Object Recognition II

Computer Vision
Fall 2019
Columbia University

Homework Grades

- Homework 1: Median=94/100, Mean=89/100
- Homework 2: Median=100/100, Mean=94/100
- For regrades: Please post private message on Piazza. **Include your uni.** When we regrade, your grade may go up or down.

Final Course Grades

- If you score 90 and above, we will guarantee you at least a A-
- If you score 80 and above, we will guarantee you at least a B-
- If you score 70 and above, we will guarantee you at least a C-
- We will also curve the course, and give you the best grade from the curve or flat score.
- Extra credit is applied after the curve.

Quick Experiment

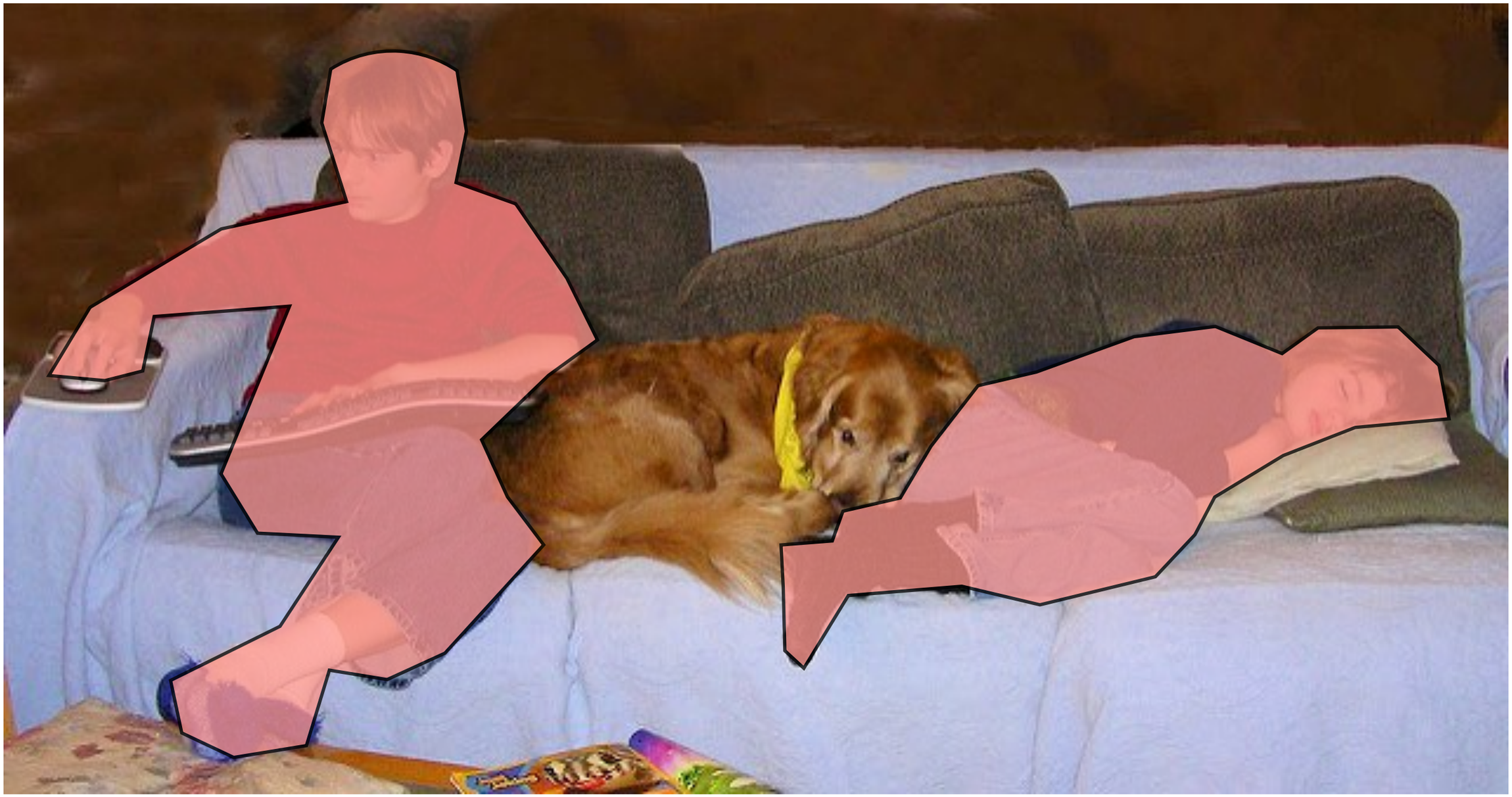
- Get pen and paper
- Draw a coffee cup

Canonical Perspective

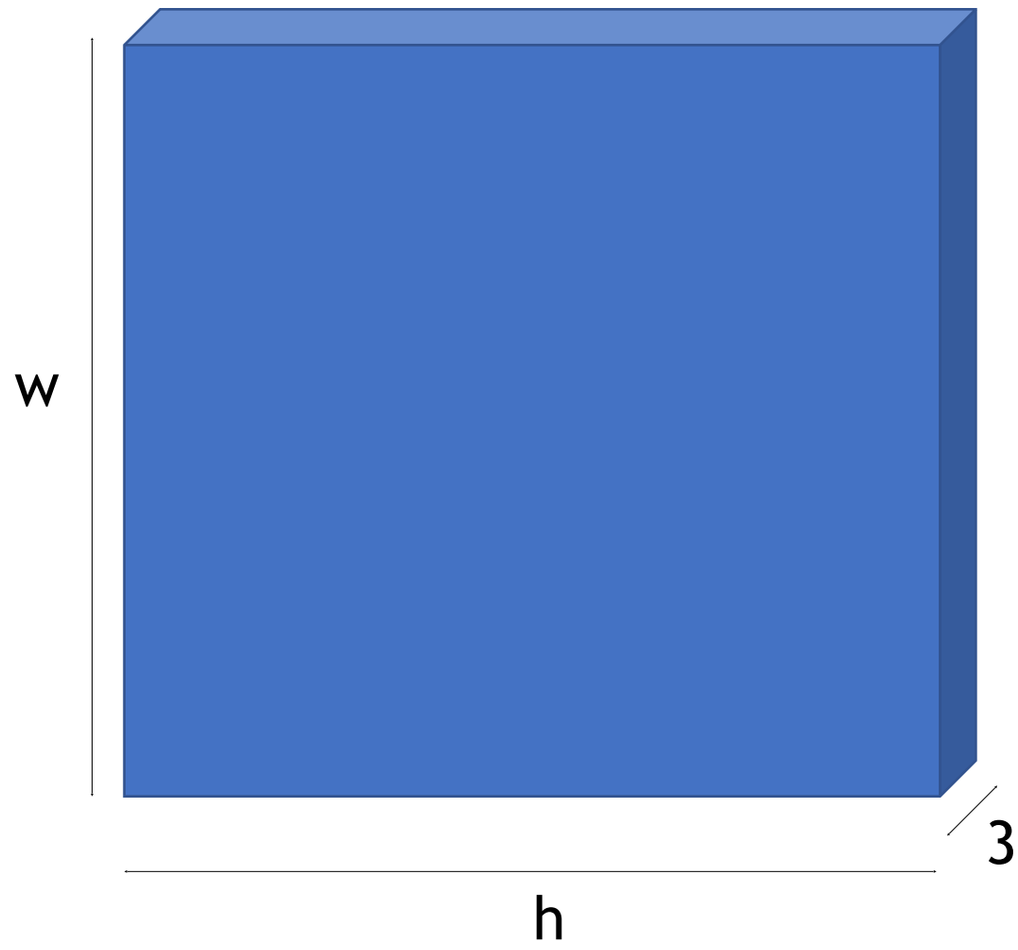
The “best,” most easily identified view of an object.
(Palmer, Rosch & Chase, 1981)

Slide by Palmer

Segmentation:
Where ***really*** are the people?

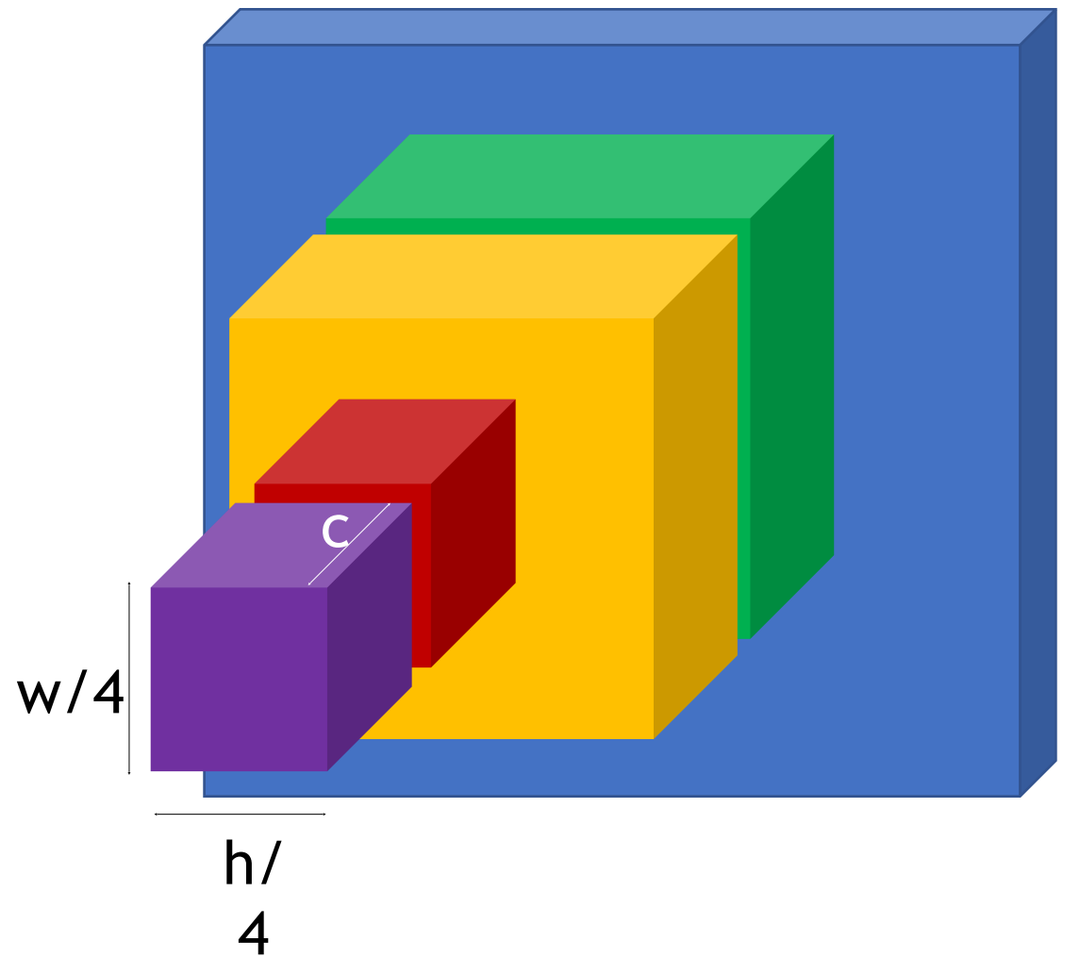


Semantic segmentation using convolutional networks



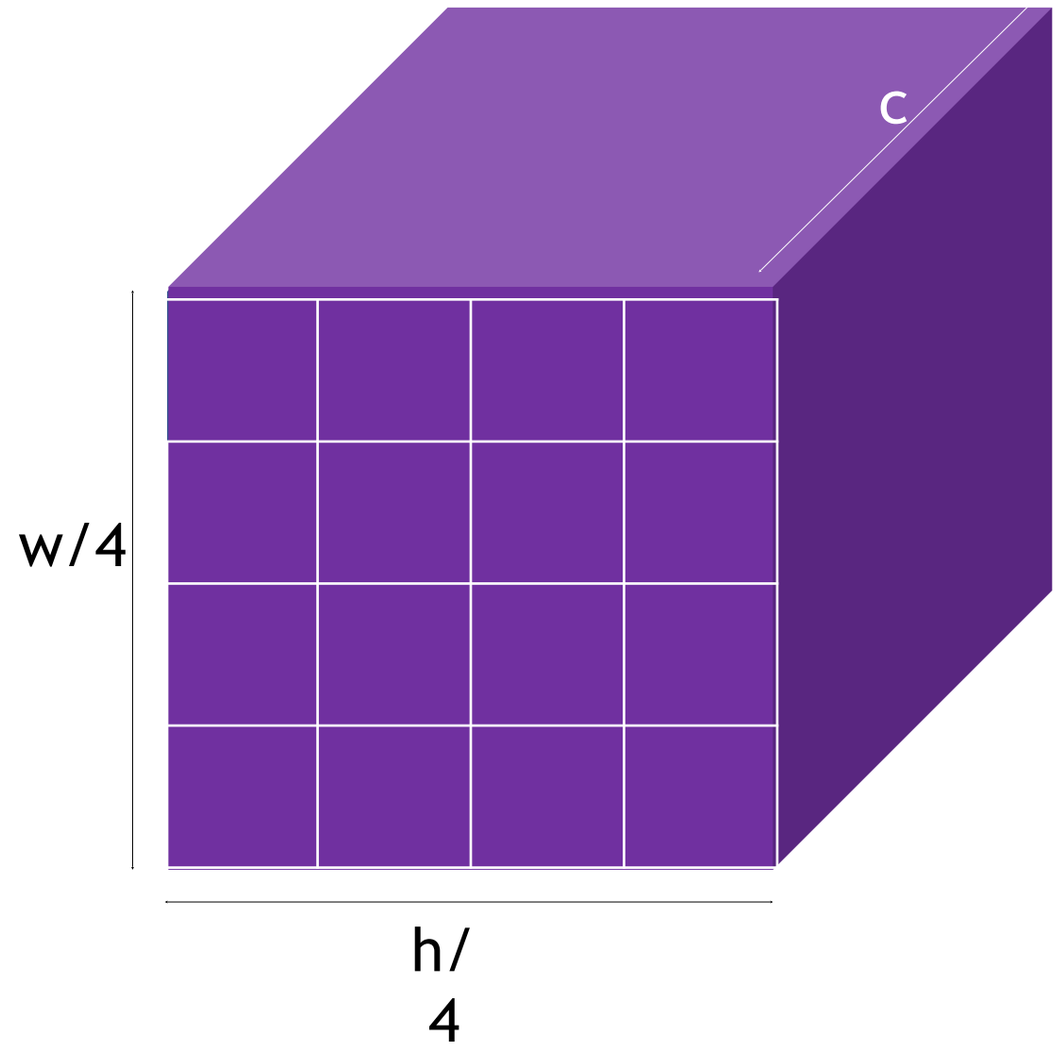
Slide credit: Bharath Hariharan

Semantic segmentation using convolutional networks



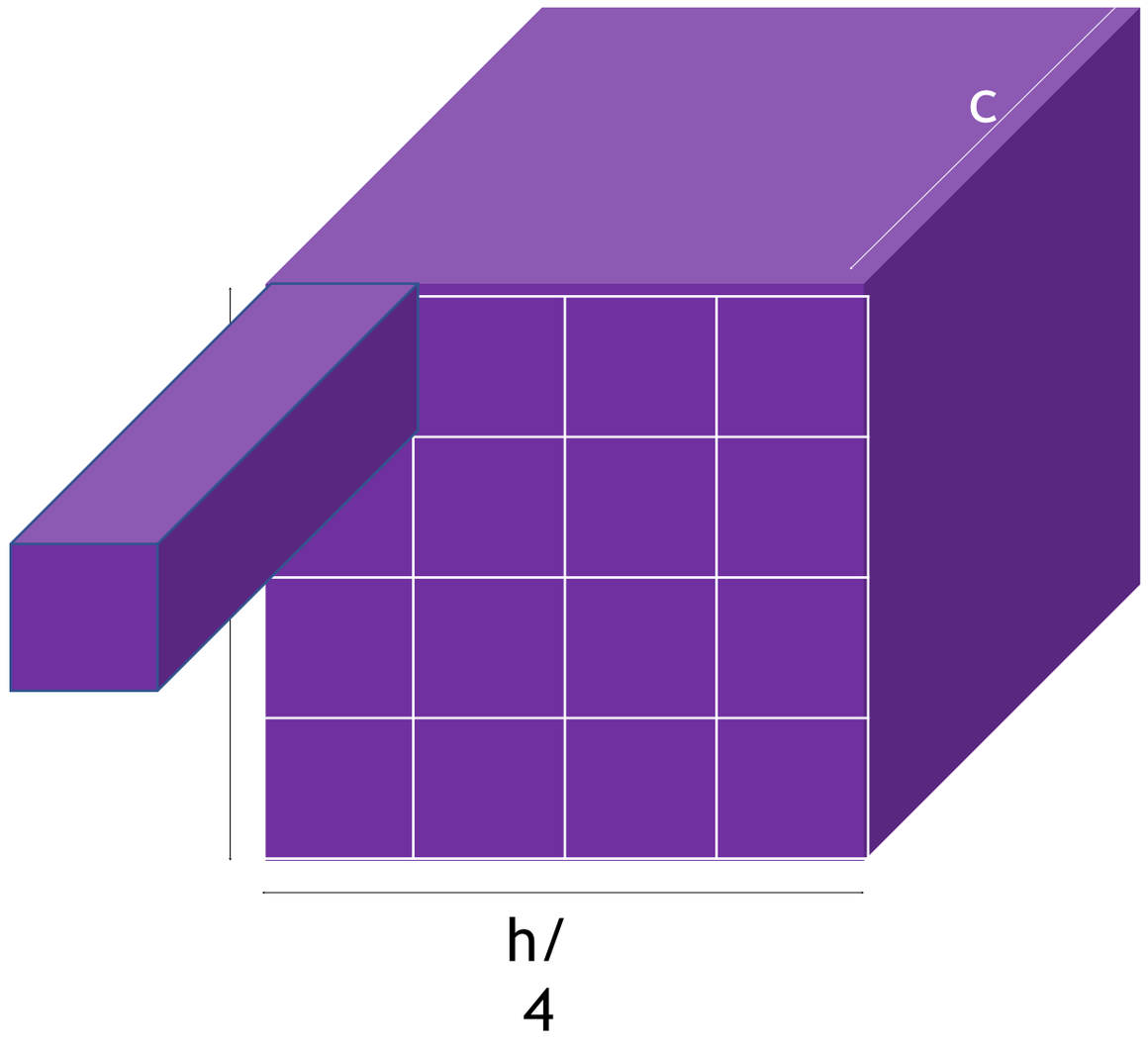
Slide credit: Bharath Hariharan

Semantic segmentation using convolutional networks



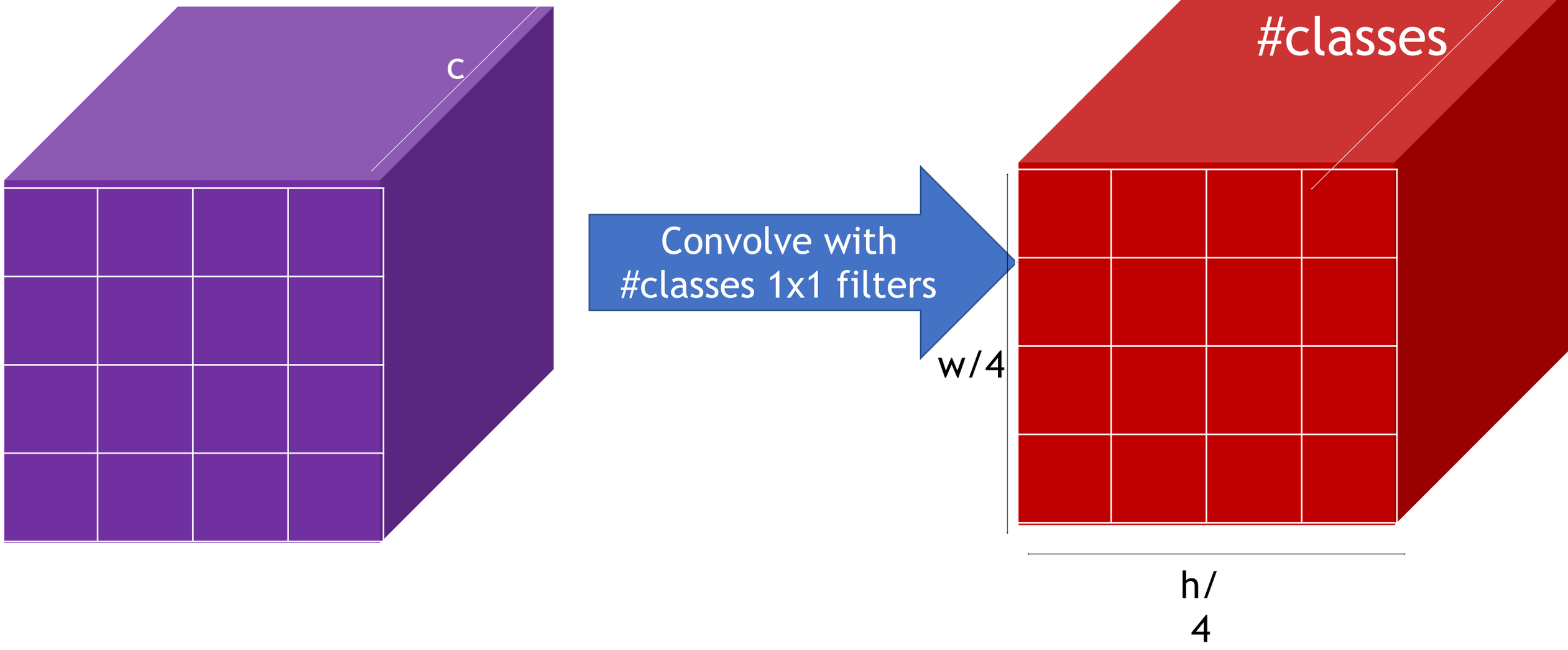
Slide credit: Bharath Hariharan

Semantic segmentation using convolutional networks



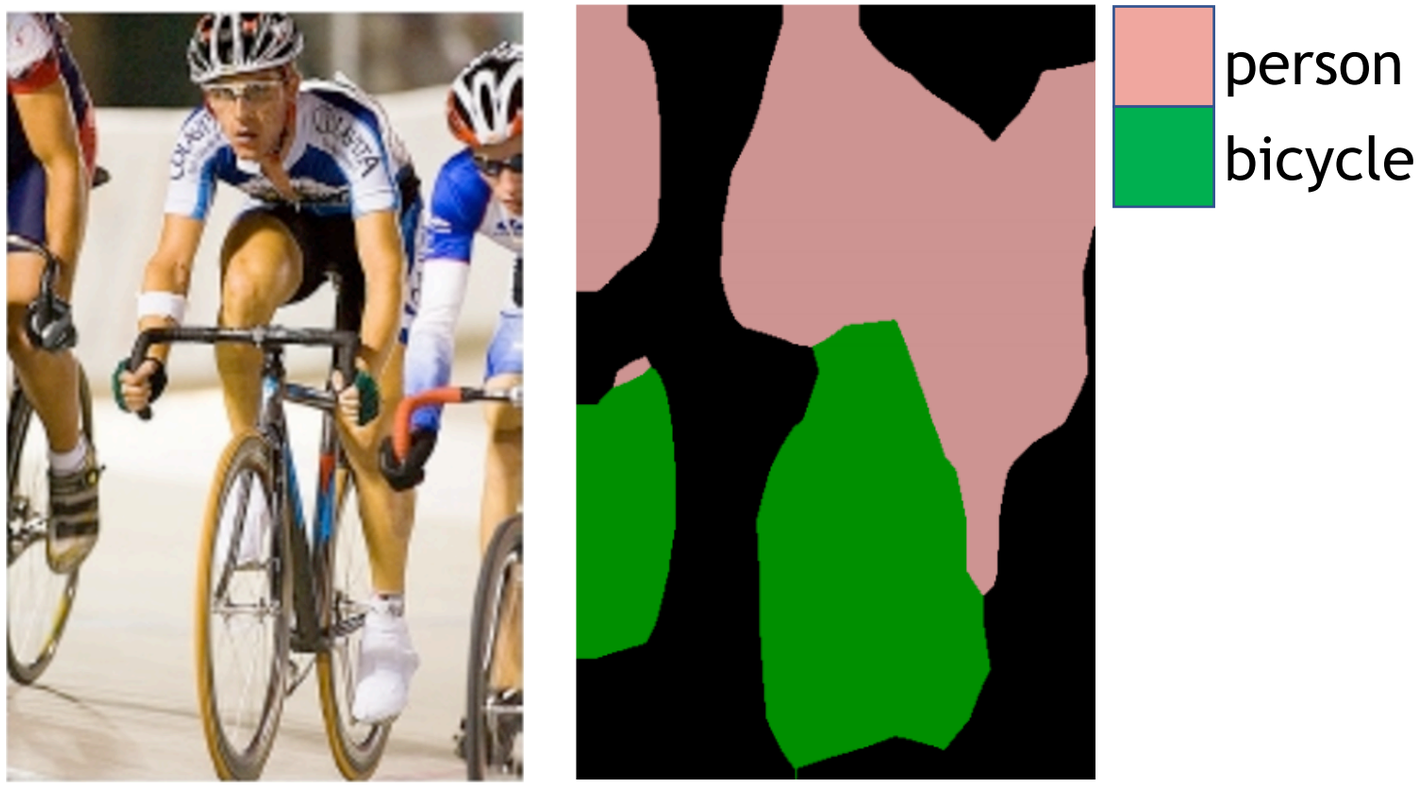
Slide credit: Bharath Hariharan

Semantic segmentation using convolutional networks



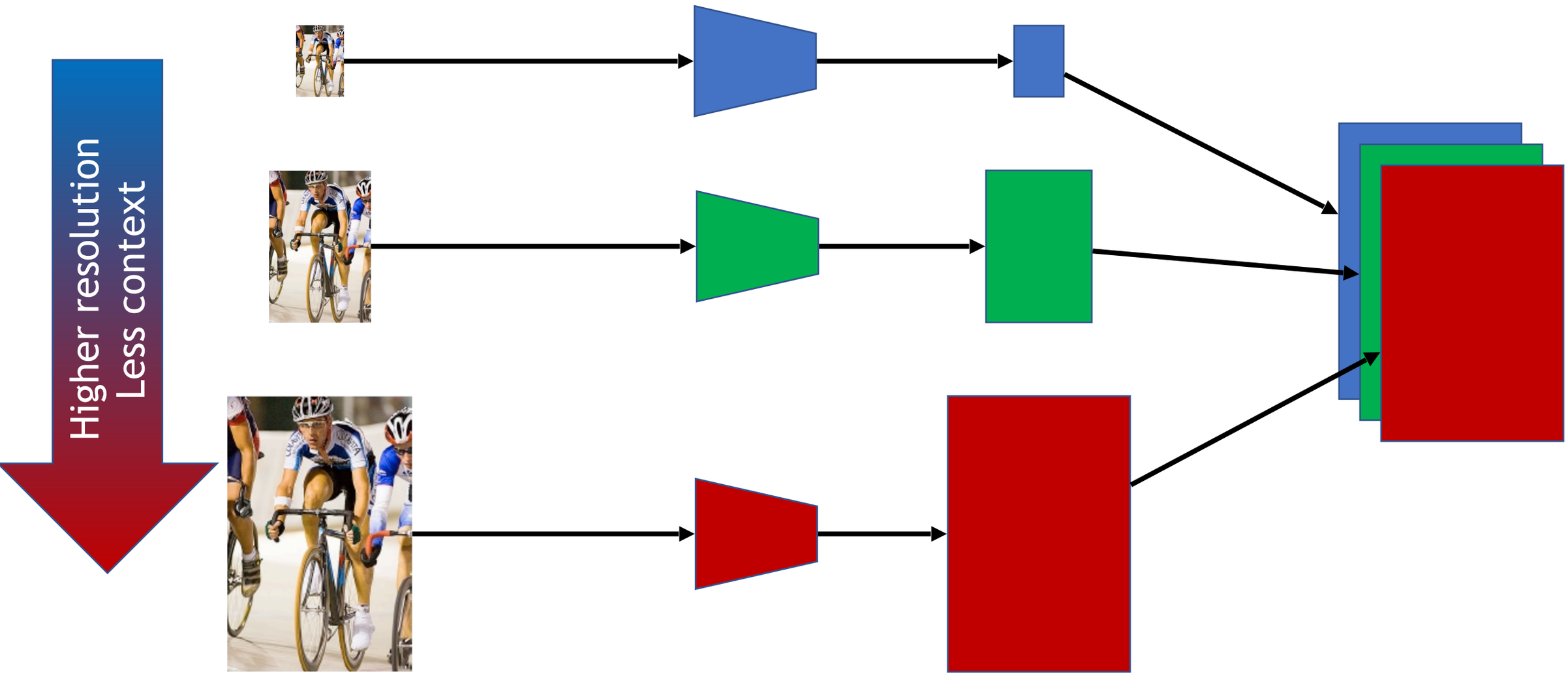
Slide credit: Bharath Hariharan

Semantic segmentation using convolutional networks



Slide credit: Bharath Hariharan

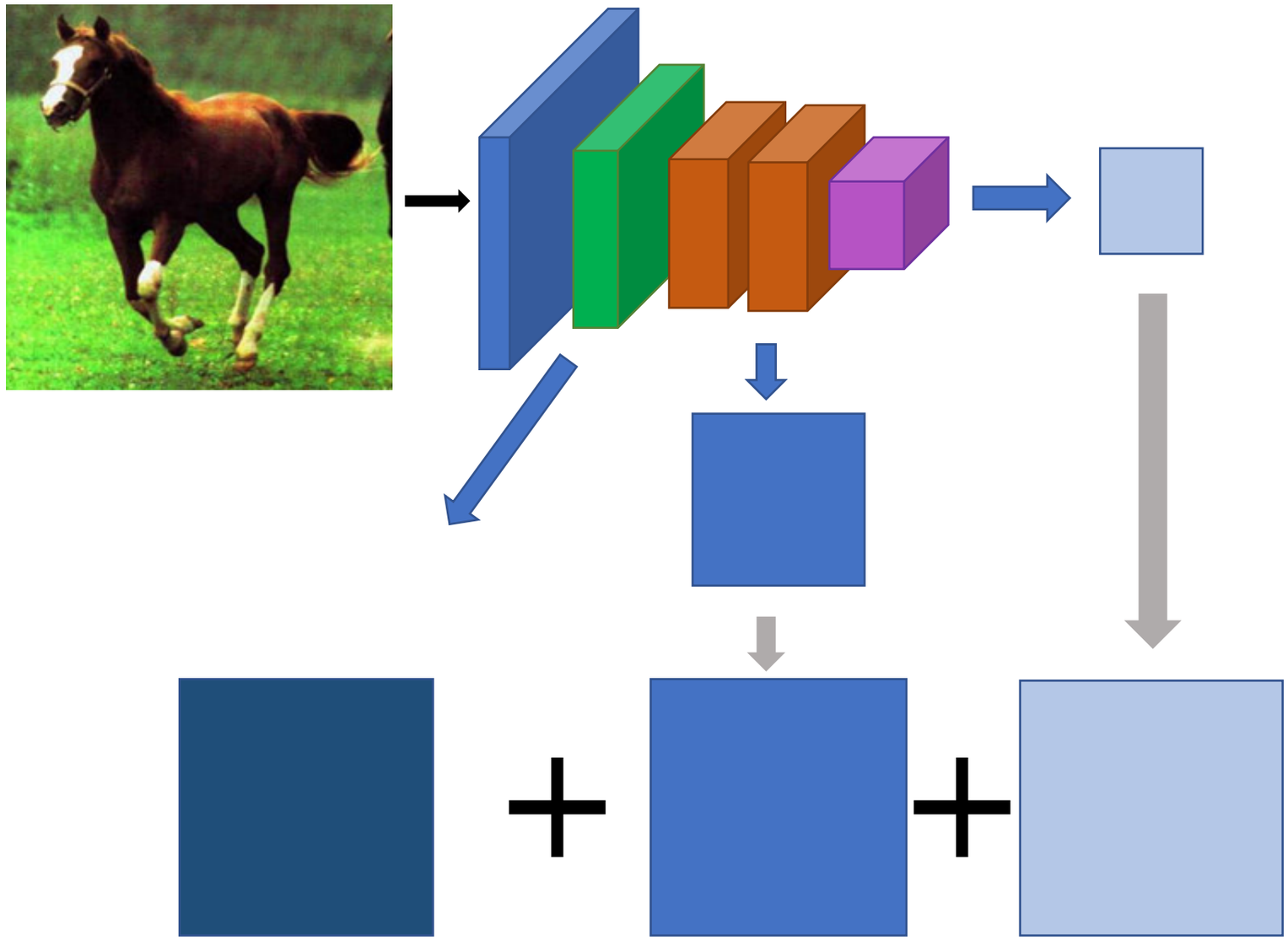
Solution 1: Image pyramids



Learning Hierarchical Features for Scene Labeling. Clement Farabet, Camille Couprie, Laurent Najman, Yann LeCun. In *TPAMI*, 2013.

Slide credit: Bharath Hariharan

Solution 2: Skip connections



Slide credit: Bharath Hariharan

Skip connections

Fully convolutional networks for semantic segmentation. Evan Shelhamer, Jon Long, Trevor Darrell. In *CVPR* 2015

Slide credit: Bharath Hariharan

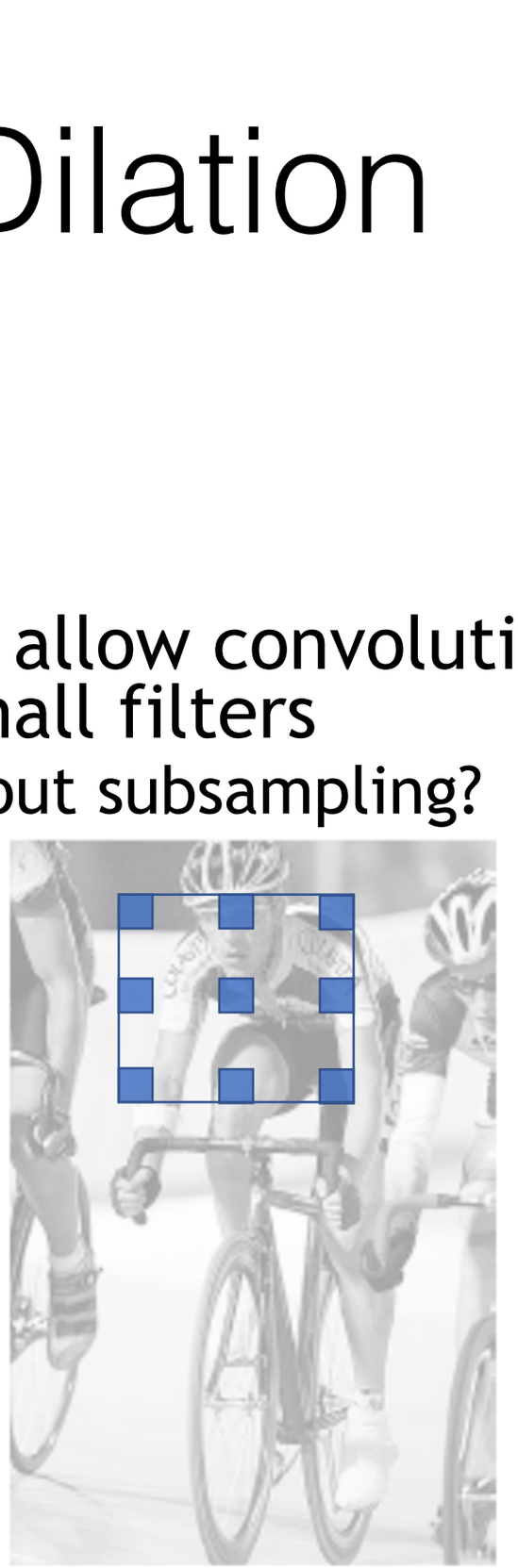
Solution 3: Dilation

- Need subsampling to allow convolutional layers to capture large regions with small filters
  - Can we do this without subsampling?

Slide credit: Bharath Hariharan

Solution 3: Dilation

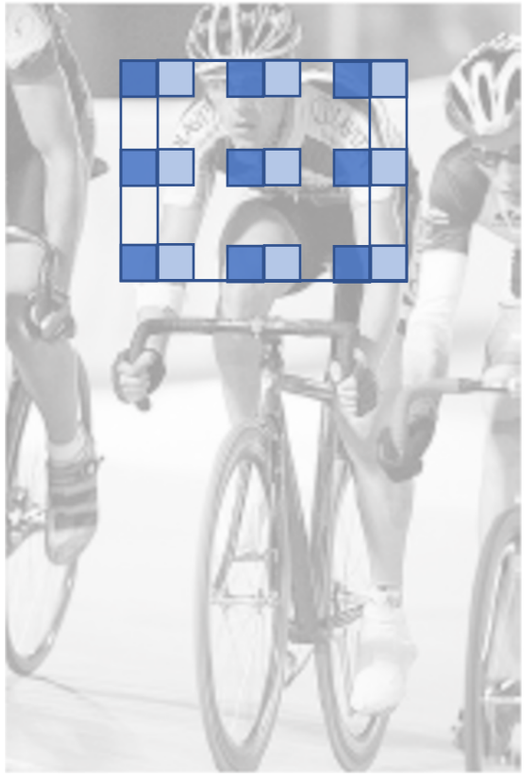
- Need subsampling to allow convolutional layers to capture large regions with small filters
  - Can we do this without subsampling?



Slide credit: Bharath Hariharan

Solution 3: Dilation

- Need subsampling to allow convolutional layers to capture large regions with small filters
  - Can we do this without subsampling?

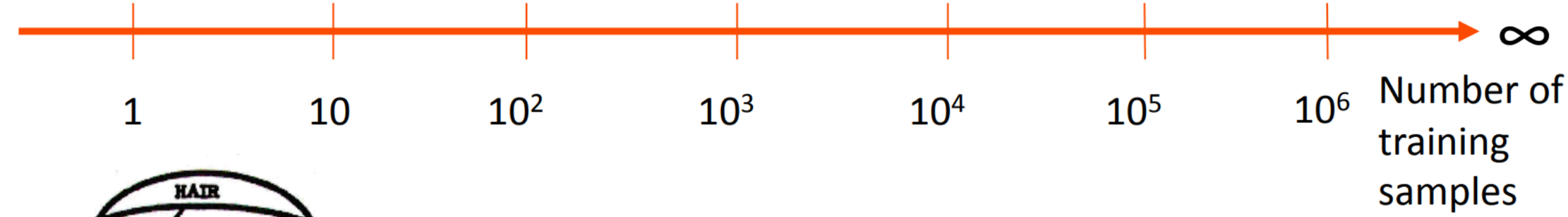


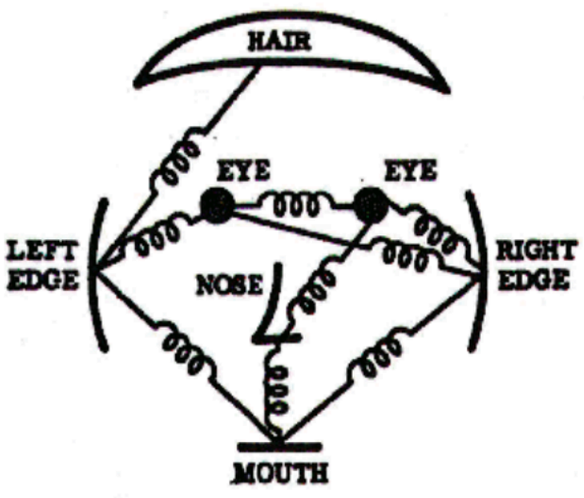
Slide credit: Bharath Hariharan

Two Extremes of Vision

Extrapolation problem
Generalization
Diagnostic features

Interpolation problem
Correspondence
Finding the differences





Slide credit: Aude Oliva

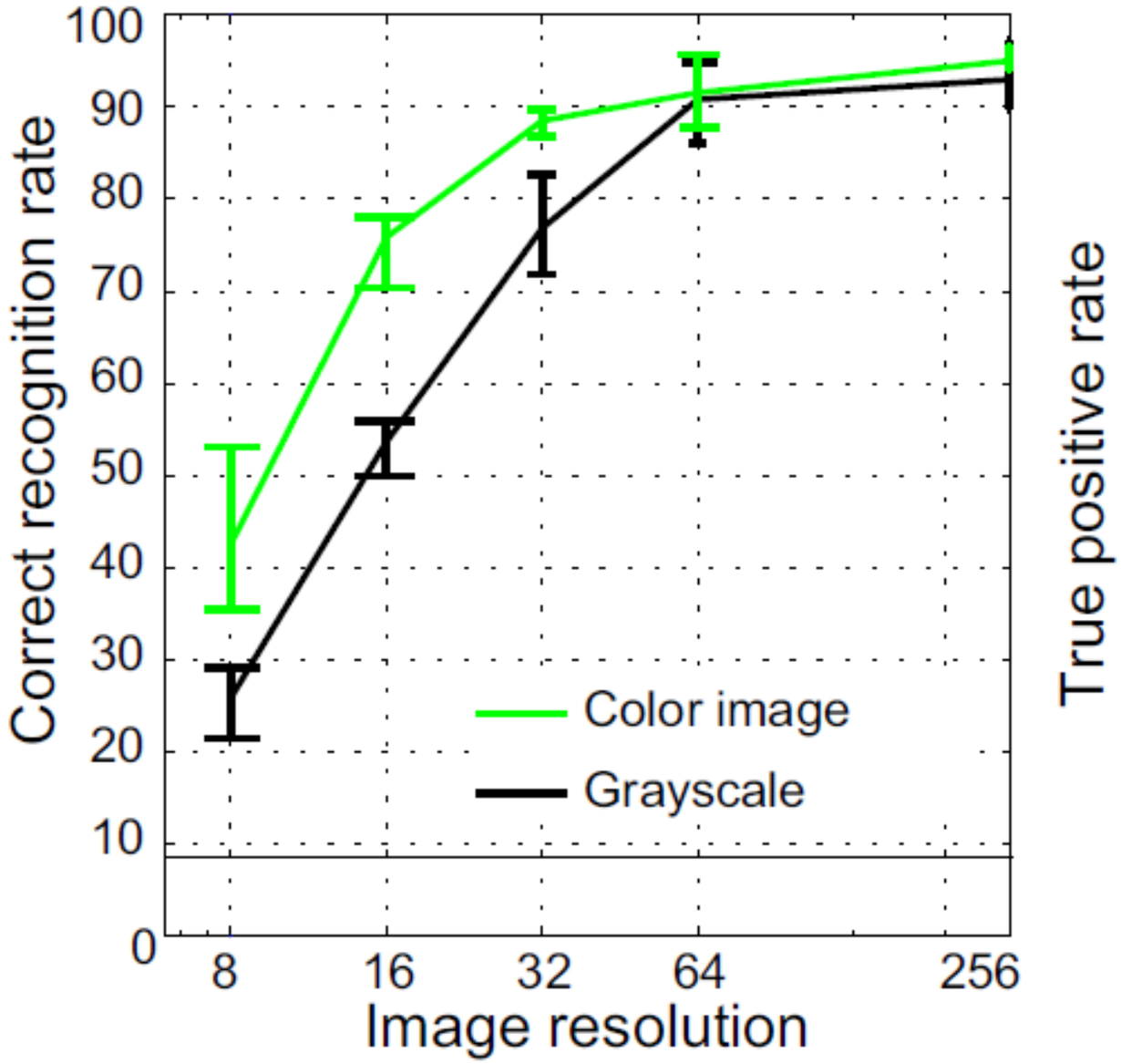
Tiny Images

80 million tiny images: a large dataset for non-parametric object and scene recognition

Antonio Torralba, Rob Fergus and William T. Freeman. PAMI 2008.
http://groups.csail.mit.edu/vision/TinyImages/

c) Segmentation of 32x32 images

Given a benchmark, resolution and human scene recognition accuracy increase to a limit



Torralba et al.

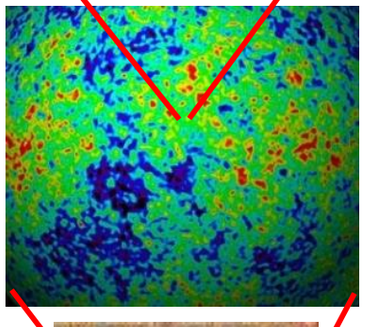
Powers of 10

Number of images on my hard drive: 10^6

Number of images seen during my first 10 years: 10^8
(3 images/second * 60 * 60 * 16 * 365 * 10 = 630,720,000)

Number of images seen by all humanity: 10^{20}
106,456,367,669 humans[1] * 60 years * 3 images/second * 60 * 60 * 16 * 365 =
1 from http://www.prb.org/Articles/2002/HowManyPeopleHaveEverLivedonEarth.aspx

Number of photons in the universe: 10^{88}



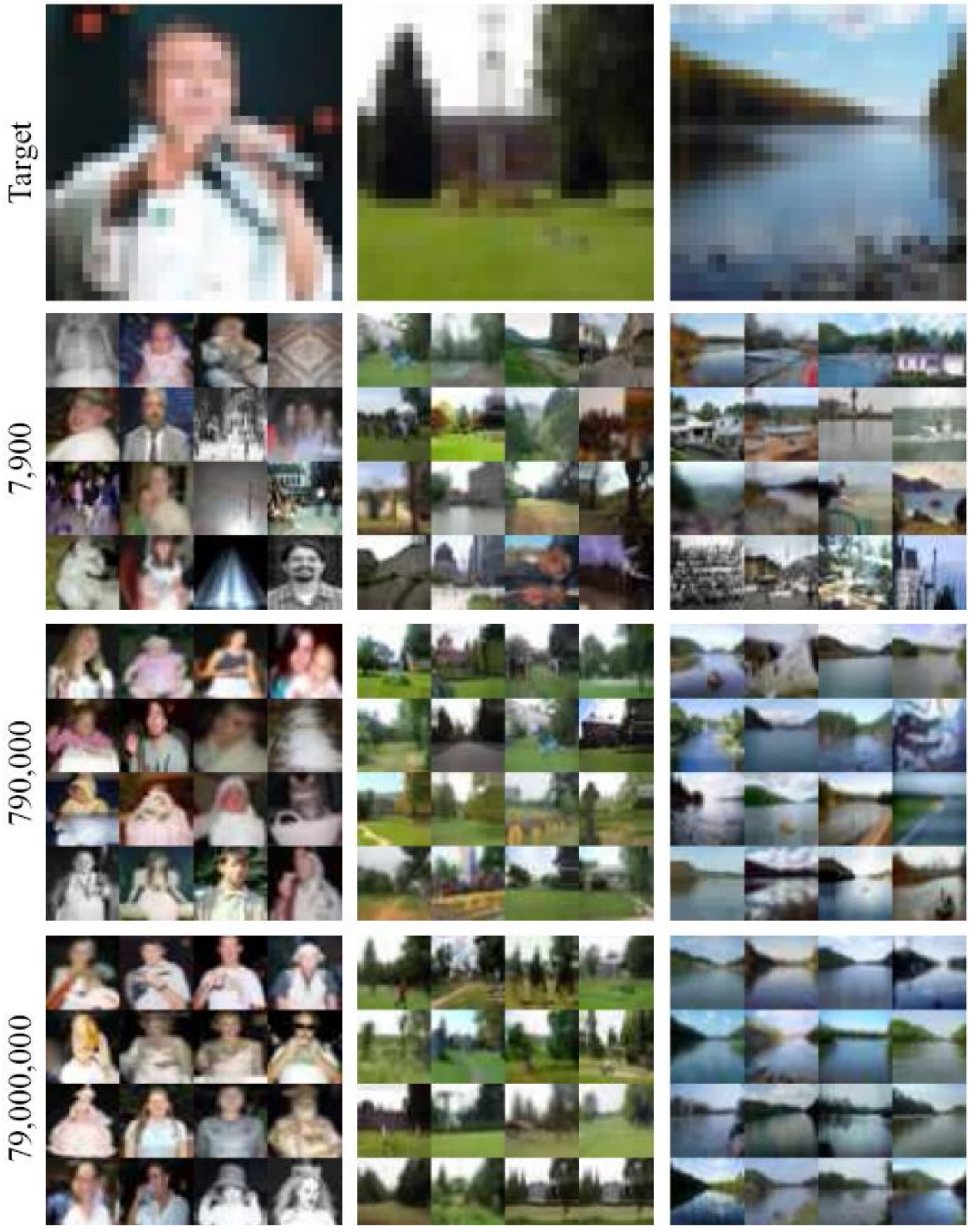
Number of all 32x32 images: 10^{7373}
$256^{32*32*3} \sim 10^{7373}$

But not all scenes are so original

Lots

Of

Images

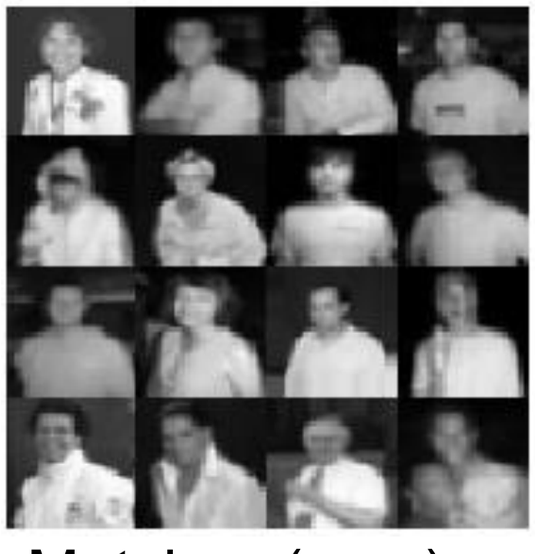


Application: Automatic Colorization

Input

Color Transfer

Color Transfer



Matches (gray)

Matches (w/ color)

Avg Color of Match

Application: Automatic Colorization

Input

Color Transfer

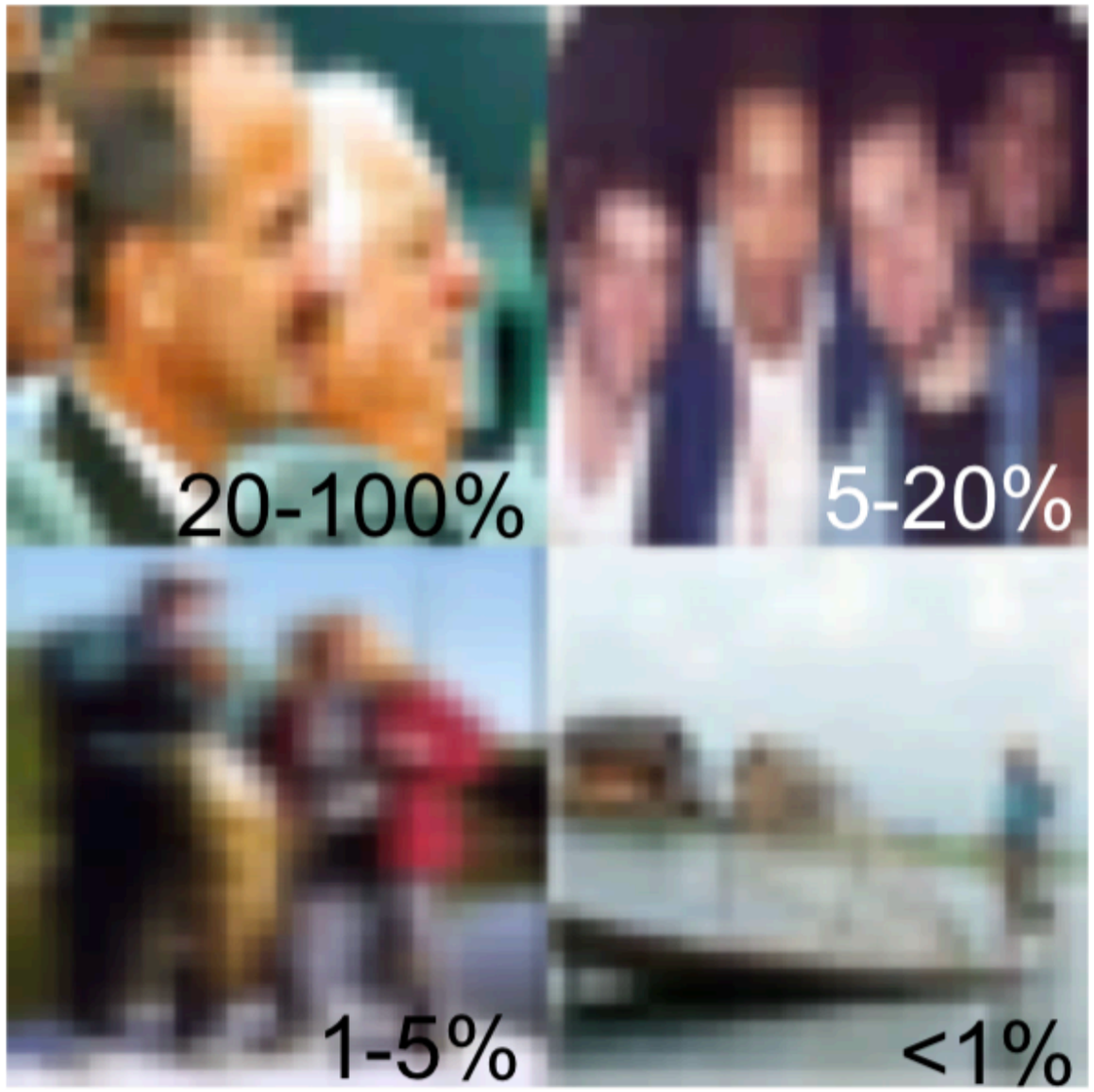
Color Transfer

Matches (gray)

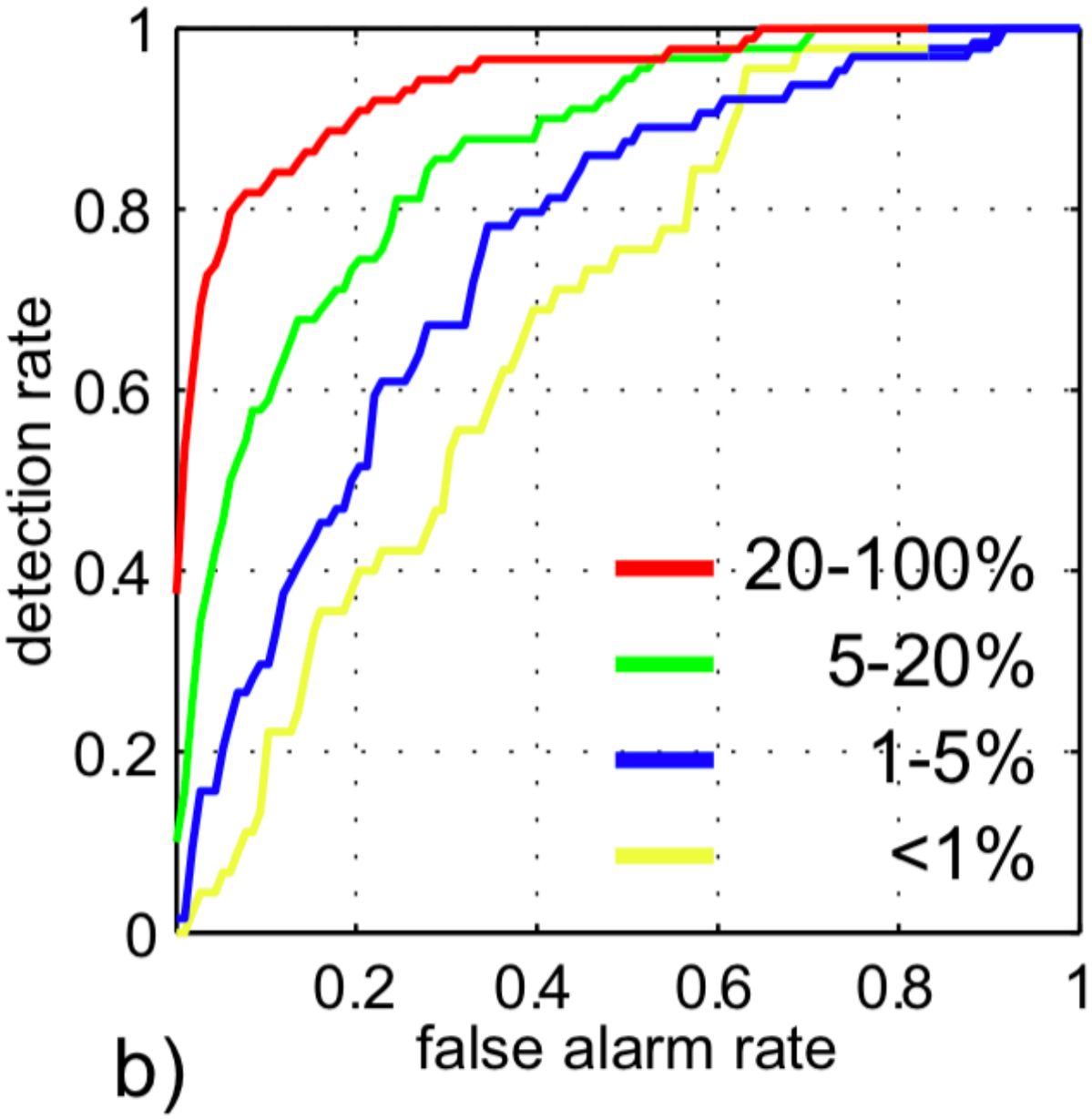
Matches (w/ color)

Avg Color of Match

Person Recognition



a)



b)

.\|

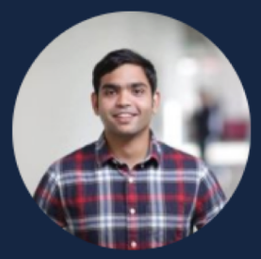
Exploring the Limits of Weakly Supervised Pretraining

Laurens van der Maaten

ECCV 2018


Dhruv Mahajan

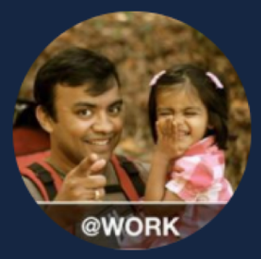

Ross Girshick


Vignesh Ramanathan


Kaiming He


Manohar Paluri


Yixuan Li



Ashwin Bharambe

facebook
Artificial Intelligence Research

https://arxiv.org/pdf/1805.00932.pdf

3,500,000,000 images!

Experiments

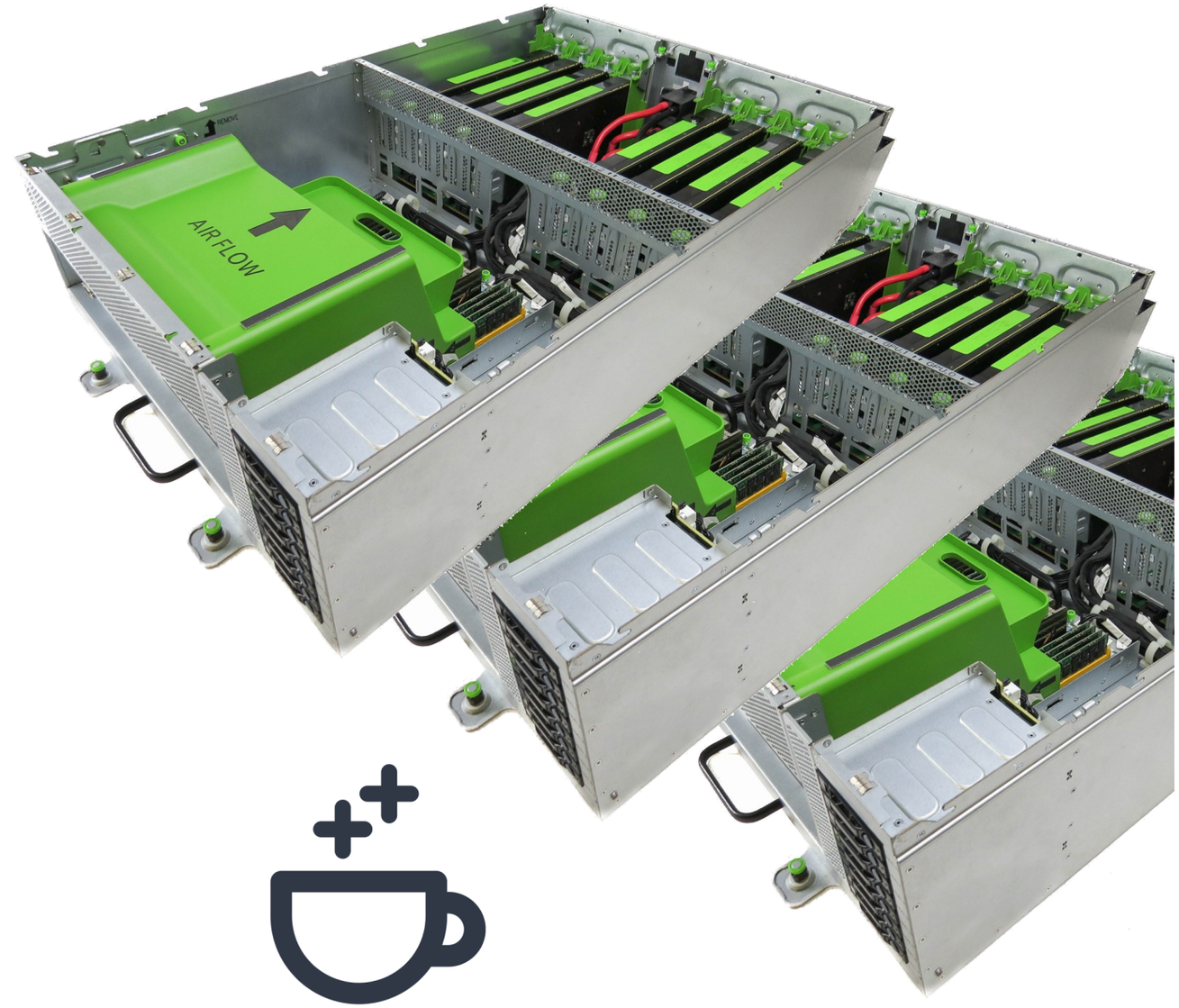
- Select a set of hashtags
- Download all public Instagram images that has at least one of these hashtags
- Use WordNet synsets to merge hashtags into canonical form (merge #brownbear and #ursusarctos)
- The final list has 17,517 hashtags

1	aar	44	accommodation			17474	yurt
2	aardvark	45	accompaniment			17475	zabaglione
3	aardwolf	46	accordion			17476	zambeziriver
4	aba	47	accoutrement			17477	zamboni
5	abaca	48	accumulator			17478	zamia
6	abacus	49	ace			17479	zantac
7	abalone	50	aceofclubs			17480	zantedeschia
8	abatis	51	aceofdiamonds			17481	zap
9	abaya	52	aceofhearts			17482	zapper
10	abbey	53	aceofspades			17483	zarf
11	abele	54	acer			17484	zea
12	abelia	55	acerjaponicum			17485	zebra
13	abies	56	acerola			17486	zebrafinch
14	abila	57	acerpalmatum			17487	zebrawood
15	abm	58	acerrubrum			17488	zebu
16	abortus	59	acetaminophen			17489	zero
17	abronia	60	acetate			17490	zeus
18	absinth	61	acheron			17491	zhujiang
19	absinthe	62	acherontia			17492	ziggurat
20	abstraction	63	acherontiaatropos			17493	zill
21	abstractionism	64	achillea			17494	zimmerframe
22	abutilon	65	achilleamillefolium			17495	zinfandel
23	abutment	66	achimenes	...		17496	zing
24	abyss	67	acid			17497	zingiber
25	abyssinian	68	acidophilus			17498	zinnia
26	acacia	69	acinonyxjubatus			17499	zipgun
27	acaciadealbata	70	acinus			17500	zipper
28	academy	71	ackee			17501	zither
29	acalypha	72	aconcagua			17502	ziti
30	acanthaceae	73	aconite			17503	ziziphus
31	acanthurus	74	aconitum			17504	zizz
32	acanthus	75	acorn			17505	zodiac
33	acanthusmollis	76	acornsquash			17506	zoloft
34	acapulcogold	77	acousticguitar			17507	zombi
35	acarus	78	acoustics			17508	zoologicalgarden
36	accelerator	79	acrididae			17509	zoom
37	accelerometer	80	acrobates			17510	zooplankton
38	access	81	acropolis			17511	zootsuit
39	accessory	82	acropora			17512	zori
40	accident	83	acrylic			17513	zoysia
41	accipiter	84	acrylicpaints			17514	zuiderzee
42	accipiternisus	85	actias			17515	zygnema
43	accipitridae	86	actiasluna			17516	zygocactus
						17517	zygoptera

facebook
Artificial Intelligence Research

Experiments

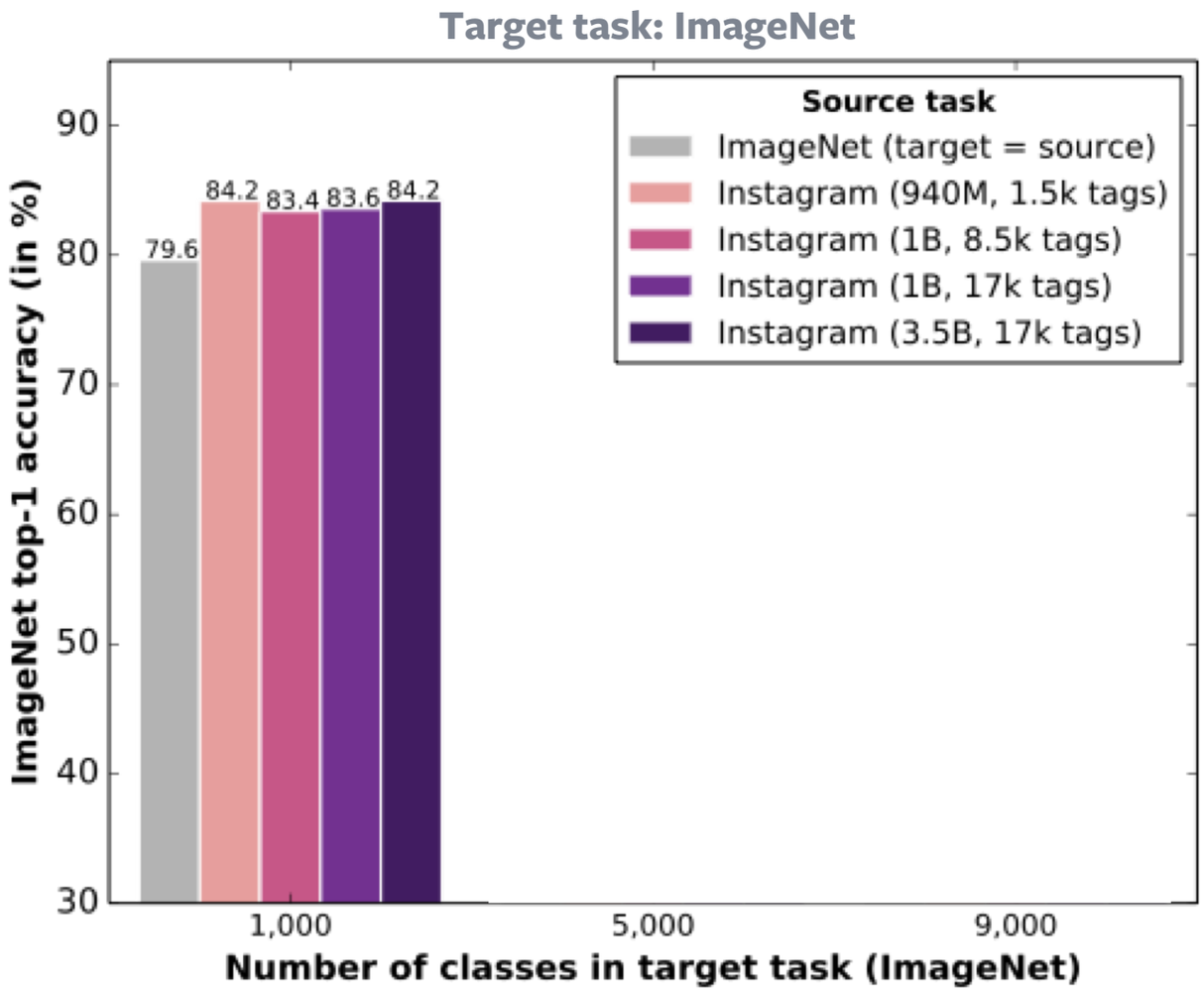
- Train ResNeXt-32xCd convolutional networks
- Use *c*-of-*K* vector to represent multiple labels
- Train to minimize multi-class logistic loss
- Distribute training batches across 336 GPUs
- Scale learning rate by batch size (*N=8,064*) after learning rate "warm-up" (Goyal *et al.*, 2017)



Fix Model;
Vary Data

- Pretrain model on ImageNet or Instagram

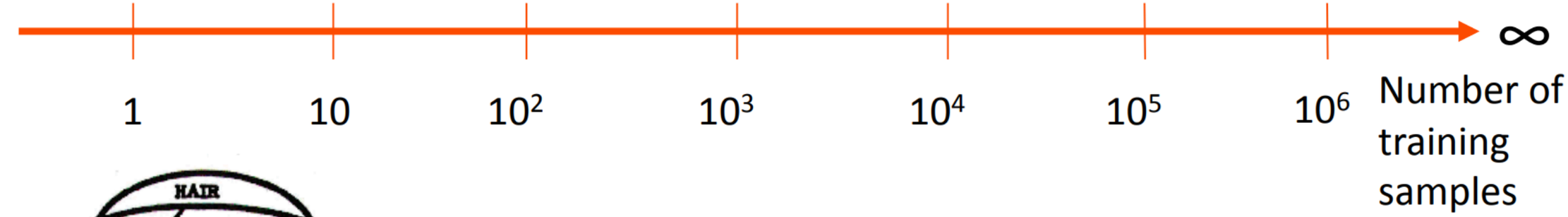
- Finetune on ImageNet

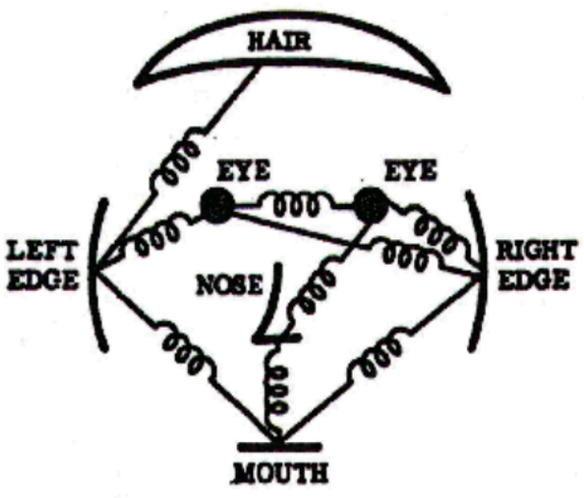


Two Extremes of Vision

Extrapolation problem
Generalization
Diagnostic features

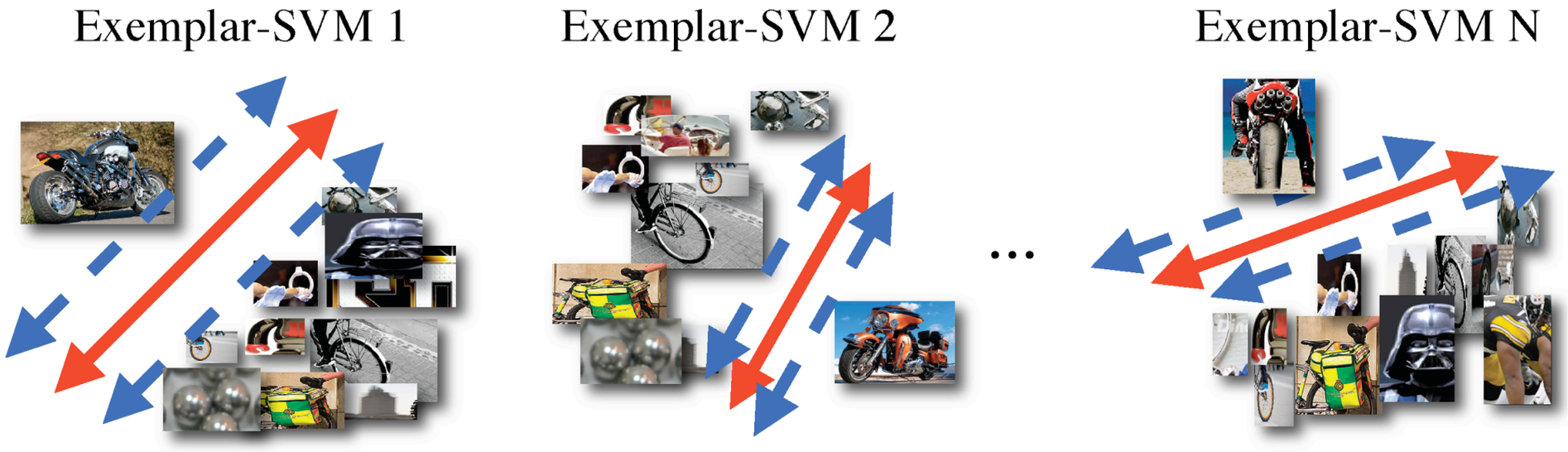
Interpolation problem
Correspondence
Finding the differences





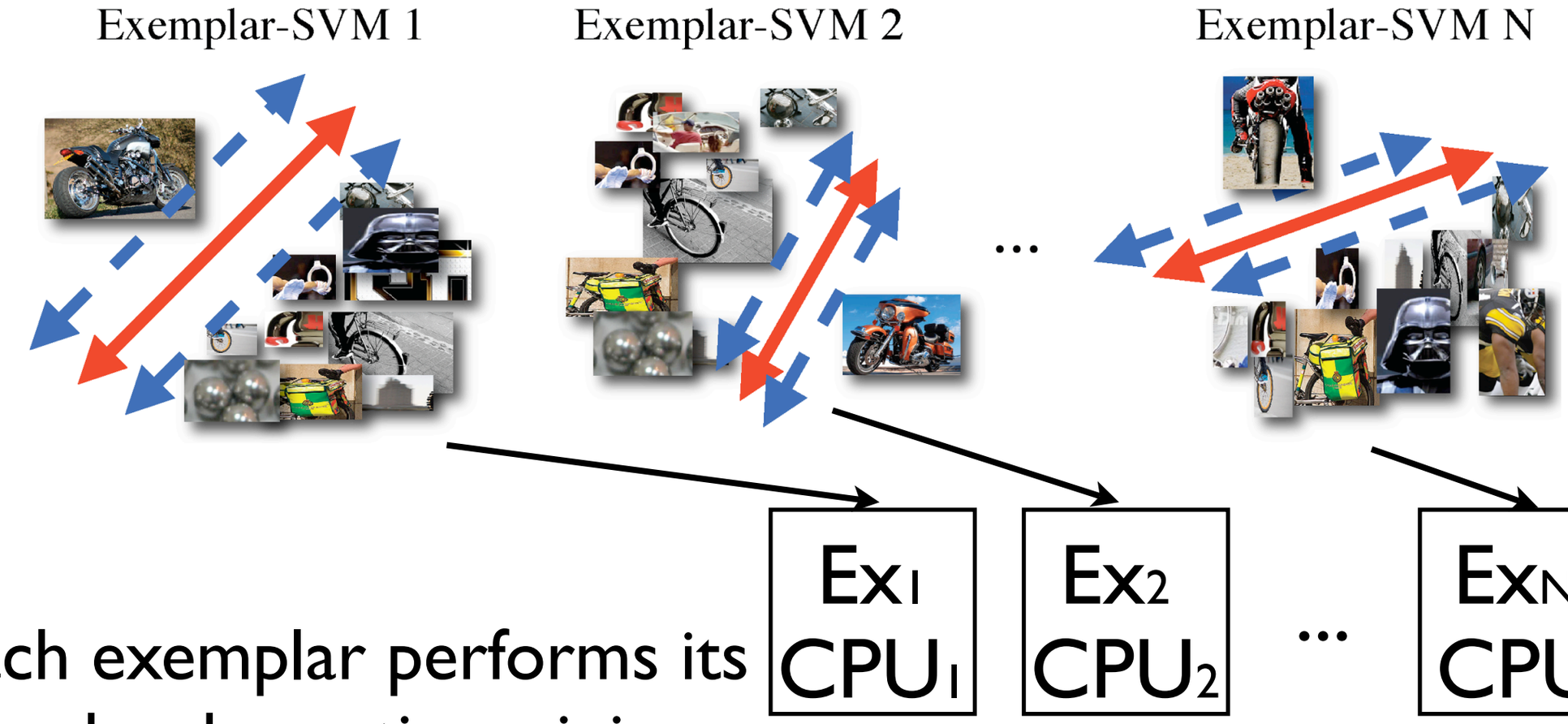
Slide credit: Aude Oliva

Exemplar-SVMs



- Learn a separate linear SVM for each instance (exemplar) in the dataset (PASCAL VOC)
- Each Exemplar-SVM is trained with a **single** positive instance
- Each Exemplar-SVM is more defined by "*what it is not*" vs. "*what it is similar to*"

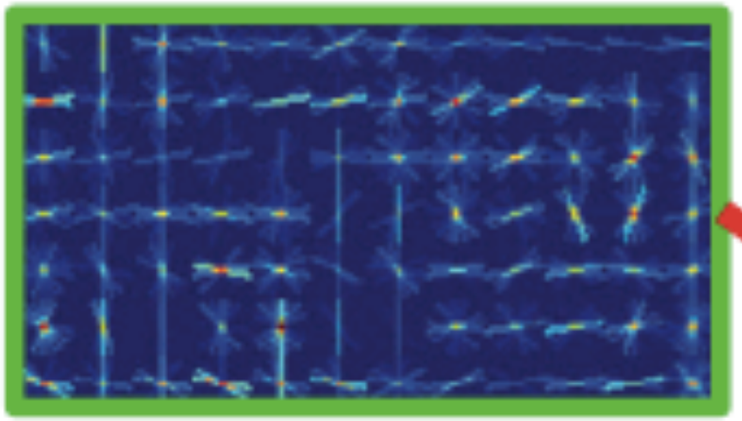
Large-scale training



- Each exemplar performs its own hard negative mining
- Solve many convex learning problems
- Parallel training on cluster

Exemplar

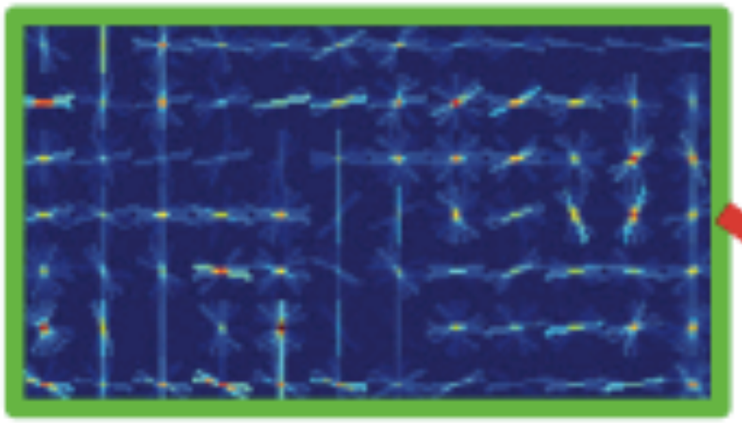
Detector **w**



Appearance

Exemplar

Detector w



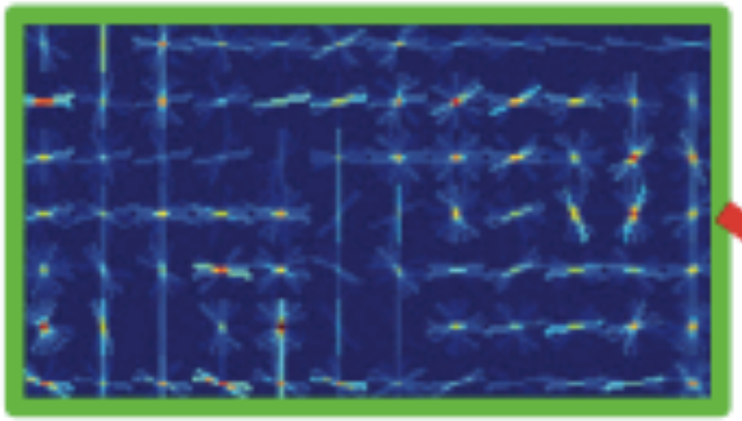
Appearance

Meta-data

Geometry

Exemplar

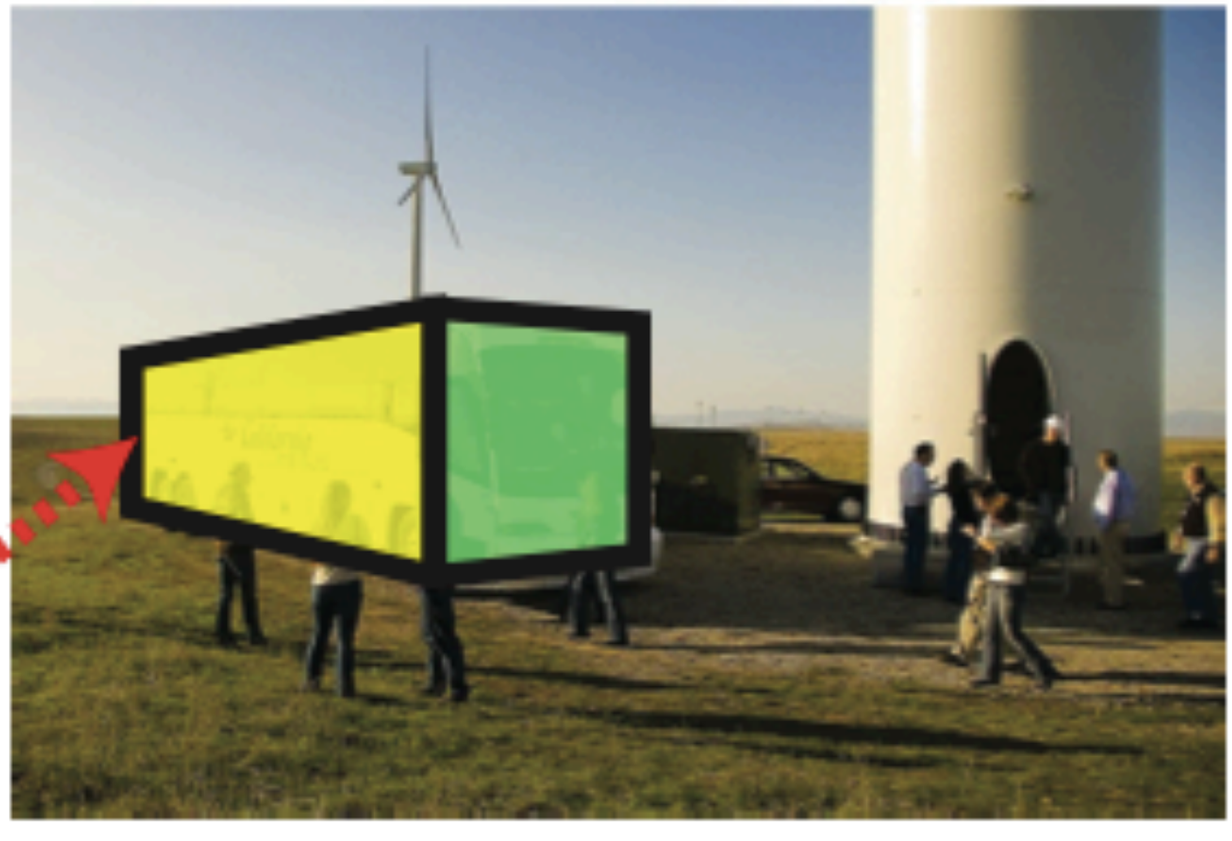
Detector w



Appearance

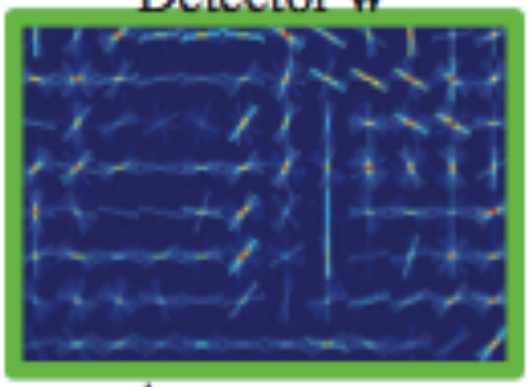
Meta-data

Geometry

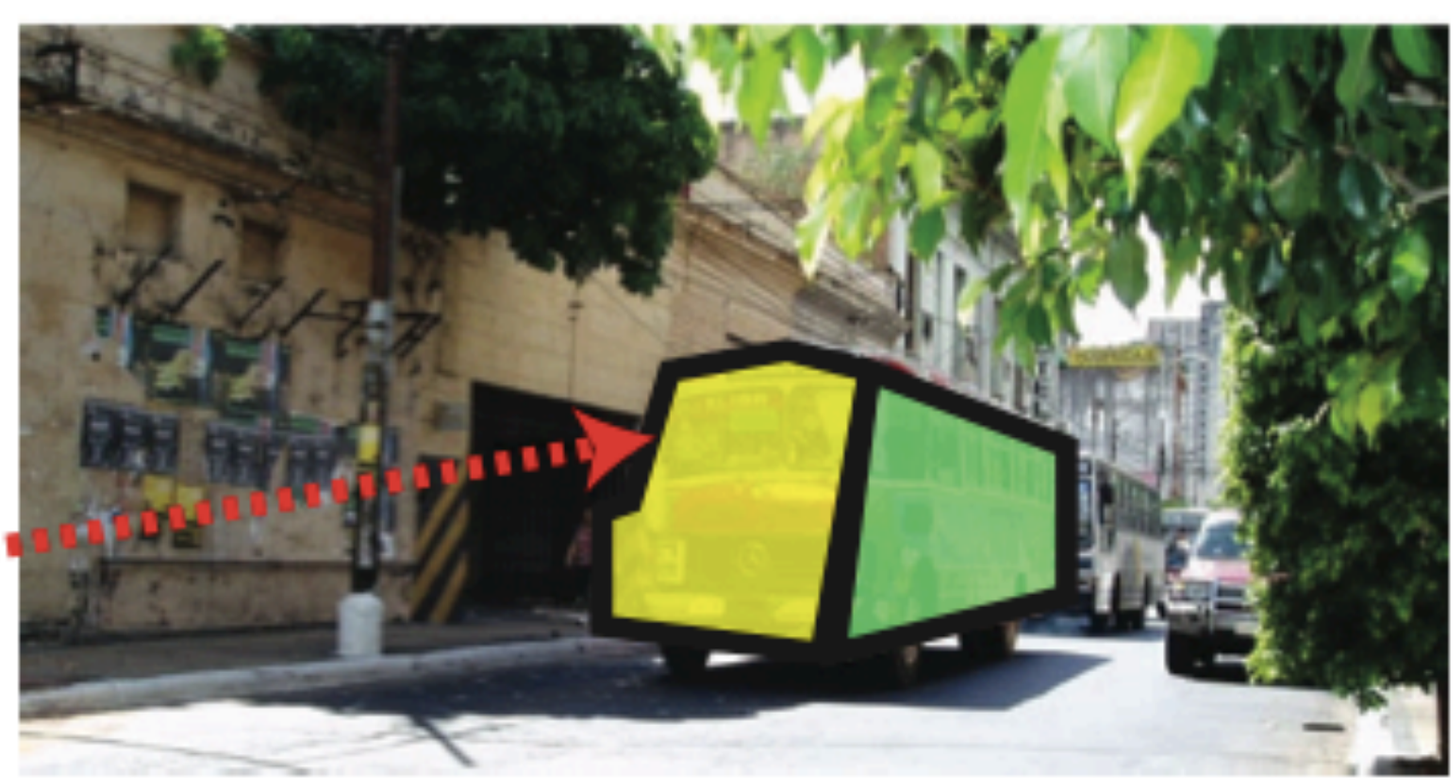


Exemplar

Detector w



Appearance

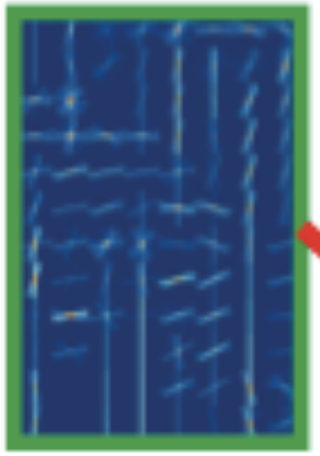


Meta-data

Geometry

Exemplar

Detector **w**



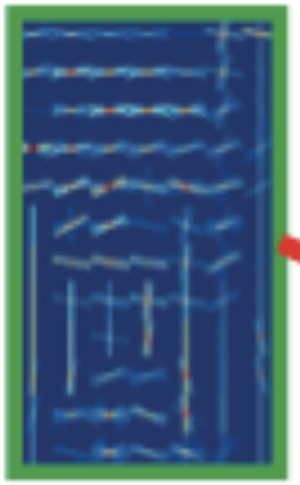
Appearance

Meta-data

3D Model

Exemplar

Detector **w**



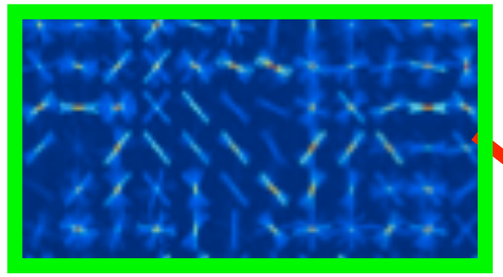
Appearance

Meta-data

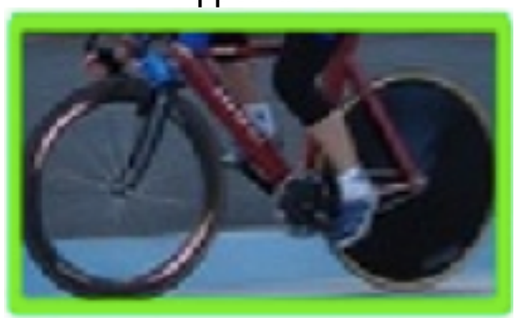
3D Model

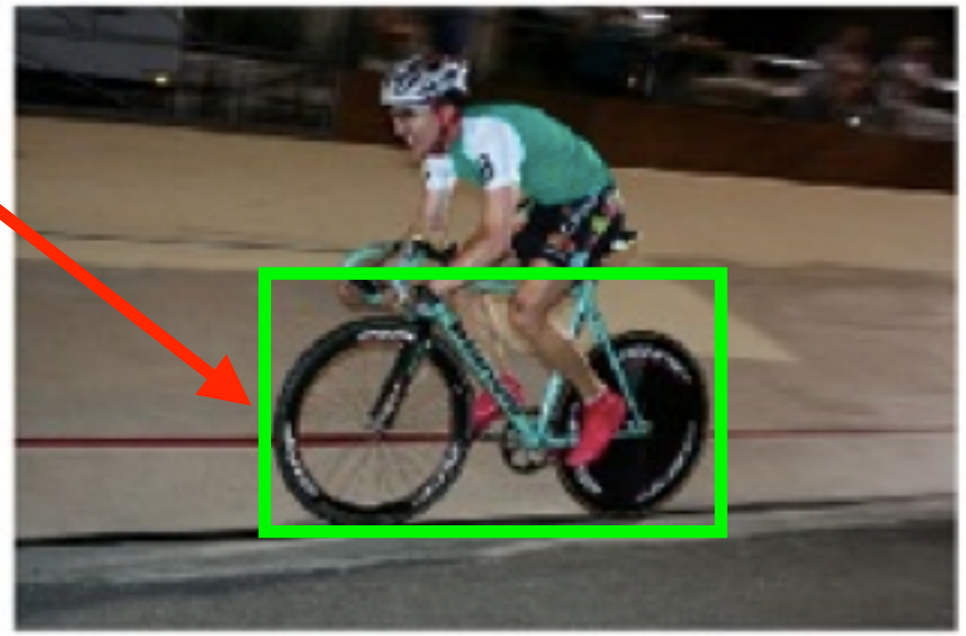
Exemplar

Detector **w**



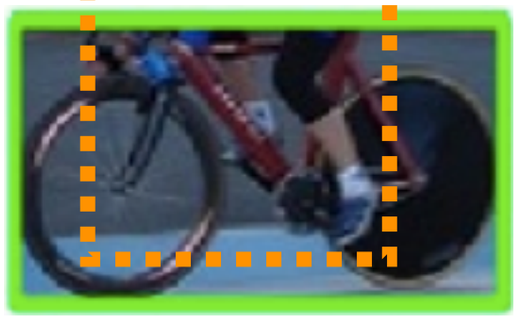
Appearance



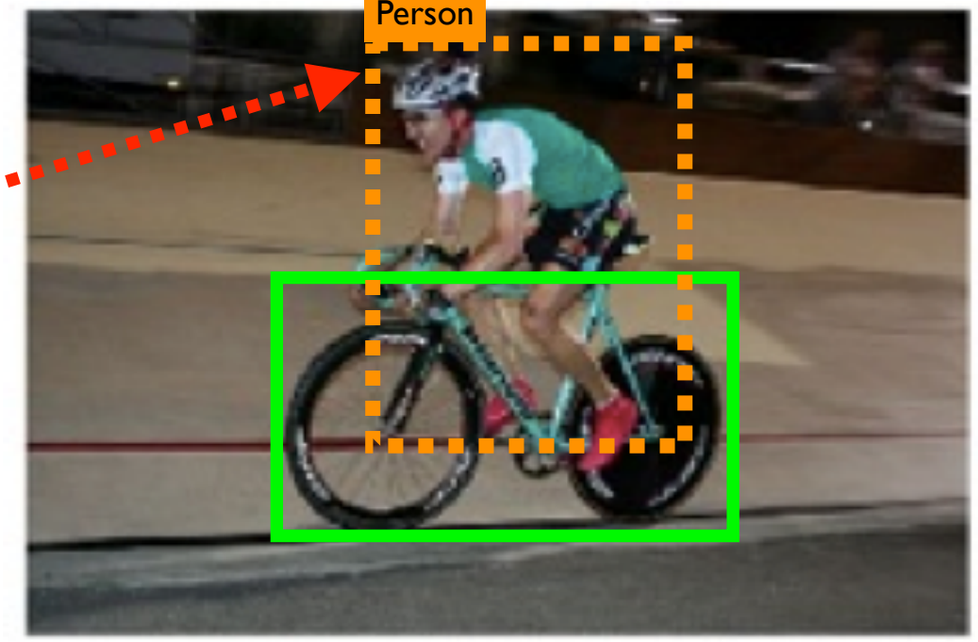


Meta-data

Person

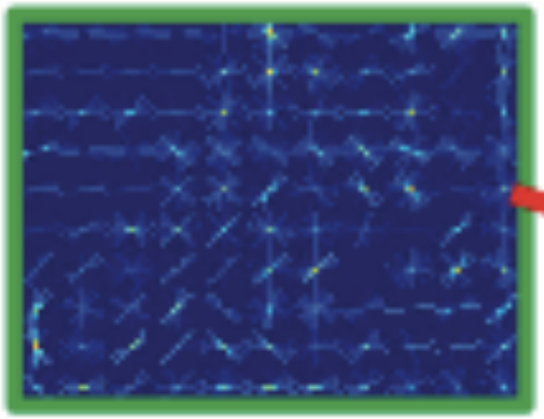


Person

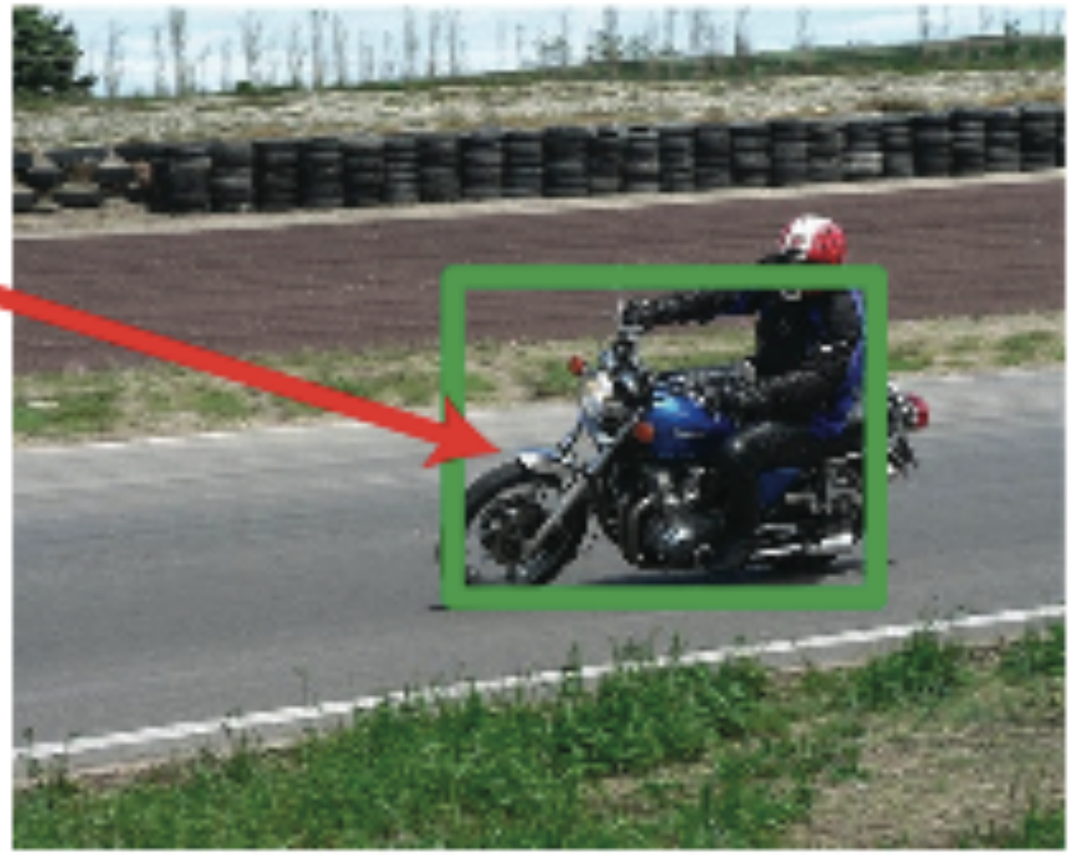


Exemplar

Detector w

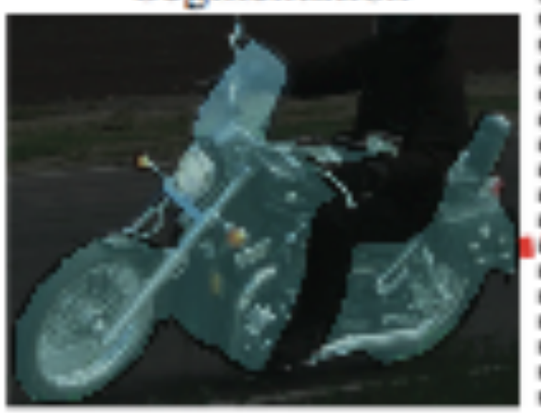


Appearance



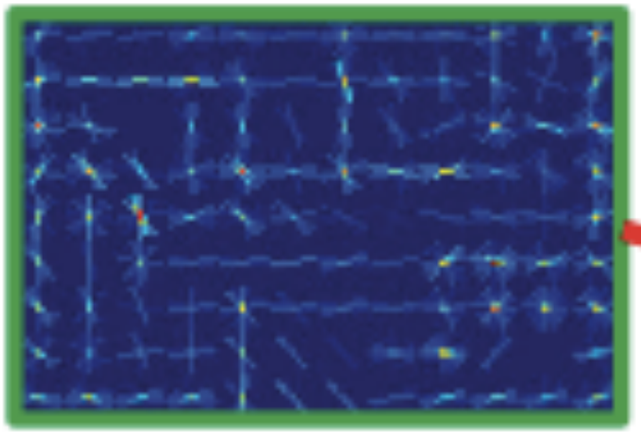
Meta-data

Segmentation



Exemplar

Detector w



Appearance

Meta-data

Segmentation

What’s this?

Photo from Coffee Creek Watershed Preserve

What’s this?

Entry-level categories
(Jolicoeur, Gluck, Kosslyn 1984)

- Typical member of a basic-level category are categorized at the expected level
- Atypical members tend to be classified at a subordinate level.

A bird

An ostrich

Classical Categorization

- Group objects by common properties

- What are birds?
  - animals, has wings, has feathers, can fly, chirps

Photo from Coffee Creek Watershed Preserve

Prototype Theory

Rosch and Lakoff

- According to the prototype view, an object will be classified as an instance of a category if it is sufficiently similar to the prototype.
- **Evidence for Prototype**:
- **Typicality ratings**: how good are robins as an example of birds
- **Production order of exemplars**: Name all the kinds of bird you can think of
- **Time to verify categorical statements**: True or false: a robin is a bird

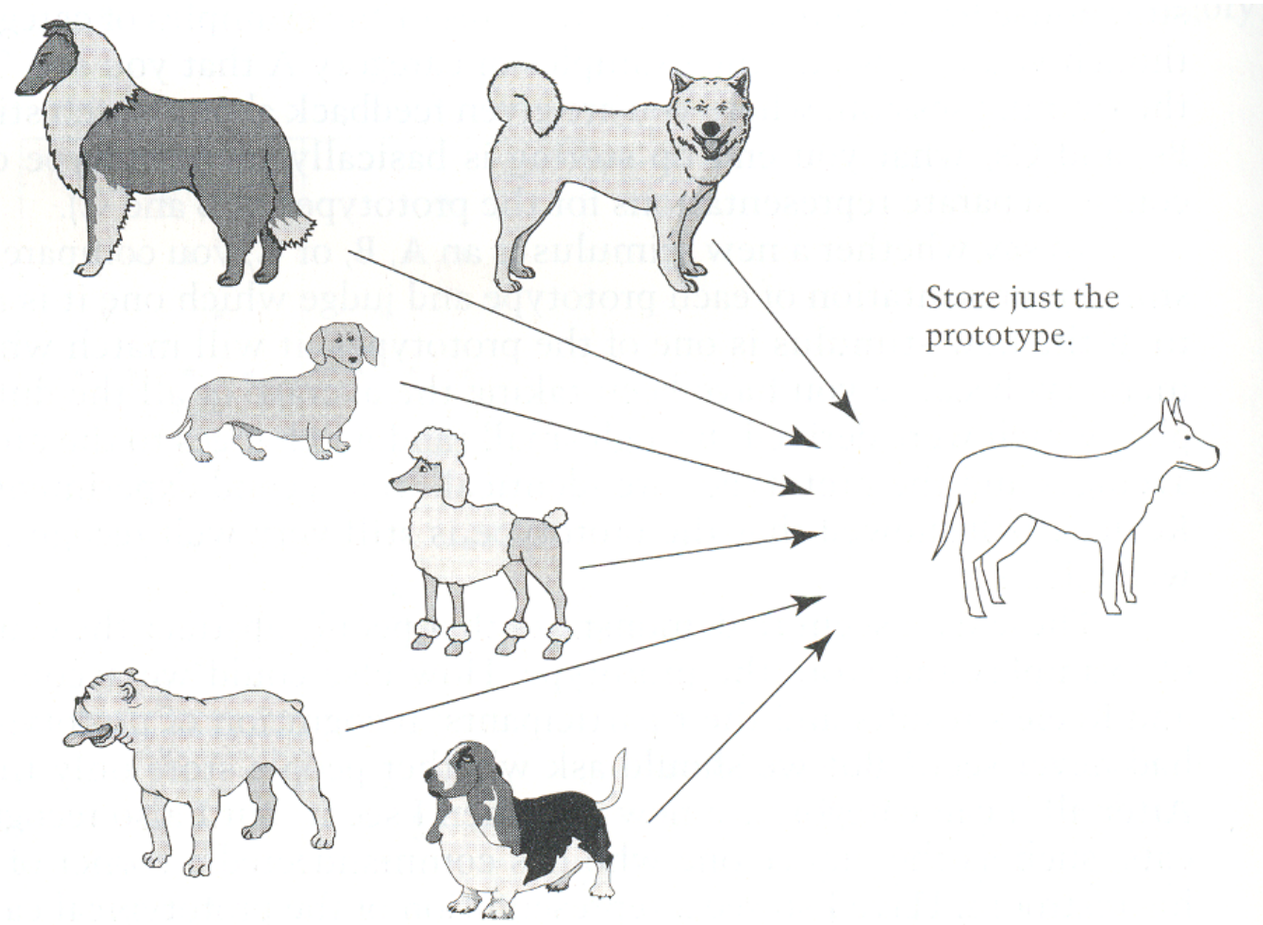
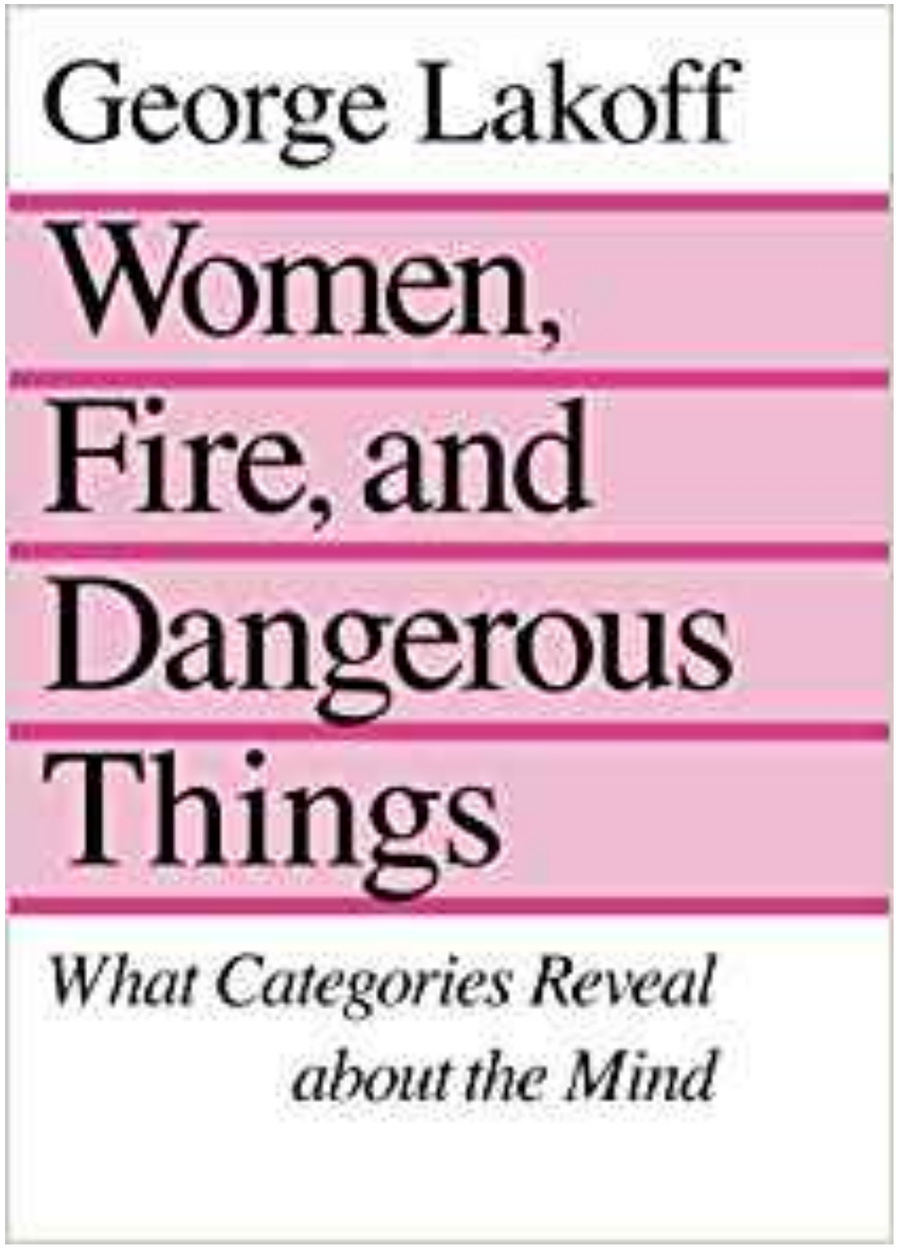


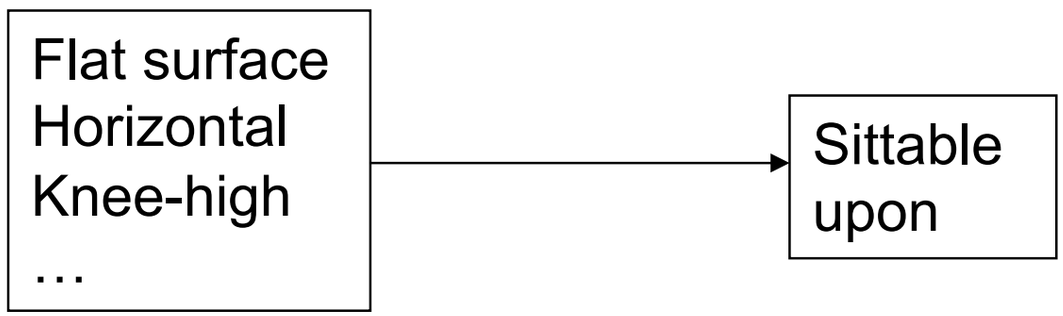
Figure 7.3. *Schematic of the prototype model. Although many exemplars are seen, only the prototype is stored. The prototype is updated continually to incorporate more experience with new exemplars.*

Dyirbal Indigenous People

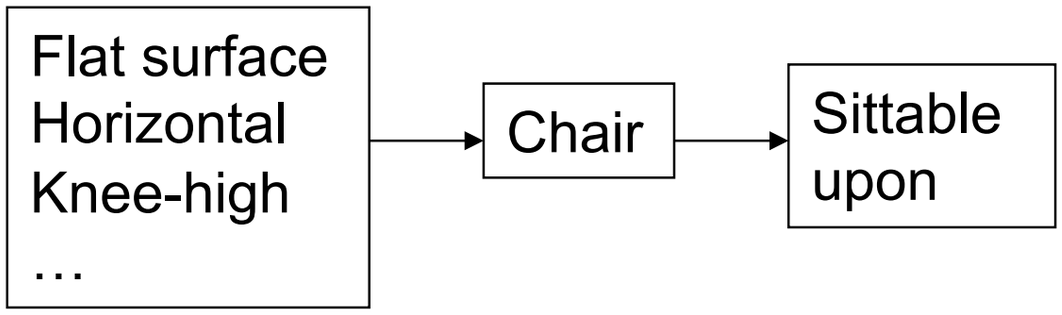


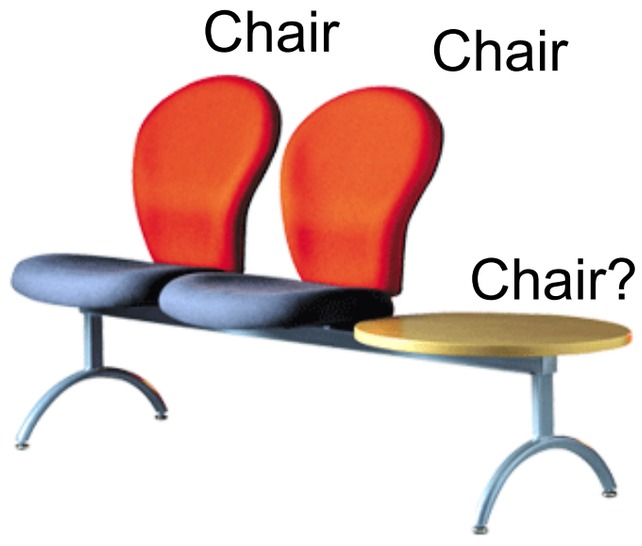
The perception of function

- Direct perception (affordances): Gibson



- Mediated perception (Categorization)





Direct perception

Some aspects of an object function can be perceived directly

- Observer relativity: Function is observer dependent

From http://lastchancerescueflint.org

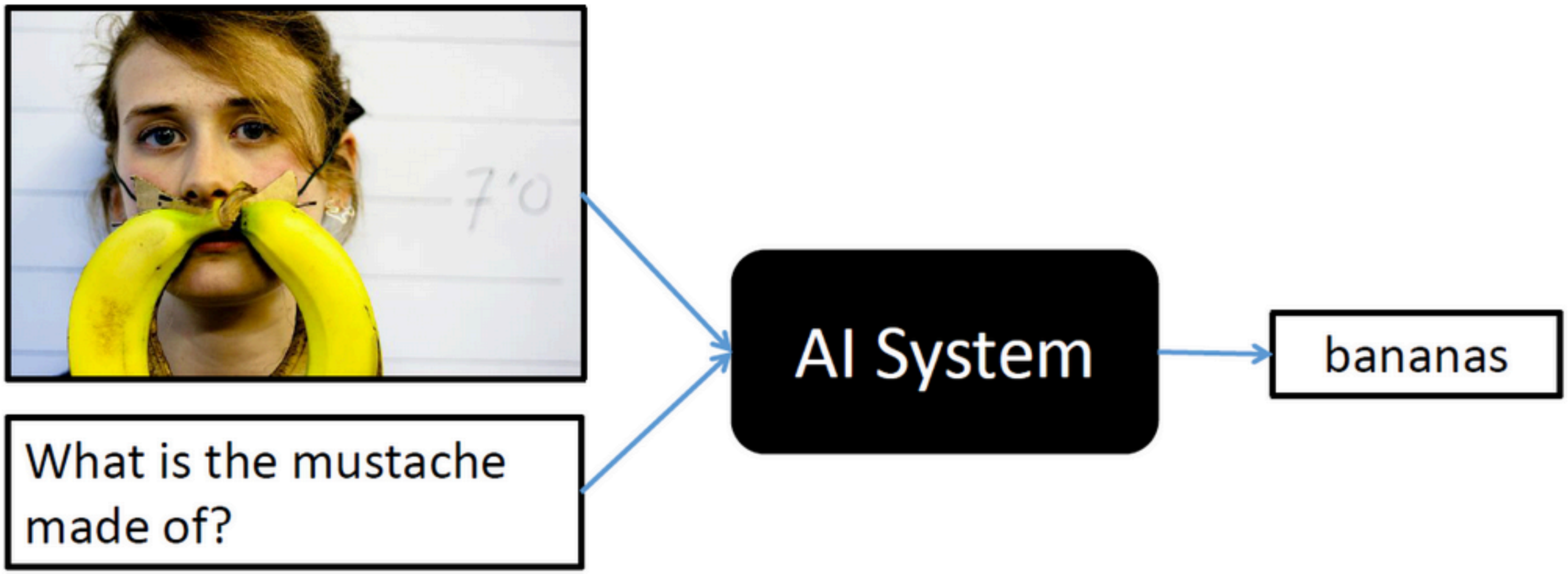
Limitations of Direct Perception

Objects of similar structure might have very different functions

Figure 9.1.2 Objects with similar structure but different functions. Mailboxes afford letter mailing, whereas trash cans do not, even though they have many similar physical features, such as size, location, and presence of an opening large enough to insert letters and medium-sized packages.

Not all functions seem to be available from direct visual information only.

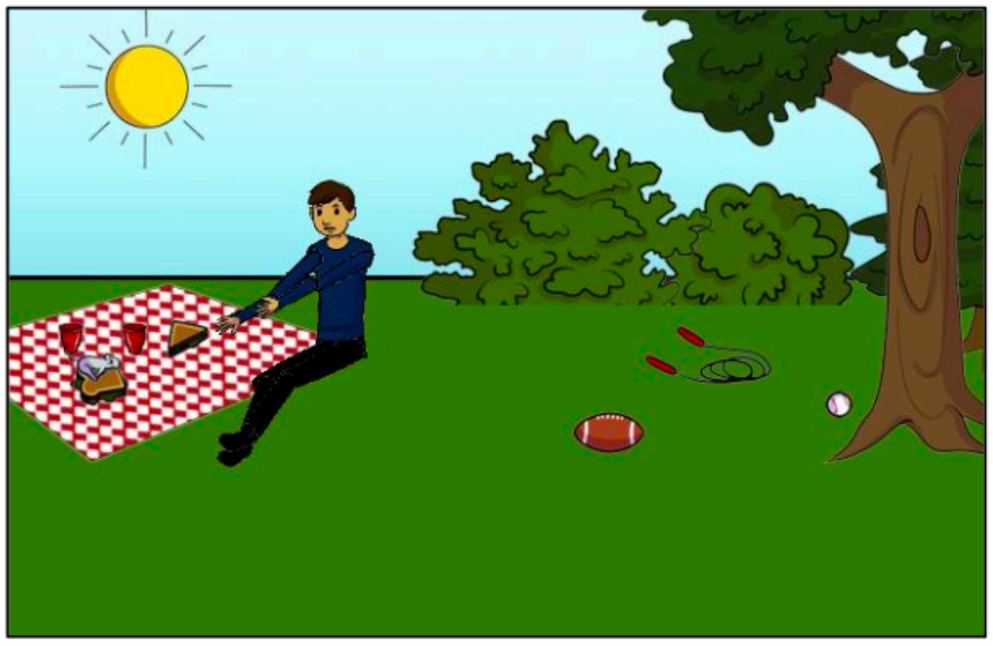
The functions are the same at some level of description: we can put things inside in both and somebody will come later to empty them. However, we are not expected to put inside the same kinds of things…



http://www.visualqa.org/challenge.html

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

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

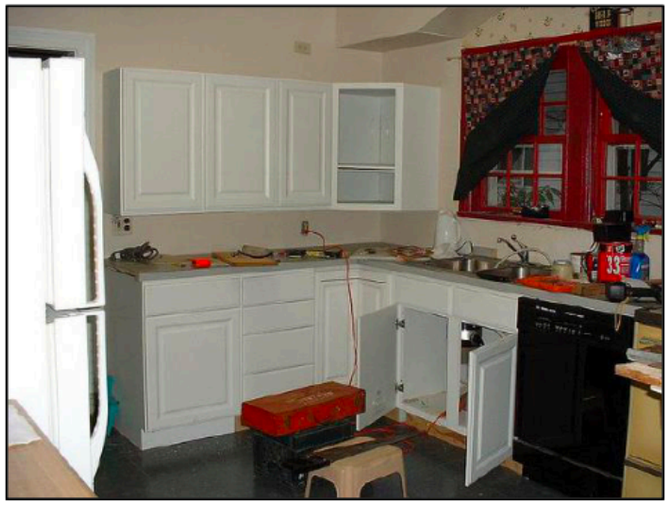


Is this person expecting company?
What is just under the tree?

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

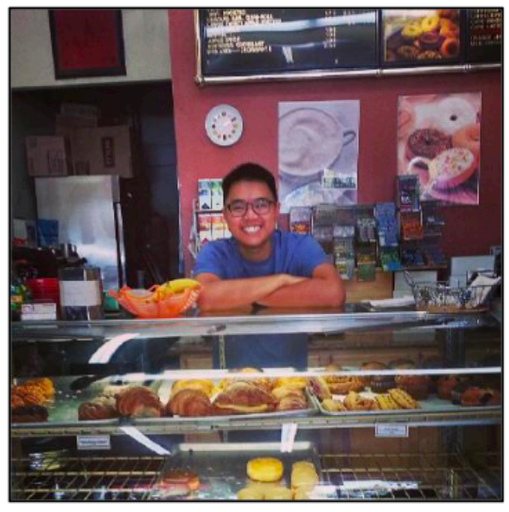
Fig. 1: Examples of free-form, open-ended questions collected for images via Amazon Mechanical Turk. Note that commonsense knowledge is needed along with a visual understanding of the scene to answer many questions.

Questions and answers collected with Amazon Mechanical Turk

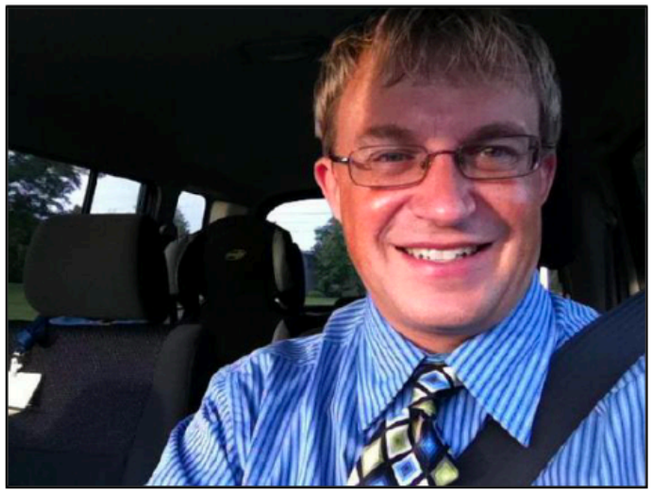


Is something under the sink broken?	yes yes yes	no no no
What number do you see?	33 33 33	5 6 7

Can you park here?	no no no	no no yes
What color is the hydrant?	white and orange white and orange white and orange	red red yellow



What kind of store is this?	bakery bakery pastry	art supplies grocery grocery
Is the display case as full as it could be?	no no no	no yes yes



Does this man have children?	yes yes yes	yes yes yes
Is this man crying?	no no no	no yes yes

Has the pizza been baked?	yes yes yes	yes yes yes
What kind of cheese is topped on this pizza?	feta feta ricotta	mozzarella mozzarella mozzarella

How many pickles are on the plate?	1 1 1	1 1 1
What is the shape of the plate?	circle round round	circle round round

Fig. 2: Examples of questions (black), (a subset of the) answers given when looking at the image (green), and answers given when not looking at the image (blue) for numerous representative examples of the dataset. See the appendix for more examples.

Architecture

Words

- Need ways to compare words

Next to the 'sofa' is a desk, and a 'person' is sitting behind it.

'armchair'	'man'
'bench'	'woman'
'chair'	'child'
'deck chair'	'teenager'
'ottoman'	'girl'
'seat'	'boy'
'stool'	'baby'
'swivel chair'	'daughter'
'loveseat'	'son'
…	…

word2vec

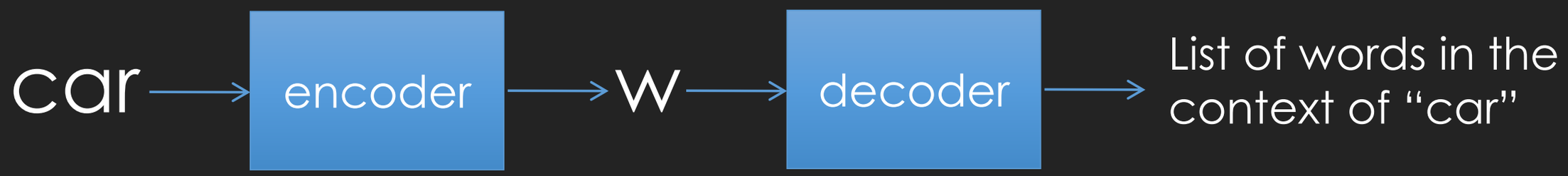
I parked the car in a nearby street. It is a red car with two doors, …

I parked the vehicle in a nearby street…

T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781, 2013

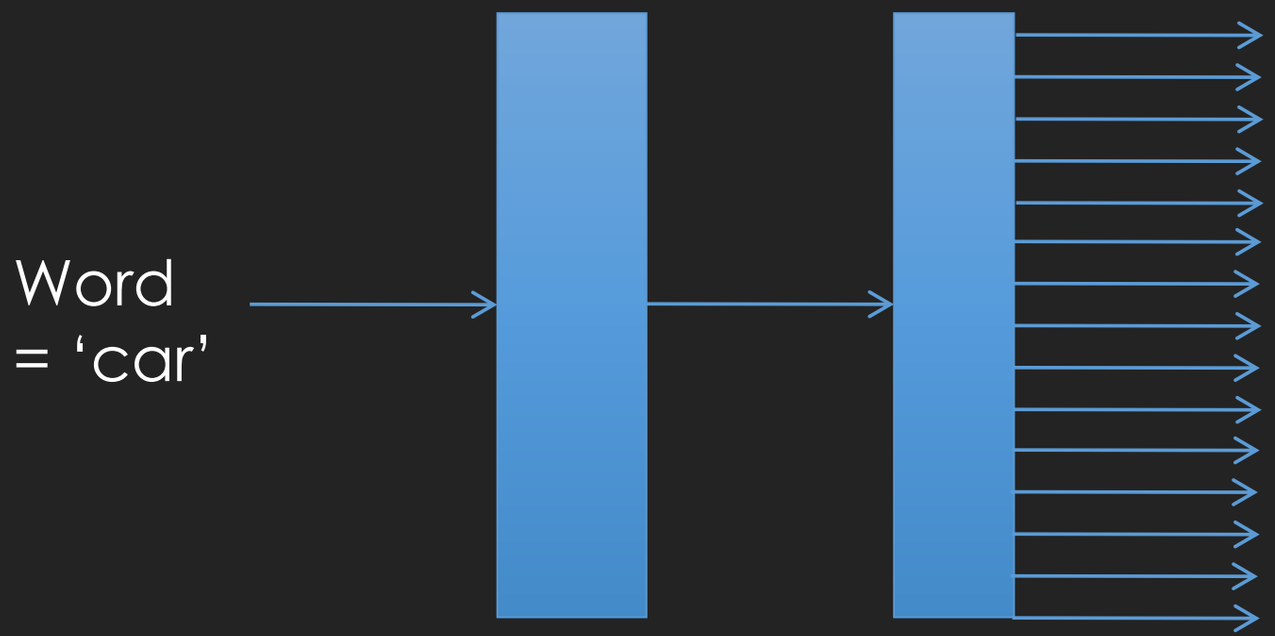
word2vec

I parked the car in a nearby street. It is a red car with two doors, …



T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781, 2013

word2vec



Output prob. That each word is in the context of the input word

Hidden layer Soft-max classifier

Encoder Decoder

T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space. arXiv:1301.3781, 2013

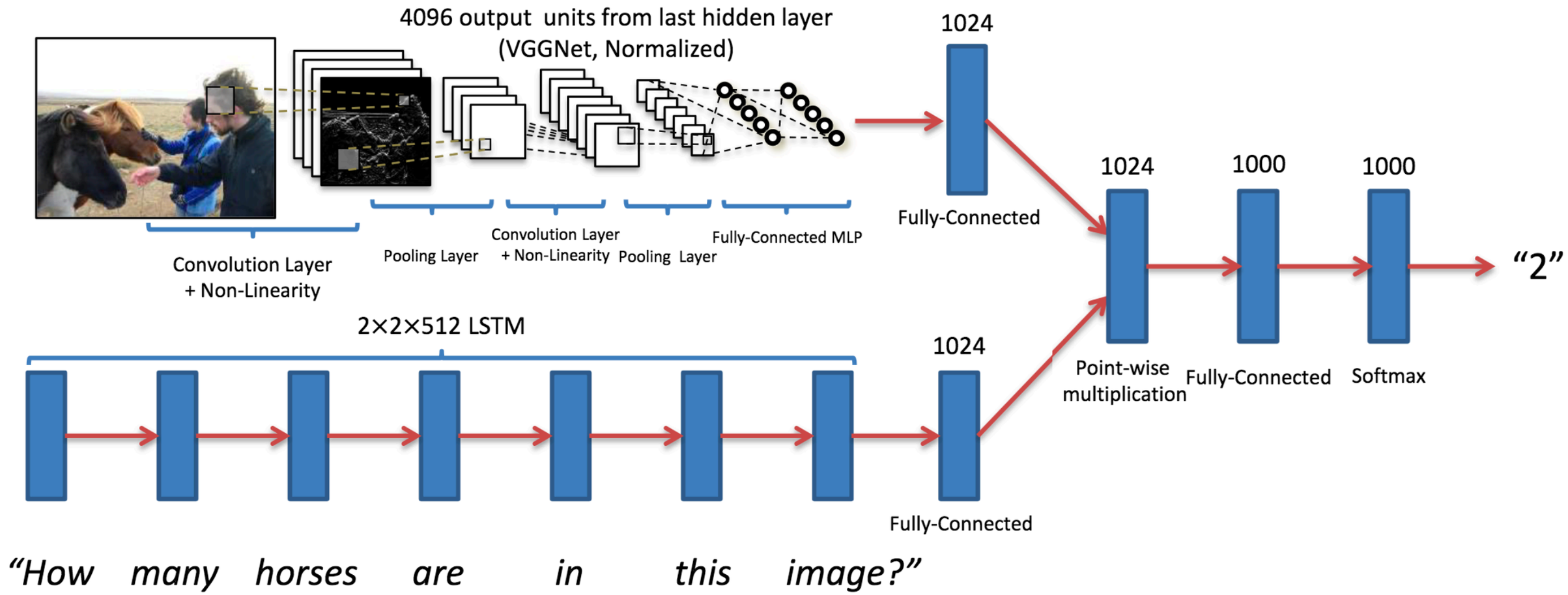
Algebraic operations with the vector representation of words

X = Vector(“Paris”) – vector(“France”) + vector(“Italy”)

Closest nearest neighbor to X is vector(“Rome”)

Architecture

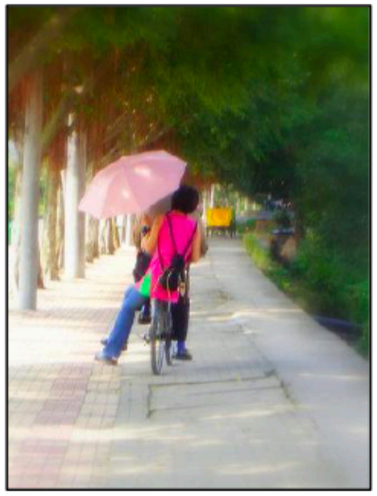
Architecture



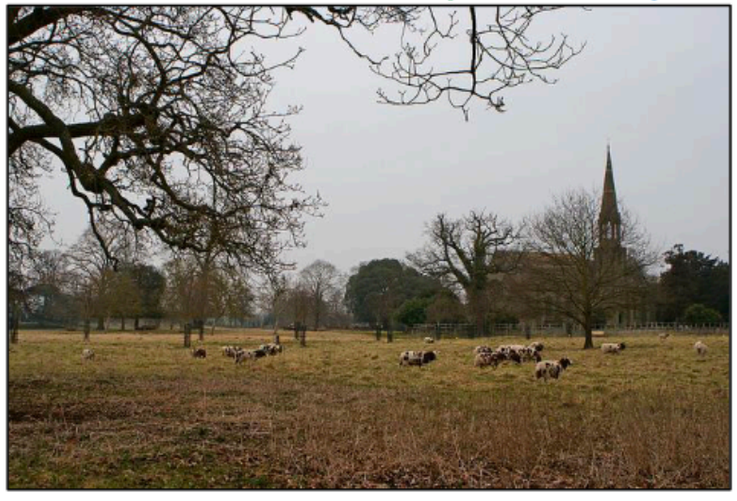
There are 1000 possible answers in this system.
Questions are unlimited.

What red objects in front are almost covered by snow?	meters parking meters parking meters	car cars shoes
Is it winter?	yes yes yes	no yes yes

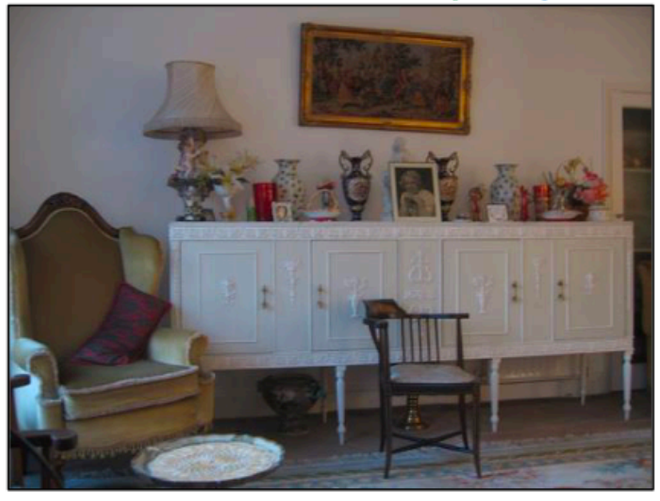
Does the car have a license plate?	yes yes yes	yes yes yes
Could the truck have a camper?	yes yes yes	yes yes yes



Is the woman on the back of the bicycle pedaling?	no no yes	no no yes
Why is the woman holding an umbrella?	sunny to block sun uncertain	it's raining it's raining to stay dry



Is this photo taken in Antarctica?	no no no	no yes yes
Overcast or sunny?	overcast overcast overcast	overcast overcast sunny



Is the picture hanging straight?	no yes yes	no yes yes
How many cabinets are on the piece of furniture?	4 4 4	3 3 6

What type of trees are here?	palm palm palm	ash oak pine
Is the skateboard airborne?	yes yes yes	no yes yes

Fig. 27: Random examples of questions (black), (a subset of the) answers given when looking at the image (green), and answers given when not looking at the image (blue) for numerous representative examples of the real image dataset.

what is on the ground?

Submit

what is on the ground?

Submit

Predicted top-5 answers with confidence:

sand

90.748%

snow

2.858%

beach

1.418%

surfboards

0.677%

water

0.528%

what color is the umbrella?

Submit

what color is the umbrella?

Submit

Predicted top-5 answers with confidence:

yellow

95.090%

white

1.811%

black

0.663%

blue

0.541%

gray

0.362%

are we alone in the universe?

Submit

are we alone in the universe?

Submit

Predicted top-5 answers with confidence:

no

78.234%

yes

21.763%

people

0.001%

birds

0.000%

out

0.000%

what is the meaning of life?

Submit

what is the meaning of life?

Submit

Predicted top-5 answers with confidence:

beach

15.262%

sand

8.537%

seagull

4.708%

tower

2.393%

rocks

1.746%

what is the yellow thing?

Submit

Predicted top-5 answers with confidence:

frisbee

79.844%

surfboard

7.319%

banana

2.844%

lemon

2.438%

surfboards

1.252%

how many trains are in the picture?

Submit

Predicted top-5 answers with confidence:

3

30.233%

5

18.270%

4

17.000%

2

11.343%

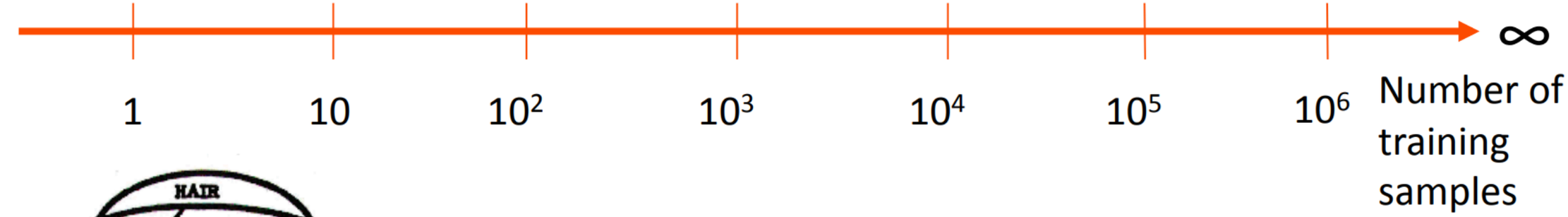
6

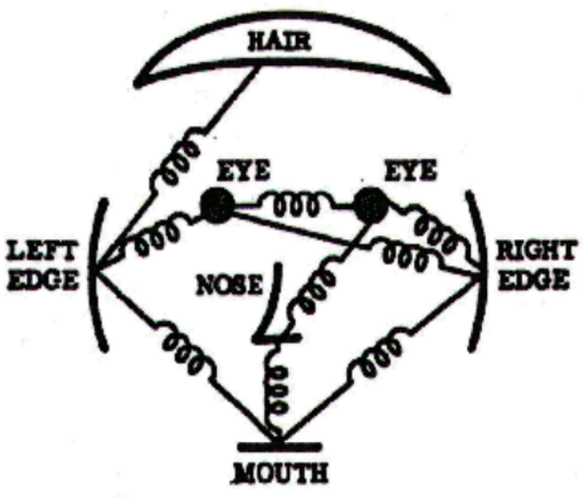
7.806%

Two Extremes of Vision

Extrapolation problem
Generalization
Diagnostic features

Interpolation problem
Correspondence
Finding the differences





Slide credit: Aude Oliva