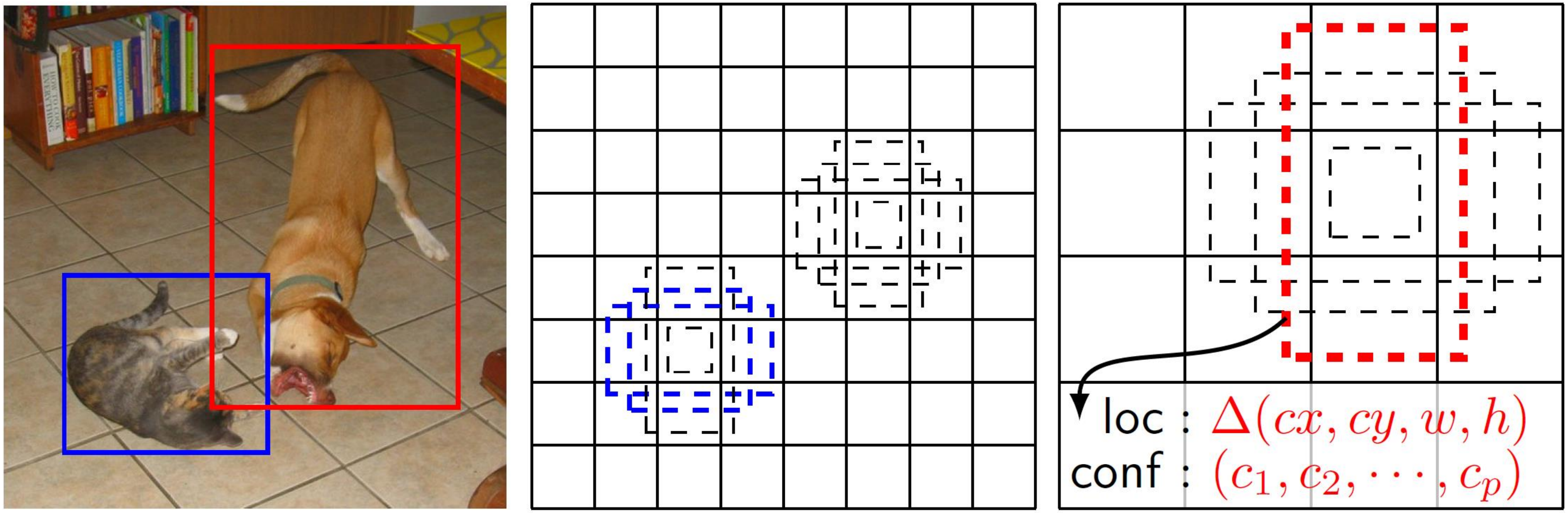
“Unsupervised” Deep Learning and Deep Style Transfer

James Hays

slides from Carl Doersch and Mark Chang

Recap from Previous Lecture

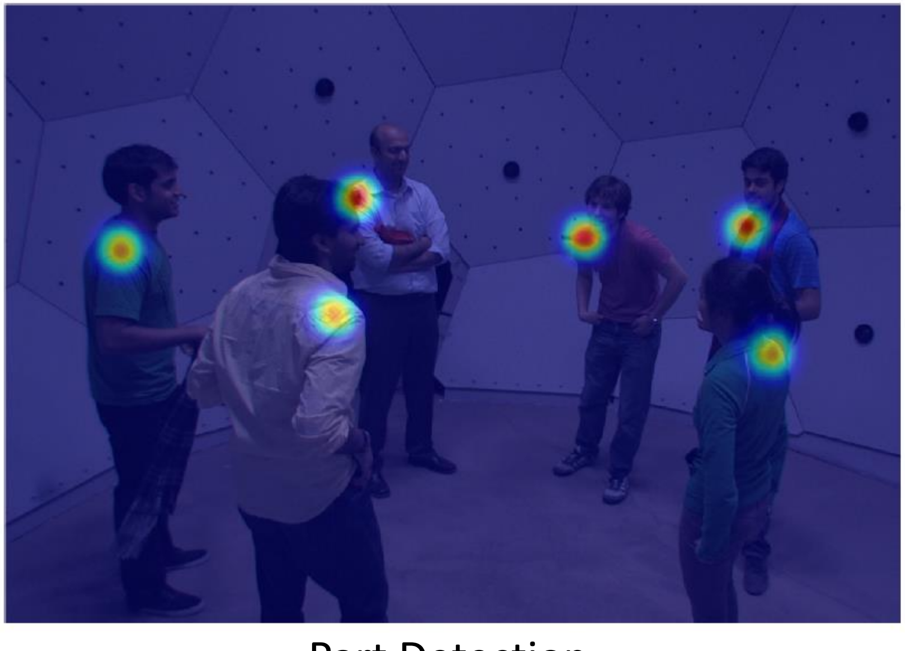
- We saw two strategies to get *structured* output while using deep learning
  - With object detection, one strategy is brute force: detect everywhere at once


loc : $\Delta(cx, cy, w, h)$
conf : $(c_1, c_2, \cdots, c_p)$

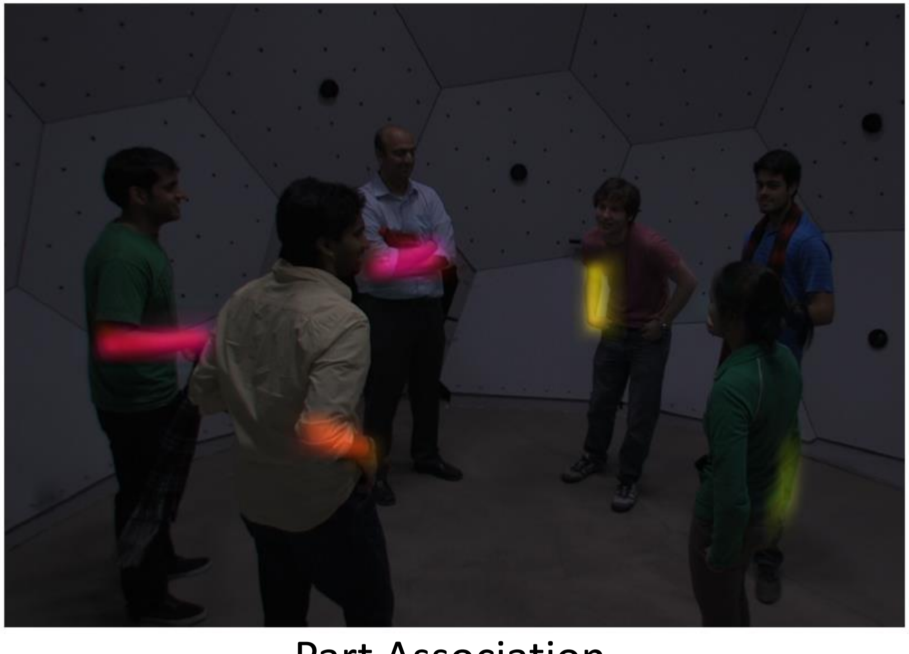
(a) Image with GT boxes (b) 8×8 feature map (c) 4×4 feature map

Recap from Previous Lecture

- We saw two strategies to get *structured* output while using deep learning
  - With pose estimation / keypoint detection, the network produces an image-based intermediate representation



Part Detection



Part Association

Recap from Previous Lecture

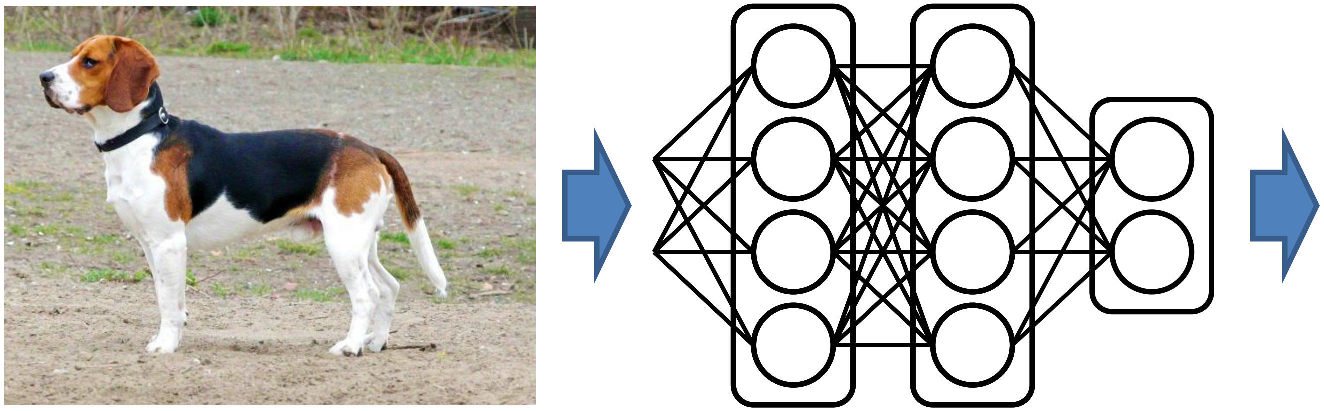
- More generally, it can pay off to get creative. Even if Deep ConvNets aren’t a natural fit for an image-related task, they might be able to learn a subtask or create a useful intermediate representation.

Unsupervised Visual Representation Learning by Context Prediction

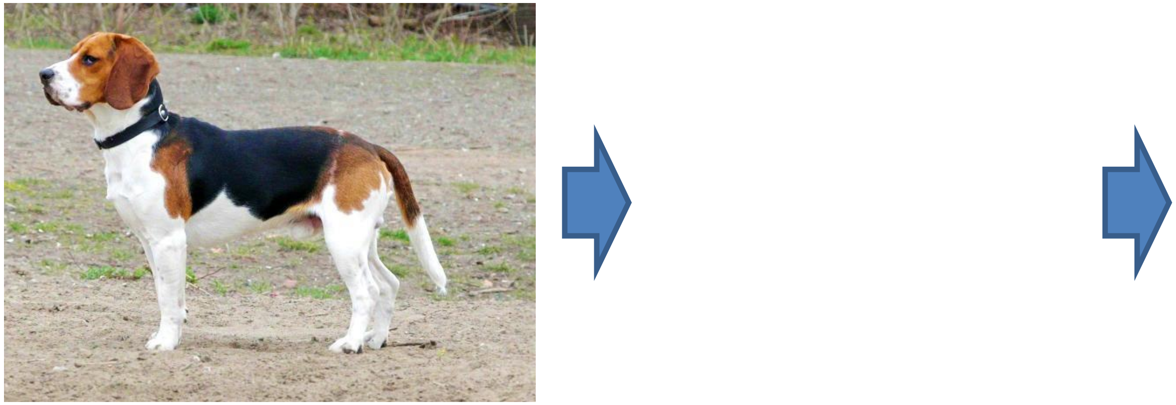
Carl Doersch
Joint work with Alexei A. Efros & Abhinav Gupta

ICCV 2015

ImageNet + Deep Learning

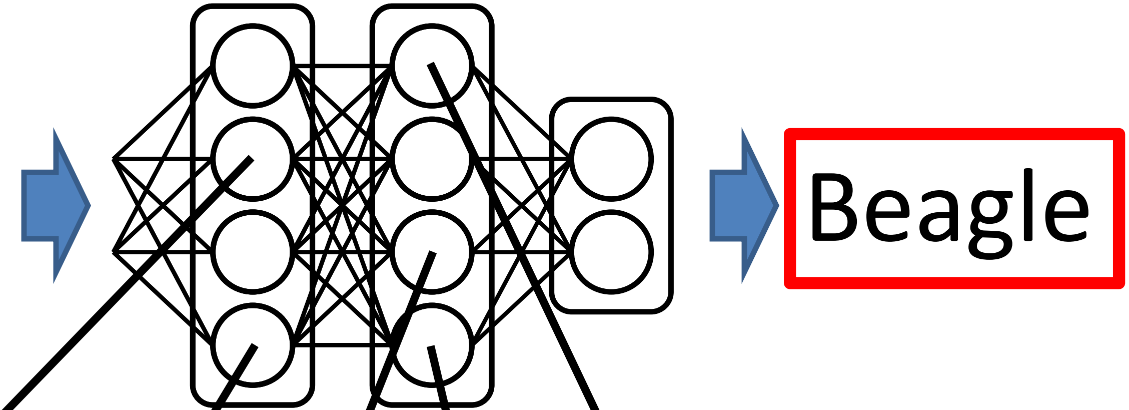


Beagle



- Image Retrieval
- Detection (RCNN)
- Segmentation (FCN)
- Depth Estimation
- …

ImageNet + Deep Learning



Beagle

Materials?

Parts?

Pose?

Geometry?

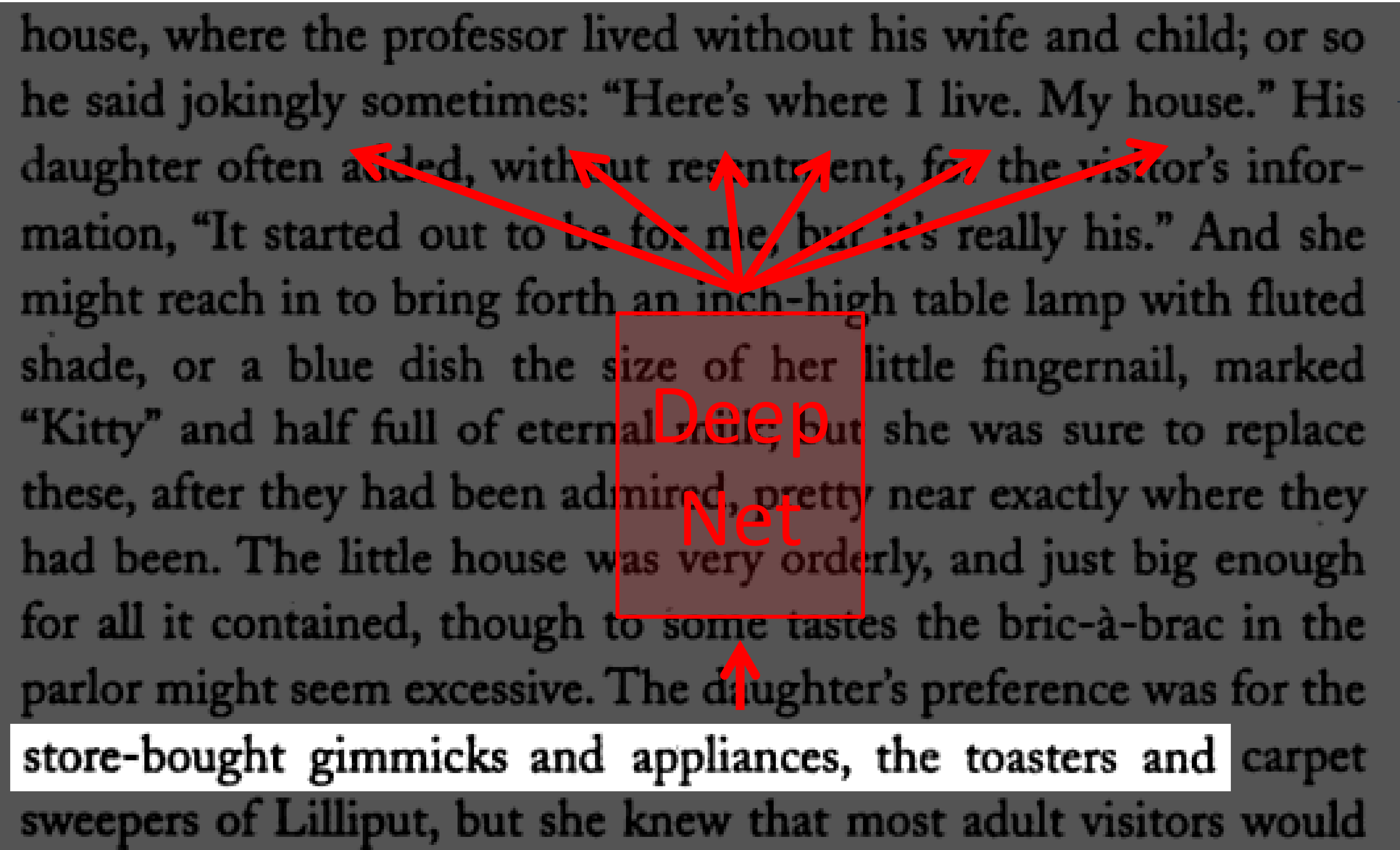
Boundaries?

Do we even need semantic labels?

Do we need this task?

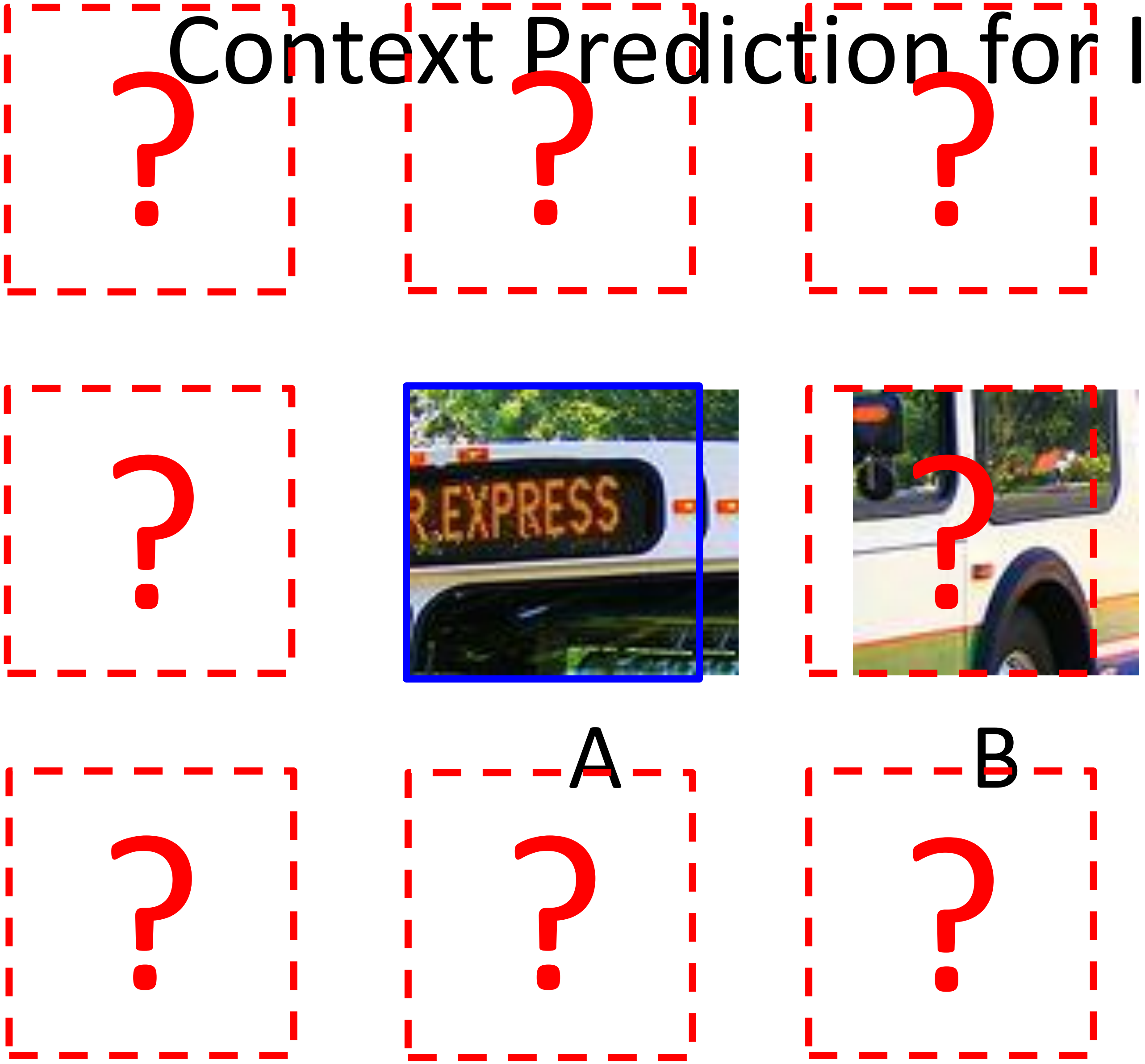
Context as Supervision

[Collobert & Weston 2008; Mikolov et al. 2013]

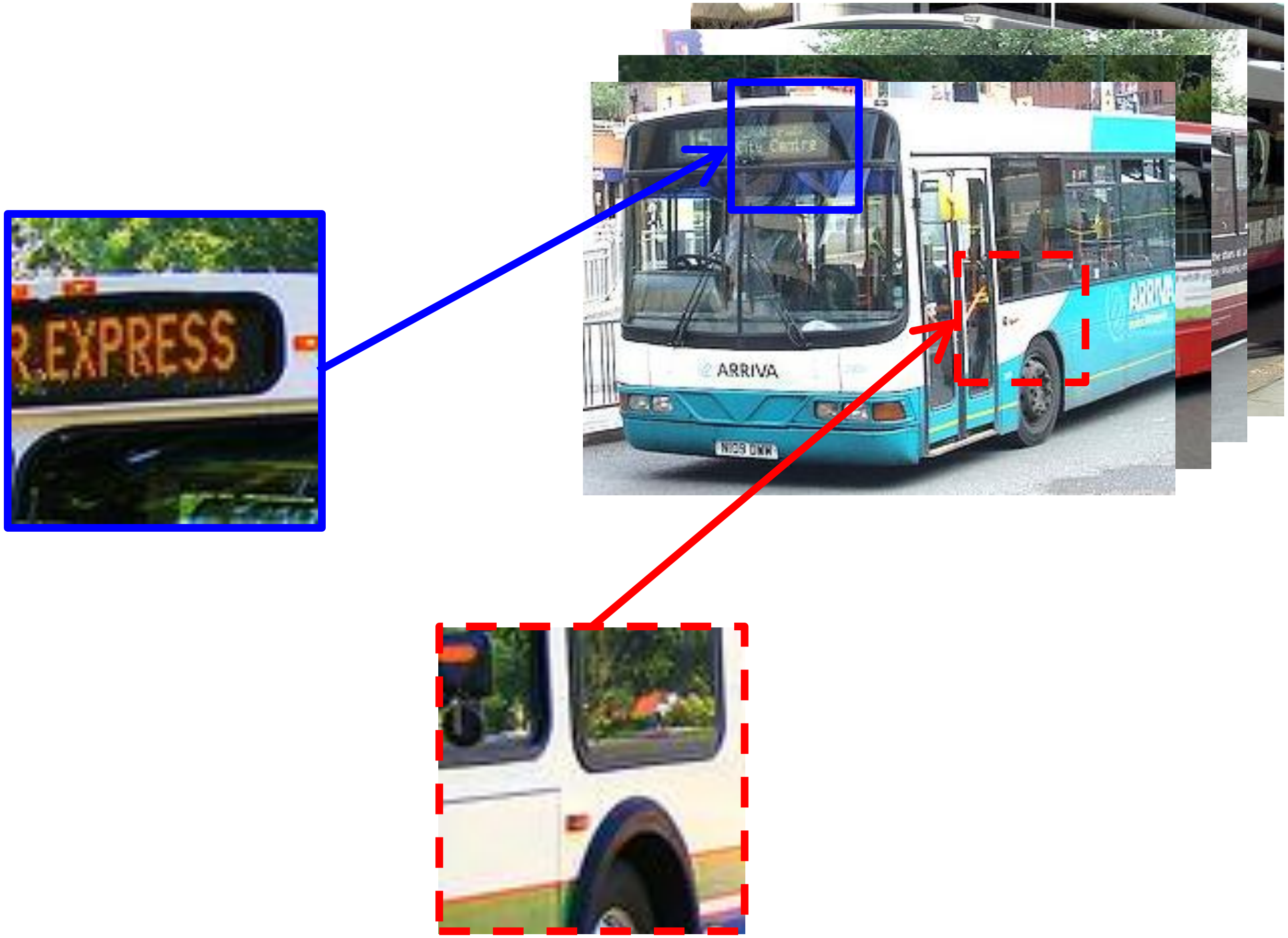


house, where the professor lived without his wife and child; or so he said jokingly sometimes: "Here's where I live. My house." His daughter often added, without resentment, for the visitor's information, "It started out to be for me, but it's really his." And she might reach in to bring forth an inch-high table lamp with fluted shade, or a blue dish the size of her little fingernail, marked "Kitty" and half full of eternal milk, but she was sure to replace these, after they had been admired, pretty near exactly where they had been. The little house was very orderly, and just big enough for all it contained, though to some tastes the bric-à-brac in the parlor might seem excessive. The daughter's preference was for the store-bought gimmicks and appliances, the toasters and carpet sweepers of Lilliput, but she knew that most adult visitors would

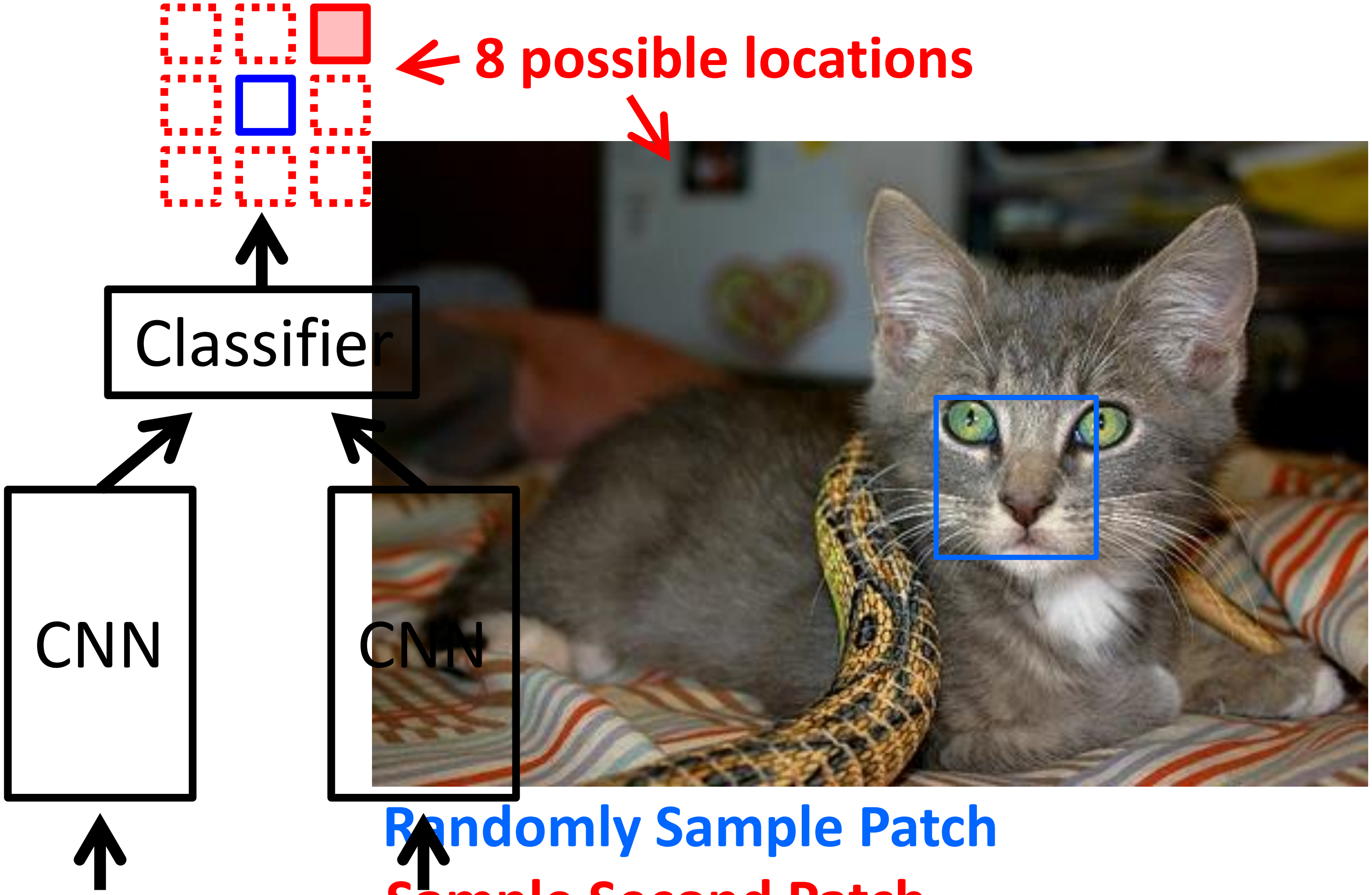
Context Prediction for Images



Semantics from a non-semantic task



Relative Position Task



8 possible locations

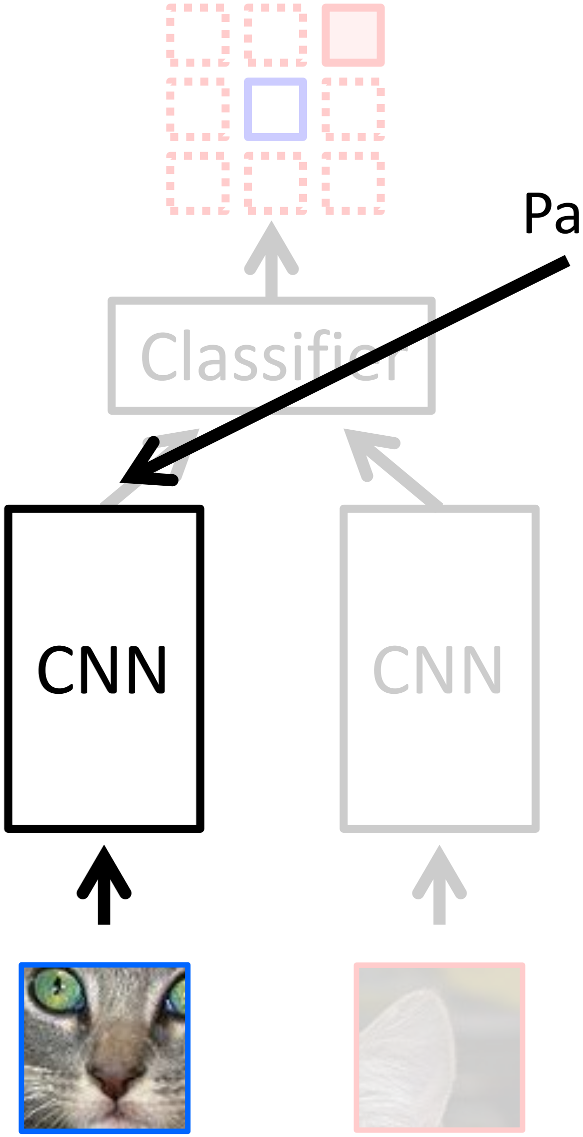
Classifier

CNN

CNN

Randomly Sample Patch

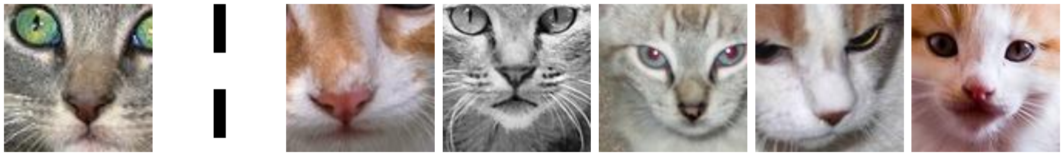
Sample Second Patch



Patch Embedding

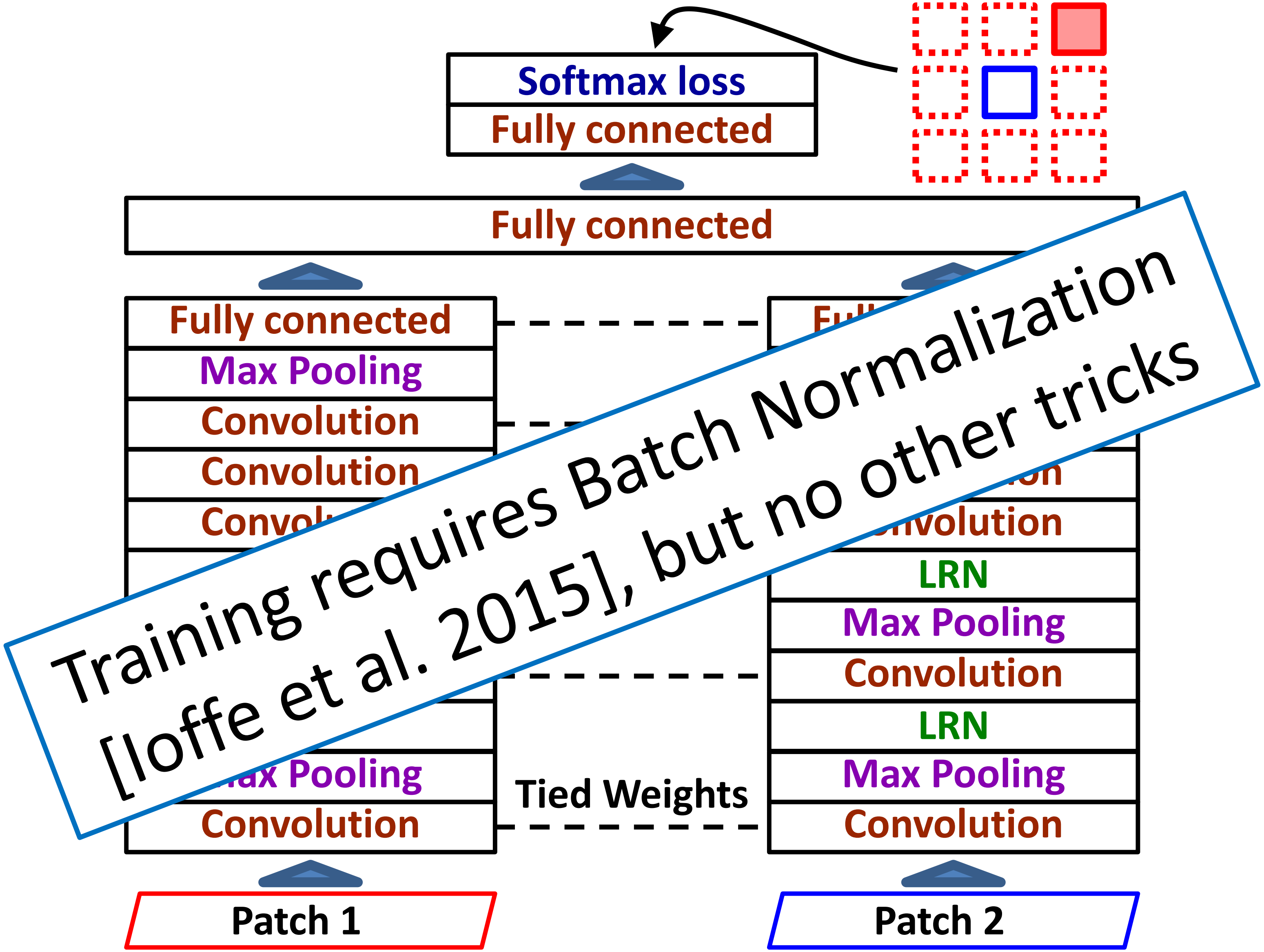
Input

Nearest Neighbors



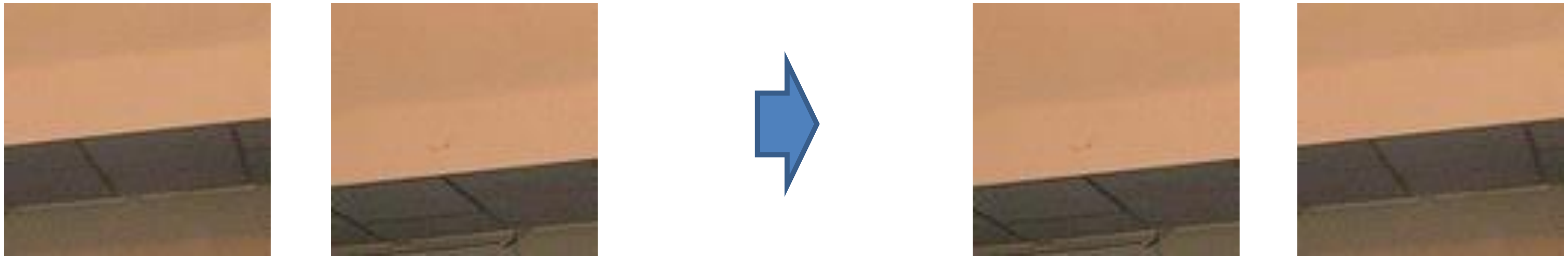
Note: connects ***across*** instances!

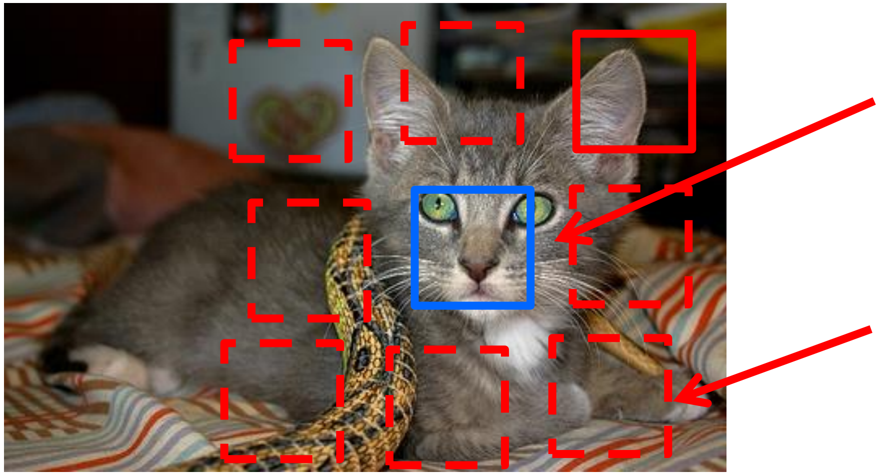
Architecture



Training requires Batch Normalization [Ioffe et al. 2015], but no other tricks

Avoiding Trivial Shortcuts

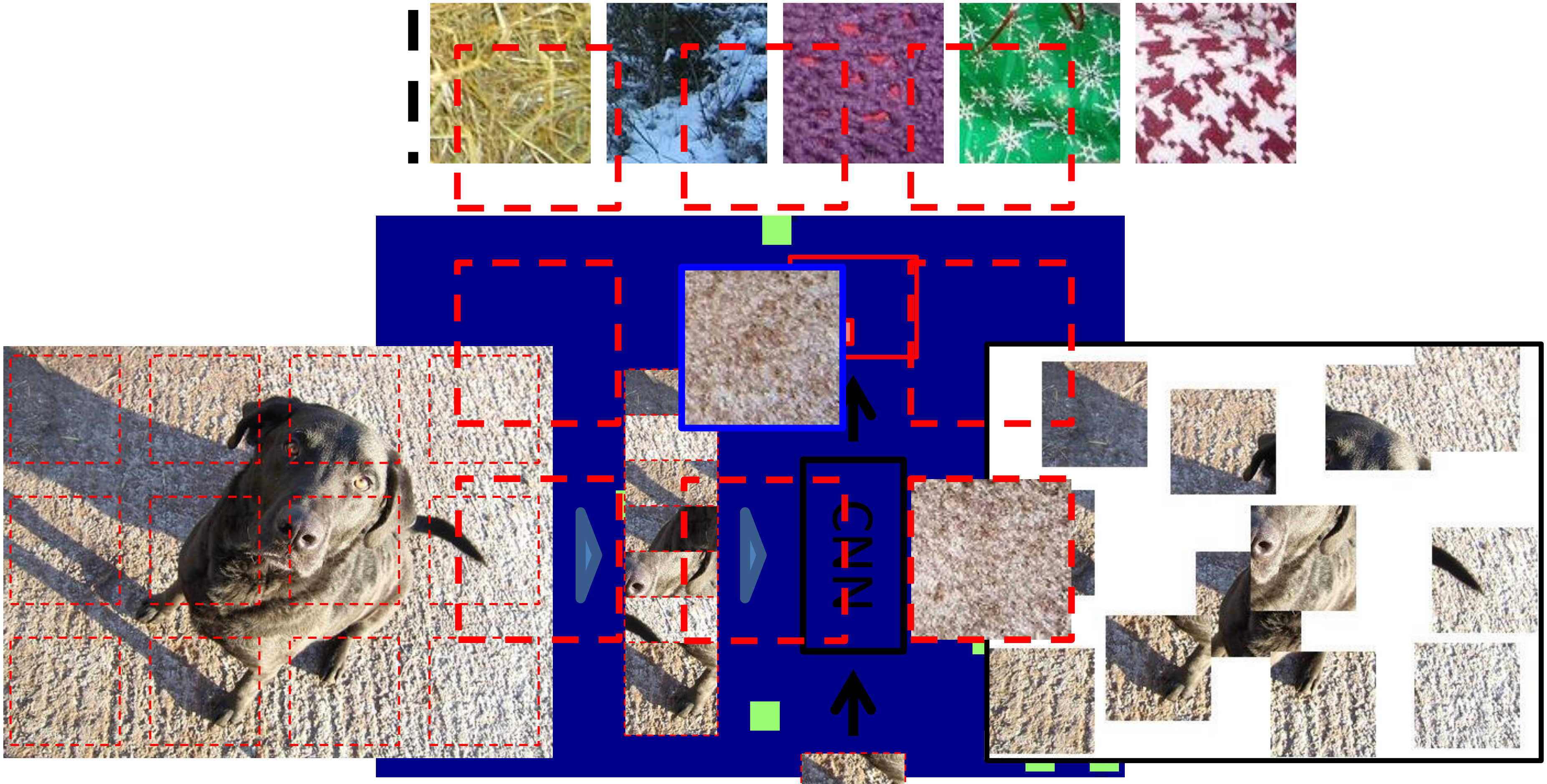




Include a gap

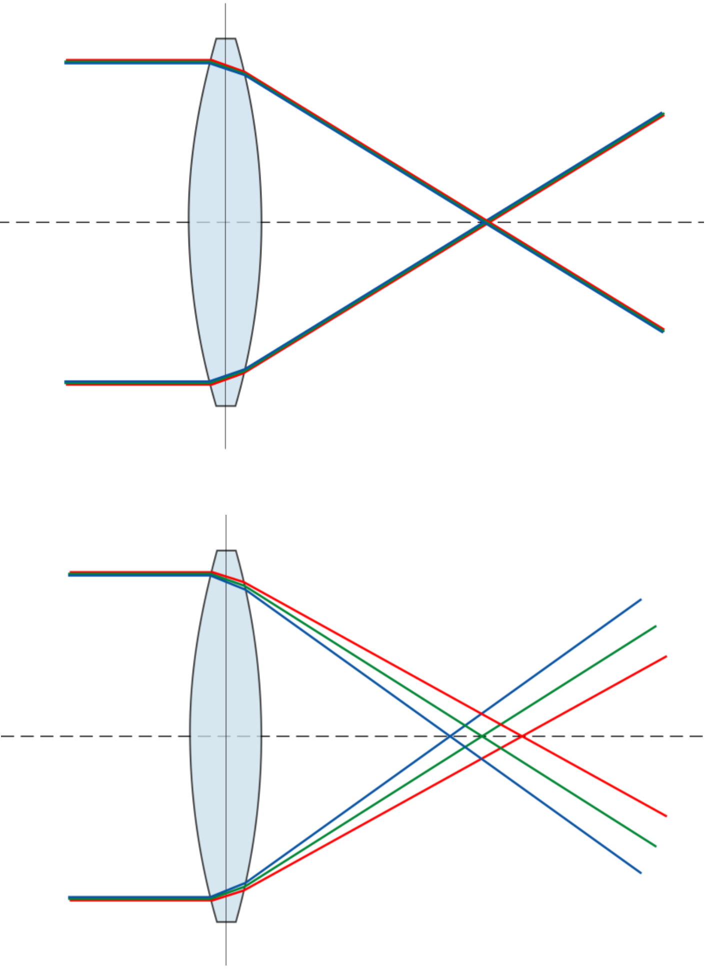
Jitter the patch locations

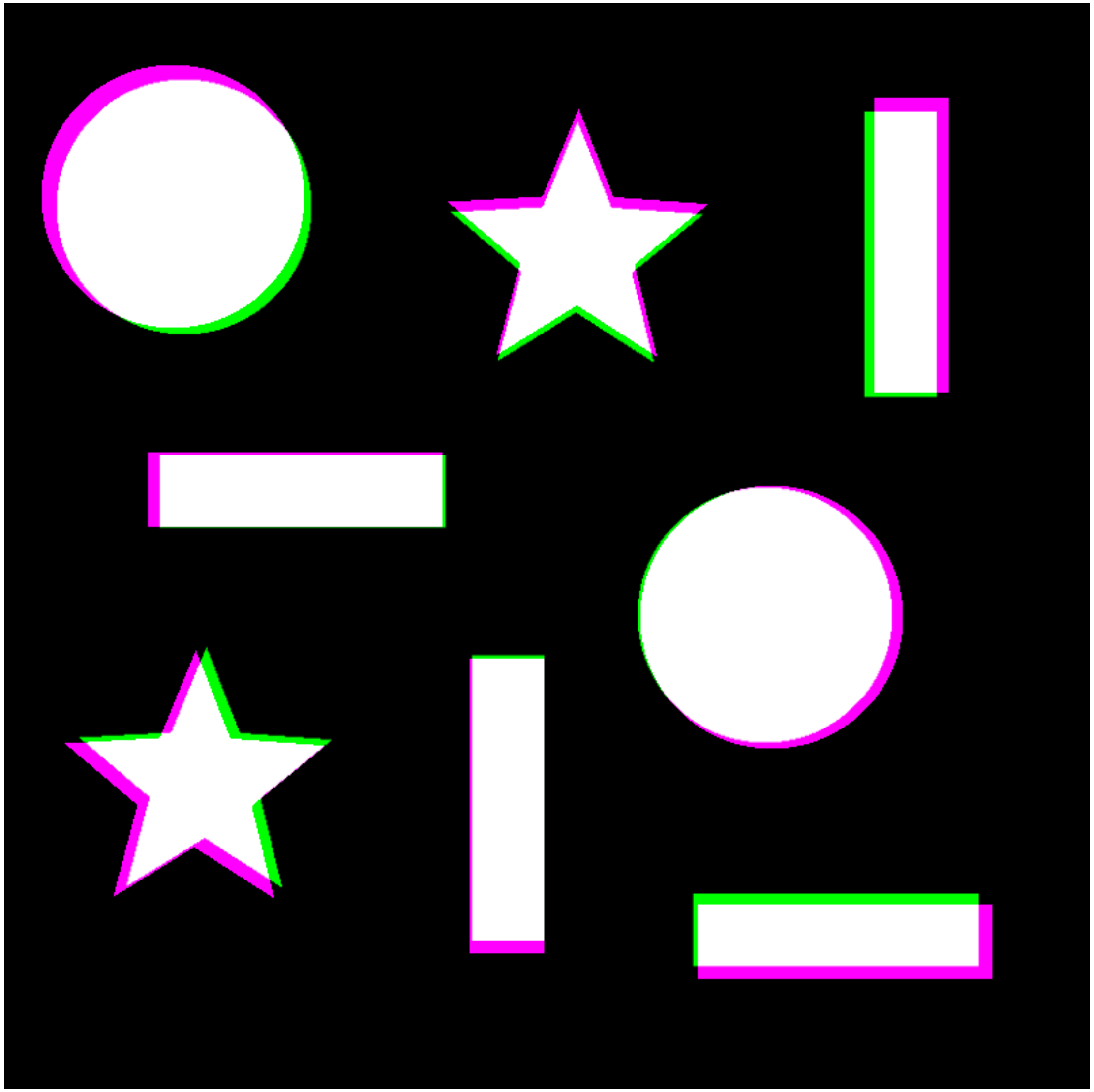
A Not-So "Trivial" Shortcut



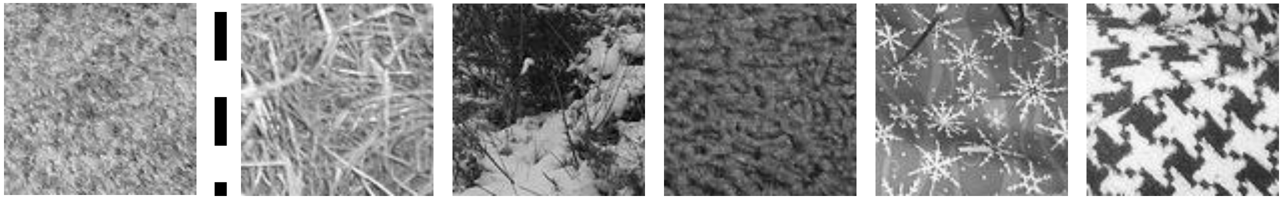
Position in Image

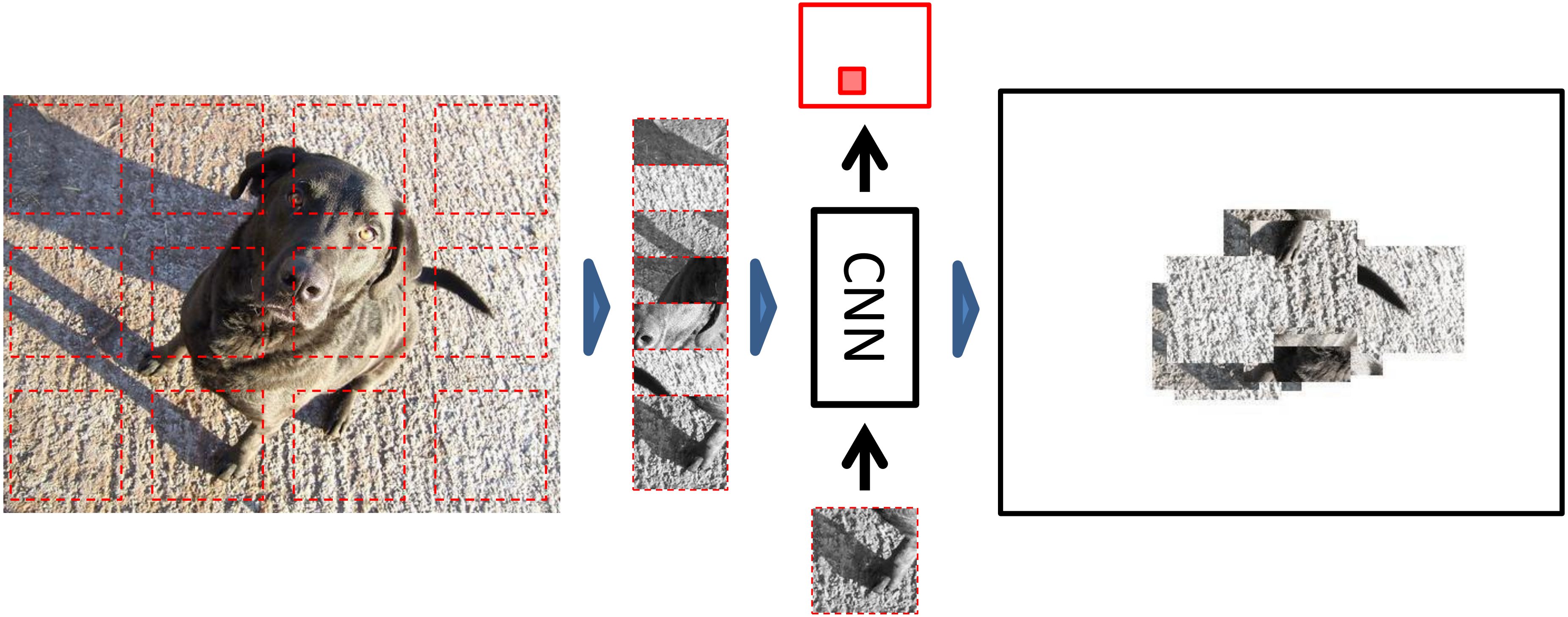
Chromatic Aberration



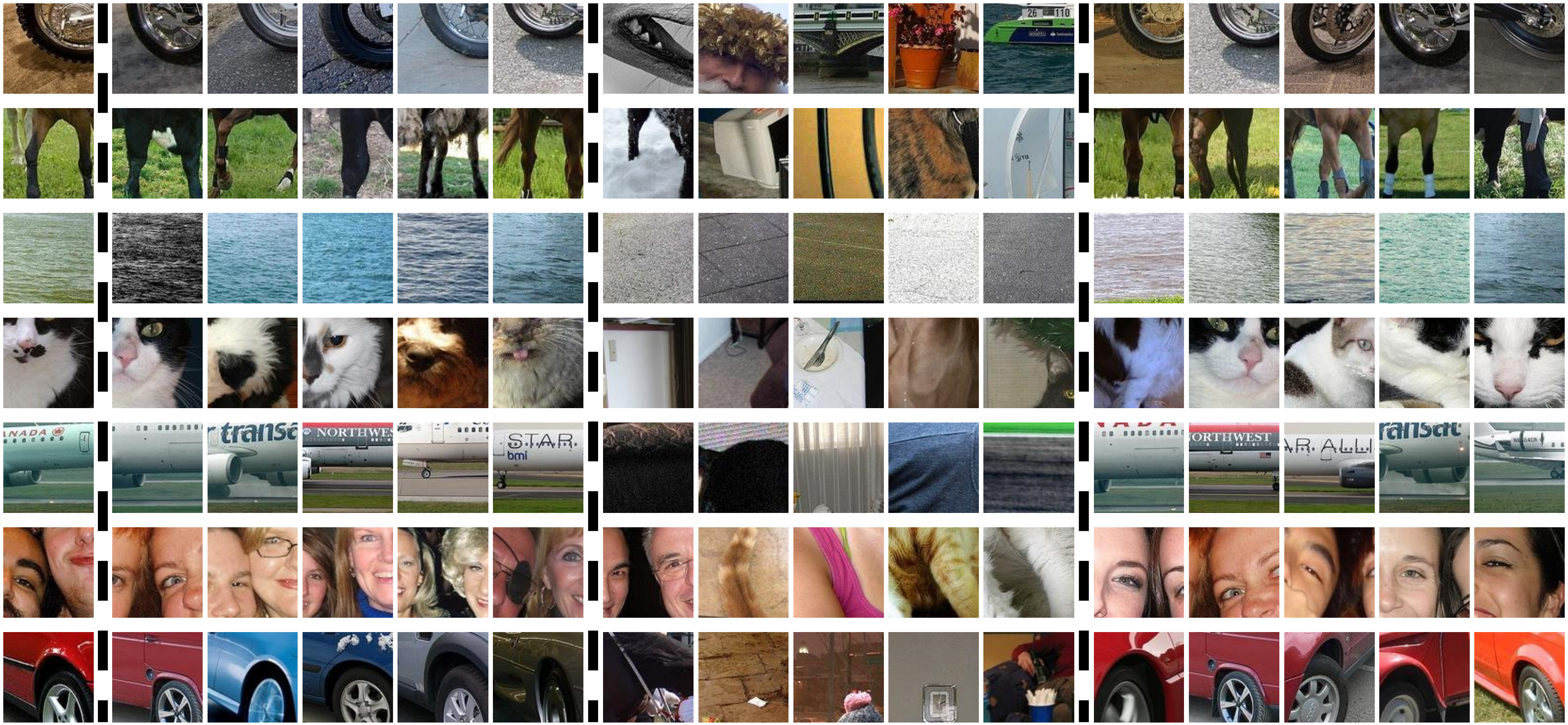


Chromatic Aberration



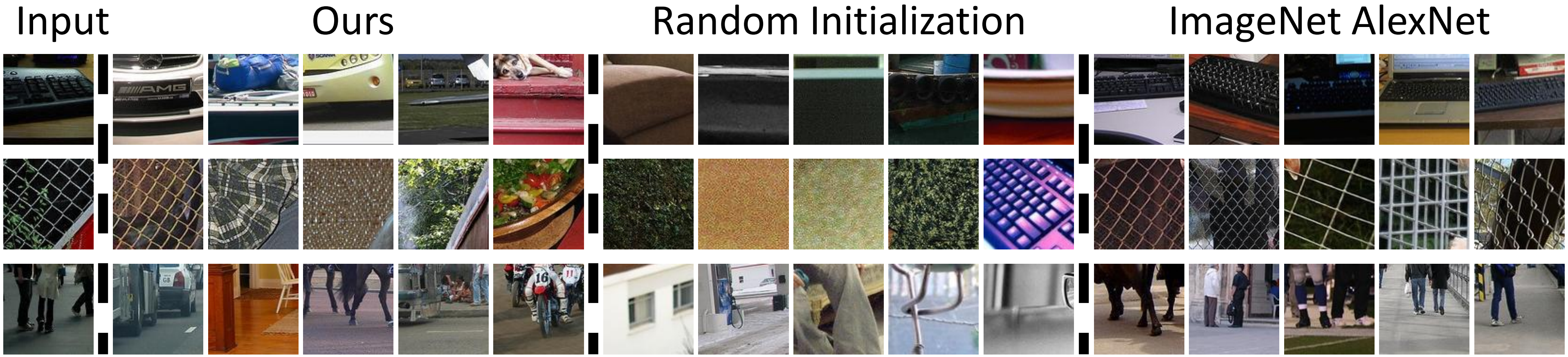


What is learned?

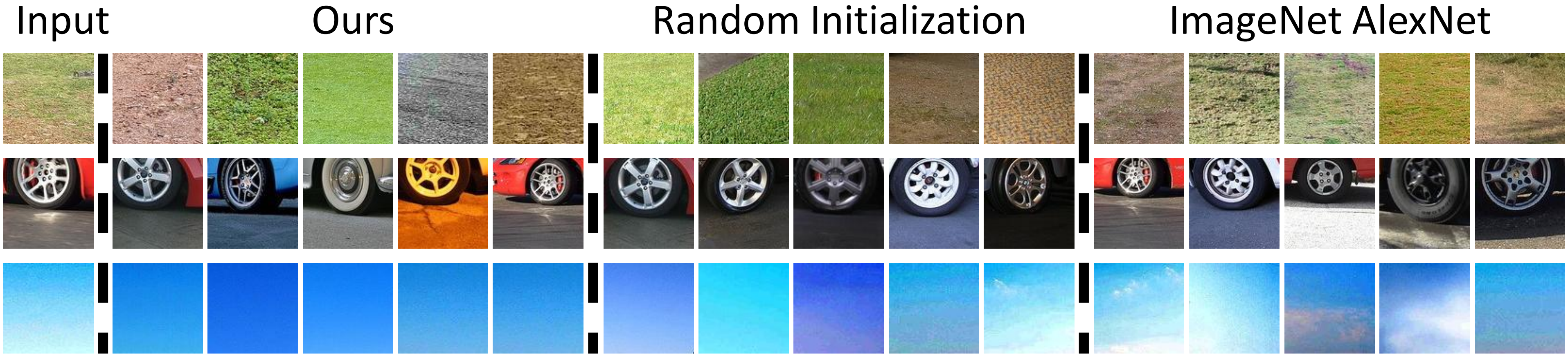


Input Ours Random Initialization ImageNet AlexNet

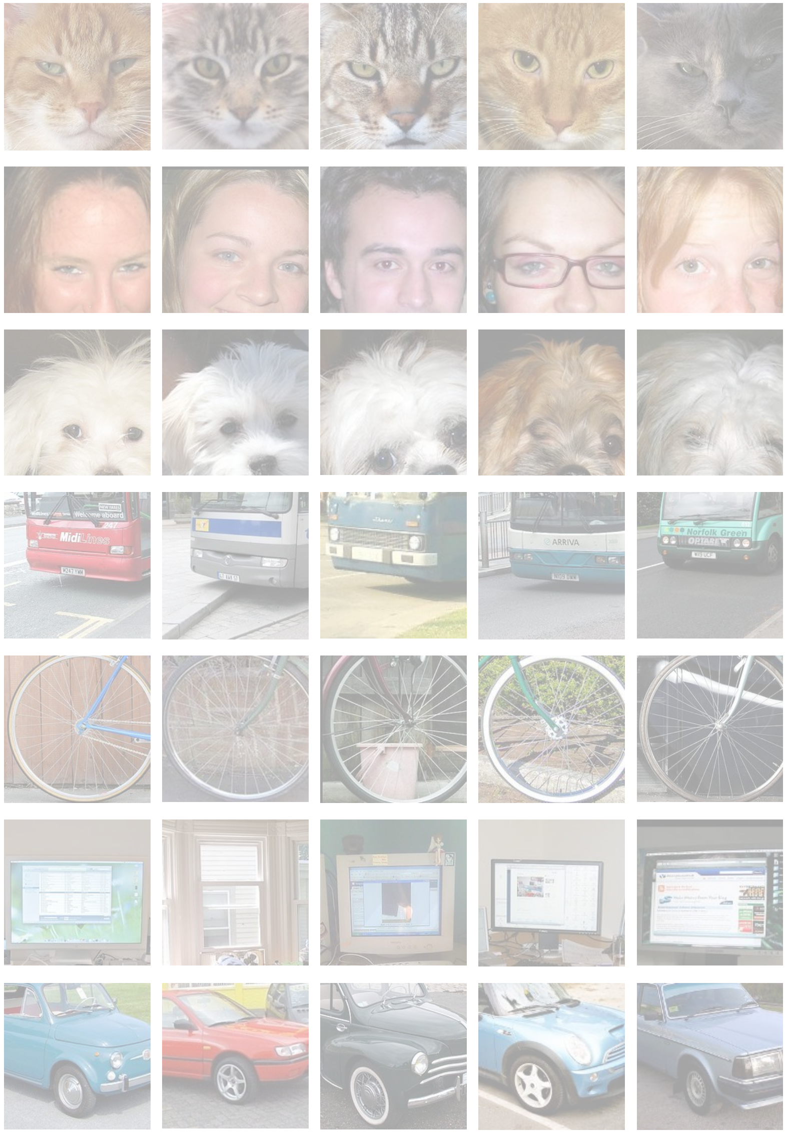
Still don’t capture everything

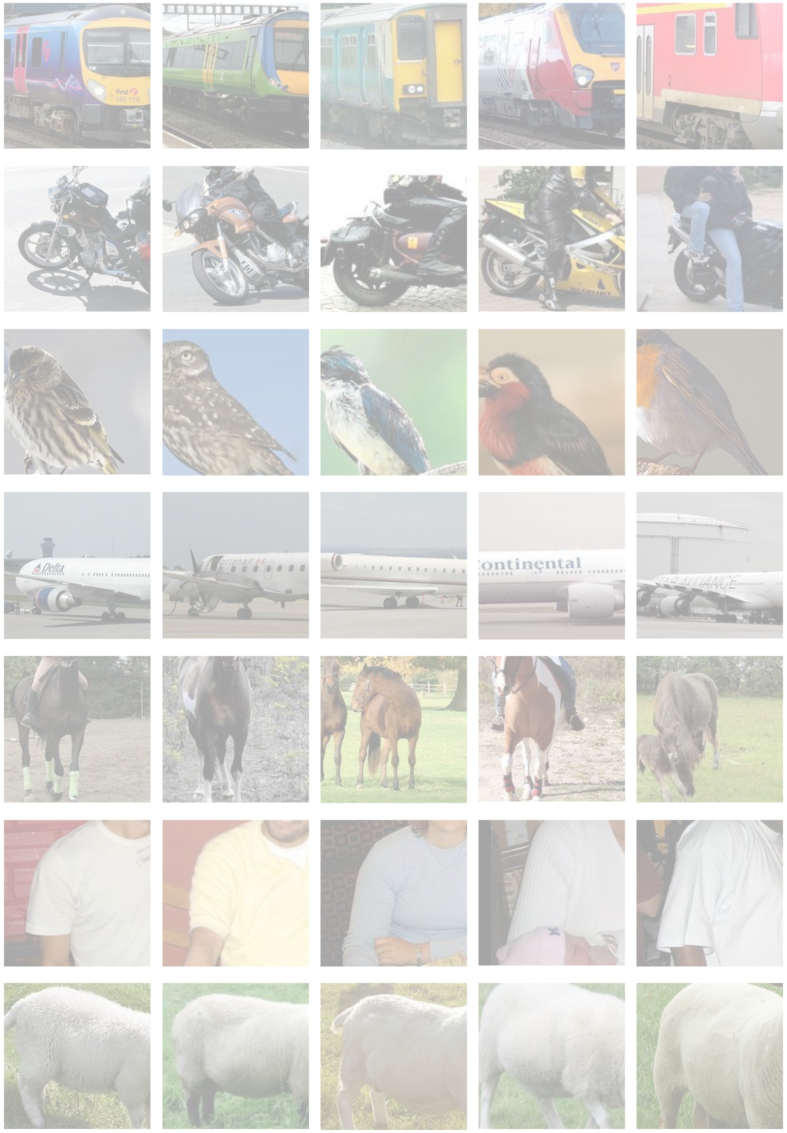


You don’t always need to learn!

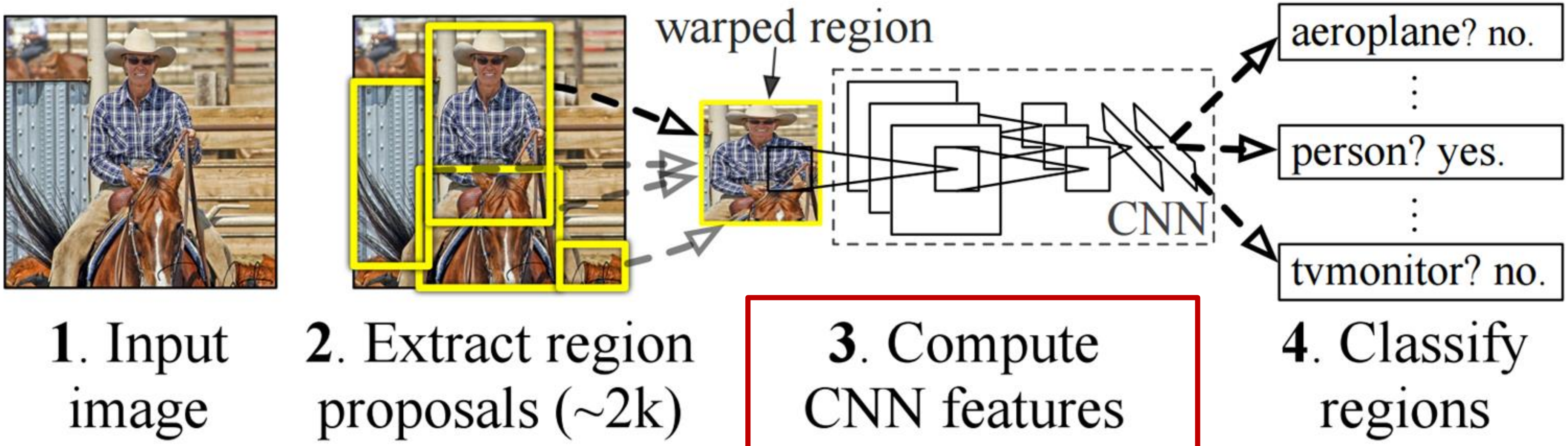


Mined from Pascal VOC2011





Pre-Training for R-CNN

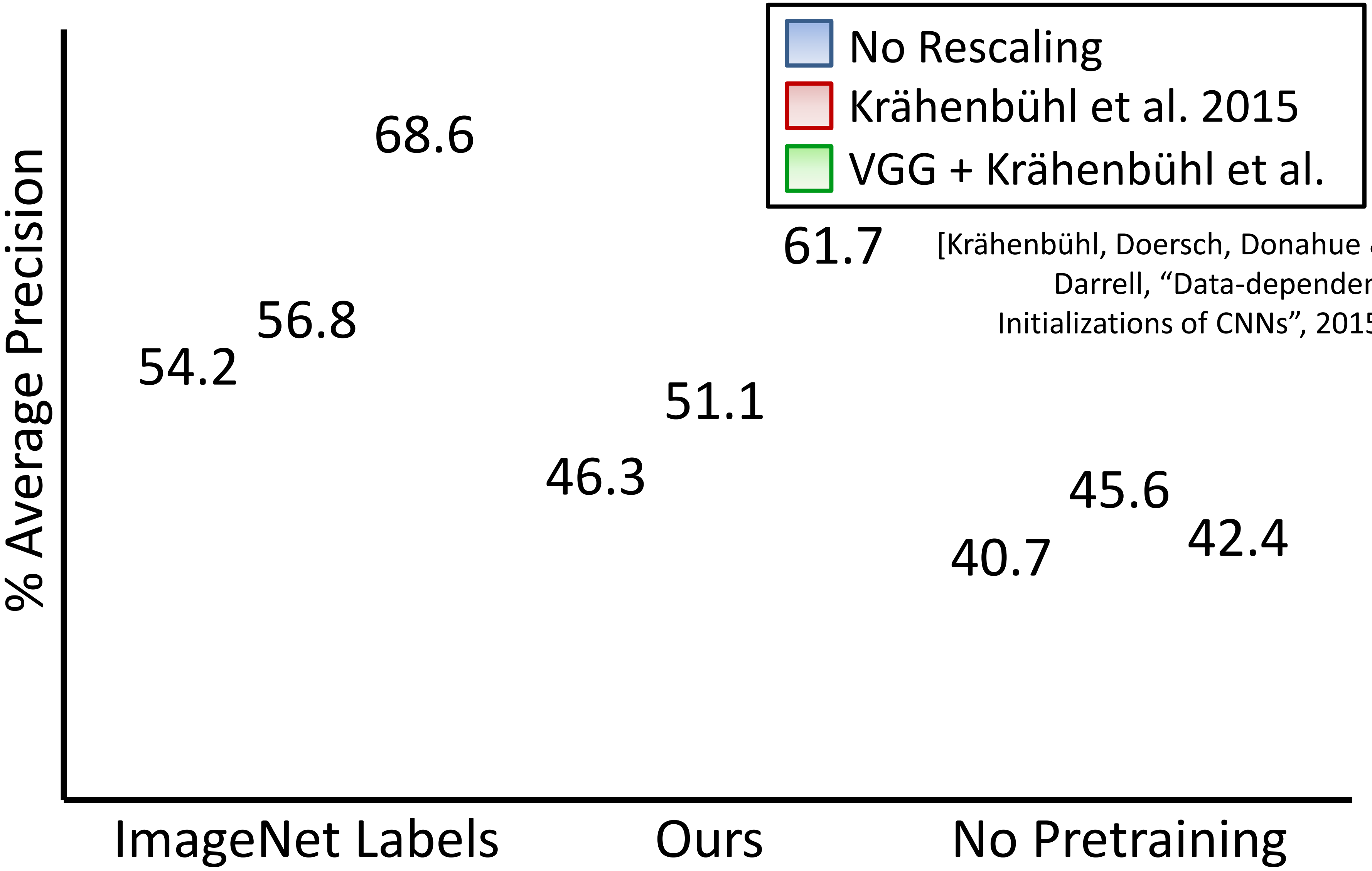


Pre-train on relative-position task, w/o labels

[Girshick et al. 2014]

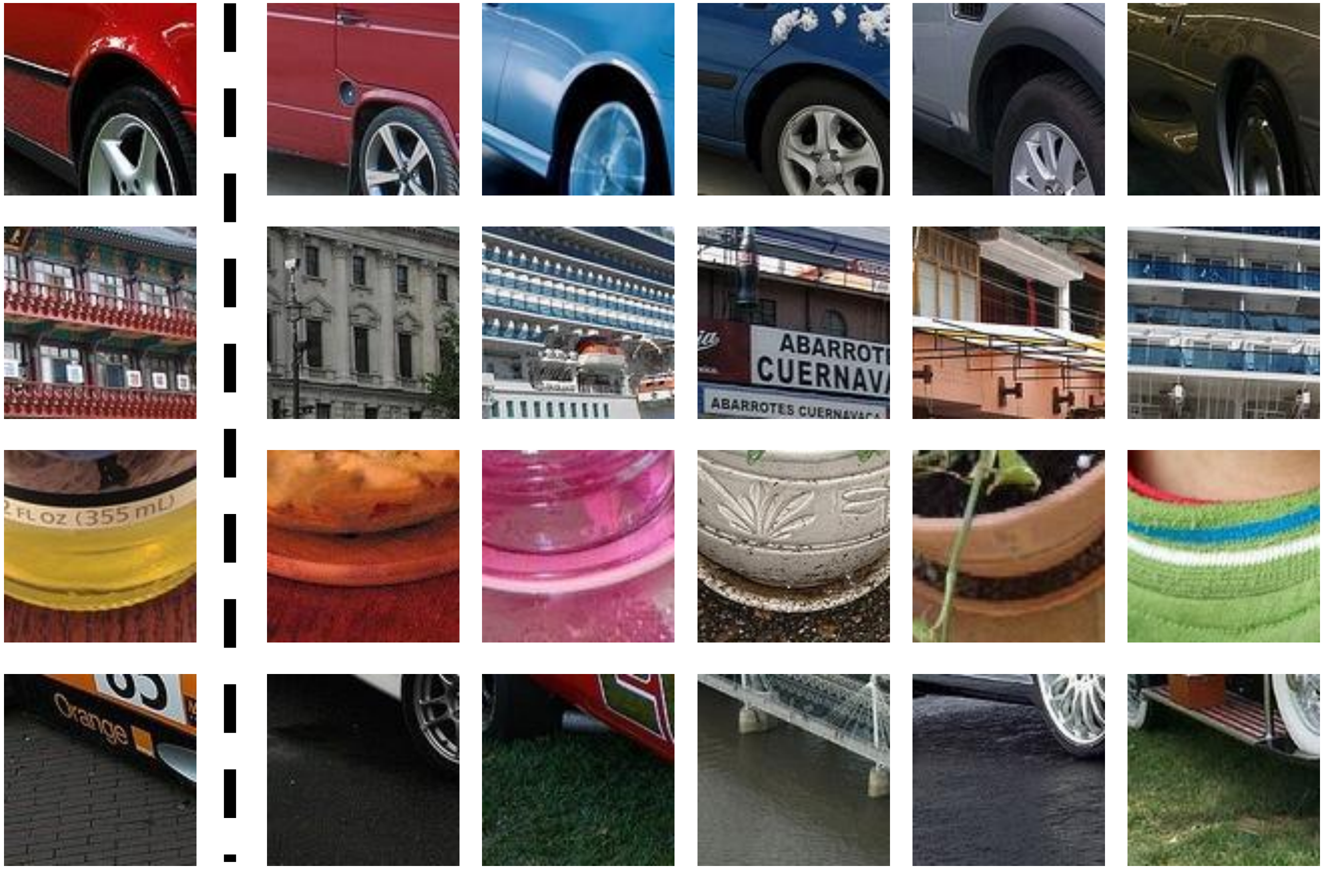
VOC 2007 Performance

(pretraining for R-CNN)

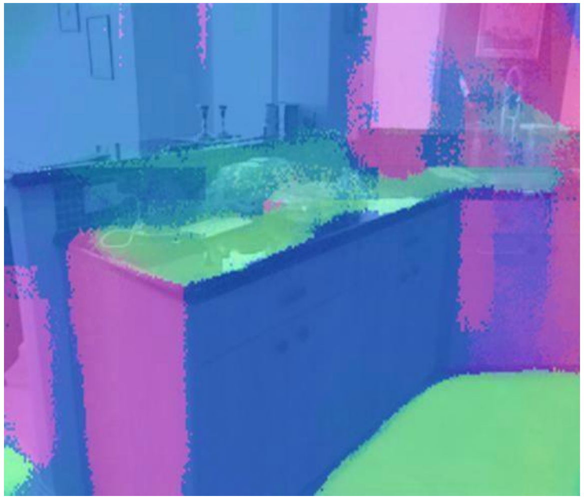


[Krähenbühl, Doersch, Donahue & Darrell, “Data-dependent Initializations of CNNs”, 2015]

Capturing Geometry?



Surface-normal Estimation

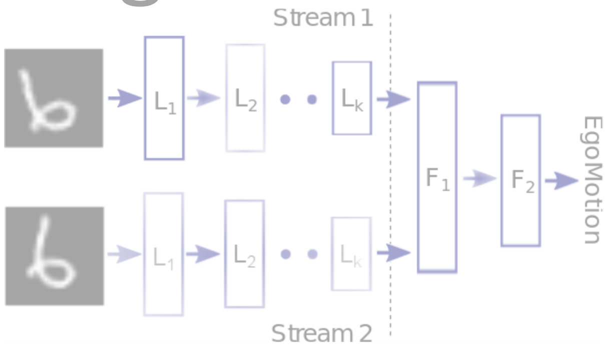


	Error (Lower Better)		% Good Pixels (Higher Better)		
Method	Mean	Median	11.25°	22.5°	30.0°
No Pretraining	38.6	26.5	33.1	46.8	52.5
Ours	**33.2**	21.3	36.0	51.2	57.8
ImageNet Labels	33.3	**20.8**	**36.7**	**51.7**	**58.1**

So, do we need semantic labels?

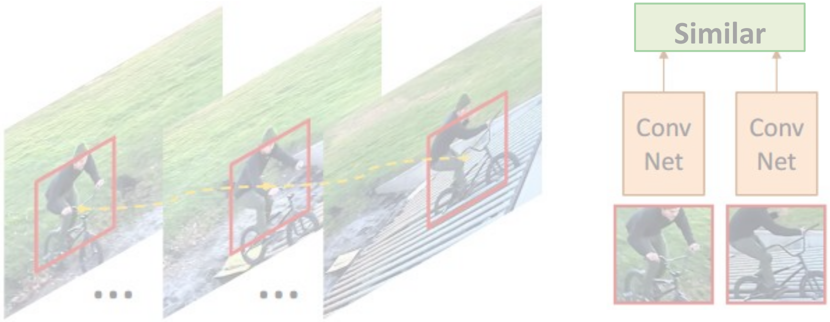
“Self-Supervision” and the Future

Ego-Motion



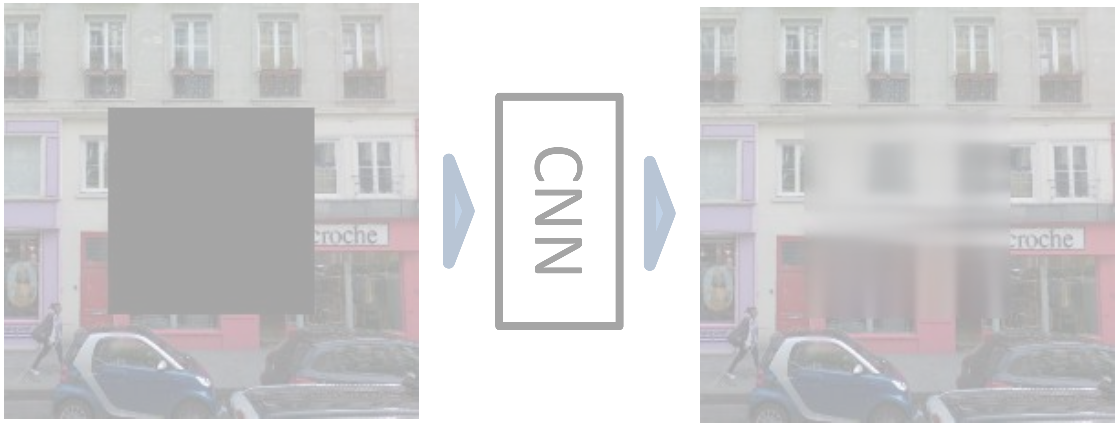
[Agrawal et al. 2015; Jayaraman et al. 2015]

Video



[Wang et al. 2015; Srivastava et al 2015; …]

Context



[Doersch et al. 2014; Pathak et al. 2015; Isola et al. 2015]

A Neural Algorithm of Artistic Style

Leon A. Gatys, Alexander S. Ecker, Matthias Bethge.

CVPR 2016.

See pdf on course website