

A Brief History of Visual Object Detection

Test of Time Award Presentation for

**“Faster R-CNN: Towards Real-Time Object Detection
with Region Proposal Networks”, NeurIPS 2015**

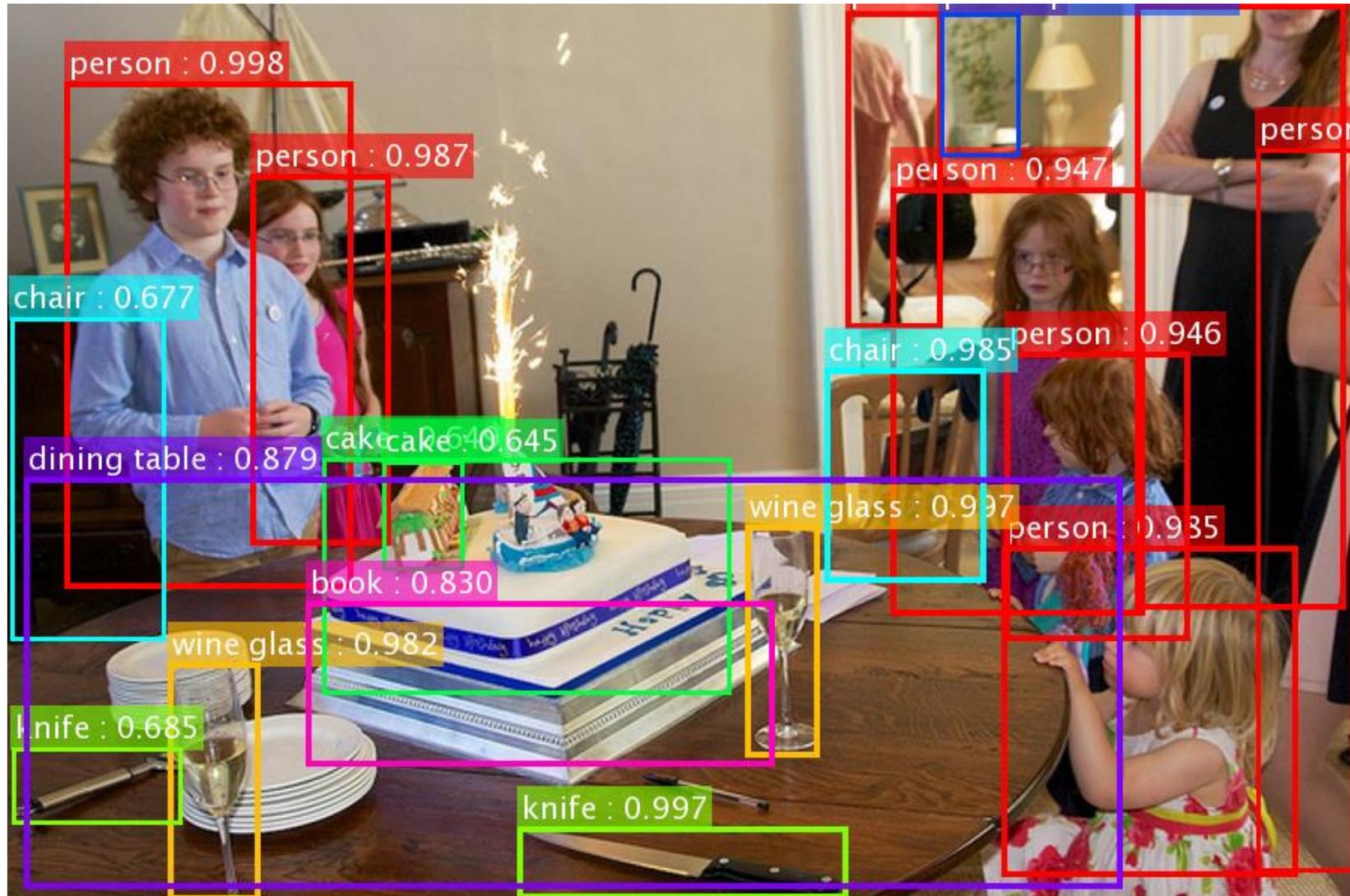
Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun

presenter: Kaiming He



What Is Object Detection?

recognition + localization



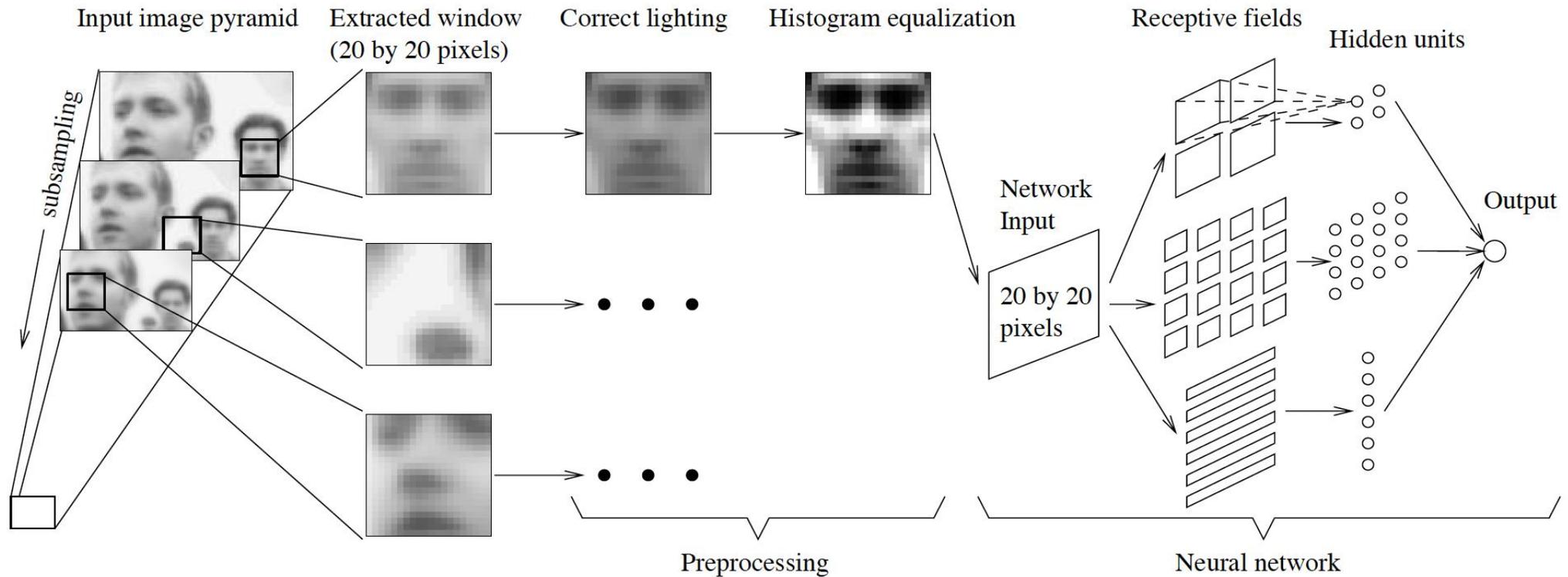
30 Years of Visual Object Detection

Neural Network-Based Face Detection

H. Rowley, S. Baluja, T. Kanade. CVPR 1996

Test of Time Award, CVPR 2006

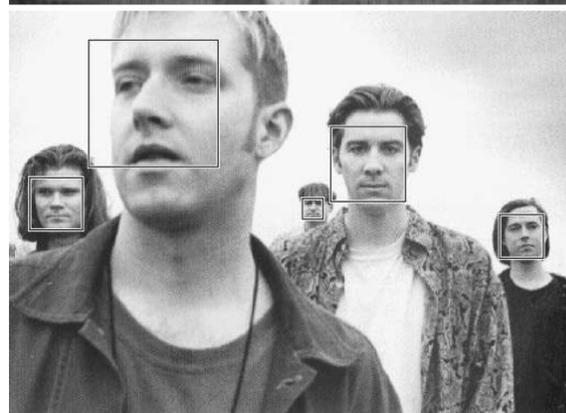
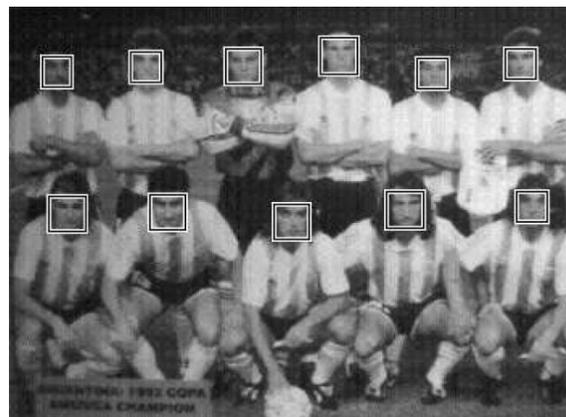
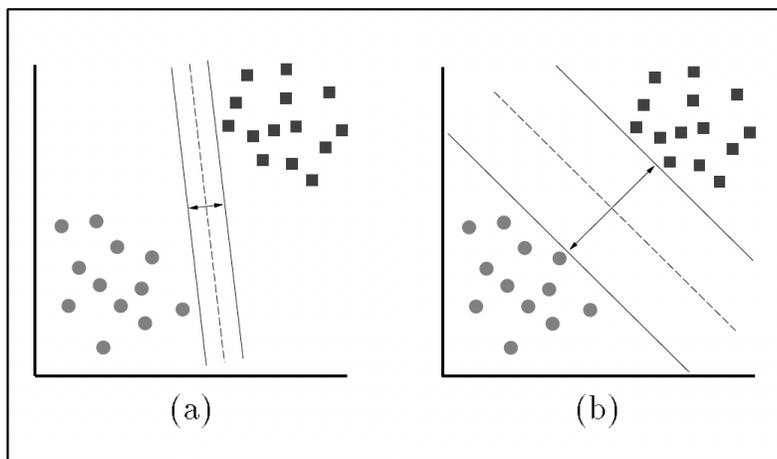
(first CV paper I've ever read)



SVM for Face Detection

E. Osuna, R. Freund, F. Girosi. CVPR 1997

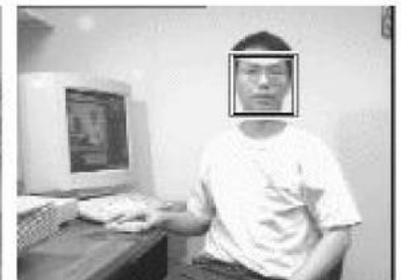
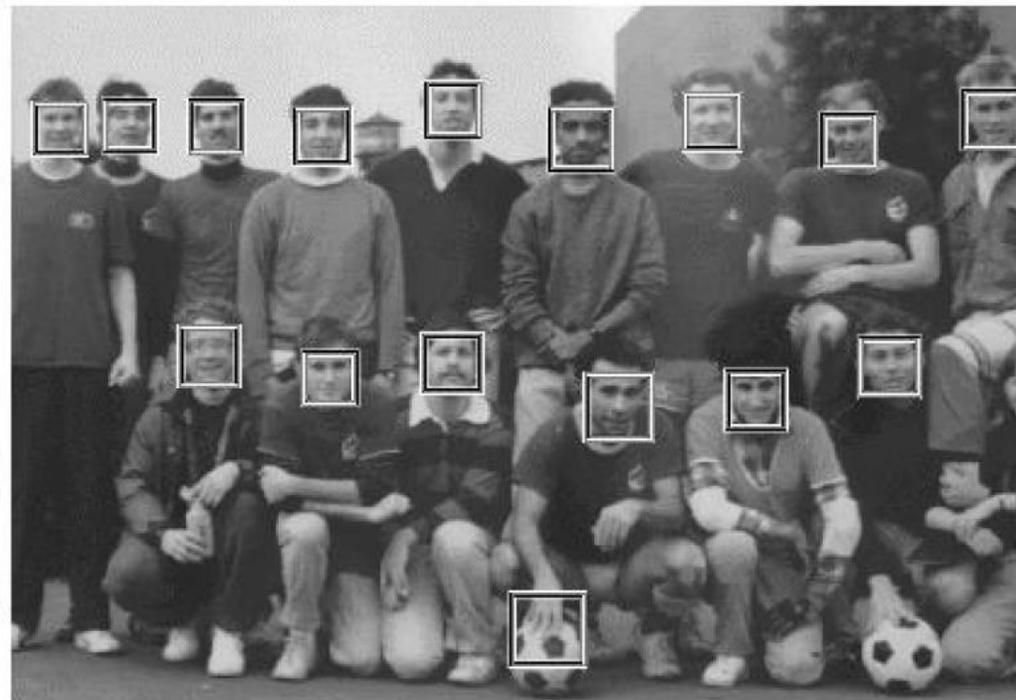
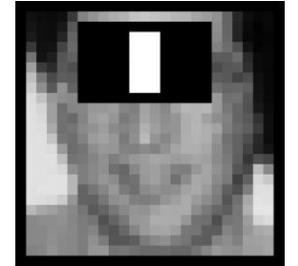
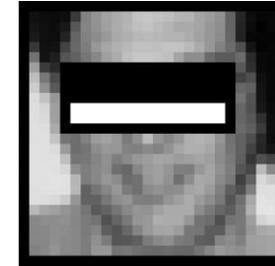
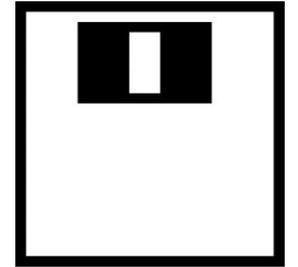
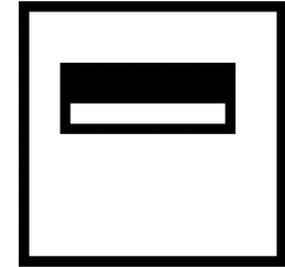
Test of Time Award, CVPR 2007



The Viola-Jones Framework

P. Viola, M. Jones. CVPR 2001

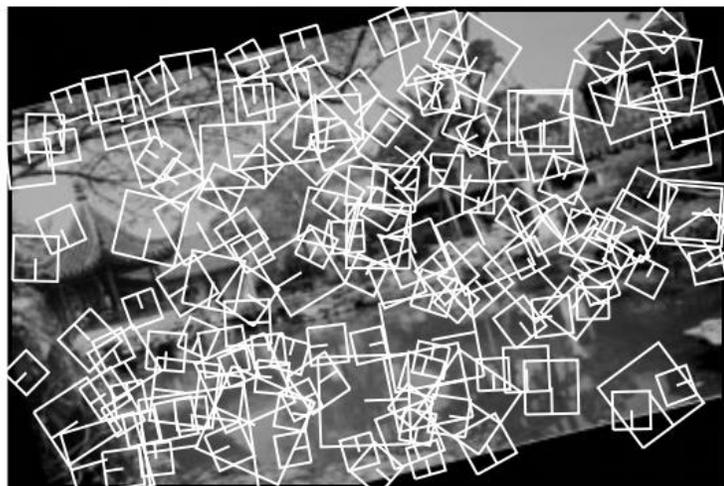
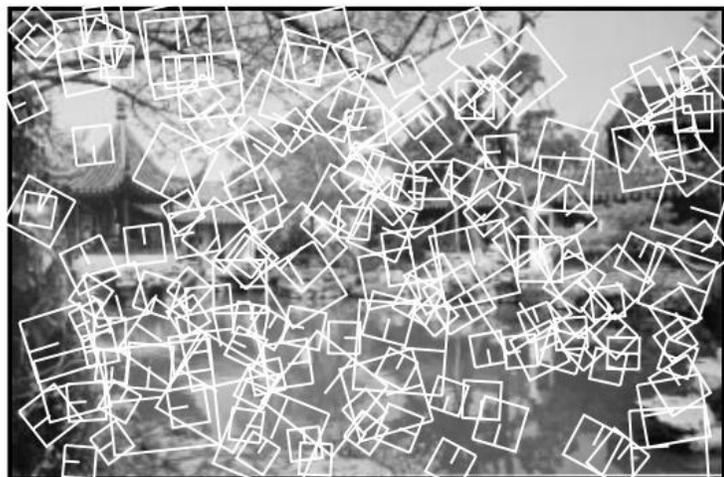
Test of Time Award, CVPR 2011



SIFT

D. Lowe. ICCV 1999

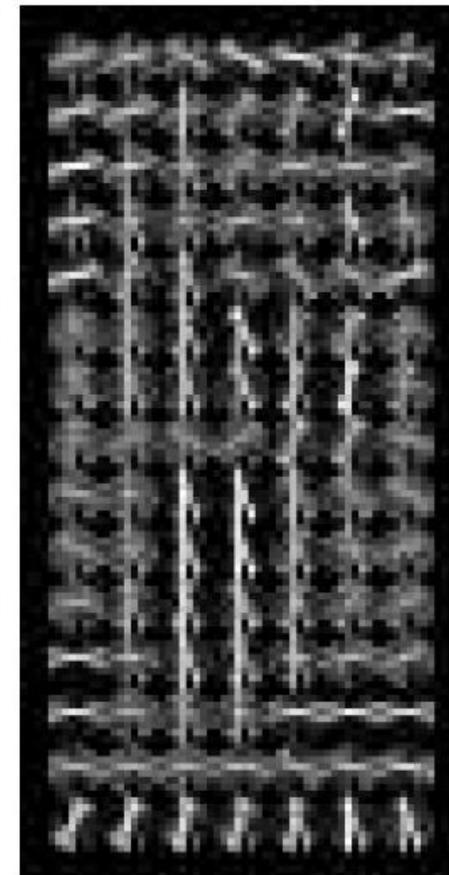
Test of Time Award, ICCV 2011



HOG

N. Dalal, B. Triggs. CVPR 2005

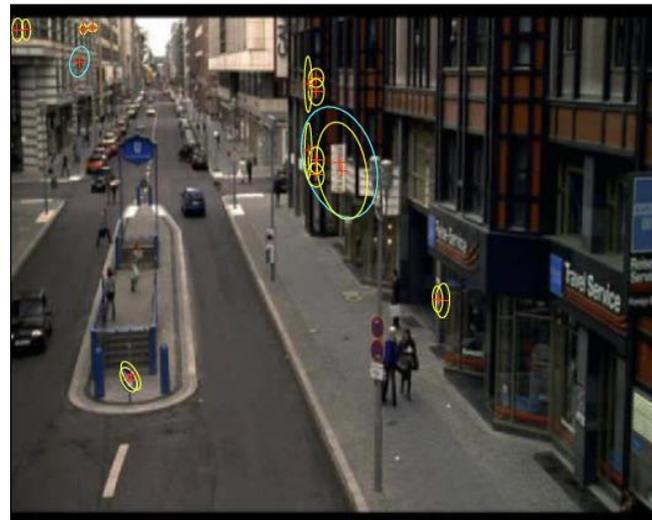
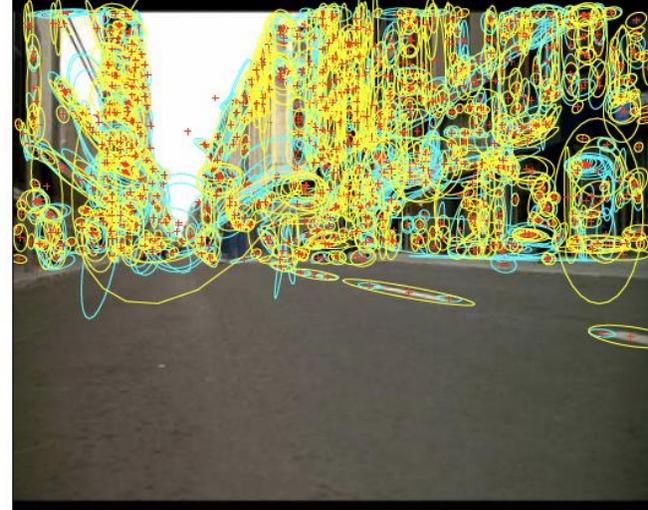
Test of Time Award, CVPR 2015



Bag of Visual Words

J. Sivic, A. Zisserman. ICCV 2003

Test of Time Award, ICCV 2017

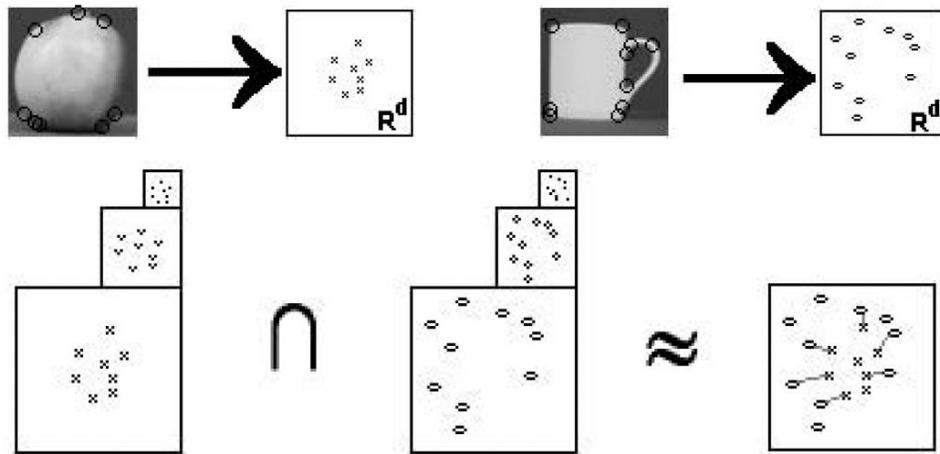


Pyramid Match Kernel

K. Grauman and T. Darrell.

ICCV 2005

Test of Time Award, ICCV 2017

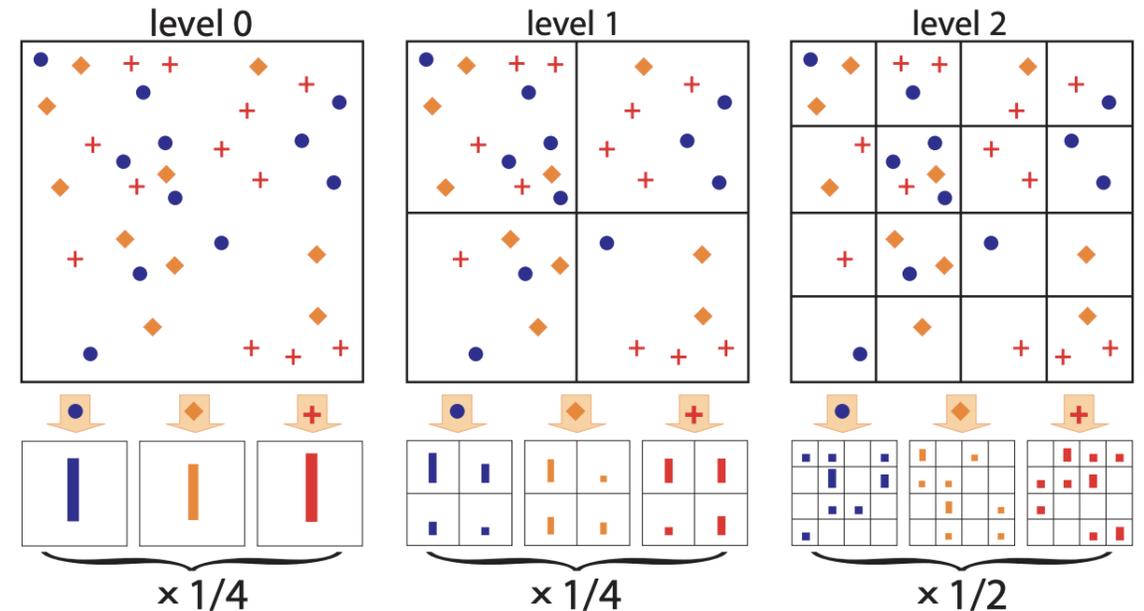


Spatial Pyramid Matching

S. Lazebnik, C. Schmid, J. Ponce.

CVPR 2006

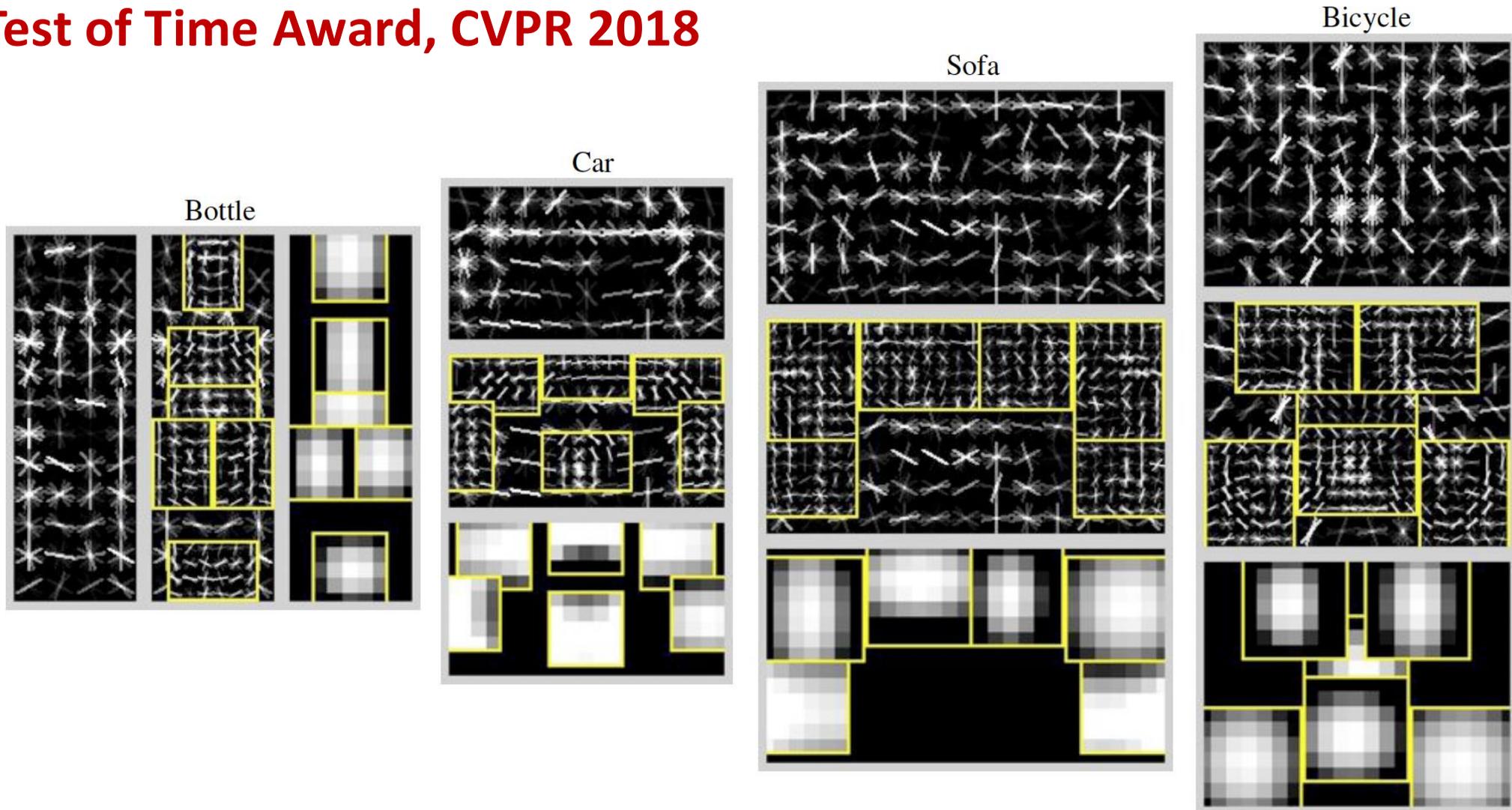
Test of Time Award, CVPR 2016



Deformable Part Model (DPM)

P. Felzenszwalb, D. McAllester, and D. Ramanan. CVPR 2008

Test of Time Award, CVPR 2018



Debate in ImageNet 2012 Workshop

Impressive ImageNet

Task: 1000-way whole-image classification

Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
<i>SIFT + FVs [7]</i>	—	—	26.2%
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	16.4%
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	15.3%

metric: classification error rate

But... does it generalize to other datasets and tasks?

See: [DeCAF Tech Report, Donahue et al.](#)

Much debate at ECCV'12 and on [LeCun's Google+](#)

the ILSVRC workshop at ECCV 2012.¹ The central issue of debate can be distilled as: **To what extent do the CNN classification results on ImageNet generalize to object detection results on the PASCAL VOC Challenge?**

In this paper, we study object detection using features computed by a large convolutional neural network, answering this important scientific question.² from the R-CNN paper

Debate summary

“So, why not just try deep learning on PASCAL VOC and make this whole thing a non-issue?”

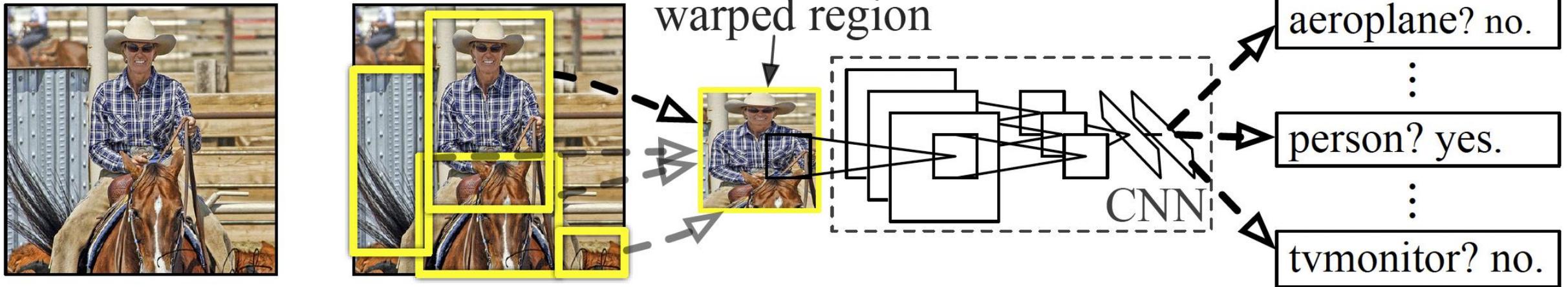
– [Alyosha Efros](#)

Last comment at <http://goo.gl/PhTvoE>
(10 months ago)

R-CNN (Region-based CNN)

R. Girshick, J. Donahue, T. Darrell and J. Malik. CVPR 2014

Test of Time Award, CVPR 2024



R-CNN (Region-based CNN)

R. Girshick, J. Donahue, T. Darrell and J. Malik. CVPR 2014

Test of Time Award, CVPR 2024



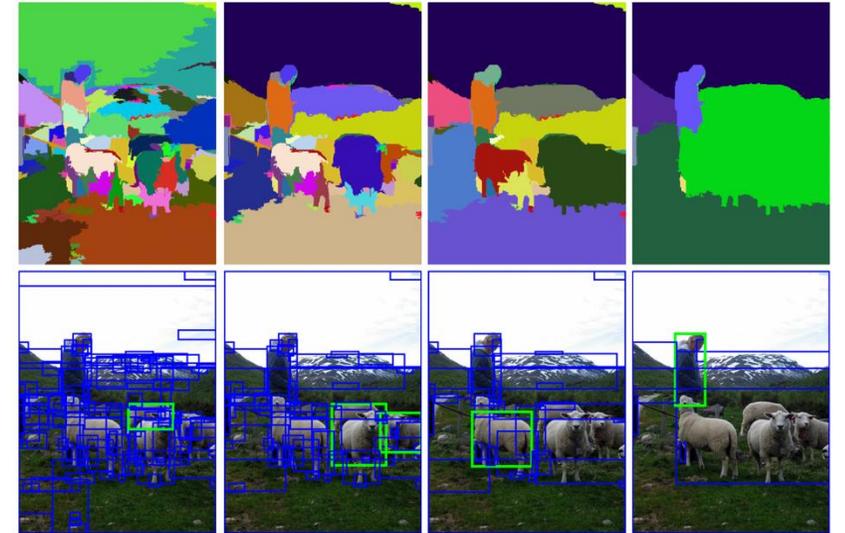
R-CNN (Region-based CNN)

R. Girshick, J. Donahue, T. Darrell and J. Malik. CVPR 2014

Test of Time Award, CVPR 2024



Selective Search

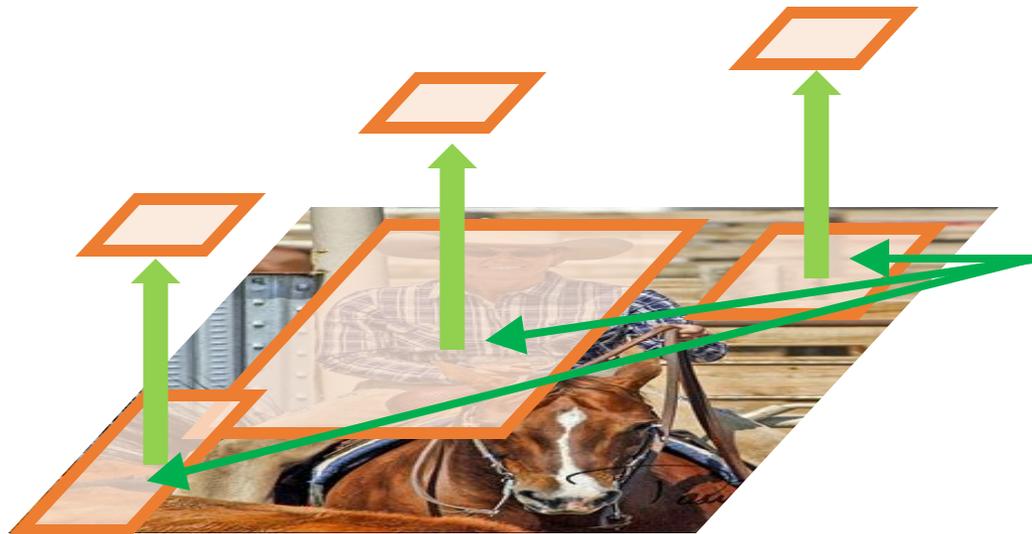


- ~2000 region proposals

R-CNN (Region-based CNN)

R. Girshick, J. Donahue, T. Darrell and J. Malik. CVPR 2014

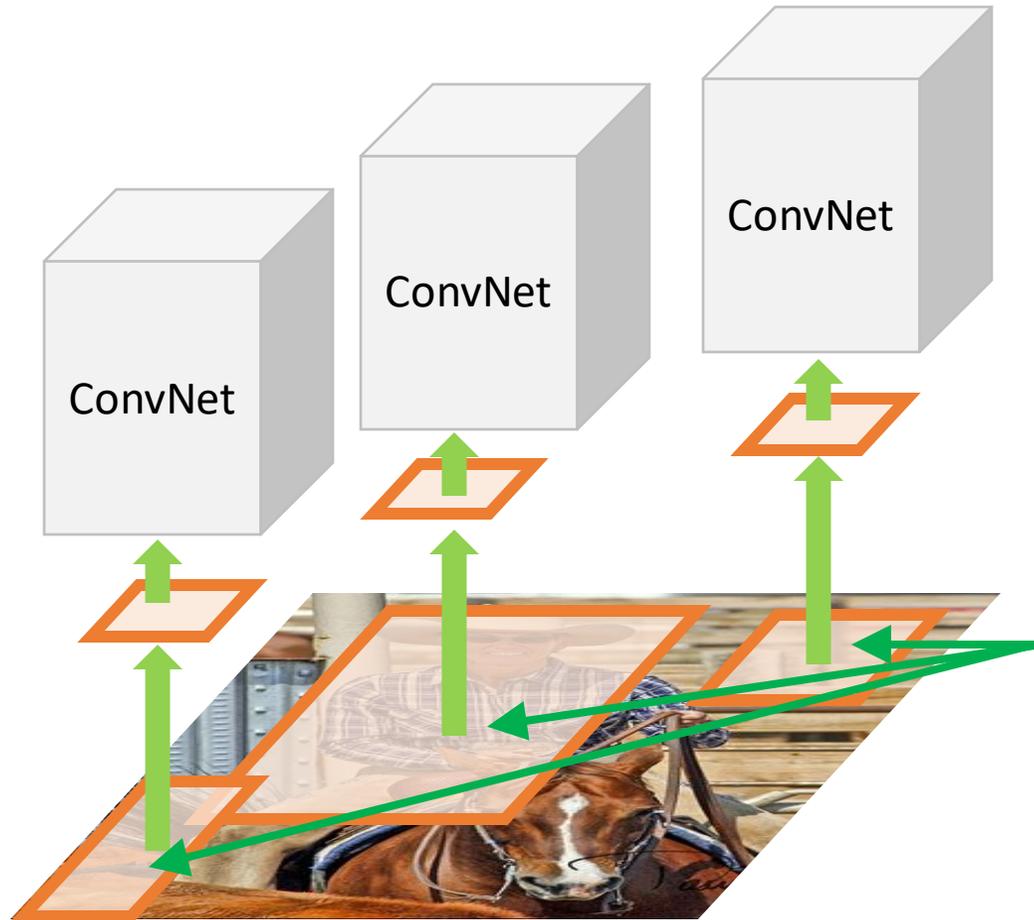
Test of Time Award, CVPR 2024



R-CNN (Region-based CNN)

R. Girshick, J. Donahue, T. Darrell and J. Malik. CVPR 2014

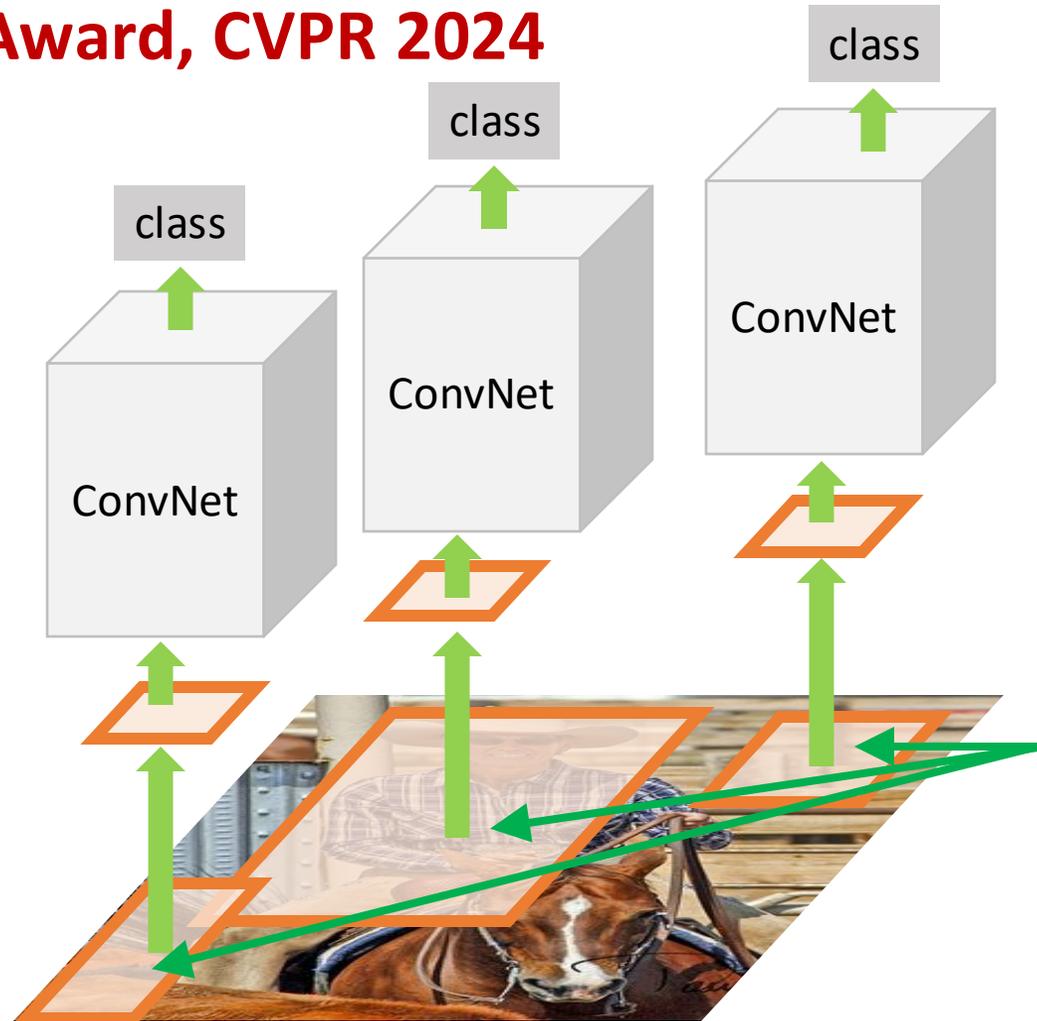
Test of Time Award, CVPR 2024



R-CNN (Region-based CNN)

R. Girshick, J. Donahue, T. Darrell and J. Maliki. CVPR 2014

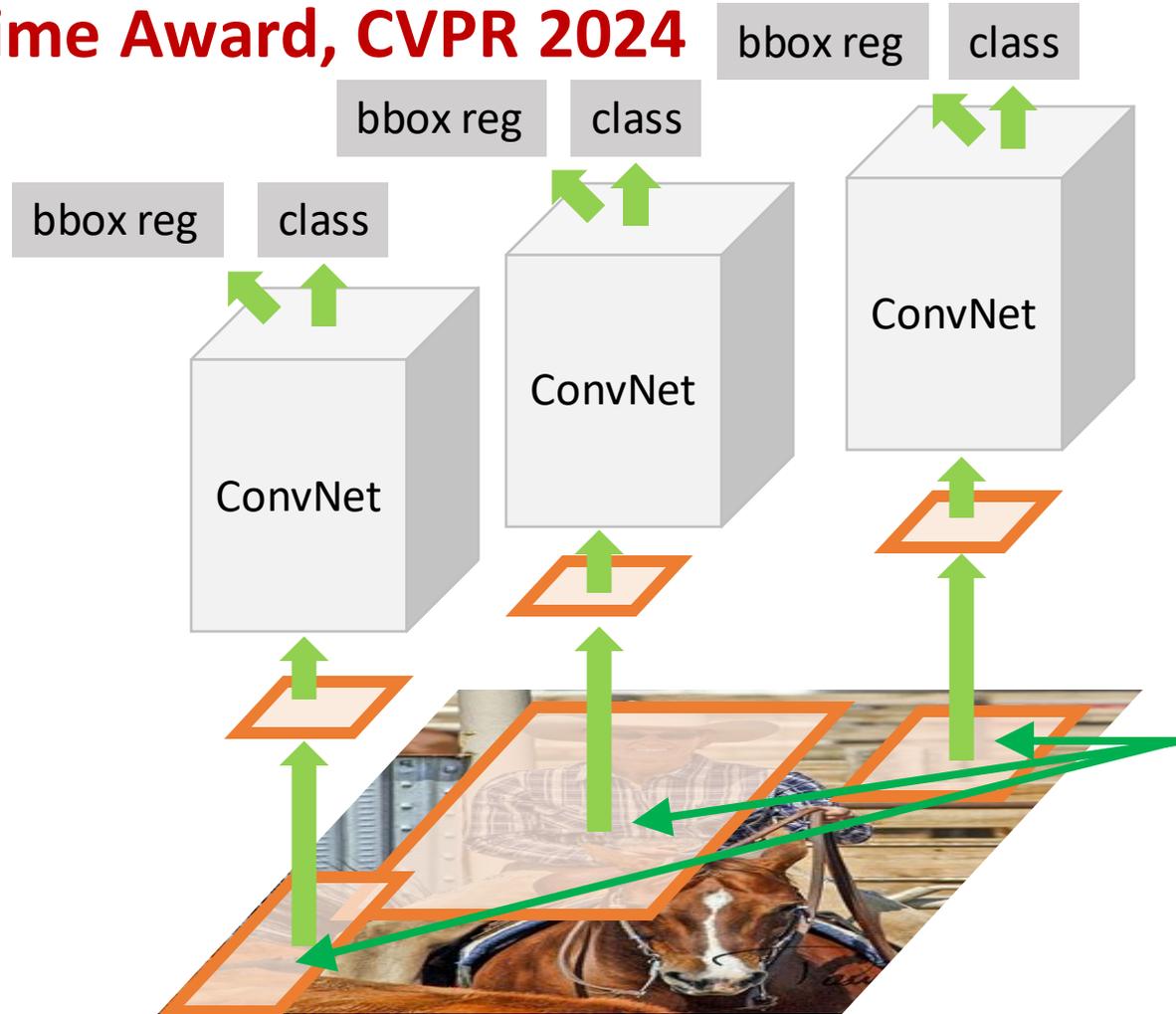
Test of Time Award, CVPR 2024



R-CNN (Region-based CNN)

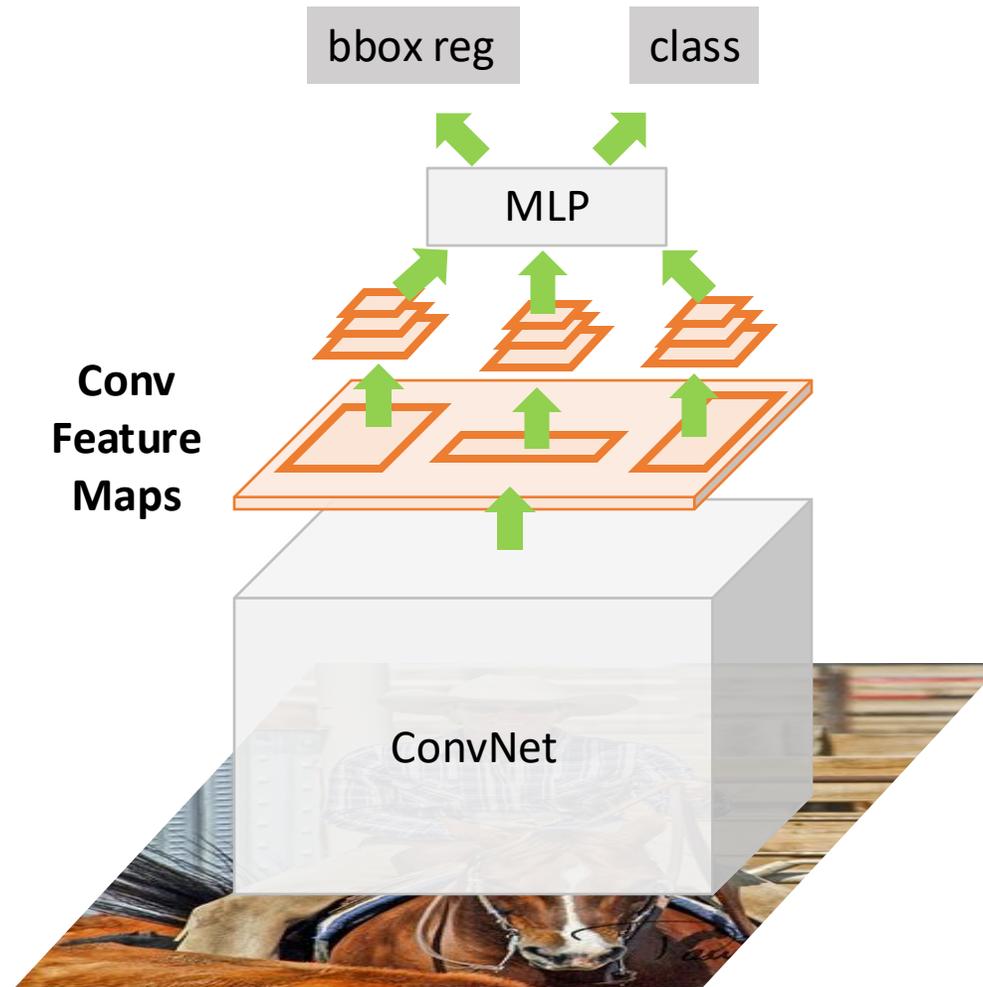
R. Girshick, J. Donahue, T. Darrell and J. Maliki. CVPR 2014

Test of Time Award, CVPR 2024



Conv Feature Maps + Spatial Pyramid Pooling

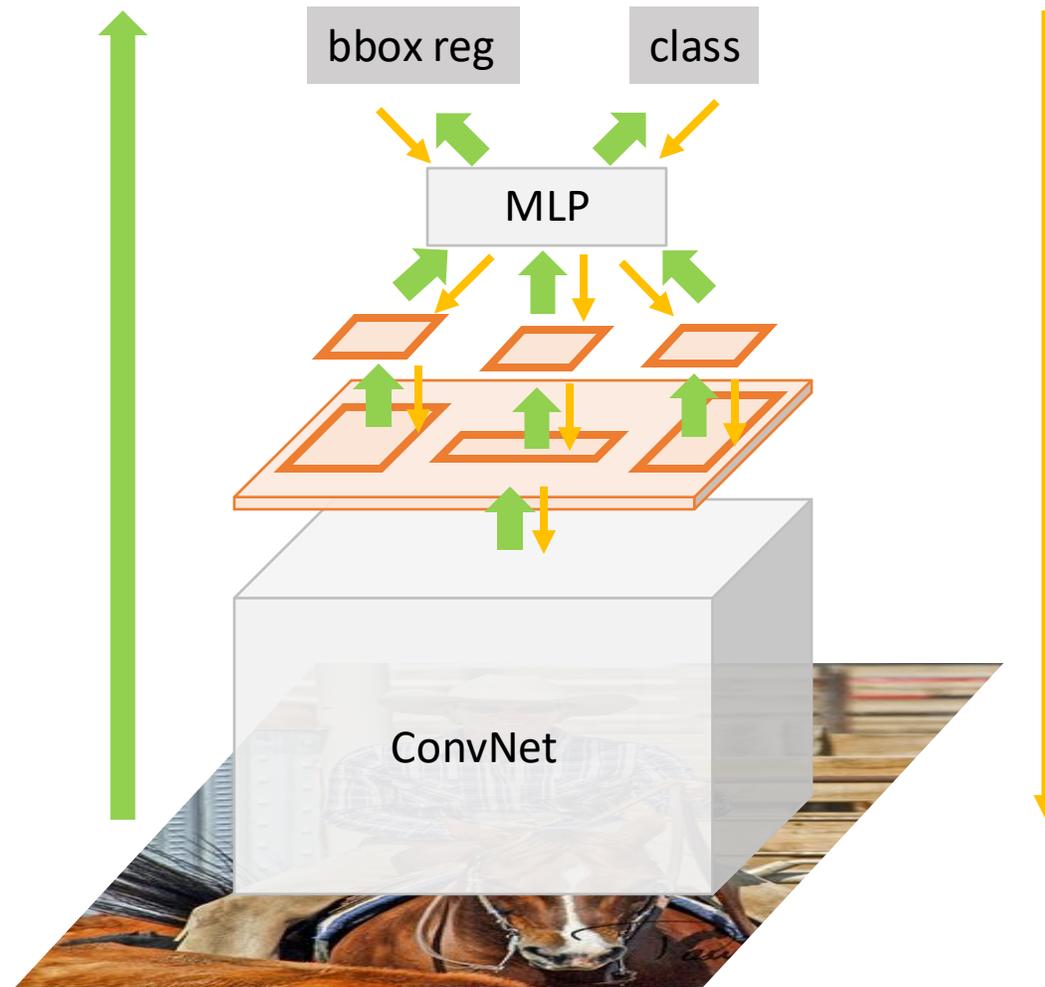
K. He, X. Zhang, S. Ren, J. Sun. ECCV 2014



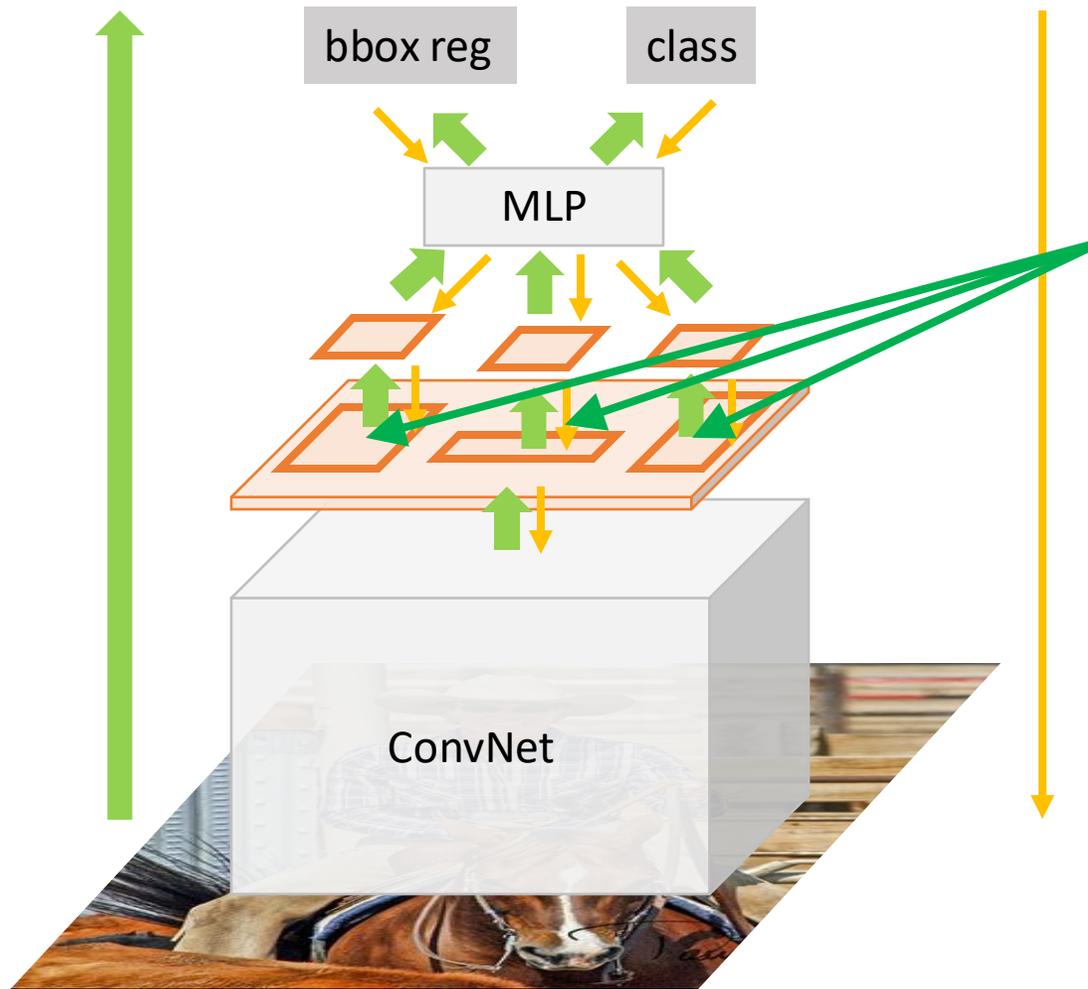
≈ analogize
HOG + SPM

Fast R-CNN

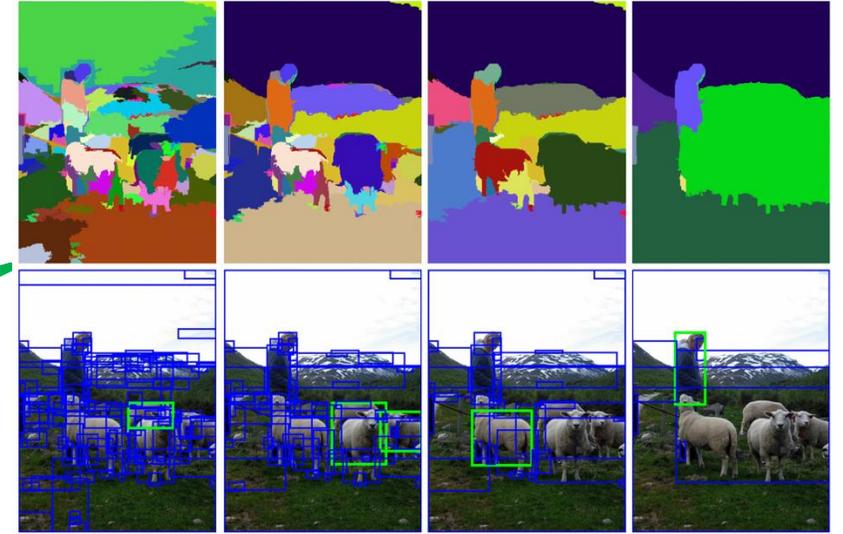
R. Girshick. ICCV 2015. **Test of Time Award, ICCV 2025**



What's the missing piece?



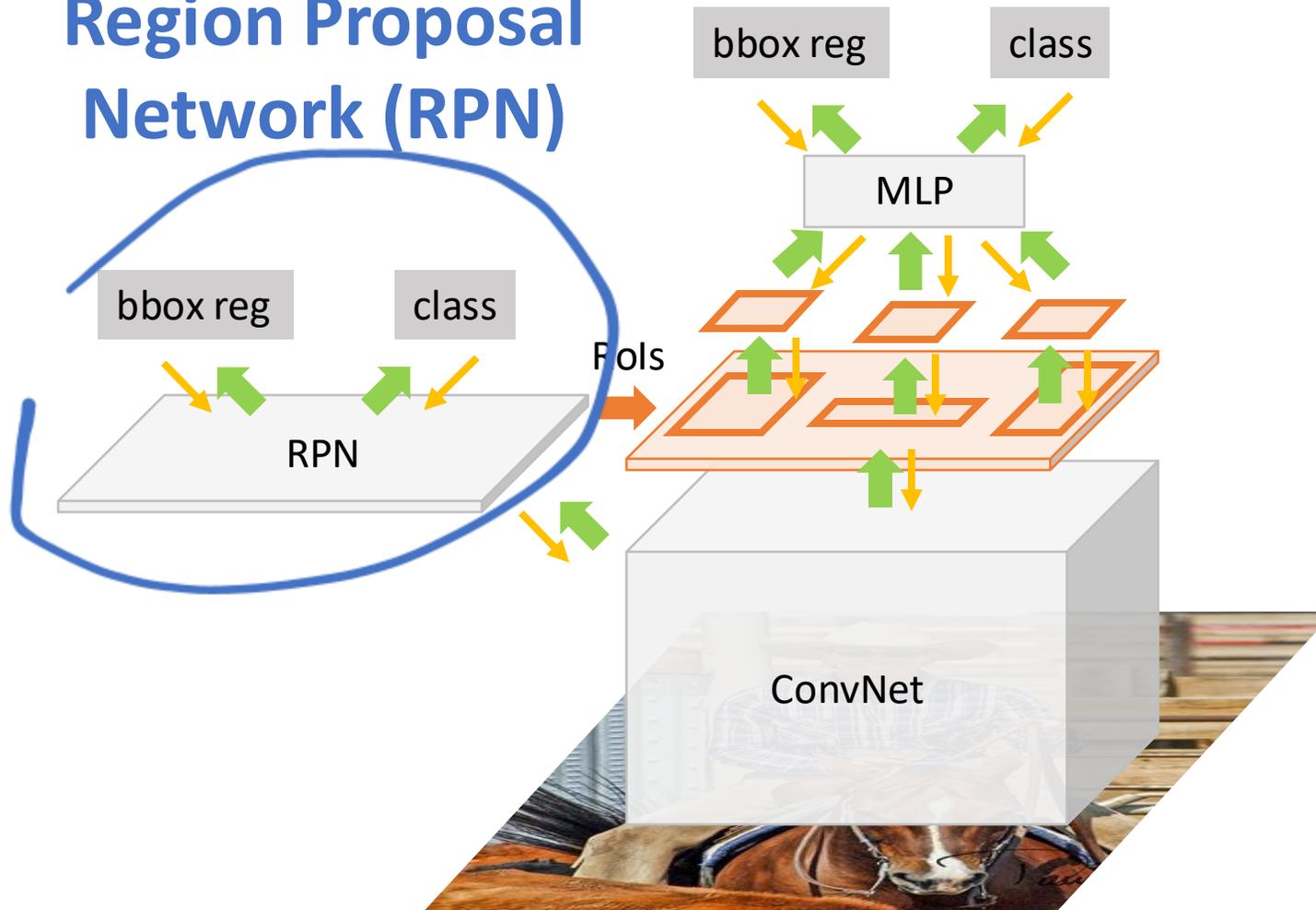
hand-crafted regions



Faster R-CNN: Everything in Deep Learning

S. Ren, K. He, R. Girshick, J. Sun. NeurIPS 2015. **This Talk**

Region Proposal Network (RPN)

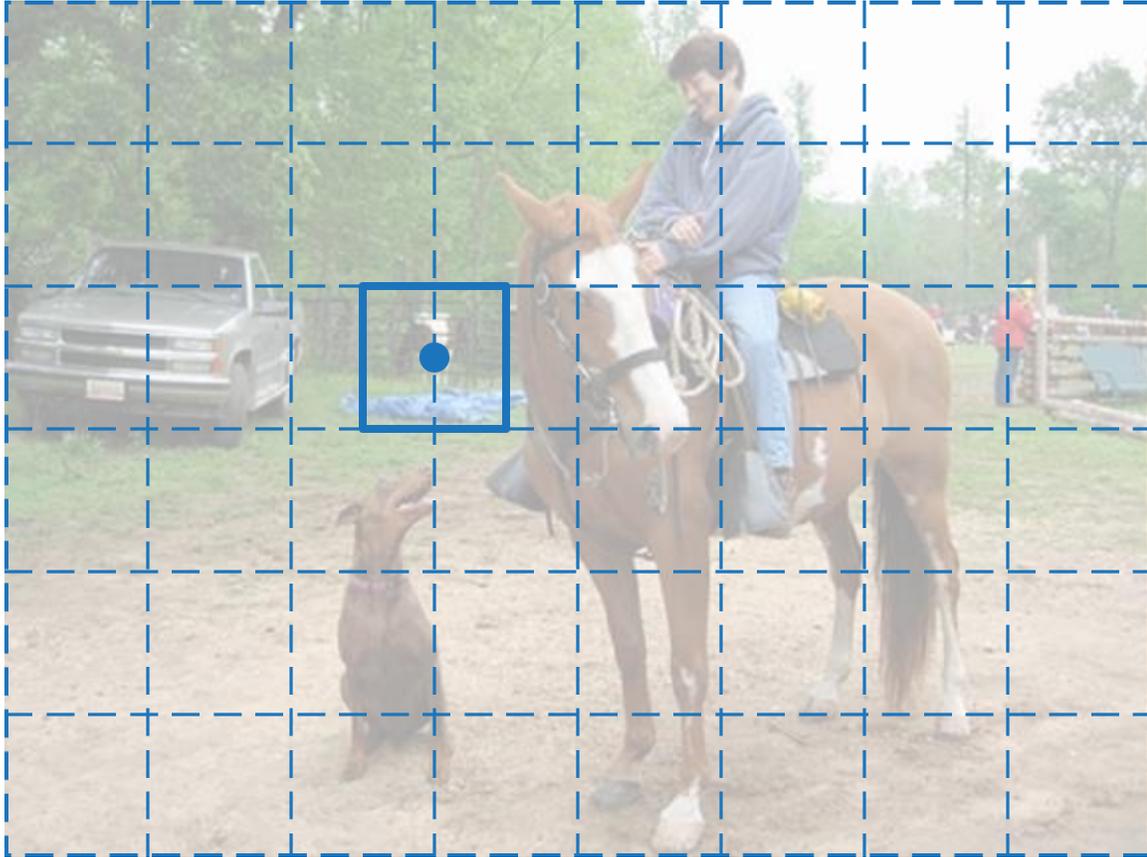


Region Proposal Network



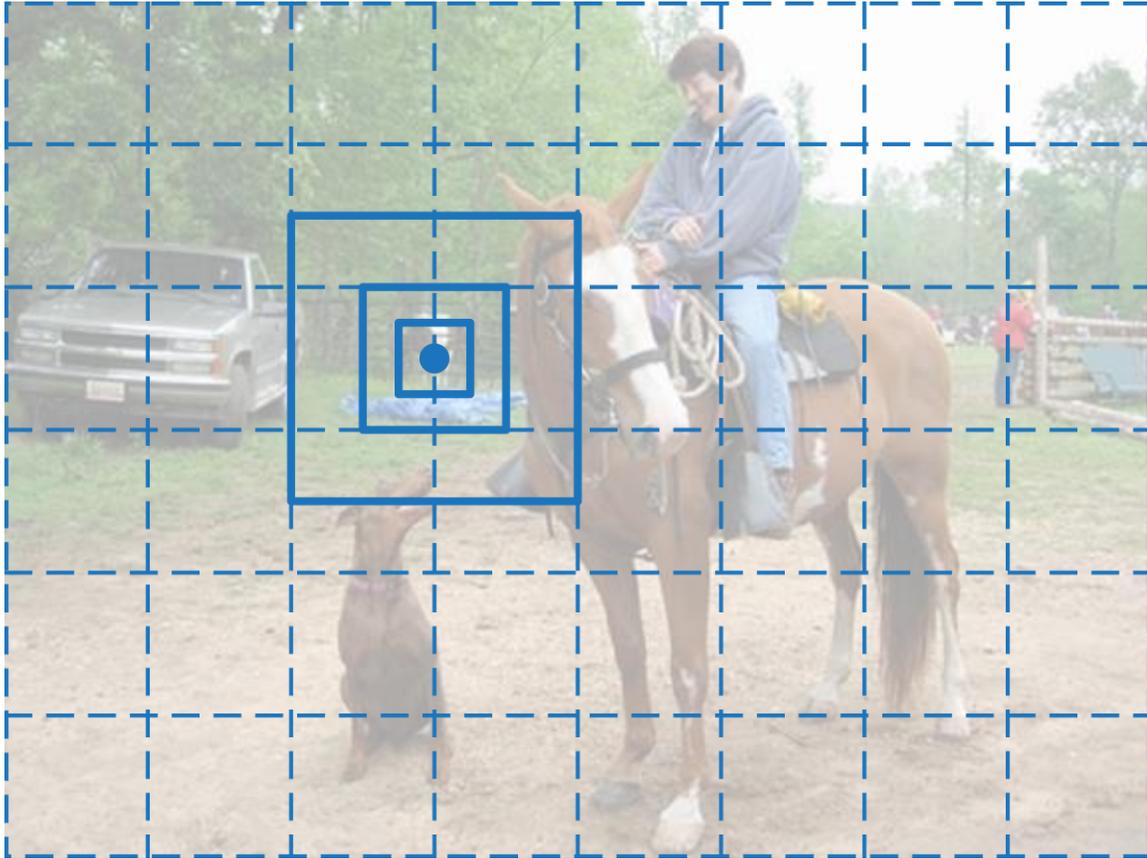
**fully conv
“anchors”**

Region Proposal Network



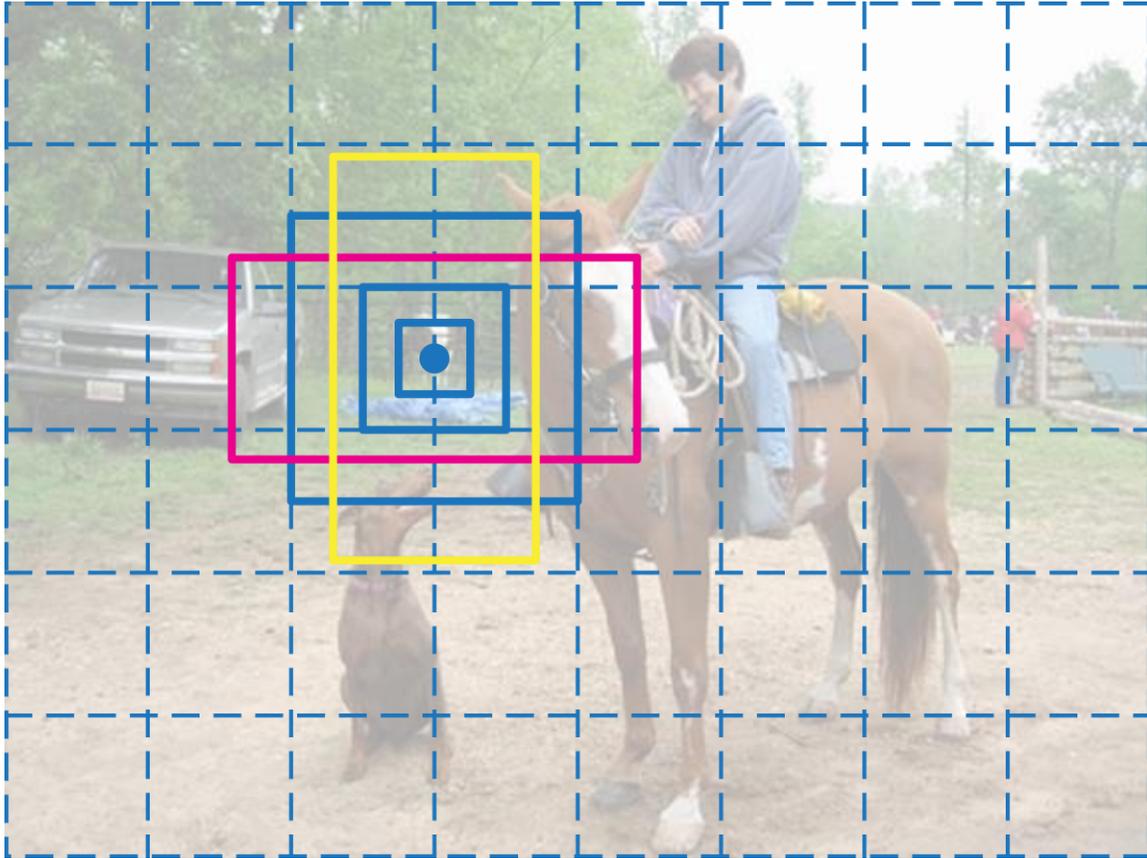
**fully conv
“anchors”**

Region Proposal Network



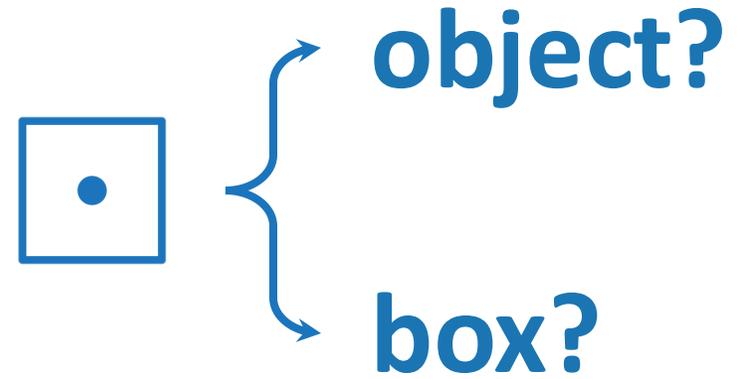
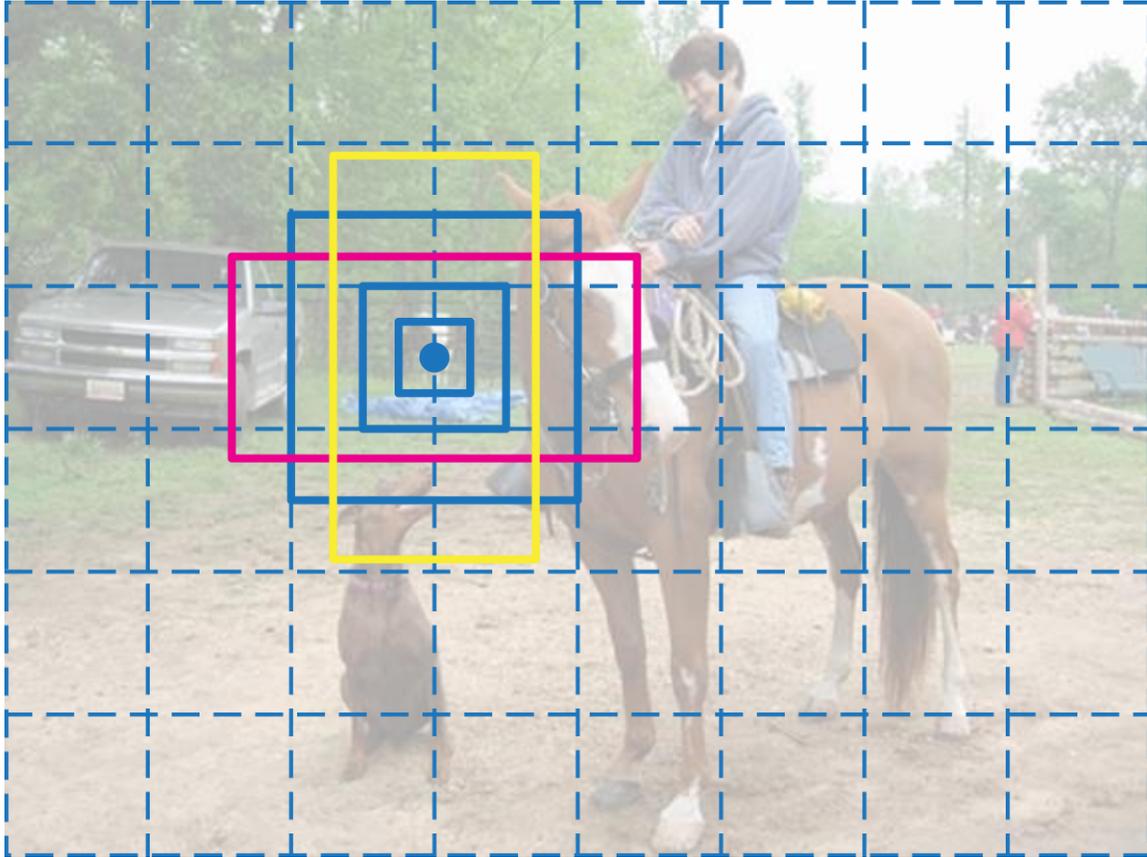
**multi-scale
“anchors”**

Region Proposal Network

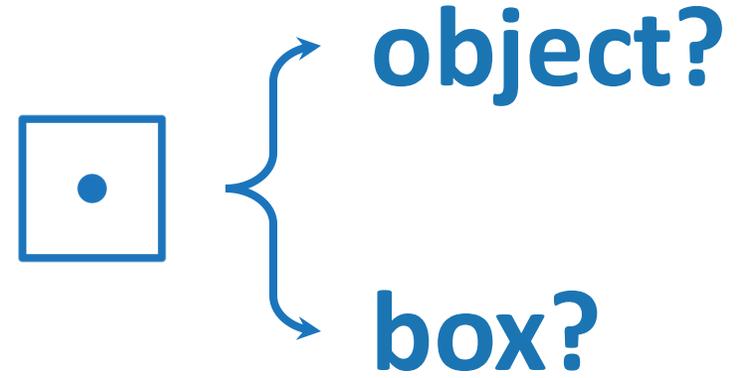
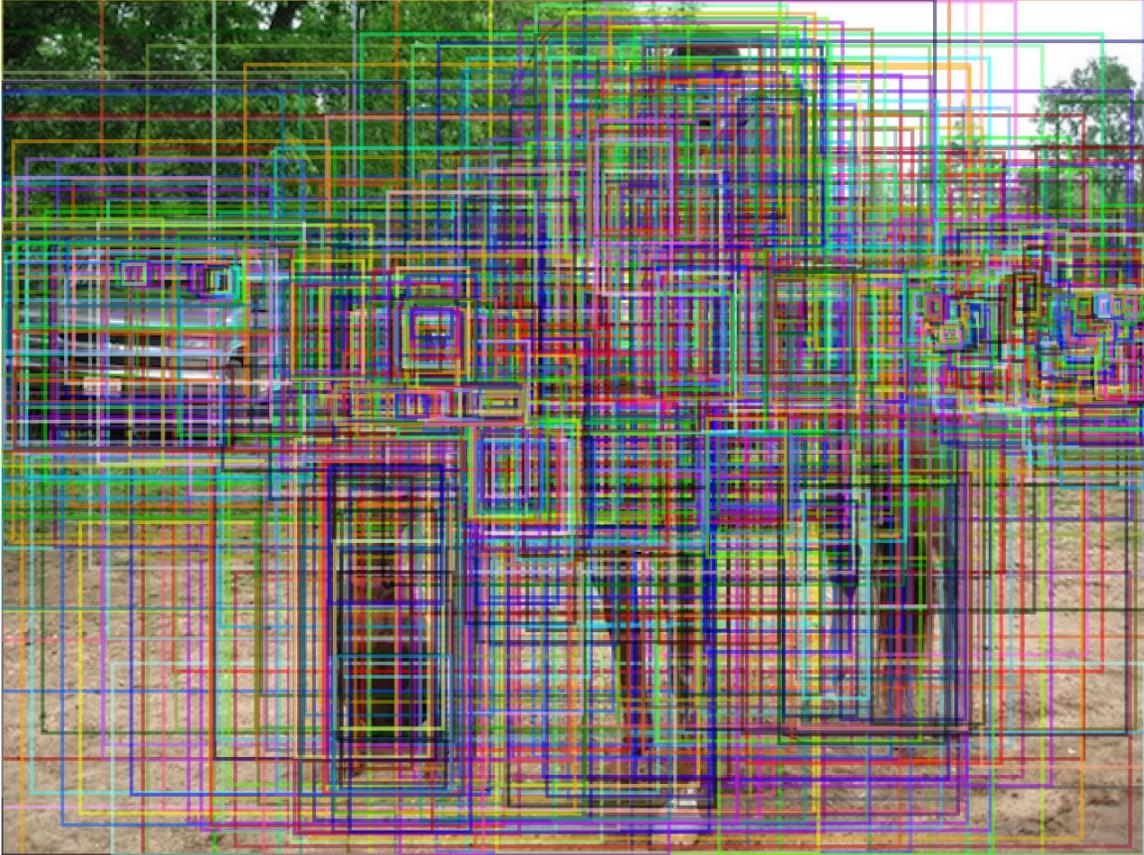


**multi-size
“anchors”**

Region Proposal Network



Region Proposal Network

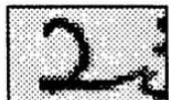
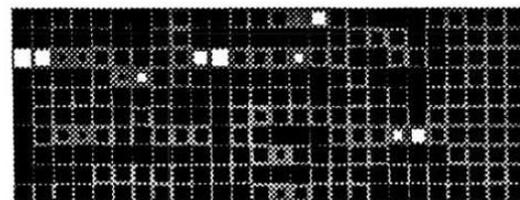
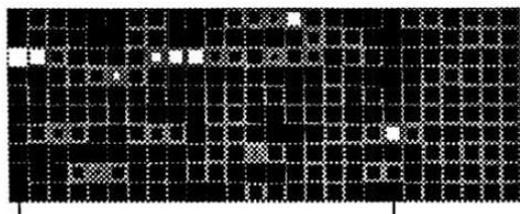


Space Displacement Net

NeurIPS 1991

Multi-Digit Recognition Using A Space Displacement Neural Network

Ofer Matan*, Christopher J.C. Burges,
Yann Le Cun and John S. Denker
AT&T Bell Laboratories, Holmdel, N. J. 07733



MultiBox

CVPR 2014

Scalable High Quality Object Detection

Christian Szegedy
Google Inc.
1600 Amphitheatre Pkwy, Mountain View, CA
szegedy@google.com

Scott Reed
University of Michigan
reedscot@umich.edu

Dimitru Erhan
dimitru@google.com

Dragomir Anguelov
dragomir@google.com

Sergey Ioffe
sioffe@google.com

Scalable Object Detection using Deep Neural Networks

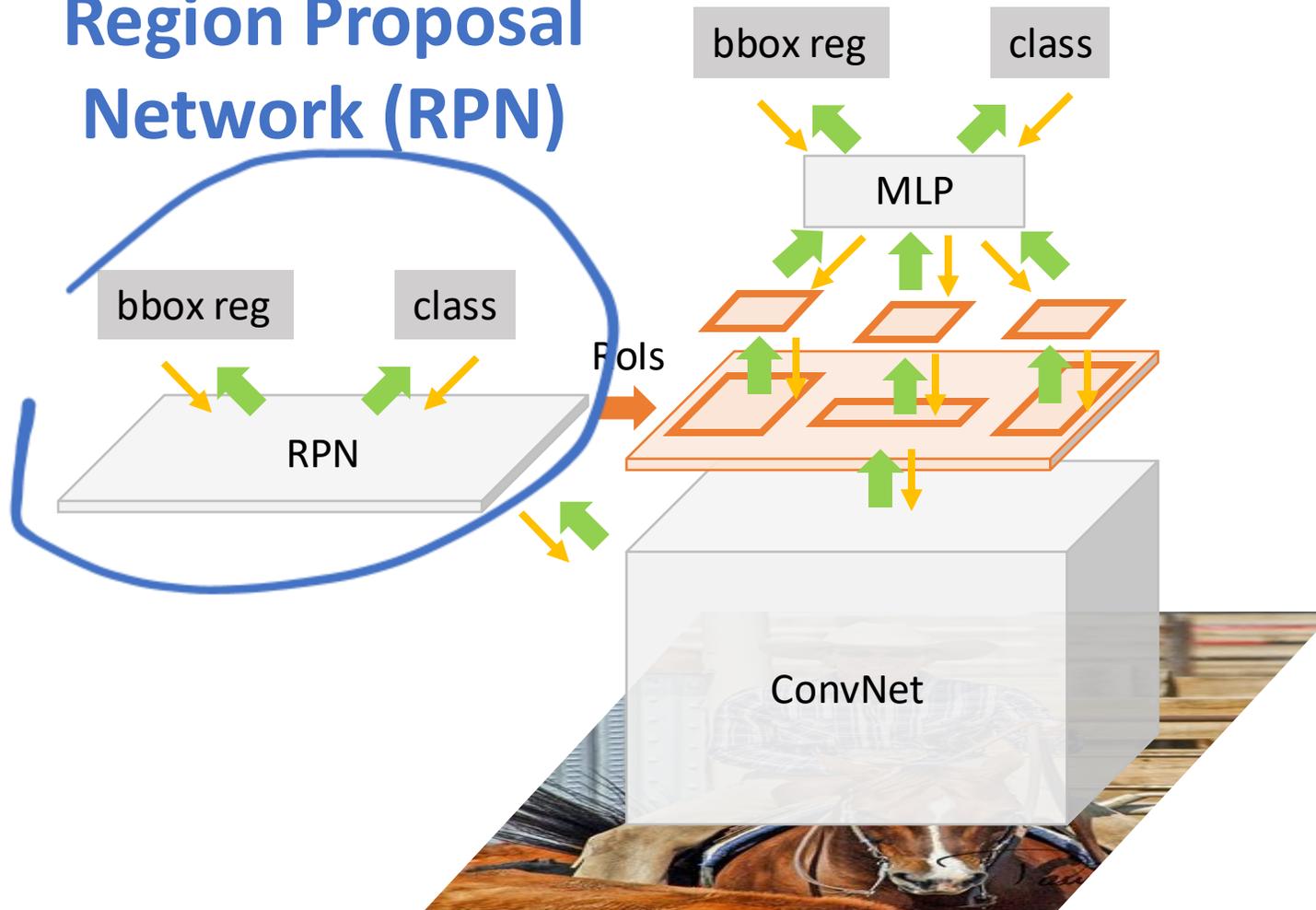
Dimitru Erhan, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov
Google, Inc.

1600 Amphitheatre Parkway, Mountain View (CA), 94043, USA
{dimitru, szegedy, toshev, dragomir}@google.com

Faster R-CNN: Everything in Deep Learning

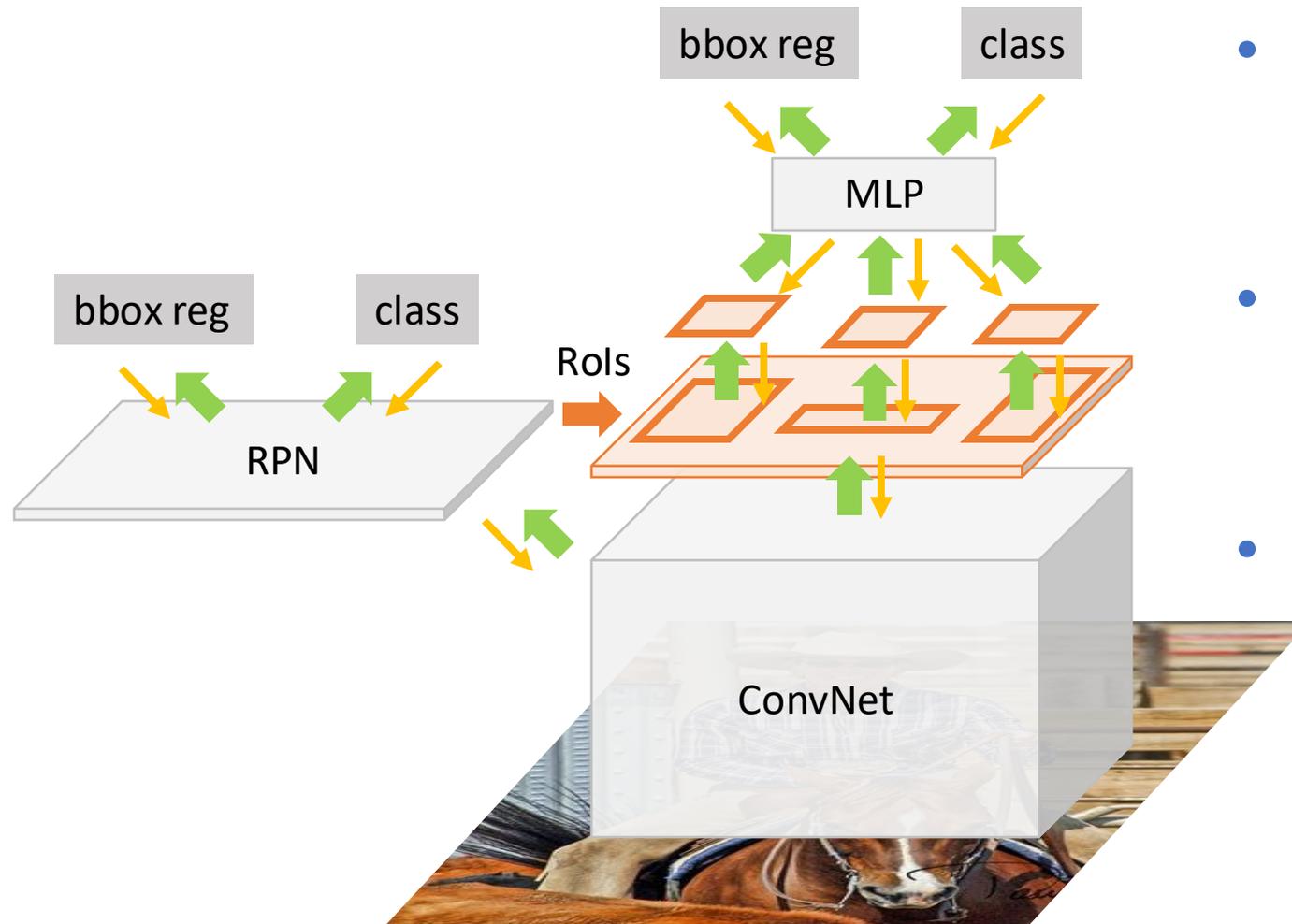
S. Ren, K. He, R. Girshick, J. Sun. NeurIPS 2015. **This Talk**

Region Proposal Network (RPN)



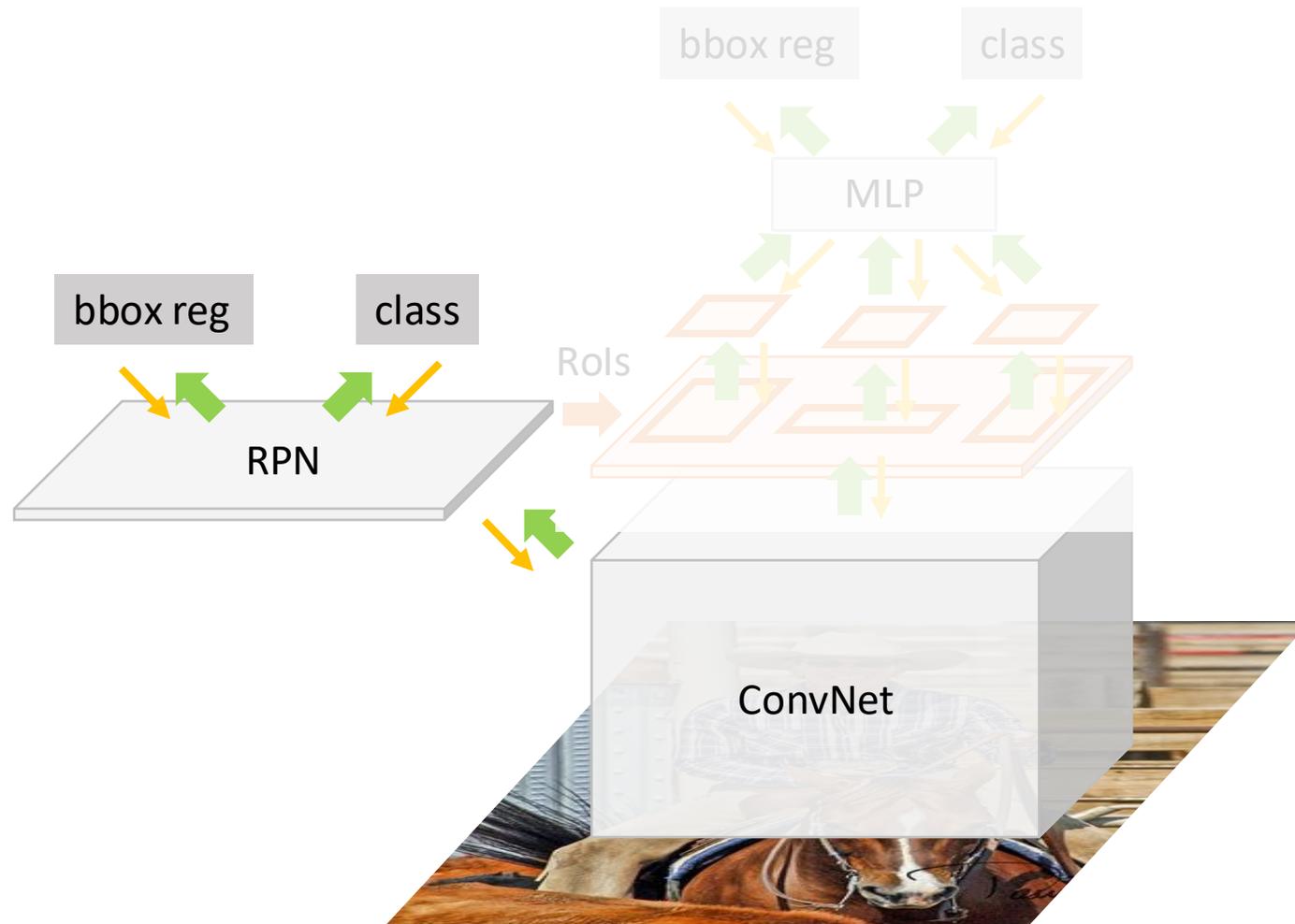
Faster R-CNN: Everything in Deep Learning

S. Ren, K. He, R. Girshick, J. Sun. NeurIPS 2015. **This Talk**



- Among first real usages of **differentiable programming**
- A conceptual shift in CV from “architectures” to “**programs**”
- Urged software to adopt better **abstractions** (e.g., PyTorch)

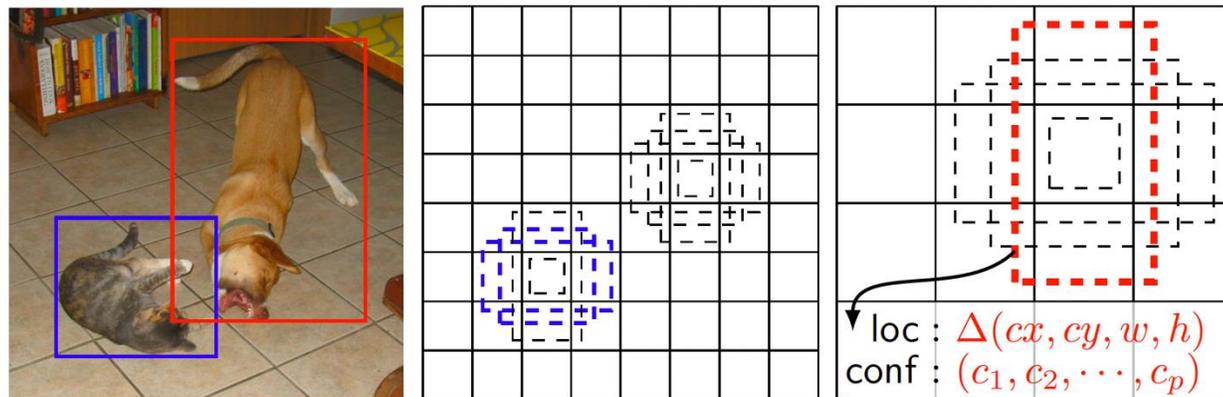
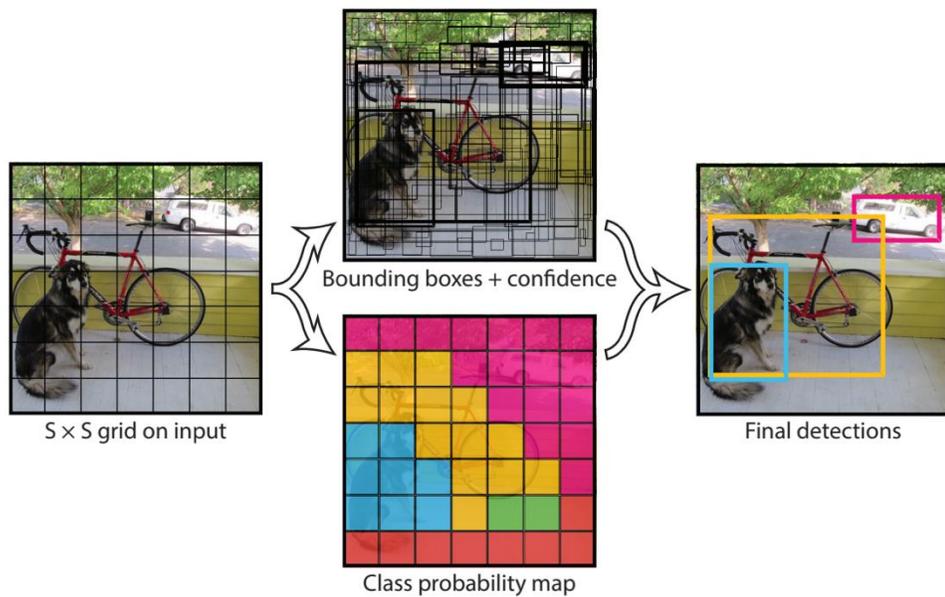
Next Chapter: One-Stage Detection



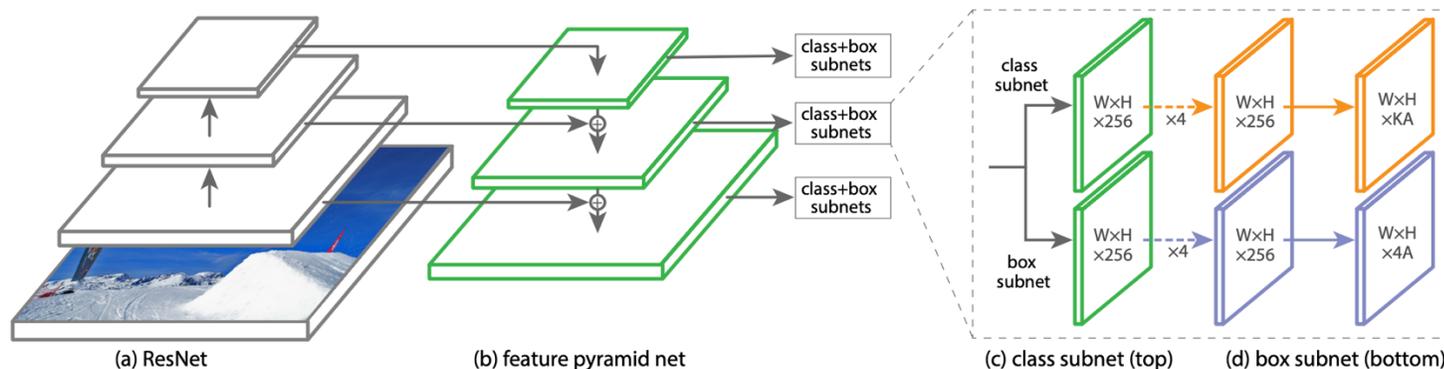
Next Chapter: One-Stage Detection

SSD, ECCV 2016

YOLO, CVPR 2016

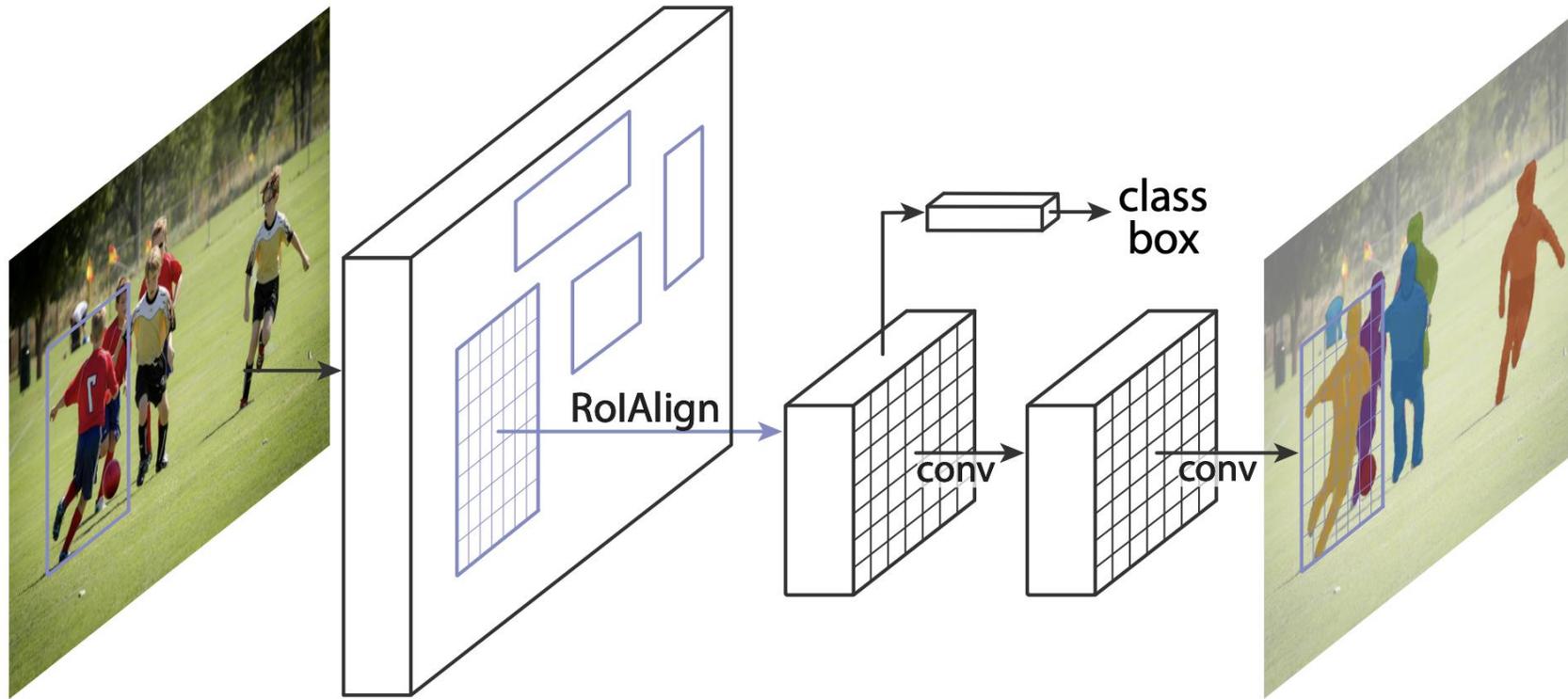


Focal Loss, ICCV 2017



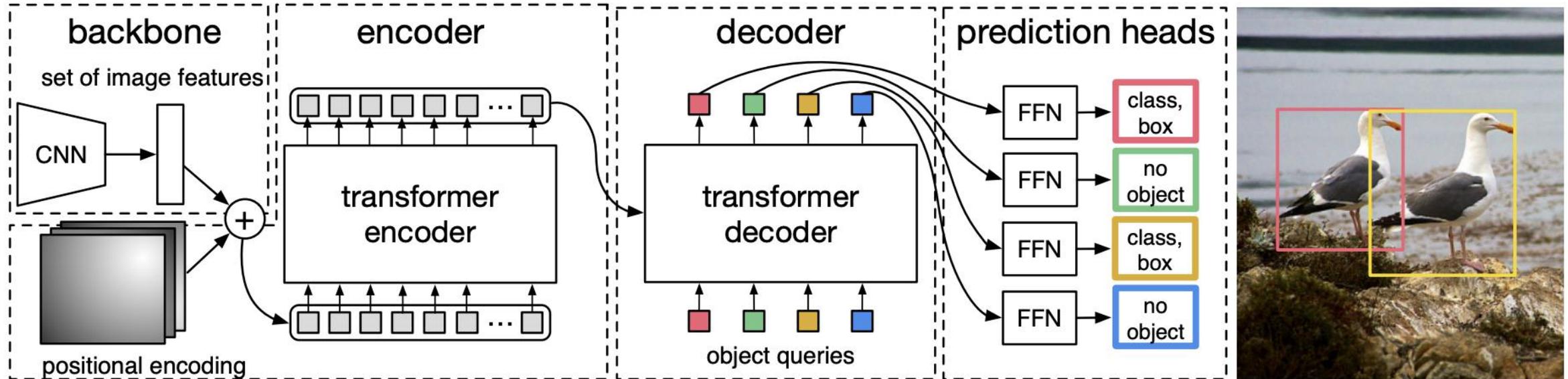
Next Chapter: Object Segmentation

Mask R-CNN, ICCV 2017



Next Chapter: Everything in Transformers

Detection Transformer (DETR) 2020



Next Chapter: Anything in the Wild

Segment Anything Model (SAM), 2023



What Have We Learned Over the Past Decades?

- ~~Write object detection papers and win Test of Time Awards :)~~

Age of Discoveries

- Into the fog
 - No oracle map
 - Is there even a destination?
-
- What's on the other side of the land?
 - And: what will it mean for us?