SSRG International Journal of Electronics and Communication Engineering Volume 6 Issue 8, 18-21, August 2019
ISSN: 2348 – 8549 /doi:10.14445/23488549/IJECE-V6I8P103 © 2019 Seventh Sense Research Group®

Comparative Study of Convolutional Neural Networks

Sagnik M[#1], Ramapraasad P[#2]

[#]Department of Electronics & Communication Engineering, School of Engineering & IT, Manipal Academy of Higher Education, Dubai Campus, Dubai, UAE-345050

Abstract**—***Artificial Neural Networks have changed the way any computer works. Starting from machine learning, we have now come a long way where detection of faces, data representations. have now come up to a stage that an observer can exclaim that the machine is "perfect". However, a field where we still have research going on and is becoming efficient daily is mobile phones. Researchers have targeted convolutional neural networks (CNN), a branch of deep learning, for building neural network models that can execute on mobile phones. An emerging player in this field is the MnasNet, which was developed by Google's brain team.*

In this review paper, details about MnasNet's facilities shall be discussed, it is advantages, and also how it fairs against various other CNN models by Google, Apple. Furthermore lastly, where next to the researchers will be looking into and the areas where the MnasNet might have to improve.

Index Terms*—Artificial Neural Network, Convolutional Neural Network, Machine Learning,*

I. INTRODUCTION

A. *Machine Learning*

MachineLearning is a data analysis method based on the automation of analytical model building. This AI branch was created with the idea that systems can understand data and patterns without much human intervention [2]. The difference with Machine learning is that the objective is to understand the data structure, compared to data mining, where the emphasis was on identifying unknown patterns from data through statistical algorithms, text analytics, and various other areas of analytics [2]. Even if there is no concrete knowledge of structure, Machine learning allows computers to probe the data.

B. *Artificial Neural Networks*

One of the most important tools used in Machine Learning, Artificial Neural Networks (ANNs) is human-brain-inspired systems, replicating the way humans learn [3]. Fig 1 represents a simple ANN with two hidden layers. Neural Networks use input and output layers, often including hidden layers that convert the inputs into something the output can use. Considering the raw data as input, it is passed down a conveyor belt where different layers work on particular features of the input. Suppose the objective is to recognize an object, the first layer analyzes the brightness of the pixel; the second layer may identify edges of the image, the third layer may recognize textures, and so on [3]. After going through each of these hidden layers, the network forms complex feature detectors, which helps figure certain elements that appear all the time in any input (Fig. 1) [3].

.

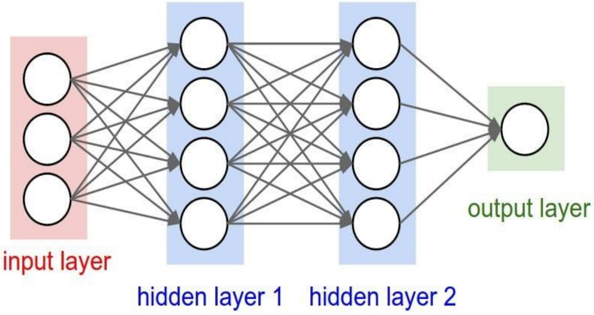

Fig. 1 A representation of how Artificial neural network works [3]

C. *Deep Learning*

Deep Learning is a subfield of machine learning. A major advantage that deep learning provides is its ability to perform automatic feature extraction from raw data [4]. Fig 2 shows that the performance of deep learning algorithms is better than other algorithms. The area of expertise in deep learning is that it excels in problems when the inputs and outputs are analog, i.e., they are not quantities in tabular columns but instead are images of pixel data, documents of text data, or files of audio data. Deep learning can be supervised, semi-supervised or unsupervised.

Deep Learning can also be considered as the stepping stones for very large Convolutional Neural Networks, which have great success on object recognition in photographs [4].

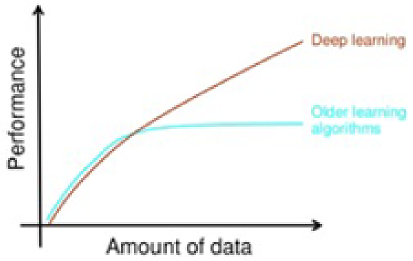

Fig. 2 Performance v/s data graph for data learning [4]

D. Convolutional Neural Networks

Convolutional Neural Networks (CNN) are similar to Artificial Neural Networks, except that they assume the input is an image. This helps to make the forward function more efficient to implement and reduces the number of parameters in the network [5]. Because of this particular assumption, CNN arranges neurons in a 3D structure in height, length, and width, as shown in Fig 3. CNN has three main layers: Convolutional Layer, Pooling Layer, and Fully-connected layer. Stacking these layers creates the Convolutional Network Architecture [5]. In particular, the convolutional and Fully-connected layers perform transformations that are functions of not only the activations in the input volume but also of the parameters (the weights and biases of the neurons), while the Pool layer implements a fixed layer [5].

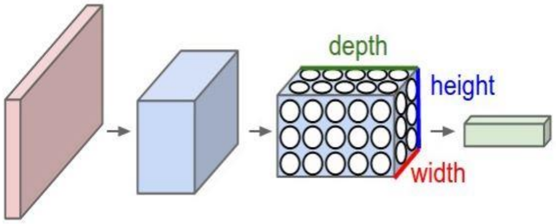

Fig. 3 Representation of CNN's working [5]

To improve CNN model efficiency, some common approaches have been 1) Quantization of the baseline into lower bit representations or 2) reducing less important filters [6]. Nevertheless, these approaches do not focus on CNN operations but instead focus on baseline models.

II. PREVIOUSLY MADE MOBILE ARCHITECTURES

E. MobileNet V1 and V2

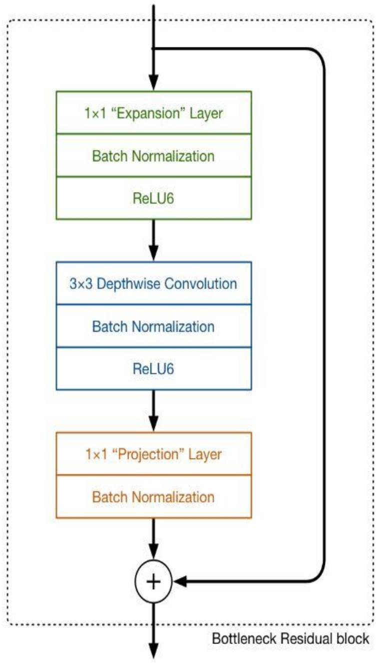

Fig. 4 Block diagram representing MobileNet V2 [7]

In Mobilenet V1 architecture, it was proposed that the convolution layer could be replaced by depthwise separable convolutions [7]. In V1, the job is split in 2: depthwise layer filters the input followed by another filter (pointwise) that combines these filtered values. Instead of having pooling layers, some of the depthwise layers reduce the spatial dimensions of the data. Mobilenet uses ReLu6 layers, which prevent activations from becoming too big [7](ReLu: Rectified Linear unit, popular activation function in neural networks).

Fig 4 shows the block diagram of MobileNetV2 architecture. In V2, there is an additional layer added to the earlier layers. This new layer makes the number of channels smaller, contrary to V1, which kept the same no. of channels. This layer, known as the 'projection layer,' projects the data consisting of higher dimensions into a tensor with a much lower no. of dimensions. This layer is also called the bottleneck layer, as it reduces the amount of data flowing through the network [7].

F. NASNet Architecture

NAS stands for Neural Architecture Search. As the name suggests, NASNet searches for the best architecture [8]. The NAS algorithm uses Recurrent

Neural Network (RNN) controller, which samples "building blocks" to form an end-to-end architecture. Some of these building blocks are convolution layers and pool layers of various dimensions. A simplified way of looking at NAS is given in Fig5 [8].

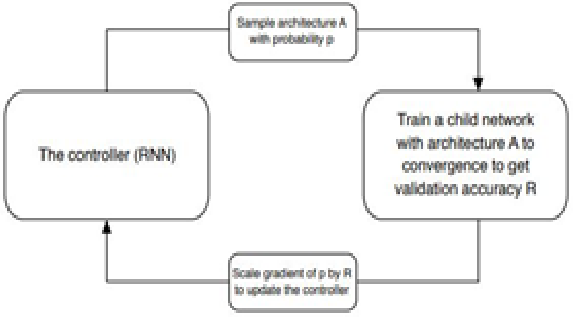

Fig. 5 Block diagram of NASNet [8]

There are two reasons why NAS has had success. One, it consists of many assumptions. Architecture is trained and tested on smaller datasets. However, if a similarly structured dataset is used on larger ones, then it should work as fast as it is in the case of smaller data.

Secondly, NAS has kept the search space limited. This is because of building architecture similar to current states [8]. Though NASNet is efficient, it is unavailable to users outside Google.

G. *ResNet Architecture*

ResNet was based on the idea of 'identity shortcut connection,' which skips one or two layers. A residual is an error present in the result [14]. Imagine a sack consisting of 8 balls, and you have been asked to identify how many balls may be present. If the person says 6, then the person is off by 2. So, the residual is 2. Residual is what should be added to match the actual [14]. Authors of ResNEt depicted that stacking layers does not degrade the architecture and keeps the performance level at par [9]. ResNet claims that rather than letting stacked layers fit the underlying mapping, it is better to fit them in a residual mapping. Fig 6 shows the block diagram of ResNet architecture.

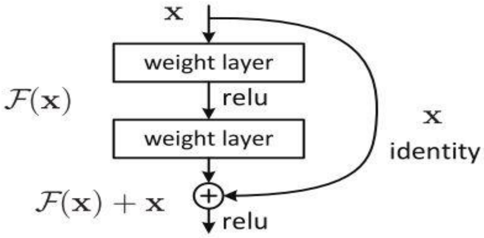
x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Fig. 6 Block diagram of ResNet Architecture [9]

Before ResNet, this idea was implemented in highway networks and also Long Term Short Memory cell. However, none of these were as efficient as it was required to be.

III. MNASNET ARCHITECTURE

Block diagram of the MnasNet model is shown in Fig 7. MnasNet is an AutoML (Automated Machine learning) inspired approach to CNN architecture. It depicts an automatic neural search in designing mobile models using reinforcement learning [12]. Reinforcement learning, a machine learning technique, allows trial and error using feedbacks [13]. MnasNet introduces a multi-objective neural architecture search approach that emphasizes searching for high accuracy CNN models with low real-world inference latency [6]. This approach, named Factorized Hierarchical Search Space, partitions CNN layers into groups and search its connections. This has a specific advantage of balancing the layers' diversity and reduces the size of the search space. Consider a network is partitioned into 'B' blocks, and each has search space of 'S' with 'N' average layers per block. Then total search space will be 'S^B', and per layer, it will be 'S^(B*N)'.

Compared to previous models that mostly employ 2 3x3 kernels in depthwise convolution layers, MnasNet employs single 5x5 kernels. The developer of MnasNet claims that 5x5 kernels have fewer add- ons than 3x3 kernels, considering input depth is less than 7.

There are three components in the MnasNet architecture: 1) an RNN-based controller used for learning and sampling model architectures, 2) a TensorFlow engaged inference engine for measuring the model speed on real mobile phones, and 3) a trainer optimized for accuracy of the models [10].

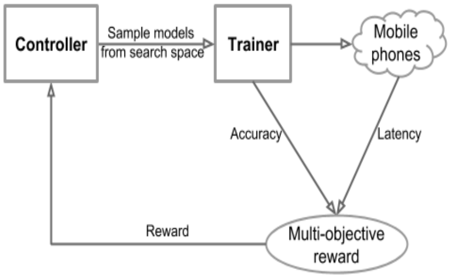

Fig. 7 Block diagram of MnasNet model [6][10]

According to the Google team results that developed MnasNet, it was found that MnasNet search is 1.5 times faster than MobileNet V2 and 2.4 times faster than NASNet [6]. Apart from this, the MnasNet had 19 times lesser parameters and 10 times fewer add-on operations, yet providing the same level of accuracy.

The Google team that created MnasNet also tested for

COCO Object Detection. COCO (Common Objects in Context) is a large-scale object detection, segmentation, and captioning dataset [11]. The team trained their MnasNet models on COCO trainval35k and evaluated them on test-dev2017. It was found out, MnasNet improves both the inference latency and the mAP (mean Average Precision) quality (COCO challenge metrics) over MobileNet V1 and V2. This fared much better than larger MnasNet models like MnasNet- 92, which enhanced the mAP quality even further and had 35 times fewer multiply-add computations [6].

IV. CONCLUSION

The MnasNet is a brilliant automated approach to solving complex mobile vision tasks with ease. With a lesser number of parameters and add-on operations, the architecture seems less complex and adjustable. For the future, increasing the search space optimizations and incorporating further operations is the team's goal and make it more user friendly and applicable to mobile vision tasks like semantic segmentation [12]. Focusing on elaborating this research and making it a more genuine approach to ease CNN models is also required.

REFERENCES

[1] SWNS: “Americans check their phones 80 times a day: Study”, New York Post, November 8, 2017. Retrieved from: [nypost.com/2017/11/08/americans-check-their-phones-80-times-a-day-study/]

[2] Hui Li, “Machine Learning – What it is and why it matters”, SAS- The power to know, 2018. Retrieved from: [sas.com/en_ae/insights/analytics/machine-learning.html]

[3] Luke Dormehl, “What is an artificial neural network?”, Digital Trends, September 13, 2018. Retrieved from: [digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/]

[4] Jason Brownlee, “What is Deep Learning”, Machine Learning Mastery, August 16, 2016. Retrieved from: [machinelearningmastery.com/what-is-deep-learning/]

[5] Karpathy, “Convolutional Neural Networks for Visual Recognition”, Github- Stanford edu. Retrieved from: [cs231n.github.io/convolutional-networks/]

[6] MingXing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, “MnasNet: Platform- Aware Neural Architecture Search for Mobile”, July 31, 2018.

[7] MatthjisHollemans, “MobileNet version 2”, Machinethink, April 22, 2018. Retrieved from: [machinethink.net/blog/mobilenet-v2/]

[8] George Seif, “Everything you need to know about AutoML and Neural Architecture Search”, Towards Data Science, August 21, 2018. Retrieved from: [towardsdatascience.com/everything-you-need-to-know-about-automl-and-neural-architecture-search-8db1863682bf]

[9] Vincent Fung, “An overview of ResNet and its Variants”, Towards Data Science, July 15, 2017. Retrieved from: [towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035]

[10] Jesus Rodriguez, “What is New in Deep Learning Research: Mobile Deep Learning with Google MnasNet”, Towards Data Science, August 13, 2018. Retrieved from: [towardsdatascience.com/whats-new-in-deep-learning-research-mobile-deep-learning-with-google-MnasNet-cf9844d30ae8]

[11] Tsung-Yi Lin, Michael Maire, Serge Belongie, LubomirBourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, PiotrDollár,“Microsoft COCO: Common Objects in Context”, May 1, 2014.

[12] Bo Chen, Quoc V. Le, Ruoming Pang, and Vijay Vasudevan, “MnasNet: Towards Automating the Design of Mobile Machine Learning Models”, August 08, 2018.

[13] Shweta Bhatt, “5 Things You Need to Know about Reinforcement Learning”, Kdnuggets, March 2018. Retrieved from: [kdnuggets.com/2018/03/5-things-reinforcement-learning.html]

[14] Anand Saha, “Decoding the ResNet architecture”, November 02, 2017.

[15] R. Vasudevan, "Neural Networks and Web Mining" SSRG International Journal of Electronics and Communication Engineering 1.1 (2014): 9-14.