CAFFE: CONVOLUTIONAL ARCHITECTURE FOR FAST FEATURE EMBEDDING

Jia Y., Shelhamer E.*, Donahue J., Karayev S., Long J., Girshick R., Guadarrama S., Darrell T.
UC Berkeley EECS
{jiayq,shelhamer,jdonahue,sergeyk,jonlong,rbg,sguada,trevor}@eecs.berkeley.edu

caffe.berkeleyvision.org

Abstract

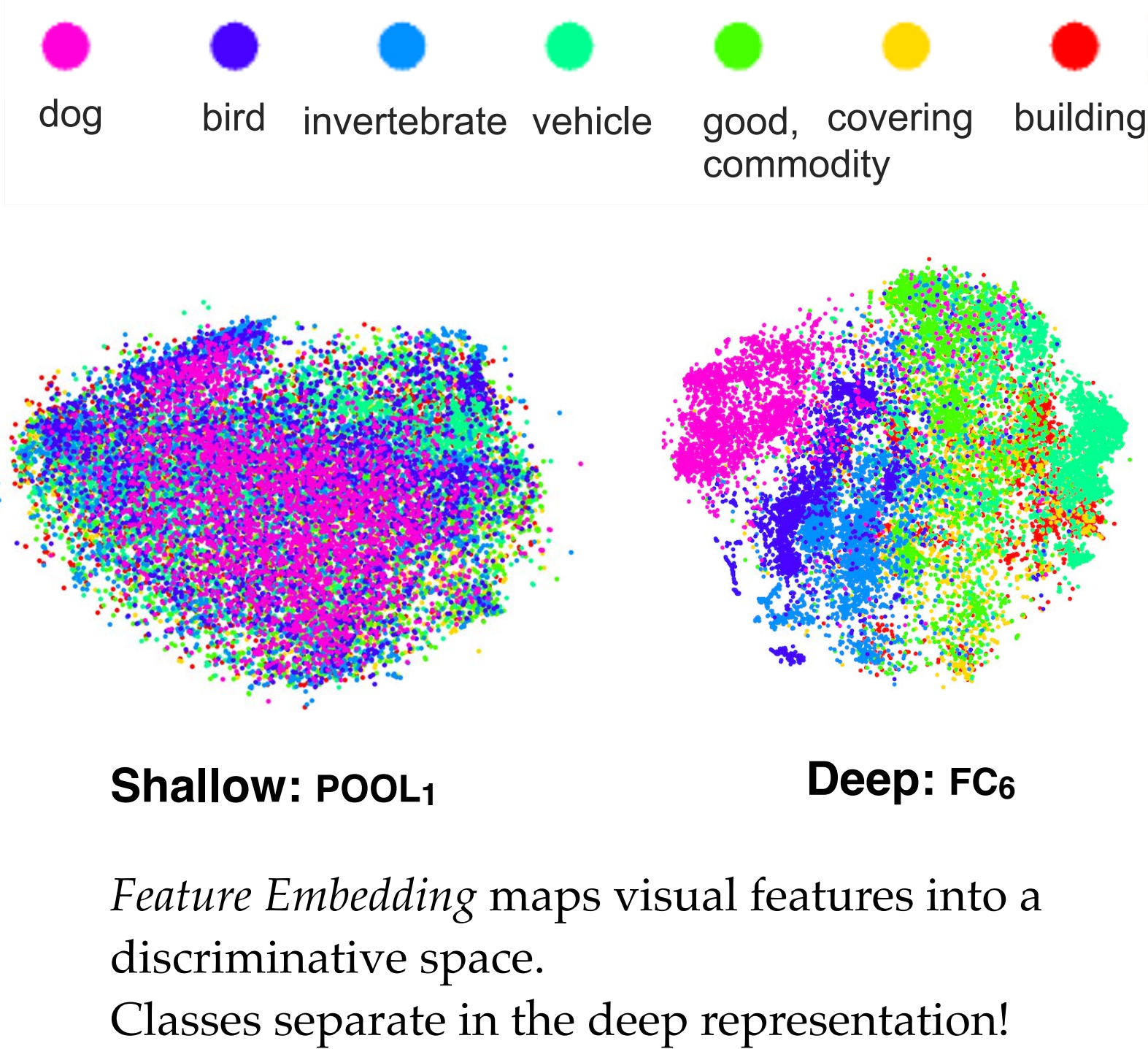Caffe provides vision scientists and practitioners with

- a framework for state-of-the-art deep learning algorithms
- a collection of reference models
- processing of > 40 million images a day
  ≈ 2.5 ms per image on a single GPU
- Python and MATLAB bindings for experiments

in a fully open-source BSD-licensed C++ / CUDA library.

Framework Comparison

Framework	License	Core language	Binding(s)	CPU	GPU	Open source	Training	Pretrained models	Development
Caffe	BSD	C++	Python, MATLAB	✓	✓	✓	✓	✓	distributed
cuda-convnet [7]	unspecified	C++	Python		✓	✓	✓		discontinued
Decaf [2]	BSD	Python		✓		✓	✓	✓	discontinued
OverFeat [9]	unspecified	Lua	C++,Python	✓				✓	centralized
Theano/Pylearn2 [4]	BSD	Python		✓	✓	✓	✓		distributed
Torch7 [1]	BSD	Lua		✓	✓	✓	✓		distributed

The Unreasonable Effectiveness of Deep Features

dog · bird · invertebrate · vehicle · good, covering commodity · building

Shallow: POOL₁   **Deep:** FC₆

Feature Embedding maps visual features into a discriminative space.
Classes separate in the deep representation!

Seamless CPU / GPU Switch

Prototype · Training · Deployment

All with the same code.

Models

LogReg ↑

LeNet →

ImageNet, Krizhevsky 2012 →
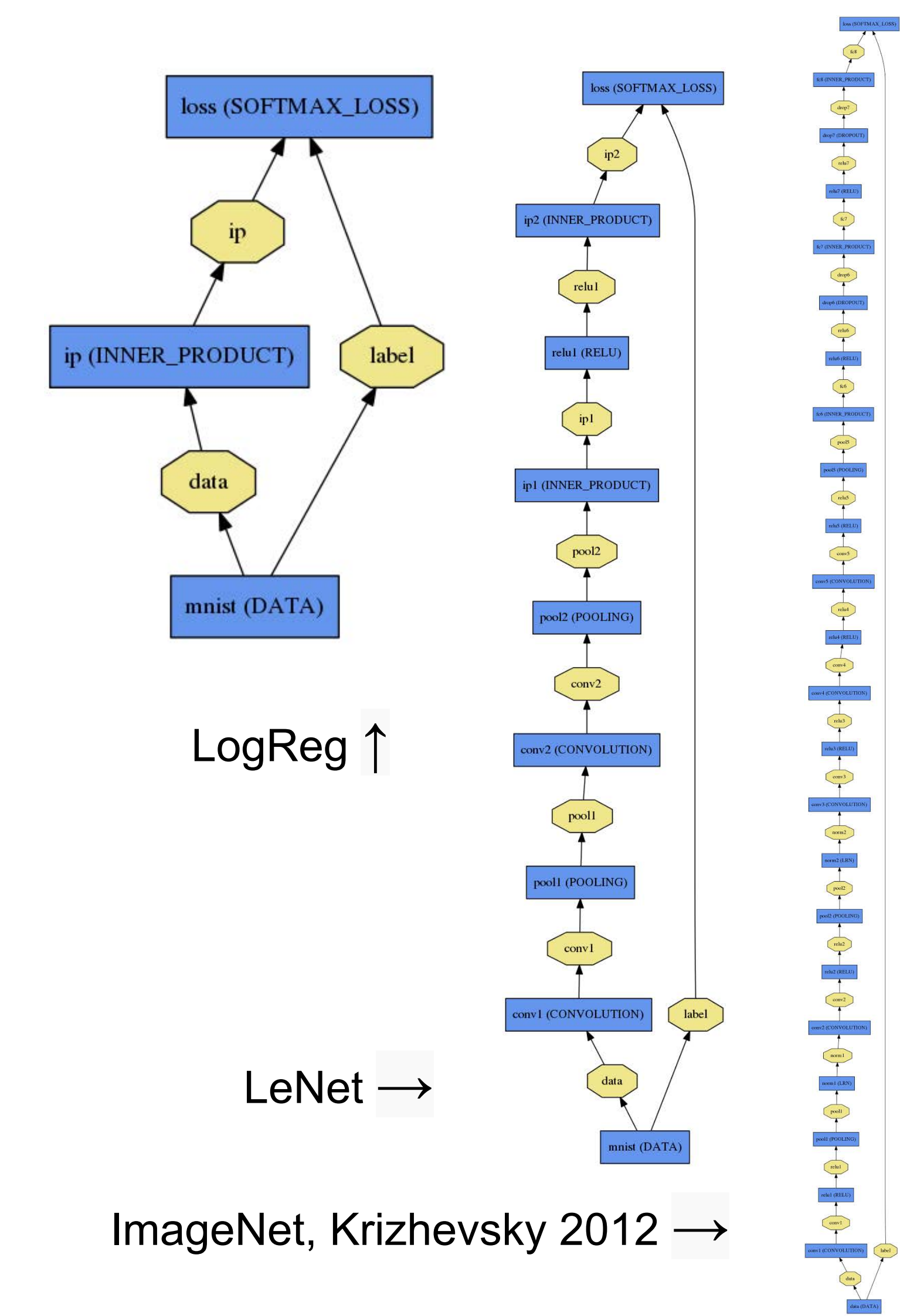
AlexNet: ImageNet Classification
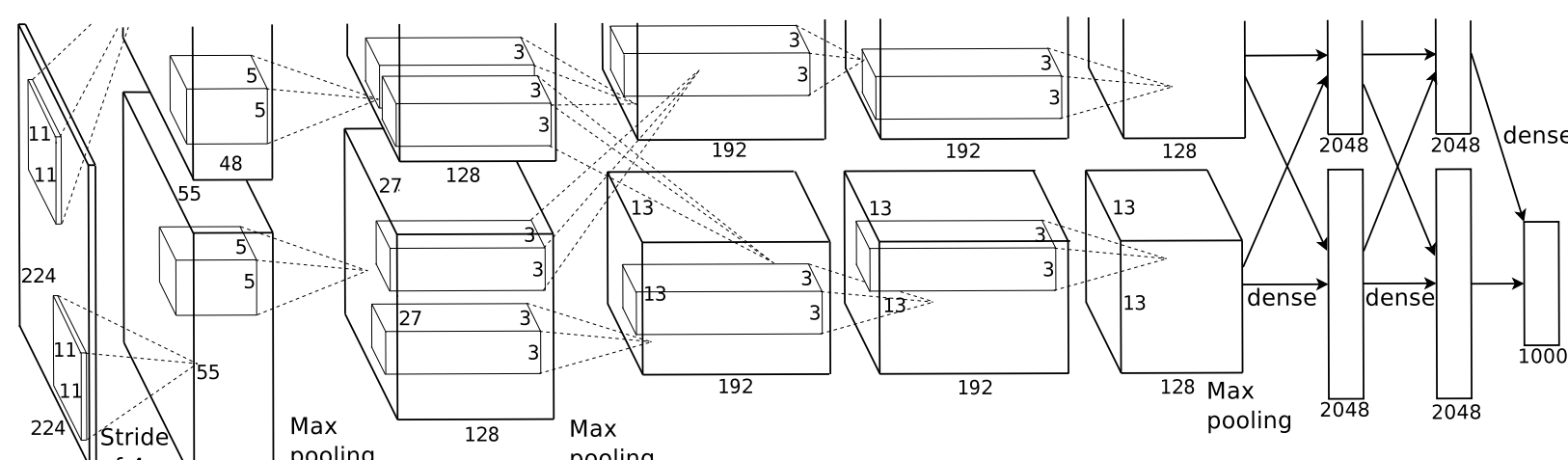
R-CNN: *Regions with CNN features*

warped region

aeroplane? no.
person? yes.
tvmonitor? no.

CNN

1. Input image   **2.** Extract region proposals (~2k)   **3.** Compute CNN features   **4.** Classify regions
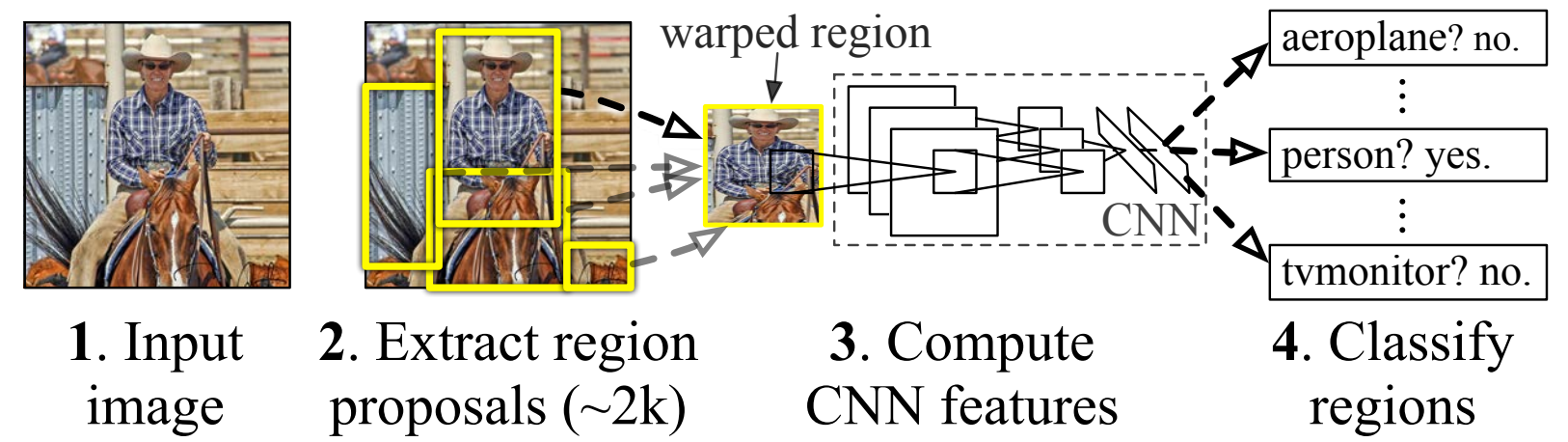
Finetune Your Own Model

Simply change a few lines in the layer definitions.

```
layers {
  name: "data"
  type: DATA
  data_param {
    source: "ilsvrc12_train_leveldb"
    mean_file: "../../data/ilsvrc12/i
    batch_size: 256
    crop_size: 227
    mirror: true
  }

layers {
  name: "fc8"
  type: INNER_PRODUCT
  blobs_lr: 1
  blobs_lr: 2
  weight_decay: 1
  weight_decay: 0
  inner_product_param {
    num_output: 1000
```

```
layers {
  name: "data"
  type: DATA
  data_param {
    source: "dogs-vs-cats-leveldb"
    mean_file: "../../data/ilsvrc12/i
    batch_size: 256
    crop_size: 227
    mirror: true
  }

layers {
  name: "fc8-dogcat"
  type: INNER_PRODUCT
  blobs_lr: 1
  blobs_lr: 2
  weight_decay: 1
  weight_decay: 0
  inner_product_param {
    num_output: 2
```

Input: difference source   Output: different classifier

References

[1] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A MATLAB-like environment for machine learning. In BigLearn, NIPS Workshop, 2011.

[2] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In ICML, 2014.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, 2014.

[4] I. Goodfellow, D. Warde-Farley, P. Lamblin, V. Dumoulin, M. Mirza, R. Pascanu, J. Bergstra, F. Bastien, and Y. Bengio. Pylearn2: a machine learning research library. arXiv preprint 1308.4214, 2013.

[5] S. Karayev, M. Trentacoste, H. Han, A. Agarwala, T. Darrell, A. Hertzmann, and H. Winnemoeller. Recognizing image style. arXiv preprint 1311.3715, 2013.

[6] A. Krizhevsky. cuda-convnet. https://code.google.com/p/cuda-convnet/, 2012.

[7] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, 2012.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[9] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In ICLR, 2014.

[10] M. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. arXiv preprint 1311.2901, 2013.

Acknowledgements

caffe.berkeleyvision.org