Introduction to Computer Vision

24/09/2019

Top row, left to right:
Image by Roger H Goun is licensed under CC BY 2.0
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain
Image is CC0 1.0 public domain

Middle row, left to right
Image by BGPHP Conference is licensed under CC BY 2.0; changes made
Image is CC0 1.0 public domain
Image by NASA is licensed under CC BY 2.0
Image is CC0 1.0 public domain

Bottom row, left to right
Image is CC0 1.0 public domain
Image by Derek Keats is licensed under CC BY 2.0; changes made
Image is public domain
Image is licensed under CC-BY 2.0; changes made

What is Computer Vision?

“What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking.”

– David Marr, Vision (1982)

“What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking.”

– David Marr, Vision (1982)

What is (computer) vision?

Image (or video)

Sensing device

Interpreting device

Interpretations

garden, spring, bridge, water, trees, flower, green, etc.

Measurement 🤝 Perception

Measurement

Pollefeys et al.

Goesele et al.

Perception

Slide credit: Kristen Grauman

Biology
Psychology
Physics
Computer Science
Engineering
Mathematics
Computer Vision
Neuroscience
optics
Cognitive sciences
Image processing
Algorithms, theory,...
Speech, NLP
Systems, architecture, ...
Robotics
Information retrieval
Machine learning

Computer Vision Tasks

- Image Classification
- Image Segmentation
- Object Detection
- Object Localization
- Edge detection
- Image denoising
- Optical flow estimation
- Tracking
- Reconstruction
- Semantic Segmentation
- Super-resolution
- Image captioning
- Visual question answering
- Activity recognition

Classification

CAT

No spatial extent

Semantic Segmentation

GRASS, CAT, TREE, SKY

No objects, just pixels

Object Detection

DOG, DOG, CAT

Instance Segmentation

DOG, DOG, CAT

Multiple Object

This image is CC0 public domain

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

1,000 object classes

1,431,167 images

Russakovsky et al. IJCV 2015

Who is wearing glasses?

man

woman

Where is the child sitting?

fridge

arms

Is the umbrella upside down?

yes

no

How many children are in the bed?

2

1

Image Captioning

"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."

Sadegh's:
1. a hot dog bun filled with mustard and ketchup.
2. a hot dog bun filled with fries and pickles.
3. a hotdog with mustard, ketchup and onion on a bun.
4. a plate of food with hot dog bun filled with cheese and vegetables.
5. a hotdog with mustard, ketchup, mustard, and onions on a bun.
6. a white plate topped with a sandwich and a salad.
7. a plate of food with a hot dog, tomatoes and sausage.
8. a plate with a sandwich and french fries on a table.
9. a plate with a sandwich and french fries.

Caption from mode: a hot dog bun filled with fries and pickles.

Microsoft: a hot dog on a bun.

Sadegh's:
1. a person holding a rainbow colored umbrella over their head .
2. a woman holding an umbrella in the rain .
3. a woman holding a colorful umbrella over her head .
4. a woman holding an umbrella in the rain .
5. a woman with a red umbrella is standing on the beach .
6. a woman in a red shirt is holding a blue umbrella .
7. a person holding an umbrella while walking down a street .
8. a woman holding an umbrella while walking on a sidewalk .
9. a little girl with a purple umbrella standing in the rain .

Caption from mode: a young woman holding an umbrella standing on a sidewalk .

Microsoft: a person holding a blue umbrella

- **Sadegh Aliakbarian**, PhD Student, Data61, CSIRO

Why study computer vision?

- https://perfectial.com/blog/computer-vision-applications/

Images and video are everywhere!

Google Image Search | Google Photos | flickr | webshots | picsearch | YouTube Broadcast Yourself

Surveillance and security

Medical and scientific images

Computer Vision is Useful and Interesting

- Autonomous driving
- Robotics vision
- 3D city modelling
- Security
- Sports
- Medical image processing
- Information retrieval
- Face detection
- Image generation
- Style transfer
- Image-to-image translation
- Colorization
- Image completion
- DeepFakes

Image Search

- https://www.macrumors.com/2017/02/16/iphone-3d-laser-scanner-facial-recognition/

https://www.theverge.com/2016/3/21/11255126/apple-new-macbook-pro-with-touch-strip-keys

https://www.theverge.com/2016/3/21/11255126/apple-new-macbook-pro-with-touch-strip-keys

https://www.wired.com/2013/07/cara-placemeter-google-glass/

https://www.wired.com/2013/07/cara-placemeter-google-glass/

3D Maps

Google Maps

Sports

1. Hawk-Eye 2. Goel line technology
3. What will Happen Next?Forecasting Player Moves in Sports Videos
4. https://indatalabs.com/blog/how-does-computer-vision-work

Gaming

Drone

- https://petapixel.com/2017/05/24/dji-unveils-spark-mini-drone-can-control-hand-gestures/

Special Effects

Real vs Fake

Large scale gan training for high fidelity natural image synthesis. ICLR 2019

Image Generation

Karras et al., "Progressive Growing of GANs for Improved Quality, Stability, and Variation", ICLR 2018

Image Style Transfer Using Convolutional Neural Networks. CVPR 2016

Deepfake

https://en.wikipedia.org/wiki/Deepfake

Deepfake

https://giphy.com/explore/deepfakes

Norwegian Bride (est late 1890s)

Image Inpainting

Image Inpainting

Underwater Image Enhancement

Saeed Anwar, Chongyi Li, and Fatih Porikli. "Deep underwater image enhancement." arXiv 2018.

Underwater Image Enhancement

Saeed Anwar, Chongyi Li, and Fatih Porikli. "Deep underwater image enhancement." arXiv 2018.

Outside border images, clockwise, starting from top left:
Image by Pop Culture Geek is licensed under CC BY 2.0; changes made
Image by the US Government is in the public domain
Image by the US Government is in the public domain
Image by Glogger is licensed under CC BY-SA 3.0; changes made
Image by Sylenius is licensed under CC BY 3.0; changes made
Image by US Government is in the public domain

Inside four images, clockwise, starting from top left:
Image is CC0 1.0 public domain
Image by Tucania is licensed under CC BY-SA 3.0; changes made
Image by Intuitive Surgical, Inc. is licensed under CC BY-SA 3.0; changes made
Image by Oyundari Zorigtbaatar is licensed under CC BY-SA 4.0

CVPR 2019 CORPORATE SPONSORS

Diamond Sponsors

Microsoft, MEGVII 旷视, 商汤 sensetime

Platinum Sponsors

amazon, ANDURIL, anyVISION A BETTER TOMORROW, Apple, ARGO, Baidu 百度, ByteDance, COGNEX, cruise, DeepBlue, facebook, Google, HUAWEI, IBM Research, inspur, intel AI, lyft, NAVER, LINE, NVIDIA, pony.ai, Qualcomm, 快手, SAMSUNG, Tencent YouTu Lab, TOYOTA RESEARCH INSTITUTE, Uber, WAYMO, wayfair, XILINX

Gold Sponsors

Adobe, AIBEE, Alibaba Group 阿里巴巴集团, Horizon Robotics, DiDi, Disney Research, HRI Honda Research Institute US, dji, IIAI Inception Institute of Artificial Intelligence, iRobot, JD.COM, Lambda, MALONG 码隆, Micron, Panasonic, PathAI, scale, SUPER ANNOTATE, voyage

Silver Sponsors

ABEJA, algolux, AMAX, appen data with a human touch, aquifi, Aurora, BAOBAB, cādence, CEVA, Cirrascale CLOUD SERVICES, cloudfactory, CVTE Dream Future 视源股份, deepen, 深睿医疗 deepwise, floure eight, FORESIGHT.AI, gumgum, GENIUS PROS 光电智能, HITACHI Inspire the Next, HOVER, 虎牙直播 huya.com, isee, iMerit, JUMIO, KINATRAX, Kitware, Matroid, meitu, MTLab Meitu Imaging & Vision Lab, MVTec Software GmbH, nuimox 纽劢科技, ObjectVideo LABS, Orbital Insight, Playment, PROPHESEE METAVISION FOR MACHINES, SAP, 眼控科技 EMDATA, SIEMENS Ingenuity for life, SIGOPT, SK telecom, SPORT LOGiQ, Tencent AI Lab, TESLA, Velodyne Lidar, Verisk, VinAI, VOXEL51, Weights & Biases

ZILLOW GROUP, zoox