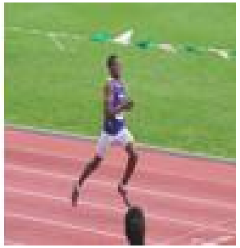
Image Classification

Datasets

- CIFAR-10: 60k images, each 32x32, in 10 classes
  - Good sanity check – models can be trained and tested quickly
- ImageNet: 1.2M images in 1000 classes
  - Well-balanced, well-curated

- Yahoo Flickr 100M: 100M images in 10k classes/tags
  - Images are noisy, unbalanced – more like the real world
- Open Images: 9M classes in 6k classes
  - Creative commons
  - Includes Image masks and bounding boxes
- MIT Places: 10M images in 400+ classes
  - Focuses on scenes rather than objects

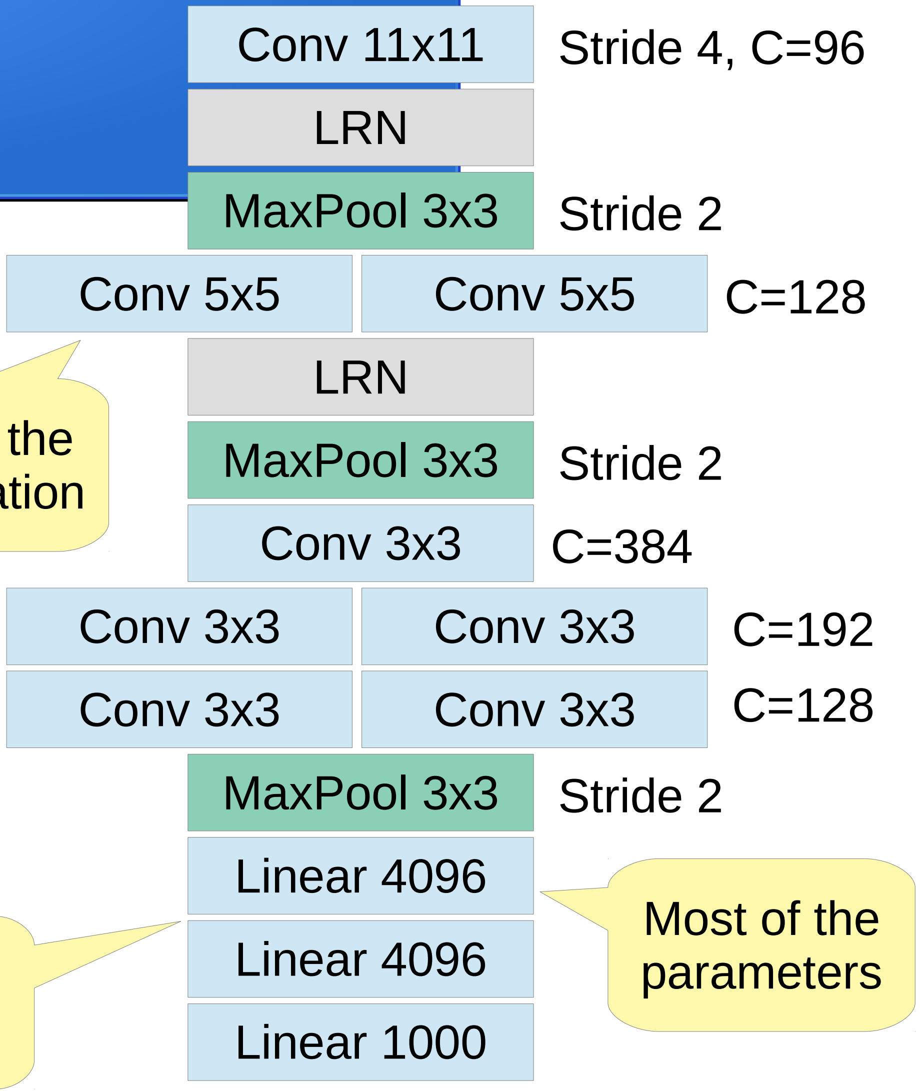


Applications

- Visual search / Reverse image search
- Testbed for new architectures
- Important first step for many other tasks
  - Until very recently, other tasks almost always started from pretrained classification networks

AlexNet

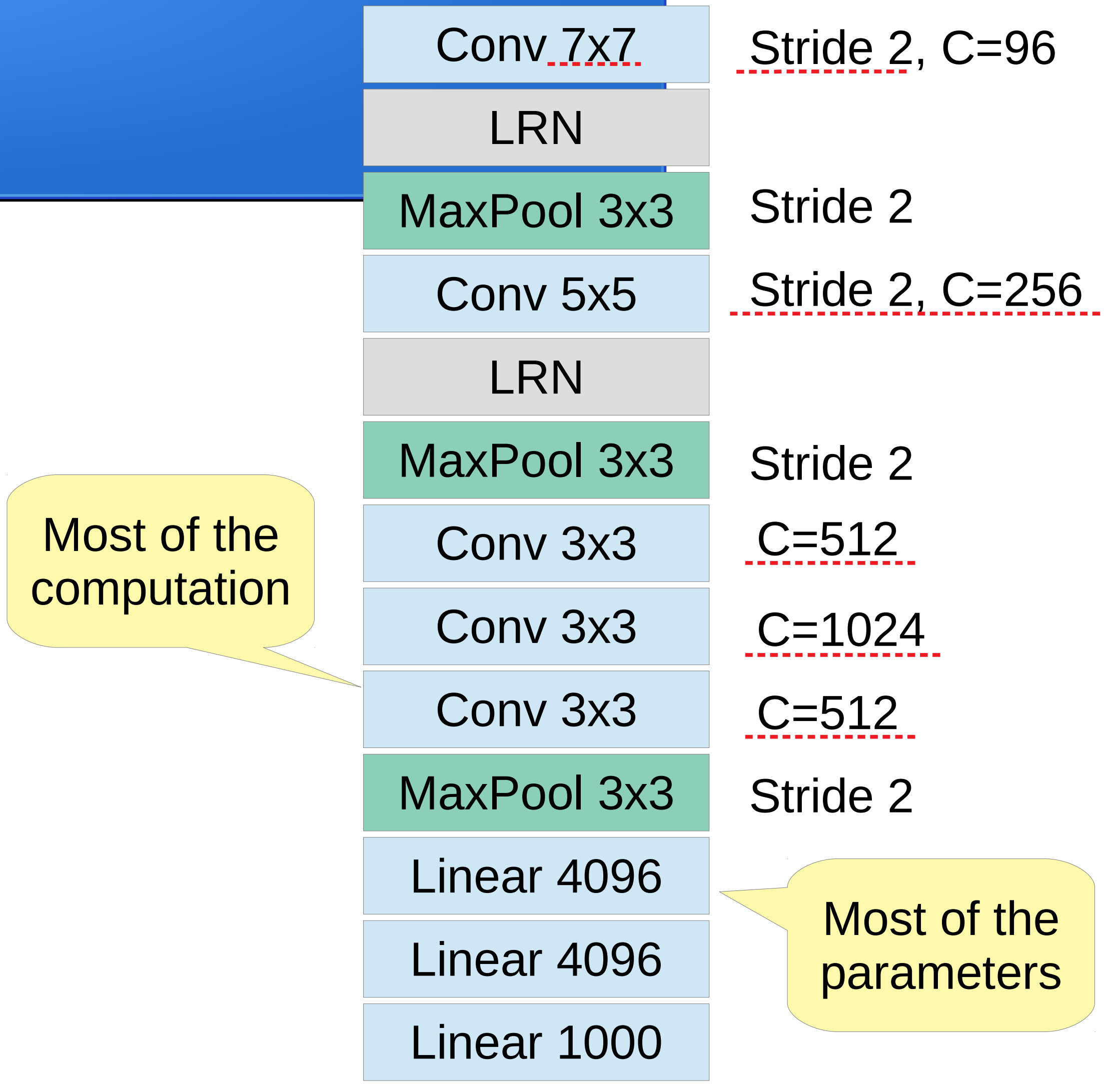
- The beginning of the deep learning revolution
- Won the ImageNet competition in 2012
  - Was the first deep network to win this competition
  - Reduced top-5 error by about 30%



Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. NeurIPS 2012

ZFNet

- Analysis of AlexNet
- Used upconvolution as an analytical tool
- Improved on AlexNet for the 2013 ImageNet competition



Matthew D. Zeiler and Rob Fergus. 2014. Visualizing and Understanding Convolutional Networks. ECCV 2014.

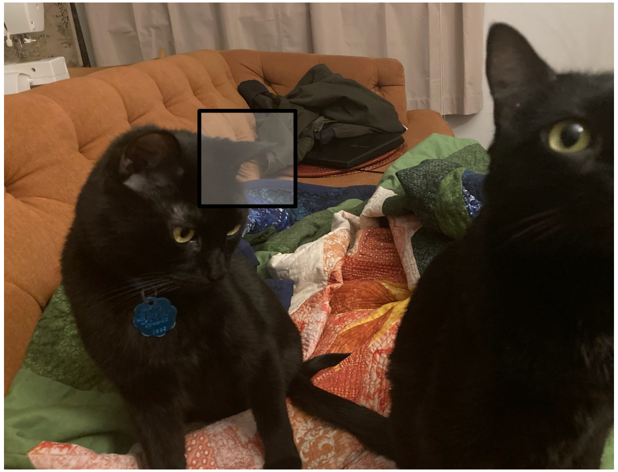
VGG

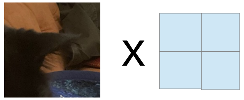
- Deeper than AlexNet/ZFNet
- 8 Layers → 16 Layers
- All 3x3 kernels
- Vanishing gradients
  - Train increasingly large prefixes of the network.
- Generalizes to new tasks

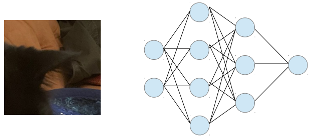
2 X	Conv 3x3	C=64
	MaxPool 2x2	
2 X	Conv 3x3	C=128
	MaxPool 2x2	
3 X	Conv 3x3	C=256
	MaxPool 2x2	
3 X	Conv 3x3	C=512
	MaxPool 2x2	
3 X	Conv 3x3	C=512
	MaxPool 2x2	
	Linear 4096	
	Linear 4096	
	Linear 1000	

Karen Simonyan and Andrew Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. ICLR 2015

Interlude: Factorization





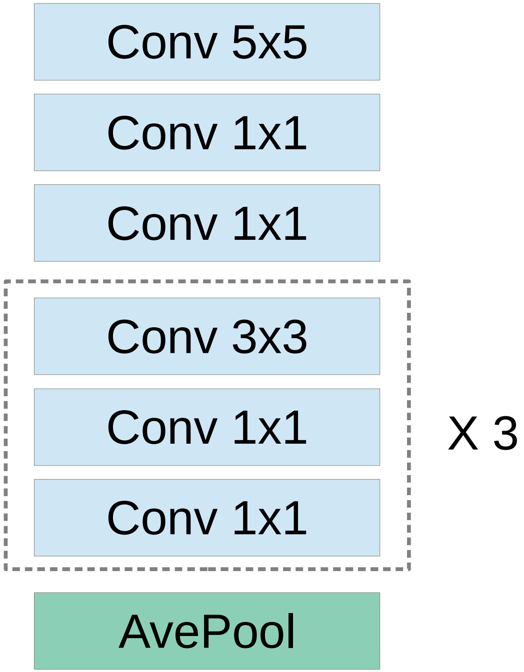


Conv 3x3

Conv 1x1

Network in Network

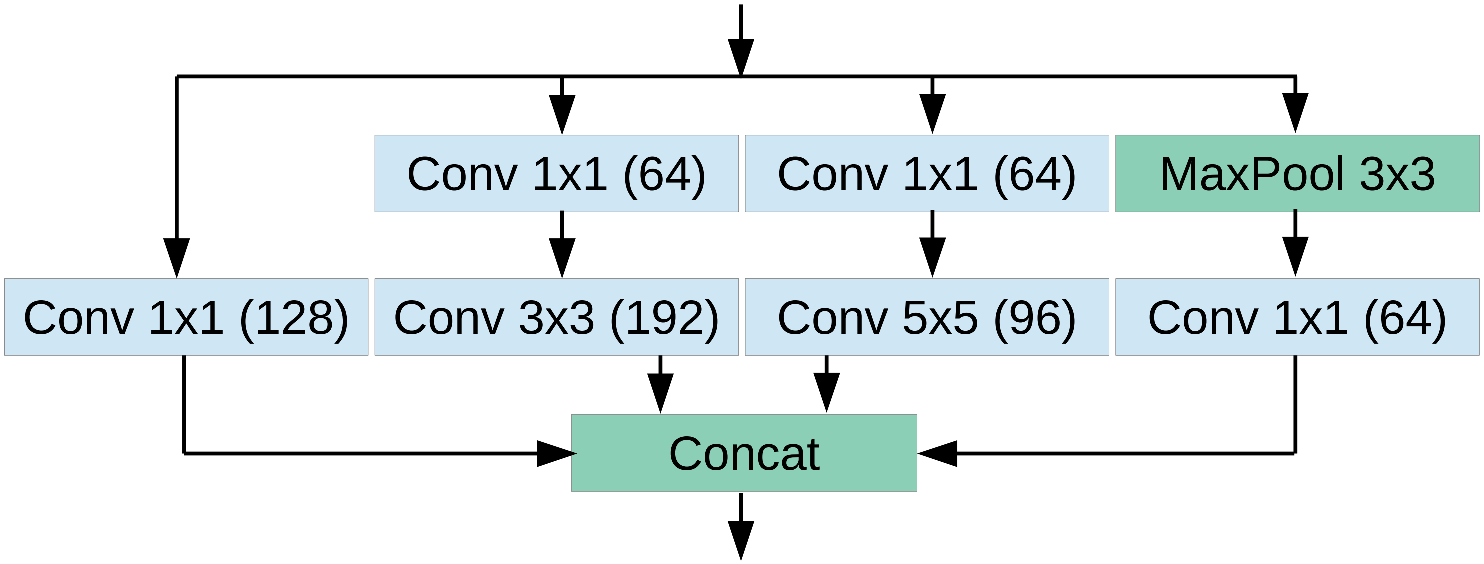
- Small architecture
- Factorized convolutions
- Global average pooling

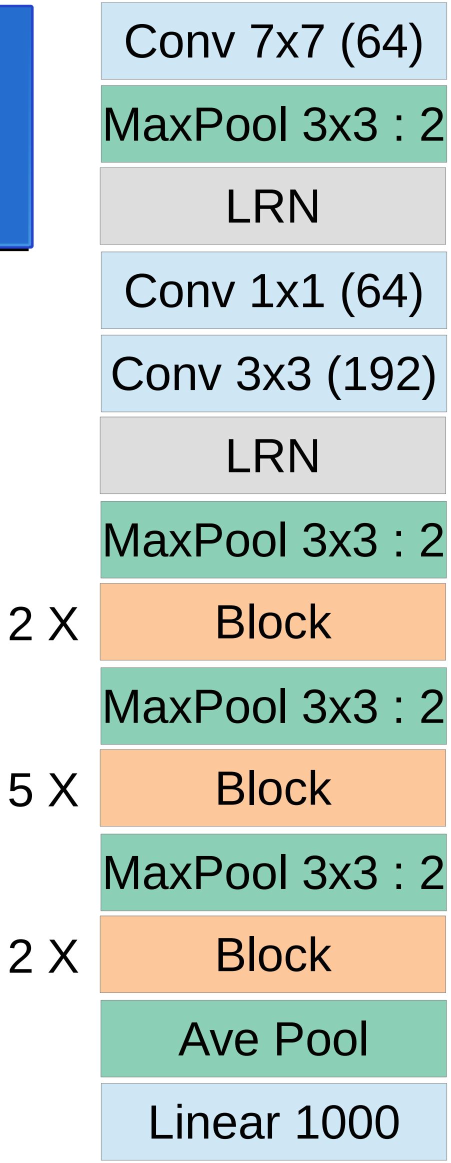


Min Lin, Qiang Chen, and Shuicheng Yan. 2014. Network in Network. ICLR 2014.

Inception / GoogLeNet

- Similar to Network-in-Network
- Multiple kernel sizes

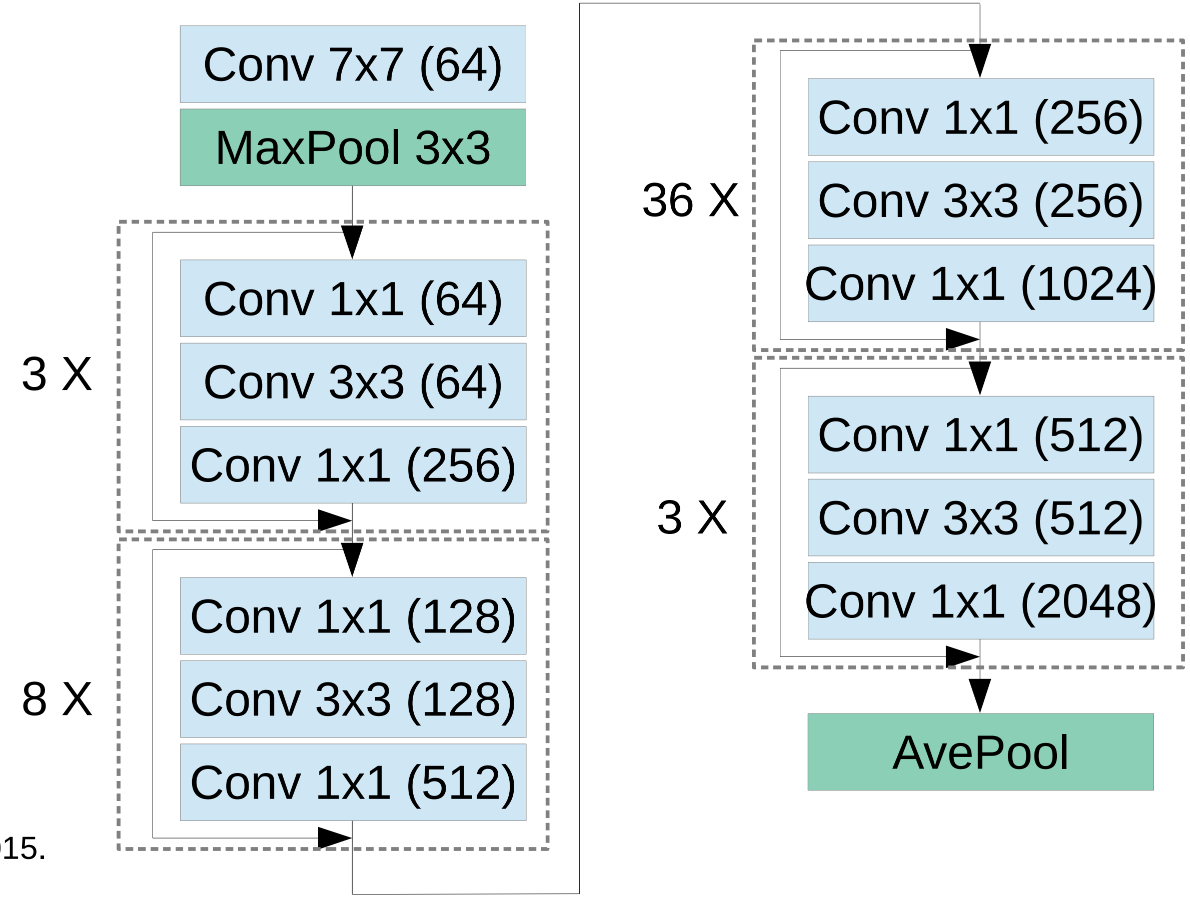




Christian Szegedy, et al. 2015. Going Deeper with Convolutions. CVPR 2014.

ResNet

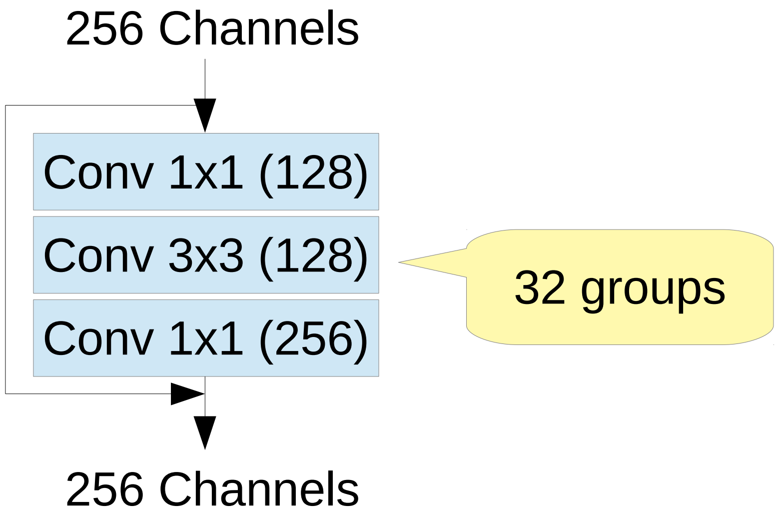
- Residual connections allow better training
- Huge jump in network depth
  - 22 layers (Inception) → 152 layers
- Variants: ResNet-18, -34, -50, -101, -152, -1001



Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015.
Deep Residual Learning for Image Recognition. CVPR 2016.

ResNeXt and Stochastic Depth

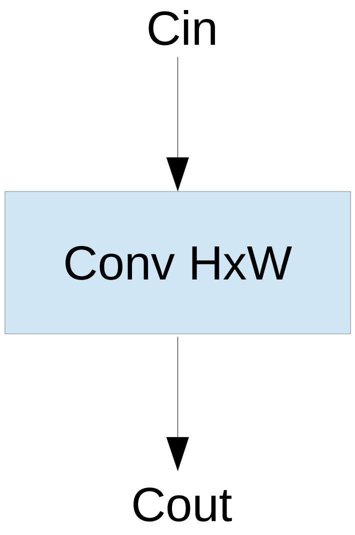
- ResNeXt
  - Use 1x1 to downsample to save parameters
  - Use grouping to save parameters
- Stochastic Depth Networks
  - Drop blocks at random



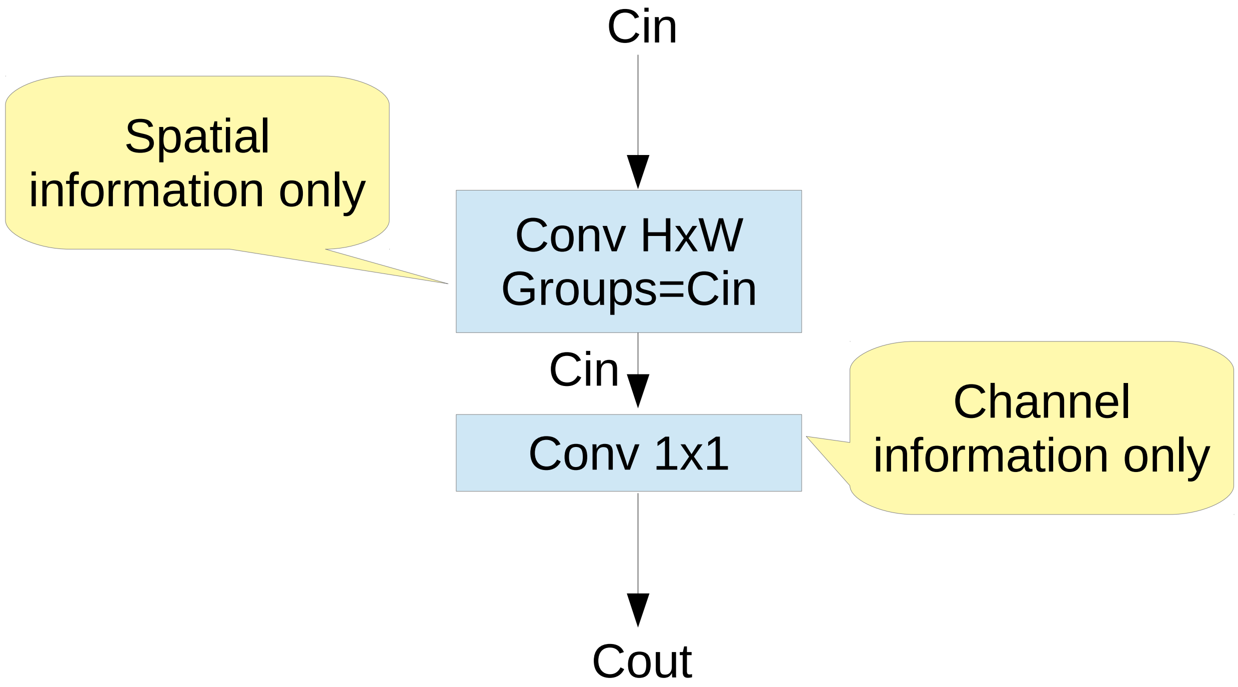
Saining Xie, Ross Girschick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2017. Aggregated Residual Transformations for Deep Neural Networks. CVPR 2017.

Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Weinberger. 2016. Deep Networks with Stochastic Depth. ECCV 2016.

Interlude 2: Factorization Boogaloo



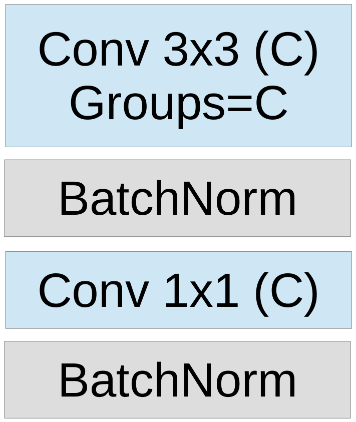
Parameters: H x W x Cin x Cout

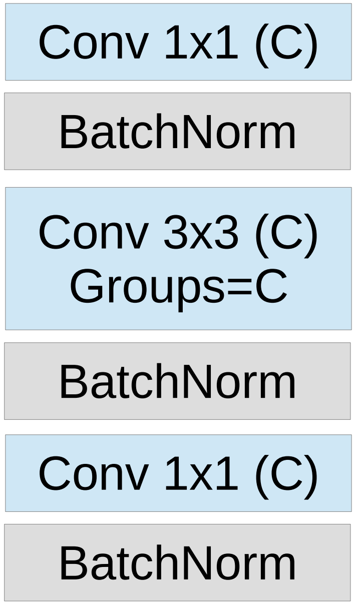


Parameters: H x W x Cin + Cin x Cout

MobileNet

- Small architecture designed to be run on phones
- Uses factorization to reduce parameters and computation





Andrew G. Howard, et al. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv.

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen. 2018. MobileNetv2: Inverted Residuals and Linear Bottlenecks. CVPR 2018.