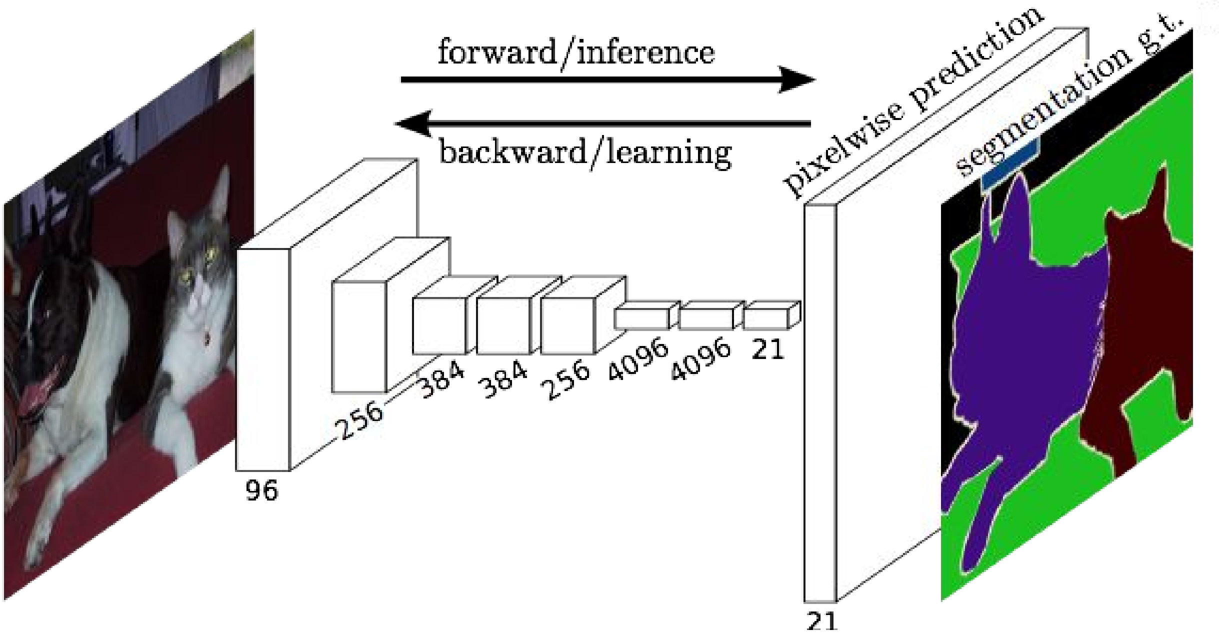
Fully Convolutional Networks for Semantic Segmentation

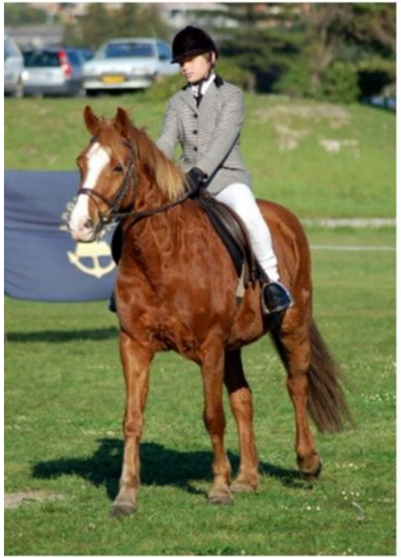


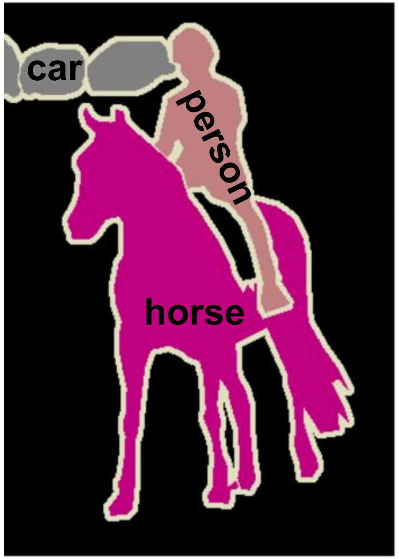
Jonathan Long* Evan Shelhamer* Trevor Darrell

UC Berkeley

Semantic Segmentation

- what kind of thing is each pixel part of?
- what kind of stuff is each pixel?



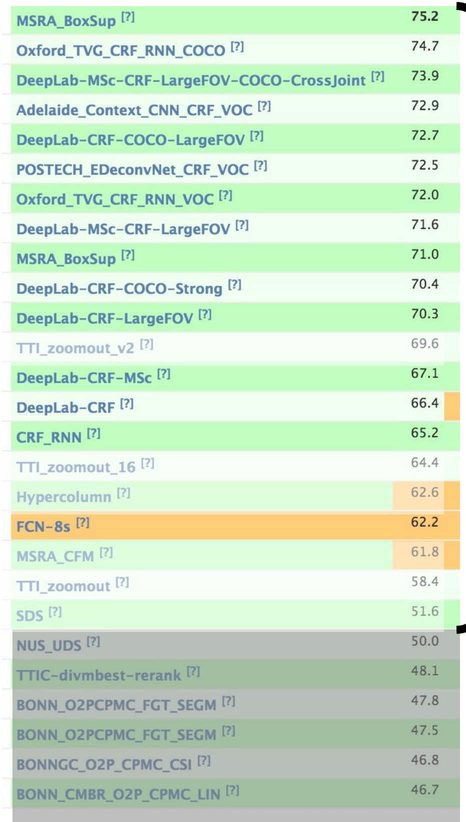


Challenges

- tension between recognition and localization
- amount of computation

Segmentation: PASCAL VOC

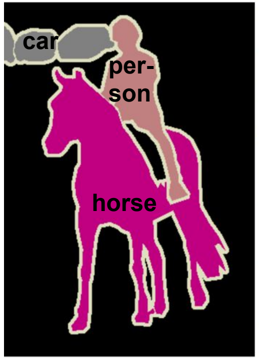
Leaderboard



MSRA_BoxSup [?]	75.2
Oxford_TVG_CRF_RNN_COCO [?]	74.7
DeepLab-MSc-CRF-LargeFOV-COCO-CrossJoint [?]	73.9
Adelaide_Context_CNN_CRF_VOC [?]	72.9
DeepLab-CRF-COCO-LargeFOV [?]	72.7
POSTECH_EDeconvNet_CRF_VOC [?]	72.5
Oxford_TVG_CRF_RNN_VOC [?]	72.0
DeepLab-MSc-CRF-LargeFOV [?]	71.6
MSRA_BoxSup [?]	71.0
DeepLab-CRF-COCO-Strong [?]	70.4
DeepLab-CRF-LargeFOV [?]	70.3
TTI_zoomout_v2 [?]	69.6
DeepLab-CRF-MSc [?]	67.1
DeepLab-CRF [?]	66.4
CRF_RNN [?]	65.2
TTI_zoomout_16 [?]	64.4
Hypercolumn [?]	62.6
FCN-8s [?]	62.2
MSRA_CFM [?]	61.8
TTI_zoomout [?]	58.4
SDS [?]	51.6
NUS_UDS [?]	50.0
TTIC-divmbest-rerank [?]	48.1
BONN_O2PCPMC_FGT_SEGM [?]	47.8
BONN_O2PCPMC_FGT_SEGM [?]	47.5
BONNGC_O2P_CPMC_CSI [?]	46.8
BONN_CMBR_O2P_CPMC_LIN [?]	46.7

} deep learning with Caffe

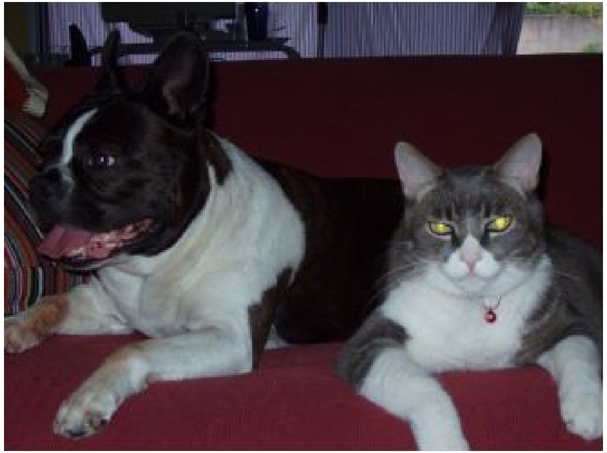
end-to-end networks lead to
50% relative improvement or 30 points absolute
and >100x speedup in 1 year!

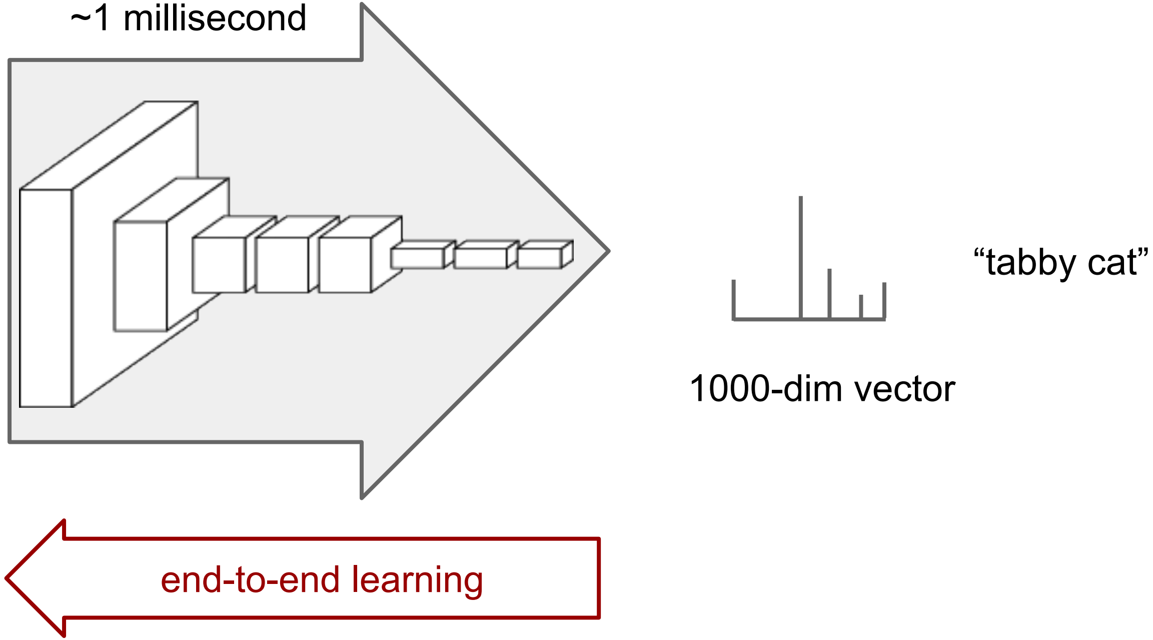


FCN:
pixelwise
convnet

state-of-the-art,
in Caffe

convnets perform classification





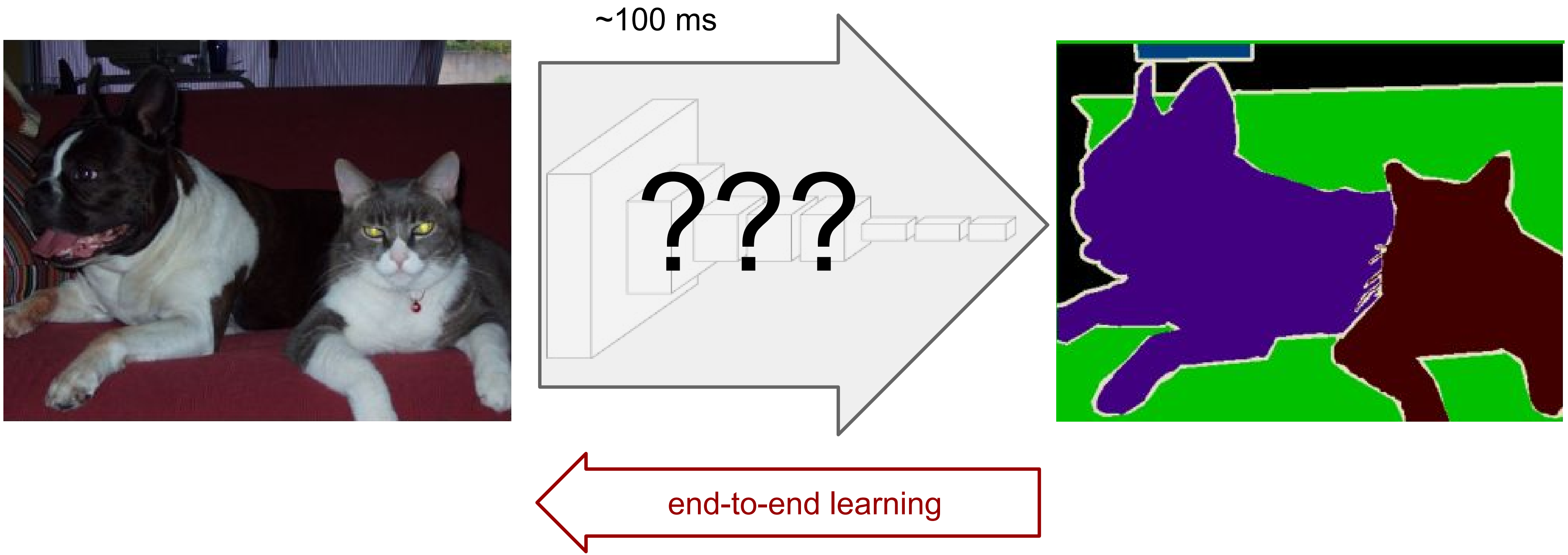
~1 millisecond

1000-dim vector

“tabby cat”

end-to-end learning

convnets perform segmentation?

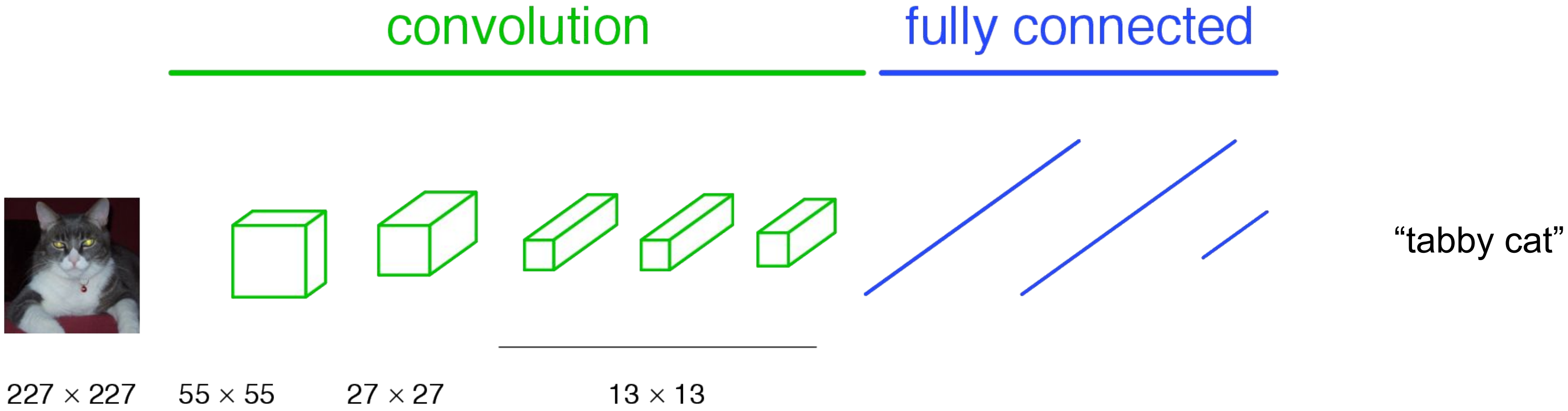


~100 ms

???

end-to-end learning

a classification network



convolution

fully connected

“tabby cat”

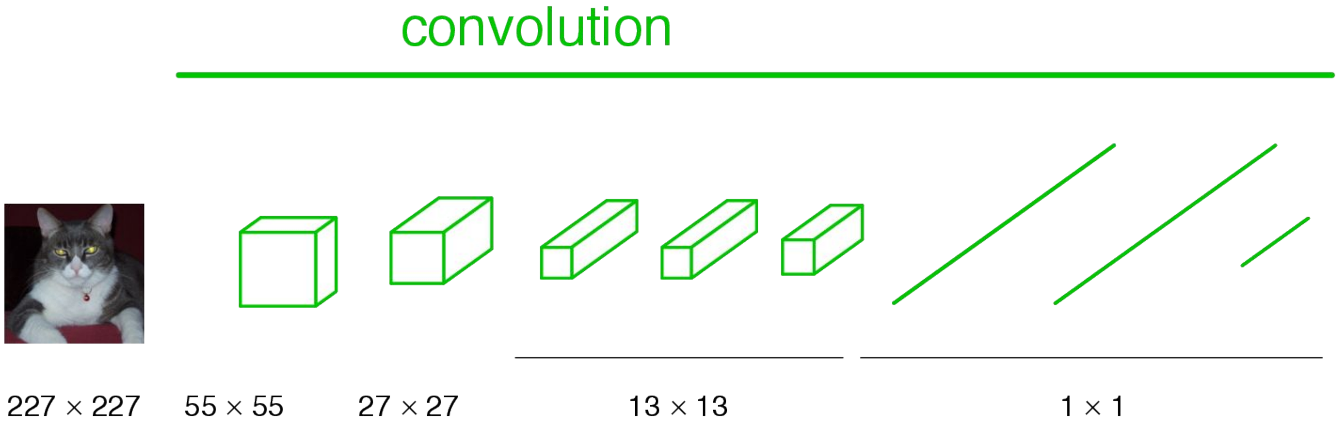
227 × 227

55 × 55

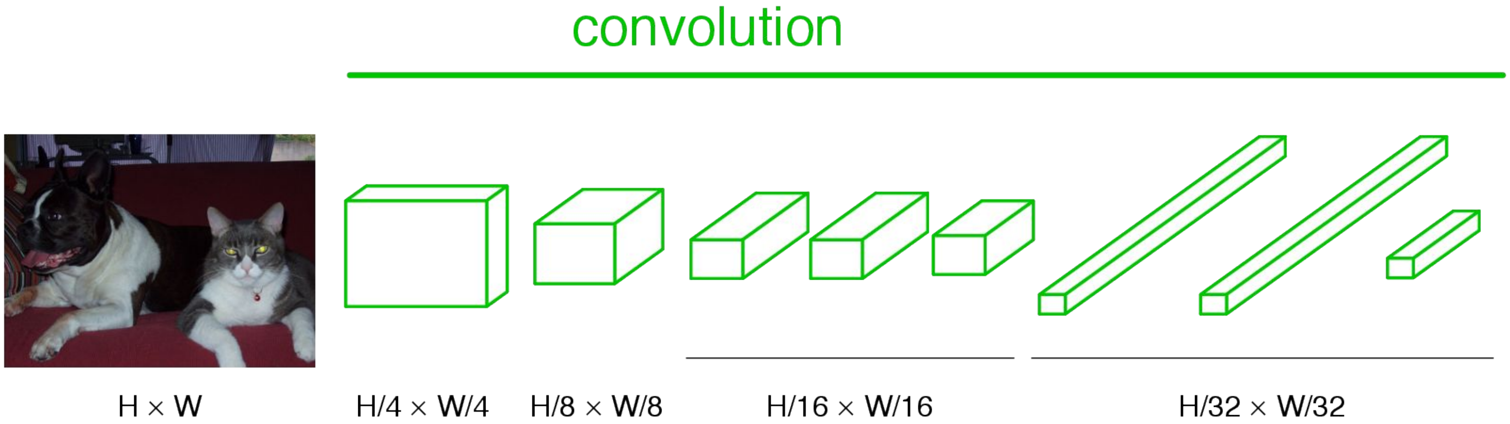
27 × 27

13 × 13

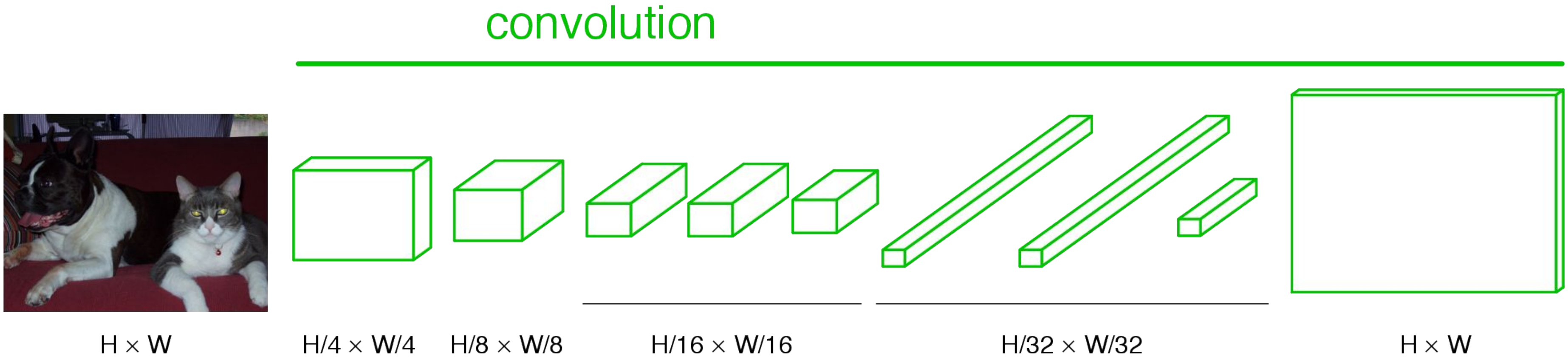
becoming fully convolutional



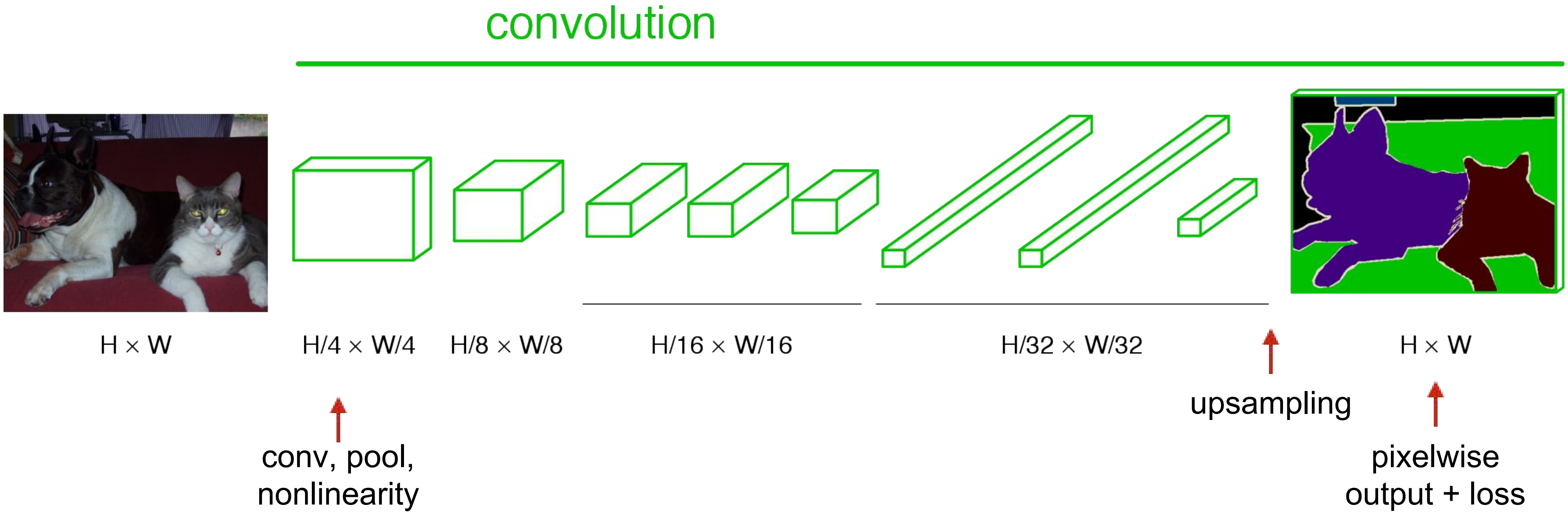
becoming fully convolutional



upsampling output



end-to-end, pixels-to-pixels network



convolution

H × W

H/4 × W/4

H/8 × W/8

H/16 × W/16

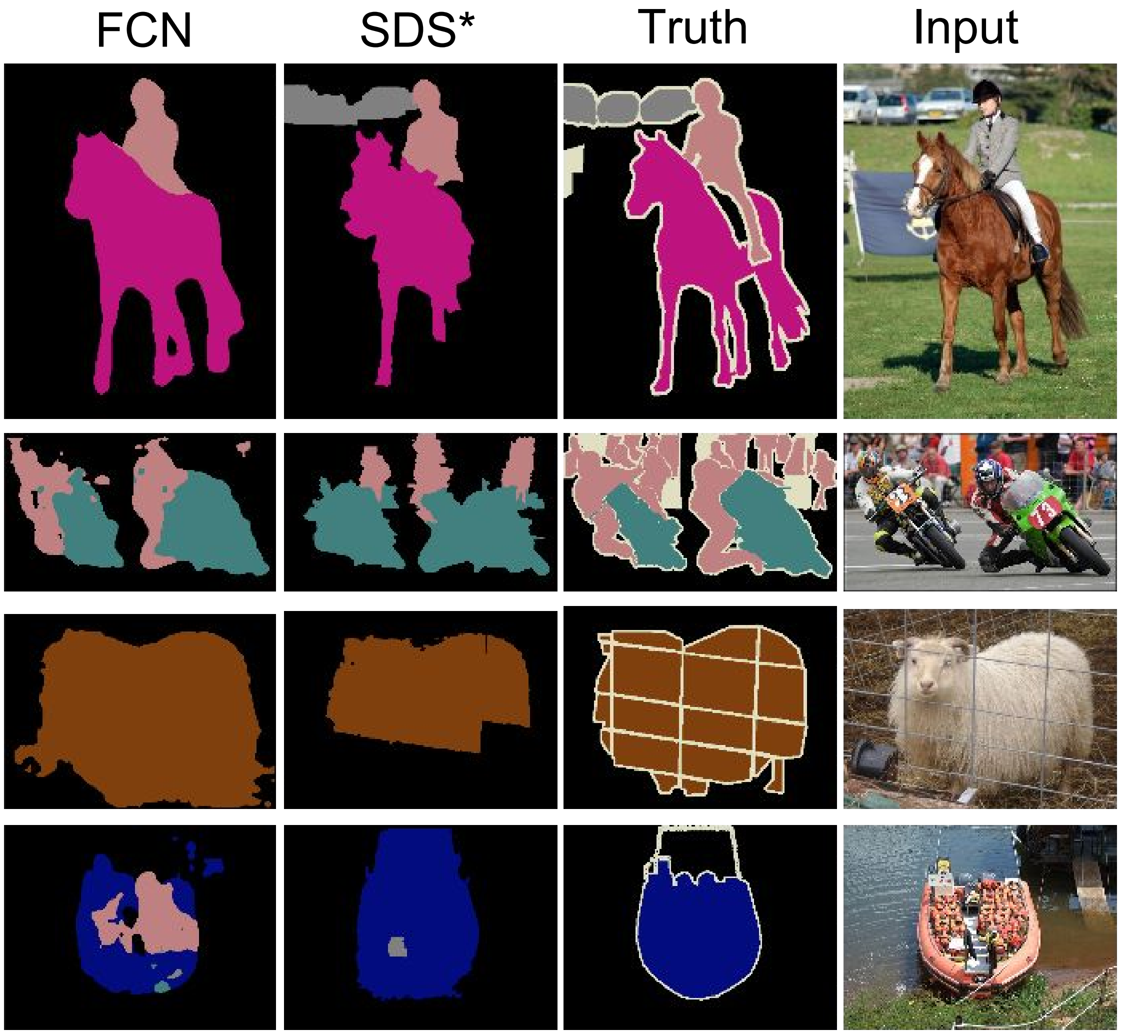
H/32 × W/32

H × W

conv, pool,
nonlinearity

upsampling

pixelwise
output + loss



Relative to prior state-of-the-art SDS:

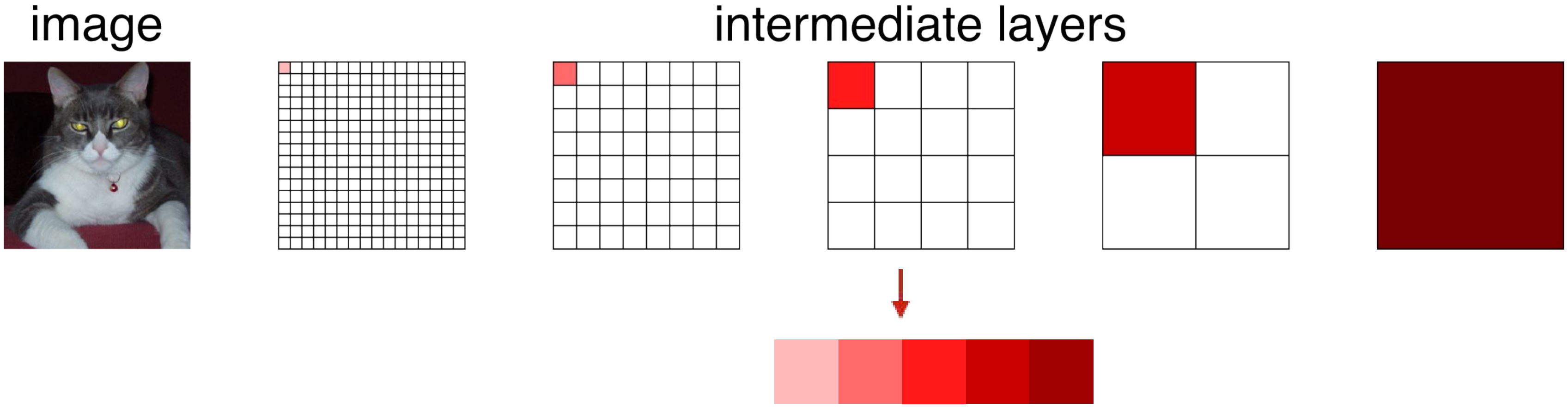
- 30% relative improvement in accuracy (67.2% on VOC 2012)

- 286× faster

*Simultaneous Detection and Segmentation
Hariharan et al. ECCV14

spectrum of deep features

combine *where* (local, shallow) with *what* (global, deep)



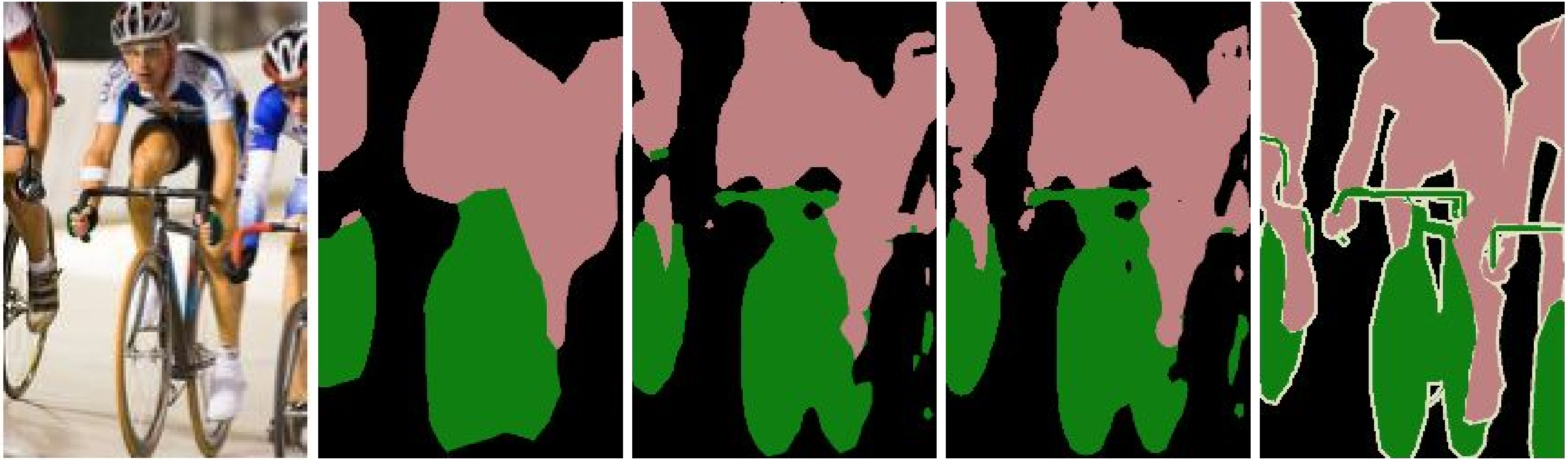
image

intermediate layers

fuse features into **deep jet**

(cf. Hariharan et al. CVPR15 “hypercolumn”)

skip layer refinement



input image

stride 32

stride 16

stride 8

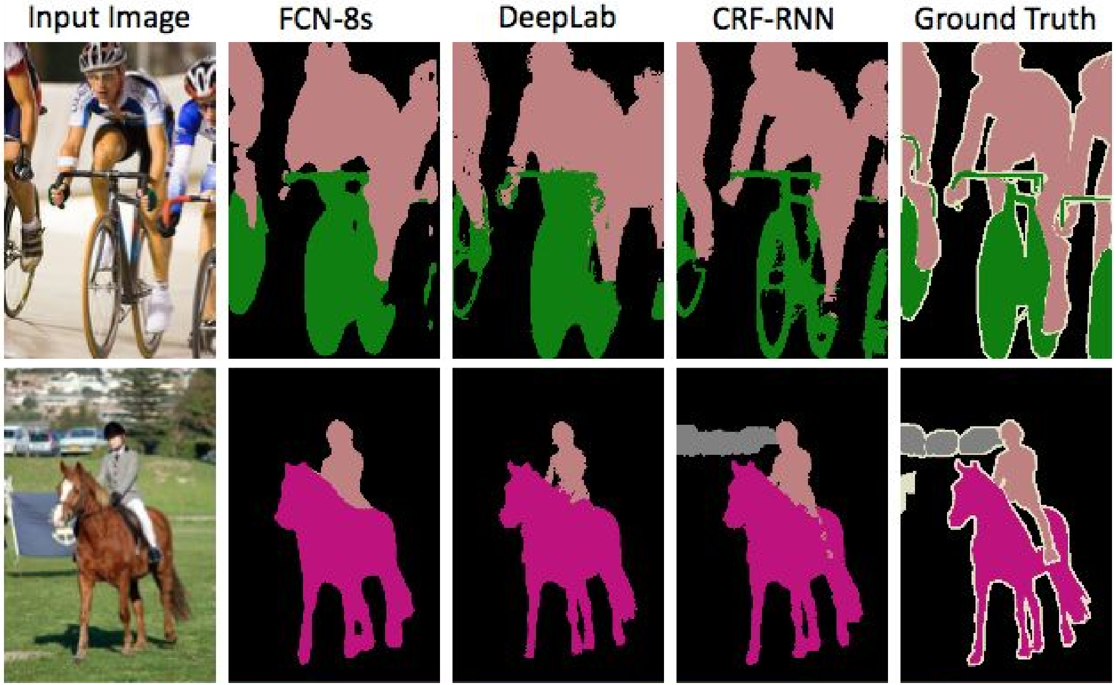
ground truth

no skips

1 skip

2 skips

graphical model refinement

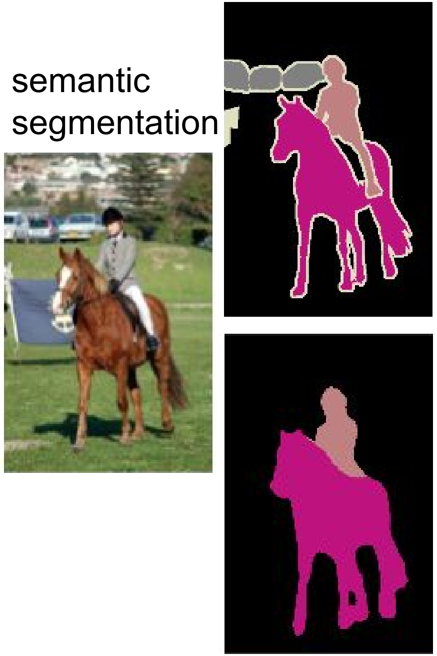


[comparison credit: CRF as RNN, Zheng* & Jayasumana* et al. ICCV 2015]

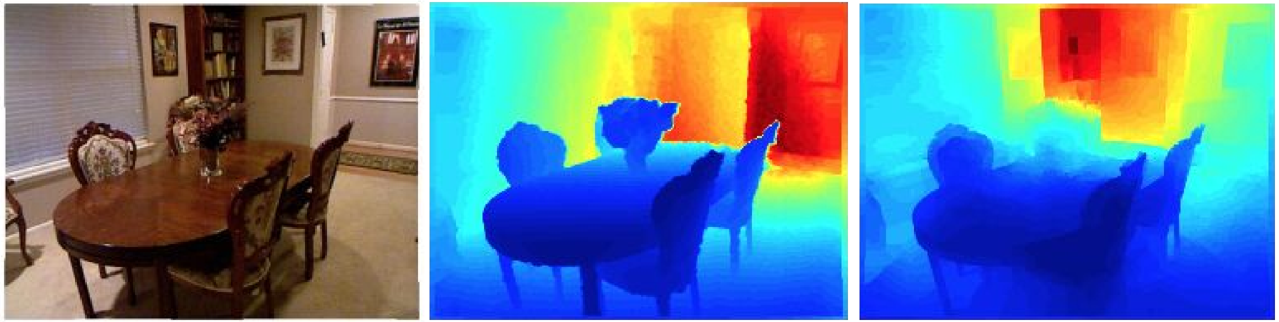
DeepLab: Chen* & Papandreou* et al. ICLR 2015.   **CRF-RNN**: Zheng* & Jayasumana* et al. ICCV 2015

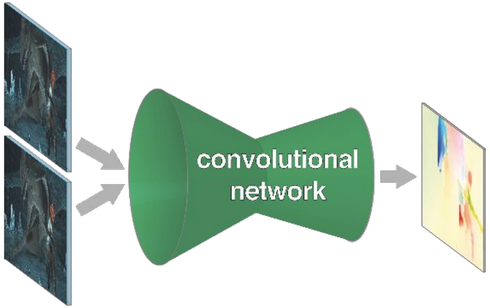
nets for many pixelwise tasks

semantic segmentation

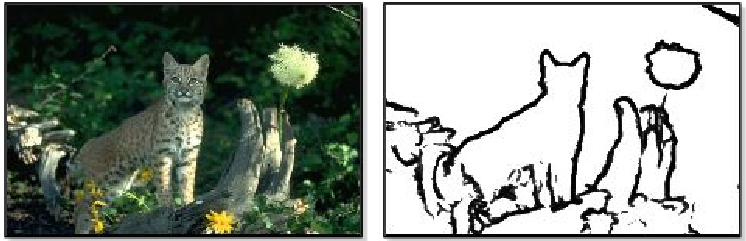


monocular depth estimation (Eigen & Fergus 2015)





optical flow Fischer et al. 2015

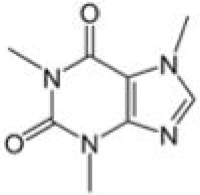


boundary prediction (Xie & Tu 2015)

conclusion

fully convolutional networks are fast, end-to-end models for pixelwise problems

- **code** in Caffe master
- **models** for PASCAL VOC, NYUDv2, SIFT Flow, PASCAL-Context



caffe.berkeleyvision.org

github.com/BVLC/caffe

fcn.berkeleyvision.org

model example
inference example
solving example