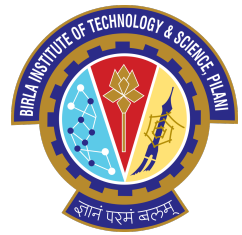
CS-F441: Selected Topics from Computer Science (Deep Learning for NLP & CV)

Lecture-KT-12: CNN Architectures, LeNet, AlexNet, ZFNet, VGG



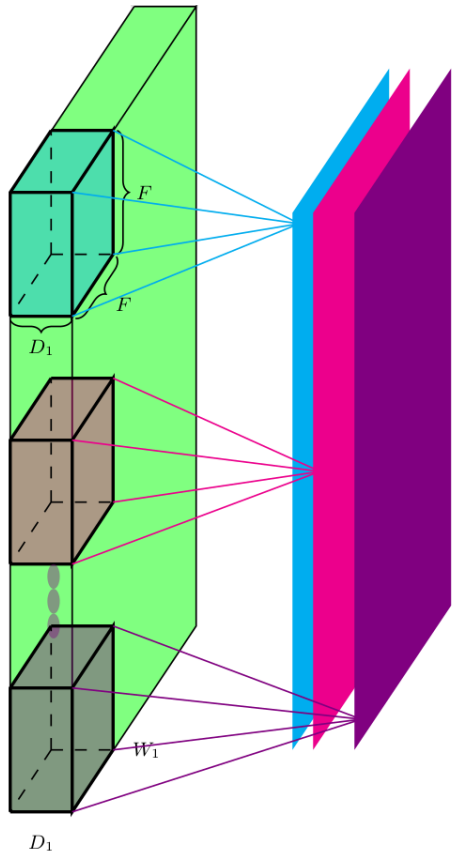
Dr. Kamlesh Tiwari,
Assistant Professor,
Department of Computer Science and Information Systems,
BITS Pilani, Rajasthan-333031 INDIA

Nov 11, 2019 (Campus @ BITS-Pilani July-Dec 2019)

Recap: Motivation

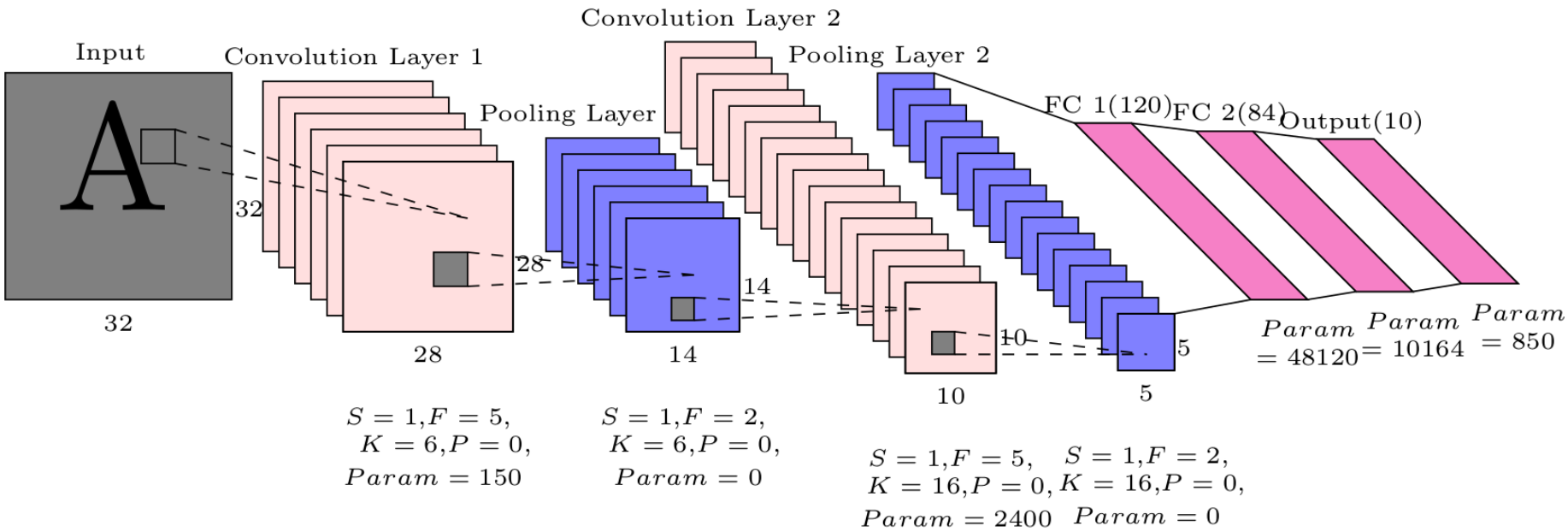
CNN are sparse FC-NN with weight sharing

- Multiple filters to get multiple feature maps
- Filter/Kernels are treated as parameters to learning them
- Filters, Padding, Stride and Pooling
- Classification pipeline has components



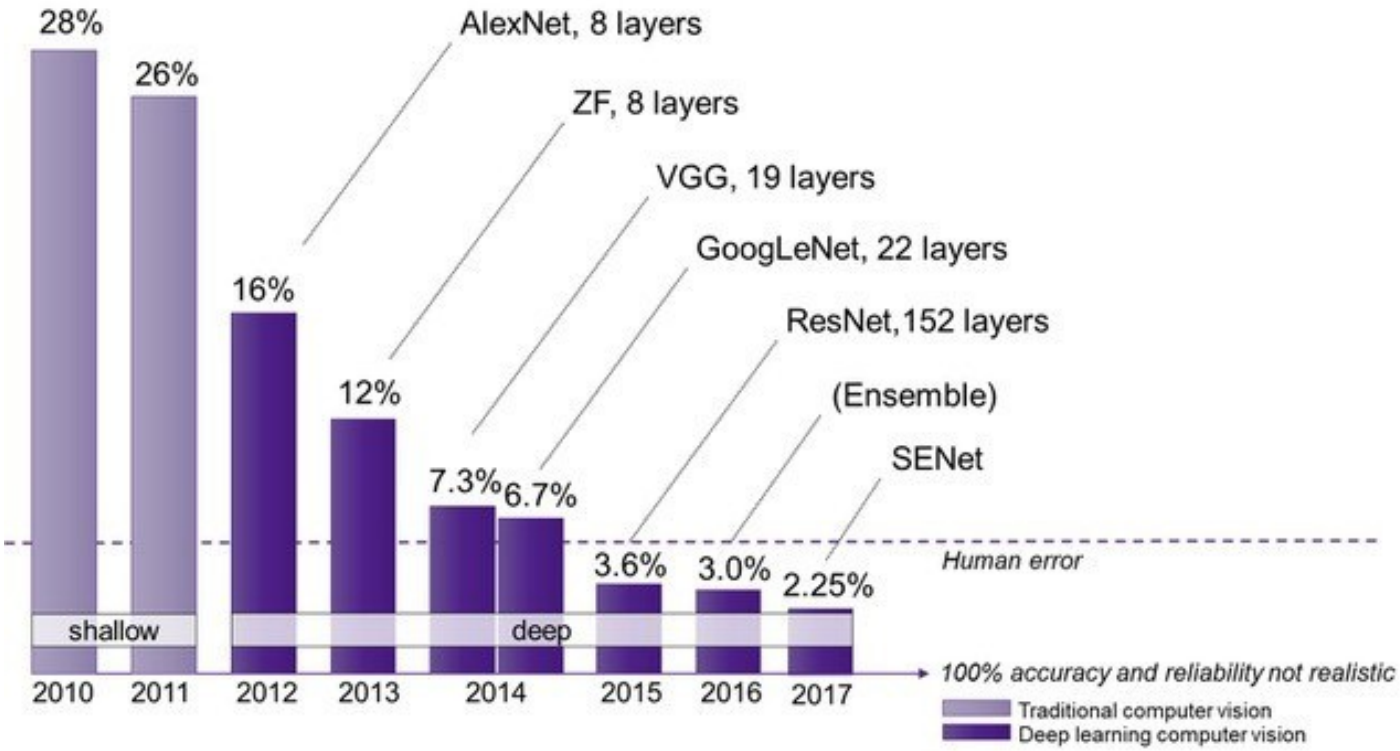
$$H, W = \frac{H - F + 2P}{S} + 1, \frac{S - F + 2P}{S} + 1$$

LeNet-5 for handwritten character recognition [1]



[1] Yann Lecun and Leon Bottou and Yoshua Bengio and Patrick Haffner, *Gradient-based learning applied to document recognition*, pp 2278–2324, IEEE-1998

ImageNet ILSVRC[2]



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

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

AlexNet [3]

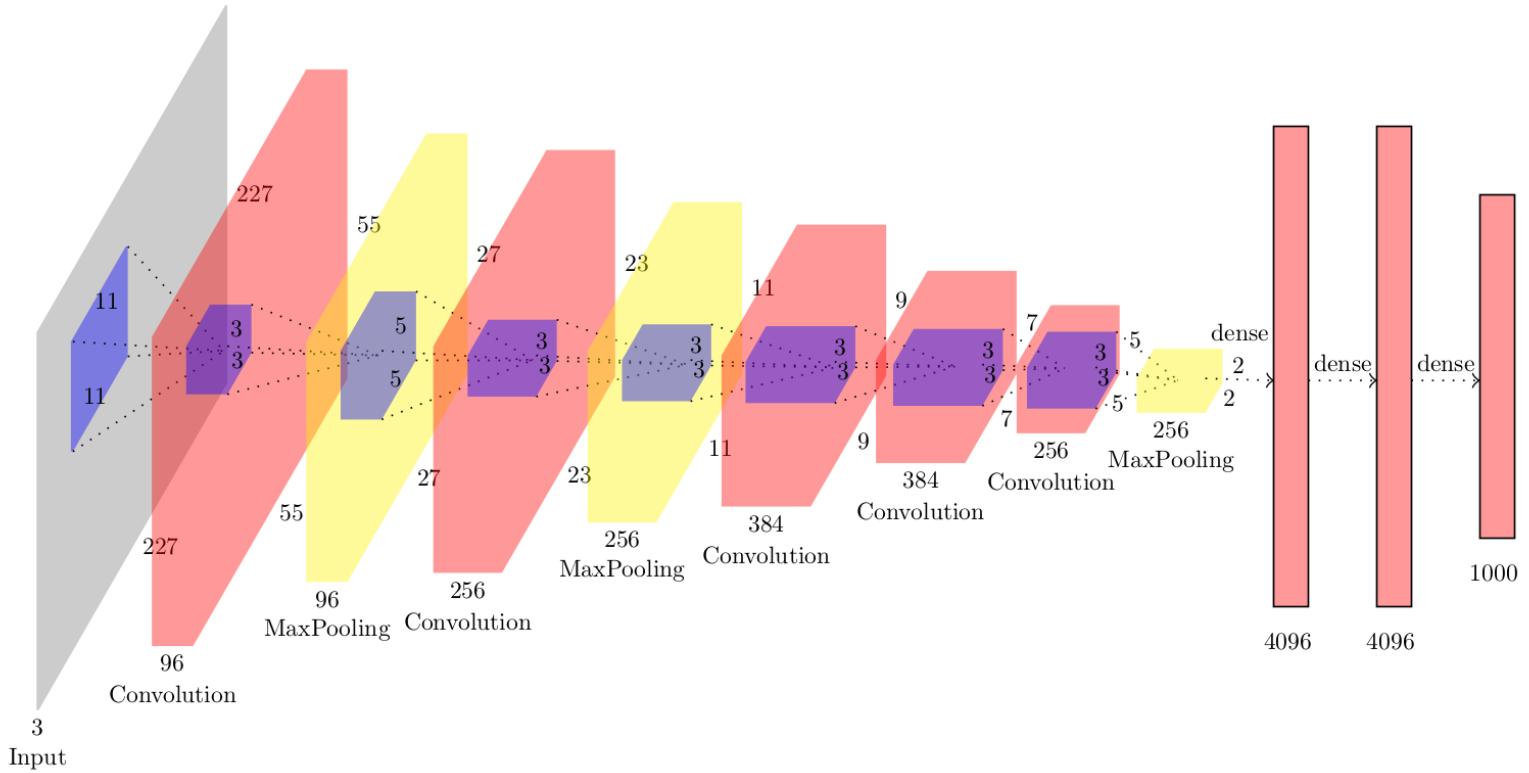


Image size 227 × 227 × 3 → [69 F=11, S=4, P=0] → [MaxPool F=3, S=2] → [256 F=5, S=1, P=0] → [MaxPool F=3, S=2] → [384 F=3, S=1, P=0] → [384 F=3, S=1, P=0] → [256 F=3, S=1, P=0] → [MaxPool F=3, S=2] → FC 4096 → FC 4096 → FC 1000

27.55 M parameters. Convolution is passes through RelU activation.

[3] Cite: 50086, **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 [4]

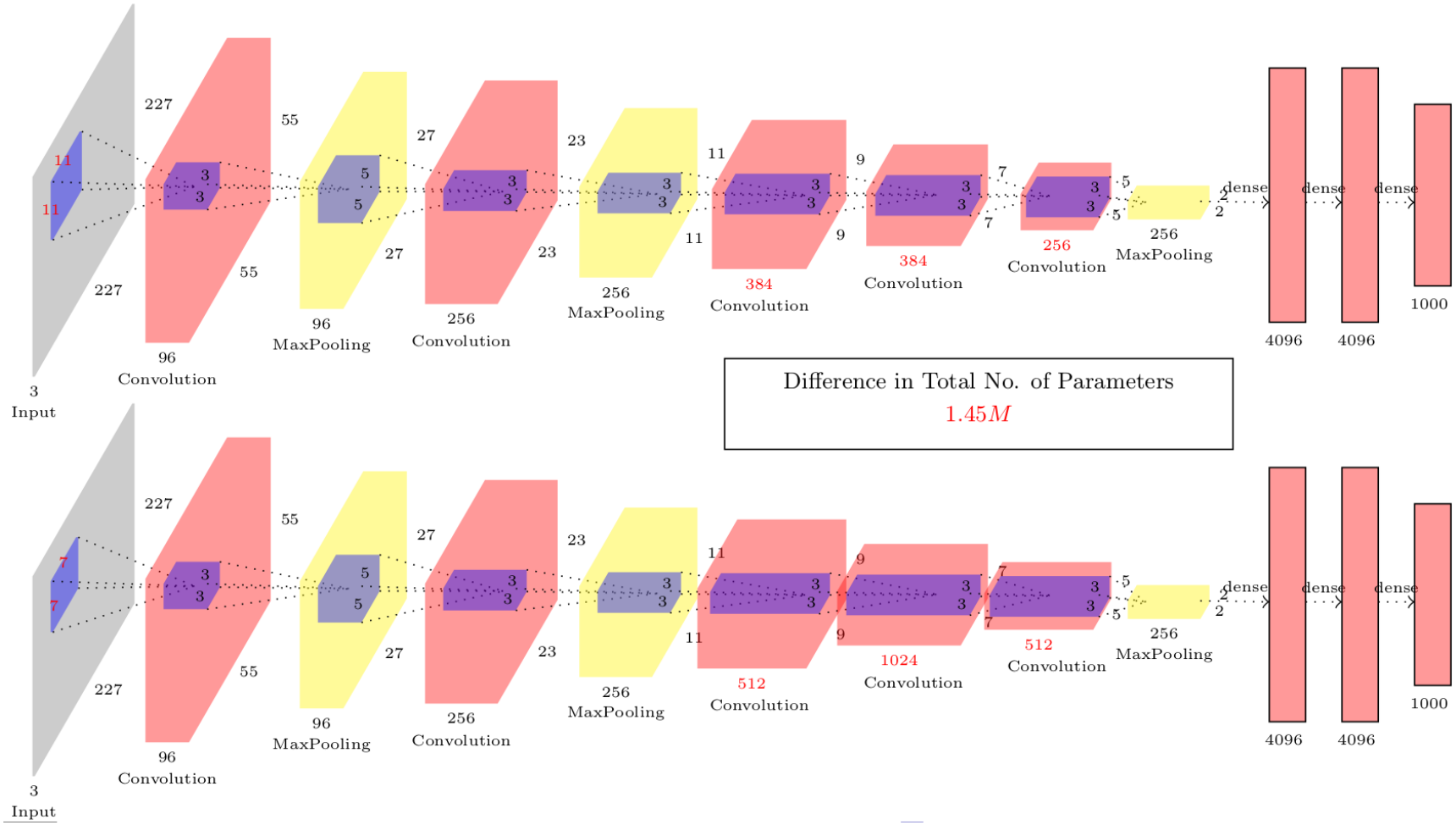
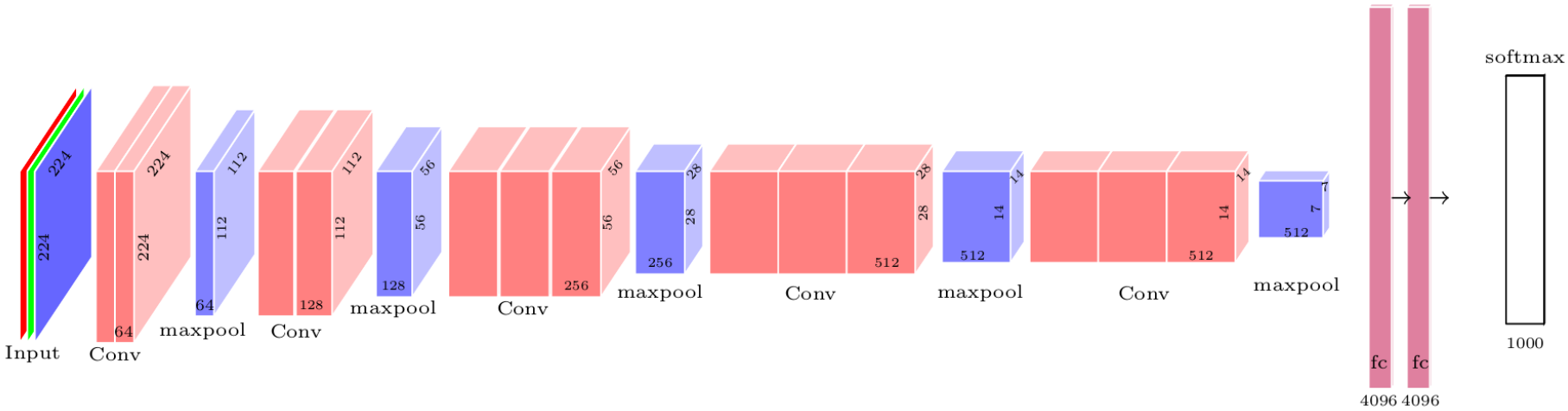


Image size 227 × 227 × 3 → [69 F=7, S=4, P=0] → [MaxPool F=3, S=2] → [256 F=5, S=1, P=0] → [MaxPool F=3, S=2] → [512 F=3, S=1, P=0] → [1024 F=3, S=1, P=0] → [512 F=3, S=1, P=0] → [MaxPool F=3, S=2] → FC 4096 → FC 4096 → FC 1000

[4] Cite 7560 **Visualizing and understanding convolutional networks**, Zeiler, Matthew D and Fergus, Rob, pages 818–833, European conference on computer vision (ECCV) Springer-2014

VGG16 [5]



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)

[5] Cite 29548 **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

Thank You!

Thank you very much for your attention[6]!

Queries ?

[6] Credit: Prof. Mitesh Khapra, Deep Learning(CS7015): Lec 11.4 CNNs for NPTEL