Object Detection

Danna Gurari
University of Colorado Boulder
Fall 2024

Review

- Last lecture: semantic segmentation
  - Motivation
  - Datasets
  - Evaluation metric
  - Fully convolutional network
  - Swin transformer
  - Discussion

- Assignments (Canvas)
  - Reading assignment was due earlier today
  - Project proposal due Wednesday
  - Reading assignment due next Monday

- Questions?

Object Detection: Today’s Topics

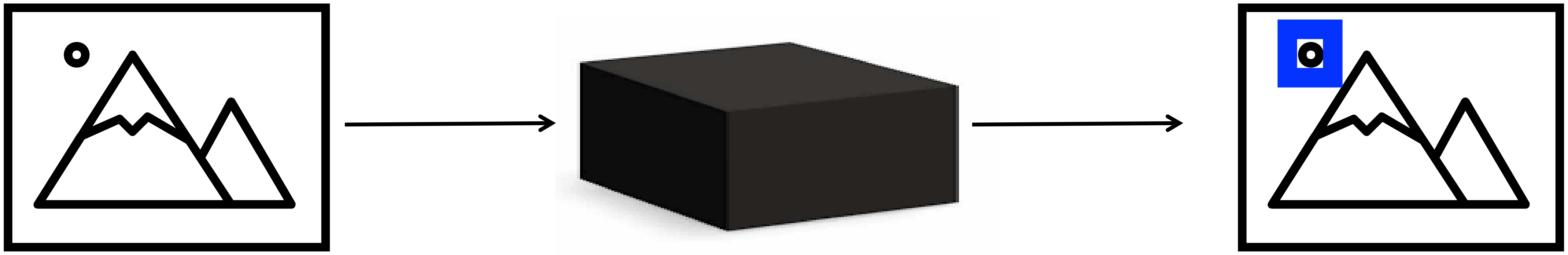
- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)

Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)

Problem Definition

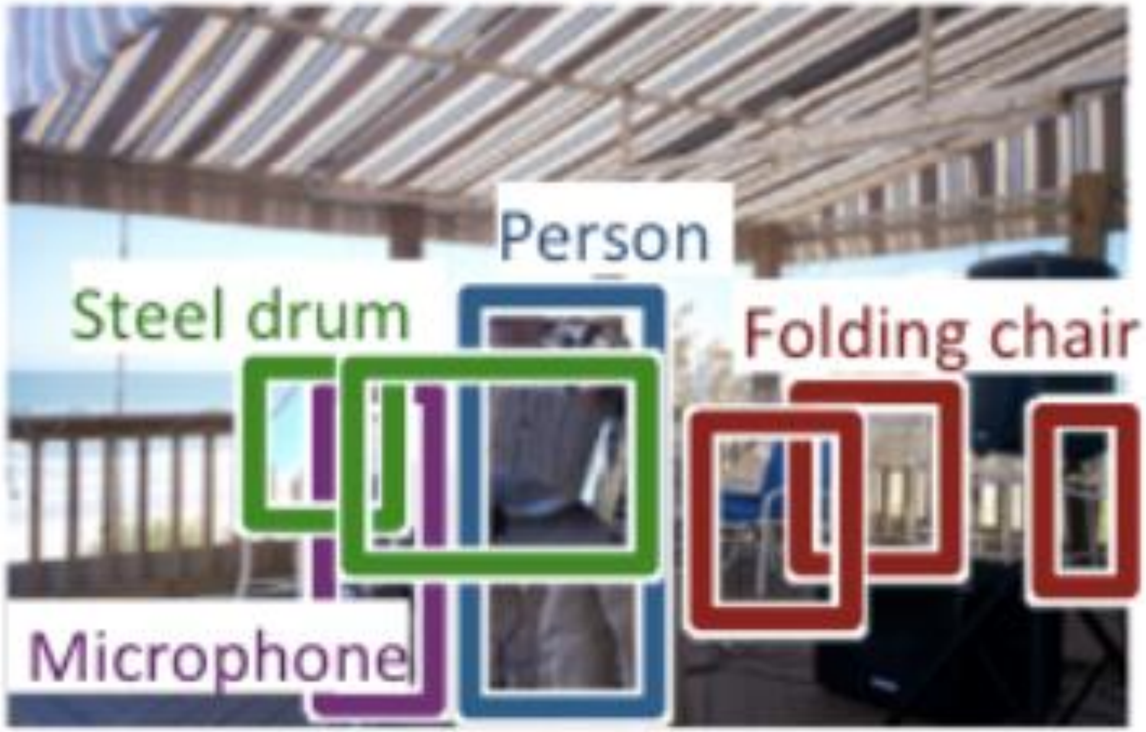
- Localize with a bounding box object(s) of interest



Focus for today's lecture

Problem: **Semantic** Object Detection

- Localize with a bounding box every instance of an object from pre-specified categories



[Russakovsky et al; IJCV 2015]

A reasonably solved problem

Problem: **Salient** Object Detection

- Localize with a bounding box the salient object(s)

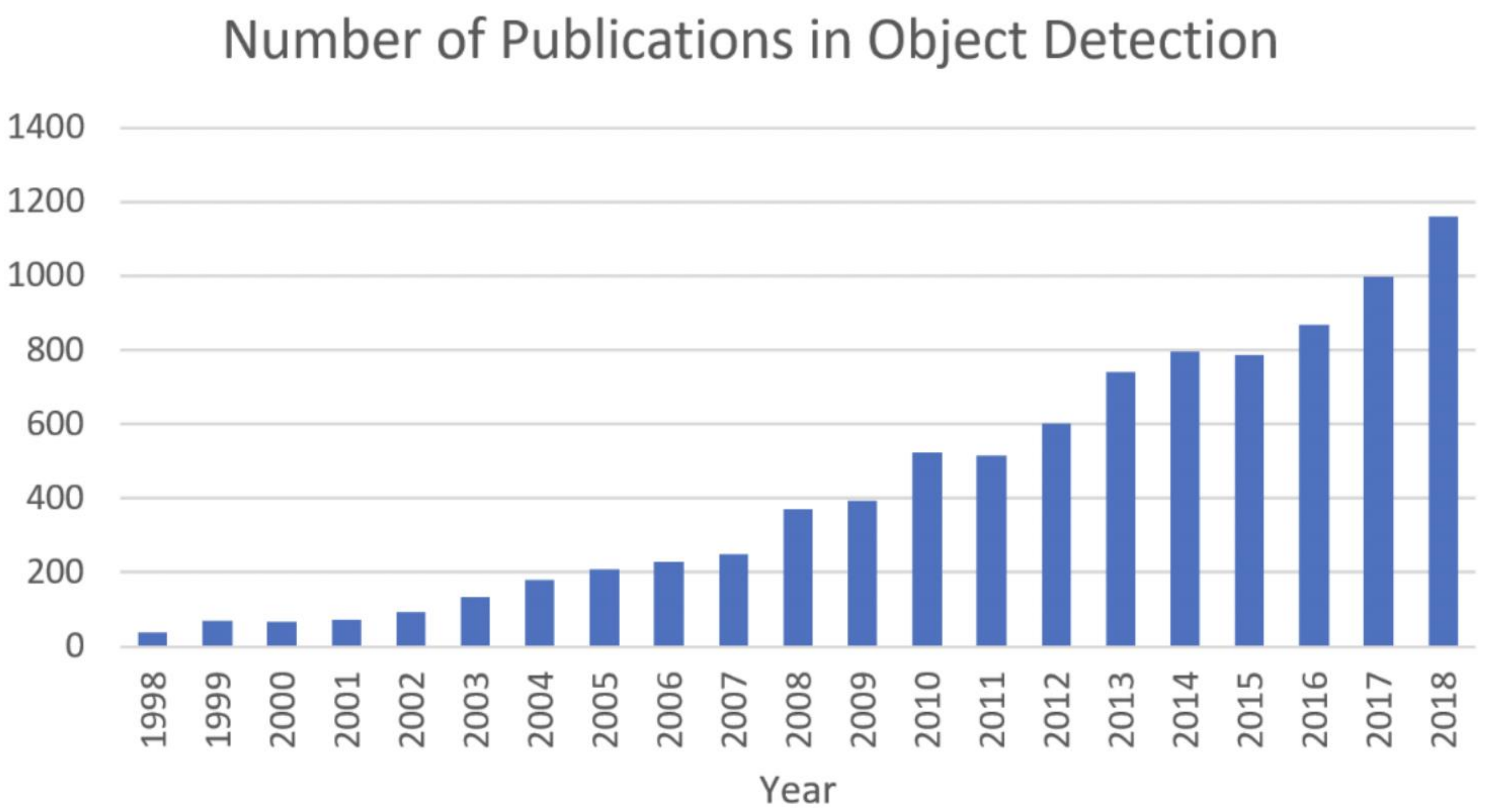
[Liu et al; CVPR 2007]

Object Detection vs Object Recognition

"How does (semantic) object detection differ from object recognition?"

- Extends object recognition of assigning labels **by also** indicating each object's location with rectangular coordinates (necessitating different model architectures and loss functions)
- Must learn an object's appearance rather than only its image context;
  - e.g., giraffes are often photographed in savannah-like landscapes

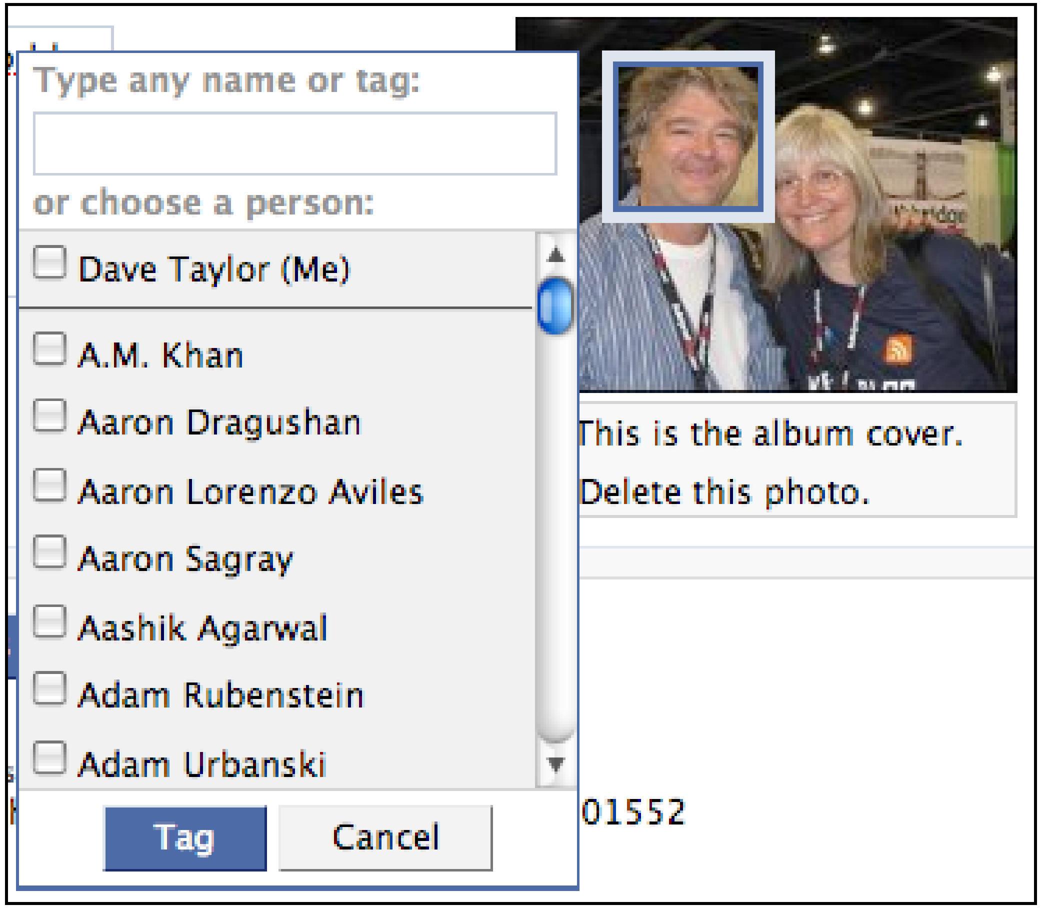
Community Research Engagement



"Data from Google scholar advanced search: allintitle: 'object detection' AND 'detecting objects'"

Zhou et al. Object Detection in 20 Years: A Survey. arXiv 2019

Application: Social Media



Face detection
(e.g., Facebook)

Application: Banking

Mobile check deposit
(e.g., Bank of America)

Application: Transportation

License Plate Detection (e.g., AllGoVision)

Application: Construction Safety

Pedestrian Detection
(e.g., Blaxtair)

http://media.brintex.com/Occurrence/121/Brochure/3435/brochure.pdf

Application: Counting

e.g., Business Traffic Analytics

Can you think of any other potential applications?

Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)

Object Detection Datasets

1945 — 

1957 — 

1966 — Birth of CV

1983 — CVPR

1987 — ICCV

1990 — ECCV

2007 — VOC

2012 — ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Recall VOC

1. Category Selection

- 20 categories chosen:
1) Initial 4 categories stem from existing dataset
2) 2006: added 6 classes
3) 2007: added 10 classes

- Categories added for more generalization and finer-grained coverage

→

2. Image Collection

- 500,000 images retrieved from Flickr with many search terms

→

3. Image Verification + Image Annotation

- University of Leeds annotation party to recruit annotators

- Annotation guidelines & real-time assistance

- Review of every annotation

- Annotate only "minority" classes at end of party to increase the count of them

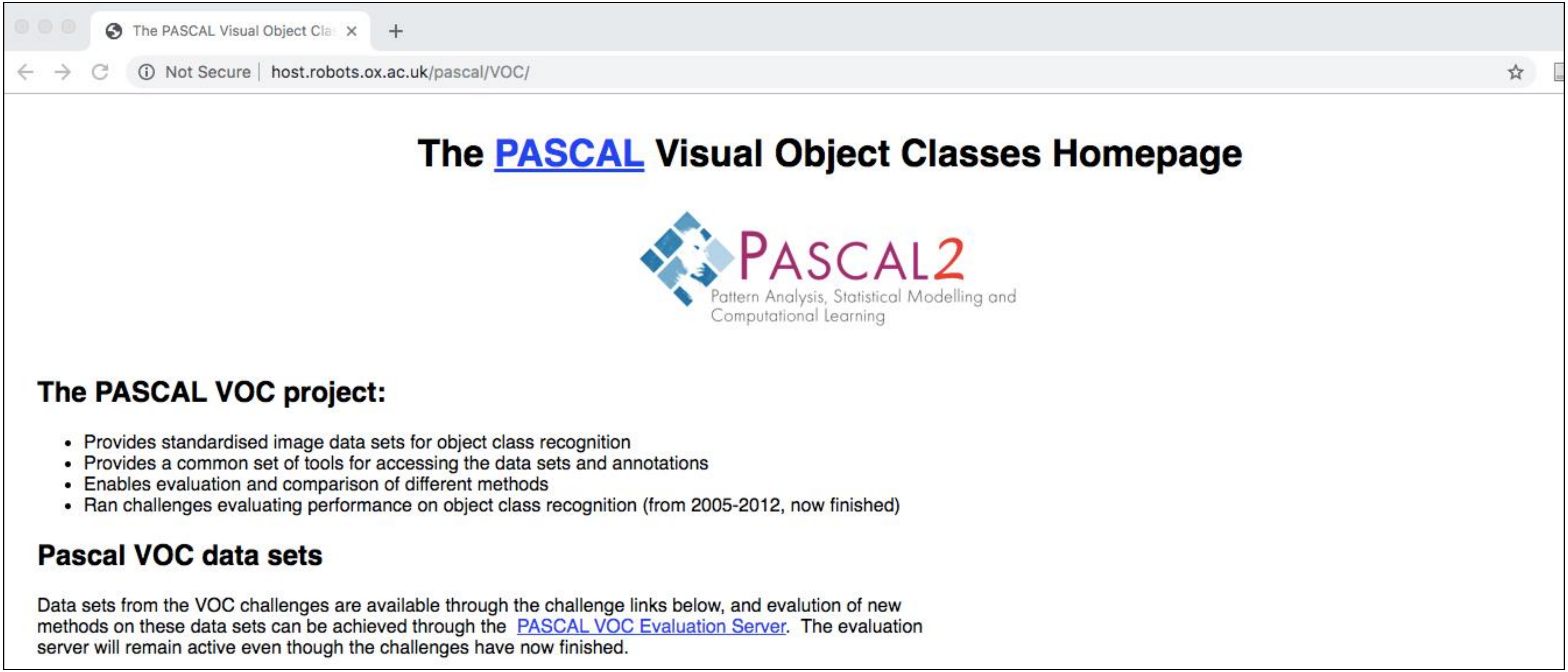
Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. IJCV 2010

VOC Guidelines:

What are potential limitations of this task design for resulting datasets (and so algorithms developed with such datasets)?

What to label	*All objects of the defined categories*, unless: • you are unsure what the object is. • the object is very small (at your discretion). • less than 10-20% of the object is visible, *such that you cannot be sure what class it is*. e.g. if only a tyre is visible it may belong to car or truck so cannot be labelled car, but feet/faces can only belong to a person. If this is not possible because too many objects, mark image as bad.
Viewpoint	Record the viewpoint of the 'bulk' of the object e.g. the body rather than the head. Allow viewpoints within 10-20 degrees. If ambiguous, leave as 'Unspecified'. Unusually rotated objects e.g. upside-down people should be left as 'Unspecified'.
Bounding box	Mark the bounding box of the visible area of the object (*not* the estimated total extent of the object). Bounding box should contain all visible pixels, except where the bounding box would have to be made excessively large to include a few additional pixels (<5%) e.g. a car aerial.
Truncation	If more than 15-20% of the object lies outside the bounding box mark as Truncated. The flag indicates that the bounding box does not cover the total extent of the object.
Occlusion	If more than 5% of the object is occluded within the bounding box, mark as Occluded. The flag indicates that the object is not totally visible within the bounding box.
Image quality/ illumination	Images which are poor quality (e.g. excessive motion blur) should be marked bad. However, poor illumination (e.g. objects in silhouette) should not count as poor quality unless objects cannot be recognised. Images made up of multiple images (e.g. collages) should be marked bad.
Clothing/mud/ snow etc.	If an object is 'occluded' by a close-fitting occluder e.g. clothing, mud, snow etc., then the occluder should be treated as part of the object.
Transparency	Do label objects visible through glass, but treat reflections on the glass as occlusion.
Mirrors	Do label objects in mirrors.
Pictures	Label objects in pictures/posters/signs only if they are photorealistic but not if cartoons, symbols etc.

Recall VOC Annual Workshop



The PASCAL Visual Object Cla ×

Not Secure | host.robots.ox.ac.uk/pascal/VOC/

The PASCAL Visual Object Classes Homepage

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

The PASCAL VOC project:

- Provides standardised image data sets for object class recognition
- Provides a common set of tools for accessing the data sets and annotations
- Enables evaluation and comparison of different methods
- Ran challenges evaluating performance on object class recognition (from 2005-2012, now finished)

Pascal VOC data sets

Data sets from the VOC challenges are available through the challenge links below, and evalution of new methods on these data sets can be achieved through the PASCAL VOC Evaluation Server. The evaluation server will remain active even though the challenges have now finished.

http://host.robots.ox.ac.uk/pascal/VOC/

ILSVRC

“ILSVRC follows in the footsteps of the PASCAL VOC challenge… which set the precedent for standardized evaluation of recognition algorithms in the form of yearly competitions.”

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too “big” in the image
3) are basic-level categories
4) backward compatible: VOC

Class name in PASCAL VOC (20 classes)	Closest class in ILSVRC-DET (200 classes)
aeroplane	airplane
bicycle	bicycle
bird	bird
boat	*watercraft*
bottle	*wine bottle*
bus	bus
car	car
cat	domestic cat
chair	chair
cow	*cattle*
dining table	*table*
dog	dog
horse	horse
motorbike	motorcyle
person	person
potted plant	*flower pot*
sheep	sheep
sofa	sofa
train	train
tv/monitor	tv or monitor

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too “big” in the image
3) are basic-level categories
4) backward compatible: VOC

→

2. Image Collection

- Subset of images from ImageNet

- More Flickr images added

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too “big” in the image
3) are basic-level categories
4) backward compatible: VOC

→

2. Image Collection

- Subset of images from ImageNet

- More Flickr images added

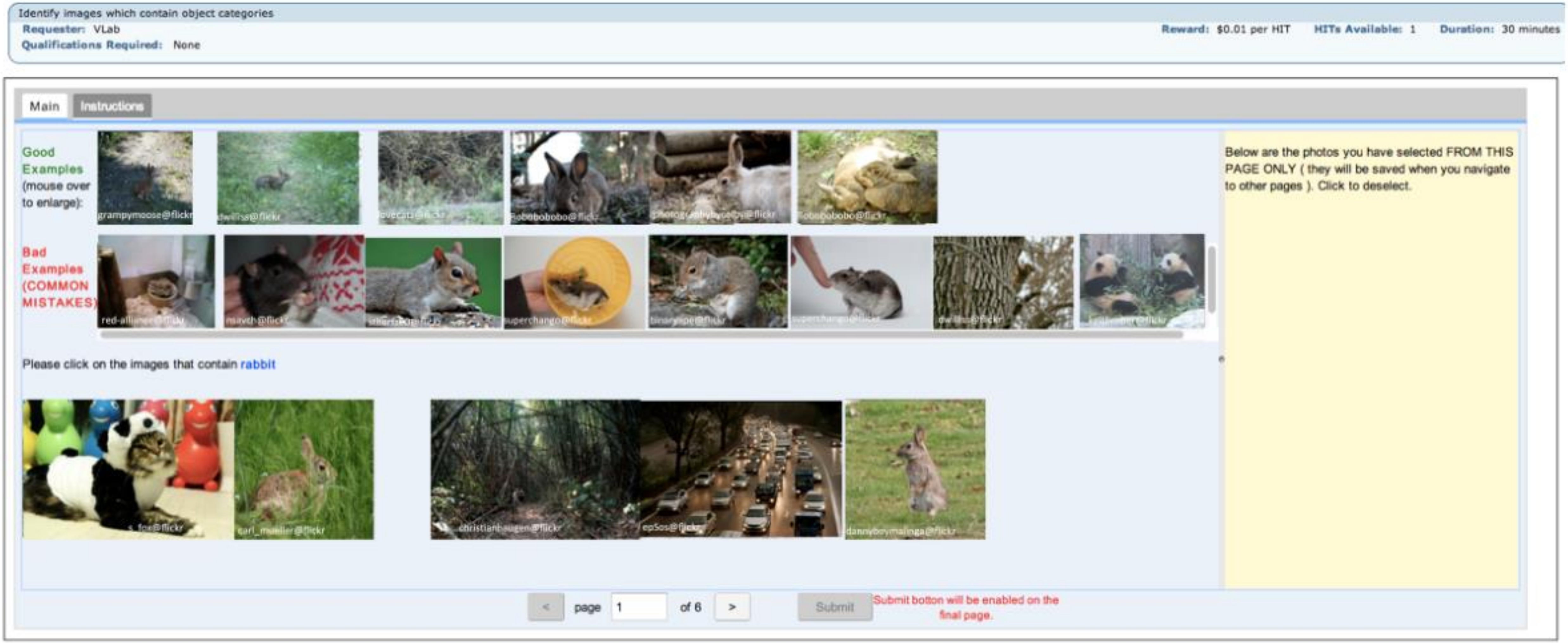
→

3. Image Verification

- Crowdsource assigning all relevant categories from 200 object categories to each image

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Recall from ImageNet: Object Presence Labeling



Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too "big" in the image
3) are basic-level categories
4) backward compatible: VOC

→

2. Image Collection

- Subset of images from ImageNet
- More Flickr images added

→

3. Image Verification

- Crowdsource assigning all relevant categories from 200 object categories to each image

→

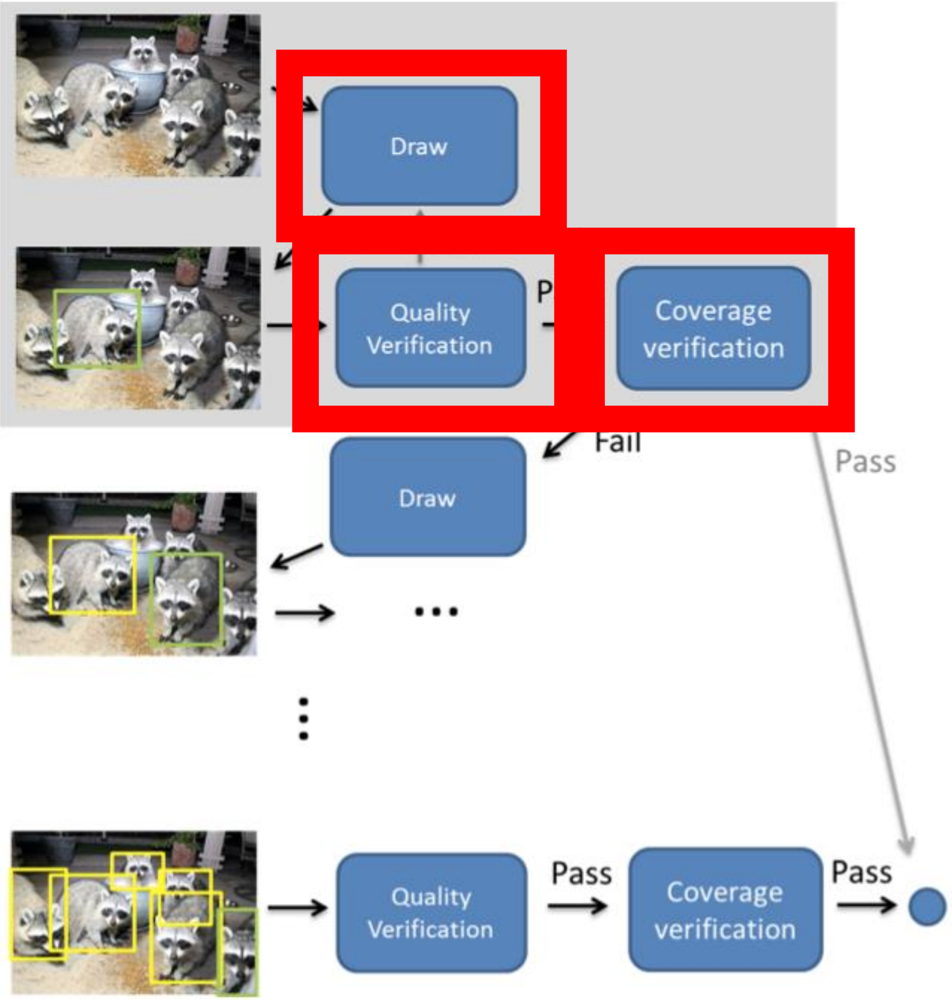
4. Image Annotation

- Crowdsource demarcating a bounding box around EVERY instance of every object category

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

ILSVRC: Efficient Object Localization

- 3 Tasks:

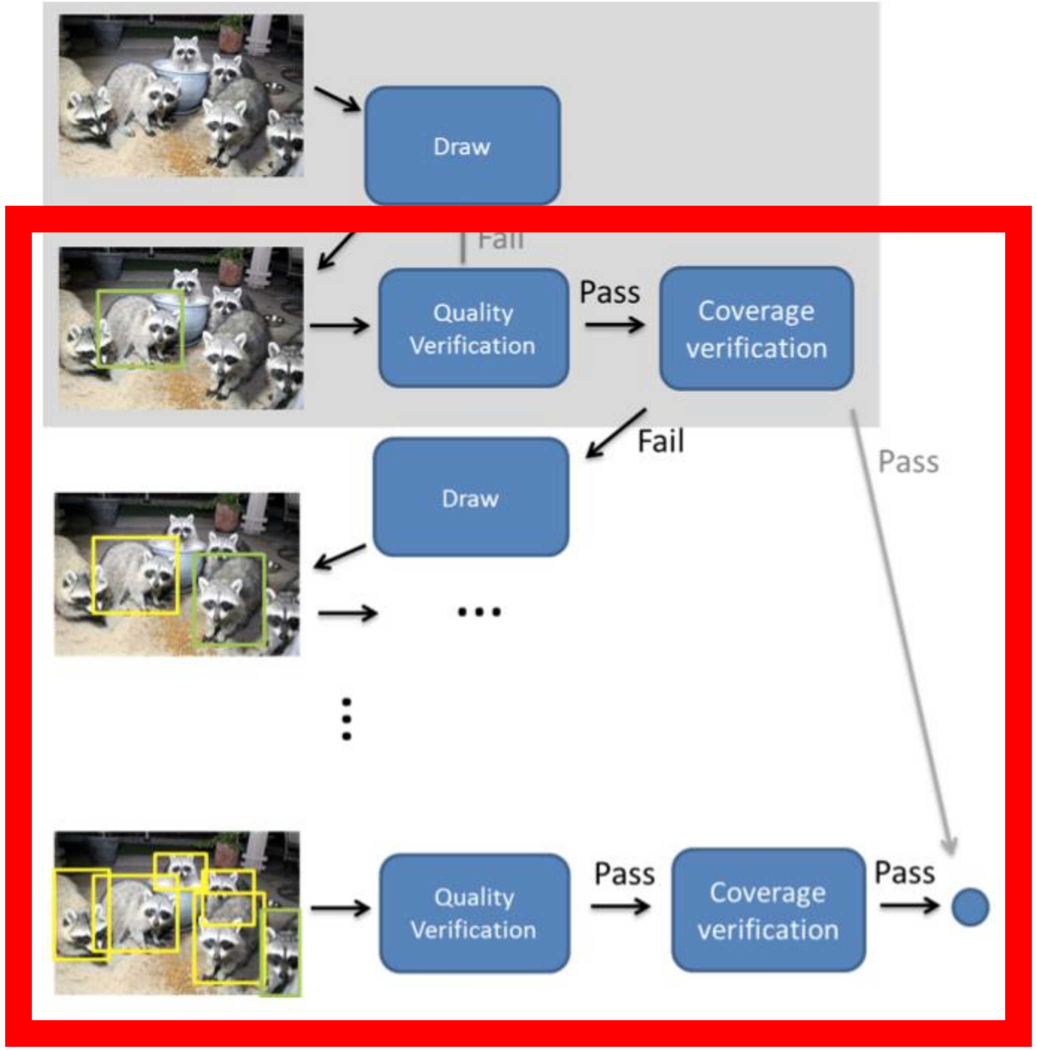


Idea: each task has fixed and predictable amount of work

Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012

ILSVRC: Efficient Object Localization

- 3 Tasks:

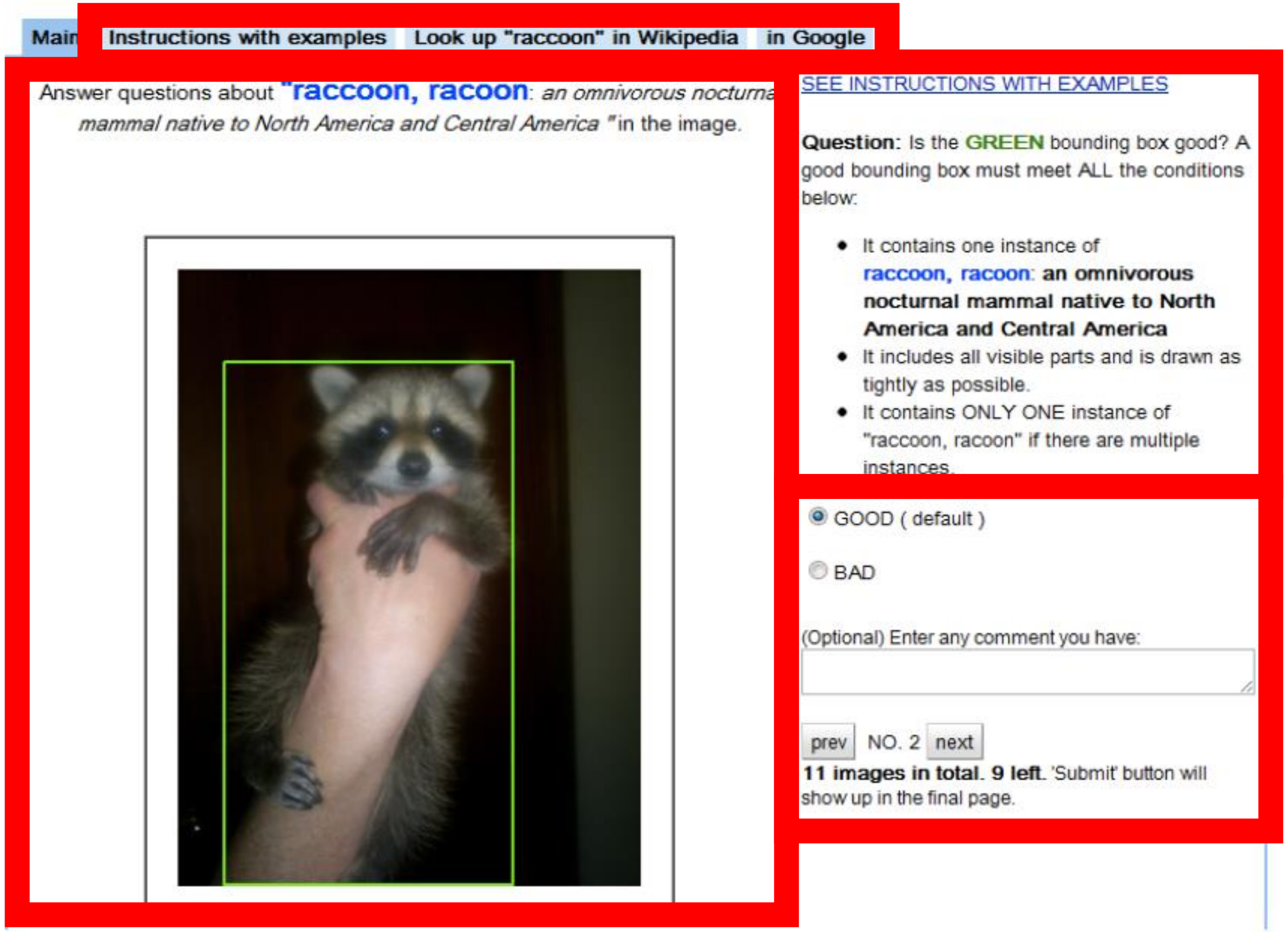


Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012

ILSVRC: Drawing **Task**

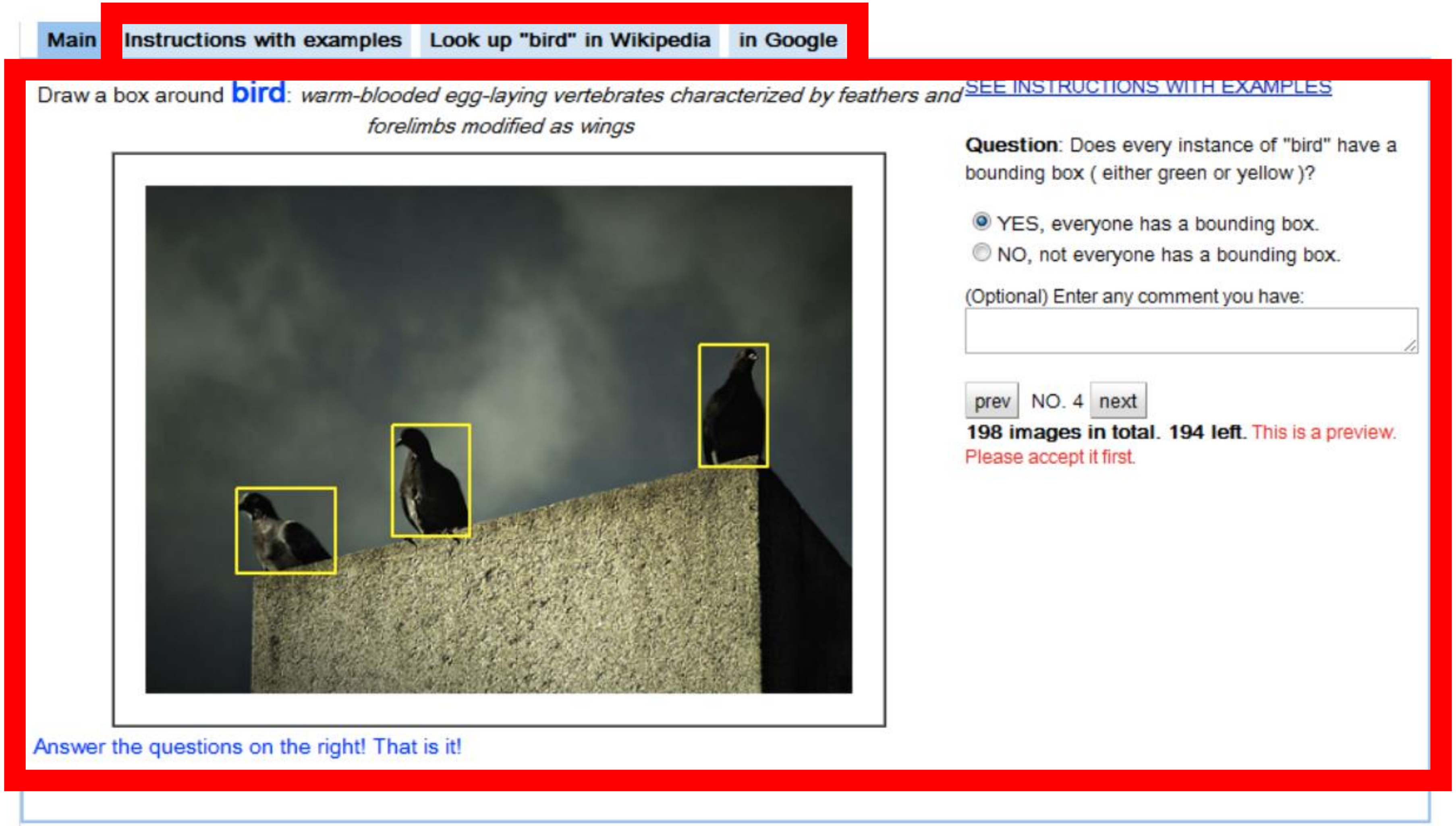
Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012

ILSVRC: Quality Verification **Task**



Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012

ILSVRC: Coverage Verification **Task**



Hao Su, Jia Deng, and Li Fei-Fei. Crowdsourcing Annotations for Visual Object Detection. AAAI 2012

ILSVRC

1. Category Selection

- 200 ImageNet classes which:
1) exclude synset overlap
2) exclude object classes too "big" in the image
3) are basic-level categories
4) backward compatible: VOC

→

2. Image Collection

- Subset of images from ImageNet
- More Flickr images added

→

3. Object presence labeling

- Crowdsource assigning all relevant categories from 200 object categories to each image

→

4. Object localization

- Crowdsource demarcating a bounding box around EVERY instance of every object category

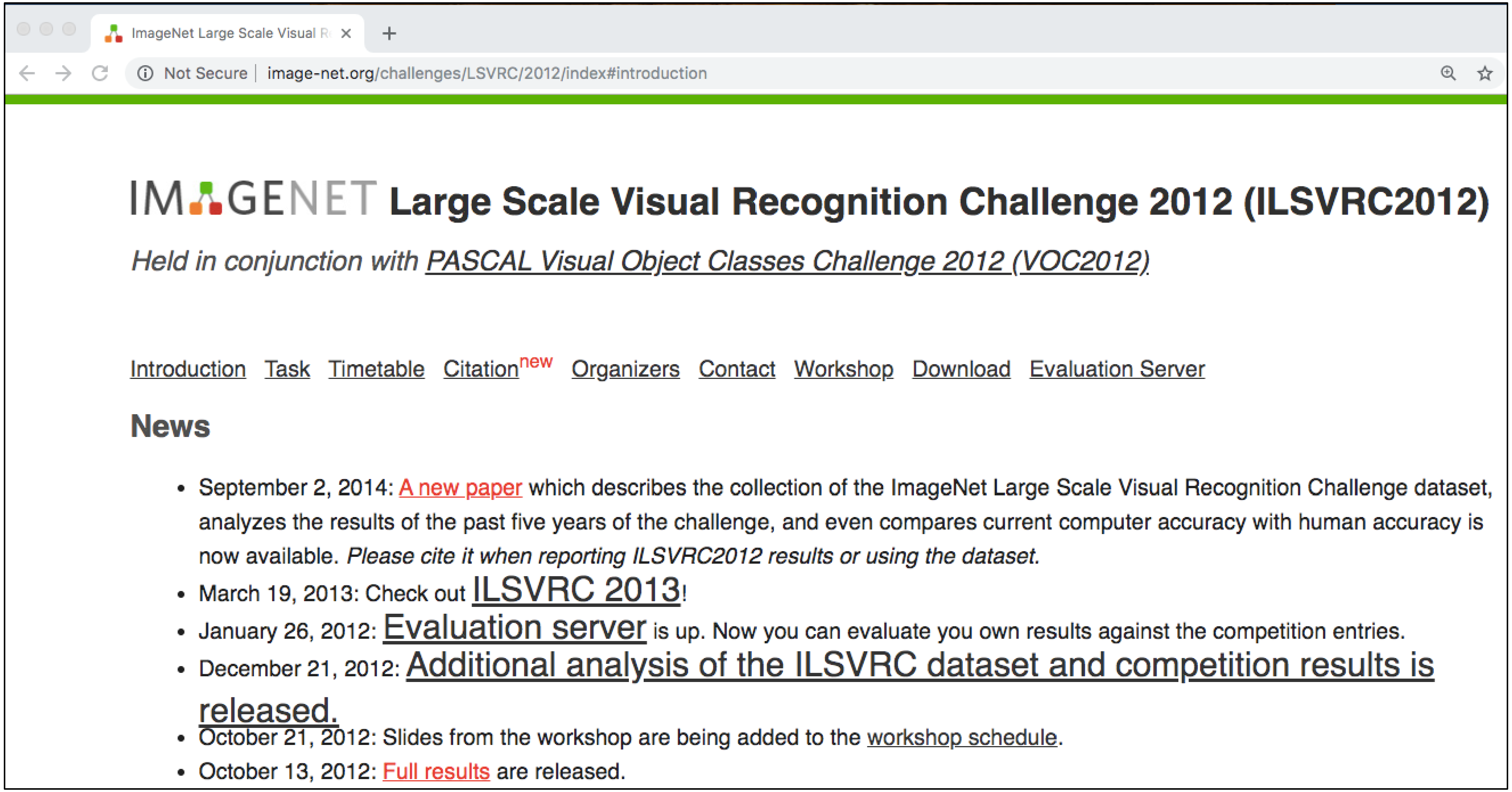
→

5. Author Review

- Ambiguous objects: BB for two categories with large overlap(~3%).
- Duplicates: >50% overlap for same object (~1%).

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei , IJCV 2015

Object Detection: ILSVRC Annual Workshop



IMAGENET Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)

Held in conjunction with PASCAL Visual Object Classes Challenge 2012 (VOC2012)

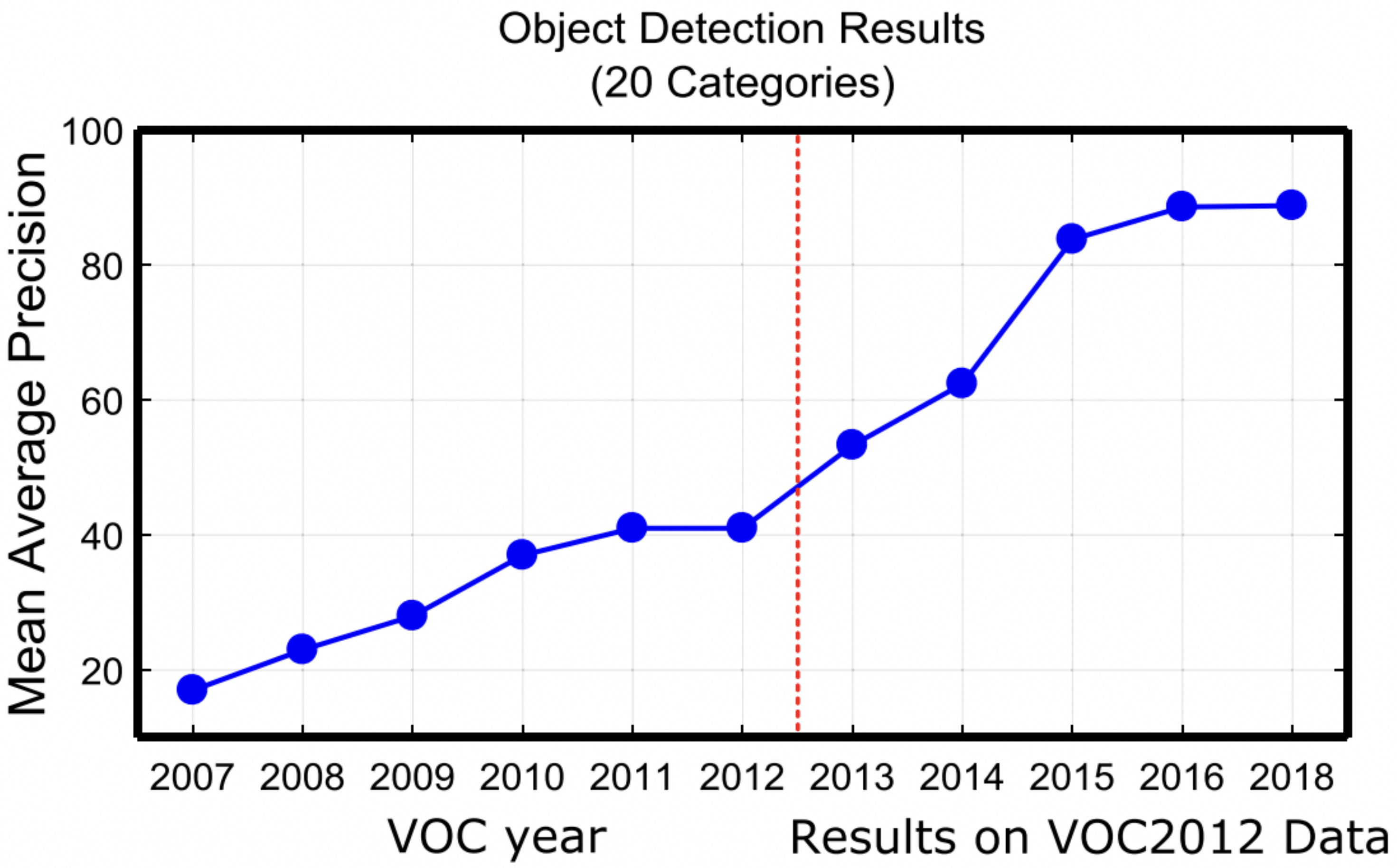
Introduction Task Timetable Citation new Organizers Contact Workshop Download Evaluation Server

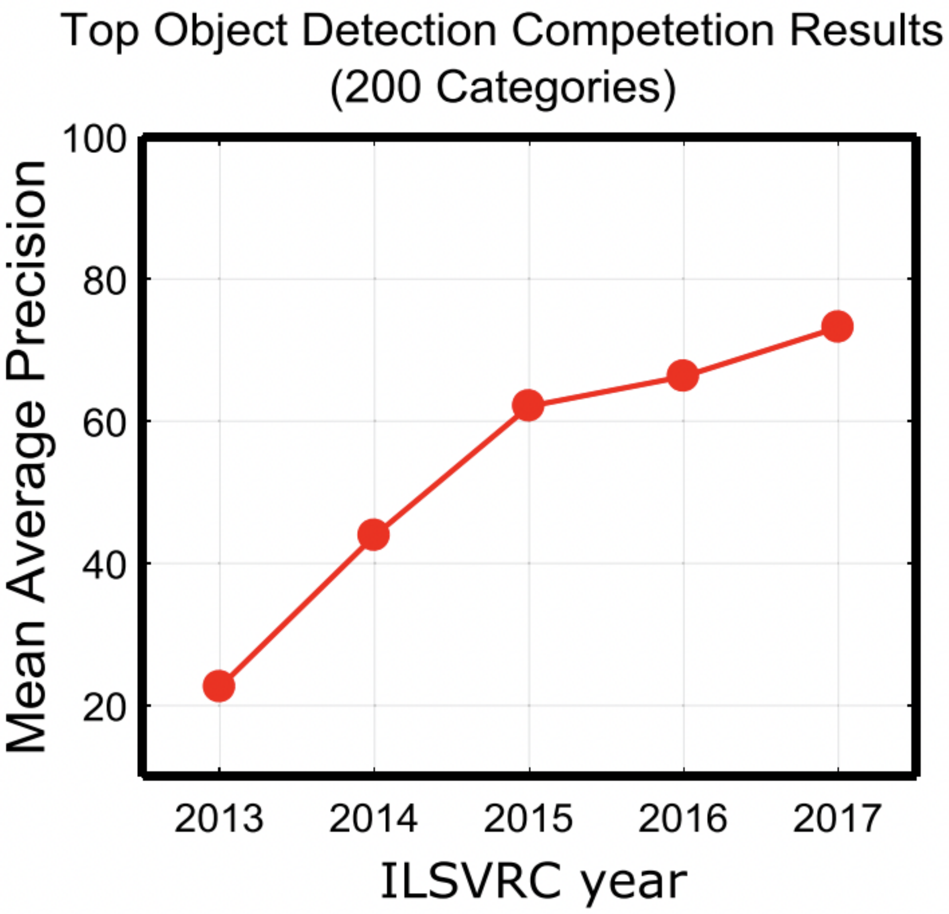
News

- September 2, 2014: A new paper which describes the collection of the ImageNet Large Scale Visual Recognition Challenge dataset, analyzes the results of the past five years of the challenge, and even compares current computer accuracy with human accuracy is now available. *Please cite it when reporting ILSVRC2012 results or using the dataset.*
- March 19, 2013: Check out ILSVRC 2013!
- January 26, 2012: Evaluation server is up. Now you can evaluate you own results against the competition entries.
- December 21, 2012: Additional analysis of the ILSVRC dataset and competition results is released.
- October 21, 2012: Slides from the workshop are being added to the workshop schedule.
- October 13, 2012: Full results are released.

http://image-net.org/challenges/LSVRC/2012/index#introduction

Turning Point: 2012 (Deep Learning Solutions)



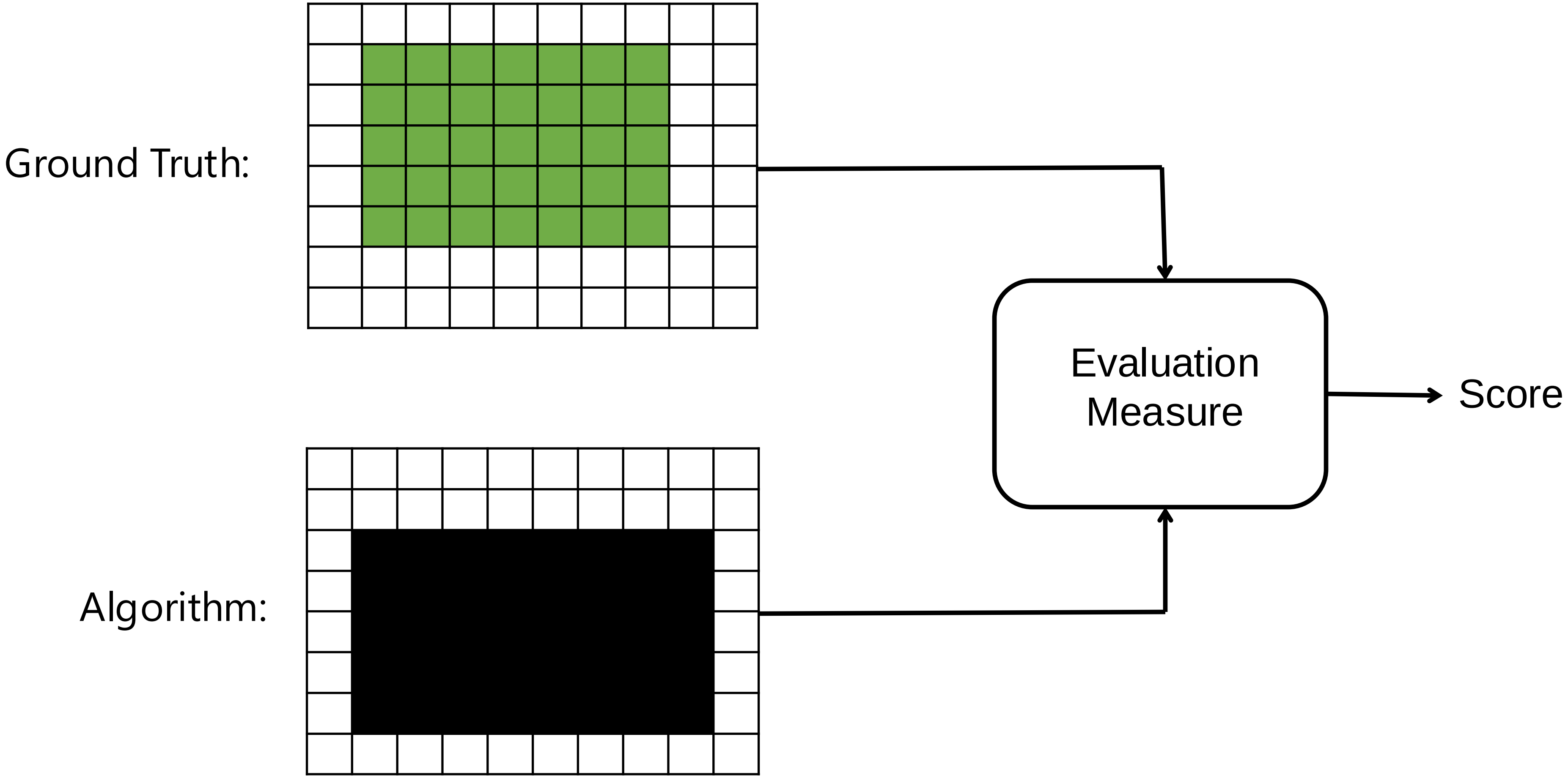


Li Liu et al. "Deep Learning for Generic Object Detection: A Survey." IJCV 2019

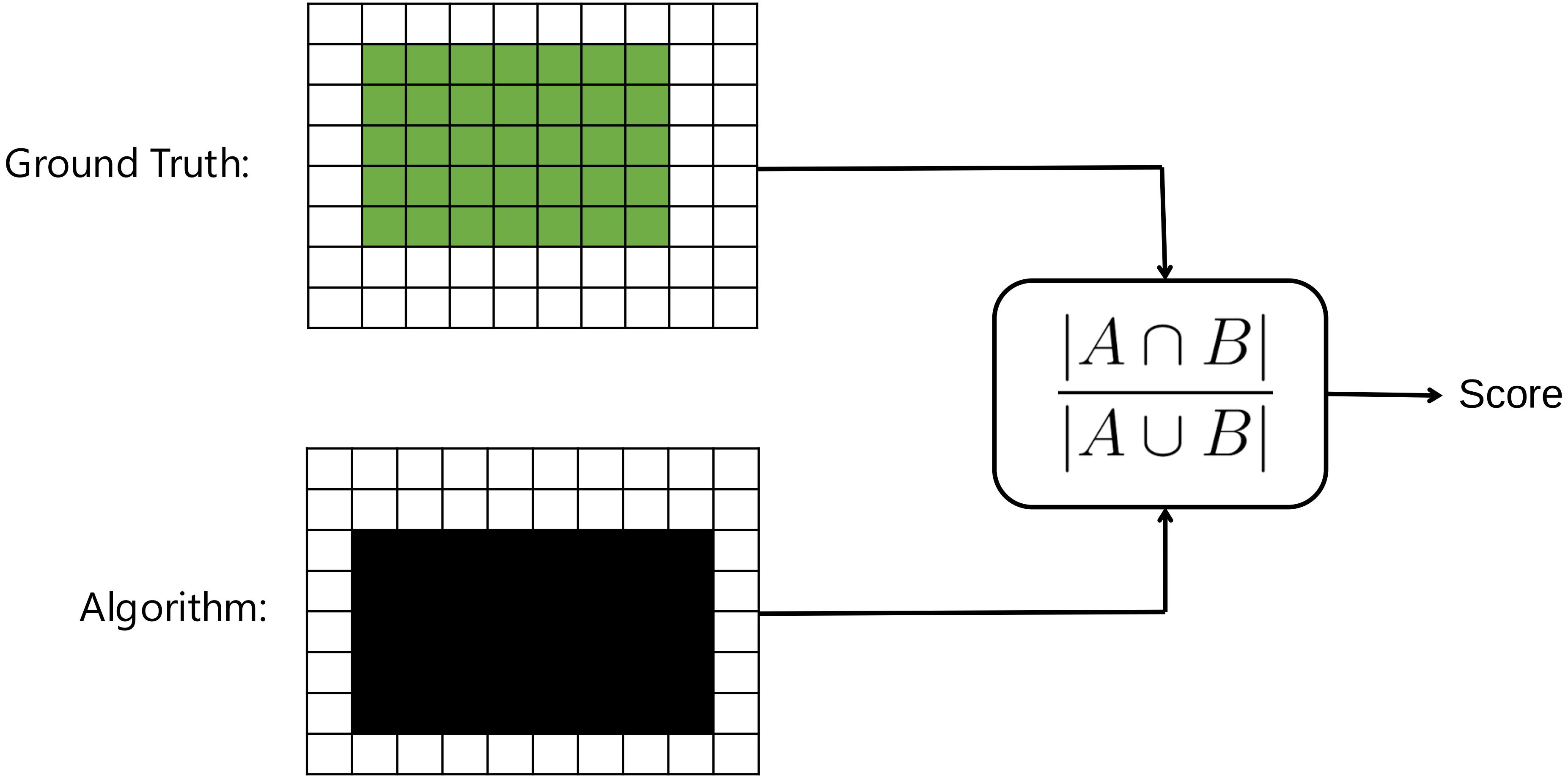
Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)

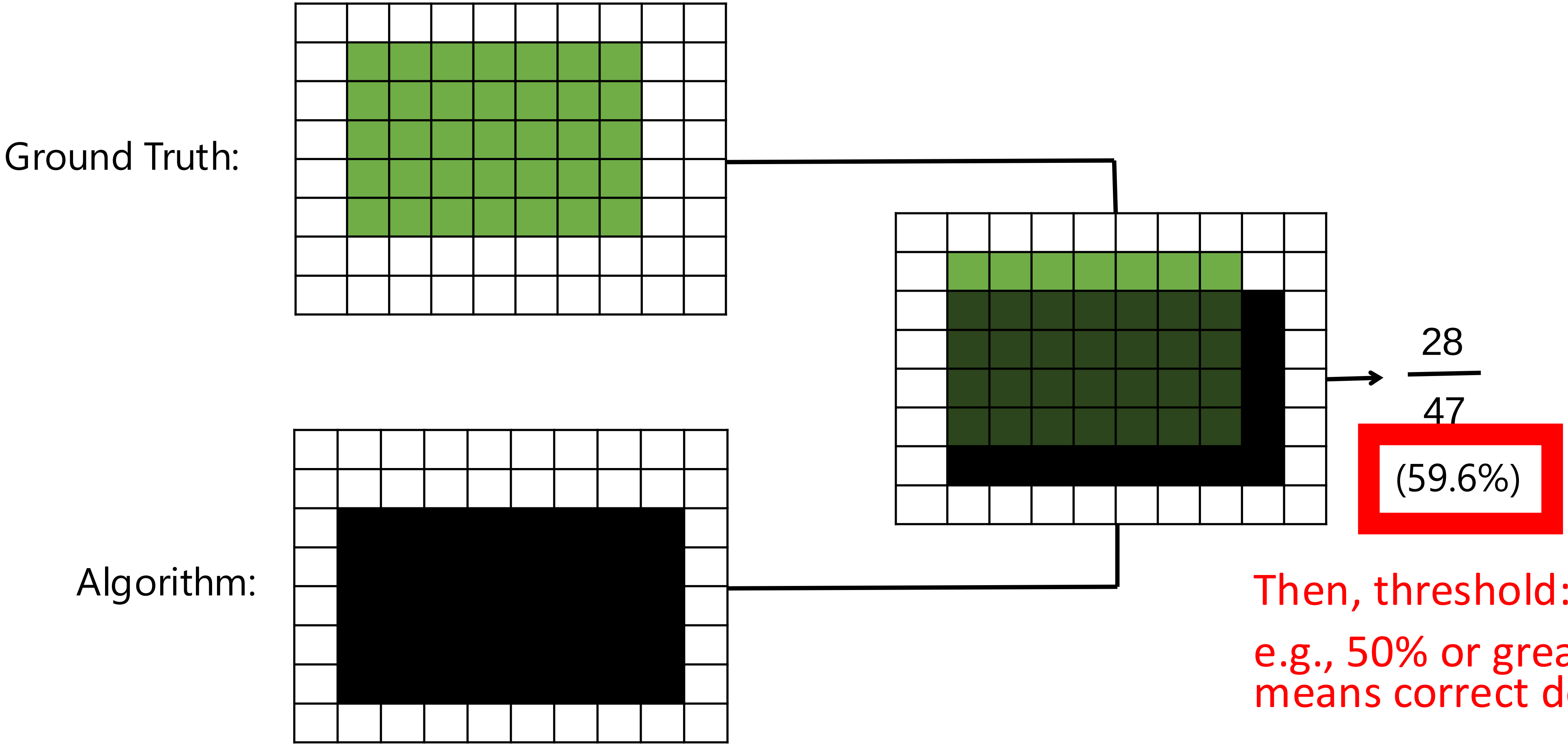
Single Object



Single Object: IoU (Intersection Over Union)



Single Object: IoU (Intersection Over Union)

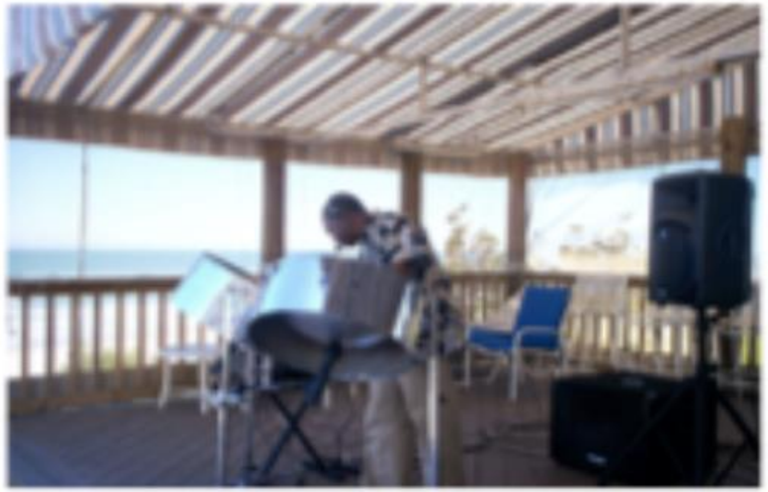


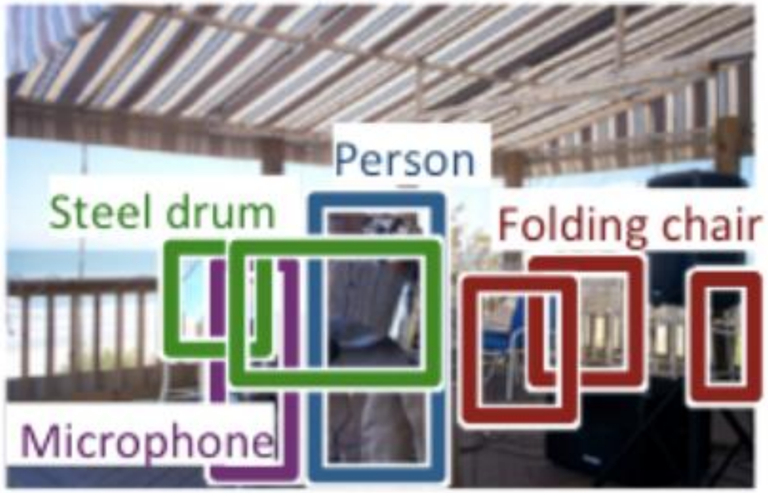
Then, threshold:

e.g., 50% or greater means correct detection!

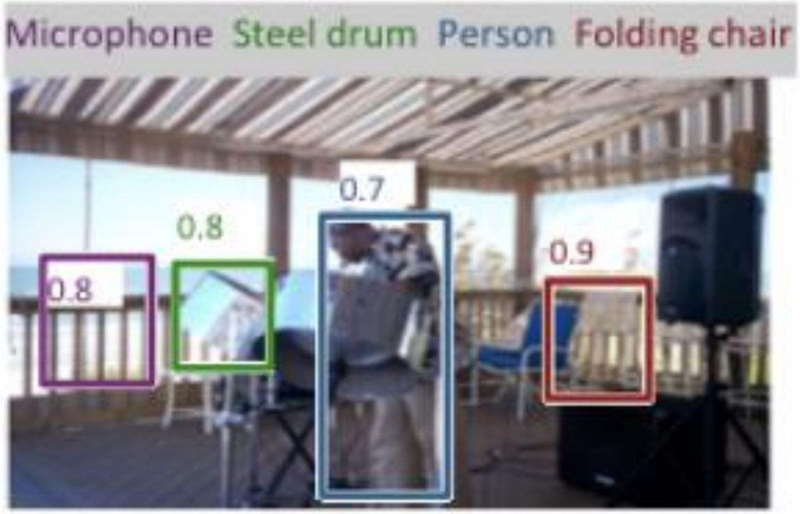
Evaluation Metric Basics: Precision and Recall

- For each object class, for detections with confidence above a confidence threshold:
  - **precision**: ability to only locate GT instances when predicting (assume 0.5 IoU threshold)





Ground truth



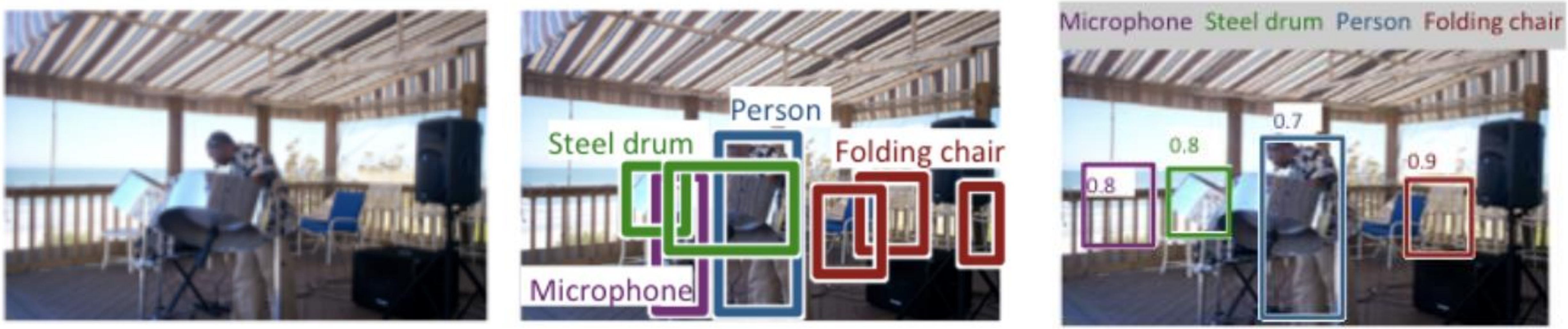
Precision: ? ? ? ?

[Russakovsky et al; IJCV 2015]

https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

Evaluation Metric Basics: Precision and Recall

- For each object class, for detections with confidence above a confidence threshold:
  - **precision**: ability to only locate GT instances when predicting (assume 0.5 IoU overlap)
  - **recall**: ability to retrieve all GT instances when predicting (assume 0.5 IoU overlap)



Ground truth

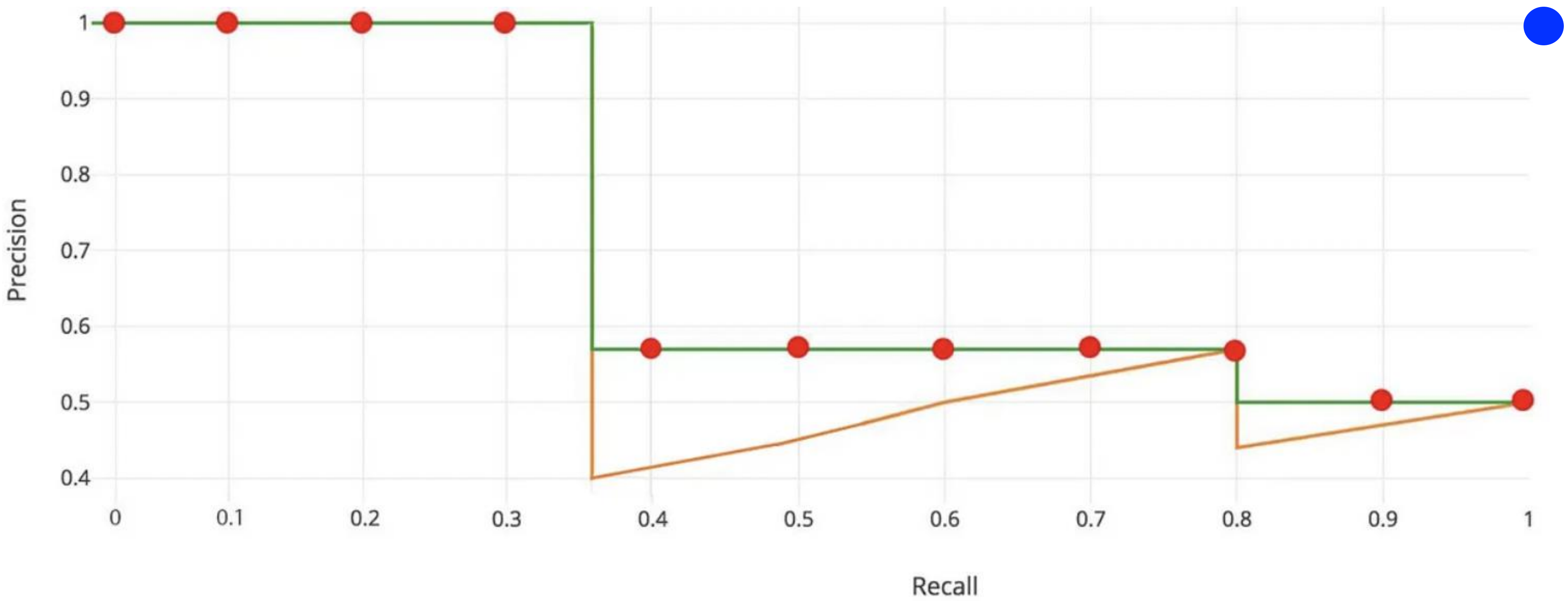
Recall: ? ? ? ?

[Russakovsky et al; IJCV 2015]

https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

Evaluation Metric Basics: Average Precision (AP)

- **Average precision (for a category)**: area under precision-recall curve created by varying the confidence level that determines a positive prediction (and using maximum precision value to right of recall value)



Plot precision-recall points with all confidence values predicted by a model for a category. **What is the optimal point for a model?**

Then, interpolate between the points and compute the area under the curve. **What is the optimal AP?**

https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

Great tutorial: Padilla et al. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. 2021

Evaluation Metric: mAP

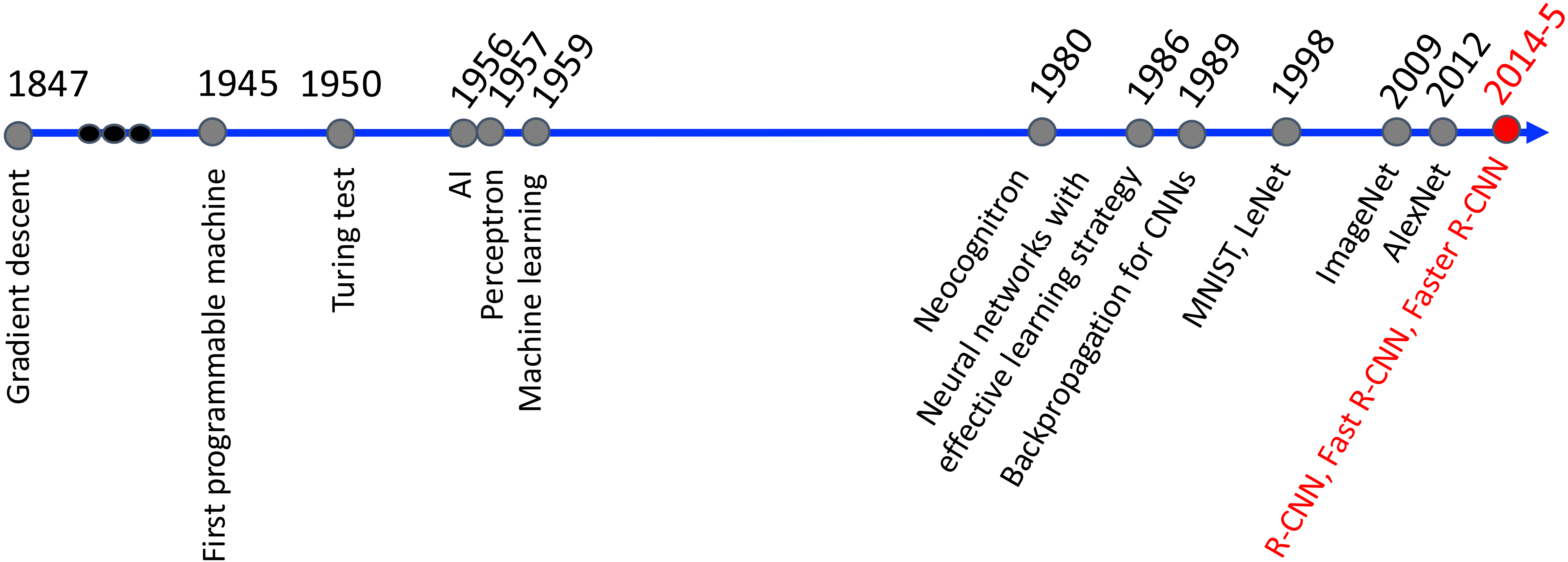
- Compute **m**ean of the **a**verage **p**recision scores for all object categories (e.g., cat, dog, …)

[Russakovsky et al; IJCV 2015]
https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)

Historical Context: R-CNN Methods



Inspiration: Sliding Window Approach

https://yourboulder.com/boulder-neighborhood-downtown/

Inspiration: Sliding Window Approach

https://yourboulder.com/boulder-neighborhood-downtown/

Inspiration: Sliding Window Approach

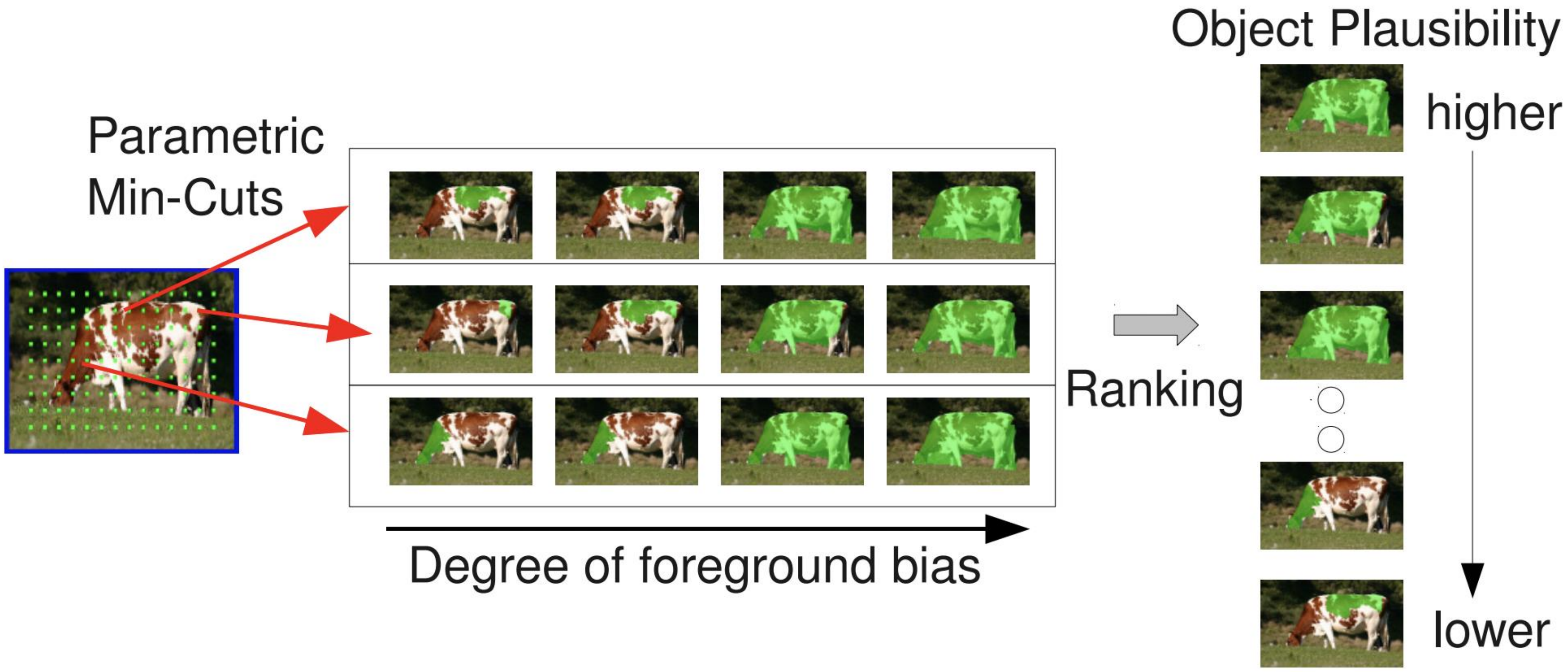
- For each object category, test different locations at…
  - Different aspect ratios (e.g., for person vs car or car viewed at different angles)
  - Different scales

- Number of regions to test? (e.g., 1920 x 1080 image)
  - Easily can explode to hundreds of thousands or millions of windows

- Key limitation
  - Very slow!

Key Idea: Use Region Proposals

Locate fewer regions than sliding windows by grouping similar pixels that are "object"-like to achieve high recall; e.g., hand-crafted methods such as CPMC and Selective Search



Carreira and Sminchisescu. Constrained Parametric Min-Cuts for Automatic Object Segmentation. CVPR 2010

Why R-CNN?

Named after the proposed technique: **R**egion proposals with **CNN** features

Idea: test a “manageable” number of image regions with diverse properties (e.g., scales, aspect ratios) if the target object type is located there very fast

Key Contributions of Faster R-CNN

1. An end-to-end trained model that learns all parts of the pipeline, including locating region proposals

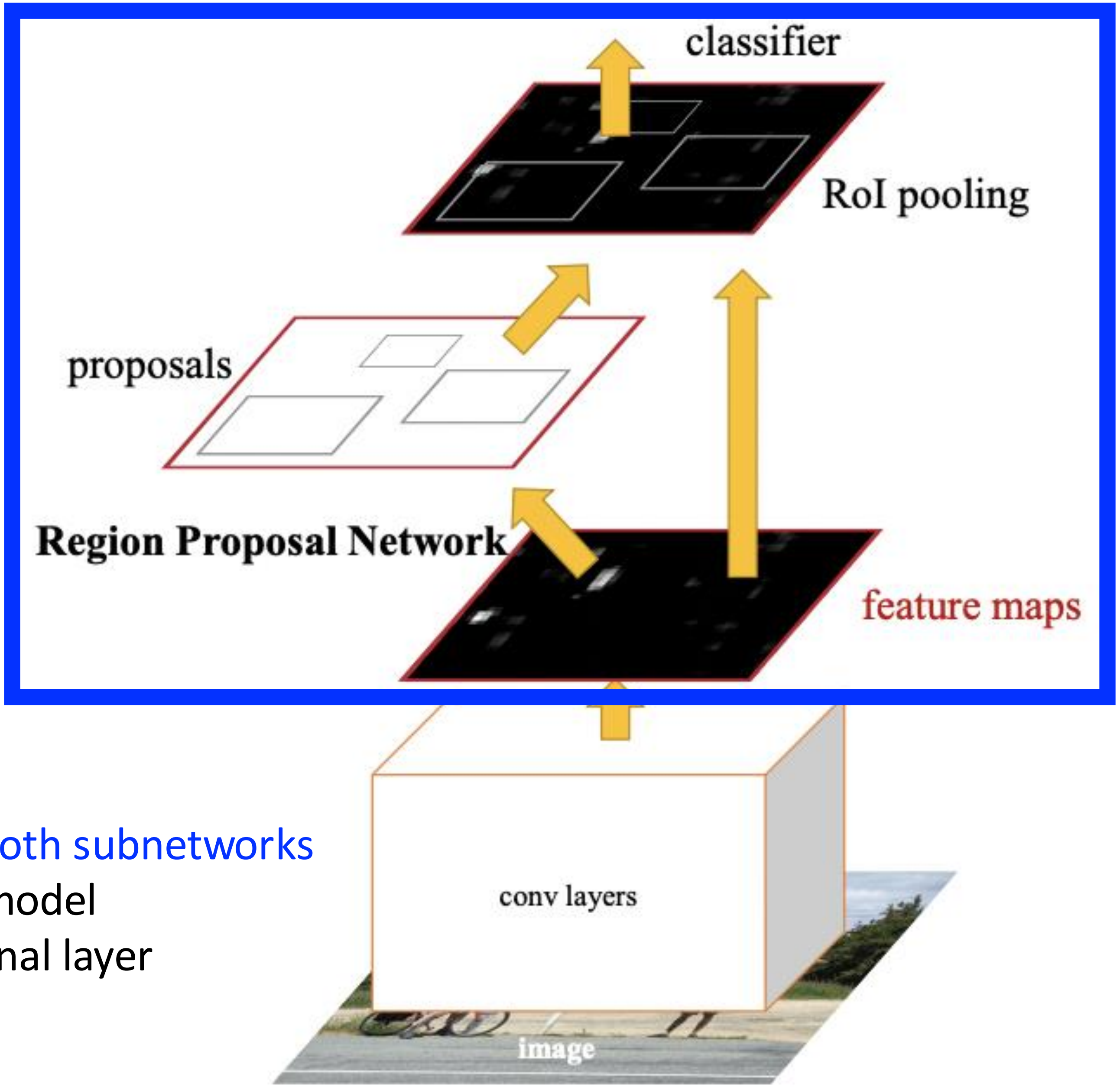
2. State of art object detection model in terms of accuracy and speed

Architecture

The single model performs two tasks:

(1) proposes image regions and then

(2) classifies category per region

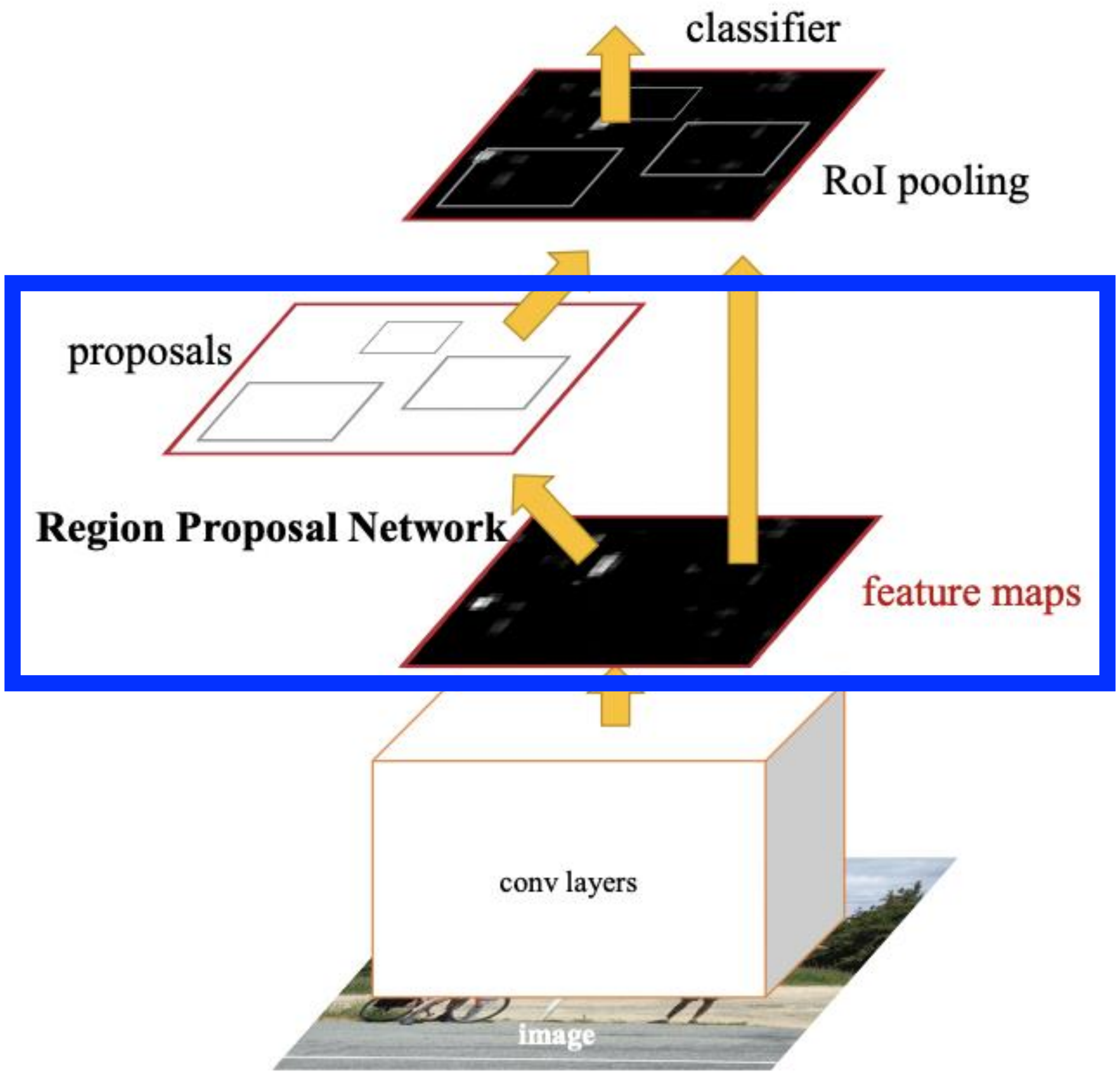


Same image representation shared for both subnetworks

- 2 architectures tested: VGG16 and ZF model
- input to subnetworks is last convolutional layer

Ren Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” Neurips 2015

Architecture

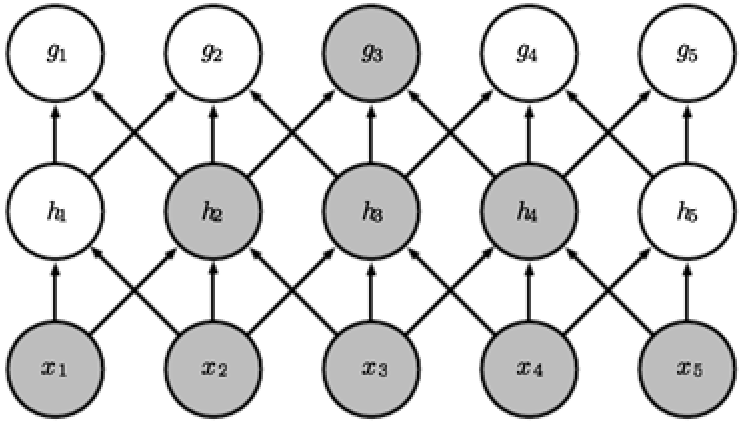


Ren Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” Neurips 2015

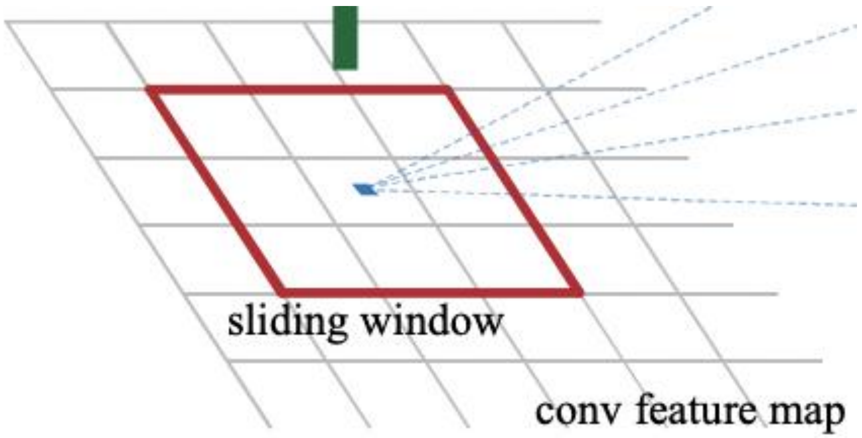
Architecture: Region Proposal Network

Input: convolutional feature map from pretrained model

Step 1: 3 x 3 convolutional filter applied to identify candidate proposals (recall, filter in the middle of an architecture maps to a larger input space, aka receptive field)



https://www.deeplearningbook.org/contents/convnets.html

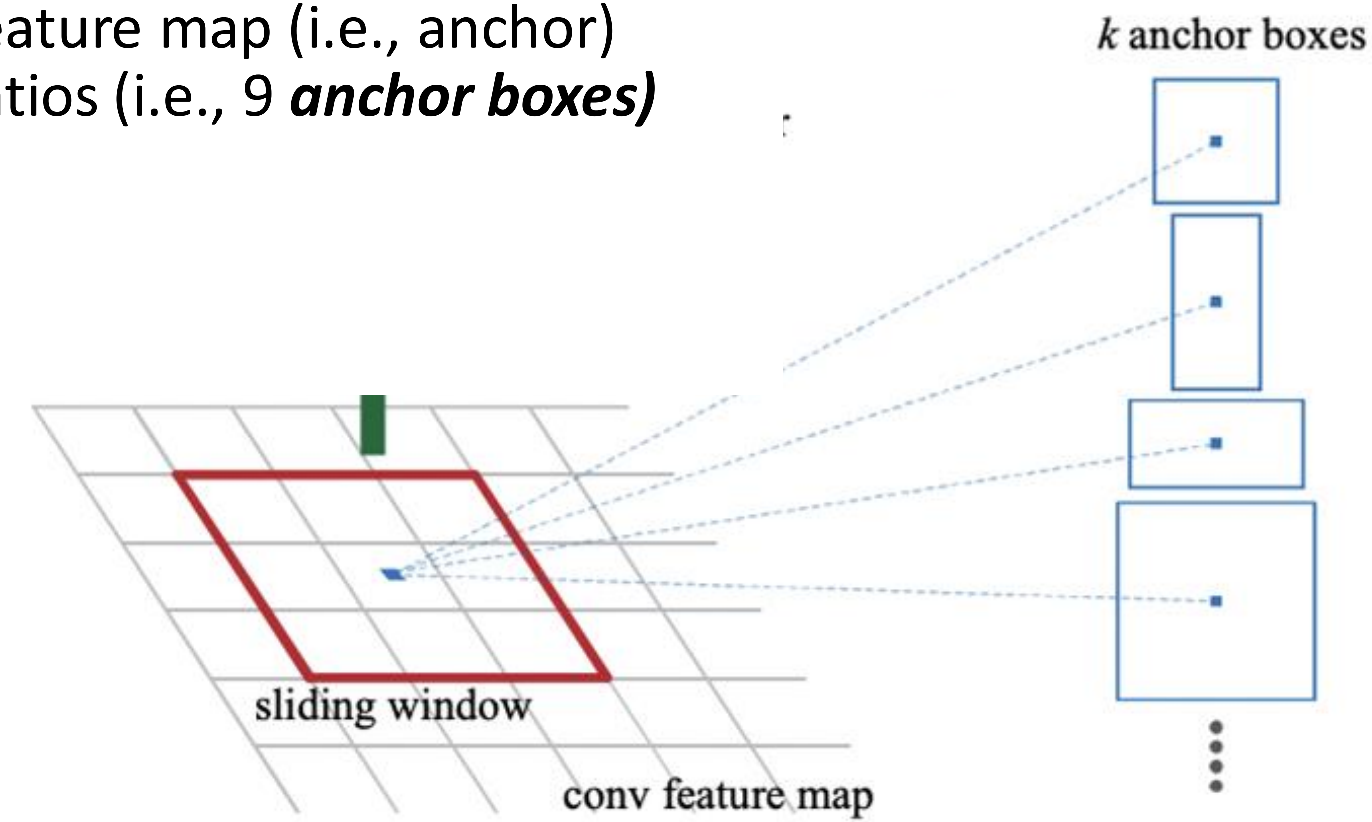


Ren Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Neurips 2015

Architecture: Region Proposal Network

Step 2: multiple scales are efficiently supported by generating for each point on the feature map (i.e., anchor) boxes with 3 scales and 3 aspect ratios (i.e., 9 ***anchor boxes)***

Each anchor box specializes in a particular shape and size (centered on each pixel)



Ren Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Neurips 2015

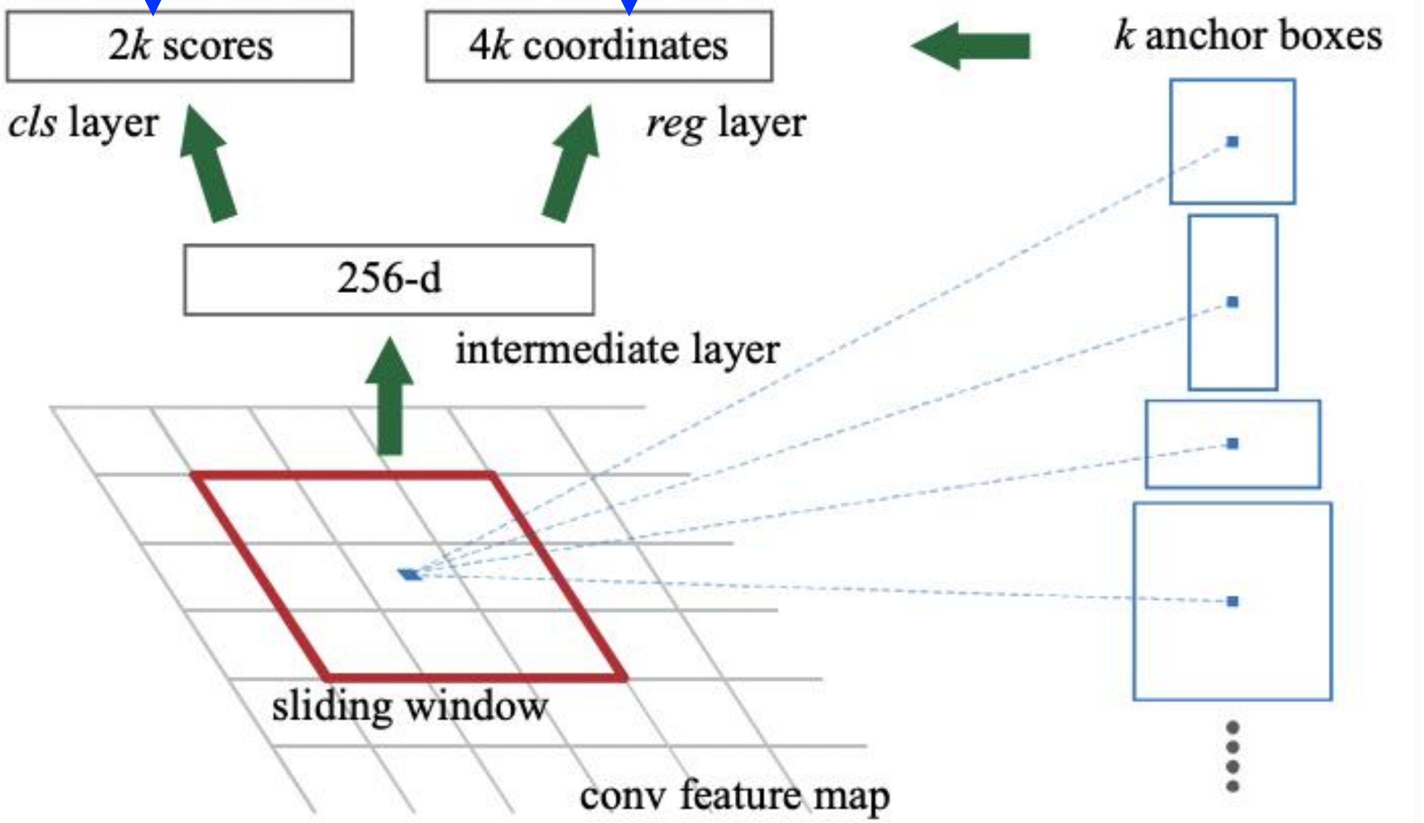
Architecture: Region Proposal Network

(*k* independent regressors learned to support *k* anchor box dimensions)

Step 3: For each region, then predict:

(1) Probability of object/not object

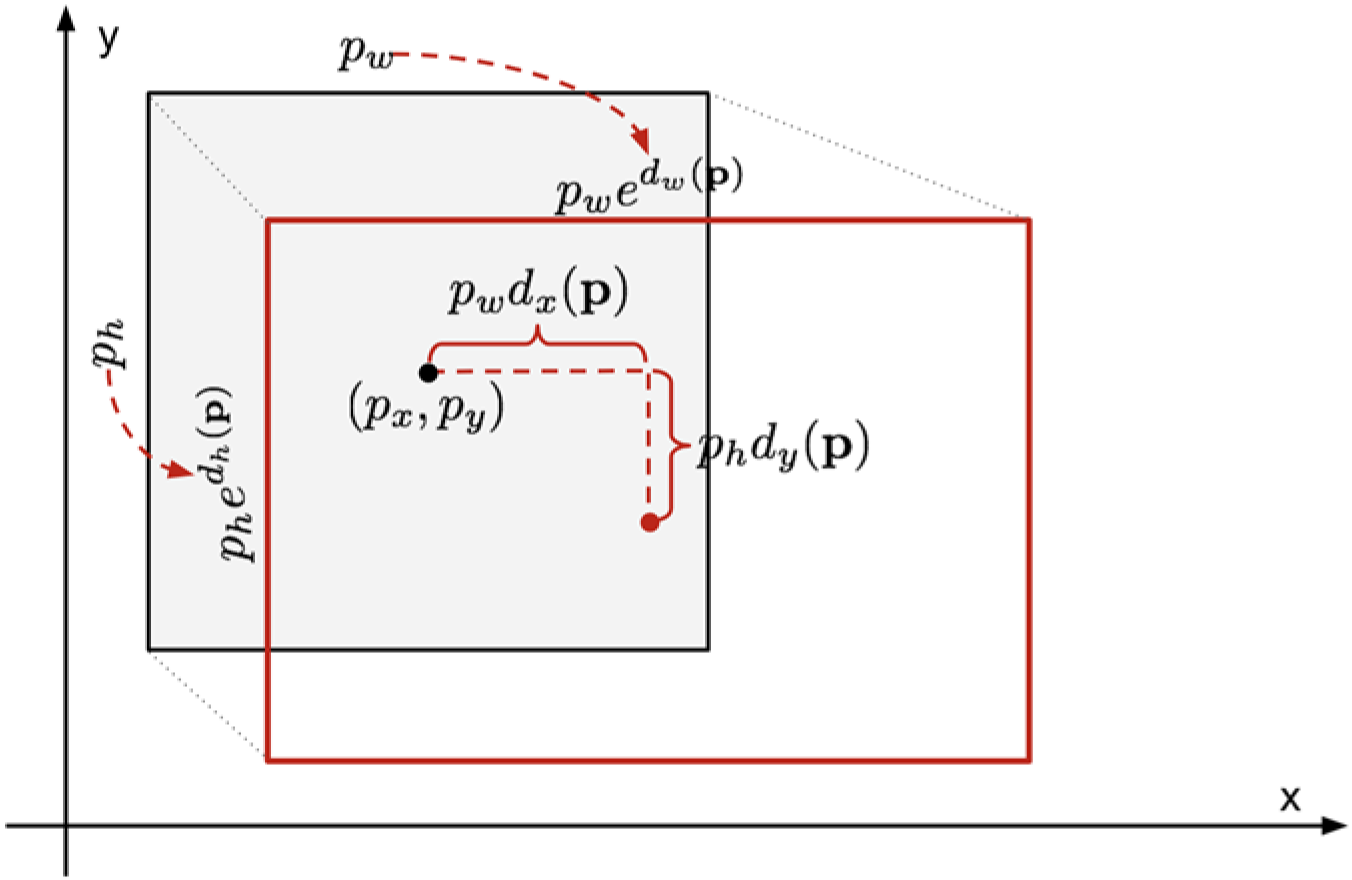
(2) Parameters to regress anchor box to GT box (center, width, and height)



Ren Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Neurips 2015

Architecture: Region Proposal Refinement

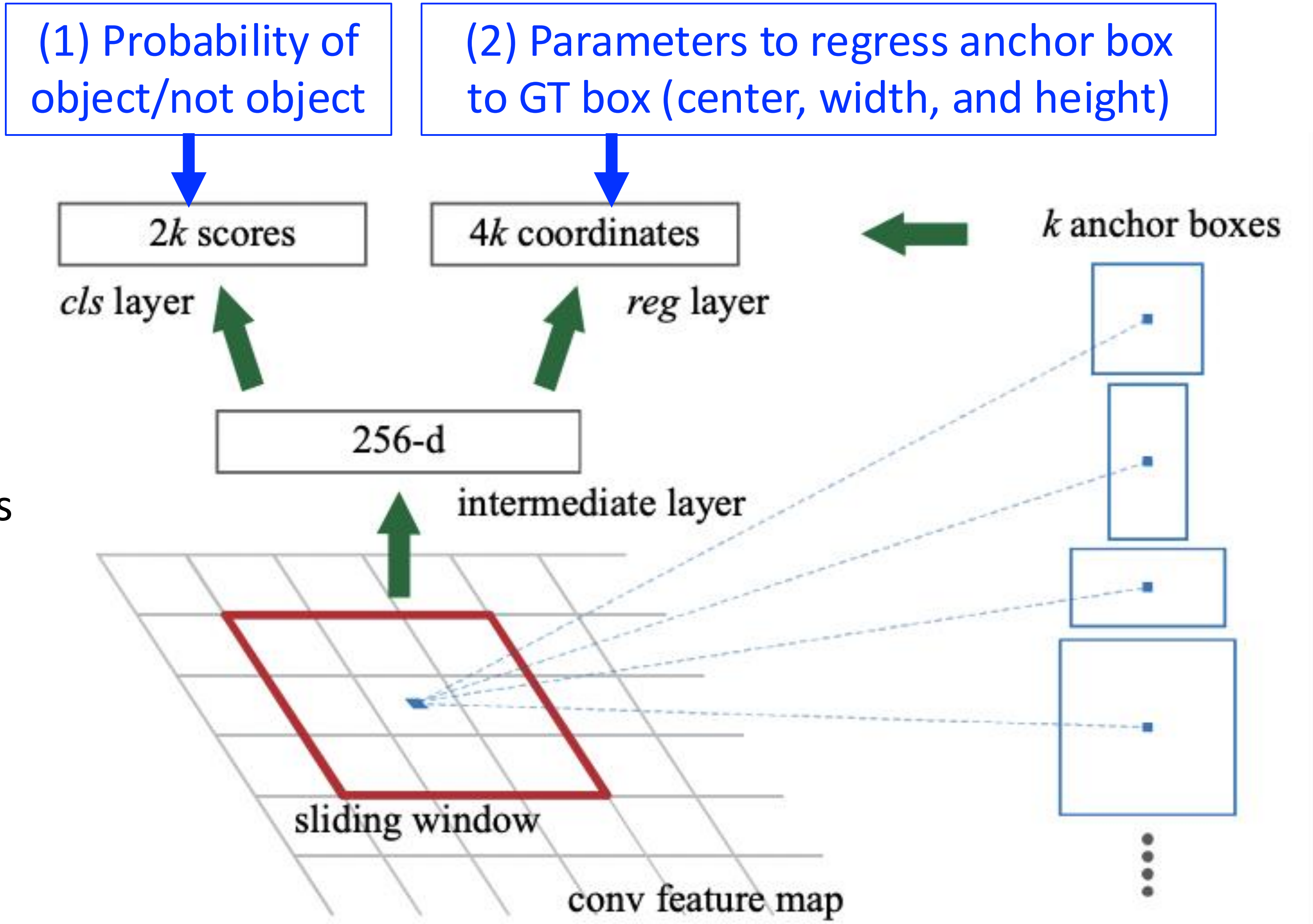
Parameters to regress original region proposal with center (p_x, p_y), width (p_w), and height (p_h) to the ground truth location: d_x, d_y, d_w, d_h



https://lilianweng.github.io/lil-log/2017/12/31/object-recognition-for-dummies-part-3.html#bounding-box-regression

Training: Region Proposal Multi-task Loss

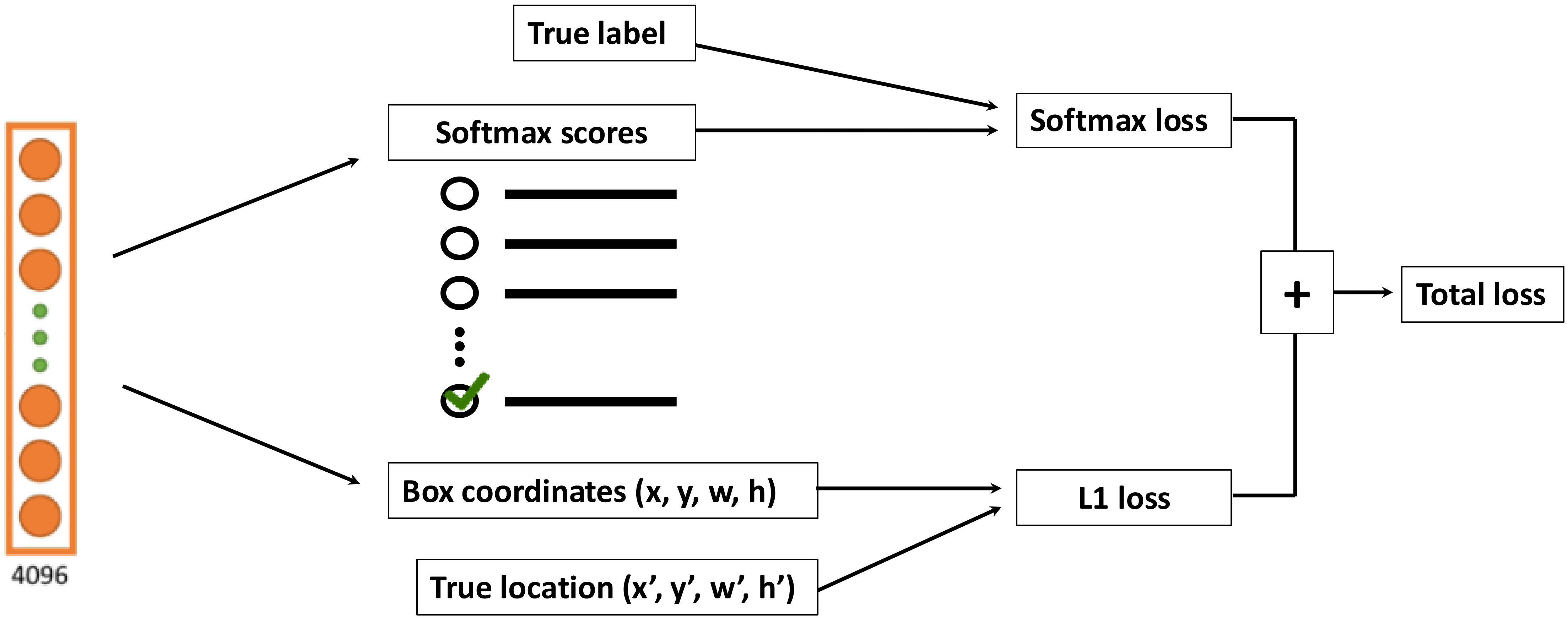
- Sample of positive anchor boxes ("objects"): anchors with IoU > 0.7 with GT (can be multiple anchors) or, when none, highest scoring one
- Sample of negative anchor boxes ("background"): non-positive anchors with IoU < 0.3 with GT
- Non-assigned anchors are ignored
- Multi-task loss: for each region proposal, use classification and, when relevant, localization losses



Ren Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Neurips 2015

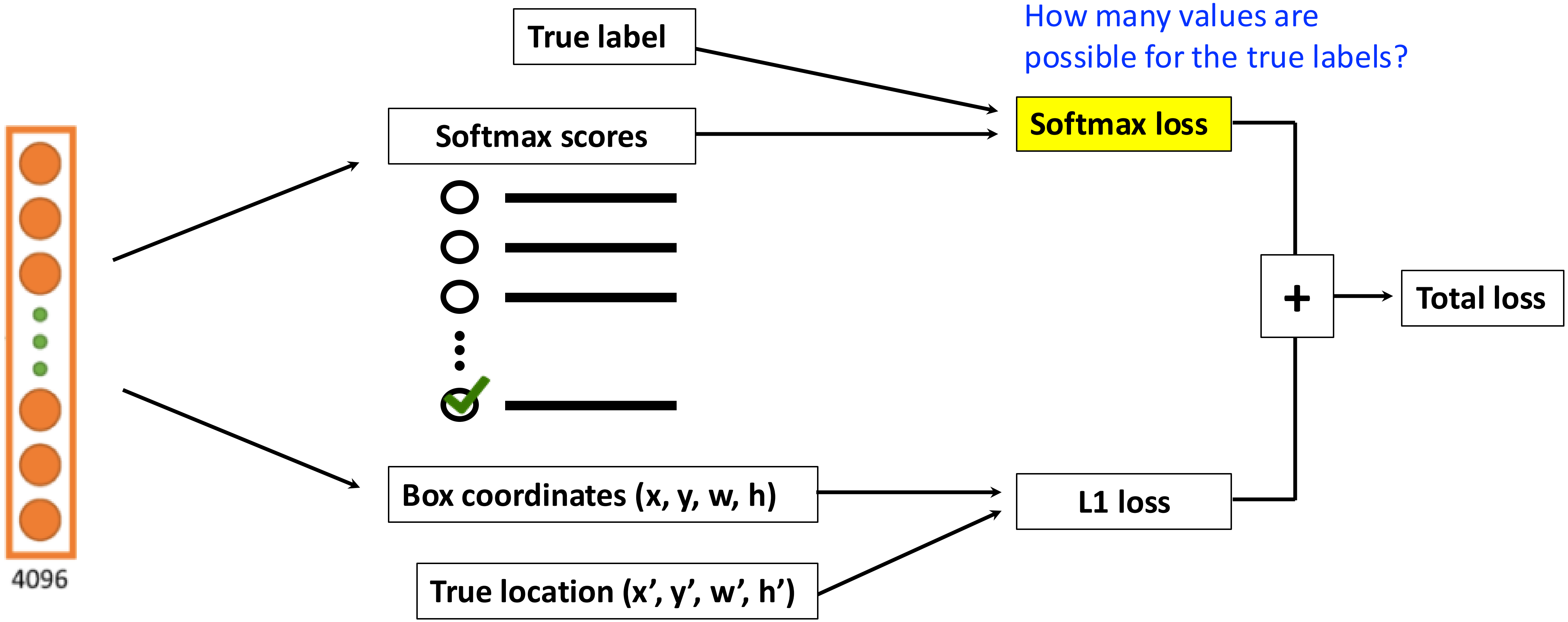
Training: Region Proposal Multi-task Loss

Sum classification and (sometimes) localization losses for each region proposal



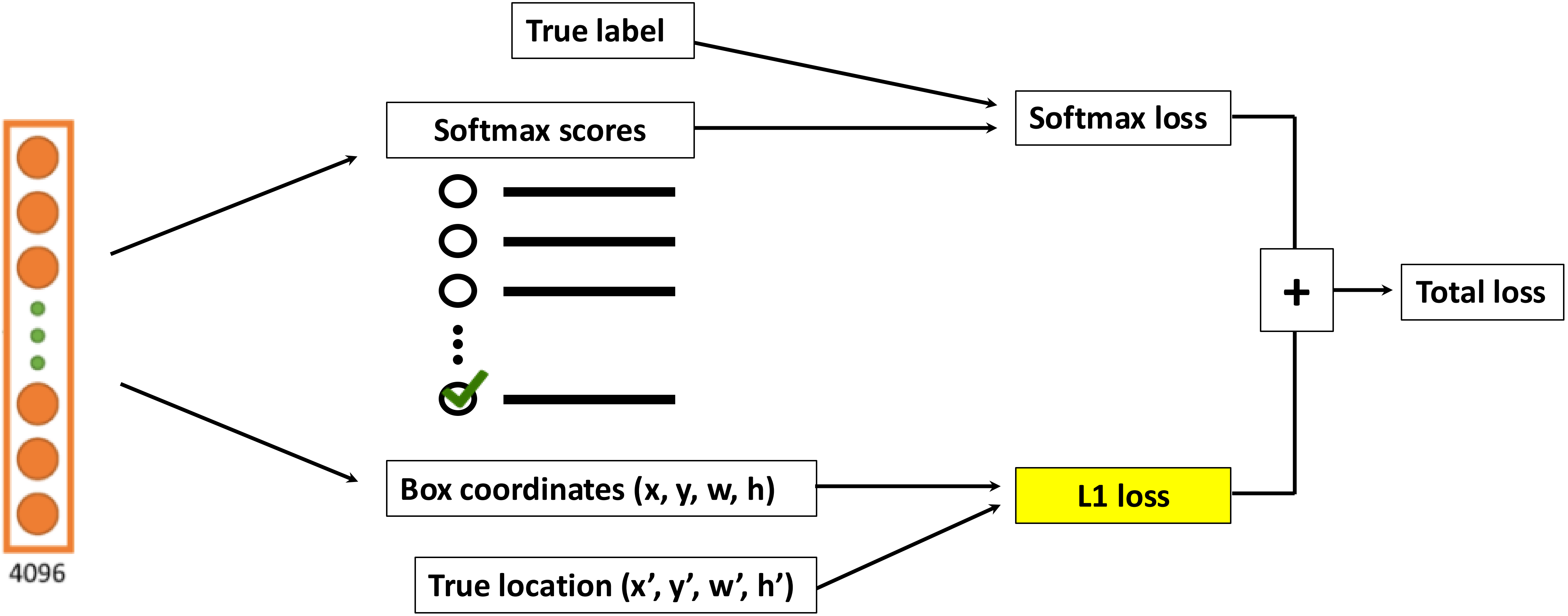
Training: Region Proposal Multi-task Loss

Sum classification and (sometimes) localization losses for each region proposal



Training: Region Proposal Multi-task Loss

Sum classification and (sometimes) localization losses for each region proposal



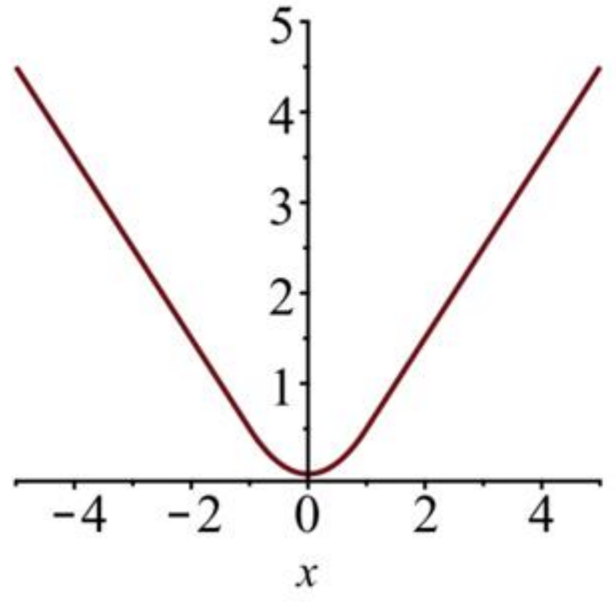
Training: Region Proposal Multi-task Loss

$$\mathcal{L}_{\text{box}}(t^u, v) = \sum_{i \in \{x,y,w,h\}} L_1^{\text{smooth}}(t_i^u - v_i) \rightarrow L_1^{\text{smooth}}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

True location for true class “u”

Predicted location for class u

Less sensitive to outliers than SSE



What is relevance of the **regression loss** when no object is present (i.e., GT negative)?

- none; regression loss disabled in such cases

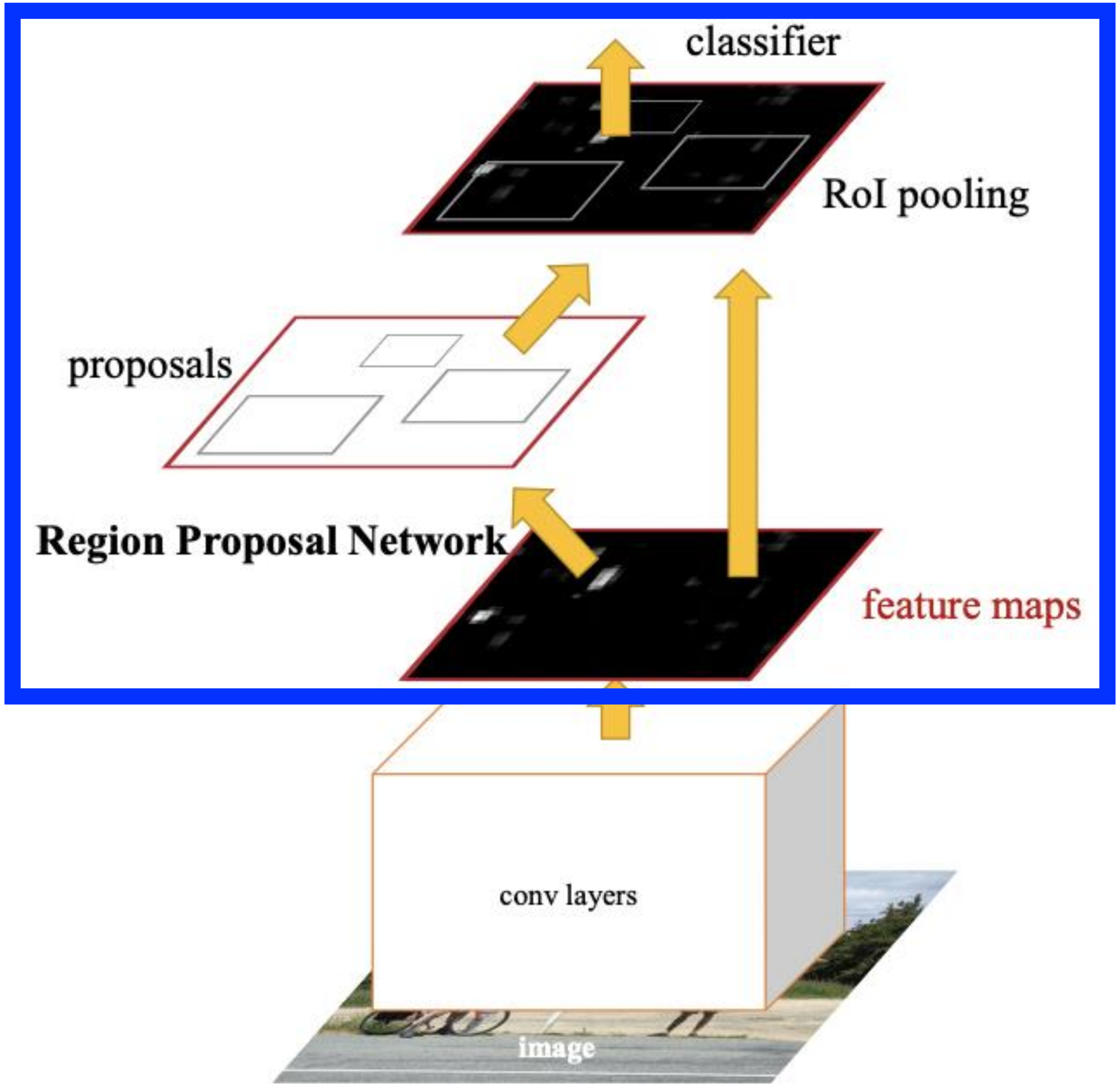
https://lilianweng.github.io/lil-log/2017/12/31/object-recognition-for-dummies-part-3.html

Architecture

The single model performs two tasks:

(1) proposes image regions and then

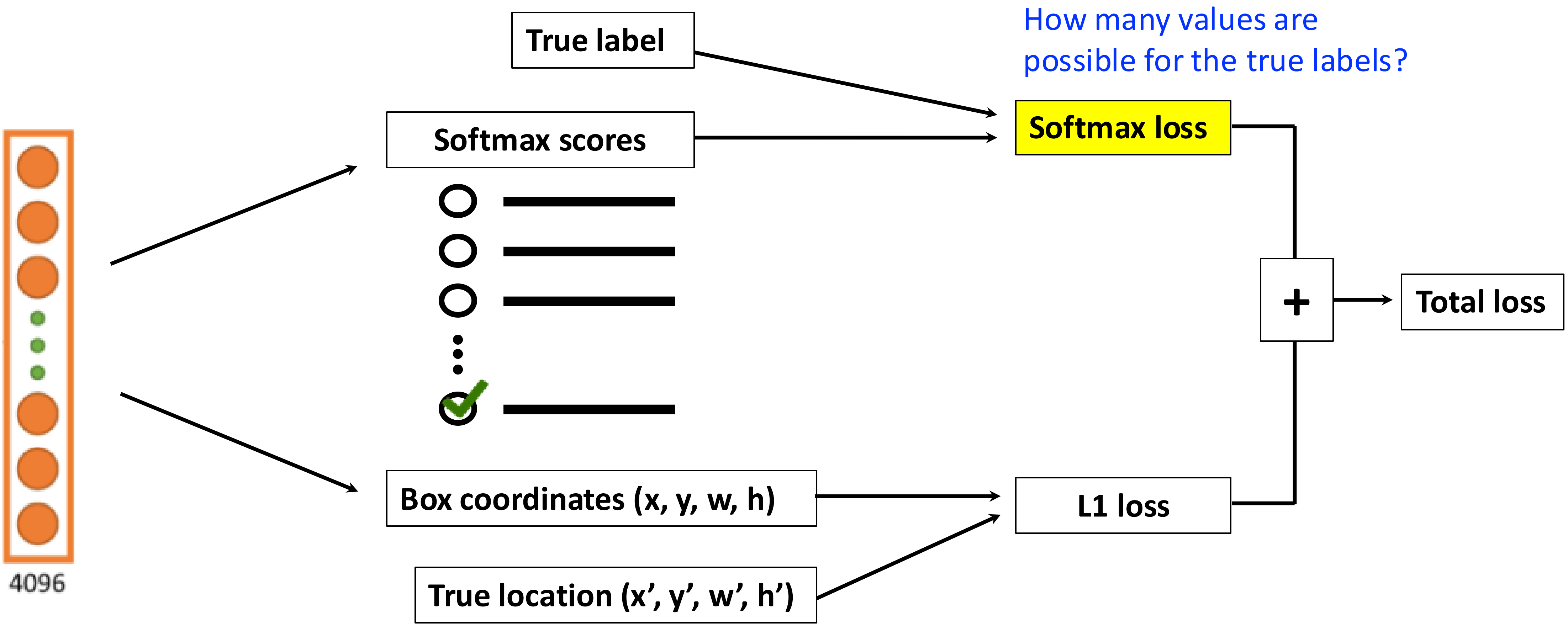
(2) classifies category per region, with architecture of the region proposal networks to predict an object's category and coordinates



Ren Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Neurips 2015

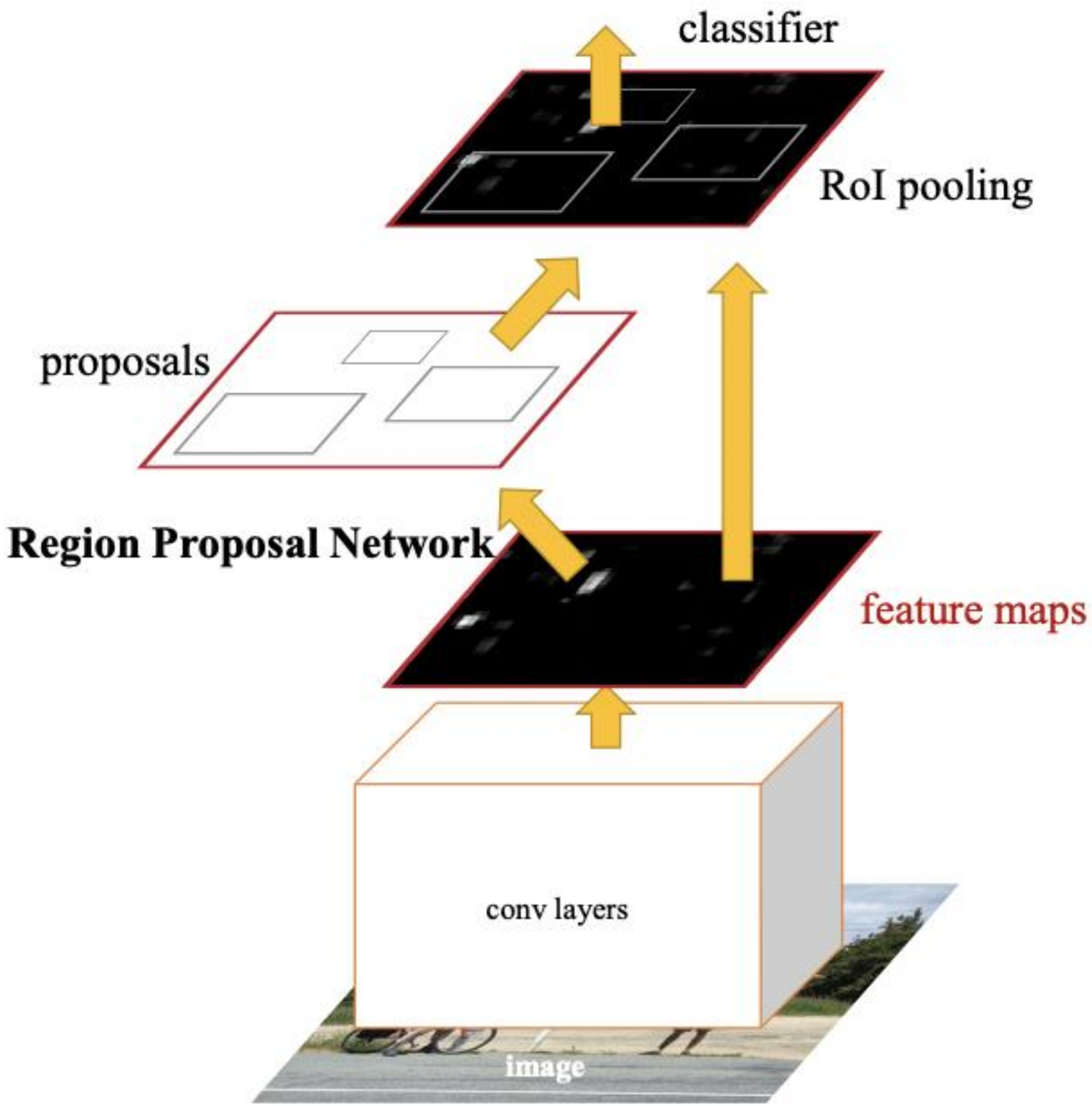
Training: Region Classifier Multi-task Loss

Sums classification and localization losses for each region proposal



Training

1. Train RPN
2. Train Faster R-CNN using proposals from pretrained RPN
3. Fine-tune layers unique to RPN
4. Fine-tune the fully connected layers of Faster R-CNN

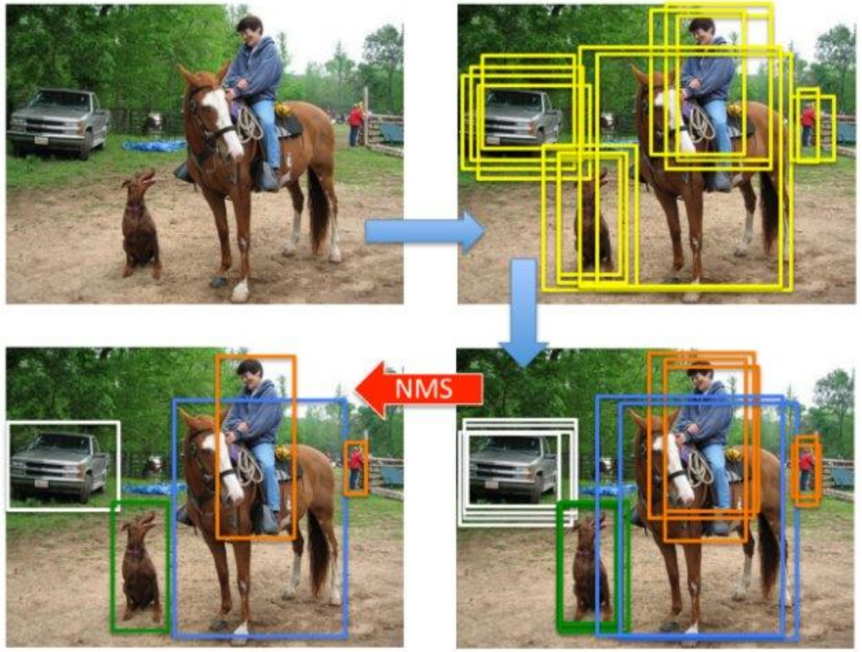


Ren Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” Neurips 2015

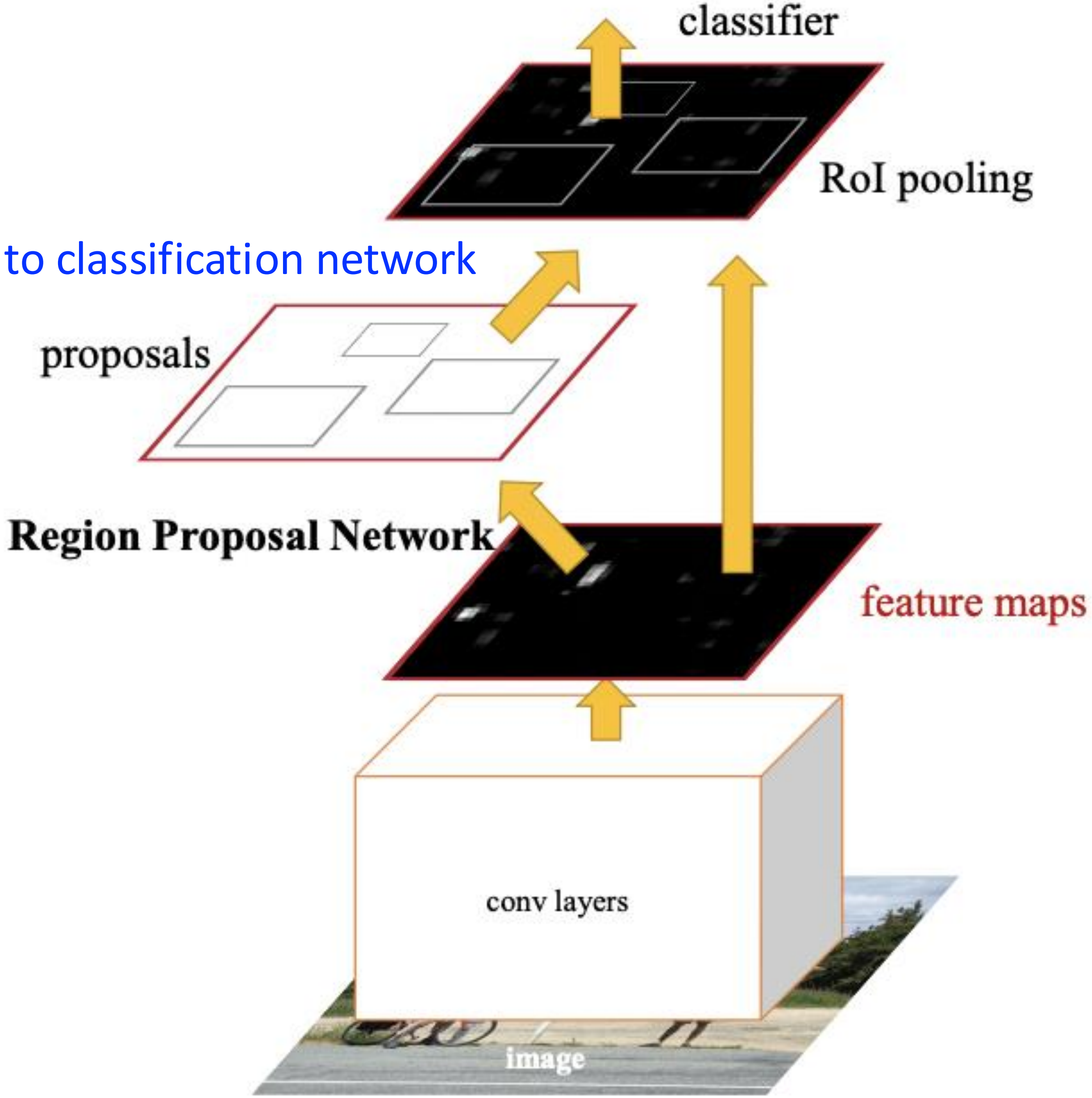
Inference

(1) Top ~300 boxes passed from RPT to classification network

(2) Non-maximum suppression: removes neighbor regions with lower confidence scores



https://towardsdatascience.com/deep-learning-method-for-object-detection-r-cnn-explained-ecdadd751d22



Ren Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” Neurips 2015

Limitations

- Requires indirect 2-stage process (predict proposals and then classify)
- Requires post-processing to remove duplicates (e.g., non-maximum suppression)
- Cannot run in real-time (i.e., relatively slow)

Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)

Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion

Why DETR?

Named after the proposed technique **DE**tection **TR**ansformer:

Carion et al. “End-to-End Object Detection with Transformers.” ECCV 2020.

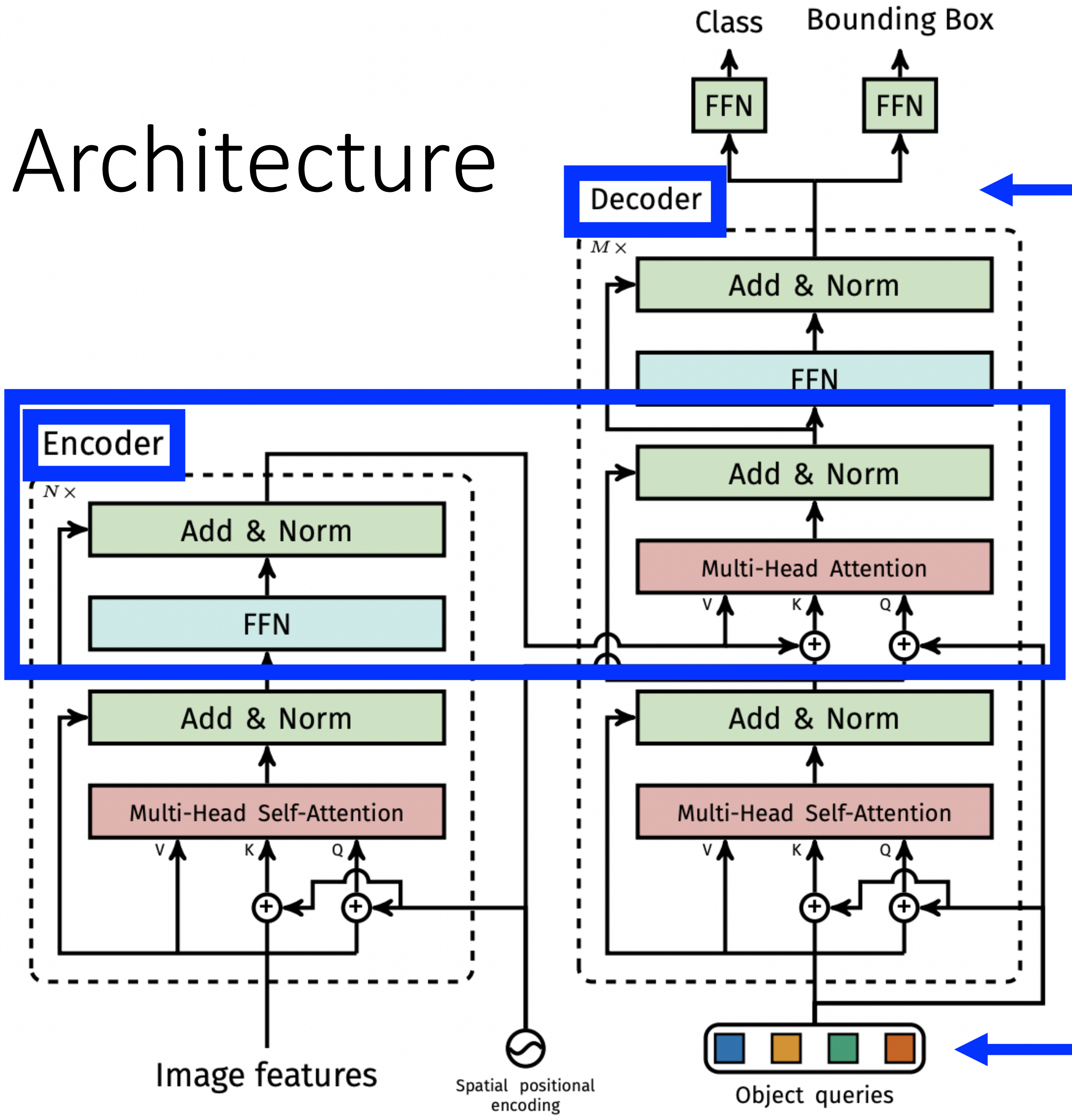
Key Contributions of DETR

1. First fully end-to-end object detection model (e.g., no post-processing)

2. First to perform object detection with the Transformer's encoder-decoder network

3. Achieves comparable performance to Faster R-CNN

Architecture



Each object query feature is transformed independently to predict a location (4 values) and class (including "no object"); same FFNs shared for all queries

Prediction is conditioned on the image by using "cross-attention" with the encoder's output representations (i.e., query comes from decoder and keys and values come from encoder)

The number of positional embeddings fed to the decoder (called "object queries) determine how many objects get detected

Carion et al. "End-to-End Object Detection with Transformers." ECCV 2020

Architecture: Encoder’s Context

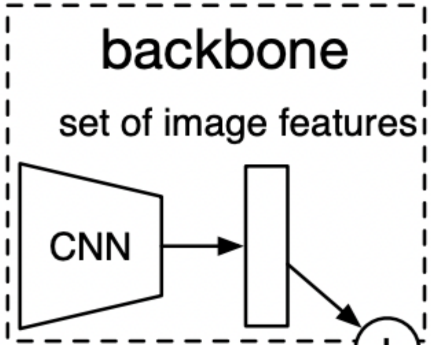
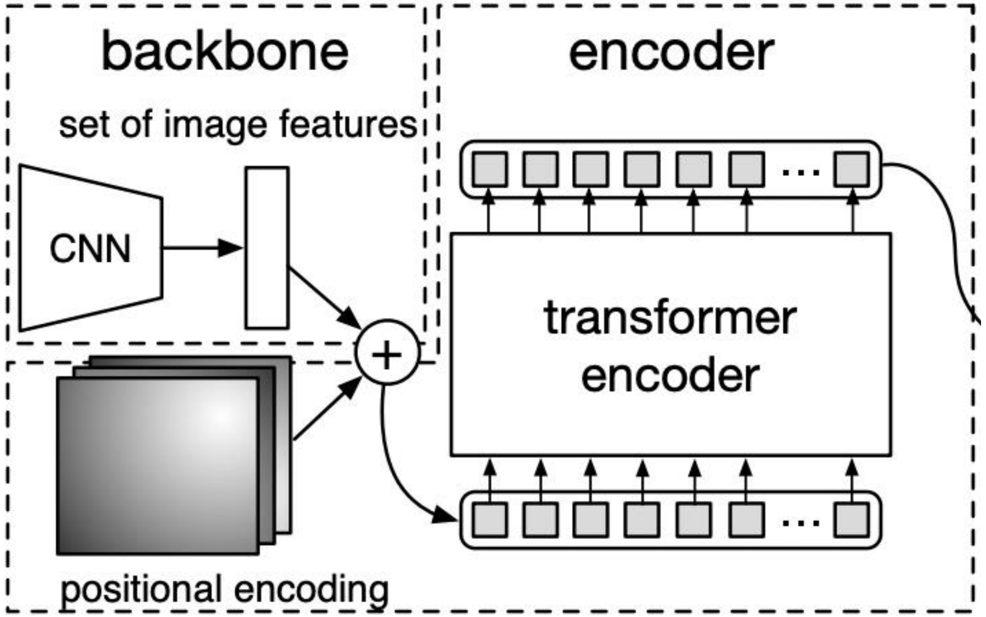


Image converted into a more compact CNN feature map, which is flattened into a set of feature vectors (e.g., CxHxW -> HxW, C)

Carion et al. “End-to-End Object Detection with Transformers.” ECCV 2020

Architecture: Encoder's Context



Features modified to embed global context
for how each each relates to all other features

(recall, positional embeddings infuse spatial
information)

Carion et al. "End-to-End Object Detection with Transformers." ECCV 2020

Training: First of Two Steps

1. Associate predictions to ground truths with Hungarian algorithm

Number of predictions (ground truth set padded with "no object" until size *N*)

Forces unique one-to-one matching between ground truths and predictions

$$\underset{\sigma \in \mathfrak{S}_N}{\arg\min} \sum_{i}^{\boxed{N}} \underbrace{\mathcal{L}_{\text{match}}\left(y_i, \hat{y}_{\sigma(i)}\right)}_{\text{Based on both classification and localization}}$$

2. Compute loss based on classification and localization (like Faster R-CNN)

$$\sum_{i=1}^{N} \left[\underbrace{-\log \hat{p}_{\hat{\sigma}(i)}(c_i)}_{\text{Classification performance}} + \underbrace{\mathbb{1}_{\{c_i \neq \varnothing\}}}_{\text{When object is present:}} \underbrace{\mathcal{L}_{\text{box}}(b_i, \hat{b}_{\hat{\sigma}}(i))}_{\text{Detection performance}} \right]$$

CVPR 2024 paper reveals (1) how DETR and Faster R-CNN can be interpreted in the same way and (2) what are the key ingredients leading to DETR's advantages

Hybrid Proposal Refiner: Revisiting DETR Series from the Faster R-CNN Perspective

Jinjing Zhao* Fangyun Wei* Chang Xu
The University of Sydney
jzha0100@uni.sydney.edu.au fwei8714@uni.sydney.edu.au c.xu@sydney.edu.au

Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)

Object Detection: Today's Topics

- Motivation
- Datasets
- Evaluation metric
- Faster R-CNN
- DETR
- Discussion (chosen by YOU ☺)


The End