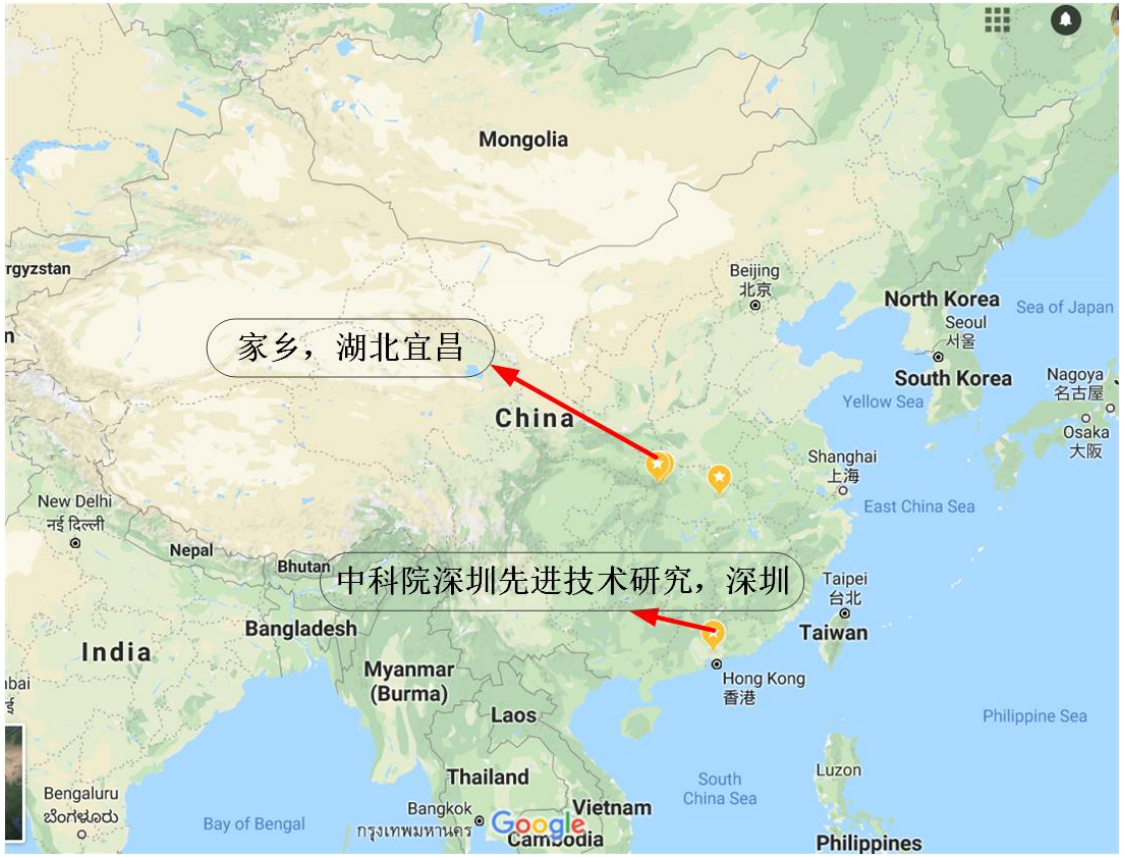

# Content

- Introduction
- The state-of-the-arts
- Applications in human-computer interaction
- Applications in video surveillance
- Conclusions

# Brief Biography

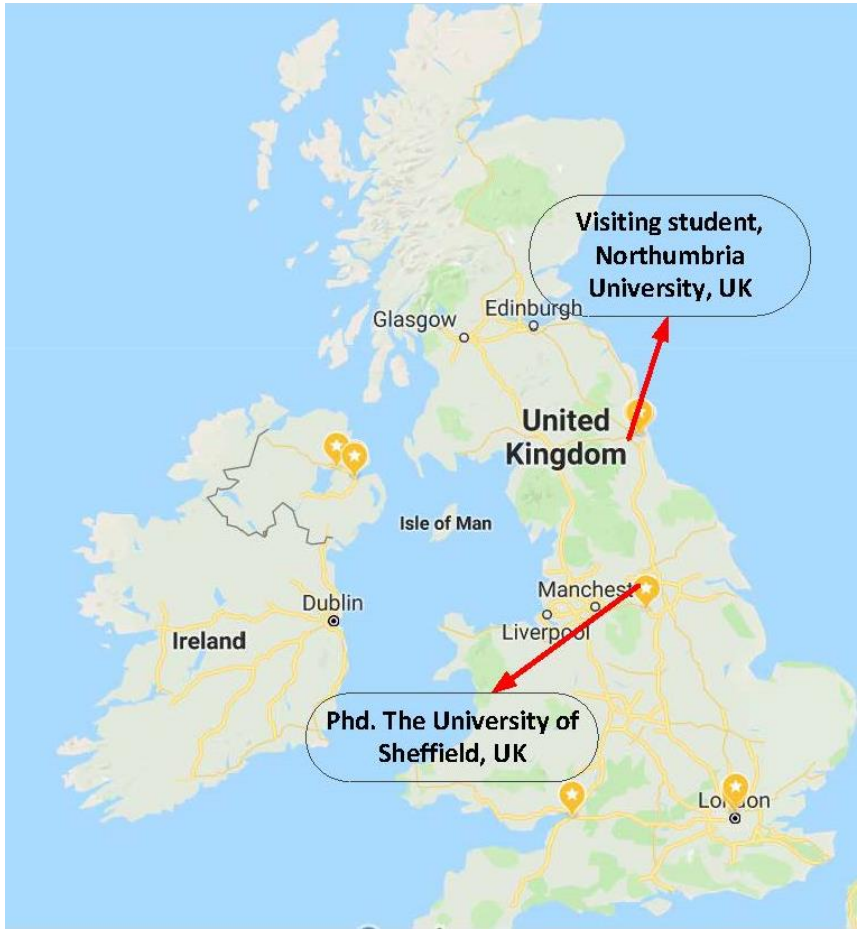
# China

- Shenzhen Institutes of Advanced Technology (SIAT), CAS, Jul. 2009 - Sep. 2012



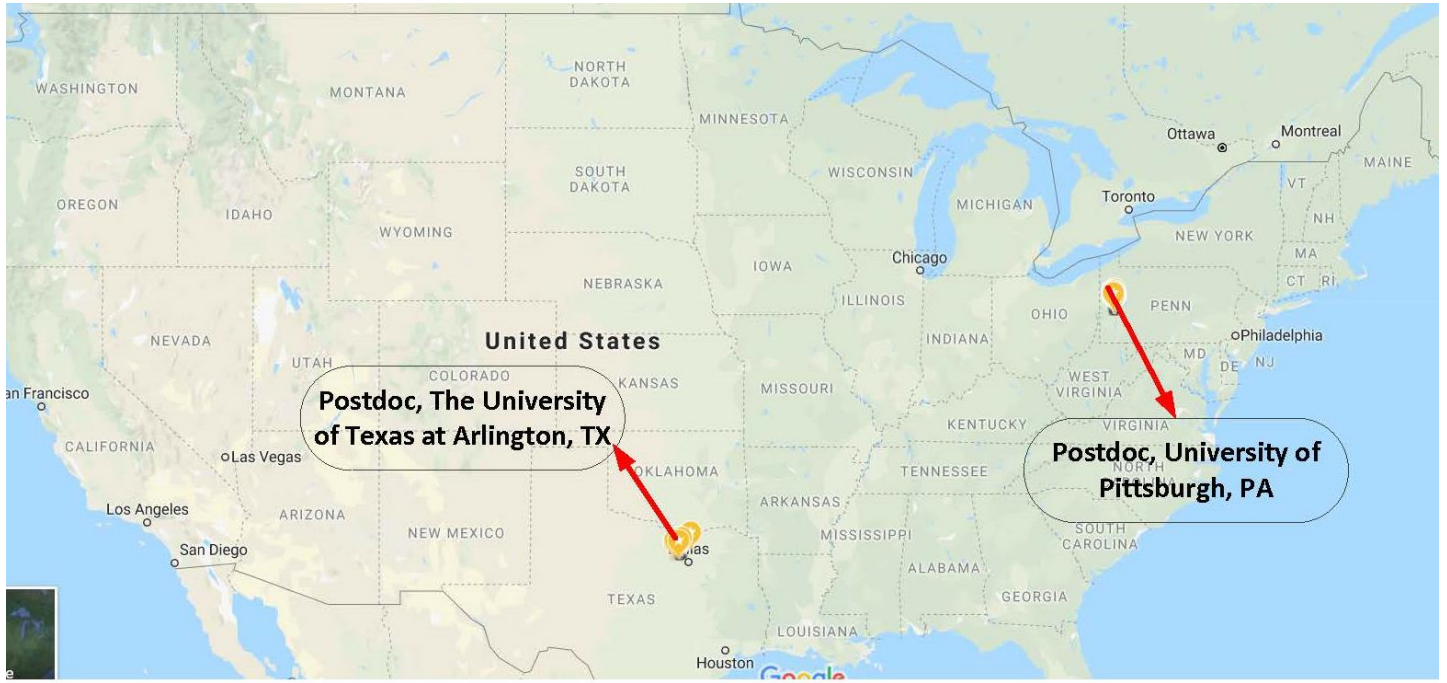
# United Kingdom

- The University of Sheffield, UK, Oct. 2012 - Otc. 2016



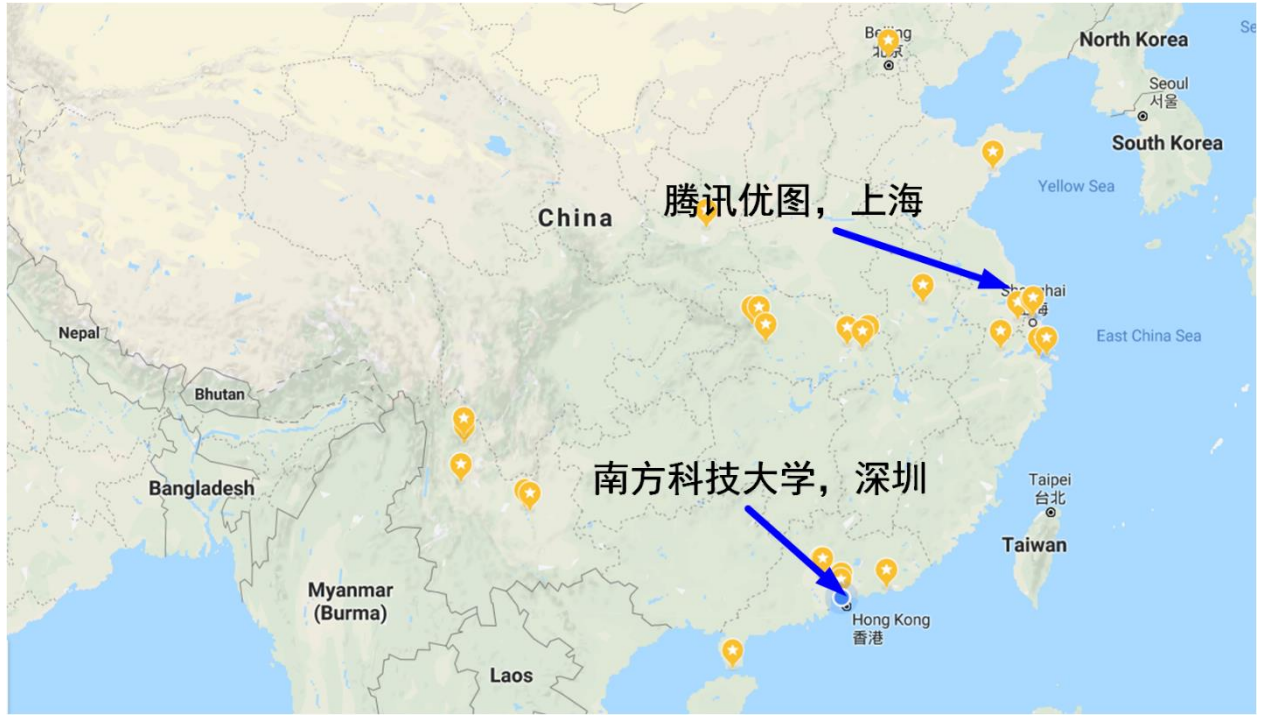
# United States

- The University of Texas, at Arlington, Texas, USA, Dec. 2016 - Aug. 2017
- University of Pittsburgh, Pittsburgh, USA, Sep. 2017 - July 2018



# China

- Southern University of Science and Technology, Shenzhen, China, Nov. 2018 - present
- Youtu Lab, Tencent, Shanghai, China, Aug. 2018 - Oct. 2018



# Summary: A circle

副教授

- 国自然面上项目-59万
- 国家优青2022年-200万
- 腾讯，国家空间中心，国防科技

- CVPR, ICCV, IJCAI, AAAI

**深圳南方科技大学**

18-现在

助理研究员

**深圳先进技术研究院**

09-12

- 优秀员工
- 手势识别系统
- CCTV报道

博士生 (EPSRC全奖)

**英国谢菲尔德大学**

12-16

- 合作企业：
- 联想-15万
- 土木工程系James Brownjohn合作
- 心理学系Allan Wing合作：
- 人体与结构协同运动检测
- 学术成果：
- IEEE TPAMI/TITS/TIP

博后研究员

**美国德州大学艾灵顿分校**

16-17

- 大规模果蝇数据的快速检索： AAAI 2018
- 合作者： Heng Huang

博后研究员

**美国匹兹堡大学**

17-18

- 电子病历EMR数据的快速检索： CVPR, IJCAI
- 合作者: Kamran Ghasedi Dizaj, Heng Huang

高级研究员

**上海腾讯优图**

18-18

- 腾讯万达智慧新零售
- 卓越研发奖（团队）

- CVPR 2019
- 超出其他算法9.5%以上
- CUKH03数据集上，目前仍然排名世界第一

跨模态

视觉

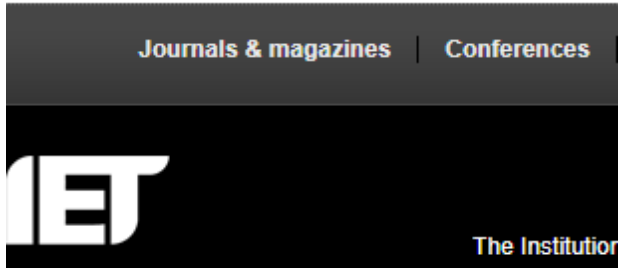
# Services

## Associate Editor
IET Image Processing


**Feng Zheng**
Southern University of Science and Technology
PR China

**Zhao Zhang**
Soochow University
PR China

**Ju Jia Zou**
Western Sydney University
Australia




## Program Committee：CCF-A类
PC 2021: ICLR,ICCV, AAAI, ICML, IJCAI, NIPS, CVPR
PC 2020: ICLR, ECCV, AAAI, ICML, IJCAI, NIPS, CVPR
PC 2019: ICLR, UAI, AAAI, ICML, IJCAI, NIPS
PC 2018: AAAI, IJCAI, NIPS
PC 2017: IJCAI

## Reviewer for leading AI journals (>10)，including：
IEEE Transactions on NNLS/CSVT/CYB/MM
Pattern Recognition, etc.


## Area Chair
ACM MM 2020, 2021 (CCF-A类)

## Local Chair
IEEE ICME 2021(CCF-B类), IJCB 2021 (CCF-C类)


# Achievements

## 发表论文：100余篇

- CCF-A类：近5年内56篇，一作&通讯30篇
- Trans期刊：16篇，一作&通讯10篇
- 谷歌学术引用：2435余次，高被引2篇


## 代表性论文（TPAMI, TIP, CVPR等）

(1) **Feng Zheng**; Y. Tang; L. Shao*; Hetero-Manifold Regularisation for Cross-Modal Hashing, **IEEE TPAMI**, 2018, 40(5): 1059-1071.

(2) **Feng Zheng**; C. Deng; X. Sun*; X. Jiang; X. Guo; Z. Yu; F. Huang; R. Ji*; Pyramidal Person Re-IDentification via Multi-Loss Dynamic Training, **CVPR**, 2019.6.

(3) **Feng Zheng**; L. Shao*; A Winner-Take-All Strategy for Improved Object Tracking, **IEEE TIP**, 2018, 27(9): 4302-4313.

(4) **Feng Zheng**; L. Shao*; Learning Cross-view Binary Identities for Fast Person Re-identification, **IJCAI**, 2016.7.

(5) X. Chen; C. Fu; Y. Zhao; **Feng Zheng***; J. Song; R. Ji; Yi Yang; Salience-Guided Cascaded Suppression Network for Person Re-identification, **CVPR**, 2020.3.

# Achievements: 12篇CVPR

Categories ▾ English ▾

| | Publication | h5-index | h5-median |
|---|---|---|---|
| 1. | Nature | 444 | 667 |
| 2. | The New England Journal of Medicine | 432 | 780 |
| 3. | Science | 401 | 614 |
| 4. | IEEE/CVF Conference on Computer Vision and Pattern Recognition | 389 | 627 |

[52] Yawen Huang, **Feng Zheng***, Danyang Wang, Weilin Huang, Matthew R. Scott, Ling Shao. Brain Image Synthesis with Unsupervised Multivariate Canonical CSC$\ell_4$Net. **CVPR** 2021. (**CCF A**)

[51] Jiaxing Chen, Xinyang Jiang, Fudong Wang, Jun Zhang, **Feng Zheng**, Xing Sun, WEI-SHI ZHENG. Learning 3D Shape Feature for Texture-insensitive Person Re-identification. **CVPR** 2021. (**CCF A**)

[38] Xuesong Chen, Xiyu Yan, **Feng Zheng***, Yong Jiang, Shutao Xia, Yong Zhao, Rongrong Ji. One-shot Adversarial Attacks on Visual Tracking with Dual Attention. Accepted by **CVPR** 2020. (**CCF A**)

[37] Yunhang Shen, Rongrong Ji, Zhiwei Chen, Xiaopeng Hong, **Feng Zheng**, Jianzhuang Liu, Mingliang Xu, Qi Tian. Noise-Aware Fully Webly Supervised Object Detection. Accepted by **CVPR** 2020. (**CCF A**)

[36] Xuesong Chen, Canmiao Fu, Yong Zhao, **Feng Zheng***, Jingkuan Song, Rongrong Ji, Yi Yang. Salience-Guided Cascaded Suppression Network for Person Re-identification. Accepted by **CVPR** 2020. (**CCF A**)

[27] **Feng Zheng**, Cheng Deng, Xing Sun, Xingyang Jiang, Xiaowei Guo, Zongqiao Yu, Feiyue Huang, Rongrong Ji. Pyramidal Person Re-IDentification via Multi-Loss Dynamic Training. Accepted by **CVPR, 2019**. (**CCF A**)

[26] Xinyi Xu, Yanhua Yang, Cheng Deng, **Feng Zheng**. Deep Asymmetric Metric Learning via Rich Relationship Mining. Accepted by **CVPR, 2019**. (**CCF A**)

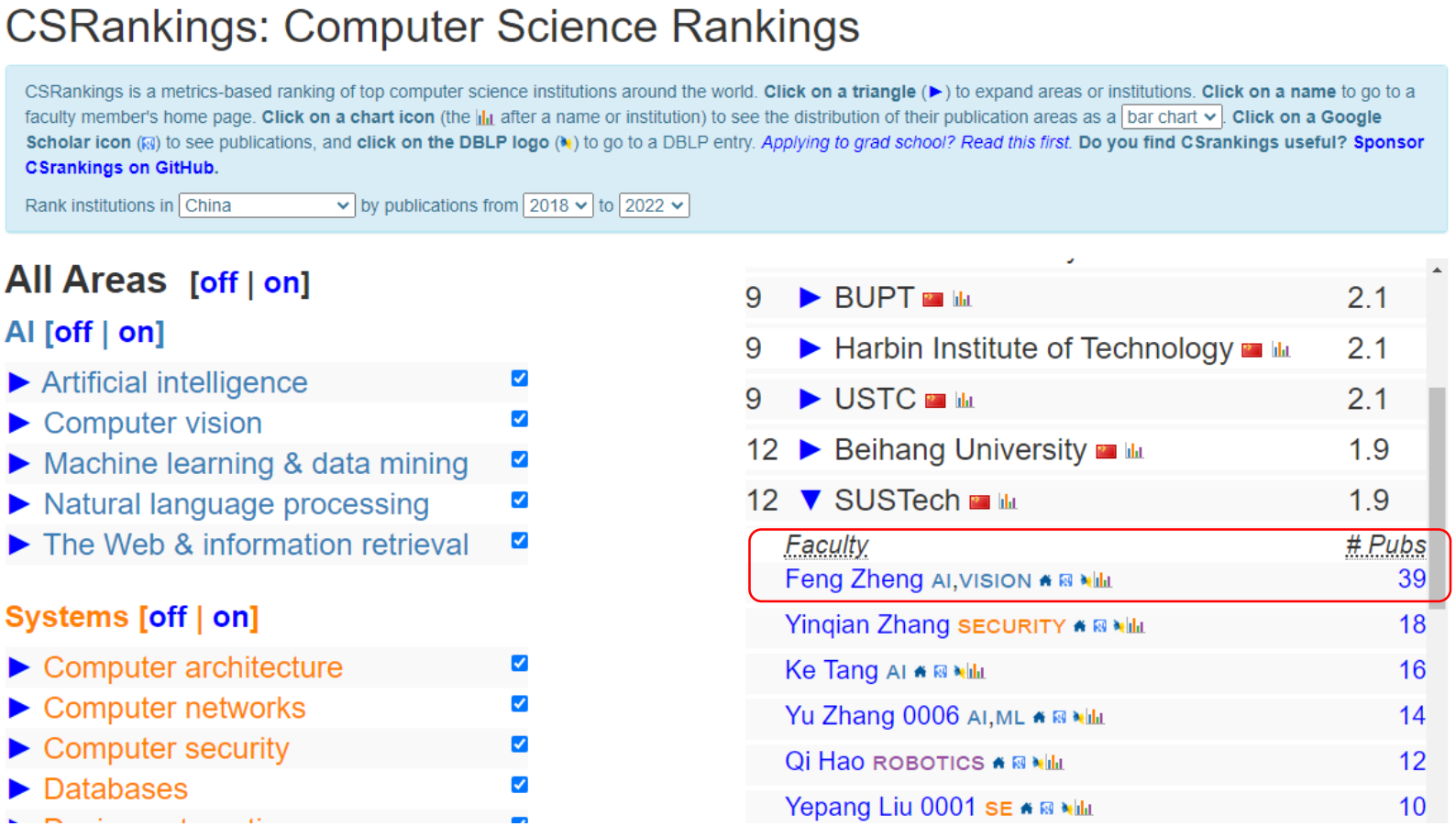
[25] Xu Yang, Cheng Deng, **Feng Zheng**, Junchi Yan, Wei Liu. Deep Spectral Clustering using Dual Autoencoder Network. Accepted by **CVPR, 2019**. (**CCF A**)

[21] Kamran Ghasedi Dizaji, **Feng Zheng**, Najmeh Sadoughi Nourabadi, Yanhua Yang, Cheng Deng, Heng Huang. Unsupervised Deep Generative Adversarial Hashing Network. **CVPR**, Salt Lake City, Utah, United States, 2018. (**CCF A**)

https://scholar.google.com/citations?view_op=top_venues&hl=en

# CSRanking

- https://csrankings.org/

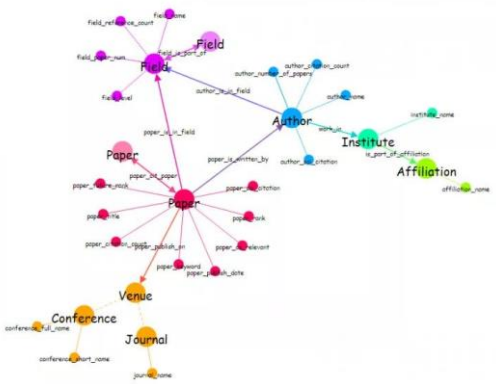


# Introduction of CS308

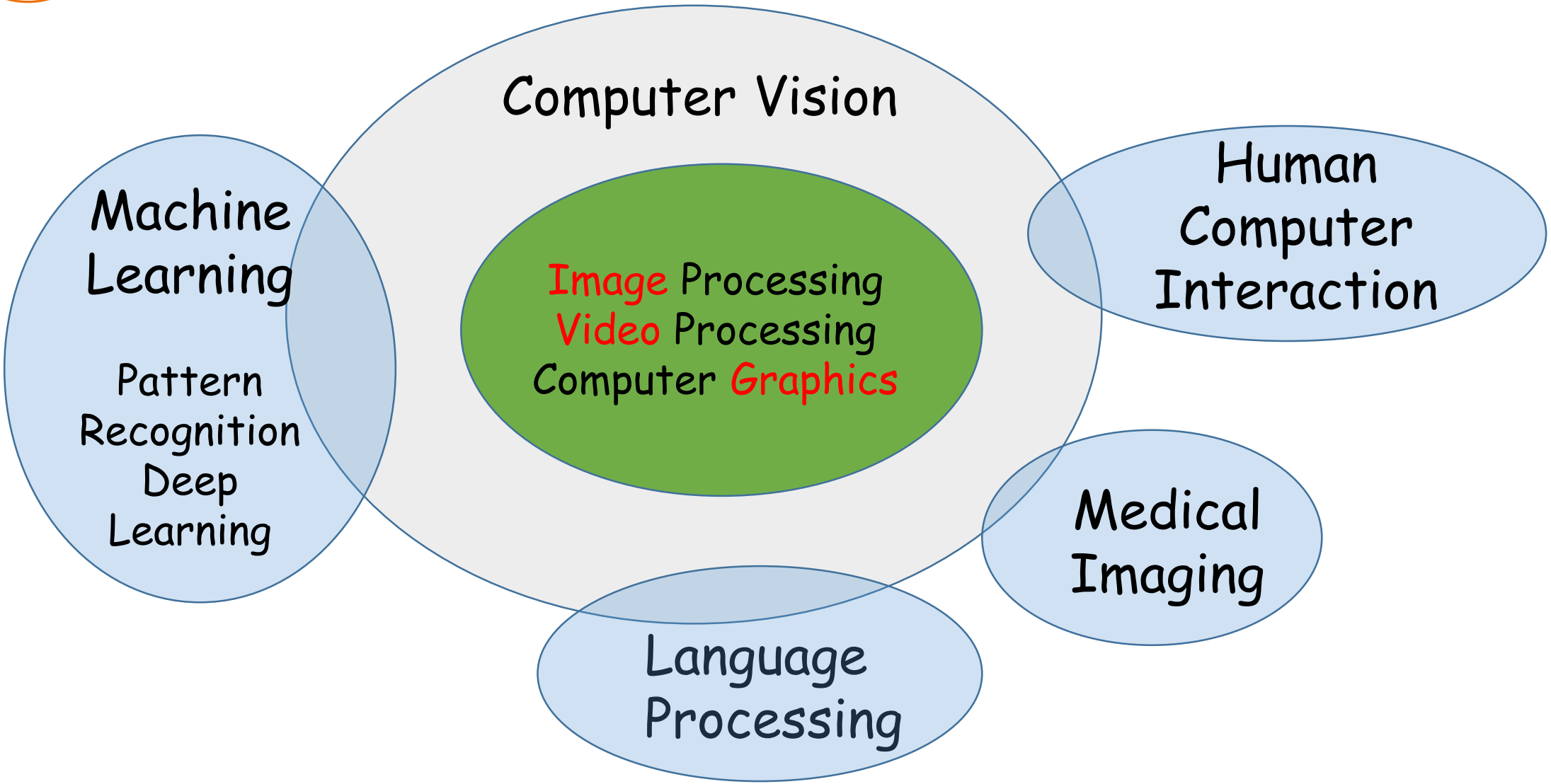
# What Is Computer Vision?

- Visual computing is the **science** and **technology** of machines that **see** (capturing, understanding and prediction).
  - ➢ Come up with computational models of the human visual system
  - ➢ Build autonomous systems which could perform some of the tasks


+

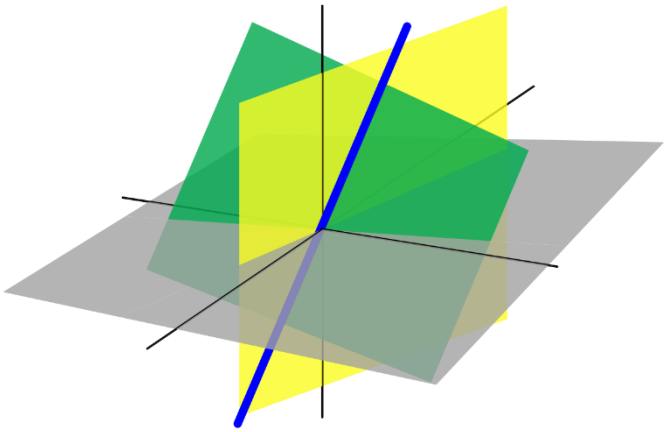


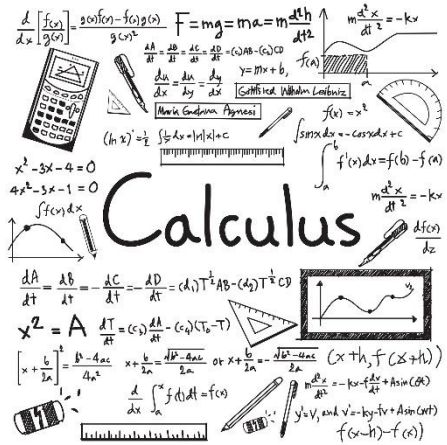
# CS308-Computer Vision

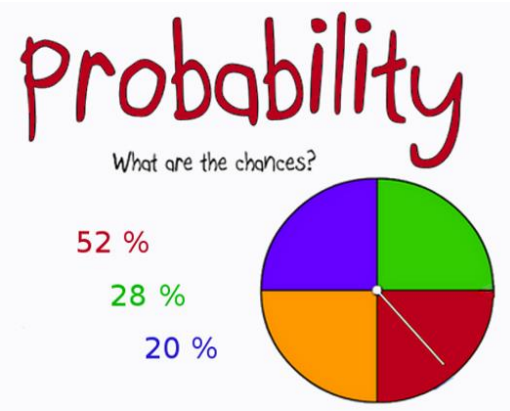


# CS308-Computer Vision

- Linear algebra
- Basic calculus
- Probability



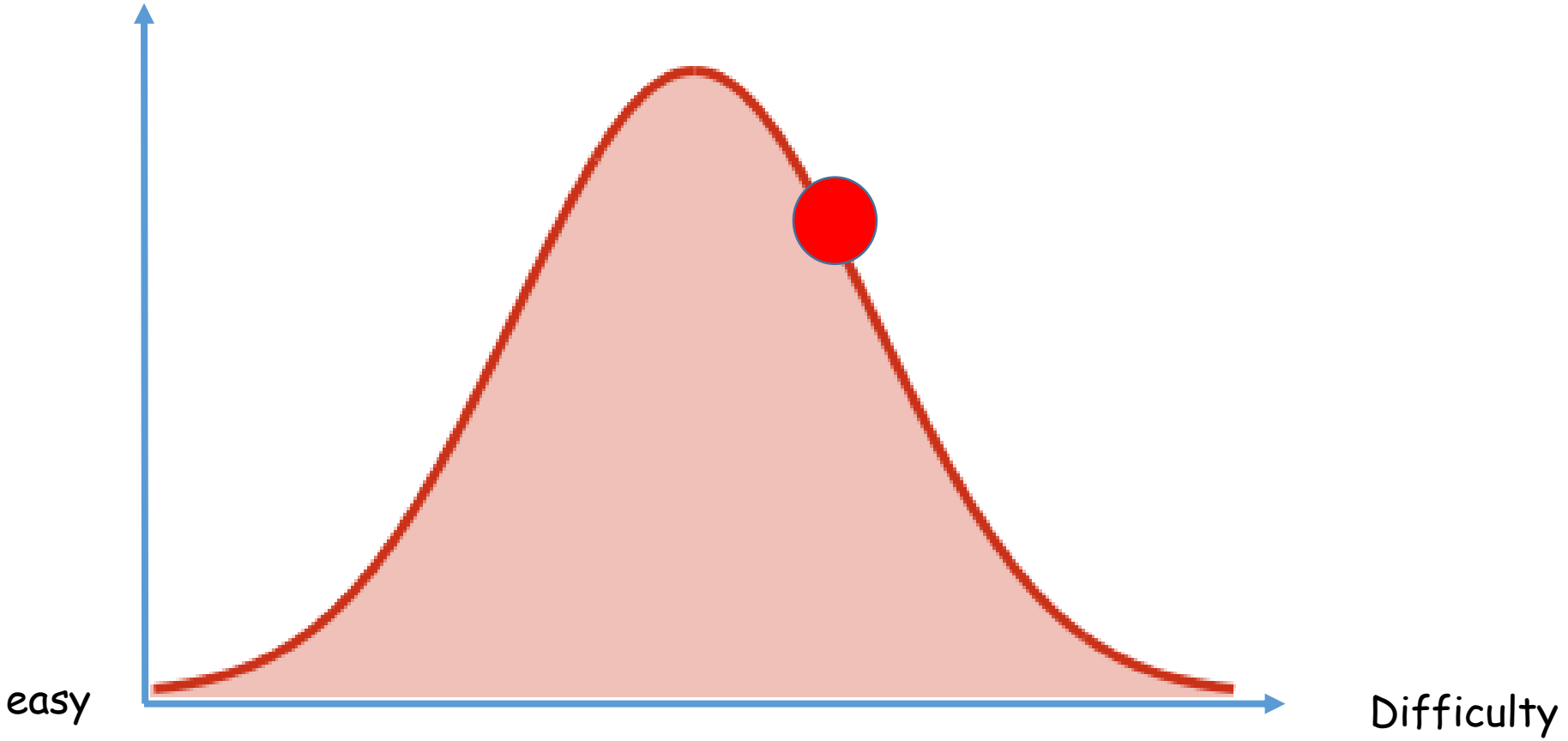



- Experience with image processing will help but is not necessary
- Experience with Python or Python-like languages will help

# CS308-Target Student

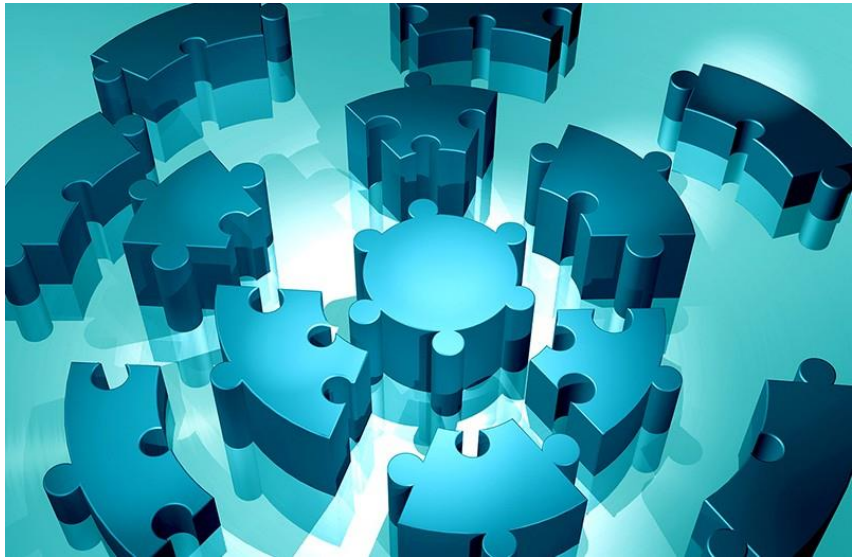
- A little higher than average ability



# Expectations

- Understand the basics of computer vision (old)
- Know research trends (new)

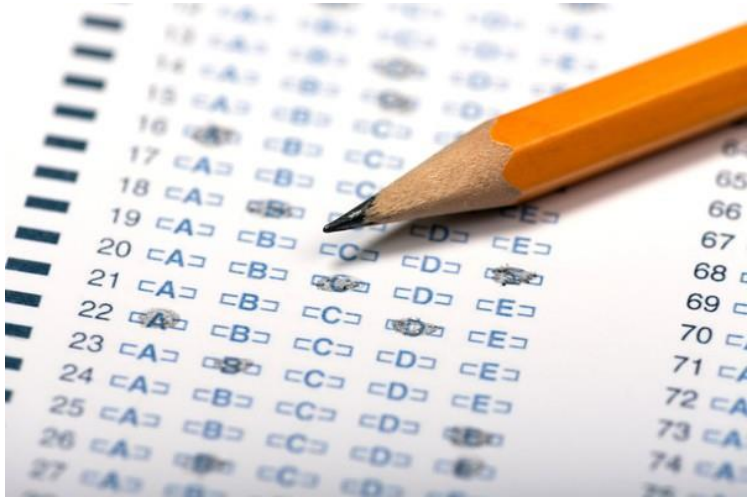
- Ability to model visual tasks
- Ability to implement the models



# Exams test you on

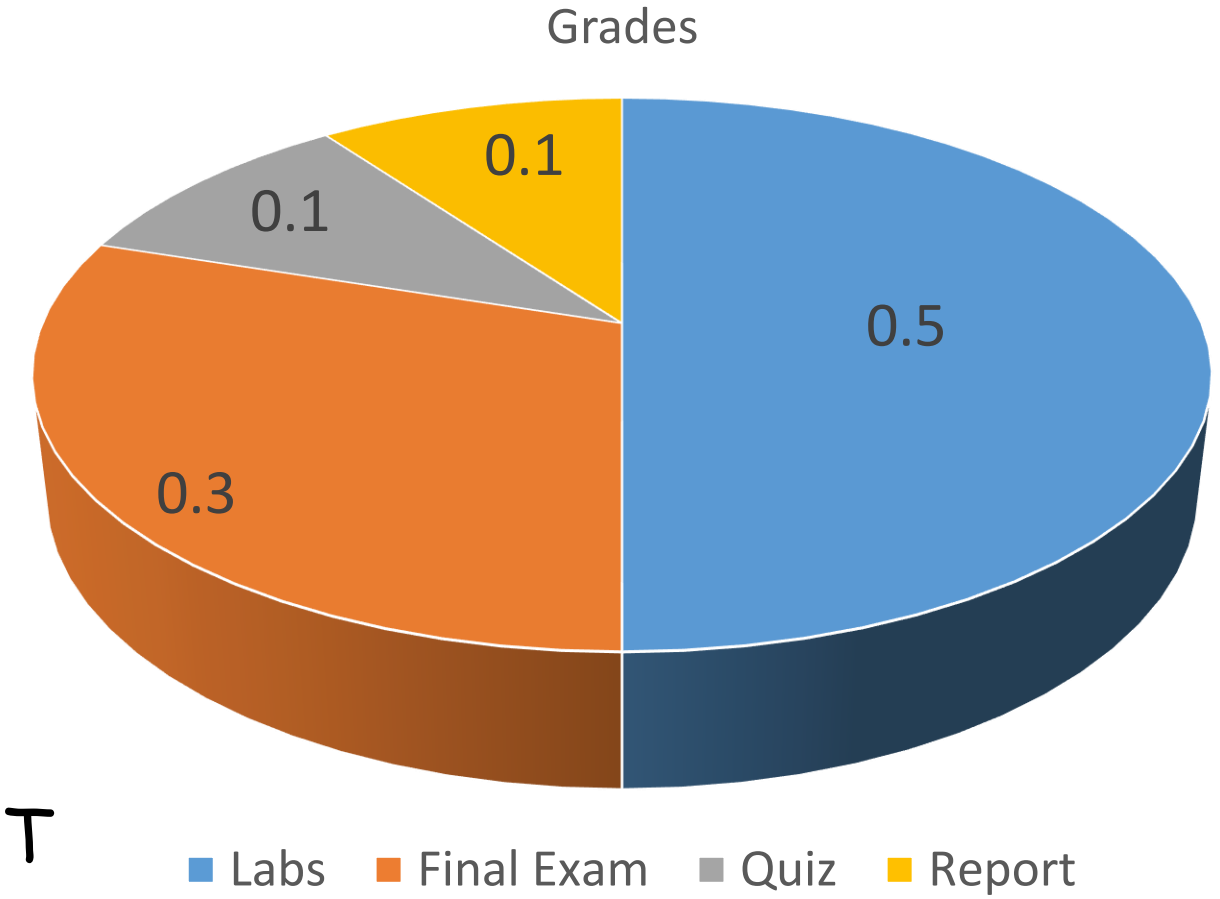
- General knowledge of visual technology

- Ability to model simple tasks
- Understand the general flow of the visual system (implement)



# Grade Component

- Middle-Term Report 10%
- Final Exam 30%
- Labs 50%
- Quiz: 10%




- LABS are VERY IMPORTANT

# Honesty

- Get code from the internet for labs/assignments is perfectly **OK**
  - When you borrow, just say it.
  - You don't need to reinvent the wheel

~~**Plagiarism**~~


- **DON'T** pretend that you are the author of something that you didn't write.

# QQ Group for CS308 Autumn

- Quick Response Code

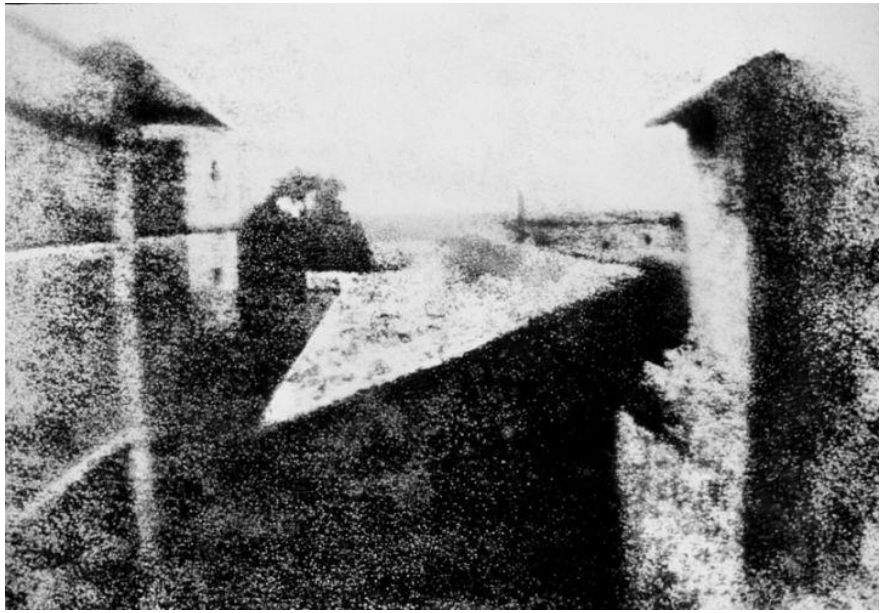
- Blackboard: Computer Vision Fall 2022

群名称：CS308-2022-Fall
群　号：187176142

# Introduction of CV

# Visual data: image

- The first photograph



Nicéphore Niépce. *View from the Window at Le Gras.* ca. 1826.

*Robert Cornelius, self-portrait, October or November 1839.*

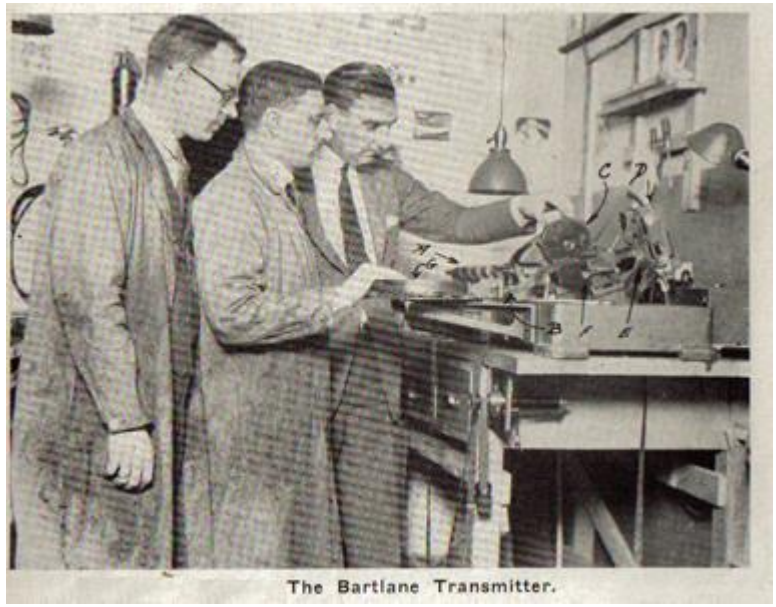
*Walden Kirsch as scanned into the SEAC computer by Russell A. Kirsch in 1957.*

# Visual data: video

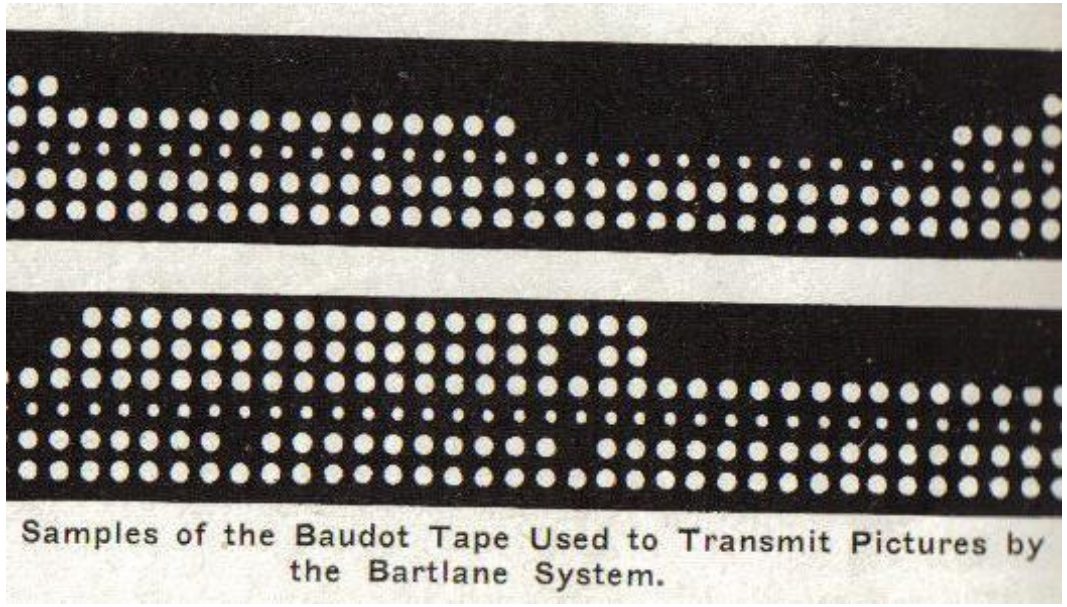
- The first video

# Visual data: digital imaging

- The first **digital** image was produced in 1920
  - Bartlane cable picture transmission system
  - Harry G. Bartholomew and Maynard D. McFarlane
  - London and New York



The Bartlane Transmitter.

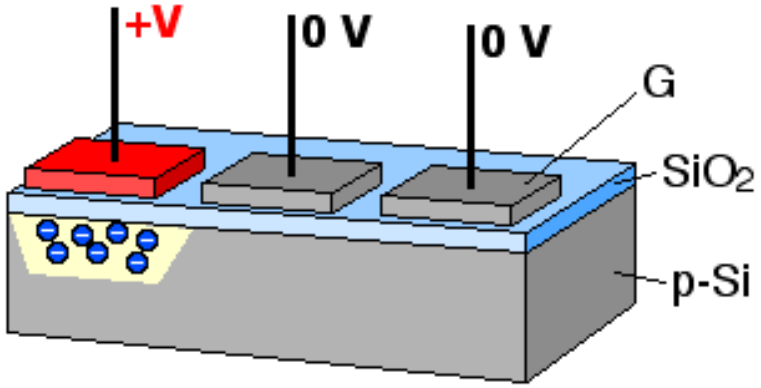


Samples of the Baudot Tape Used to Transmit Pictures by the Bartlane System.

http://www.digicamhistory.com/1906_1920.html

# Visual data: digital imaging

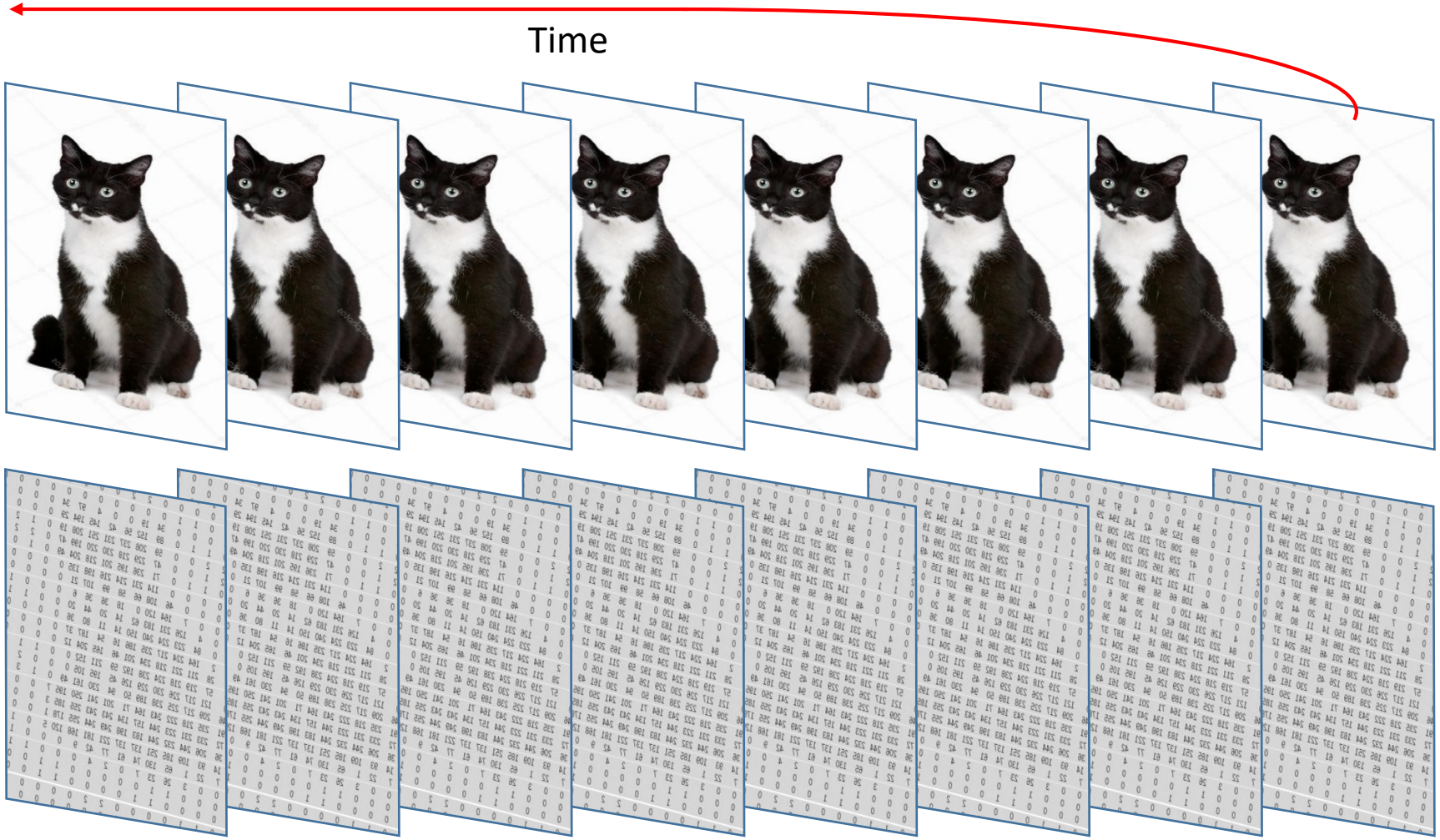
- Charge-coupled device(CCD)
  - AT&T Bell Labs(1969) by Willard Boyle and George E. Smith
  - A piece of lens
  - A Capacitor array (the photoactive region)
  - A control circuit

# Visual data: matrix

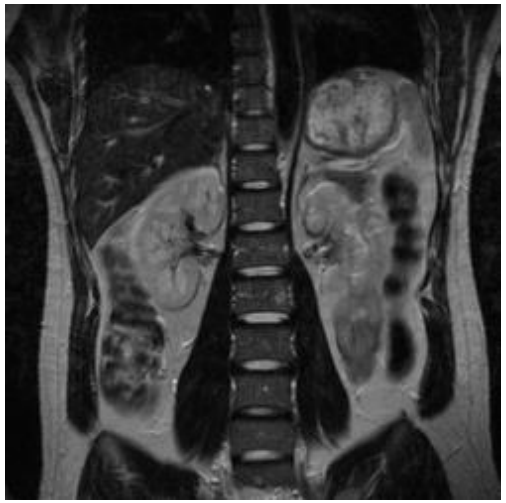
| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 34 | 19 | 0 | 0 | 4 | 97 | 34 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 1 | 0 | 89 | 152 | 56 | 42 | 145 | 194 | 29 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 2 | 2 | 0 | 0 | 59 | 208 | 237 | 231 | 251 | 208 | 19 | 0 | 1 | 2 | 3 |
| 0 | 0 | 2 | 2 | 1 | 0 | 0 | 47 | 229 | 218 | 230 | 220 | 199 | 47 | 0 | 2 | 2 | 1 |
| 0 | 1 | 1 | 0 | 1 | 1 | 0 | 71 | 236 | 195 | 201 | 218 | 204 | 49 | 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 114 | 231 | 214 | 216 | 198 | 135 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 46 | 108 | 66 | 58 | 99 | 107 | 21 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 7 | 164 | 120 | 0 | 18 | 36 | 36 | 6 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 4 | 126 | 231 | 183 | 62 | 14 | 20 | 44 | 20 | 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 1 | 0 | 84 | 223 | 224 | 240 | 150 | 14 | 11 | 80 | 36 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 2 | 164 | 224 | 217 | 235 | 186 | 16 | 54 | 187 | 37 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 28 | 211 | 221 | 218 | 234 | 201 | 46 | 165 | 204 | 12 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 57 | 219 | 218 | 224 | 238 | 192 | 59 | 211 | 152 | 0 | 0 | 1 | 1 | 0 | 1 |
| 0 | 0 | 0 | 121 | 217 | 226 | 230 | 229 | 126 | 45 | 195 | 105 | 0 | 0 | 0 | 2 | 3 | 3 |
| 0 | 0 | 46 | 209 | 217 | 225 | 238 | 189 | 50 | 94 | 230 | 161 | 49 | 0 | 1 | 3 | 1 | 0 |
| 0 | 0 | 91 | 235 | 218 | 222 | 243 | 164 | 71 | 201 | 241 | 250 | 195 | 7 | 0 | 0 | 0 | 0 |
| 1 | 0 | 72 | 233 | 231 | 223 | 244 | 157 | 134 | 243 | 243 | 255 | 185 | 3 | 0 | 0 | 0 | 0 |
| 0 | 0 | 36 | 206 | 244 | 232 | 244 | 183 | 198 | 249 | 244 | 255 | 178 | 1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 14 | 93 | 109 | 185 | 251 | 137 | 137 | 221 | 181 | 168 | 120 | 5 | 0 | 1 | 1 | 0 |
| 0 | 0 | 7 | 22 | 1 | 65 | 130 | 74 | 61 | 77 | 42 | 9 | 0 | 0 | 0 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 3 | 26 | 23 | 7 | 0 | 2 | 4 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |

# Visual data: sequence of matrix



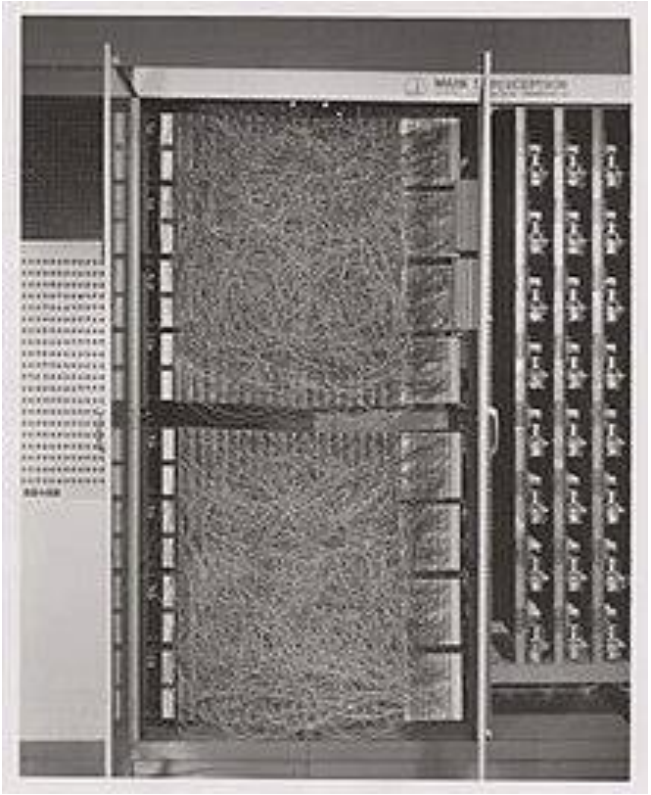
# Visual data: more

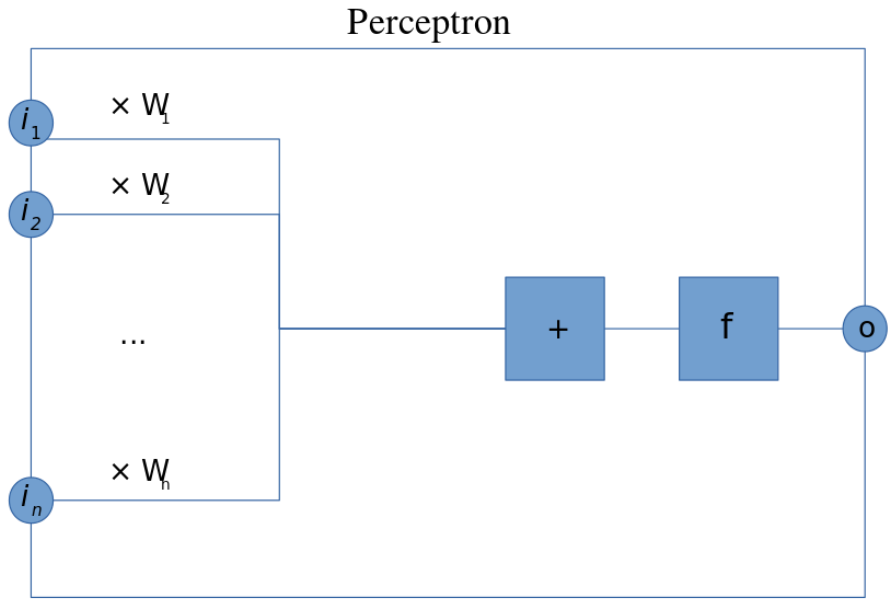
- Depth image (Time of flight, structured-light)
- Ultrasound imaging
- Magnetic resonance imaging



# Early work

- Frank Rosenblatt (1957): using "Perceptron" machine to sort images into very simple categories like **triangle** and **square**

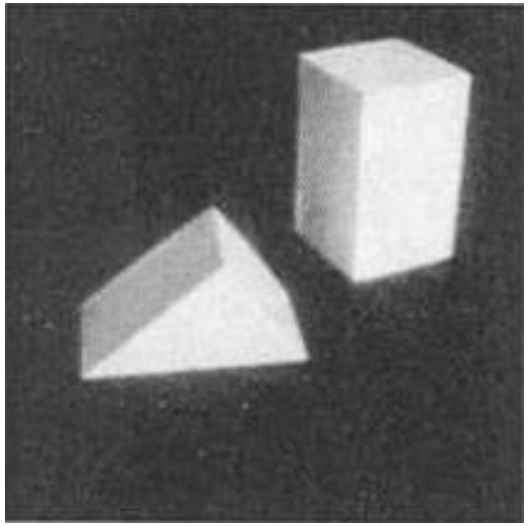




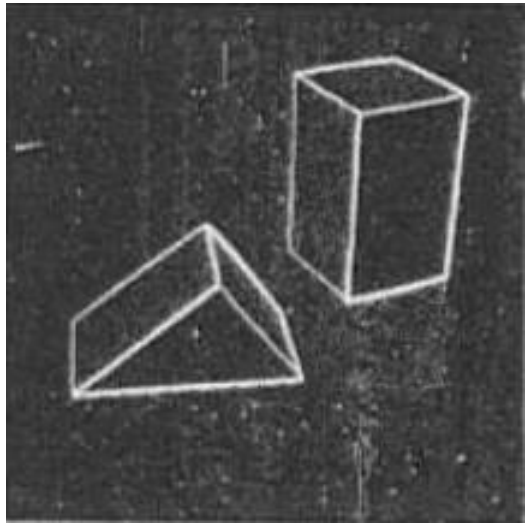

$$o = f\left(\sum_{k=1}^{n} i_k \cdot W_k\right)$$

# Early work

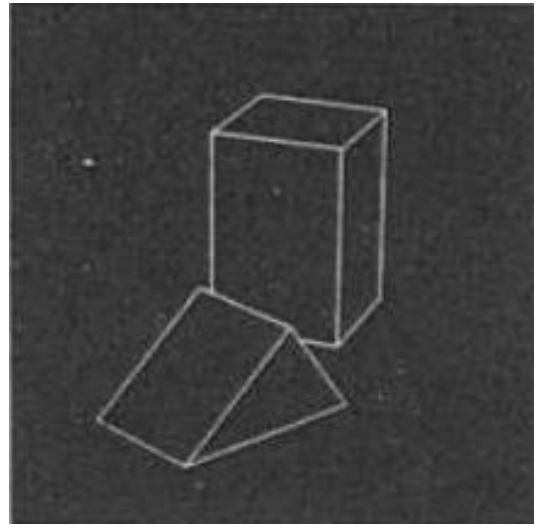
- Larry Roberts at MIT(1960): extracting 3D geometrical information from 2D perspective views of blocks
  - Machine perception of three-dimensional solids
  - Father of computer vision



Input image



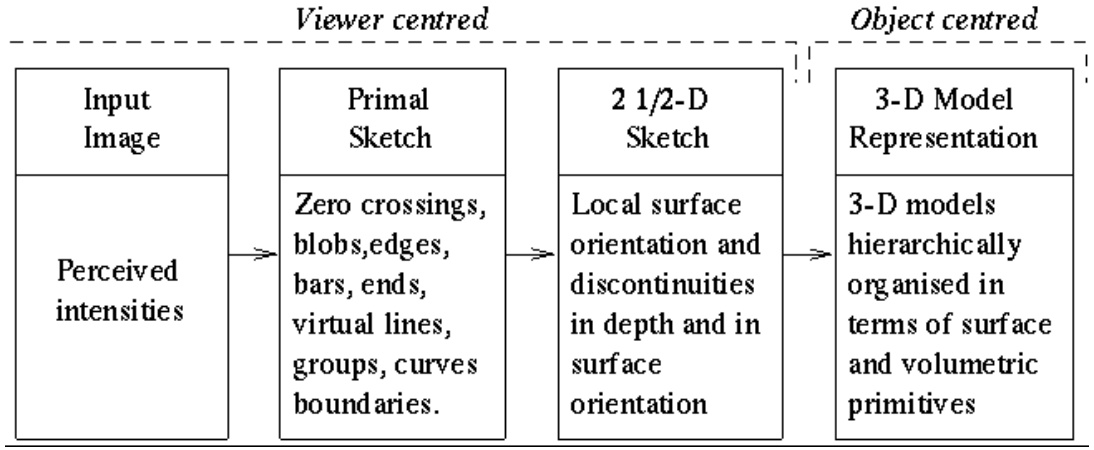
2x2 gradient operator

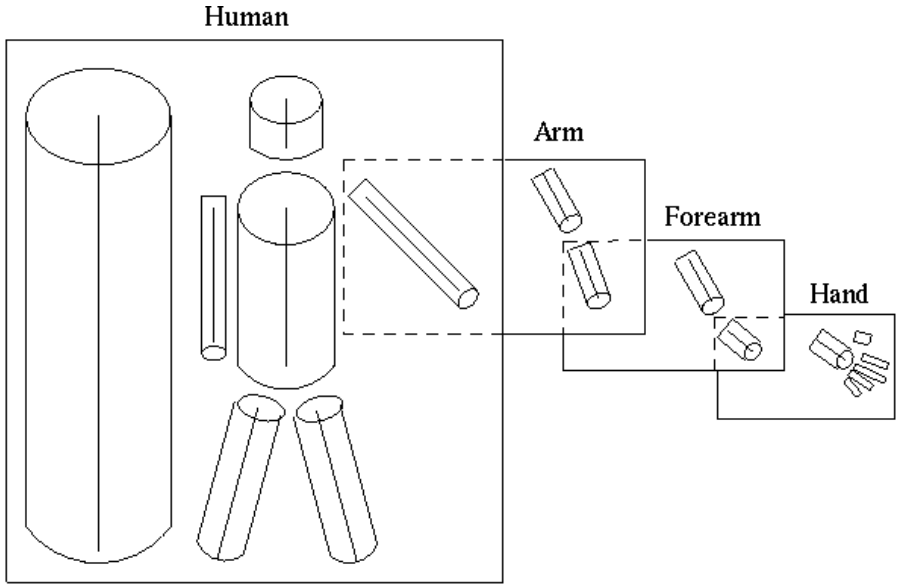


Computed 3D model

# Early work

- Marvin Minsky at MIT(1966): connecting a camera to a computer
- David Marr at MIT(1978): proposing a bottom-up approach to scene understanding



# Developing


https://papers.nips.cc › paper › 4824-imagenet-cla... ▾ PDF

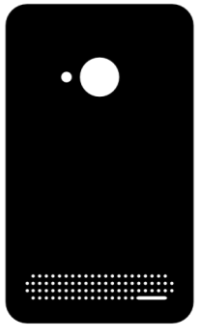
ImageNet Classification with Deep Convolutional Neural Nets

by A Krizhevsky · Cited by 86442 — **ImageNet Classification with Deep Convolutional Neural Networks**. Alex Krizhevsky. University of Toronto kriz@cs.utoronto.ca. Ilya Sutskever.

Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. NeuIPS, 2012.

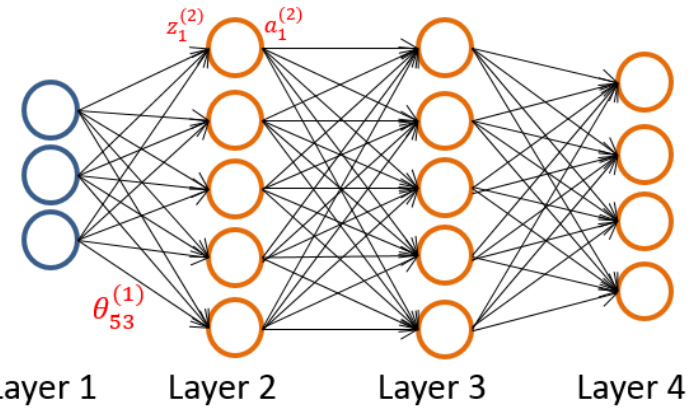
# Why it bring about a renaissance?

- Mobile technology with built-in cameras (data)



- Computing power (devices)
- Mass storage

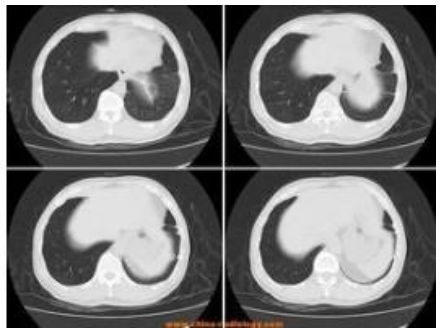

- New algorithms (models)
  - Support vector machine
  - Convolutional neural networks
  - Transformer



# Why it matters

- Applications of visual computing

Safety



Health


Security

Facility

Fun

Industry

# The state-of-the-arts

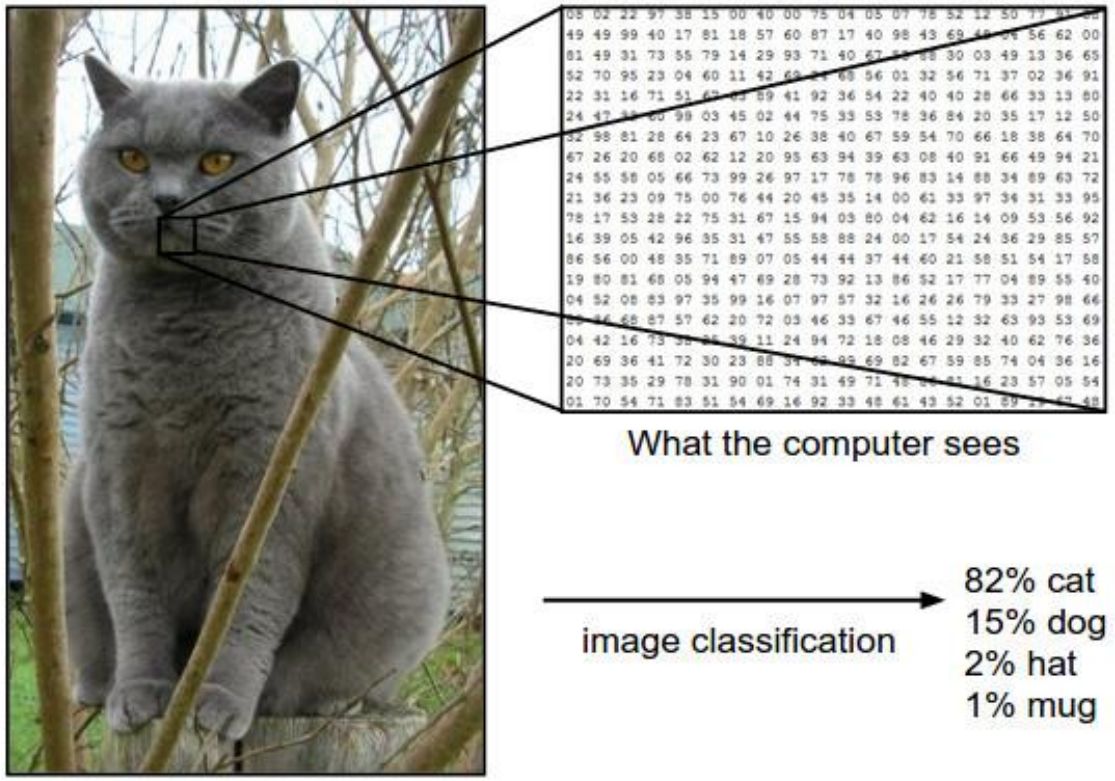
# Areas of visual computing

- Image classification
- Object detection
- Image segmentation
- Pose estimation
- Visual language
  - Image captioning
  - VQA…
- Object tracking
- Object identification
- View synthesis (3D)

……

# Image classification

- What is image classification?
  - Building a model (function) from hypothesis space
  - Image to label
  - Matrix to number



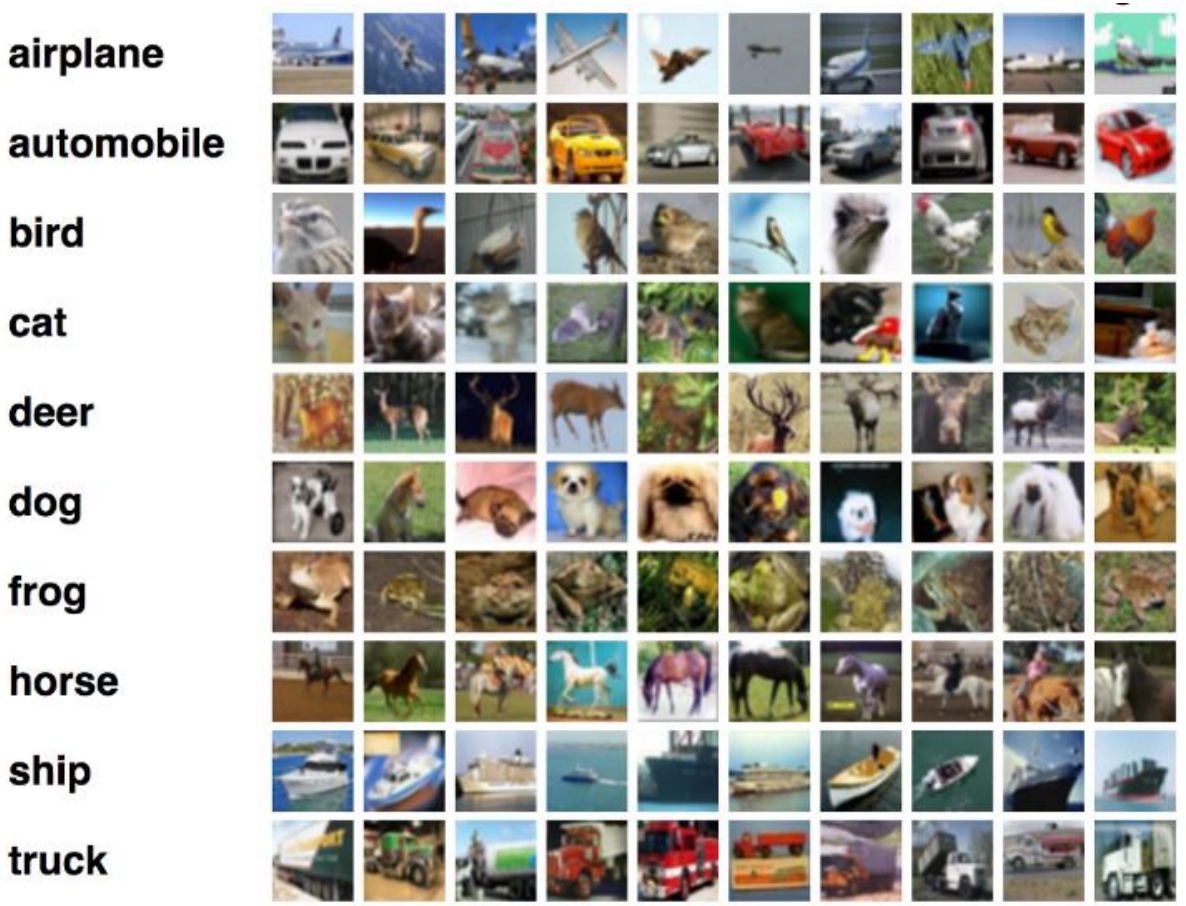
# Image classification

- Task 1: MNIST database of handwritten digits
  - ICML 2013: 99.79%
  - It has a training set of 60,000 examples, and a test set of 10,000 examples

http://yann.lecun.com/exdb/mnist/

# Image classification

- The CIFAR-10 dataset
  - arXiv 2015: 96.53%
  - 60000 32x32 colour images
  - 10 classes,
  - 6000 images per class.
  - 50000 training images
  - 10000 test images.

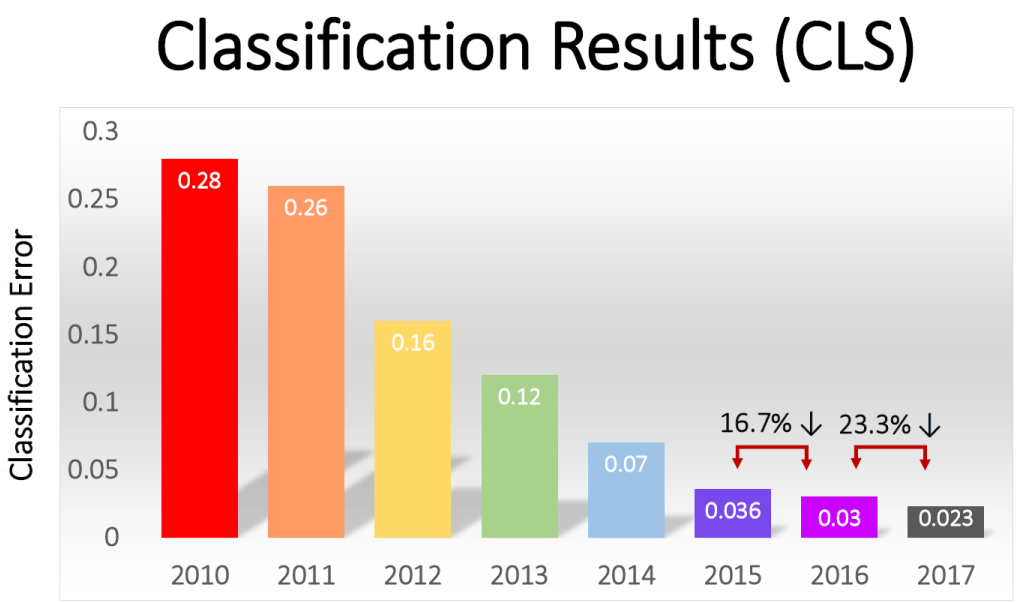



https://www.cs.toronto.edu/~kriz/cifar.html

# Image classification

- IMAGENET: Large Scale Visual Recognition Challenge
  - 1000 classes: 1M train images and 100K test images

http://www.image-net.org/




ImageNet: A large-scale hierarchical image database - IEEE ...

by J Deng · 2009 · Cited by 31462 — We introduce here a new **database** called "**ImageNet**", a **large-scale** ontology of **images** built upon the backbone of the WordNet structure. **ImageNet**...

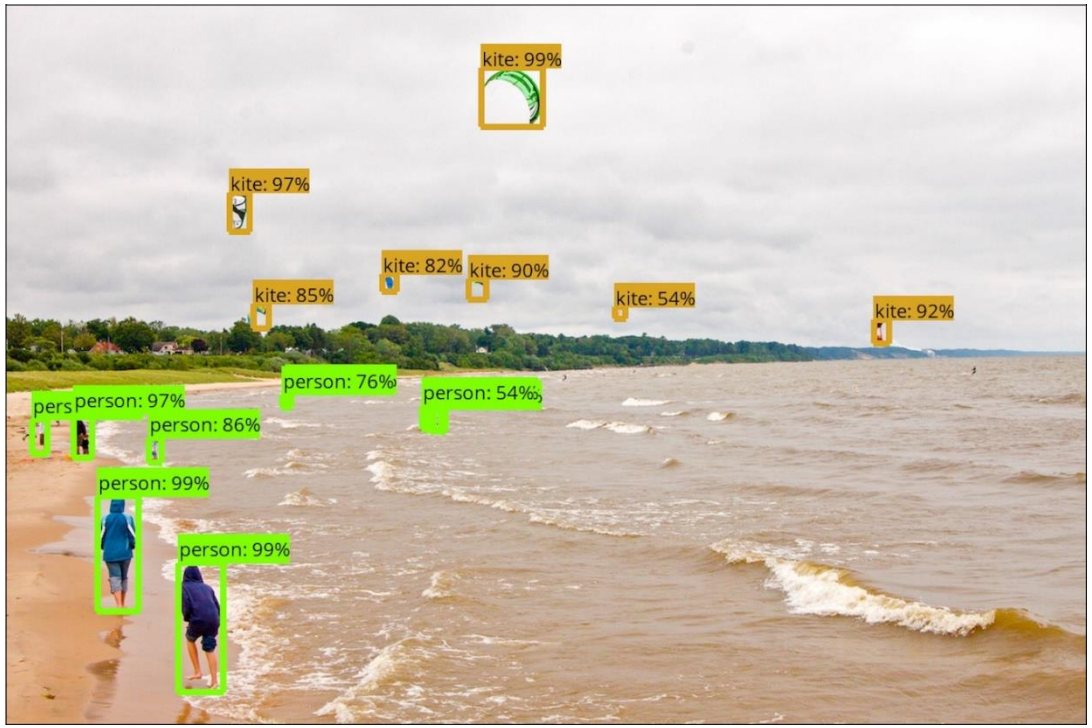
**Date Added to IEEE Xplore:** 18 August 2009 **DOI:** 10.1109/CVPR.2009.5206848

**Date of Conference:** 20-25 June 2009 **INSPEC Accession Number:** 10836047

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, **ImageNet: A Large-Scale Hierarchical Image Database.** *IEEE Computer Vision and Pattern Recognition (CVPR), 2009.*

# Object detection

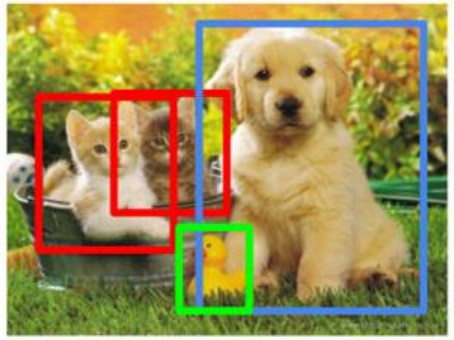
- What is object detection?
  - Input: image
  - Output: locations of objects




**Classification**

CAT

**Object Detection**



CAT, DOG, DUCK

# Object detection

- The frameworks

Faster R-CNN

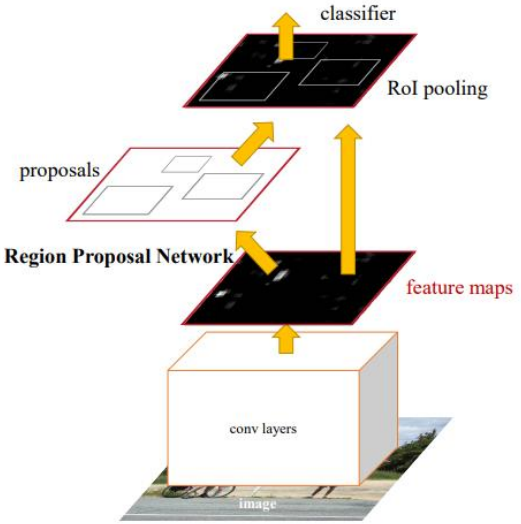
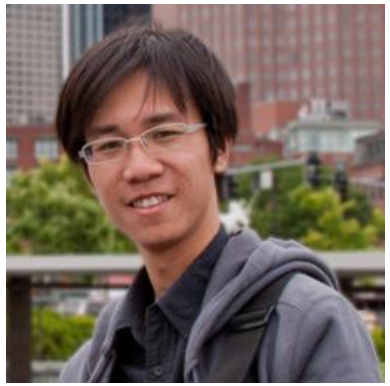



Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.



http://kaiminghe.com/

Cited by

| | All | Since 2015 |
|---|---|---|
| Citations | 145231 | 142905 |
| h-index | 53 | 53 |
| i10-index | 61 | 61 |

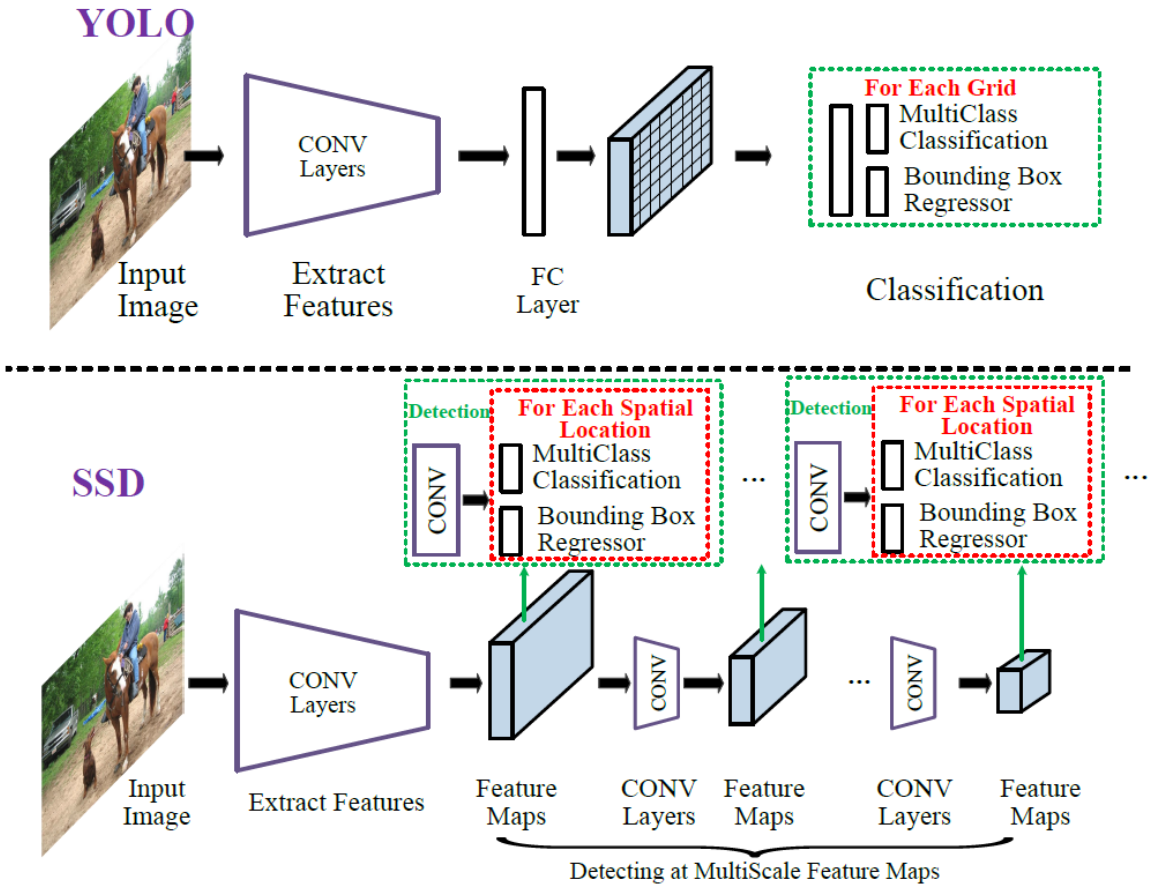
https://arxiv.org › cs

Faster R-CNN: Towards Real-Time Object Detection with ...

by S Ren · 2015 · Cited by 47205 — An RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end…

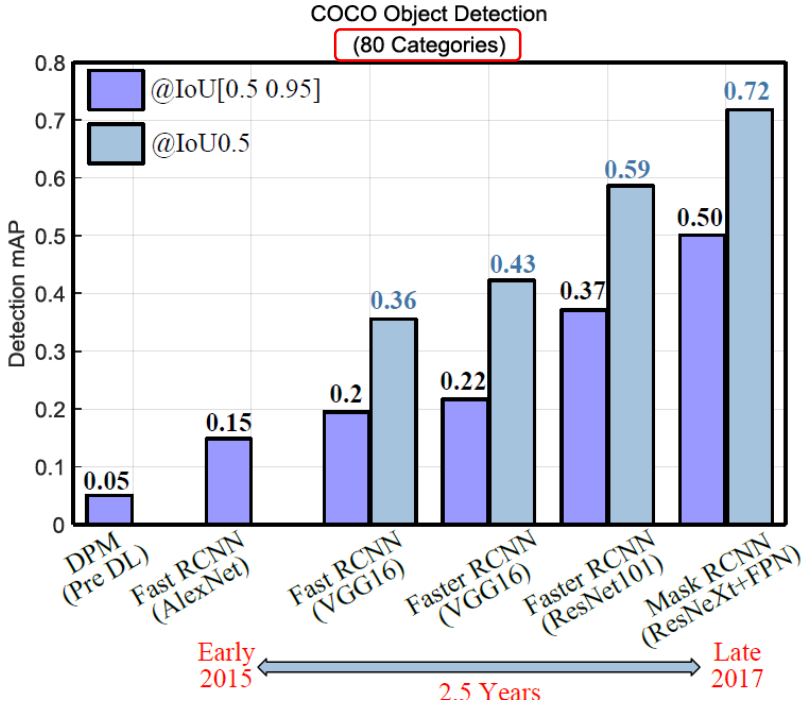
Cite as: arXiv:1506.01497

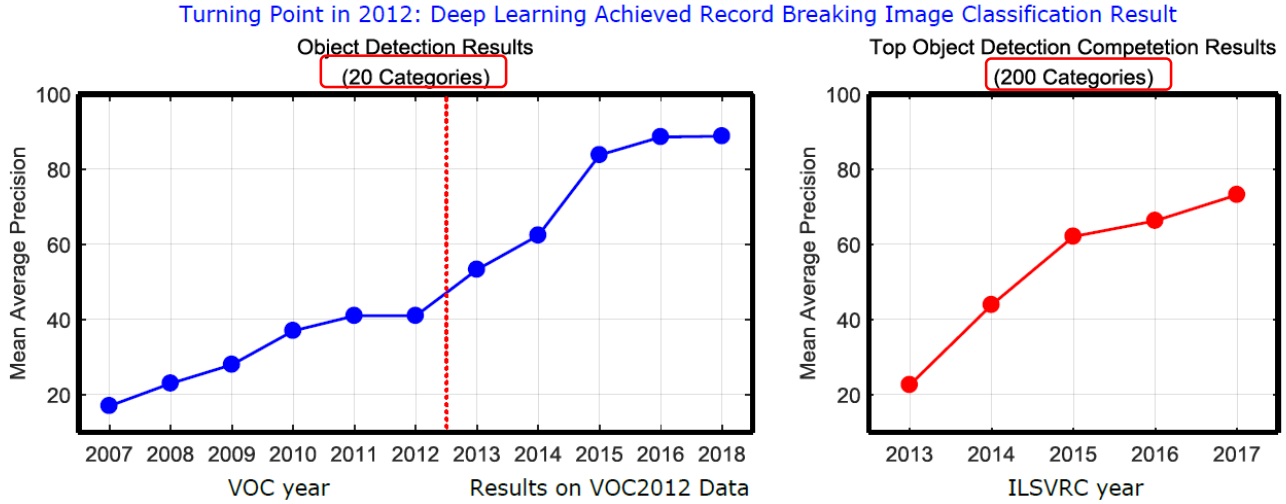
You've visited this page 2 times. Last visit: 22/12/20

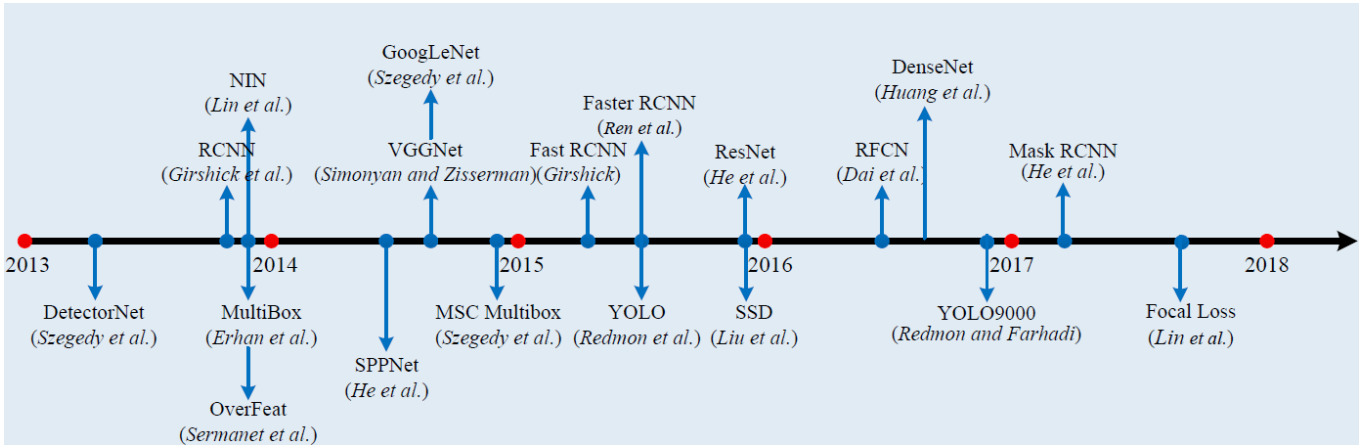


# Object detection

- Milestones



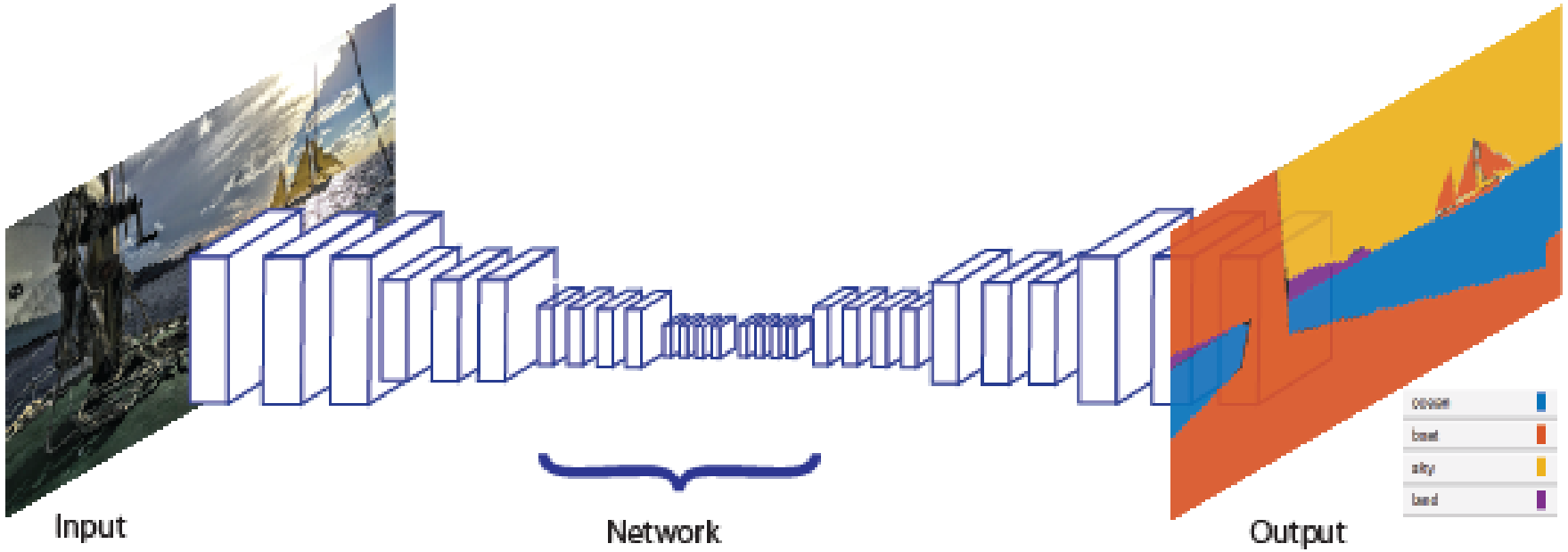




https://www.youtube.com/watch?v=9DQGuD1sHBI

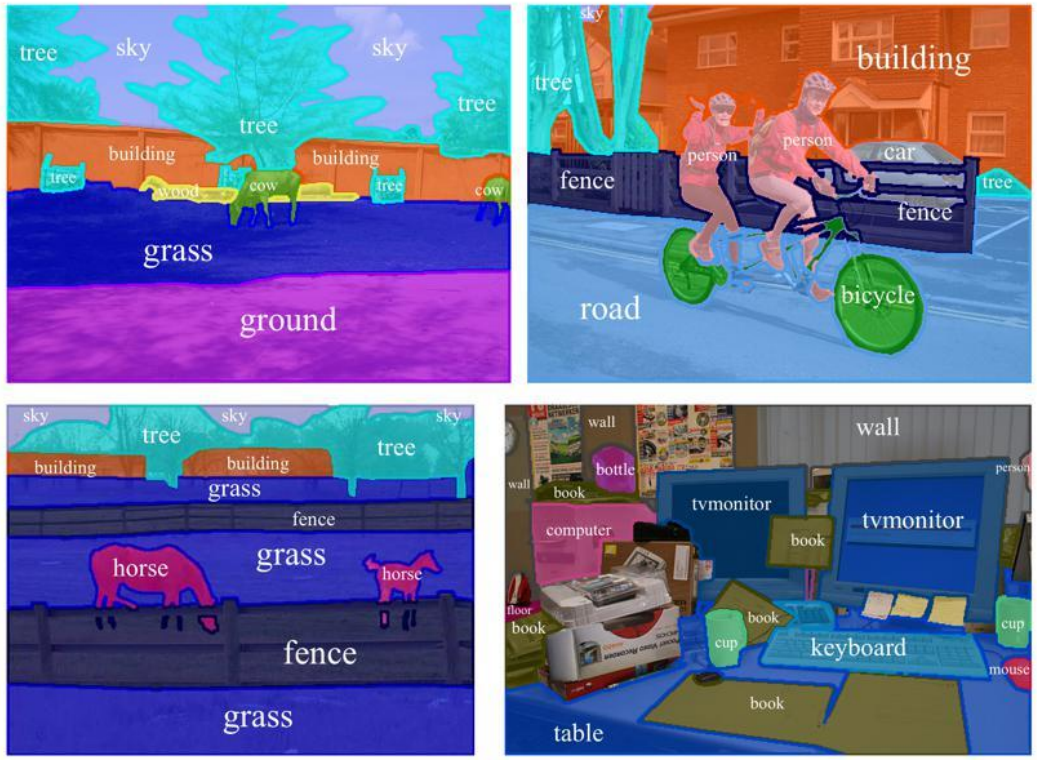
# Image segmentation

- What is segmentation?
  - Input: image
  - Output: regions, structures



# Image segmentation

- What is **semantic** segmentation?
  - Idea: recognizing, understanding what's in the image.
  - "Two men riding on a bike in front of a building on the road. And there is a car"



# Image segmentation

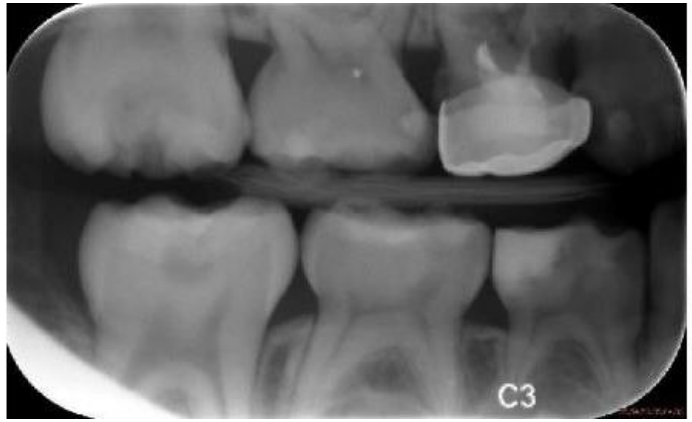
- Why semantic segmentation?
  - Robot vision and understanding
  - Autonomous driving
  - Medical purposes
- Mask R-CNN

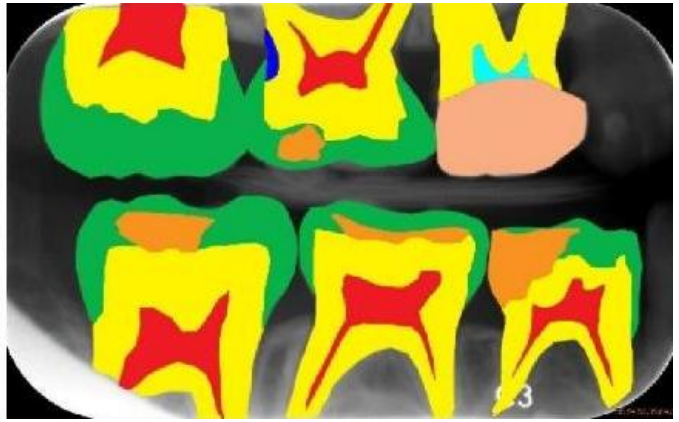
https://arxiv.org › cs

[1703.06870] Mask R-CNN - arXiv

by K He · 2017 · Cited by 20599 — Abstract: We present a conceptually simple, flexible, and general framework for object instance segmentation.

Cite as: arXiv:1703.06870

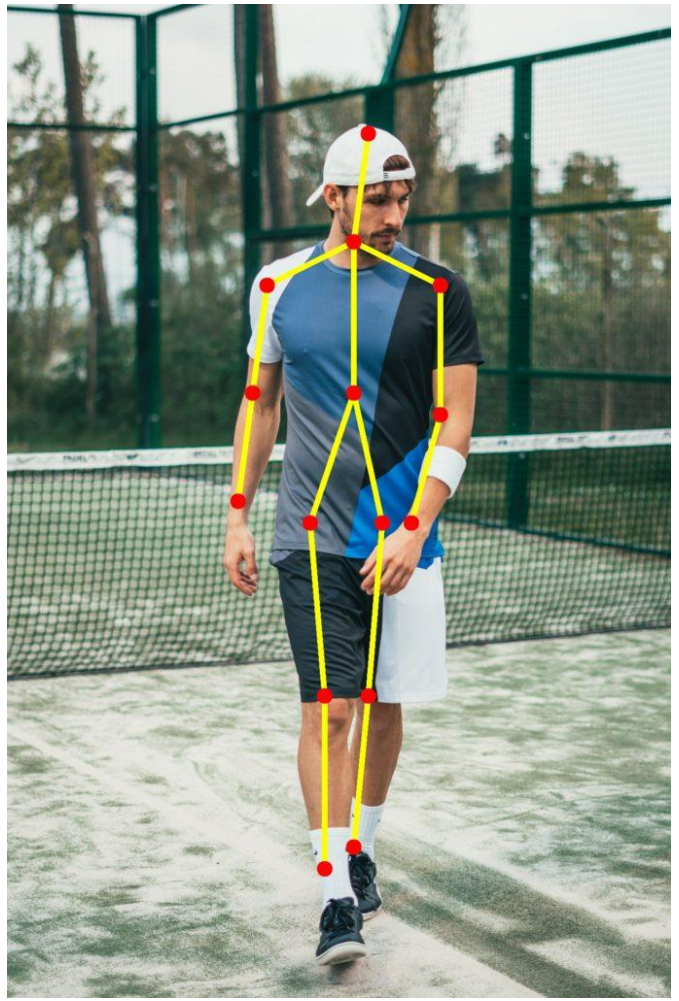


https://www.youtube.com/watch?v=OOT3UIXZztE

# Pose estimation

- What is pose estimation?
  - Keypoint Detection
  - Input: image
  - Output: configuration



https://lmb.informatik.uni-freiburg.de/projects/hand3d/

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

# SOTA-Visual and Language

# Image captioning

- What is image captioning?
  - It is the process of generating textual description of an image
  - It uses both Natural Language Processing and Computer Vision to generate the captions

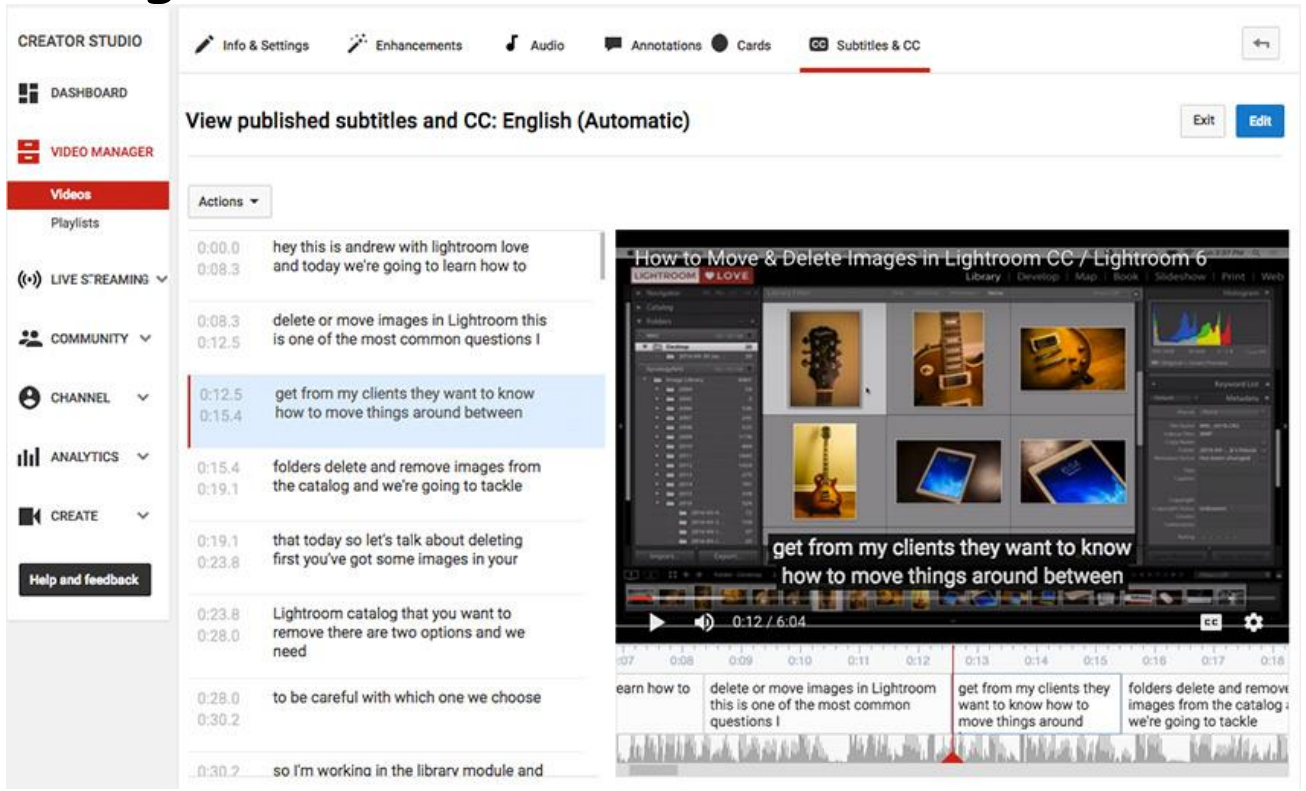
"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

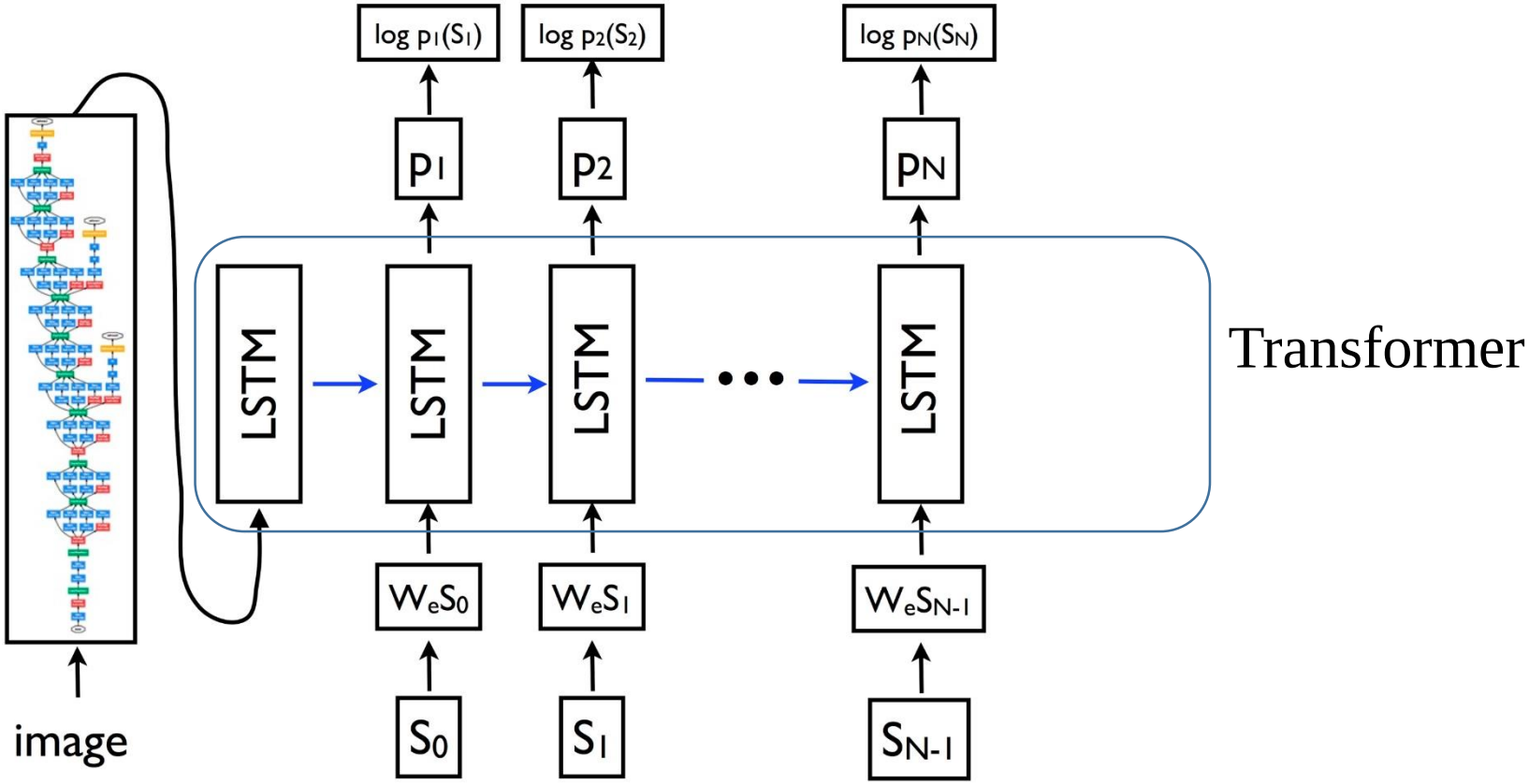
# Image captioning

- Why caption generation?
  - Generating summaries for YouTube videos
  - Captioning unlabeled images
  - Semantic search



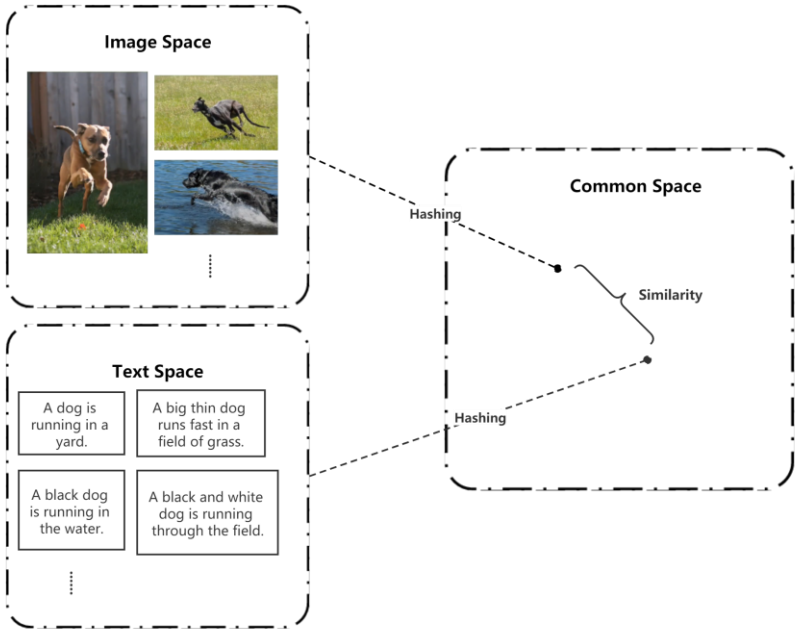
# Image captioning

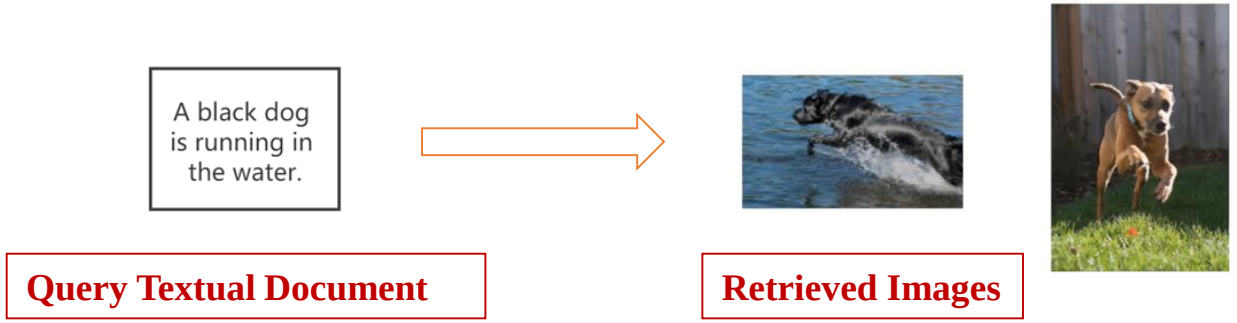
- Framework of image captioning



# Cross-modal retrieval

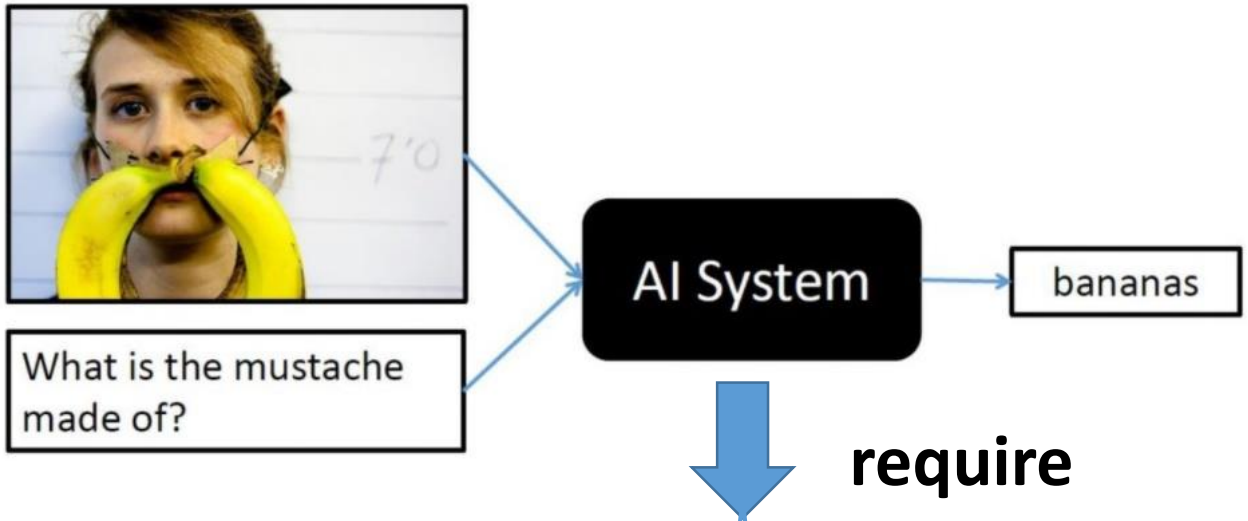
- Cross-modal Retrieval
  - ➢ Support similarity search for multi-modal data, e.g., the retrieval of images in response to a query textual document or vice versa.



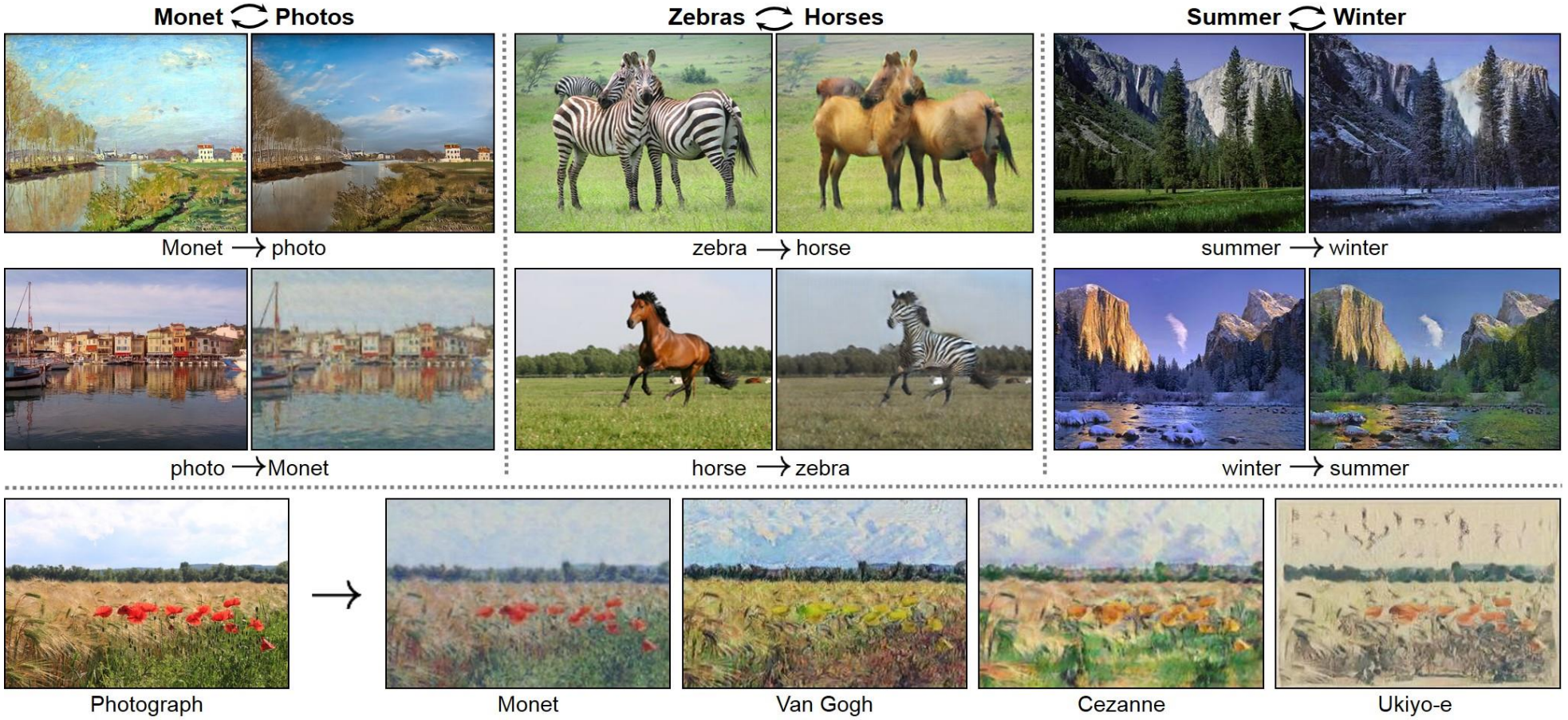
# Visual Question Answering (VQA)

- Given an image and a question (text) about the image
  - Aim to provide an accurate natural language answer

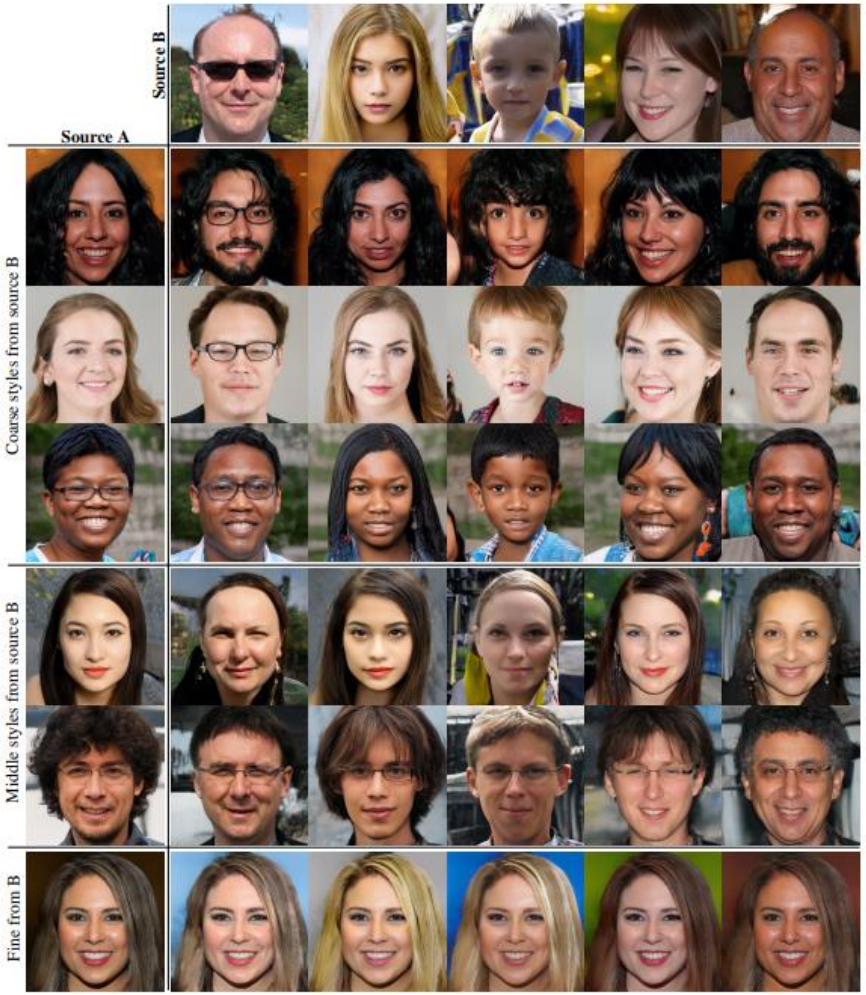


# Generative Images

# Image generation (CycleGAN)



# Image generation






朱茵→杨幂

**A Style-Based Generator Architecture for Generative Adversarial Networks**

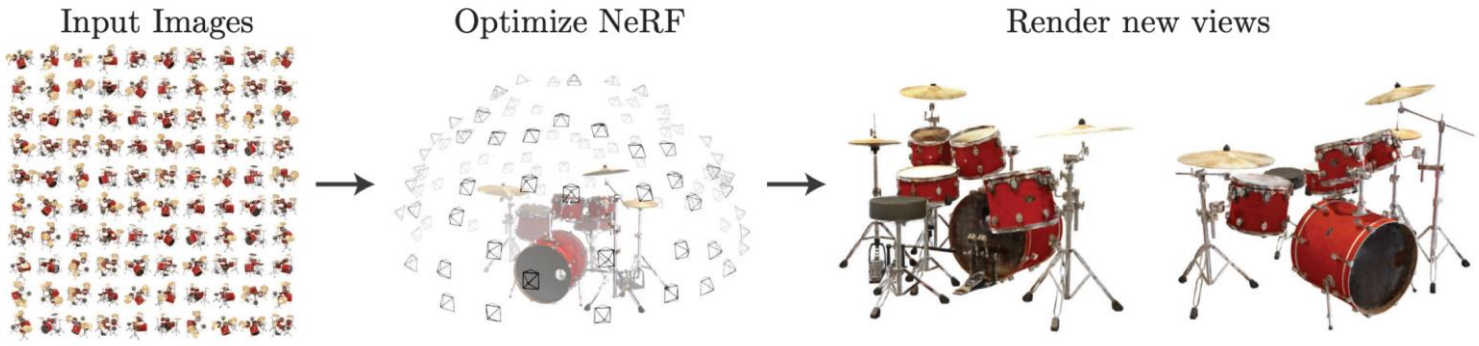
Tero Karras
NVIDIA
tkarras@nvidia.com

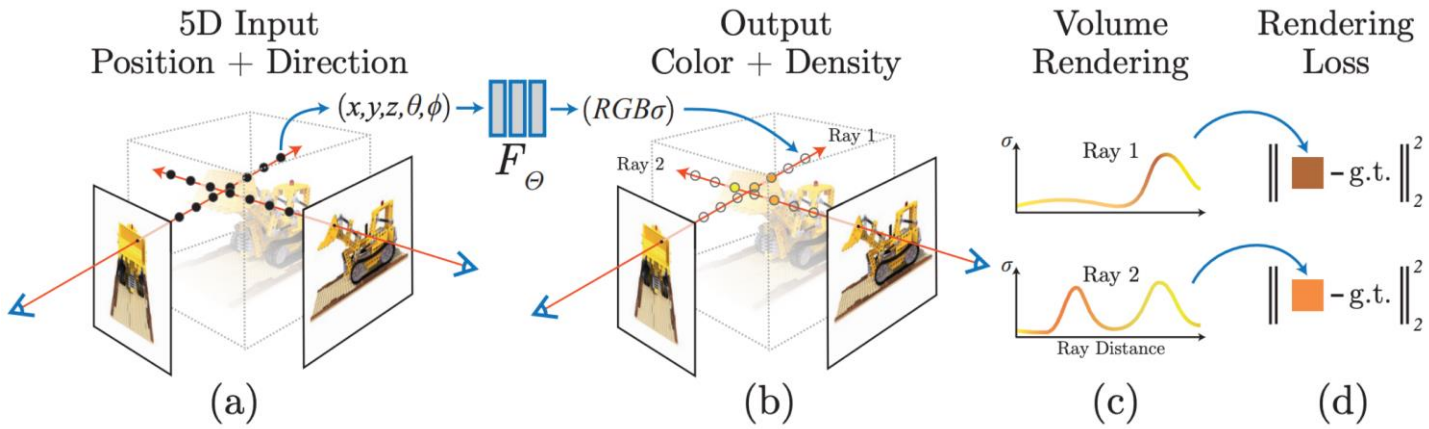
Samuli Laine
NVIDIA
slaine@nvidia.com

Timo Aila
NVIDIA
taila@nvidia.com

# Neural Radiance Field (NeRF)

- Mildenhall et al. ECCV 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis



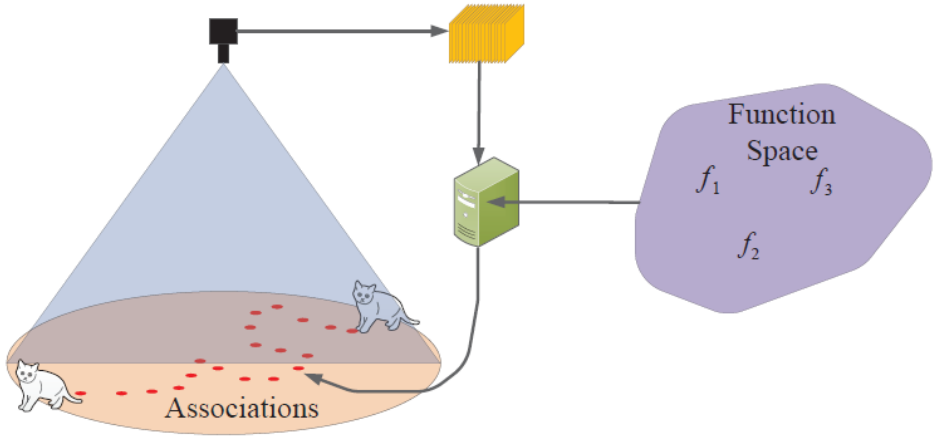

# Neural Radiance Field (NeRF)

- Videos in another slide

# Applications in human-computer interaction

# Object tracking

- What is object tracking?
  - Input: video
  - Output: trajectory




# Object tracking

- Human-structure interaction
  - Inertial sensors: Opal (APDM )
  - Marker-based sensors: Qualisys, VICON, Codamotion
  - Video-based marker-less techniques

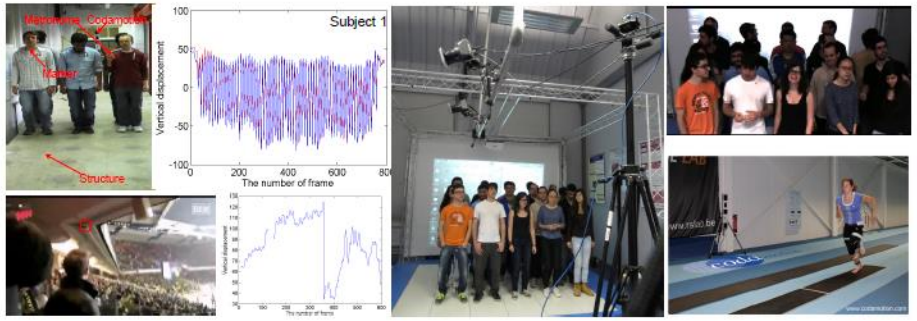



Figure How individuals combine visual and tactile information from other members to synchronise their actions as a group.

Collaborators: James Brownjohn, Vito Racic, Department of Civil and Structural Engineering, University of Sheffield
Mark Elliott: School of Psychology, University of Birmingham

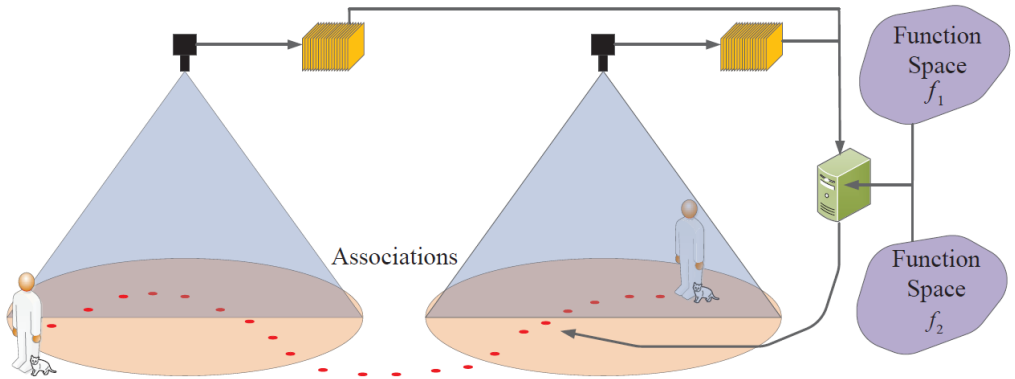
# Object tracking

- Human-computer interaction
- Human-mobile interaction

# Applications in video surveillance

# Object re-identification

- What is object re-identification?
  - Input: images (multi-camera)
  - Output: associations




# Object re-identification

- Vehicle re-identification

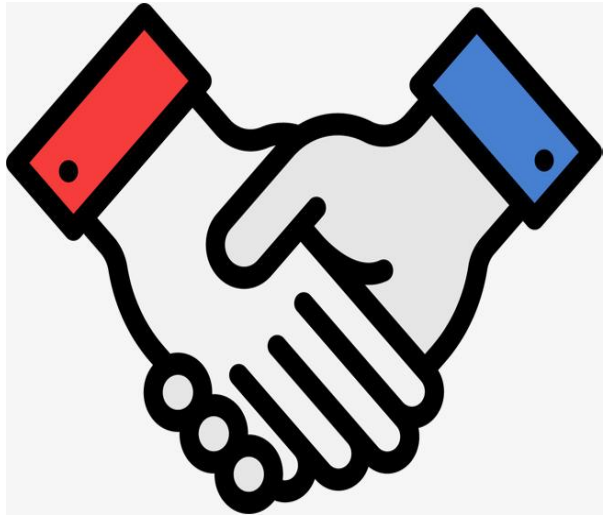
# Conclusions

# Conclusions

- Visual computing is very significant

- Visual computing is very interesting

- Visual computing is easy to study



# Thanks

zhengf@sustc.edu.cn