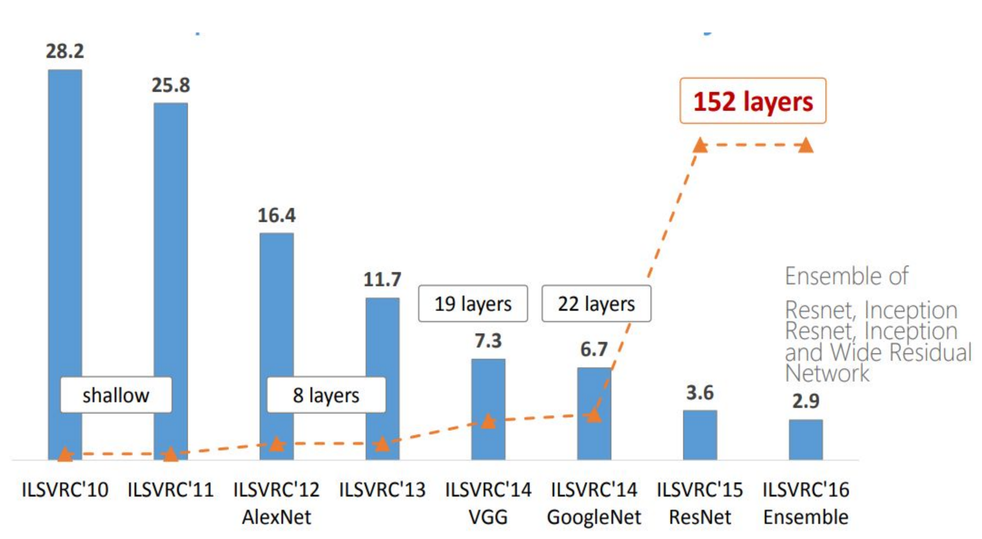
CNNs for real-time object detection on mobiles

Rajat Chand
rajatc@uw.edu
CSE 576 (Spring '18)

Introduction



ImageNet Classification Top-5 error (%)

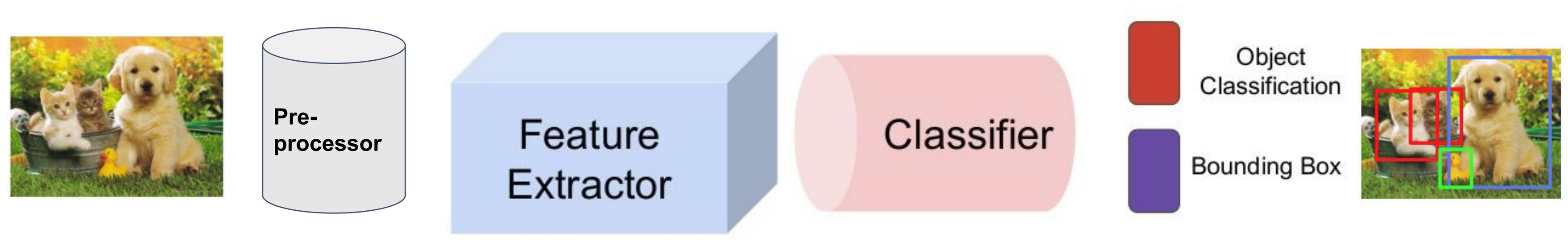
Strategies to make CNNs more efficient

- Shallow networks
- Compressing pre-trained networks
- Designing compact layers
- Quantizing parameters
- Network binarization

Model: Leeco Le 2
Chipset: Snapdragon 652
CPU: 4 x 1.8 GHz Cortex- A72 +
4 x 1.4 GHz Cortex-A53
GPU: Adreno 510
Memory: 3 GB RAM
Storage: 32 GB
OS: Android 6

Classification | **Classification + Localization** | **Object Detection**

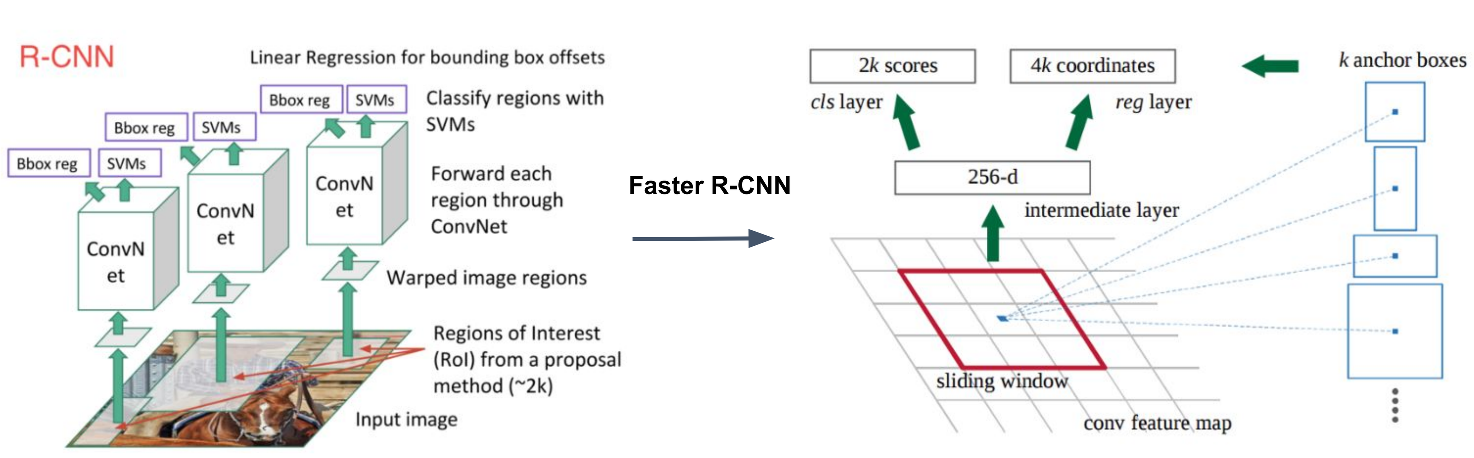
Object Detection Pipeline

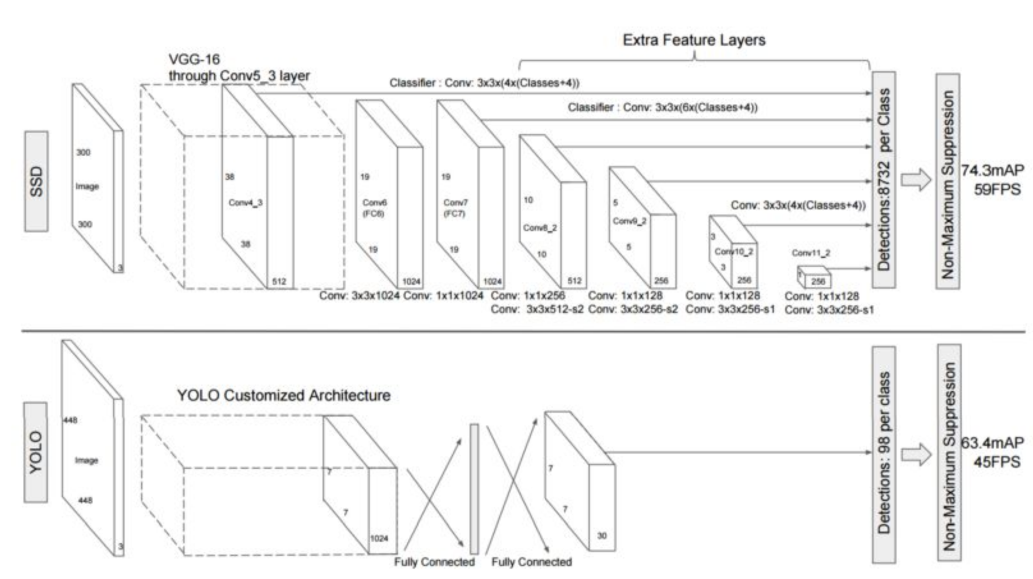


- **Image-resizing**
- **Image pixel normalization**
- **Data - Augmentation**
  - **flip**
  - **Random rotations**

- **VGG**
- **Inception**
- **Resnet**
- **Inception Resnet**
- **Googlenet**
- **MobileNet**

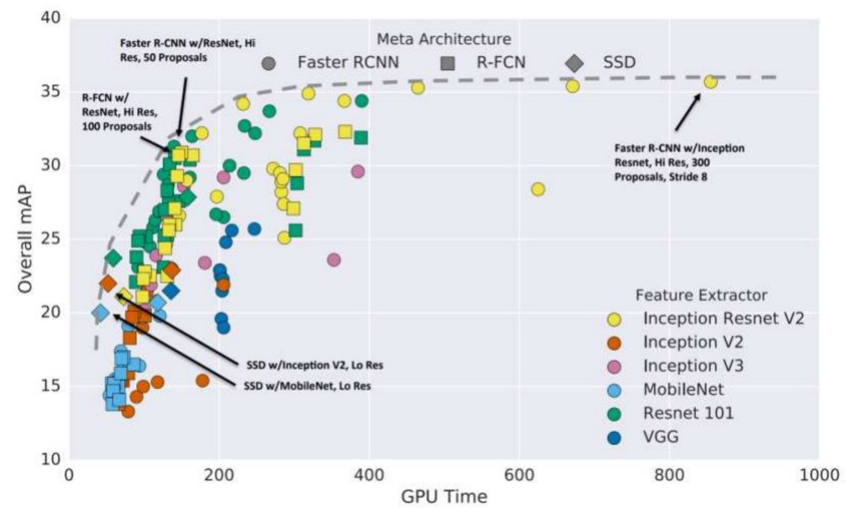
- **Single Shot Detector (SSD and YOLO)**
- **Faster R-CNN**
- **R-FCN**

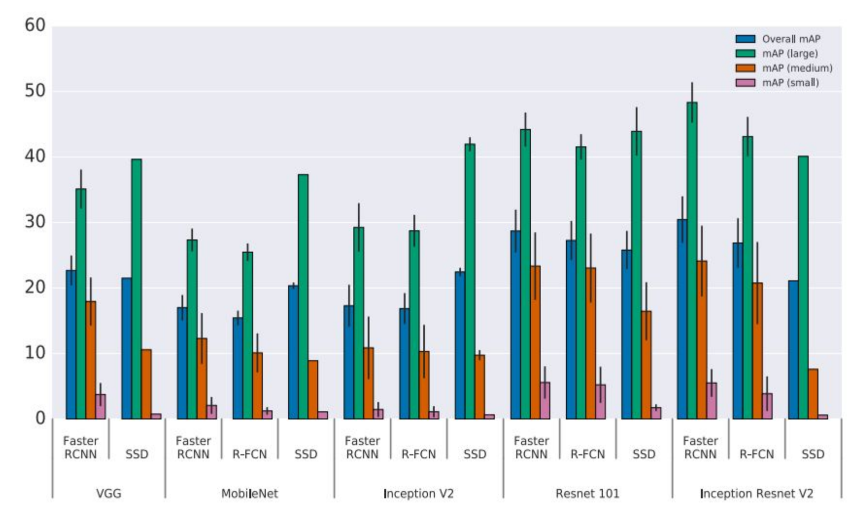




Evaluation Metrics and Test

- **Mean Average Precision (mAP)**
  - thresholding based on intersection of union score
  - average over all class predictions
  - higher the better
- **Inference time:** Preprocessing + prediction time per image
- **Memory usage:** Model's gpu/cpu usage while prediction





Transfer Learning and Implementation

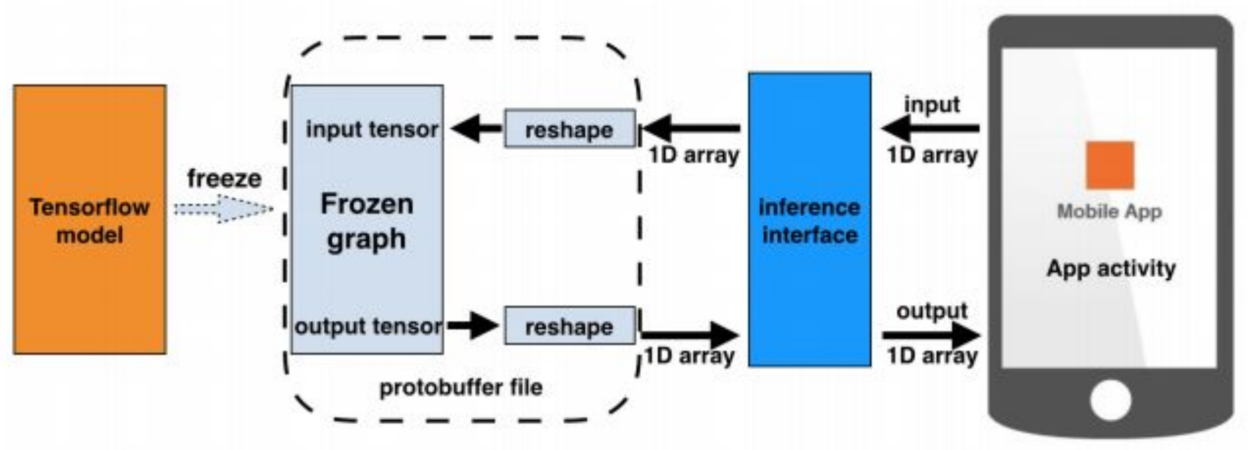
"Don't Be A Hero"
- Andrej Karpathy

Dataset

- 7481 labelled images
- 1240 x 375 pixels
- 11 classes

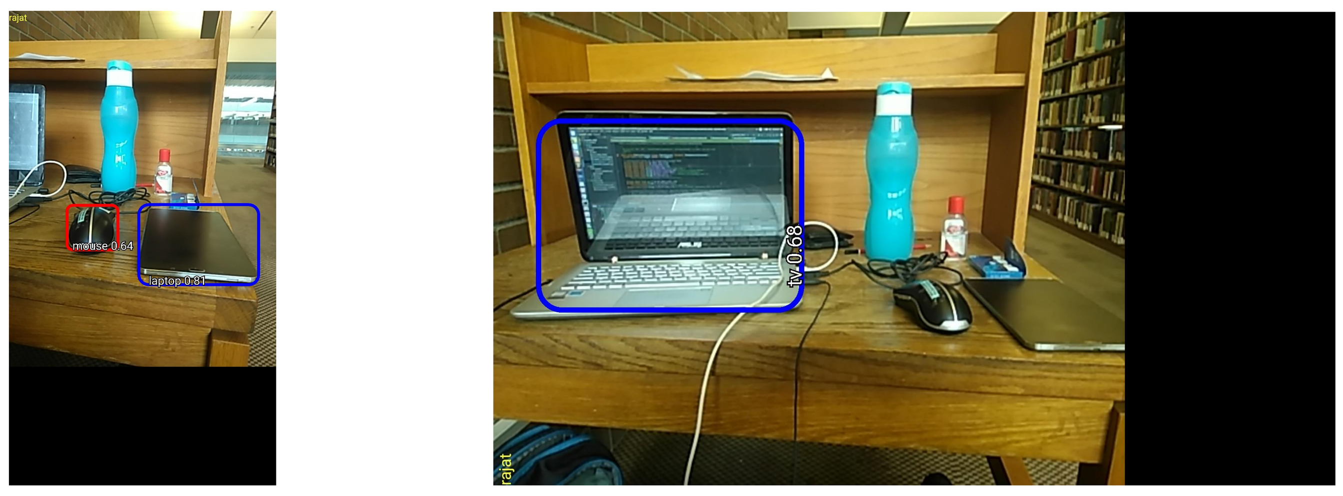
Transfer Learning	**Models**
• Finetune existing pre-trained models • Freeze the base extractor and train the classifier • Train a linear classifier at the final layer	• Mobilenet-SSD • Tiny-yolo • Incepted-SSD • Inception-Faster R-CNN

Android App Architecture

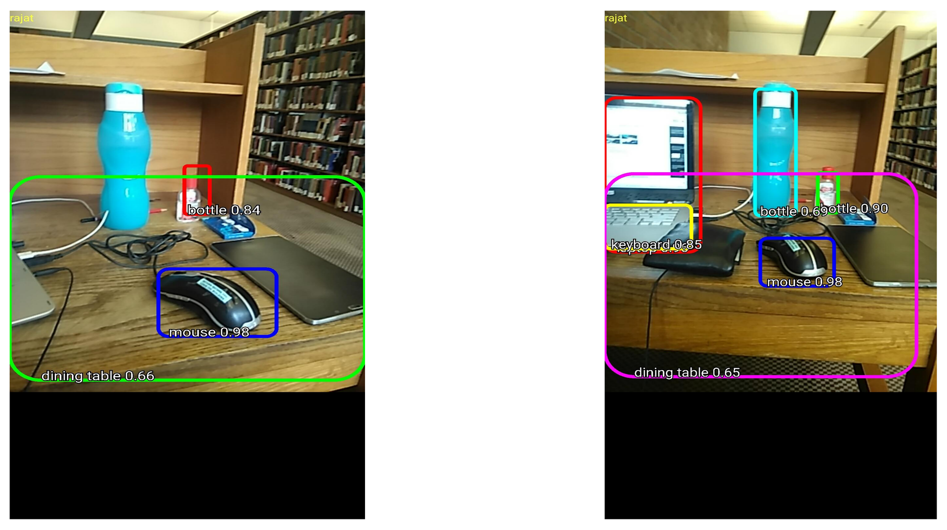


Results and Discussions

- **Mobilenet based SSD trained on COCO dataset**



- **Inception based R-CNN trained on COCO dataset**



Work in Progress

1. Benchmarking the models on the mobile device
2. Fine-tuning MobileNet-SSD on the Kitti dataset

References

1. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S. and Murphy, K., 2017, July. Speed/accuracy trade-offs for modern convolutional object detectors. In *IEEE CVPR*.
2. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
3. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J. and Keutzer, K., 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.
4. Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2016. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
5. Redmon, J. and Farhadi, A., 2017. YOLO9000: better, faster, stronger. *arXiv preprint*.
6. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., 2016, October. Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.
7. Geiger, A., Lenz, P. and Urtasun, R., 2012, June. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 3354-3361). IEEE.
8. He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
9. CSE/EE 576 Lecture Slides
10. CS231n Lecture Slides