Lec1 课程引论

AIPKU

人工智能引论实践课 计算机视觉小班

主讲人：刘家瑛

部分讲义来源与致谢

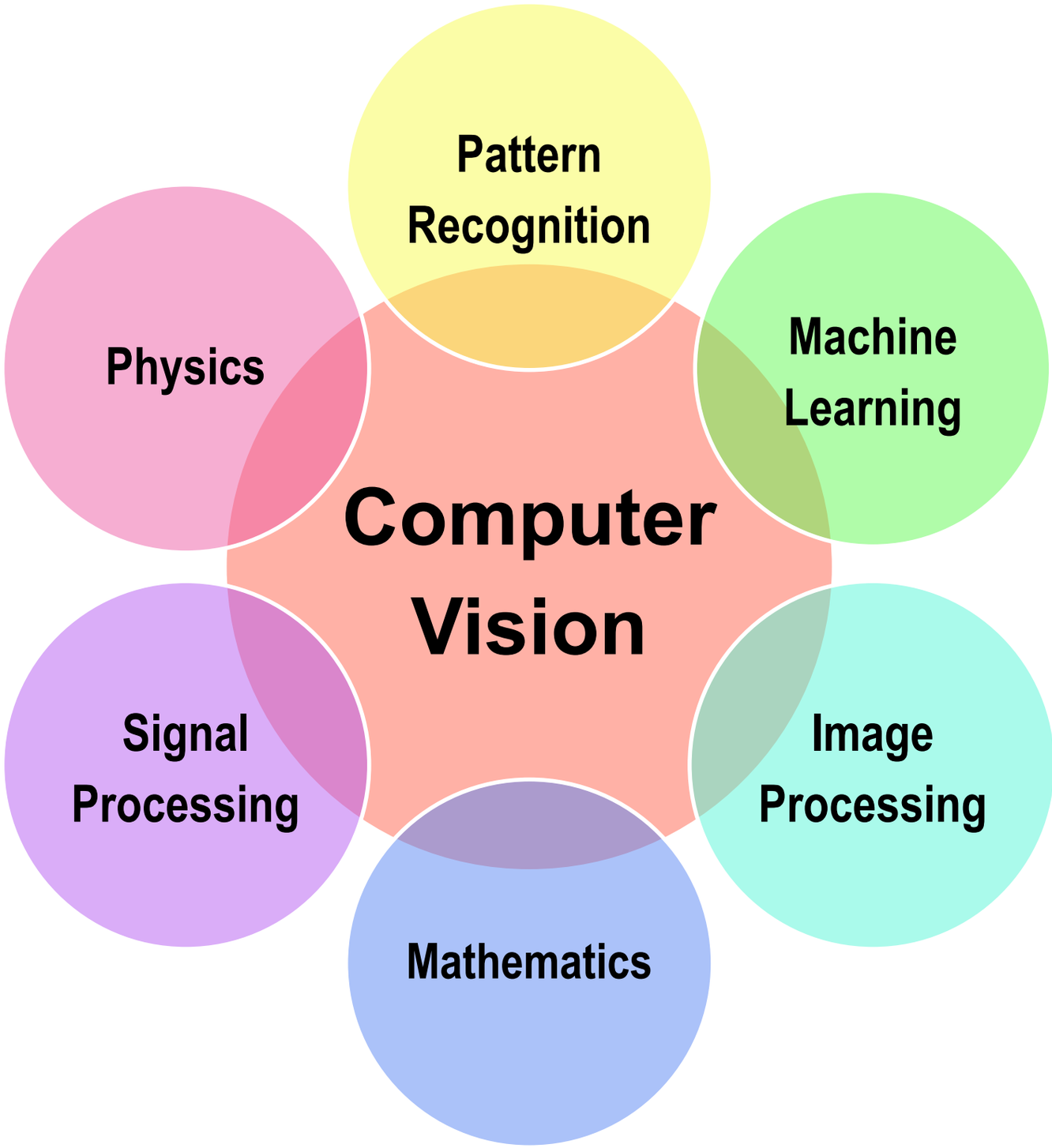
1. Fei-Fei Li, Justin Johnson, Serena Yeung. Stanford University CS231n: Deep Learning for Computer Vision
2. 梅涛. 从感知智能到认知智能的视觉计算
3. Stanford Institute for Human-Centered Artificial Intelligence. AI Index Report 2022
4. Joseph Redmon, Ali Farhadi. YOLOv3: An Incremental Improvement. arXiv 2018
5. Lorenzo Porzi, Samuel Rota Bulò, Aleksander Colovic, Peter Kontschieder. Seamless Scene Segmentation. CVPR 2019
6. Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. ICCV 2017
7. Andrew Brock, Jeff Donahue, Karen Simonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis. ICLR 2019
8. Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Nikolai Yakovenko, Andrew Tao, Jan Kautz, Bryan Catanzaro. Video-to-Video Synthesis. NeurIPS 2018

What is Computer Vision?

Vision is the process of discovering from images what is present in the world, and where it is.

- David Marr, Vision [1982]

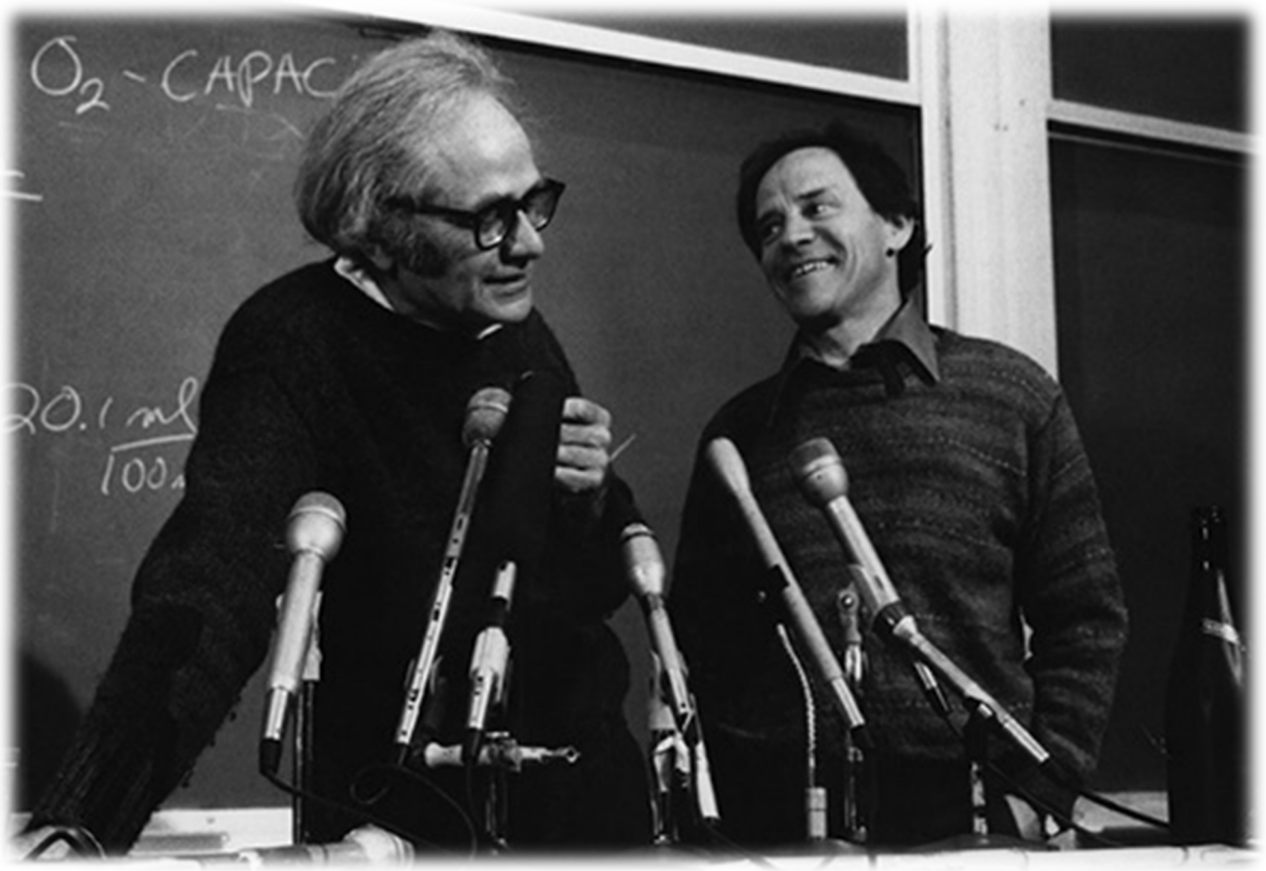
Related Areas



Nobel Prize

- Nobel Prize in Physiology or Medicine 1981
- Awarded to David H. Hubel and Torsten N. Wiesel

"For their discoveries concerning information processing in the visual system."



David Hubel (Left) and Torsten Wiesel (Right) after receiving news of the 1981 Nobel Prize in Medicine. Photo courtesy of Harvard University, Joe Wrinn, photographer.

Origins of Computer Vision

- An MIT undergraduate summer project

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".

Computer Vision: 50 Years of Progress

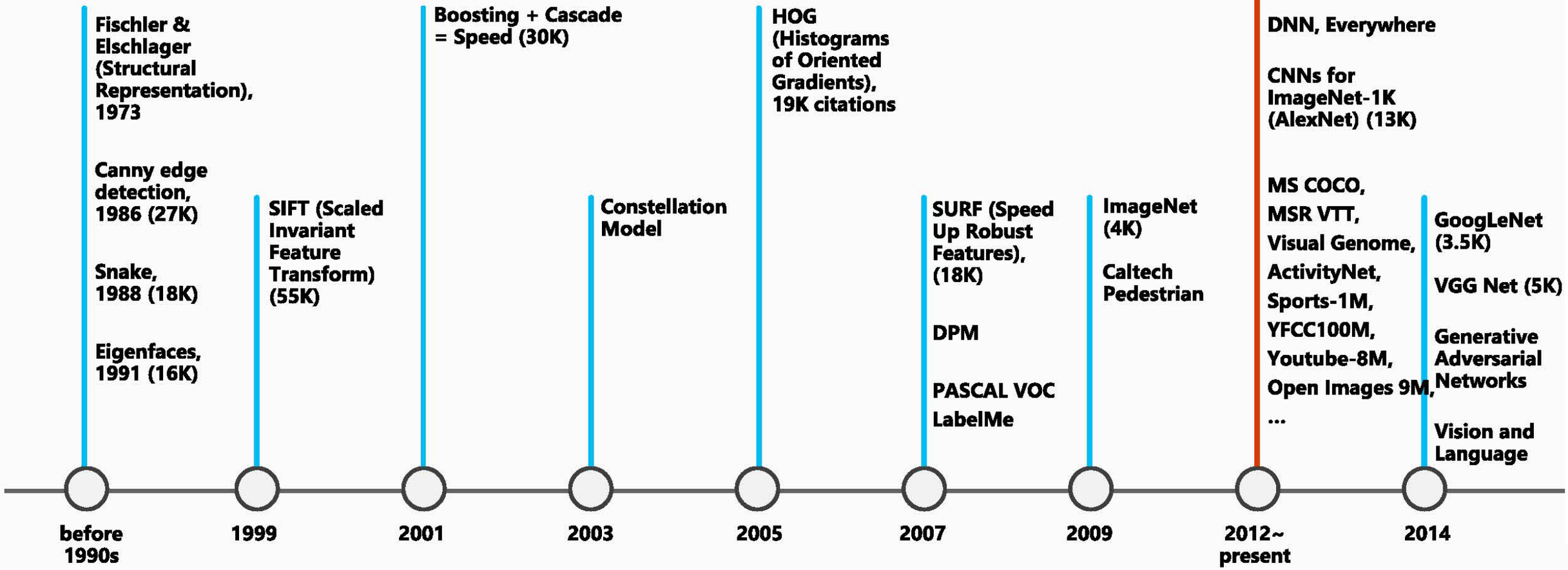
I: image
f: visual feature
y: label/class

feature engineering + model learning

$\mathbf{f} = f(I)$ $\quad y = g(\mathbf{f}, \theta)$

deep learning

$y = g(I, \theta)$



Source: Tao Mei

Computer Vision: 50 Years of Progress

I: image
f: visual feature
y: label/class

feature engineering $\mathbf{f} = f(I)$ **+ model learning** $y = g(\mathbf{f}, \theta)$ → **deep learning** $y = g(I, \theta)$

before 1990s
- Fischler & Elschlager (Structural Representation), 1973
- Canny edge detection, 1986 (27K)
- Snake, 1988 (18K)
- Eigenfaces, 1991 (16K)

1999
- SIFT (Scaled Invariant Feature Transform) (55K)

2001
- Boosting + Cascade = Speed (30K)

2003
- Constellation Model

2005
- HOG (Histograms of Oriented Gradients), 19K citations

2007
- SURF (Speed Up Robust Features), (18K)
- DPM
- PASCAL VOC
- LabelMe

2009
- ImageNet (4K)
- Caltech Pedestrian

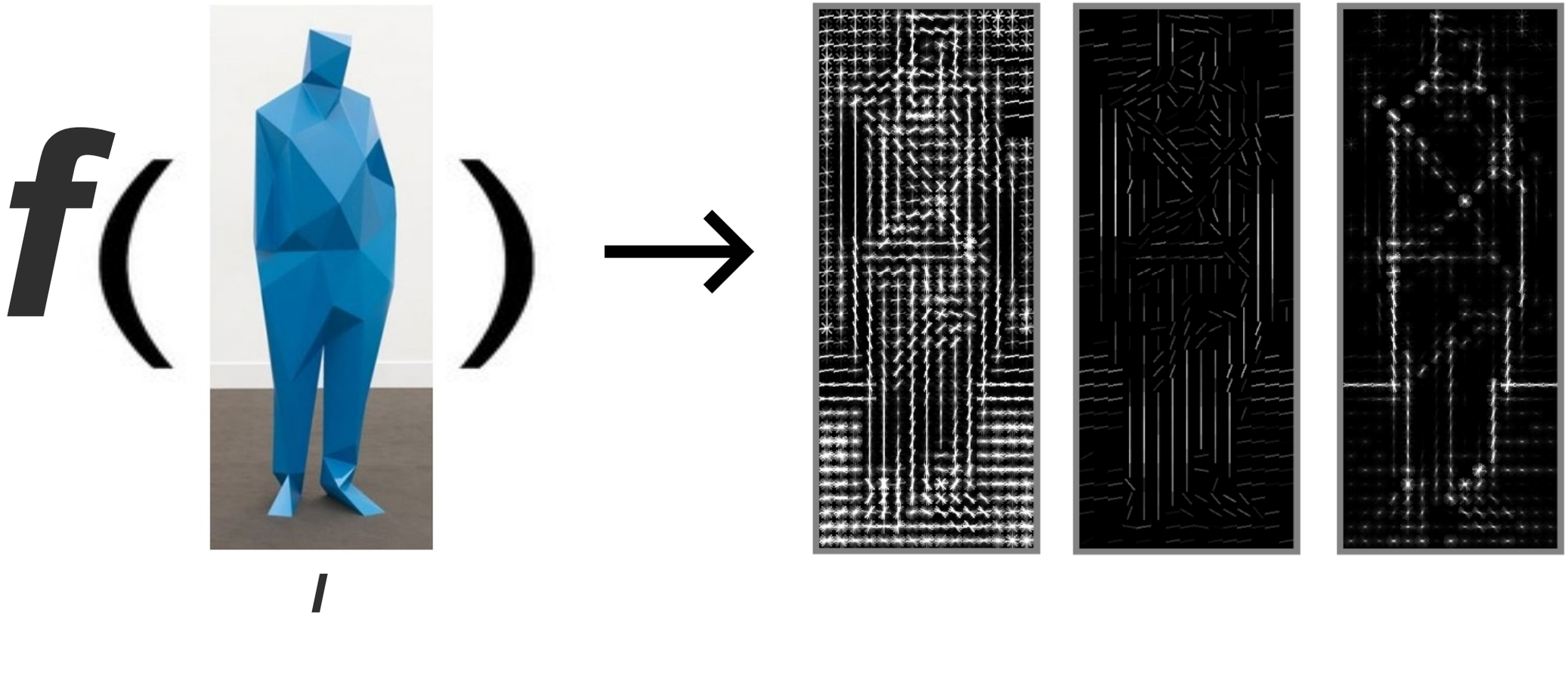
2012~ present
- DNN, Everywhere
- CNNs for ImageNet-1K (AlexNet) (13K)
- MS COCO,
- MSR VTT,
- Visual Genome,
- ActivityNet,
- Sports-1M,
- YFCC100M,
- Youtube-8M,
- Open Images 9M,
- ...

2014
- GoogLeNet (3.5K)
- VGG Net (5K)
- Generative Adversarial Networks
- Vision and Language

Source: Tao Mei

Feature Engineering

- **Feature Detector**
  - A sub-system of a vision system that detects presence or absence of certain features in a visual scene



Computer Vision: 50 Years of Progress

I: image
f: visual feature
y: label/class

feature engineering + model learning
$\mathbf{f} = f(I)$ $y = g(\mathbf{f}, \theta)$

deep learning
$y = g(I, \theta)$

before 1990s
- Fischler & Elschlager (Structural Representation), 1973
- Canny edge detection, 1986 (27K)
- Snake, 1988 (18K)
- Eigenfaces, 1991 (16K)

1999
- SIFT (Scaled Invariant Feature Transform) (55K)

2001
- Boosting + Cascade = Speed (30K)

2003
- Constellation Model

2005
- HOG (Histograms of Oriented Gradients), 19K citations

2007
- SURF (Speed Up Robust Features), (18K)
- DPM
- PASCAL VOC
- LabelMe

2009
- ImageNet (4K)
- Caltech Pedestrian

2012~ present
- DNN, Everywhere
- CNNs for ImageNet-1K (AlexNet) (13K)
- MS COCO,
- MSR VTT,
- Visual Genome,
- ActivityNet,
- Sports-1M,
- YFCC100M,
- Youtube-8M,
- Open Images 9M,
- ...

2014
- GoogLeNet (3.5K)
- VGG Net (5K)
- Generative Adversarial Networks
- Vision and Language

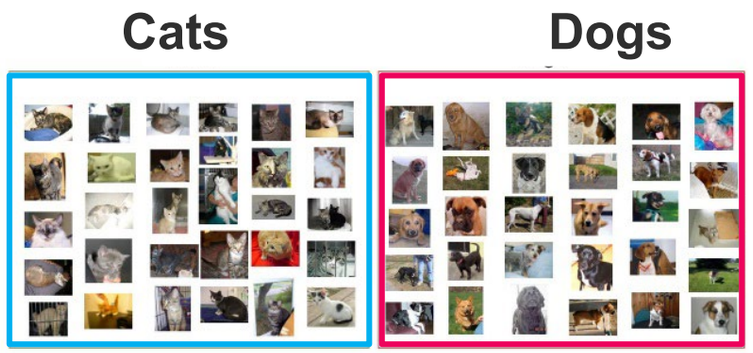
2

Source: Tao Mei

Model Learning

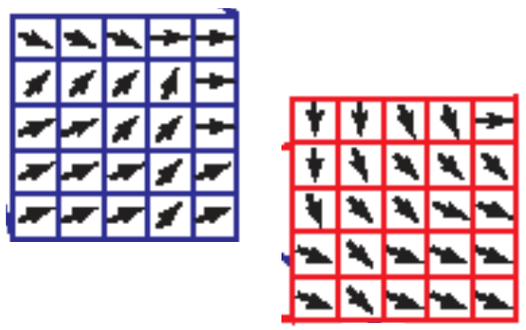
- An example of two-class image classification

Data collection



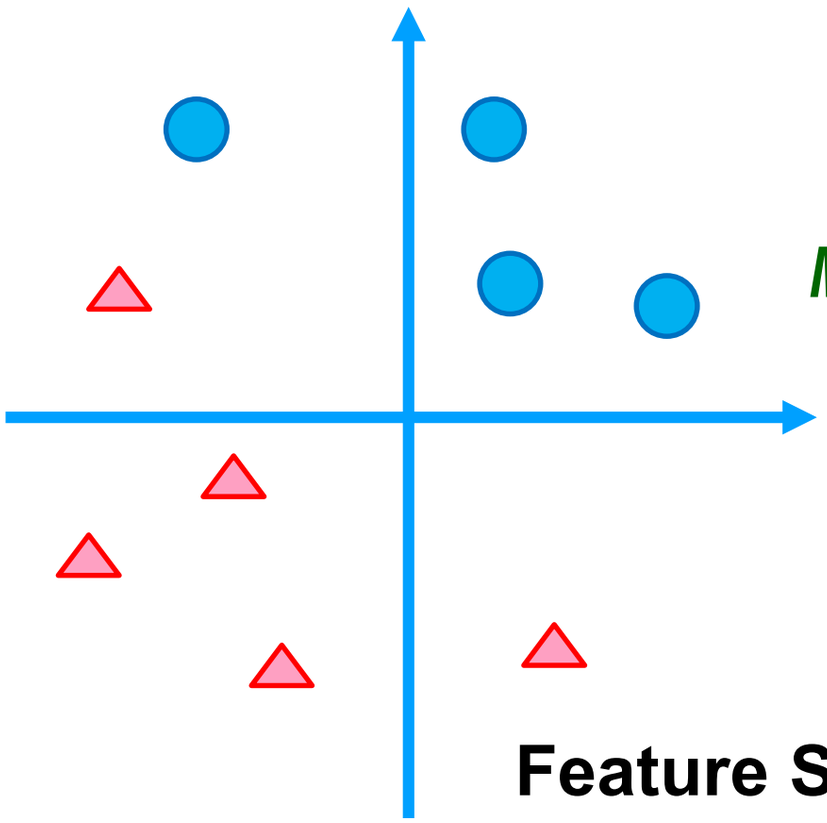
Sample of cats & dogs images

Feature extraction



Visual Features

Model learning

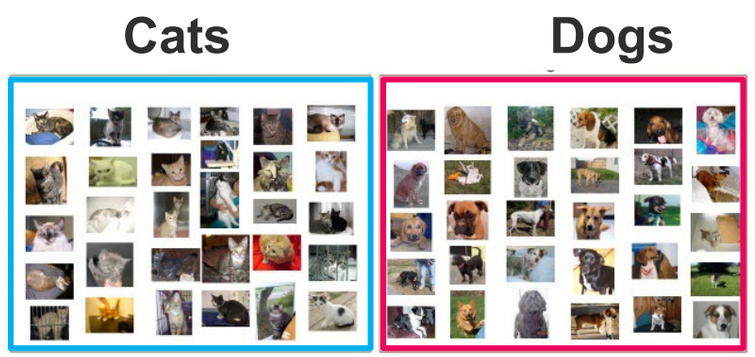


Feature Space

Model Learning

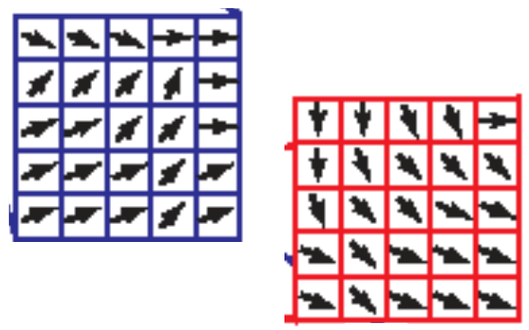
- An example of two-class image classification

Data collection



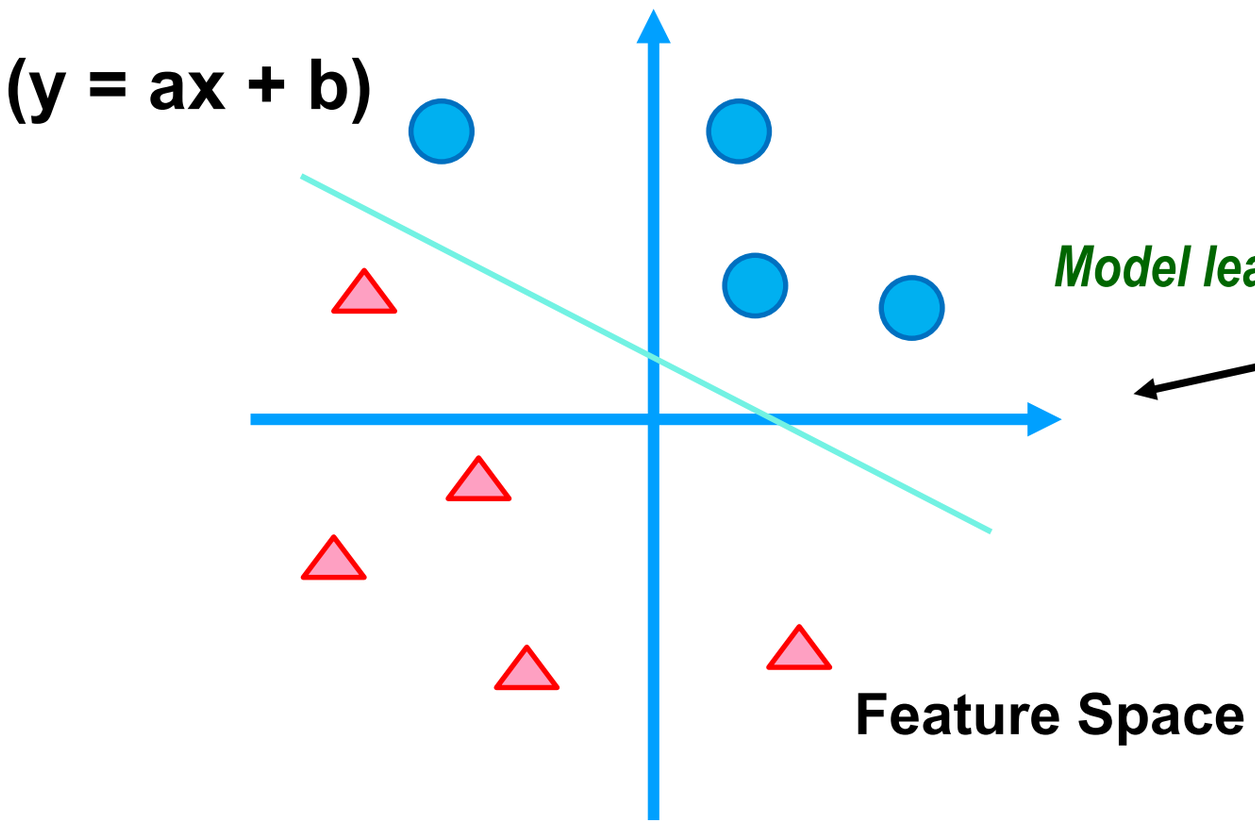
Sample of cats & dogs images

Feature extraction



Visual Features

Model learning

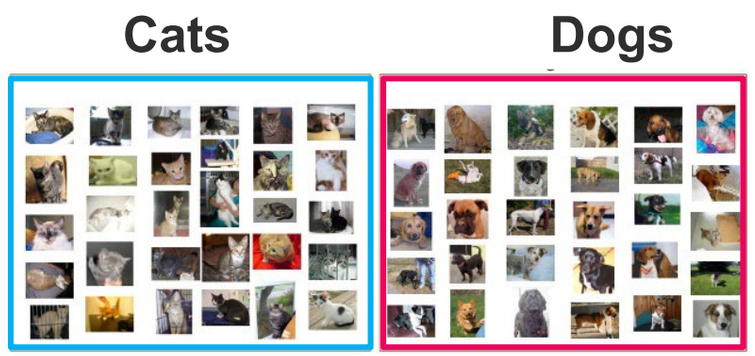


Feature Space

Model Learning

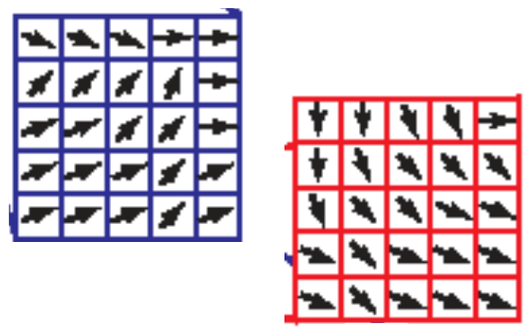
- An example of two-class image classification

Data collection



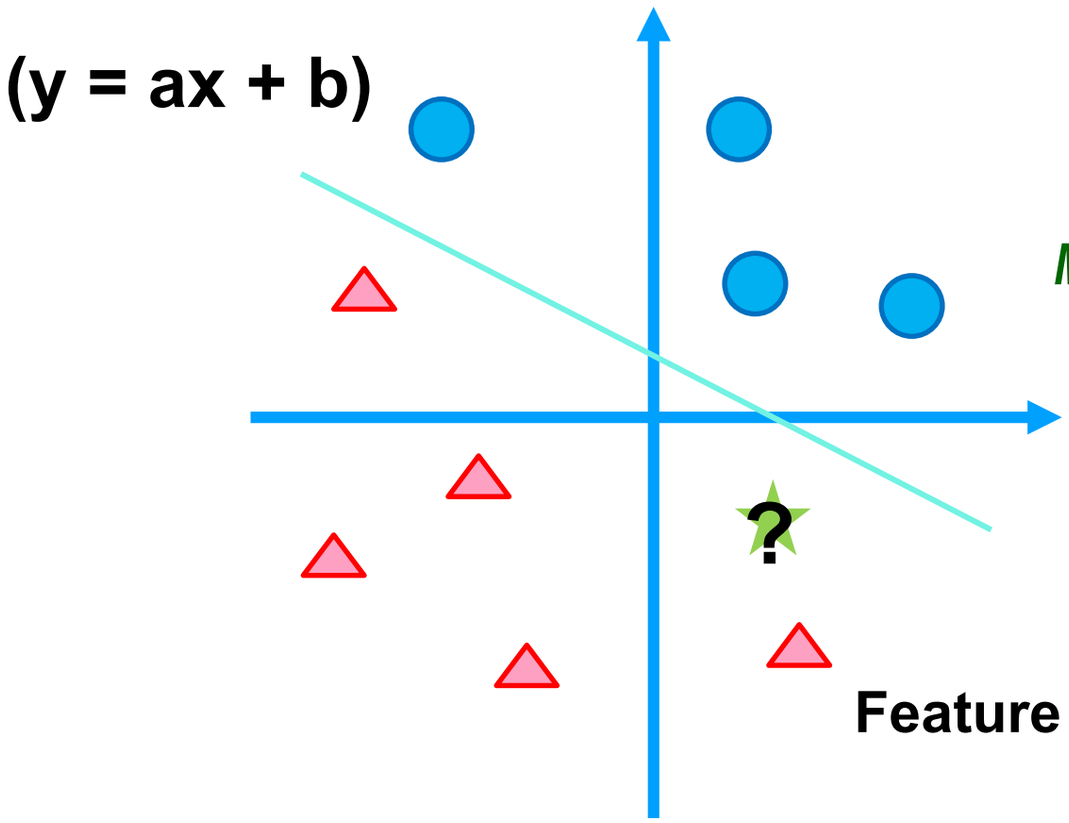
Sample of cats & dogs images

Feature extraction



Visual Features

Model learning

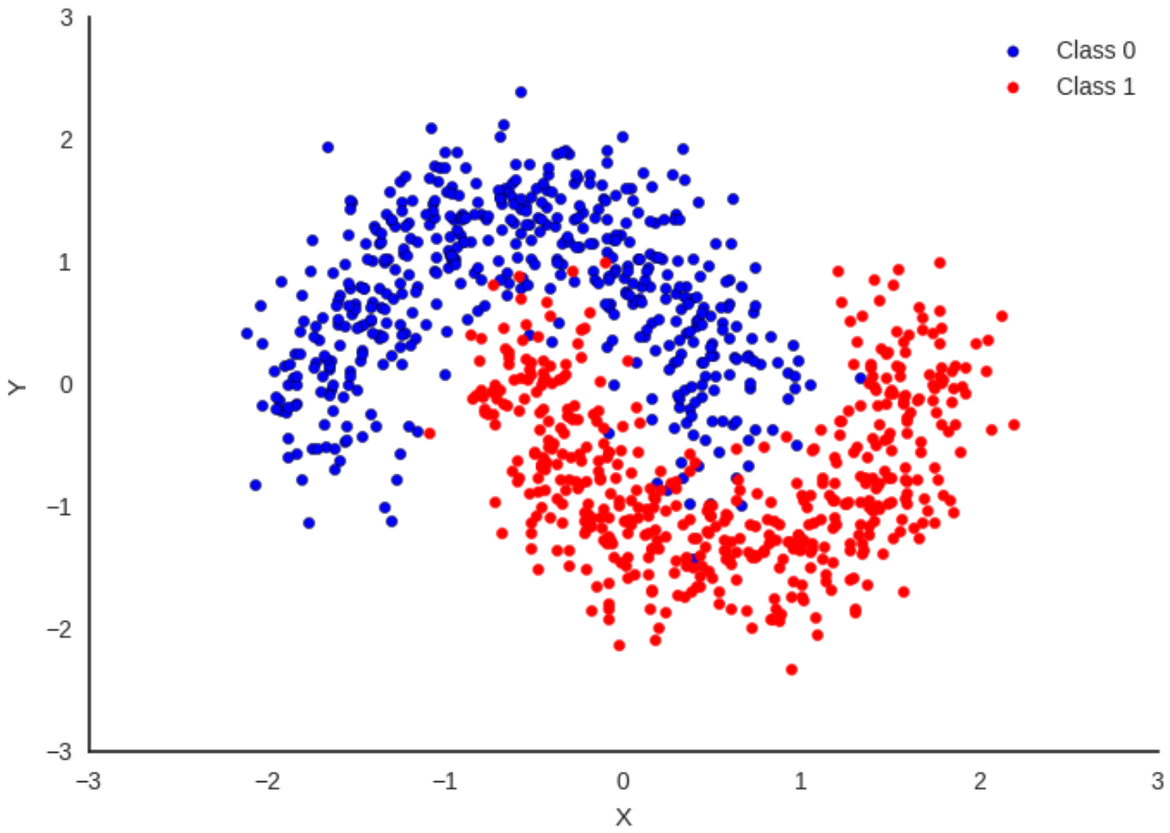


Feature Space

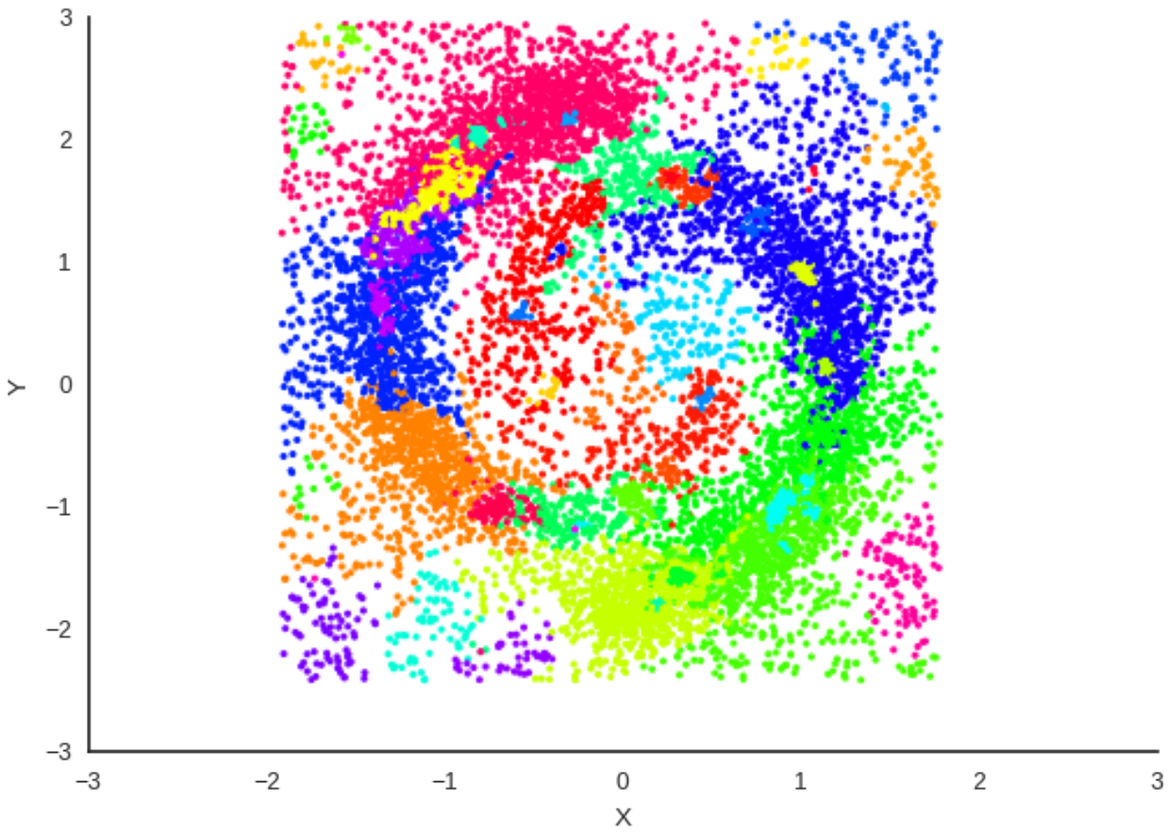
Data from Real-world

- Data from Real-world Problems → Complex Structure

2-class data



***N*-class data**



High Diversity of Images

- Inter-Class Variability

High Diversity of Images

- Intra-class Variability

Class I

Class II

Computer Vision: 50 Years of Progress

I: image
f: visual feature
y: label/class

feature engineering + model learning

$\mathbf{f} = f(I)$ $\quad y = g(\mathbf{f}, \theta)$

deep learning

$y = g(I, \theta)$

before 1990s
- Fischler & Elschlager (Structural Representation), 1973
- Canny edge detection, 1986 (27K)
- Snake, 1988 (18K)
- Eigenfaces, 1991 (16K)

1999
- SIFT (Scaled Invariant Feature Transform) (55K)

2001
- Boosting + Cascade = Speed (30K)

2003
- Constellation Model

2005
- HOG (Histograms of Oriented Gradients), 19K citations

2007
- SURF (Speed Up Robust Features), (18K)
- DPM
- PASCAL VOC
- LabelMe

2009
- ImageNet (4K)
- Caltech Pedestrian

2012~ present
- DNN, Everywhere
- CNNs for ImageNet-1K (AlexNet) (13K)
- MS COCO,
- MSR VTT,
- Visual Genome,
- ActivityNet,
- Sports-1M,
- YFCC100M,
- Youtube-8M,
- Open Images 9M,
- ...

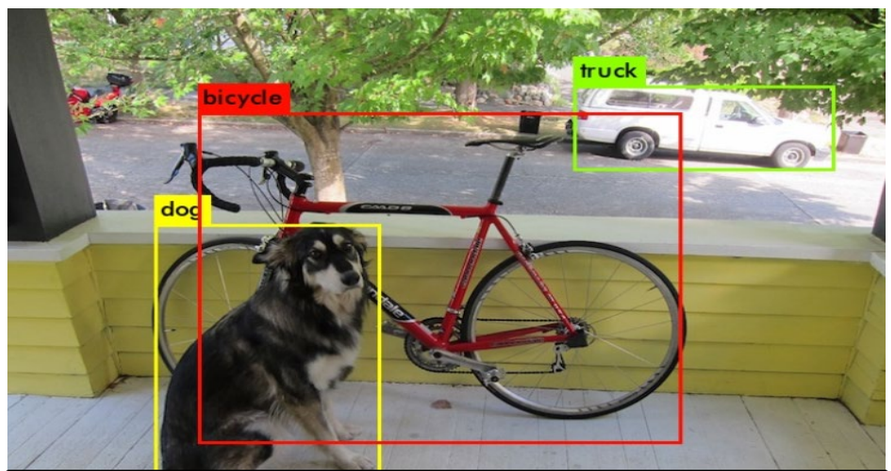
2014
- GoogLeNet (3.5K)
- VGG Net (5K)
- Generative Adversarial Networks
- Vision and Language

Source: Tao Mei

计算机视觉·总览

- **判别问题:** 让计算机认识世界

图像识别



物体检测

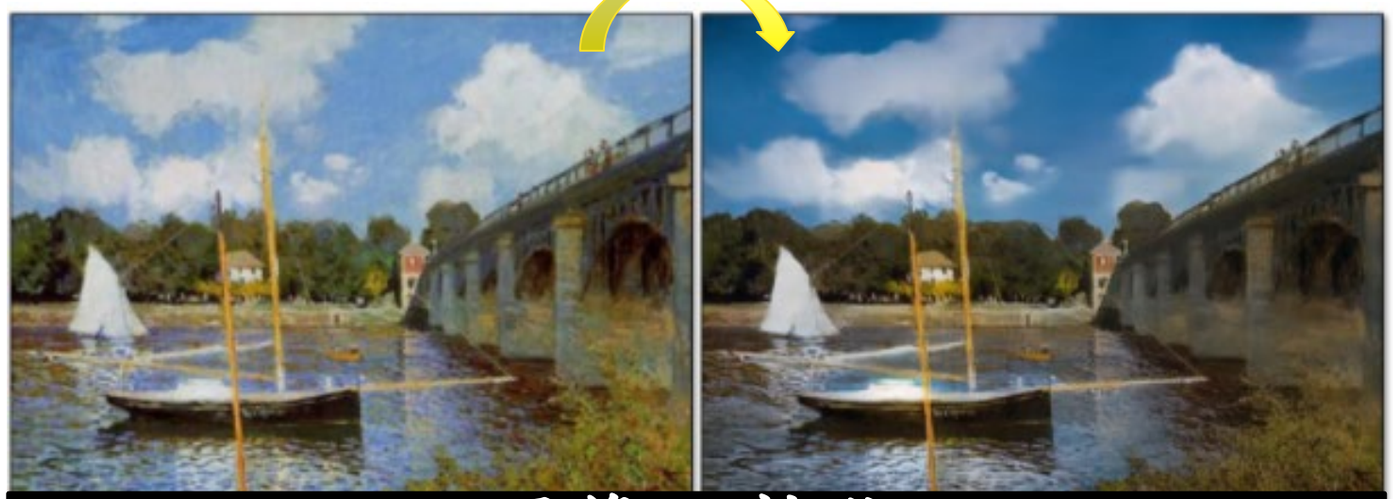
语义分割

...

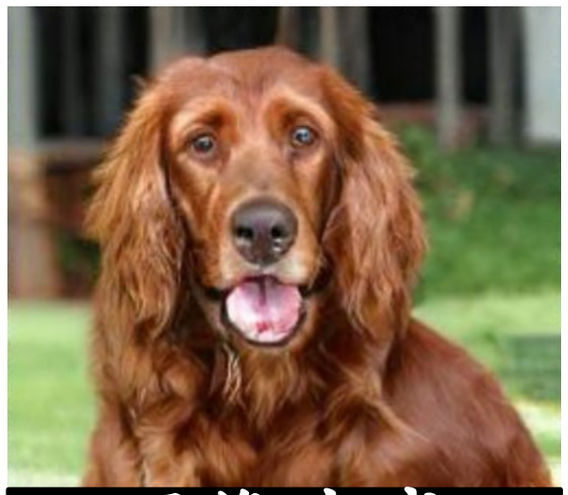
YOLOv3: An Incremental Improvement, arXiv 2018.
Seamless Scene Segmentation, CVPR 2019.

计算机视觉·总览

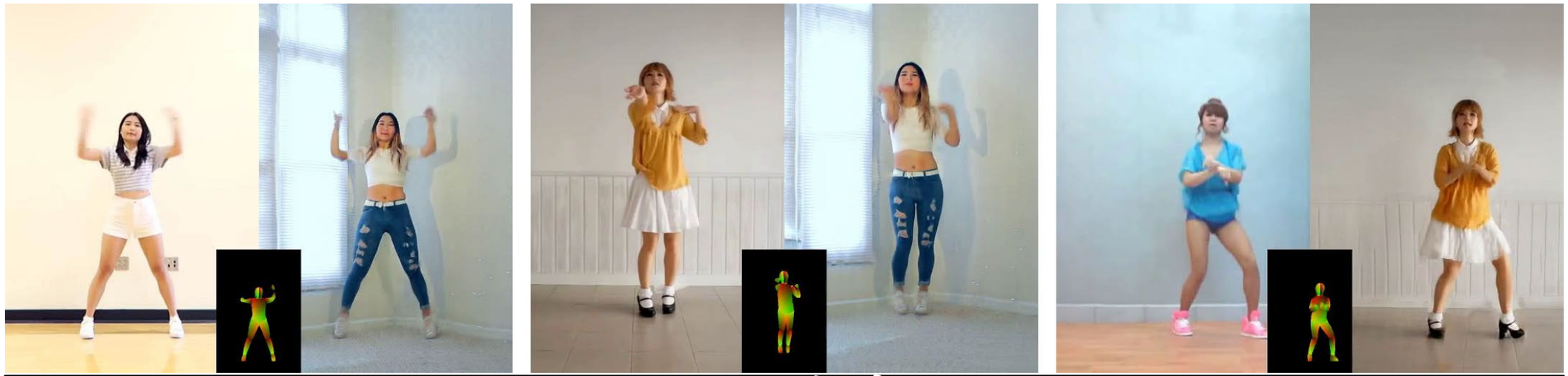
- **生成问题:** 让计算机理解世界



图像风格化



图像生成



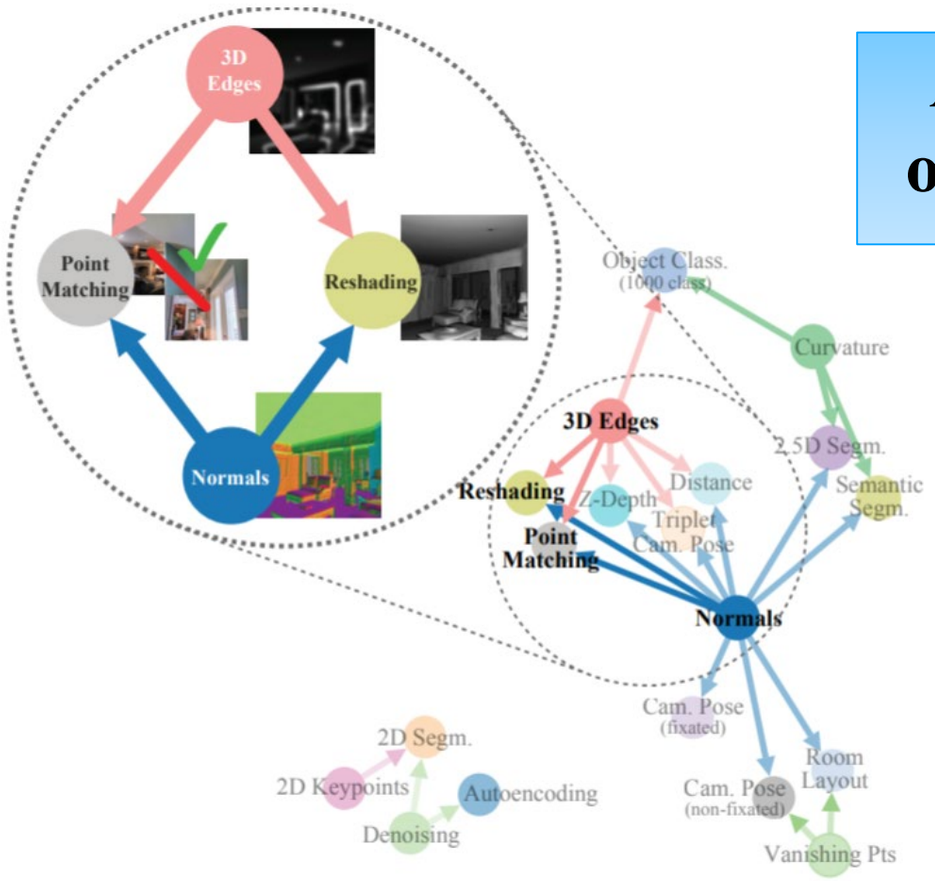
视频生成

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, ICCV 2017.
Large Scale GAN Training for High Fidelity Natural Image Synthesis, ICLR 2019.
Video-to-Video Synthesis NeurIPS, 2018.

计算机视觉 · 总览

■ 他山之石，可以攻玉

- 任务交叉



Taskonomy: Disentangling Task Transfer Learning, CVPR 2018.

- 领域交叉

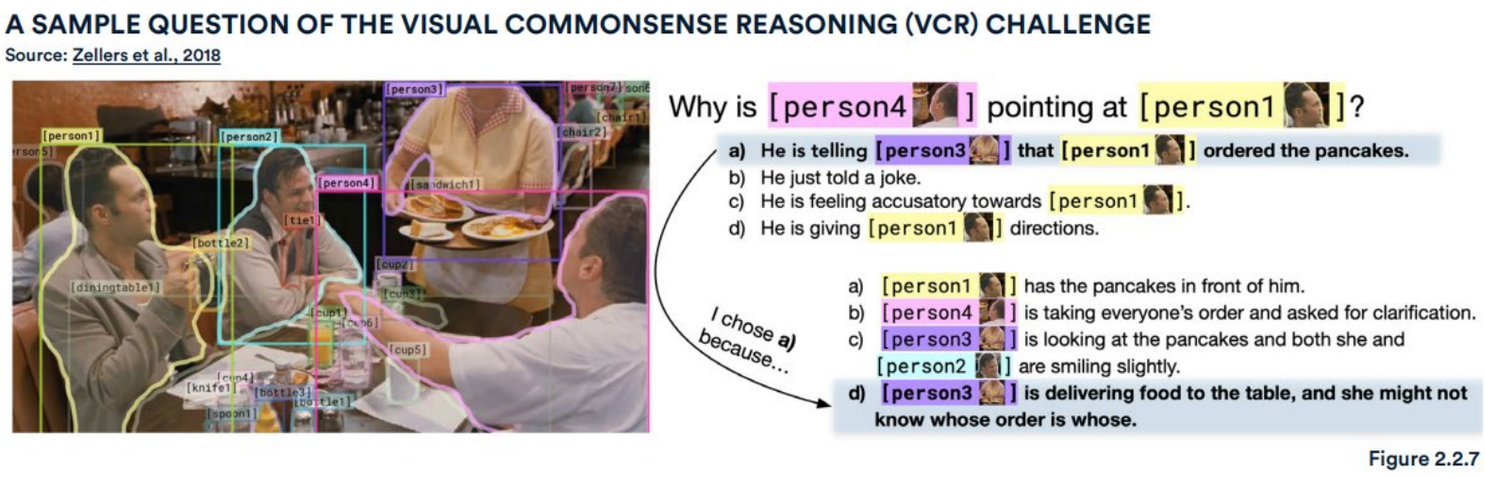
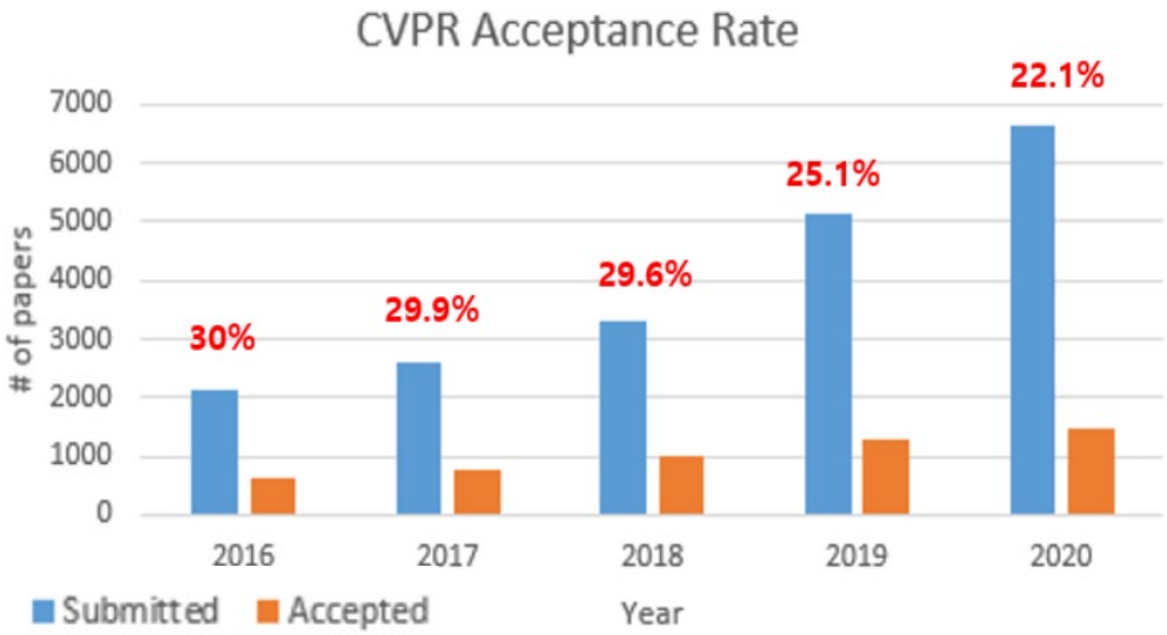


Figure 2.2.7

计算机视觉 · 总览

国际顶级会议: CVPR / ICCV / ECCV

- IEEE/CVF Conference on Computer Vision and Pattern Recognition
- 每年召开, 录取率20% - 30%, 口头报告录用率约为 5%
- CVPR 2021投稿量: 7015, 录取量: 1663, 录取率: 23.7%
- 投稿DDL: 每年11月中
- Rebuttal: 次年1月 - 2月
- 录用通知: 次年2月 - 3月



CVPR 2021

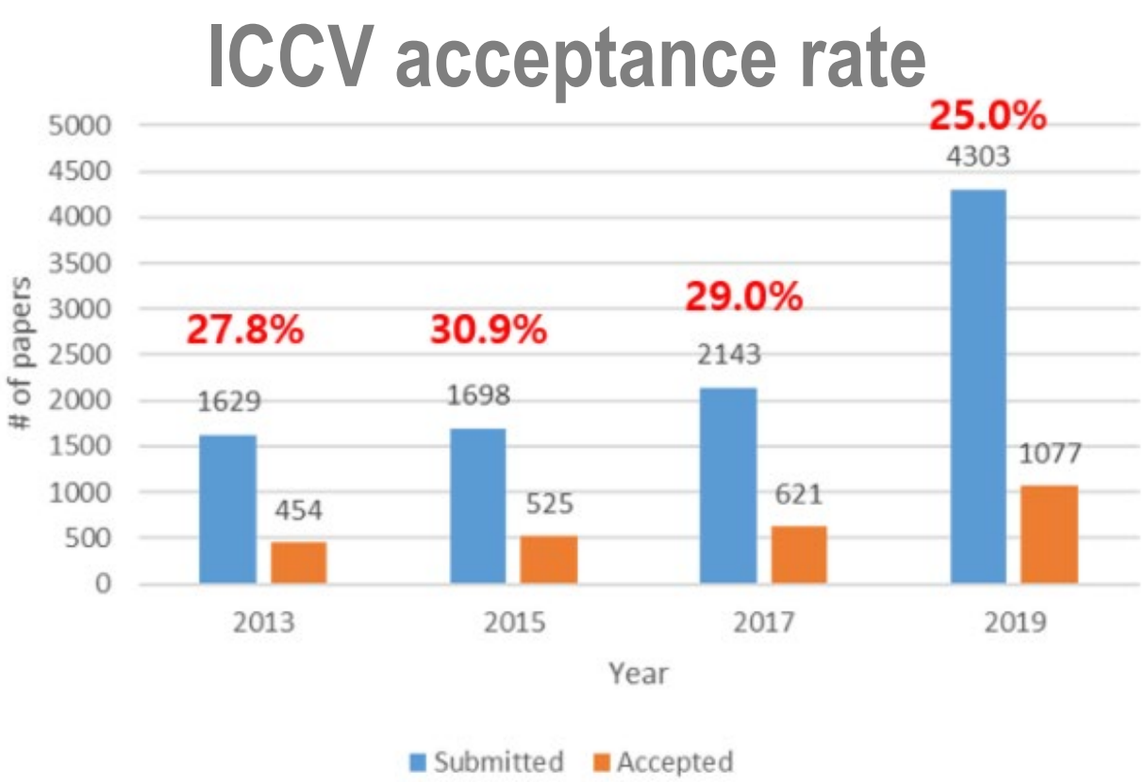
CVPR 2020

https://github.com/hoya012/CVPR-2020-Paper-Statistics

计算机视觉 · 总览

国际顶级会议: CVPR / ICCV / ECCV

- IEEE/CVF International Conference on Computer Vision
- 每两年 (奇数年) 召开, 录取率20% - 30%, 口头论文率约为2% - 3%
- 投稿DDL: 当年3月左右 (ICCV 2021现在审稿阶段)
- Rebuttal: 当年6月左右
- 录用通知: 当年7月左右

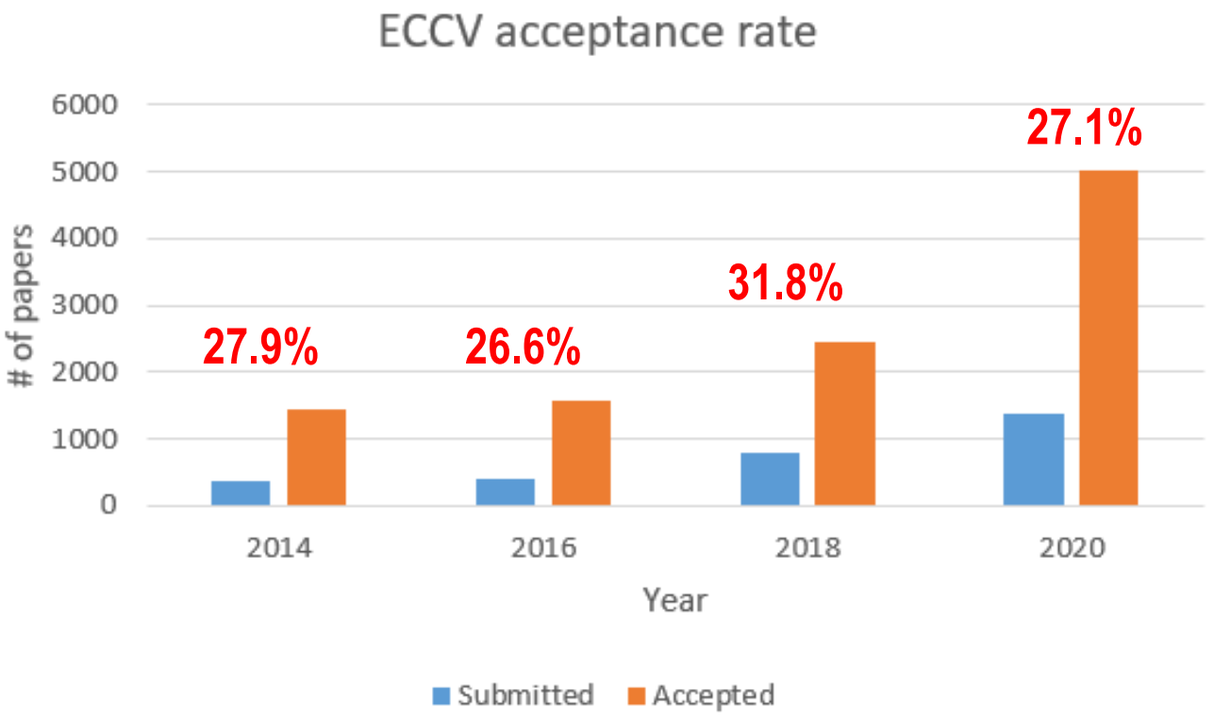


https://github.com/hoya012/ICCV-2019-Paper-Statistics

计算机视觉 · 总览

国际顶级会议: CVPR / ICCV / ECCV

- European Conference on Computer Vision
- 每两年 (偶数年) 召开, 录取率20% - 30%, 口头论文率约为2% - 3%
- 投稿DDL: 当年3月左右
- Rebuttal: 当年6月左右
- 录用通知: 当年7月左右



https://github.com/lixin4ever/Conference-Acceptance-Rate

Artificial Intelligence Index Report

Artificial Intelligence
Index Report 2022

Table of Contents

https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf

Vision – Object Recognition

IMAGENET CHALLENGE: TOP-1 ACCURACY

Source: Papers with Code, 2021; arXiv, 2021 | Chart: 2022 AI Index Report

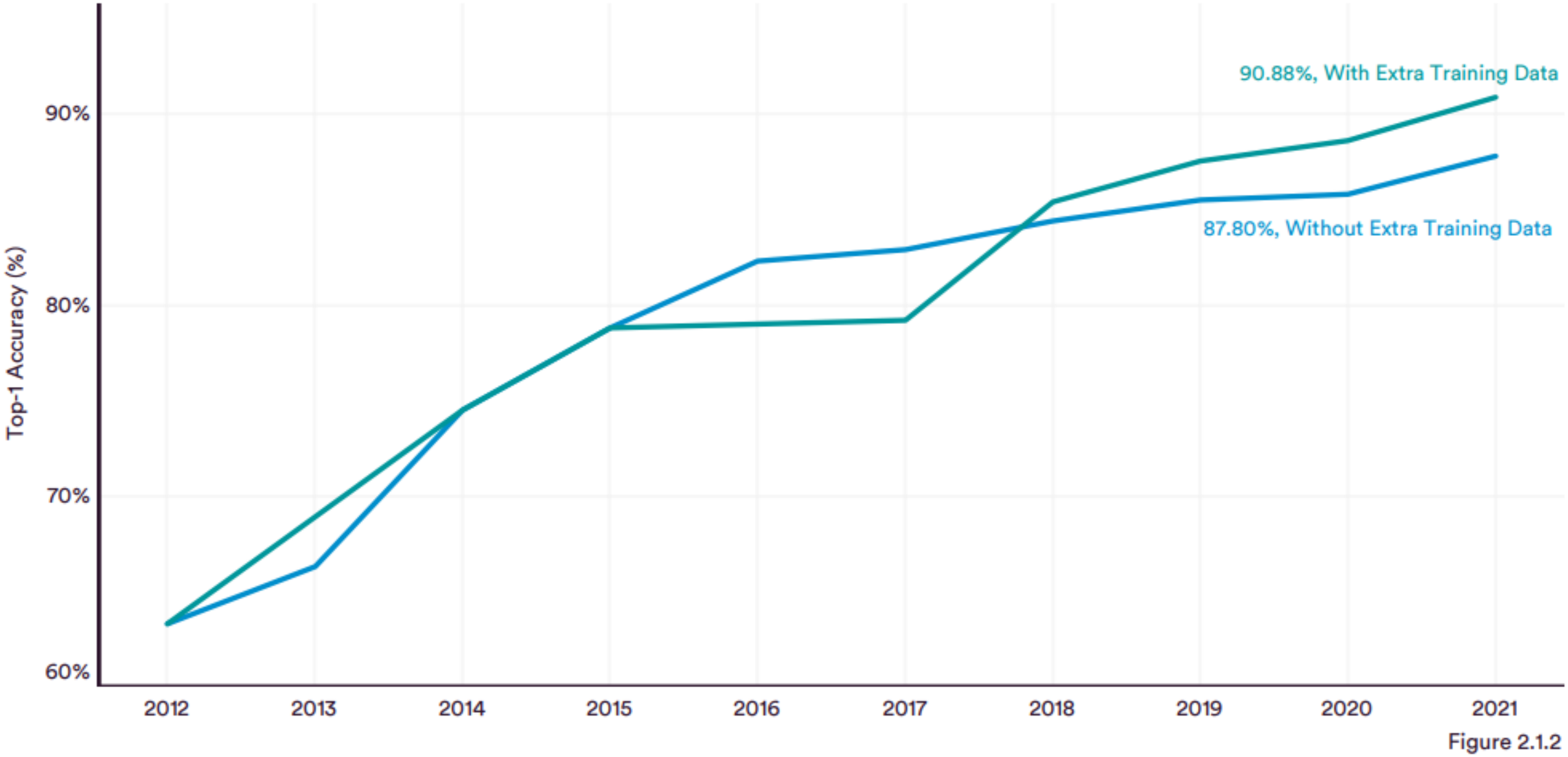


Figure 2.1.2

Vision – Object Recognition

IMAGENET CHALLENGE: TOP-5 ACCURACY

Source: Papers with Code, 2021; arXiv, 2021 | Chart: 2022 AI Index Report

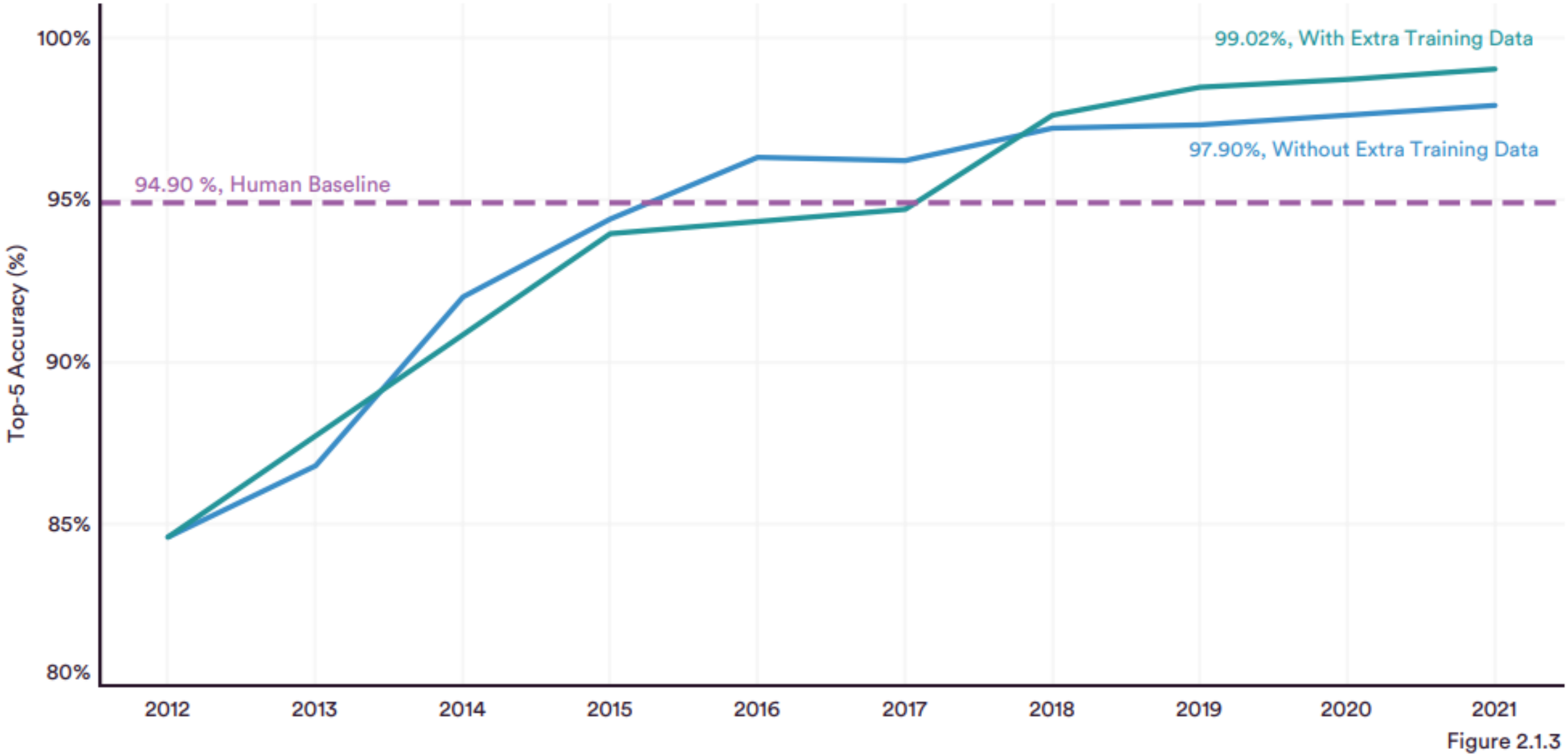
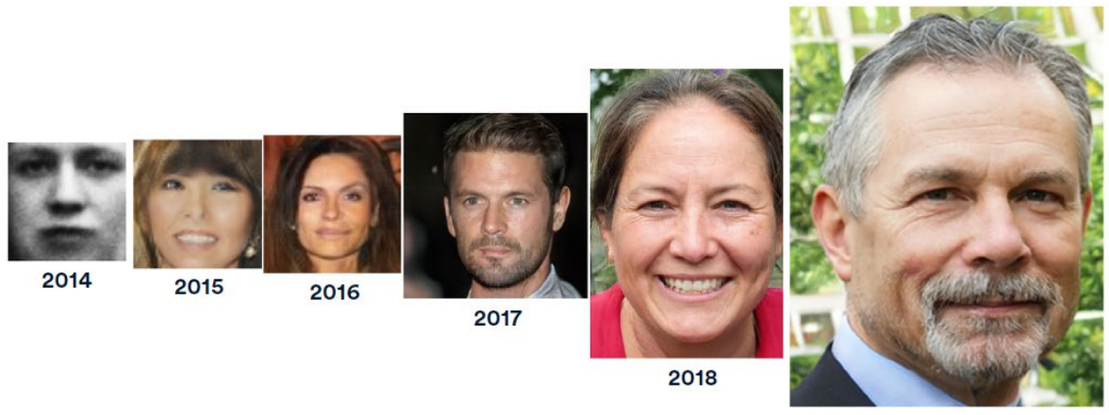


Figure 2.1.3

Vision – Face Generation

GAN PROGRESS ON FACE GENERATION

Source: Goodfellow et al., 2014; Radford et al., 2016; Liu & Tuzel, 2016; Karras et al., 2018; Karras et al., 2019; Goodfellow, 2019; Karras et al., 2020; AI Index, 2021



GAN PROGRESS ON FACE GENERATION

Source: Goodfellow et al., 2014; Radford et al., 2016; Liu & Tuzel, 2016; Karras et al., 2018; Karras et al., 2019; Goodfellow, 2019; Karras et al., 2020; AI Index, 2021; Vahdat et al., 2021

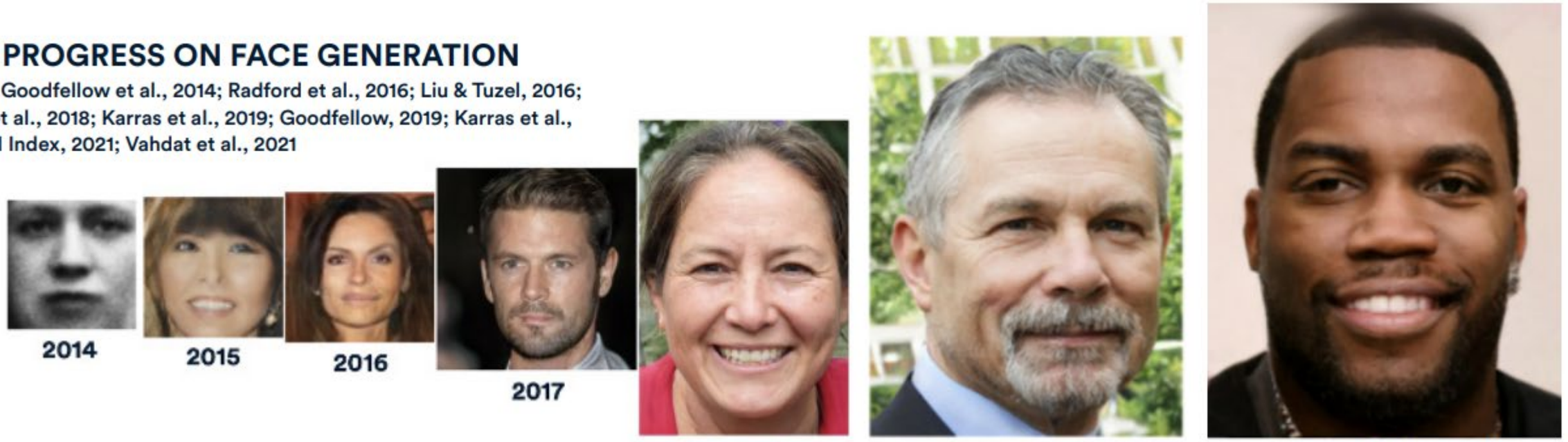
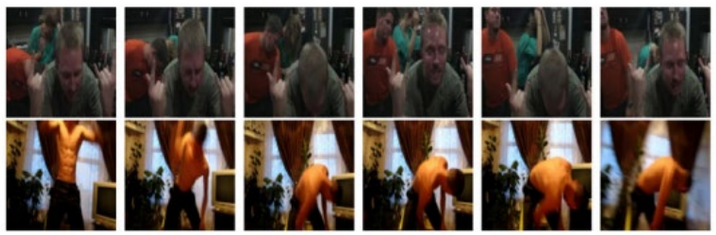


Figure 2.1.4

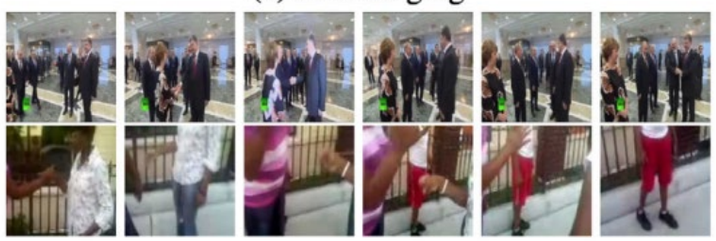
Vision – Action Analytics

EXAMPLE CLASSES FROM THE KINETICS DATASET

Source: Kay et al., 2017



(a) headbanging



(c) shaking hands

(b) stretching leg

(d) tickling

KINETICS-400, KINETICS-600, KINETICS-700: TOP-1 ACCURACY

Source: Papers with Code, 2021; arXiv, 2021 | Chart: 2022 AI Index Report