Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning

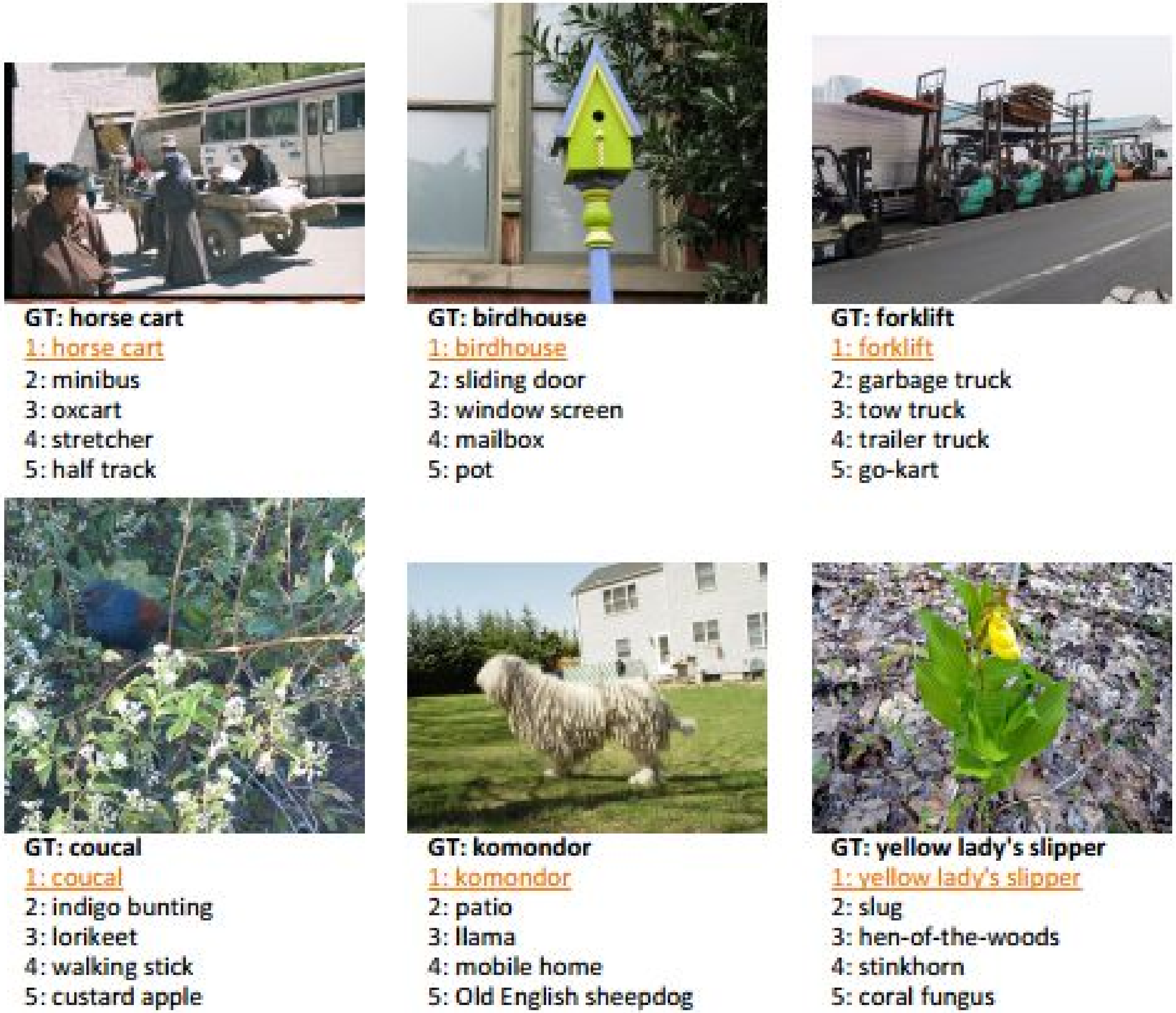
Christian Szegedy, Sergey Ioffe and Vincent Vanhoucke

Presented by: Iman Nematollahi

Outline

- Introduction
- Previous architectures:
  - Inception-v1: Going deeper with convolutions
  - Inception-v2: Batch Normalization
  - Inception-v3: Rethinking the Inception architecture
  - Deep Residual Learning for Image Recognition
- Inception-v4
- Inception-ResNet
- Experimental Results

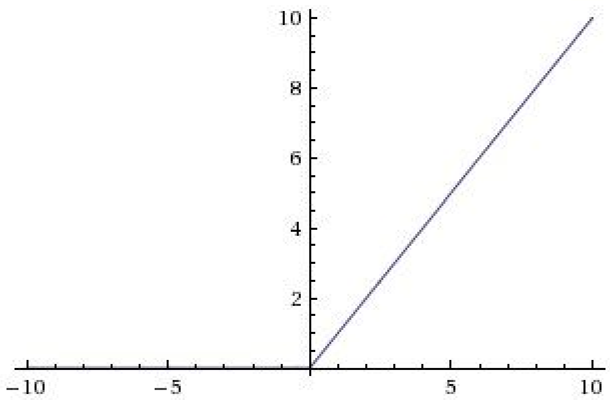
Introduction



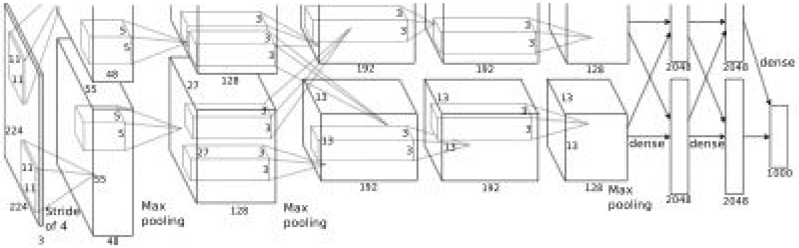
http://www.iamwire.com/2015/02/microsoft-researchers-claim-deep-learning-system-beat-humans/109897

Background

IMAGENET



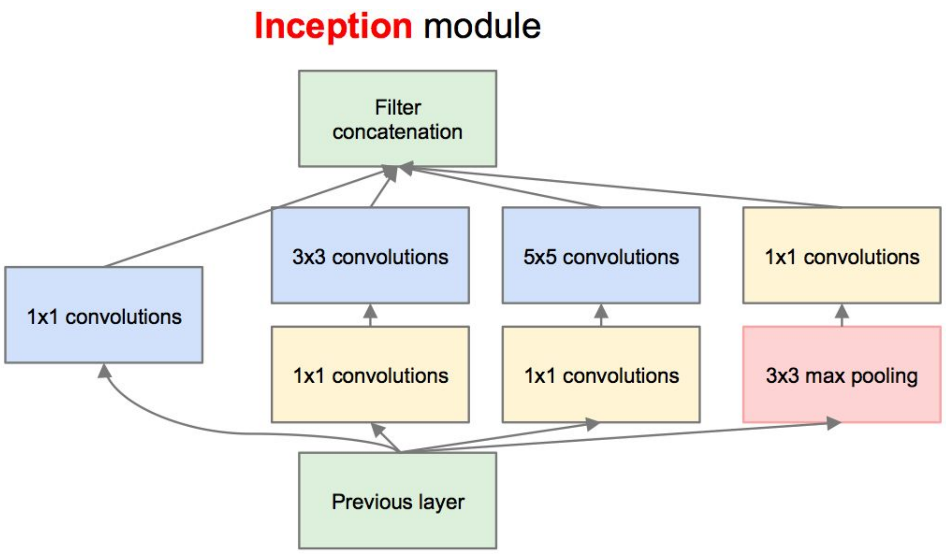
ReLU activation function



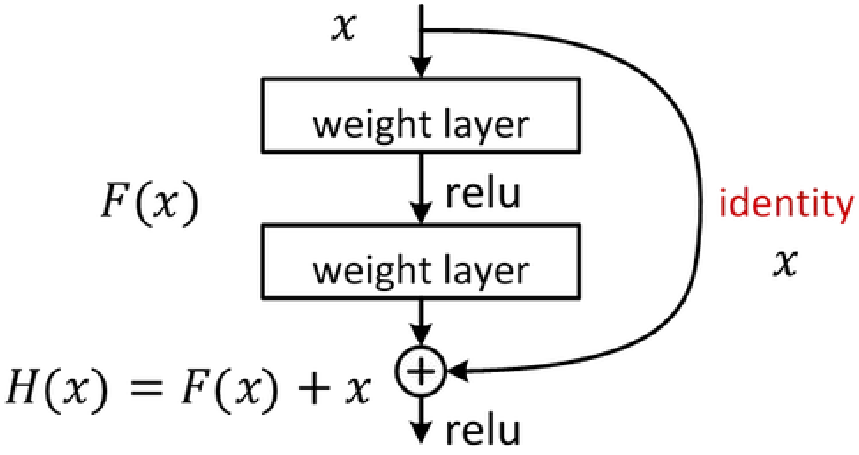
Alex-net architecture

Two Powerful Networks

- Inception Network



- Deep Residual Network



Inception-v1

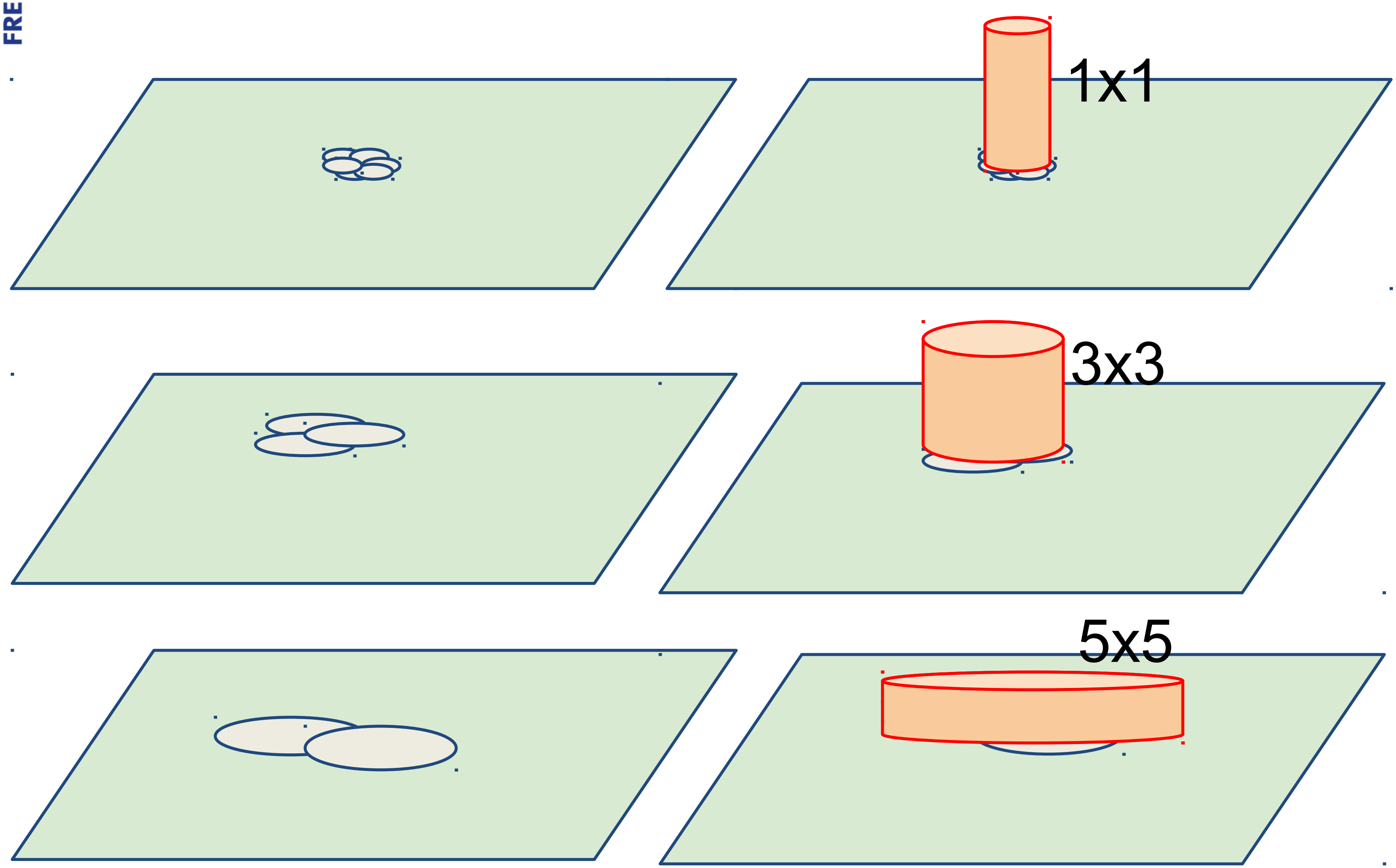
Drawbacks of going deeper:

1. Overfitting
2. Increased use of computational resources

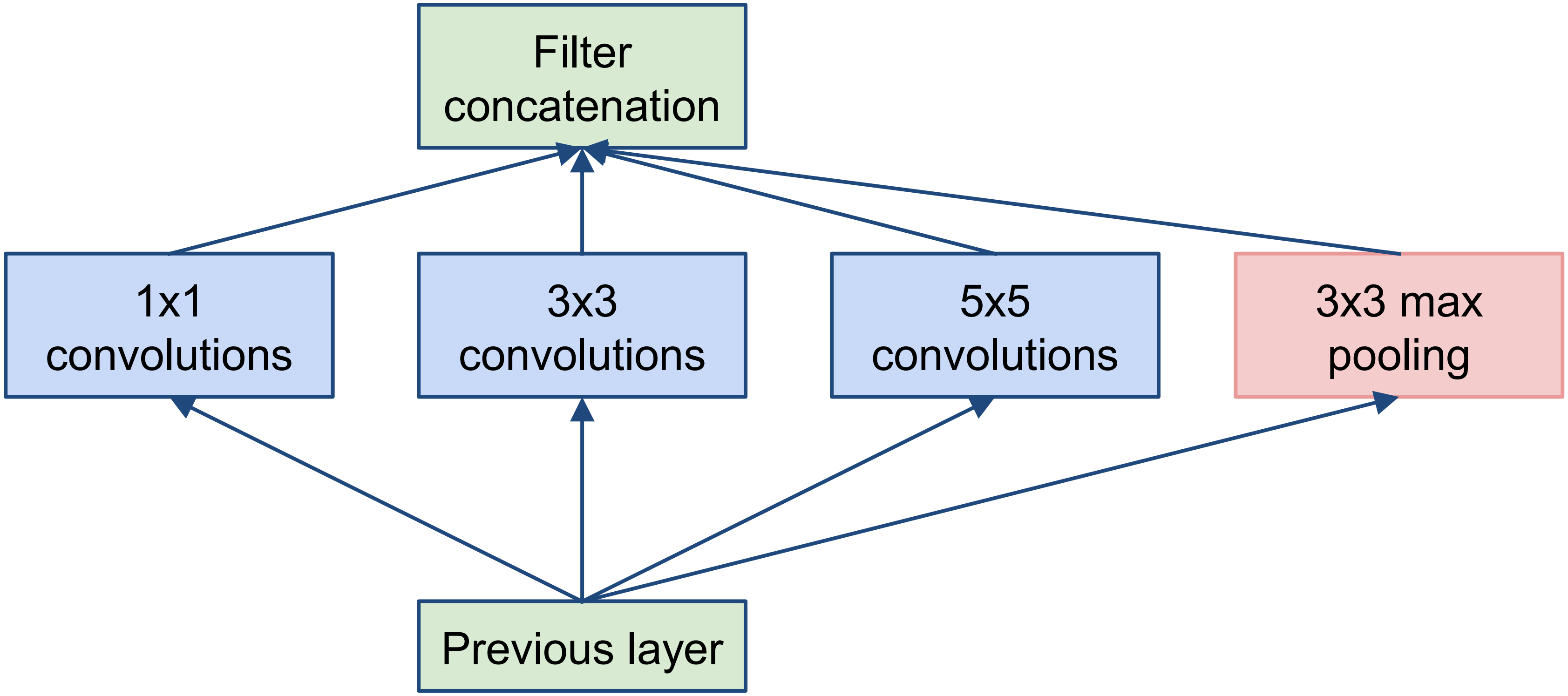
Proposed solution:

- Moving from fully connected to sparsely connected architectures
- Clustering sparse matrices into relatively dense submatrices

Inception-v1: Going deeper with

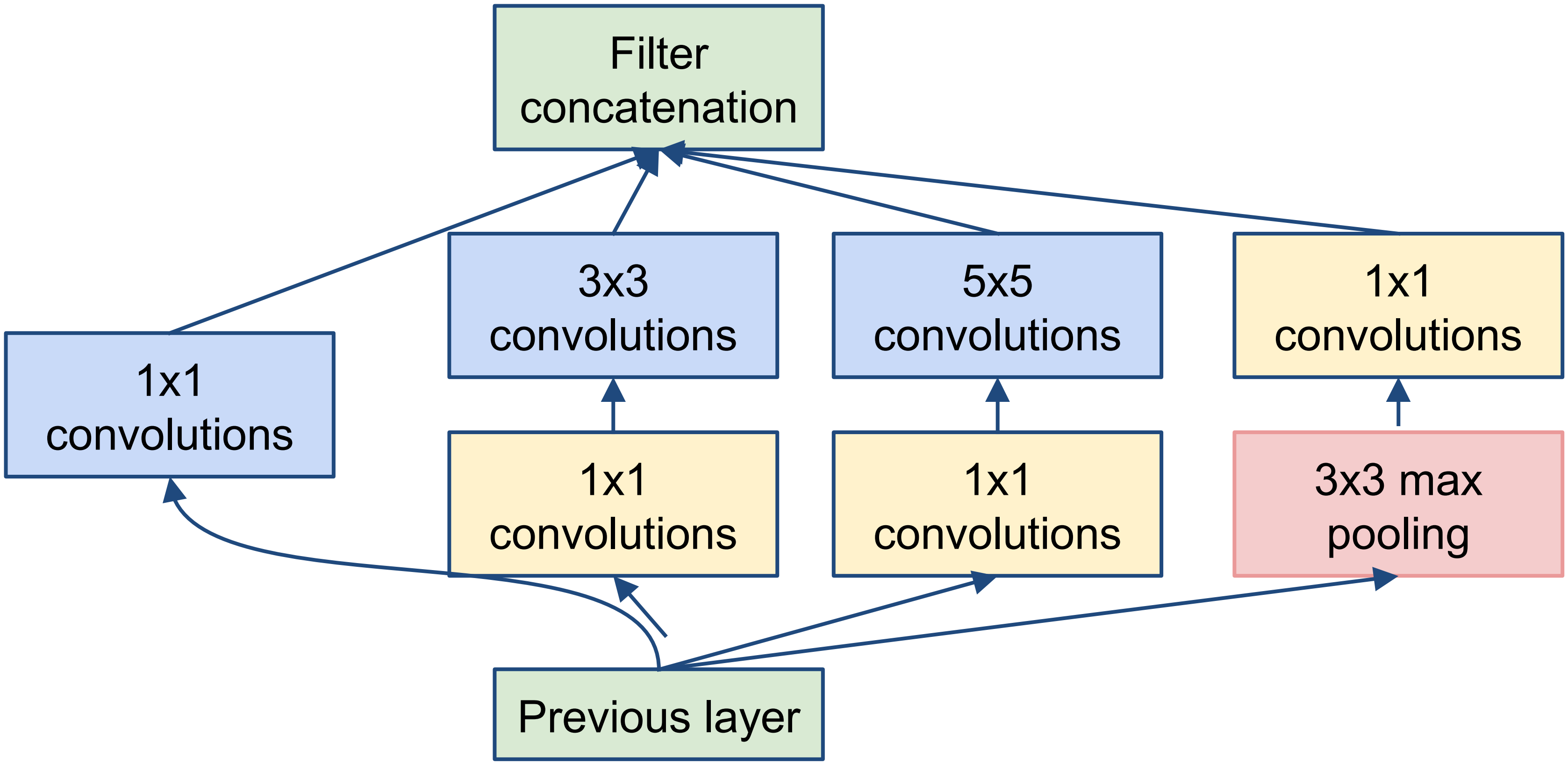


Inception-v1: Going deeper with



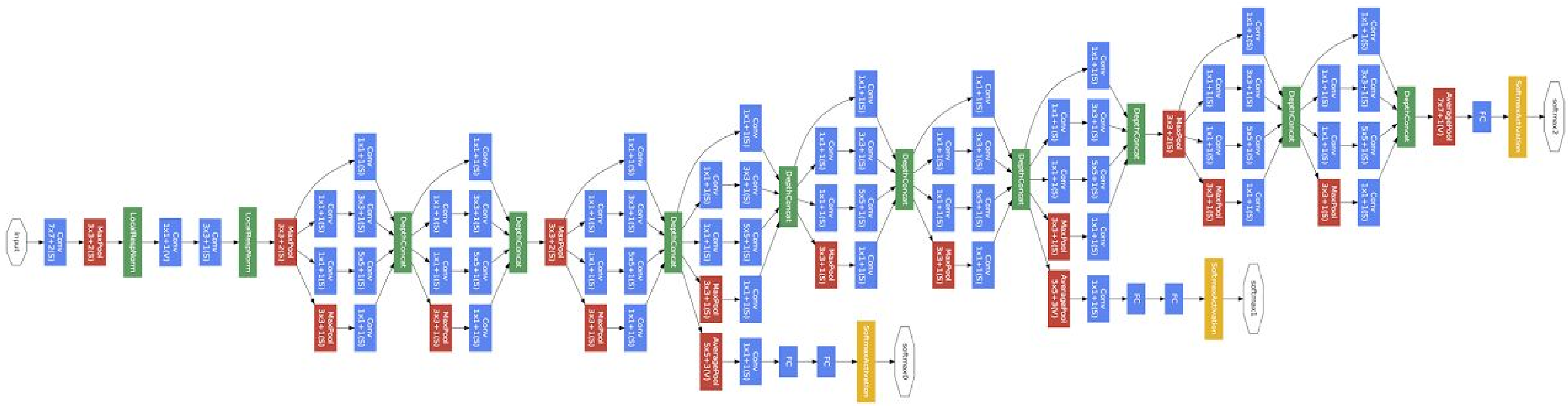
Inception-v1: Going deeper with

Inception module



Inception-v1: Going deeper with

GoogLeNet

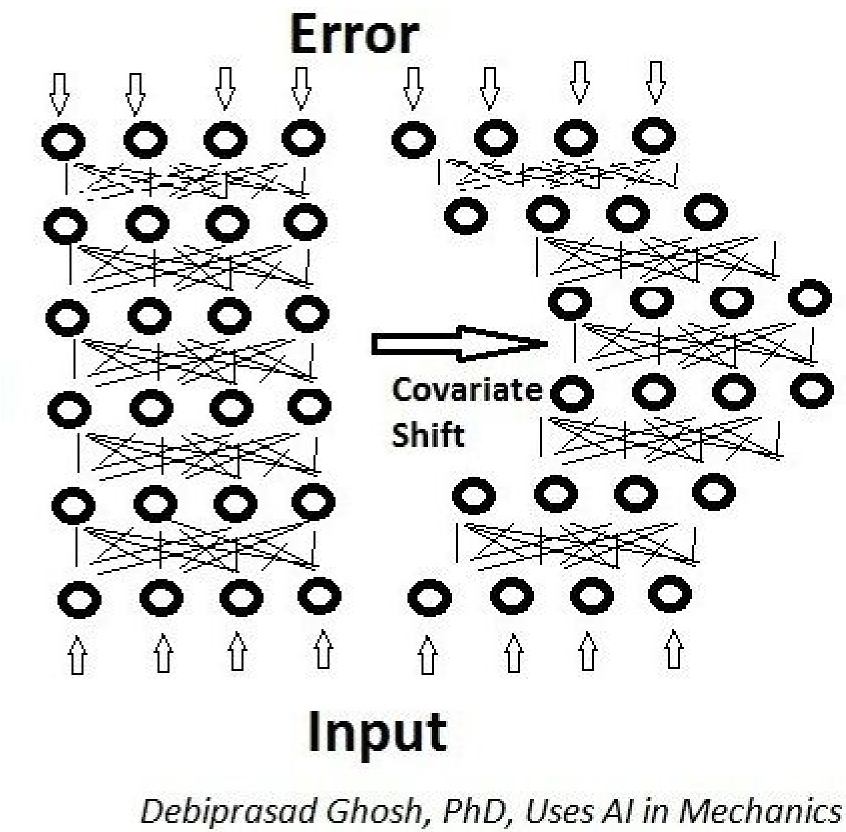


Convolution
Pooling
Softmax
Other

Inception-v2: Batch Normalization

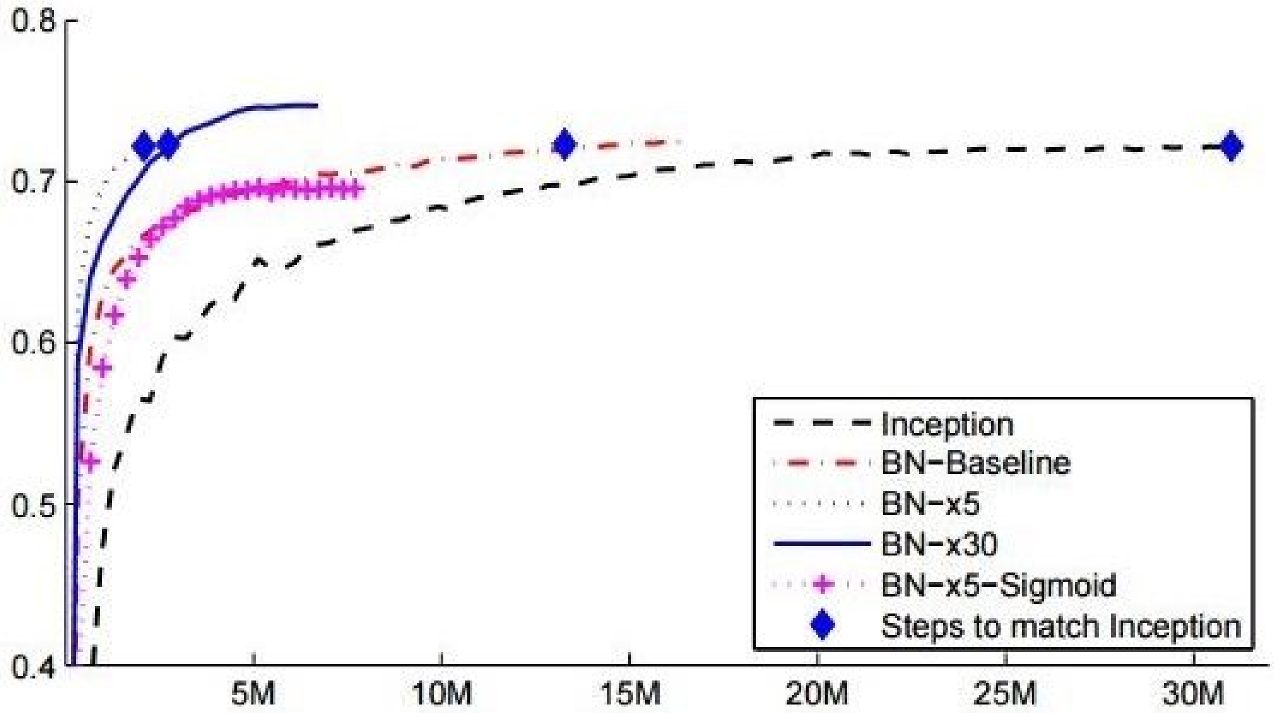
- Problem of internal covariate shift

- Introducing Batch Normalization:
  - Faster learning
  - Higher overall accuracy



https://www.quora.com/Why-does-batch-normalization-help

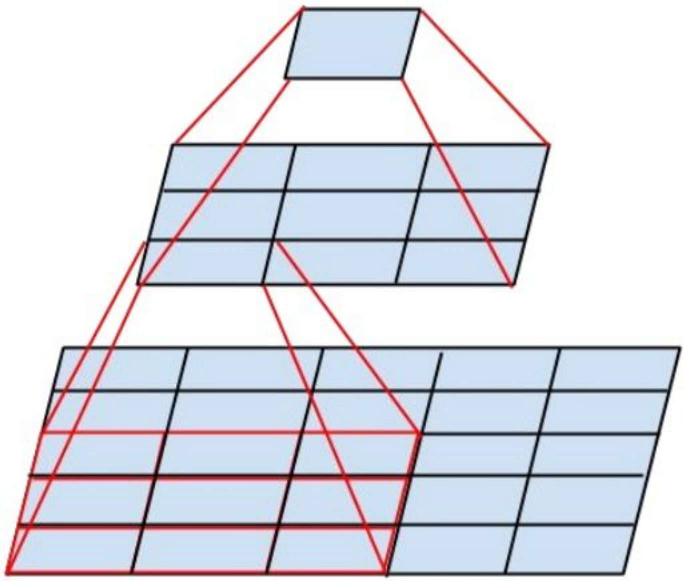
Inception-v2: Batch Normalization



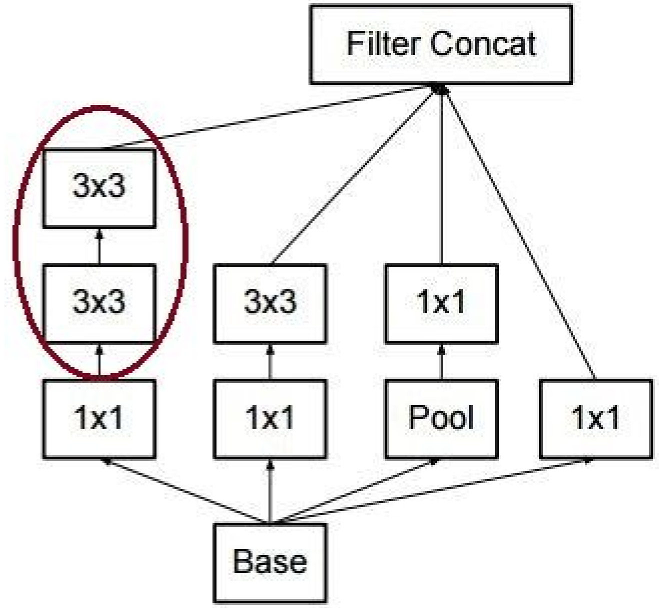
Model	Steps to 72.2%	Max accuracy
Inception	$31.0 \cdot 10^6$	72.2%
BN-Baseline	$13.3 \cdot 10^6$	72.7%
BN-x5	$2.1 \cdot 10^6$	73.0%
BN-x30	$2.7 \cdot 10^6$	74.8%
BN-x5-Sigmoid		69.8%

Inception-v3: Rethinking the Inception

Idea: Scale up the network by factorzing the convolutions

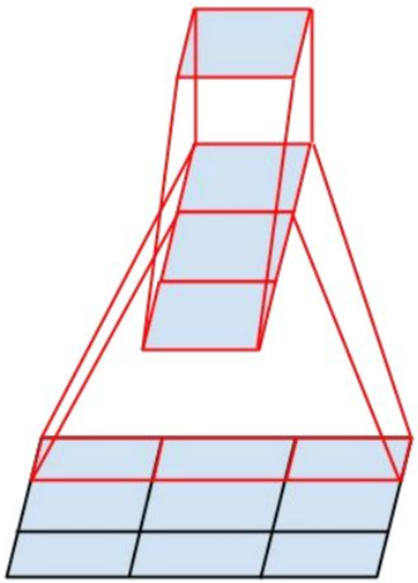


Replacing 5*5 Convolution
by two 3*3 convolutions

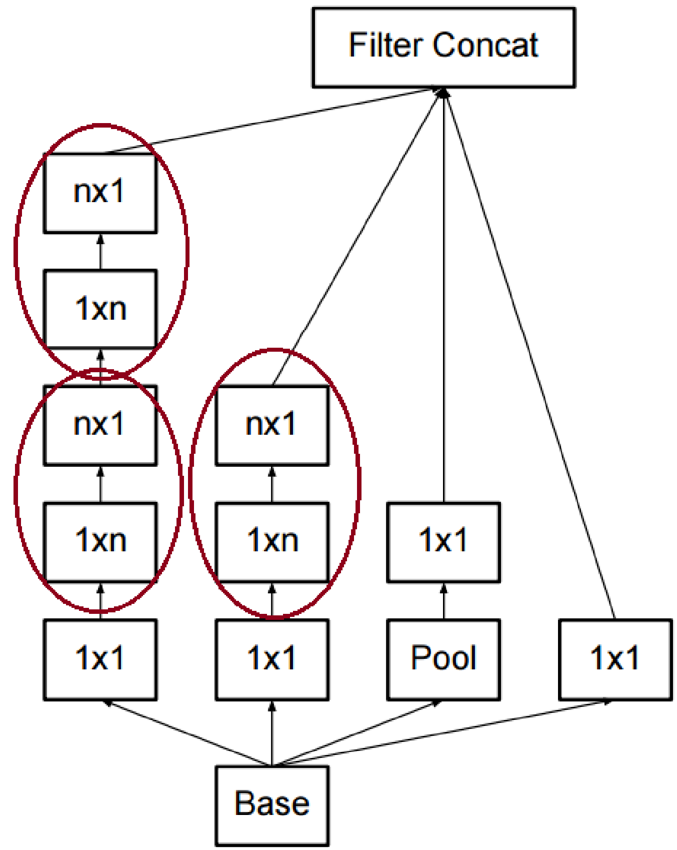


Inception-v3: Rethinking the Inception

Idea: Scale up the network by factorzing the convolutions



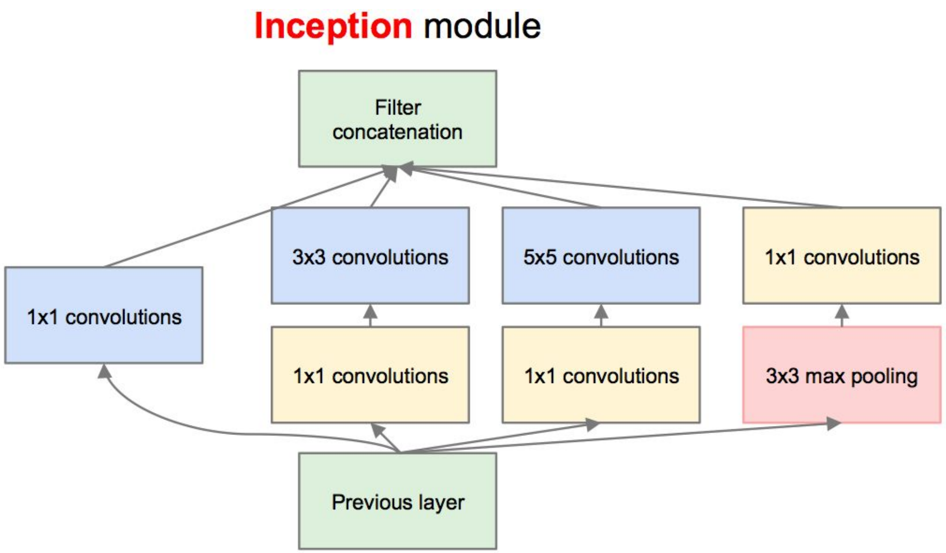
Replacing the 3 × 3 convolutions. The lower layer of this network consists of a 3 × 1 convolution with 3 output units.



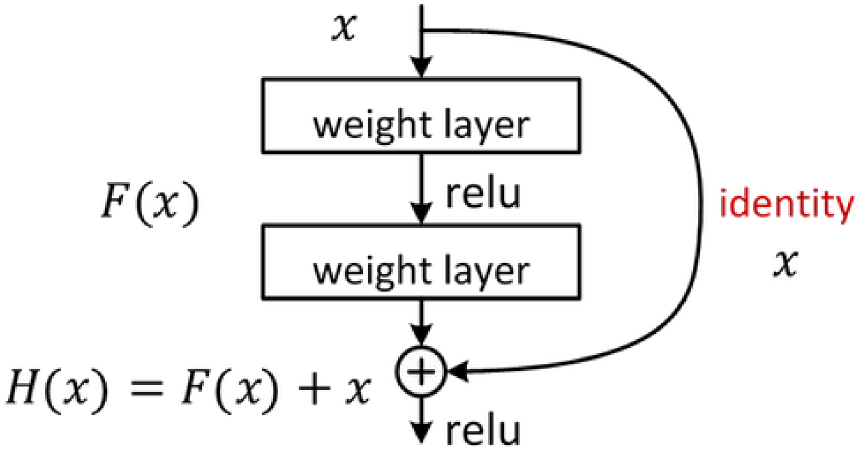
Inception modules after the factorization of the n × n convolutions. In Inception-v3: n = 7

Two Powerful Networks

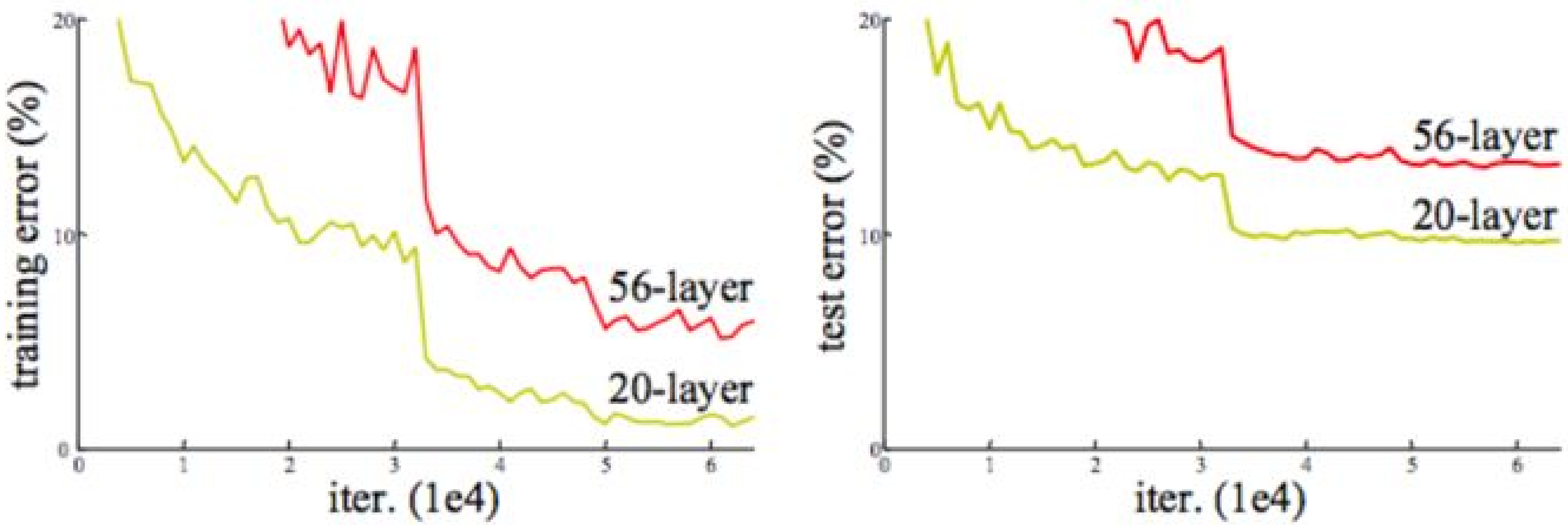
- Inception Network



- Deep Residual Network



Degradation Problem



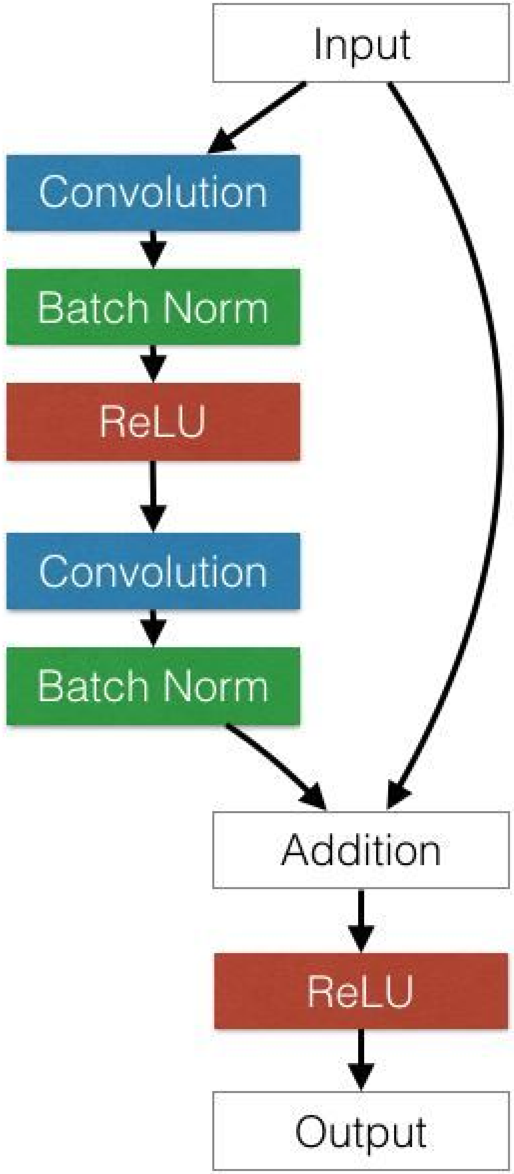
Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error.

Deep Residual Learning for Image

Extremely Deep Network:

152 layer

- **Easier to optimize**
- **More accurate**



New architectures

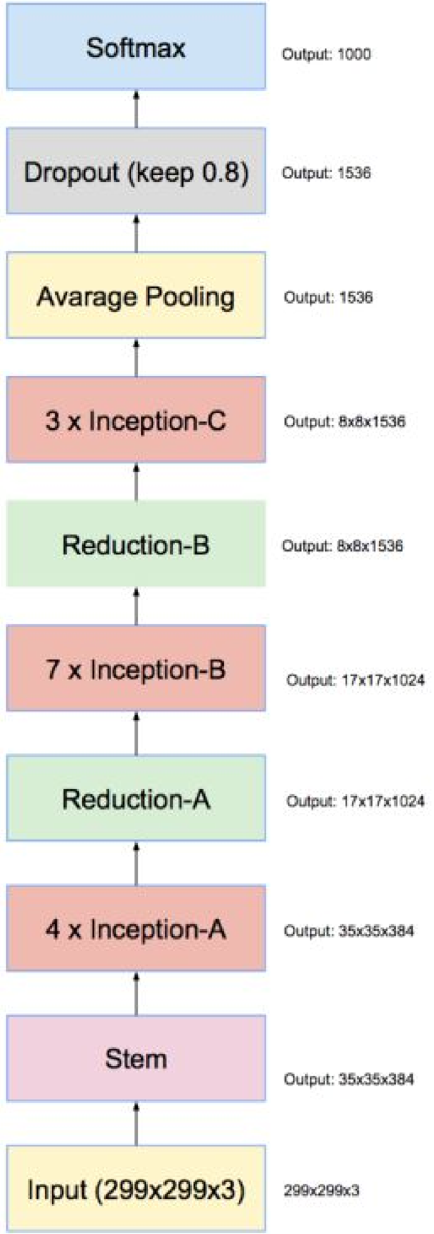
- Investigating an updated verion of Inception network with and without residual connections:
  - Inception-v4
  - Inception-ResNet-v1
  - Inception-ResNet-v2

Results in:

- Accerelation of training speed
- Improvement in accuracy

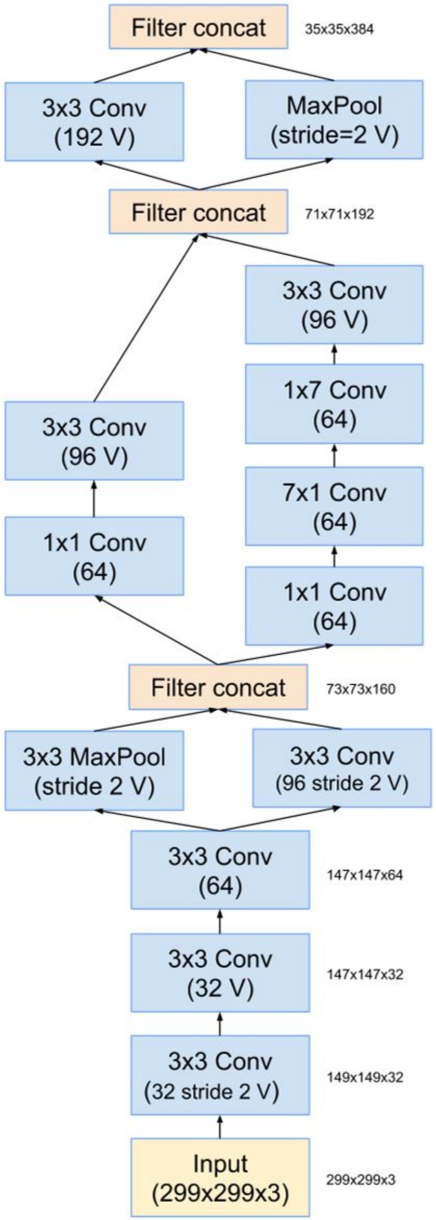
Inception-v4

- Uniform simplified architecture
- More Inception modules
- DistBelief replaced by TensorFlow



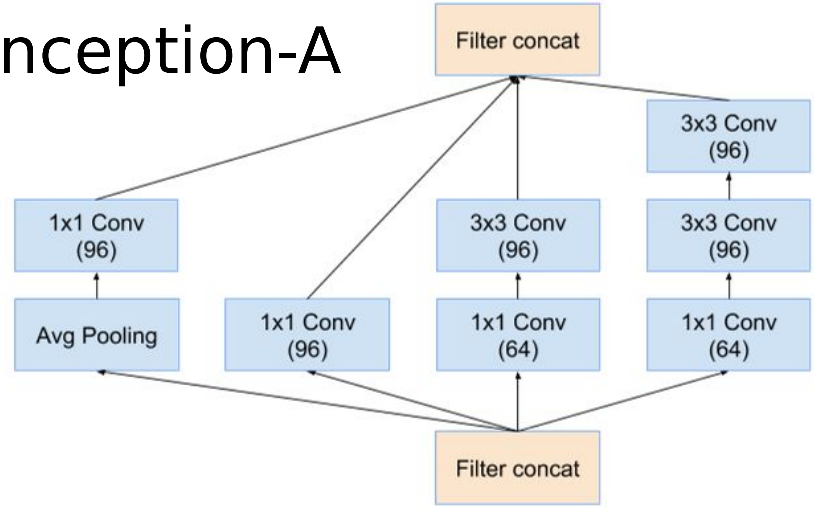
Inception-v4

Stem of Inception-v4

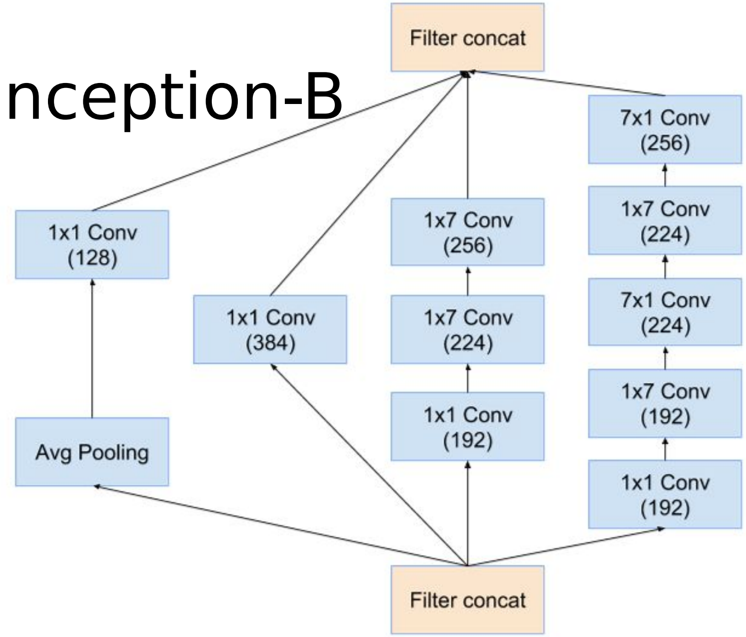


Inception-v4

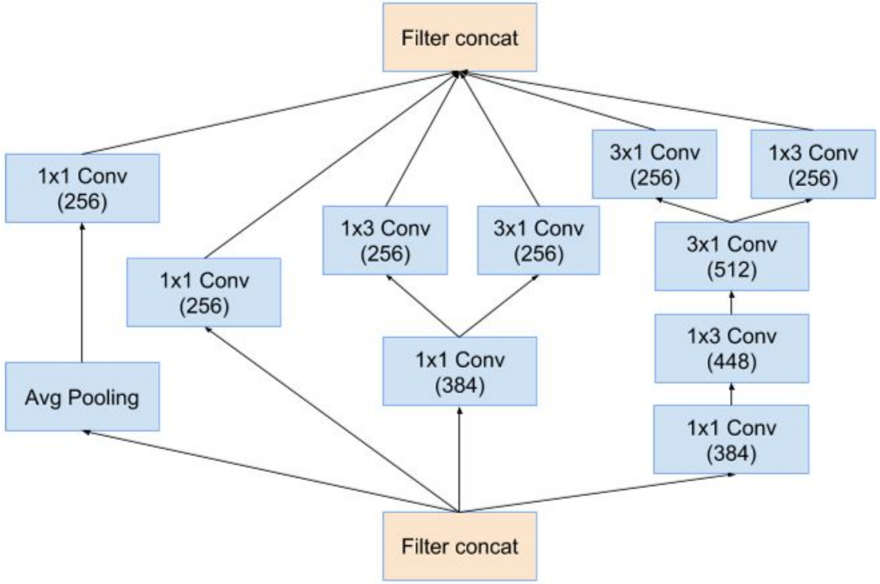
Inception-A



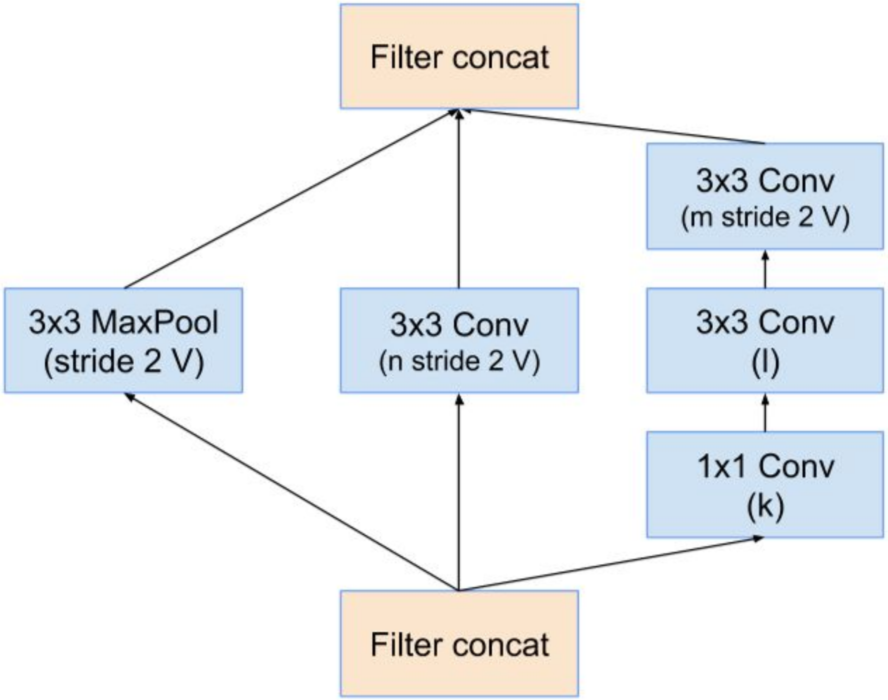
Inception-B

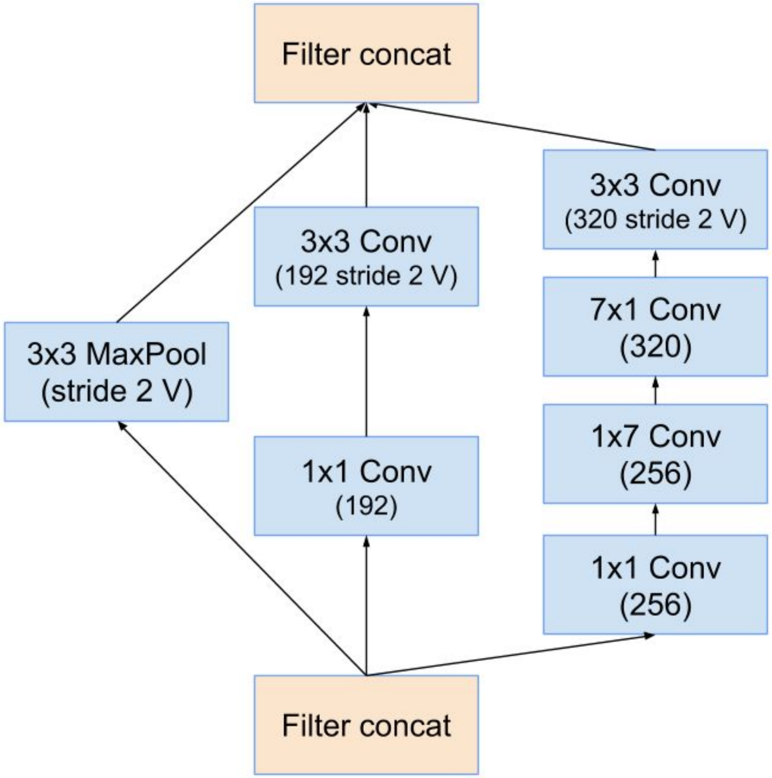


Inception -C



Inception-v4





Reduction-A

K=192, l=224, m=256, n=384

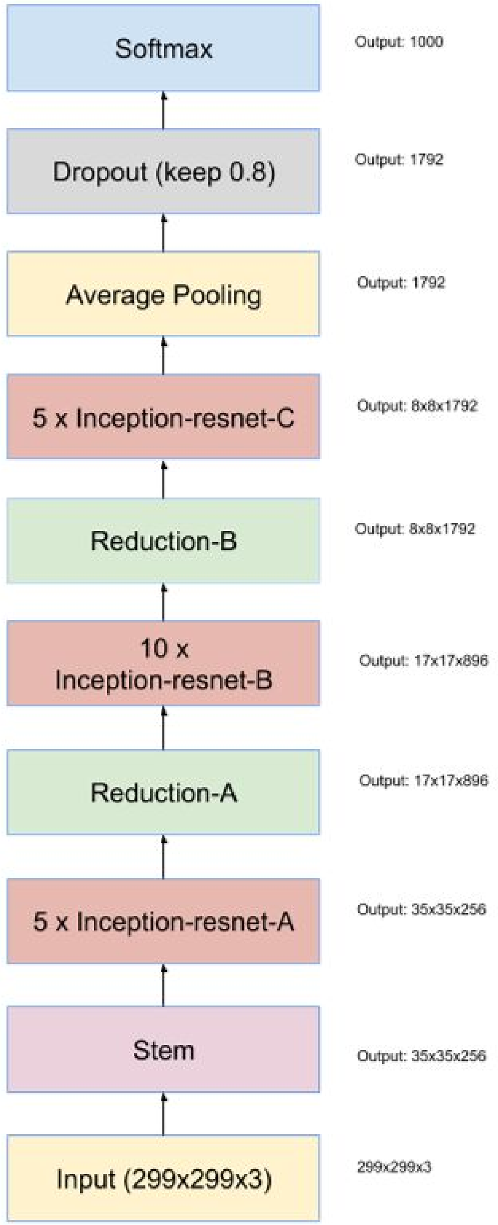
Reduction-B

Inception-ResNet-v1 and v2

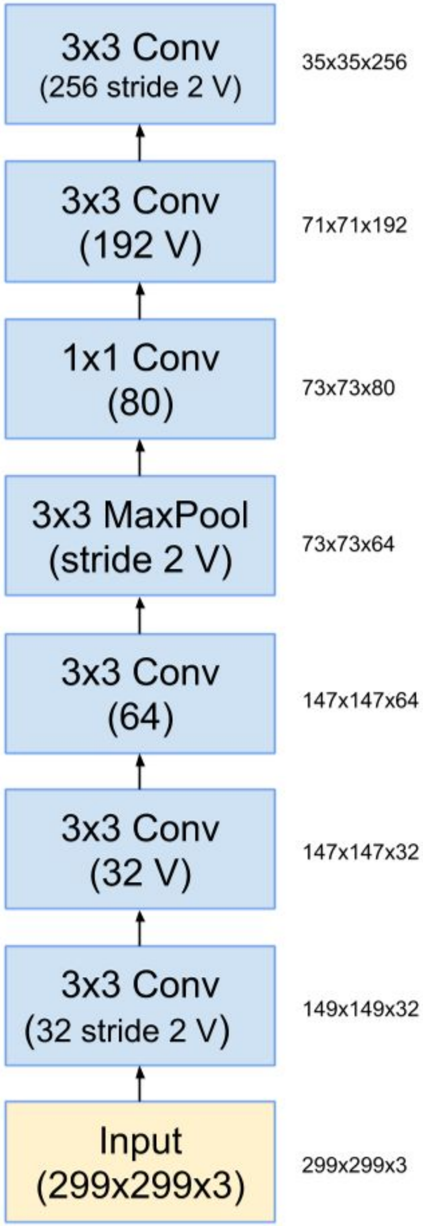
Computational cost:

Inception-ResNet-v1 ≈ Inception-v3

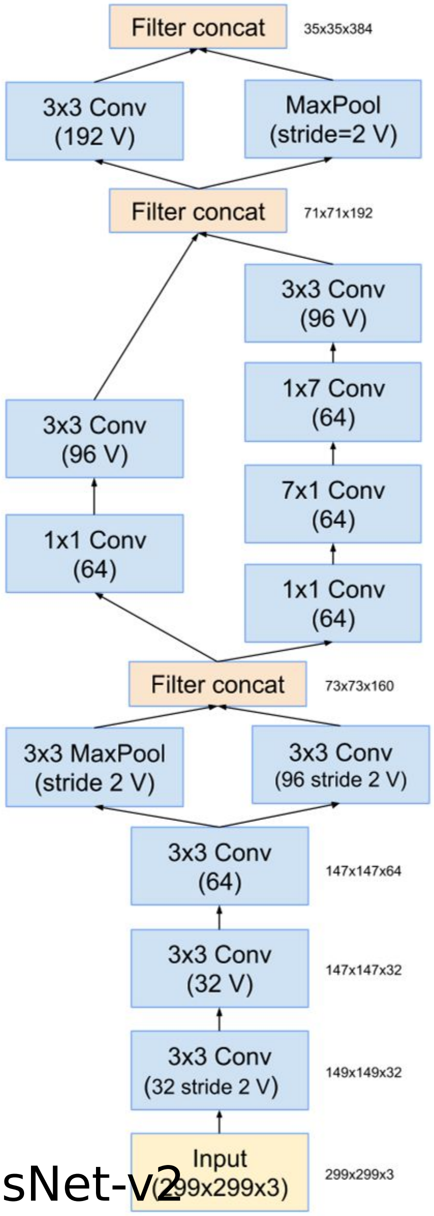
Inception-ResNet-v2 ≈ Inception-v4



Inception-ResNet-v1 and v2

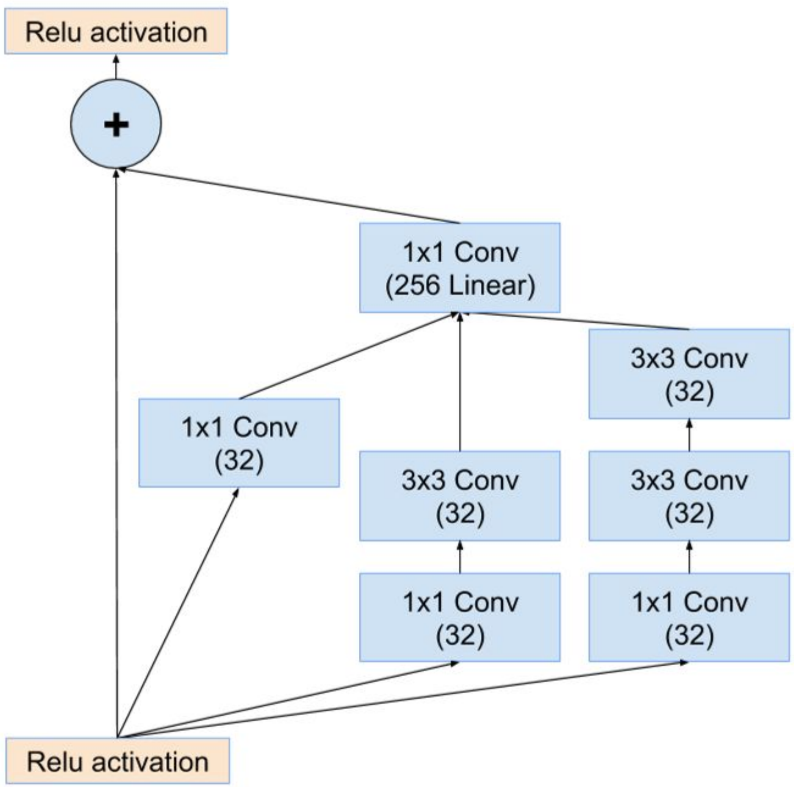


Stem of Inception-ResNet-v1

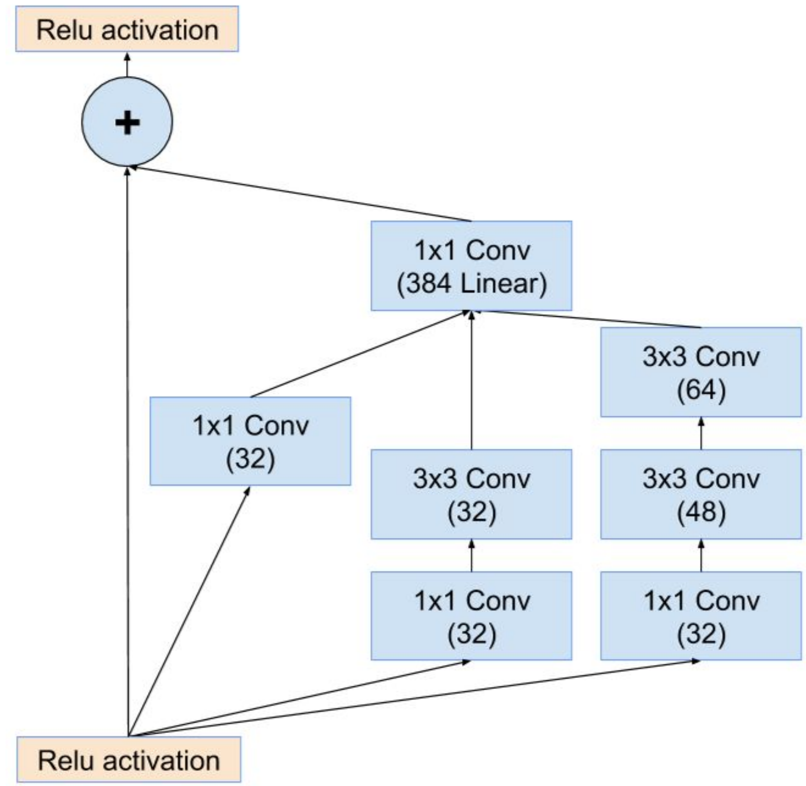


Stem of Inception-ResNet-v2

Inception-ResNet-v1 and v2

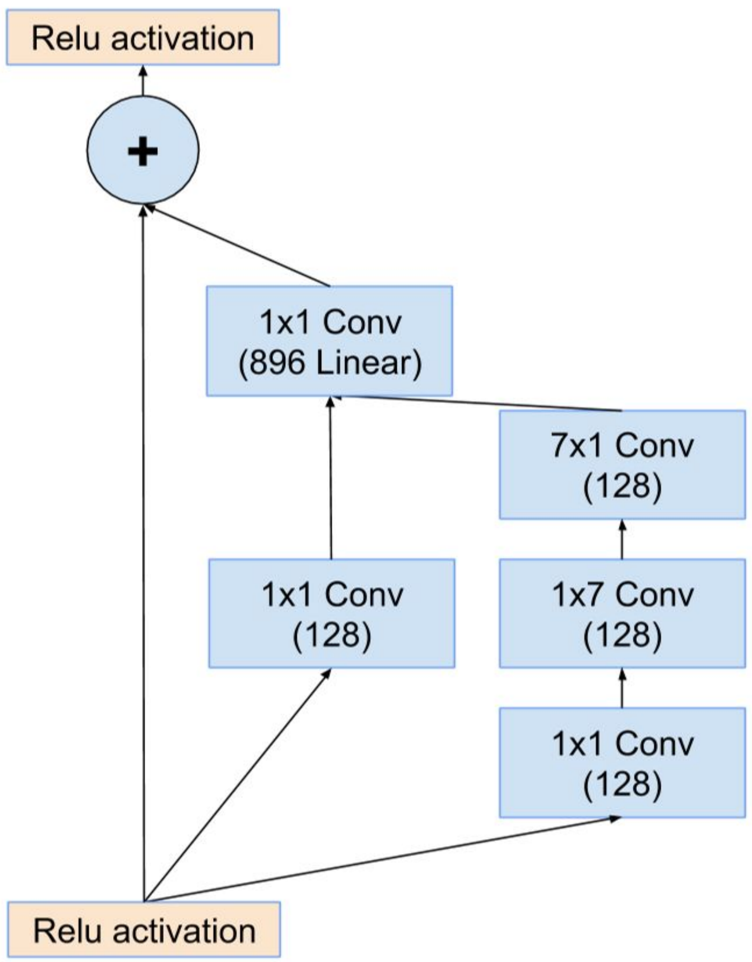


Inception-ResNet-A in v1

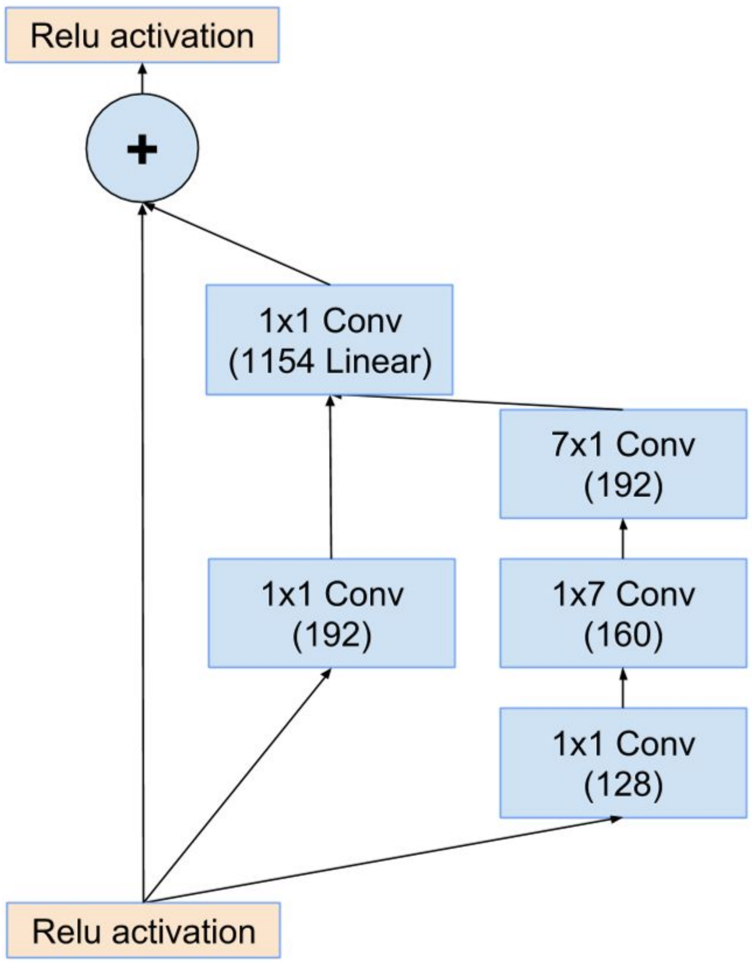


Inception-ResNet-A in v2

Inception-ResNet-v1 and v2

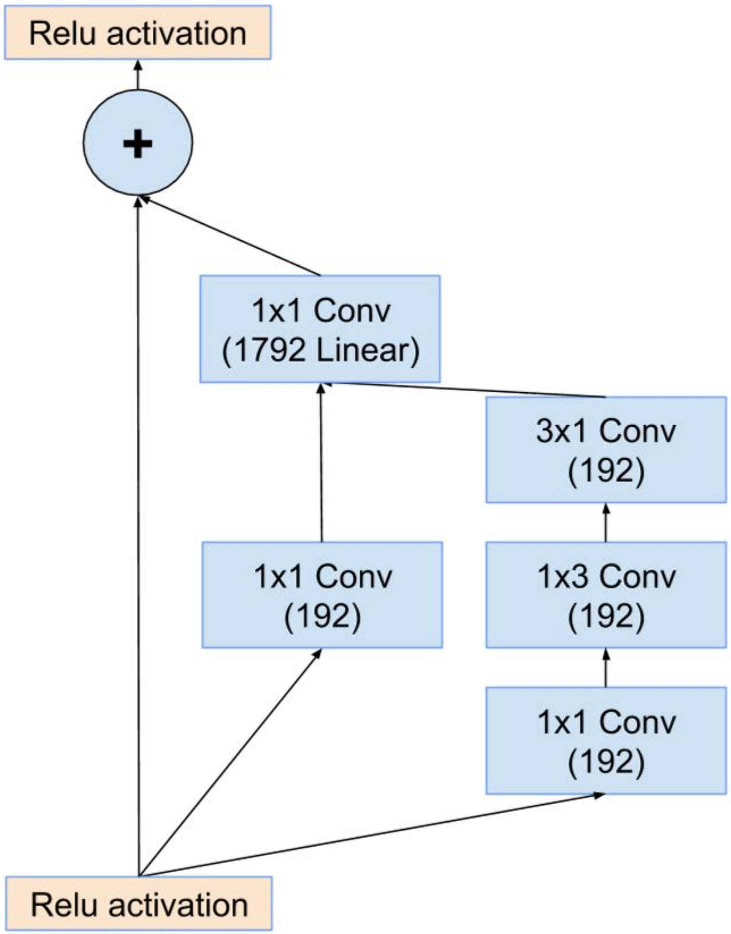


Inception-ResNet-B in v1

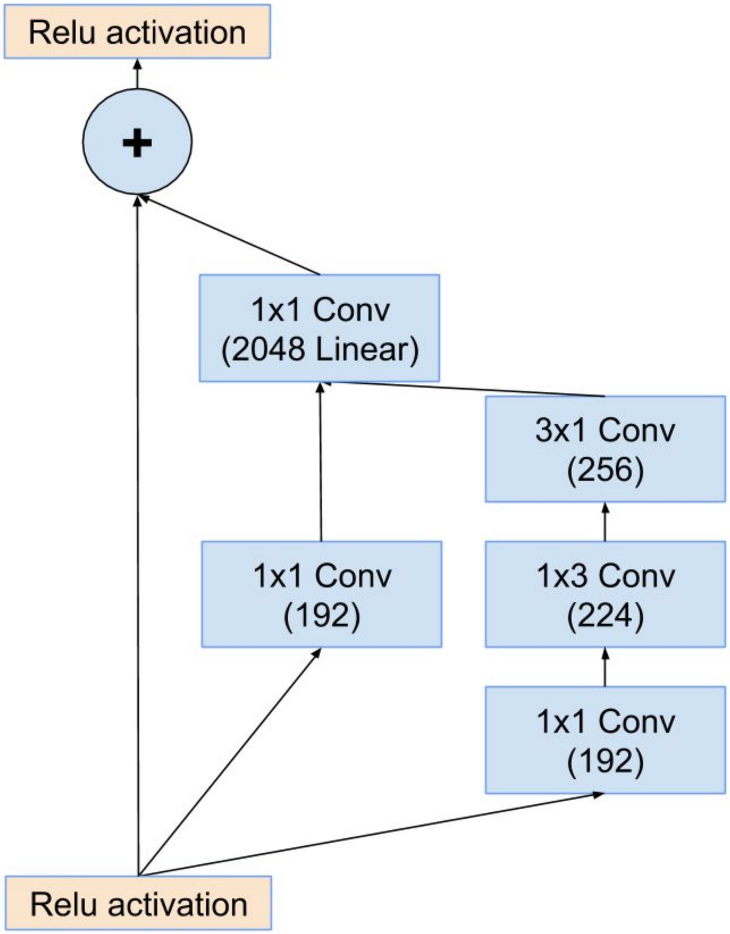


Inception-ResNet-B in v2

Inception-ResNet-v1 and v2

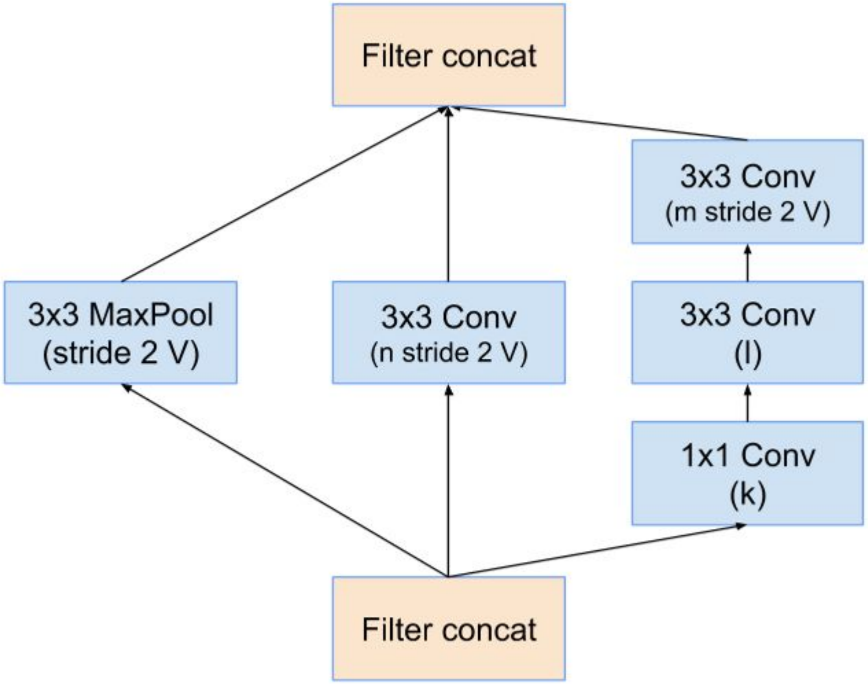


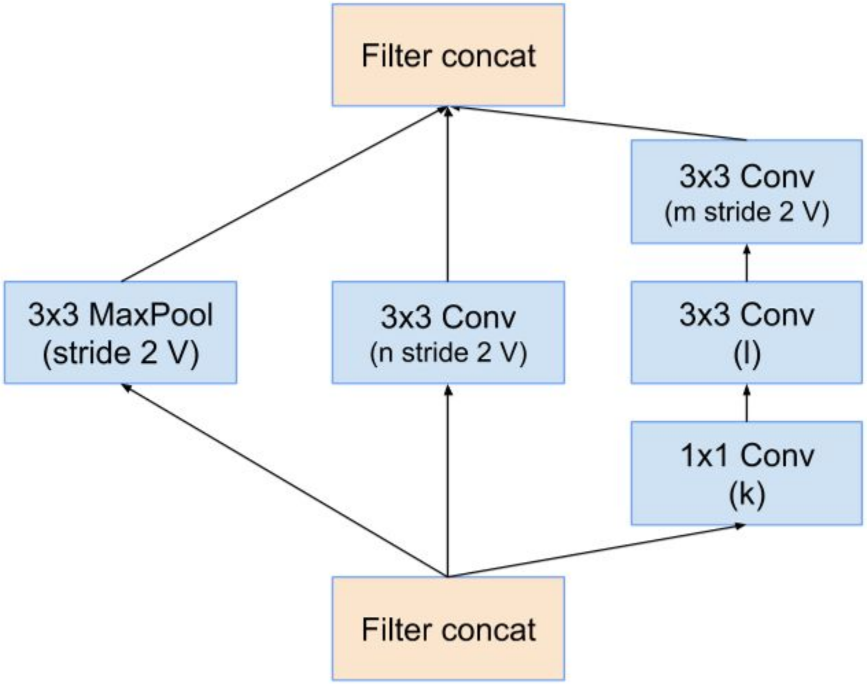
Inception-ResNet-C in v1



Inception-ResNet-C in v2

Inception-ResNet-v1 and v2

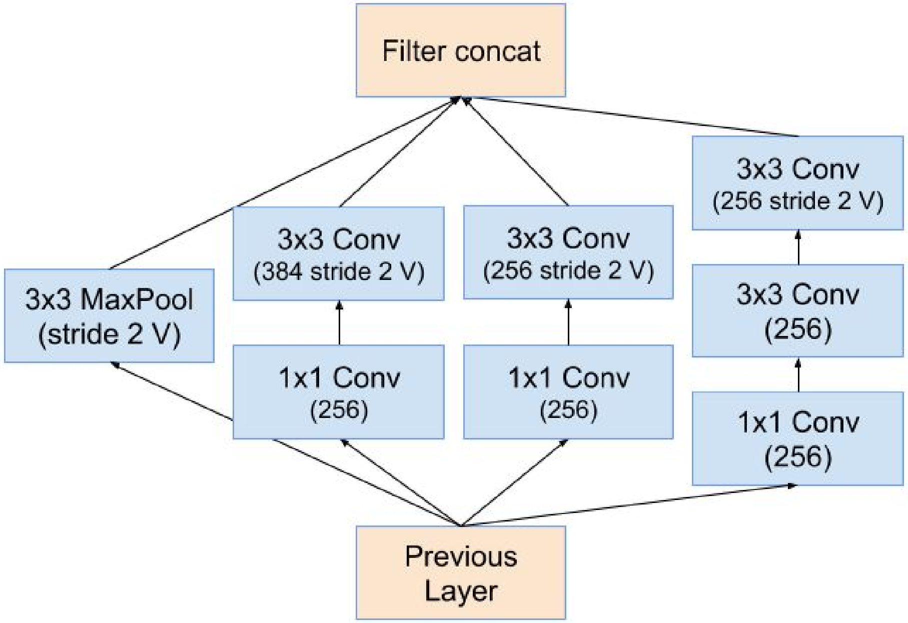




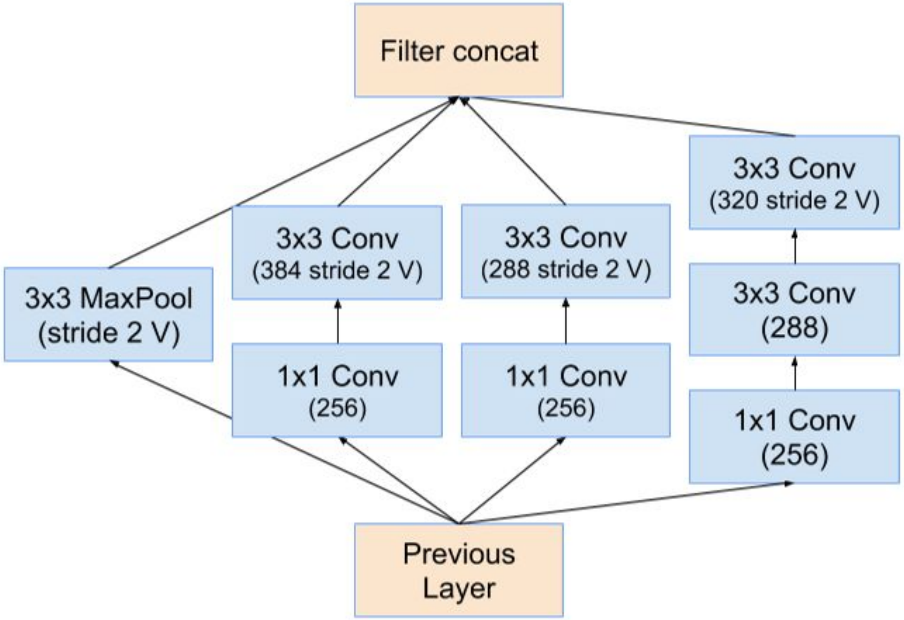
Reduction-A v1
K=192, l=192, m=256, n=384

Reduction-A v2
K=256, l=256, m=384, n=384

Inception-ResNet-v1 and v2



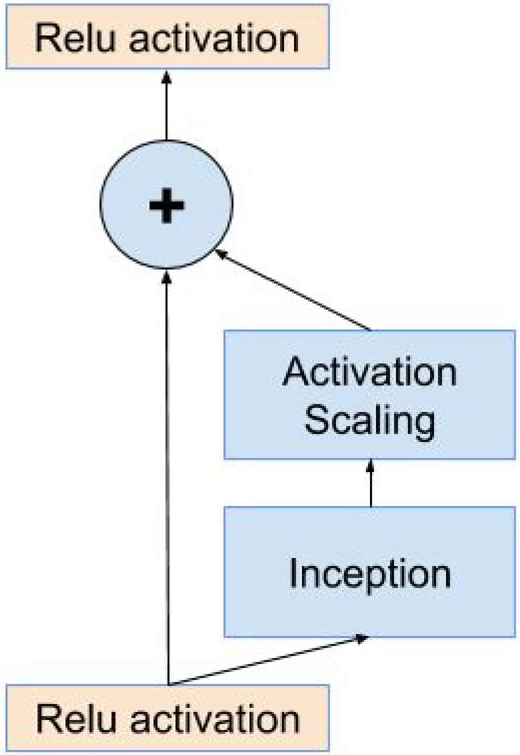
Reduction-B v1



Reduction-B v2

Inception-ResNet-v1 and v2

„If the number of filters exceeded 1000, the residual variants started to exhibit instabilities“



Training Methodology

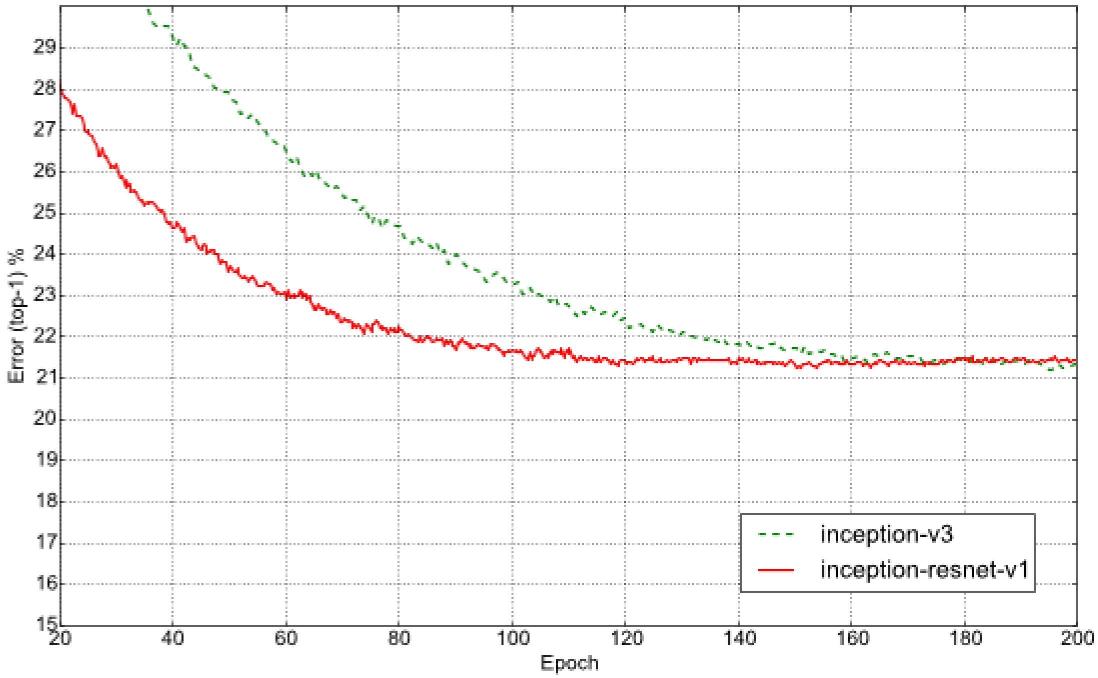
- TensorFlow
- 20 replicas running each on a NVidia Kepler GPU
- RMSProp with decay of 0.9 and $\varepsilon = 1.0$
- learning rate of 0.045, decayed every two epochs using an exponential rate of 0.94

Experimental Results

Network	Top-1 Error	Top-5 Error
BN-Inception [6]	25.2%	7.8%
Inception-v3 [15]	21.2%	5.6%
Inception-ResNet-v1	21.3%	5.5%
Inception-v4	20.0%	5.0%
Inception-ResNet-v2	19.9%	4.9%

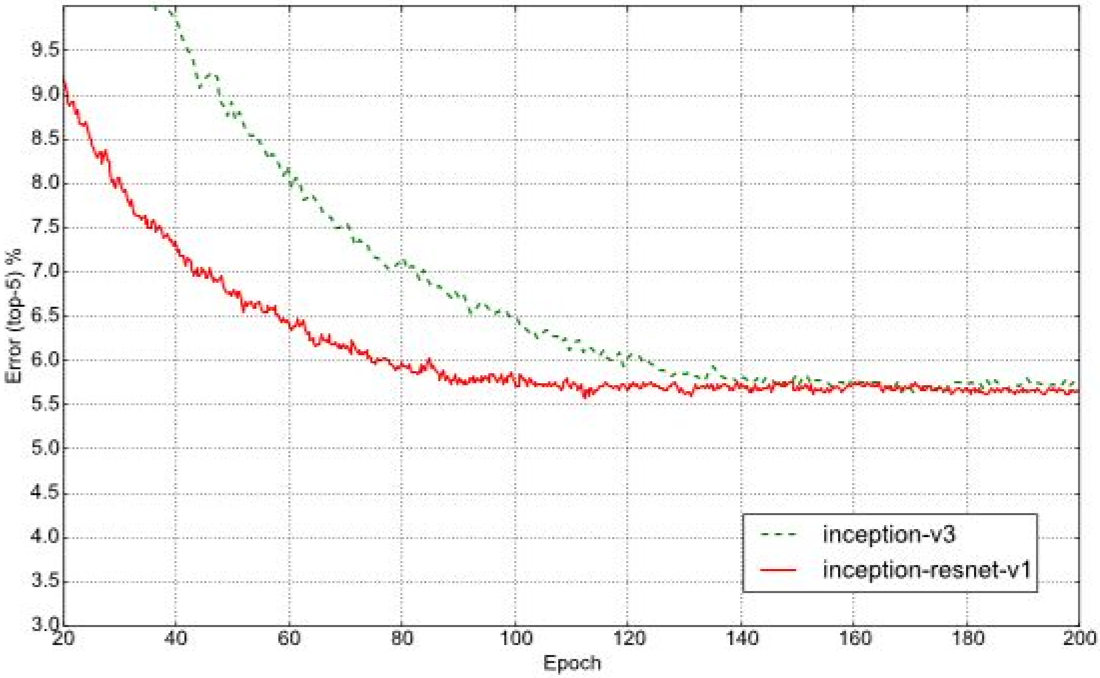
Single crop - single model experimental results. Reported on the non-blacklisted subset of the validation set of ILSVRC 2012.

Experimental Results



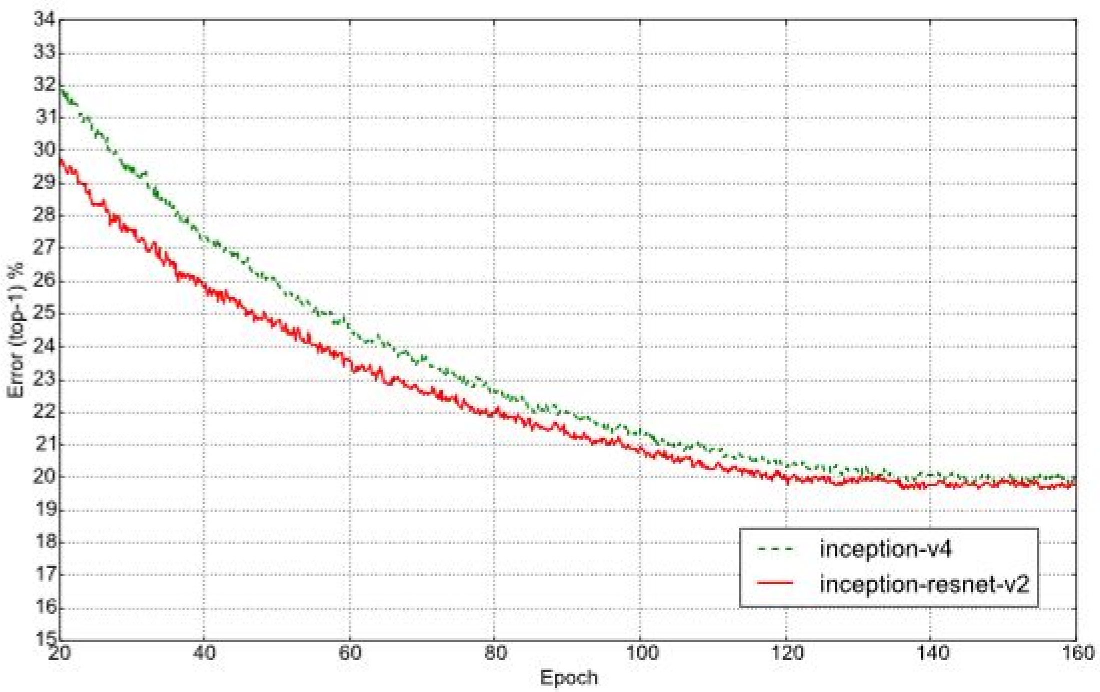
Top-1 error evolution during training of pure Inception-v3 Vs Inception-resnet-v1. The evaluation is measured on a single crop on the non-blacklist images of the ILSVRC-2012 validation set.

Experimental Results



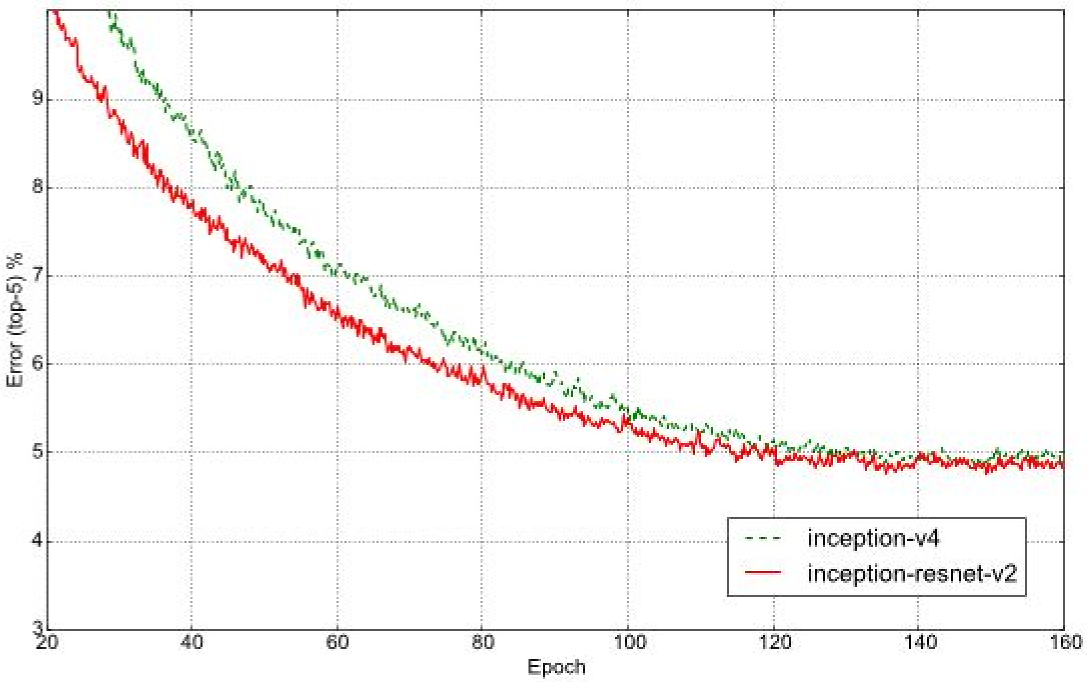
Top-5 error evolution during training of pure Inception-v3 Vs Inception-resnet-v1. The evaluation is measured on a single crop on the non-blacklist images of the ILSVRC-2012 validation set.

Experimental Results



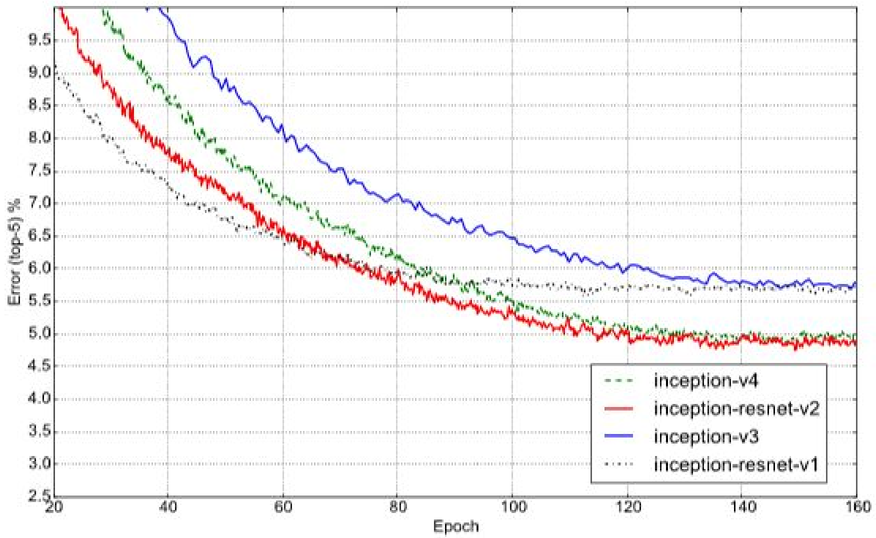
Top-1 error evolution during training of pure Inception-v4 Vs Inception-resnet-v2. The evaluation is measured on a single crop on the non-blacklist images of the ILSVRC-2012 validation set.

Experimental Results

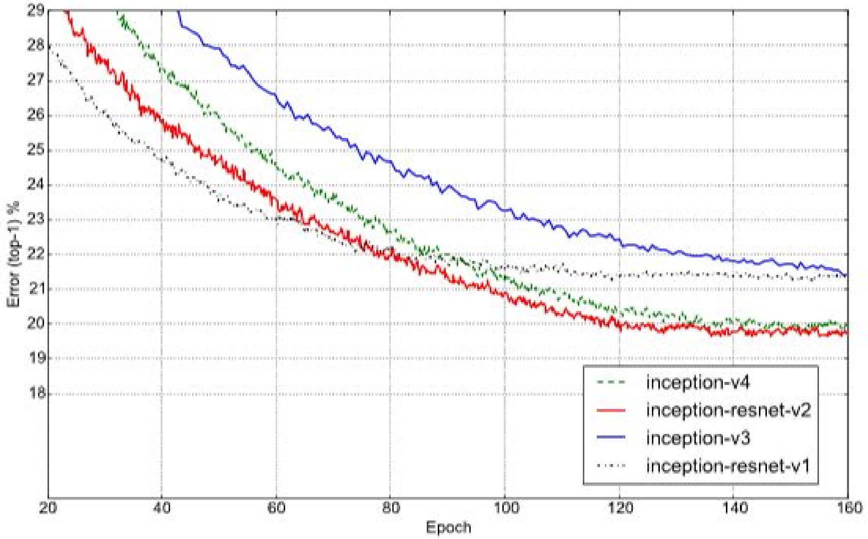


Top-5 error evolution during training of pure Inception-v4 Vs Inception-resnet-v2. The evaluation is measured on a single crop on the non-blacklist images of the ILSVRC-2012 validation set.

Experimental Results



Top-5 error evolution of all four models (single model, single crop)



Top-1 error evolution of all four models (single model, single crop)

Experimental Results

Multi crops evaluations - single model experimental results

Network	Crops	Top-1 Error	Top-5 Error
ResNet-151 [5]	10	21.4%	5.7%
Inception-v3 [15]	12	19.8%	4.6%
Inception-ResNet-v1	12	19.8%	4.6%
Inception-v4	12	18.7%	4.2%
Inception-ResNet-v2	12	18.7%	4.1%

Network	Crops	Top-1 Error	Top-5 Error
ResNet-151 [5]	dense	19.4%	4.5%
Inception-v3 [15]	144	18.9%	4.3%
Inception-ResNet-v1	144	18.8%	4.3%
Inception-v4	144	17.7%	3.8%
Inception-ResNet-v2	144	17.8%	3.7%

Experimental Results

Exceeds state-of-the-art single frame performance on the ImageNet validation dataset

Network	Models	Top-1 Error	Top-5 Error
ResNet-151 [5]	6	–	3.6%
Inception-v3 [15]	4	17.3%	3.6%
Inception-v4 + $3\times$ Inception-ResNet-v2	4	16.5%	3.1%

Ensemble results with 144 crops/dense evaluation. Reported on the all 50000 images of the validation set of ILSVRC 2012.

Concolution

- Three new architectures:
- Inception-resnet-v1
- Inception-resnet-v2
- Inception-v4

- Introduction of residual connections leads to dramatically improved training speed for the Inception architecture.

References

- A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
- C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015.
- S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of The 32nd International Conference on Machine Learning, pages 448–456, 2015.
- C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. arXiv preprint arXiv:1512.00567, 2015.
- K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385

The End

Thank you