



NEUROSCIENCE-INSPIRED ARTIFICIAL VISION FEATURE PARALLELISM AND DEEP LEARNING MODELS, A COMPARATIVE STUDY II DEPTH



 **Marwa Yousif Hassan¹⁺**
Othman O. Khalifa²
 **Aisha Hassan Abdalla³**

^{1,2,3}Department of Electrical and Computer Engineering, IIUM Kuala Lumpur, Malaysia.
¹Email: marwayuosif@gmail.com



(+ Corresponding author)

ABSTRACT

Article History

Received: 11 June 2019

Revised: 17 July 2019

Accepted: 23 August 2019

Published: 9 October 2019

Keywords

Feature Parallelism Model
Deep learning
Machine Learning
Neuroscience
Cortical Column
Laminar Organization
Computer vision
Visual system
Depth analysis
FPM.

This study originates a new model, the Feature Parallelism Model (FPM), and compares it to deep learning models along depth, which is the number of layers that comprises a machine learning model. It is the number of layers in the horizontal axis, in the case of FPM. We found that only 6 layers optimize FPM's performance. FPM has been inspired by the human brain and follows some organizing principles that underlie the human visual system. We review here the standard practice in deep learning, which is opting in to the deepest model that the computational resources allow up to hundreds of layers, seeking better accuracies. We have implemented FPM using 5, 6, 7, and 8 layers and observed accuracy as well as training time for each. We show that much less depth is needed for FPM, down to 6 layers. This optimizes both accuracy and training time for the model. Moreover, in a previous study we have proposed the model and have shown that while FPM uses less computational resources proved by 21% reduction in training time, it performs as well as deep learning regarding models' accuracy.

Contribution/ Originality: This study originates a new model, the Feature Parallelism Model (FPM), and compares it to deep learning models along depth, which is the number of layers that comprises a machine learning model.

1. INTRODUCTION

Nature has continuously inspired inventions throughout history. Deep learning is one field that has been inspired by neuroscience and the brain. Deep learning systems have led to great performances of artificial vision [1], [2]. However, the well-known neuroscience theories have not yet been utilized by artificial vision, not to mention the undiscovered ones.

Nowadays the go to technology both in industry and academia to solve computer vision problems is deep learning. It is a machine learning method that is in a way inspired by the hierarchal architecture of the brain. That is, multiple layers stacked one after another. To learn a task, information is processed in one layer before is handled to subsequent layers in the hierarchy.

Feature Parallelism Model (FPM) [3]; [4] has addressed the parallel nature of the brain. It conceptualizes unutilized science facts about the human visual system such as the Feature Integration Theory of visual attention “FIT” Nakayama and Silverman [5], Treisman and Gelade [6]; Treisman [7] and Quinlan [8].

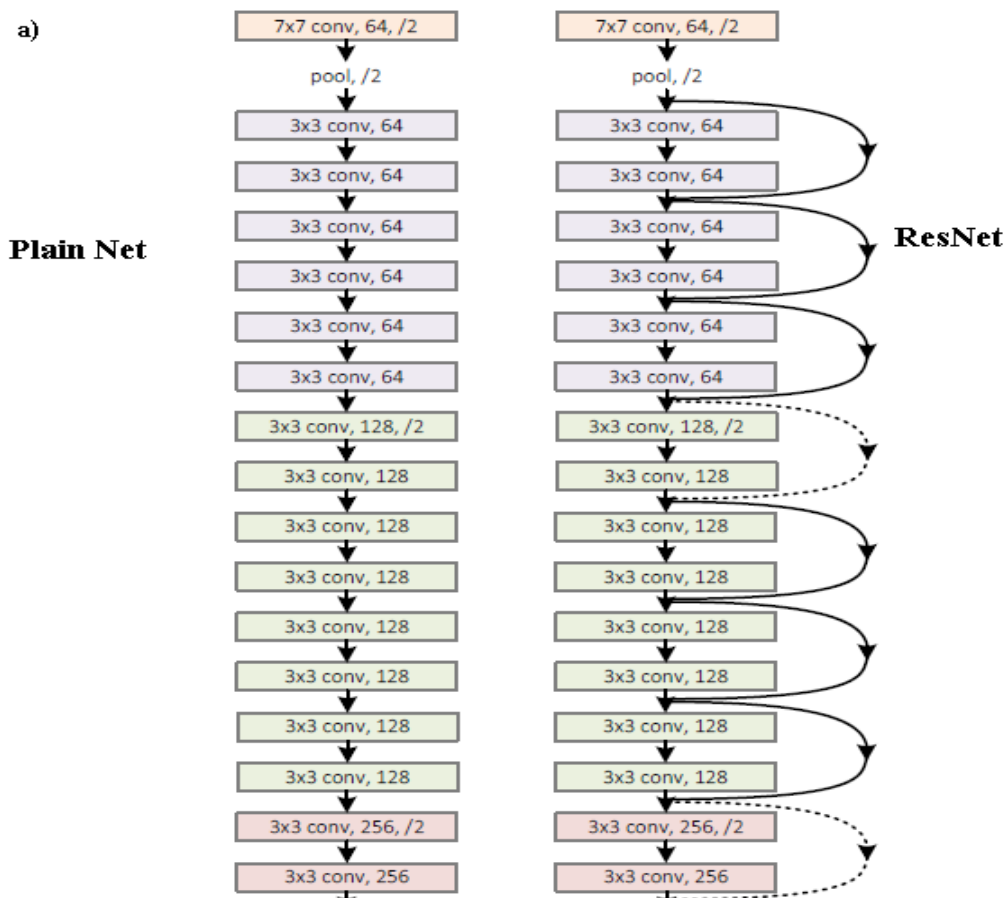
In this work, we compare between the two models, i.e. deep learning and FPM, along depth, which is the number of layers that comprises a deep learning model. It is the number of layers in the horizontal axis, in the case of FPM. In section 2, we discuss the significance of depth in deep learning models. Section 3 introduces FPM. In section 4, we discuss depth in FPM and then we conclude the paper in section 5.

2. SIGNIFICANCE OF DEPTH IN DEEP LEARNING

It has been widely accepted among researchers from academia and industry, that more layers means better accuracy for their deep learning models. They believe this is due to adding more nonlinearities with layers, which capture the complexity of the brain. Deep learning is inspired by the human brain, at least intuitively.

Researchers in Simonyan and Zisserman [9] have found that deeper models, up to 19 layers, produce better accuracies. However, others [10] found out that this is true up to a certain level, 20 layers in their case. This is when using normal convolutional neural networks, plain nets, as they named it Figure 1. After that level, test error would have begun to increase if more layers are added. They have proposed a new ConvNets scheme at the time, which is ResNet [10]. Figure 1 shows the difference between a plain net and a ResNet.

They have shown that under ResNet, deeper networks, up to 200 layers, add more accuracy to the model and are easier to train [10]; [11] Figure 2. Figure 2- a shows how deeper networks contribute to better results in ImageNet Large Scale Visual Recognition Competition (ILSVRC) over time. Figure 2- b shows the effect of depth in enhancing accuracy for ResNet architecture. And Figure 2- c shows some of those results in a table.



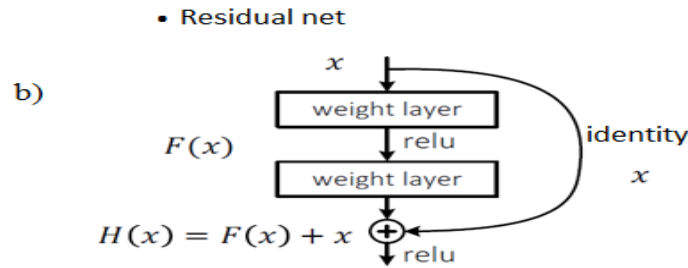


Figure-1. ResNet architecture [10].
 (a) Difference between a plain net and ResNet.
 (b) A single unit of ResNet.

Source: He, et al. [10].

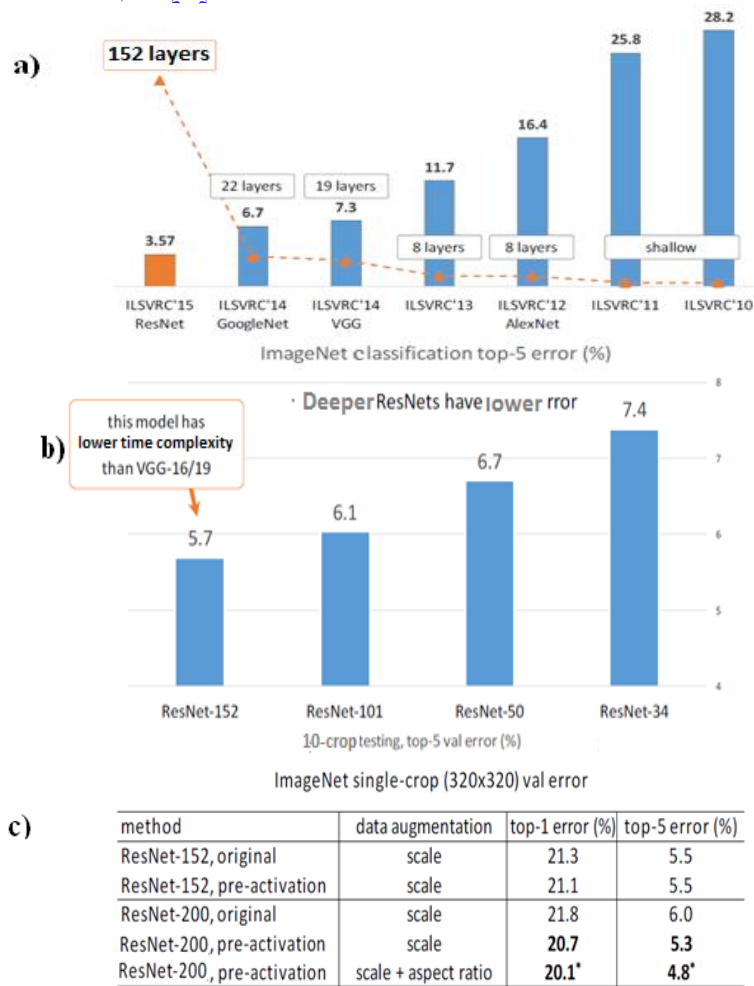


Figure-2. Revolution of depth- ResNet He, et al. [10] and He, et al. [11].

Source: He, et al. [10] and He, et al. [11].

3. FEATURE PARALLELISM MODEL (FPM) - DESIGN

The principle of parallel processing of different features- such as shape, color and motion- has been around for a while since the development of the feature integration theory (FIT) in the early 80's of the last Century. However, to the best of our knowledge, that concept have not been implemented in modern computer vision systems, until recently in Feature Parallelism Model.

General “feature parallelism” model for object recognition is designed and depicted in Figure 3 features such as color, shape and motion, are processed independently in parallel. Figure 4 shows that within each feature dimension, e.g. shape, there are parallel paths for sub- dimensions of that feature. For face recognition, the feature “Shape” is subdivided into 3 parallel sub- features, which are, texture, parts, and edges Figure 4.

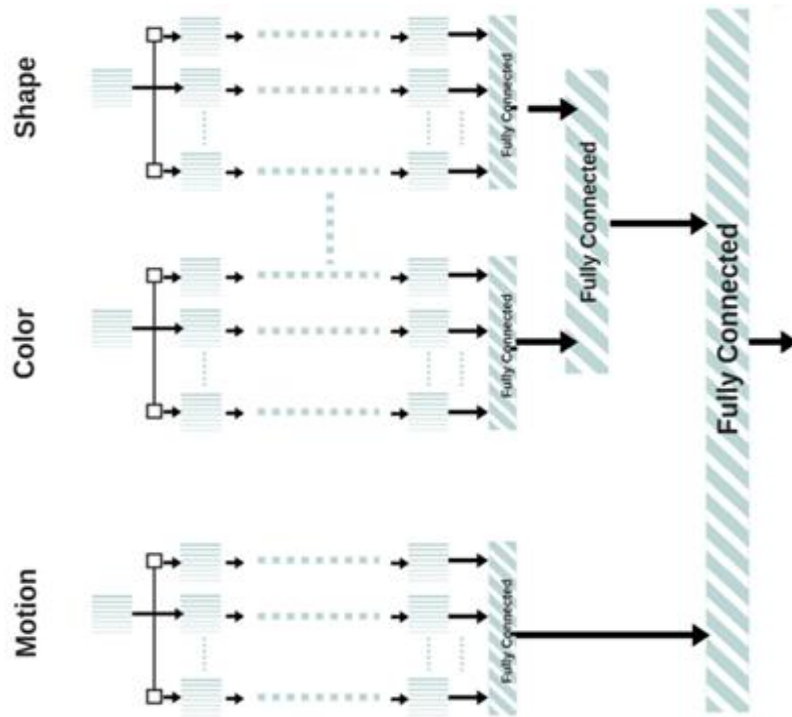


Figure-3. "General feature parallelism" model for object recognition [3].
Source: Marwa, et al. [3].

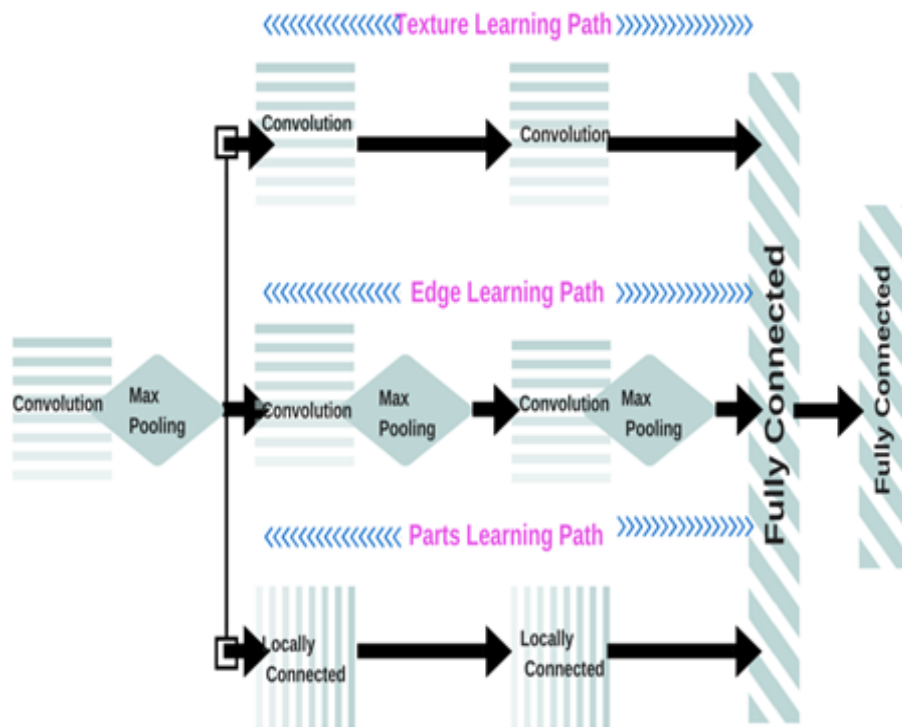


Figure-4. "Feature parallelism" model for face recognition along the "shape" feature [3].
Source: Marwa, et al. [3].

4. DEPTH IN FEATURE PARALLELISM MODEL

In this section, first we present a theoretical basis regarding depth in next section, and then we investigate the effect of depth on FPM in section 4.2.

4.1. Theory

Here, we discuss a theoretical basis regarding depth in neuroscience inspired artificial vision. We discuss next the functional architecture of the cerebral cortex where laminar organization and cortical columns are further elaborated.

4.1.1. Functional Architecture of the Cerebral Cortex

4.1.1.1. Laminar Organization

Differences in thickness, sizes and the shapes of neurons of the layers of the cerebral cortex have led researchers more than a century ago to identify about fifty distinct areas of the cortex. This classification was the basis to link different functions to these areas. However, the design principle that underlies the six-layered structure of the cerebral cortex remains a mystery. This six-layered structure is uniform across different species and different cortical areas [12]; [13].

Number of layers in previous deep learning models is considered a hyper-parameter [9]; [14] it's chosen either experimentally or arbitrary. A hyper-parameter is a variable that need to be adjusted and chosen prior to training. It is defined as a variable to be set prior to the actual application of the learning algorithm to the data, a one that is not directly selected by the algorithm itself [15].

4.1.1.2. Cortical Columns

A cortical column is the basic information processing unit of the cortex [16]. It refers to cells in any vertical cluster that share the same tuning for a given receptive field's attribute [17]. It is the narrow chain of neurons extending vertically across the six cellular layers, perpendicular to the surface of the cortex [16]. To explain more, consider the orientation selectivity of V1. One cortical column is responsible for the processing of only one orientation (say, the vertical orientation) and not responsible for another (e.g. the horizontal or the 45° orientation) Mountcastle [16], Horton and Adams [17] and Costa and Martin [18]. While Information inside a column is processed serially across its six layers, we can consider columns as the elementary parallel processing units of the cortex.

Some have argued that a cortical column is a structure without a function [17]. Here we can argue that dividing the cortex into small parallel units of similar tuning attributes, i.e. 'columns', may significantly reduce processing time and make learning faster.

4.2. Investigating the Effect of Depth on Feature Parallelism Model

We have used 136x136 images, ImageNet ILSVRC object recognition dataset, Nvidia K20 GPU card and Cuda-convnets2 deep learning framework. Our reference model is the 6-layers model.

We have trained all models for 45 epochs. We found that the optimum number of layers is six; with 68% top-1 error rate Figure 5 and 45% top-5 error rate Figure 6. Training time for the six-layered model was around 8.080 seconds Figure 7. Training time for the 5-layered model is about 8.450 seconds Figure 10, with 69% top-1 error rate Figure 8 and 46% top-5 error rate Figure 9. Results for the 7-layers model were 73% top-1 error rate Figure 11 52% top-5 error rate Figure 12 and training time of about 24.060 seconds Figure 13. The model with 8 layers has results as follows: 78% top-1 error rate Figure 14 58% top-5 error rate Figure 15 and around 27.350 seconds training time Figure 16.

4.2.1. 6-Layers End- to- End (Reference Model)

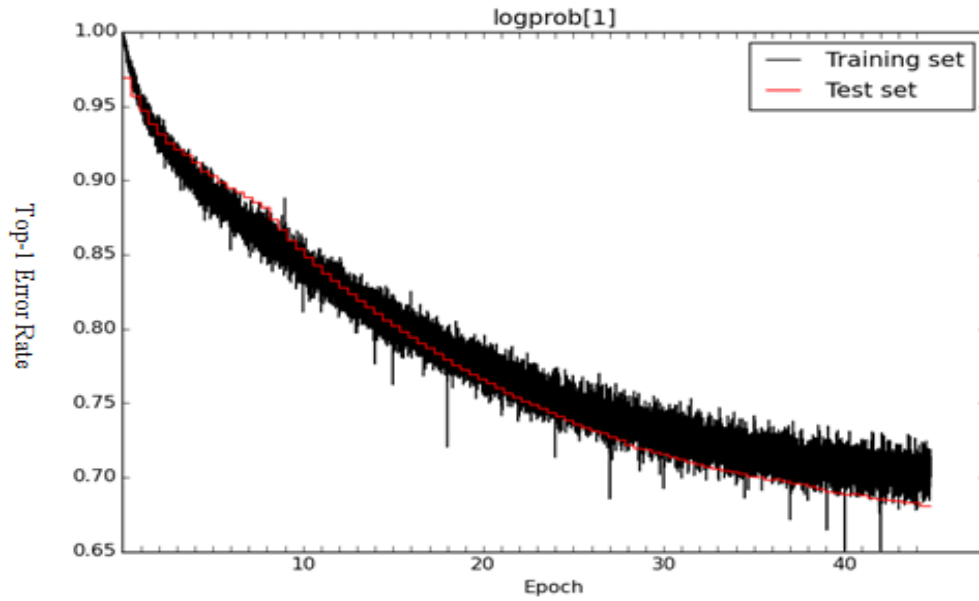


Figure-5. Top-1 error rate: 6- layer. Around 68%.

Source: Hassan, et al. [4].

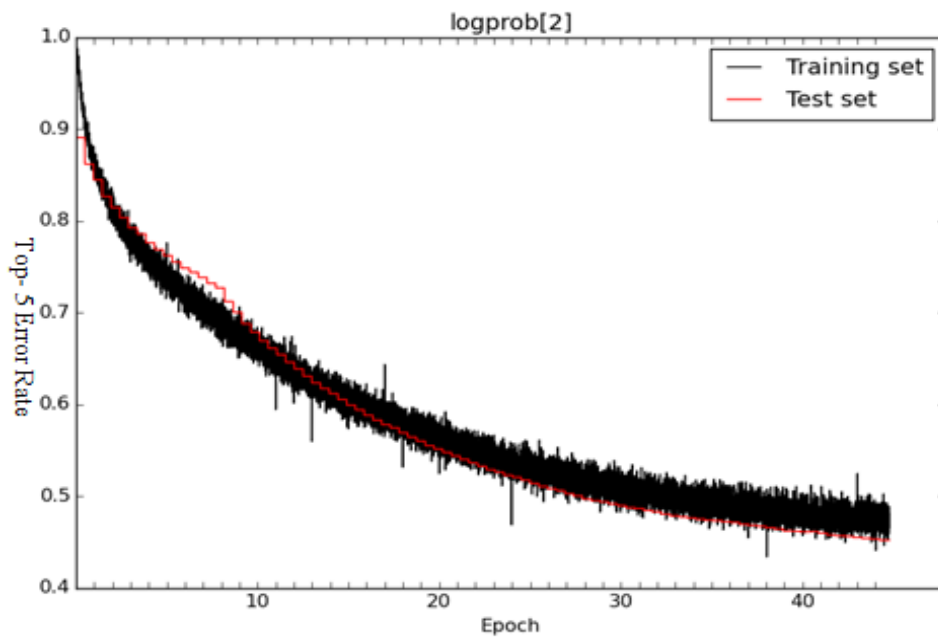


Figure-6. Top-5 error rate: 6- layers. Around 45%.

Source: Hassan, et al. [4].

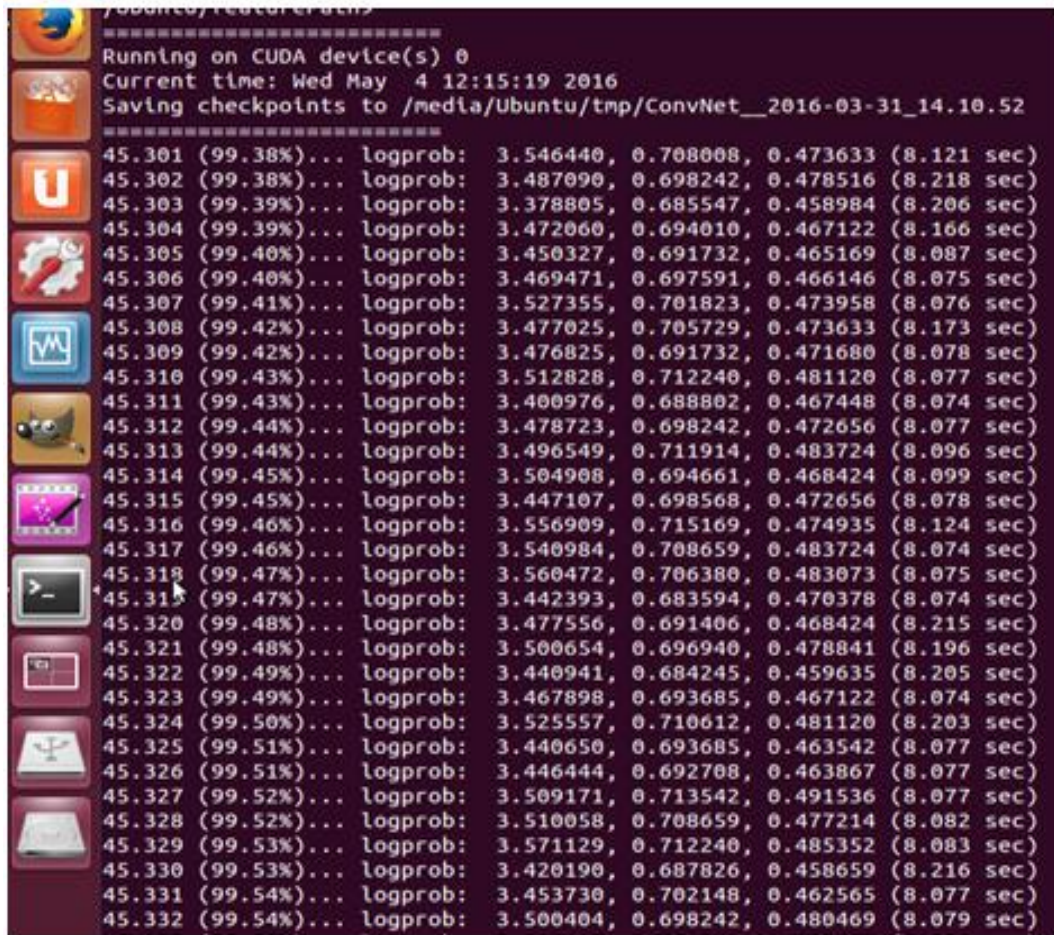


Figure-7. 6- Layers: Training time. Around 8.080 seconds.

Source: Hassan, et al. [4].

4.2.2. 5- Layers End- to- End

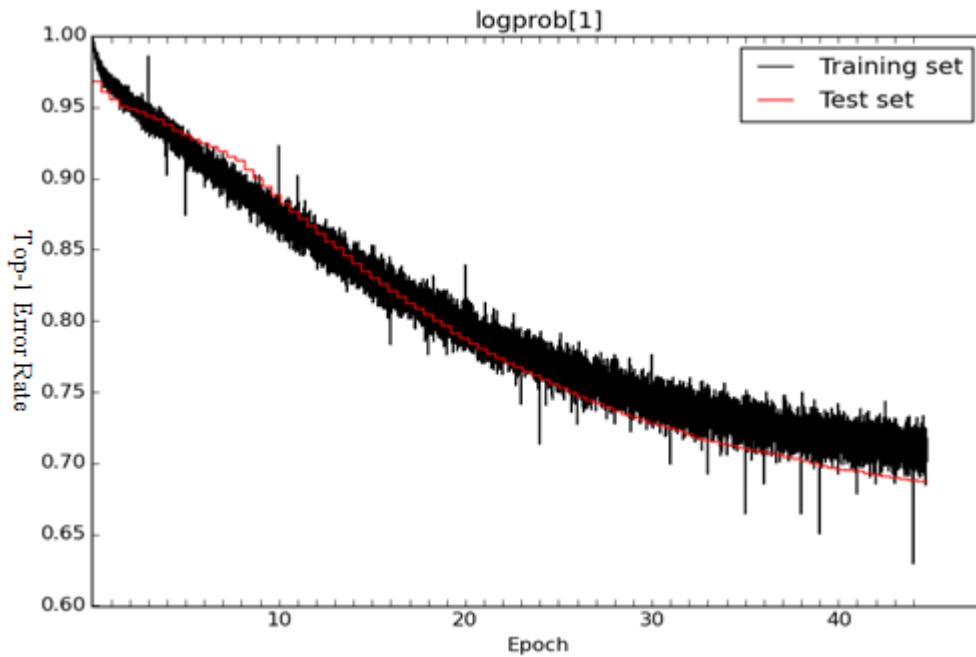


Figure-8. 5- Layers: Top-1 error rate. Around 69%, 1% more error rate.

Source: Hassan, et al. [4].

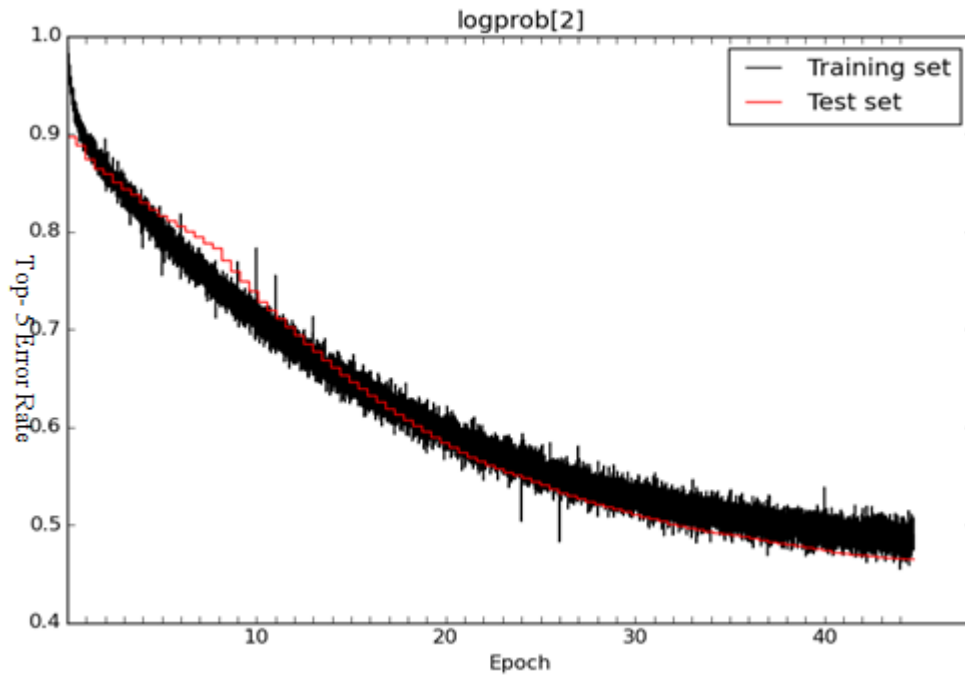


Figure-9. 5- Layers: Top- 5 error rate. Around 46%, 1% more error rate.

Source: Hassan, et al. [4].

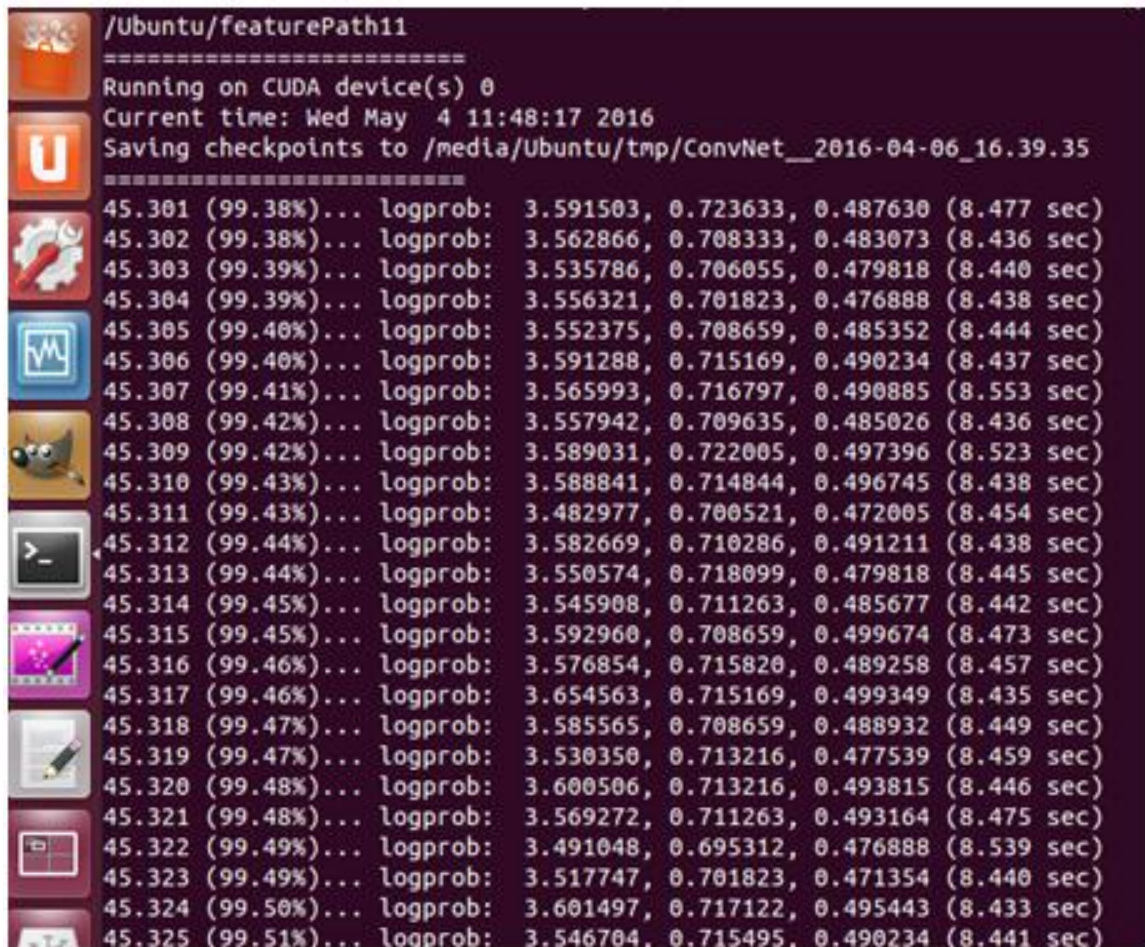


Figure-10. 5- Layers: Training time. Around 8.450 seconds, a few milliseconds more training time.

Source: Hassan, et al. [4].

4.2.3. 7- Layers End- to- End

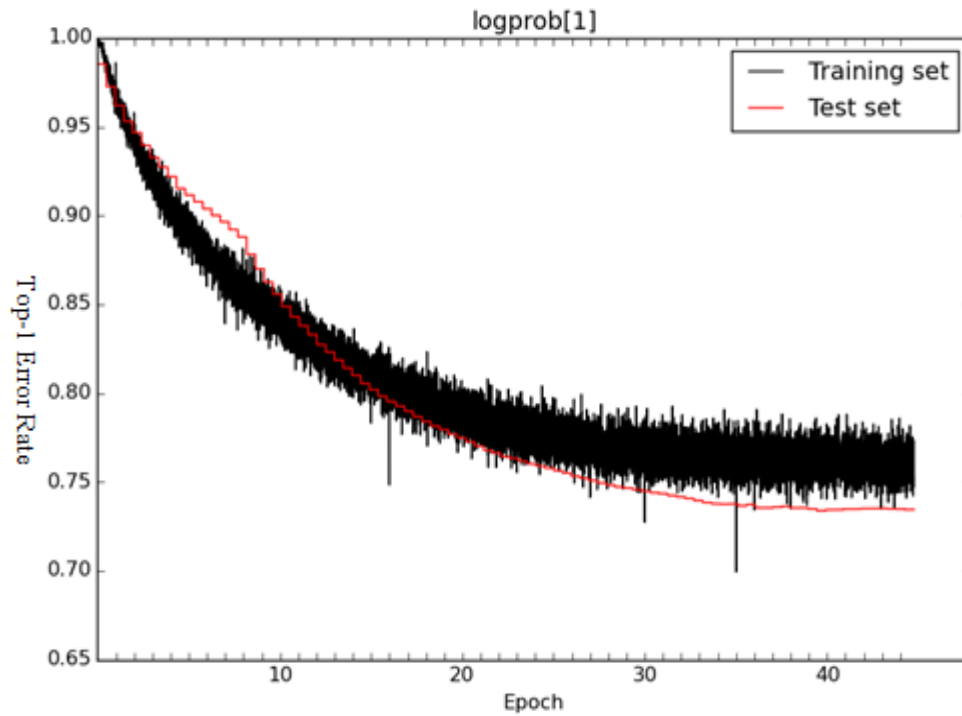


Figure-11. 7- Layers: Top-1 error rate. Around 73%, 5% more error rate.

Source: Hassan, et al. [4].

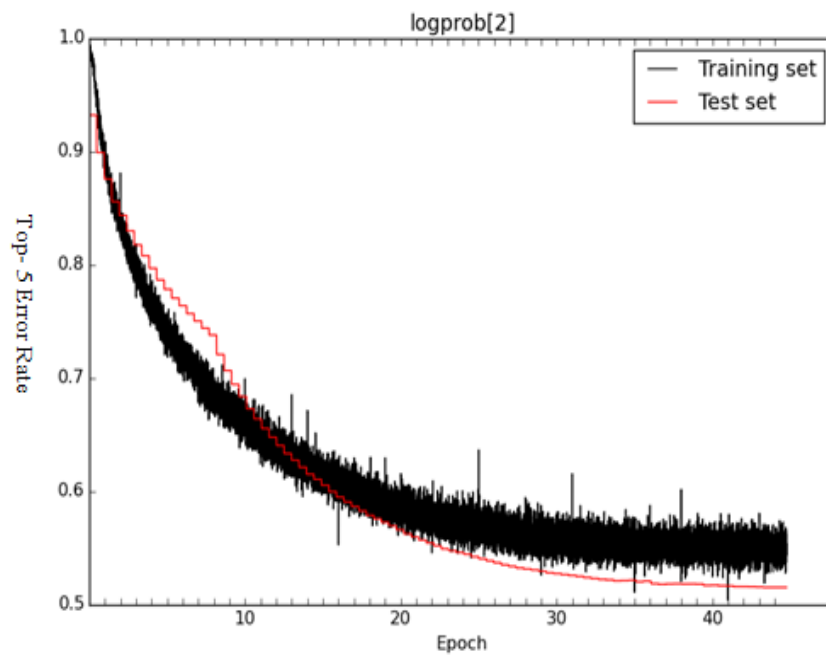


Figure-12. 7- Layers: Top-5 error rate. Around 52%, 7% more error rate.

Source: Hassan, et al. [4].

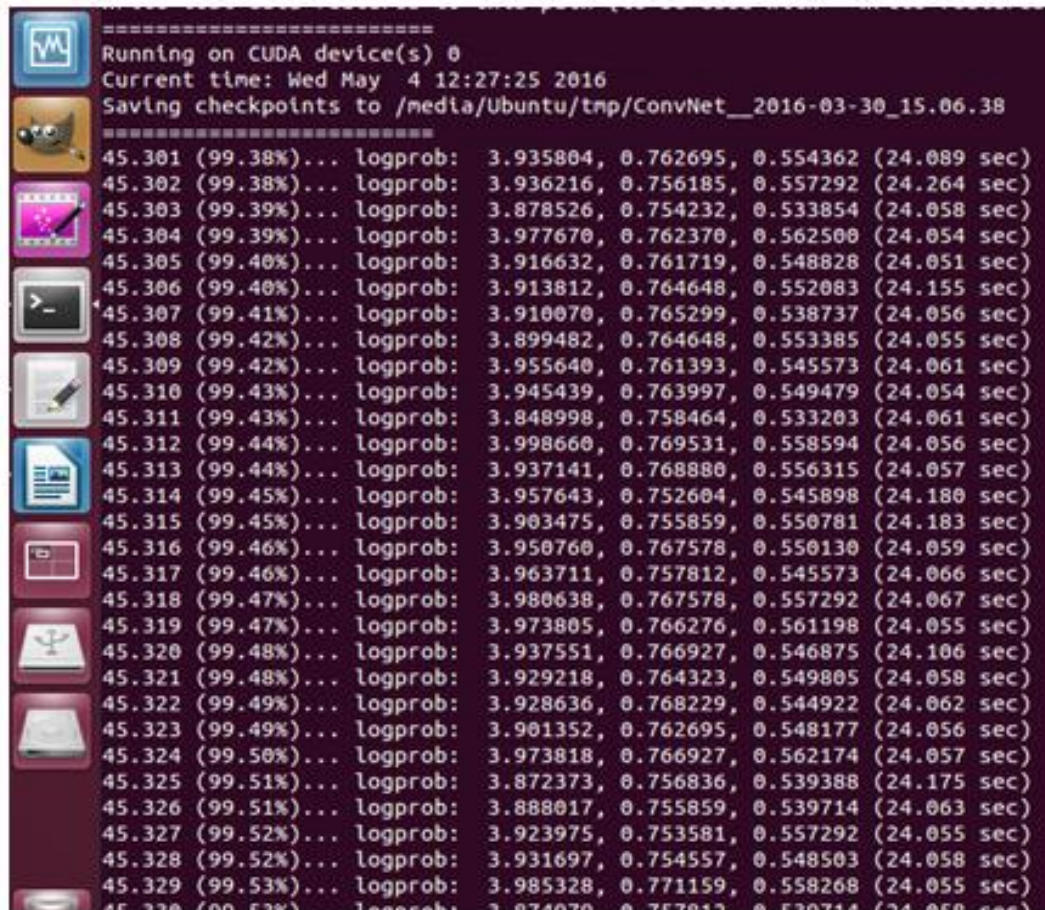


Figure-13. 7- Layers: training time. around 24.060 seconds, almost 3 times the training time of the 6- layered model. Source: Hassan, et al. [4].

4.2.4. 8- Layers End- to- End

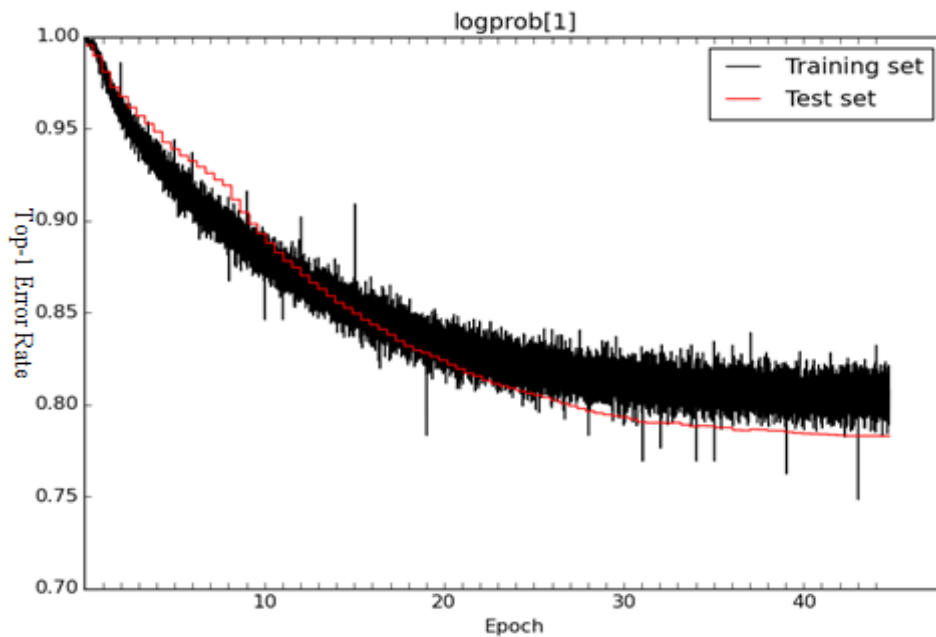


Figure-14. 8- Layers: Top-1 error rate. Around 78%, 10% more error rate compared to 6- layered model. Source: Hassan, et al. [4].

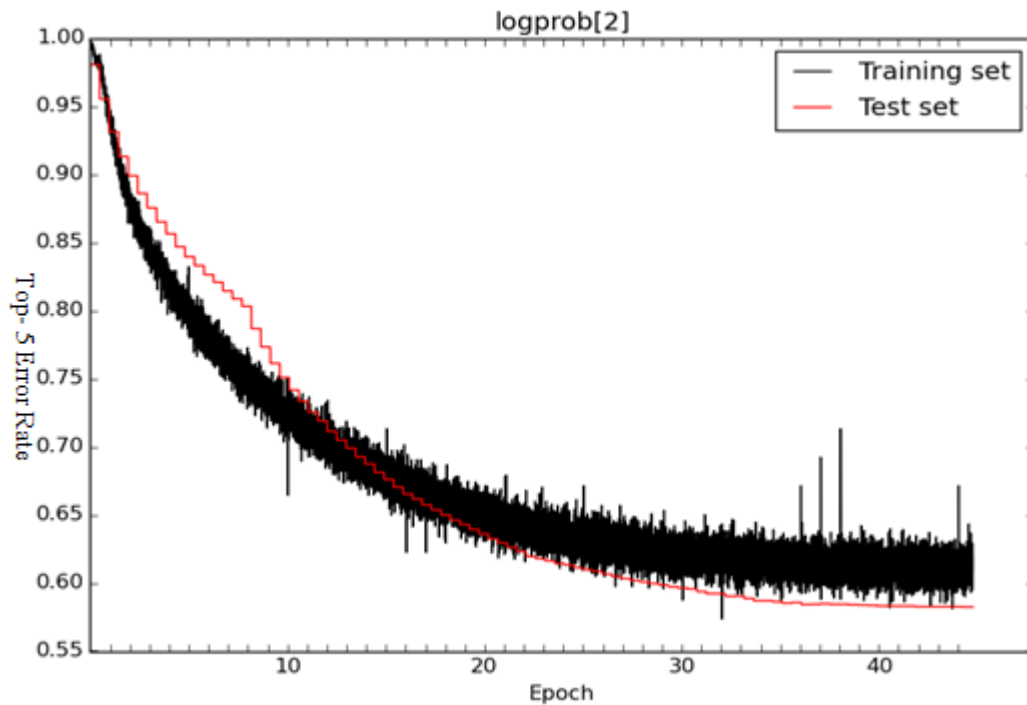


Figure-15. 8- Layers: Top-5 error rate. Around 58%, 13% more error rate, compared to 6- layered model.
Source: Hassan, et al. [4].

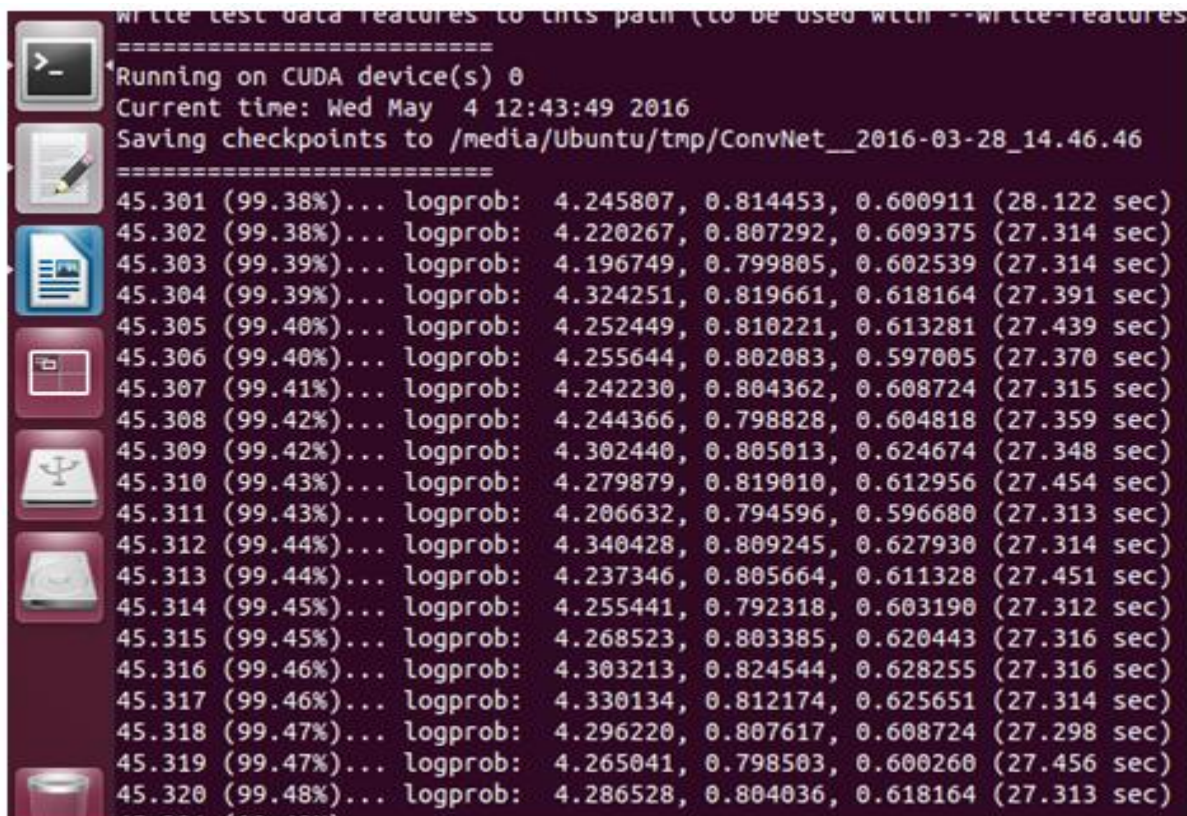


Figure-16. 8- Layers: Training time. Around 27.350 seconds, 19 seconds more training time, compared to that of 6- layered model.
Source: Hassan, et al. [4].

5. ANALYSIS AND CONCLUSION

When Feature Parallelism Model was investigated regarding depth, training time has dropped by few milliseconds while going from 5 to 6 layers; and error rate slightly drops by 1%. However, when going from 6 to 7 layers, training time increases significantly from 8.080 seconds to 24.060 seconds; almost three times that of the 6-

layers model. It continues to increase by a smaller rate when going from 7 to 8 layers to around 27.350 seconds. Similarly, top-1 error rate increases from 68% to 73%, a 5% increase when going from 6 to 7 layers; and top-5 error rate increases from 45% to 52%, a 7% increase. Also, when going from 7 to 8 layers we have noticed a 5% increase in top-1 error rate from 73% to 78%, and a 6% increase in top-5 error rate from 52% to 58%. The optimum number of layers has been found to be six layers.

Many studies have been conducted to decipher the wisdom behind cortical columns and the six-layered architecture of the cortex. Many theories have been developed. In this study we suggest a one possible wisdom, that is, the organization of parallel processing units into six-layer architecture appears to significantly enhance processing time and optimizes accuracy rates.

We conclude in this paper that while deep learning models encourage putting more layers, sometimes 200 layers and above, our study regarding depth in FPM suggests putting only 6 layers. Above 6, both accuracy and training time deteriorate significantly. Those results are biologically plausible, as they conform to the biological fact that the cerebral cortex is organized in 6-layers. Hence, the organization of parallel processing units into 6-layers, either in our brains or in artificial vision systems, enhances both accuracy and processing time.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Contributors/Acknowledgement: All authors contributed equally to the conception and design of the study.

REFERENCES

- [1] A. Krizhevsky and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," presented at the Advances in Neural Information Processing Systems (NIPS), 2012.
- [2] Y. Taigman, M. A. Ranzato, T. Aviv, and M. Park, "DeepFace: Closing the Gap to human-level performance in face verification abstract," presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.
- [3] Y. H. Marwa, O. K. Othman, B. A. T. Azhar, and H. A. Aisha, "A novel neuroscience-inspired architecture: For computer vision applications," presented at the 2016 Conference of Basic Sciences and Engineering Studies (SGCAC), 2016.
- [4] M. Y. Hassan, A. O. Shuriye, A.-H. Abdallah, and M. JE, "The feature parallelism model of visual recognition," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 12, pp. 171-186, 2017. Available at: <https://doi.org/10.14257/ijmue.2017.12.2.13>.
- [5] K. Nakayama and G. H. Silverman, "Serial and parallel processing of visual feature conjunctions," *Nature*, vol. 320, pp. 264-265, 1986. Available at: <https://doi.org/10.1038/320264a0>.
- [6] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97-136, 1980. Available at: [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5).
- [7] A. Treisman, "Perceptual grouping and attention in visual search for features and for objects," *Journal of Experimental Psychology. Human Perception and Performance*, vol. 8, pp. 194-214, 1982. Available at: <https://doi.org/10.1037/0096-1523.8.2.194>.
- [8] P. Quinlan, "Visual feature integration theory: Past, present, and future," *Psychological Bulletin*, vol. 129, pp. 643-673, 2003. Available at: <https://doi.org/10.1037/0033-2909.129.5.643>.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," presented at the 3rd International Conference on Learning Representations (ICLR), San Diego, USA, 2015.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," presented at the European Conference on Computer Vision. Springer, Cham, 2016.

- [12] S. Grossberg, *Laminar cortical architecture in visul perception*. In *The Handbook of Brain Theory and Neural Networks, Second Edi.*, M. A. Arbib, Ed. London, Massachusetts: The MIT Press Cambridge, 2003.
- [13] R. D. Raizada and S. Grossberg, "Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system," *Cerebral Cortex*, vol. 13, pp. 100-113, 2003. Available at: <https://doi.org/10.1093/cercor/13.1.100>.
- [14] S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, and V. D. Calhoun, "Deep learning for neuroimaging: A validation study," *Frontiers in Neuroscience*, vol. 8, pp. 1-11, 2014.
- [15] Y. Bengio, *Practical recommendations for gradient-based training of deep architectures*. In *Neural networks: Tricks of the trade*. Berlin, Heidelberg: Springer, 2012.
- [16] V. B. Mountcastle, "The columnar organization of the neocortex," *Brain: A Journal of Neurology*, vol. 120, pp. 701-722, 1997. Available at: <https://doi.org/10.1093/brain/120.4.701>.
- [17] J. C. Horton and D. L. Adams, "The cortical column: A structure without a function," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, pp. 837-862, 2005. Available at: <https://doi.org/10.1098/rstb.2005.1623>.
- [18] N. M. D. Costa and K. Martin, "Whose cortical column would that be?," *Frontiers in Neuroanatomy*, vol. 4, p. 16, 2010.

Views and opinions expressed in this article are the views and opinions of the author(s), Journal of Asian Scientific Research shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.