Recognizing Novel Objects in Novel Surroundings with Single-shot Detectors

Prof. Alexander C. Berg

May 22, 2018

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Shopagon

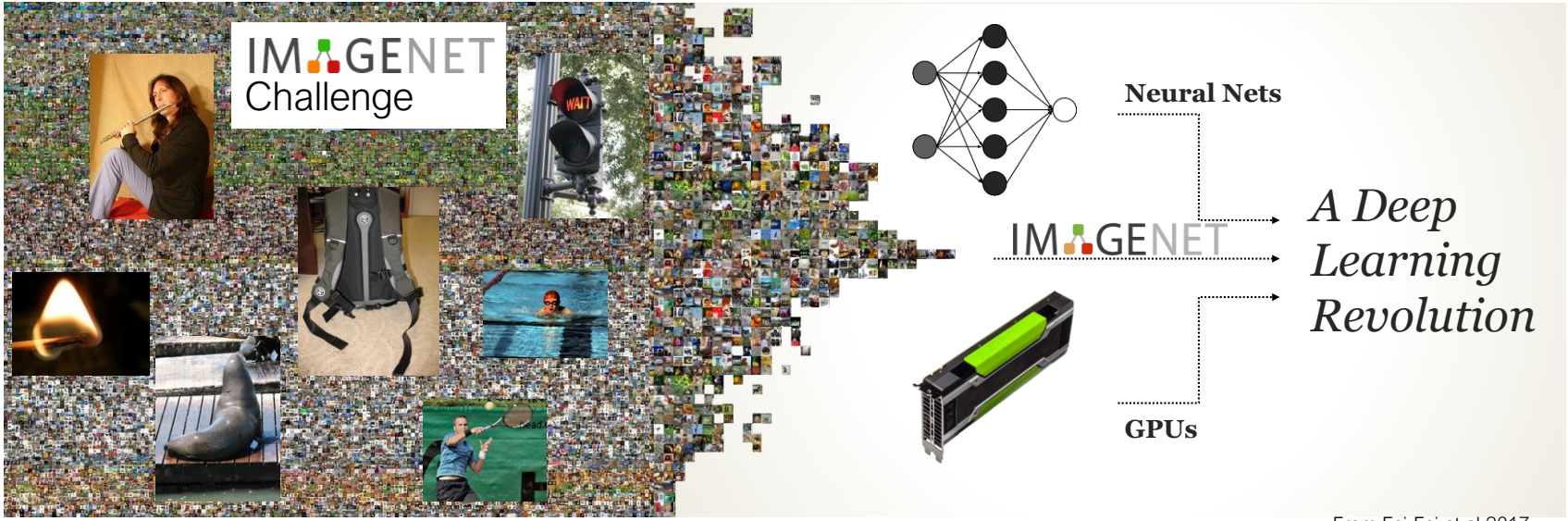
My research focuses on recognition in computer vision

My research focuses on recognition in computer vision

ImageNet Challenge
ILSVRC 2010-2017

My research focuses on recognition in computer vision

ImageNet Challenge
ILSVRC 2010-2017



From Fei-Fei et al 2017

My research focuses on recognition in computer vision

ImageNet Challenge
ILSVRC 2010-2017

Low-Power Image Recognition Challenge
LPIRC 2015-
(held at DAC and CVPR)

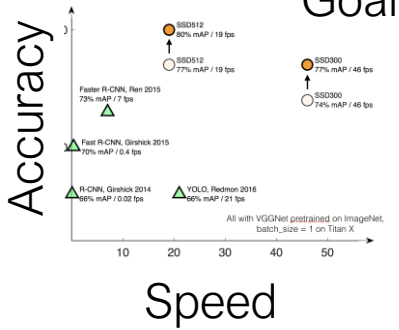
My research focuses on recognition in computer vision

ImageNet Challenge
ILSVRC 2010-2017

Low-Power Image Recognition Challenge
LPIRC 2015-
(held at DAC and CVPR)

Single-Shot Detector (SSD)
2016-

Goal!



Pushing the state-of-the-art for speed-vs-accuracy in object detection.

Connection between research and practice

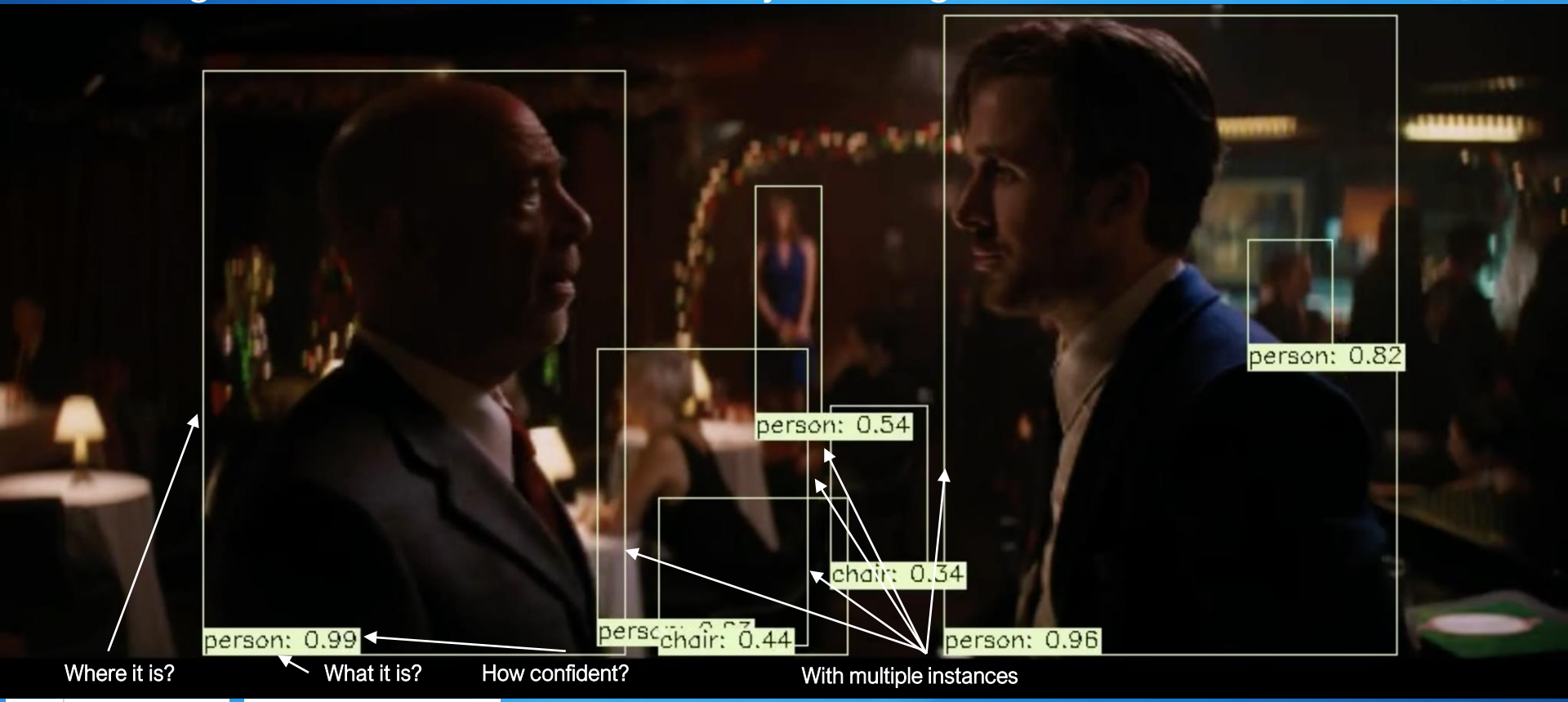
- Research focuses on learnability --- this is usually the bottleneck
- Practice can focus more on efficient evaluation (inference)
- Delay between research and deployment in systems
  - Work on improving accuracy & efficient implementations
  - Very little delay in cloud-based vision (SSD, IKSVM 1-2 months)
  - Somewhat longer delay for embedded systems
    - different power/compute regime
    - different targets, may need new training data

Hey! That's what this talk is supposed to be about!

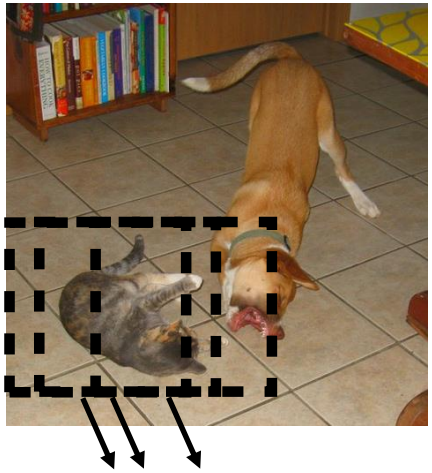
Today's talk: detectors and adapting to new targets

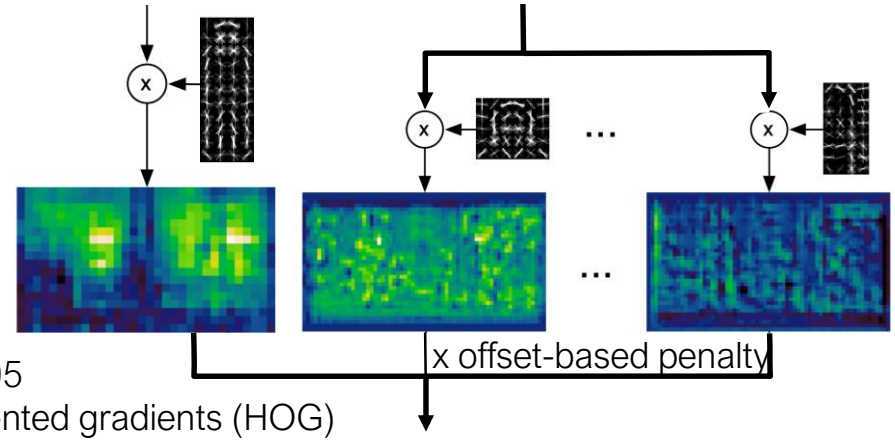
Review of how deep-learning based detectors work

Object detection
—recognition and localization of objects, e.g.



Early 2000s: Sliding windows (convolve->classify)

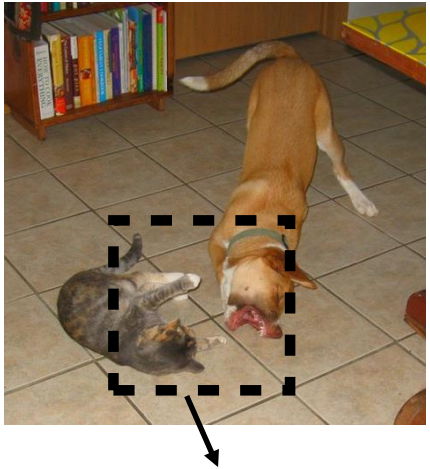




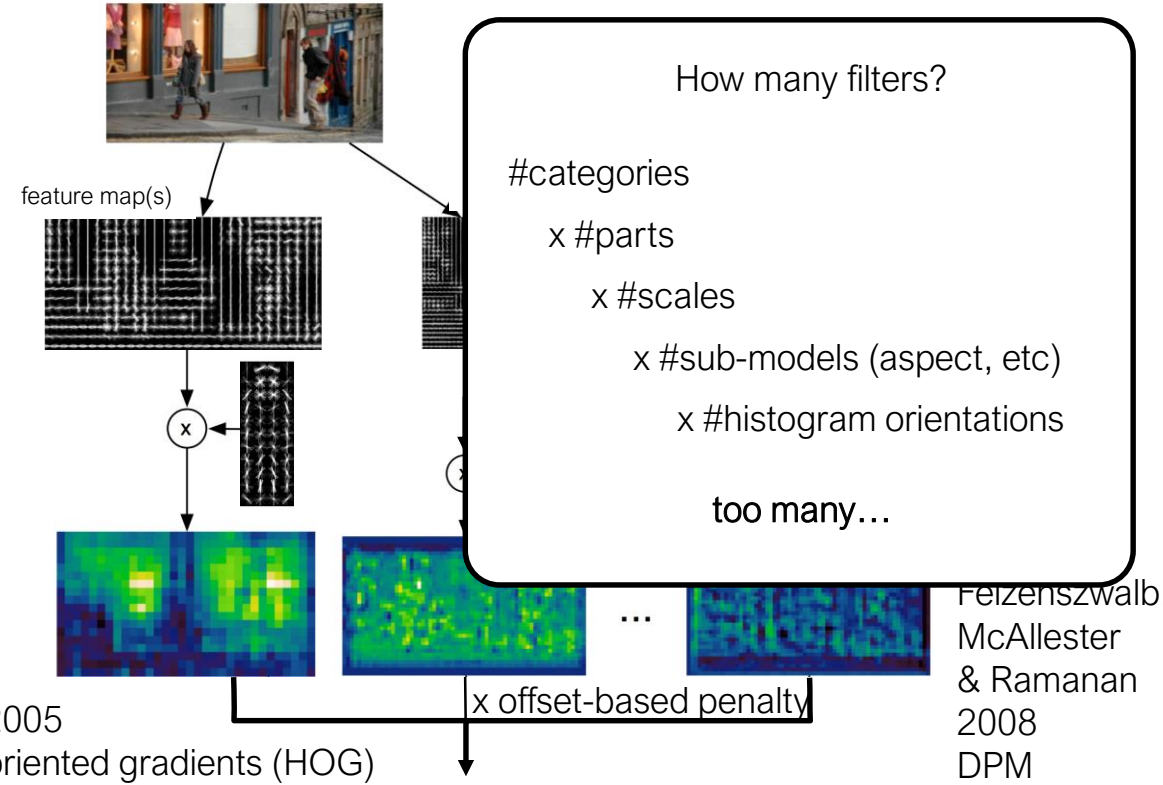
Dalal & Triggs 2005
Histograms of oriented gradients (HOG)

Felzenszwalb, McAllester & Ramanan 2008
DPM

Early 2000s: Sliding windows (convolve->classify)



Is it a cat? No



How many filters?

#categories

x #parts

x #scales

x #sub-models (aspect, etc)

x #histogram orientations

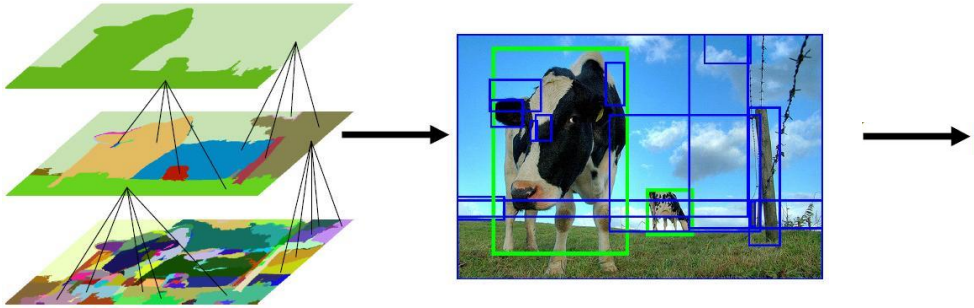
too many…

Dalal & Triggs 2005
Histograms of oriented gradients (HOG)

Felzenszwalb, McAllester & Ramanan 2008
DPM

Early 2010s: Bottom-up-segmentation-inspired ROI proposals and powerful classifiers (propose+classify)

Selective search framework



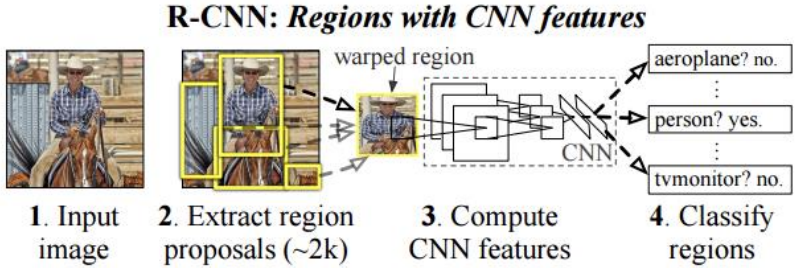
Aggregate features per box & classify

van de Sande et al. 2011

First effective detector for many object categories

(derived from much older ideas)

Perceptual organization
- figure/ground
- pop out

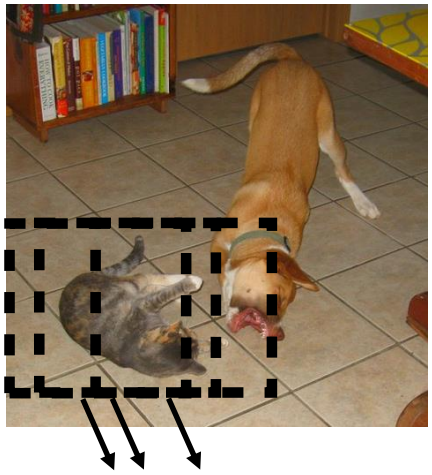


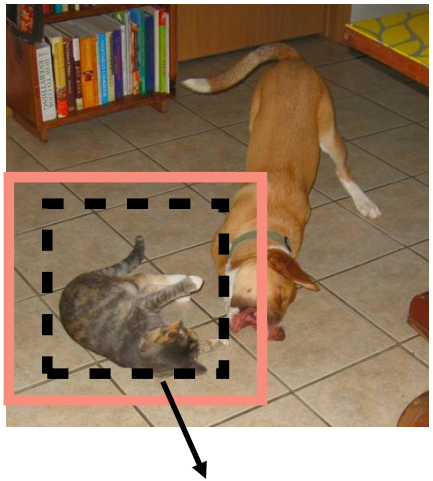
Girshick et al. 2014

Much better (deep) classifiers in the same selective search framework

Return to sliding window!

2016: Sliding windows revisited

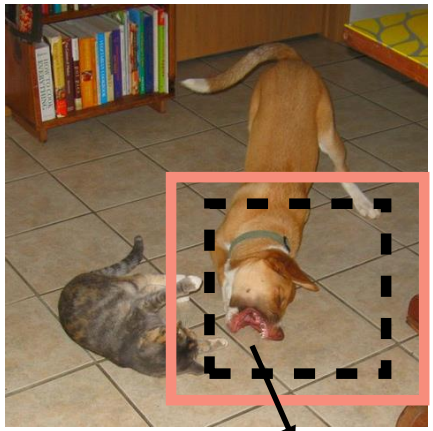




2016: Sliding windows revisited

More powerful deep-learning-based approaches allow more complex predictions, reducing computational complexity for the same accuracy.

Is it a cat? No

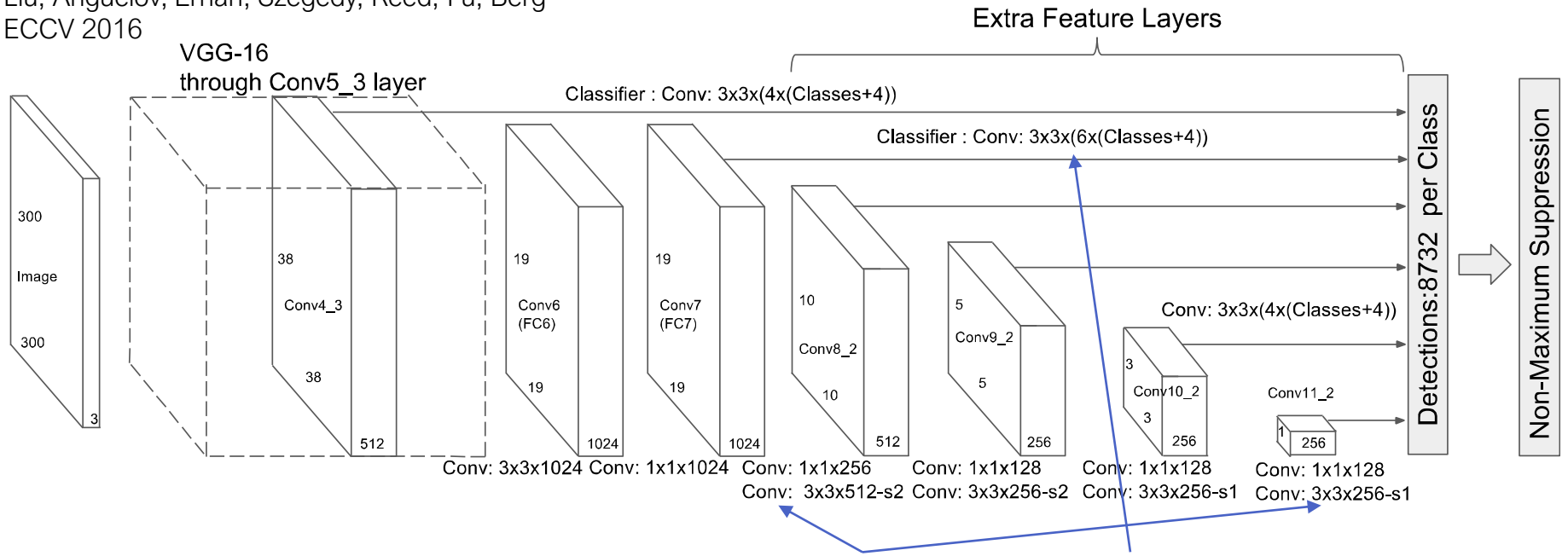


dog: 0.4 cat: 0.2

Single Shot Detector (SSD) approach

SSD: Single Shot MultiBox Detector

Liu, Anguelov, Erhan, Szegedy, Reed, Fu, Berg
ECCV 2016



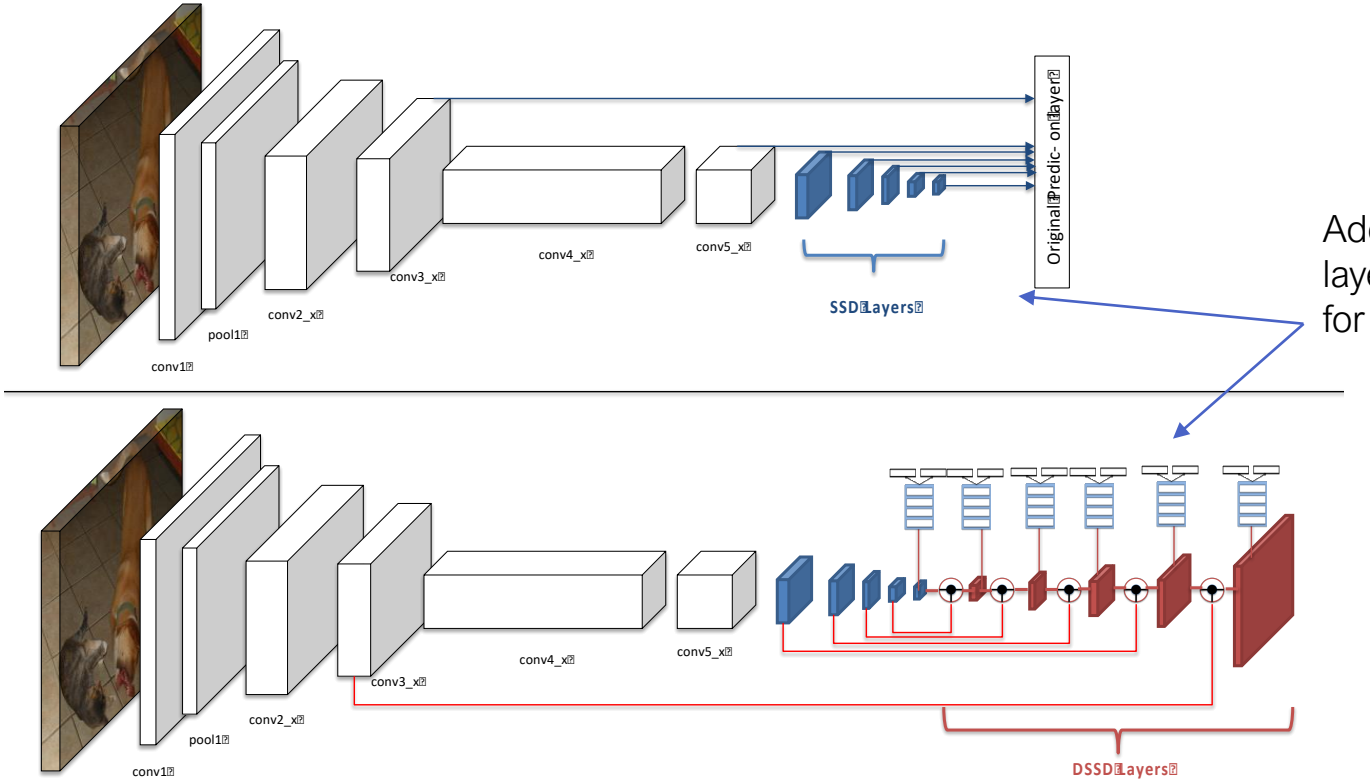
Input Image

Standard convolutional network (VGG, GoogleNet, MobileNet, etc.)

Added SSD feature layers

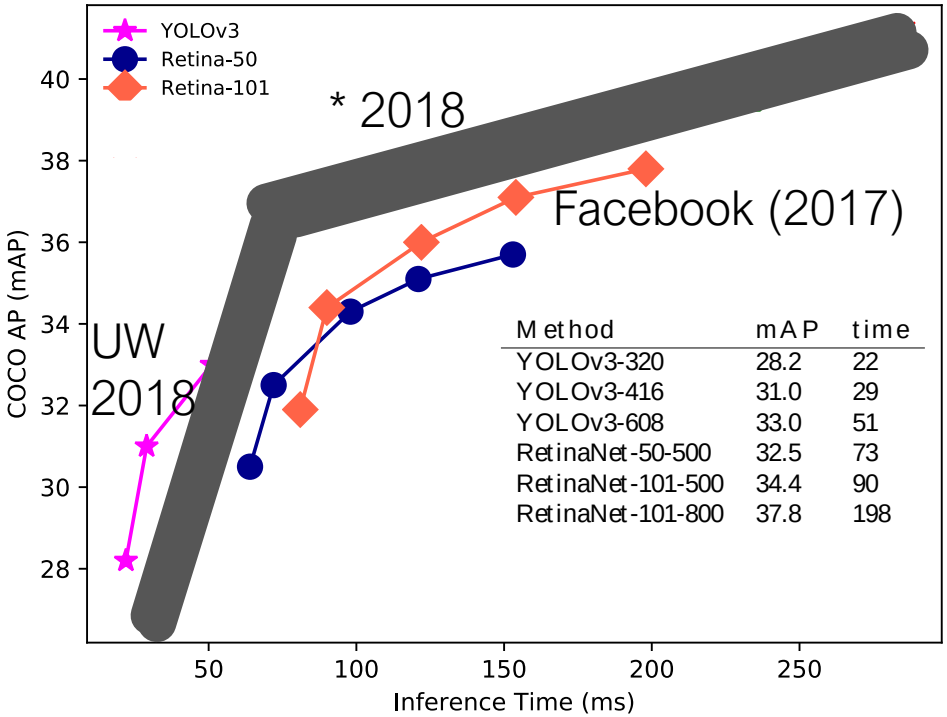
SSD uses 3x3x#maps filters for predicting detection scores and bounding box regression

Single Shot Detector (SSD) approach extended



Adding deconvolutional layers improves accuracy for small objects

Current detection accuracy vs compute time: all Single-Shot



Method	mAP	time
YOLOv3-320	28.2	22
YOLOv3-416	31.0	29
YOLOv3-608	33.0	51
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	37.8	198

Take home message:

Single Shot Detectors are doing well for general object detection with large amounts of training data.

What about specific instances or novel objects?

Training detectors on instances or novel objects

Approaches for training detectors with few examples

- Make more examples
  - Computer graphics
  - Composite images
  - Find more examples
- Train a detector with few examples
  - Similarity-based detection
    - Similar to tracking

Approaches for training detectors with few examples

- Make more examples
  - Computer graphics
  - Composite images
  - Find more examples
- Train a detector with few examples
  - Similarity-based detection
    - Similar to tracking

Playing for Benchmarks
Richter, Hayder, Koltun
International Conference on Computer Vision (ICCV) 2017

Approaches for training detectors with few examples

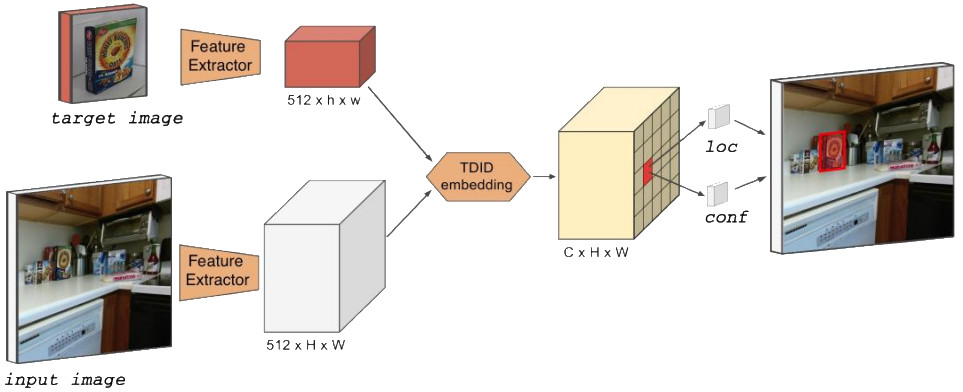
- Make more examples
  - Computer graphics
  - Composite images
  - Find more examples
- Train a detector with few examples
  - Similarity-based detection
    - Similar to tracking…

Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection
Dwibedi, Misra, Hebert
International Conference on Computer Vision (ICCV) 2017

Synthesizing Training Data for Object Detection in Indoor Scenes
Georgakis, Mousavian, Berg, Kosecka
Robotics Science and Systems (RSS) 2017

Approaches for training detectors with few examples

- Make more examples
  - Computer graphics
  - Composite images
  - Find more examples
- Train a detector with few examples
  - Similarity-based detection →
    - Similar to tracking



Target Driven Instance Detection
Ammirato, Fu, Shvets, Kosecka, Berg
2018

Approaches for training detectors with few examples

- Make more examples
  - Computer graphics
  - Composite images
  - Find more examples
- Train a detector with few examples
  - Similarity-based detection
    - Similar to tracking…

Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection
Dwibedi, Misra, Hebert
International Conference on Computer Vision (ICCV) 2017

Synthesizing Training Data for Object Detection in Indoor Scenes
Georgakis, Mousavian, Berg, Kosecka
Robotics Science and Systems (RSS) 2017

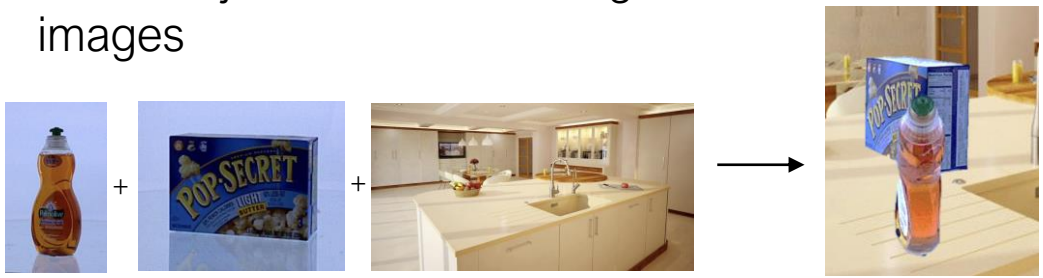
Composing new example training images

BigBIRD
(Berkeley Instance Recognition Dataset)
Singh, Sha, Narayan, Achim, Abbeel
ICRA 2014

Clean object images

Scene images

Composite training image



Can significantly reduce over-fitting in training by using multiple compositing approaches!

Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection
Dwibedi, Misra, Hebert
International Conference on Computer Vision (ICCV) 2017

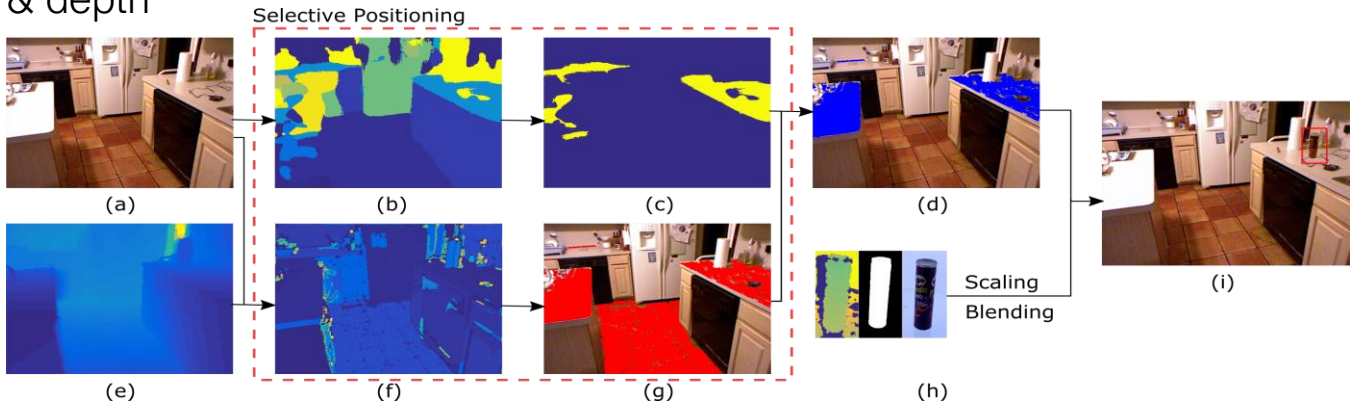
Composing new example training images

BigBIRD
(Berkeley Instance Recognition Dataset)
Singh, Sha, Narayan, Achim, Abbeel
ICRA 2014

Scene image & depth

Scene understanding

More useful compositing

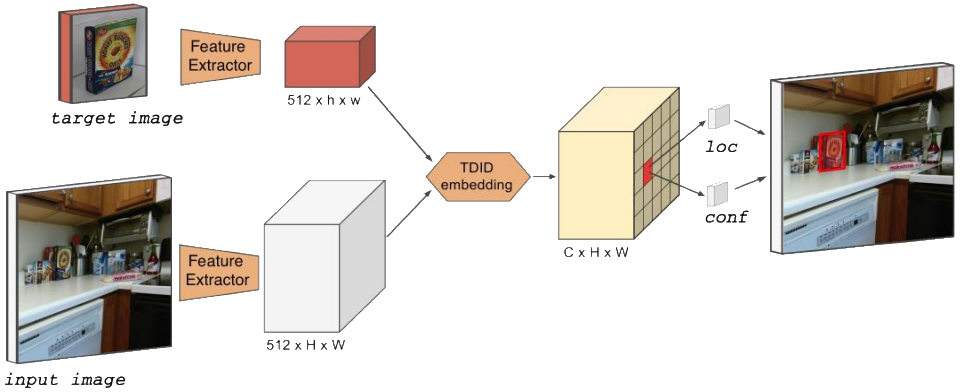


It helps to understand scene layout and better place objects in training…

Synthesizing Training Data for Object Detection in Indoor Scenes
Georgakis, Mousavian, Berg, Kosecka
Robotics Science and Systems (RSS) 2017

Approaches for training detectors with few examples

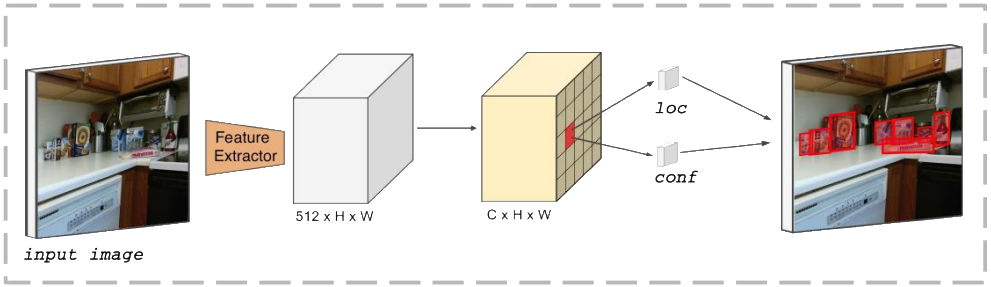
- Make more examples
  - Computer graphics
  - Composite images
  - Find more examples
- Train a detector with few examples
  - Similarity-based detection ⟶
    - Similar to tracking



Target Driven Instance Detection
Ammirato, Fu, Shvets, Kosecka, Berg
2018

Approaches for training detectors with few examples

ce Detection
:s, Kosecka, Berg



Faster R-CNN region proposal

To be refined with second-stage classifier

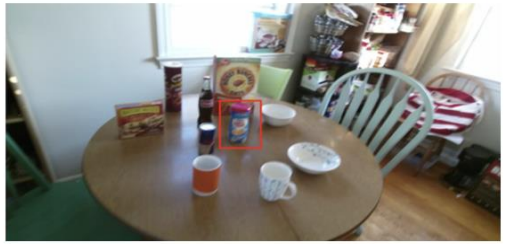
Approaches for training detectors with few examples

Target Driven Instance Detection
Ammirato, Fu, Shvets, Kosecka, Berg
arXiv 2018

Target

Image

Previously unseen



Agnostic region proposal map

Target-driven region proposal map

Thanks to all my co-authors

Recent object synthesis, detection, and robotics bunch

Older Crowd

Broadening the focus of work on object detection:

+ Speed-vs-accuracy
+ Instances
+ Few training examples
+ Outputs for applications:
  +Pose +Masks +Fiducials

References and papers can be found in the slides above and in the papers here: http://acberg.com

LPRIC: http://rebootingcomputing.ieee.org/lpirc