Fully Convolutional Networks for Semantic Segmentation

JONATHAN LONG, EVAN SHELHAMER, TREVOR DARRELL

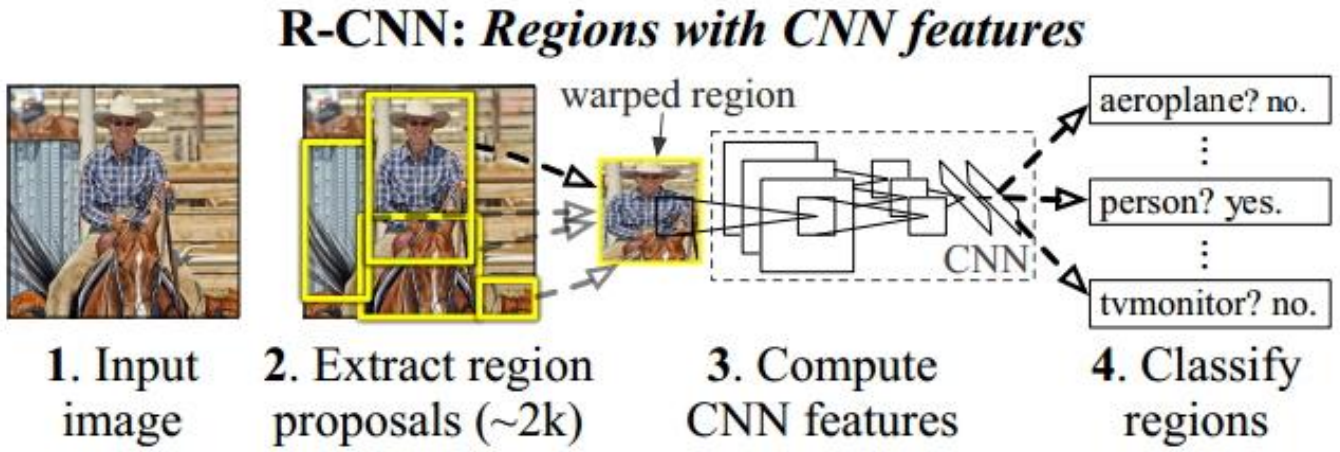
CHANDANA AMANCHI, AMIN ANVARI

Background

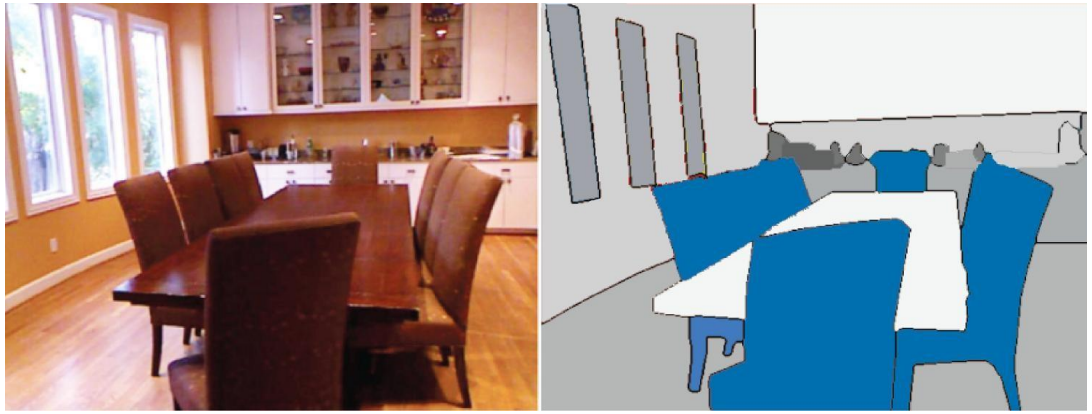
(a) Siberian husky (b) Eskimo dog

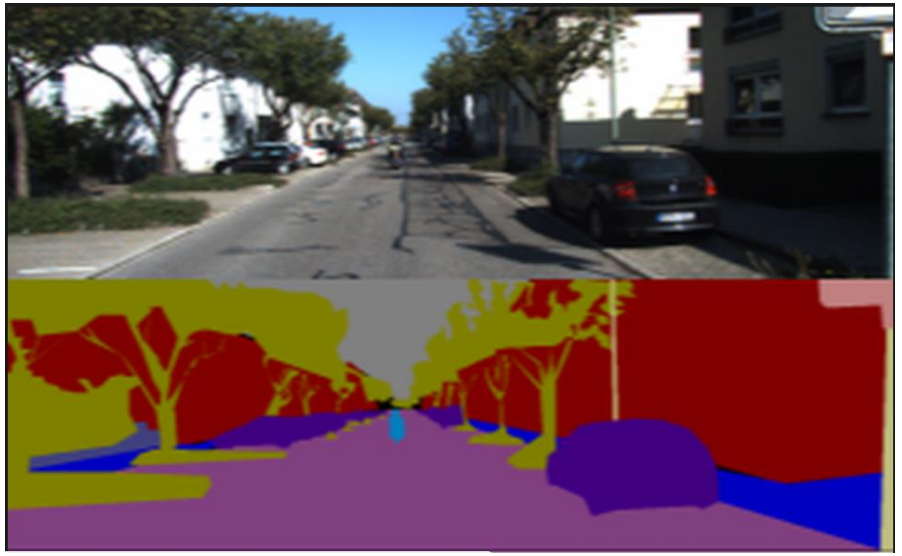
Classification

Bounding box object detection

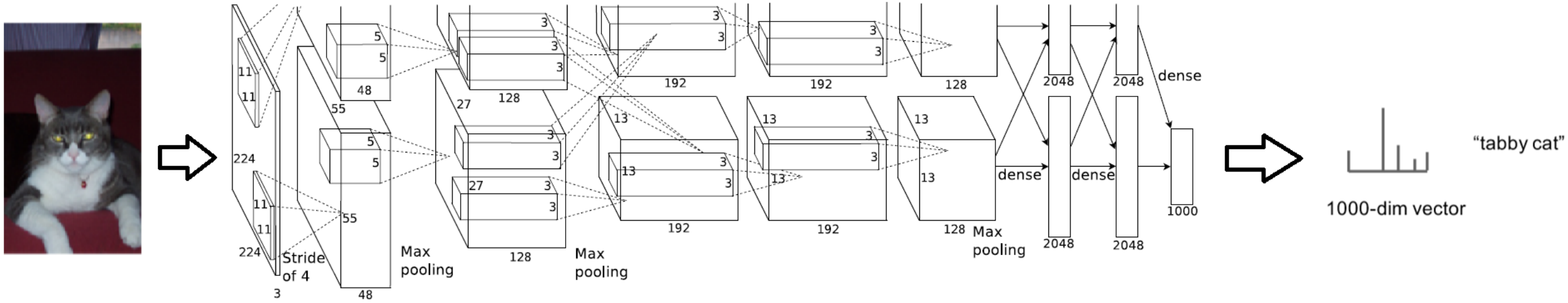


Semantic Segmentation

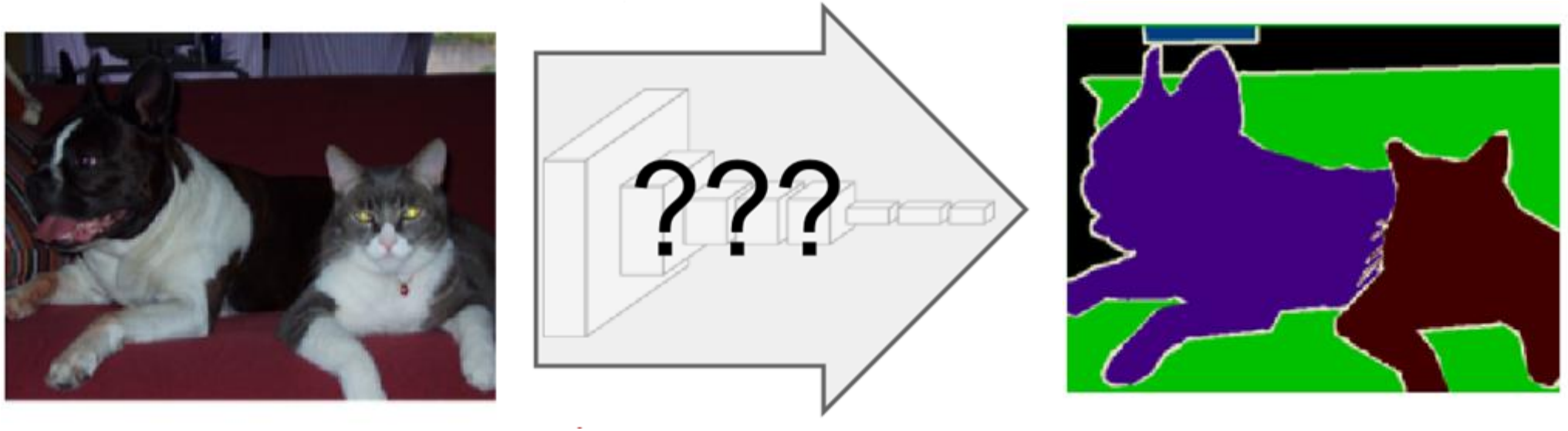




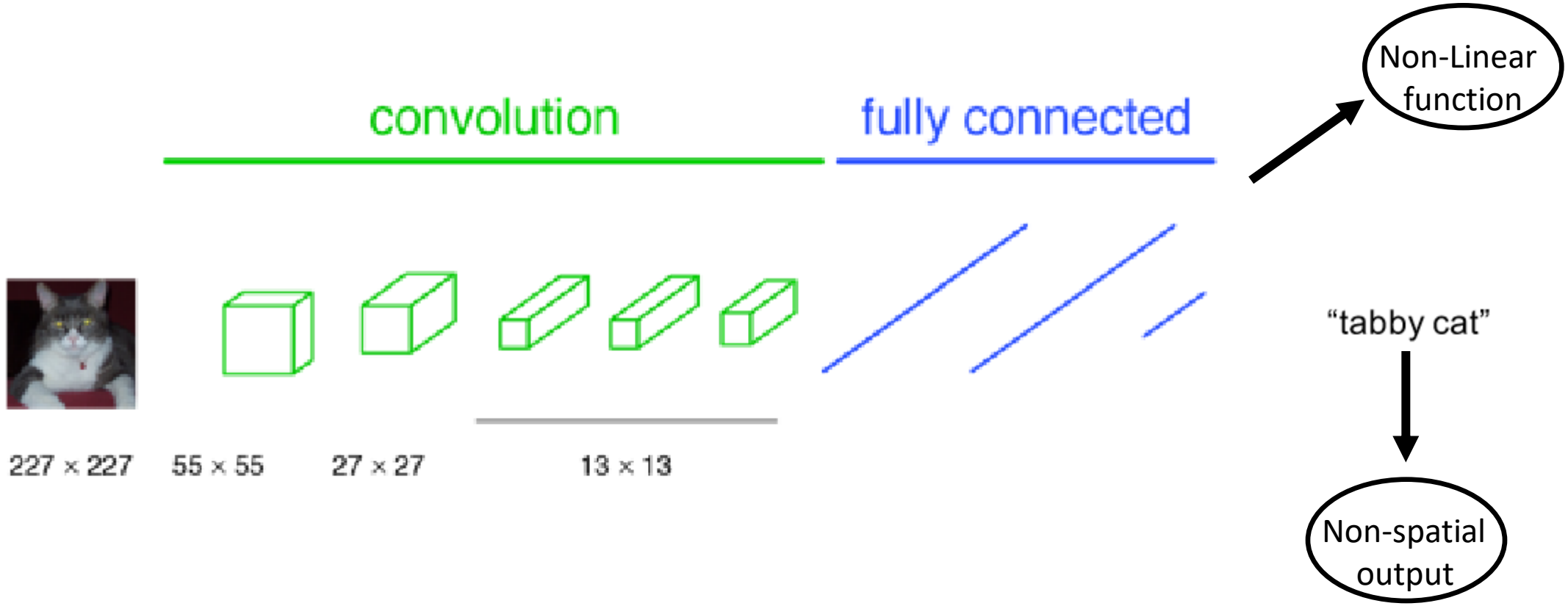
Classification networks



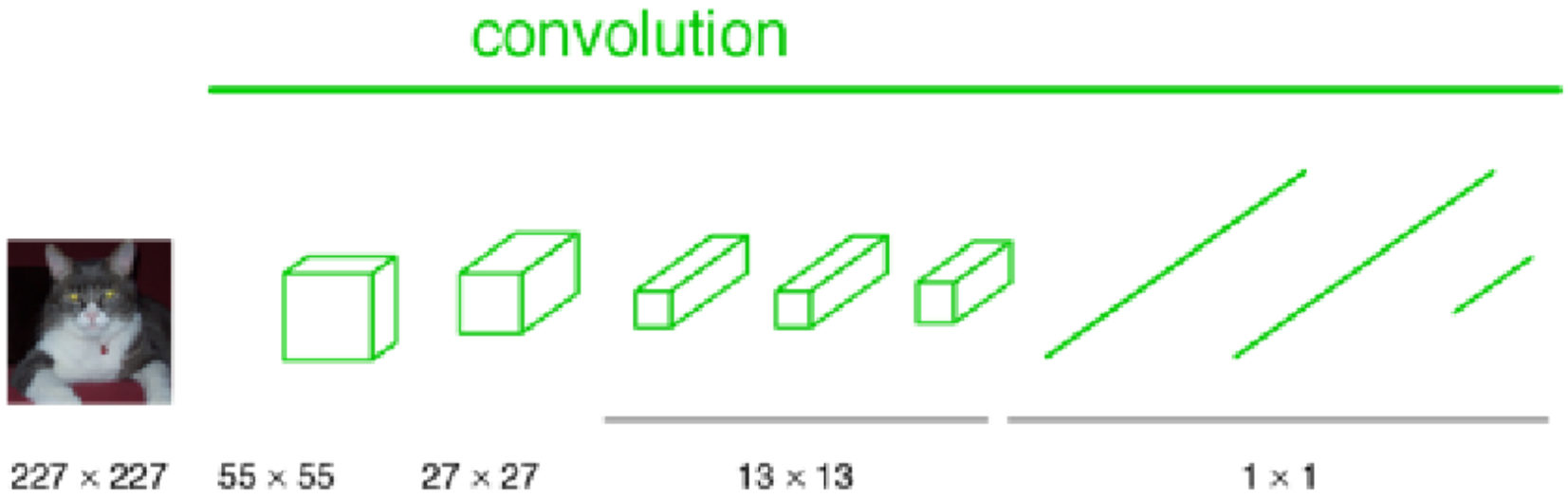
Network for Semantic Segmentation?



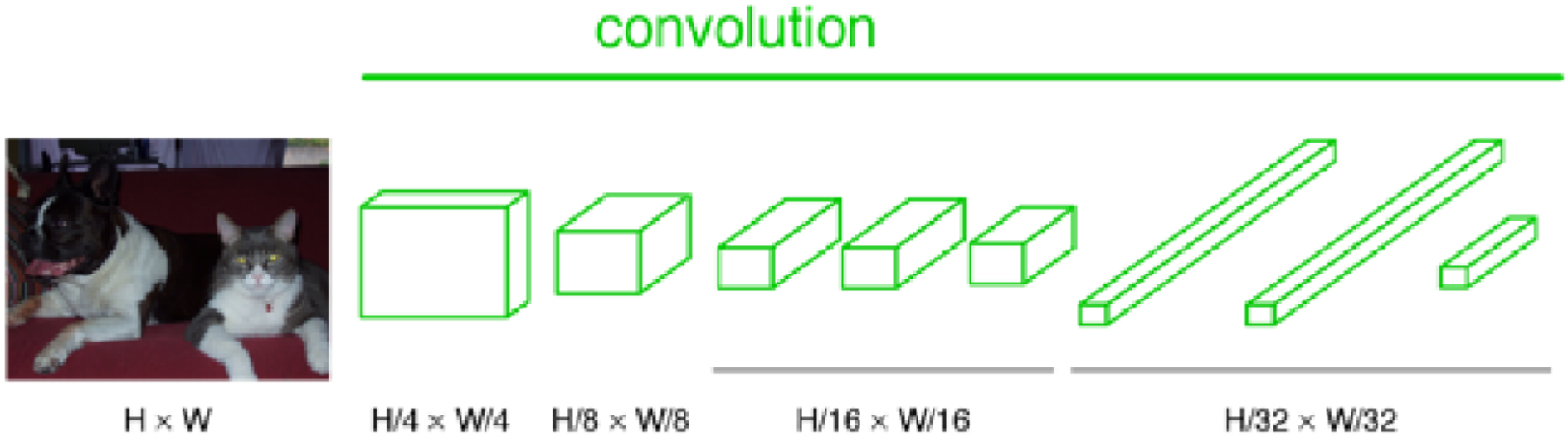
Classification Networks



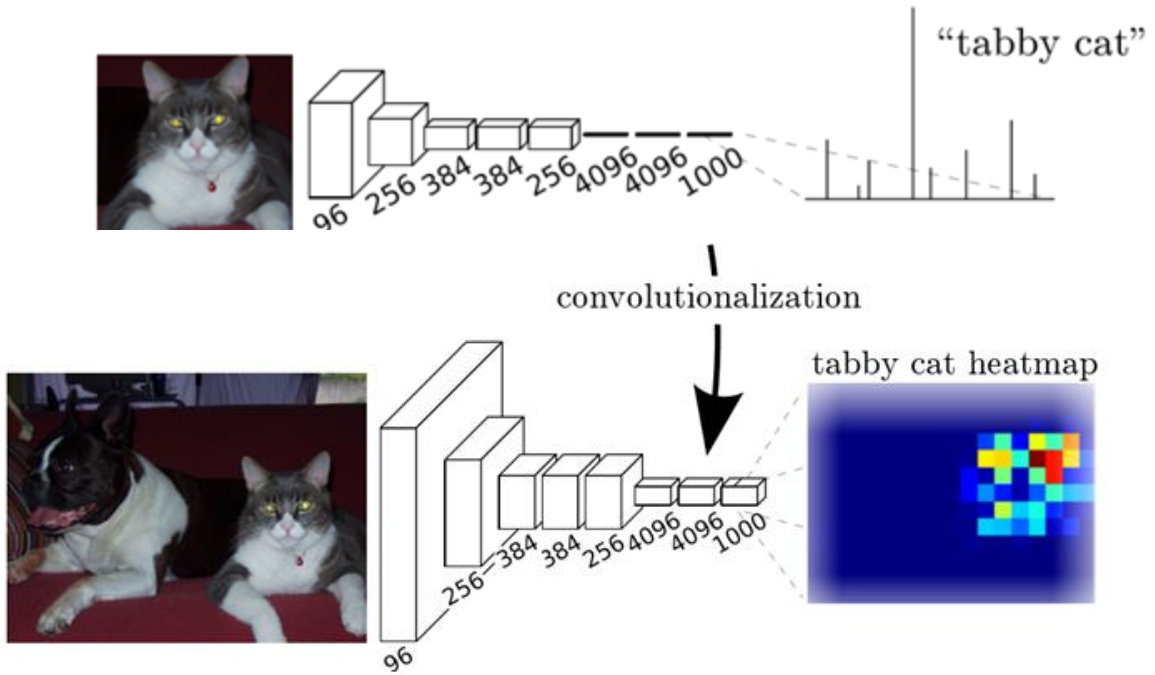
Classification -> Full Convolution



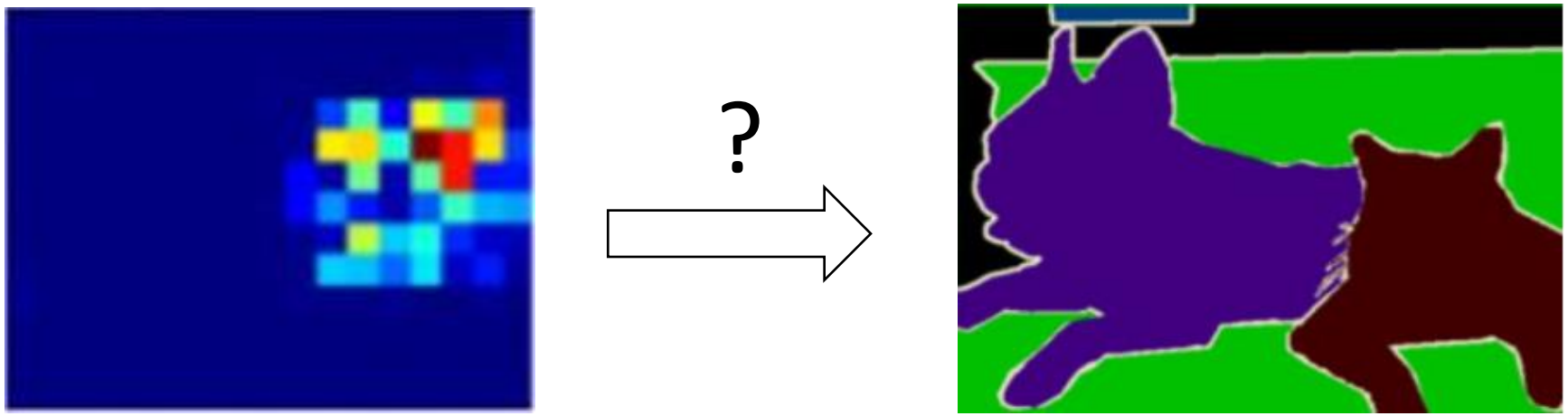
Fully Convolution Networks



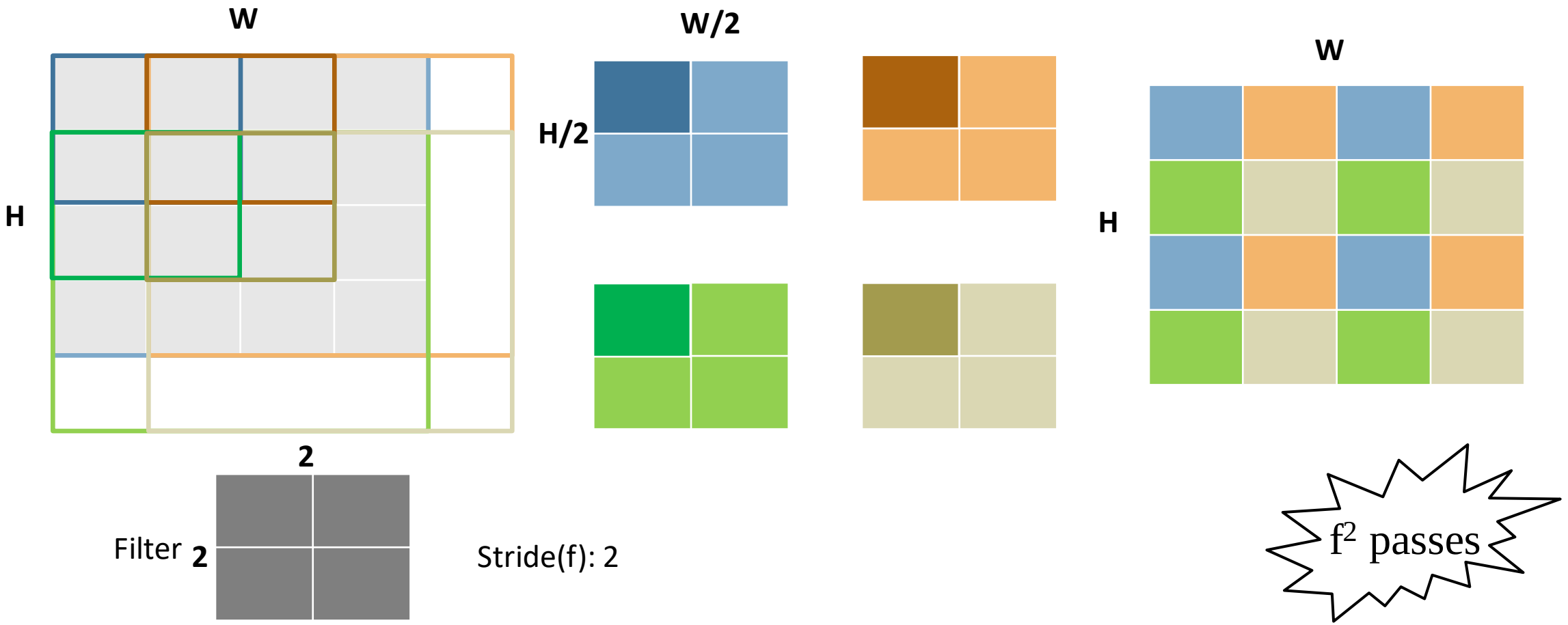
Fully Convolution Networks



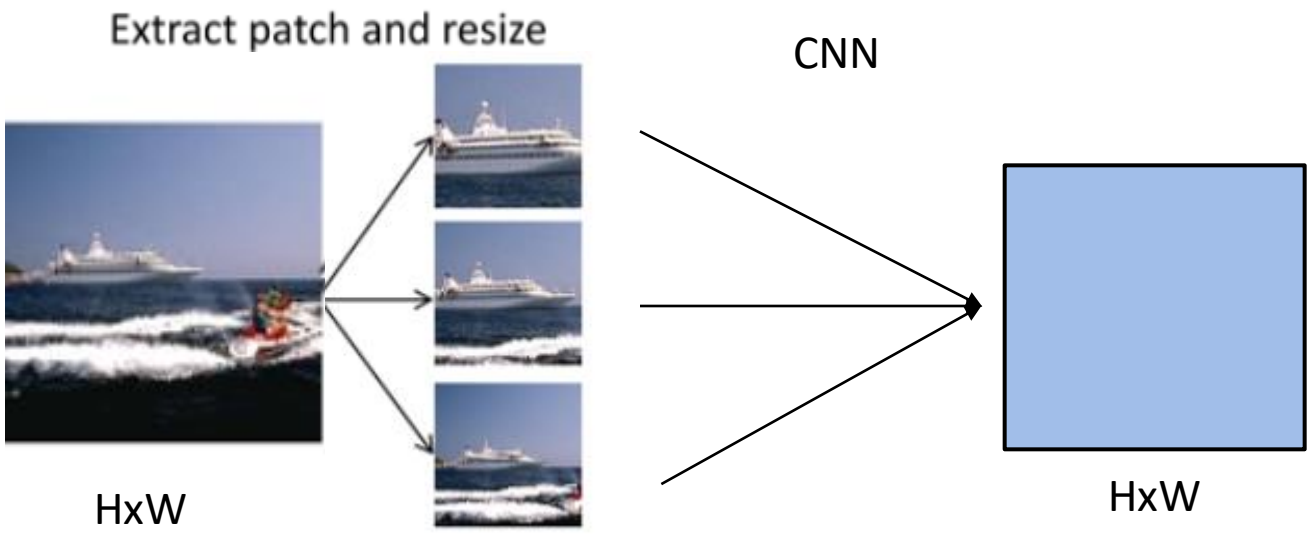
End-to-end Dense predictions



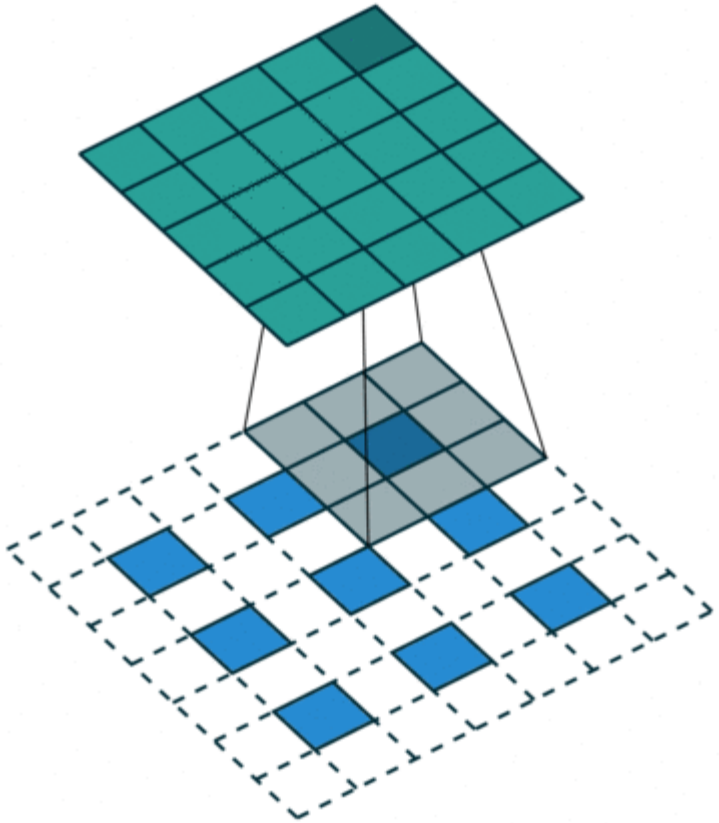
Shift and Stitch



Patch wise Training

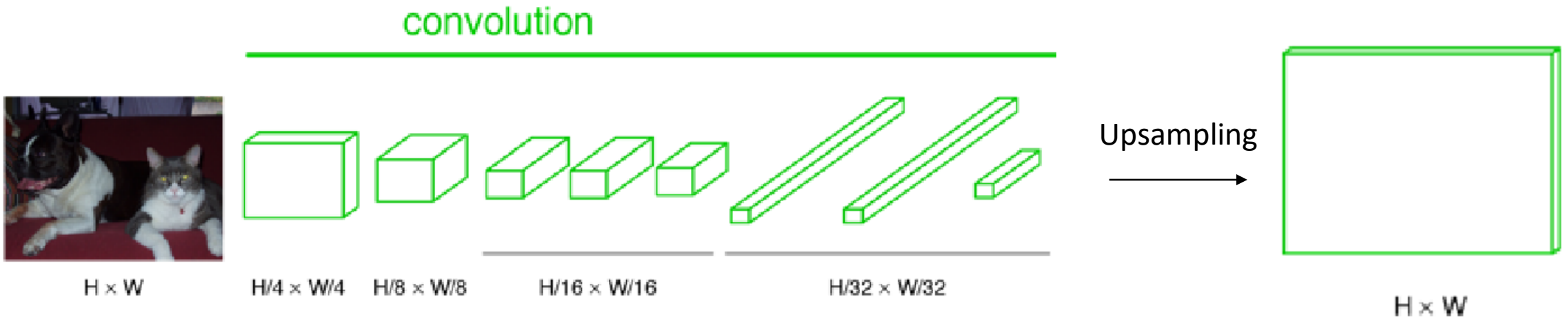


Upsampling

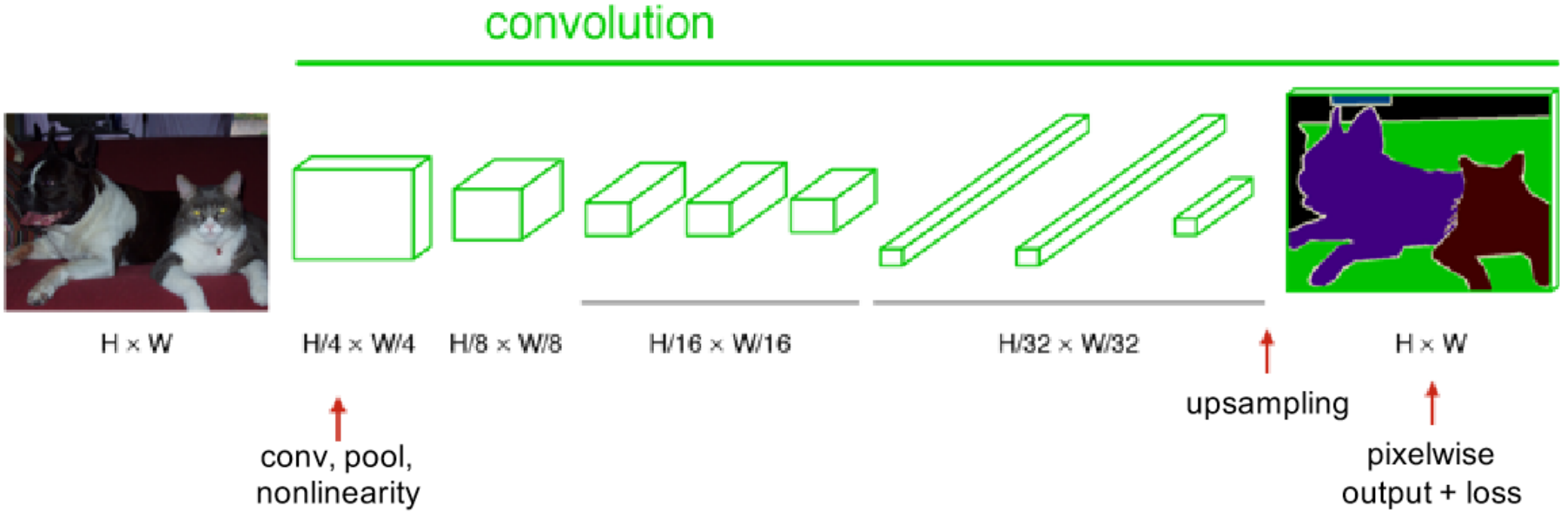


SOURCE: HTTPS://GITHUB.COM/VDUMOULIN/CONV_ARITHMETIC

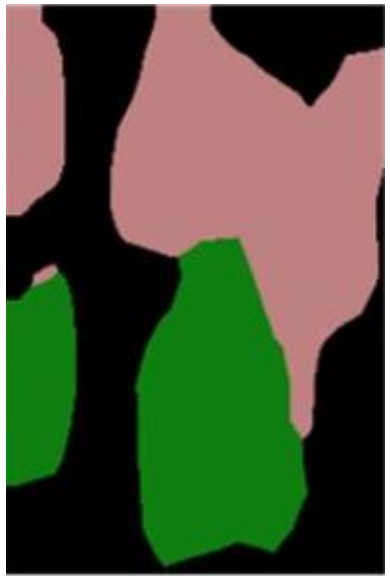
Upsampling



Fully Convolution Networks



Spatial Precision of the O/P

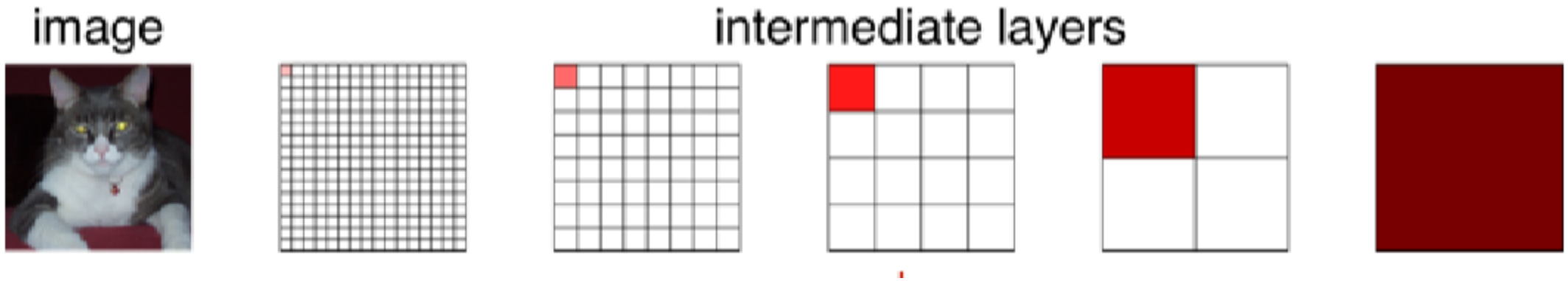


FCN – 32s

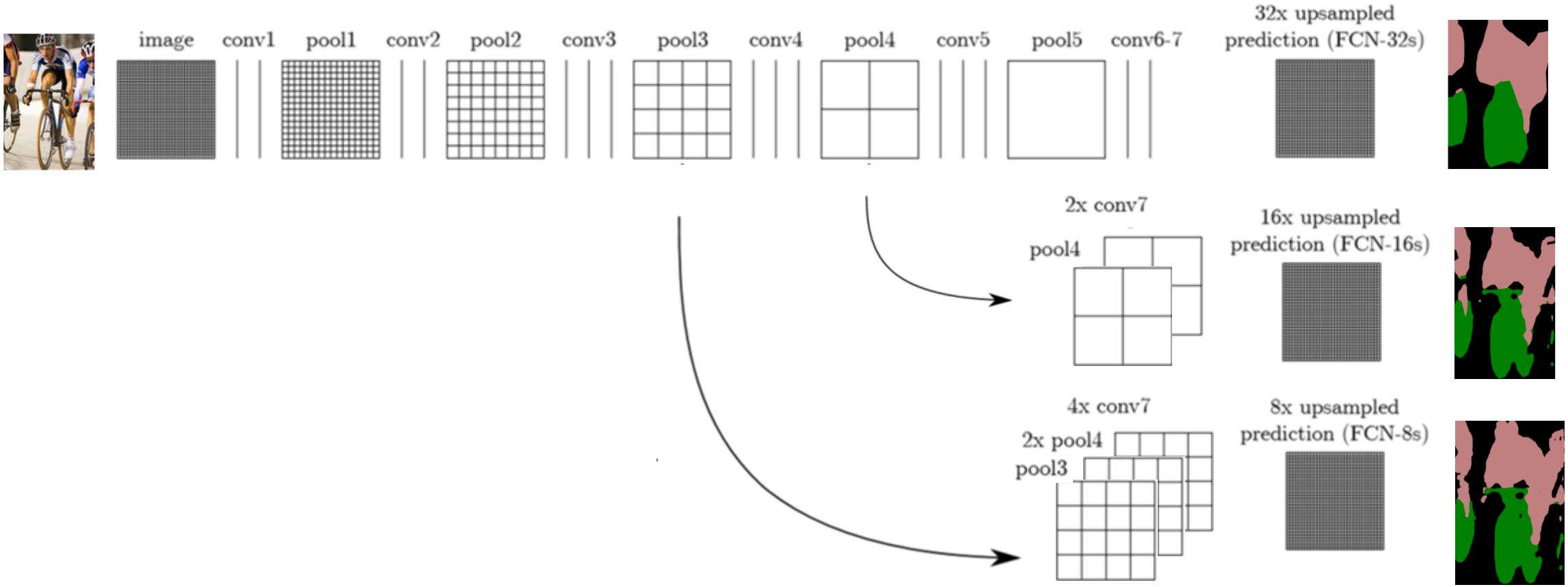
Image

Ground Truth

Combining What and Where



Skip Architecture



Pros

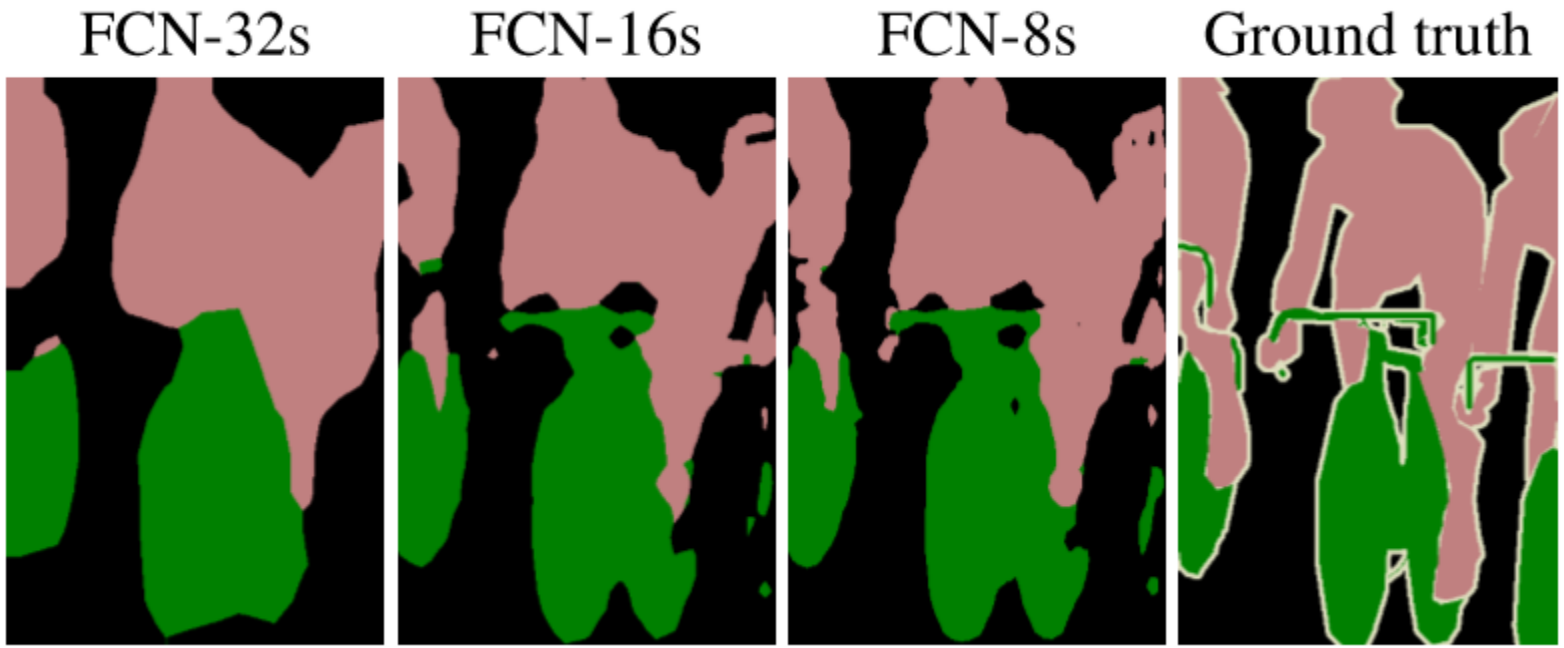
Well structured paper

Supervised pre-training

Intuitive idea

Combining what & where

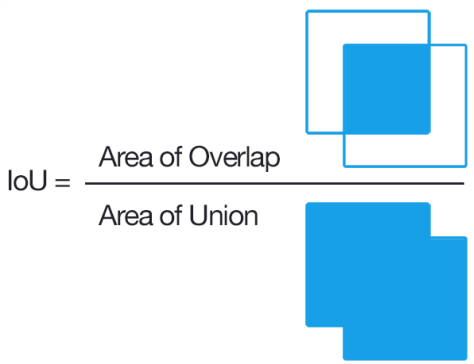
Combining what & where



Discussion of approaches for dense predictions

Good Results

Results



mean IU: $(1/n_{\text{cl}}) \sum_i n_{ii} / \left(t_i + \sum_j n_{ji} - n_{ii}\right)$

	FCN-AlexNet	FCN-VGG16	FCN-GoogLeNet[4]
mean IU	39.8	**56.0**	42.5
forward time	50 ms	210 ms	59 ms
conv. layers	8	16	22
parameters	57M	134M	6M
rf size	355	404	907
max stride	32	32	32

Results

	mean IU VOC2011 test	mean IU VOC2012 test	inference time
R-CNN [12]	47.9	-	-
SDS [17]	52.6	51.6	~ 50 s
FCN-8s	**62.7**	**62.2**	~ **175 ms**

PASCAL VOC

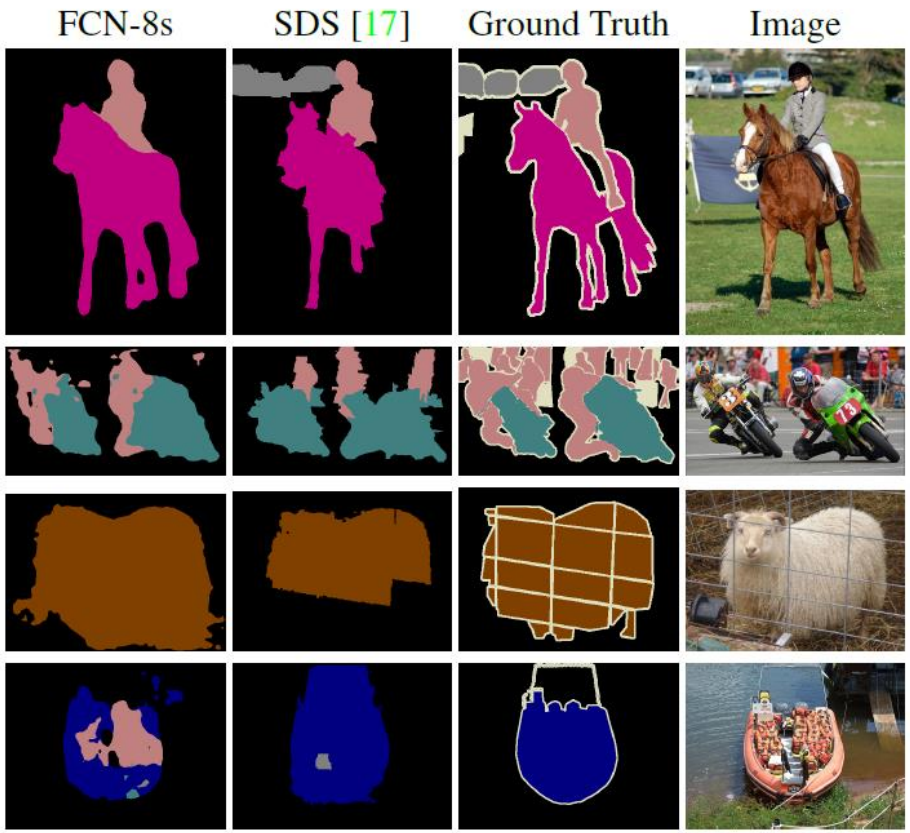
	pixel acc.	mean acc.	mean IU	f.w. IU
Gupta *et al.* [15]	60.3	-	28.6	47.0
FCN-32s RGB	60.0	42.2	29.2	43.9
FCN-32s RGBD	61.5	42.4	30.5	45.5
FCN-32s HHA	57.1	35.2	24.2	40.4
FCN-32s RGB-HHA	64.3	44.9	32.8	48.0
FCN-16s RGB-HHA	**65.4**	**46.1**	**34.0**	**49.5**

NYUDv2

	pixel acc.	mean acc.	mean IU	f.w. IU	geom. acc.
Liu *et al.* [25]	76.7	-	-	-	-
Tighe *et al.* [36]	-	-	-	-	90.8
Tighe *et al.* [37] 1	75.6	41.1	-	-	-
Tighe *et al.* [37] 2	78.6	39.2	-	-	-
Farabet *et al.* [9] 1	72.3	50.8	-	-	-
Farabet *et al.* [9] 2	78.5	29.6	-	-	-
Pinheiro *et al.* [31]	77.7	29.8	-	-	-
FCN-16s	**85.2**	**51.7**	39.5	76.1	**94.3**

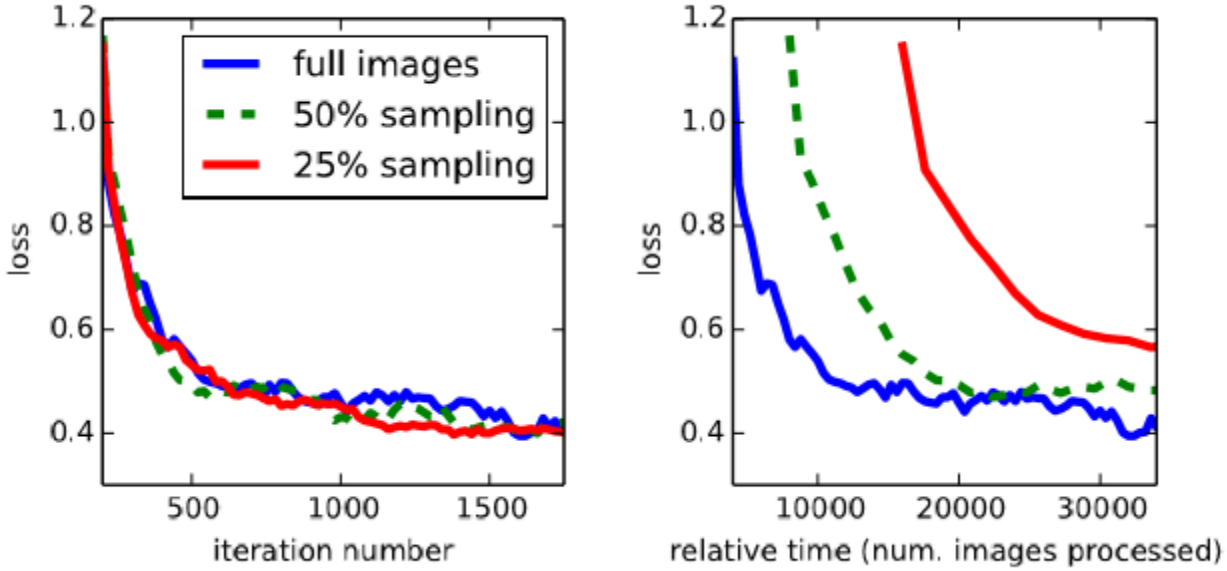
SIFT Flow

Results – PASCAL VOC



Speed improvement

Whole image Vs Patchwise training



Heavily used in practice

Google fully convolutional networks for semantic segmentation

Scholar

Articles

Case law

My library

Any time

Since 2017

Fully convolutional networks for semantic segmentation

J Long, E Shelhamer, T Darrell - Proceedings of the IEEE …, 2015 - cv-foundation.org

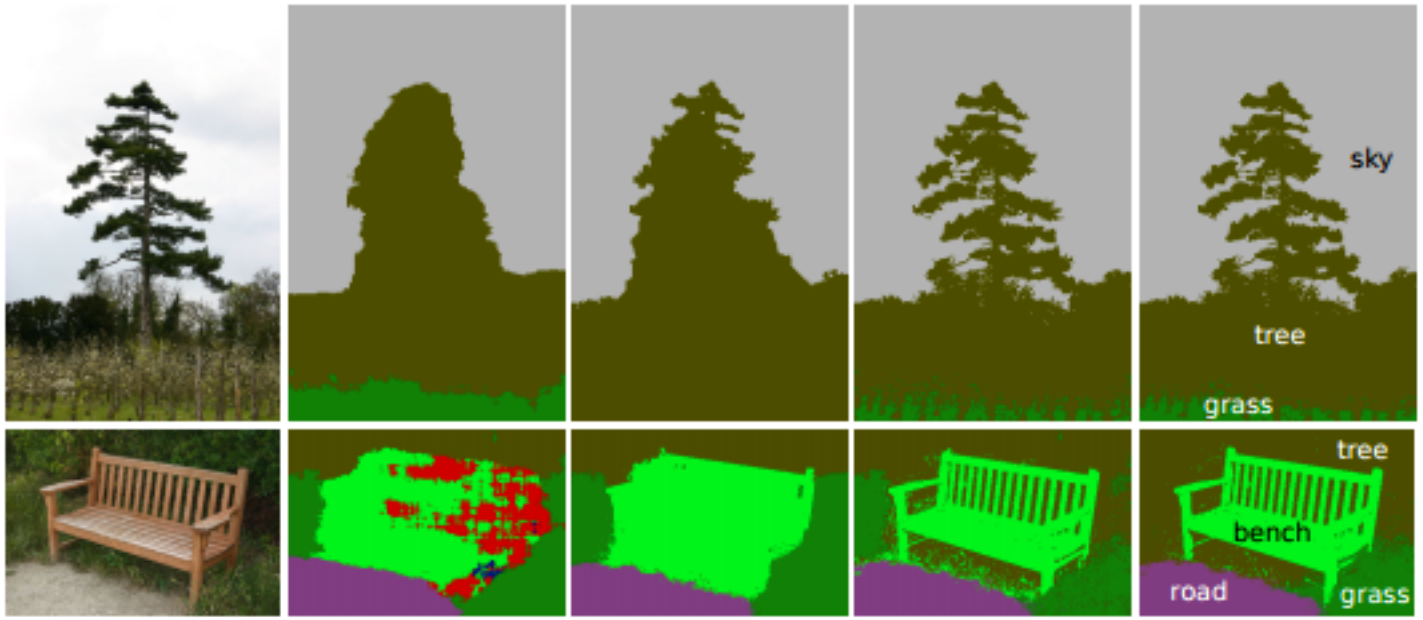
Abstract Convolutional networks are powerful visual models that yield hierarchies of features. We show that convolutional networks by themselves, trained end-to-end, pixels-to-pixels, exceed the state-of-the-art in semantic segmentation. Our key insight is to build" fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. We define and detail the space of fully ...

Cited by 2439 Related articles All 21 versions Cite Save

Cons

Cons1: What about Fully connected CRF?

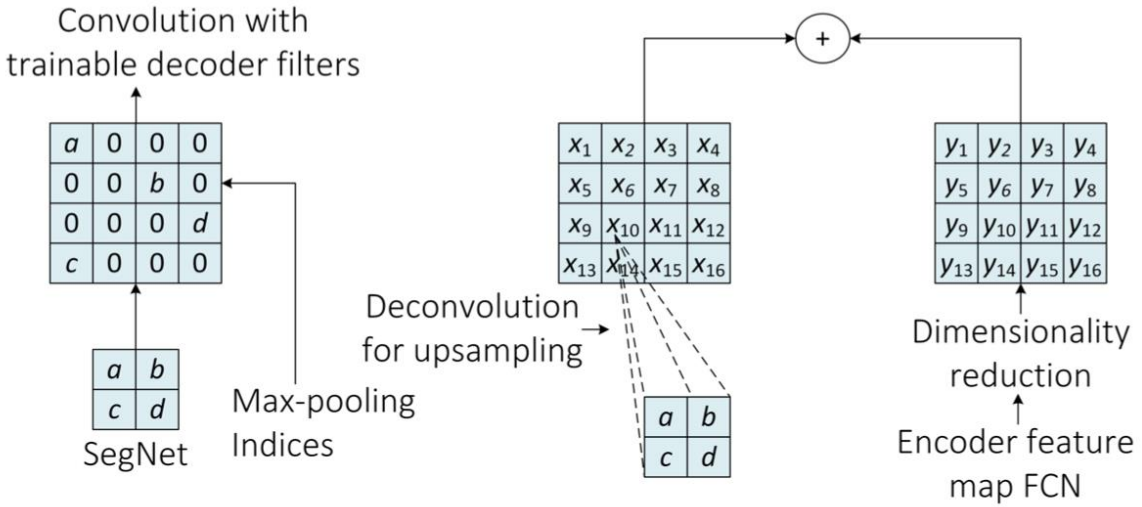
(Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, Krähenbühl and Koltun)



(a) Image (b) Unary classifiers (c) Robust P^n CRF (d) Fully connected CRF, MCMC inference, 36 hrs (e) Fully connected CRF, our approach, 0.2 seconds

Cons2: Qualitative Claims -> possible explanations?

Cons3: Execution time vs Memory footprint vs. Accuracy trade-off



SegNet vs. FCN

Cons4: Future directions?

Questions?