Javed et al. LGURJCSIT 2023



LGU Research Journal of Computer Science & IT SSN: 2521-0122 (Online) ISSN: 2519-7991 (Print)

doi: 10.54692/lgurjcsit.2023. 0701361 Vol.7 Issue 1, January – March 2023

Algorithmic as well as Space and Time Comparison of Various Deep Learning Algorithms

Waheed Javed^{1*}, Gulnaz Parveen²

¹Computer Sciences, Pakistan institute of Applied and Social Sciences Kasur, Pakistan ²Computer Sciences, University of Management & Technology, Pakistan

Email: waheedjaved146@gmail.com

ABSTRACT:

Deep Learning is an artificial intelligence subfield within machine learning. Deep Learning has been used in various applications like computer vision, natural language processing, speech recognition, social network filtering, neural machine translation, etc. Deep Learning, Convolutional Neural Network (CNN) is a set of deep neural networks mainly designed for image analysis. Deep Learning strong ability is mainly due to multiple feature extraction. This paper will discuss and compare AlexNet, VGGNet-16, and Residual Networks (ResNet 50,101,152).

KEYWORDS: Deep Learning, CNN, AlexNet, ResNet, VGG.

1. INTRODUCTION

Deep Learning is a sub-classification of AI that has given machines knowledge. Deep Learning has recently been popular for applications like pattern recognition, natural language processing, object recognition, and otherimage processing. However, it is mostly utilized in Object Recognition activities such as Driver Assistance Systems, Autonomous Driving Systems, and Real-World Target Detection. [1]. Deep Learning may be done with a variety of Neural Network Architectures in general [2]. Feature extraction and processing are the primary functions of these networks. Deep Learning uses four different types of neural networks. Auto-encoders are made up of Generative Adversarial Networks (GANs), Unsupervised Pre-trained Networks Convolutional Neural (UPNs), Networks (CNNs), Recursive Neural Networks, Recurrent Neural Networks, and Unsupervised Pre-trained Networks.[3]

Convolutional Neural Networks (CNNs) are among the finest learning algorithms for interpreting image information, with outstanding results in image segmentation, classification, detection, and retrieval. [4]. The convolutional neural network is the most frequently used deep learning framework. It has gained applications in object tracking, posture estimation, text detection and identification, activity recognition, scene labelling, and other fields.

We shall explore several CNN current architectures that are utilized in Deep Learning in this article. The construction of CNN, several CNN models, and their comparison are all covered in this paper.

2. CONVOLUTIONAL NEURAL NETWORK

CNN's are essentially the same as regular neural organizations, which may be thought of as a collection of neurons arranged in a non-cyclic pattern. A mystery layer neuron is connected to a subgroup of neurons in the back layer, which is a big capability from a neural organization.

It can learn to incorporate favorably as a result of this poor arrangement. The extraction of significant neural organizations in various evened out includes, for example, the pre-arranged channels of the first layer can set edges or concealing masses, the second layer as specific shapes, the following layer channels learn object parts, and the last layerscan recognize the objects. [5]

2.1. Structure of CNN

We learned that CNNs are mostly employed for image recognition and organization. It may also be used to recognizefaces, objects, and traffic signs [3]. Figure 1 depicts CNN's strategy.



Figure 1: Structure of CNN

Feature extractors and a classifier were two crucial components of the CNN setup. Each layer of the association obtains the yield from its previous layer and sends it as the commitment to the subsequent layer in the segment extraction layers. Convolution, max-pooling, and grouping are the three types of layers used in the CNN setup. Convolutional layers and max-pooling layers are the two layers in the association's low and focus level.

Convolutions performed are on the even-numbered layers, whereas max-pooling activities are performed on the odd- numbered layers. The convolution and max-pooling layers' yield center points are collected into a 2D plane called feature arranging. Each layer's plane is derived from a mixture of about one plane from previous levels. The centers of a planeare linked to a small portion of each previous layer's associated planes. By using the convolution method on the information nodes, each center of the convolution layer eliminates the arrangements from the data images. Features produced by lower-level layers are used to create higher-level components. The portions of arrangements are reduced depending on the size of the piece for the convolutional and max-pooling assignments independently as the components stimulate to the most significant layer or level.

In any case, the number of component maps is frequently increased to achieve better data picture arrangements and ensure request accuracy. The output of the CNN's last layer is utilized as a promise to a completely unrelated organization known as the plan layer. The depiction layer is made up of feed-forward neural connections, which have a superior execution.

The isolated arrangements are regarded as commitments in the course of action layer for the component of the weight matrix of the last neural association. Regardless, the completely linked layers are costly to the degree as they are limited by association or learning.

Nowadays, there are a couple of new techniques, including typical pooling and overall ordinary pooling that is used as analternative to totally related organizations. The score of the species is not settled in the top request layer using a fragile max layer. Considering the best score, the classifier gives a yield for looking at classes[6]. Mathematical nuances on different layers of CNNs are discussed underneath.

2.2. Convolutional Layer

To extract components from the information image, a convolution layer is used. By learning picture highlights using small squares of information data, convolution saves the spatial relationship between pixels. The image might be referred to as a network.

2.3. Pooling Layer

The CNN association has Conv layers and pooling layers, as well as the ability to limit the spatial component of order maps (without sacrificing information) and the number of limitations in the net, therefore decreasing the computational complexity in everyday use. This effectively eliminates the problem of over-fitting. Max pooling, ordinary pooling, stochastic pooling, ghost pooling, spatial pyramid pooling, and multi-scale demand less pooling are by far the most common pooling exercises.

2.4. Fully Connected Layer

Completely associated layer is generally utilized toward the finish of the organization for characterization. As pooling and convolution, it is a worldwide activity. The score of each class is calculated using the discarded configurations from a convolutional layer as the process advances. It necessitates dedication from the feature extraction phases as well as the layer. As a sensitive max depiction, the completely associated feed-forward neural layers are employed.

2.5. Activation Function

Enactment capacities are non-direct capacities that take on a solitary number and do some numerical procedure on it. Enactment work goes about as a choice capacity and helps in learning of examples. The great determination of an actuation func-tion can speed up the learning cycle. Many similar capabilities exist, but the most commonly used are sigmoid, tanh, maxout, Wash, ReLU, and variants of ReLU, such as cracked ReLU, ELU, and PReLU, which are used for non-straight component mixes. . ReLU and its variations are given need as they help in conquering the inclination issue.

3. CNN ARCHITECTURE

In this part, we will examine different CNN algorithms and their designs. CNN's are developed from time to time to workon their presentation. Each design has a different number of layers, several convolution channels, and unpredictability.

3.1. Alexn

AlexNet was presented by Krizhevesky et al. [7] that drawn out the learning limit of the CNN by making it more profound and by carrying out various boundary enhancements strate-gies. first time this organization demonstrated that the provisions got by learning can go pastagainst all the customary AI, Breaks past PC vision draws near. Picture distinguishing proof and grouping activities, which is the stage in history where contribution in profound learning has immediately grown up. The design of AlexNetis displayed in Figure 2.



Figure 2: The architecture of AlexNet

The essential format of AlexNet design showing its five convolutions and three completely associated layers. Convolution, max-pooling, Nearby Reaction Standardization (LRN), and completely associated (FC)layer in it.

The first convolutional layer uses Nearby Reaction Standardization (LRN) to conduct convolution and max- pooling using 96 relative channels that are 11 11 in size. The most effective pooling exercises are done with three channels and a stage size of two. In the second layer, with 5 5 channels, comparable workouts are performed. With 384, 384, and 296 component maps, 3 3 directives are utilized solely in the third, fourth, and fifth convolutional layers.

With dropout, two fully associated (FC) layers are used, followed by a SoftMax layer at the end. For this model, two organizations with similar architecture and a similar number of component maps are produced in the same way. This organization is the first to introduce two concepts: Local Response Normalization (LRN) and dropout. LRN may be used in two different ways: first, on single channel or component maps, where a NN fix is picked from a comparable element map and standardized based on the local characteristics; and second, on multichannel or component maps. Second, LRN may be used on several channels or element maps at the same time (neighborhood along with the third measurement however a solitary pixel or area).

3.2. VGGNET

In the realm of Deep Learning, VGGNet[8] is the most often used design. VGGNET won the ILSVRC 2014 challenge and was given the moniker VGGNet by the network. Karen Simonyan and Andrew Zisserman presented VGGNet, which looked at how a small number of channelsmay work on an exhibition.



Figure 3: The architecture of VGGNet

VGG supplanted the 11x11 and 5x5 channels with 3x3 channels layer and clarified that (3x3) little

size channels candecrease the impact of the (5x5 and 7x7)large size channel. VGGNet comprises 16 and 19 convolutional layers profound contrasted with AlexNet to run the connection of VGG16 is a convolution neural net (CNN) design that was used to win the 2014 ILSVR(Imagenet) competition. It is regarded as one of the most stunning vision models plans ever developed. VGG16 is a convolution neural net (CNN) arrangement that won the ILSVR(Imagenet) competition in 2014. It is regarded as one of the most outstanding vision model arrangements to date.

The VGG network aids in identifying articles that are subject to produce probabilities of various classifications in which an image could appear. VGG-16 takes a 224 244 3 picture as a piece of information and transmits just 3 3 convolutions and 2 2 pooling throughout the whole association design [9].

Figure 5 depicts the schematic orchestration of VGG-16. Throughout the whole arrangement, it sticks to this convolution and max pool layer shape. It features two FC (completely related layers) at the end, followed by a softmaxfor yield. The number 16 in VGG16 denotes that there are 16 layers with loads. This organization is very massive, witha limit of roughly 138 million (approximately).

3.3. ResNet

In 2015, Kaiming et al. presented ResNet. The 2015- ILSVRC competition was won by ResNet's 152-layer substantial CNN. By displaying the leftover learning in CNNs and devising an effective approach for the preparation profound organizations,

ResNet revolutionized the CNN compositional notion. It has lingering learning system, which allows the layers to acquire remaining capacities based on the information sources they have access to, rather than learning unreferenced capacities. In comparison to AlexNet and VGG, the residual network was 20 to 100 times more advanced and had less computing complexity than previously proposed networks.

On the picture 34 layers representation job, ResNet with 50/101/152 layers had fewer errors. ResNet on picture acknowledgment and confinement undertakings has demonstrated great execution and that illustrative profundityis for the most part centered for some, visual acknowledgment errands The main disadvantage of theresidual network is that due to a large number of parameters, it is much expensive to evaluate. The number of parameters can be minimized by finishing the first Fully-Connected layer.

4. Experiments and Result 4.1. Dataset

For the purpose of comparison, we have taken data set from' CelebA dataset' source.for gender classification. CelebA contains 218*178, we resize images to 256*256. Target classes to classify images are male and female. SourcelinkCelebA:http://mmlab.ie.cuhk.edu.hk/ projects/ CelebA.html.

4.2. Implementation Results of various architectures on the CelebA face dataset to train and test gender classifier

4.2.1. Implementation Results of VGG16

epoen:	000/000	L	Darcu	1900/2044	1	COSL:	0.11
Epoch:	003/003	L	Batch	1950/2544		Cost:	0.02
Epoch:	003/003	l	Batch	2000/2544	I.	Cost:	0.20
Epoch:	003/003	L	Batch	2050/2544	I.	Cost:	0.02
Epoch:	003/003	L	Batch	2100/2544	1	Cost:	0.06
Epoch:	003/003	Ĺ	Batch	2150/2544	Í.	Cost:	0.08
Epoch:	003/003	Ĺ	Batch	2200/2544	Í.	Cost:	0.23
Epoch:	003/003	Ĺ	Batch	2250/2544	Í.	Cost:	0.09
Epoch:	003/003	l	Batch	2300/2544	I.	Cost:	0.12
Epoch:	003/003	1	Batch	2350/2544		Cost:	0.05
Epoch:	003/003	Ĺ	Batch	2400/2544	Í.	Cost:	0.13
Epoch:	003/003	Ĺ	Batch	2450/2544	Í.	Cost:	0.24
Epoch:	003/003	İ.	Batch	2500/2544	1	Cost:	0.16
Epoch:	003/003	L	Train:	95.849%	1	/alid:	96.6
Time e	lapsed: 5	4.	.64 mir	1			
Total 1	Training	Тi	ime: 54	1.64 min			

Figure 5: Training Time of VGG16

20 40 60 80 100 120 20 40 60 80 100 120 Probability Female 90.72% Test accuracy: 95.48% Figure 6: Accuracy Result of VGG16

```
4.2.2. Implementation Results of Resnet50

Epoch: 010/010 | Batch 0000/0318 | Cost: 0.0238

Epoch: 010/010 | Batch 0050/0318 | Cost: 0.0251

Epoch: 010/010 | Batch 0100/0318 | Cost: 0.0144

Epoch: 010/010 | Batch 0150/0318 | Cost: 0.0133

Epoch: 010/010 | Batch 0250/0318 | Cost: 0.0441

Epoch: 010/010 | Batch 0250/0318 | Cost: 0.0358

Epoch: 010/010 | Batch 0300/0318 | Cost: 0.0277

Epoch: 010/010 | Train: 99.374% | Valid: 97.966%

Time elapsed: 37.70 min

Total Training Time: 37.70 min
```

Figure 7: Training Time of Resnet50



Figure 8: Accuracy Result of Resnet50

4.2.3. Implementation R	Results of Resnet101
-------------------------	----------------------

Epoch:	010/010	Batch	0400/1272	Cost	t: 0.0135
Epoch:	010/010	Batch	0450/1272	Cost	t: 0.0488
Epoch:	010/010	Batch	0500/1272	Cost	t: 0.007S
Epoch:	010/010	Batch	0550/1272	Cost	t: 0.0217
Epoch:	010/010	Batch	0600/1272	Cost	t: 0.0188
Epoch:	010/010	Batch	0650/1272	Cost	t: 0.0215
Epoch:	010/010	Batch	0700/1272	Cost	t: 0.0211
Epoch:	010/010	Batch	0750/1272	Cost	t: 0.0182
Epoch:	010/010	Batch	0800/1272	Cost	t: 0.0526
Epoch:	010/010	Batch	0850/1272	Cost	t: 0.0395
Epoch:	010/010	Batch	0900/1272	Cost	t: 0.0396
Epoch:	010/010	Batch	0950/1272	Cost	t: 0.0294
Epoch:	010/010	Batch	1000/1272	Cost	t: 0.0291
Epoch:	010/010	Batch	1050/1272	Cost	t: 0.0118
Epoch:	010/010	Batch	1100/1272	Cost	t: 0.0166
Epoch:	010/010	Batch	1150/1272	Cost	t: 0.0559
Epoch:	010/010	Batch	1200/1272	Cost	t: 0.0168
Epoch:	010/010	Batch	1250/1272	Cost	t: 0.0127
Epoch:	010/010	Train	: 99.183%	Valid	1: 98.017%
Time e	lapsed: 1	21.09 m	in		
Total '	Training '	Time: 1	21.09 min		

Figure 9: Training Time of Resnet101



Figure 10: Accuracy Result of Resnet101

4.2.4. Implementation Results of Resnet152

Epoch:	010/010	Batch	0550/1272	Cost:	0.0327
Epoch:	010/010	Batch	0600/1272	Cost:	0.0302
Epoch:	010/010	Batch	0650/1272	Cost:	0.0150
Epoch:	010/010	Batch	0700/1272	Cost:	0.0185
Epoch:	010/010	Batch	0750/1272	Cost:	0.0202
Epoch:	010/010	Batch	0800/1272	Cost:	0.0233
Epoch:	010/010	Batch	0850/1272	Cost:	0.0087
Epoch:	010/010	Batch	0900/1272	Cost:	0.0147
Epoch:	010/010	Batch	0950/1272	Cost:	0.0120
Epoch:	010/010	Batch	1000/1272	Cost:	0.0200
Epoch:	010/010	Batch	1050/1272	Cost:	0.0427
Epoch:	010/010	Batch	1100/1272	Cost:	0.0141
Epoch:	010/010	Batch	1150/1272	Cost:	0.0519
Epoch:	010/010	Batch	1200/1272	Cost:	0.0281
Epoch:	010/010	Batch	1250/1272	Cost:	0.0487
Epoch:	010/010	Train	99.060%	Valid:	98.062%
Time e	lapsed: 1	78.71 m:	in		
Total '	Training '	Fime: 1	78.71 min		

Figure 11: Training Time of Resnet152



Test Accuracy: 92.72 % Probability Female: 99.99%

Figure 12: Accuracy Result of Resnet152

4.3. Computational Parameters of Various Deep CNN Algorithms

The examination of newly proposed models dependent on network boundaries, and greatest number of associations are given in Table 1.

4.4. Comparison of Algorithms on the basis of Time, Probability and Accuracy

The comparison of VGG16, Resnet50, Resnet101, Resnet152 models based on time, accuracy and probability are given in Table 2.

Table 1: Computational Parameters for VGG-16 RESNET-50 RESNET101 RESNET152 Deep CNN Models

Methods	VGG- 16	ResNet- 50	ResNet- 101	ResNet- 152
Input size	218x178	218x178	218x178	218x178
Learning rate	0.001	0.001	0.001	0.001
Batch size	128	128	128	128
Epochs	10	10	10	10
Number of Conv Layers	16	50	101	152
Filter Size	3	1,3,7	1,3,7	1,3,7
Stride	1	1,2	1,2	1,2
Number of Weights	14.7Ms	23.5M	44.6M	60.2M

Table 2: Comparson of for VGG-16 RESNET-50 RESNET101 RESNET152

Comparison	VGG -16	Resnet - 50	Resnet - 101	Resnet - 152
Time(min)	54.64	37.70	121.09	178.71
Accuracy	95.48%	97.40%	97.52%	97.72
Probability	90.72%	100%	99.99%	99.99

5. CONCLUSION

In this paper, we have talked about various sorts of convolutional neural networks. Various sorts are Alexnet, VGG-16, ResNet-50, ResNet-101, ResNet-152. Development of various sorts of CNN models is examined. We have implemented VGG-16, ResNet- 50, ResNet-101, and ResNet-152 and played out their examination based onprecision, time, and probability of gender classification results. From the Comparison table. it tends to be said that Convolution Neu-ral Network changes yield dependent on layers determination. The usefulness worked with numerous stages for highlights extraction and grouping.ResNet-152 gives better accuracy while selecting 152-layers as compared to VGG- 16, ResNet-50, ResNet-101 CNN features but its computational time is higher than as compared to other models.

REFERENCES

[1] Srinivas SS Kruthiventi, Kumar Ayush, and R Venkatesh Babu. "Deepfix: A fully convolutional neural networkfor predicting human eye fixations". In: IEEE Transactions on Image Processing 26.9, pp. 4446–4456, 2017.

[2] Zhang Bo and Zhang Ling. "The analysis and improve- ment of artificial neural network models". In: 1997 IEEE International Conference on Intelligent Processing Sys- tems (Cat. No. 97TH8335). Vol. 1. IEEE, pp. 7–13, 1997.

[3] R HARIDAS and RL Jyothi. "Convolutional neural networks: A comprehensive survey". In: International Journal of Applied Engineering Research 14.3, pp. 780–789, 2019.

[4] Asifullah Khan et al. "A survey of the recent architec- tures of deep convolutional neural networks". In: arXiv preprint arXiv:1901.06032, 2019.

[5] Neena Aloysius and M Geetha. "A review on deep convolutional neural networks". In: 2017 International Conference on Communication and Signal Processing (ICCSP). IEEE, pp. 0588–0592, 2017.

[6] Md Zahangir Alom et al. "A state-of-the-art survey on deep learning theory and architectures". In: Electronics 8.3, p. 292, 2019.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: Advances in neural information process- ing systems, pp. 1097–1105, 2012.

[8] Javed, W., Danish, M., Tahir, H., & Bashir, H. (2021, November). Vehicle Tracking System and Speed Estimation Utilizing Optimal Technique. In 2021 International Conference on Innovative Computing (ICIC) (pp. 1-6). IEEE.

[9] Jaderberg, M., Simonyan, K., & Zisserman, A. Spatial transformer networks. Advances in neural information processing systems, 28, 2017-2025, 2015.

[10] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recogni- tion". In: arXiv preprint arXiv:1409.1556, 2014.

[11] http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html