Convolutional neural network architectures for image classification and detection

Mohammad G. Syed[1,2], Stephen McKenna[2], Mohan V. [3],Emanuele Trucco[2]

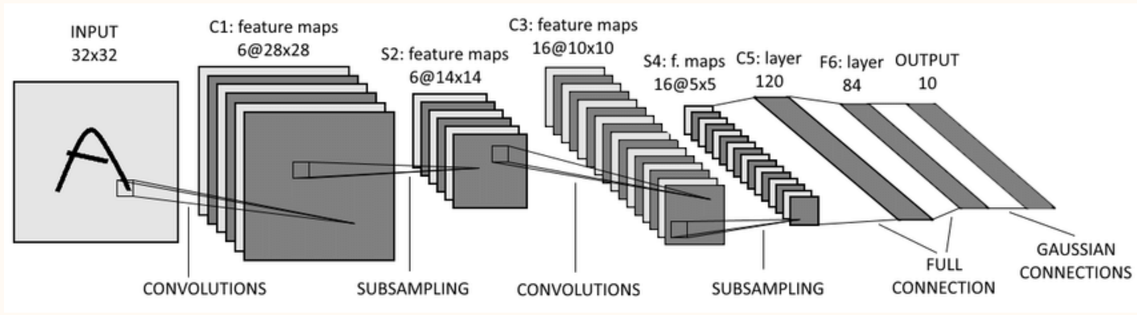
[1]School of Medicine, University of Dundee, UK; [2]School of Computing, University of Dundee, UK; [3]Madras Diabetes Research Foundation, Chennai, India.

1. Abstract

- CNNs are very well suited for image classification and detection.
- As an initial exercise for my PhD on deep-learning discovery of features in retinal images specifically associating to clinical outcomes, some notable CNN architectures are studied.
- ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is an annual competition since 2010 for visual recognition task . It provided around 1.2M/50K/100K images as training/validation/testing data respectively spread across 1000 distinct image categories for competitions.
- Different CNN architectures participated in ILSVRC competition are discussed.

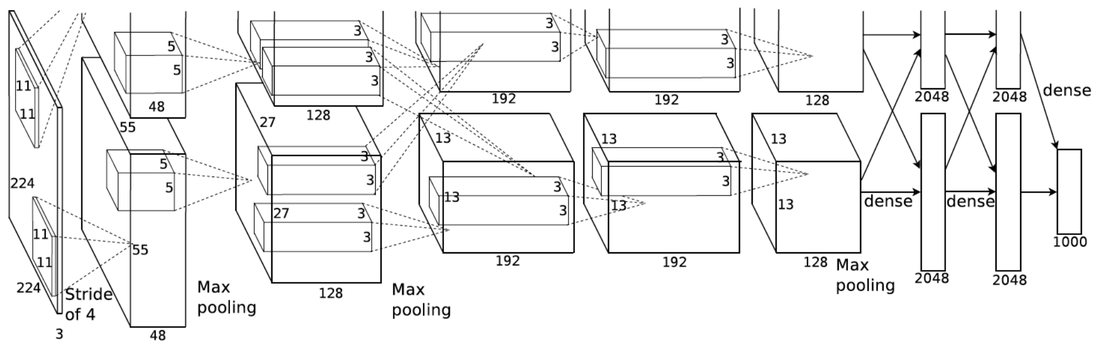
2. LeNet-5

- LeNet-5 is a CNN developed by LeCun et al. in 1998. It became standard structure of CNN.
- It is a 7-layered architecture with 60K trainable parameters.



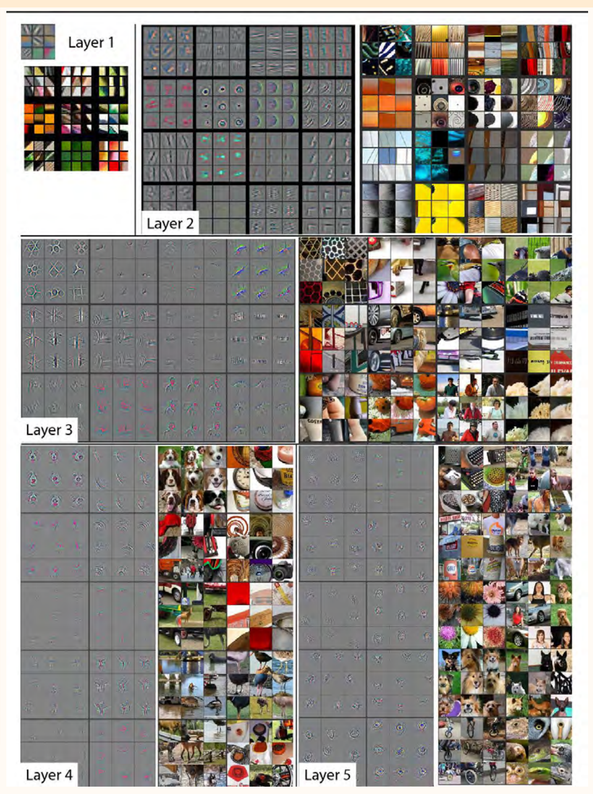
3. AlexNet

- By Krizhevsky et al. & ILSVRC 12 winner.
- Used 2 NVIDIA GPUs.
- 8-layered architecture with 60M parameters.
- Achieved top-5 error rate of 16.4% on test.



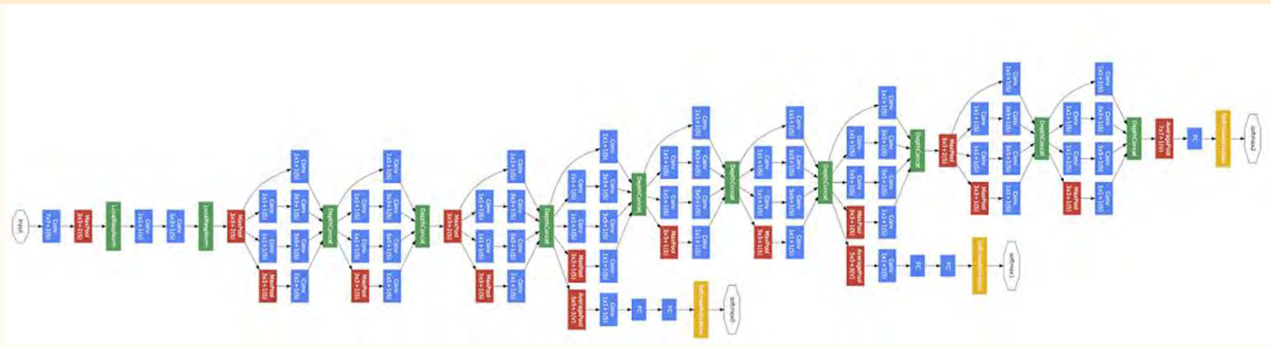
4. ZFNet

- Developed by Zeiler and Fergus in 2013.
- ZFNet architecture is similar to AlexNet, with a few modifications (in kernel size, stride and feature maps).
- Used deconvolution for visualizing and understanding Convolutional Networks.
- Winner of ILSVRC 13 with improved hyper-parameters.
- Achieved top-5 error rate of 11.7% on test data.



5. GoogLeNet

- Developed by Szegedy et al. in 2014 and is the winner of ILSVRC 14.
- Introduced Inception modules with dimensionality reduction using 1x1 convolutions.
- Used only 3 small kernels of size 1x1, 3x3 and 5x5 through out the networks.
- It is 22-layered deep architecture with 4M parameters.
- Achieved top-5 error rate of 6.67% on test data.



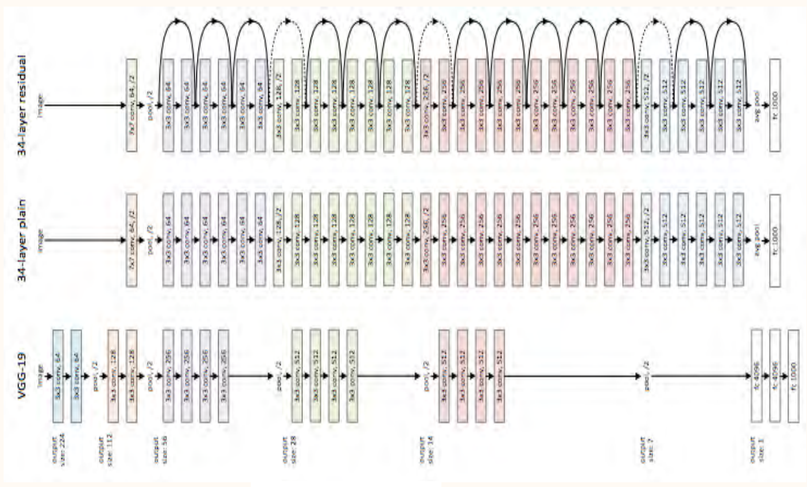
6. VGGNet

- Developed by Simonyan and Zisserman; runners in ILSVRC 14.
- 11 to 19 layered architecture.
- Top-5 error rate of 7.3% on test data.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

7. ResNet

- Developed by He K. et al in 2015; winner in ILSVRC 15.
- Introduced a residual learning framework by shortcuts.
- Trained very deep architecture up to 152 layered.
- Achieved top-5 error rate of 3.37% on test data.



8. Summary

- With the increase in computational power and large training data, CNN are growing deeper & achieving greater accuracy in vision tasks.
- We plan to use deep learning to identify retinal feature associated with diabetic phenotype and genotype, leveraging the GoDARTS bioresource.
- We also plan to use the deconvolution technique of ZFNet authors in visualizing the discovered retinal features.

Acknowledgement

The research was commissioned by the National Institute for Health Research using Official Development Assistance (ODA) funding [INSPIRED 16/136/102].

Disclaimer *: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.*