International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Handwritten Bangla Digit Recognition using Capsule Network

Toshiba Kamruzzaman[1], Mashrief Bin Zulfiquer[2]

[1,2]Student, Dept. of Electrical & Computer Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh

---

Abstract - *A capsule is a set of neurons. It stores instantiating parameters of an object such as position, scale, angle of view, rotation deformation, velocity, albedo, hue, texture on in a high dimensional vector space. In this paper, we exploit CapsuleNet for handwritten Bangla numeral recognition. From experiments, we have achieved 99.91% recognition rate on the handwritten Bangla numerical data set NumtaDB which comprise more than 85,000 images of hand-written Bengali digits*

Key Words: **Capsule network, Image recognition, Bangla handwritten digit.**

1. INTRODUCTION

Handwritten digit recognition is among one of the core applications of artificial intelligence in our life, which aimed at postal address interpretation, robotics, number plate recognition and evaluation of handwritten signature. As a branch of image processing, it has acquired the attraction of researchers and many models were proposed in this regard.

Bangla is one among the foremost spoken languages within the world, and is spoken by over 250 million native speakers.[1] It is the national language of Bangladesh, and a outsized number of individuals in India speak in Bangla similarly. Bangla characters will find many applications, and play an important role in Bangla language processing. Due to its demographic assortment, writing patterns of Bangla script comprise of critical shapes and varied sizes. It consists of 50 characters, 10 numerical digits, more than 200 compound characters. So, recognizing Bangla handwritten digits is complex and resilient.

The NumtaDB dataset is one of the largest and diverse dataset consisting of more than 85000 handwritten digits[2].The dataset is a combination of six datasets that were gathered from different sources and at different times and collected from over 2700 contributors containing blurring, noise, rotation, translation, shear, zooming, height/width shift, brightness, contrast, occlusions, and superimposition.

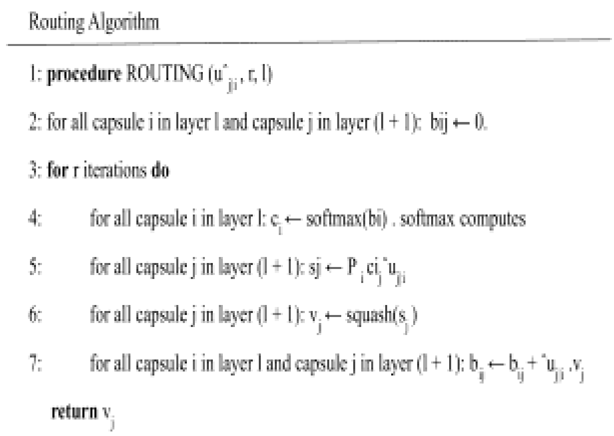
The novelty of this paper comes in deploying a capsule network for training and testing the handwritten digit recognition in the Bangla Language. The rest of the paper is organized as follows: Firstly, the structure of a typical capsule network is described. Then a framework of our system being formed for the Bangla digit identification. After that, experimental arrangements with all prerequisites are discussed. Finally, the paper is terminated by conclusion and cited references.

2. CAPSULE NETWORK (CAPSNET)

2.1 Working principle of Capsule

A capsule network consists of several layers of capsules. The set of capsules in layer L is denoted as Ω_L . Each capsule has a 4x4 pose matrix, M, and an activation probability, a. These are like the activities in a standard neural net: they depend on the current input and are not stored. In between each capsule i in layer L and each capsule j in layer L + 1 is a 4x4 trainable transformation matrix, Wij . These Wij s (and two learned biases per capsule) are the only stored parameters and they are learned discriminatively. The pose matrix of capsule i is transformed by W_{ij} to cast a vote $V_{ij} = M_i W_{ij}$ for the pose matrix of capsule j. The poses and activations of all the capsules in layer L + 1 are calculated by using a non-linear routing procedure which gets as input V_{ij} and a_i for all $i \in \Omega_L$, $j \in \Omega_{L+1.}$ [3] Therefore, the cost to activate the parent capsule willl be-

$cost^h_{ij} = -\ln(P^h_{i|j})$



```
Routing Algorithm
1: procedure ROUTING (û_{j|i}, r, l)
2: for all capsule i in layer l and capsule j in layer (l + 1): bij ← 0.
3: for r iterations do
4:     for all capsule i in layer l: c_i ← softmax(bi) . softmax computes
5:     for all capsule j in layer (l + 1): sj ← P_i cij ^u_{j|i}
6:     for all capsule j in layer (l + 1): v_j ← squash(s_j)
7:     for all capsule i in layer l and capsule j in layer (l + 1): b_ij ← b_ij + ^u_{j|i} .v_j
   return v_j
```

Fig-1: CapsuleNet Algorithm for Bangla Digit Recognition

2.2. Spread Loss of Capsule

Finally, the "spread loss", L, is used to maximize the gap between the activation of the target class at and the activation of the other classes, considering a margin, m, so that if the squared distance between them is smaller than m, the loss of this pair is set to zero [4]

$L_i = \max(0, m(a_t a_i)^2)$; $L \sum_{i \neq t} L_i$

3. Methodology

3.1. Datasets

The data in our study is provided from NumtaDB which contains handwritten Bangla images, which contains 60000 training-data, as well as 20000 test-data samples, in grayscale with 32 □2 bits resolution. For cross - validation,5000 data samples are considered for each epoch. The hardware used for the experiments, was a laptop with CPU 4700MQ, Core i5- 2.2GHz, with 8GB RAM. The training time cost was about 95 ~100 minutes for every epoch.

Fig-2: Sample input digits

3.1. Preprocessing of digits

A. Resizing and Grayscaling

180□80 pixels is the original size of our Dataset which were too large for preprocessing semantically. So we reduce the size of images to 28□8 pixels. Moreover , RGB images were converted to GRAY scale images.

B. Interpolation

We have used inter- LANCZ0S4 interpolation after resizing images.

C. Removing Blur from Images

HB filters were used to deblur our image

D. Sharpening Images

The details of an image can be emphasized by using a high-pass filter. We have used Kirsch Compass Masks

E. Removing Noise from Images

We remove salt and pepper noise from NumtaDB images. We used the Median filter which removes only the noise without disturbing the edges

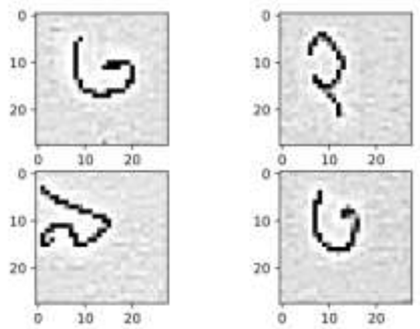


Fig-3: Image after preprocessing

3.2. Regularization

To stop over-fitting, regularization was done. We scaled down reconstruction loss by a factor of 0.0005 so that the margin loss is not influenced.

4. Result evaluation

The proposed model is applied to different datasets and get a pretty best accuracy on train, test and validation sets over other models which is shown below

4.1. Model Performance

The accuracy has reached the highest value of 99.91%, and loss has significantly reduced to 0.0001. the accuracy stopped increasing after 50 epochs. The validation and training accuracy are functional.

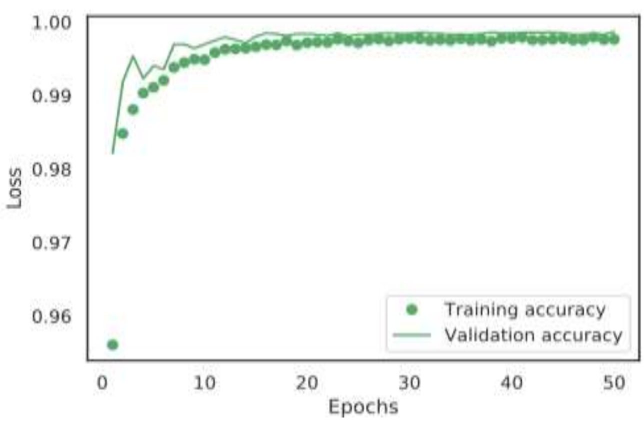


Fig-4: Training and Validation Accuracy per epoch

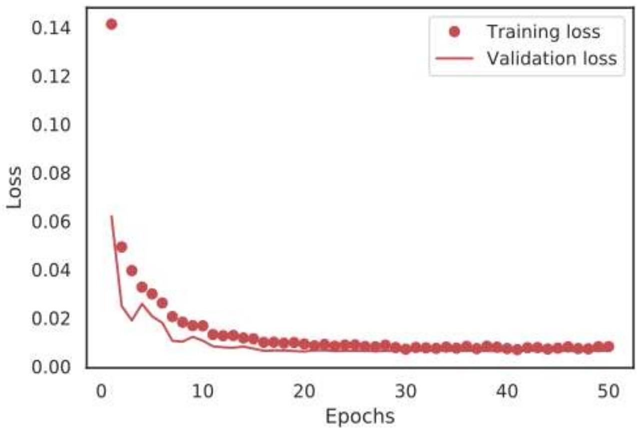


Fig -5: Training and validation loss per epoch

4.2Confusion matrix of the proposed architecture

We have normalized the confusion matrix for clear understanding. From Confusion matrix, we can say that-

Table-1: Overall Statics

Accuracy	0.9991
95% CL	(0.9985-0.9996
Kappa	0.9980
Mcnemar's Test P-Value	NA

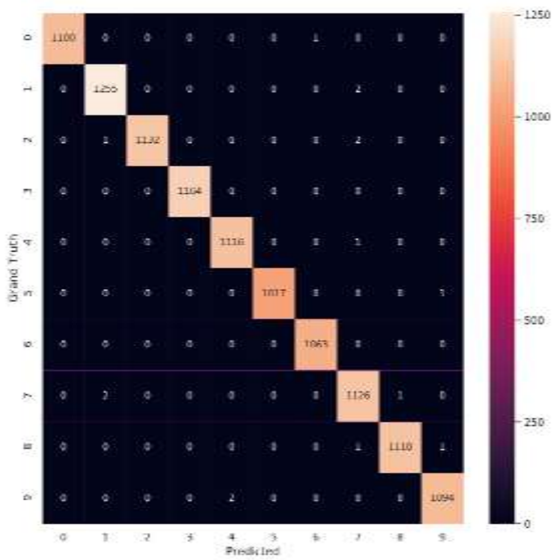


Fig -6: Confusion Matrix

5. Conclusion and future work

In this paper, we have presented a CapsuleNet for recognition of handwritten Bangla digits. The Capsulenet gives excellent results on segmentation tasks and out-performs other models and was very lightweight. For the Bangla digits recognition, we have received 99.91% accuracy which is better than all the other CNN models. Variation was observed in the overall classification accuracy by altering the number of hidden layers and batch size. In future, compound digits will be evaluated and handwritten mathematical signs and equations will be analyzed.

REFERENCES

1. Masica, C. P. (1993). The indo-aryan languages. Cambridge University Press.

2. Alam, S., Reasat, T., Doha, R. M., & Humayun, A. I. (2018). NumtaDB-Assembled Bengali Handwritten Digits. arXiv preprint arXiv:1806.02452.

3. Hinton, G. E., Sabour, S., & Frosst, N. (2018). Matrix capsules with EM routing.

4. Neill, J. O. (2018). Siamese capsule networks. arXiv preprint arXiv:1805.07242.