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

Services

Associate Editor
IET Image Processing

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)

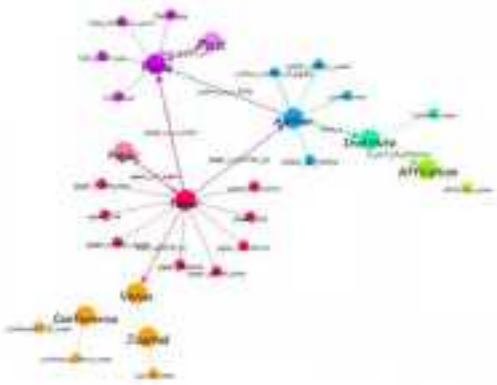
Local Chair
IEEE ICME 2021(CCF-B类), IJCB 2021 (CCF-C类)

Introduction of CS308

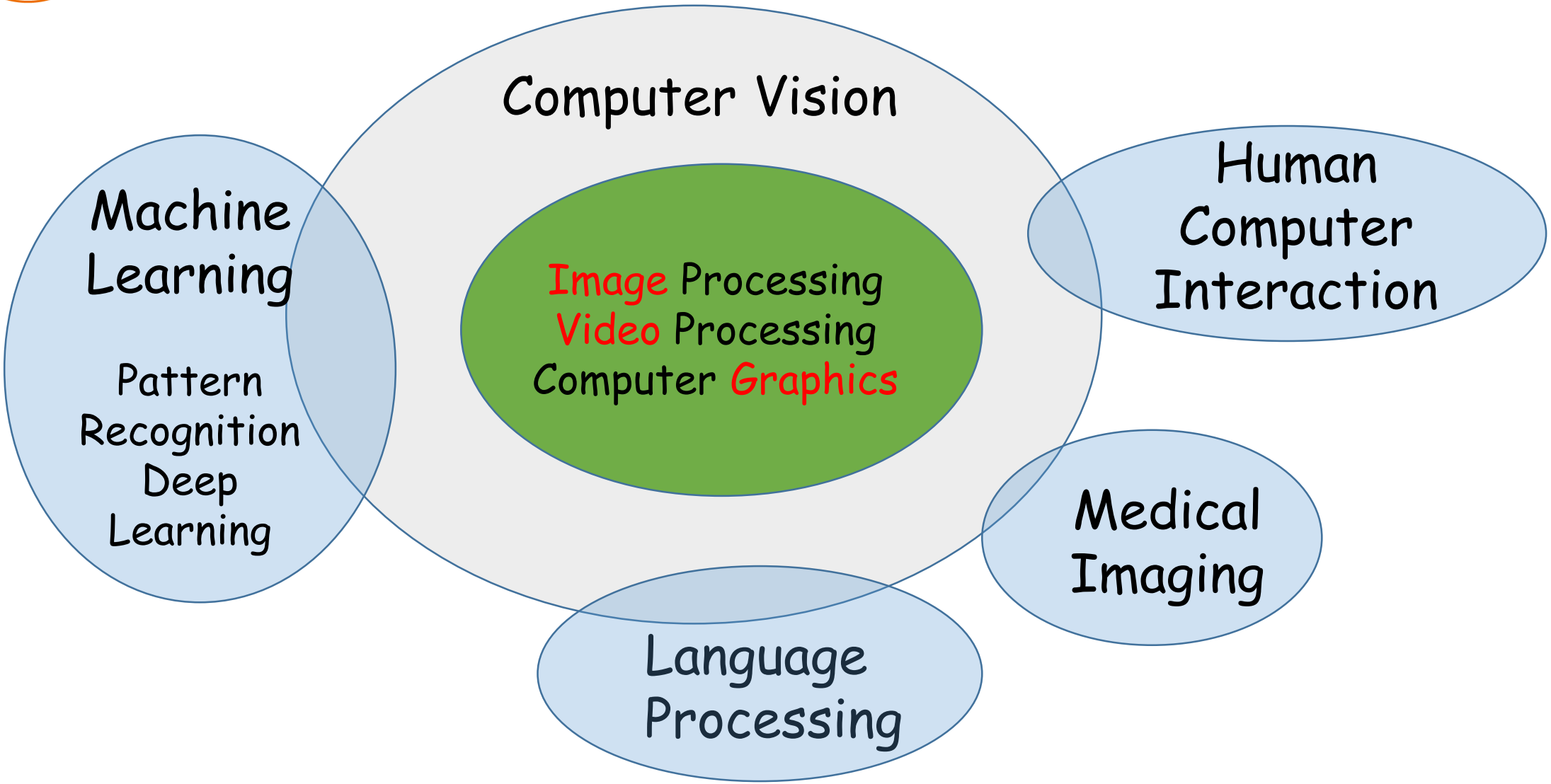
What Is Computer Vision?

- 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

+

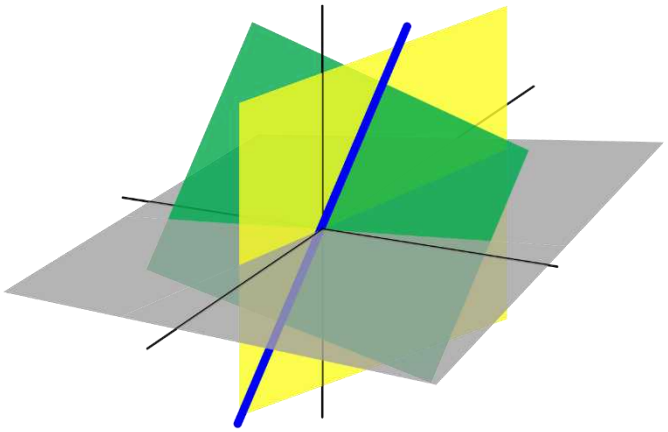


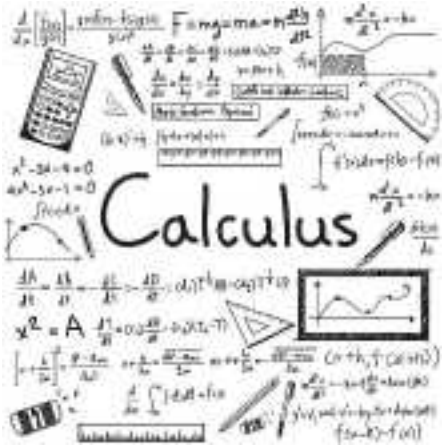
CS308-Computer Vision



CS308-Computer Vision

- Linear algebra
- Basic calculus
- Probability

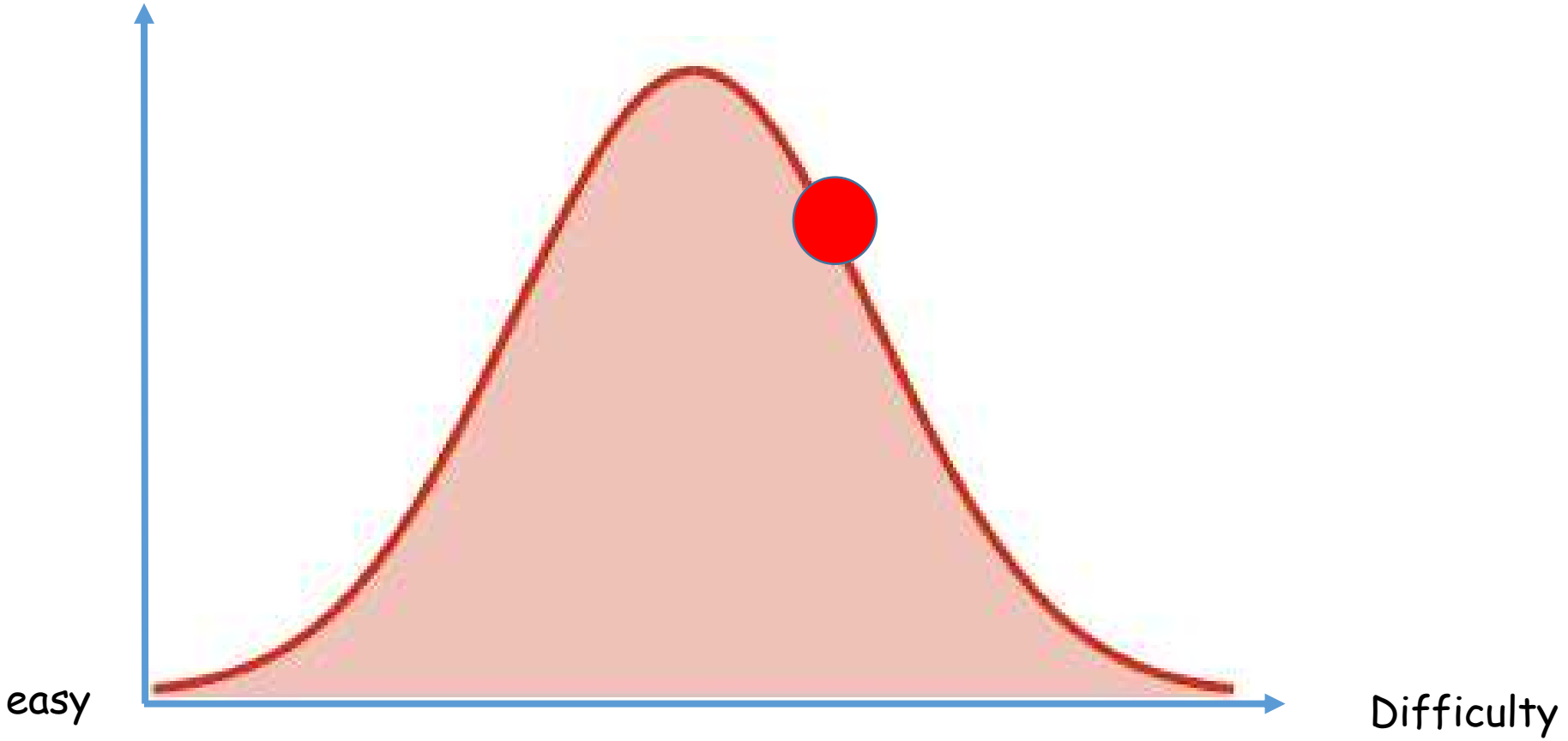




- Experience with image processing will help but is not necessary
- Experience with Python or Python-like languages will help

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

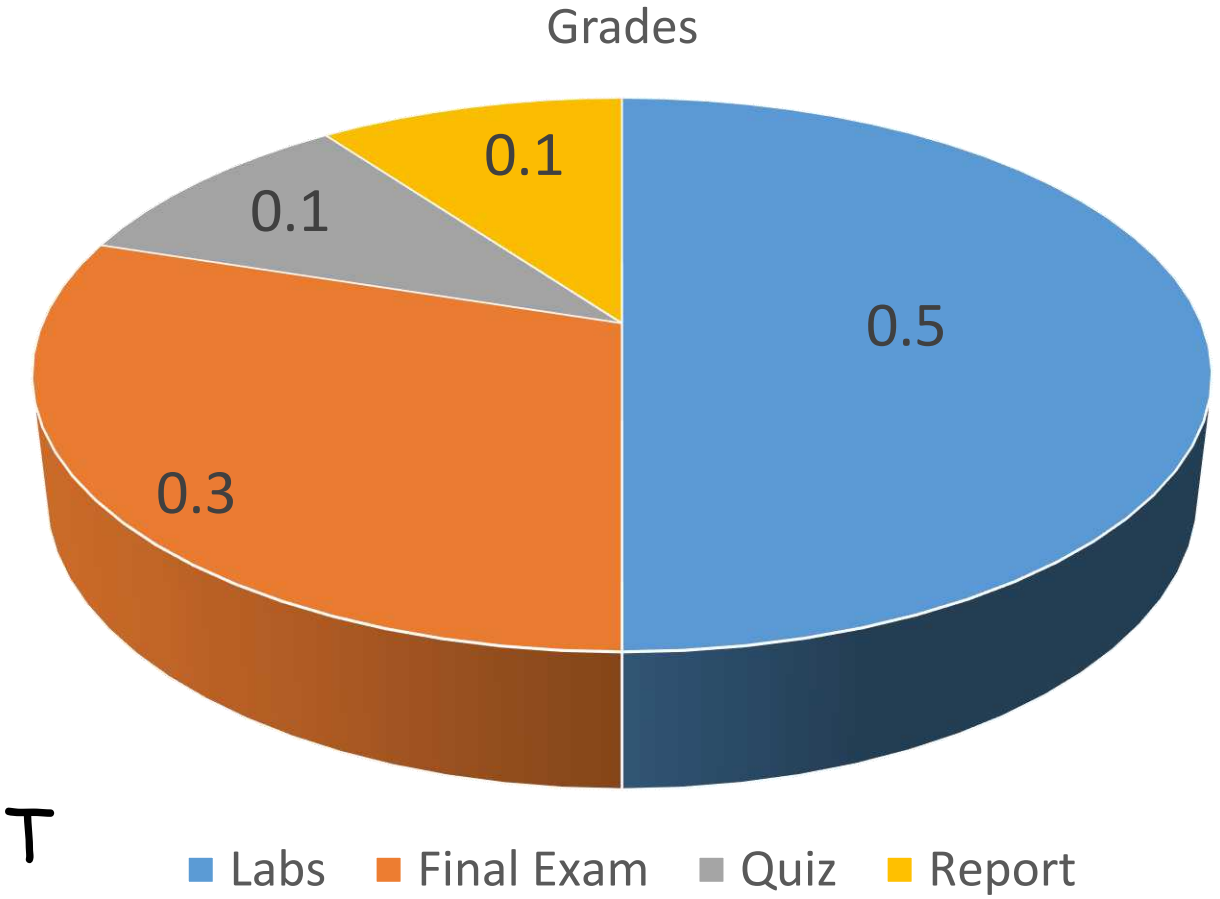
Exams test you on

- 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

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.

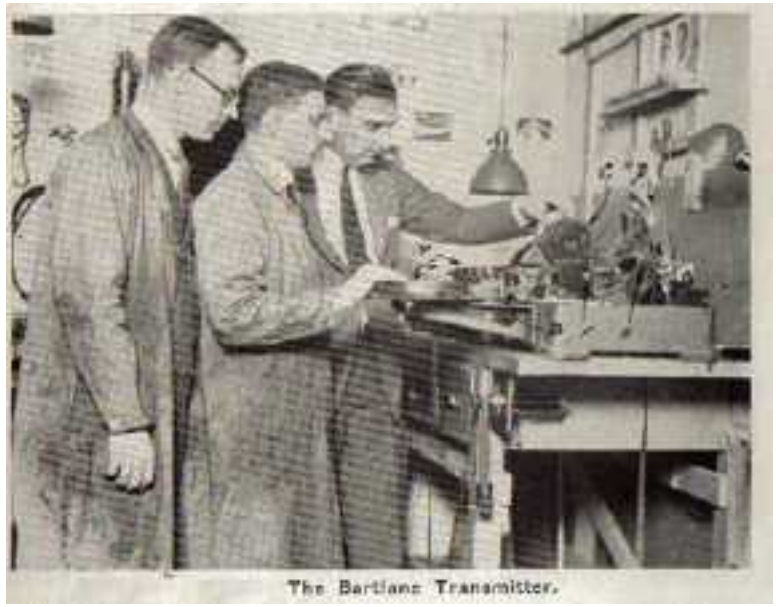
Walden Kirsch as scanned into the SEAC computer by Russell A. Kirsch in 1957.

Visual data: video

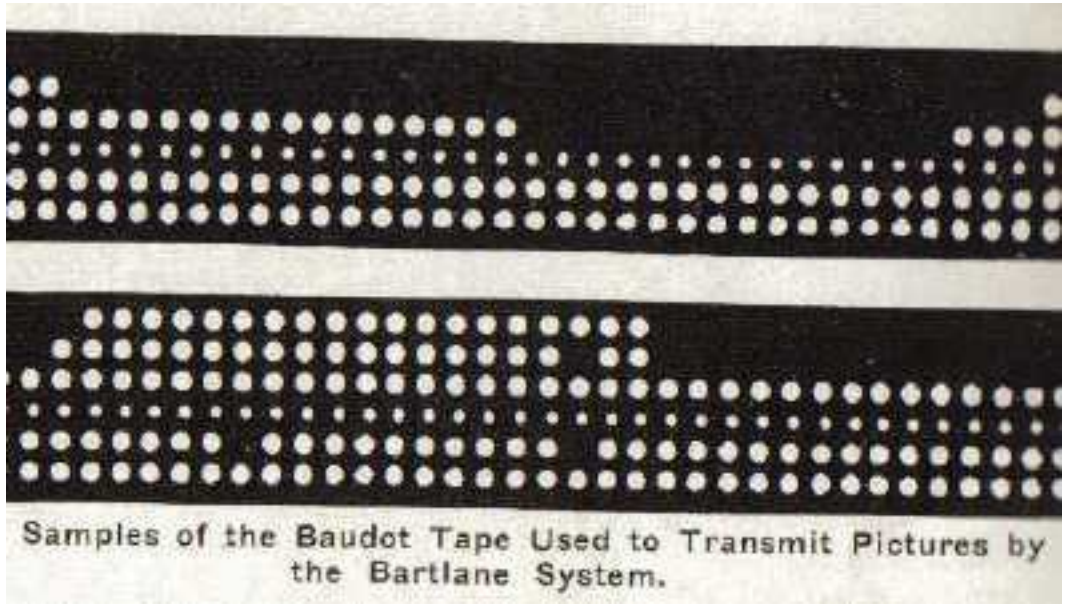
- The first video

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.

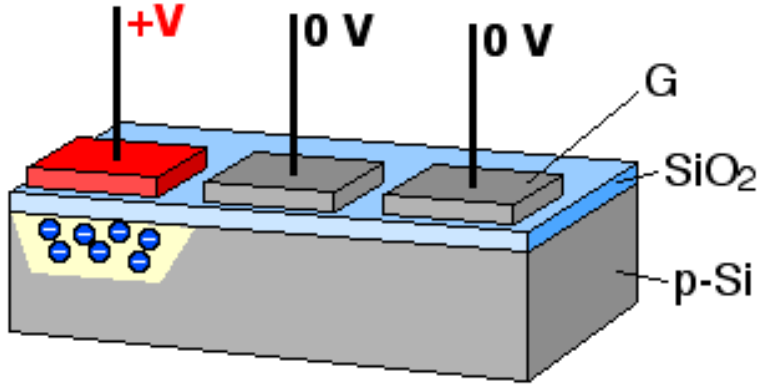


Samples of the Baudot Tape Used to Transmit Pictures by the Bartlane System.

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

Visual data: digital imaging

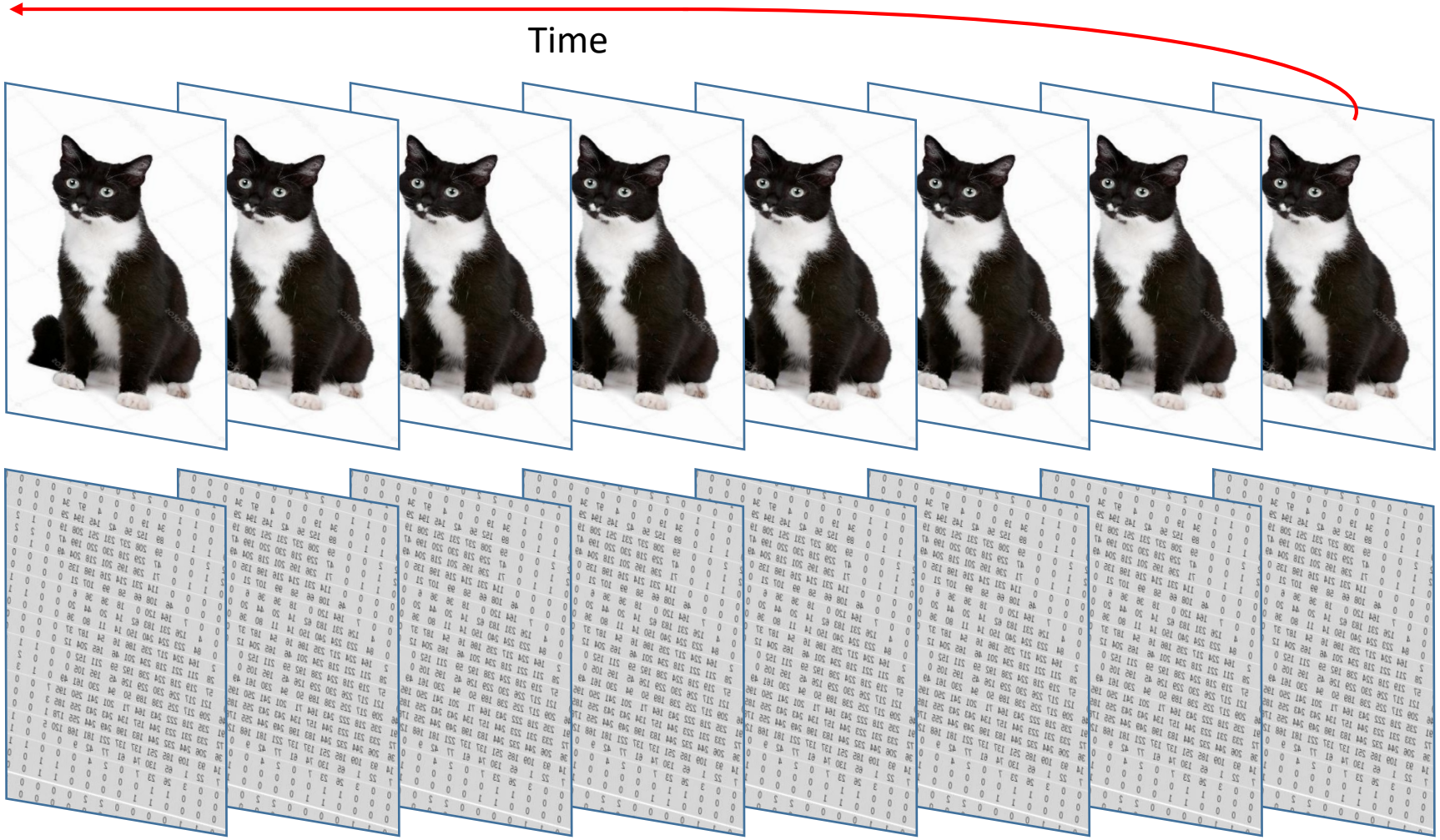
- Charge-coupled device(CCD)
  - AT&T Bell Labs(1969) by Willard Boyle and George E. Smith
  - A piece of lens
  - A Capacitor array (the photoactive region)
  - A control circuit



Visual data: matrix

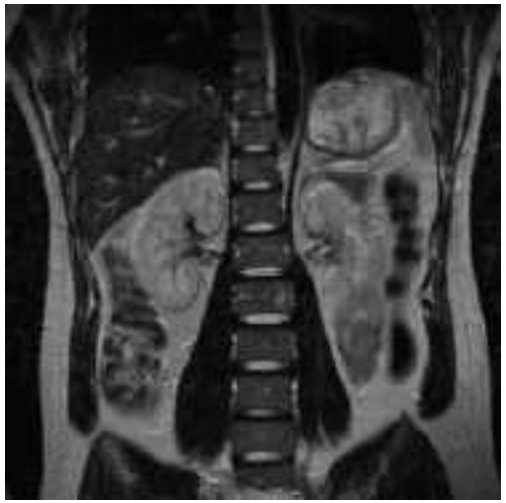
0	0	0	1	1	0	0	0	0	2	2	0	0	0	0	1	1	1
0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	1	0	0
0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	34	19	0	0	4	97	34	0	0	0	0
0	0	0	0	1	1	0	89	152	56	42	145	154	29	0	0	0	0
0	0	0	2	2	0	0	59	208	237	231	251	208	19	0	1	2	3
0	0	2	2	1	0	0	47	229	238	230	220	156	47	0	2	2	1
0	1	1	0	1	1	0	71	236	195	201	218	204	49	0	1	0	0
1	1	0	0	0	0	0	134	231	214	216	198	135	0	0	0	0	0
1	1	0	0	0	0	46	188	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	166	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	196	106	0	0	0	2	3	3
0	0	46	209	217	225	236	189	58	94	230	161	49	0	1	3	1	0
0	0	91	235	218	222	243	164	71	200	241	250	156	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	188	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



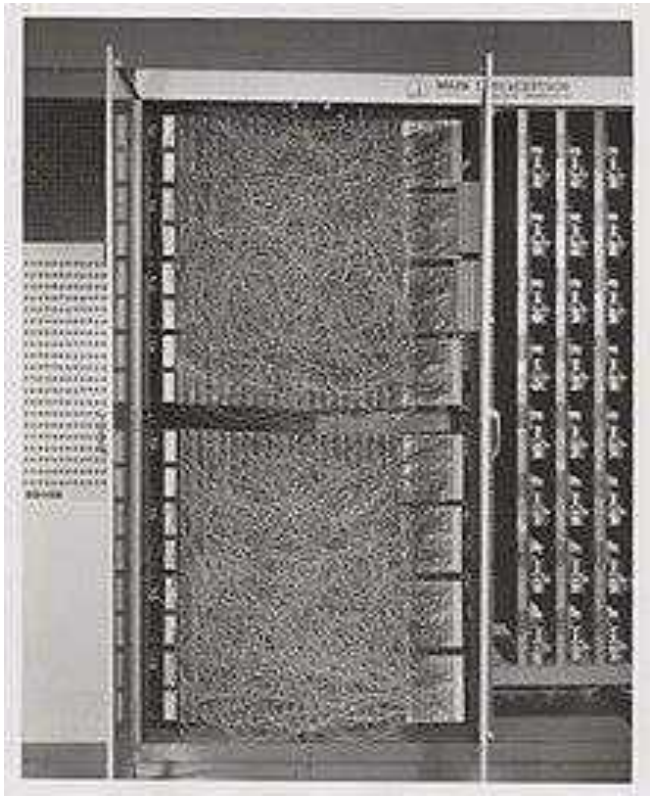
Visual data: more

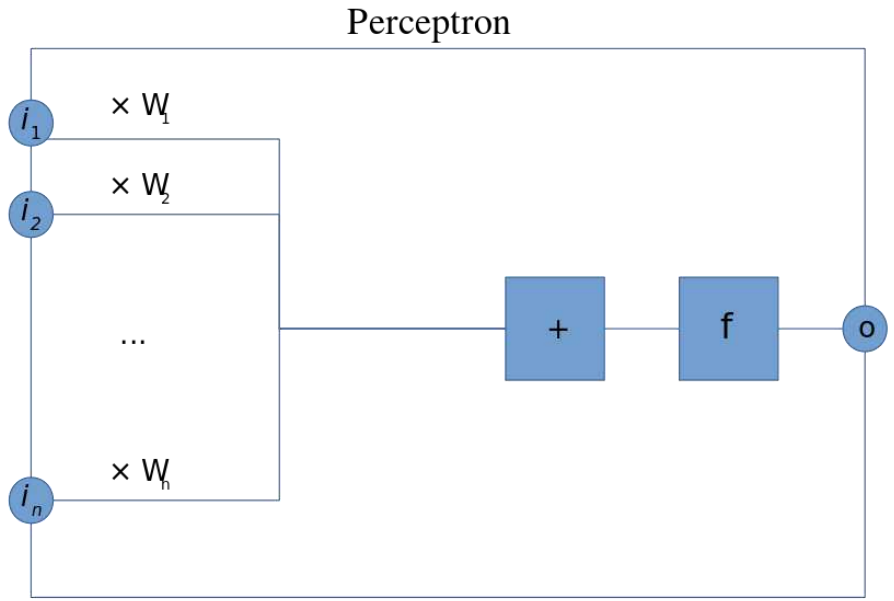
- Depth image (Time of flight, structured-light)
- Ultrasound imaging
- Magnetic resonance imaging



Early work

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

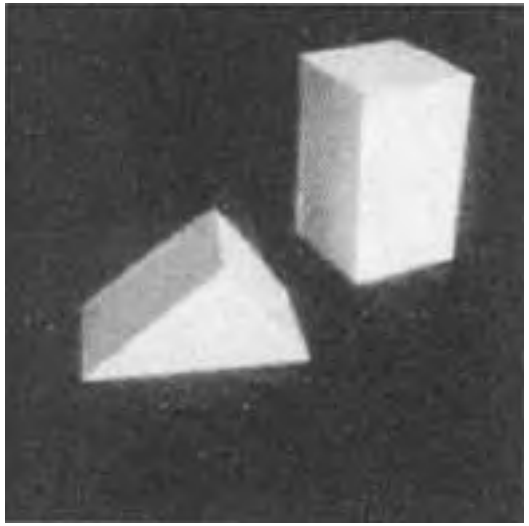




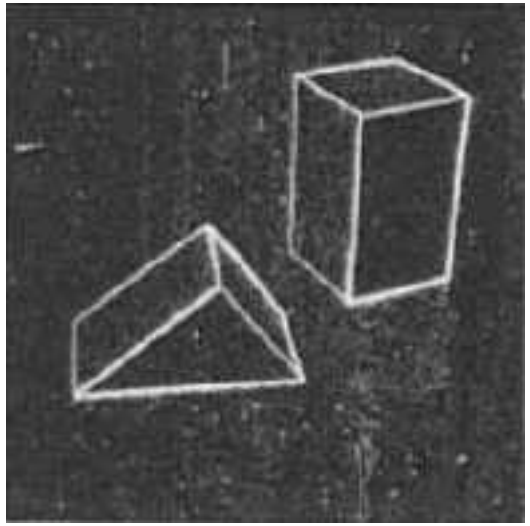
$$o = f\left(\sum_{k=1}^{n} i_k \cdot W_k\right)$$

Early work

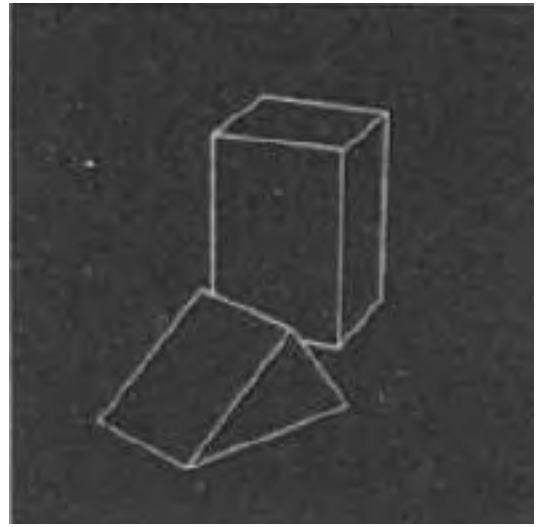
- Larry Roberts at MIT(1960): extracting 3D geometrical information from 2D perspective views of blocks
  - Machine perception of three-dimensional solids
  - Father of computer vision



Input image



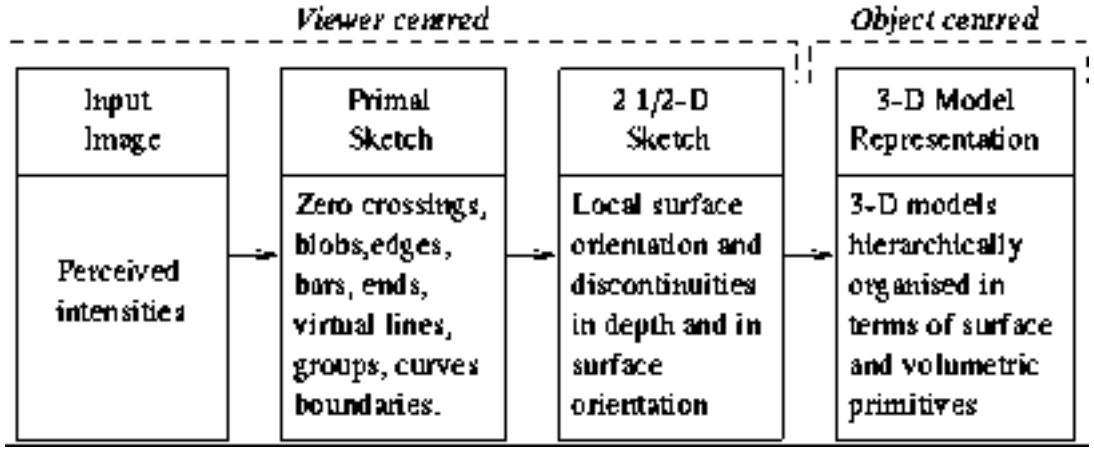
2x2 gradient operator

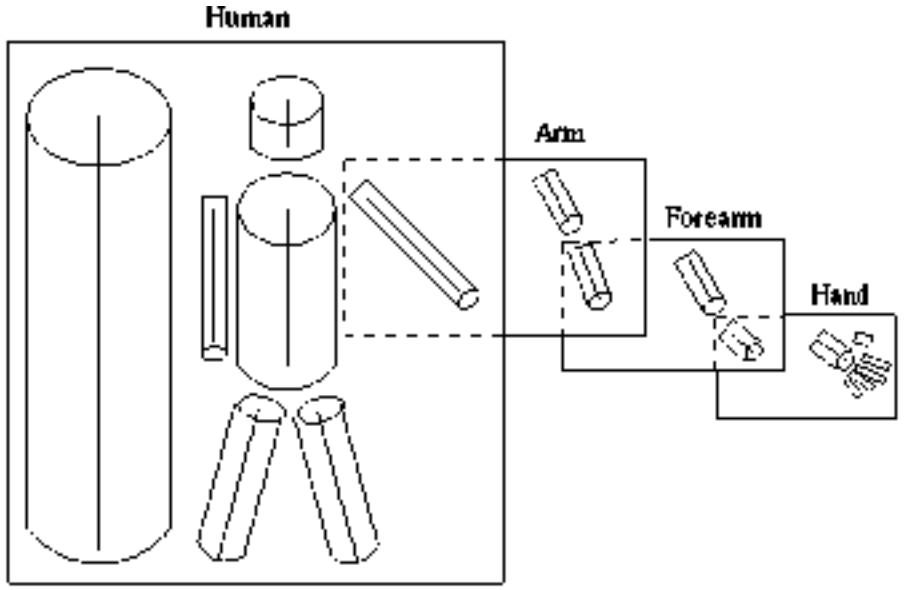


Computed 3D model

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

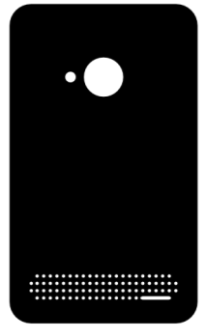
ImageNet Classification with Deep Convolutional Neural Nets

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.

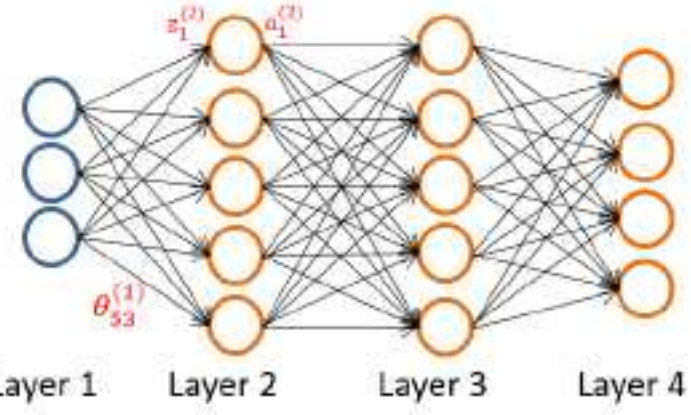
Why it bring about a renaissance?

- Mobile technology with built-in cameras (data)



- Computing power (devices)
- Mass storage

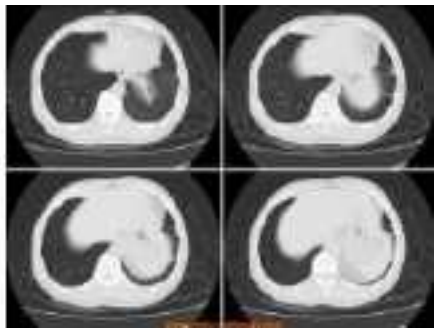
- New algorithms (models)
  - Support vector machine
  - Convolutional neural networks
  - Transformer



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

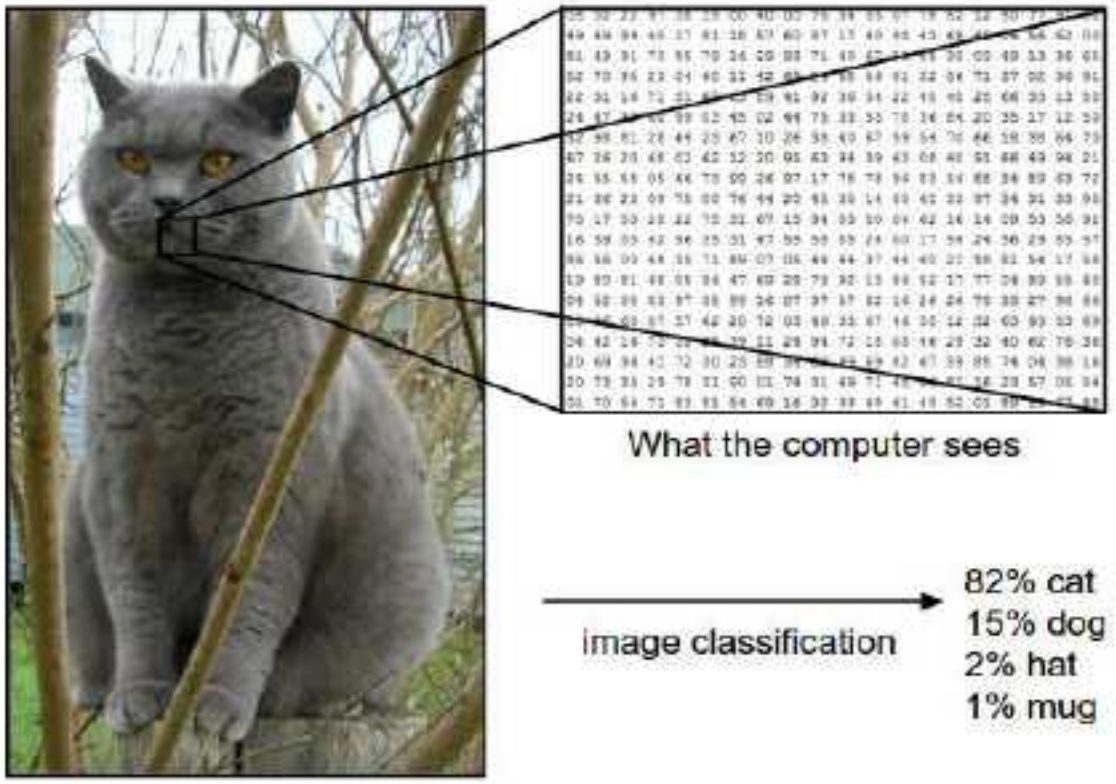


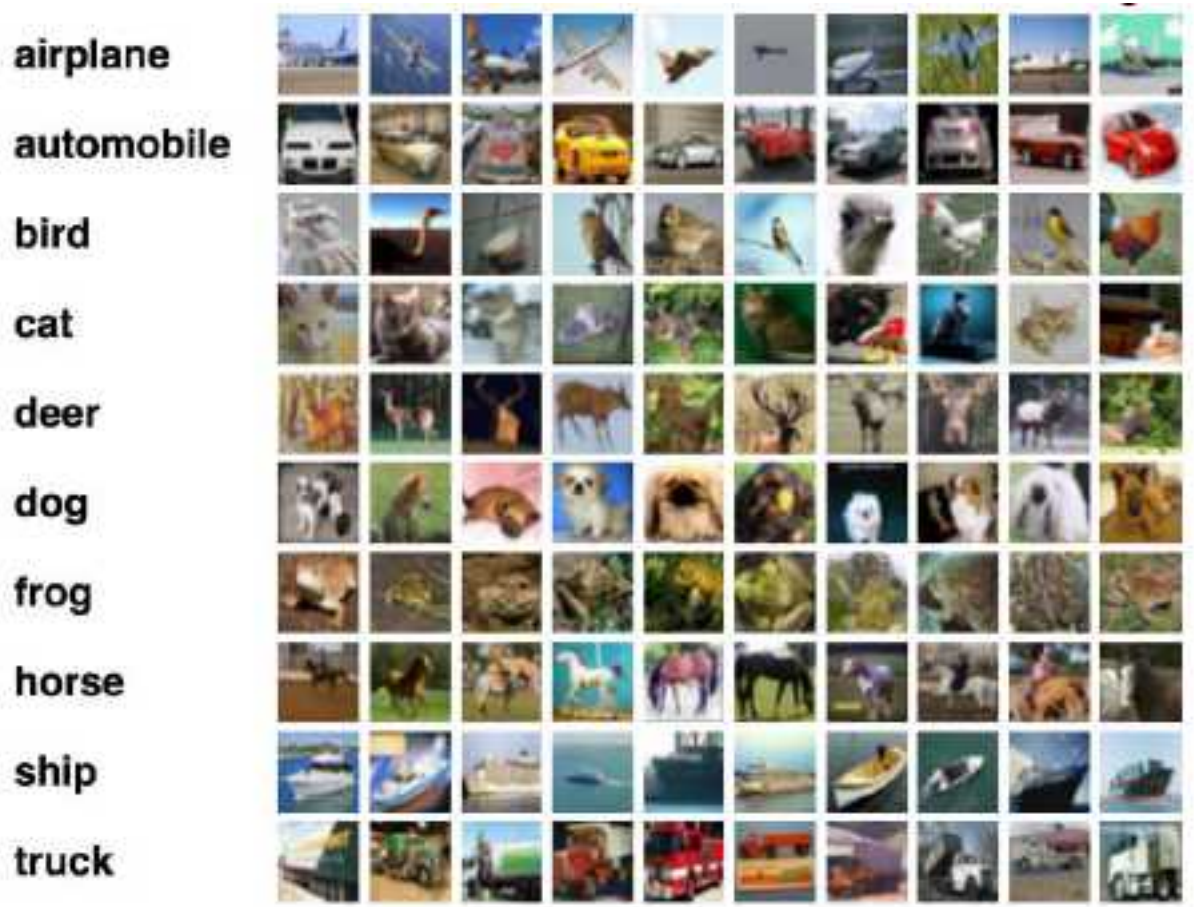
Image classification

- 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.

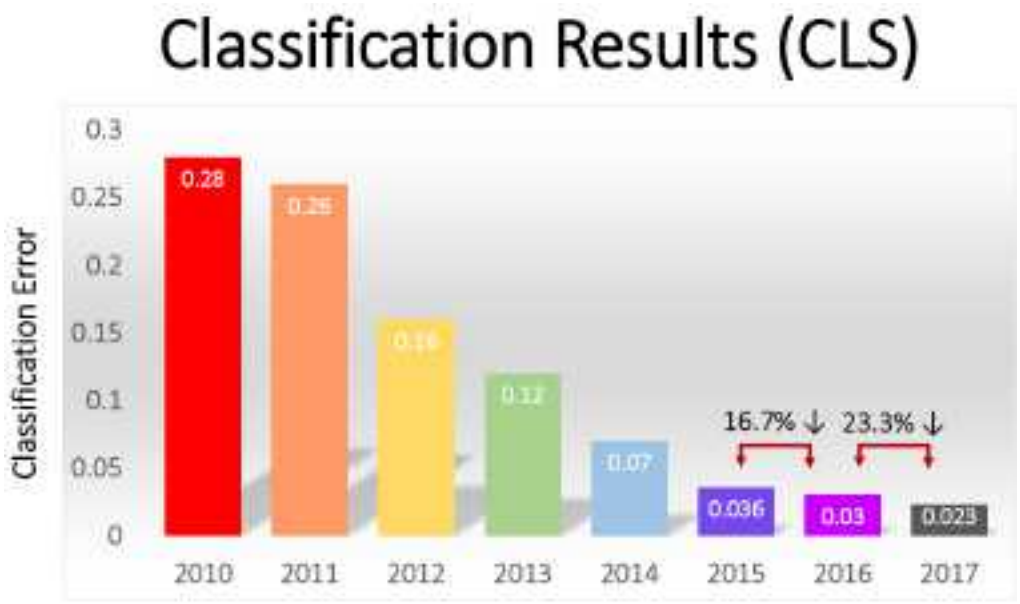


https://www.cs.toronto.edu/~kriz/cifar.html

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...

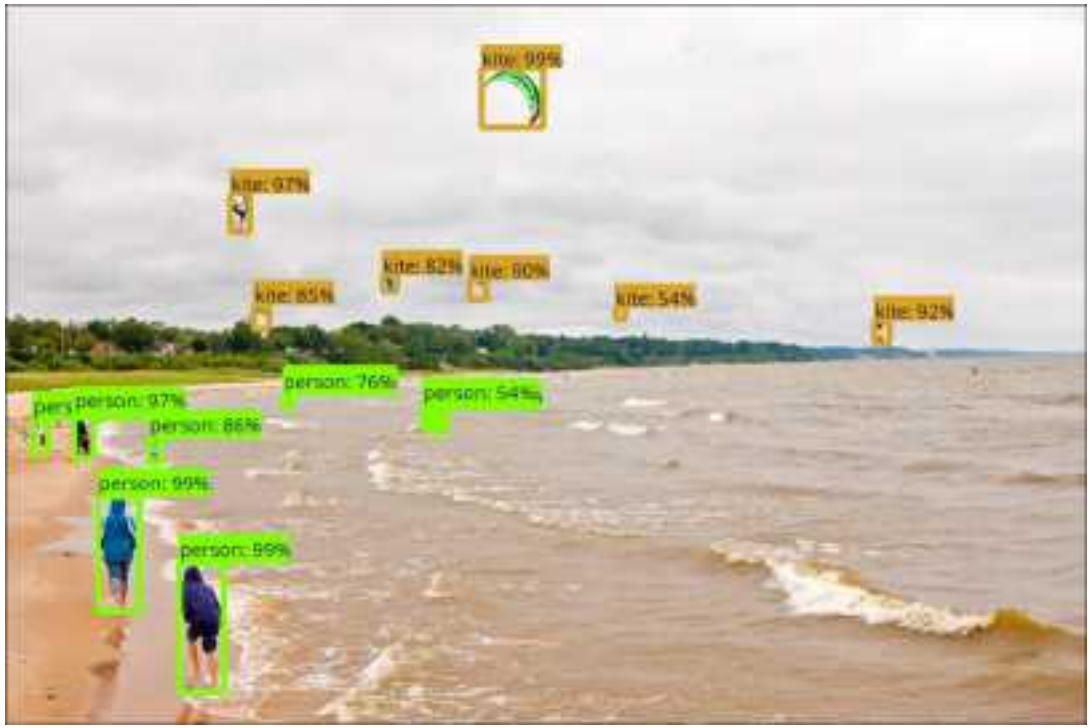
Date Added to IEEE Xplore: 10 August 2009 DOI: 10.1109/CVPR.2009.5206848

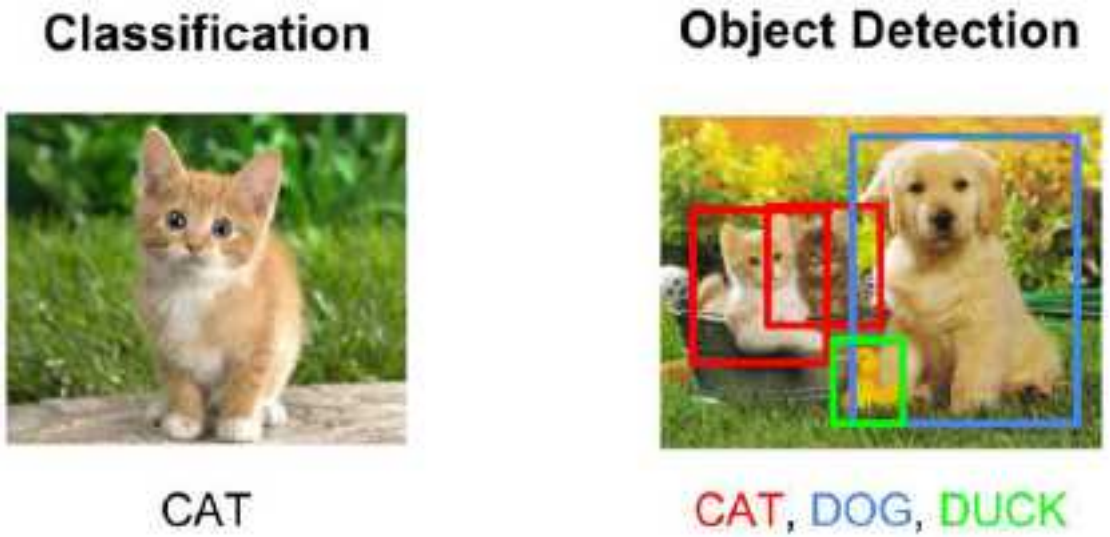
Date of Conference: 20-25 June 2009 INSPEC Accession Number: 10936047

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





Object detection

- The frameworks

  Faster R-CNN

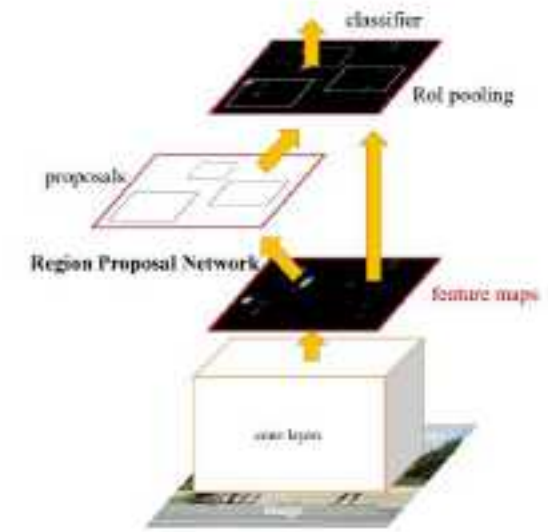


Figure 2: Faster R-CNN is a single, unified network for object detection. The RPN module serves as the 'attention' of this unified network.

http://kaiminghe.com/

Cited by

	All	Since 2015
Citations	145231	142905
h-index	53	53
i10-index	81	81

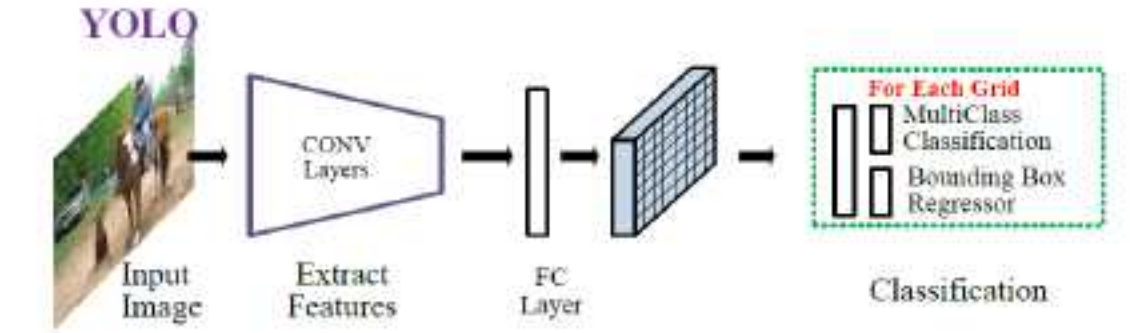
https://arxiv.org › cs

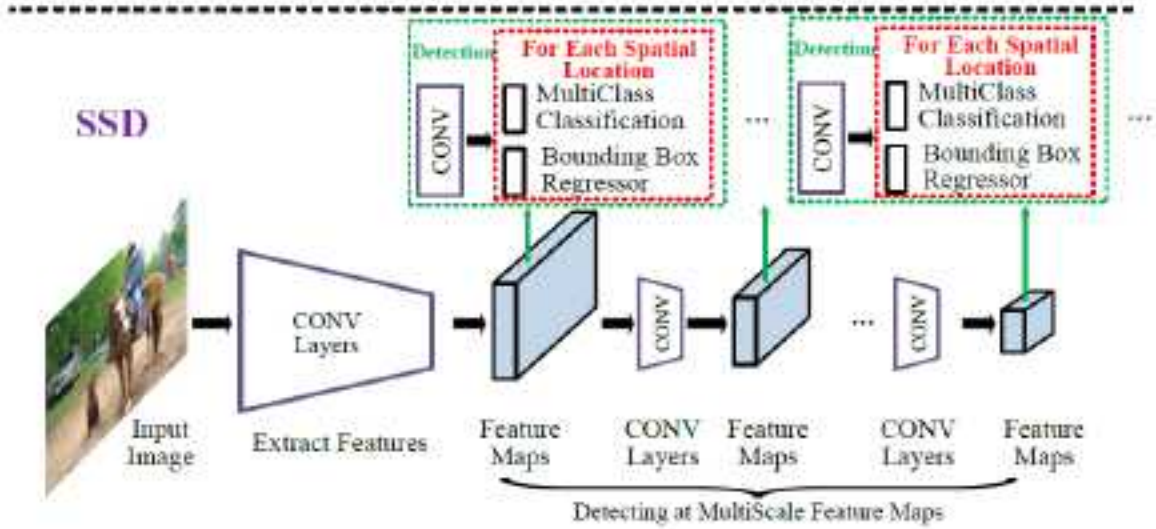
Faster R-CNN: Towards Real-Time Object Detection with ...

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

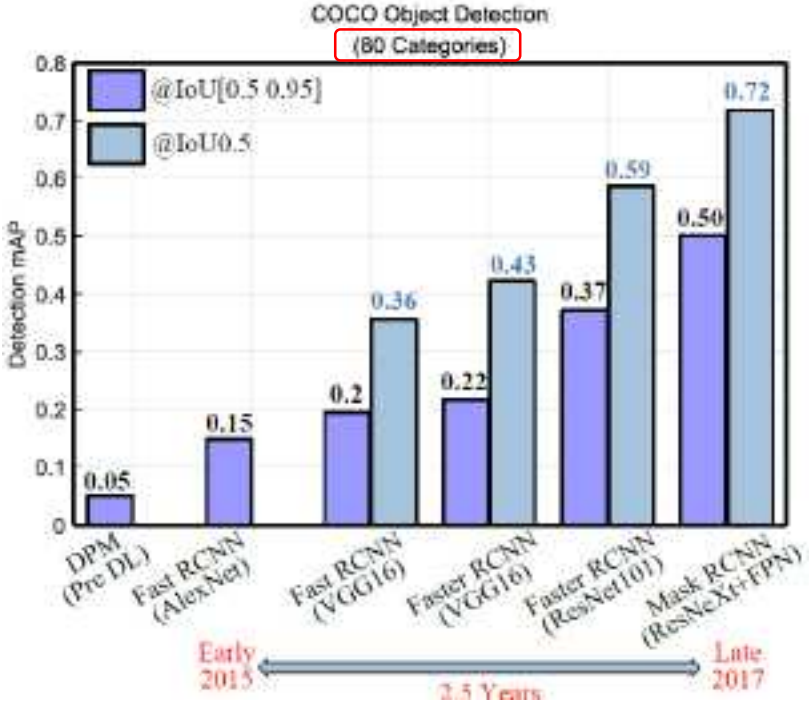
You've visited this page 2 times. Last visit: 2/21/20

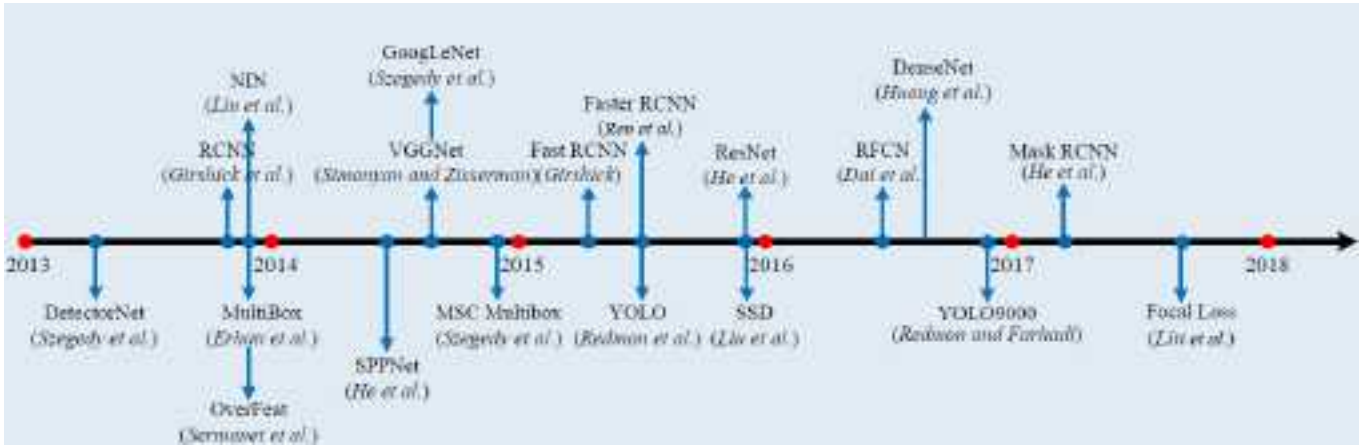




Object detection

- Milestones





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

Image segmentation

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

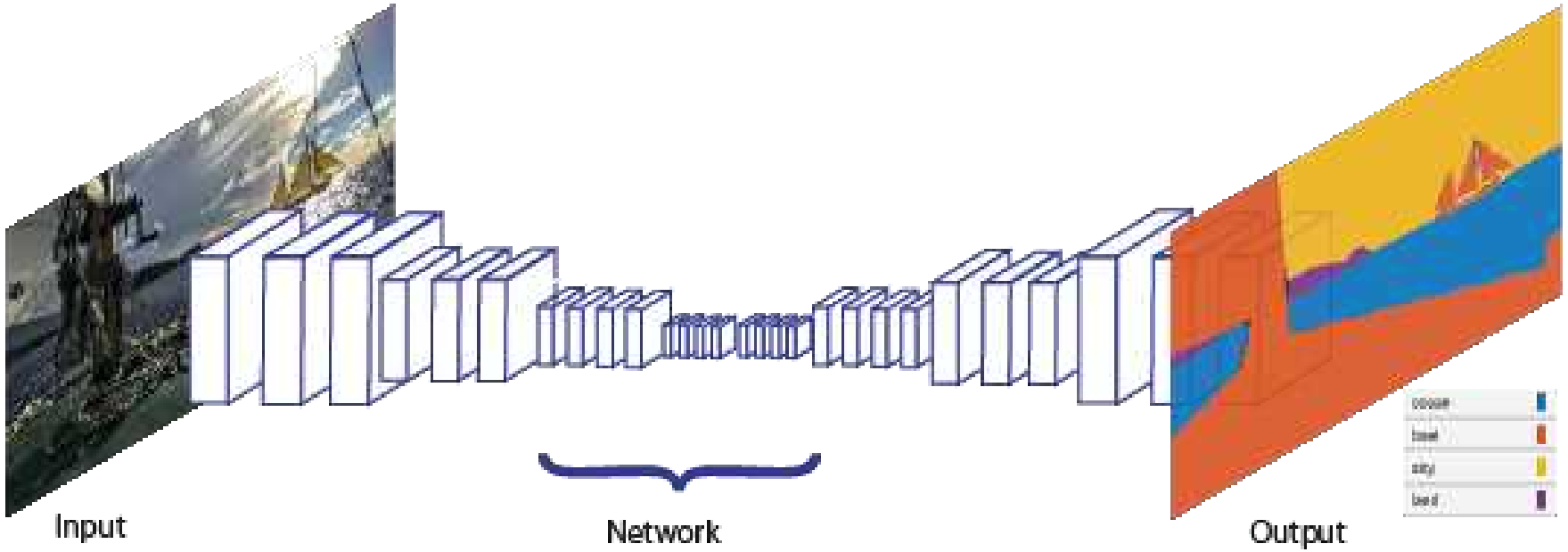


Image segmentation

- 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

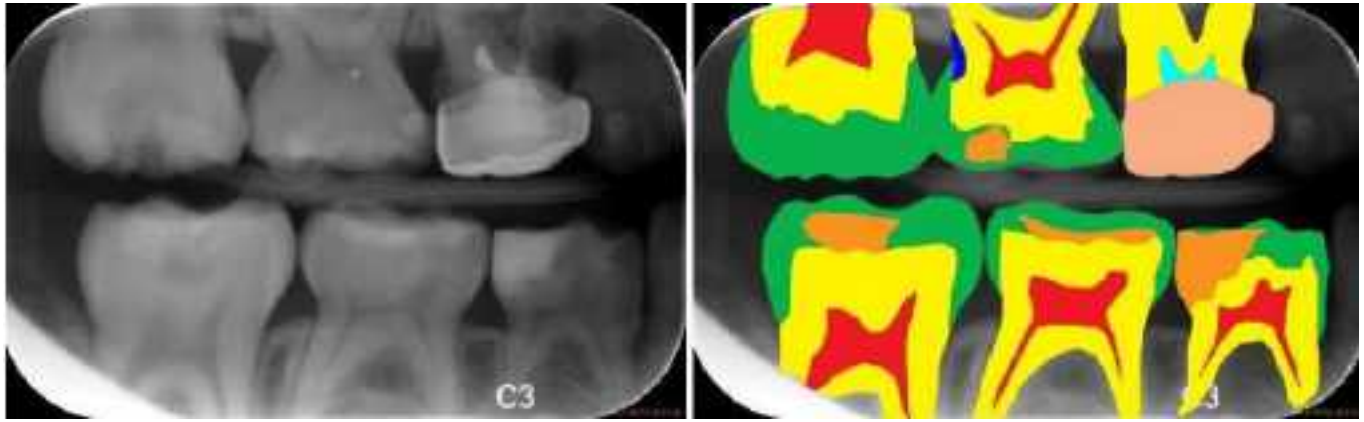
- Why semantic segmentation?
  - Robot vision and understanding
  - Autonomous driving
  - Medical purposes
- Mask R-CNN

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."

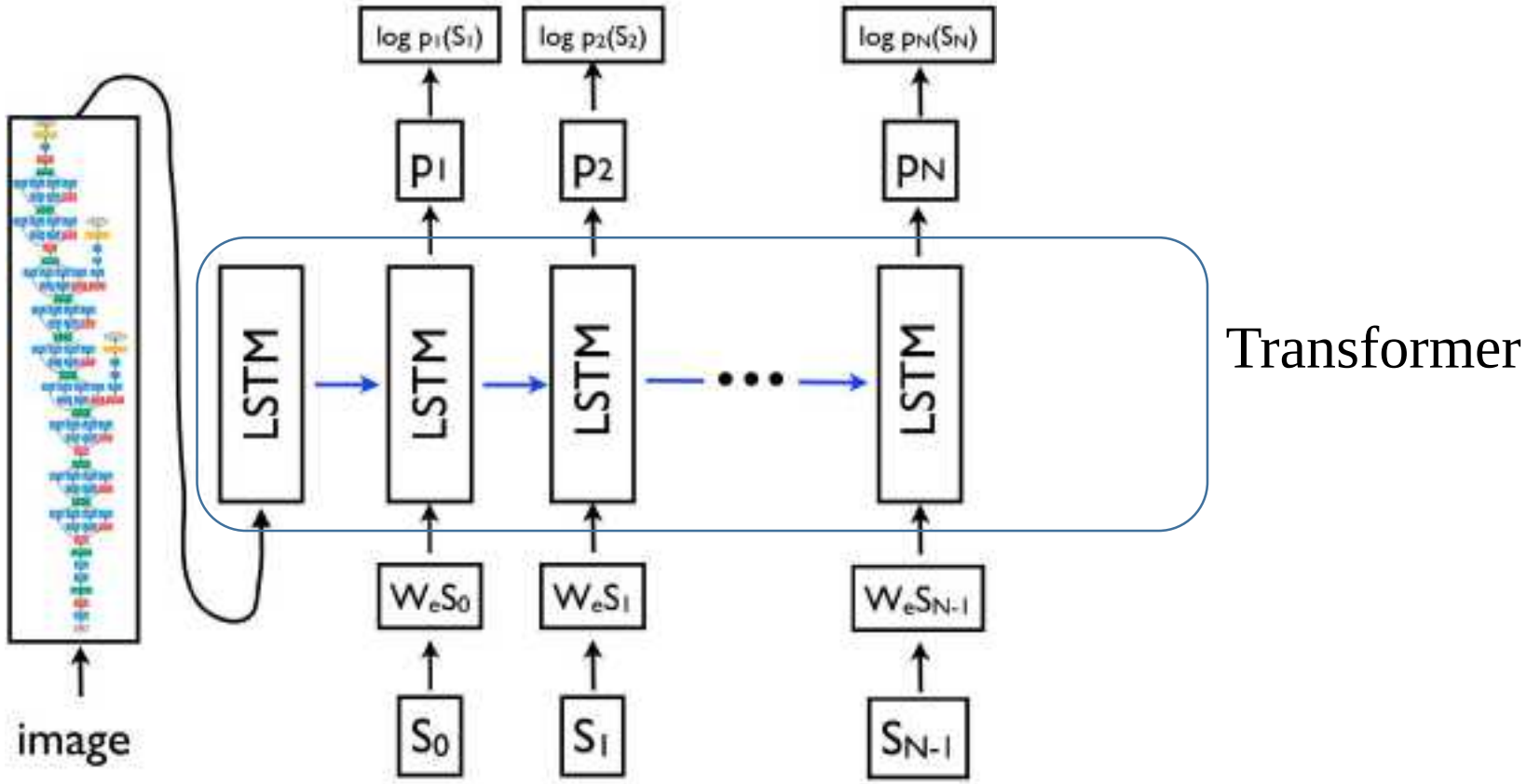
"two young girls are playing with lego toy."

Image captioning

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

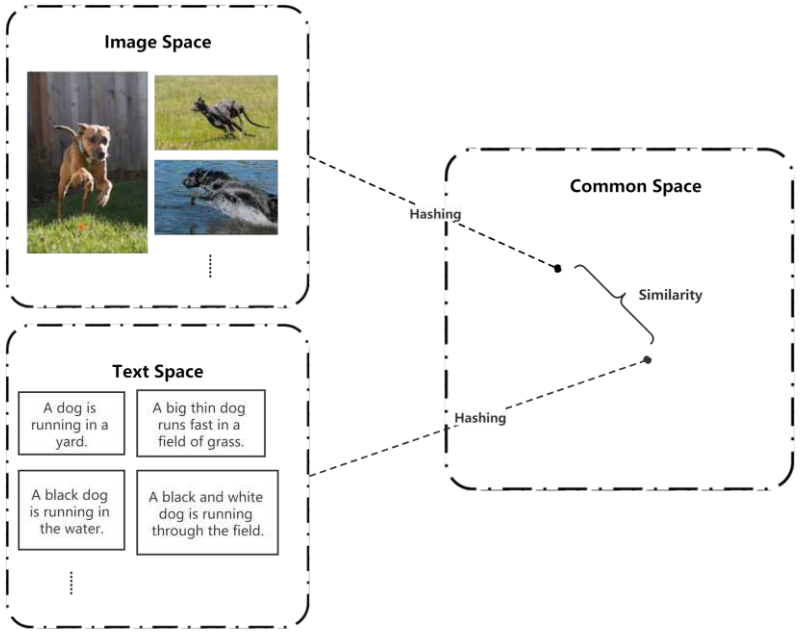
Image captioning

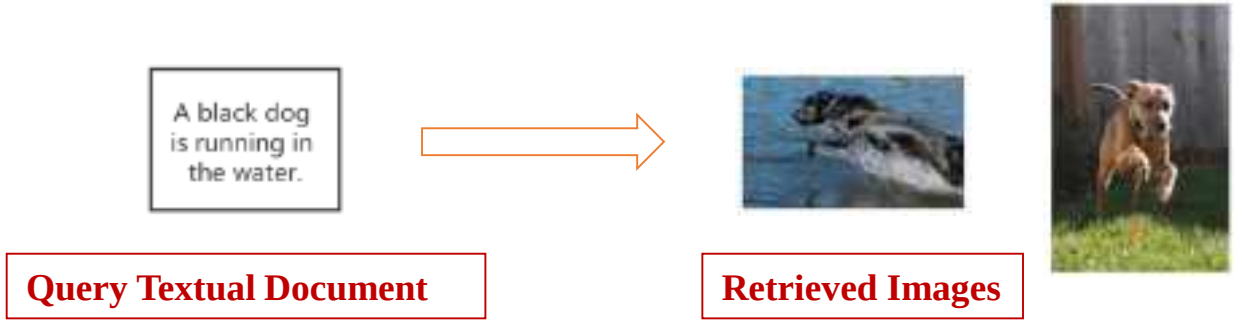
- Framework of image captioning



Cross-modal retrieval

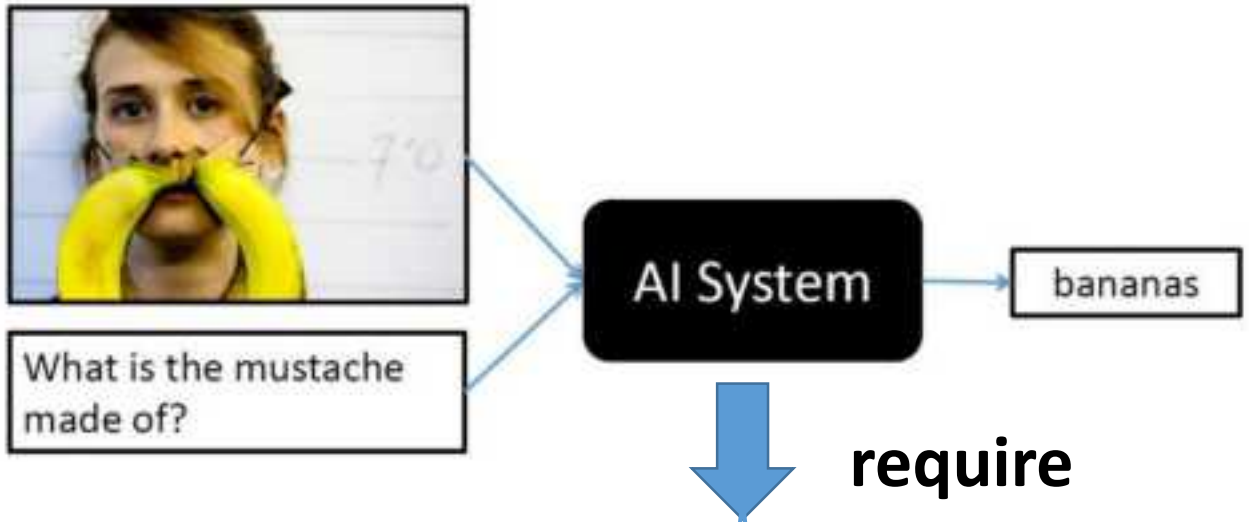
- Cross-modal Retrieval
  - ➢ Support similarity search for multi-modal data, e.g., the retrieval of images in response to a query textual document or vice versa.





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)

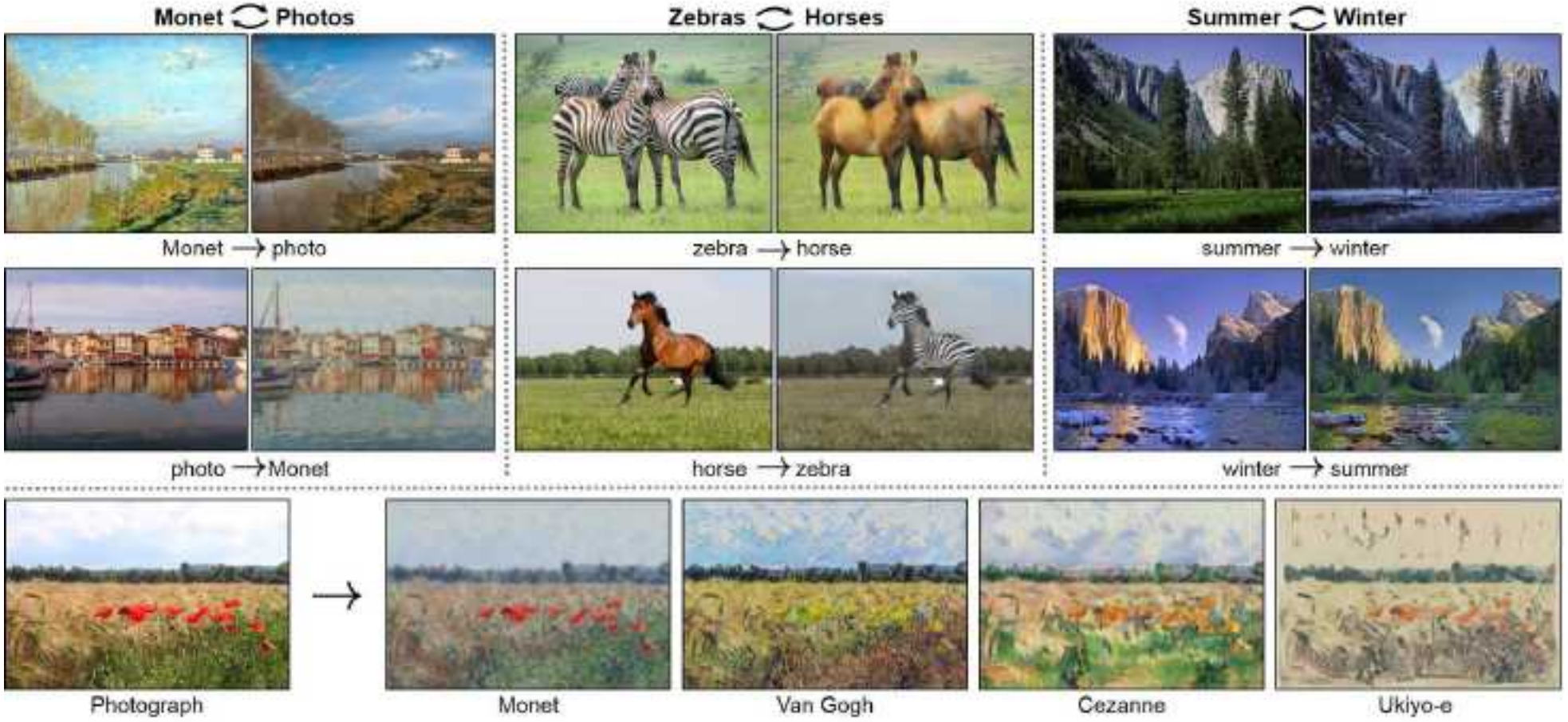
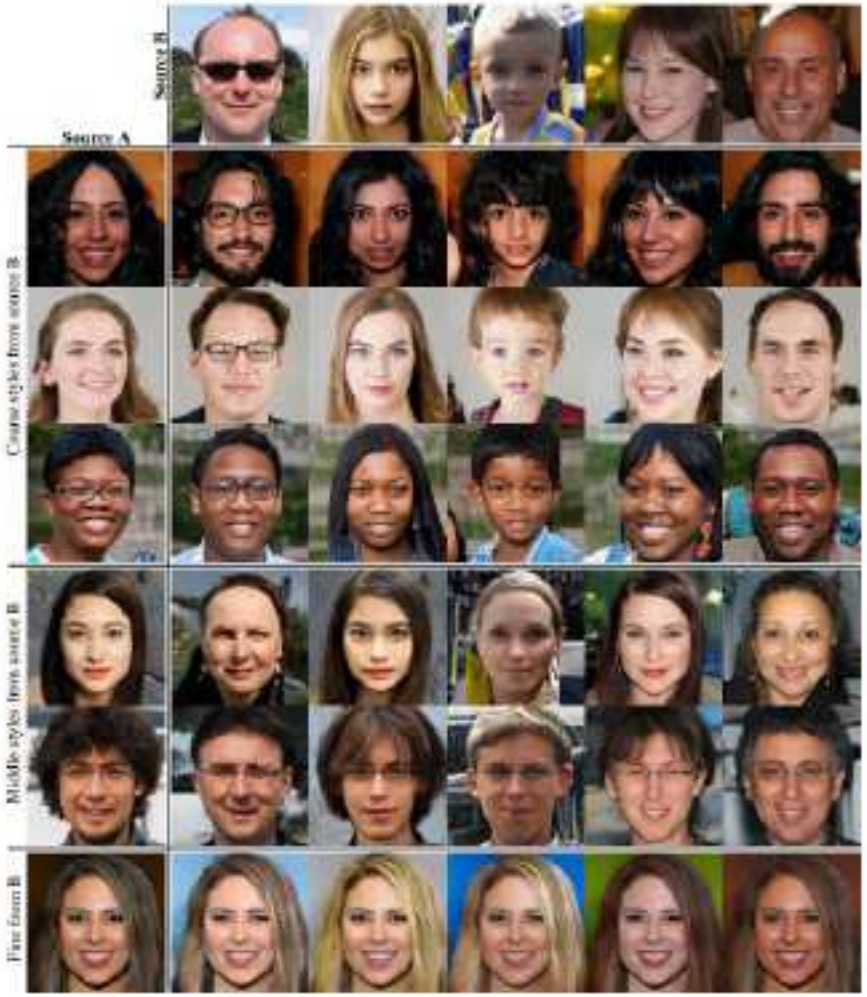


Image generation



朱茵→杨幂

A Style-Based Generator Architecture for Generative Adversarial Networks

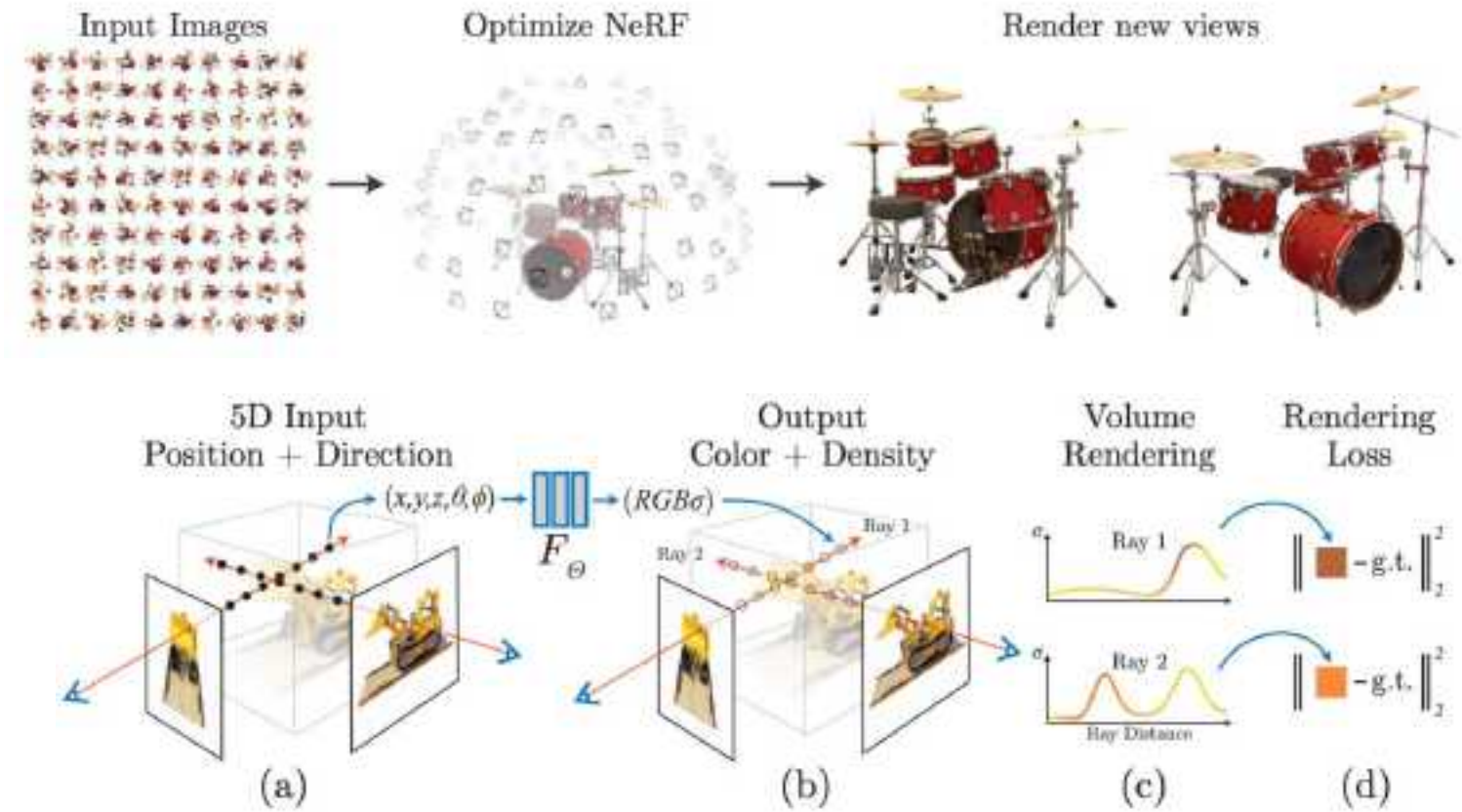
Tero Karras
NVIDIA
tkarras@nvidia.com

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



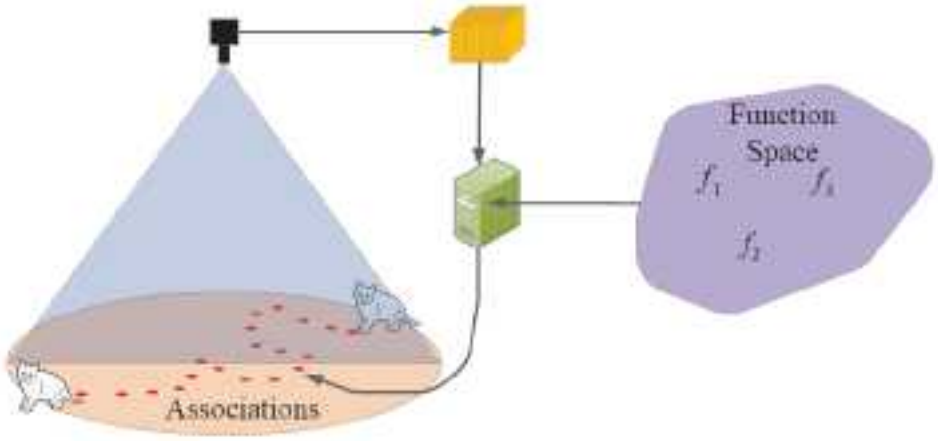
Neural Radiance Field (NeRF)

- 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

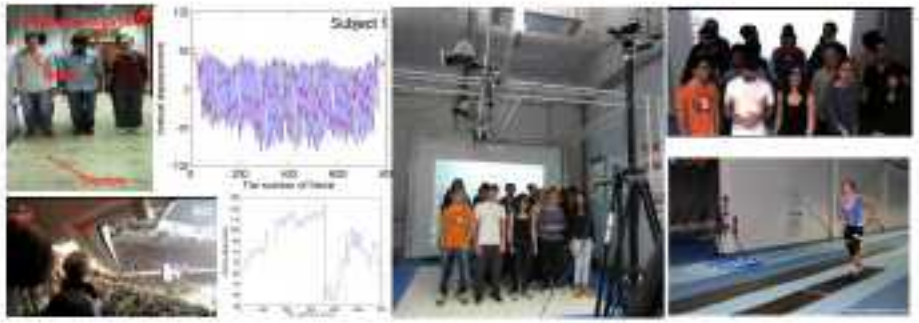


Figure How individuals combine visual and tactile information from other members to synchronise their actions as a group.

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

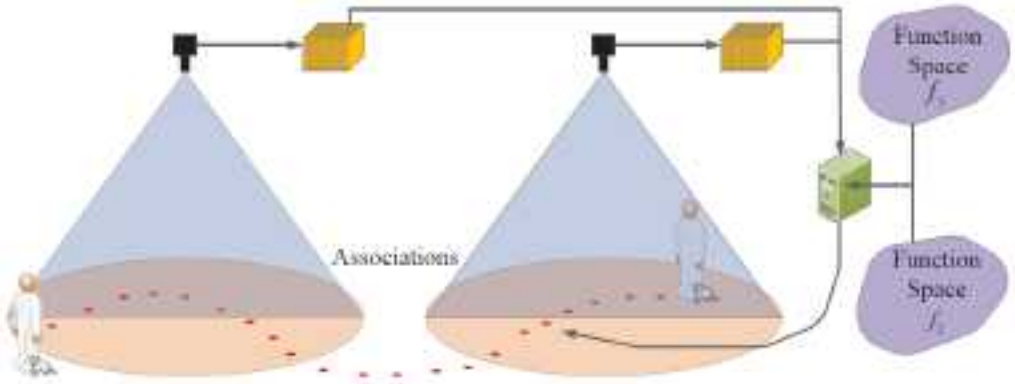
Object tracking

- 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