JCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

A Study on Capsule Networks with the Comparative Analysis of Capsule Networks and CNN

Anusha Mehta[1*], **V. D. Parmar**[2]

[1,2]Dept. of Information and Technology, Shantilal Shah Engineering College, Gujarat Technological University, Bhavnagar, India

[*]*Corresponding Author: anumehta2194@gmail.com, Tel.: 0278-2200034*

DOI: https://doi.org/10.26438/ijcse/v7i4.105108 | Available online at: www.ijcseonline.org

Accepted: 14/Apr/2019, Published: 30/Apr/2019

Abstract— Artificial intelligence, with the emergence of machine learning and deep learning techniques, is growing up with breath neck speed. With the evaluation of the deep convolutional neural network, applications like image classification, object recognition and detection become easier. Recently, a new network deep learning architecture named Capsule Network is introduced to overcome some spatial and rotational limitations of CNN by using the concepts of capsules and the dynamic routing algorithm. Capsules are a group of neurons that generates activity vector whose length predicts the class of image and the orientation defines the pose parameters related to the image. Capsule networks have resulted in state of the art performance on various dataset such as MNIST. The paper defines the architecture and working of the capsule network, along with the comparative analysis of CNN and Capsule network on the various dataset. Along with this, the paper specifies the hands-on experiments done on capsule networks and the future scope with capsule networks.

Keywords— Capsule networks, convolutional neural networks, deep learning, dynamic routing algorithm, image classification.

I. INTRODUCTION

Convolutional neural networks can be considered as the pioneer of today's deep learning. The emergence in various design and models of CNN are the reasons why deep learning applications are very popular and prominent today. But, CNN is also having some limitations and drawbacks related to spatial information and rotational invariance that should be overcome.

Capsule networks, the comparatively new deep learning network architecture is inspired by the inverse graphics and hierarchical mapping concepts of the human brain. In the human brain, low-level features and the spatial relationship between object features are responsible for making the high-level feature or object prediction. Capsule networks use the same concept with the dynamic routing algorithm. Plus, in computer graphics, the image is created by instantiation parameters such as height, width, angle, etc. In inverse graphics, the instantiation parameters are defined from the image, this concept is used in capsule networks for equivariance.

Capsule networks are made up of capsules which is the group of neurons. The neurons are nested together to make a caps ule. Capsule generates an activity vector rather than scalar values as in CNN. The activity vector's length gives the probability of which the object exists and the orientation gives the instantiation parameters such as pose, hue, texture, etc. Capsule networks work on routing by agreement algorithm i.e. low-level capsule will bet on the high-level capsule for their existence in high-level capsules. According to that, whose prediction vector will be larger, that will give feedback. This is referred to as a dynamic routing algorithm.

The organization of the paper is as follows: Section I contains an introduction, section II contains the related technology, section III contains working of capsule network, section IV contains the related work on the capsule network with the comparative analysis, section V contains the experiments and results, section VI contains the benefits and limitations of capsule network and section VII contains the conclusion and future work.

II. RELATED TECHNOLOGY

The capsule networks are an extension of convolutional neural networks with some changes in architecture and overcome the limitations of CNN. For comparison, convolutional neural networks are used.

A. Convolutional neural network

The convolutional neural network is an artificial neural network with the convolutional operation and more hidden layers. The traditional convolutional neural network is made up of the main three layers: convolutional layer, pooling layer, and fully connected layer. CNN is mainly divided into two parts: the feature extraction part and the classification part.

Convolutional layer performs on input data with the filter and stride to provide feature maps of images. Numerous convolutions are performed on the image with different filter size and strides generate different feature maps [1]. The feature maps together considered as the output of the convolutional layer. In convolutional layer, the activation is also performed for non-linearity. Various activity functions can be used such as sigmoid, softmax, ReLU.

After convolutional layer, pooling layer reduces the dimensionality to reduce the number of parameters and computation in the network which accordingly reduce the training time and controls overfitting [1]. The major disadvantage of using pooling layer is it loses the spatial information related to positions of features in the image which may create trouble in testing results. Finally, the fully connected layer flatten the data and classify the images to a particular class.

B. Limitations of the convolutional neural network

Pooling layers of convolutional neural network use sub-sampling which loses the precise spatial relationship. Because of that if the position of the feature in image changes, CNN will not test such images properly. Plus, CNN is rotationally invariant. So, if images are rotated or translated then it may results poor in testing. Such limitations can be overcome by capsule networks because capsule networks use a dynamic routing algorithm and they are equivariant. Capsule networks have a 16-dimensional vector that stores the pose parameters and orientation details.

III. CAPSULE NETWORK

A. The architecture of Capsule network

The capsule network architecture is made up of two parts: encoder and decoder. The encoder part is useful for generating a 16-dimensional vector from an image that contains instantiation parameters. The decoder part recreates the image from a 16-dimensional vector of a correctly predicted capsule. It forces capsule to learn features that are useful for reconstructing the original image.

The layer structure [2] for capsule network is as follows:

- Convolutional layer
- Primary capsule layer
- Digit capsule layer
- Decoder with a fully connected layer

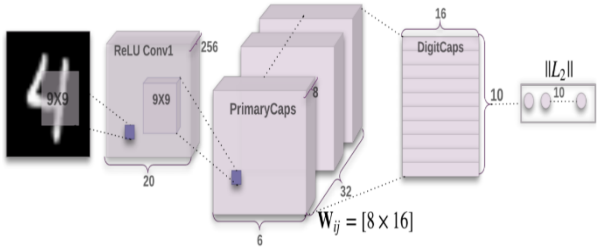

Figure 1. Capsule Network Encoder [2]

The encoder part of capsule network is responsible for the prediction of the class label related to the image. The classification part of capsule network is its encoder network. There are mainly three layers: convolutional layer, primary capsule layer, and routing capsule/ digit capsule layer.

Here, the input image first passes from a simple convolutional layer with respective kernel size and stride just like CNN. The next layer is the primary capsule layer. This layer takes basic features from the convolutional layer and produces combinations of the features. The capsules are reshaped to 8D vector here. Next, it will pass to the digit capsule layer and generates 16D capsule per digit class. The routing algorithm works between the primary capsule and digit capsule layer. The length of 16D vector defines the probability of the respective class for the classification purpose.

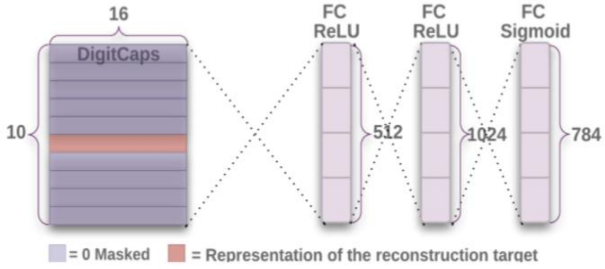

Figure 2. Capsule Network Decoder [2]

The decoder part is responsible for the reconstruction of the image. The output of digitCaps is fed into a decoder. Decoder consists of 3 fully connected layers that model pixel intensity [2]. The reconstruction from 16D output capsule generated here. This is how the basic architecture of capsule network is. The capsule network model is a multi-input model and follows the concept of auto-encoders [3].

B. Characteristics of Capsule Network

1) *Better Connection:* Capsule networks are made up of capsules which are a group of neurons and nested together for a better connection. All the features are connected and passed to one layer to another layer to increase computational efficiency.
2) *Routing by Agreement:* In a capsule network, each child node is connected to all the parents' node. Here, whose

prediction vector is large is considered to be having strong bonding, and makes a strong agreement of routing. This reduces backpropagation errors.

3) *Capsules are Equivariant:* Capsule networks provide better results with the affine transformation of images because of equivariance. That means when an object changes its dimension, activity vector will also change its dimension without affecting the probability of vector.

IV. RELATED WORK

Capsule networks are first introduced by Hinton in the paper “Dynamic Routing between Capsules” [2] in 2017. The paper covers overall working and mathematics behind capsule networks. The capsule networks output vectors instead of scalar values. The paper defines state-of-the-art performance on MNIST dataset with the comparison of baseline CNN. The experiments are done on testing data, such as affine transformation and capsule nets result better for transformation than CNN.

The paper "Matrix Capsules with EM routing" [4] is the next version of the capsule networks that improvise the capsule network performance from vector to matrix capsule and apply EM routing to classify images with different viewpoints. The dataset used was a smallNORB dataset. The baseline CNN model is 2 convolutional layer with max pooling and fully connected layer. The capsule network provides state-of-the-art results with this dataset rather than CNN and also robust for the adversarial attacks.

The comparison of CNN and Capsule Networks are also defined in the paper "Pushing the limits of Capsule Network" [5]. The paper uses various dataset such as MNIST, fashion-MNIST, SVHN, etc. Here, for comparison, it uses the AlexNet CNN model with the routing capsule network model. The testing process is done on the affine transformation of images in which capsule networks perform better. The problem arises in the reconstruction of images from SVHN and also iterations in the routing algorithm is not affecting the whole network.

To know the rotational views comprehension in capsule networks, Sellpy's dataset has been used. The images are photographs of clothes with a white background. These images are divided into two parts such as Sellpy Face Forward (SFF) and Sellpy Rotated Objects (SRO). The dataset is tested on traditional CNN architecture with different numbers of convolutional layers such as CNN-1, CNN-2, etc. Here, it is observed that the error rate for capsule networks is lesser than that of CNN, and capsule networks perform well on SRO images rather than SFF images [7].

According to this, capsule networks are suited for the affine transformation of images and the parameters may lesser than that of CNN. But CNN also provides state of the art results on a dataset without affine transformation. The comparative survey on the various dataset is as follows:

Table 1. Comparison of various model of capsule network and CNN

Dataset	Model Setup		Accuracy (%)	
	Proposed model	*Baseline CNN*	*Proposed CapsNet*	*Baseline CNN*
MNIST	CapsNet	Traditional CNN	99.23	99.22
AffNIST	CapsNet	Traditional CNN	79	66
SVHN	CapsNet	AlexNet	91.06	87.43
Fashion-MNIST	CapsNet	AlexNet	89.80	83.00

V. EXPERIMENTS AND RESULTS

Capsule networks are an immature field to be developed right now. The results show that major concern is the transformations are predicted even if it is not in training set and the output vector represents pose parameters.

A. Dataset used

In the experiment, the dataset used is standard CIFAR10 dataset. The capsule network gives a state-of-the-art performance on datasets like MNIST but this dataset consists of comparatively simple greyscale images with a similar background. CIFAR10 is a dataset having total 60000 of RGB images belonging to 10 classes such as an airplane, horse, bird, etc. The dataset is divided into 50000 training examples and 10000 testing examples. The image size is (32, 32, 3) taken in this experiment. The shape of the images is 32x32 and the type is RGB image with the varying background.

B. Software and Hardware Requirements

For programming, keras deep learning framework [8] with the tensorflow backend is used. Keras is an open-source deep learning library and it is written in python language. The programming language is python with modules such as numpy, matplotlib, opencv, etc. The dataset CIFAR10 is preloaded in keras library. For faster training process, Google Colaboratory free GPU service is used. The GPU specs of google colab are Nvidia Tesla K80.

C. Model Setup and Results

The model used is a capsule network with dynamic routing. The training setup is 50 epochs with Adam optimizer and batch size 100. The routing iterations are set to 3. After training, the accuracy of capsule network for CIFAR10 dataset is 66.05% and the testing accuracy is 65.56%. For comparison, the CNN model used is AlexNet. The accuracy of this model is 75.77% and sometimes more than that too.

Here is where the limitations of capsule network are highlighted.

Capsule networks result in poor for RGB large images and the reconstruction of images is also poor. Though the error rate is less in capsule network and parameters are also less so it may possible that extending the capsule network with more convolutional layers can improve the accuracy.

D. Enhancement in Capsule Network model

For increasing the accuracy for CIFAR10 dataset in capsule network, the number of convolution layers is increased. For this experiment, we are using a pre-trained VGG model instead of a first single convolutional layer of capsule network. Other layers such as the primary capsule layer and digit capsule layers will remain as it is. After this setup, when the capsule network is trained, the accuracy is increased to 85.07%, which is more than baseline CNN.

From results, it is assumed that accuracy for RGB images in capsule networks can be increased by using more number of convolutional layers at the beginning. The trained model can be used as well as transfer learning can also be helpful and worth exploring with capsule networks.

VI. BENEFITS AND LIMITATIONS OF CAPSULE NETWORK

A. Benefits

- Viewpoint Equivariance.
- Spatial information is utilized
- Less number of parameters
- Works well with affine transformation
- Information loss is less.
- Full connection due to the routing algorithm
- Less number of images are required for training
- Performs better for greyscale simple images

B. Limitations

- Uncertainty in training large images
- Poor performance in varying background image data
- Slow training process

VII. CONCLUSION AND FUTURE WORK

Capsule networks provide state-of-the-art performance on simple greyscale images like MNIST but for larger and complex images, it is not that promising like CNN. To improve accuracy, more convolutional layers can be added. Capsule networks perform well for affine transformation than CNN. It is observed that capsule networks are also promising in object segmentation [5] and white box adversarial attacks. With compared to such attacks, Capsule networks are robust than CNN. According to the survey, we can conclude that capsule network can be further enhanced for major computer vision tasks.

The future scope includes modification in existing capsule network layers for improving accuracy in varying background of images. The capsule network model is comparatively simpler than the latest CNN models and uses only one convolutional layer, so making the network deeper can help increase accuracy. The routing algorithm uses iterations that is taking too much training time, so training time is also a major constraint to be taken into account. These are the future direction in which the ability of capsule networks can be increased.

REFERENCES

[1] K. O'Shea, R. Nash. "*An introduction to convolutional neural networks*." arXiv preprint arXiv:1511.08458 (2015).

[2] S. Sabour, N. Frosst, and GE. Hinton. "*Dynamic routing between capsules."* In the Proceedings of 2017 NIPS Conference on Advances in neural information processing systems, pp. 3856-3866. 2017.

[3] GE. Hinton, A. Krizhevsky, SD. Wang. "*Transforming auto-encoders*." In International Conference on Artificial Neural Networks, Springer, Berlin, Heidelberg, pp. 44-51, 2011.

[4] GE. Hinton, S. Sabour, N. Frosst. "*Matrix capsules with EM routing."* Open review (2018).

[5] P. Nair, R. Doshi, S. Keselj. "*Pushing the limits of capsule networks*." Technical note (2018).

[6] R. LaLonde, U. Bagci. "*Capsules for object segmentation."* arXiv preprint arXiv:1804.04241 (2018).

[7] M. Engelin. "*CapsNet Comprehension of Objects in Different Rotational Views: A comparative study of capsule and convolutional networks*." (2018).

[8] A. Gulli, and S.Pal. "*Deep Learning with Keras"*. Packt Publishing Ltd, Birmingham - Mumbai, pp. 71-105, 2017.

AUTHORS PROFILE

Ms. Anusha Mehta is Bachelor of engineering in Information and Technology from Government Engineering College - Bhavnagar, affiliated with Gujarat Technological University, Gujarat. She is currently pursuing Masters of Engineering in Information and Technology from Shantilal Shah Engineering College - Bhavnagar, affiliated with Gujarat Technological University, Gujarat. Her main research work focus is on deep learning, capsule networks, and convolutional neural networks. Research is going on under the guidance of V. D. Parmar at Shantilal Shah Engineering College – Bhavnagar, Gujarat.

Mr. V. D. Parmar is Masters of engineering in Information and Technology. He is currently working as Asst. Professor at Shantilal Shah Engineering College – Bhavnagar affiliated with Gujarat Technological University, Gujarat. His main research work focus is on artificial intelligence, machine learning, and deep learning.