Machine learning for remote sensing II

Prof. Biplab Banerjee

Computer Vision

Make computers understand images and video.

What kind of scene?

Where are the cars?

How far is the building?

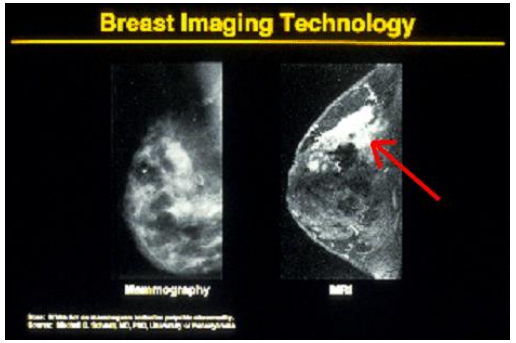
…

Vision is really hard

- Vision is an amazing feat of natural intelligence
  - Visual cortex occupies about 50% of Macaque brain
  - More human brain devoted to vision than anything else

Why computer vision matters

Safety

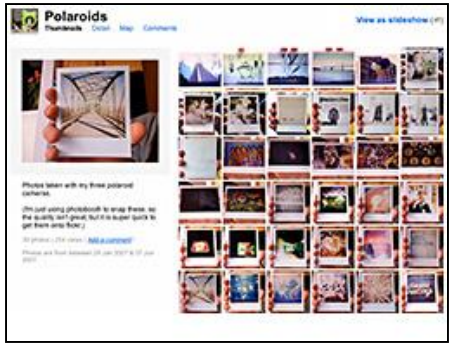


Health

Security

Comfort

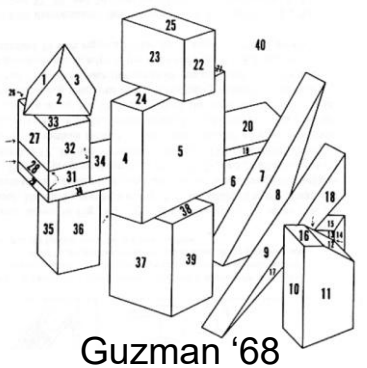
Fun



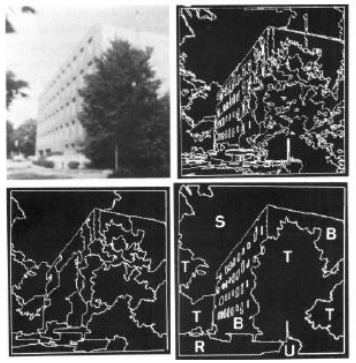
Access

Ridiculously brief history of computer vision

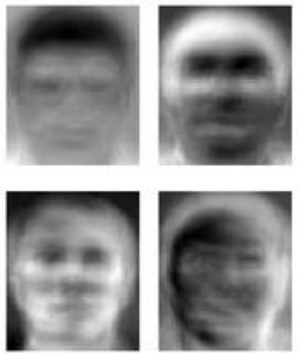
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

How vision is used now

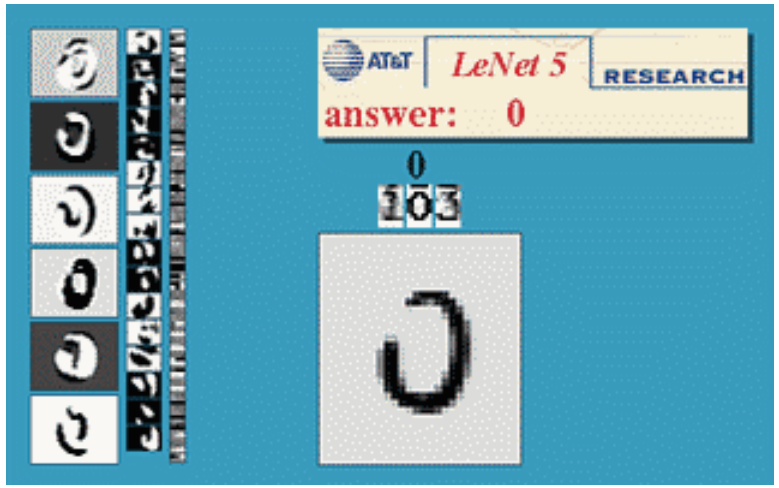
- Examples of state-of-the-art

Some of the following slides by Steve Seitz

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



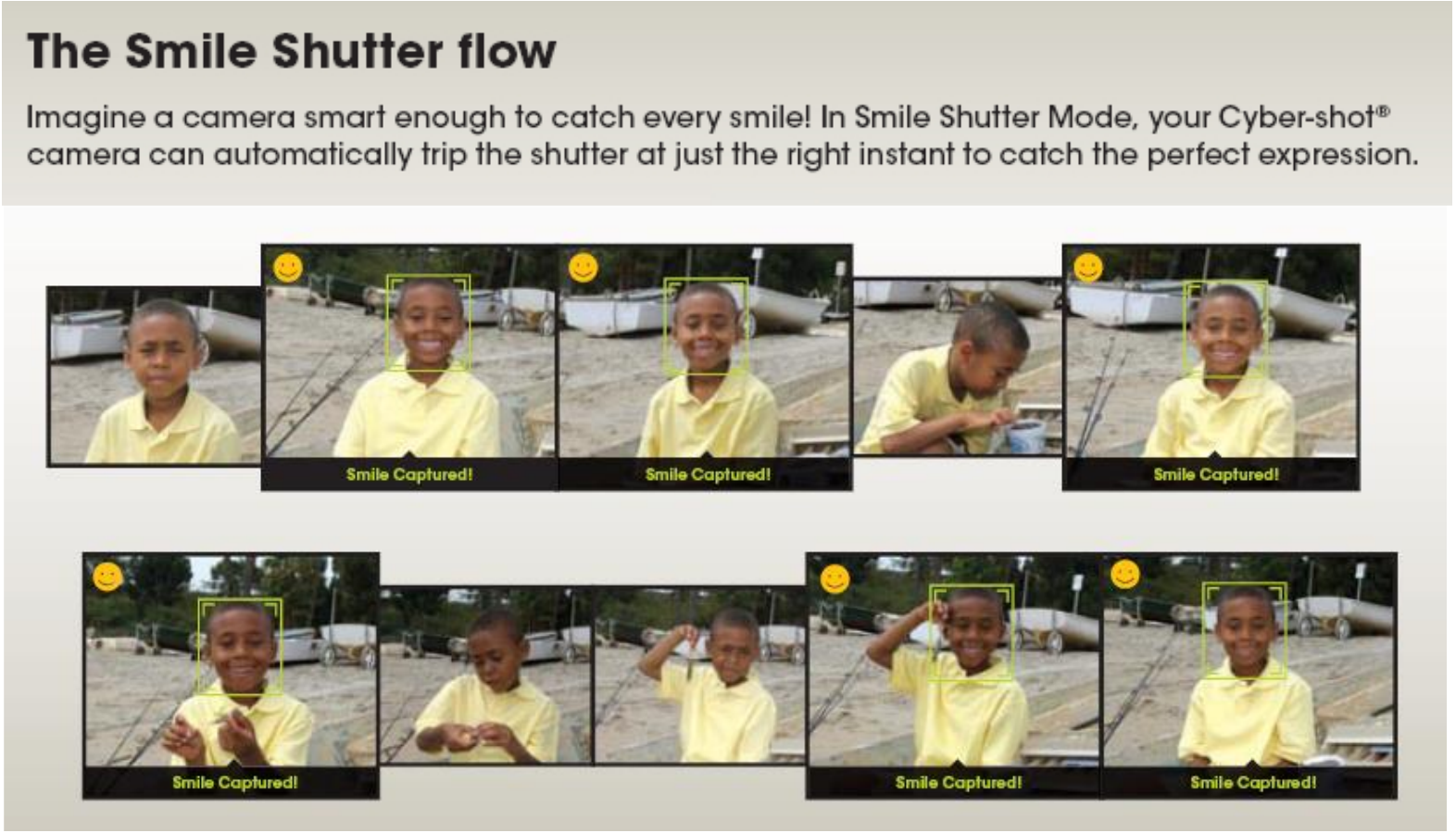
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Smile detection



[Sony Cyber-shot® T70 Digital Still Camera]

3D from thousands of images

Building Rome in a Day: Agarwal et al. 2009

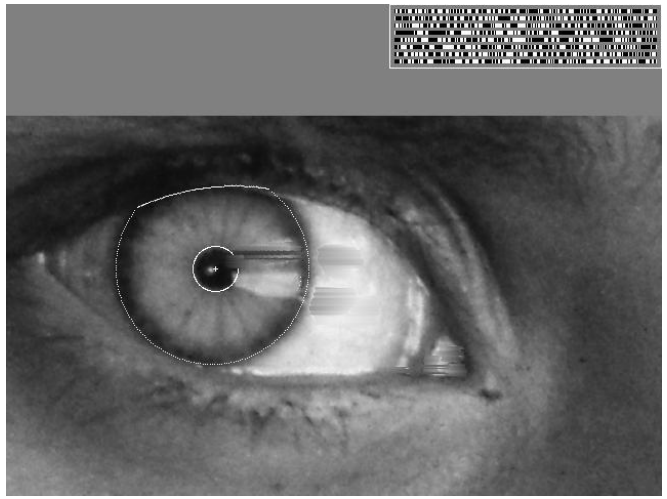
Object recognition (in supermarkets)

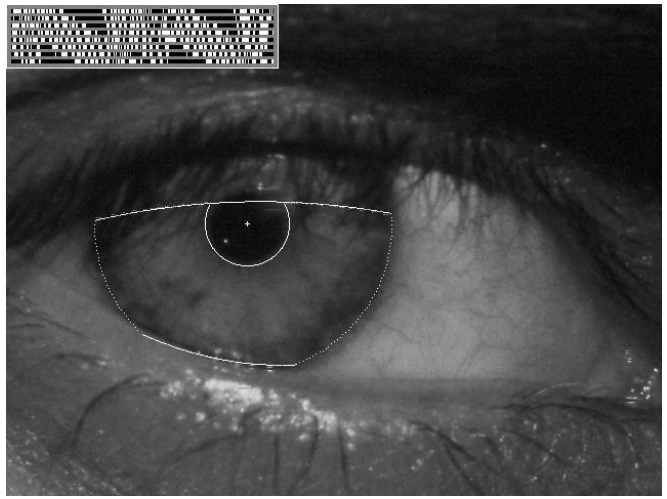
LaneHawk by EvolutionRobotics

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… “

Vision-based biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story wikipedia





Login without a password…

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

[Google Goggles]()

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html

Smart cars

Slide content courtesy of Amnon Shashua

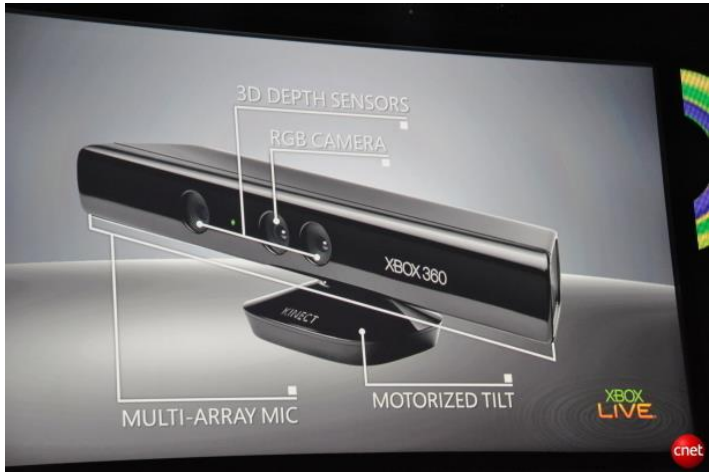
- Mobileye
  - Vision systems currently in high-end BMW, GM, Volvo models
  - By 2010: 70% of car manufacturers.

Google cars

http://www.nytimes.com/2010/10/10/science/10google.html?ref=artificialintelligence

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

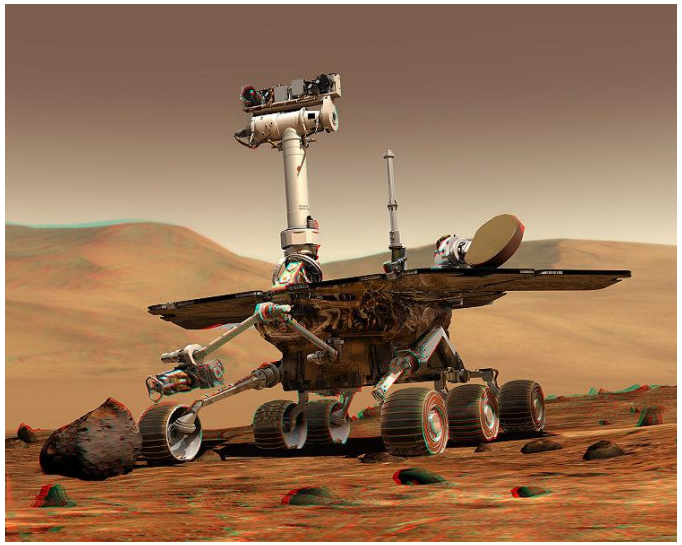
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “Computer Vision on Mars” by Matthies et al.

Industrial robots

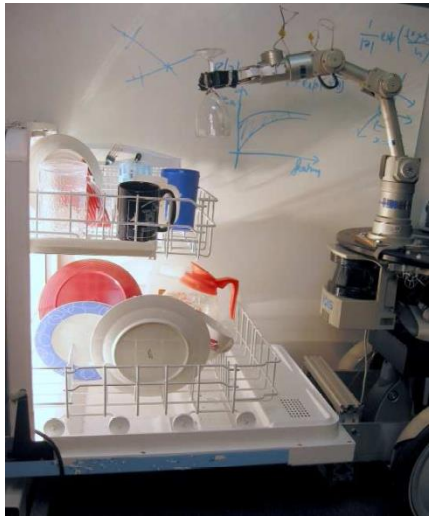
Vision-guided robots position nut runners on wheels

Mobile robots



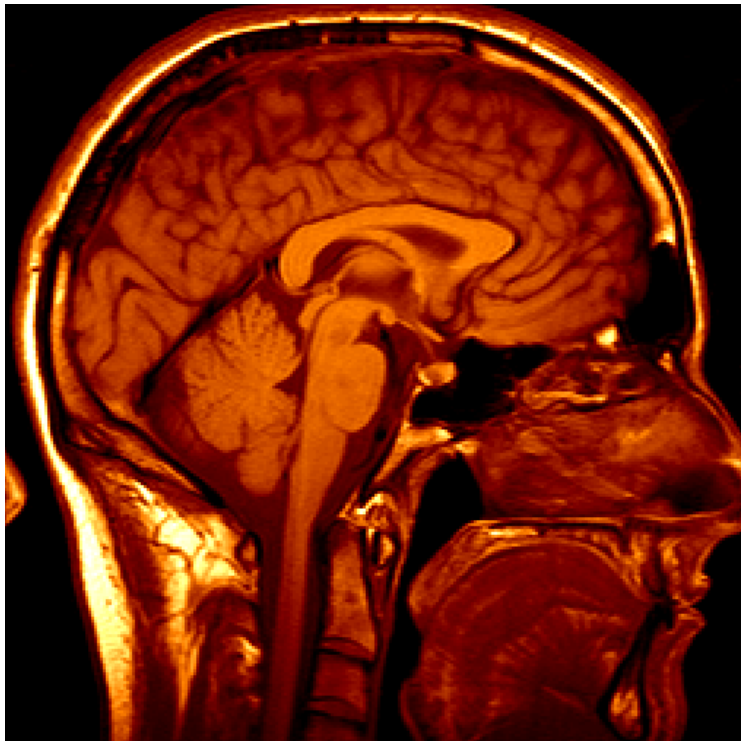
NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

http://www.robocup.org/



Saxena et al. 2008
STAIR at Stanford

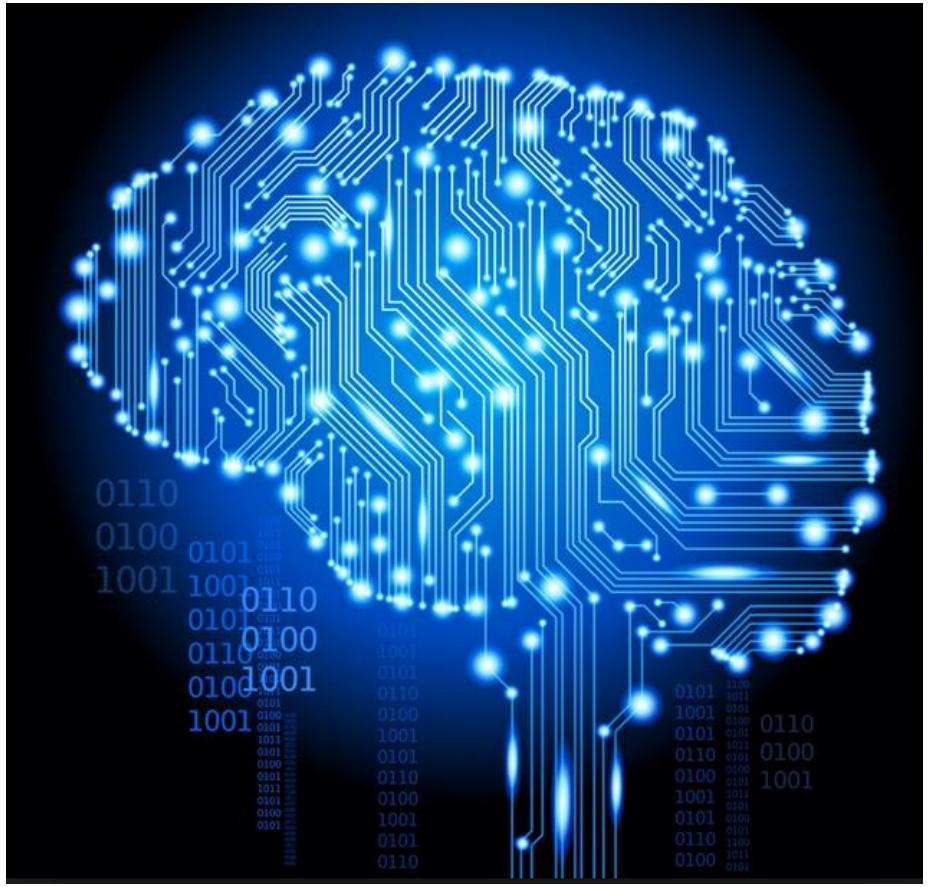
Medical imaging



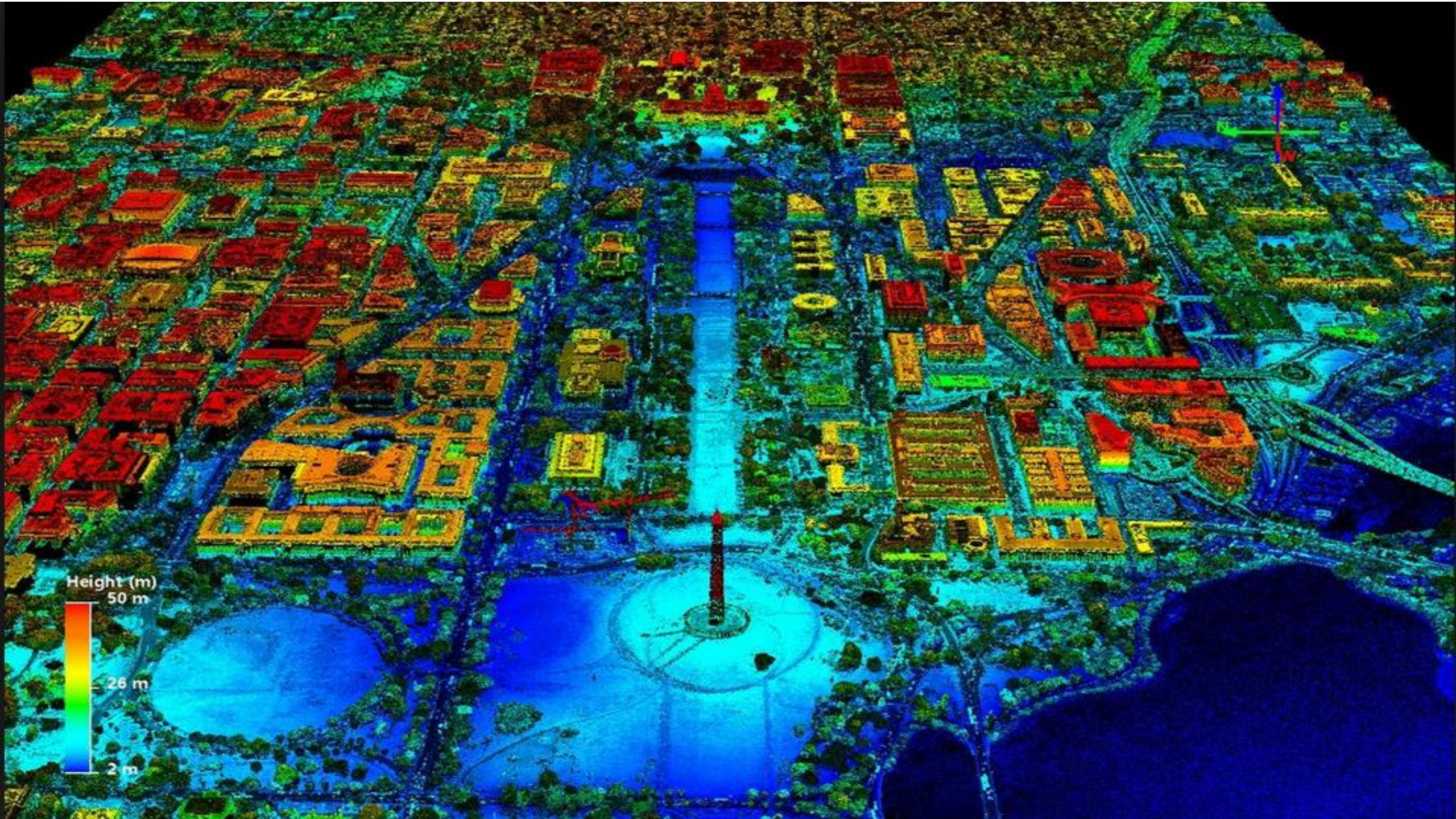
3D imaging
MRI, CT

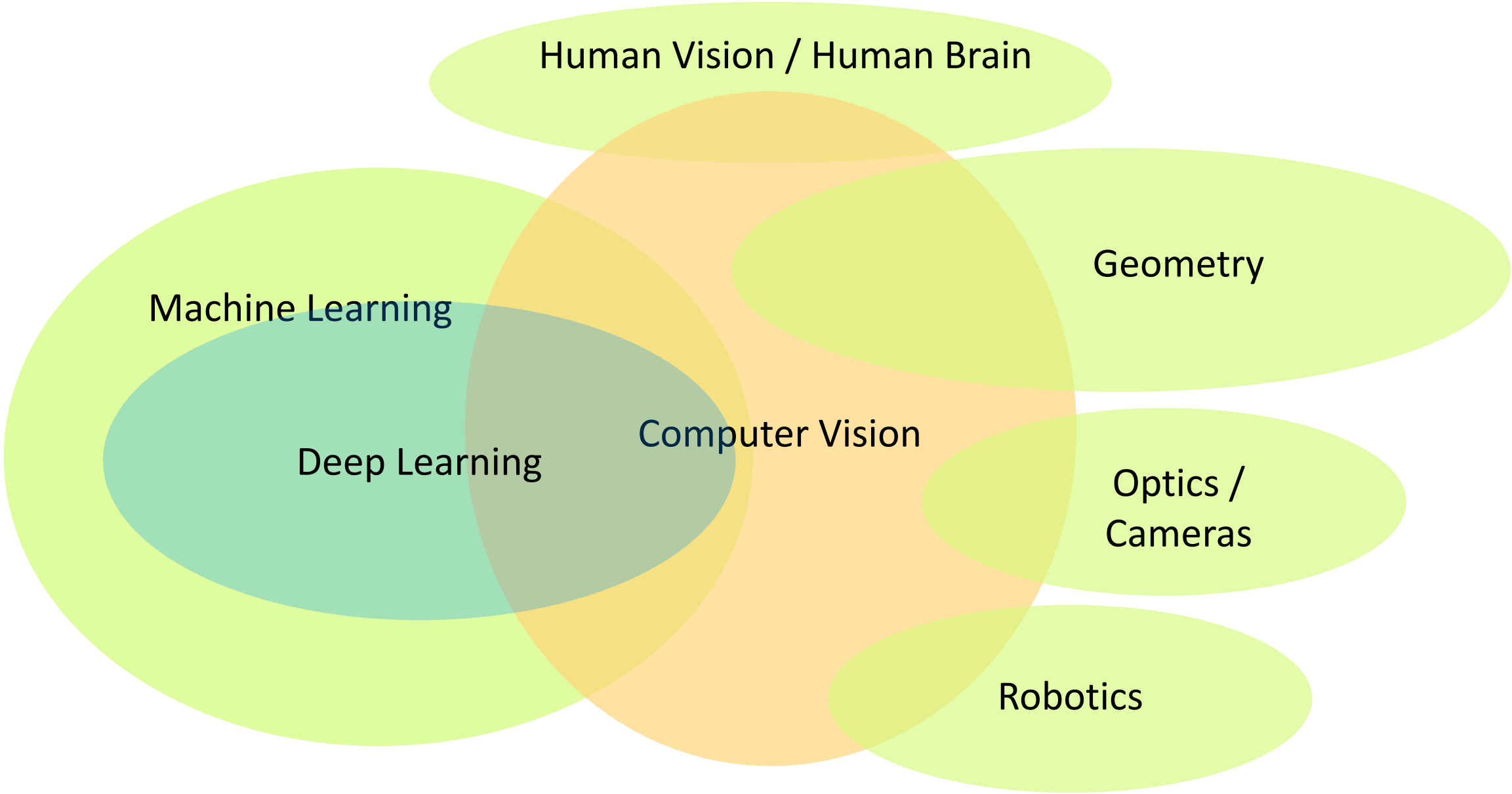
Image guided surgery
Grimson et al., MIT

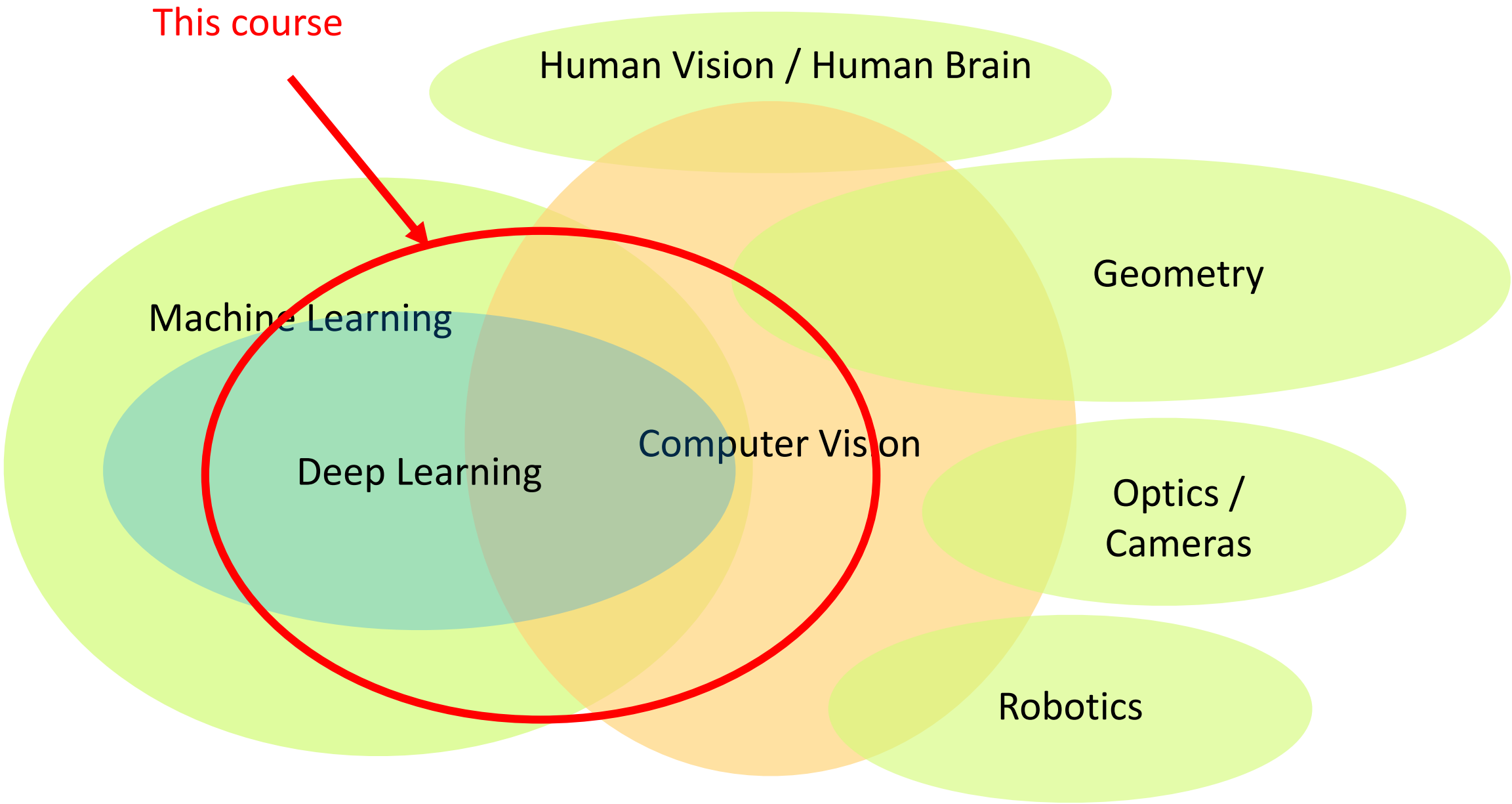
Deep learning for visual inference




The European Camera of the Year 1983.
Fortunately, fine photography is a universal language.
PENTAX
PENTAX


Height (m)
50 m
26 m
2 m
N
S
W


Human Vision / Human Brain
Machine Learning
Geometry
Deep Learning
Computer Vision
Optics /
Cameras
Robotics

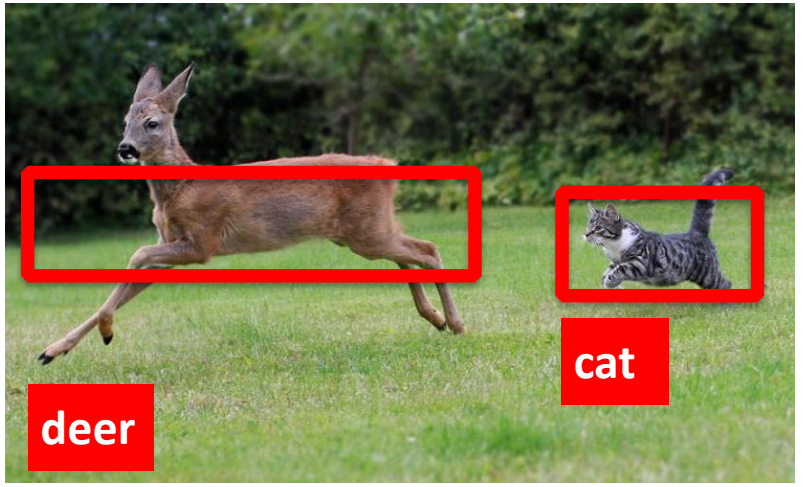

This course
Human Vision / Human Brain
Machine Learning
Geometry
Deep Learning
Computer Vision
Optics /
Cameras
Robotics

Relationship with Other Fields

- Image Processing: Image → Image

Relationship with Other Fields

- Computer Vision: Image → Knowledge



Relationship with Other Fields

- Computer Graphics: Knowledge → Image

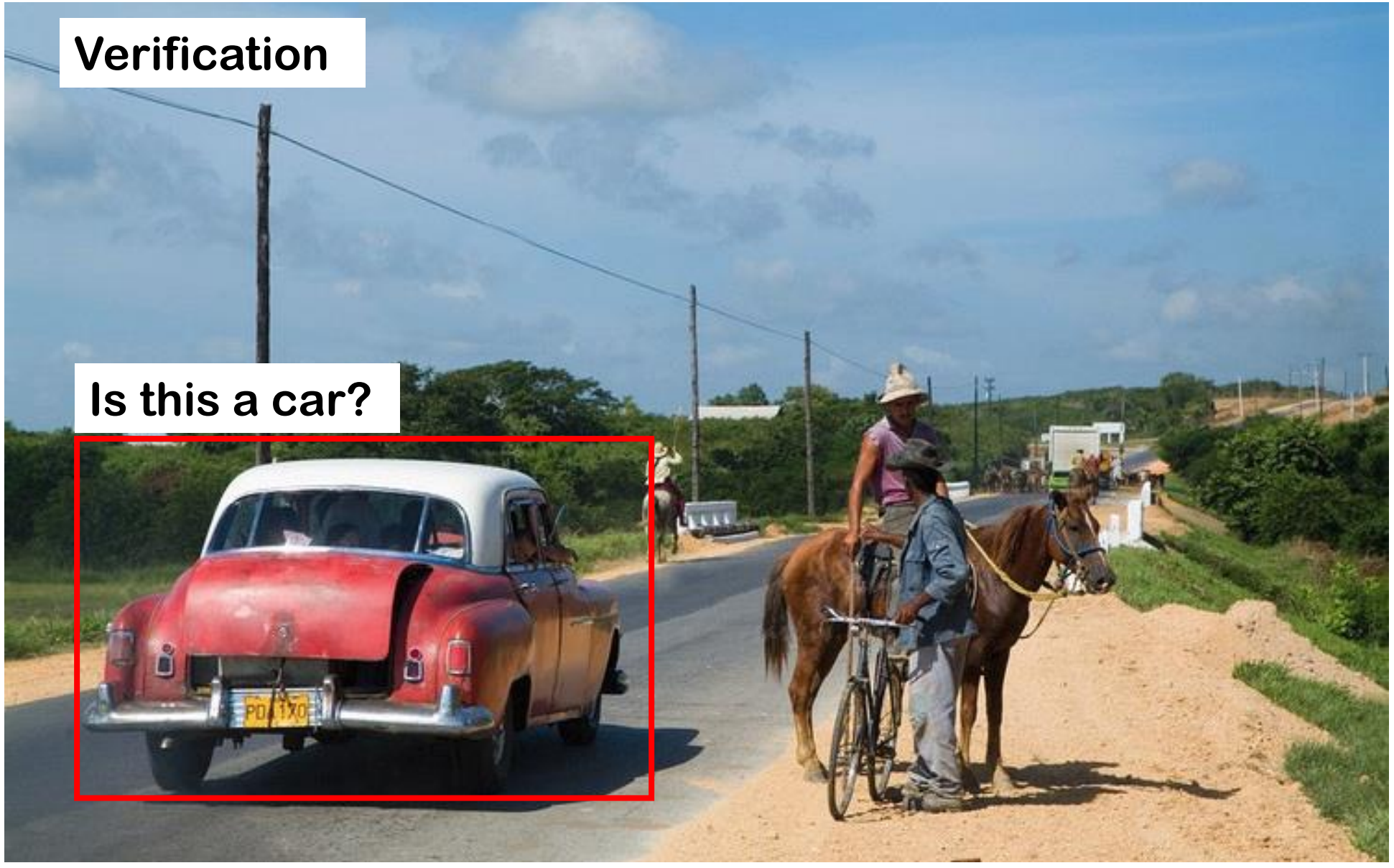
Vertices, Locations, Objects,
Shapes, Colors, Material properties,
Lighting settings, Camera settings, etc.

→

Visual Recognition?

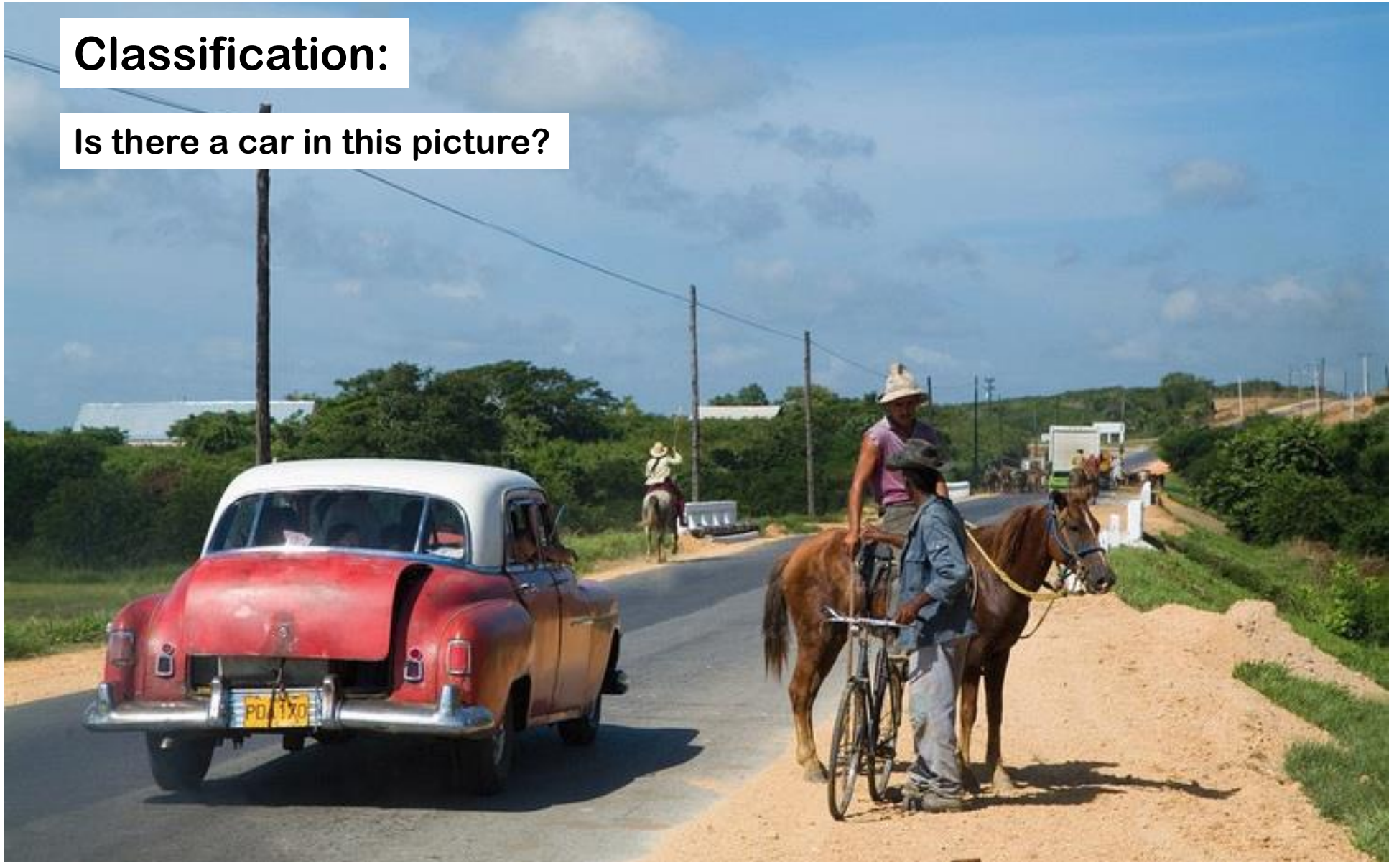
- What does it mean to “see”?
  - “What” is “where”, Marr 1982

- Get computers to “see”


Verification
Is this a car?

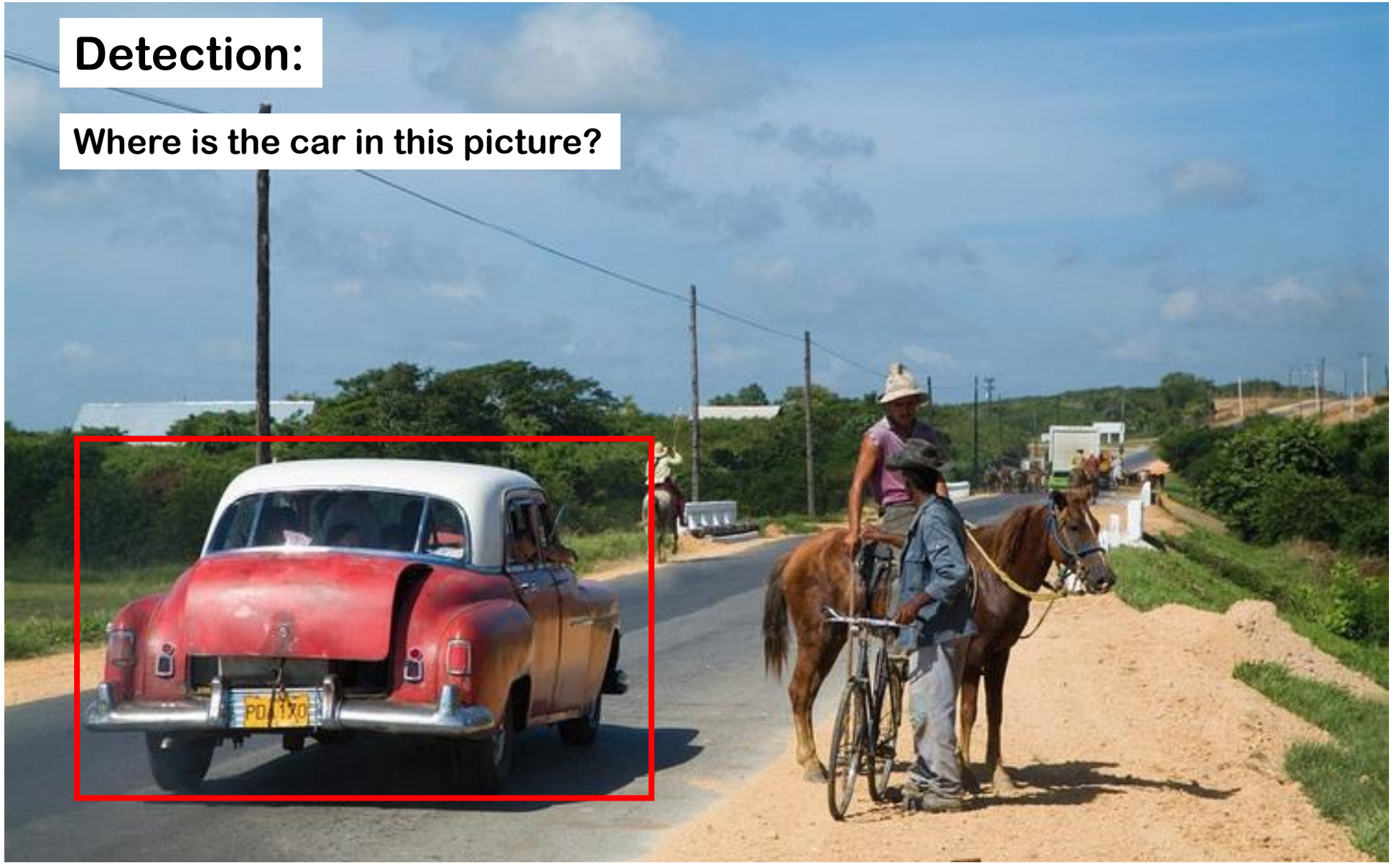
Classification:

Is there a car in this picture?

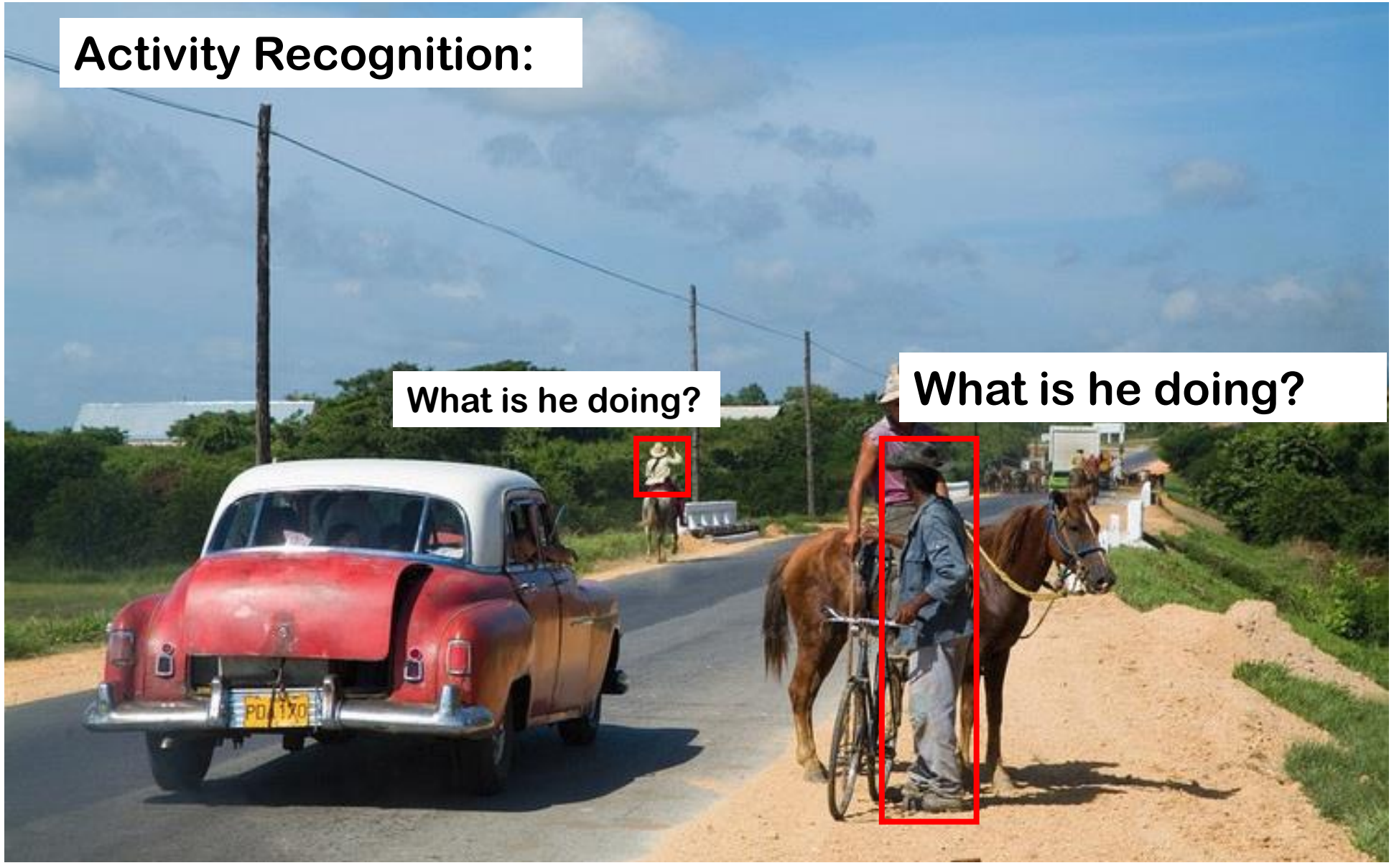


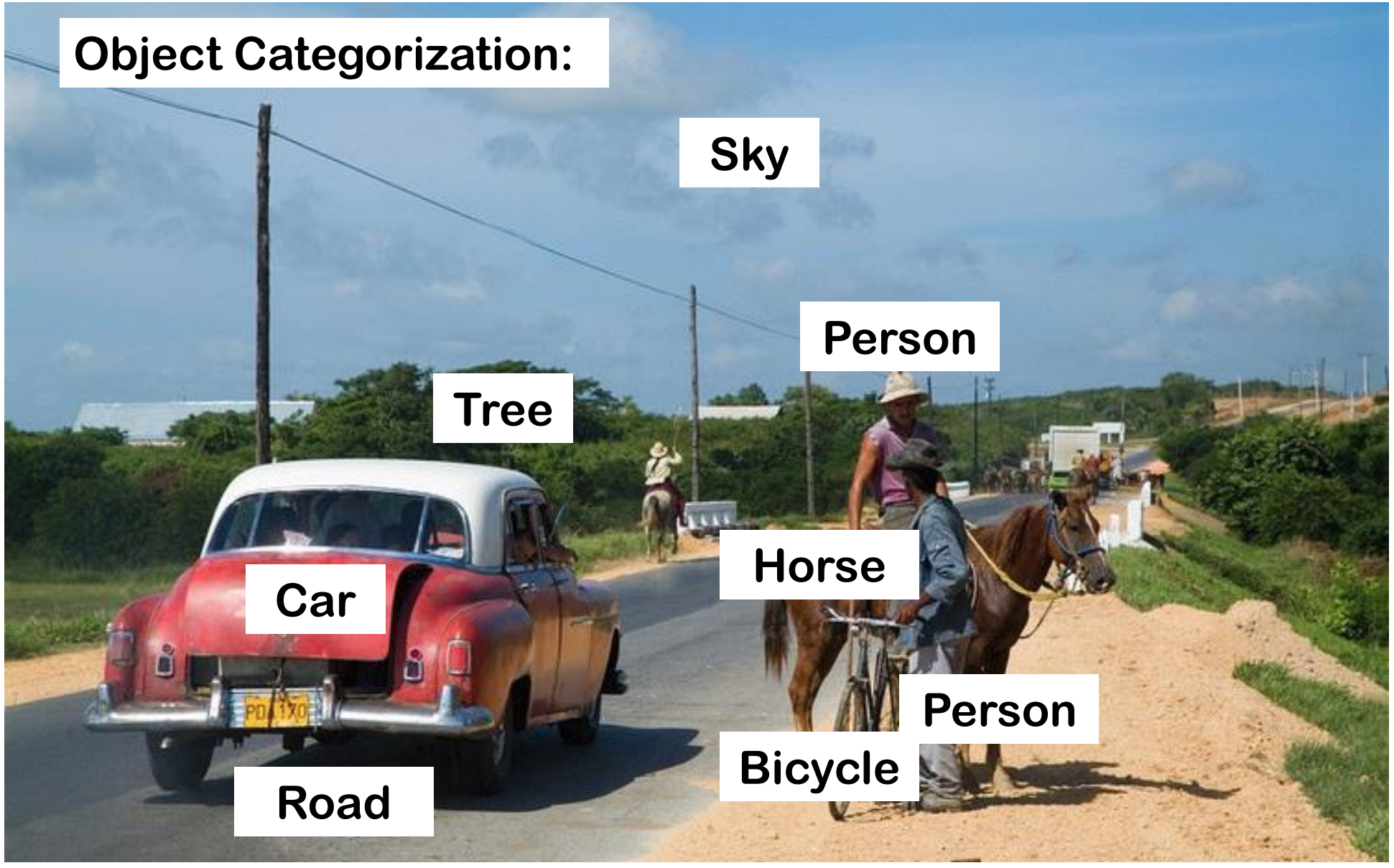
Detection:

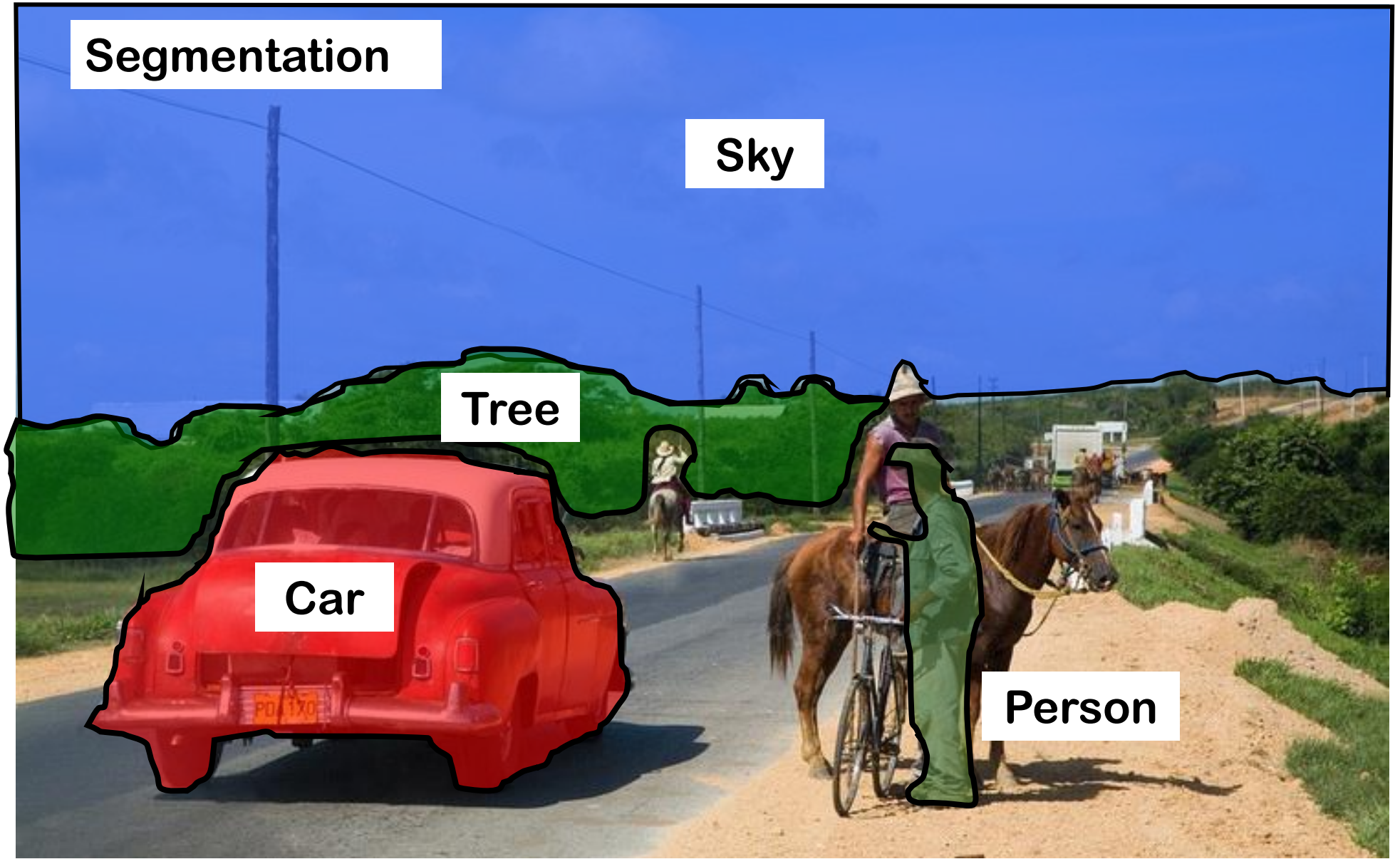
Where is the car in this picture?



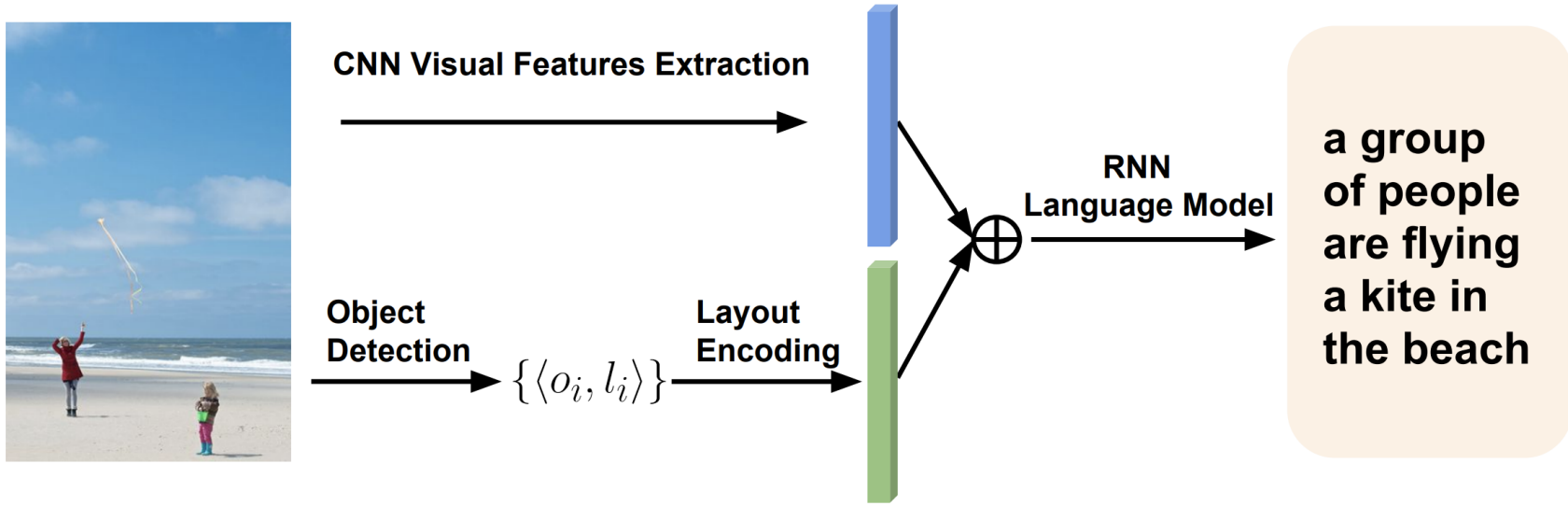

Pose Estimation:


Activity Recognition:
What is he doing?
What is he doing?

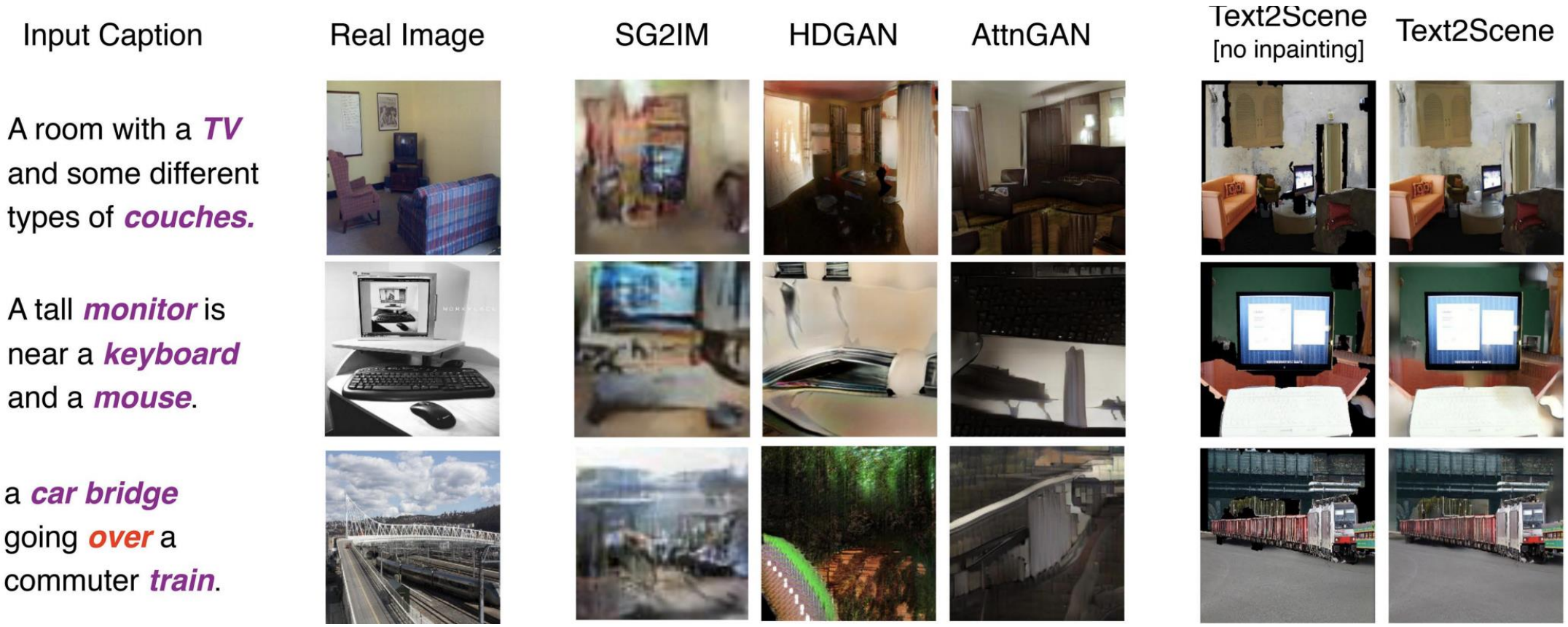

Object Categorization:
Sky
Person
Tree
Horse
Car
Person
Bicycle
Road


Segmentation
Sky
Tree
Car
Person

Describing Images with Language

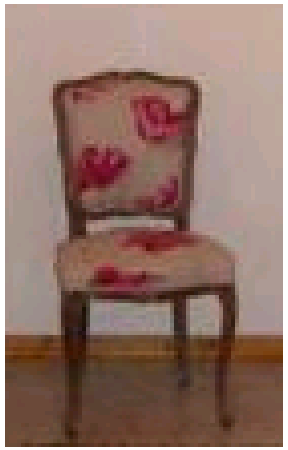


Text-to-Image Synthesis: Text2Scene



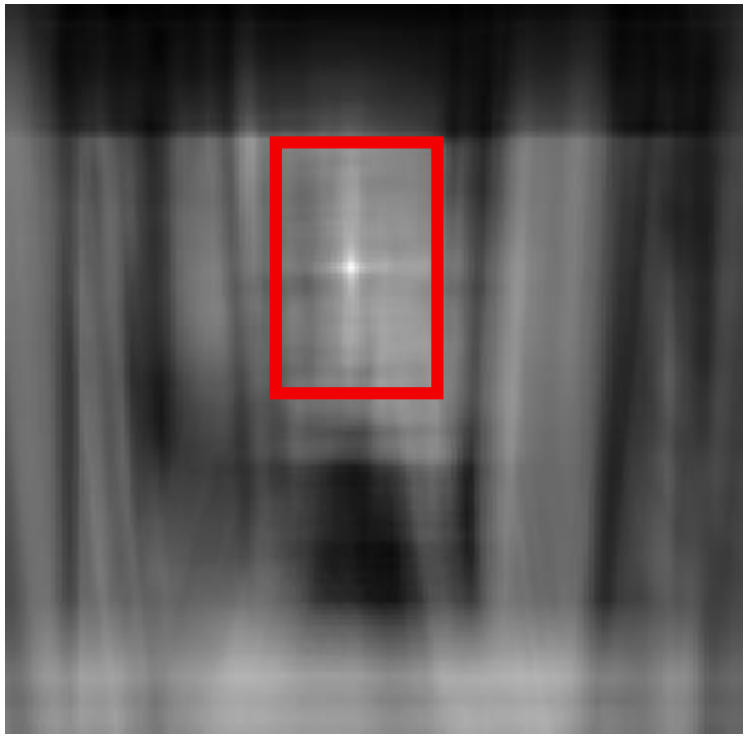
Object recognition
Is it really so hard?

This is a chair



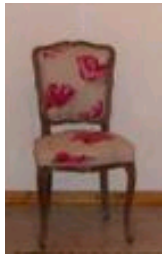
Find the chair in this image

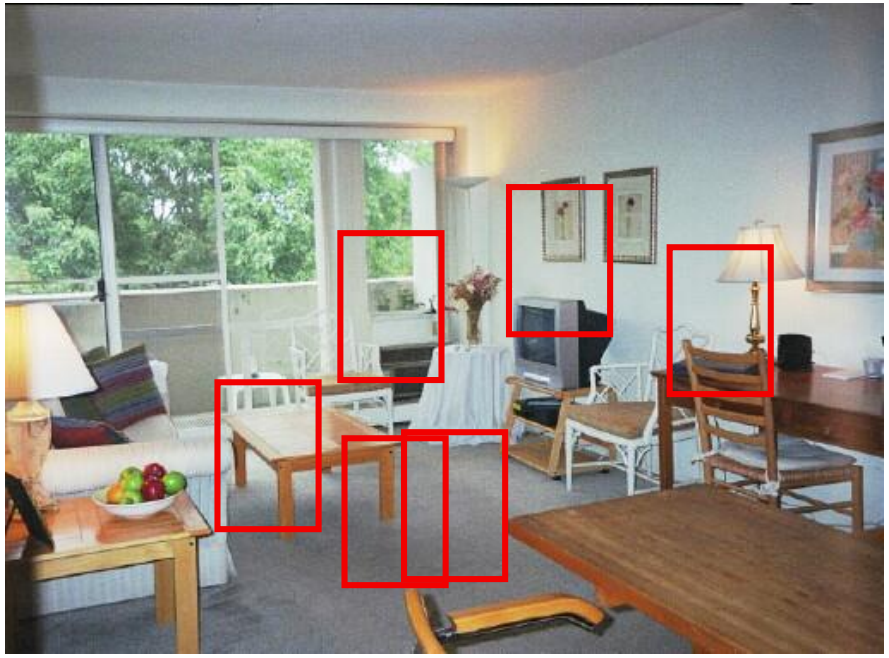
Output of normalized correlation

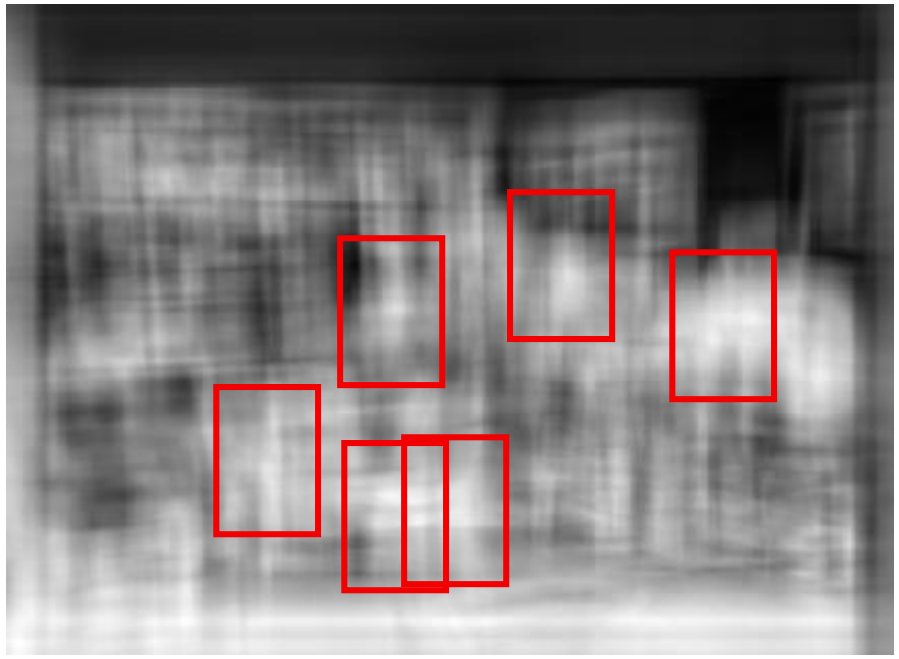


Object recognition
Is it really so hard?

Find the chair in this image



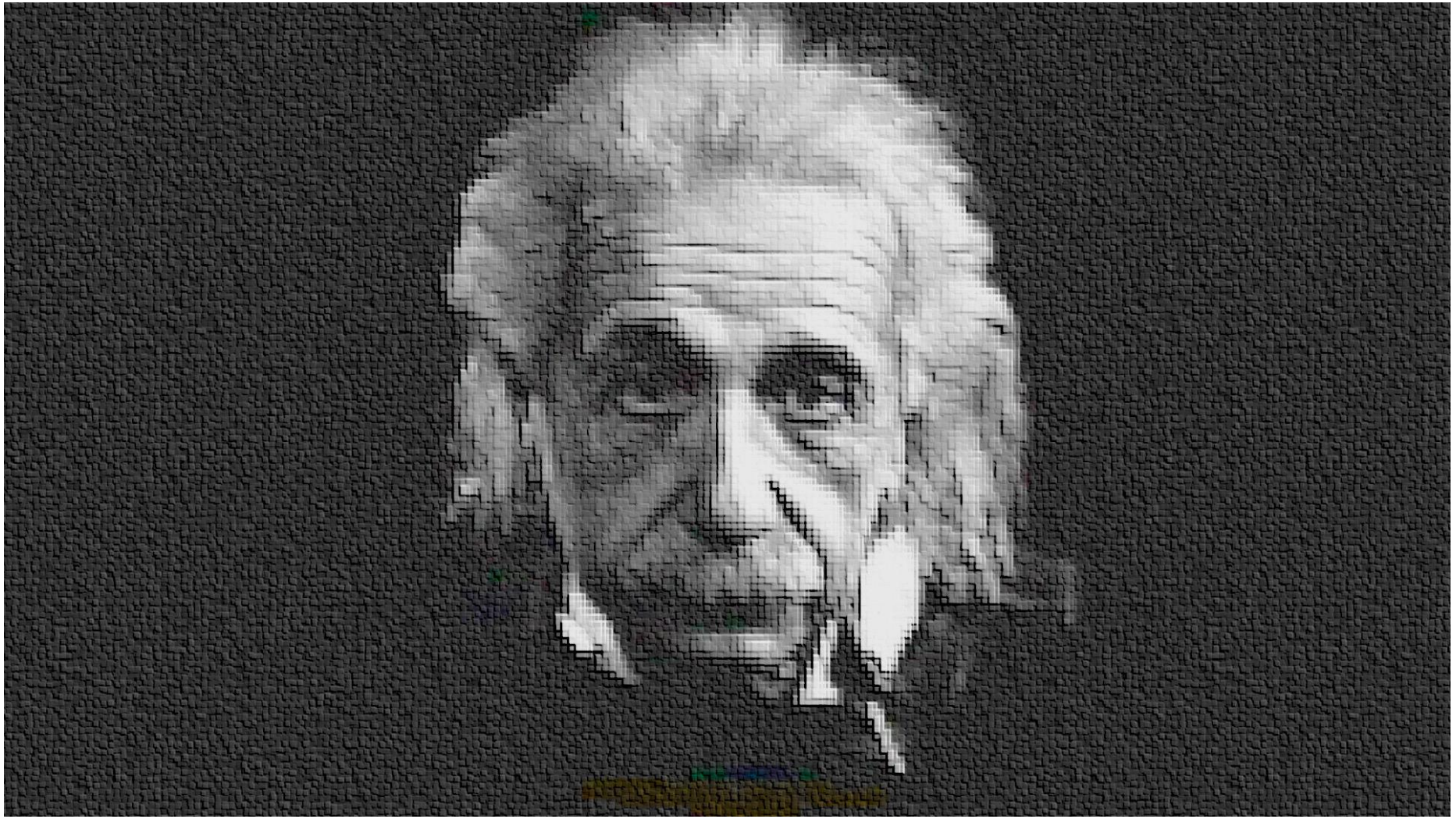




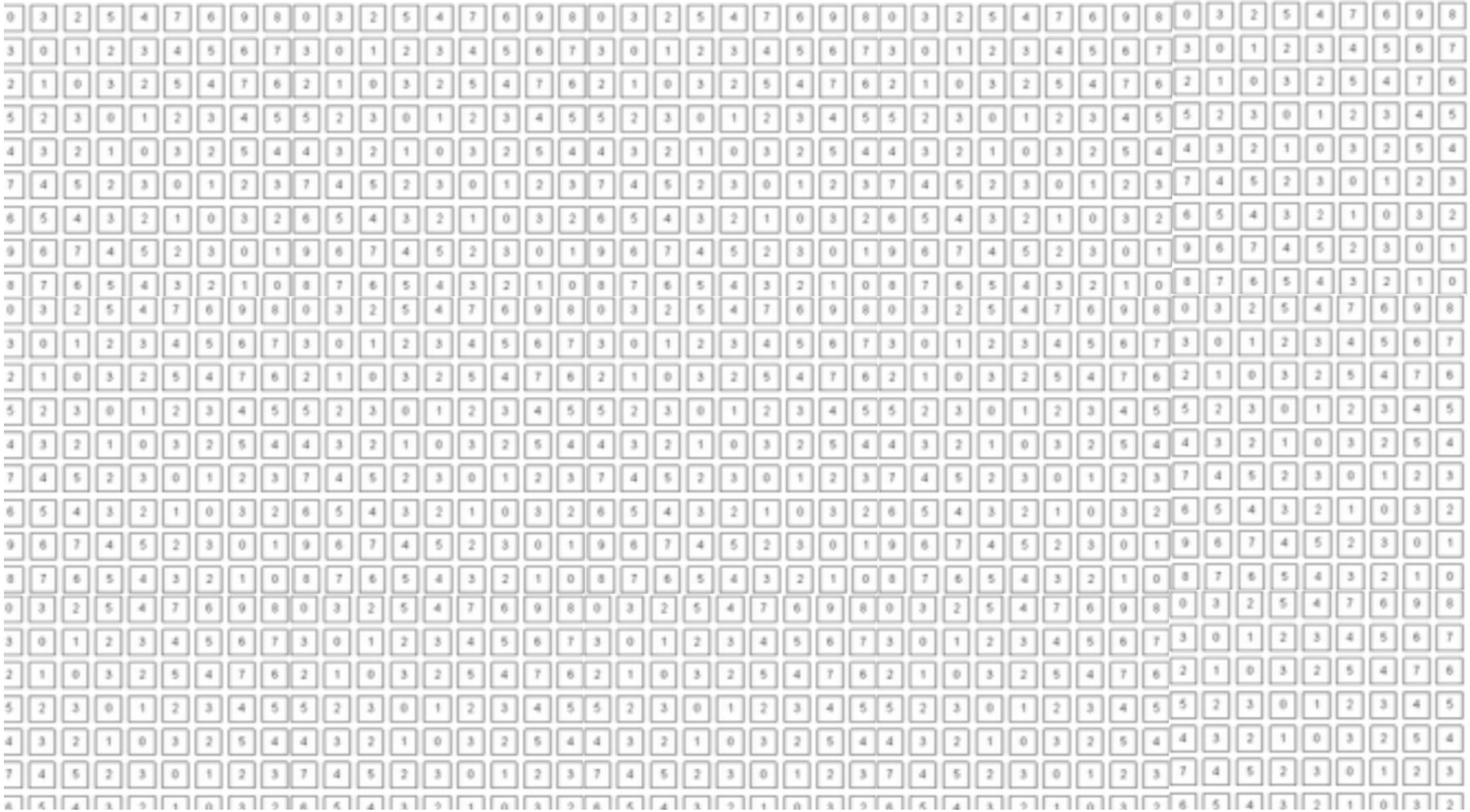
Pretty much garbage

Simple template matching is not going to make it

This is an image to us



This is an image to a computer



Challenges 1: view point variation

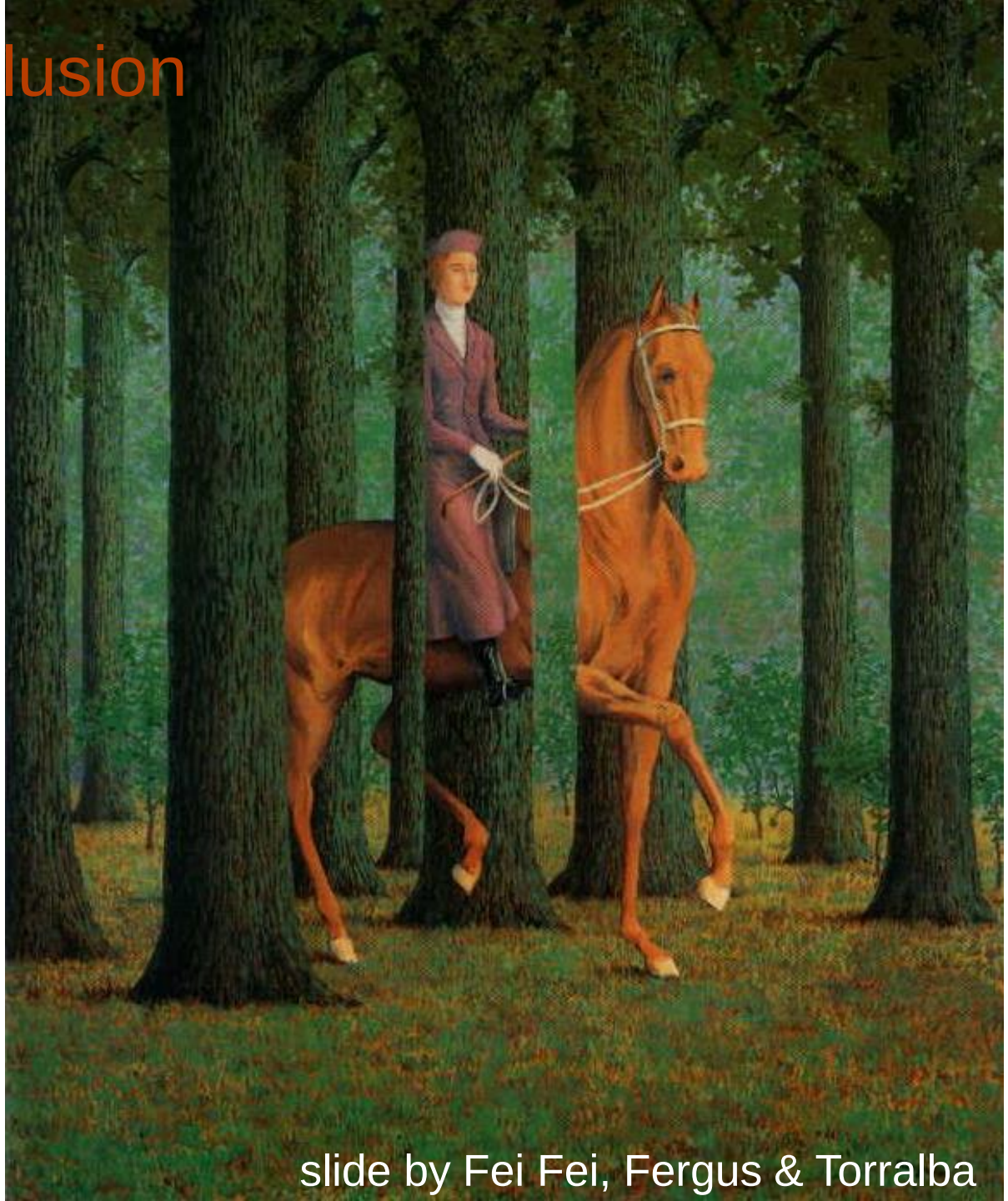
Michelangelo 1475-1564

slide by F

Challenges 2: illumination

slide credit: S. Ullman

Challenges 3: occlusion



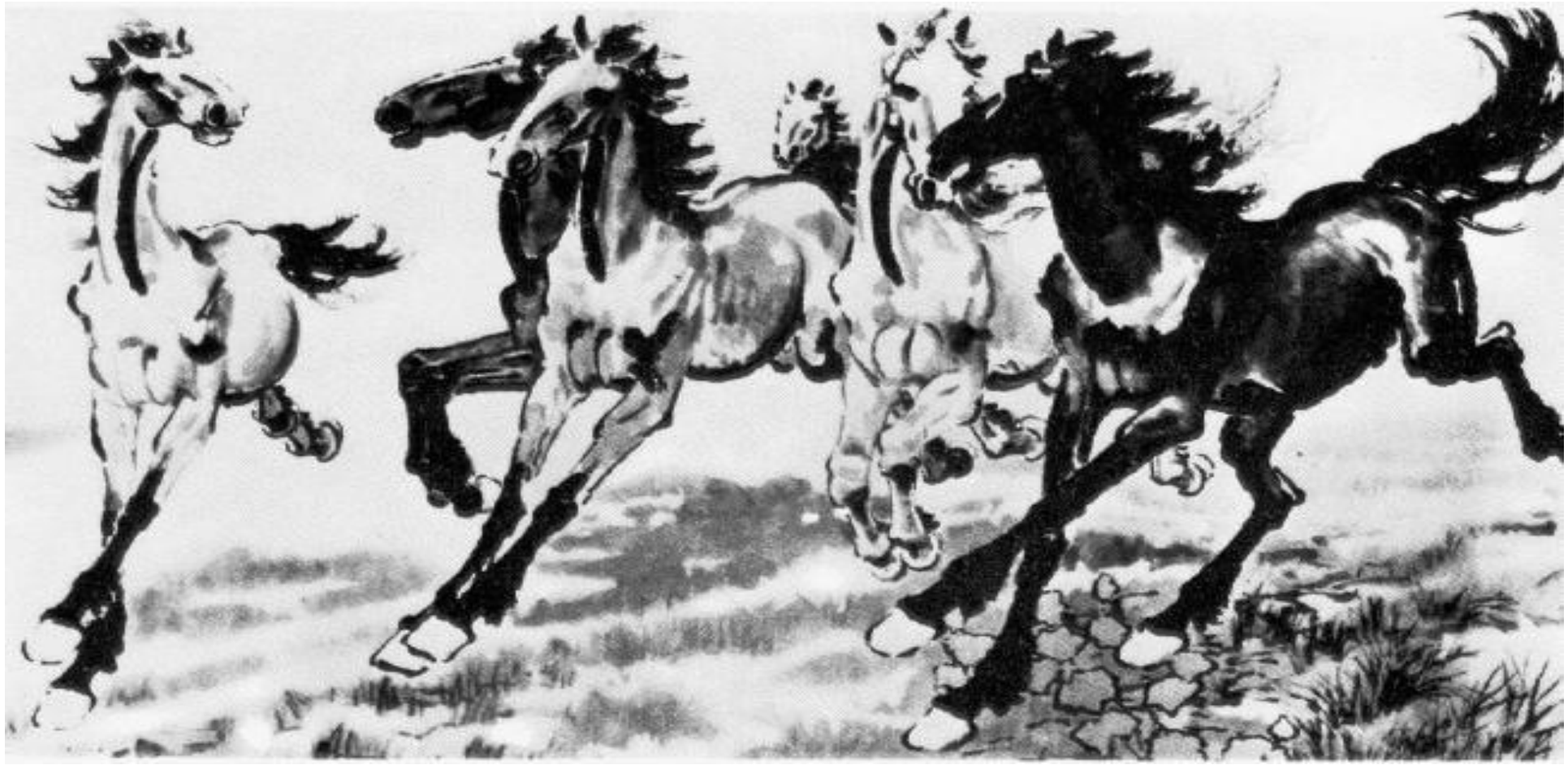
Magritte, 1957

slide by Fei Fei, Fergus & Torralba

Challenges 4: scale

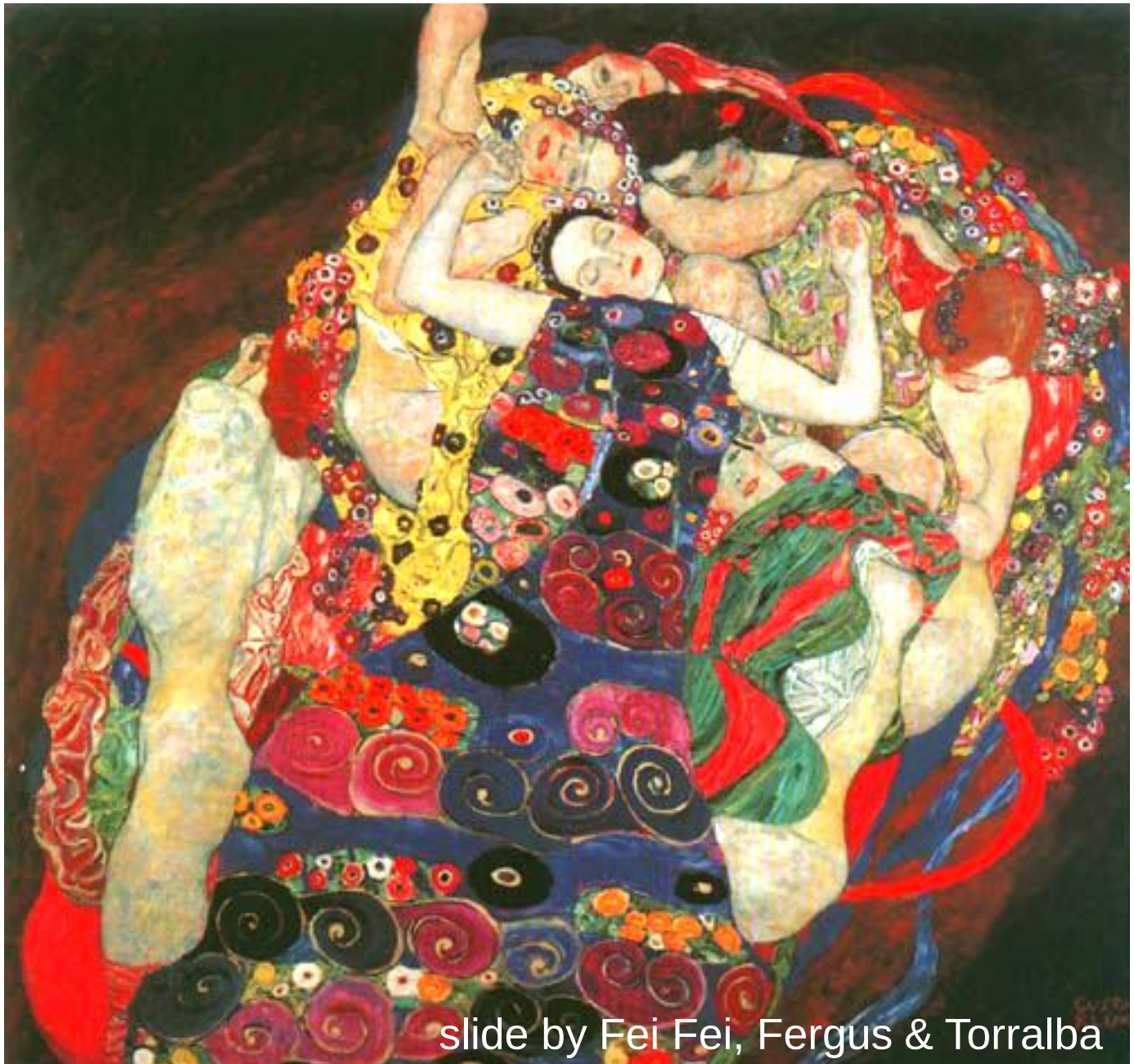
slide by Fei Fei, Fergus & Torralba

Challenges 5: deformation



Xu, Beihong 1943

Challenges 6: background clutter

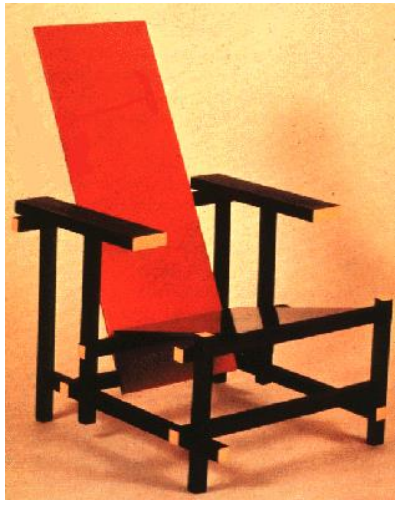


Klimt, 1913

slide by Fei Fei, Fergus & Torralba

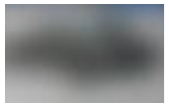
Challenges 7: object intra-class variation

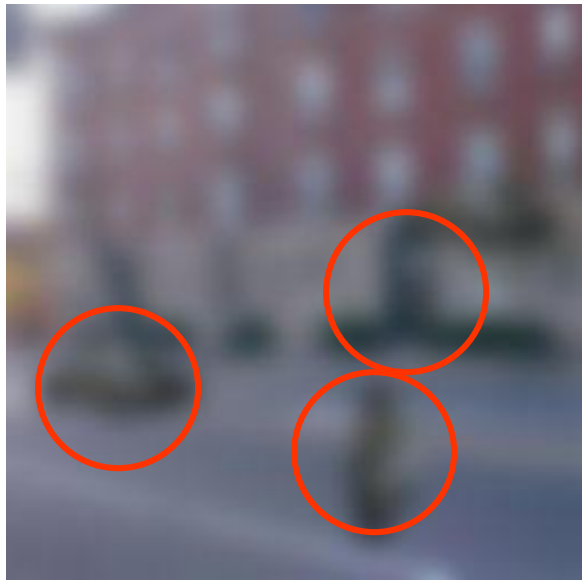


slide by Fei-Fei, Fergus & Torralba

Challenges 8: local ambiguity





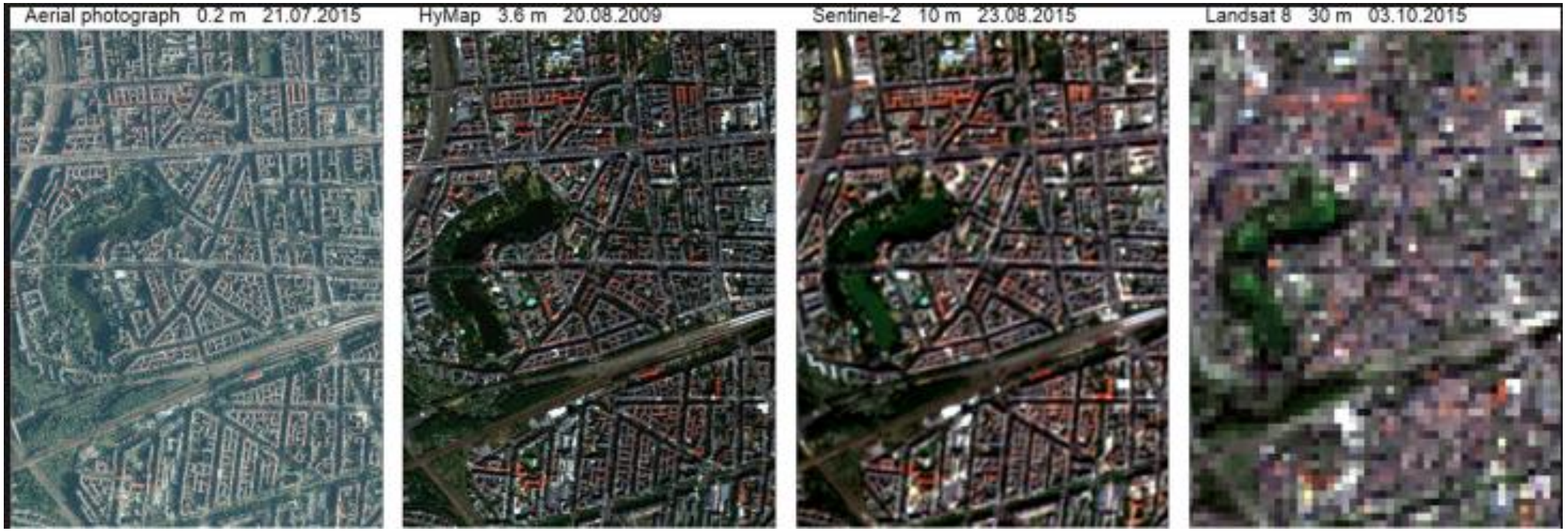
slide by Fei-Fei, Fergus & Torralba

Challenge 9 – labels are ambiguous

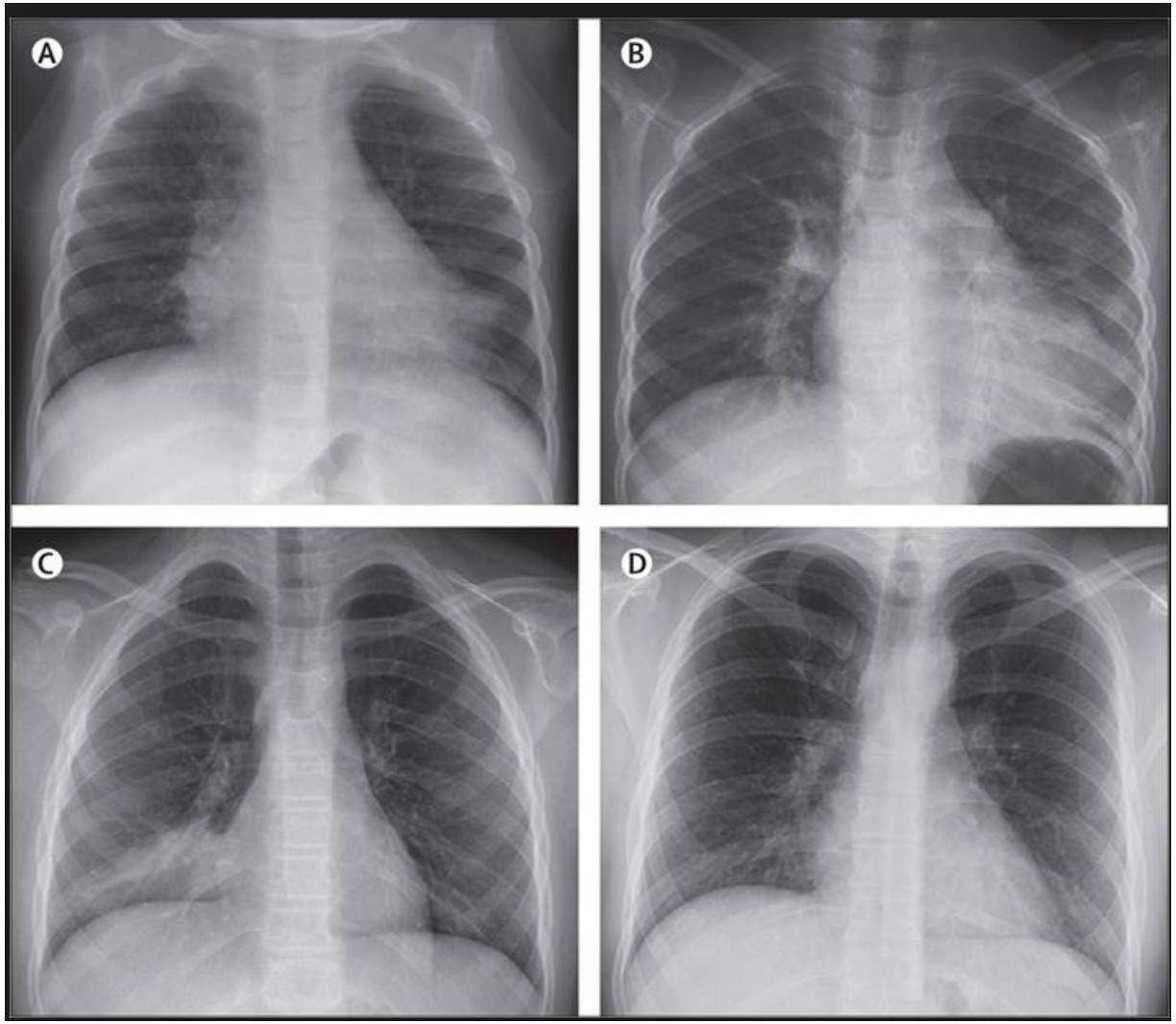
Superordinates: animal, vertebrate
Basic Level: bird
Entry Level: bird
Subordinates: American robin

Superordinates: animal, vertebrate
Basic Level: bird
Entry Level: penguin
Subordinates: Chinstrap penguin

Challenge 10 – different spatial resolution



Challenge 11 – medical images









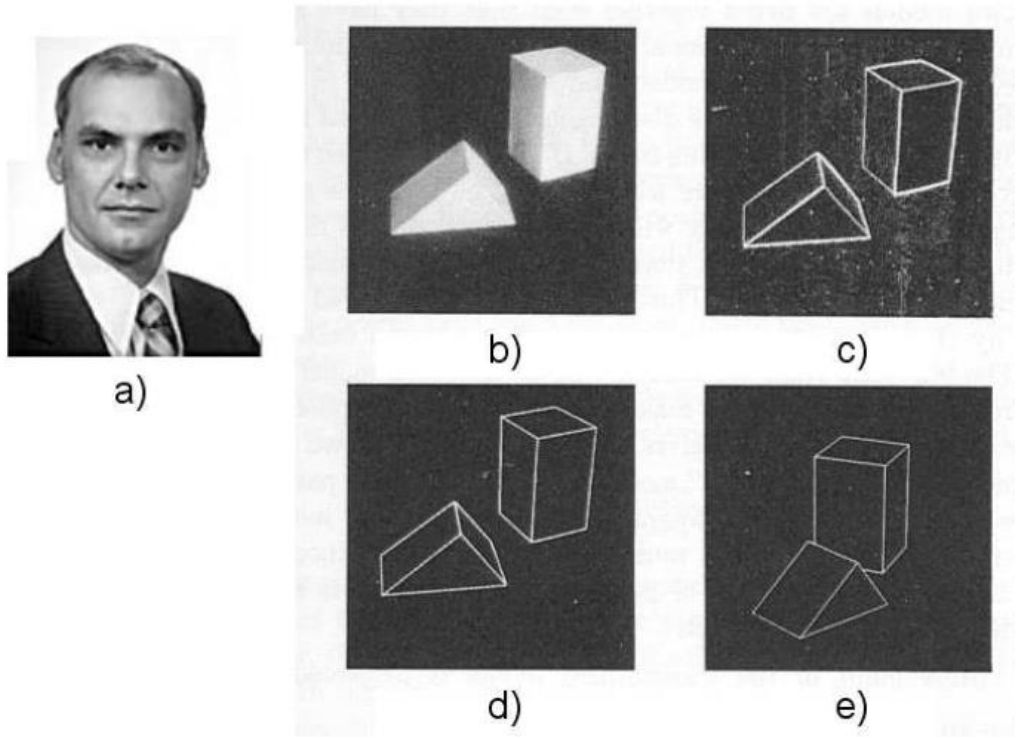




History of ideas in recognition

- **1960s – early 1990s: the geometric era**
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

Recognition as an alignment problem: Block world



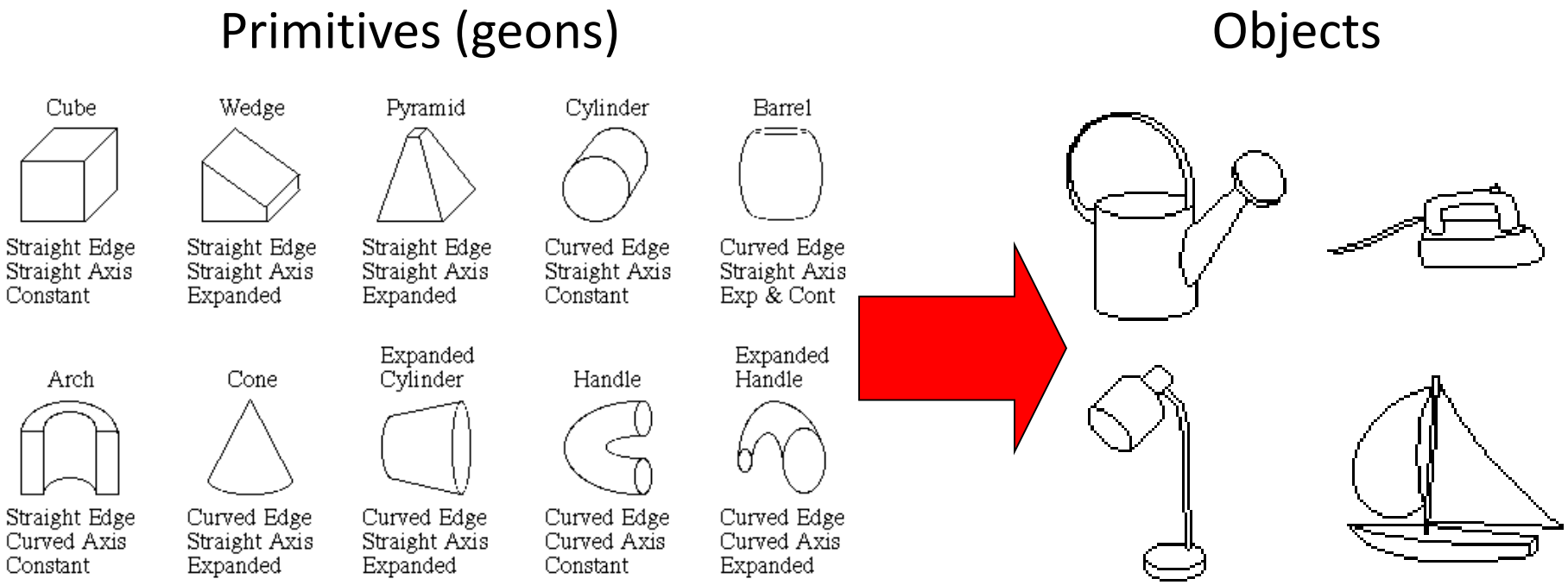
L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Fig. 1. A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b)A blocks world scene. c)Detected edges using a 2x2 gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

J. Mundy, Object Recognition in the Geometric Era: a Retrospective, 2006

Recognition by components

Biederman (1987)

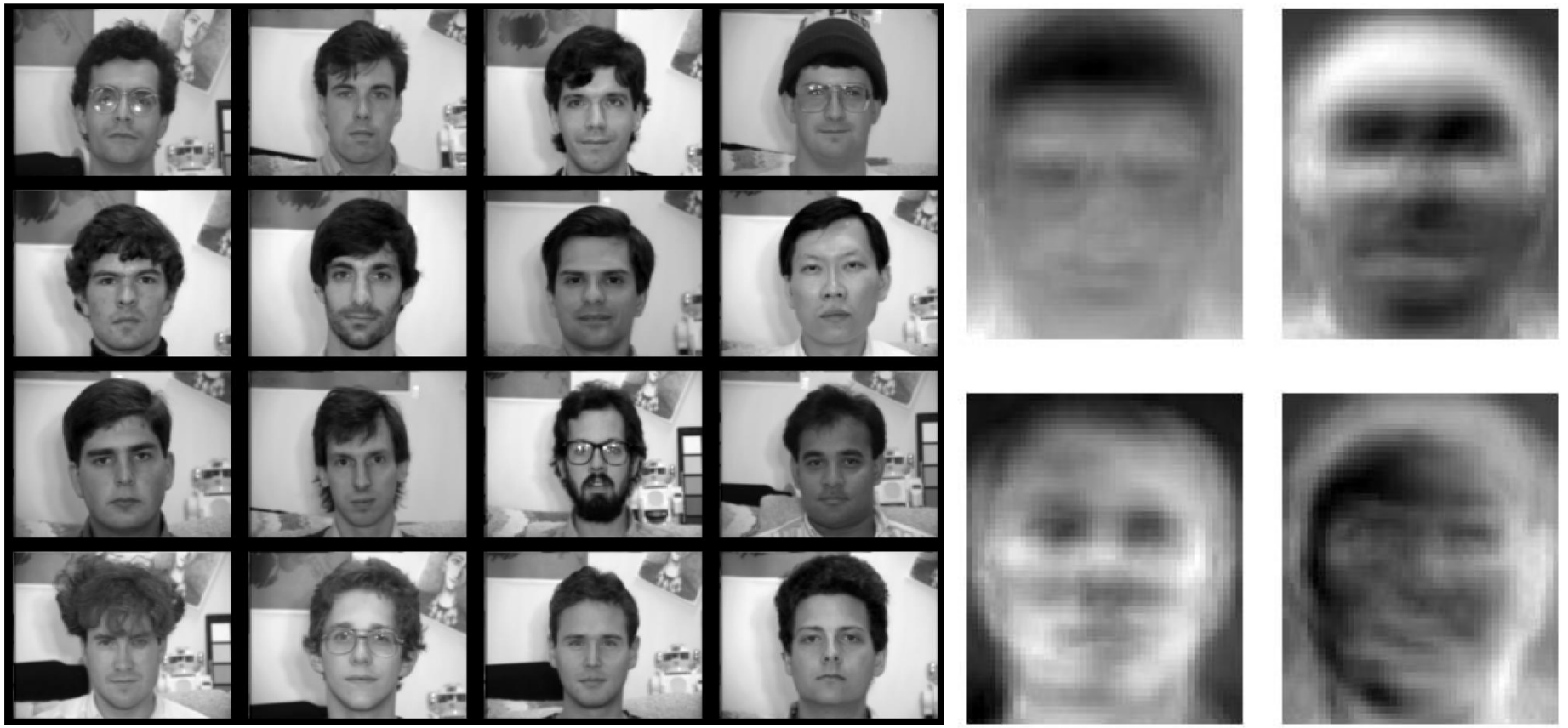


http://en.wikipedia.org/wiki/Recognition_by_Components_Theory

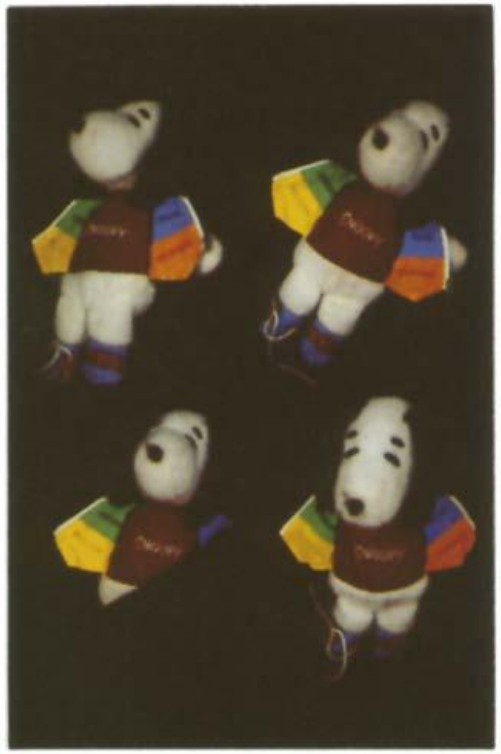
History of ideas in recognition

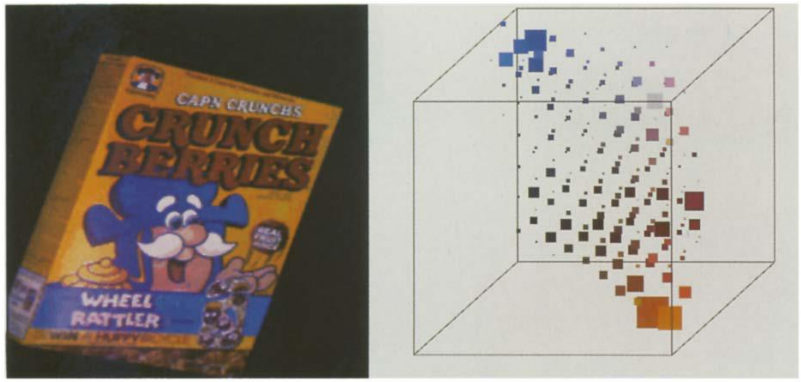
- 1960s – early 1990s: the geometric era
- **1990s: appearance-based models**
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

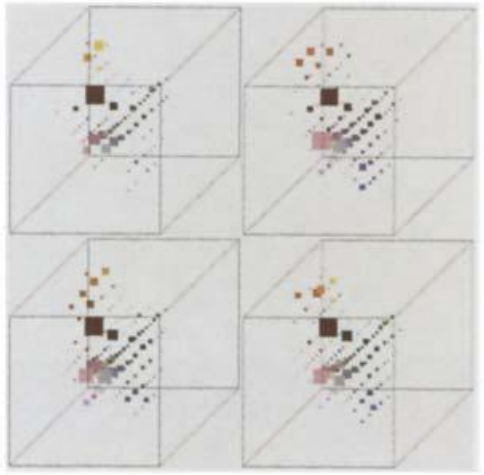
Eigenfaces (Turk & Pentland, 1991)



Color Histograms





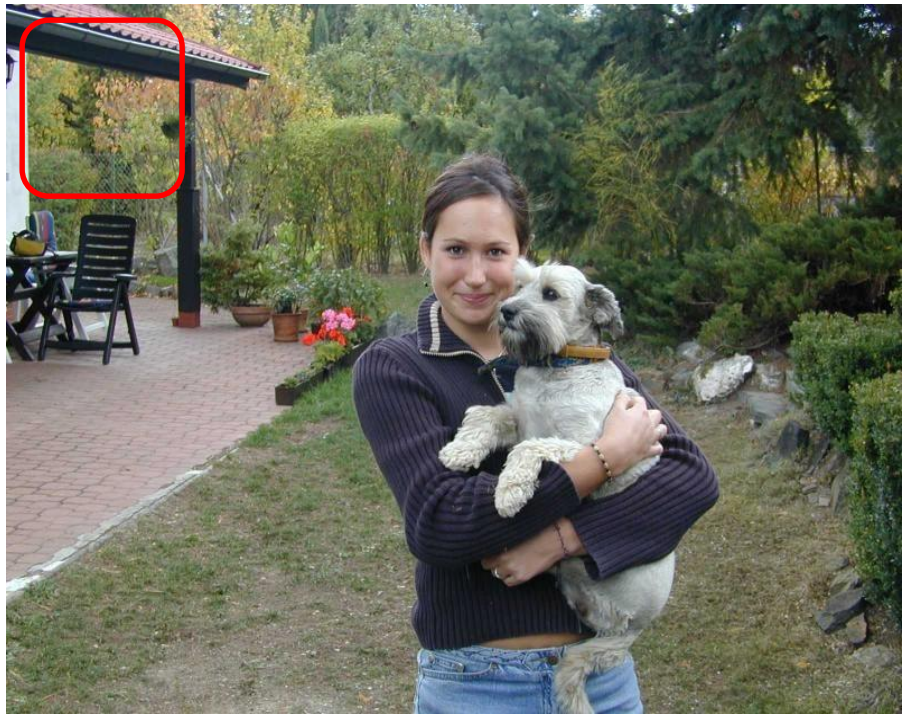


Swain and Ballard, Color Indexing, IJCV 1991.

History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- **Mid-1990s: sliding window approaches**
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

Sliding window approaches



History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- **Late 1990s: local features**
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

Local features for object instance recognition

D. Lowe (1999, 2004)

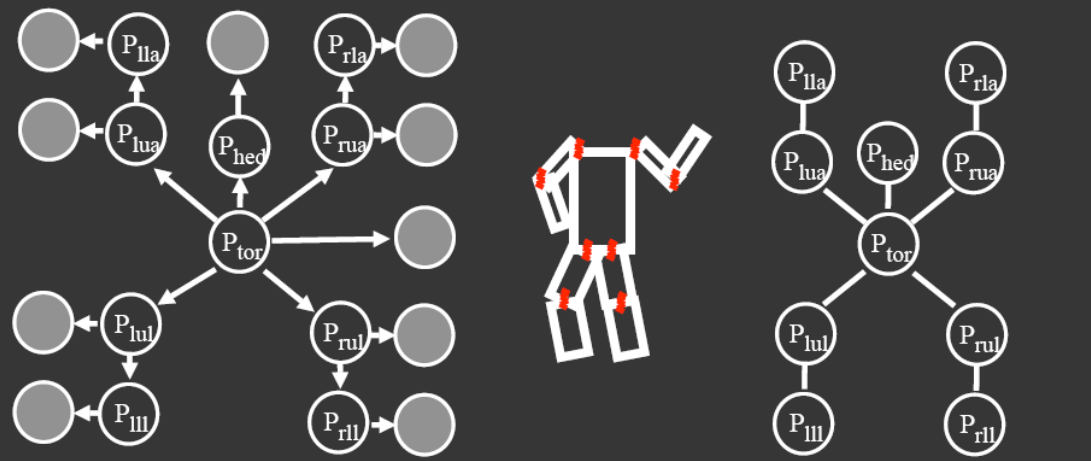
History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- **Early 2000s: parts-and-shape models**
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

Representing people

Pictorial structure model

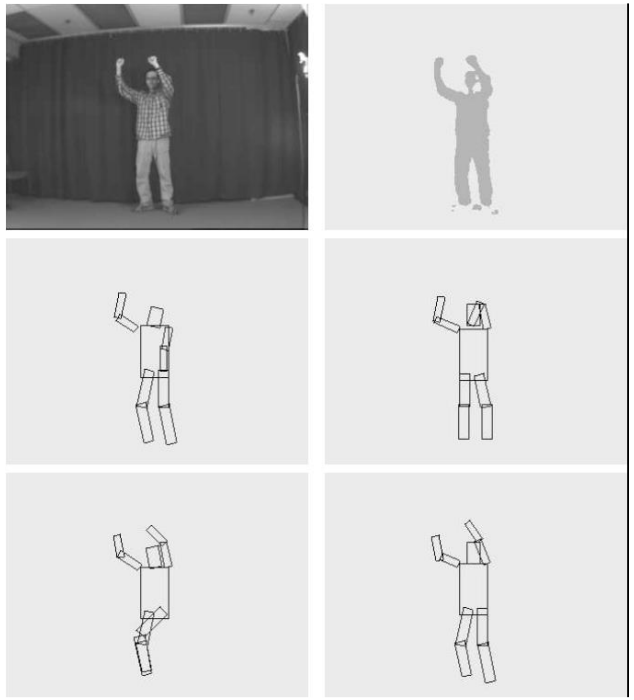
Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)



$$\Pr(P_{tor}, P_{arm}, \ldots | \mathrm{Im}) \propto \prod_{i,j} \Pr(P_i \mid P_j) \prod_i \Pr(\mathrm{Im}(P_i))$$

part geometry

part appearance



History of ideas in recognition

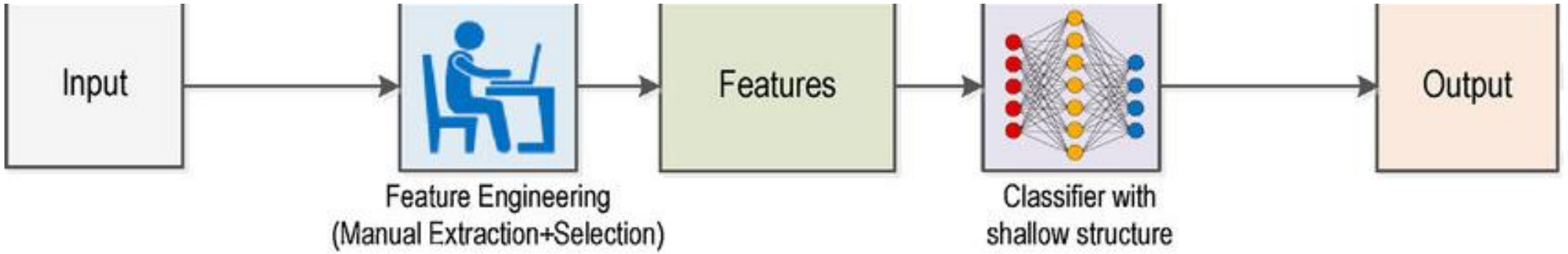
- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- **Mid-2000s: bags of features**
- Present trends: combination of local and global methods, data-driven methods, context

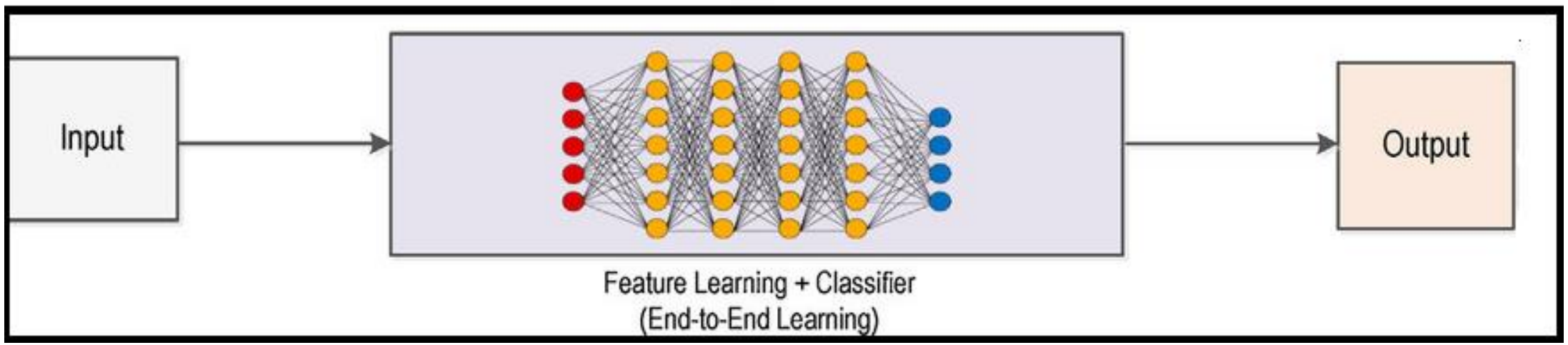
Bag-of-words models

Object → **Bag of ‘words’**

Deep Learning and Vision

- Deep Learning has been a great disruption into the field of Computer Vision. Has made a lot of new things work!
- Many deep learning methods being applied to vision these days.
- This is a deep learning course for visual inference. We will briefly review some pre-deep learning methods, and then mostly deep learning.


Input
Features
Output
Feature Engineering
(Manual Extraction+Selection)
Classifier with
shallow structure


Input
Output
Feature Learning + Classifier
(End-to-End Learning)

Objectives

- Understanding foundational concepts for representation learning using neural networks
- Becoming familiar with state-of-the-art models for tasks such as image classification, object detection, image segmentation, scene recognition, etc.
- Obtain practical experience in the implementation of visual recognition models using deep learning.

Course syllabus

- Image recognition
  - Bag of words – the state of the art before deep learning
  - Deep learning methods
    - Discussions on different layers
    - Discussions on regularization and loss functions
    - Discussions on optimization techniques
    - Case study on Imagenet – evolution of models trained on Imagenet
    - Fine-grained object classification

Course syllabus

- Object detection
  - SVM + HOG based object detection before deep learning
  - R-CNN family
  - YOLO, SSD
- Image segmentation
  - Semantic segmentation
  - Instance segmentation
- Some details of CNN for videos
- Some advanced CNN models, e.g, Bayesian CNN, Siamese/Triplet CNN

Course syllabus

- Recurrent networks
  - Training RNN – back-propagation through time
  - LSTM
  - GRU
- Attention model in deep learning
- Transformers

Course syllabus

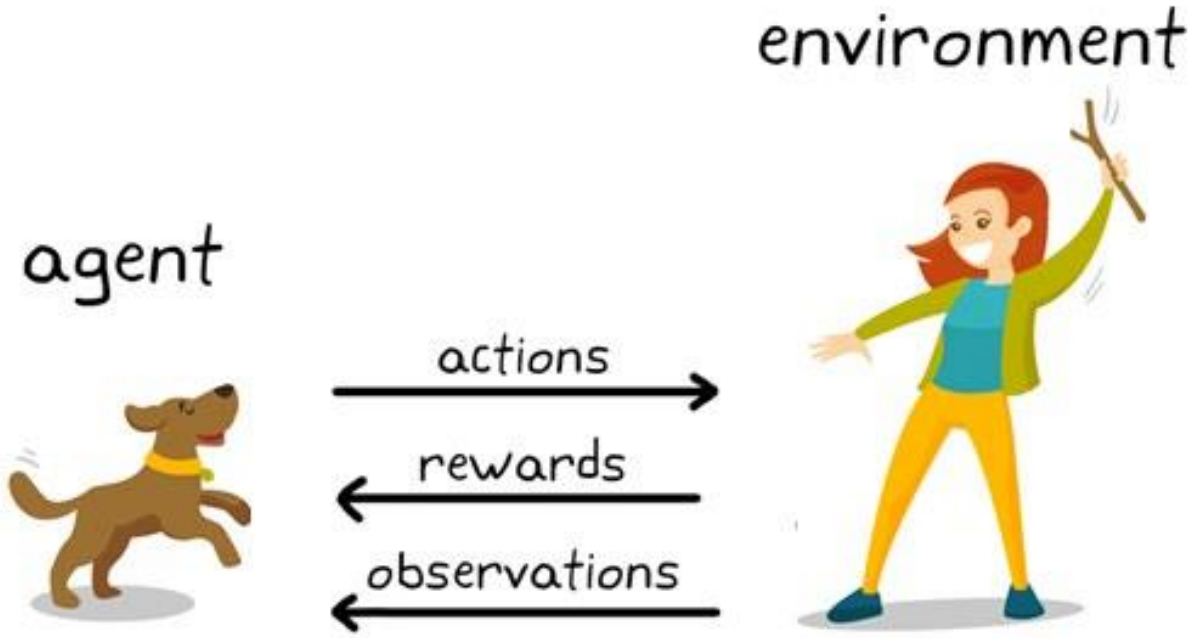
- Deep Representation learning
  - Regularization in encoder-decoder model
    - Sparse
    - Denoising
    - Contrastive
    - Stacked
  - Self-supervised learning

Course syllabus

- Deep generative models
  - Variational auto-encoder
  - GANs
    - Vanilla GAN
    - LS GAN
    - Info GAN
    - W-GAN…
  - Pixel CNN / RNN

Deep Reinforcement Learning (if time permits)

- Deep Q Learning
- Policy Gradient and REINFORCE
- Actor-Critic models



Evaluation

- Home work – 20%
  - 2 paper reviews – recent papers from ICCV/CVPR/NeurIPS/ICML/ICLR/ECCV/BMVC/TPAMI/IJCV/TNNLS (>=2018) – 10 x 2 = 20 % (team of 2 people)
- 1 Kaggle competition of image classification – 10%
- 1 course project – implement a paper, propose a new problem and implement, write a blog about a very recent paper – 25% (team of at least 3 people)
- Quiz (3) – 15 x 3 = 45%

Requirements for audit

- All Quizzes
- Kaggle

E - Office hours & TA

- Thursday 4-5PM (with appointment)

- Names will be given later

Resources

- Online lectures (Stanford, MIT, NYU, IITM)
- Deep learning book by GoodFellow
- http://neuralnetworksanddeeplearning.com/
- Deep Learning Methods and Applications by Deng & Yu
- https://www.purestorage.com/content/dam/purestorage/pdf/whitepapers/oreilly-deep-learning-book.pdf
- Papers

Pytorch, Tensorflow, Keras

- Pytorch – zero to GAN
  - https://www.youtube.com/watch?v=GIsg-ZUy0MY
- Keras by deepLizard
  - https://www.youtube.com/watch?v=qFJeN9V1ZsI
- Tensorflow tutorial
  - https://www.youtube.com/watch?v=tPYj3fFJGjk