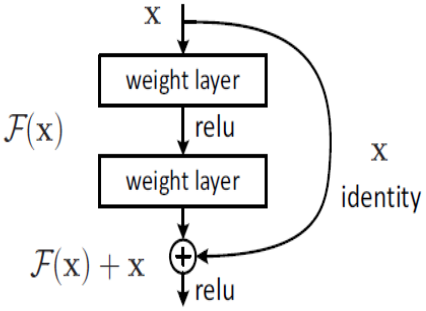
ResNet

Microsoft Research

Is learning better networks as easy as stacking more layers?

Few problems arise when plain networks get deeper and deeper:
1) Vanishing/exploding gradients.
2) More difficult to optimize in feasible time.
Residual nets can counter these problems very efficiently.

->Skip Connections-Building block of ResNets



Instead of fitting the desired mapping(H(x)) ,the network learns a residual mapping (H(x)-x) and the original input is then added as a skip connection to the output.

Intuition:
A deep network can imitate a shallow network by pushing all the weights of residual mapping to zero and hence,learn an identity mapping.

Network Architecture:

->3x3 filters were used.
For same output feature map size,number of filters were kept same.If output size was halved,then the number of filters were doubled(to ensure constant time complexity).
->Instead of FC layers,**Global average pooling** was used followed by a 1000 way softmax classifier.

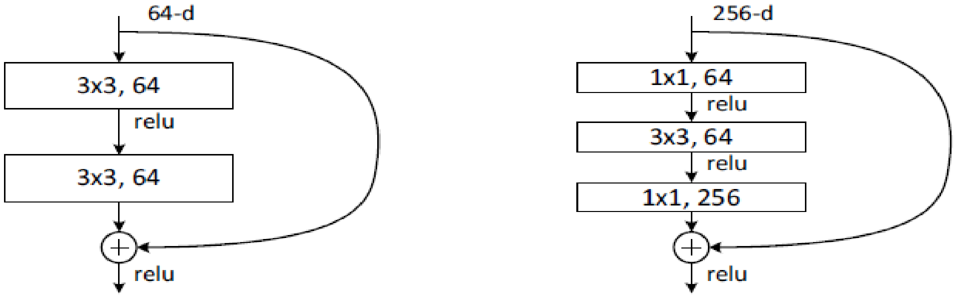
->Skip connections(Matching the dimensions):
If the output has the same size as input,identity shortcuts are directly added.
If the output size is smaller than input size,then:
1) Shortcuts are zero padded.(No extra parameters are required)
(OR)
2) Projection of shortcuts(using 1x1 conv) can be used. (Extra parameters are required)

Bottleneck Layers:

To reduce the time complexity of the model,a stack of 3 layers was used.

- 1x1 conv was added at the start to decrease the dimensions.
- 3x3 conv was added in the middle.
- 1x1 conv was added in the end to increase the dimensions.

Repeating bottleneck layers,101 and 152 layered ResNets were formed.



Training and Testing:

1)Image augmentation -similar to AlexNet.
2)RGB mean subtraction and color augmentation were used.
3)BN,Weight decay and momentum were also used.
4)10 crop testing was used.

Performance:

->ILSVRC 2015(First place):

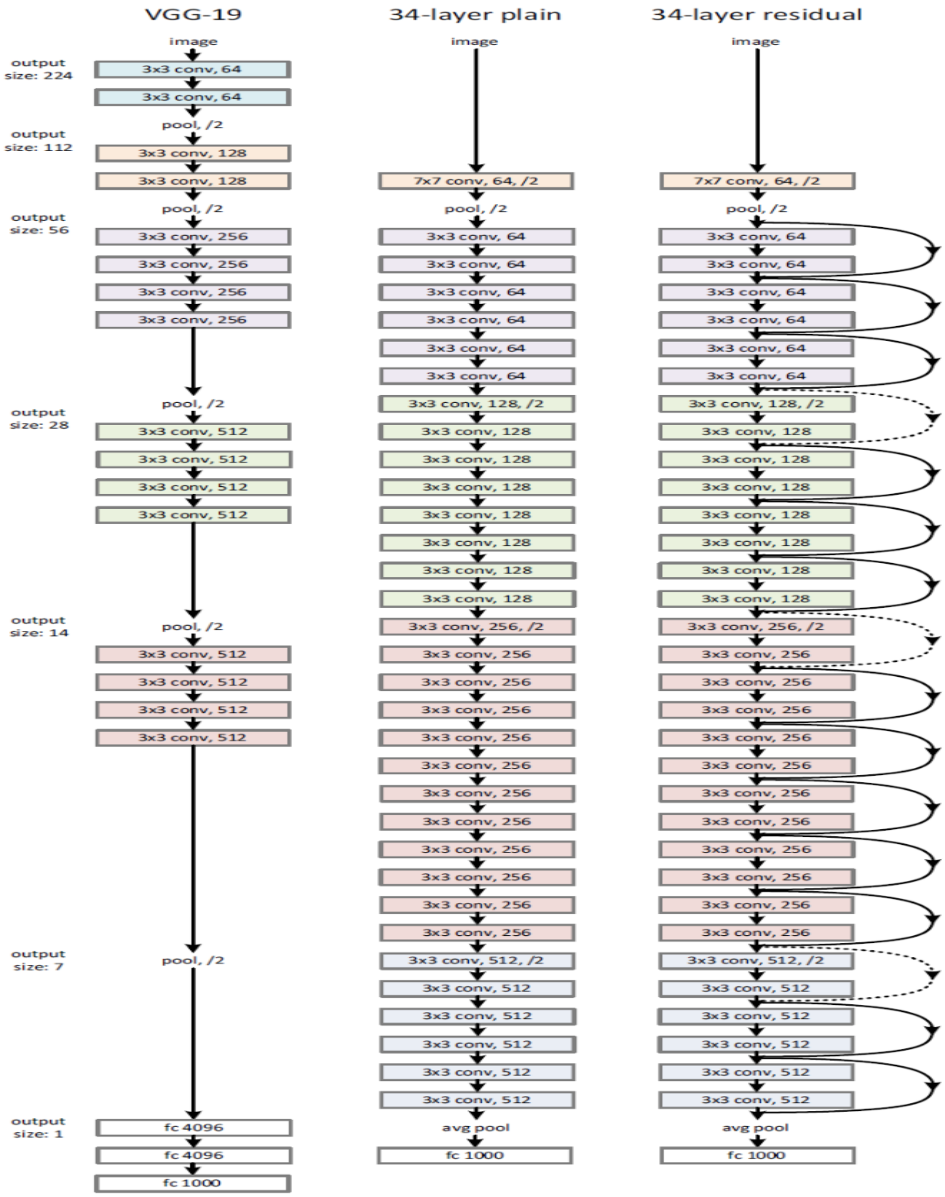
With 10-Crop Testing + Fully Conv with single model, ResNet-152 obtained **4.49%** error rate(Top 5%). With a 6-model ensemble ,the error rate was **3.57%.**

->Winner of ImageNet Detection, Localization, COCO Detection and COCO Segmentation.

Other details:

Analysis of layer responses:The residual functions were found to be closer to zero,learning mapping closer to identity.

Exploring 1000 layers: Despite using 1202 layers,the model achieved <0.1% training error(no optimization difficulty).But the model became prone to overfitting.



ayer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\left[\begin{array}{c}3\times3, 64\\3\times3, 64\end{array}\right]\times2$	$\left[\begin{array}{c}3\times3, 64\\3\times3, 64\end{array}\right]\times3$	$\left[\begin{array}{c}1\times1, 64\\3\times3, 64\\1\times1, 256\end{array}\right]\times3$	$\left[\begin{array}{c}1\times1, 64\\3\times3, 64\\1\times1, 256\end{array}\right]\times3$	$\left[\begin{array}{c}1\times1, 64\\3\times3, 64\\1\times1, 256\end{array}\right]\times3$
conv3_x	28×28	$\left[\begin{array}{c}3\times3, 128\\3\times3, 128\end{array}\right]\times2$	$\left[\begin{array}{c}3\times3, 128\\3\times3, 128\end{array}\right]\times4$	$\left[\begin{array}{c}1\times1, 128\\3\times3, 128\\1\times1, 512\end{array}\right]\times4$	$\left[\begin{array}{c}1\times1, 128\\3\times3, 128\\1\times1, 512\end{array}\right]\times4$	$\left[\begin{array}{c}1\times1, 128\\3\times3, 128\\1\times1, 512\end{array}\right]\times8$
conv4_x	14×14	$\left[\begin{array}{c}3\times3, 256\\3\times3, 256\end{array}\right]\times2$	$\left[\begin{array}{c}3\times3, 256\\3\times3, 256\end{array}\right]\times6$	$\left[\begin{array}{c}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{array}\right]\times6$	$\left[\begin{array}{c}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{array}\right]\times23$	$\left[\begin{array}{c}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{array}\right]\times36$
conv5_x	7×7	$\left[\begin{array}{c}3\times3, 512\\3\times3, 512\end{array}\right]\times2$	$\left[\begin{array}{c}3\times3, 512\\3\times3, 512\end{array}\right]\times3$	$\left[\begin{array}{c}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{array}\right]\times3$	$\left[\begin{array}{c}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{array}\right]\times3$	$\left[\begin{array}{c}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{array}\right]\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9