Course Logistics

CS 698O: Visual Recognition

Vinay P. Namboodiri
vinaypn@iitk.ac.in

Instructor Details

- Vinay P. Namboodiri
- # RM 406, RM building, CSE
- Office Hours: Tuesday and Wednesday 10 - 11 am
- Preferably email: vinaypn@iitk, Subject: CS6980….

About the Course

- Visual Recognition
- A graduate elective
- Objectives: Obtain different perspectives on understanding visual recognition
  - Problems
  - Approaches
  - Advances

Lectures

- Monday, and Thursday
- Mon 10:30 -11:50, Thu 10:30-11:50,
- Venue: KD 101

Grading

- Weightage (Tentative):
- Mid-Sem - 25%
- End-Sem 25%
- Assignments 20% (Programming, Paper Review)
- Project 30% (atleast 2 stages)

Course Outline

- Introduction
- Exact instance retrieval
- Classification
- Detection
- Segmentation
- Weak Supervision
- Active Learning
- Domain Adaptation
- Unsupervised Representation learning
- Vision and Language

Course Outline

- Introduction
- Exact instance retrieval
- Classification
- Detection
- Segmentation
- Weak Supervision
- Active Learning
- Domain Adaptation
- Unsupervised Representation learning
- Vision and Language

Course Outline

- Introduction
- Exact instance retrieval
- Classification
- Detection
- Segmentation
- Weak Supervision
- Active Learning
- Domain Adaptation
- Unsupervised Representation learning
- Vision and Language

Traditional Feature Based

Course Outline

- Introduction
- Exact instance retrieval
- Classification
- Detection
- Segmentation
- Weak Supervision
- Active Learning
- Domain Adaptation
- Unsupervised Representation learning
- Vision and Language

Deep learning based

Course Outline

- Introduction
- Exact instance retrieval
- Classification
- Detection
- Segmentation

Deep learning based

- Weak Supervision
- Active Learning
- Domain Adaptation
- Unsupervised Representation learning
- Vision and Language

Tentative set of advanced topics

Course Material

- Lecture slides that will be posted online
- Course will be based mainly on research papers

- Reference books:
- Deep Learning by Ian Goodfellow, Yoshua Bengio and Aaron Courville Available online
- Computer Vision: Algorithms and Applications by Richard Szeliski Available online
- Computer Vision: Models, Learning, and Inference by Simon J.D. Prince Available online
- Computer Vision: A Modern Approach by Forsyth and Ponce Indian edition available

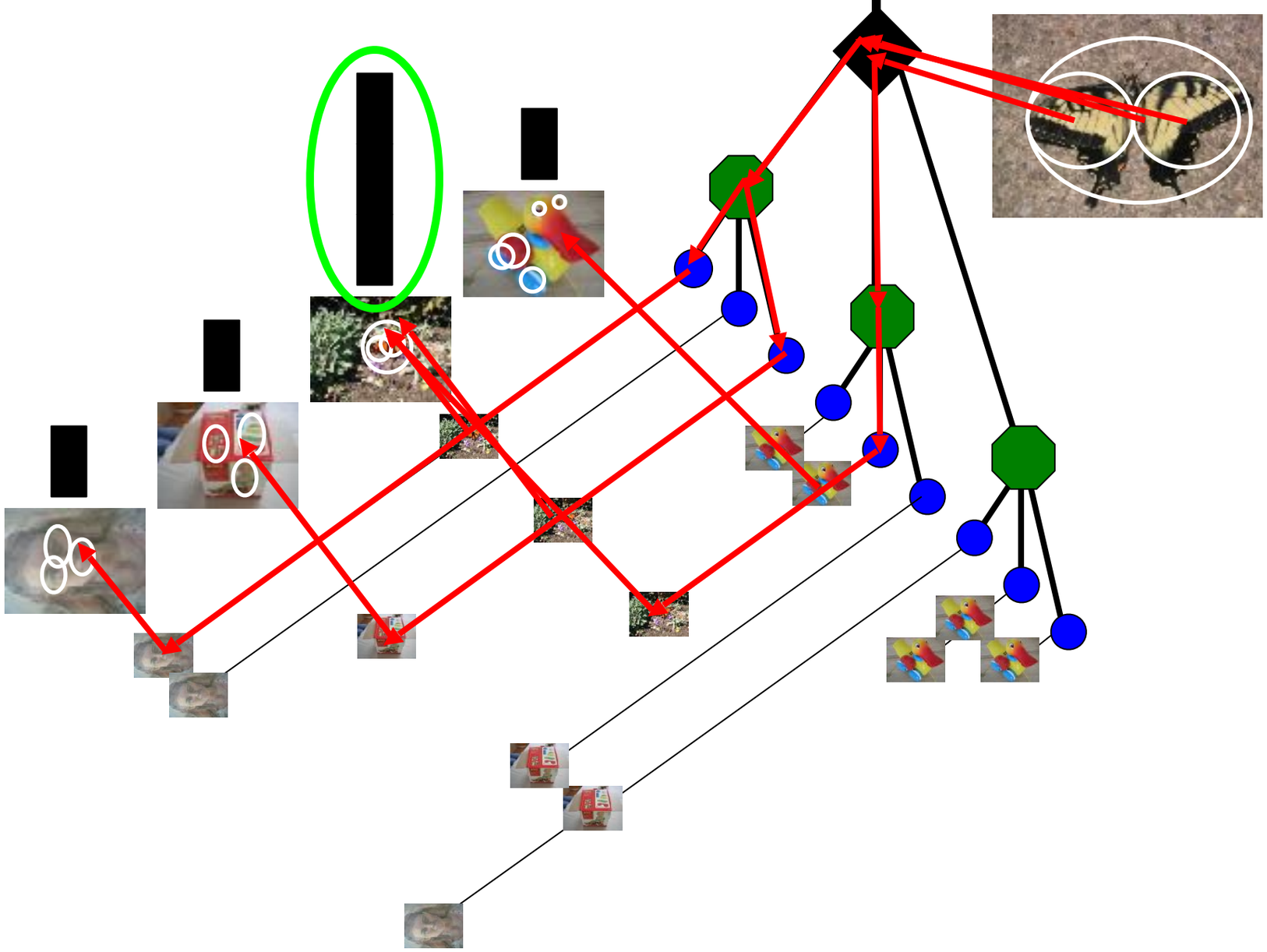
Introduction

CS 698O: Visual Recognition

Vinay P. Namboodiri
vinaypn@iitk.ac.in

What is Visual Recognition?

Semantic Representation



Instance Recognition

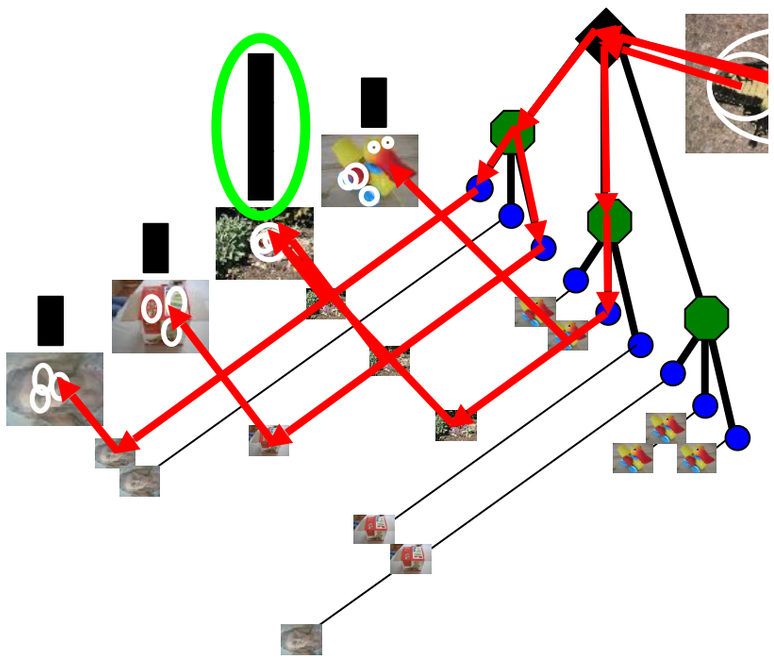
Scalable recognition with a Vocabulary tree
Nister, Strewenius, CVPR 2006

Semantic Representation

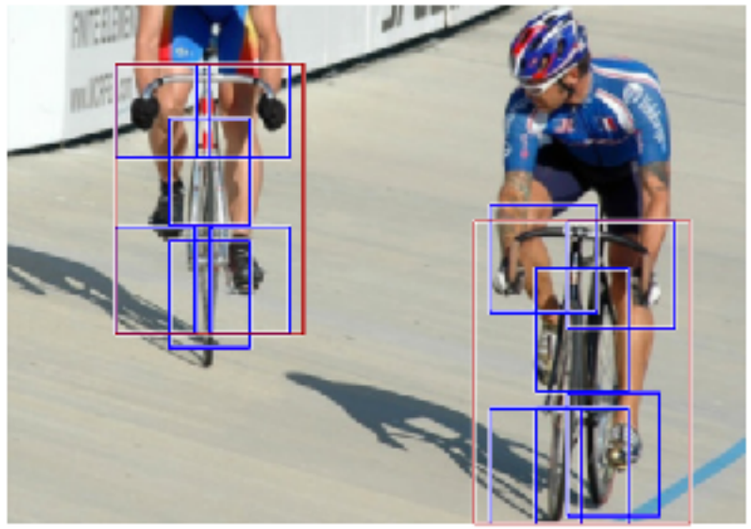
Object Classification
ImageNet

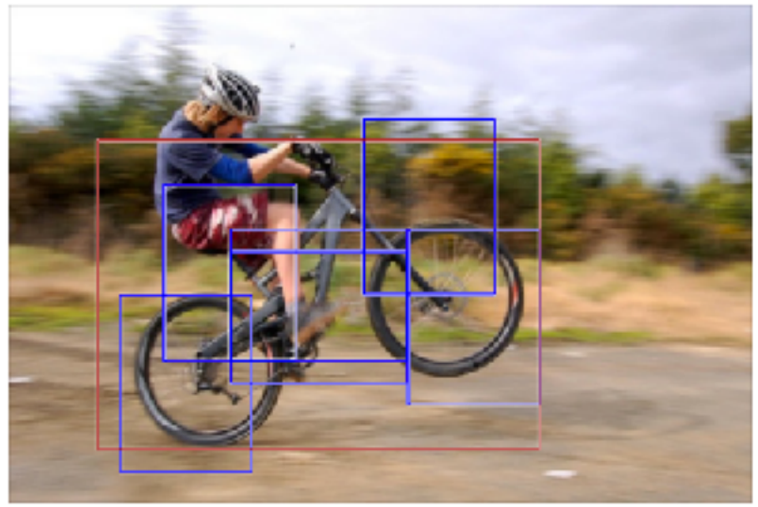
Image credit: Karpathy

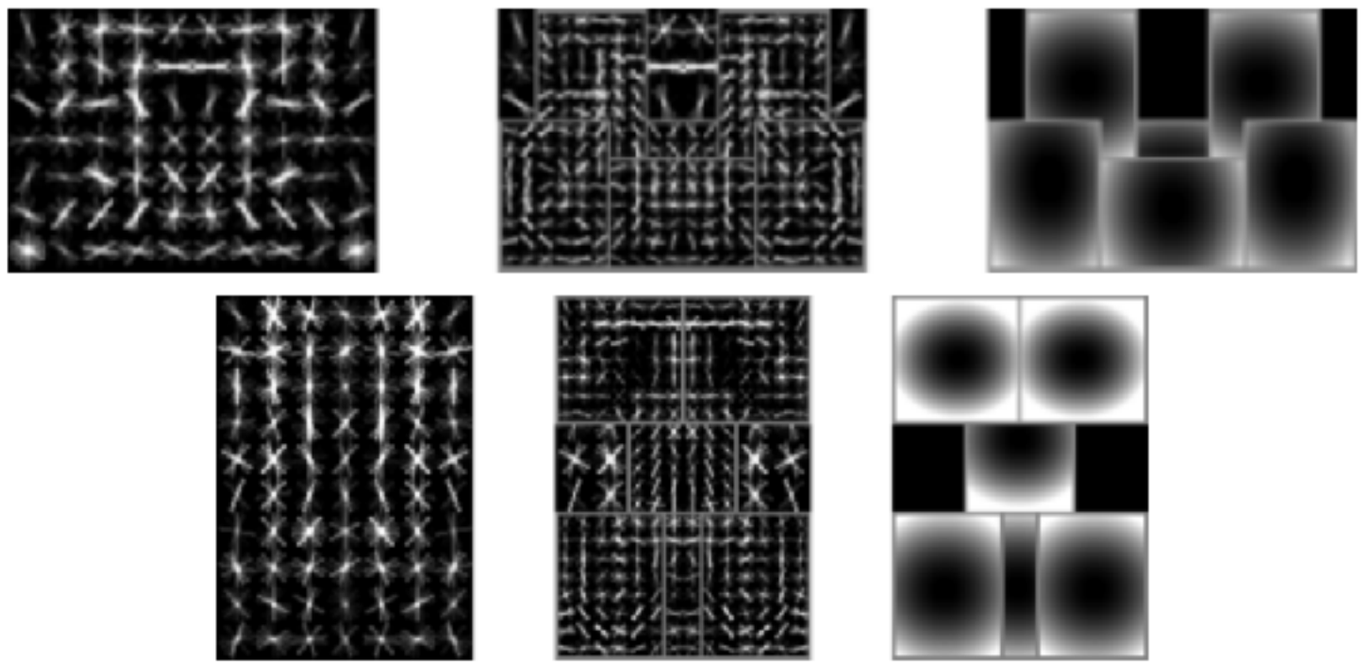
Semantic Representation



Instance



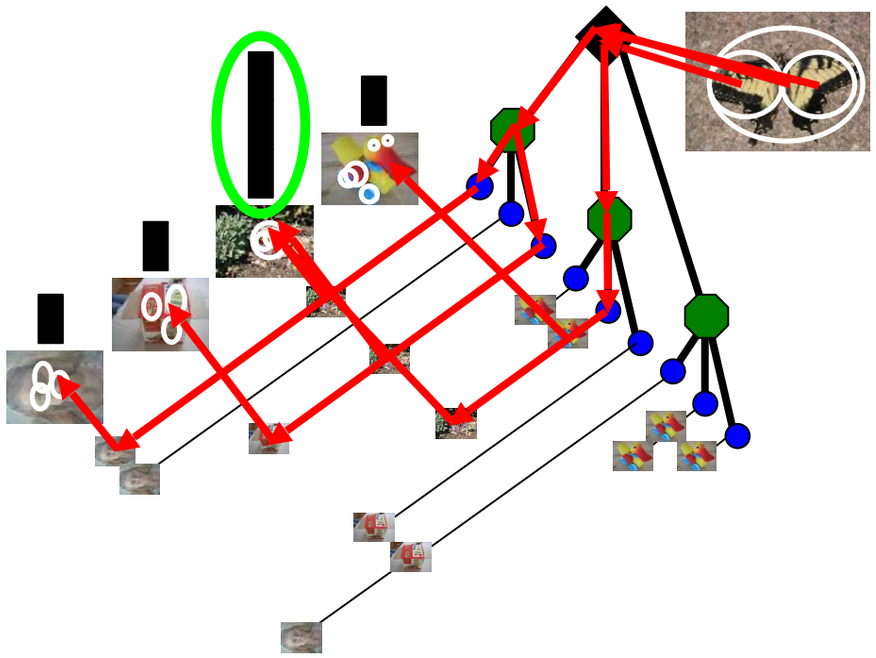




Object Detection

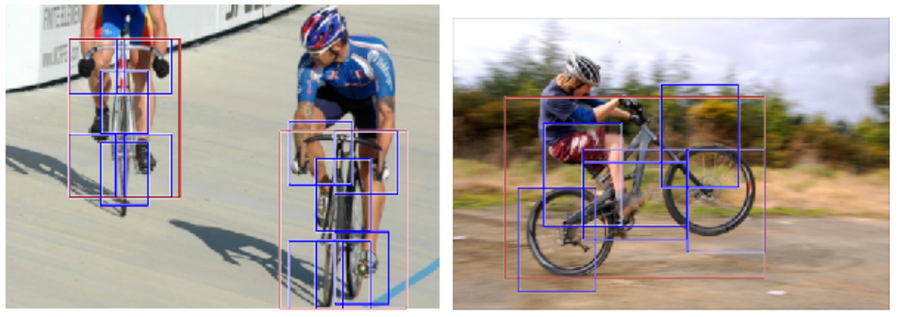
Object Detection with Discriminatively Trained Part Based Models
P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan
PAMI 2010

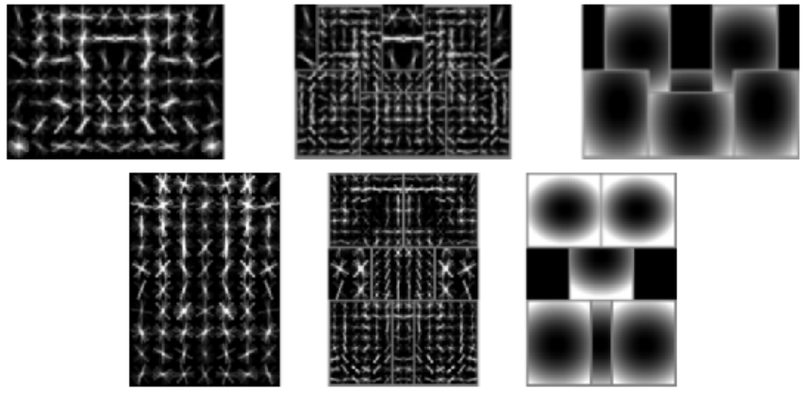
Semantic Representation



Instance

Object





Object Detection

Visual Recognition

Why understand Visual Recognition?

Motivation

- Intellectual curiosity
- Algorithms for general visual perception (also enable general machine learning methods)
- Applications

Intellectual Challenge

Intellectual Challenge

Figure credit: Hakan Bilen

Intellectual Challenge

- Making machines *see*

- Extracting semantic information from signals

Table 1

3	120	23	33
6	34	45	56
1	59	67	90
90	99	23	84
200	121	89	55

Figure credit: Hakan Bilen

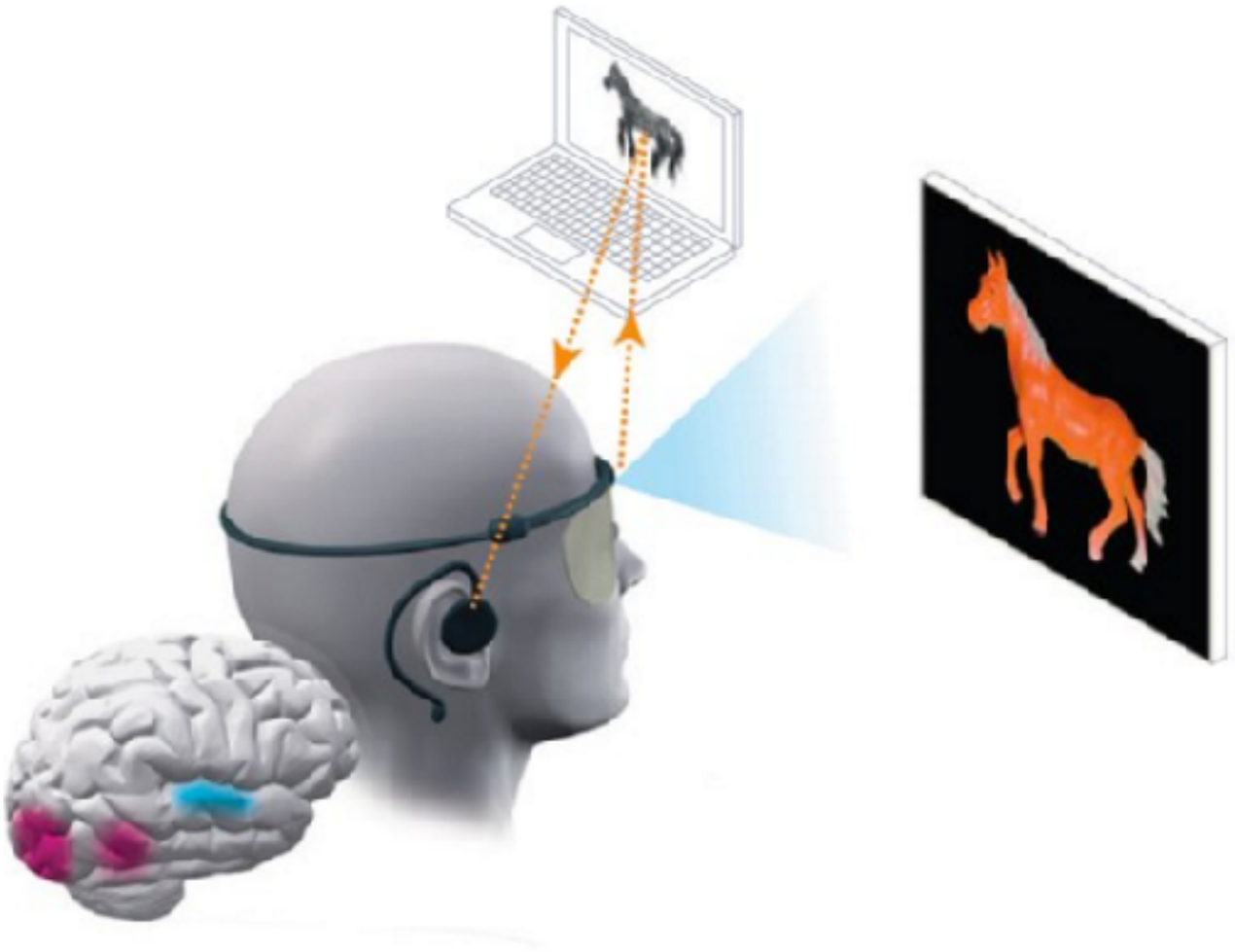
Algorithms

- Segmentation (Graph partitioning, Non-parametric density estimation)
- Denoising (Sparsity)
- Template Matching
- Deep Neural networks

Applications

Self driving cars

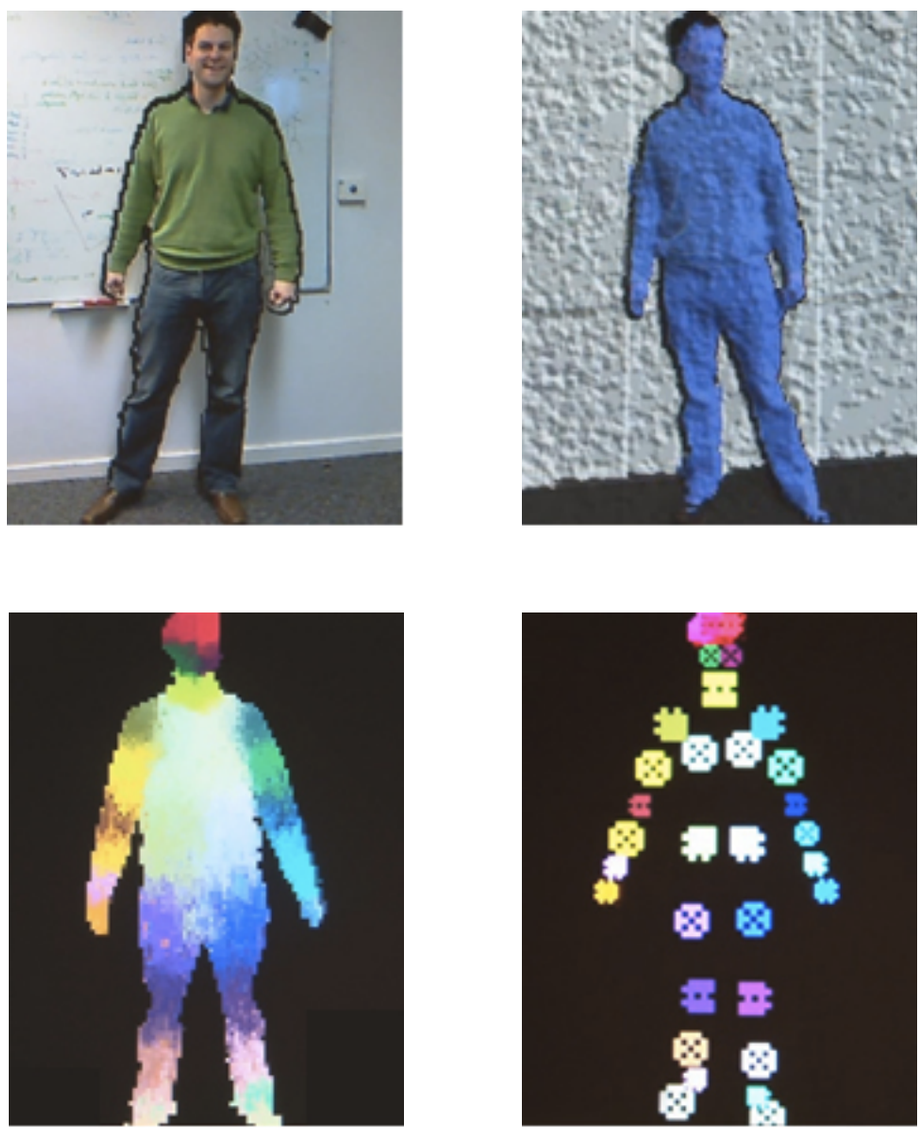
Assistive Vision



Applications

Surveillance

Applications



Human-Computer Interfaces

Challenges

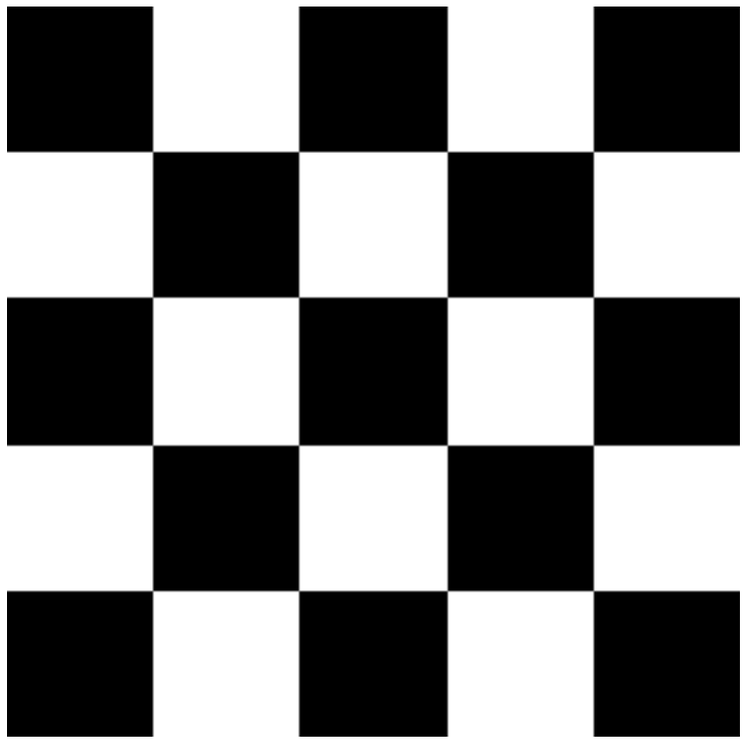
Challenges

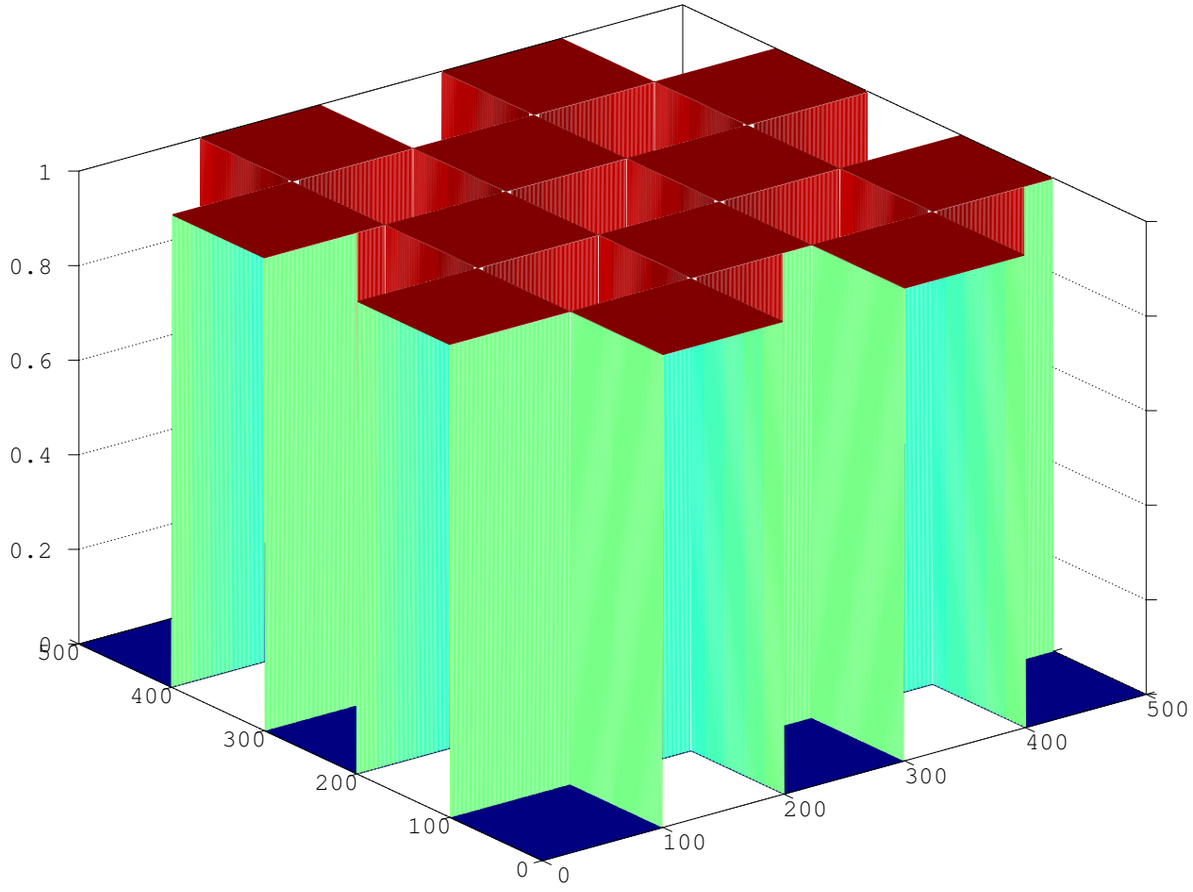
Challenges

Challenges

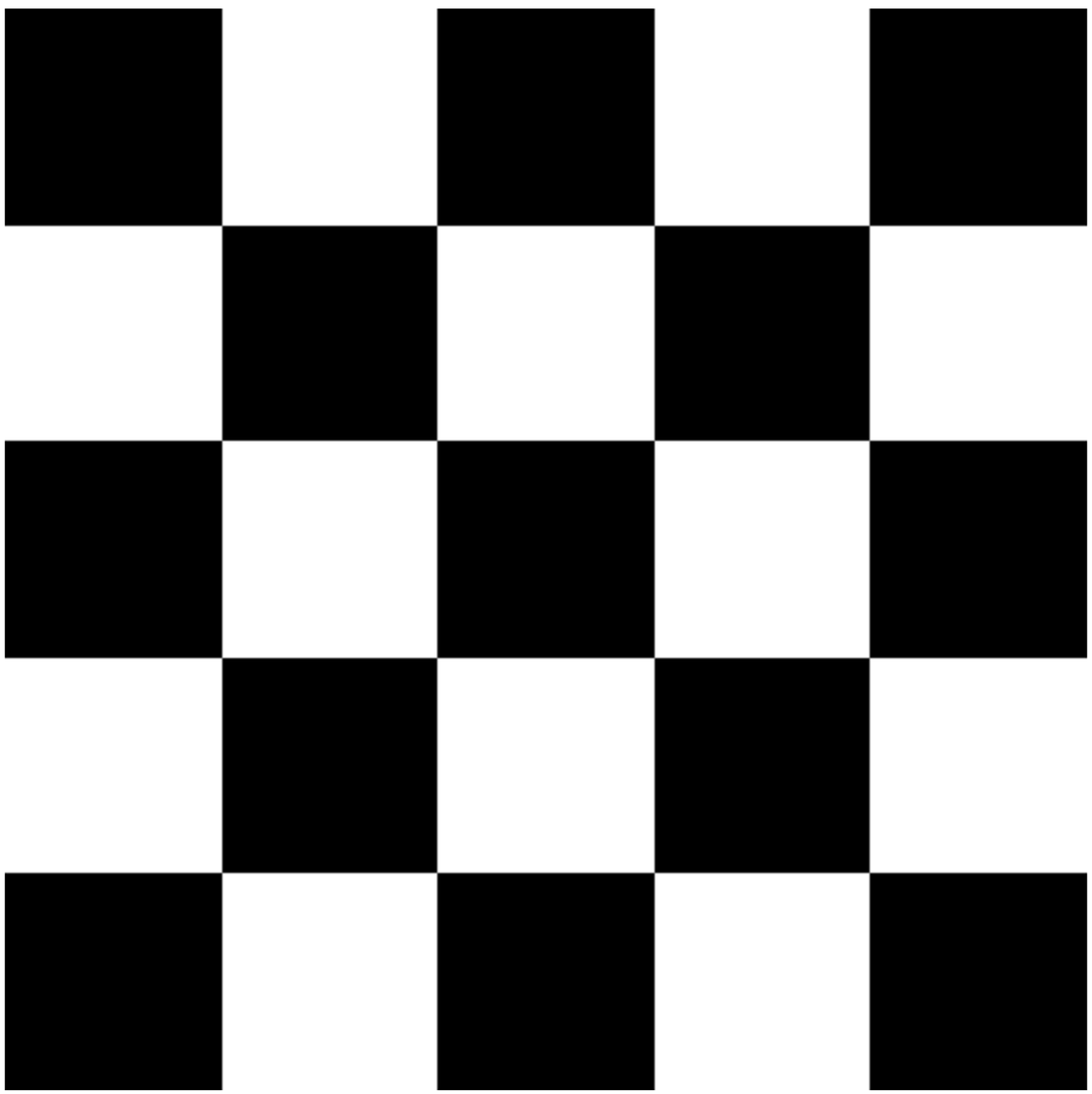
An image

- A rectangular grid of picture elements (pixels). Each pixel can be a scalar or a vector(3 or 4 element)



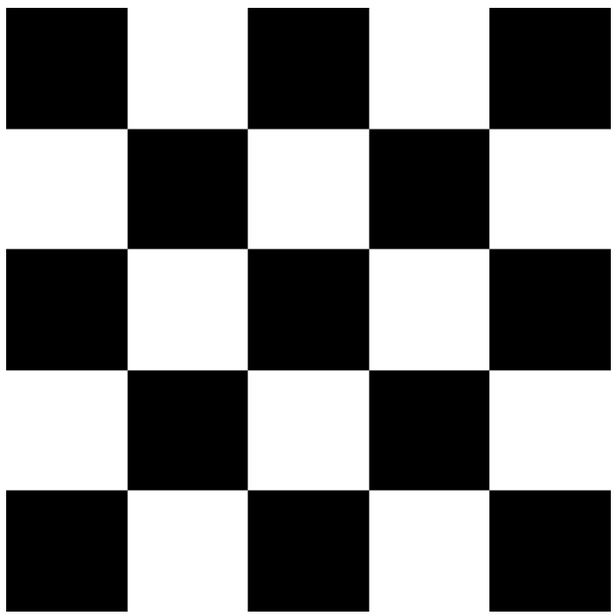


Interpreting an image



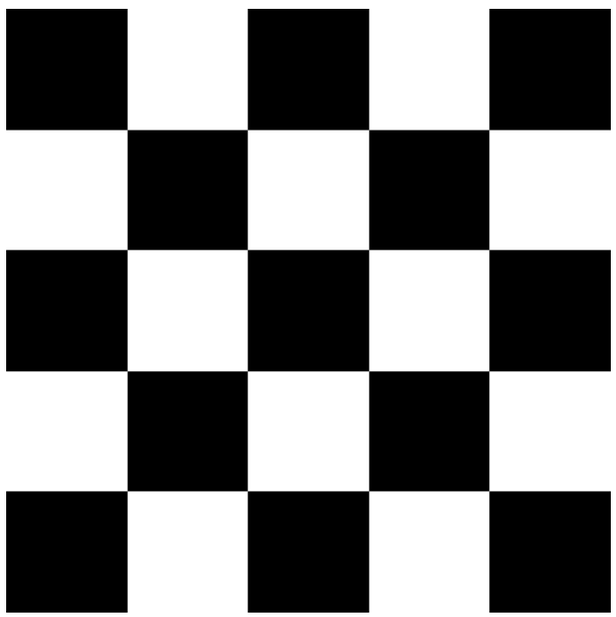
Not a very tough task to interpret this image

Interpreting an image



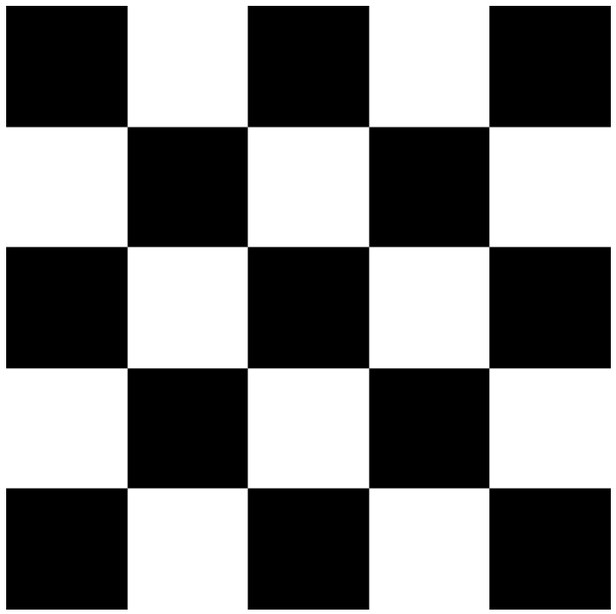
However, same checkerboard could be oriented differently, and say extracting the checkerboard pattern would not be trivial with geometric variation

Interp



Checkerboard on cookies!

Interpreting an image

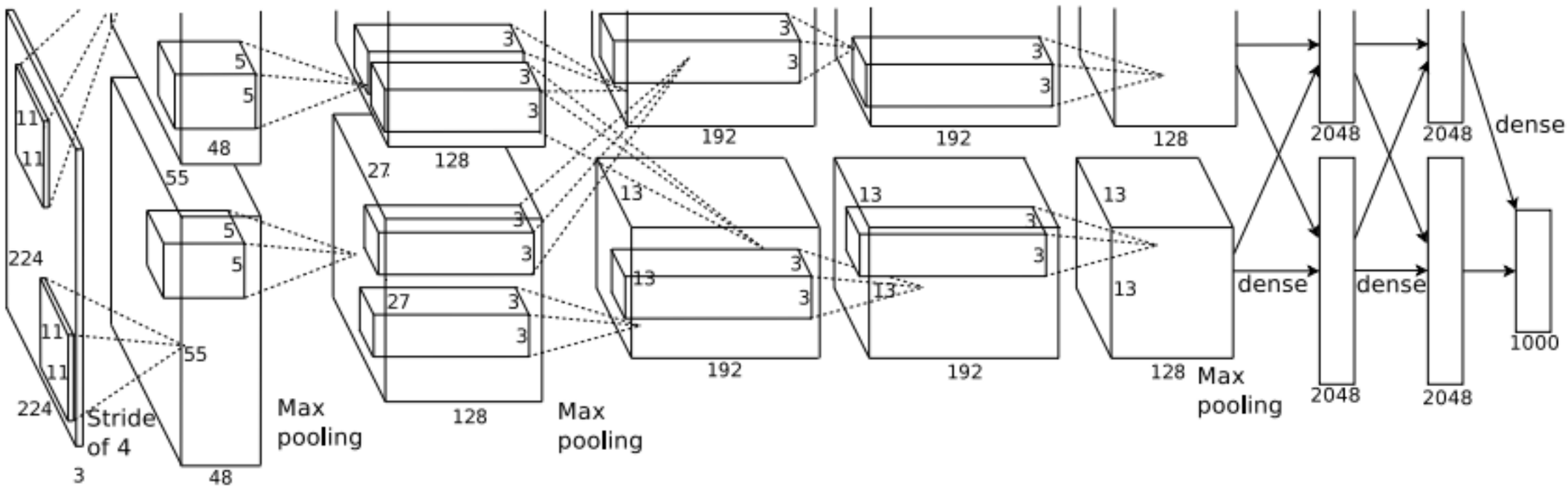


Checkerboard pattern variants

Image interpretation

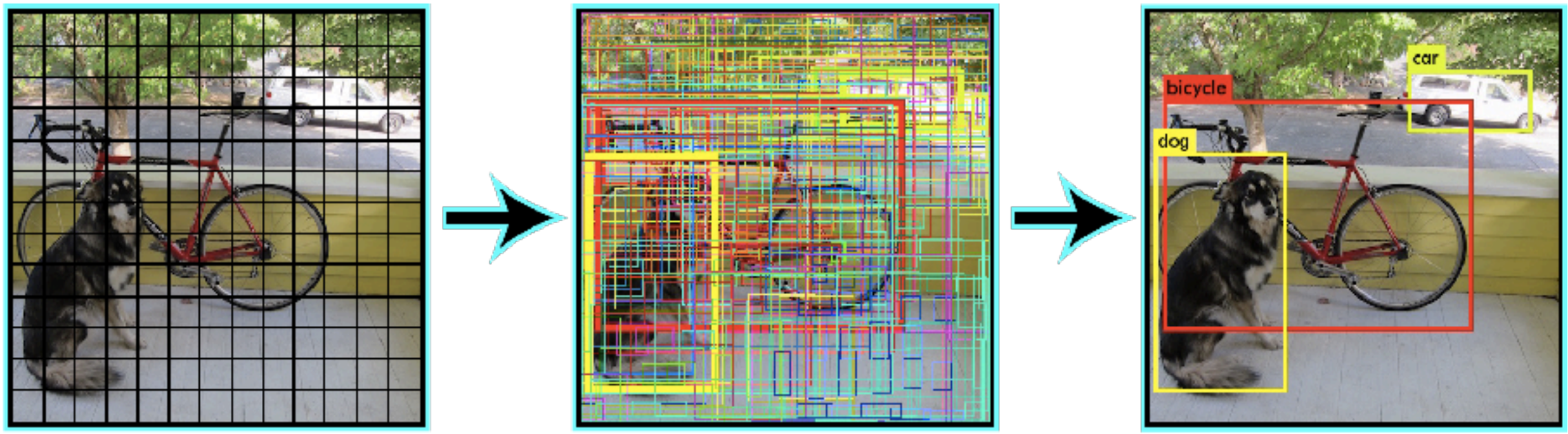
The Adventure of the dancing men
- Sherlock Holmes by Arthur Conan Doyle

Recent Successes



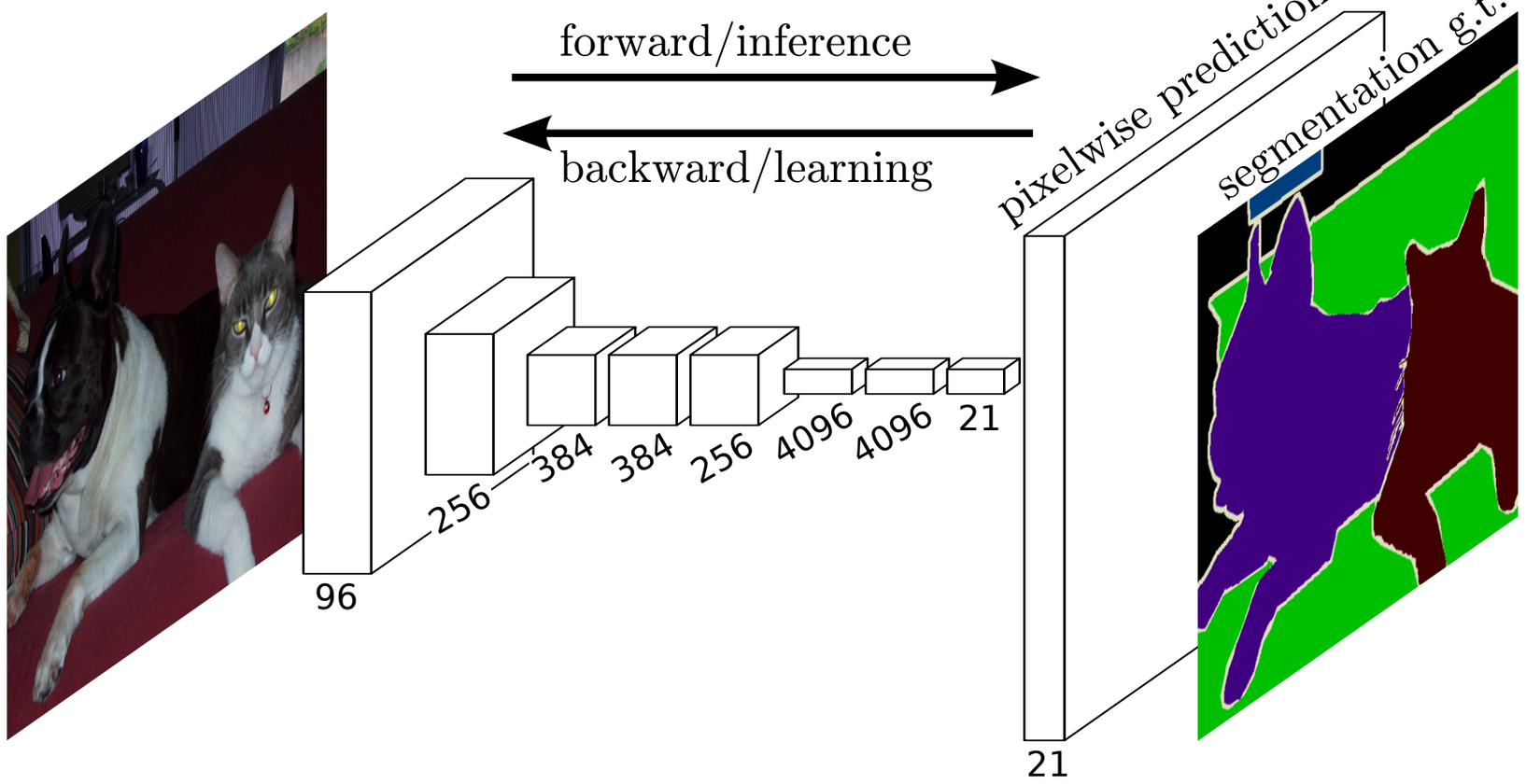
ImageNet Classification with Deep Convolutional Neural Networks
Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton
NIPS 2012

Recent Successes



You Only Look Once: Unified, Real-Time Object Detection
Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi
CVPR 2016

Recent Successes



Fully Convolutional Networks for Semantic Segmentation

Jon Long*, Evan Shelhamer*, Trevor Darrell

CVPR 2015

Conclusion

- Study of visual recognition is one of the classical and interesting problems that is fascinating
- Solving this enables many applications
- This could enable us to move towards real developments in *AI*

Instance Recognition

CS 698O: Visual Recognition

Course Outline

- Introduction
- **Exact instance retrieval**
- Classification
- Detection
- Segmentation
- Weak Supervision
- Active Learning
- Domain Adaptation
- Unsupervised Representation learning
- Vision and Language

Traditional learning based

Tentative set of advanced topics

Problem

- Given a bounding box in an image, extract similar regions in the image for a database of images

- Analogy: Given a term or set of terms, look up and retrieve pages that are most relevant for the term

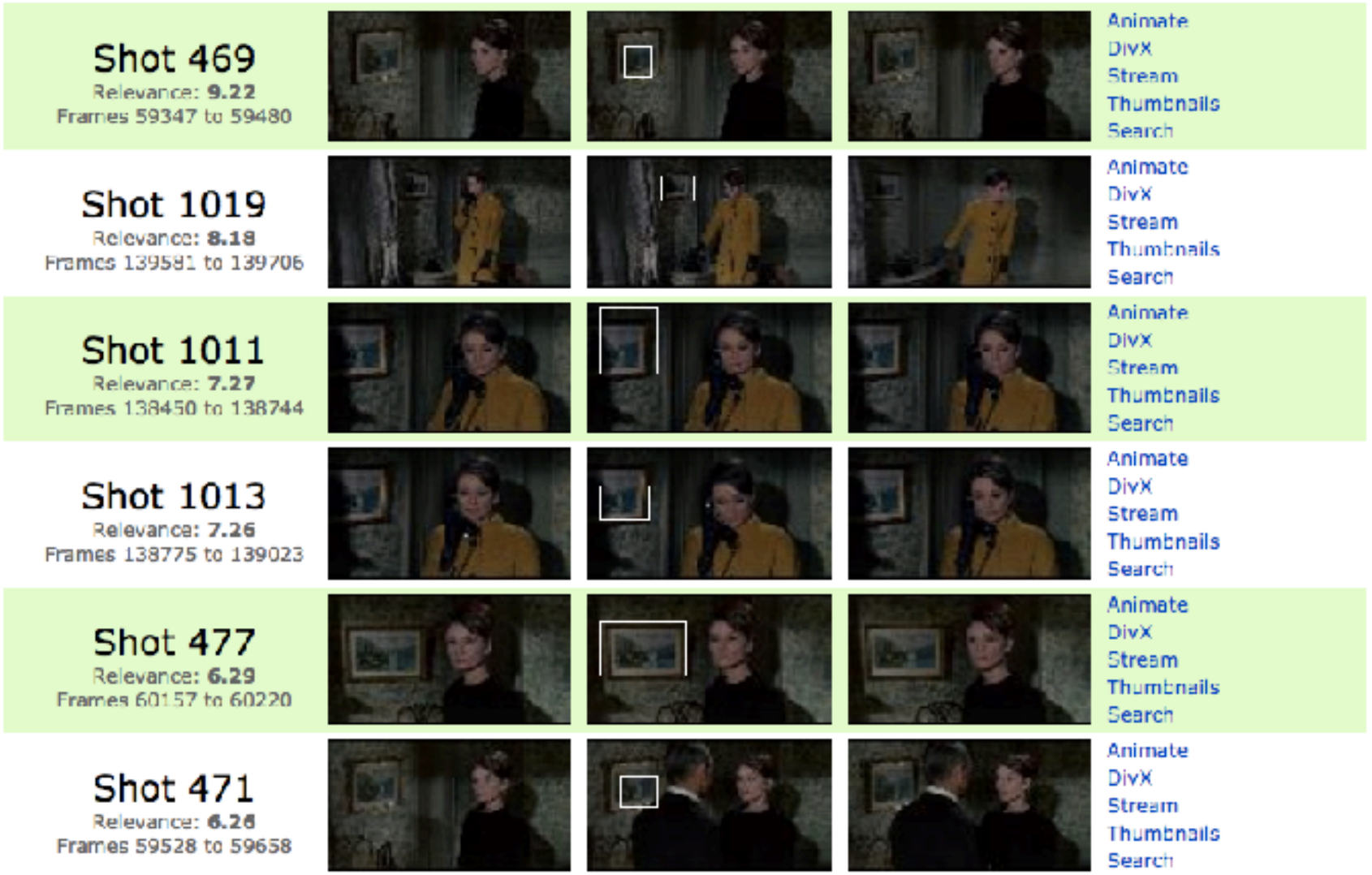
- Assumption: The bounding box can be found without much variation in the database

Example

Video Google: A Text Retrieval Approach to Object Matching in Videos
Josef Sivic and Andrew Zisserman
ICCV 2003

link to demo: http://www.robots.ox.ac.uk/~vgg/research/vgoogle/index.html

Example



Video Google: A Text Retrieval Approach to Object Matching in Videos
Josef Sivic and Andrew Zisserman
ICCV 2003

Approach

- Text retrieval systems
- Documents are parsed into words
- Words are stemmed
- Stored in an inverted file index
- Documents are matched using TF-IDF score

Example

- The **advances** in **image recognition** extend far beyond cool social apps. **Medical startups** claim they'll soon be able to use computers to read X-rays, MRIs, and CT scans more **rapidly** and **accurately** than **radiologists**, to **diagnose cancer** earlier and less invasively, and to accelerate the search for life-saving **pharmaceuticals**. Better image recognition is crucial to unleashing **improvements** in **robotics**, **autonomous drones**, and, of course, **self-driving cars**—a development so momentous that we made it a cover story in June

An excerpt from Fortune magazine on Deep learning http://fortune.com/ai-artificial-intelligence-deep-machine-learning/

Example

- The **advance** in **image recognition** extend far beyond cool social apps. **Medical startup** claim they'll soon be able to use computers to read X-rays, MRIs, and CT scans more **rapid** and **accurate** than **radiologists**, to **diagnose cancer** earlier and less invasively, and to accelerate the search for life-saving **pharmaceutical**. Better image recognition is crucial to unleashing **improve** in **robotics**, **autonomy drone**, and, of course, **self-driving car**—a development so momentous that we made it a cover story in June

An excerpt from Fortune magazine on Deep learning http://fortune.com/ai-artificial-intelligence-deep-machine-learning/

Example

advance image recognition Medical startup

rapid accurate radiologists diagnose cancer

pharmaceutical

improve robotics autonomy drone self-driving car

An excerpt from Fortune magazine on Deep learning http://fortune.com/ai-artificial-intelligence-deep-machine-learning/

Words - Visual Words?