DenseNet: Efficient Computation of Deep Neural Networks for Object Detection

Forrest Iandola, Matt Moskewicz, Sergey Karayev, Ross Girshick, Kurt Keutzer, and Trevor Darrell
forresti@eecs.berkeley.edu

Introduction

- Deep learning has revolutionized object classification [4] and detection [5]
- The state-of-the-art object detector [5] computes CNN descriptors for each candidate window, which is slow
- We achieve a 10x speedup in feature extraction by computing CNN pyramids instead of computing per-window features

Driving Applications

In addition to traditional object detection…

Acknowledgements

Thanks to Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergio Guadarrama, and the BVLC team for their suggestions and their work on the Caffe project.

2005-2012: Object Detection with HOG [1] and Parts Models (e.g. [2])

- Object detection accuracy: ~33% mAP on PASCAL VOC 2007
- Speed: 4 frames per second at 640x480 on a multicore CPU [7]
- HOG computation and sliding-window template matching have similar amounts of overhead

2013: Region-based CNNs (R-CNN) for Object Detection [5]

- State-of-the-art object detection on PASCAL VOC 2007:
  - pool5 layer 44% mAP
  - fc7 layer 54% mAP
- Speed: 0.1 frames per second on an NVIDIA K20 GPU
- Computation time is dominated by ConvNet feature extraction on region proposals

2014: Dense CNNs descriptors for efficient object detection

- Speed: 1 frame per second on an NVIDIA K20 GPU
  - 10x speedup over R-CNN
  - PASCAL VOC 2007 object detection results:
    - pool5 layer 42% mAP (near R-CNN state-of-art)

References