iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VIII Month of publication: August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37378

# Comparative Performance Evaluation of VGG-16 and Capsnet using Kannada MNIST

Dr. Ramya C[1], Rajput Rohan Singh[2], Rahul Kumar S[3], Jai Kumar S[4], Bandam Purna Sai Vijaychandra[5]
[1, 2, 3, 4, 5]*Computer Science Department, MVJ College of Engineering, Bangalore, India,560067*

***Abstract: Handwriting recognition is an important problem in character recognition. It is much more difficult especially for regional languages such as Kannada. In this regard there has been a recent surge of interest in designing convolutional neural networks (CNNs) for this problem. However, CNNs typically require large amounts of training data and cannot handle input transformations. Capsule networks, which is referred to as capsNets proposed recently to overcome these shortcomings and posed to revolutionize deep learning solutions. Our particular interest in this work is to recognize kannada digit characters, and making capsnet robust to rotation and transformation. In this paper, we focus to achieve the following objectives :1. Explore whether or not capsnet is capable of providing a better fit for the digit images; 2. Adapt and incorporate capsNets for the problem of kannada MNIST digit classification problem at hand; 3. develop a real time application to take handwritten input from the user and recognize the digit; 4. Compare the capsnet with other models on various parameters.***
***Keywords*: Capsule Networks, Deep Learning, Convolutional Neural Networks (CNNs), Kannada MNIST, VGG-16**

## I. INTRODUCTION

The convolutional neural networks have large learning ability and can infer the nature of an input image without any prior knowledge, which makes them an appropriate method for image classification (Krizhevsky et al., 2012). Although CNNs are considered as the pioneers of deep learning, they have some limitations and problems. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better. A well-known fact is that ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters.And also it needs large amounts of data for training. Due to these reasons, the interest of researchers is shifted toward capsule neural networks.

To overcome the limitations of CNN, Sabour and Hinton et al. have recently proposed Capsule networks (Sabour et al., 2017). In computer graphics, the image is created by some parameters like width, height and angle. In inverse graphics, these parameters are defined from the image and used for equivariance in the capsule. Capsule networks use the dynamic routing algorithm for recognizing the object successfully by the capsules consisting of a group of neurons. For object recognition, classification and segmentation, a more robust model was obtained by using the capsule structure, dynamic orientation and squash function in the capsule network. The purpose of our work is to classify images of the Kannada-MNIST data set using the capsule network and VGG-16 model. The rest of this paper is arranged as follows: In Section 2, related works are discussed. In Section 3, after showing the properties of the data set, VGG-16 and capsule networks are explained in detail. In Section 4, a real time application to take handwritten input from the user and recognize the digit is explained. In section 5 obtained results and observations are evaluated comparatively and conclusion is placed at Section 6.

## II. RELATED WORKS

Capsules are a new network architecture in Deep Learning and are producing amazing results by comparison to convolutional neural networks and traditional neural networks. The applications carried out with the capsule network reached by literature review are listed below. Mandal et al., have implemented capsule network and LeNet and AlexNet architectures on handwritten Indic digits and character datasets. They have also combined capsule networks wtih other networks such as LeNet and AlexNet. Their study has showed that AlexNet with capsule networks were achieved the best performance on most of the dataset (Mandal et al., 2018). Haque et al., have proposed a model for testing effectiveness of capsule network on Bangla handwritten recognition. Their results show that capsule networks offer acceptable accuracy (Haque et al., 2018). Nair et al., have used several datasets such as MNIST, fashion-MNIST, SVHN, etc. for comparing CNN and Capsule networks. For comparison they utilized the AlexNet model with the routing capsule network model. As a result, they found that capsule networks were able to perform better than AlexNet on more complicated datasets but they were not as good at dealing with deformations (Nair et al., 2018). Engelin has used Sellpy's dataset to test the rotational views comprehension in capsule networks.

This dataset includes images of clothes with a white background and are divided into two parts such as Sellpy Face Forward (SFF) and Sellpy Rotated Objects (SRO). They tested on traditional CNN architecture with different numbers of convolutional layers and observed that the error rate for capsule networks is lesser than that of CNN. As well the results show that capsule networks perform well on SRO images rather than SFF images (Engelin, 2018). Mehta and Parmar, have used CIFAR10 dataset for comparing the CNN and Capsule network. According to their results, the accuracy of the capsule network is better than CNN but for larger and complex images, it is not that promising like CNN (Mehta and Parmar, 2019).
Mukhometzianov and Carrillo investigated the performance of capsule networks with three well-known classifiers (Fisherfaces, LeNet, and ResNet). They tested the accuracy of classification on four datasets that includes images of faces, traffic signs, and everyday objects. The evaluation results show that CapsNet appears to be a new important image classification technique but requires significant computational resources for simple architectures Mukhometzianov and Carrillo, 2018).

## III.PROPOSED METHODOLOGY

### *A. Methods and Dataset*

In this part all the methods and dataset used is explained here.

1) *Dataset:* In this research, we used a new handwritten digits dataset called Kannada-MNIST (Prabhu, 2019). With nearly 60 million speakers worldwide, Kannada is the official language of Karnataka State in India. This dataset contains a training set and a test set that respectively consist of 60000 28x28 gray scale sample images and 10000 sample images equally distributed into the 10 classes. And also contains a more challenging test dataset called DigMNIST dataset that consists of 10240 28x28 gray-scale images.

Fig. 1. Sample images of kannada MNIST dataset.

### *B. Methods*

1) *Vgg16:* We can say that Vgg 16 is easy to understand and implement. It only contains a convolution and pooling layer. The architecture of Vgg 16 looks similar to the architecture of stack. the main aim of this model is to create a deep neural network by using a stacked representation of Conv and pooling layers.VGG is a classical convolutional neural network architecture. It was based on an analysis of how to increase the depth of such networks. The network utilises small 3 x 3 filters. Otherwise, the network is characterized by its simplicity: the only other components being pooling layers and a fully connected layer.

2) *VGG16 – Convolutional Network for Classification and Detection:* VGG16 is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition ". The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes. It was one of the famous models submitted to ILSVRC-2014. It makes the improvement over AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another. VGG16 was trained for weeks and was using the NVIDIA Titan Black GPU.

Architecture of VGG16

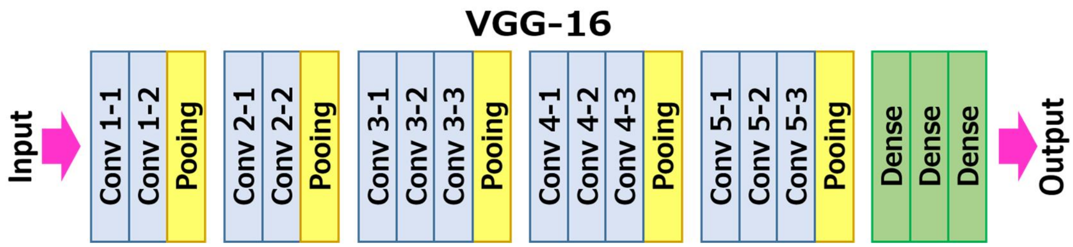



Fig. 2. VGG-16 Architecture

We have a final dense or output layer with 10 nodes of the size which classify between 10 classes of image net.

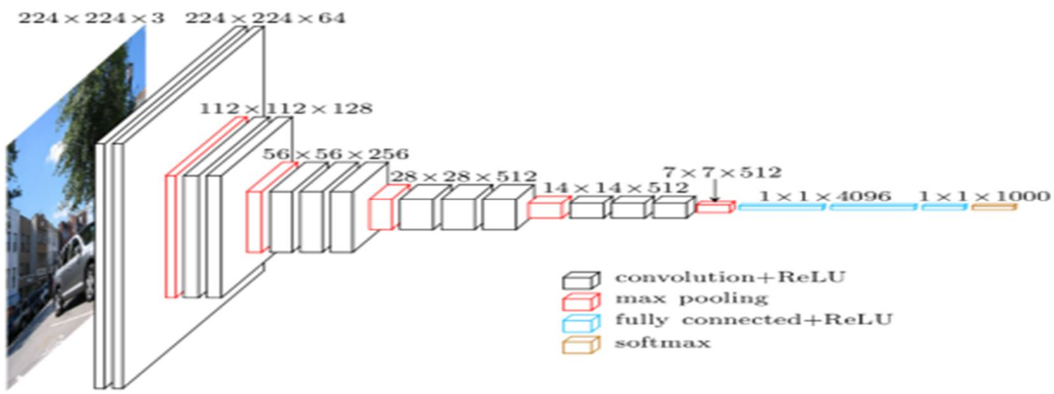



Fig. 3. VGG-16 layers

3) *CapsNet:* Capsule network (CapsNet) which has recently emerged as a more advanced architecture than its ancestor, namely CNN (Convolutional Neural Network). We could see the superiority of the capsule network on recognizing the handwritten digits, using all the previously applied methods. The power of the system against the noisy input images is further investigated, and the outputs have been proven to be robust against various noise effects and transformations. The output of the system, clearly outperforms the results achieved by its ancestors, as well as other previously presented recognition algorithms. Capsule neural network is an enhanced model of classical CNN. Firstly, the capsules have been introduced by Hinton et al. (2017) for addressing the limitations of CNN (Sabour et al., 2017). The main weakness of the CNN is mostly related to the pooling layers. Because the pooling layers of convolutional neural networks use sub-sampling which loses the precise spatial relationship. If the images are rotated or tilted, CNN will not test such images properly. Capsule networks can overcome such limitations because they use a dynamic routing algorithm and have a 16-dimensional vector that stores the pose parameters and orientation details (Mehta and Parmar, 2019). Capsules with transformation matrices allow networks to learn part-whole relationships automatically. A Capsule-Net is structured in several layers too. The lowest layer capsule is called the primary capsule and receives a small area of the image as input and tries to determine the presence and pose of a particular pattern. Simultaneously the upper layer capsules called routing capsules, trace massive and highly complex objects. With a few convolutional layers, the primary capsule layer is applied, thereafter the output is remodelled to a vector, where these vectors are given values ranging from 0 to 1 to represent a probability using a squashing function. This generates the output of the primary capsules. An algorithm called routing by agreement is used in the subsequent layers for tracing objects and their pose. This algorithm maintains a routing weight for each connection, when there is an agreement the routing weight increases otherwise it decreases (Sabour et al., 2017).

4) *Application:* To test the model in real world use, we tried to make a web application using streamlit. The user simply draws a kannada digit on the sketch pad. The web application converts it into image and converts it into gray scale images, resizes it to 28x28 pixels and feeds it to the saved model in the backend. The user gets the predicted output, computed by model, with a string output and a bar chart.

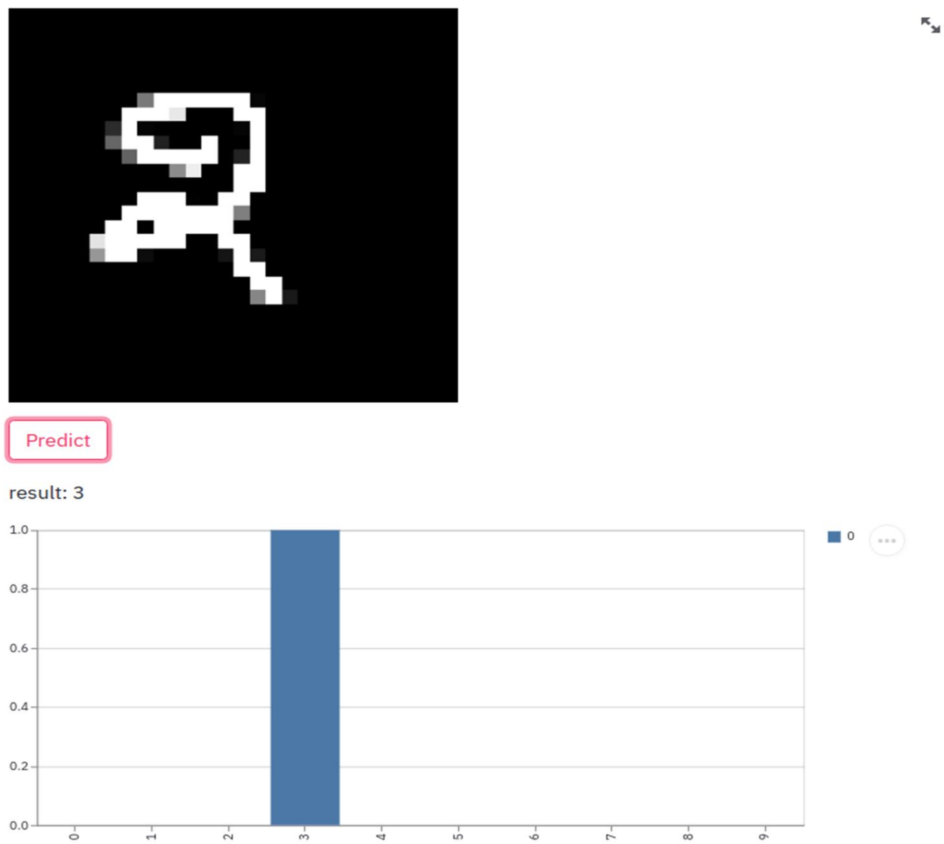



Fig. 4 Application

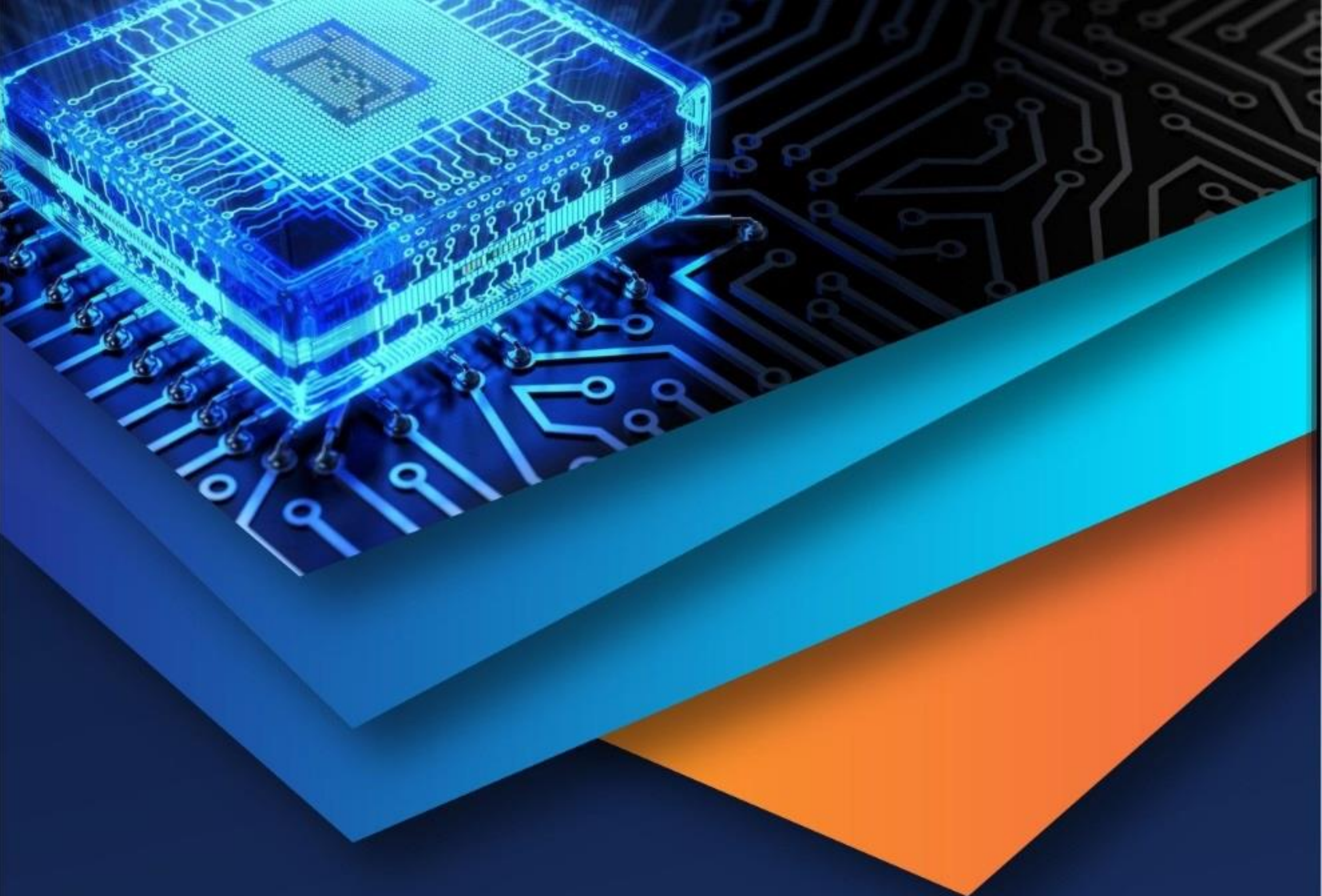
## IV. CONCLUSIONS

In this study we analyzed the performance of VGG-16 and the capsule network model, which is very new in the literature. For measuring the performance, we used a Kannada-MNIST dataset. Experimental results showed that the overall accuracy of Capsule networks obtained by the combination of the best conditions was found to be 97% while the overall accuracy of vgg-16was found to be 95% on testing data. As a result of the analysis carried out that Capsule networks achieves better results than CNN. In the future it is aimed to measure the performance of the model on different datasets.

## REFERENCES

[1] Dombetzki, L. A. (2018). An overview Over Capsule Networks. 1-7.
[2] E, G. (2020, 07 27). https://analyticsindiamag.com.
[3] Retrieved from https://analyticsindiamag.com: https://analyticsindiamag.com/understanding-capsule-net-with-its-implementation-in-computer-vision/
[4] Geron, A. (2018, 02 06). Retrieved from https://www.oreilly.com/: https://www.oreilly.com/content/introducing-capsule-networks/
[5] He, k., Zhang, X., Ren, S., & Sun, J. (2015). Surpassing Human level Performance on ImageNet Classification. 1-11.
[6] Kuo, J. (2016). Understanding Convolutional Neural Networks with A Mathematical Model. 1-21.
[7] LeCun, Y., Buttou, L., Bengio, Y., & Haffner, P. (2018). Gradient-Based Learning Applied to Document Recognition. PROC. OF THE IEEE, 1-46.
[8] Prabhu, V. U. (2019). KANNADA-MNIST: A NEW HANDWRITTEN DIGITS DATASET FOR THE KANNADA LANGUAGE. 1-21.
[9] Sabour, S., & Frosst, N. (2017). Dynamic Routing Between Capsules. 1-11.
[10] Wang, D., & Liu, Q. (2018). An OPTIMIZATION VIEW ON DYNAMIC ROUTING BETWEEN CAPSULES. Workshop track-ICLR, 1-4.
[11] Wu, J., & Group, L. (2017). Introduction to Convolutional Neural Networks. 1-31.
[12] Xi, E., Bing, S., & Jin, Y. (2017, 12 10). Retrieved from https://arxiv.org: https://arxiv.org/abs/1712.03480
[13] Zhao, W., Ye, J., Yang, M., Lei, Z., Zhang, S., & Zhou, Z. (2018). Investigating Capsule Networks with Dynamic routing for text classification. 1-10

[14] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules
[15] In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, pages 3856–3866. Curran Associates, Inc., 2017.
[16] Yann LeCun, Corinna Cortes and Chris Burges,“MNIST handwritten digit database”,IEEE 20- Nov-2019.
[17] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms,” pp. 1–6, 2017.
[18] G. Cohen, S. Afshar, J. Tapson, and A. Van Schaik, “EMNIST: Extending MNIST to handwritten letters,” Proc. Int. Jt. Conf. Neural Networks, vol. 2017-May, pp.
[19] Nimisha Jain, Kumar Rahul, Ipshita Khamaru. AnishKumar Jha, Anupam Ghosh (2017). “Hand Written Digit Recognition using Convolutional Neural Network (CNN)”, International Journal of Innovations & Advancement in Computer Science, IJIACS, ISSN 2347 – 8616, Volume 6.
[20] Mart´ın Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for largescale machine learning. In OSDI, volume 16, pages 265–283, 2016.
[21] Handwritten Digit Recognition using various Neural Network Approaches",International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 2, pp. 78-80, 2015.
[22] Comparisions of deep learning algorithms for MNIST in real-time environment. Akmaljon Palvanov and Young Im Cho- Gachon University, Seongnam , Korea June 2018
[23] Simonyan, K., & Zisserman, A. (2014, 9 4). Very Deep Convolutional Networks For Large Scale Image Recognition. Retrieved from https://arxiv.org/: https://arxiv.org/abs/1409.1556




10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY