

(EE-641)/Deep Learning (EE-842)

Lecture 04

Dr S M Wasif

`syed.wasif@uog.edu.pk`

Updated: 2024/11/22

Department of Electrical Engineering,
University of Gujrat



Table of Contents

1. Deeper Convolutional Neural Networks (CNNs)
2. Visual Geometry Group Network (VGGNet)
3. Inception / GoogLeNet
4. Residual Network (ResNet)

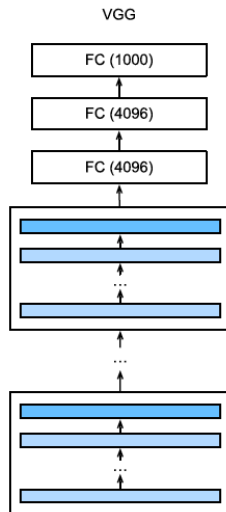
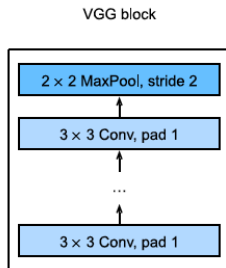
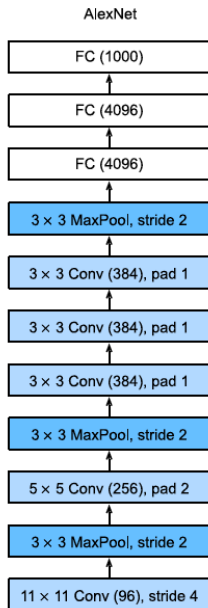
Deeper Convolutional Neural Networks (CNNs)

Visual Geometry Group Network (VGGNet)

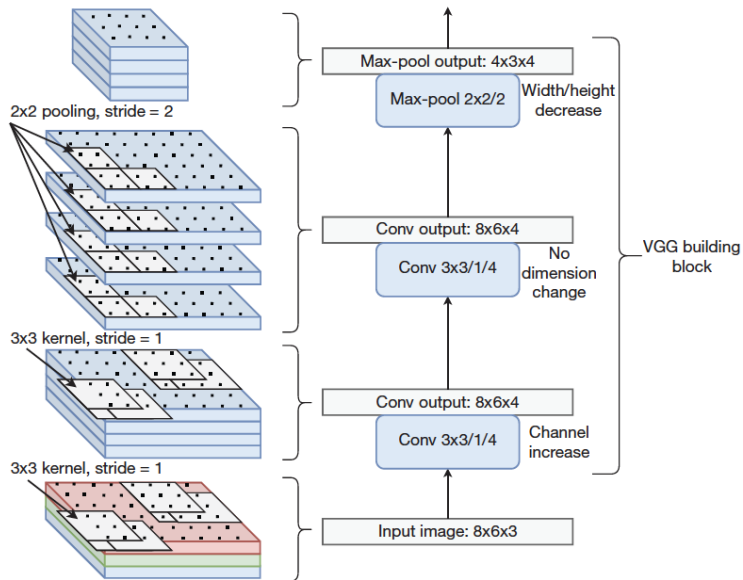
Visual Geometry Group Network (VGGNet)

- ⊙ Problem with a simple CNN is rapid spatial resolution decrease.
- ⊙ Imposes a hard limit of $\log_2 d$ convolution layers on the network of d dimensions.
- ⊙ Depth of VGGNet can be adjusted without adjustment of any other parameter.
- ⊙ VGGNet uses fixed kernel of size 3×3 and stride 1 to maintain width and height dimensions across multiple layers.
- ⊙ VGGNet reduced the filter size but increased depth.
- ⊙ Architectures with greater depth and reduced filter size require fewer parameters.

Visual Geometry Group Network (VGGNet)



Visual Geometry Group Network (VGGNet)



Visual Geometry Group Network (VGGNet)

- ⊙ Increased depth led to greater sensitivity with initialization known as cause to instability.
- ⊙ This problem was solved by using pretraining *i.e.*, a shallow architecture (11-layer subset of the architecture was trained first) was trained and then further layers are added.
- ⊙ VGGNet achieved a top-5 error of only 7.3% (*i.e.*, show in column D) in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) contest.
- ⊙ Terms: **C₃D₆₄** refers to convolution with filter size 3×3 and batch size of 64; max-pooling layer is referred to as **M**; local response normalization is **LRN**; **S** for softmax; **FC₄₀₉₆** refers to a fully connected layer with 4096 units.

Visual Geometry Group Network (VGGNet)

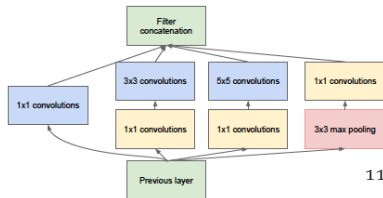
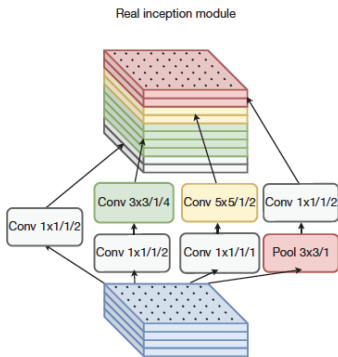
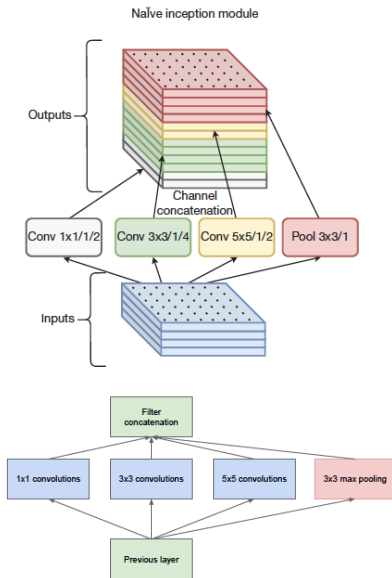
Name:	A	A-LRN	B	C	D	E
# Layers	11	11	13	16	16	19
	C3D64	C3D64	C3D64	C3D64	C3D64	C3D64
		LRN	C3D64	C3D64	C3D64	C3D64
	M	M	M	M	M	M
	C3D128	C3D128	C3D128	C3D128	C3D128	C3D128
			C3D128	C3D128	C3D128	C3D128
	M	M	M	M	M	M
	C3D256	C3D256	C3D256	C3D256	C3D256	C3D256
	C3D256	C3D256	C3D256	C3D256	C3D256	C3D256
				C1D256	C3D256	C3D256
						C3D256
	M	M	M	M	M	M
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
				C1D512	C3D512	C3D512
						C3D512
	M	M	M	M	M	M
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
	C3D512	C3D512	C3D512	C3D512	C3D512	C3D512
				C1D512	C3D512	C3D512
						C3D512
	M	M	M	M	M	M
	FC4096	FC4096	FC4096	FC4096	FC4096	FC4096
	FC4096	FC4096	FC4096	FC4096	FC4096	FC4096
	FC1000	FC1000	FC1000	FC1000	FC1000	FC1000
	S	S	S	S	S	S

Inception / GoogLeNet

Inception / GoogLeNet

- ⊙ GoogLeNet also referred to as Inception architecture as it uses a building block called Inception module.
- ⊙ An Inception architecture is a network within a network.
- ⊙ The Inception module used in GoogLeNet provides the ability to work with multiple receptive field sizes.
- ⊙ Inception architecture can be broken into three parts:
 1. **Stem:** the first two or three convolutions that operate on the image and extract low-level features.
 2. **Body:** perform convolution with multiple-size kernels in parallel, to be able to detect features at different sizes at the same time.
 3. **Head:** maps the features obtained to the required classification, segmentation, detection, or tracking problem.

Inception / GoogLeNet



- ⊙ The Inception use lower weights by reducing number of channels with 1×1 convolution before 3×3 and 5×5 convolutions.
- ⊙ The general idea behind the Inception module is two-fold:
 1. Multi-level feature extractor
 2. Dimensionality reduction
- ⊙ GoogLeNet is made up of nine inception modules that are arranged sequentially.

Inception / GoogLeNet

type	patch size/ stride
convolution	$7 \times 7 / 2$
max pool	$3 \times 3 / 2$
convolution	$3 \times 3 / 1$
max pool	$3 \times 3 / 2$
inception (3a)	
inception (3b)	
max pool	$3 \times 3 / 2$
inception (4a)	
inception (4b)	
inception (4c)	
inception (4d)	
inception (4e)	
max pool	$3 \times 3 / 2$
inception (5a)	
inception (5b)	
avg pool	$7 \times 7 / 1$
dropout (40%)	
linear	
softmax	

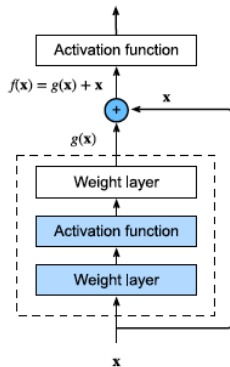
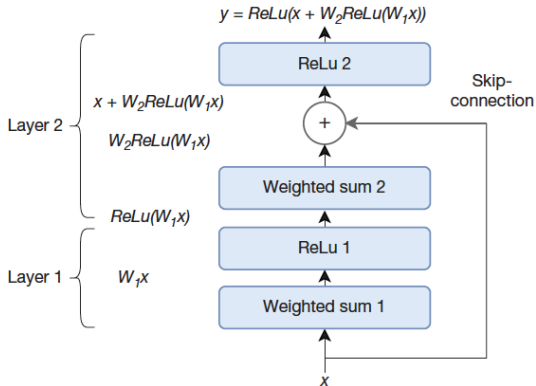


Residual Network (ResNet)

Residual Network (ResNet)

- ⊙ Very deep networks are hard to train – increasing the network depth from 18 to 34 layers, the training error increased.
- ⊙ **Accuracy Degradation Problem:** after some critical depth, the output loses the information of the input, so the correlation between the input and output starts diverging resulting in an increase in inaccuracy.
- ⊙ **ResNet:** a residual module to train Convolutional Neural Networks to depths previously thought impossible.

Residual Network (ResNet)

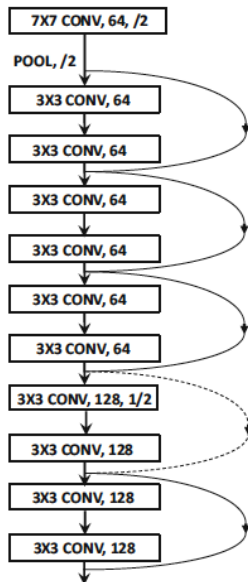


Residual Network (ResNet)



- ⊙ **Identity mapping / Skip connections:** the process of taking the original input to the module and adding it to the output of a series of operations.
- ⊙ The **residual** input is added to the output of a series of layer operations.
- ⊙ The connection between the input and the addition node is called the **shortcut**.
- ⊙ A residual layer attempts to approximate:

$$y = \text{ReLu}(x + W_2 \text{ReLu}(W_1 x))$$

Residual Network (ResNet)



References

-  Zhang A, Lipton ZC, Li M, Smola AJ. .
Dive into Deep Learning
Cambridge University Press; 2023.
-  Goodfellow, I., Yoshua Bengio, and Aaron Courville.
Deep Learning (Adaptive Computation and Machine
Learning Series)
The MIT Press, Cambridge, Massachusetts, London,
England, 2016.

THE
END