18AIC301J: DEEP LEARNING TECHNIQUES

B. Tech in ARTIFICIAL INTELLIGENCE, 5th semester

Faculty: **Dr. Athira Nambiar**

Section: A, slot:D

Venue: TP 804

Academic Year: 2022-22

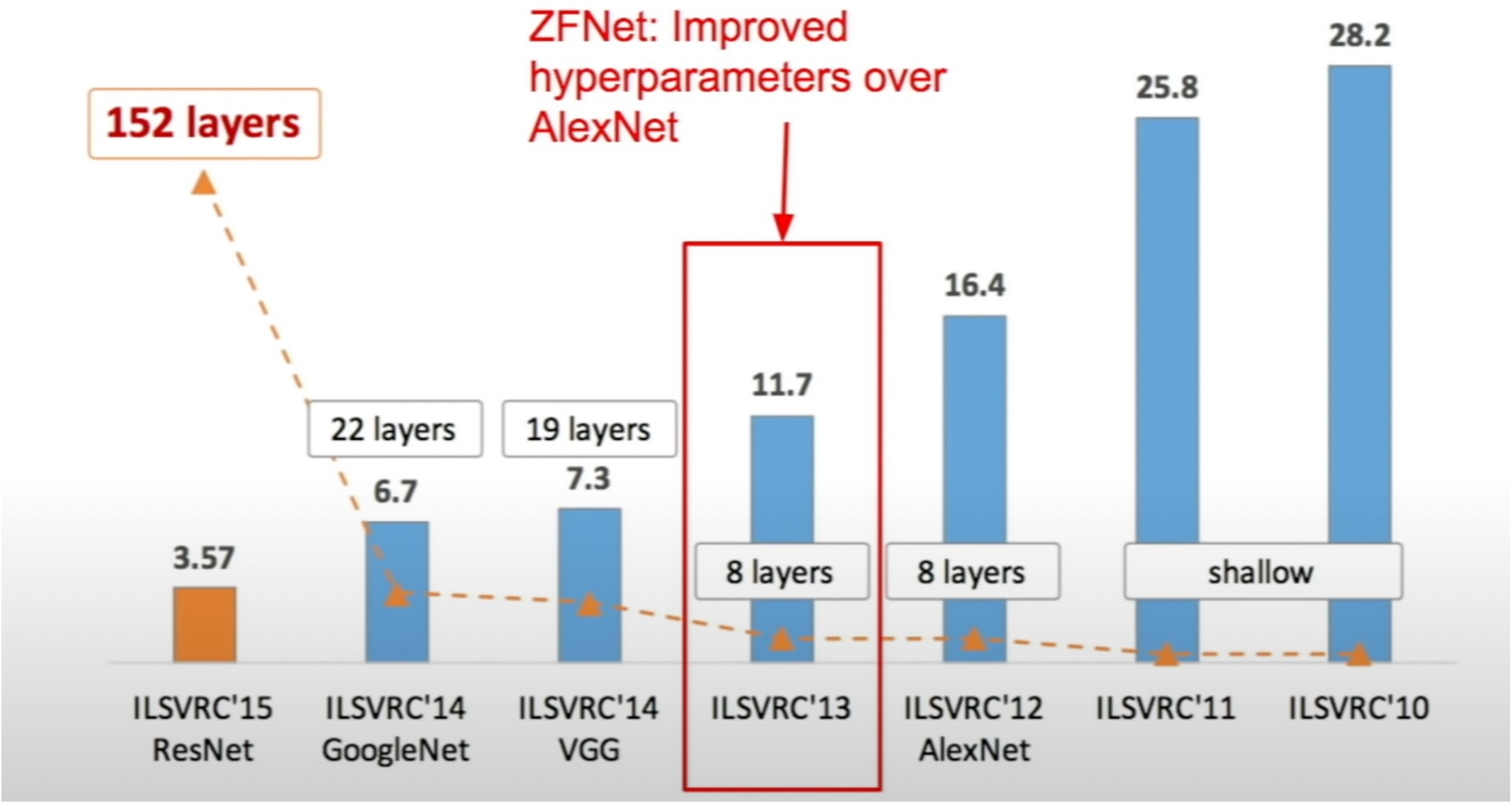
UNIT-3

One hot representation of words, Distributed representation of words
SVD for learning word Representations, Continuous bag of words model, Skip-gram model, Hierarchical Softmax
Implement skip gram model to predict words within a certain range before and after the current word
Introduction to Convolution Neural Networks, Kernel filters
The convolution operation with Filters, padding and stride, Multiple Filters, Max pooling and non-linearities
Implement LeNet for image classification
Classic CNNs architecture- The ImageNet challenge, Understanding Alex Net architecture
ZFNet, The intuition behind GoogleNet, Average pooling, Residual CNN-ResNet architecture
Implement ResNet for detecting Objects.

UNIT-3

One hot representation of words, Distributed representation of words
SVD for learning word Representations, Continuous bag of words model, Skip-gram model, Hierarchical Softmax
Implement skip gram model to predict words within a certain range before and after the current word
Introduction to Convolution Neural Networks, Kernel filters
The convolution operation with Filters, padding and stride, Multiple Filters, Max pooling and non-linearities
Implement LeNet for image classification
Classic CNNs architecture- The ImageNet challenge, Understanding Alex Net architecture
ZFNet, The intuition behind GoogleNet, Average pooling, Residual CNN-ResNet architecture
Implement ResNet for detecting Objects.

ZFNet



ZFNet: Improved hyperparameters over AlexNet
152 layers
22 layers
19 layers
8 layers
8 layers
shallow
3.57
6.7
7.3
11.7
16.4
25.8
28.2
ILSVRC'15 ResNet
ILSVRC'14 GoogleNet
ILSVRC'14 VGG
ILSVRC'13
ILSVRC'12 AlexNet
ILSVRC'11
ILSVRC'10

ZFNet

ZFNet (2013)

- ❑ **ZFNet** the winner of the competition ILSVRC 2013 with **14.8%** Top-5 error rate
- ❑ **ZFNet** built by Matthew Zeiler and Rob Fergus
- ❑ **ZFNet** has the same global architecture as Alexnet, that is to say 5 convolutional layers, two fully connected layers and an output softmax one. The differences are for example better sized convolutional kernels.
- ❑ **ZFNet** used filters of size 7x7 and a decreased stride value, instead of using 11x11 sized filters in the first layer (which is what AlexNet implemented).

ZFNet

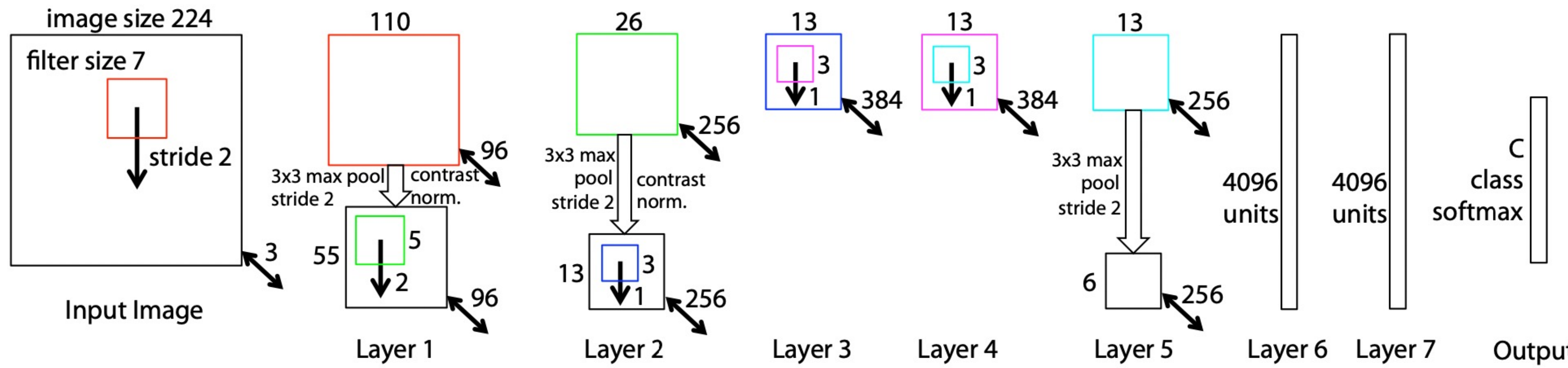


Figure 3. Architecture of our 8 layer convnet model. A 224 by 224 crop of an image (with 3 color planes) is presented as the input. This is convolved with 96 different 1st layer filters (red), each of size 7 by 7, using a stride of 2 in both x and y. The resulting feature maps are then: (i) passed through a rectified linear function (not shown), (ii) pooled (max within 3x3 regions, using stride 2) and (iii) contrast normalized across feature maps to give 96 different 55 by 55 element feature maps. Similar operations are repeated in layers 2,3,4,5. The last two layers are fully connected, taking features from the top convolutional layer as input in vector form ($6 \cdot 6 \cdot 256 = 9216$ dimensions). The final layer is a C-way softmax function, C being the number of classes. All filters and feature maps are square in shape.

AlexNet but:
CONV1: change from (11x11 stride 4) to (7x7 stride 2)
CONV3,4,5: instead of 384, 384, 256 filters use 512, 1024, 512

ZFNet

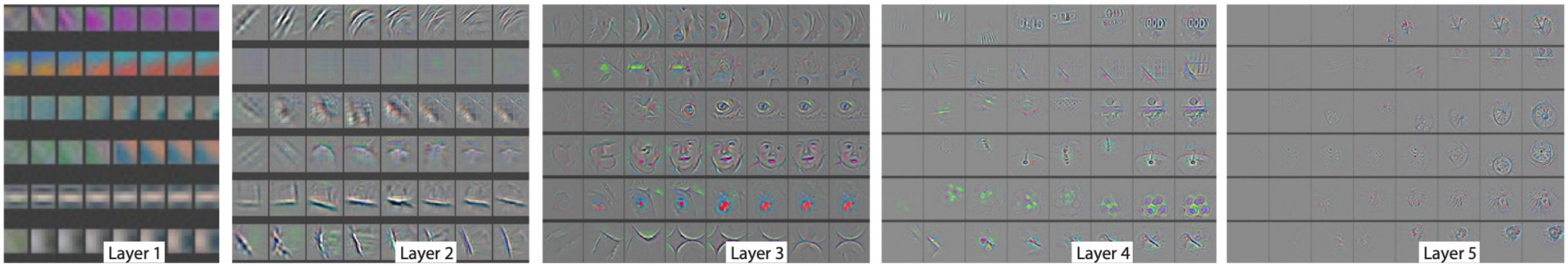
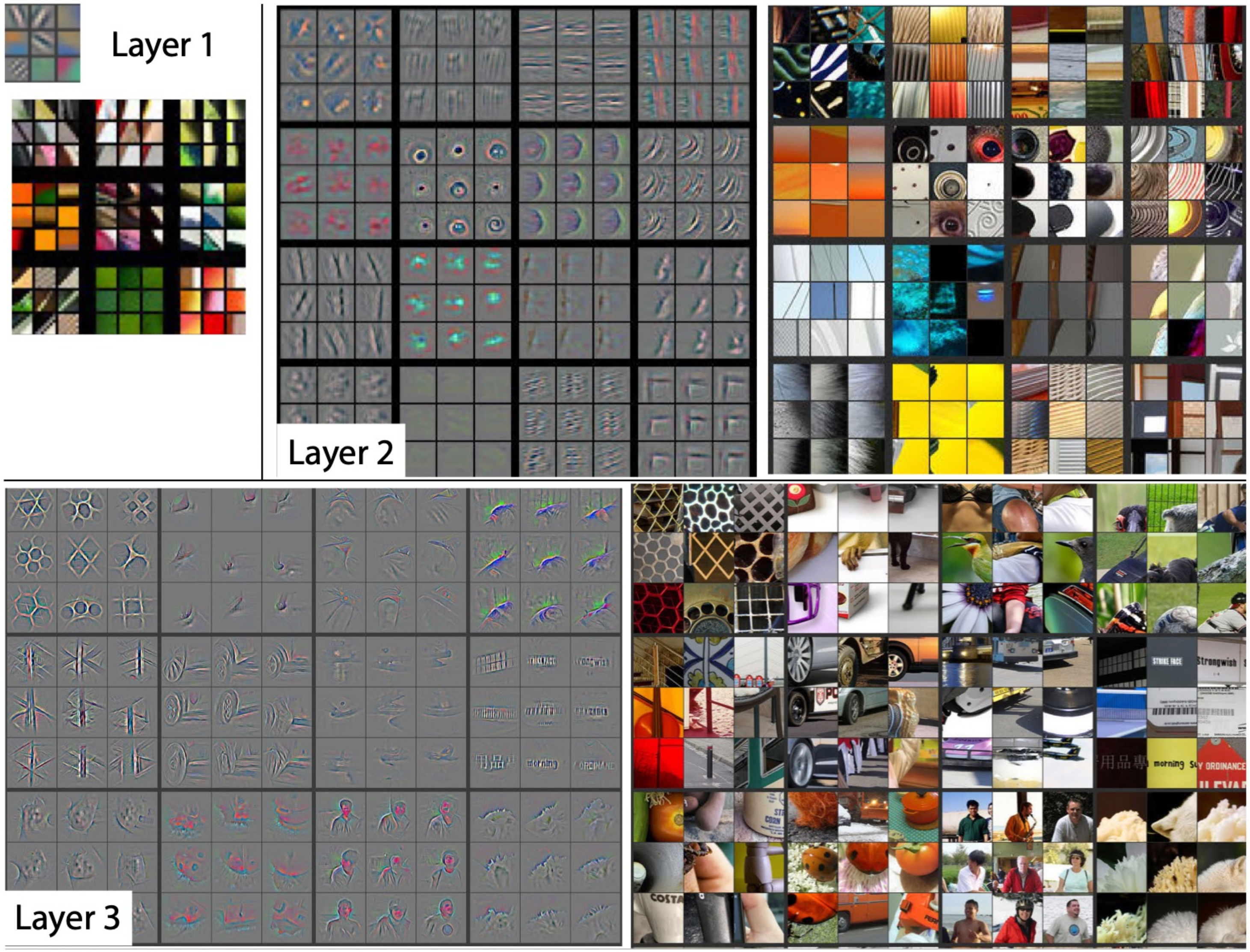


Figure 4. Evolution of a randomly chosen subset of model features through training. Each layer's features are displayed in a different block. Within each block, we show a randomly chosen subset of features at epochs [1,2,5,10,20,30,40,64]. The visualization shows the strongest activation (across all training examples) for a given feature map, projected down to pixel space using our deconvnet approach. Color contrast is artificially enhanced and the figure is best viewed in electronic form.

ZFNet



Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." In *European conference on computer vision*, pp. 818-833. Springer, Cham, 2014.

ZFNet

- ❑ **ZFNet** trained on a GTX 580 GPU for **twelve days.**
- ❑ Developed a visualization technique named Deconvolutional Network "deconvnet" because it maps features to pixels.

- Unfold the secrets of how neural networks see our world!
- ZFNet is a classic convolutional neural network. The design was motivated by visualizing intermediate feature layers and the operation of the classifier.

The intuition behind GoogleNet

GoogLeNet (2014)

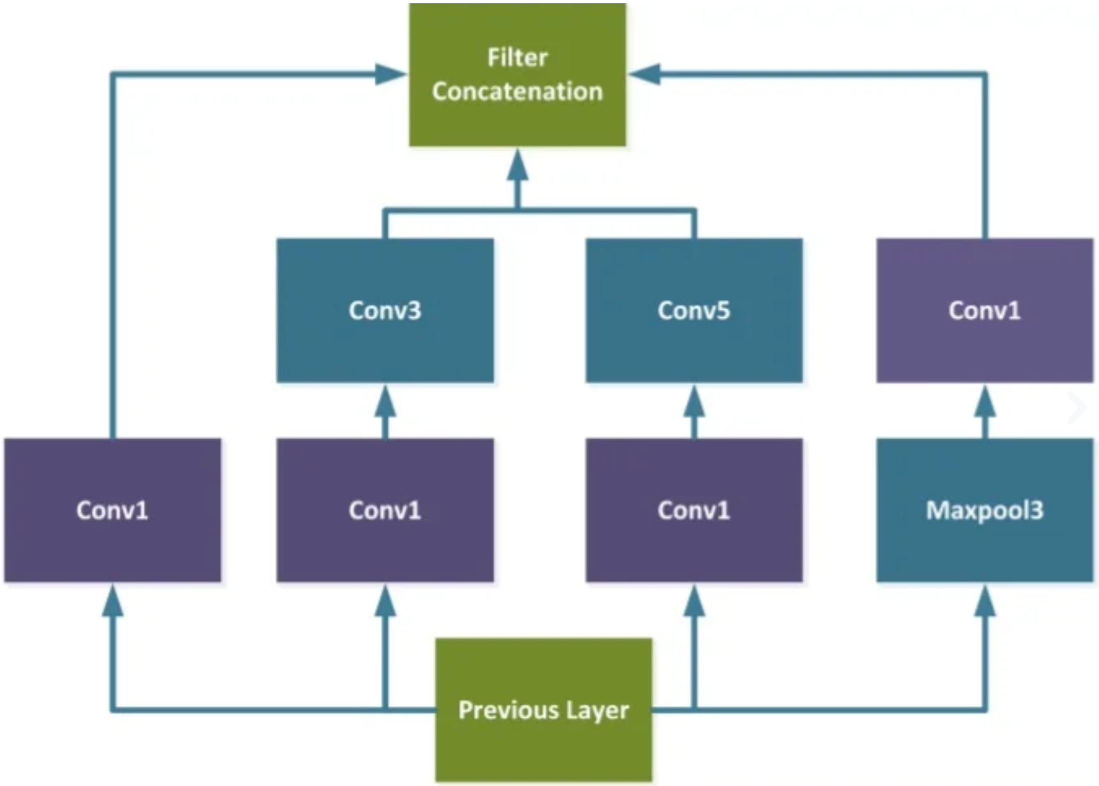
- **GoogleNet** is the winner of the competition ILSVRC 2014 with **6.7%** Top-5 error rate.
- **GoogleNet** Trained on "a few high-end GPUs **with in a week**"
- **GoogleNet** uses 12x fewer parameters than AlexNet
- **GoogleNet** use an average pool instead of fully connected layers, to go from a 7x7x1024 volume to a 1x1x1024 volume. This saves a huge number of parameters.

The intuition behind GoogleNet

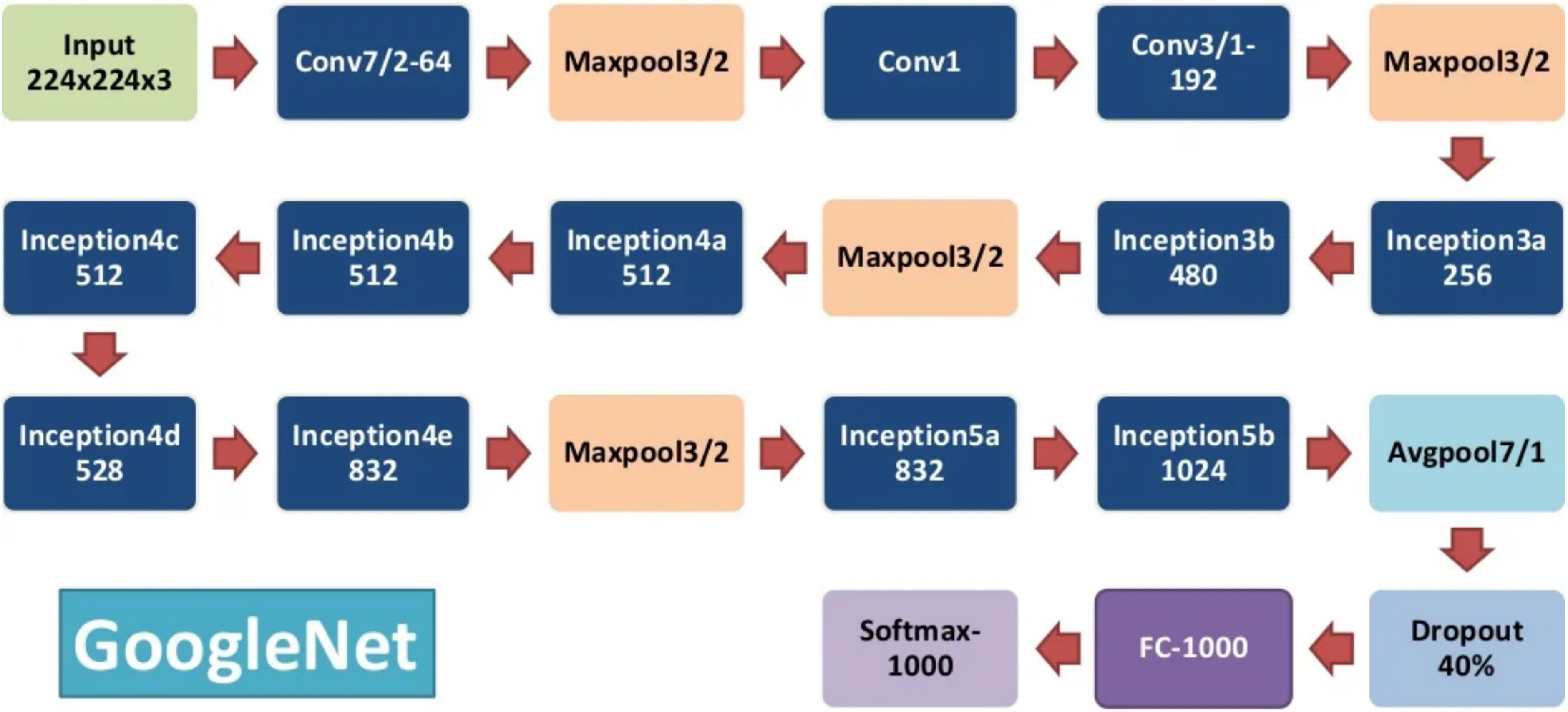
- ❑ **GoogleNet** used 9 Inception modules in the whole architecture
- ❑ This 1x1 convolutions (bottleneck convolutions) allow to control/reduce the depth dimension which greatly reduces the number of used parameters due to removal of redundancy of correlated filters.
- ❑ **GoogleNet** has 22 Layers deep network
- ❑ **GoogleNet** use an average pool instead of using FC-Layer, to go from a 7x7x1024 volume to a 1x1x1024 volume. This saves a huge number of parameters.
- ❑ **GoogleNet** use inexpensive Conv1 to compute reduction before the expensive Conv3 and Conv5
- ❑ Conv1 follow by Relu to reduce overfitting

The intuition behind GoogleNet

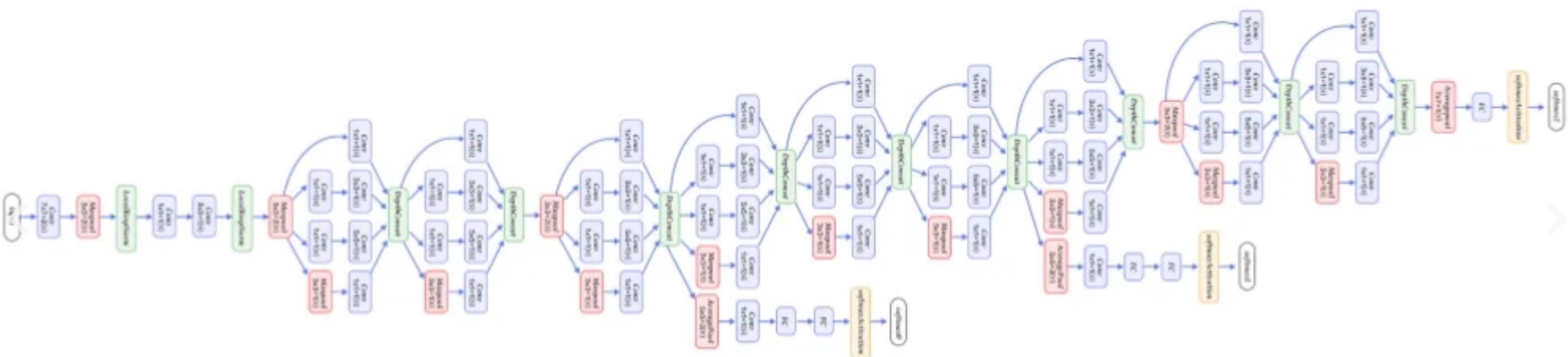
- ❑ Inception module



The intuition behind GoogleNet



The intuition behind GoogleNet



Residual CNN

Residual Networks:

- It has been proved that adding more layers to a Neural Network can make it more robust for image-related tasks. But it can also cause them to lose accuracy. That's where Residual Networks come into place.

- if we add more than 30 layers to the network, then its performance suffers and it attains a low accuracy.

- This is contrary to the thinking that the addition of layers will make a neural network better.
- This is not due to overfitting, because in that case, one may use dropout and regularizationtechniques to solve the issue altogether.

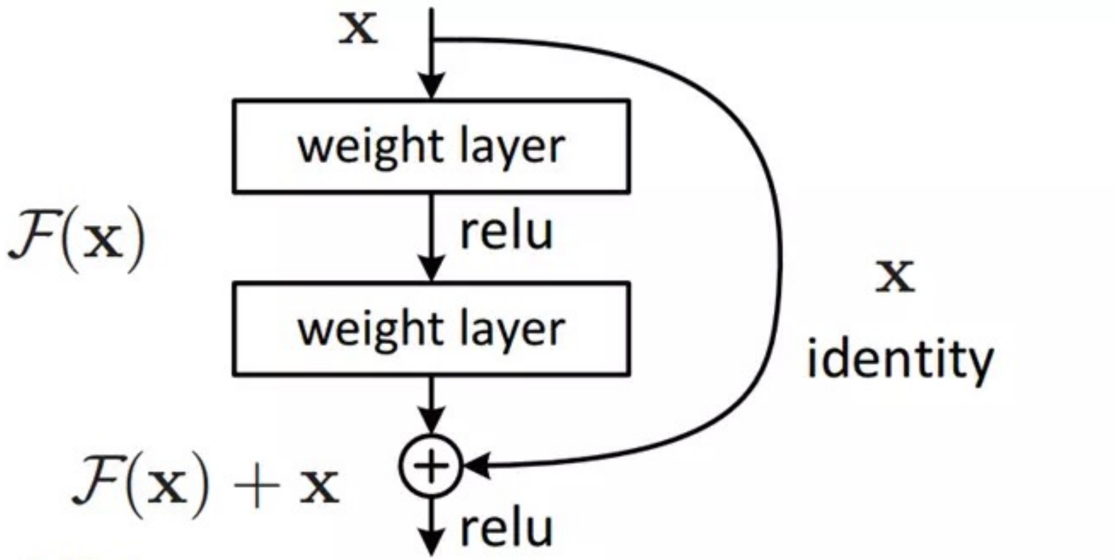
- It's mainly present because of the popular vanishing gradient problem.

Residual CNN

Residual Networks:

- The **ResNet152** model with 152 layers won the ILSVRC Imagenet 2015 test while having lesser parameters than the **VGG19** network, which was very popular at that time.

- A residual network consists of residual units or blocks which have ***skip connections***, also called ***identity connections***.

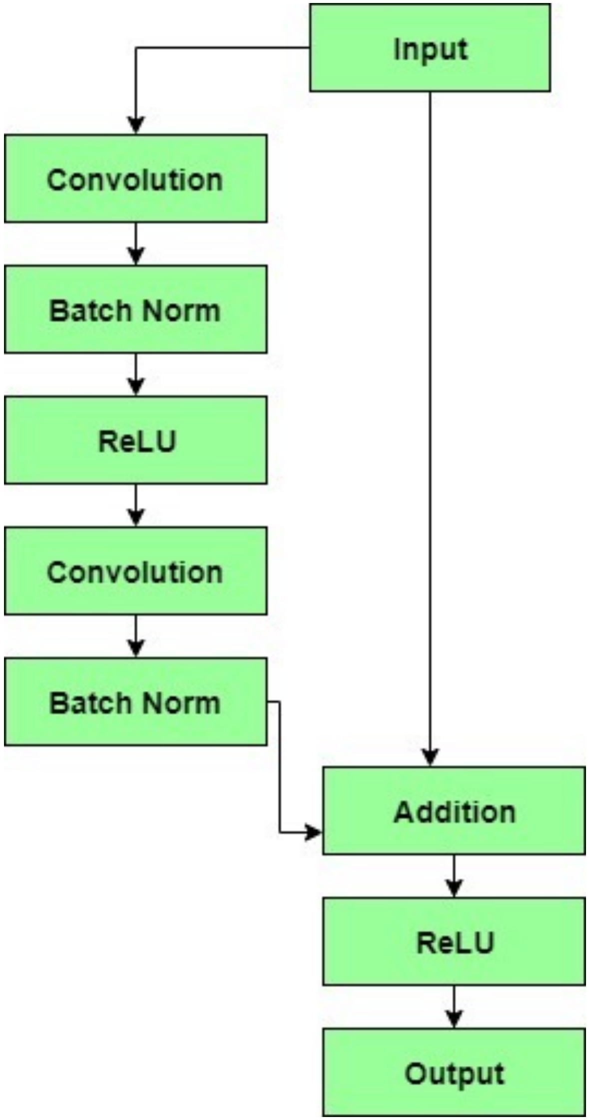


A residual block

Residual CNN

The output of the previous layer is added to the output of the layer after it in the residual block. The hop or skip could be *1, 2 or even 3*. When adding, the dimensions of *x* may be different than *F(x)* due to the convolution process, resulting in a reduction of its dimensions. Thus, we add an additional *1 x 1* convolution layer to change the dimensions of *x*.

Residual CNN-ResNet architecture



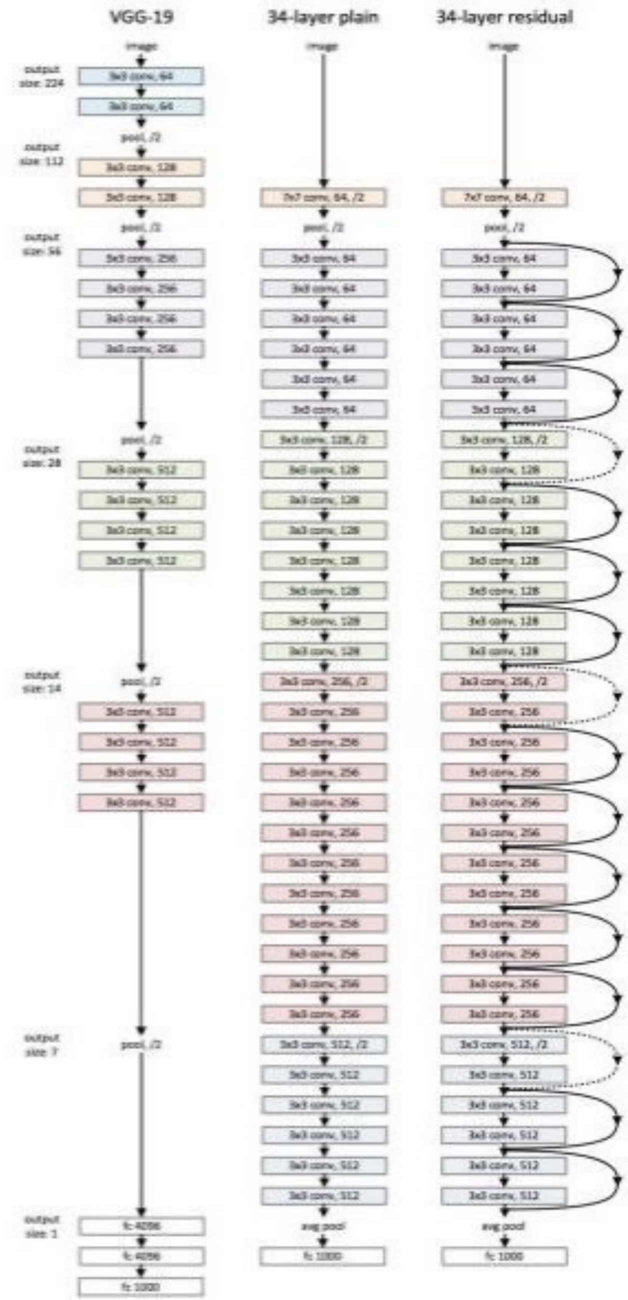
- A residual block has a *3 x 3* convolution layer followed by a batch normalization layer and a ReLU activation function.
- This is again continued by a *3 x 3* convolution layer and a batch normalization layer.
- The skip connection basically skips both these layers and adds directly before the ReLU activation function.
- Such residual blocks are repeated to form a residual network.

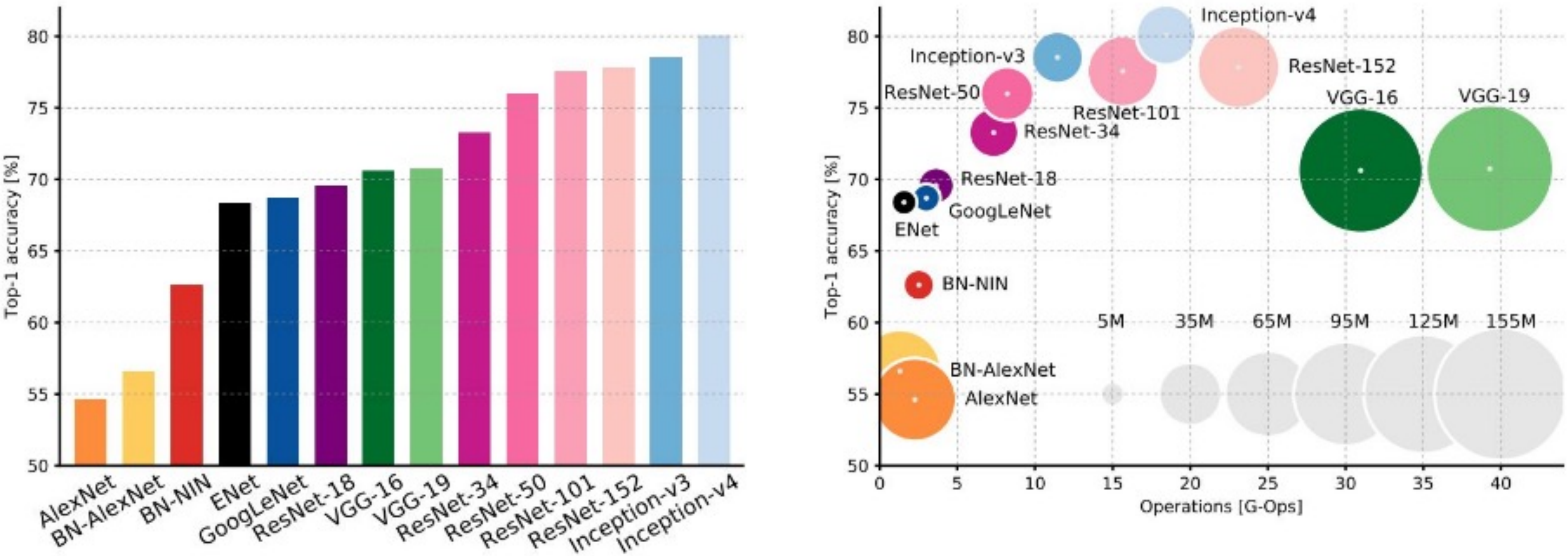
ResNet architecture

- After an in-depth comparison of all the present CNN architectures was done, the ResNet stood out by holding the lowest top 5% error rate at *3.57%* for classification tasks, overtaking all the other architectures.

- Even humans do not have much lower error rates.

- Comparison of *34* layer ResNet with VGG19 and a *34* layer plain network:





An Analysis of Deep Neural Network Models for Practical Applications, 2017.

Year	CNN	Developed by	Place	Top-5 error rate	No. of parameters
1998	LeNet(8)	Yann LeCun et al			60 thousand
2012	AlexNet(7)	Alex Krizhevsky, Geoffrey Hinton, Ilya Sutskever	1st	15.3%	60 million
2013	ZFNet()	Matthew Zeiler and Rob Fergus	1st	14.8%	
2014	GoogLeNet(19)	Google	1st	6.67%	4 million
2014	VGG Net(16)	Simonyan, Zisserman	2nd	7.3%	138 million
2015	ResNet(152)	Kaiming He	1st	3.6%	

Learning Resources

- Charu C. Aggarwal, Neural Networks and Deep Learning, Springer, 2018.
- Eugene Charniak, Introduction to Deep Learning, MIT Press, 2018.
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
- Michael Nielsen, Neural Networks and Deep Learning, Determination Press, 2015.
- Deng & Yu, Deep Learning: Methods and Applications, Now Publishers, 2013.

- https://youtu.be/DAOcjicFr1Y
- https://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/
- https://coderzpy.com/cnn-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more/

Thank you