

CS 2770: Computer Vision
Introduction

Prof. Adriana Kovashka
University of Pittsburgh
January 7, 2020

About the Instructor



Born 1985 in
Sofia, Bulgaria



Got BA in 2008 at
Pomona College, CA
(Computer Science &
Media Studies)



Got PhD in 2014
at University of
Texas at Austin
(Computer Vision)

Course Info

- **Course website:** http://people.cs.pitt.edu/~kovashka/cs2770_sp20
- **Instructor:** Adriana Kovashka (kovashka@cs.pitt.edu)
 - Use "CS2770" at the beginning of your Subject
- **Office:** Sennott Square 5325
- **Class:** Tue/Thu, 11am-12:15pm
- **Office hours:**
 - Tuesday 12:20pm-1:30pm and 4:20pm-5:30pm
 - Thursday 2pm-2:50pm and 4:20pm-5:30pm
- **TA:** Narges Honarvar Nazari (nah114@pitt.edu)
- **TA's office:** Sennott Square 5501
- **TA's office hours:** TBD (Do this Doodle by end of Jan. 11: [link](#))

Types of computer vision

- Lower-level vision
 - Analyzing textures, edges and gradients in images, without concern for the semantics (e.g. objects) of the image
- Higher-level vision (our focus)
 - Making predictions about the semantics or higher-level functions of content in images (e.g. objects, attributes, styles, motion, etc.)
 - Involves machine learning

Course Goals

- To learn the basics of low-level image analysis
- To learn the modern approaches to classic high-level computer vision tasks
- To get experience with some computer vision techniques
- To learn/apply basic machine learning (a key component of modern computer vision)
- To get exposure to emerging topics and recent research
- To think critically about vision approaches, and to see connections between works and potential for improvement

Textbooks

- [Computer Vision: Algorithms and Applications](#) by Richard Szeliski
- [Visual Object Recognition](#) by Kristen Grauman and Bastian Leibe
- More resources available on course webpage
- Your notes from class are your best study material, slides are *not* complete with notes

Programming Languages

- Homework: Matlab or Python
- Projects: Whatever language you like

Matlab

(Download for free from Pitt Software Downloads.)

http://www.cs.pitt.edu/~kovashka/cs2770_sp18/tutorial.m

http://www.cs.pitt.edu/~kovashka/cs2770_sp18/myfunction.m

http://www.cs.pitt.edu/~kovashka/cs2770_sp18/myotherfunction.m

<https://people.cs.pitt.edu/~milos/courses/cs2750/Tutorial/>

http://www.math.udel.edu/~braun/M349/Matlab_probs2.pdf

<http://www.facstaff.bucknell.edu/maneval/help211/basicexercises.html>

Ask the TA or instructor if you have any problems.

Python/NumPy/SciPy

<http://cs231n.github.io/python-numpy-tutorial/>

<https://docs.scipy.org/doc/numpy/user/numpy-for-matlab-users.html>

Ask the TA or instructor if you have any problems.

Computing Resources

- Graphics Processing Unit (GPU)-equipped servers provided by Pitt's Center for Research Computing (CRC)
- Login details soon
- Set up SSH and VPN soon

Course Structure

- Lectures
- Three programming assignments
- Course project
- Two exams
- Participation

Policies and Schedule

See course website!

Project milestones

- February: project proposals (2-3 pages)
- March: status report presentations (incl. literature review)
- April: final presentations

Logistics

- Form teams of three
- Proposal: write-up due on CourseWeb
- First presentation: 10 min
- Second presentation: 12 min

Project proposal

- Pick among the suggestions in later slides, OR propose your own – need to answer the same questions in both cases
- What is the problem you want to solve? Why is it important?
- What related work exists? This is important so you don't reinvent the wheel, and also so you can leverage prior work
- What data is available for you to use? Is it sufficient? How will you deal with it if not?
- What is your baseline method?
- What is a slightly more interesting method you can try? You don't have to have all details, just an idea
- How will you evaluate your method?

Project proposal rubric

One point for answering each of the questions below:

- What do you propose to do?
- What have others attempted in this space, i.e. what is the relevant literature?
- Why is what you are proposing interesting?
- Why is it challenging?
- Why is it important?
- What data do you plan to use?
- What is your high-level idea of how your method will work?
- In what ways is this method novel?
- How will you evaluate the method, i.e. what metrics are you going to use, and what baselines are you going to compare to?
- Give a (1) conservative and (2) an ambitious schedule of milestones for your project.

Status/first report presentation

- Assumption is you've done a complete literature review, and decided on your method
- Ideally you've begun your experiments but not completed them

Status/first presentation grading rubric

All questions except the last one scored on a scale of 1 to 5, 5=best:

- How well did the authors (presenters) explain what problem they are trying to solve?
- How well did they explain why this problem is important?
- How well did they explain why the problem is challenging?
- How thorough was the literature review?
- How clearly was prior work described?
- How well did the authors explain how their proposed work is different than prior work?
- How clearly did the authors describe their proposed approach?
- How novel is the proposed approach?
- How challenging and ambitious is the proposed approach?

Final presentation

- You've previously already talked about your method in depth
- Now review your problem and method (briefly) and describe your experiments and findings
- You need to analyze the results, not just show them

Final presentation grading rubric

All questions except the last one scored on a scale of 1 to 5, 5=best:

- To what extent did the authors develop the method as described in the first presentation? (1-10)
- How well did the authors describe their experimental validation?
- How informative were the figures used?
- Were all/most relevant baselines and competitor methods included in the experimental validation?
- Were sufficient experimental settings (e.g. datasets) tested?
- To what extent is the performance of the proposed method satisfactory?
- How informative were the conclusions the authors drew about their method's performance relative to other methods?
- How sensible was the discussion of limitations?
- How interesting was the discussion of future work?

Tips for a successful project

- From your perspective:
 - Learn something
 - Try something out for a real problem

Tips for a successful project

- From your classmates' perspective:
 - Hear about a topic in computer vision we haven't covered in depth
 - Hear about challenges and how you handled them, that they can use in their own work
 - Listen to an engaging presentation on a topic they care about

Tips for a successful project

- From my perspective:
 - Hear about the creative solutions you came up with to handle challenges
 - Hear your perspective on a topic that I care about
 - Co-author a publication with you, potentially with a small amount of follow-up work – a really good deal, and looks good on your CV!

Tips for a successful project

- Summary
 - Don't reinvent the wheel – your audience will be bored
 - But it's ok to adapt an existing method to a new domain/problem...
 - If you show interesting experimental results...
 - You analyze them and present them in a clear and engaging fashion

Possible project topics

- Common-sense reasoning
- Vision and language interactions
- Object detection
- Self-supervised learning
- Domain adaptation
- ...

See CourseWeb for details

Should I take this class?

- It will be a lot of work
 - But you will learn a lot
- Some parts will be hard and require that you pay close attention
 - Use instructor's and TA's office hours!
- Some aspects are open-ended are there are no clear correct answers
 - You will learn/practice reading research papers

Questions?

Plan for Today

- Introductions
- Intro quiz
- What is computer vision?
 - Why do we care?
 - What are the challenges?
- Overview of topics
 - What is current research like?
- Linear algebra blitz review

Introductions

- What is your name?
- What one thing outside of school are you passionate about?
- Do you have any prior experience with computer vision or machine learning?
- What do you hope to get out of this class?
- **Every time you speak, please remind me your name, and say it slowly**

Intro quiz

- Socrative.com
- Room: KOVASHKA

Computer Vision

What is computer vision?



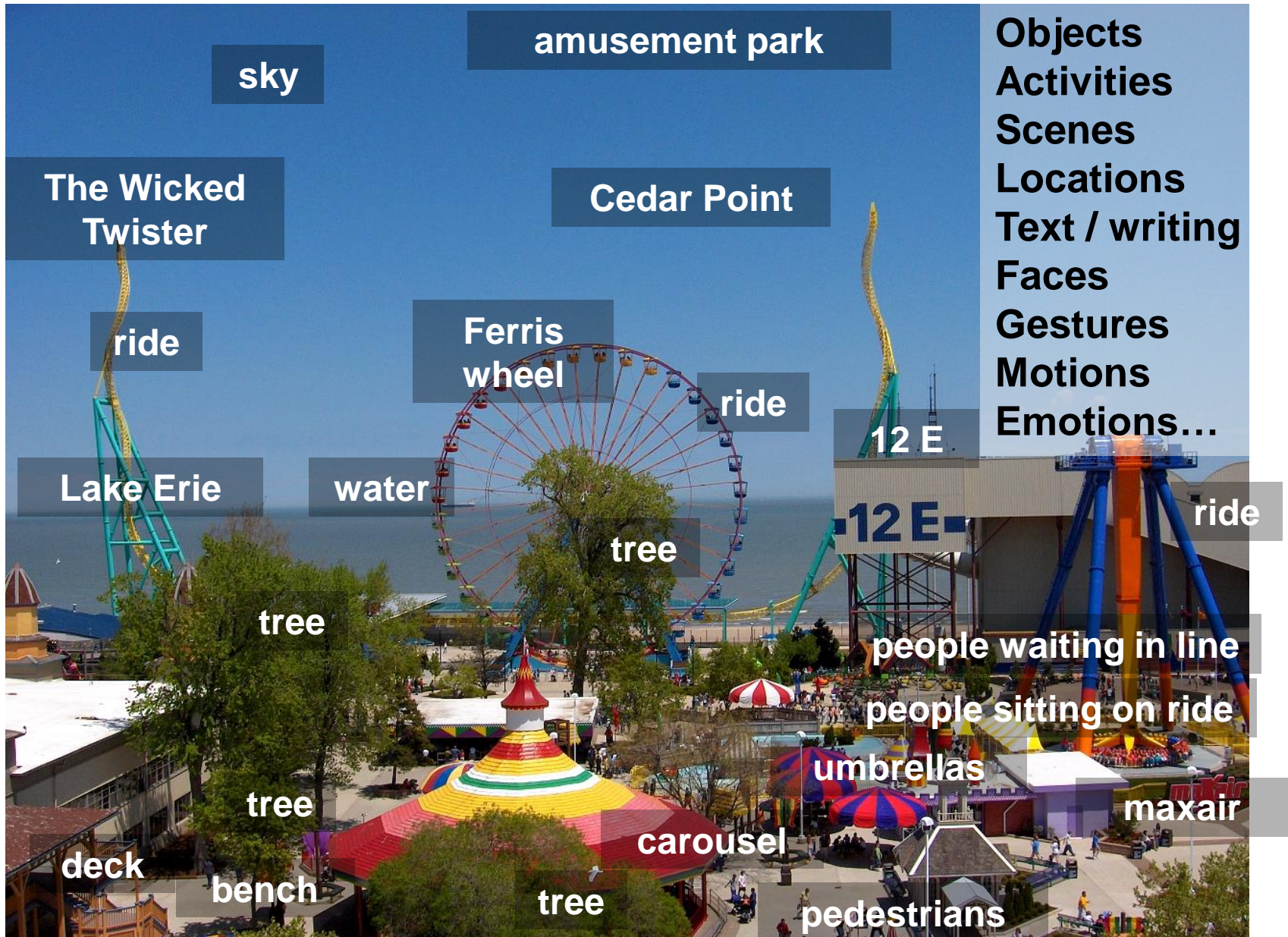
Done?

"We see with our brains, not with our eyes" (Oliver Sacks and others)

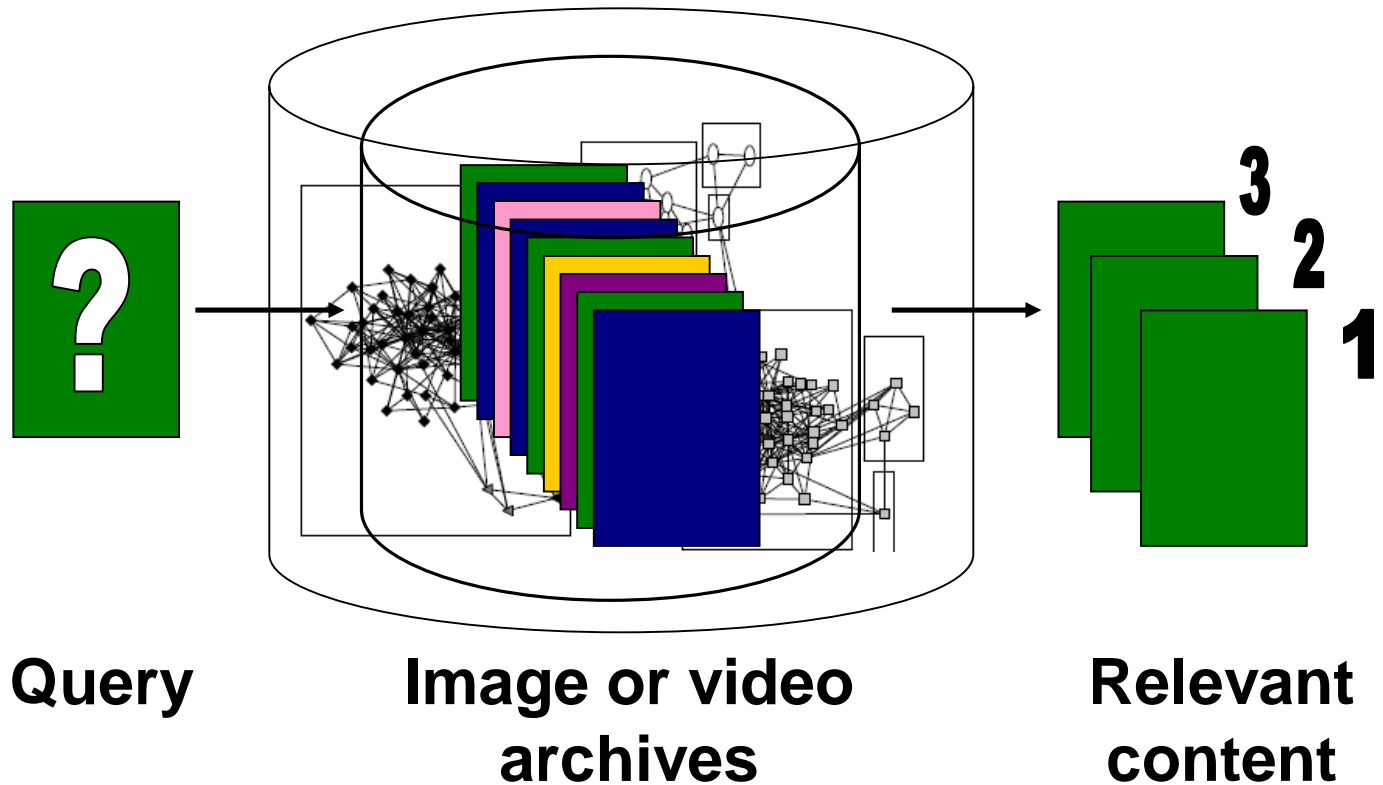
What is computer vision?

- Automatic understanding of images/video, e.g.
 - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities
 - Algorithms to mine, search, interact with visual data
 - Computing properties of the 3D world from visual data

Vision for recognition



Visual search, organization



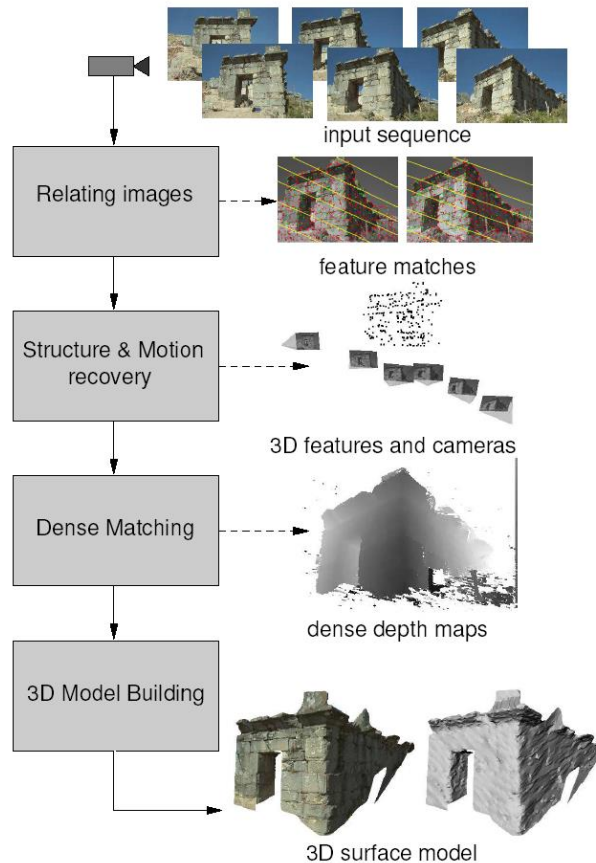
Vision for measurement

Real-time stereo



Pollefeys et al.

Structure from motion

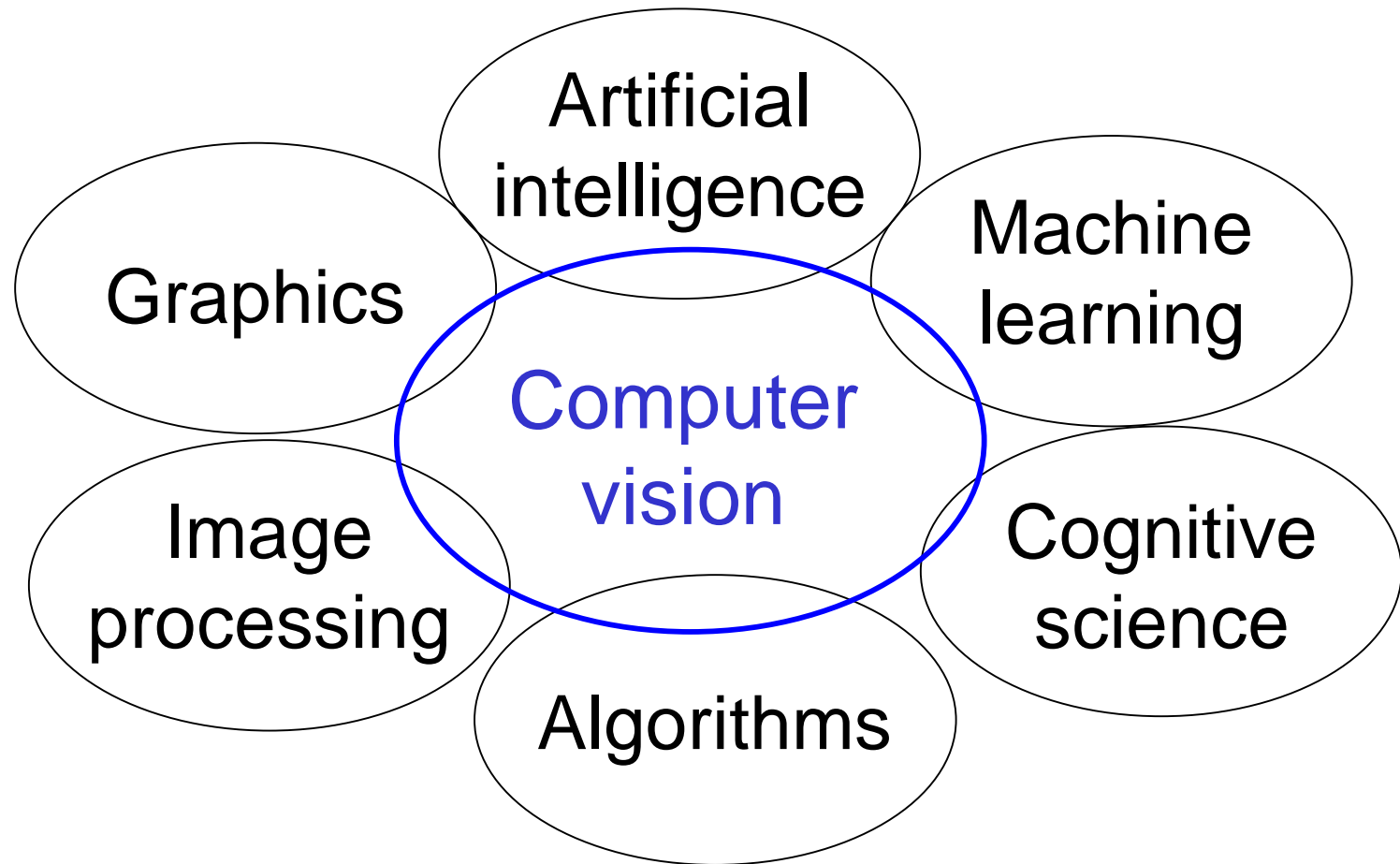


Multi-view stereo for community photo collections

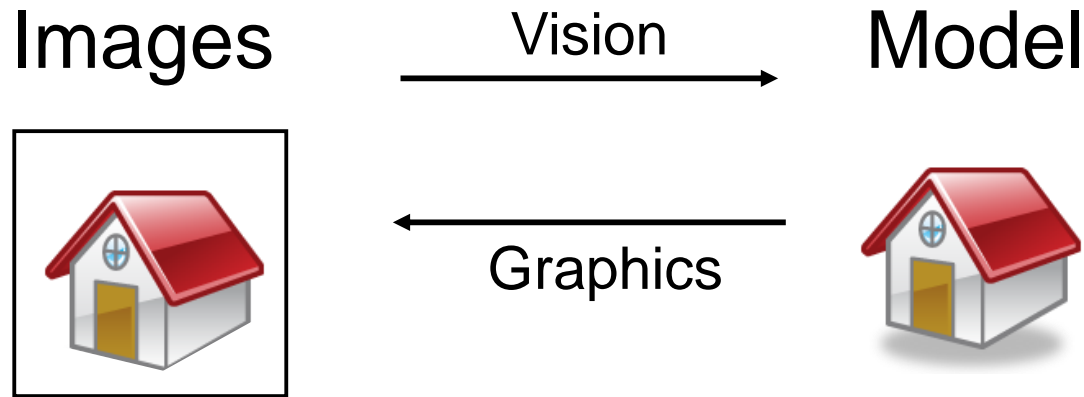


Goesele et al.

Related disciplines



Vision and graphics



Inverse problems: analysis and synthesis.

450k hours uploaded to YouTube daily
95 mil photos uploaded to Instagram daily
10 bil images indexed by Google

-
- A collage of five photographs arranged in a 2x3 grid, with the bottom-right cell empty. The top-left photo shows a baby crawling on a white surface. The top-right photo shows a couple embracing. The middle-left photo shows two men fishing on a boat. The middle-right photo shows a couple with a horse. The bottom-left photo shows four graduates in caps and gowns.

A collage of 10 images showing Mexican celebrities and athletes. The top row includes Salma Hayek, a man with arms raised, a news anchor with a radiation symbol, and a football player. The bottom row includes Al Pacino and Aladdin, a man and woman in a car, a news anchor with 'Medicare Coverage' text, and a baseball player sliding.



shutterstock™



A collage of 10 images illustrating various applications of imaging technology. The images include: a 3D reconstruction of a vascular tree; a sagittal MRI scan of a human spine; a microscopic view of a biological structure with a color-coded overlay; a false-color satellite or aerial image of a coastal area; a deep-space astronomical image showing a bright star and nebulae; a cross-sectional MRI scan of a human head; a microscopic view of a biological structure with red and blue color coding; a 3D reconstruction of a porous, lattice-like structure; a grayscale aerial photograph of a city with a river; and a false-color astronomical image of a nebula.

Adapted from Lana Lazebnik

Why vision?

- As image sources multiply, so do applications
 - Relieve humans of boring, easy tasks
 - Human-computer interaction
 - Perception for robotics / autonomous agents
 - Organize and give access to visual content
 - Description of image content for the visually impaired
 - Fun applications (e.g. transfer art styles to my photos)

What tasks are currently feasible
for computer vision systems?

Faces and digital cameras



Camera waits for everyone to smile to take a photo [Canon]



Setting camera focus via face detection

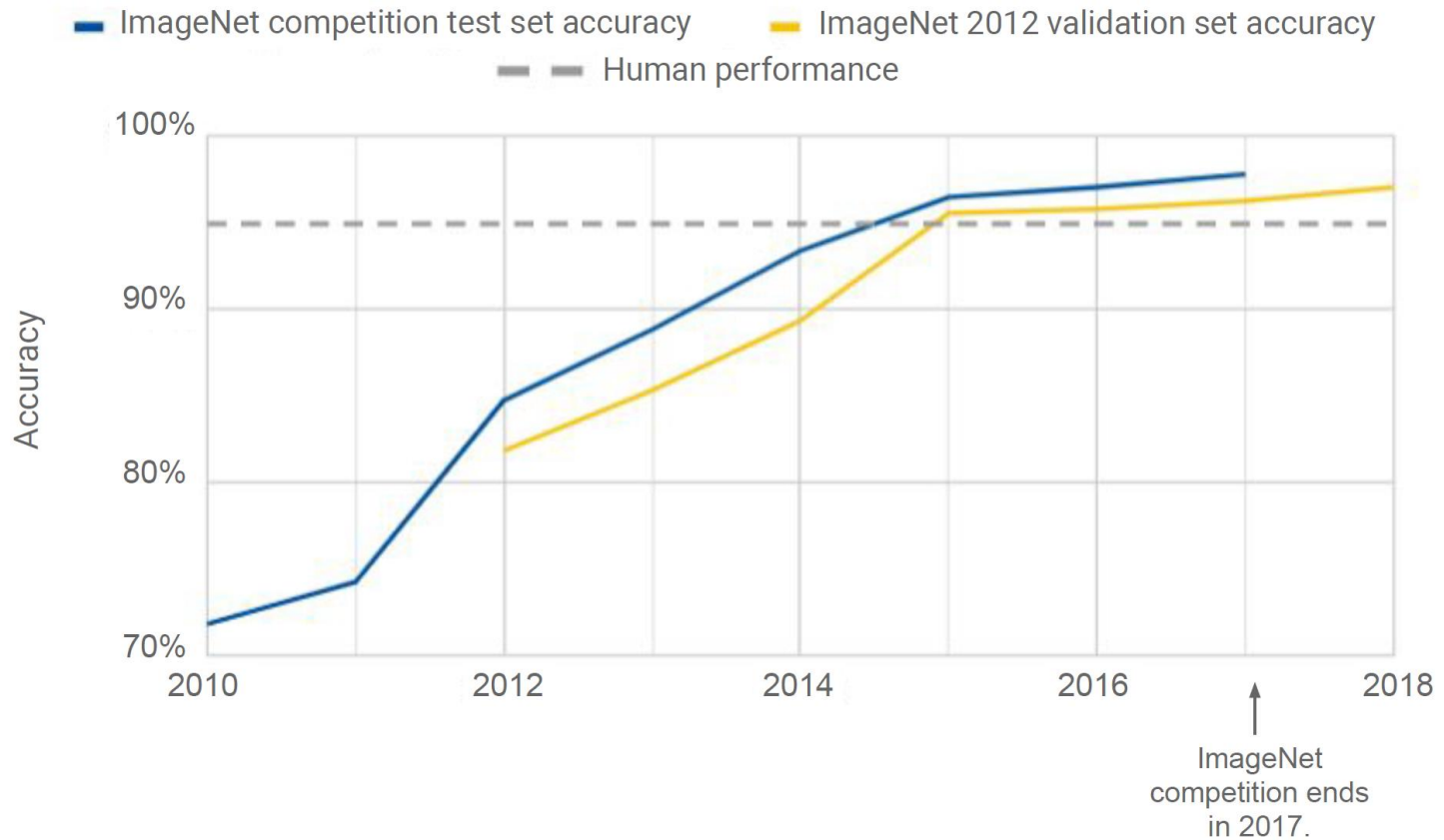
Face recognition



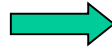
Object classification

ImageNet (2010 –2018)

Source: ImageNet; see appendix



Linking to info with a mobile device



Situated search
Yeh et al., MIT



MSR Lincoln



kooaba



Exploring photo collections



Photo Tourism

Exploring photo collections in 3D

Microsoft



(a)



(b)



(c)

Snaveley et al.

Interactive systems

KINECT
for XBOX 360.



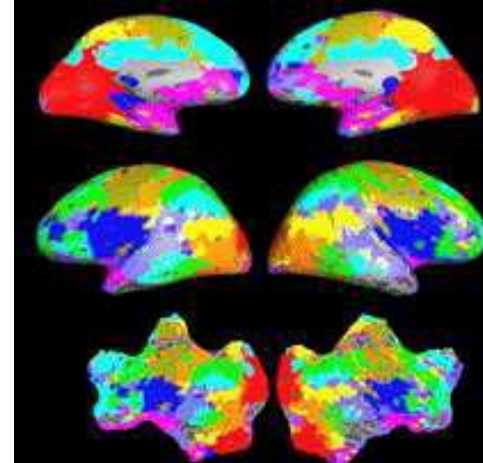
Shotton et al.



Vision for medical & neuroimages



Image guided surgery
MIT AI Vision Group



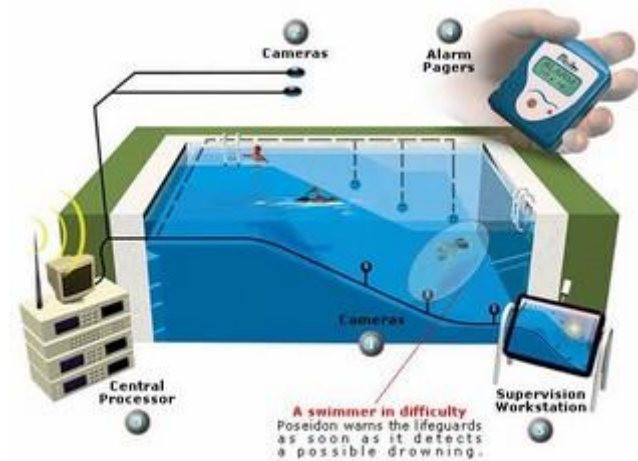
fMRI data
Golland et al.



Safety & security



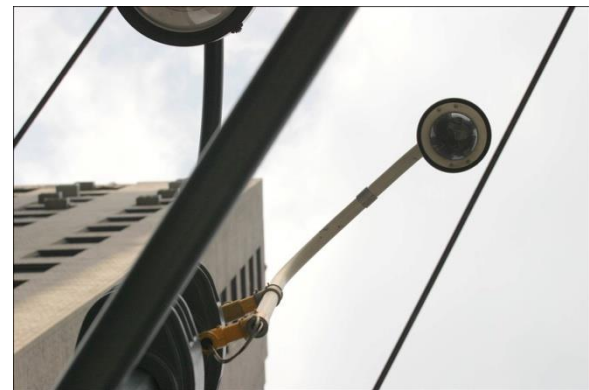
Navigation,
driver safety



Monitoring pool
(Poseidon)



Pedestrian detection
MERL, Viola et al.



Surveillance

Healthy eating



FarmBot.io
[YouTube Link](#)

Im2calories by Myers et al., ICCV 2015
[figure source](#)



Self-training for sports?

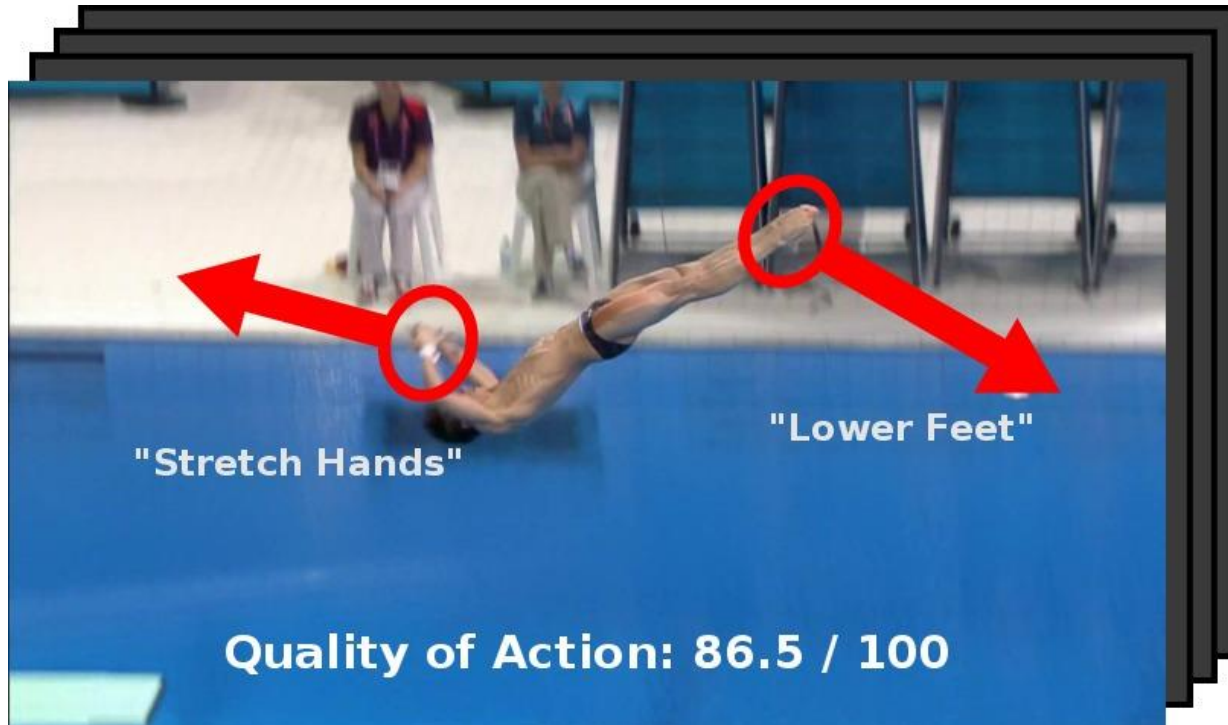
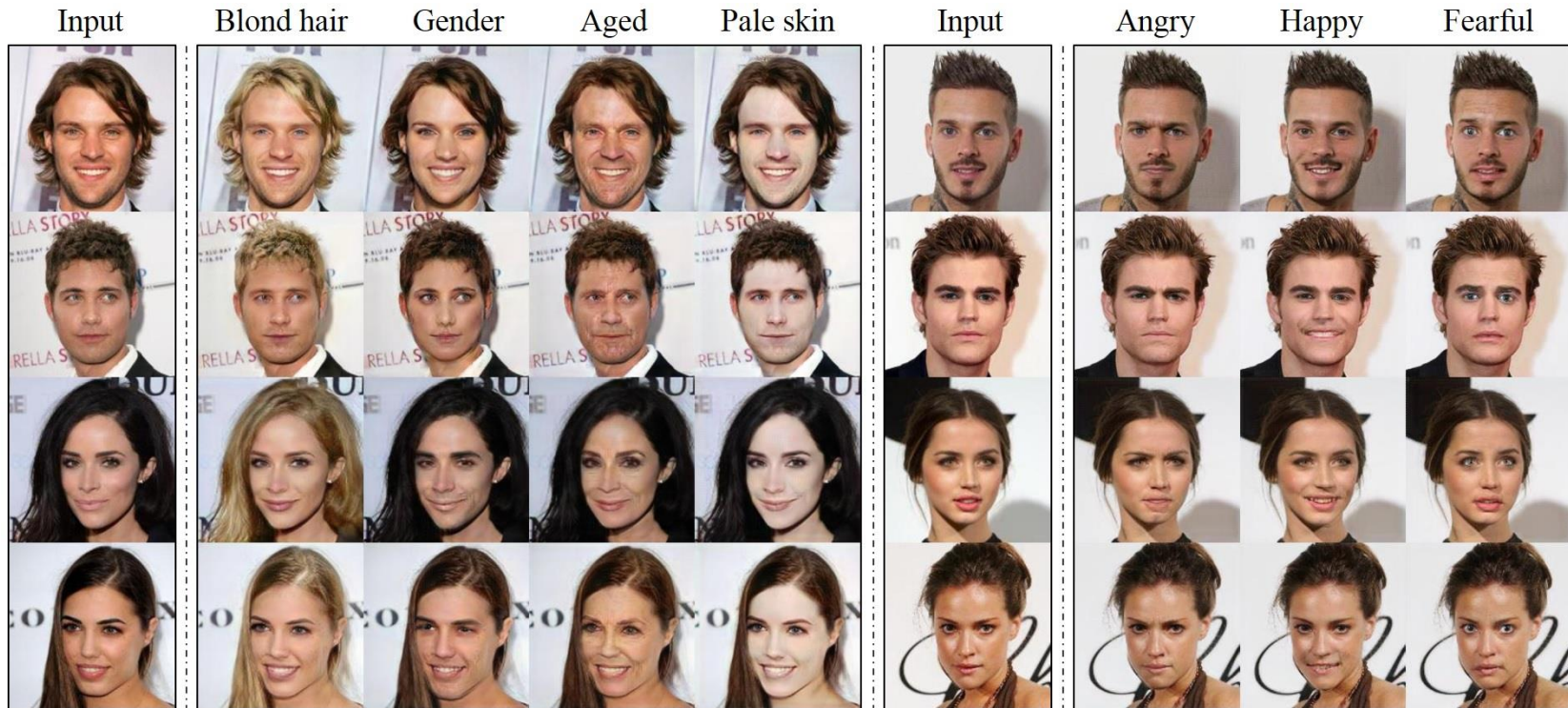


Image generation



Choi et al., CVPR 2018

this small bird has a pink breast and crown, and black primaries and secondaries.



this magnificent fellow is almost all black with a red crest, and white cheek patch.



Reed et al., ICML 2016

Seeing AI

[YouTube link](#)



Microsoft Cognitive Services: Introducing the Seeing AI project

Obstacles?

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

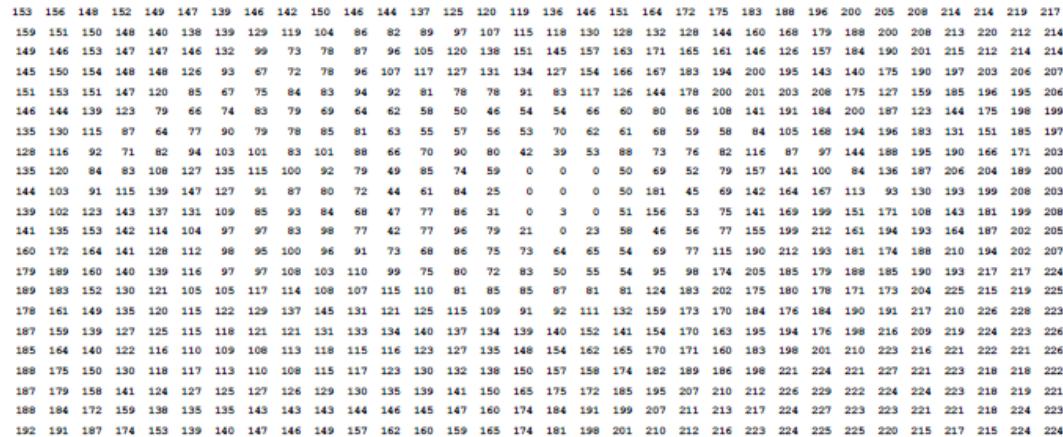
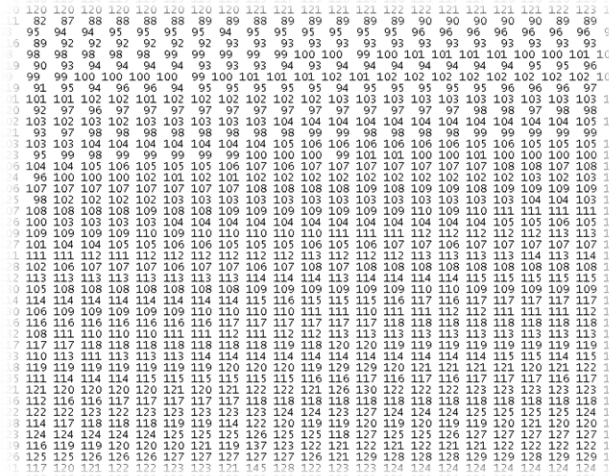
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Read more about the history: Szeliski Sec. 1.2

Why is vision difficult?

- Ill-posed problem: real world much more complex than what we can measure in images
 - 3D \rightarrow 2D
 - Dynamic \rightarrow static (many tasks)
- Impossible to literally “invert” image formation process with limited information
 - Need information outside of this particular image to generalize what image portrays (e.g. to resolve occlusion)

What the computer gets



Why is this problematic?

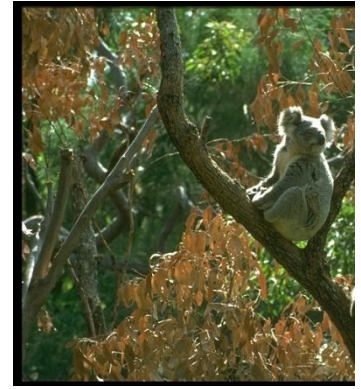
Challenges: many nuisance parameters



Illumination



Object pose



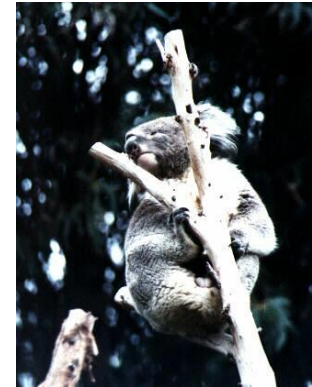
Clutter



Occlusions



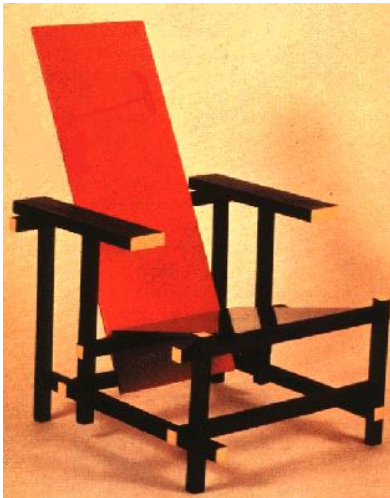
**Intra-class
appearance**



Viewpoint

Think again about the pixels...

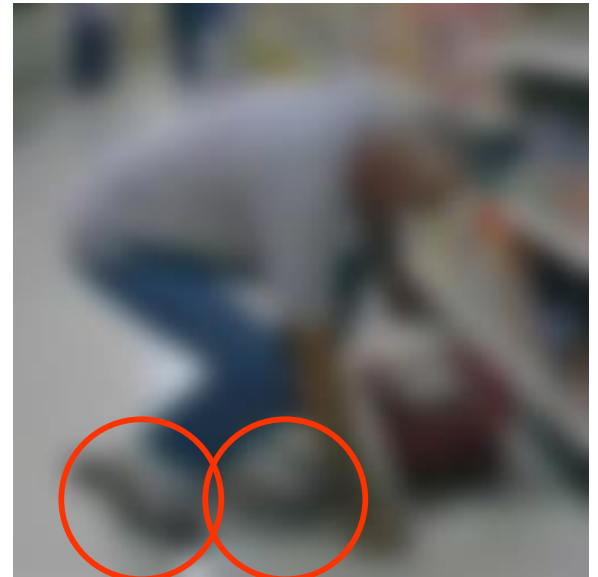
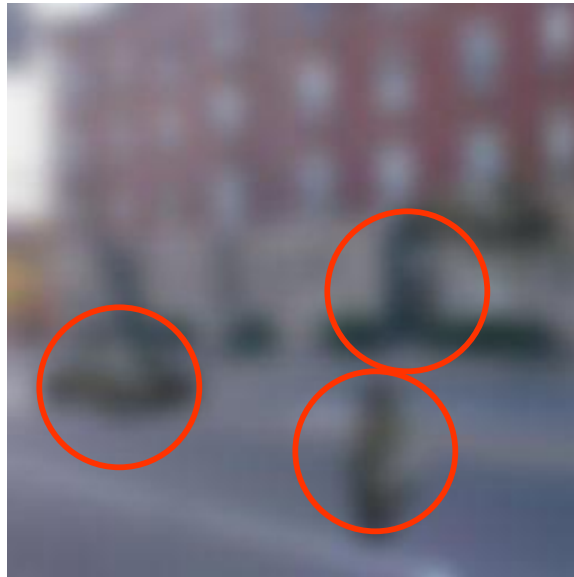
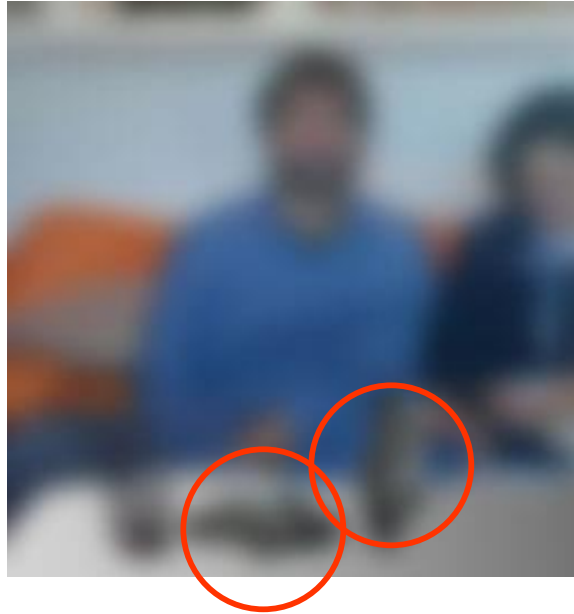
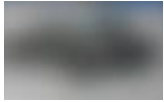
Challenges: intra-class variation



CMOA Pittsburgh



Challenges: importance of context



Challenges: Complexity

- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search
- 1.424 billion smart camera phones sold in 2015
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

Challenges: Limited supervision



Challenges: Vision requires reasoning



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



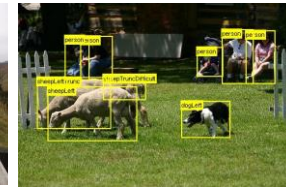
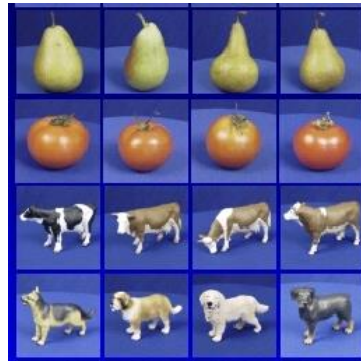
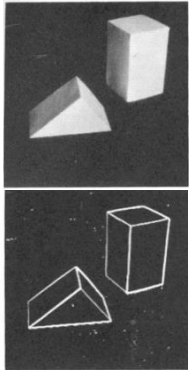
Is this person expecting company?
What is just under the tree?



Does it appear to be rainy?
Does this person have 20/20 vision?

Evolution of datasets

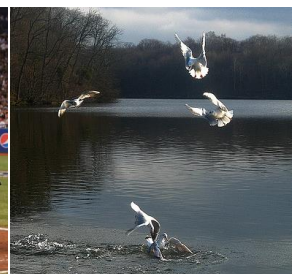
- Challenging problem → active research area



PASCAL:
20 categories, 12k images



ImageNet:
22k categories, 14mil images

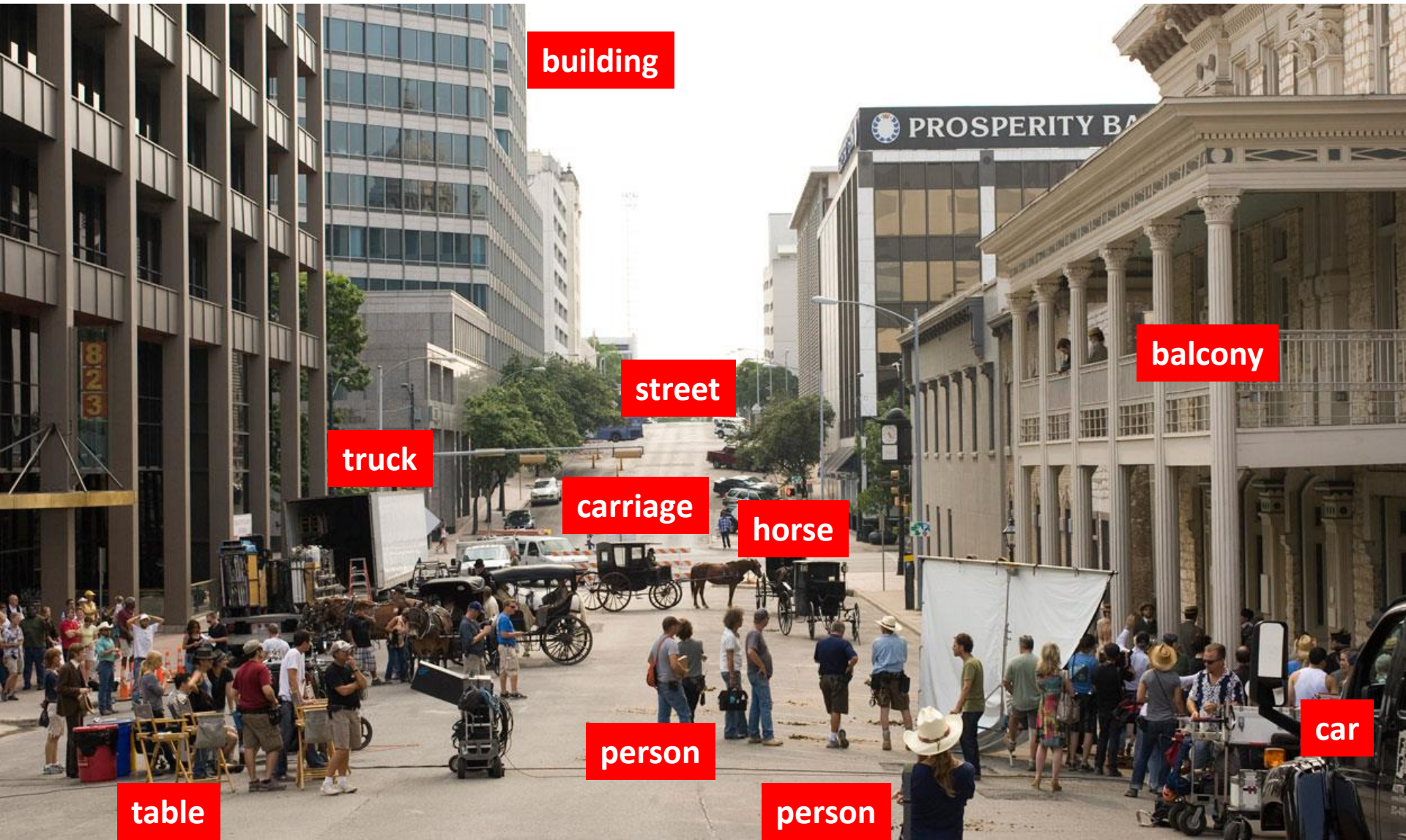


Microsoft COCO:
80 categories, 300k images

Some Visual Recognition Problems: Why are they challenging?



Recognition: What objects do you see?



building

balcony

street

truck

carriage

horse

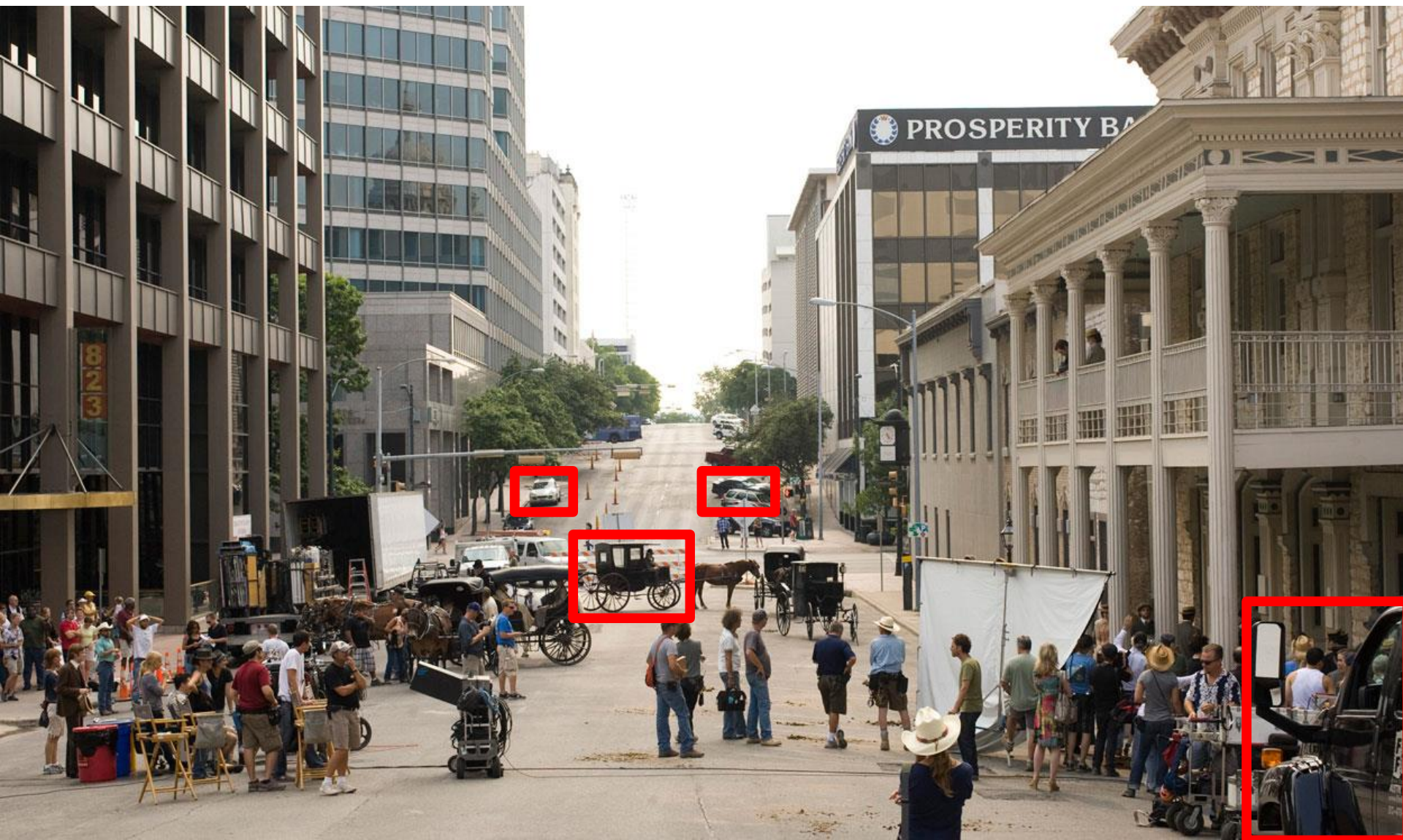
car

person

person

table

Detection: Where are the cars?



Activity: What is this person doing?



Scene: Is this an indoor scene?



Instance: Which city? Which building?



Visual question answering:

Why is there a carriage in the street?

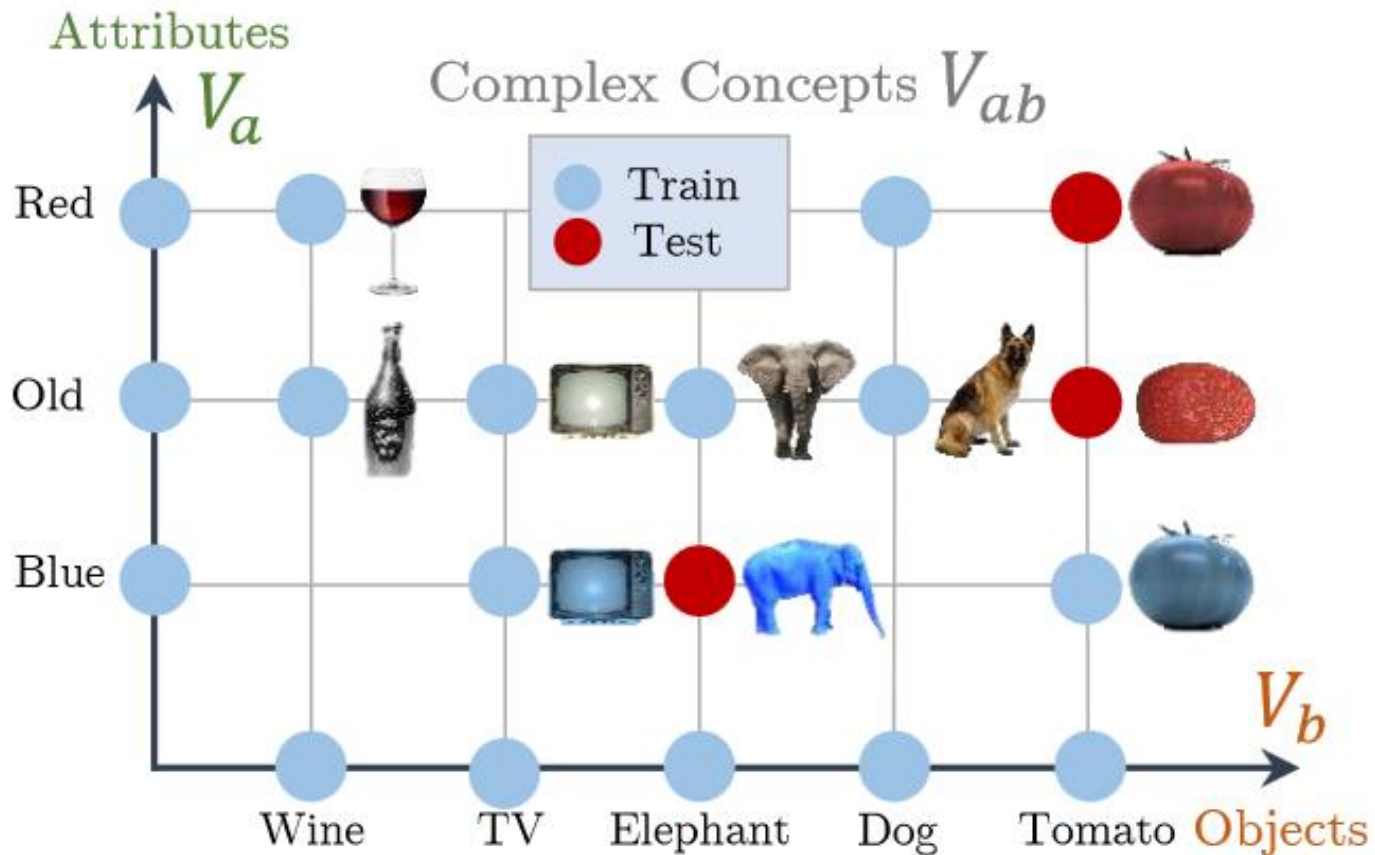


What tasks are computer vision researchers actively working on?

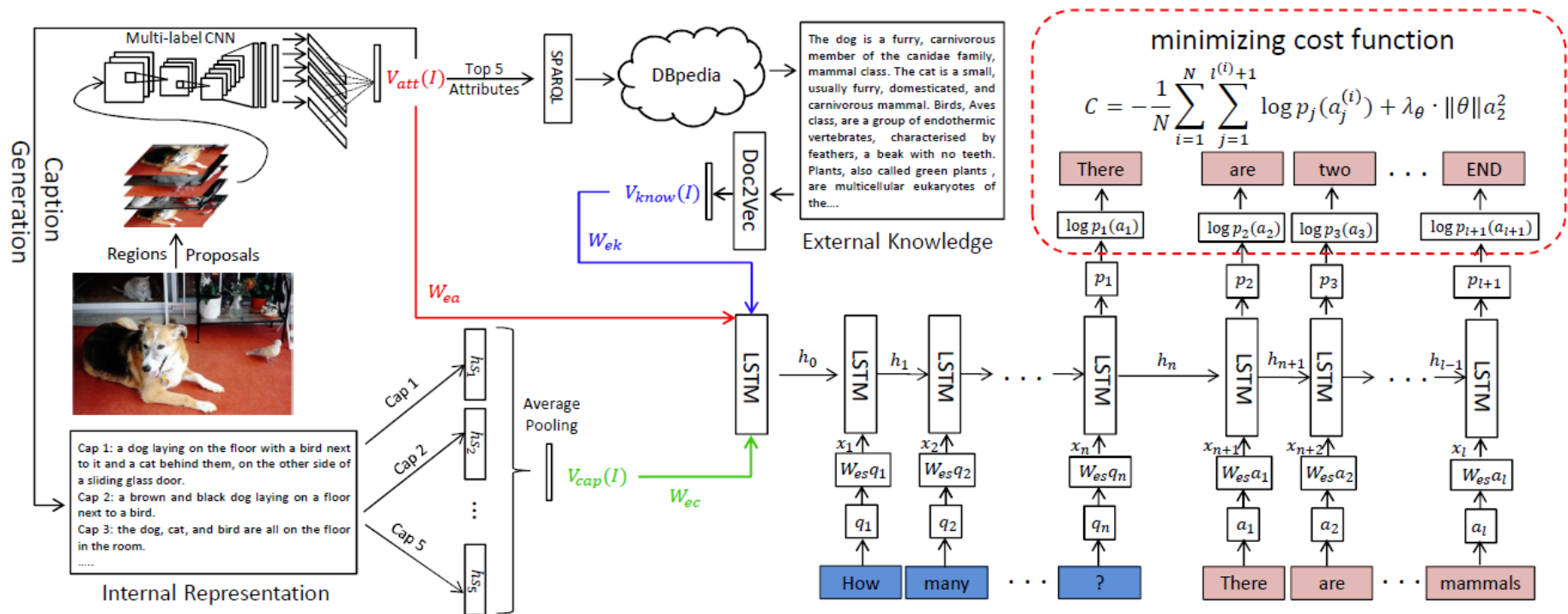
Discover and Learn New Objects from Documentaries

 <p>The elephant are about to march through them. The spiders themselves have a span as wide as a</p>	 <p>But the love serenade is over once a dog arrives.</p>	 <p>Australian camels appear sick and emaciated.</p>
 <p>Tigers are one of the few cats that actually enjoy swimming.</p>	 <p>Male koalas play no role in parenting.</p>	 <p>About 50 animals have died in just three months, including this adult orangutan on the day we</p>
 <p>Unlike mechanics, langurs are the friends of spotted deer.</p>	 <p>There's a turf war going on and the koalas are losing. (dog)</p>	 <p>The mayor has declined offers of assistance and expert advice from animal welfare groups. (elephant)</p>

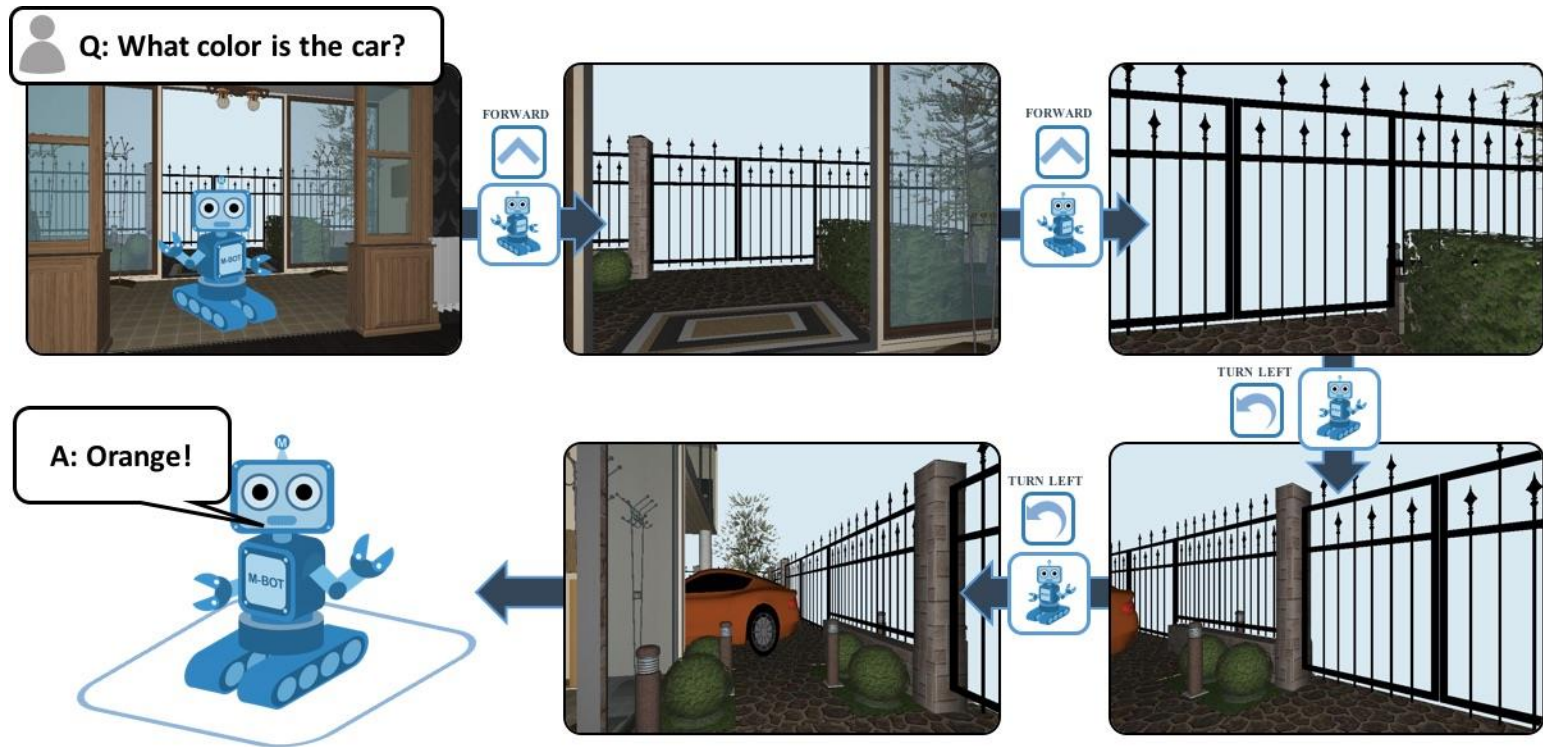
From Red Wine to Red Tomato: Composition With Context



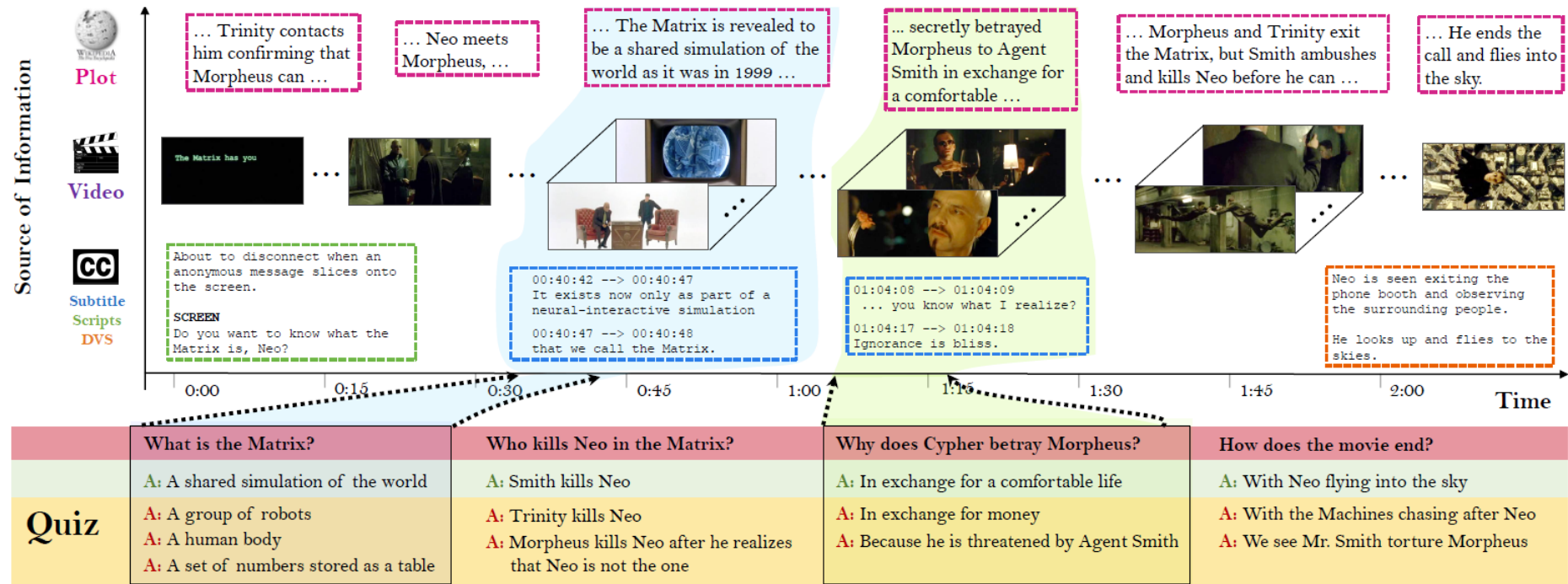
Ask Me Anything: Free-form Visual Question Answering Based on Knowledge from External Sources



Embodied Question Answering



MovieQA: Understanding Stories in Movies through Question-Answering



Automatic Understanding of Image and Video Advertisements

Zaeem Hussain, Mingda Zhang, Xiaozhong Zhang, Keren Ye, Christopher Thomas,
Zuha Agha, Nathan Ong, Adriana Kovashka

University of Pittsburgh



Understanding advertisements is more challenging than simply recognizing physical content from images, as ads employ a variety of strategies to persuade viewers.



Here are some sample annotations in our dataset.



What's being advertised in this image?

Cars, automobiles

What sentiments are provoked in the viewer?

Amused, Creative, Impressed, Youthful, Conscious

What strategies are used to persuade viewer?

Symbolism, Contrast, Straightforward, Transferred qualities

What should the viewer do, and why should they do this?

- I should buy Volkswagen because it can hold a big bear.
- I should buy VW SUV because it can fit anything and everything in it.
- I should buy this car because it can hold everything I need.

More information available at <http://cs.pitt.edu/~kovashka/ads>

We collect an advertisement dataset containing 64,832 images and 3,477 videos, each annotated by 3-5 human workers from Amazon Mechanical Turk.

Image	Topic	204,340	Strategy	20,000
	Sentiment	102,340	Symbol	64,131
	Q+A Pair	202,090	Slogan	11,130
Video	Topic	17,345	Fun/Exciting	15,380
	Sentiment	17,345	English?	17,374
	Q+A Pair	17,345	Effective	16,721

Social Scene Understanding: End-To-End Multi-Person Action Localization and Collective Activity Recognition

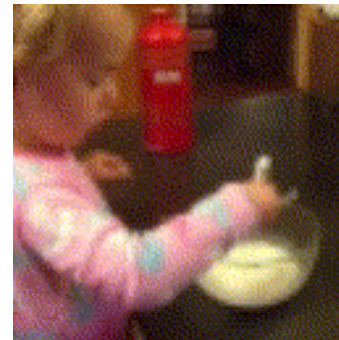
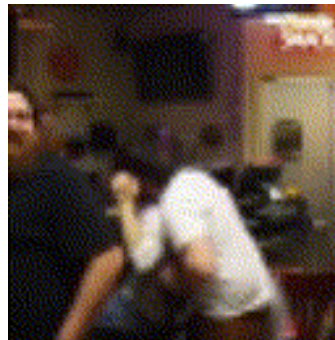
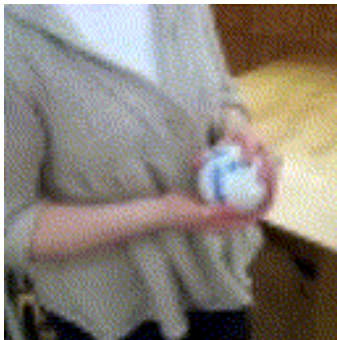
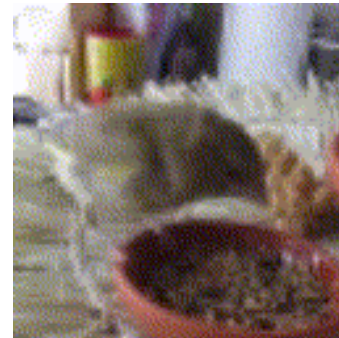


Anticipating Visual Representations from Unlabeled Video

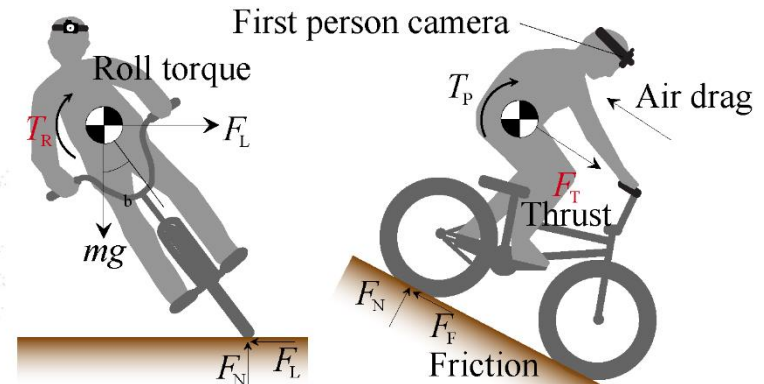
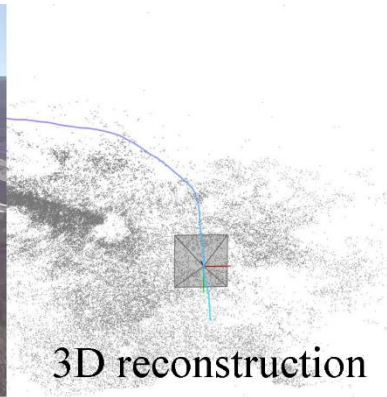


Figure 5: **Example Action Forecasts:** We show some examples of our forecasts of actions one second before they begin. The left most column shows the frame before the action begins, and our forecast is below it. The right columns show the ground truth action. Note that our model does not observe the action frames during inference.

Generating the Future with Adversarial Transformers



Force from Motion: Decoding Physical Sensation from a First Person Video



Scribbler: Controlling Deep Image Synthesis with Sketch and Color

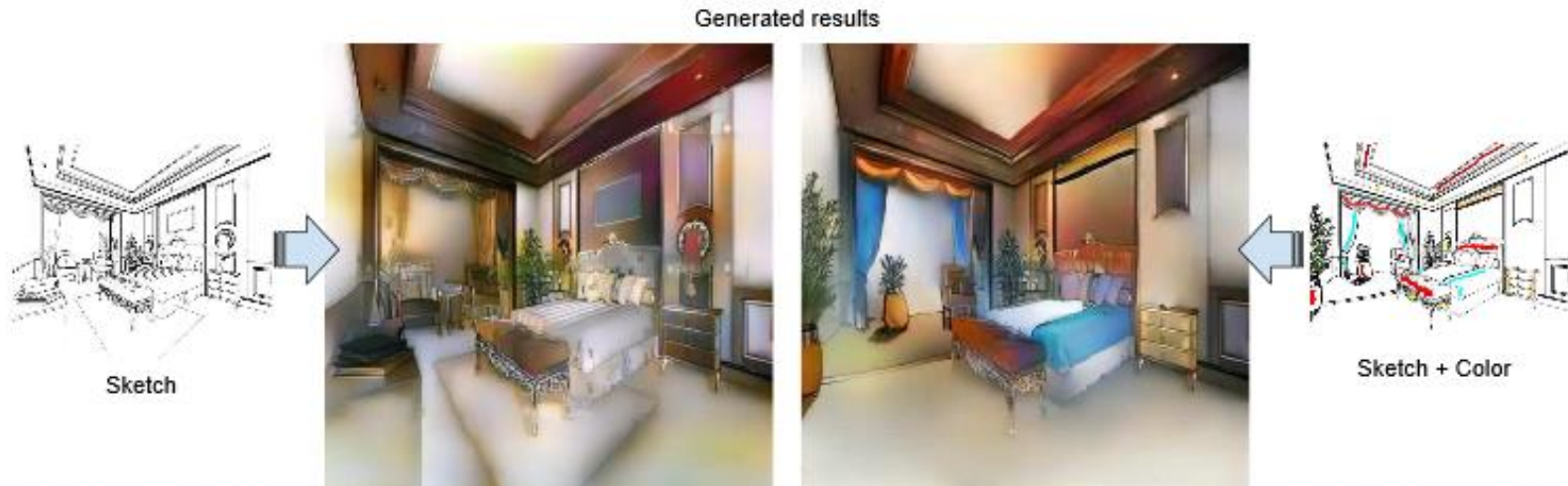
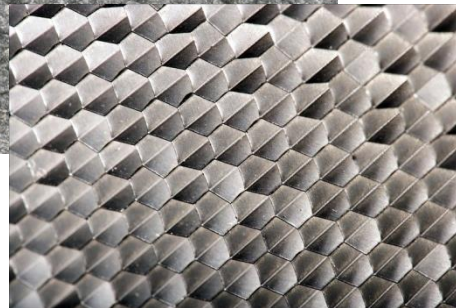
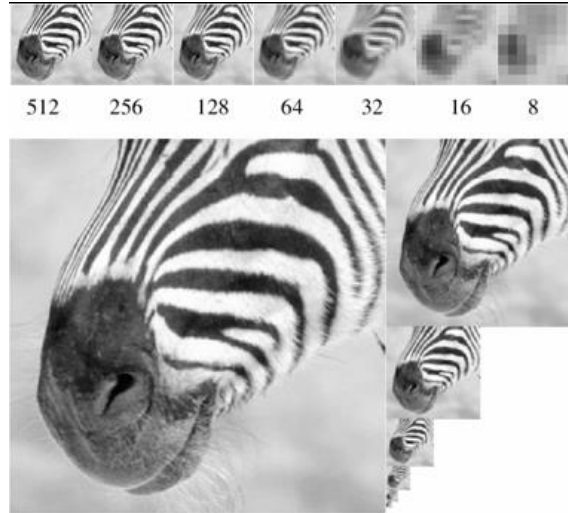


Figure 1. A user can sketch and scribble colors to control deep image synthesis. On the left is an image generated from a hand drawn sketch. On the right several objects have been deleted from the sketch, a vase has been added, and the color of various scene elements has been constrained by sparse color strokes. For best resolution and additional results, see scribbler.eye.gatech.edu

What are we going to talk about?

The Basics

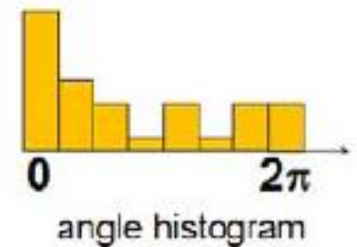
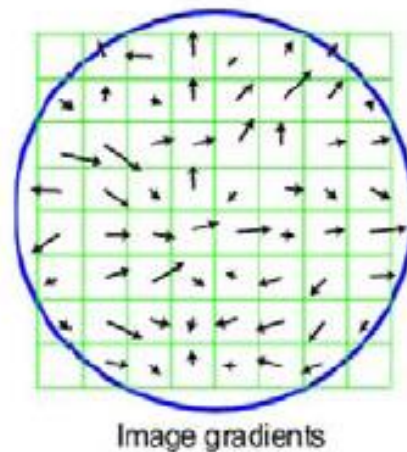
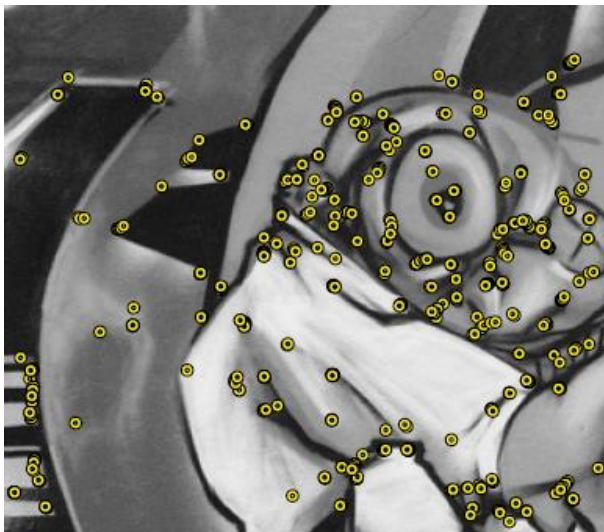
Features and filters



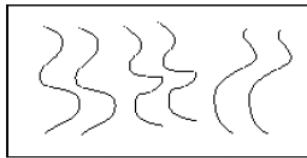
- Transforming and describing images; textures, colors, edges

Features and filters

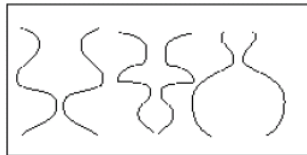
- Detecting distinctive + repeatable features
- Describing images with local statistics



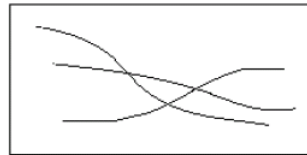
Grouping and fitting



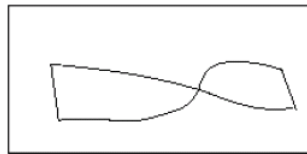
Parallelism



Symmetry



Continuity

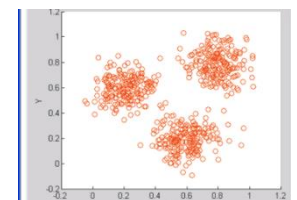
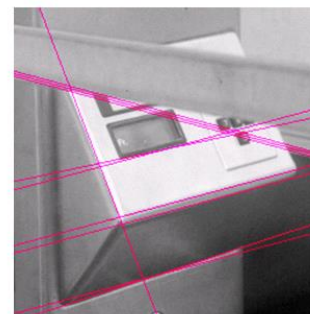


Closure



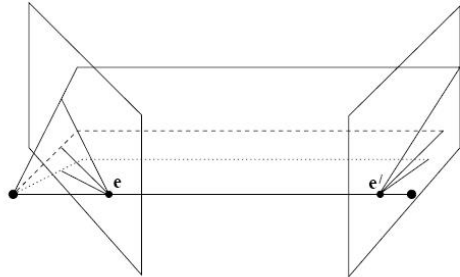
[fig from Shi et al]

- Clustering, segmentation, fitting; what parts belong together?

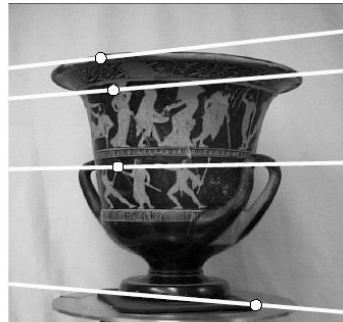


Multiple views

- Multi-view geometry, matching, invariant features, stereo vision



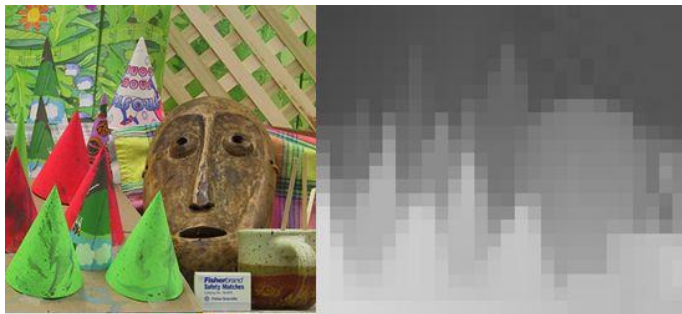
a



Hartley and Zisserman



Lowe



Kristen Grauman



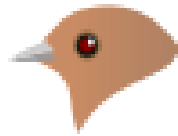
Fei-Fei Li

Image categorization

- Fine-grained recognition



Generalist



Insect catching



Grain eating



Coniferous-seed eating



Nectar feeding



Chiseling



Dip netting



Surface skimming



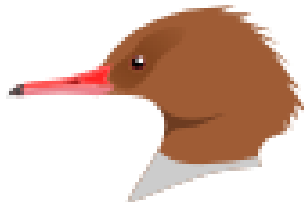
Scything



Probing



Aerial fishing



Pursuit fishing



Scavenging



Raptorial



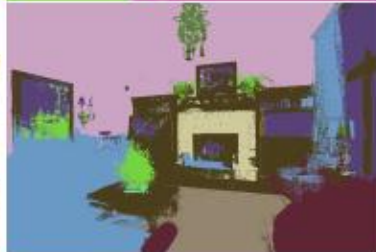
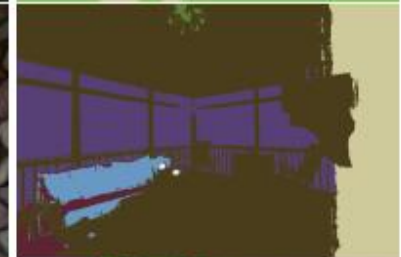
Filter feeding

[Visipedia Project](#)

Image categorization

- Material recognition

brick	food	painted	tile
carpet	glass	paper	stone
ceramic	hair	plastic	water
fabric	leather	polishedstone	wood
foliage	metal	skin	



[[Bell et al. CVPR 2015](#)]

Image categorization

- Image style recognition



HDR



Macro



Baroque



Rococo



Vintage



Noir



Northern Renaissance



Cubism



Minimal



Hazy



Impressionism



Post-Impressionism



Long Exposure



Romantic



Abs. Expressionism



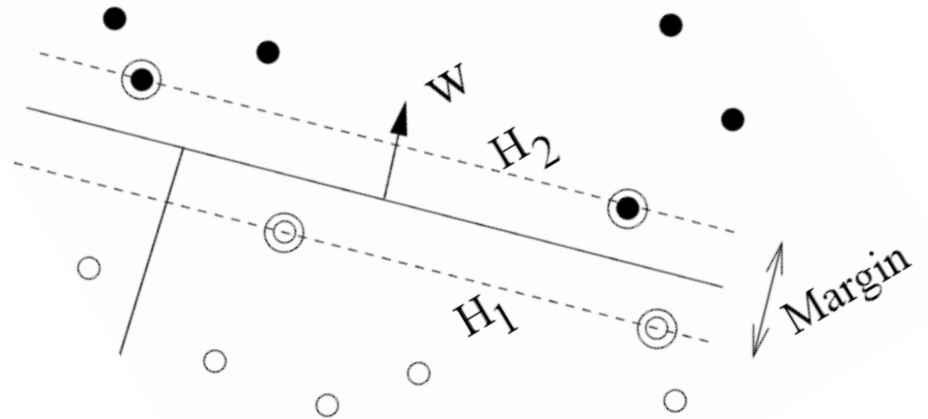
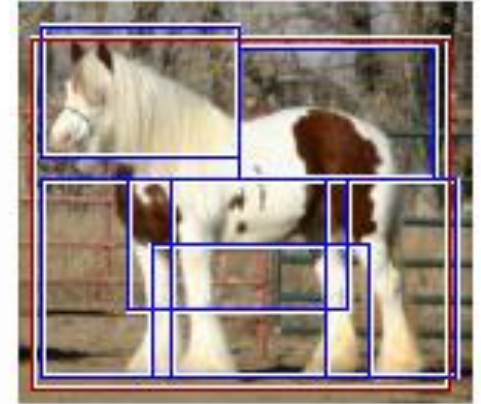
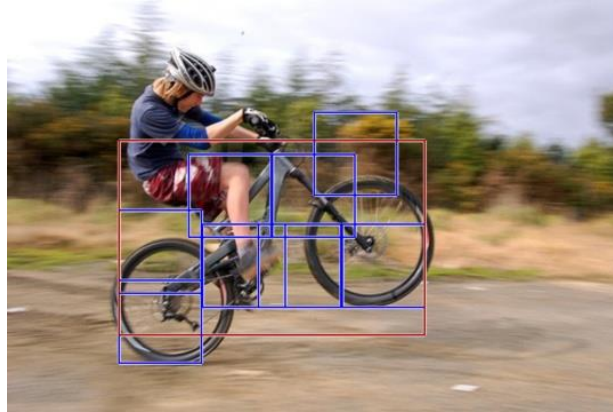
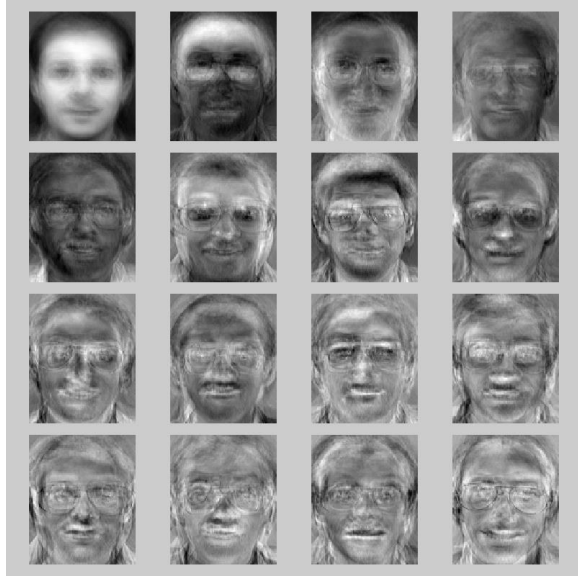
Color Field Painting

Flickr Style: 80K images covering 20 styles.

Wikipaintings: 85K images for 25 art genres.

[[Karayev et al. BMVC 2014](#)]

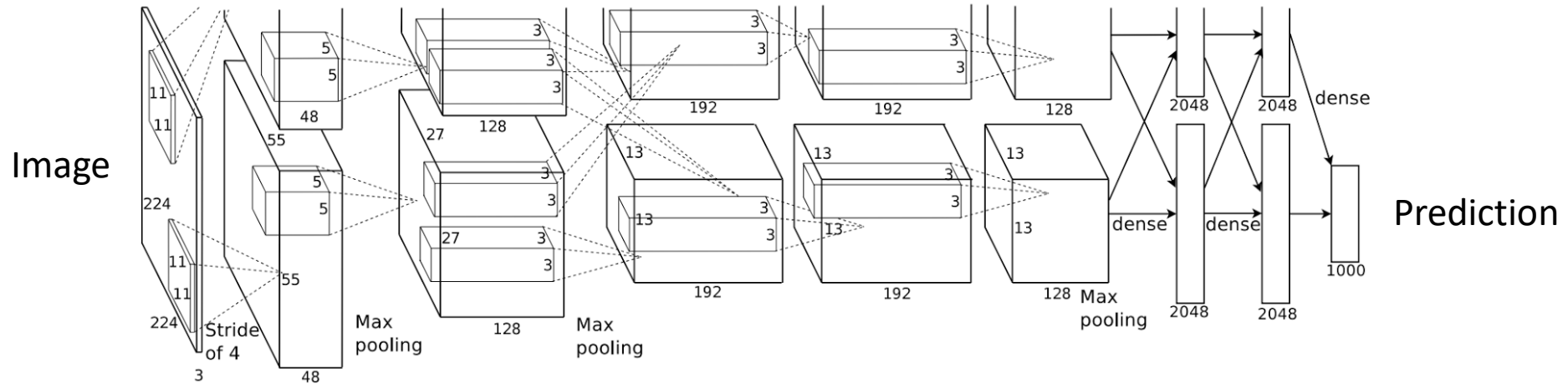
Visual recognition and SVMs



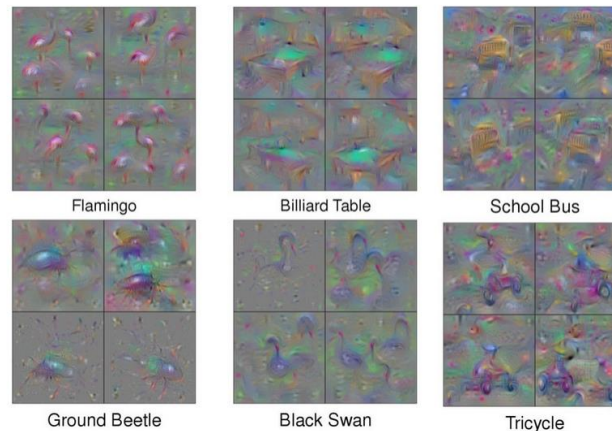
- Recognizing objects and categories, learning techniques

Convolutional neural networks (CNNs)

- State-of-the-art on many recognition tasks



Krizhevsky et al., NIPS 2012

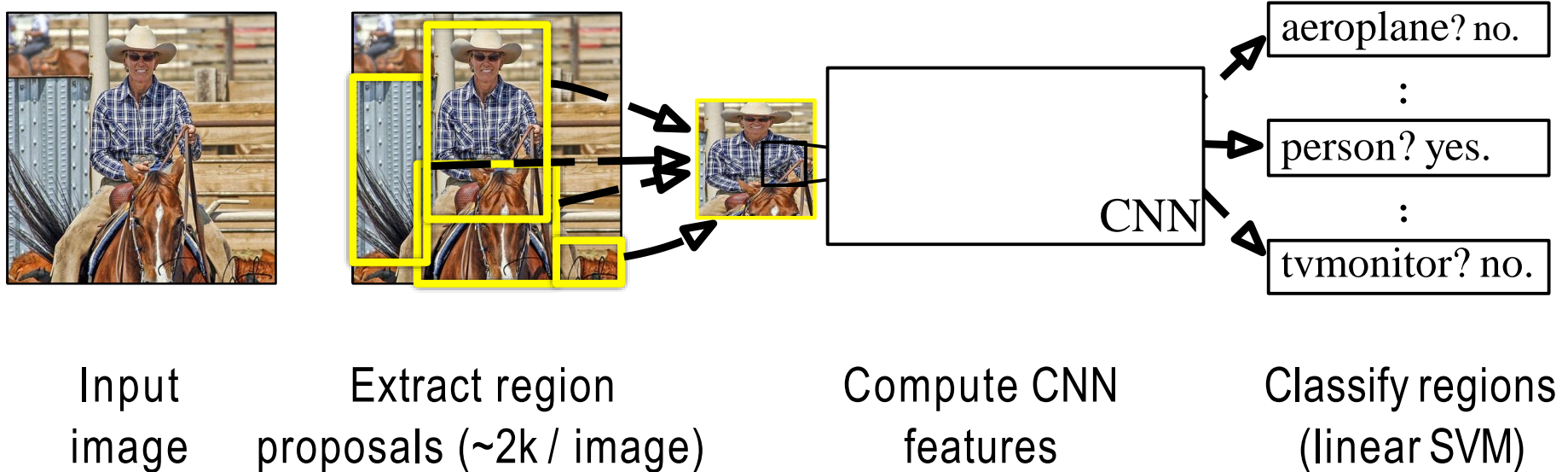


Yosinski et al., ICML DL workshop 2015

The Classics

Object Detection

Regions with CNN features



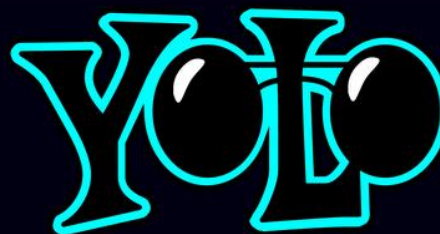
Accurate object detection in real time

	Pascal 2007 mAP	Speed	
DPM v5	33.7	.07 FPS	14 s/img
R-CNN	66.0	.05 FPS	20 s/img
Fast R-CNN	70.0	.5 FPS	2 s/img
Faster R-CNN	73.2	7 FPS	140 ms/img
YOLO	69.0	45 FPS	22 ms/img

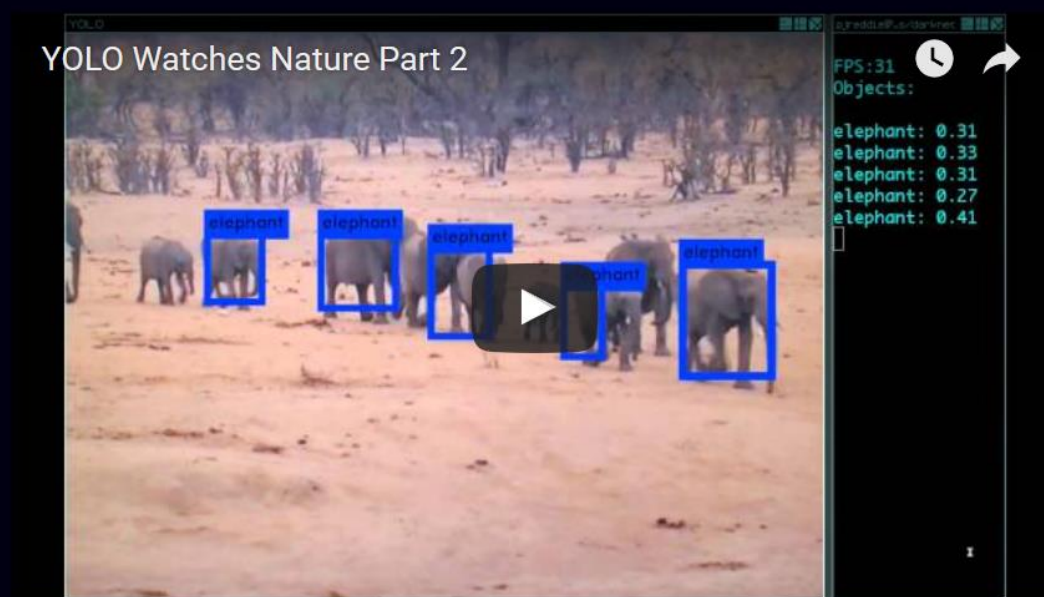


2 feet
→

Our ability to detect objects has gone
from 34 mAP in 2008
to 73 mAP at 7 FPS (frames per second)
or 63 mAP at 45 FPS
in 2016



YOLO: Real-Time Object Detection



You only look once (YOLO) is a system for detecting objects on the
Pascal VOC 2012 dataset. It can detect the 20 Pascal object classes:

- person
- bird, cat, cow, dog, horse, sheep
- aeroplane, bicycle, boat, bus, car, motorbike, train
- bottle, chair, dining table, potted plant, sofa, tv/monitor

Recognition in novel modalities

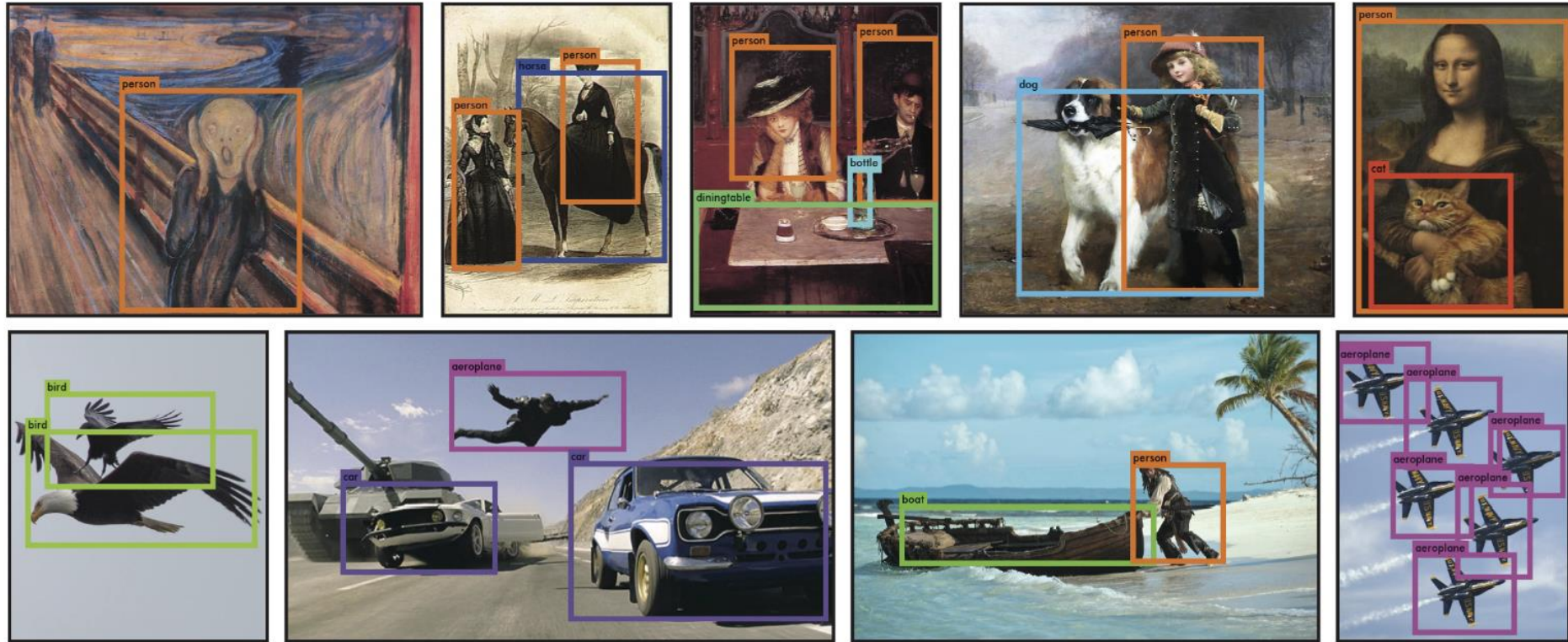


Figure 6: Qualitative Results. YOLO running on sample artwork and natural images from the internet. It is mostly accurate although it does think one person is an airplane.

Vision and Language

Image Captioning

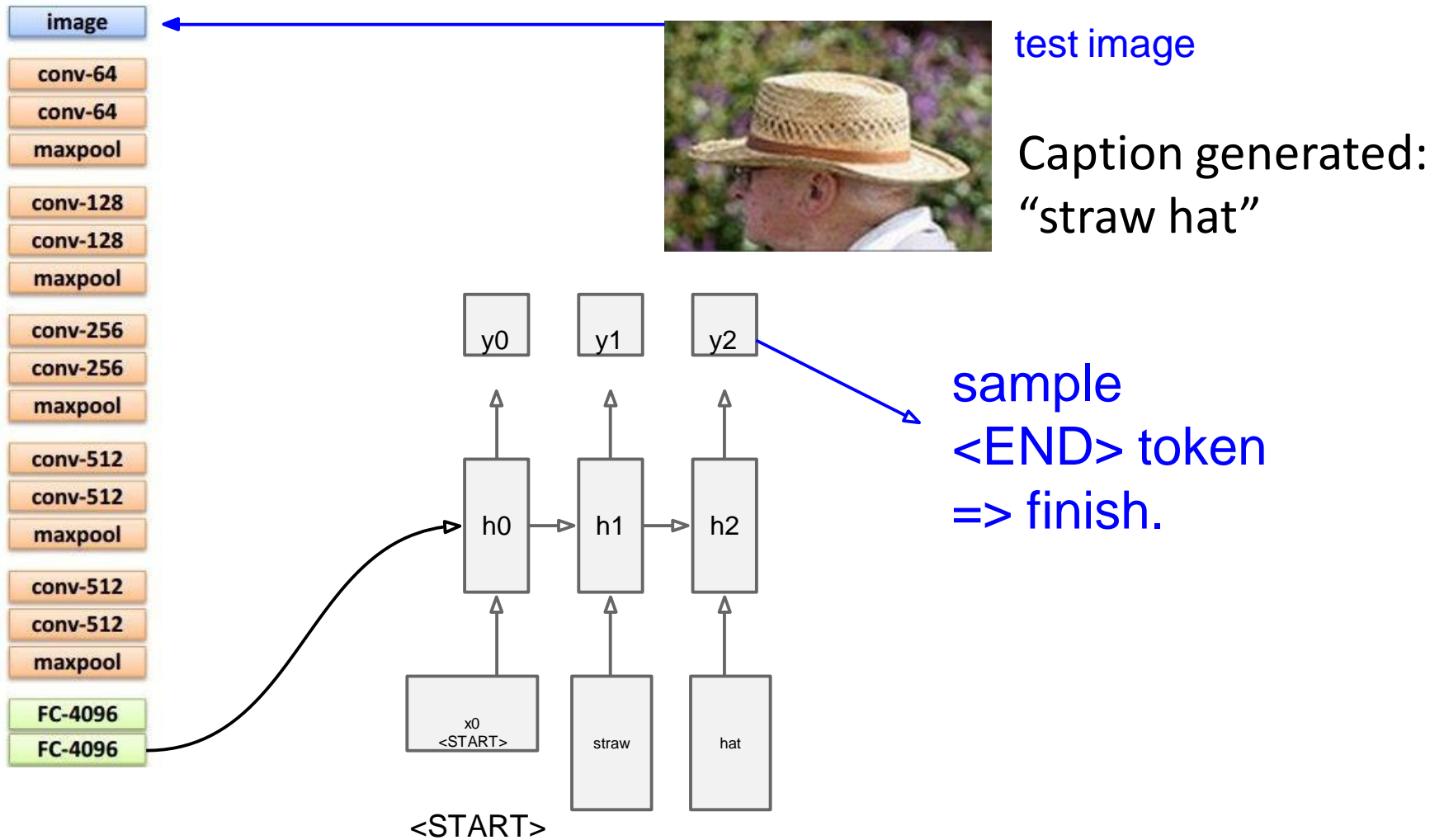


Image Captioning



"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."



"boy is doing backflip on wakeboard."



"a young boy is holding a baseball bat."



"a cat is sitting on a couch with a remote control."



"a woman holding a teddy bear in front of a mirror."



"a horse is standing in the middle of a road."

Visual Question Answering (VQA)

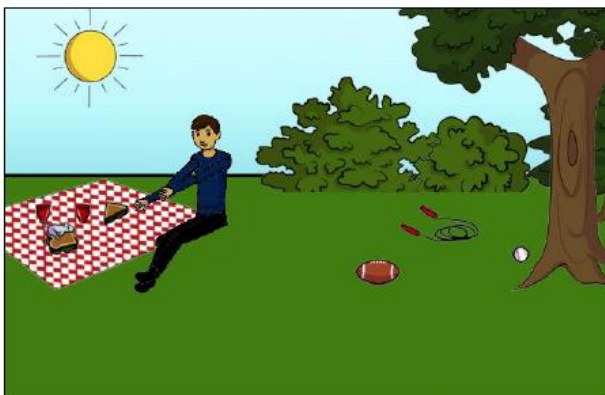
Task: Given an image and a natural language open-ended question, generate a natural language answer.



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?

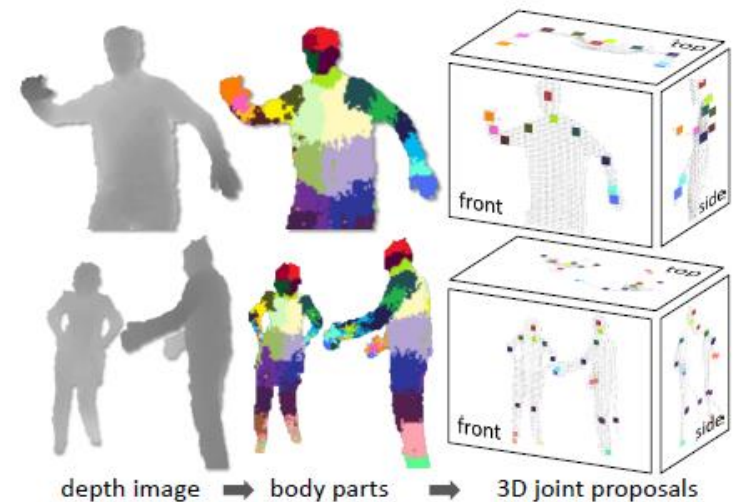
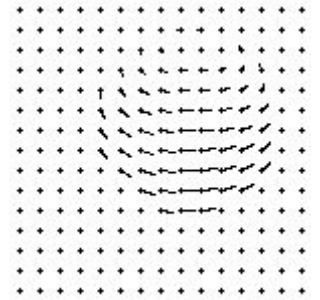


Does it appear to be rainy?
Does this person have 20/20 vision?

Video and Motion

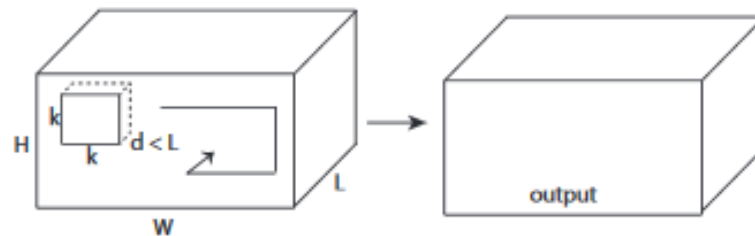
Tracking and Pose

- Tracking objects, video analysis
- Automatically annotating human pose (joints)



Recognizing Actions

- Actions in movies, sports, first-person views



Emergent Topics

Self-Supervised Learning

Context Prediction for Images

1

2

3

4



5



A

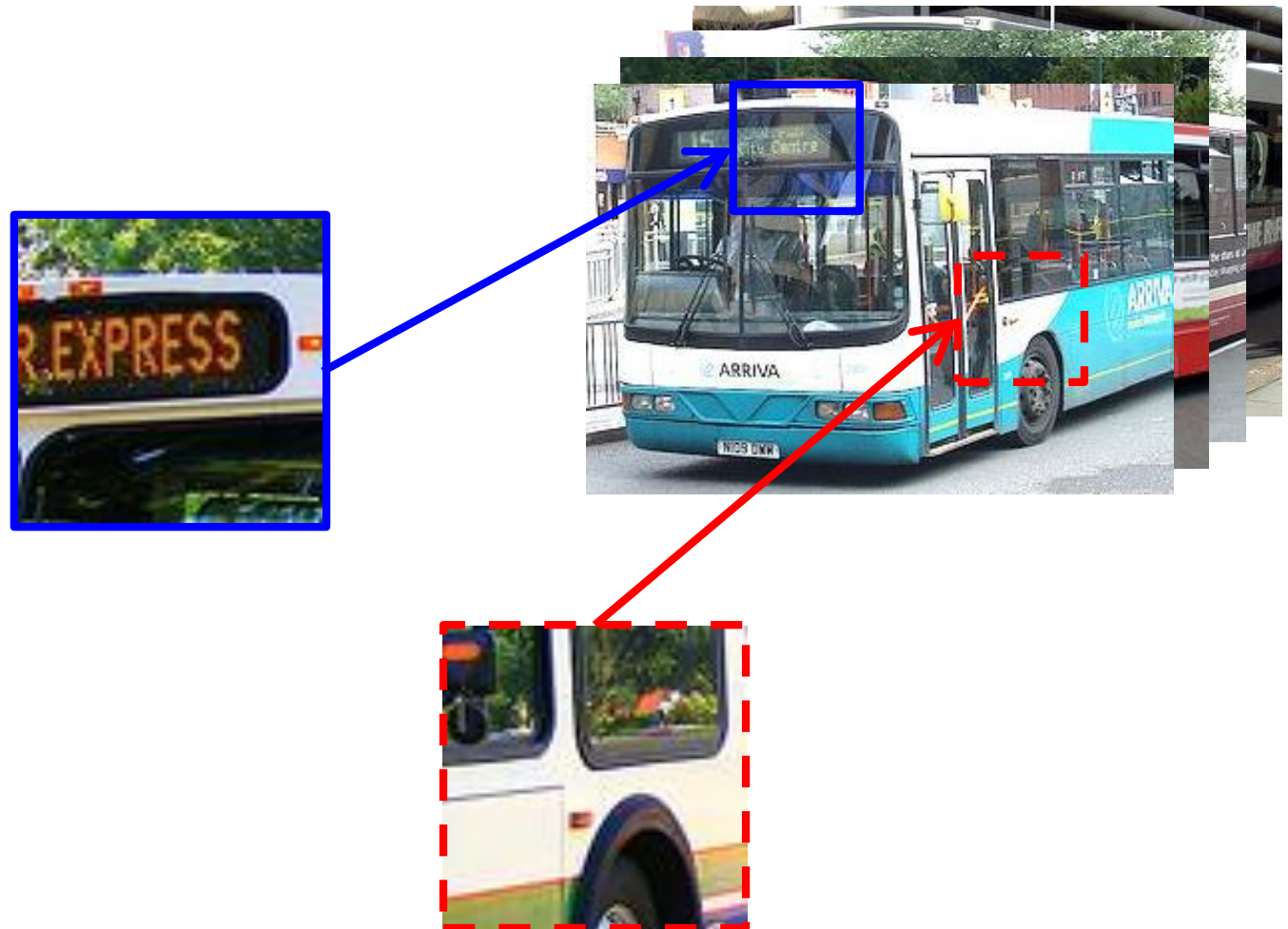
B

6

7

8

Semantics from a non-semantic task



Embodied learning

Status quo: Learn from “disembodied” bag of labeled snapshots.



Goal: Learn in the context of **acting** and **moving** in the world.



Generative Models

Celebrities Who Never Existed

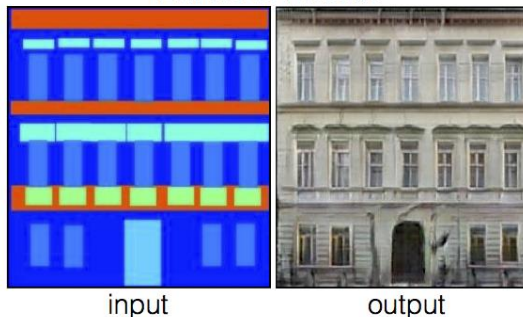


Image-to-Image Translation with Conditional Adversarial Nets

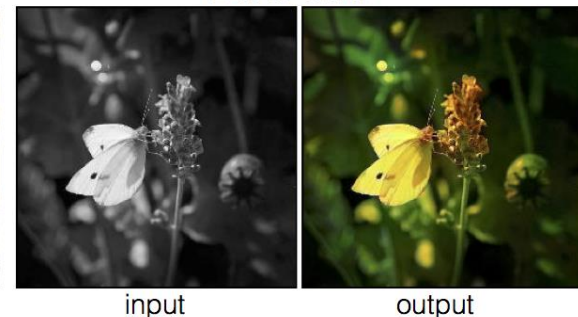
Labels to Street Scene



Labels to Facade



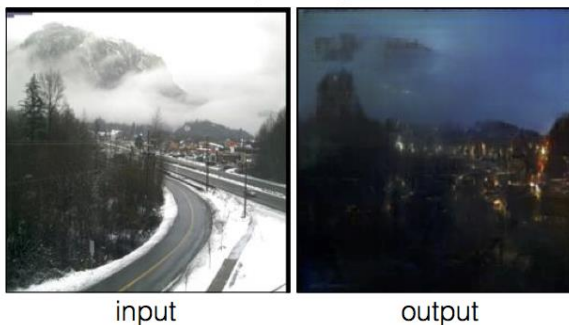
BW to Color



Aerial to Map



Day to Night



Edges to Photo



Is computer vision solved?

- Given an image and a training set in the domain of interest, we can guess what object is shown with nearly perfect accuracy
- But we can *reason* and answer questions about visual data with about 65% accuracy

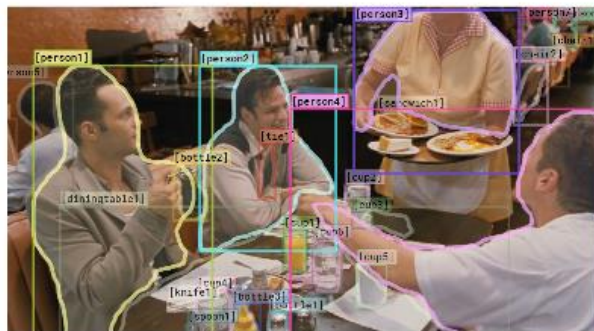
From Recognition to Cognition: Visual Commonsense Reasoning

Rowan Zellers[♦] Yonatan Bisk[♦] Ali Farhadi[♥] Yejin Choi^{♦♥}

[♦]Paul G. Allen School of Computer Science & Engineering, University of Washington

[♥]Allen Institute for Artificial Intelligence

visualcommonsense.com

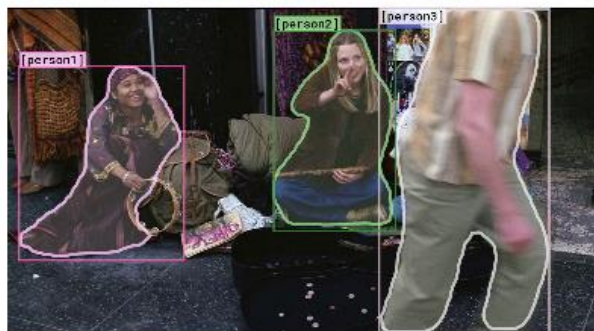


Why is [person4] pointing at [person1]?

- a) He is telling [person3] that [person1] ordered the pancakes.
- b) He just told a joke.
- c) He is feeling accusatory towards [person1].
- d) He is giving [person1] directions.

I chose a)
because...

- a) [person1] has the pancakes in front of him.
- b) [person4] is taking everyone's order and asked for clarification.
- c) [person3] is looking at the pancakes and both she and [person2] are smiling slightly.
- d) [person3] is delivering food to the table, and she might not know whose order is whose.



How did [person2] get the money that's in front of her?

- a) [person2] is selling things on the street.
- b) [person2] earned this money playing music.
- c) She may work jobs for the mafia.
- d) She won money playing poker.

I chose b)
because...

- a) She is playing guitar for money.
- b) [person2] is a professional musician in an orchestra.
- c) [person2] and [person1] are both holding instruments, and were probably busking for that money.
- d) [person1] is putting money in [person2]'s tip jar, while she plays music.

Figure 1: **VCR**: Given an image, a list of regions, and a question, a model must answer the question and provide a *rationale* explaining why its answer is right. Our questions challenge computer vision systems to go beyond recognition-level understanding, towards a higher-order cognitive and commonsense understanding of the world depicted by the image.

Why does it seem like it's solved?

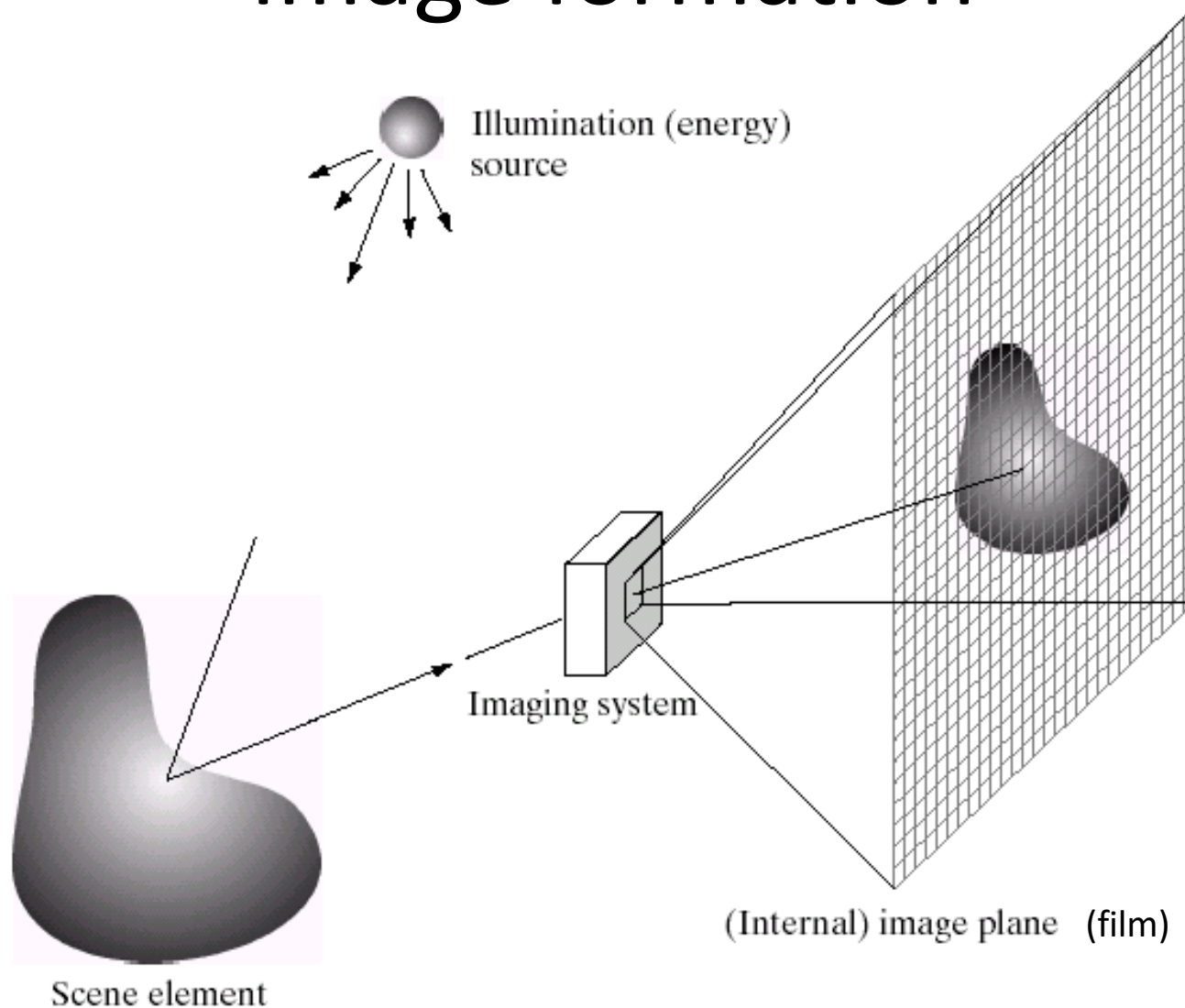
- Deep learning makes excellent use of massive data (labeled for the task of interest)
 - But it doesn't work well without massive data
 - Methods don't easily generalize to new domains
 - Logical inferences, especially ones relying on common sense, are very challenging
 - It's hard to understand how methods work

Linear Algebra Review

What are images?

- Matlab and Python/NumPy treat images as matrices of numbers
- To proceed, let's talk very briefly about how images are formed

Image formation



Digital camera

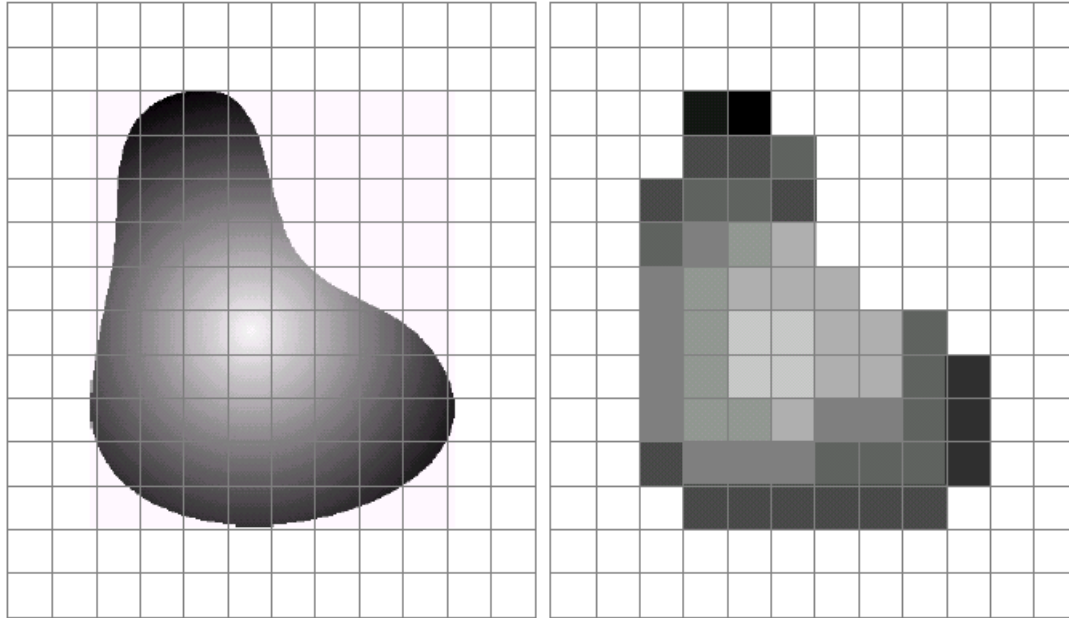


A digital camera replaces film with a sensor array

- Each cell in the array is light-sensitive diode that converts photons to electrons

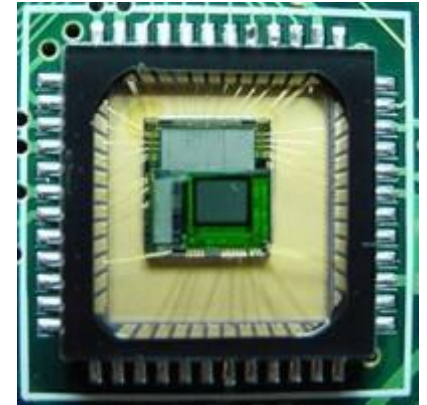
<http://electronics.howstuffworks.com/cameras-photography/digital/digital-camera.htm>

Digital images



a b

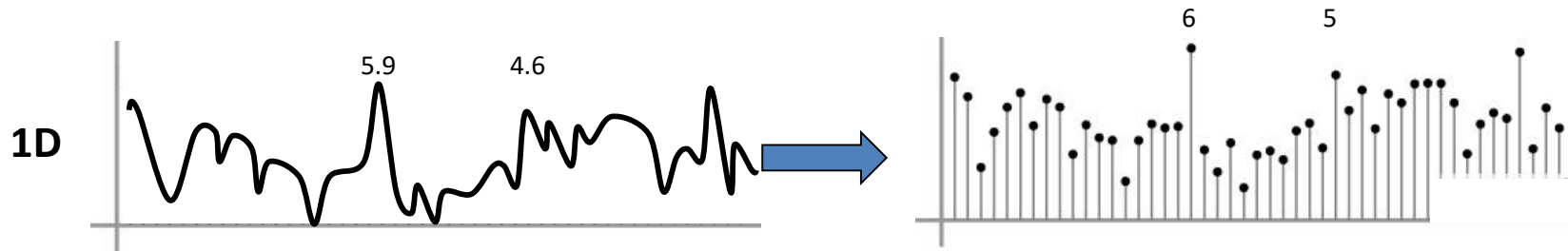
FIGURE 2.17 (a) Continuous image projected onto a sensor array. (b) Result of image sampling and quantization.



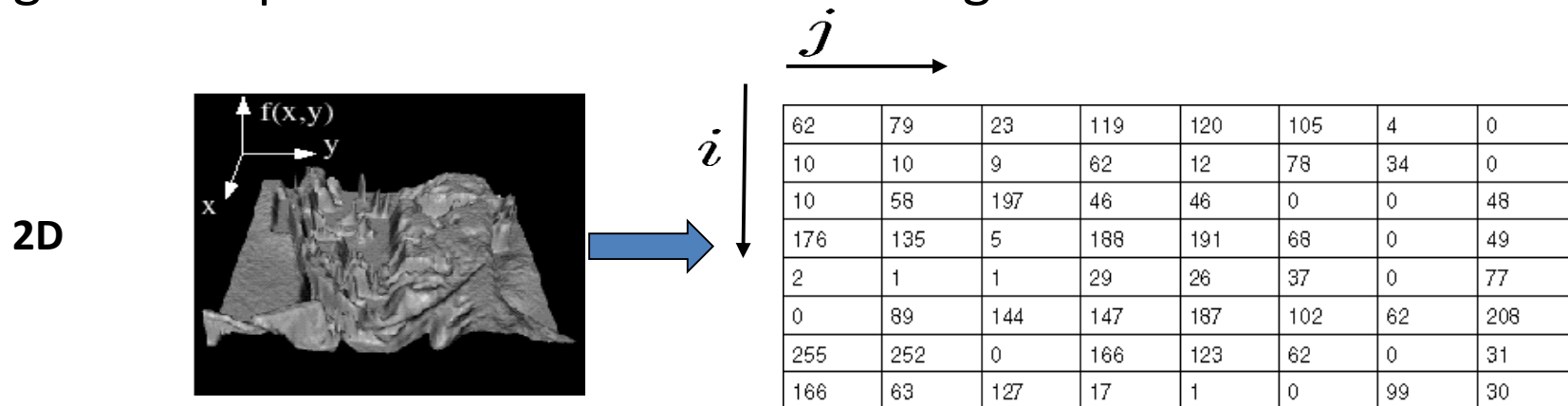
- **Sample** the 2D space on a regular grid
- **Quantize** each sample (round to nearest integer)

Digital images

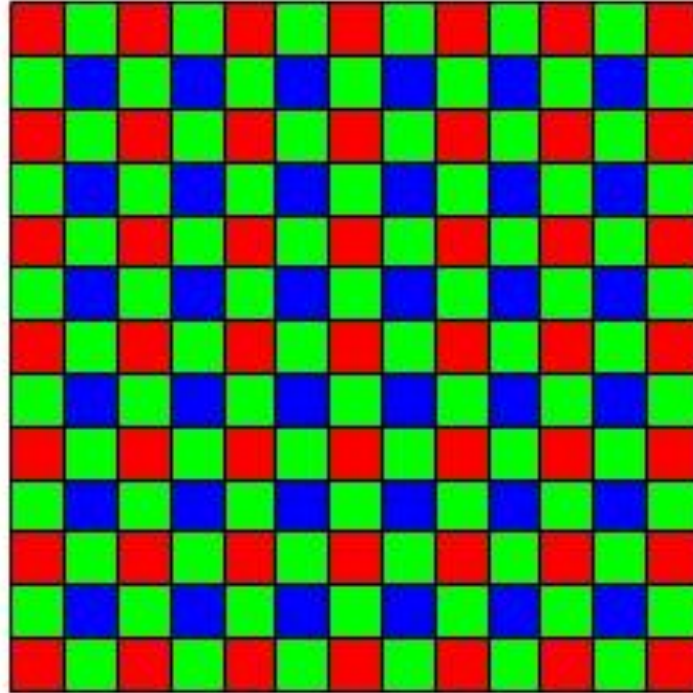
- **Sample** the 2D space on a regular grid
- **Quantize** each sample (round to nearest integer)
- What does quantizing signal look like?



- Image thus represented as a matrix of integer values.



Digital color images



Bayer filter

© 2000 How Stuff Works

Digital color images

Color images,
RGB color space:

Split image into
three channels



R



G



B

Images in Matlab

- Color images represented as a matrix with multiple channels (=1 if grayscale)
- Suppose we have a NxM RGB image called “im”
 - `im(1,1,1)` = top-left pixel value in R-channel
 - `im(y, x, b)` = y pixels **down**, x pixels **to right** in the bth channel
 - `im(N, M, 3)` = bottom-right pixel in B-channel
- `imread(filename)` returns a uint8 image (values 0 to 255)
 - Convert to double format with `double` or `im2double`

row

column

</

Vectors and Matrices

- Vectors and matrices are just collections of ordered numbers that represent something: movements in space, scaling factors, word counts, movie ratings, pixel brightnesses, etc.
- We'll define some common uses and standard operations on them.

Vector

- A column vector $\mathbf{v} \in \mathbb{R}^{n \times 1}$ where

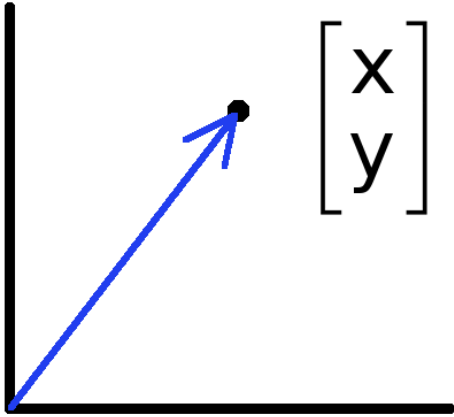
$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

- A row vector $\mathbf{v}^T \in \mathbb{R}^{1 \times n}$ where

$$\mathbf{v}^T = [v_1 \quad v_2 \quad \dots \quad v_n]$$

T denotes the transpose operation

Vectors have two main uses



- Vectors can represent an offset in 2D or 3D space
- Points are just vectors from the origin
- Data can also be treated as a vector
- Such vectors don't have a geometric interpretation, but calculations like "distance" still have value

Matrix

- A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is an array of numbers with size $m \downarrow$ by $n \rightarrow$, i.e. m rows and n columns.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

- If $m = n$, we say that \mathbf{A} is square.

Matrix Operations

- Addition

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} a + 1 & b + 2 \\ c + 3 & d + 4 \end{bmatrix}$$

- Can only add matrices with matching dimensions, or a scalar to a matrix.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + 7 = \begin{bmatrix} a + 7 & b + 7 \\ c + 7 & d + 7 \end{bmatrix}$$

- Scaling

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \times 3 = \begin{bmatrix} 3a & 3b \\ 3c & 3d \end{bmatrix}$$

Matrix Operations

- Inner product (*dot* · product) of vectors
 - Multiply corresponding entries of two vectors and add up the result
 - We won't worry about the geometric interpretation for now

$$\mathbf{x}^T \mathbf{y} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i \quad (\text{scalar})$$

Inner vs outer vs matrix vs element-wise product

- \mathbf{x}, \mathbf{y} = column vectors ($n \times 1$)
- \mathbf{X}, \mathbf{Y} = matrices ($m \times n$)
- x, y = scalars (1×1)

- $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$ = inner product ($1 \times n \times n \times 1 = \text{scalar}$)
- $\mathbf{x} \otimes \mathbf{y} = \mathbf{x} \mathbf{y}^T$ = outer product ($n \times 1 \times 1 \times n = \text{matrix}$)

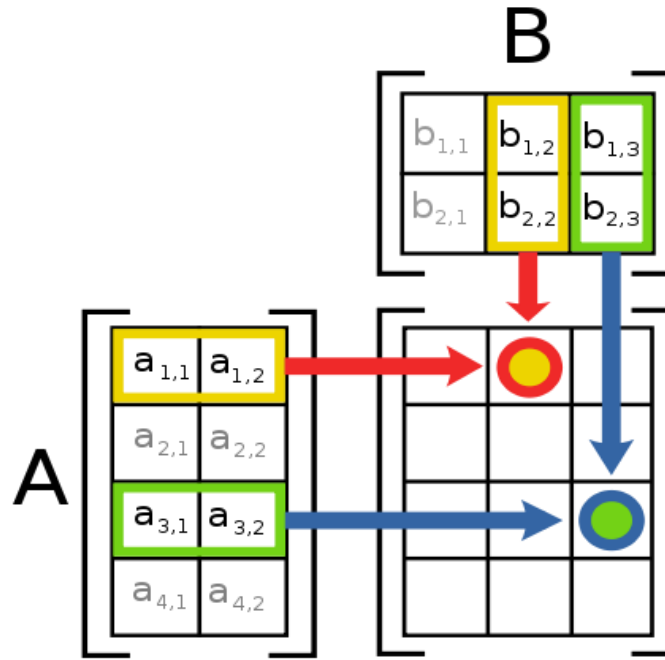
- $\mathbf{X} * \mathbf{Y}$ = matrix product
- $\mathbf{X} .* \mathbf{Y}$ = element-wise product

Matrix Multiplication

- Let X be an $a \times b$ matrix, Y be an $b \times c$ matrix
- Then $Z = X * Y$ is an $a \times c$ matrix
- Second dimension of first matrix, and first dimension of second matrix have to be the same, for matrix multiplication to be possible
- Practice: Let X be an 10×5 matrix. Let's factorize it into 3 matrices...

Matrix Operations

- Multiplication
- The product AB is:



- Each entry in the result is (that row of A) dot product with (that column of B)

Matrix Operations

- Multiplication example:

Diagram illustrating the dot product of two vectors A and B .

Vector A is $\begin{bmatrix} 0 & 2 \end{bmatrix}$ and Vector B is $\begin{bmatrix} 1 & 5 \end{bmatrix}$.

The dot product calculation is shown as:

$$\begin{bmatrix} 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 5 \end{bmatrix} = 0 \cdot 1 + 2 \cdot 5 = 10$$

The result 10 is highlighted in red.

$$0 \cdot 3 + 2 \cdot 7 = 14$$

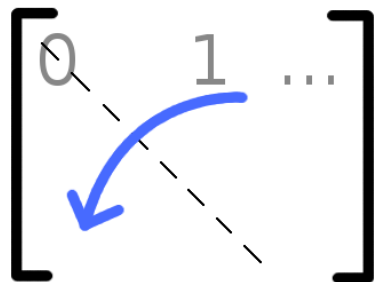
- Each entry of the matrix product is made by taking the dot product of the corresponding row in the left matrix, with the corresponding column in the right one.

Matrix Operation Properties

- Matrix addition is commutative and associative
 - $A + B = B + A$
 - $A + (B + C) = (A + B) + C$
- Matrix multiplication is associative and distributive but *not* commutative
 - $A(B * C) = (A * B)C$
 - $A(B + C) = A * B + A * C$
 - $A * B \neq B * A$

Matrix Operations

- Transpose – flip matrix, so row 1 becomes column 1


$$\begin{bmatrix} 0 & 1 & \dots \\ 2 & 3 & \\ 4 & 5 & \end{bmatrix}^T = \begin{bmatrix} 0 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix}$$

- A useful identity:

$$(ABC)^T = C^T B^T A^T$$

Special Matrices

- Identity matrix \mathbf{I}
 - Square matrix, 1's along diagonal, 0's elsewhere
 - $\mathbf{I} \cdot [\text{another matrix}] = [\text{that matrix}]$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Diagonal matrix
 - Square matrix with numbers along diagonal, 0's elsewhere
 - A diagonal \cdot [another matrix] scales the rows of that matrix

$$\begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 2.5 \end{bmatrix}$$

Norms

- L1 norm

$$\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$$

- L2 norm

$$\|\mathbf{x}\| := \sqrt{x_1^2 + \cdots + x_n^2}$$

- L^p norm (for real numbers $p \geq 1$)

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}$$

Your Homework

- Read entire course website
- Do first reading
- Fill out Doodle for TA's office hours
- Sign up for Piazza
- Start thinking about your project!