Computer Vision

Introduction

Dr. Mrinmoy Ghorai

Indian Institute of Information Technology

Sri City, Chittoor

Welcome to CV Spring 2021









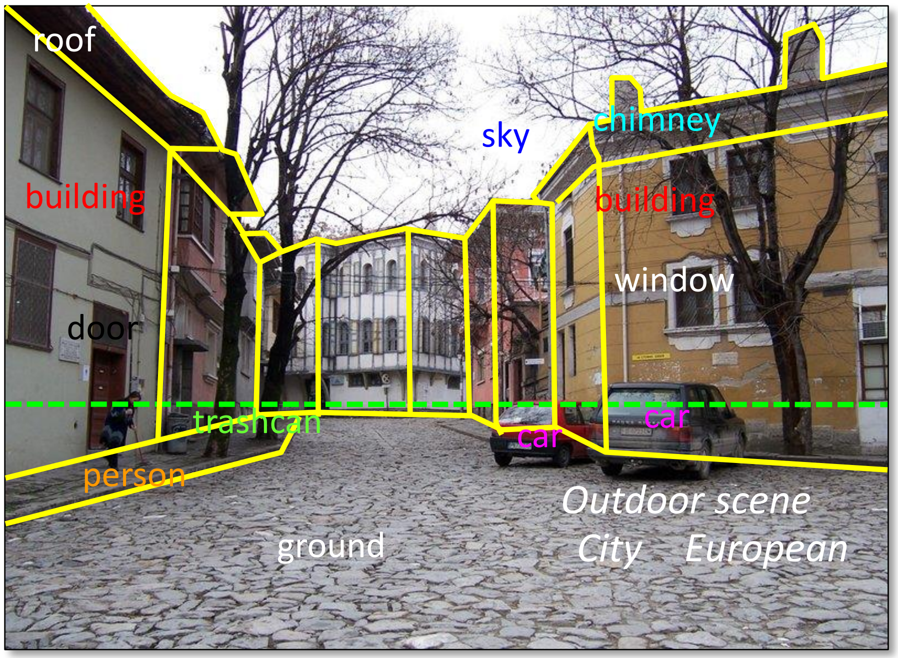


Today's Agenda

- Introduction to computer vision
- Course overview

What is computer vision?

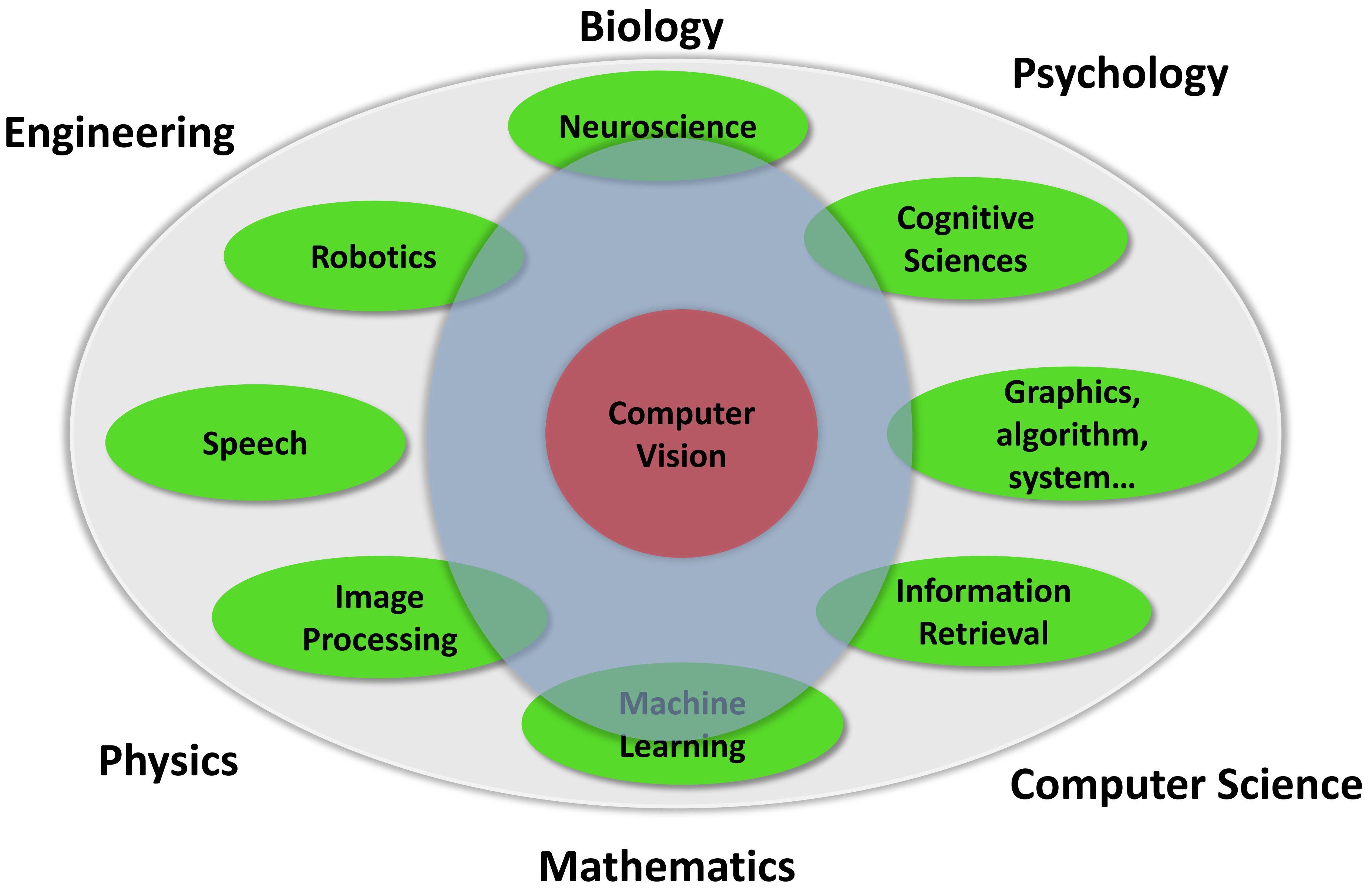
- Make computers understand images and videos.
- What kind of information can be extracted?



Geometric information

Semantic information

What is it related to?



The goal of computer vision

- Every image tells a story

- perceive the “story” behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What event?

The goal of computer vision

- To extract “meaning” from pixels

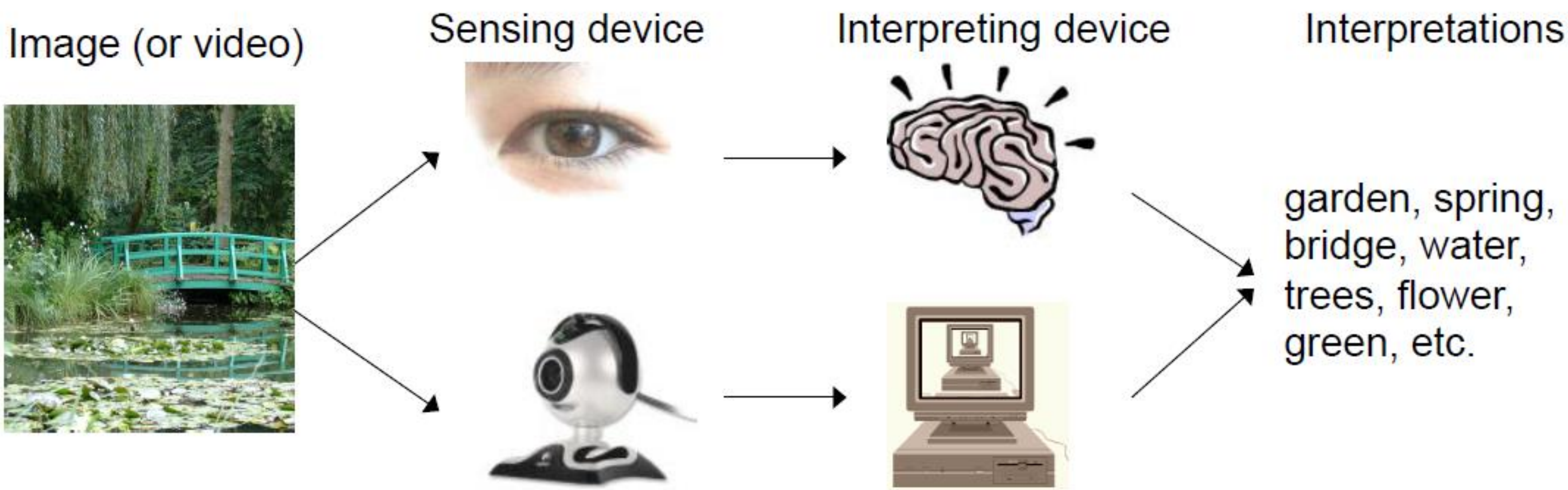
What we see

0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: **S Narasimhan**

Human and Computer Vision



Slide adapted from **Fei-Fei Li**

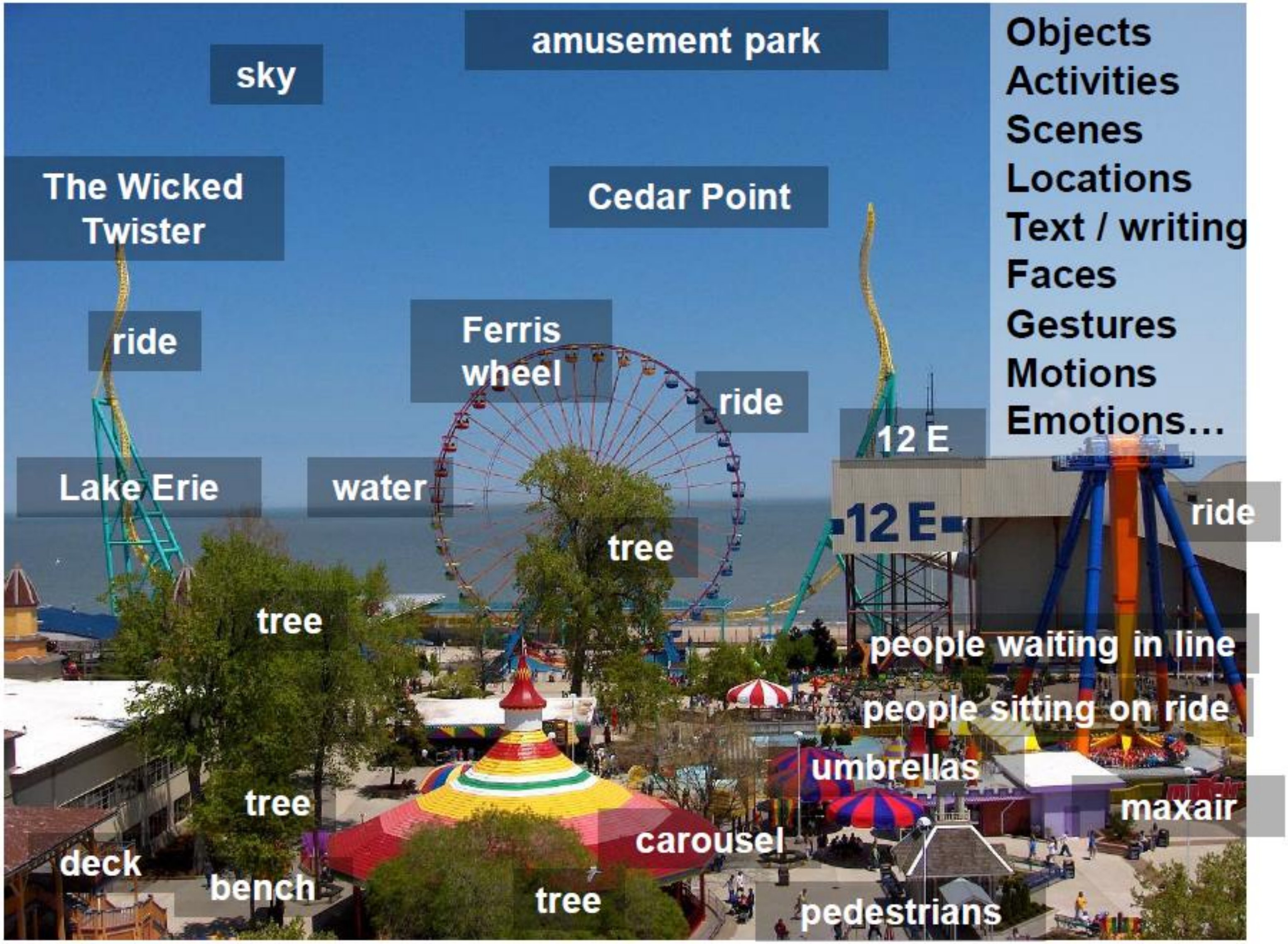
Can the computer match human perception?

- Yes and no (mainly no)
  - –computers can be better at “easy” things
  - –humans are much better at “hard” things
- But huge progress has been made
  - –Especially in the last 10 years
  - –What is considered “hard” keeps changing

What kind of information can we extract from an image?

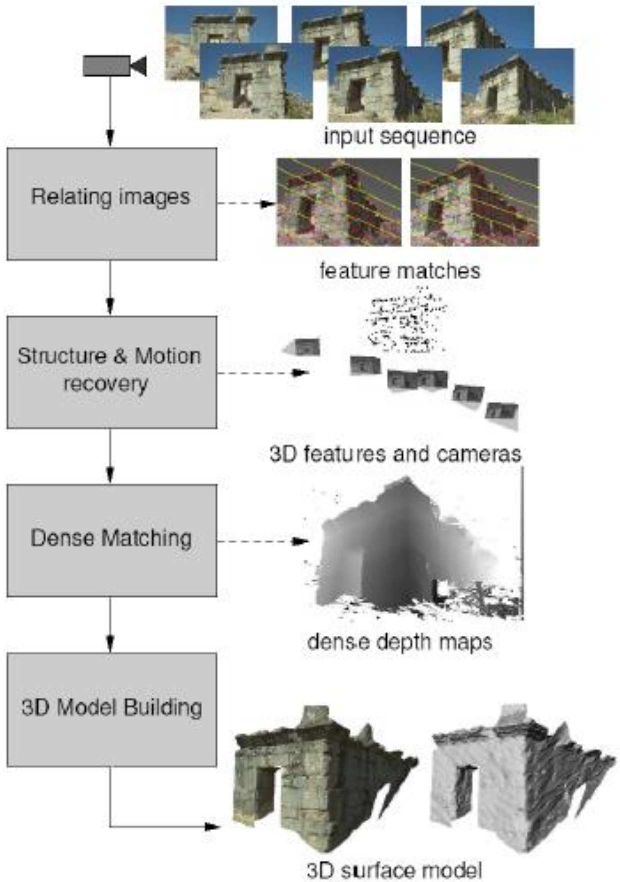
- Semantic information
- Metric 3D information

Vision as a source of semantic information

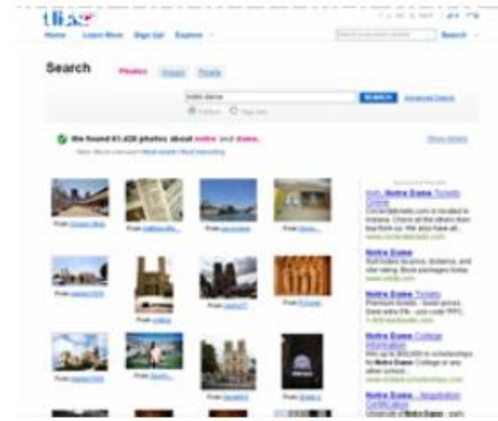


Slide from **Kristen Grauman**

Vision as measurement device



Pollefeys et al.



Goesele et al.

Why study computer vision?

- Vision is useful
- Vision is interesting
- Vision is difficult
  – Half of primate cerebral cortex is devoted to visual processing
  – Achieving human-level image understanding is probably “AI-complete”

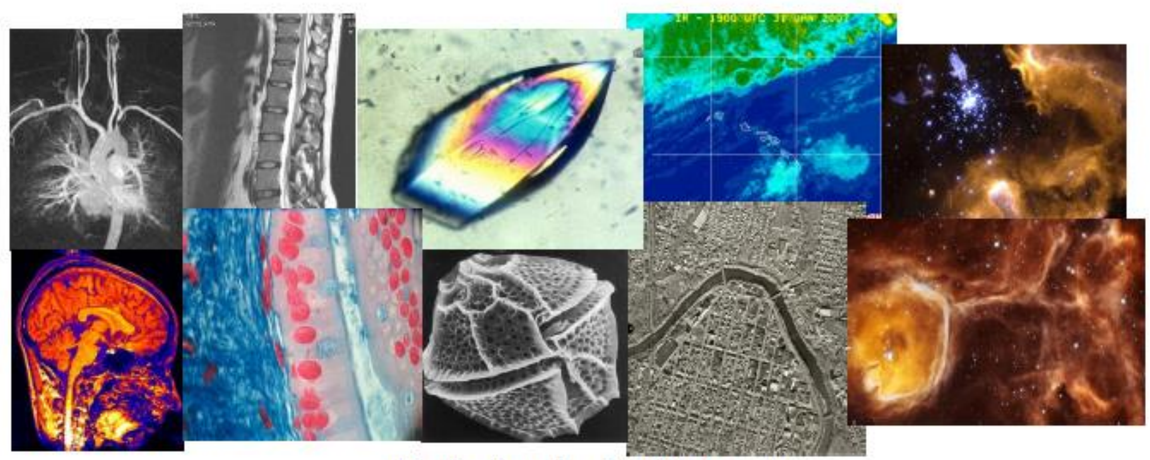
Slide adapted from **SVETLANA LAZEBNIK**

Why study computer vision?

- Millions of images being captured all the time

Google Image Search · Picasa · flickr · webshots · picsearch · YouTube Broadcast Yourself

Surveillance and security



Medical and scientific images

Visual data on the internet

- Flickr
  - 10+ billion photographs
  - 60 million images uploaded a month
- Facebook
  - 250 billion+
  - 300 million a day
- Instagram
  - 55 million a day
- YouTube
  - 100 hours uploaded every minute

Reconstruction: 3D from photo collections

Colosseum, Rome, Italy

San Marco Square, Venice, Italy

Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. Seitz, [The Visual Turing Test for Scene Reconstruction], 3DV 2013

[YouTube Video]

Slide adapted from **SVETLANA LAZEBNIK**

Reconstruction: 3D from photo collections

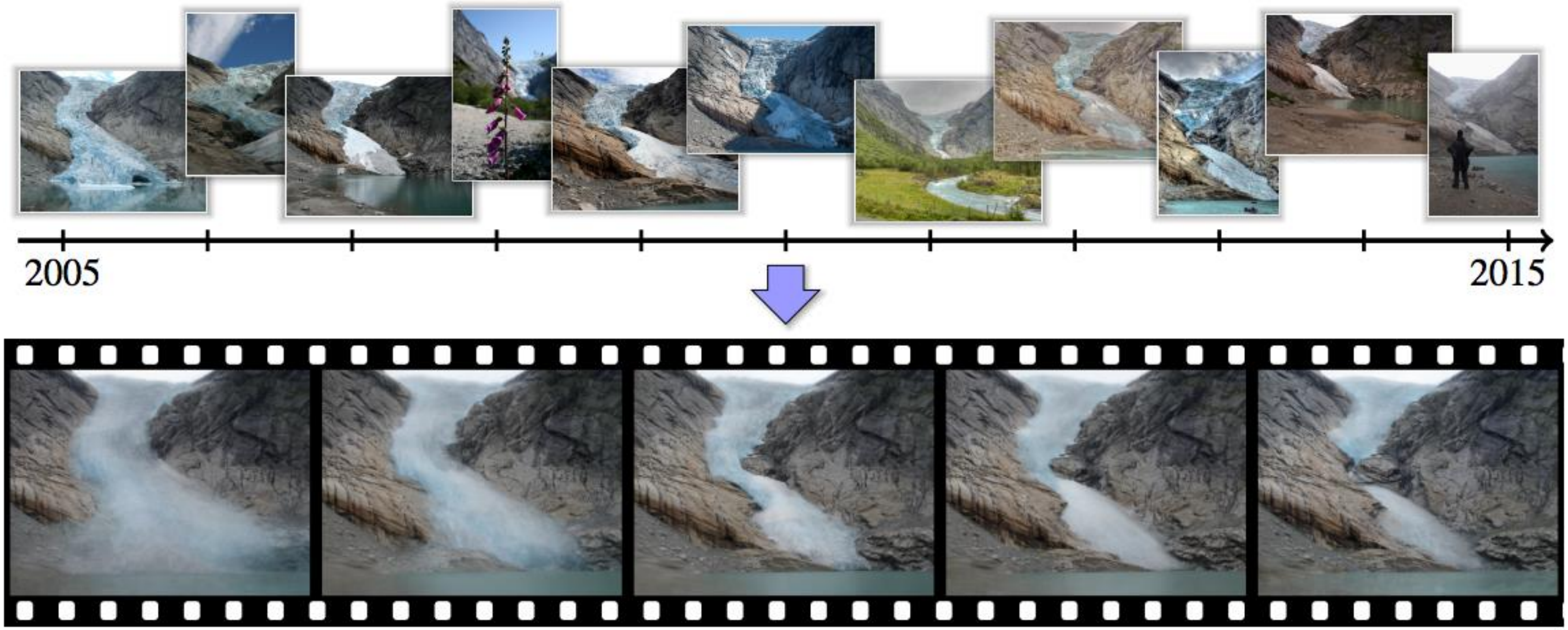


Figure 1: *We mine Internet photo collections to generate time-lapse videos of locations all over the world. Our time-lapses visualize a multitude of changes, like the retreat of the Briksdalsbreen Glacier in Norway shown above. The continuous time-lapse (bottom) is computed from hundreds of Internet photos (samples on top).* *Photo credits:* `Aliento Más Allá`, `jirihnidek`, `mcxurxo`, `elka_cz`, *Juan Jesús Orío, Klaus Wißkirchen,* `Daikrieg`, `Free the image`, `dration` *and Nadav Tobias.*

R. Martin-Brualla, D. Gallup, and S. Seitz, Time-Lapse Mining from Internet Photos, SIGGRAPH 2015

YouTube Video

Slide adapted from **SVETLANA LAZEBNIK**

Reconstruction: 4D from depth cameras

Figure 1: Real-time reconstructions of a moving scene with DynamicFusion; both the person and the camera are moving. The initially noisy and incomplete model is progressively denoised and completed over time (left to right).

R. Newcombe, D. Fox, and S. Seitz, [DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time](), CVPR 2015

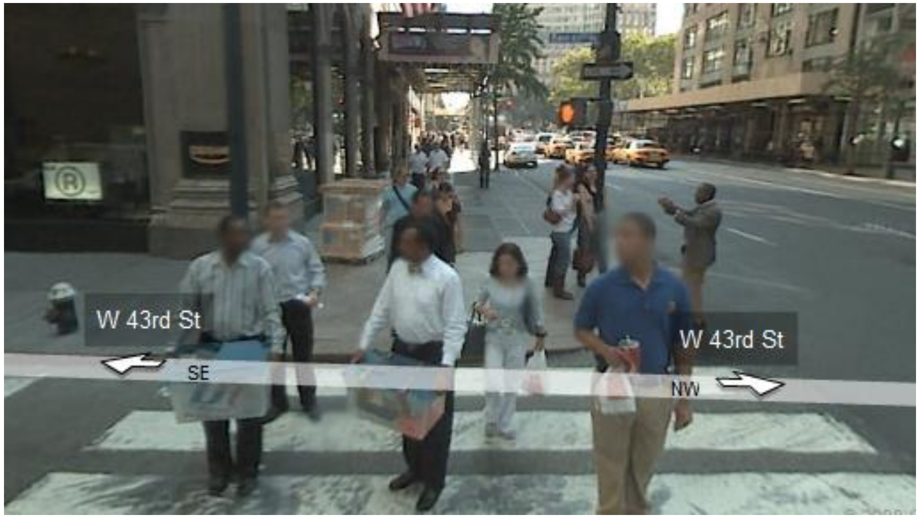
[YouTube Video]()

Slide adapted from **SVETLANA LAZEBNIK**

Recognition: “Simple” patterns

Slide adapted from **SVETLANA LAZEBNIK**

Recognition: Faces



Slide adapted from **SVETLANA LAZEBNIK**

Concerns about face recognition

[Beijing bets on facial recognition in a big drive for total surveillance]() – Washington Post, 1/8/2018

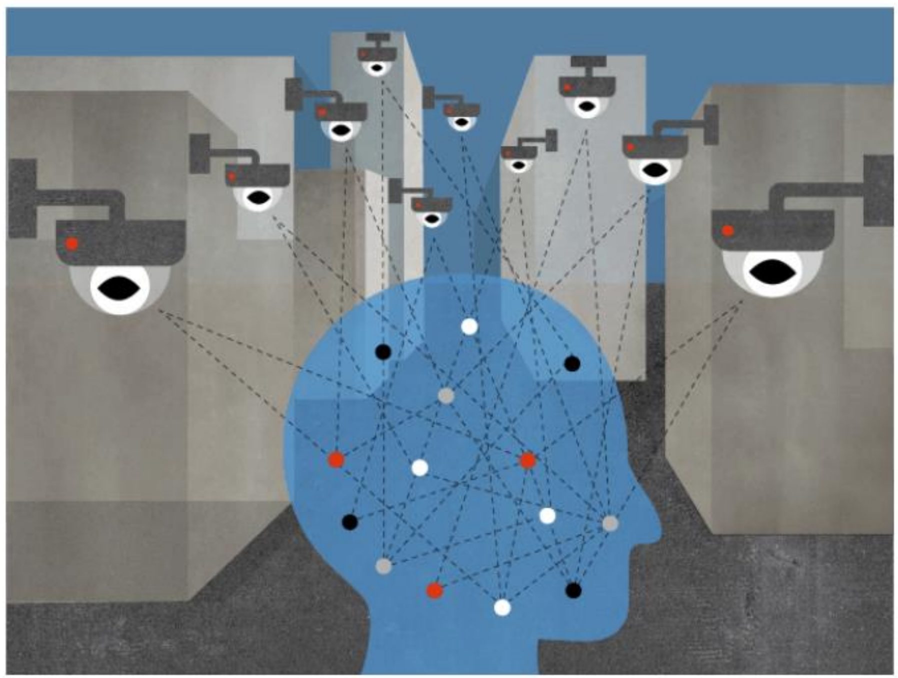
Concerns about face recognition

THE NEW YORKER

ANNALS OF TECHNOLOGY DECEMBER 17, 2018 ISSUE

SHOULD WE BE WORRIED ABOUT COMPUTERIZED FACIAL RECOGNITION?

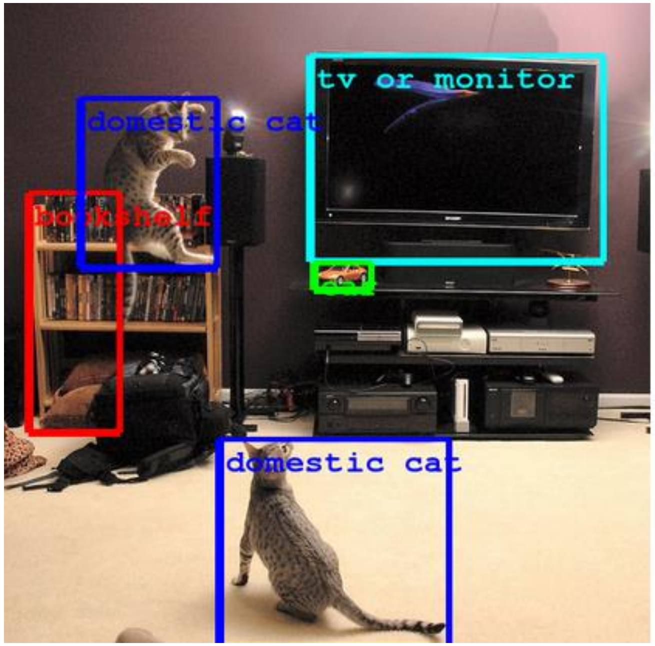
The technology could revolutionize policing, medicine, even agriculture—but its applications can easily be weaponized.

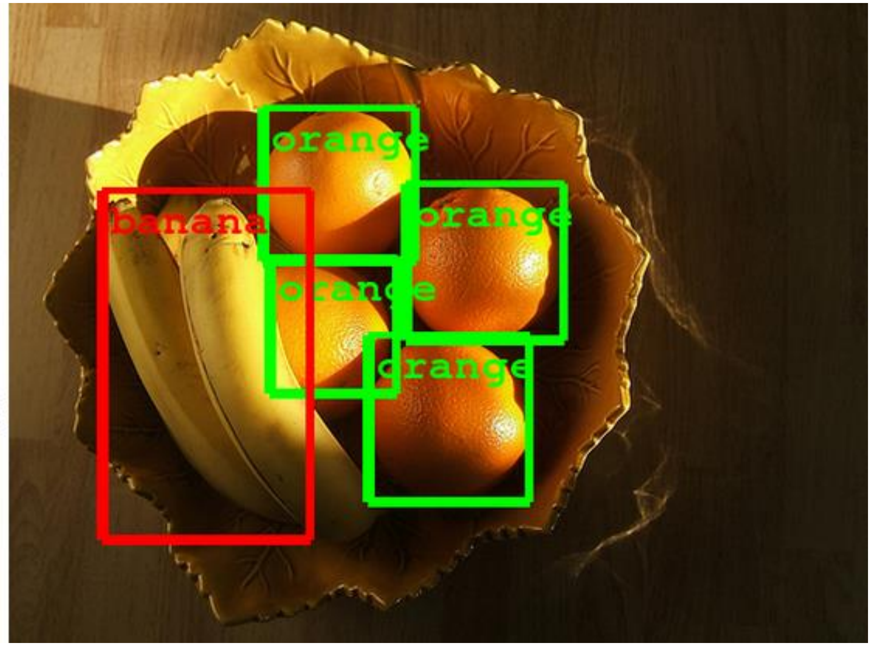


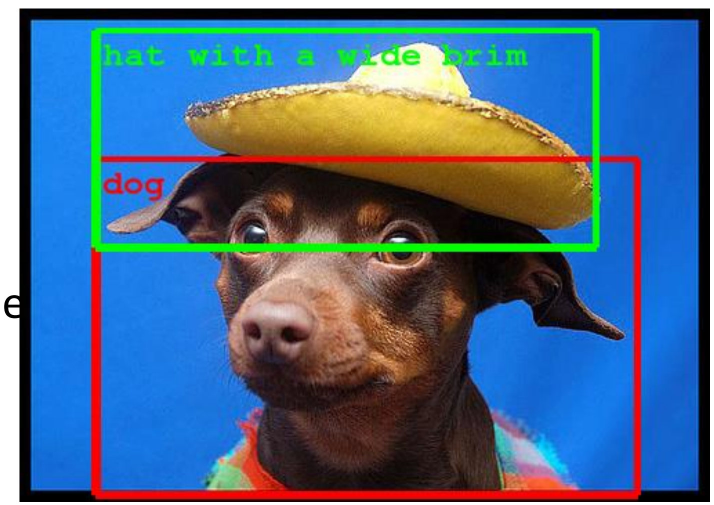
Many U.S. cities won't disclose their police departments' surveillance methods.

https://www.newyorker.com/magazine/2018/12/17/should-we-be-worried-about-

Recognition: General categories





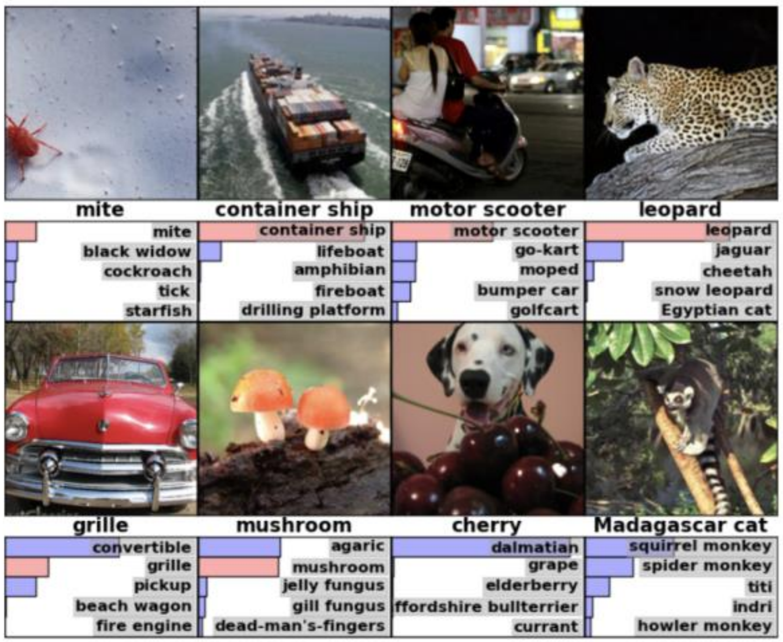


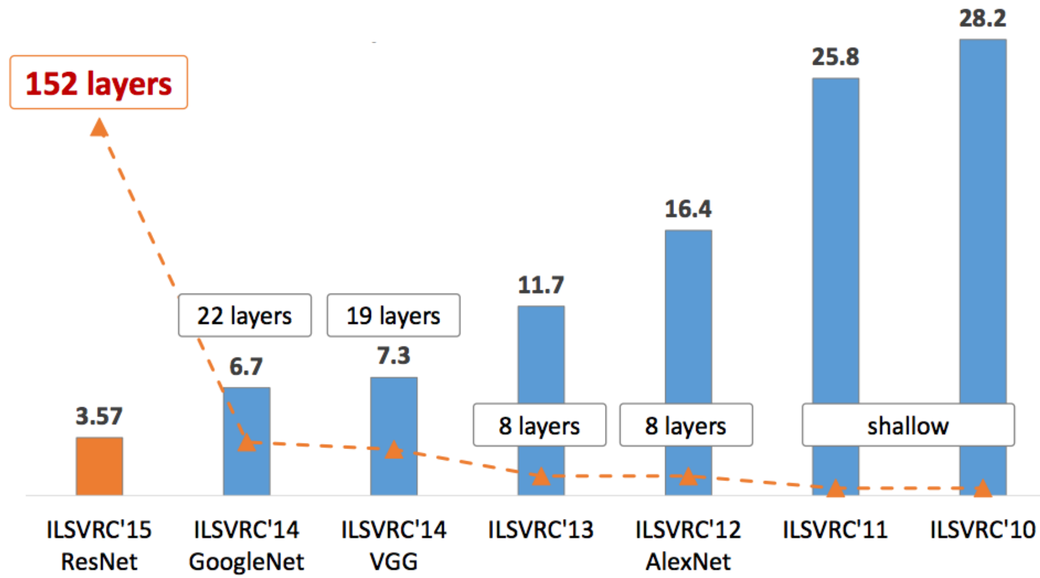
- Computer Eyesight Gets a Lot More Accurate, NY Times Bits blog, August 18, 2014
- Building A Deeper Understanding of Images, Google Research Blog, September 5, 2014

Slide adapted from **SVETLANA LAZEBNIK**

Recognition: General categories

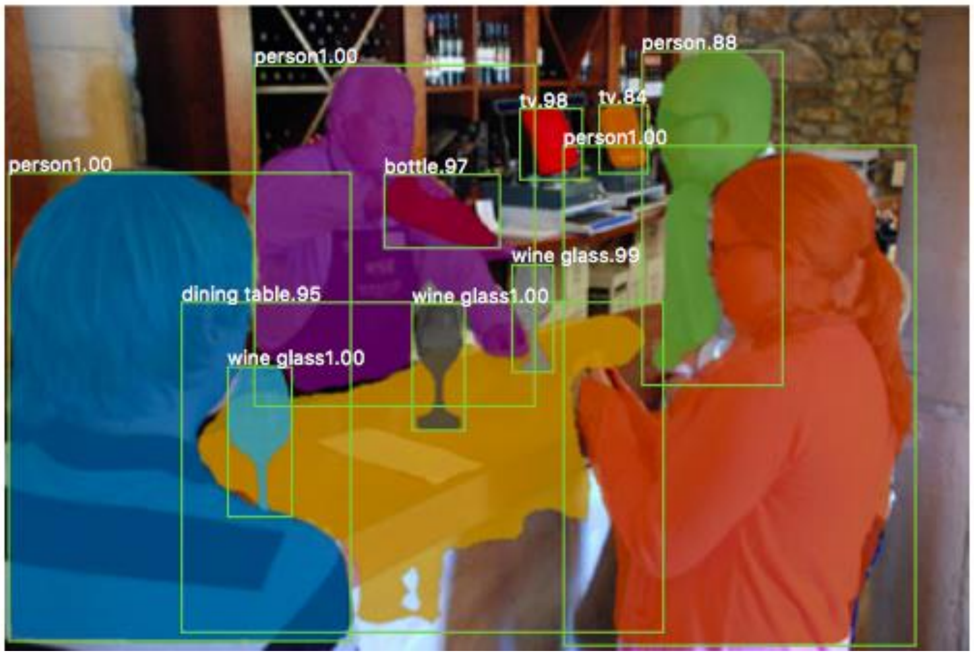
- ImageNet challenge

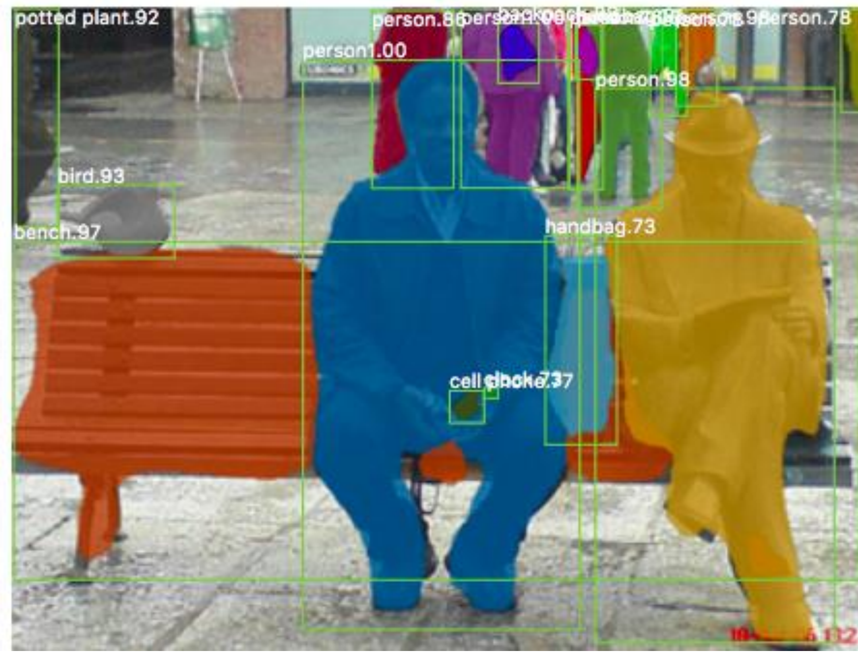




Slide adapted from **SVETLANA LAZEBNIK**

Object detection, instance segmentation





K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask R-CNN,
ICCV 2017 (Best Paper Award)

Slide adapted from **SVETLANA LAZEBNIK**

Image generation

- Faces: 1024x1024 resolution, CelebA-HQ dataset

T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, ICLR 2018
Follow-up work

Slide adapted from **SVETLANA LAZEBNIK**

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv 2018

Slide adapted from **SVETLANA LAZEBNIK**

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

A. Brock, J. Donahue, K. Simonyan, [Large scale GAN training for high fidelity natural image synthesis](), arXiv 2018

Slide adapted from **SVETLANA LAZEBNIK**

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

Easy classes

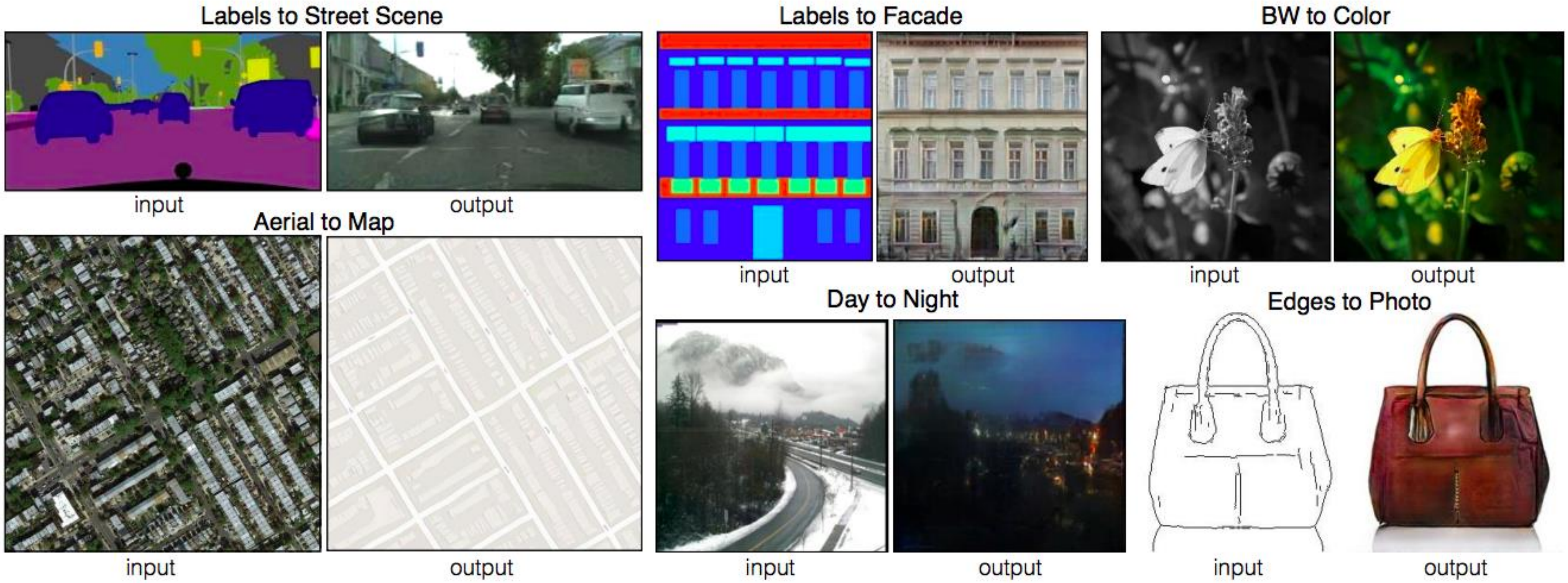
Difficult classes

A. Brock, J. Donahue, K. Simonyan, [Large scale GAN training for high fidelity natural image synthesis](), arXiv 2018

Slide adapted from **SVETLANA LAZEBNIK**

Image generation

- Image-to-image translation

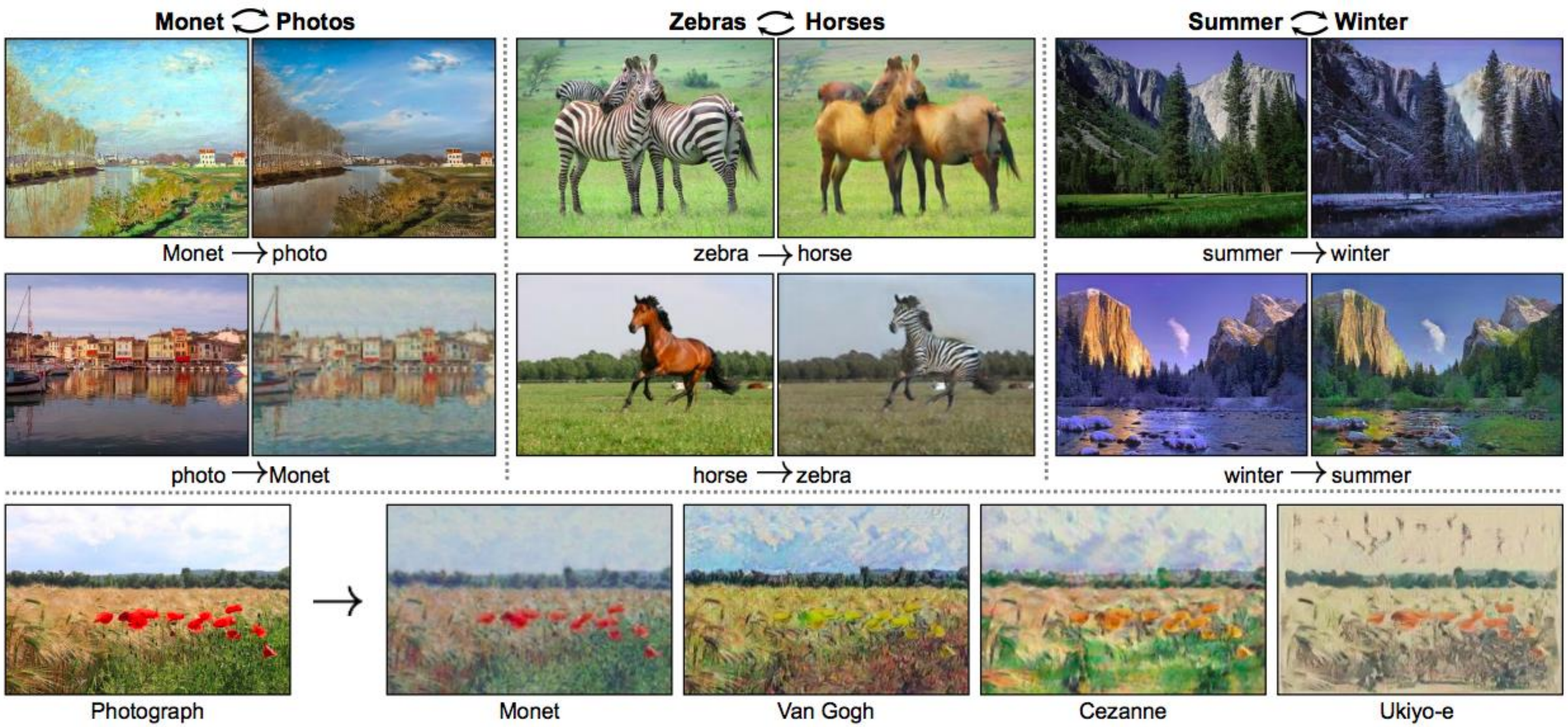


P. Isola, J.-Y. Zhu, T. Zhou, A. Efros, [Image-to-Image Translation with Conditional Adversarial Networks](), CVPR 2017

Slide adapted from **SVETLANA LAZEBNIK**

Image generation

- Unpaired image-to-image translation



J.-Y. Zhu, T. Park, P. Isola, A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, ICCV 2017

Slide adapted from **SVETLANA LAZEBNIK**

Unsupervised image-to-image translation

Figure 4: Dog breed translation results.

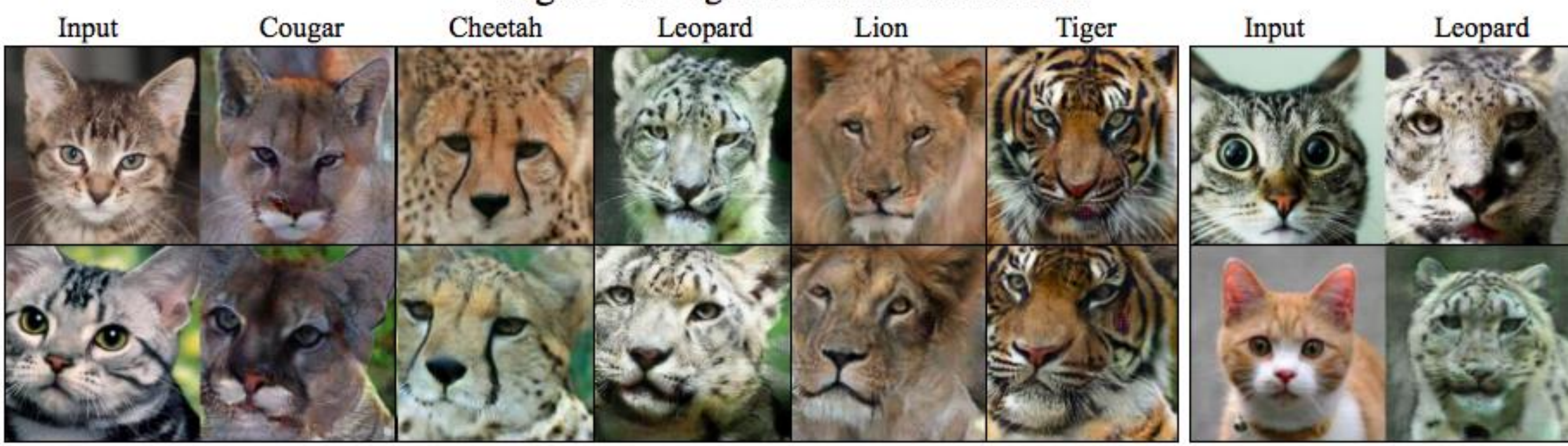


Figure 5: Cat species translation results.

M.-Y. Liu, T. Breuel, and J. Kautz, [Unsupervised Image-to-Image Translation Networks](), NIPS 2017

Unsupervised image-to-image translation

M.-Y. Liu, T. Breuel, and J. Kautz, Unsupervised Image-to-Image Translation Networks, NIPS 2017

Thank you: Question?