CS F425: Deep Learning

19

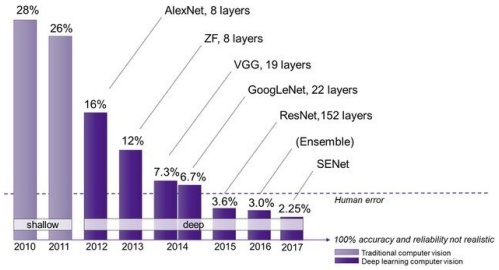
CNN Le-NET, ResNET DenseNET, U-NET

Dr. Kamlesh Tiwari
Assistant Professor, Department of CSIS,
BITS Pilani, Pilani Campus, Rajasthan-333031 INDIA

Mar 03, 2023 ON-CAMPUS Campus @ BITS-Pilani [Jan-May 2023]

http://ktiwari.in/dl

ImageNet ILSVRC[1]



- (2009) 22K category, 14M images
- Challenge 1000 class, 1431167 images
- HoG, LBP, SVM ...

[1] Imagenet large scale visual recognition challenge http://www.image-net.org/challenges/LSVRC/

AlexNet [2]

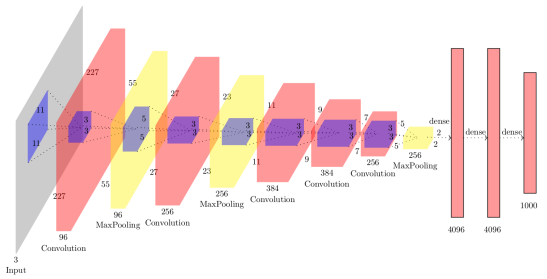


Image size 227 × 227 × 3 → [96 F=11, S=4, P=0] → [MaxPool F=3, S=2] → [256 F=5, S=1, P=0] → [MaxPool F=3, S=2] → [384 F=3, S=1, P=0] → [384 F=3, S=1, P=0] → [256 F=3, S=1, P=0] → [MaxPool F=3, S=2] → FC 4096 → FC 4096 → FC 1000

27.55 M parameters. Convolution is passes through ReLU activation.

[2] Cite: 128212, **Imagenet classification with deep convolutional neural networks**, Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E, In: Advances in neural information processing systems pages 1097–1105, NIPS-2012

ZFNet [3]

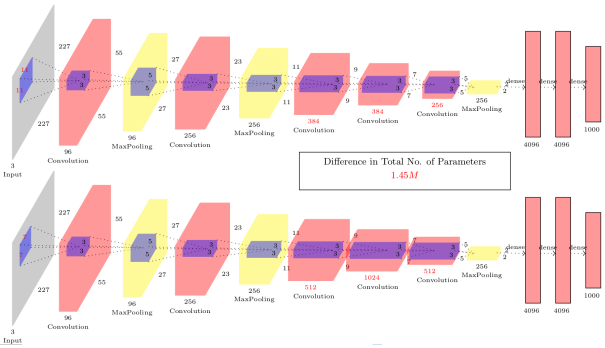
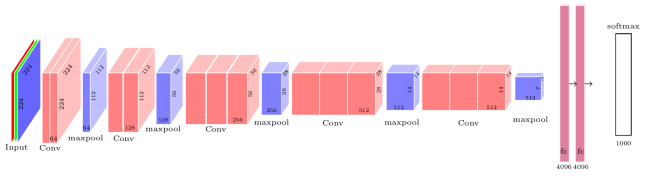


Image size 227 × 227 × 3 → [69 F=7, S=4, P=0] → [MaxPool F=3, S=2] → [256 F=5, S=1, P=0] → [MaxPool F=3, S=2] → [512 F=3, S=1, P=0] → [1024 F=3, S=1, P=0] → [512 F=3, S=1, P=0] → [MaxPool F=3, S=2] → FC 4096 → FC 4096 → FC 1000

[3] Cite 18586 **Visualizing and understanding convolutional networks**, Zeiler, Matthew D and Fergus, Rob, pages 818–833, European conference on computer vision (ECCV) Springer-2014

VGG16 [4]

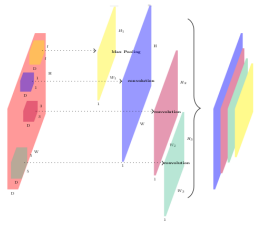


1. Kernel size is always 3 × 3
2. 16M parameters in pre-FC and 122 in FC. First FC layer is huge
3. Layers represents abstract representation and can be reused (FC or Conv)

[4] Cite 96789 **Very deep convolutional networks for large-scale image recognition**, Simonyan, Karen and Zisserman, Andrew, pages 818–833, arXiv preprint arXiv(2014):1409.1556 - In: ICLR-2015

GoogLeNet [5]

- Recall scale invariance in SIFT
- Multiple filters of different size is a good idea
- With $W \times H \times D$ input and $F \times F \times D$ Filter and $S = 1$ and no padding, output is of size $(W - F + 1) \times (H - F + 1)$
- Each value needs $F \times F \times D$ computation

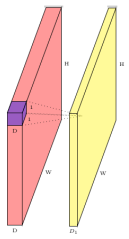


Can we reduce this computation a bit?

Idea is to have 1 × 1 computation

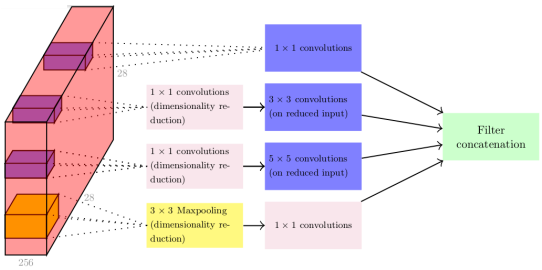
[5] Cite: 47942, **Going deeper with convolutions**, Szegedy, Christian and Liu, Wei and Jia, Yangqing and Sermanet, Pierre and Reed, Scott and Anguelov, Dragomir and Erhan, Dumitru and Vanhoucke, Vincent and Rabinovich, Andrew, In: Proceedings of the IEEE conference on computer vision and pattern recognition pages 1–9, CVPR-2015

1×1 convolution



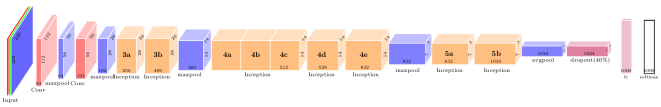
- 1×1 is $1 \times 1 \times D$
- They produce one output place
- By using D_1 such 1×1 convolution output becomes $F \times F \times D_1$
- We have $D_1 < D$

Inception Block: Multiple Convolutions



1. 1×1 convolution
2. 1×1 convolution followed by 3×3
3. 1×1 convolution followed by 5×5
4. 3×3 maxpool followed by 1×1
5. Appropriate padding is done to make things of same size

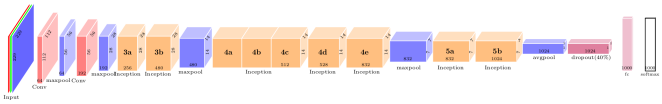
GoogLeNet



1. Input is RGB 229×229
2. Each inception module have very specific configuration.

(3a)	$192 \times 28 \times 28$	64	96	128	16	32	32
(3b)	$256 \times 28 \times 28$	28	128	192	32	96	61
(4a)	$48 \times 14 \times 14$	192	96	208	16	48	96
(4b)	$512 \times 14 \times 14$	160	112	224	24	64	64
(4c)	$512 \times 14 \times 14$	128	128	256	24	64	64
(4d)	$512 \times 14 \times 14$	112	144	228	32	64	64
(4e)	$528 \times 14 \times 14$	256	160	320	32	128	128
(5a)	$832 \times 7 \times 7$	256	160	320	32	128	128
(5b)	$832 \times 7 \times 7$	384	192	384	48	124	128

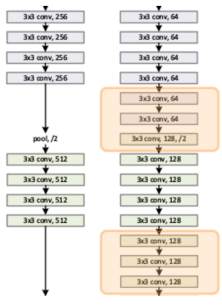
GoogLeNet



- VGGNET has $512 \times 7 \times 7$ size at pre-FC this was an issue to connect with 4096
- GoogLeNet applies a average pool. Gives 49 time reduction. has 1024 values only
- Dropout and connect to 1000
- 12 times less connections as compared to AlexNet
- 2 times more computation as compared to AlexNet
- Very high accuracy. Error reduced from 16% -to- 6.7%

ResNet [6]

If a shallow neural network works well. What would happen if we add more layers?

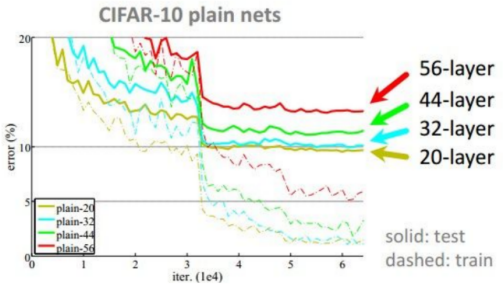


- Deep network should also work well (It would learn identity in new layers)

[6] Cite: 32871, **Deep residual learning for image recognition**, He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, In: IEEE conference on computer vision and pattern recognition, pages 770–778, CVPR-2016

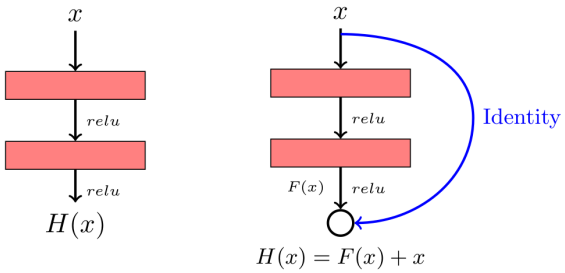
ResNet

But, in practice it was not happening

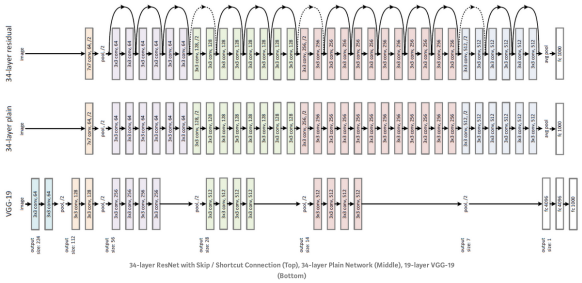


Why? Identity is one of the solution in large domain.

Let me tell this to the network



ResNet Comparison [7]



152-layer deep net was better than human. Only 3.6% error rate ImageNet Classification [a]

[a] Better than the 2nd best system **ImageNet Detection:** 16% **ImageNet Localization:** 27% **COCO Detection:** 11% **COCO Segmentation:** 12%

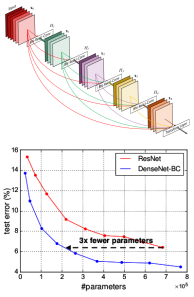
[7] ResNet — Winner of ILSVRC 2015 (Image Classification, Localization, Detection) Sik-Ho Tsang

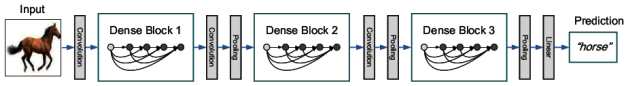
ResNet Hyper-parameters and Issues

- Training takes huge time
- Batch Normalization
- Zavier/2 initialization
- SGD and momentum
- Small learning rate 0.1
- Mini-batch size 256
- Weight decay
- No Dropout

DenseNet [8]

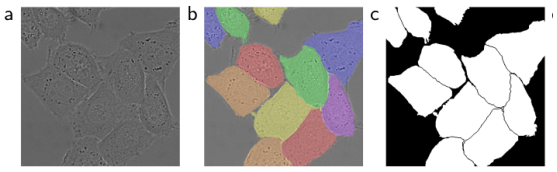
- Connect each layer to every other layer in feed-forward fashion
- It alleviates vanishing-gradient problem, strengthen feature propagation
- With feature reuse, substantially reduce number of parameters
- CIFAR10-3.6%, CIFAR100-19.6%,





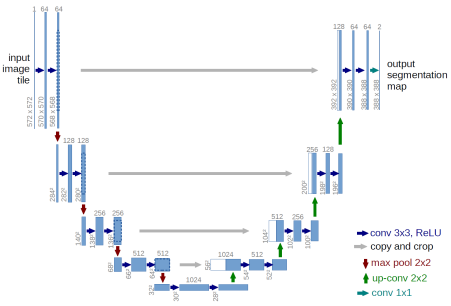
[8] **Cite 32471**, Huang, Gao and Liu, Zhuang and Van Der Maaten, Laurens and Weinberger, Kilian Q, **Densely connected convolutional networks**, In IEEE CVPR vision and pattern recognition, pp 4700–4708, 2017

Segmentation



- ISBI challenge for segmentation of neuronal structures in electron microscopic stacks
- Works with very few training images (30/application) and touching boundary. Yield more precise segmentation
- Data augmentation is essential (mainly shift, rotation and elastic deformation)

U-Net[9]



- ISBI DIC-HeLa achieved 77.6% iou as compared to 46.0% second
- ISBI Cell tracking 2015, achieved 92% IoU as compared to 83% second

[9] **Cite 57528** O. Ronneberger and P.Fischer and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Medical Image Computing and Computer-Assisted Intervention (MICCAI), LNCS-9351, pages 234–241, Springer-2015

Thank You!

Thank you very much for your attention!