# GOOGLENET: GOING DEEPER WITH CONVOLUTIONS

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed,
Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich

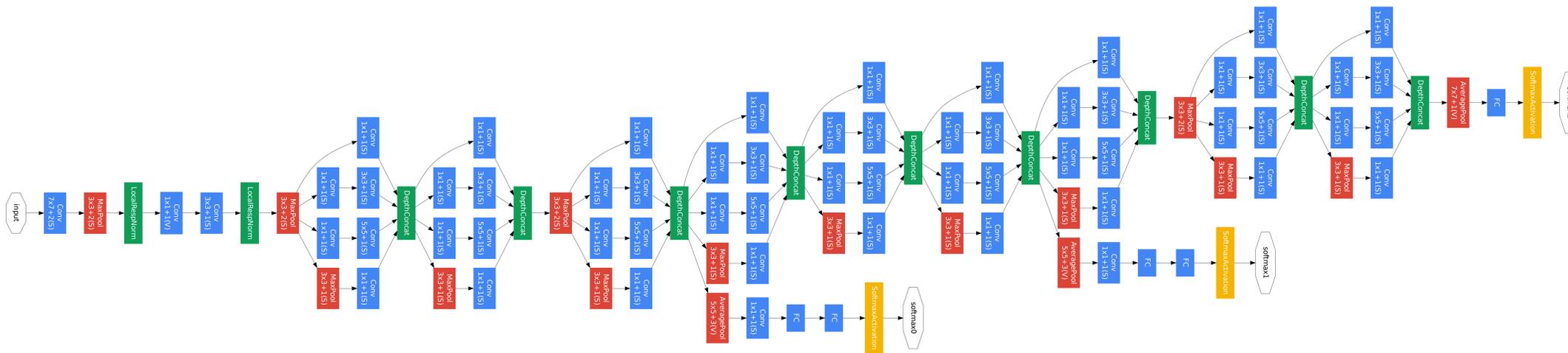THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

## GOOGLENET



**Figure 1:** GoogLeNet [1], a much deeper, wider, and sparser network, with 12 times fewer parameters than AlexNet [2]. (Figure is from [1])
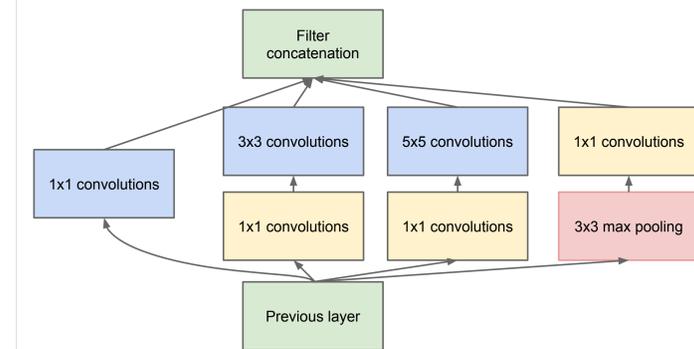
## INCEPTION MODULE



**Figure 2:** Inception module, the building block for GoogLeNet. 1x1 convolution does dimension reduction and accounts for rectified linear units (ReLU). 3x3 and 5x5 convolution deals with different scales, and 3x3 max pooling introduces invariance. (Fig. from [1])

## OVERVIEW

- Feature representation is crucial for vision. Change from hand-crafted feature (e.g. SIFT, Bag-of-Words) to feature learned from data (e.g. CNN) can improve performance on various tasks.

- CNN becomes a hot topic for vision task since 2012. Going deeper and wider is the trend.

- Computation time is important. Using 1x1 convolution as dimension reduction makes GoogLeNet possible to train fast enough and a success in ILSVRC2014 classification and detection track.

- Detection includes two major blocks: region proposal + CNN (R-CNN [3]).

- Pretrain CNN using classification images, and fine-tune on detection images using bounding boxes can improve detection performance dramatically using GoogLeNet.
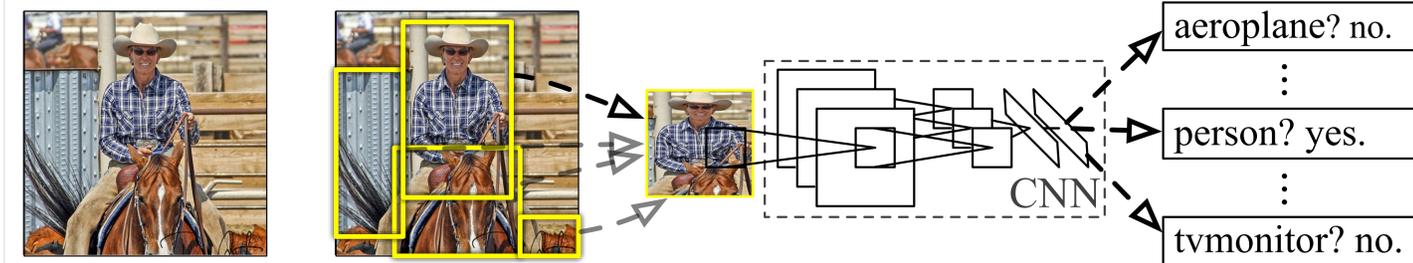
## R-CNN FOR DETECTION



aeroplane? no.

person? yes.

tvmonitor? no.

**Figure 3:** R-CNN for detection. (Figure is from [3])

## CLASSIFICATION AND DETECTION EXAMPLES



Groundtruth:
- **coffee mug**

GoogLeNet:
- **table lamp**
- **lamp shade**
- printer
- projector
- desktop computer

Groundtruth:
- **Police car**

GoogLeNet:
- laptop
- hair drier
- binocular
- ATM machine
- seat belt

Groundtruth:
- **hay**

GoogLeNet:
- **sorrel (horse)**
- **hartebeest**
- **Arabian camel**
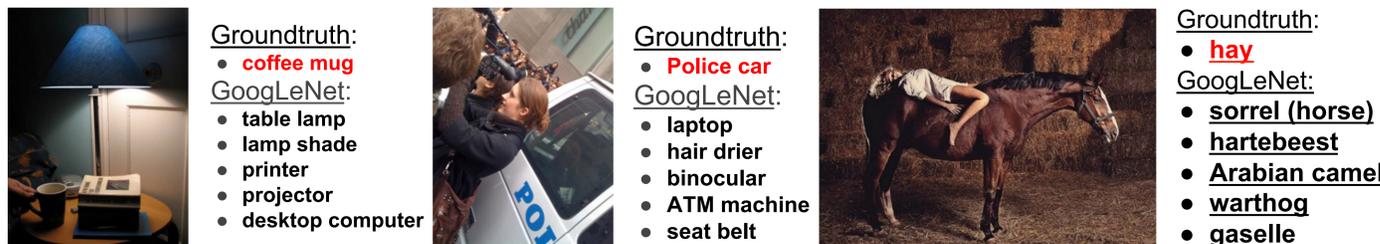- **warthog**
- **gaselle**

**Figure 4:** Classification examples using GoogLeNet.



**Figure 5:** Detection examples using GoogLeNet.

## CLASSIFICATION RESULTS

| Place | Team Name | Top5 Error |
| --- | --- | --- |
| 2012 1st | SuperVision | 15.3% |
| 2013 1st | Clarifai | 11.2% |
| 2014 3rd | MSRA | 8.1% |
| 2014 2nd | VGG | 7.3% |
| 2014 1st | **GoogLeNet** | **6.7%** |

**Table 1:** ILSVRC classification results through years.

## DETECTION RESULTS

| Place | Team Name | mAP |
| --- | --- | --- |
| 2013 1st | UvA-Euvision | 22.6% |
| 2014 3rd | Deep Insight | 40.5% |
| 2014 2nd | CUHK DeepID-Net | 40.7% |
| 2014 1st | **GoogLeNet** | **43.9%** |

**Table 2:** ILSVRC detection results through years.

## REFERENCES

[1] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[3] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv preprint arXiv:1311.2524*, 2013.