# Object Recognition

Seminar

Rita Osadchy

# So what does object recognition involve?

# Verification: is that a bus?

# Detection: locate the cars in the image

# Verification: is that a picture of Mao?

# Object categorization

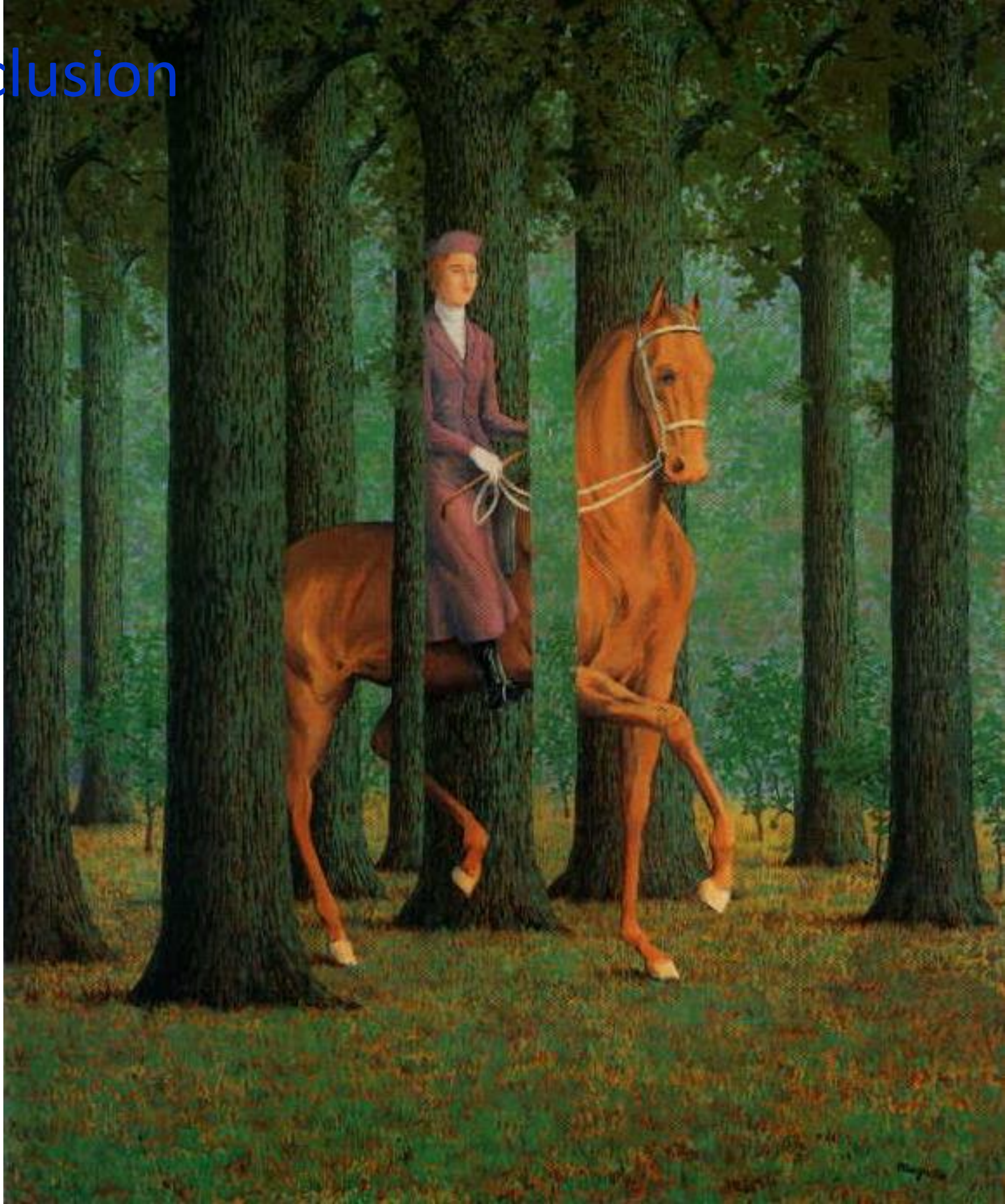
# Challenges 1: view point variation

Michelangelo 1475-1564

# Challenges 2: illumination

slide credit: S. Ullman

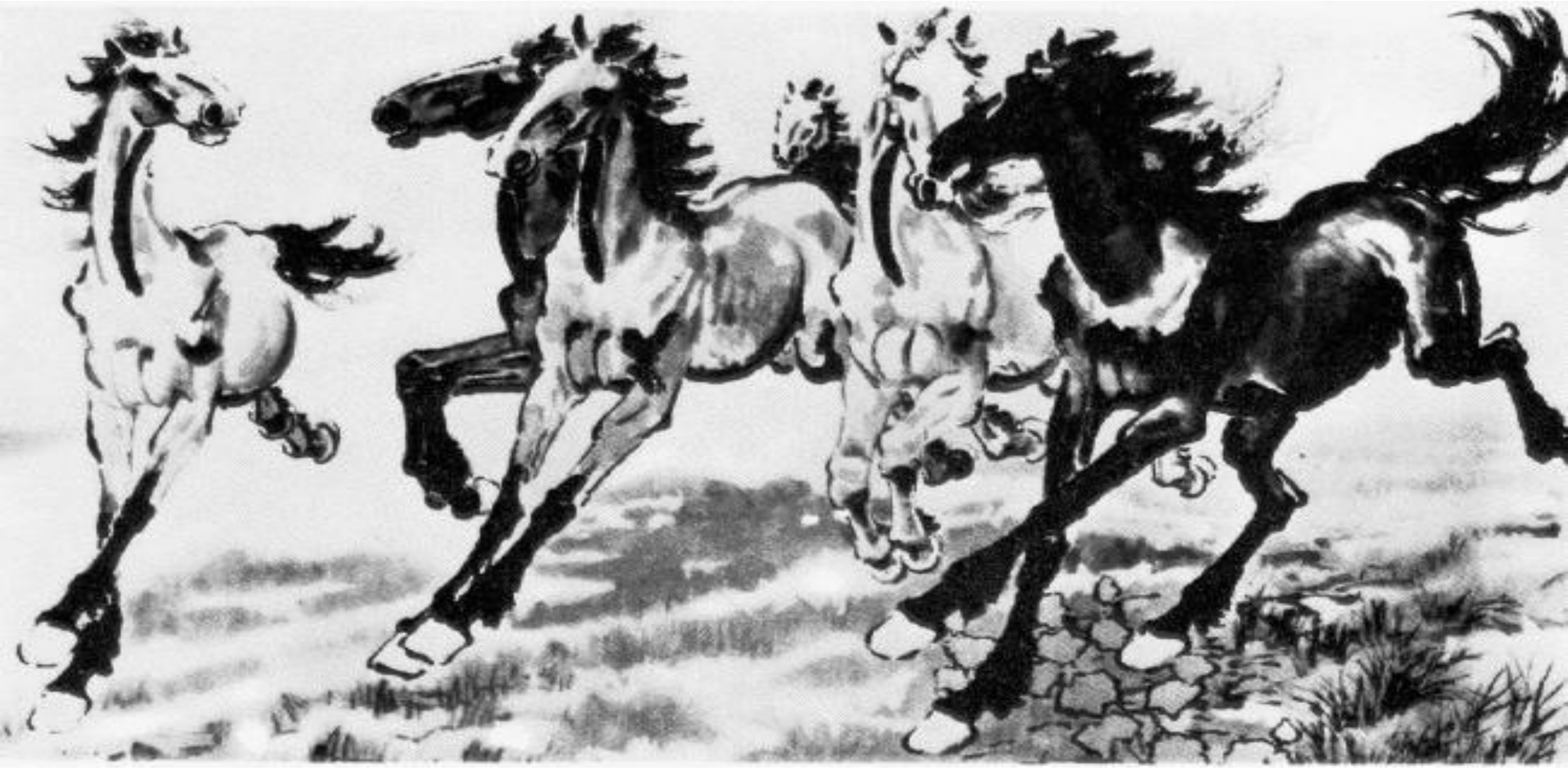
# Challenges 3: occlusion



Magritte, 1957

# Challenges 4: scale

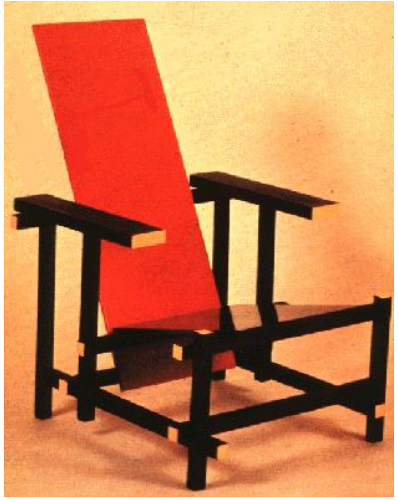
# Challenges 5: deformation



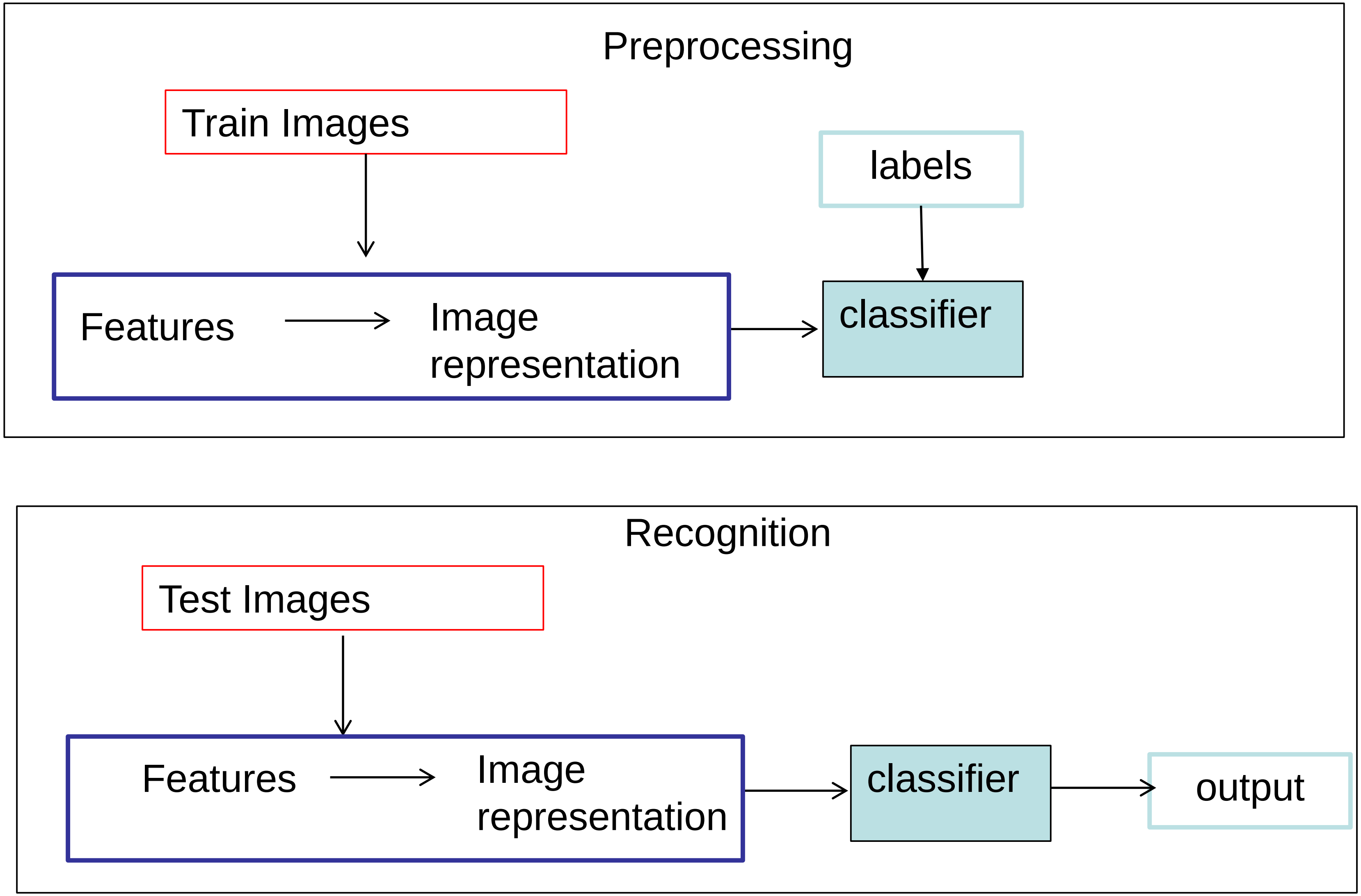
Xu, Beihong 1943

# Challenges 7: intra-class variation

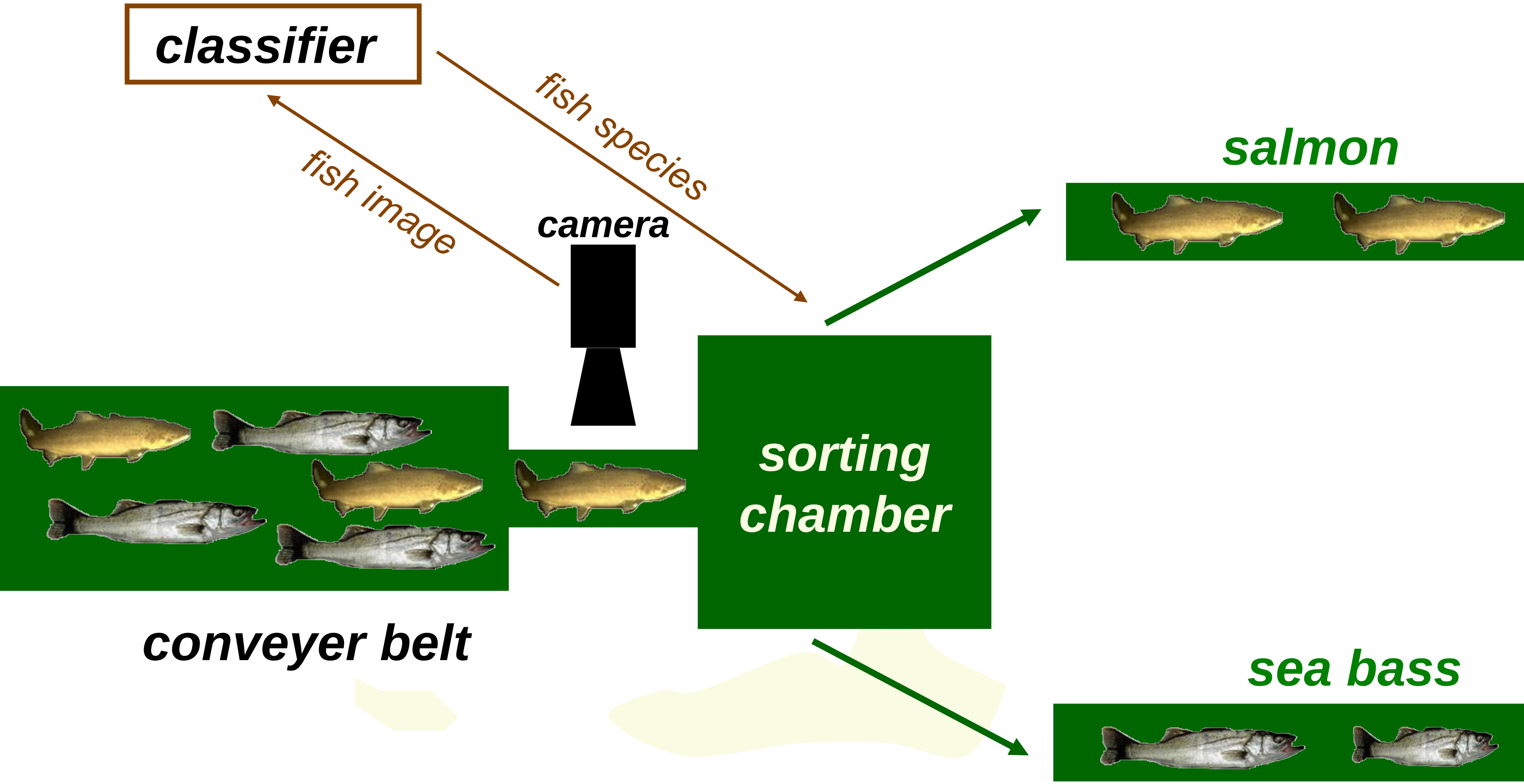


# Recognition Steps



# Object Recognition System



# How to design a PR system?

- Collect data and classify by hand

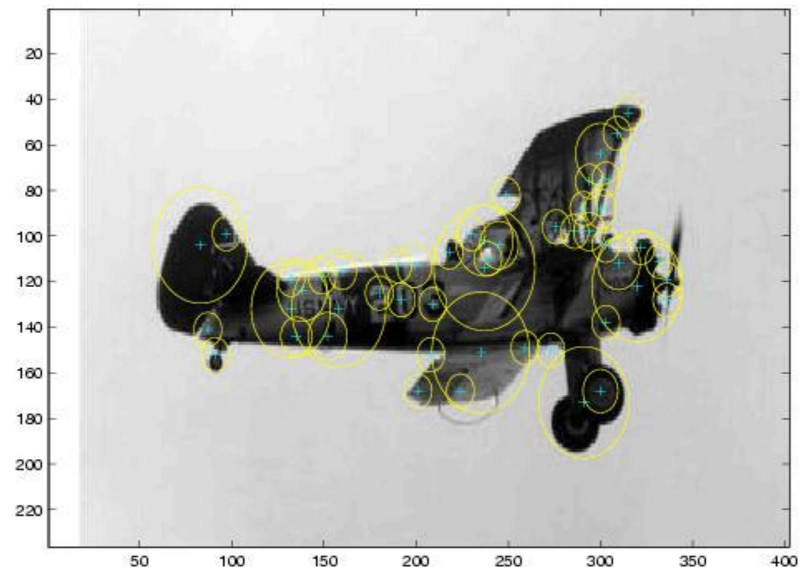
*salmon* *sea bass* *salmon* *salmon* *sea bass* *sea bass*

- Preprocess by segmenting fish from background

- Extract possibly discriminating features
  - length, lightness,width,number of fins,etc.
- Classifier design
  - Choose model
  - Train classifier on part of collected data (training data)
- Test classifier on the rest of collected data (test data) i.e. the data not used for training
  - Should classify new data (new fish images) well

# Interest Point Detectors

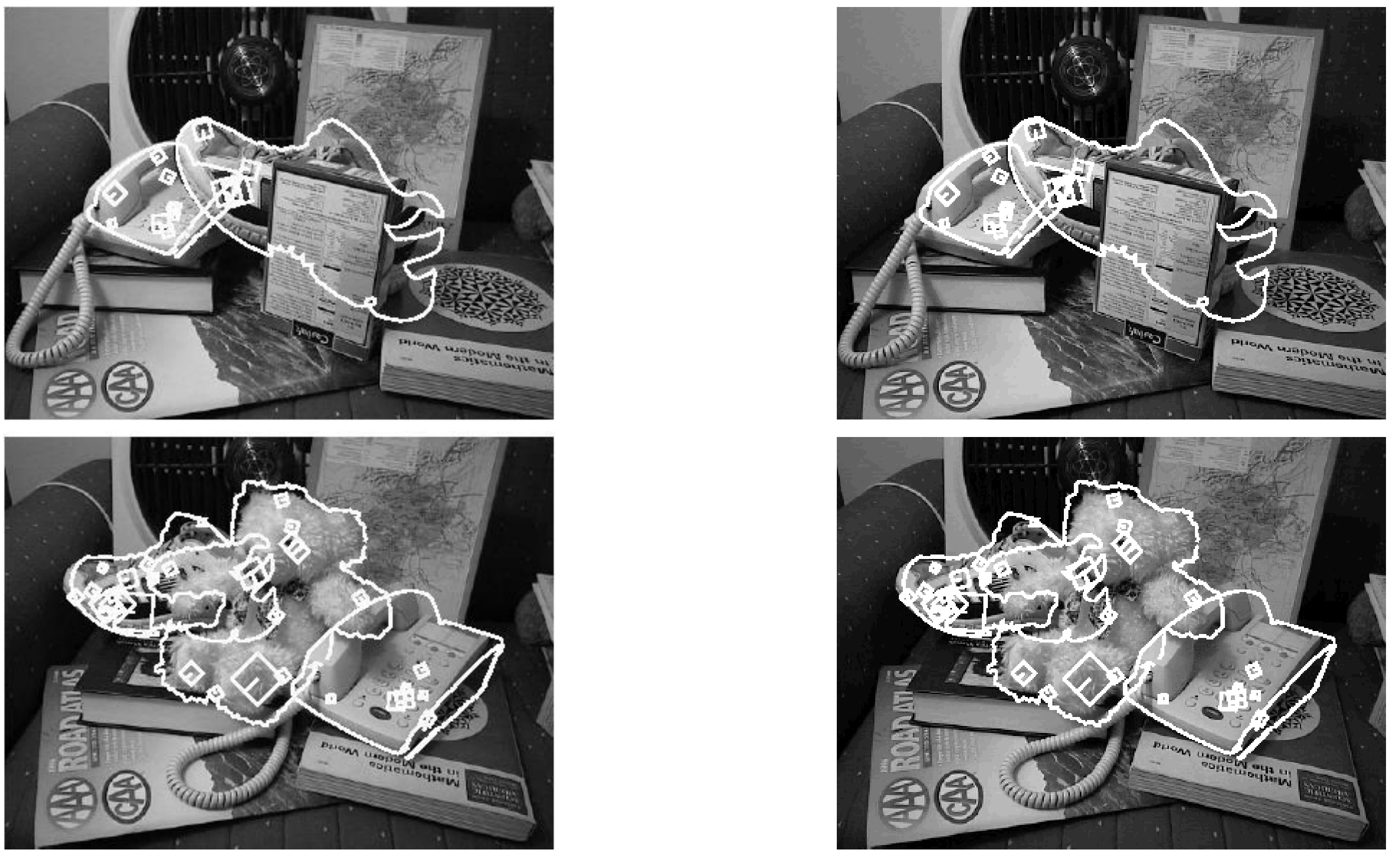
- Basic requirements:
  - – Sparse
  - – Informative
  - – Repeatable
- Invariance
  - – Rotation
  - – Scale (Similarity)
  - – Affine



# Recognizing Specific Objects

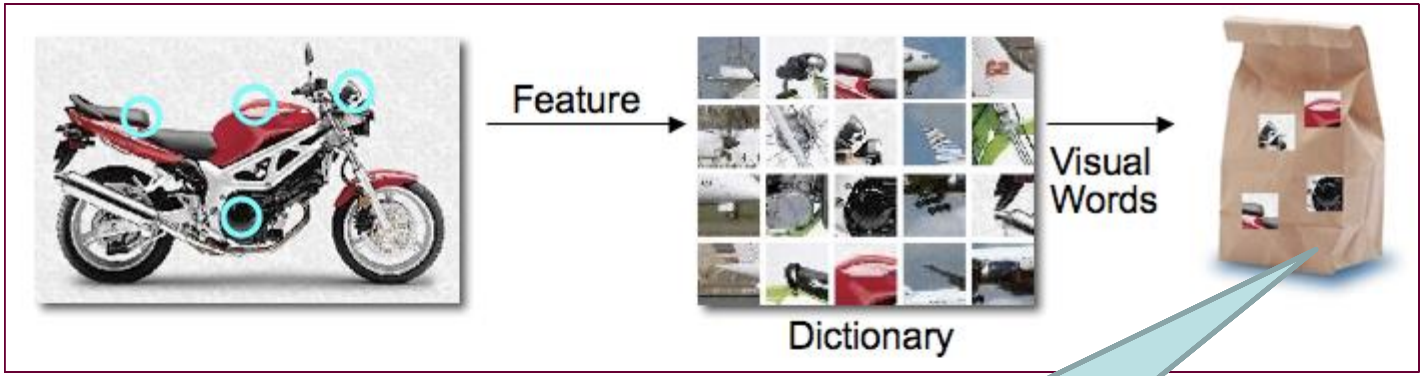
Learned models of local features, and got object outline from

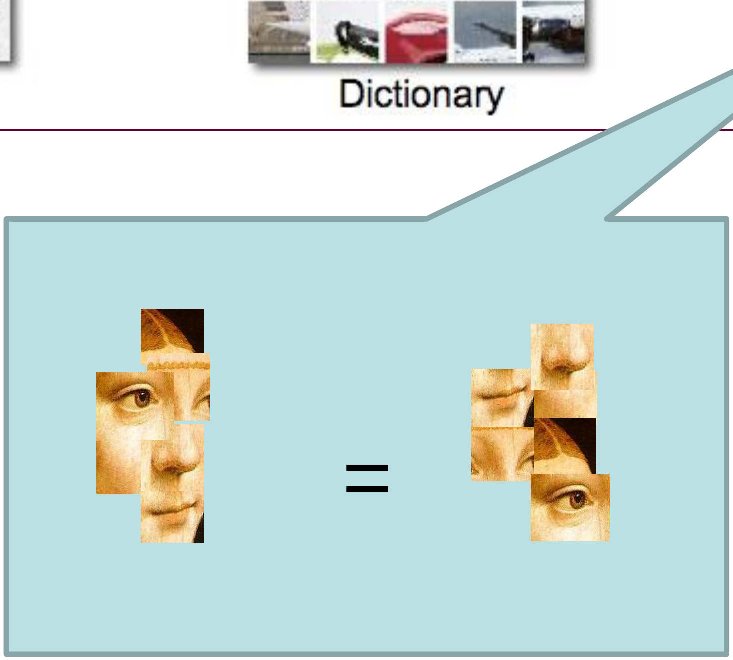
Objects may then be found under occlusion and 3D rotation



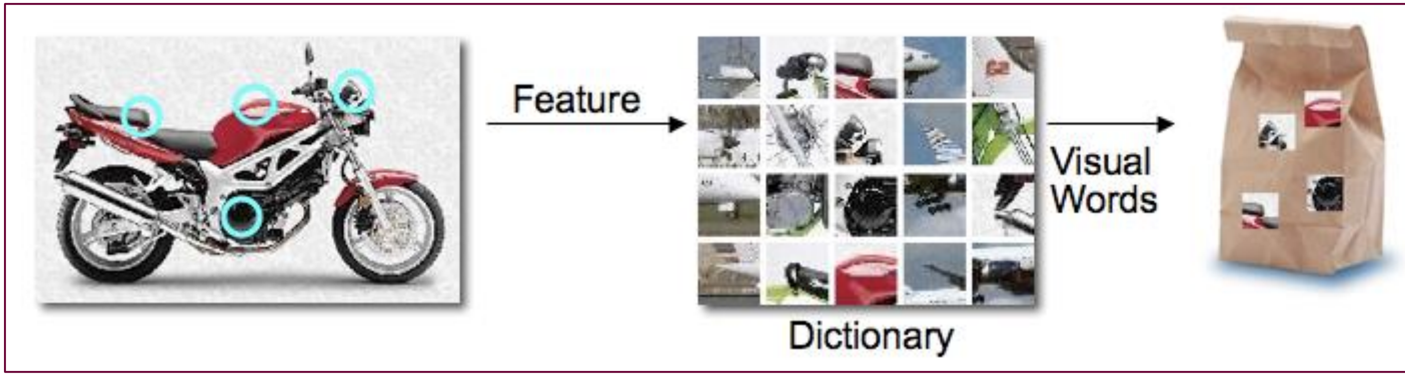
Images from: David Lowe, **Object Recognition from Local Scale-Invariant Features** Proc. of the International Conference on Computer Vision, Corfu (Sept. 1999)

# Bag of Features



# Bag of Features




Pros: fairly flexible and computationally efficient

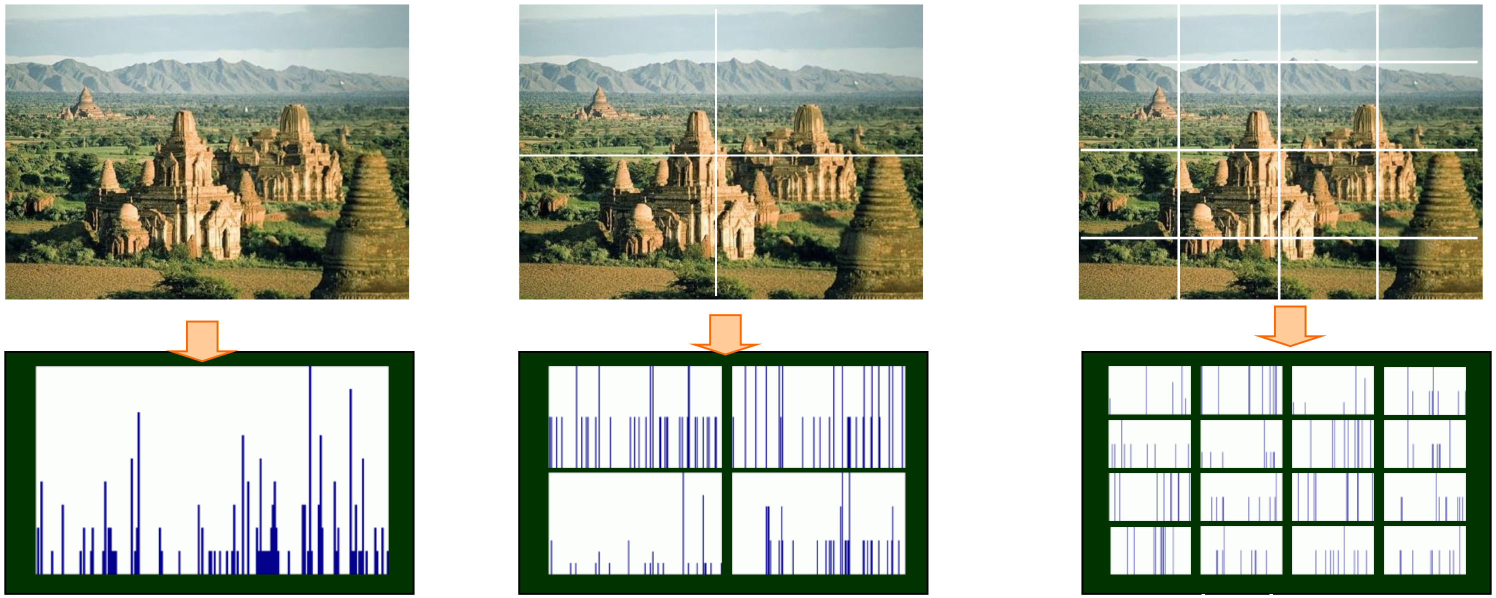
Cons: problems with large clutter

Different objects, but Similar representations;


Similar objects, different representations;

# Beyond Bags of Features

- Computing bags of features on sub-windows of the whole image.



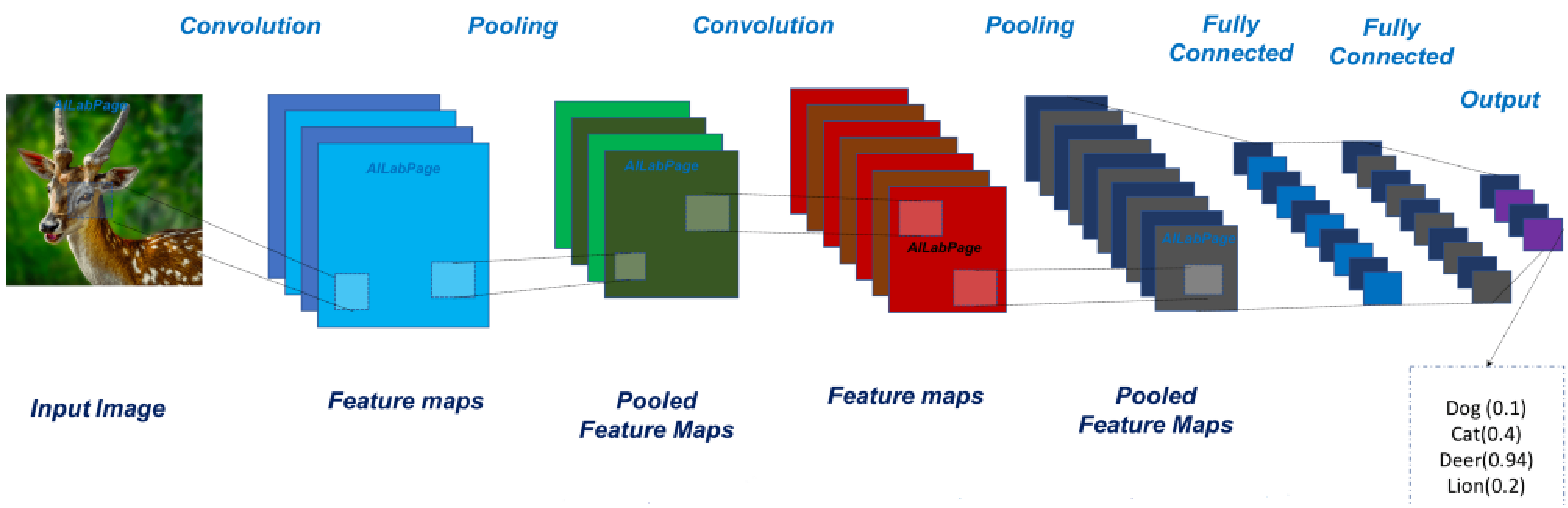
Lazebnik, Schmid & Ponce (CVPR 2006)

# Convolutional Neural Networks

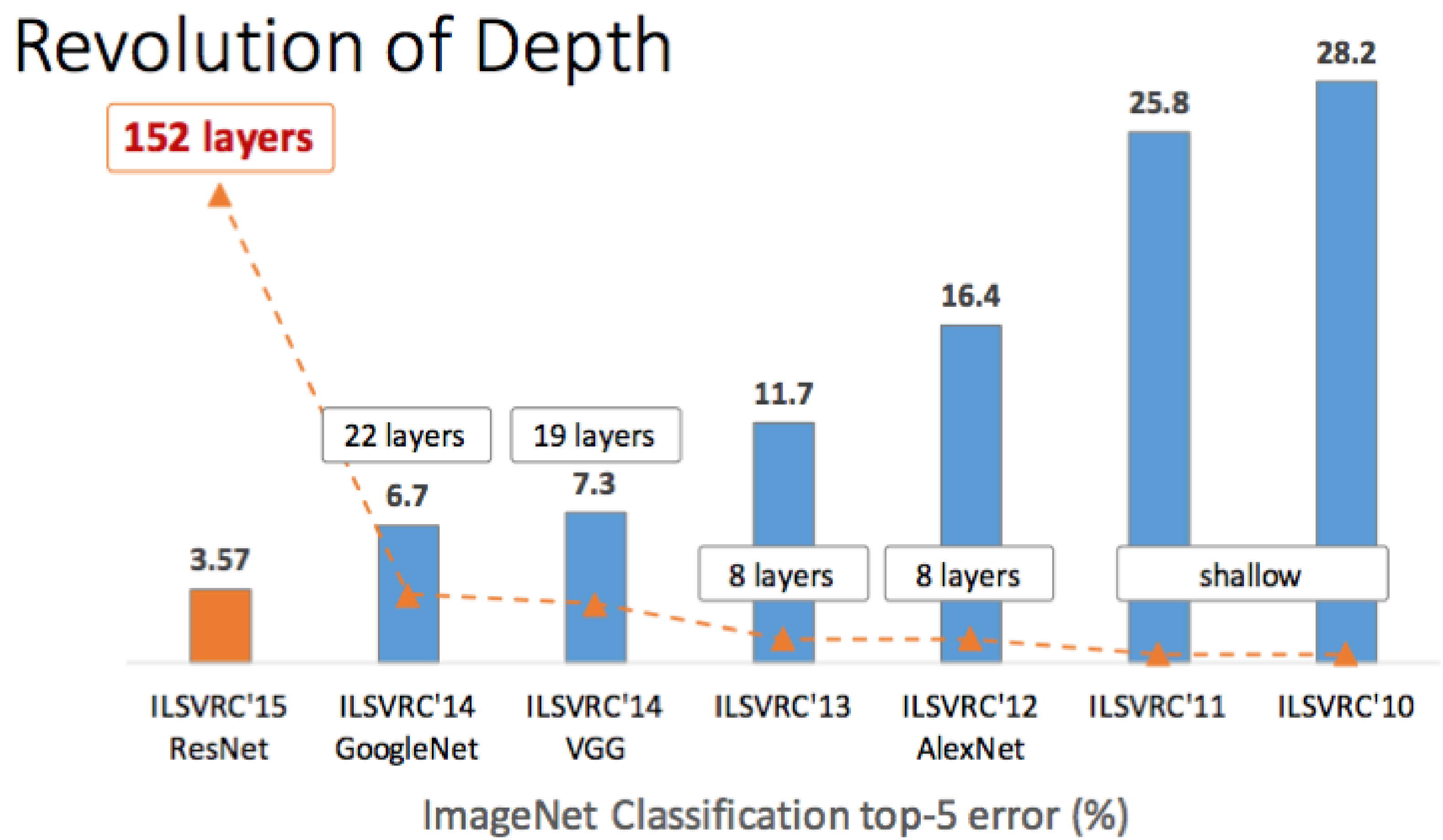
- Learn all in one deep architecture:
  - low level features
  - high level representations
  - context
  - classifiers
- Efficient Classification
- Efficient Detection
- Scalable to very large sets and large number of categories

Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks

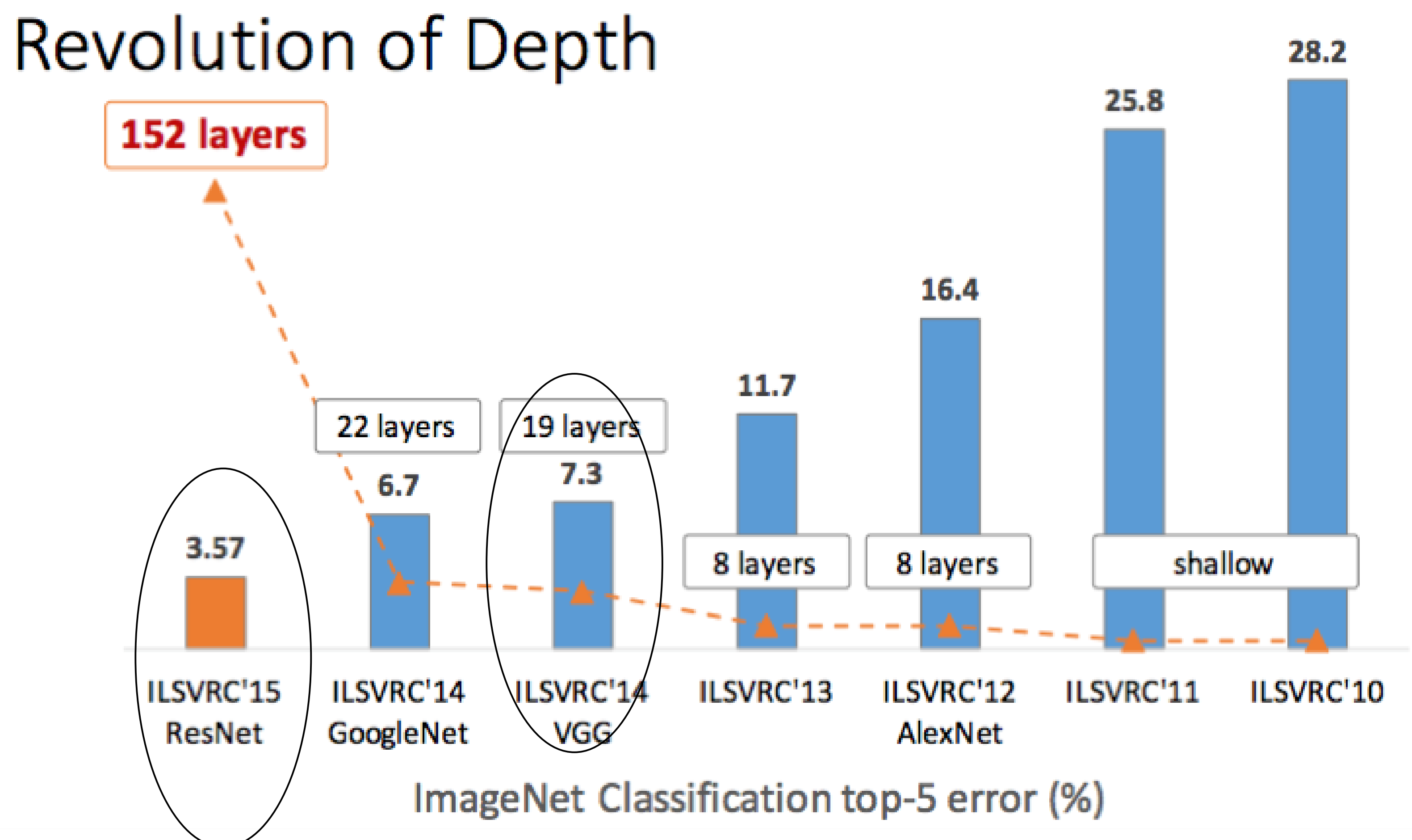
# Convolutional Neural Networks



# Very Deep Networks



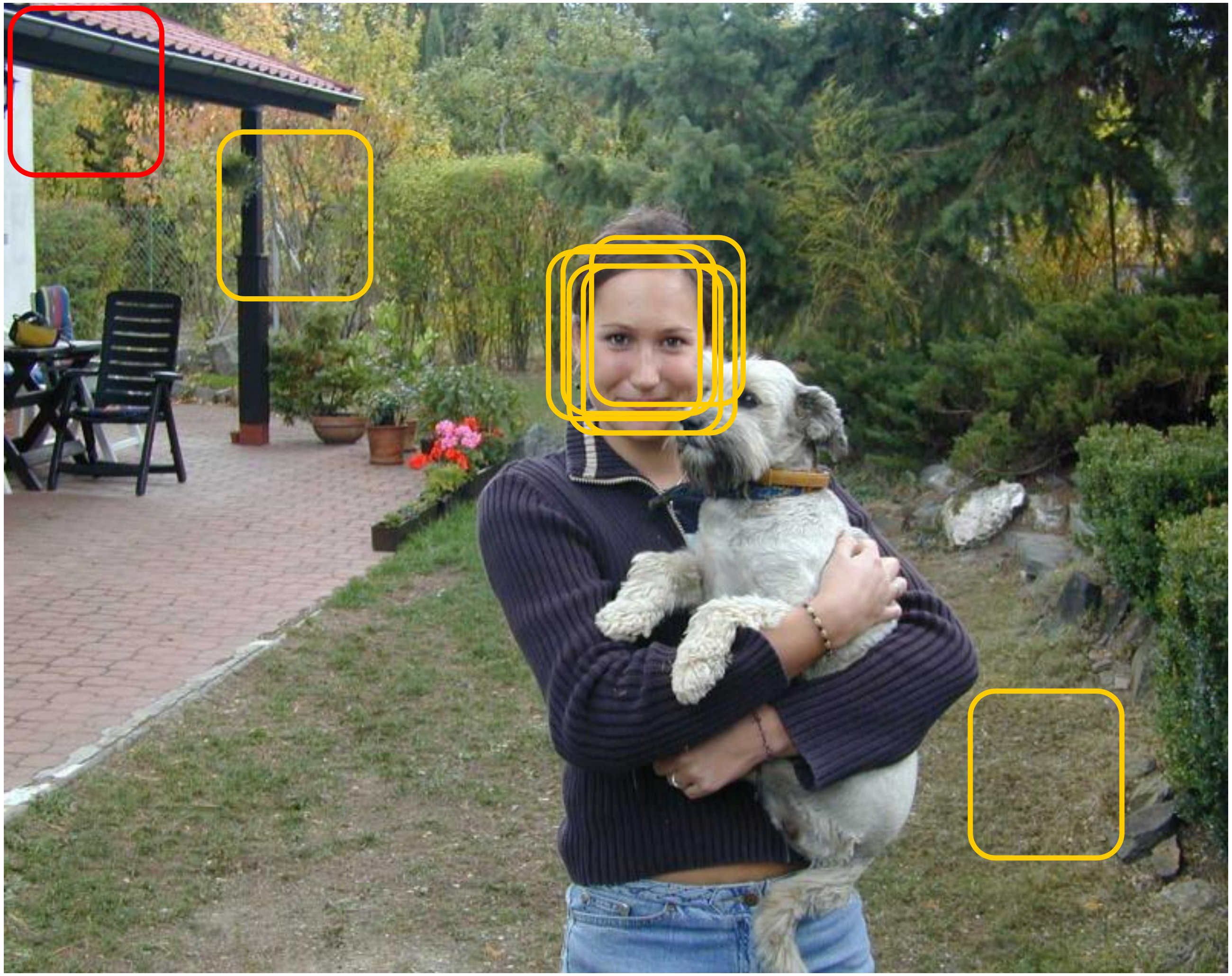
# Very Deep Networks



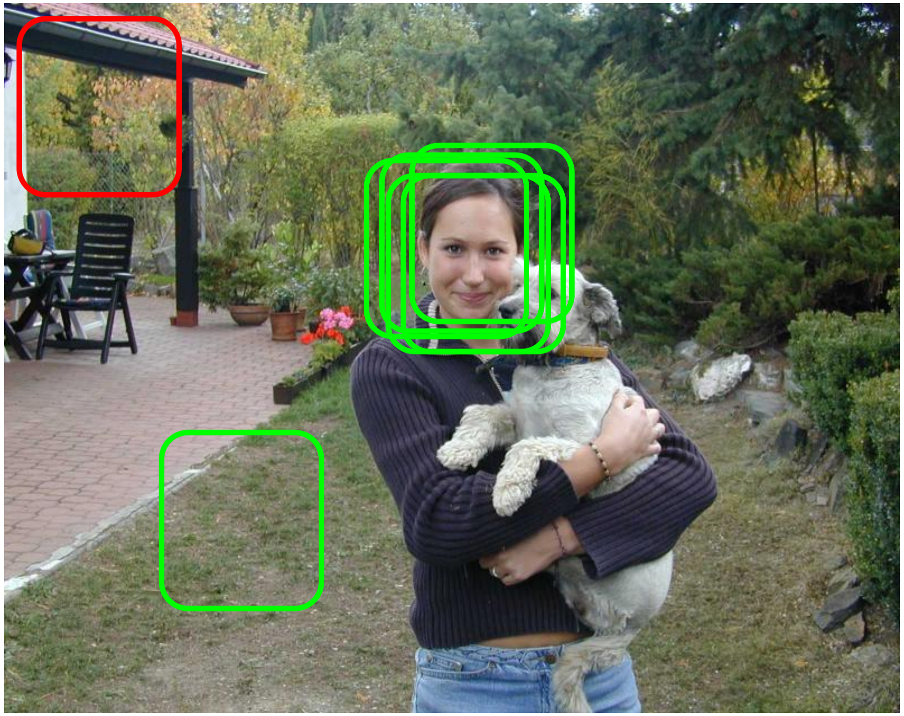

- K. Simonyan and A. Zisserman Very Deep Convolutional Networks for Large-Scale Image Recognition
- K. He, X. Zhang, S. Ren, and J. Sun: Deep Residual Learning for Image Recognition

# Detection

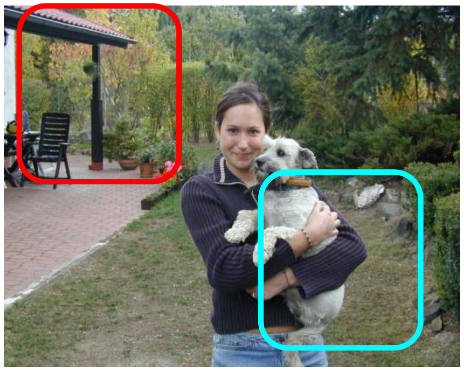
Apply classifier at Scale / position range to search over



# Detection

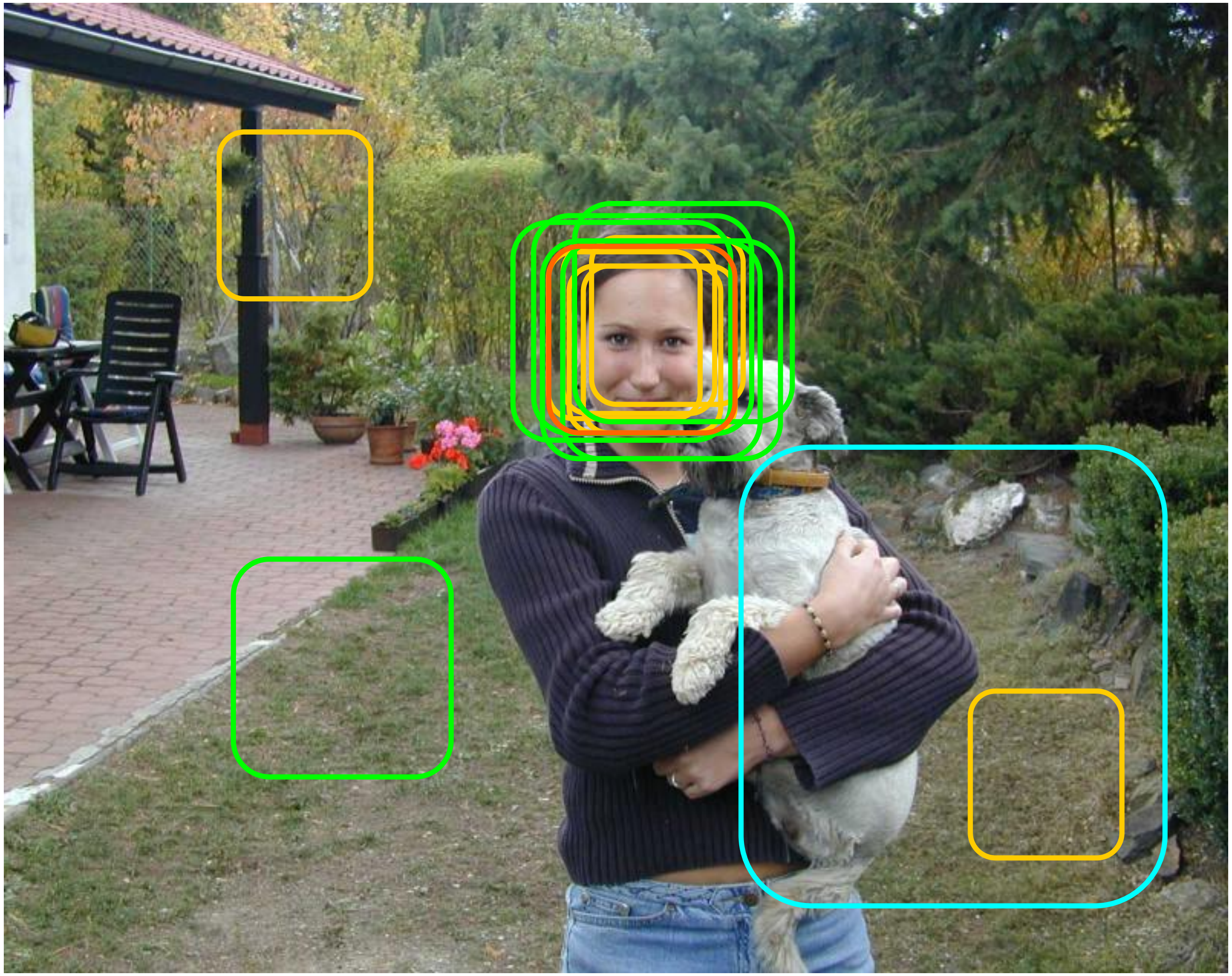


# Detection

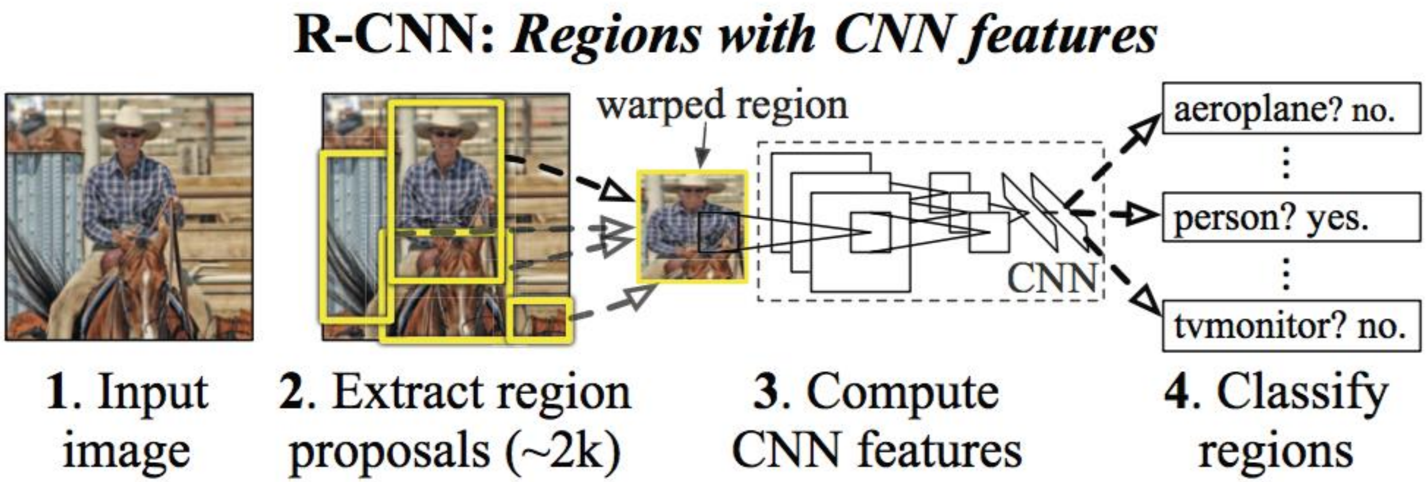


# Detection

– Combine detection over space and scale.



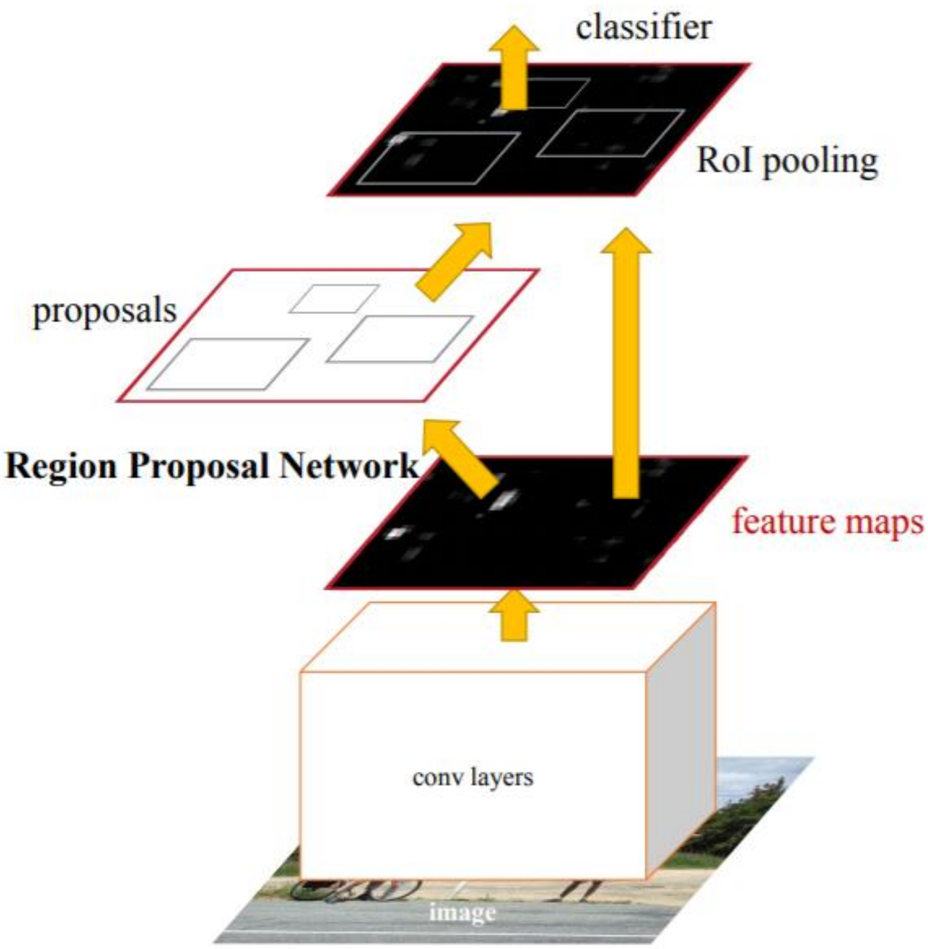
# Deep Learning in Object Detection




SVM classifier

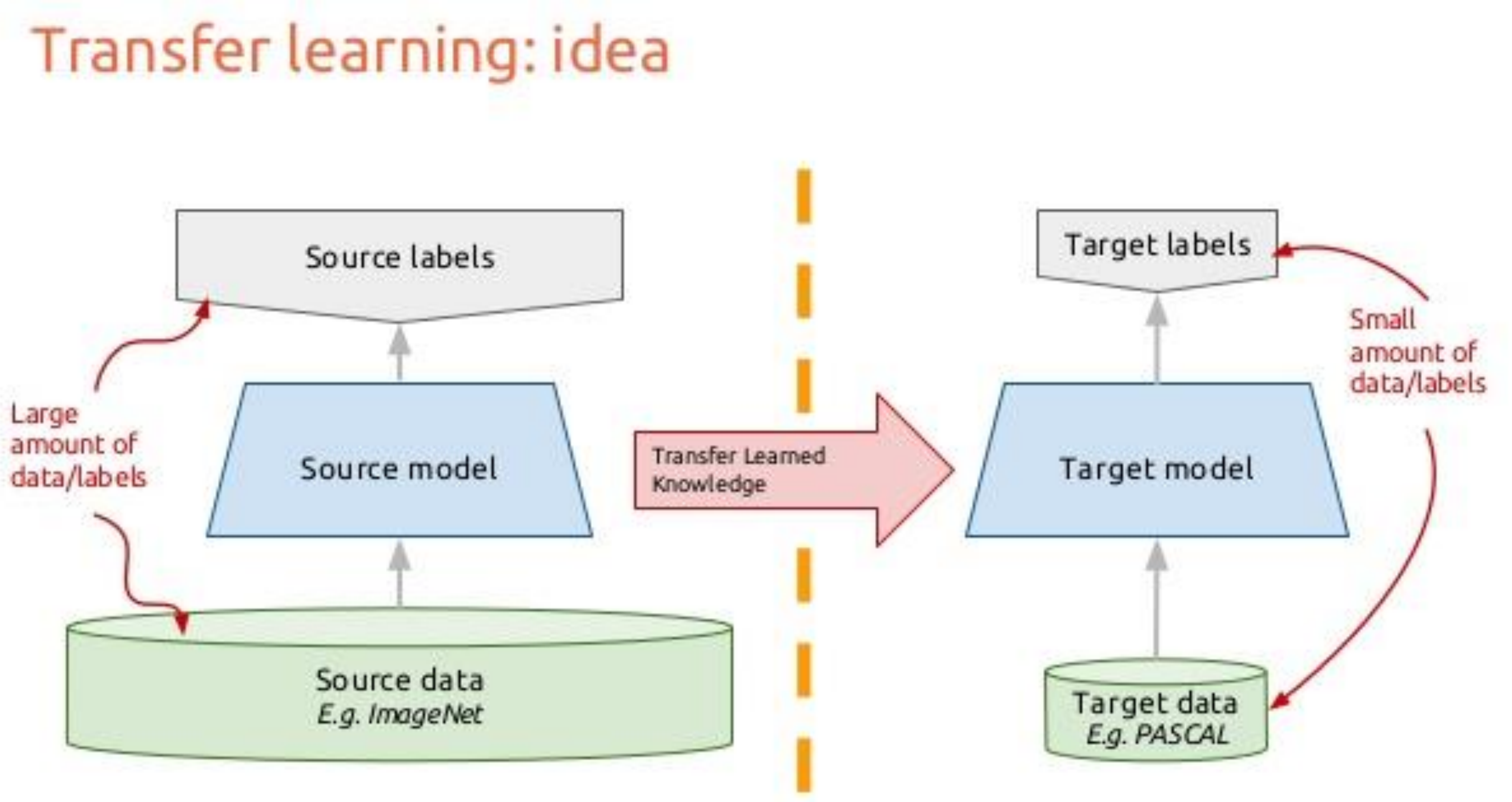
Rich feature hierarchies for accurate object detection and semantic segmentation. R. Girshick et al.

# Faster R-CNN



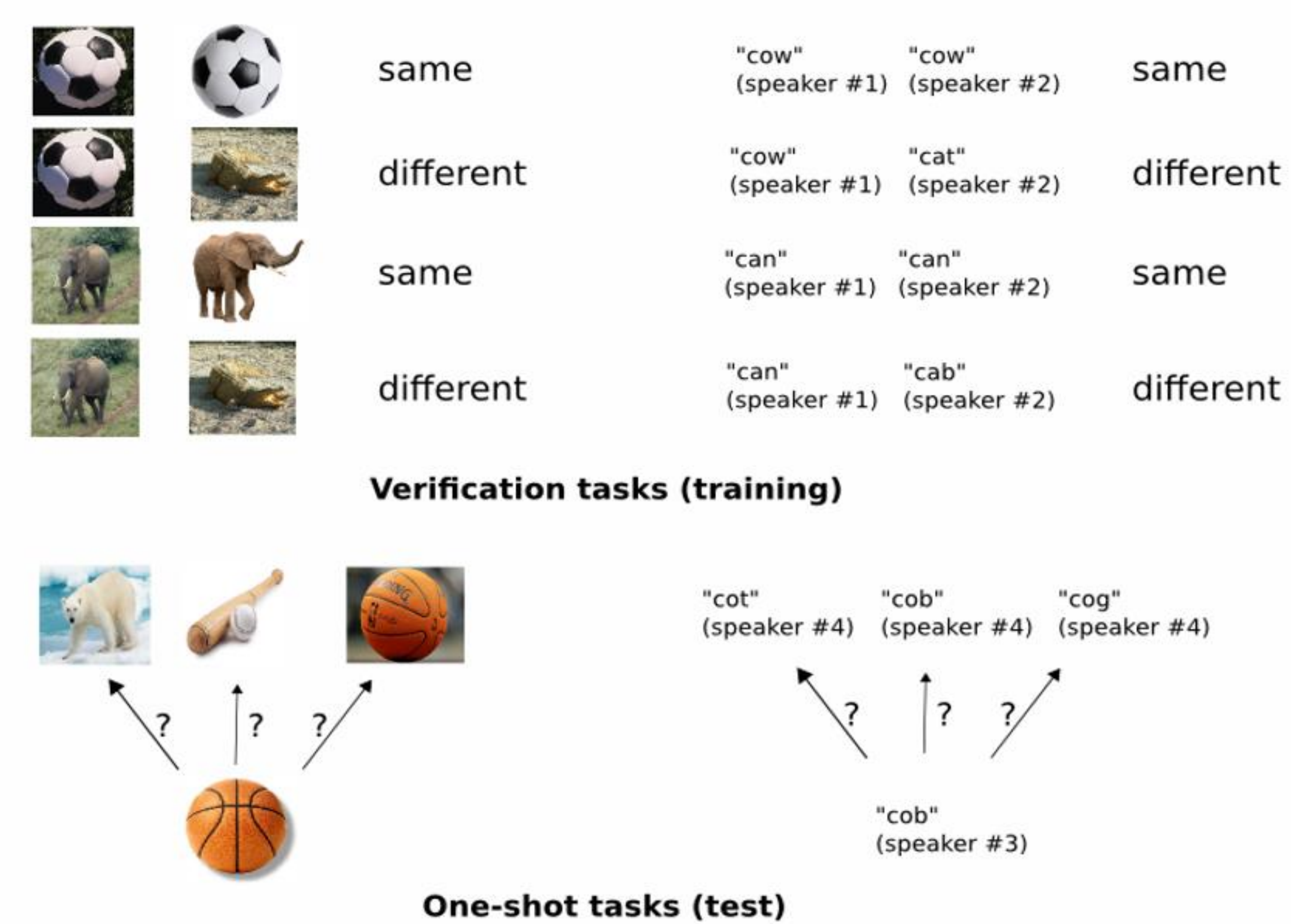

Faster R-CNN is a single, unified network for object detection. The RPN module serves as the ‘attention’ of this unified network.

# Transfer Learning




J. Donahue et.al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition.

# One Shot Learning




G. Koch, R. Zemel, and R. Salakhutdinov Siamese Neural Networks for One-shot Image Recognition

# Few-Shot Learning

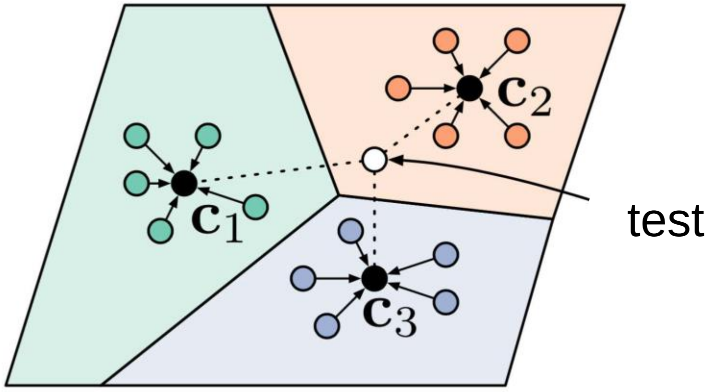
- Prototypical Networks

$$p_\phi(y = k \mid \mathbf{x}) = \frac{\exp(-d(f_\phi(\mathbf{x}), \mathbf{c}_k))}{\sum_{k'} \exp(-d(f_\phi(\mathbf{x}), \mathbf{c}_{k'}))}$$

$$\mathbf{c}_k = \frac{1}{|S_k|} \sum_{(\mathbf{x}_i, y_i) \in S_k} f_\phi(\mathbf{x}_i)$$

$$S_k = \{(\mathbf{x}_i, y_i) | y_i = k, (\mathbf{x}_i, y_i) \in D_{train}\}$$

$$\phi \equiv \Theta$$




- Maps examples to embedding such that examples of a given class are close together
- Calculates a prototype (mean vector) for every class
- Maps test instances to the same embedding
- Uses softmax over distance to prototype for label prediction

Prototypical Networks for Few-shot Learning (2017)
Jake Snell, Kevin Swersky, Richard S. Zemel

# Describing Objects with Attributes

- Shift the goal of recognition from naming to describing


Discover/detect new categories

Describing Objects by Their Attributes, A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth

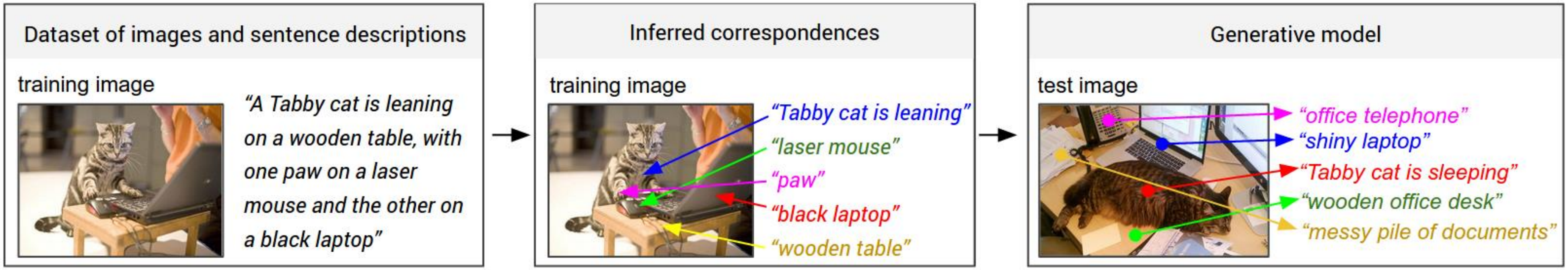
# Improvement

| Attributes | Presence | | Rating | |
|---|---|---|---|---|
| | walrus | polar bear | walrus | polar bear |
| Spot | no | no | less relevant | irrelevant |
| Blue | no | no | irrelevant | less relevant |
| Swim | yes | yes | highly relevant | relevant |
| Coastal | yes | yes | relevant | highly relevant |

Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer, C. Lampert, H. Nickisch, and S. Harmeling

Lecture 8

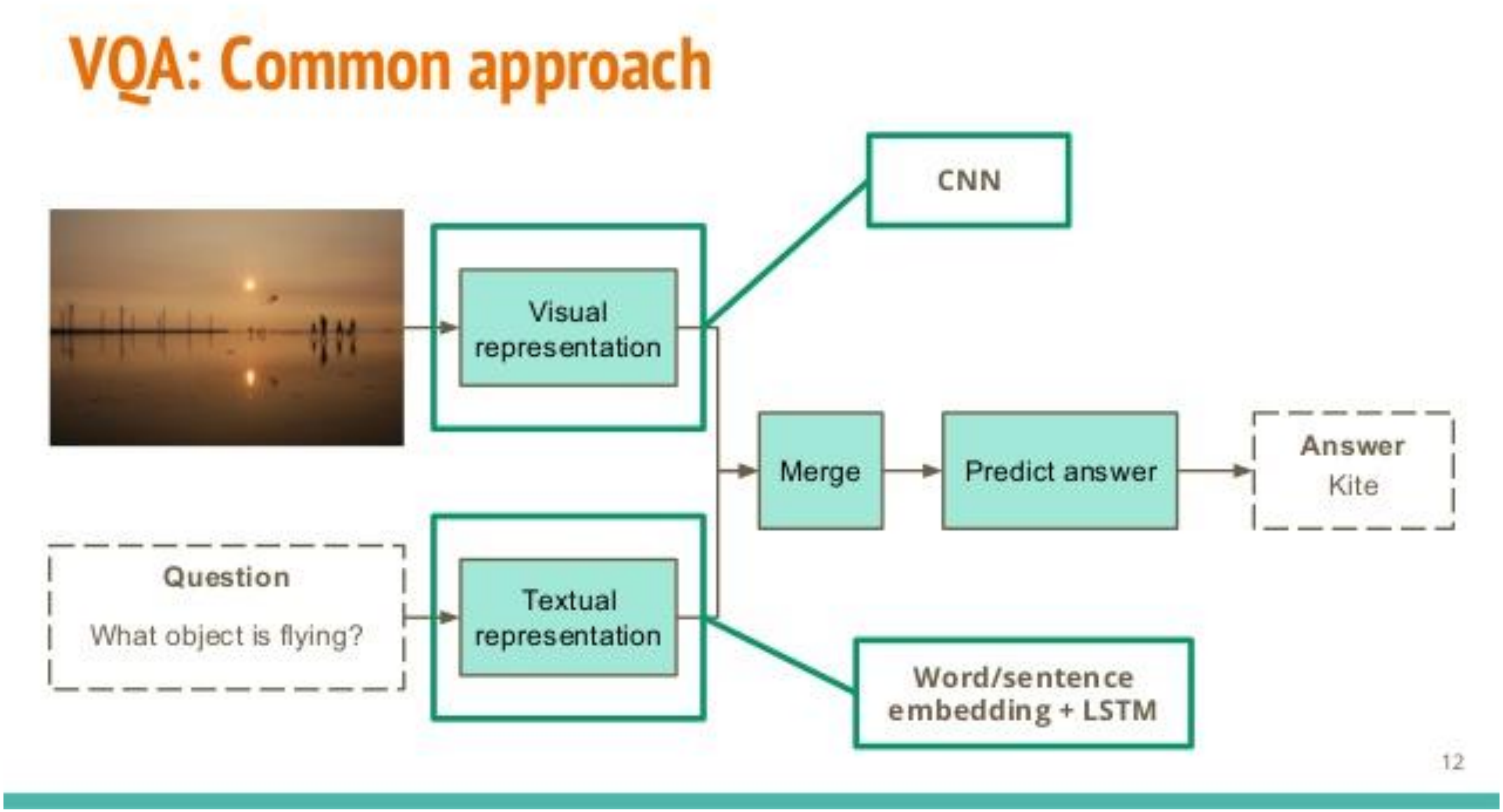
# Image Descriptions (Captioning)




Uses CNNs and RNNs

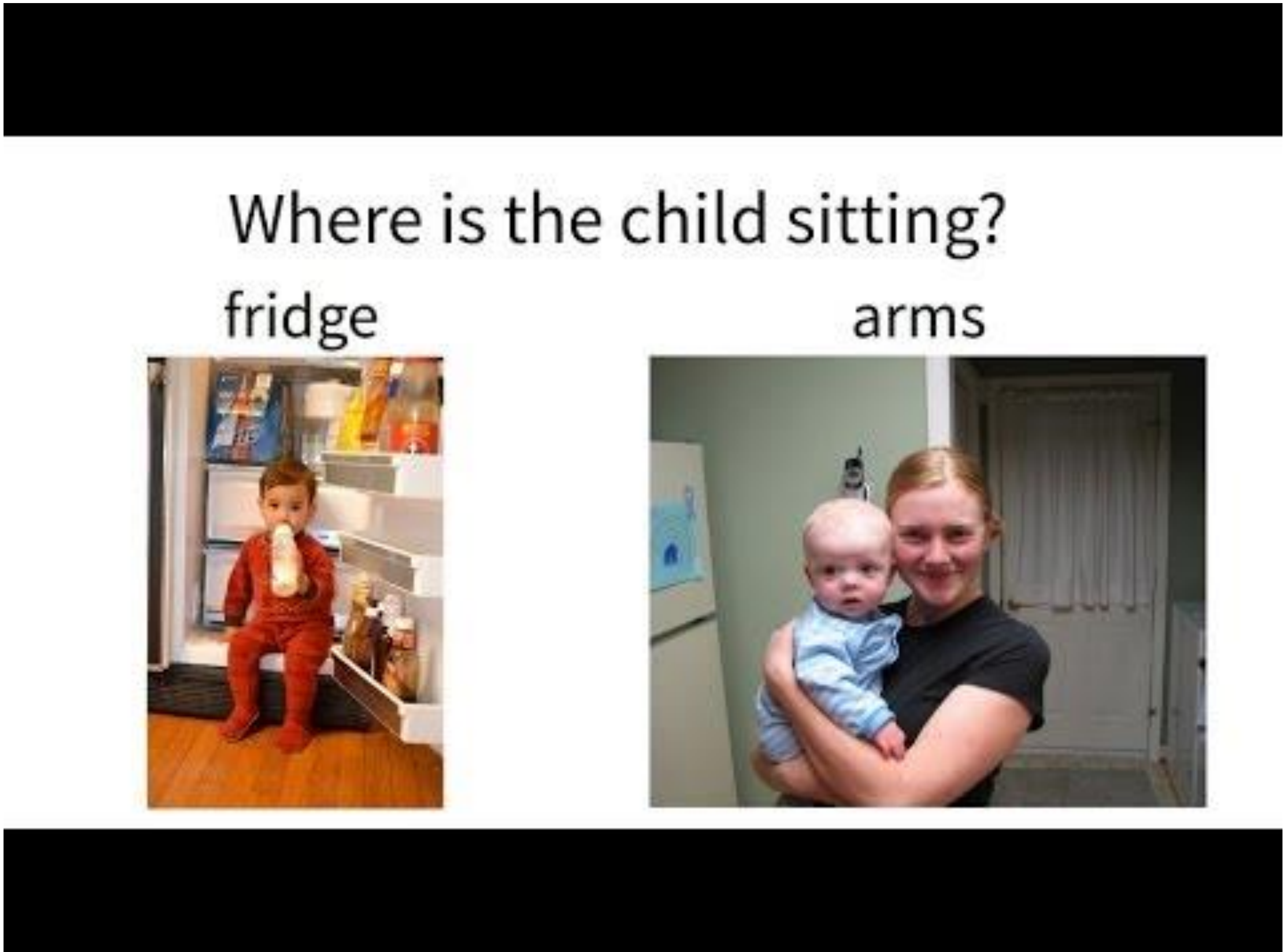
Deep Visual-Semantic Alignments for Generating Image Descriptions, A. Karpathy, Li Fei-Fei.

# VQA: Visual Question Answering



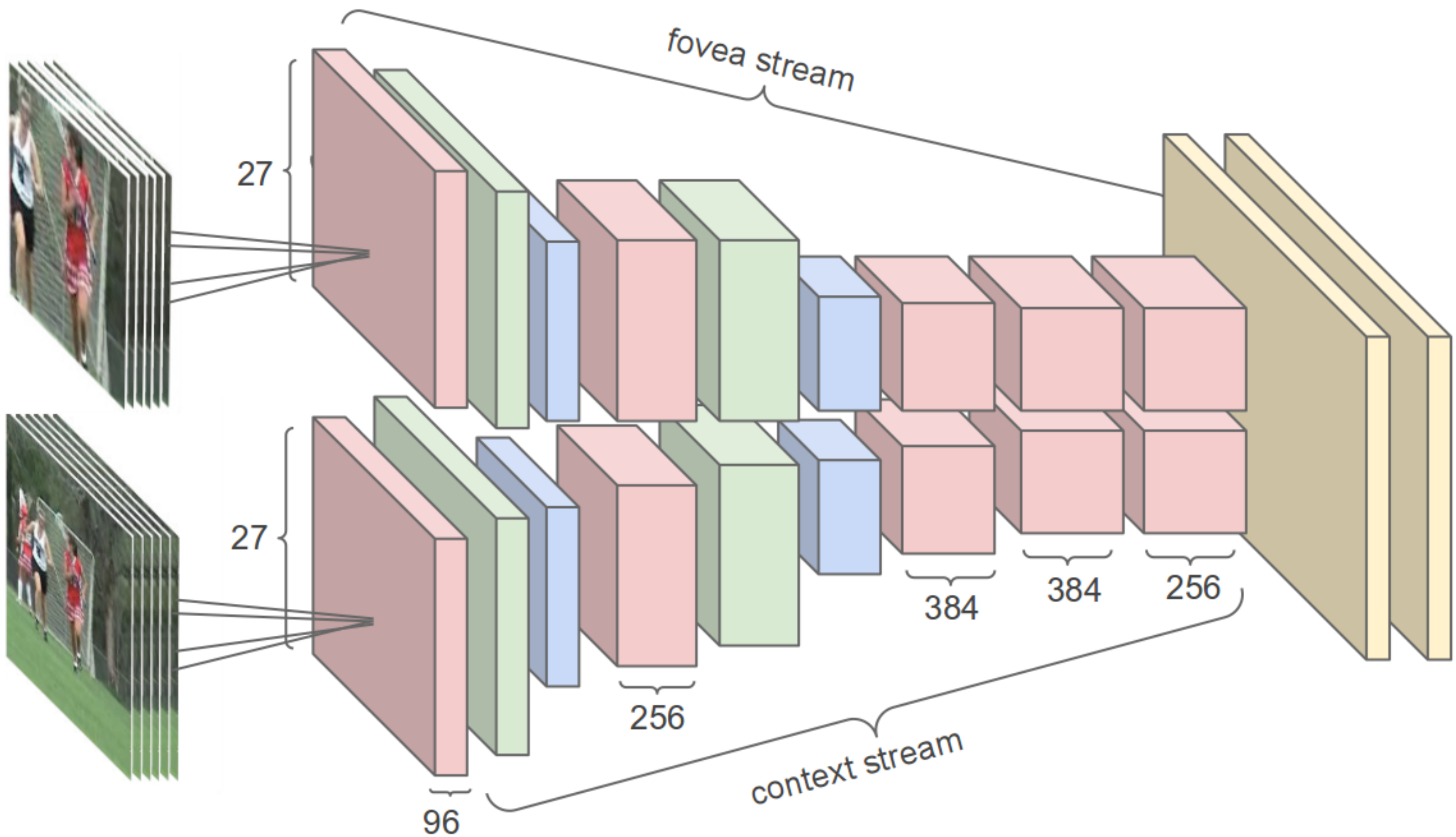

VQA: Visual Question Answering. Antol et al.

# Results



# Video Classification

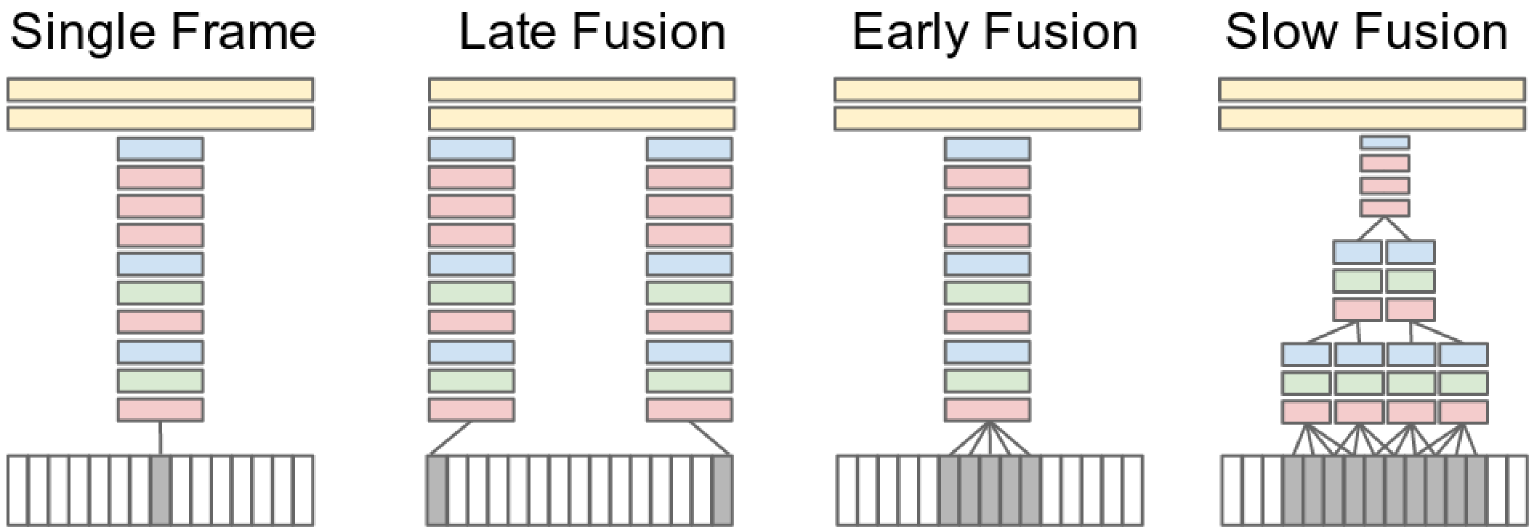
- Using CNN – Naïve Approach




Kapathy et al.: Large-scale Video Classification with Convolutional Neural Networks

# Video Classification

- Using CNN – Naïve Approach

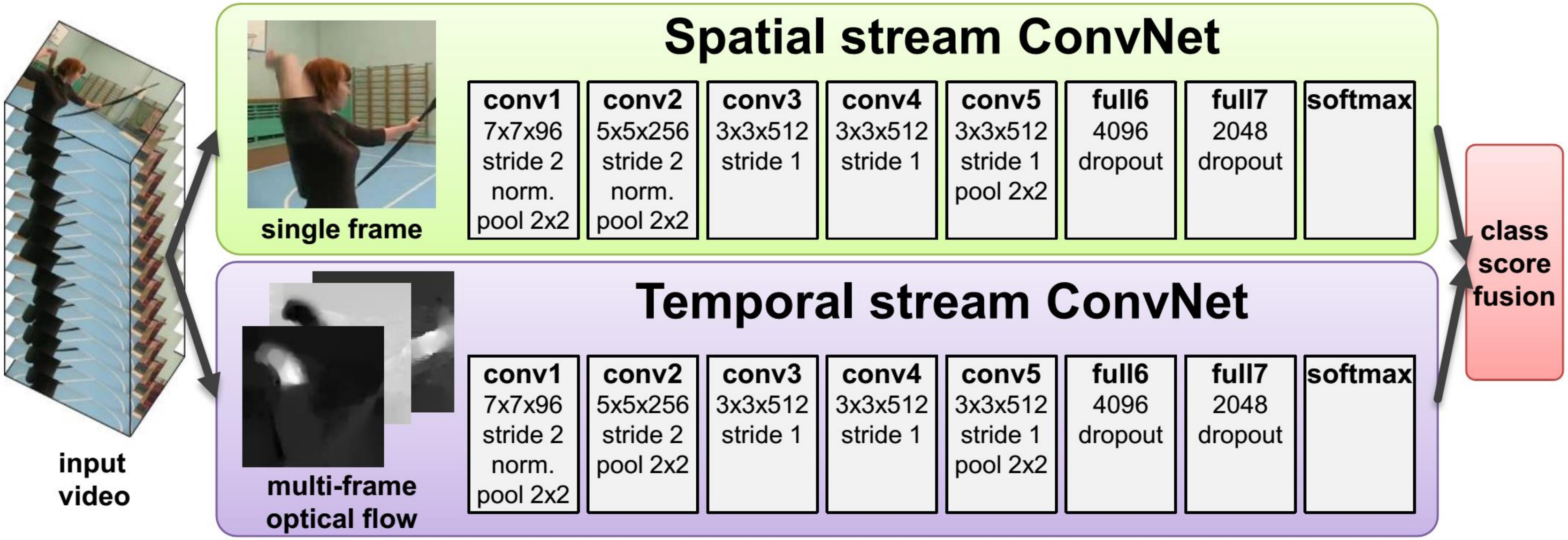



Temporal fusion

Kapathy et al.: Large-scale Video Classification with Convolutional Neural Networks

# Video Classification

- Modern Approaches




K. Simonyan, A. Zissermann: Two-Stream Convolutional Networks for Action Recognition in Videos.

Lecture 10

# Multi-task training

Lecture 10

# Continual Learning

Training Set at Time $T_1$

Training Set at Time $T_2$

Training Set at Time $T_3$

...

# Continual Learning

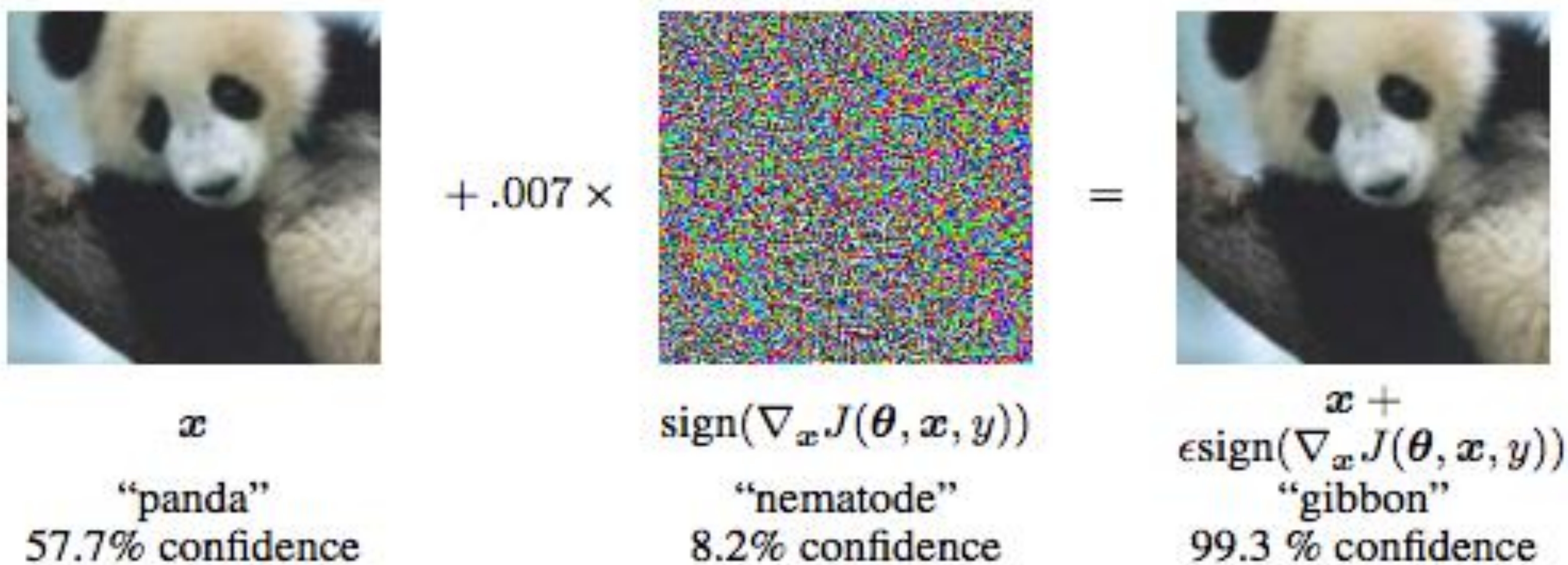
- Tasks are learned sequentially over time.
- At time $T_i$, the data of the tasks $T_1 \ldots T_{i-1}$ is no longer available.
- Leads to forgetting of previously learned tasks.
- Termed “Catastrophic Forgetting”.

Possible Solution:

- Measure the importance of learnable parameters in NN for the task, constrain their change in learning a new task.
- This could work because deep networks are highly overparametrized

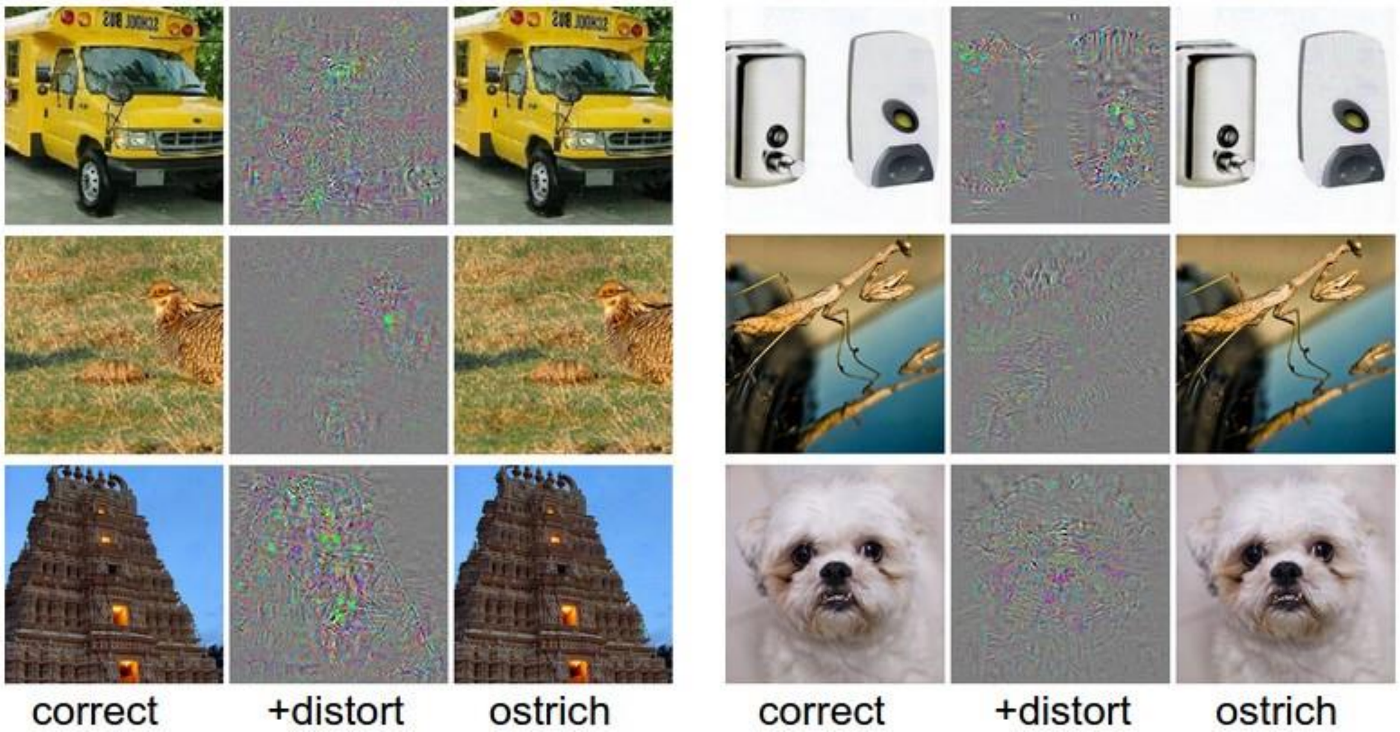
James Kirkpatrick et al., “Overcoming catastrophic forgetting in neural networks”

# Adversarial Examples




Explaining and Harnessing Adversarial Examples by Goodfellow et al.

# Adversarial Examples: Imperceptible Noise




Intriguing properties of neural networks by Szegedy et al.