International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.095

Relative Scrutiny of Different Capsule Network Architectures

Mr. B. Thiyagarajan
Research Scholar
Department of Computer Science and Engineering
Pondicherry Technological University
Puducherry. India
thiyagarajan.b@pec.edu

Dr. M. Thenmozhi
Assistant Professor
Department of Computer Science and Engineering
Pondicherry Technological University
Puducherry. India
thenmozhi@ptuniv.edu.in

Abstract—The integration of computer vision and machine learning techniques has led to significant advancements in pattern recognition and image categorization. One of the most sophisticated machine learning techniques for encoding features based on hierarchical relationships is the capsule network. Unlike convolutional neural networks (CNNs), which lose much spatial location information and require extensive training data, capsule networks use inverted graphics and a network of capsules to represent objects as separate pieces and establish their interconnections. This paper aims to present an overview of various capsule network architecture designs employed in different applications, highlighting their pros and cons. The objective is to provide readers with a comprehensive understanding of the current state-of-the-art capsule network topologies.

Keywords—Capsule Neural Network, CNN, Deep Learning, Image Classification

I. INTRODUCTION

Computer vision is a crucial field of artificial intelligence that has numerous real-time applications, such as security, character recognition, object segmentation, and image recognition [1-2]. Convolutional neural networks (CNN) and recurrent neural networks (RNN) are the primary deep learning techniques used in computer vision, as traditional symbolic AI approaches are insufficient for complex real-time problems [5]. Among these techniques, CNNs are the most widely used and effective solution for image classification, image identification, and picture recognition tasks, ranging from simple to complex objects [6-7]. Nonetheless, CNNs suffer from the pooling process, which leads to the loss of vital details like object position and posture [8]. Additionally, CNNs lack rotational invariance and require extensive training data [9]. To overcome these limitations, alternative methods such as reinforcement learning [10] and end-to-end connected layers have been proposed. However, these approaches did not yield significant improvements, leading to the development of Capsule Network (CapsNet).The CapsNet has been shown to improve model accuracy by up to 45% compared to CNNs. This paper aims to review the shortcomings of CNNs while highlighting the positive results of CapsNet in the literature. Therefore, the primary contributions of this paper are:

- Inspire researchers by presenting cutting-edge capsule models.
- Investigate potential future research fields related to capsule networks.
- Provide a comparison of the most advanced CapsNet architectures to aid in selecting the most suitable model for specific applications.
- Examine the variables that influence the performance of these architectures with modifications and applications, allowing for a better understanding of the strengths and weaknesses of CapsNet.

This study aims to clarify the fundamental concepts of capsule networks, which have gained popularity as a recent research area. Additionally, we provide a comparative analysis of CapsNet designs used in various applications to overcome the limitations of CNNs, including their advantages, disadvantages, and potential future approaches.

The paper is structured as follows: Section 1 outlines the study's objectives and provides background information on the topic under review. In Section 2, we summarize literature survey of CapsNet architectures, including their limitations, modifications, and applications. Section 3 describes the performance analysis. Finally, Section 4 concludes the paper.

A. Convolutional Neural Network (CNN)

Let's discuss the attributes of an image required for its recognition by a CNN. Specifically, let's consider a gray scale 2x2 pixel image, with each pixel represented by an 8-bit value ranging from 0 to 255. These values denote the intensity of the corresponding pixel, where 0 represents black and 255 represents white. The gray scale range between black and white spans from 0 to 255. Figure 1(a) illustrates the computer's interpretation of the gray scale image, while Figure 1(b) depicts the computer's representation of its features.

Pixel 1	Pixel 2
Pixel 3	Pixel 4

Fig. 1.(a) 2 X 2 Gray scale image

Deep Learning (DL) algorithms designed for image categorization are referred to as Convolutional Neural Networks (CNNs). Their primary purpose is to interpret image data, assign importance to various features, and differentiate between different classes. A CNN is composed of a fully connected layer with an activation function, a pooling layer, and a convolutional layer [11].

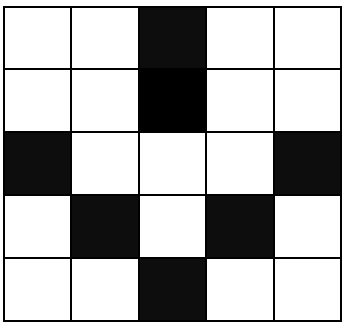
The input image is scanned by the convolutional layer to extract low-level characteristics, such as edges. This layer performs convolution between the input image and a set of

DOI: rp-9788770040723.095

learnable filters, resulting in the production of feature maps that highlight different patterns and features of the image. The Rectified Linear Unit (RELU) activation function [12] is applied to introduce nonlinearity in the model's computations and improve its accuracy.

The pooling layer, also known as down-sampling, is utilized to reduce memory requirements and recognize the same object in multiple images. Different types of pooling methods such as maximum, minimum, sum, and average pooling are employed based on the requirements of the model.

Figure 2 demonstrates the fundamental organizational structure of a CNN. It includes multiple convolutional and pooling layers that extract and reduce the image features' dimensions, respectively. The fully connected layers at the end use these features to classify the image into different categories.



0	**0**	**1**	**0**	**0**
0	**0**	**1**	**0**	**0**
1	**0**	**0**	**0**	**1**
0	**1**	**0**	**1**	**0**
0	**0**	**1**	**0**	**0**

Fig. 1.(b) Image Representation on a computer

The main issue with the pooling method in CNNs is the loss of valuable features from the input image. As a result, CNNs lack equivalence, meaning that different inputs with similar features may result in different outputs. This lack of invariance is a significant limitation of CNNs, which require substantial amounts of training data and processing time to improve their accuracy [13].

Furthermore, CNNs are susceptible to misclassifying objects when the input pixels are perturbed. Even small perturbations can significantly affect the output of the network, leading to incorrect classifications [14].

B. Capsule Network (CapsNet)

Capsule Network, also known as CapsNet, is a deep learning architecture designed to encode the connections between different entities, such as scales, location, pose, and orientation, to improve object recognition [15]. Unlike traditional Convolutional Neural Networks (CNNs), which may misclassify an image containing mouth, nose, and eyes as a face, CapsNet can effectively identify such images as not being a face by learning the relationships between different features. CapsNet is a type of neural network that can also generate inverted visuals. When detecting an object, CapsNet breaks it down into components and creates a hierarchical relationship between all the components to represent that object.

The implementation of CapsNet includes three major components: input layer, hidden layer, and output layer. The input layer processes the input image, and the hidden layer uses dynamic routing to capture the relationships between different features. Finally, the output layer produces a vector that represents the likelihood of the input image belonging to a particular class [16].

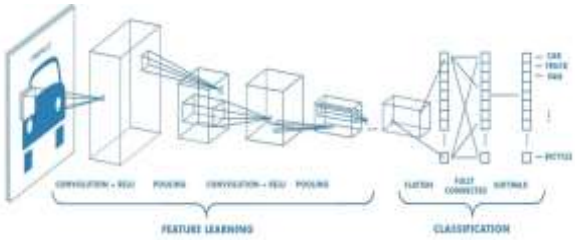
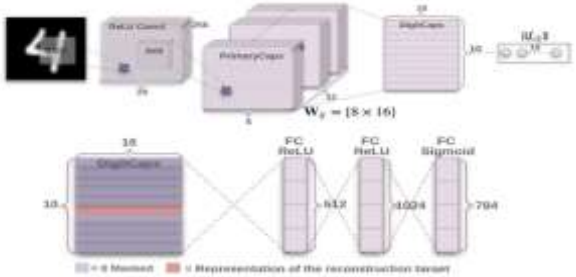


Fig .2. Basic Structure of CNN [15]

The Capsule Network was developed by Sabour and Hinton in 2017, with the goal of improving object recognition. The network includes two convolutional layers, with the first layer consisting of 256 channels and 99 filters. The stride for this layer is 1, and it uses the RELU activation function.

The second layer of the Capsule Network is a convolutional layer with 6632 capsules, and it uses a stride of 2. Each major capsule in this layer consists of 8 convolutional units and uses a 99 kernel. The squashing function is employed as the activation function in this layer.

The last layer of the Capsule Network is called the DigitCaps layer, which is a fully connected layer made up of 16D capsules in size 10. These capsules gather information from every capsule in the network and use it to classify data into ten different categories.



DOI: rp-9788770040723.095

Fig.3. Structure of Capsule Network [15]

Figure 3 shows the structure of the Capsule Network, which includes the different layers and capsules used in the network. The unique architecture of the Capsule Network allows it to encode the connections between different features, making it effective in object recognition tasks.

C. Modification in CapsNet

The Capsule Network has been subject to various modifications to enhance its performance beyond its initial implementation. Some researchers have suggested using densely connected convolutional layers instead of the original convolutional layer in CapsNet to generate a more discriminative feature map. However, this can lead to the vanishing gradient problem, which can be addressed by adding a dense connection to create a feature connection between each layer in a feed forward manner. Additionally, CapsNet can be improved by adding more convolutional and fully connected layers to perform better on datasets beyond MNIST. The initial routing technique in CapsNet used the SoftMax algorithm to normalize routing coefficients, but other routing strategies have been proposed to enhance CapsNet's performance against adversarial attacks, including Self-Routing, Expectation-Maximization Routing, Variational Bayes Routing, and Inverted Dot-Product. Despite these improvements, CapsNet is still susceptible to deception like CNN. Other methods, such as representing an entity with a matrix rather than a vector, and coupling CapsNet with advanced Convolution Network modules like skip connections and dense connections, have also been proposed to increase CapsNet's parameters and improve its performance. These novel routing strategies and architectures can enhance the resilience of affine translation in CapsNet.

II. LITERATURE SURVEY

In 2017, Hinton et al. introduced the Capsule Network (CapsNet) as a solution to overcome the limitations of CNNs. The CapsNet not only extracts and learns information about visual features, but it also learns how these features relate to one another, resulting in a more accurate model. Tests were performed on the MNIST digit dataset, and the CapsNet model outperformed the latest CNN models.

In [17], the authors proposed Failure to detect gastrointestinal (GI) tract diseases early can have severe consequences, including the development of cancer and even death. Traditional procedures for detecting these diseases are often painful and cannot cover the entire GI tract. Wireless capsule endoscopy (WCE) offers a painless and efficient alternative, but generating a large number of images makes it challenging to identify abnormalities. RAt-CapsNet offers a methodical approach to identifying GI abnormalities in WCE images and has the potential to become a promising diagnostic tool.

Researchers conducted a study that was published in [18], The Multilevel Capsule Weighted Aggregation Network (MCWANet) is proposed in this paper. The MCWANet utilizes a decoupled dynamic filter (DDF), a new multilevel capsule encoding module, and a new capsule sorting pooling (CSPool) method to extract and fuse multilevel and multiscale features, resulting in strong semantic feature representations. Experiments on two challenging datasets, AID and NWPU-RESISC45, demonstrate that the proposed MCWANet performs competitively in RSSC.

In [19], the authors proposed FRCapsNet is a new CapsNet proposed in this paper that aims to deal with the heavy computational burden of traditional routing algorithms. The proposed fast routing algorithm allows low-level capsule information to be sent to all high-level capsules simultaneously, reducing computational costs. Future work aims to connect capsules of different levels in a convolutional way, which would reduce the number of trainable parameters.

The authors suggested this in [20], a novel image classification model named Dense Caps is proposed. The model is based on dense capsule layer connection and hierarchical feature combination, composed of multiple dense capsule blocks. This is the first attempt at capsule-level dense connection, and the paper conducts an in-depth study on the feature capsule redundancy problem.

The authors suggested this in [21]. Botnet detection is the process of identifying botnets in network traffic. This paper proposes the LSTM-Capsule Net model, which combines the k-means routing algorithm with the original dynamic routing algorithm for ablation experiments. The proposed model offers better feature map clustering and generalization ability than the original dynamic routing algorithm. Experiments show that the LSTM-Capsule Net model performs better than comparison models in the DGA domain name recognition task and in multi-classification of the DGA domain name family and recognition of Real-Dataset and Gen-Dataset.

In [22], researchers proposed a capsule network called MIXCAPS, which eliminates the need for fine annotation by using a combination of experts and automatically dividing the dataset using a gating network based on convolution. The model achieved an accuracy of 92.88%, sensitivity of 93.2%, and specificity of 92.3% while being independent of pre-defined hand-shaped traits.The proposed method's unique design enables it to achieve high accuracy in object recognition tasks while eliminating the need for hand-crafted features.

The proposed [23] CapsPhase is trained using the SCSN earthquakes. The input is divided into S-probability, P-probability and noise probability. It is tested using STEAD & Japanese dataset. Here, convolutional, primary capsule and digit capsule are mainly used. Convolutional is used to create feature maps (4 s-three components). Primary capsule is used to produce the combination of feature maps. Dynamic routing is used to keep the spatial relations of the output. Digit capsule layer produces the P-Wave and S-Wave arrival time. Median filler is used here to smooth the output.

The author proposed [24] the method of feature extraction using Self-attention generative capsule network which is optimized with Sunflower optimization algorithm

DOI: rp-9788770040723.095

to overcome the high over fitting problem in previous methods done by others. This method makes use of the NIH chest x-ray image dataset from Kaggle to train and extract the features to detect the lung cancer and other lung related diseases.

The author [25] makes use of the optimized hybrid deep learning model to detect the liver disease called liver cirrhosis. Many modalities for the detection of this disease is done but it lacked the higher detection accuracy. So to overcome that, 1232 MRI images collected from hospital is used to train using two deep learning algorithms, CNN and capsule network (HCNN-CNN) are integrated to detect the liver disease with high accuracy rate.

Numerous alternative methods are utilized in Q-CapsNet to find color and facial form, according to the authors' proposal [26]. Quaternion algorithms are initially paired with capsule network layers. The input from facial color is then converted into capsules using a quaternion routing approach and a convolutional layer (RGB). Quaternion convolutional would incorporate the quaternion matrices in order to extract geometric descriptions and capture the internal dependencies between the color channels. The RGB input color matrix and the length and width input geometry are then provided in the QConvCaps layer, where they are combined into a single output. Quaternion routing algorithms of the QConvCaps Layers and QFCCaps Layers are utilized to combine it into a single prediction.

TABLE I. PERFORMANCE SUMMARY OF SOME CAPSULE NETWORK

Ref. No.	Author Name	Methods	Advantages	Disadvantages	Dataset
17	M.D. Jahin	A Novel CNN	Achieved better classification of images.	Insufficient Dataset to compare the images.	• Kvasir Capsule
18	Chunyuan Wang et.al	A Novel Capsule Network for RSSC name MCWANet	Achieved better Accuracy 95.89 and 92.15%	Further improve its classification	• AID • NWPU-RESISC45
19	R. Zeng and Y. Song	Fast Routing algorithm	Achieved better performance in 71.2% of better classification accuracy.	Reduced the Computational cost	• MNIST • CIFAR10
20	Guangcong Sun et.al	DenseCaps	More accurate and trained models	Data imbalanced in dataset	• MNIST • Fashion-MNIST • CIFAR-10 • SVHN
21	Xiang, C et.al	Capsnet	Accuracy can be increased by varying the number of feature map	Accuracy can be increased by varying the number of feature map	• MNIST
22	Afshar, P., Mohammadi, A. and Plataniotis, K.N	MIXCAPS	Lung nodule malignancy prediction	The MNIST used in this paper is simplistic image and additional experiments needed using complex datasets. CapsGAN have the ability in capture geometric transformations.	• LIDC dataset • IDRI dataset
23	Omar M. Saad, Yangkang Chen	capsule & digit capsule layers	No loss of data about position, location & texture. Robust performance even with background noise.	Cannot be done for continuous data. Training time is more. Low resolution	Network(SCSN), Stansford Earthquake Dataset(STEAD), Japanese Seismic Data
24	N.B. Mahesh Kumar	Self-Attention Generative Adversarial Capsule Network	Reduced high over fitting problem	Classification method is challenging	NIH chest X-ray image dataset
25	H. Shaheen	Optimized hybrid deep learning model	Achieved higher detection accuracy	Classification method is challenging	1232 MRI images from hospital
26	Yu Zhou , Lianghai Jin, Guangzhi Ma and Xiangyang Xu	Quaternion technique is used to process the colour information like different skin colours and illumination variation.	Q-CapsNet achieves higher accuracies than the Capsule Network	Low resolution facial images cannot be recognized	• MMI • Oulu-CASIA • RAF-DB • SFEW

III. PERFORMANCE ANALYSIS

In this section, we compare CapsNet's accuracy, number of parameters, and network speed. We choose the most appropriate hyper-parameters for this comparison. We carried out multiple trials to identify the best set of hyper-parameters.

TABLE II. CAPSNET: ACCURACY OF THE DATASET

Dataset	Accuracy (CapsNet)
MNIST	99.63%
F-MNIST	91.35%
CIFAR-10	71.63%
CIFAR -100	72.65%
SVHN	93.04%

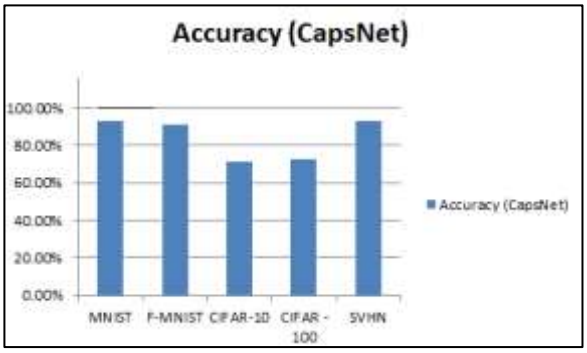


Fig. 4. CapsNet : Accuracy of the Dataset

The performance of the capsule network design is examined using the dataset, an older MNIST dataset, the CIFAR-10 dataset, and iterations. It should increase performance based on the noval capsule architecture. We are currently developing a novel architecture to evaluate multiple types of image classification datasets with high accuracy. The outcomes of the various datasets are shown in the table and figure above.

IV. CONCLUSION

The Capsule Network was introduced as a solution to the limitations of the classic CNN algorithm and has shown great promise so far. However, to fully realize its potential, further research and development are needed. In this study, we evaluated the effectiveness of various algorithms that have an impact on the field of computer vision. We specifically focused on examining how the current capsule network architecture was put into practice and provided more information about its achievements, drawbacks, and possible modifications.

The findings of this research will be beneficial to the computer vision community as they can use the successes and failures of the capsule network to develop a more reliable machine vision algorithm. By considering the structure of the capsule network and exploring ways to improve it, we can work towards achieving more accurate and efficient object recognition and classifications.

REFERENCES

[1] Q.S. Sun, S.G. Zeng, Y. Liu, P.A. Heng and D.S. Xia, "A new method of feature fusion and its application in image recognition," Pattern Recognition, vol. 38, no. 12, pp. 2437-2448, 2005.

[2] M.U. Haq, A. Shahzad, Z, Mahmood, A.A. Shah, N. Muhammad and T. Akram, "Boosting the face recognition performance of ensemble based LDA for pose, non-uniform illuminations, and low-resolution images," KSII Transactions on Internet and Information Systems (TIIS), vol. 13, no. 6, pp. 3144-3164, 2019.

[3] D.L. Pham, C. Xu and J.L. Prince, "Current methods in medical image segmentation," Annual Review of Biomedical Engineering, vol. 2, no. 1, pp. 315-337, 2000.

[4] V. K. Govindan and A.P. Shivaprasad, "Character recognition - a review," Pattern recognition, vol. 23, no. 7, pp. 671-683, 1990, DOI: 10.1109/ICEngTechnol.2017.8308186.

[5] M.K. Patrick, A.F. Adekoya, A.A. Mighty and B.Y. Edward, "Capsule networks – a survey," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 1, pp. 1295-1310, 2022.

[6] A. Krizhevsky, I. Sutskever and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol. 60, no. 6, pp. 84–90, 2012.

[7] 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, pp. 770-778, 2016.

[8] E. Xi, S. Bing and Y. Jin, "Capsule network performance on complex data," pp. 1712.03480, 2017.

[9] C. Xiang, L Zhang, Y. Tang, W. Zou and C. Xu, "MS-CapsNet: A novel multi-scale capsule network," IEEE Signal Processing Letters, vol. 25, no. 12, pp. 1850-185, 2018.

[10] K. Sreelakshmi, S. Akarsh, R. Vinayakumar and K.P. Soman, "Capsule neural networks and visualization for segregation of plastic and non-plastic wastes," In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) IEEE, pp. 631-636, 2019.

[11] Y. Li, K. Fu, H. Sun and X. Sun, "An aircraft detection framework based on reinforcement learning and convolutional neural networks in remote sensing images," Remote Sensing, vol. 10, no. 2, p. 243, 2018.

[12] J. Wu, "Introduction to convolutional neural networks. National Key Lab for Novel Software Technology," Nanjing University, China, vol. 5, no. 23, p. 495, 2017.

[13] C.C.J. Kuo, "Understanding convolutional neural networks with a mathematical model. Journal of Visual Communication and Image Representation," vol. 41, pp. 406-413, 2016, https://doi.org/10.48550/arXiv.1609.04112

[14] A. Shahroudnejad, P. Afshar, K.N. Plataniotis and A. Mohammadi, "Improved explainability of capsule networks: Relevance path by agreement," In 2018 IEEE Global Conference on Signal and Information Processing (Globalsip), IEEE, pp. 549-553, 2018.

[15] J. Su, D.V. Vargas and K. Sakurai, "One pixel attack for fooling deep neural networks," IEEE Transactions on Evolutionary Computation, vol. 23, no. 5, pp. 828-841, 2019.

[16] S. Saha, "A comprehensive guide to convolutional neural networks—the ELI5 way," Towards Data Science, p. 15, 2018.

[17] M.D. Jahin, "CapsCovNet: Rat-CasNet: A deep learning Network utilizing attention and regional information for abnormality detection in wireless," in IEEE Journal of Transactions on Engineering in Health and Medicine, vol. 10, no. 6, p. 3300108, 2022.

[18] Chunyuan Wang et al., "Multilevel Capsule weighted aggregation network based on a decoupled dynamic filter for remote sensing classification", IEEE Access, vol. 9, p. 125309, 2022.

[19] R. Zeng and Y. Song, "A Fast Routing Capsule Network with Improved Dense Blocks," in IEEE Transactions on Industrial Informatics, vol. 18, no. 7, pp. 4383-4392, 2022.

[20] Rajesh, M., & Sitharthan, R. (2022). Image fusion and enhancement based on energy of the pixel using Deep Convolutional Neural Network. Multimedia Tools and Applications, 81(1), 873-885.

[21] C. Xiang, L. Zhang, Y. Tang, W. Zou and C. Xu, "MS-CapsNet: A novel multi-scale capsule network," IEEE Signal Processing Letters, vol .25, no. 12, pp. 1850-1854, 2018, https://doi.org/10.1109/LSP.2018.2873892.

[22] Omar M. Saad and Yangkang Chen, "CapsPhase: Capsule Neural Network for Seismic Phase Classification and Picking," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, p. 5904311, 2022.

[23] Pazhani. A, A. J., Gunasekaran, P., Shanmuganathan, V., Lim, S., Madasamy, K., Manoharan, R., & Verma, A. (2022). Peer–Peer Communication Using Novel Slice Handover Algorithm for 5G Wireless Networks. Journal of Sensor and Actuator Networks, 11(4), 82.

[24] H. Shaheen et al., "An efficient classification of cirrhosis liver disease using hybrid convolutional neural network-capsule network," Biomedical Signal Processing and Control, Elsevier, pp. 1746-8094, 2022.

[25] P. Afshar, F. Naderkhani, A. Oikonomou, M.J. Rafiee, A. Mohammadi and K.N. Plataniotis, "MIXCAPS: A capsule network-based mixture of experts for lung nodule malignancy prediction," Pattern Recognition, vol. 116, p. 107942, 2021.

[26] Yu Zhou et al., "Quaternion Capsule Neural Network With Region Attention For Facial Expression Recognition in Color Images", IEEE Transactions On Emerging Topics In Computational Intelligence, vol. 6, no. 4, pp. 893–912, 2022.