IEEE TENSYMP 2020
Technology for Impactful Sustainable Development
5-7 June 2020, Dhaka, Bangladesh
IEEE region 10
IEEE Bangladesh Section
5-7 June, Dhaka, Bangladesh

EsharaGAN: An Approach to Generate Disentangle Representation of Sign Language using InfoGAN

Fairuz Shadmani Shishir[1], Tonmoy Hossain[2], Faisal Muhammad Shah[3]
Department of CSE
Ahsanullah University of Science and Technology

IF YOU WISH TO SHOW AFFILIATION LOGOS,

PUT THEM ONLY HERE

Outline

- Introduction
- Motivation
- Background Study
- Problem Statement
- Challenges
- Proposed Methodology
- Result Analysis
- Conclusion

Introduction

- Sign Language is an imperative methods of communicating with deaf people
- Approximately 13 million people in Bangladesh are suffering from irregular degrees of hearing loss[1]
- Bangla Sign Language Controls the Visual–manual modality method
- It is important to automatically classify and generate Bangla sign digits for better communication with the deaf people
- Various types of classification algorithm and neural network model was employed for this purpose

Introduction (contd.)

- Previously, capsule network was developed for classification[2]
- For better classification, a large number is needed for the purpose
- For this purpose, generation model plays extensive role in developing images aligned with the dataset
- EsharaGAN is a Bangla sign digit generation model
- It is based on InfoGAN (Information Maximizing Generative Adversarial Networks)

Motivation

- Well adaptation of automated sign digits classification in the perspective of Bangladesh
- Developing practical application for the deaf people
- On the perspective to the people who are unable speak

Background Study

- Kingma et al. proposed a model based on a stochastic variational inference that scales to large dataset[3]. It is a lower bound estimator that can minimize the training time and marginal log-likelihood

- Evtimova et al. implemented a model that intensifies the mutual properties of InfoGAN[4]. Providing noise to the input data, this model can stabilize its training time. But sometimes adding unreasonable noise to the data can intensify the loss function and generate more noisy and hazy images from the generator.

- Gorijala et al. introduced a model called Variational InfoGAN based on varying the visual descriptions by fixing the latent description[5]

Problem Statement

- EsharaGAN is developed to generate bangla sign digit
- It is developed using InfoGAN (Information Maximizing Generative Adversarial Networks)
- Maximising the mutual information between latent variables and observational variables is the fundamental working principle of InfoGAN

Challenges

- Absence of Bangla Sign Digit generation Model
- Absence of Density modeling in sign digit dataset.
- All the generative models are computationally expensive.

Proposed Methodology

- Proposed model is build on InfoGAN model
- InfoGAN maximizes the mutual information between the subset of the latent variables and the actual observations
- Ensembled with image processing followed by a modified adversarial network
- Basic image pre-processing technique is employed to get better noise-free images and then fed into the proposed network
- The model constituted 13 network layers with a generator and discriminator

Proposed Methodology (contd.)

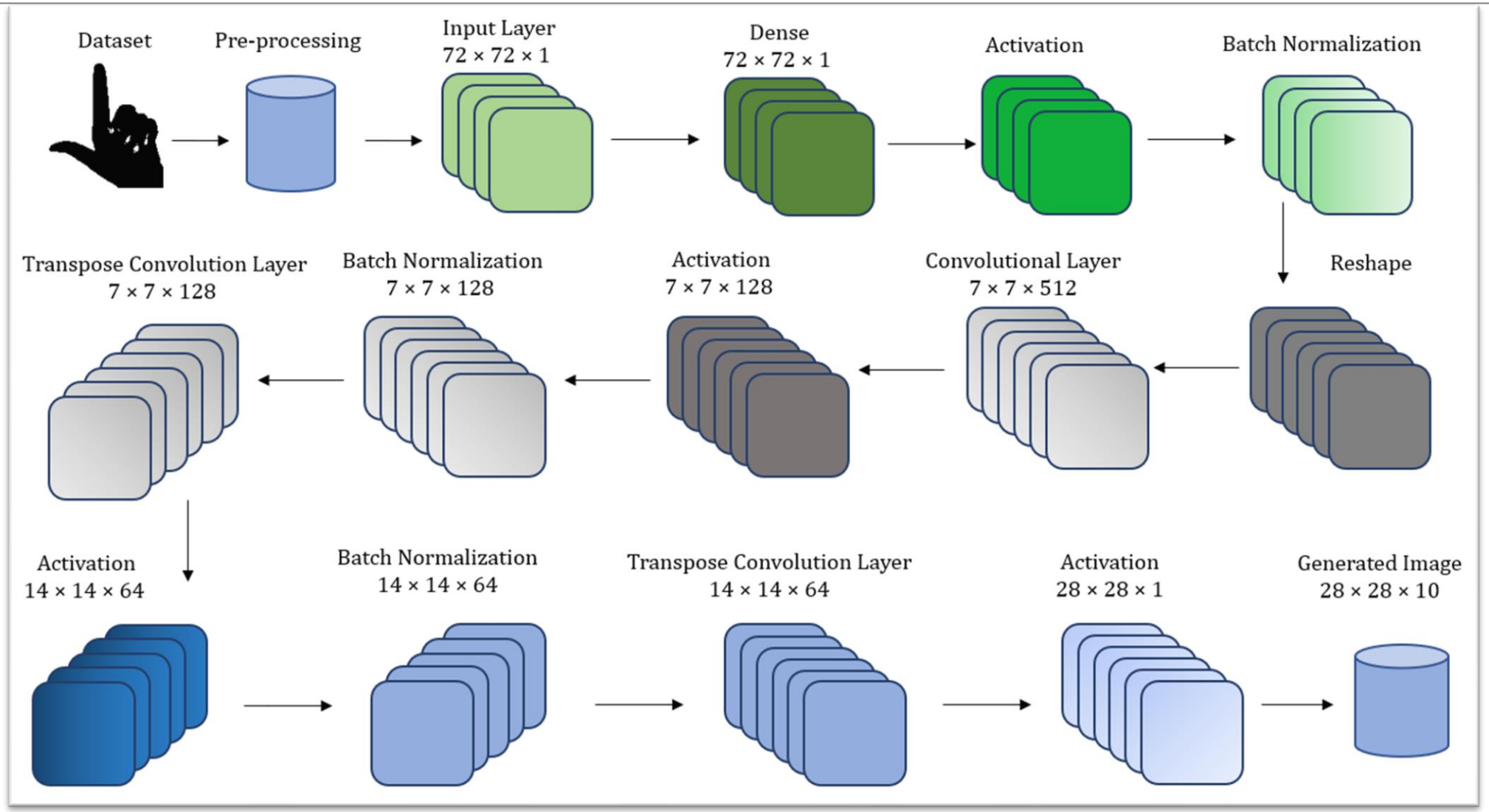

Fig. 1. Proposed Adversarial Network to generate Bangla Sign Digits

Proposed Methodology (contd.)

- IsharaLipi dataset[6] is adopted for the proposed model
- Data augmentation such as—translation, scaling, rotation etc. is done on the input data
- After that, some pre-processing technique is employed to get better input images then fed the images into the network after converting it to a vector
- After executing the 13 layers, we get the generated output images from the proposed network architecture

Fig. 1. Proposed Adversarial Network to generate Bangla Sign Digits

Result Analysis (Dataset)

- We adopted IsharaLipi Dataset for the proposed generative model
- A total of 1000 images partitioning into 10 classes each of 100 images
- All images are converted into 28×28 pixels
- Images are labeled after binarization
- Converted the image pixels into a CSV file

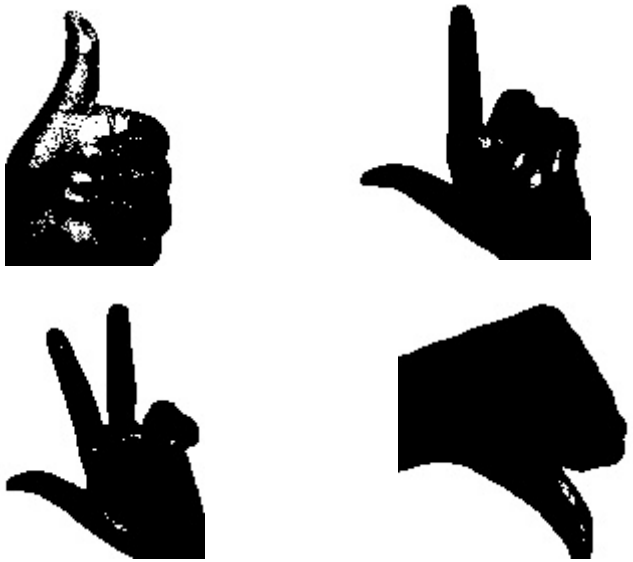

Fig. 2. Some images from the IsharaLipi Dataset

Result Analysis (Hyper-parameter setup)

For training the model, we need to set the hyper-parameter of the model accurately

No	Hyper-parameter	Value
1	Initialization_kernel	glorot_uniform
2	Initialization_bias	Zeros
3	Learning_rate	0.001
4	Optimizer	Adam
5	Batch size	64
6	Epoch	10000
7	Latent Dimension	62

Table I: Hyper-parameter of the proposed architecture

Result Analysis (Loss Function)

For the evaluation of the model, a loss function is generated from the model

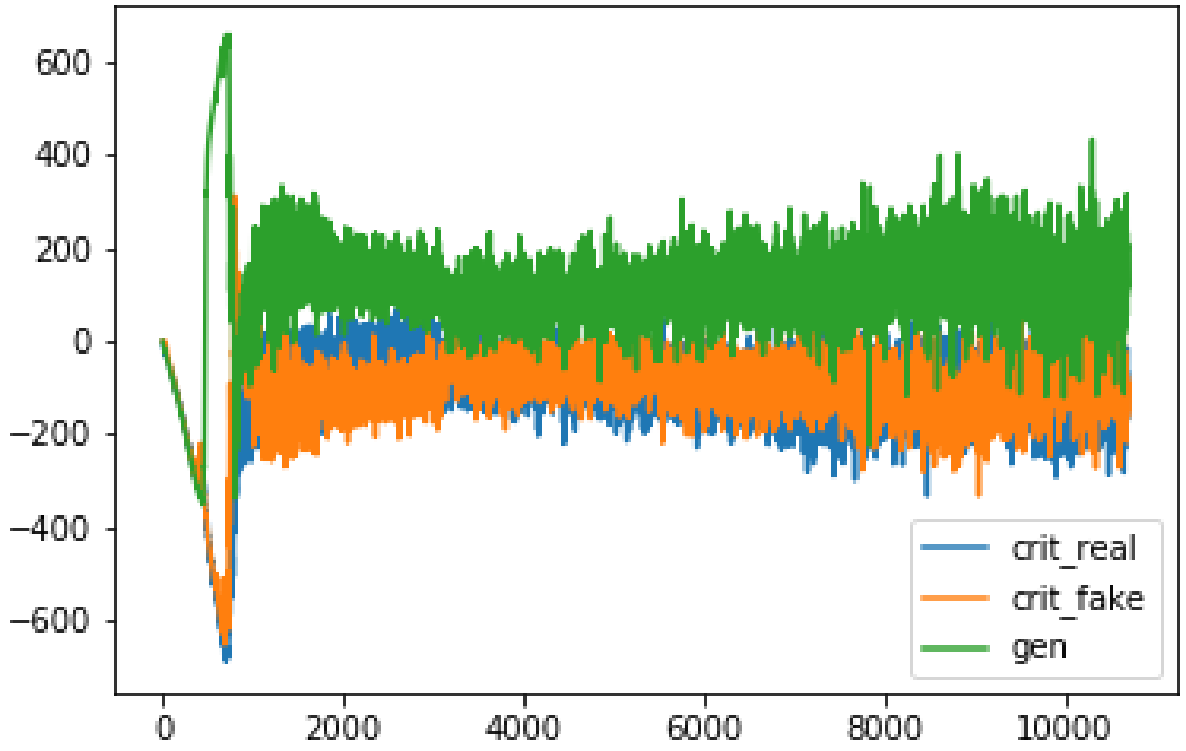

Fig. 3. Loss function curve of the proposed architecture

Result Analysis (Inception Score)

- Inception score (IS) is an objective function to evaluate the generated images
- The inception score of the model is **8.77** for 10 class

Result Analysis (Generated Image)

- Executing the proposed 13 layer architecture, the output from the architecture will be the generated synthetic images

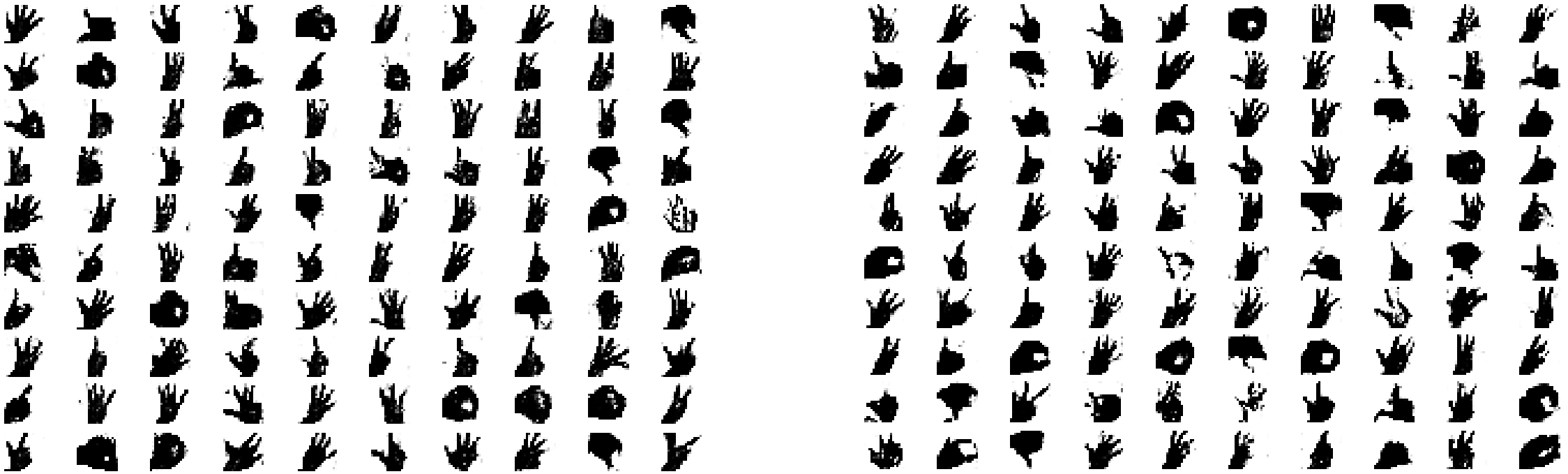

Fig. 4. Generated images from the proposed model

Conclusion

- Here we proposed a novel model to generate Bangla Sign Digit images .
- The model gives us an exceptional result as the inception score is 8.77.
- In the future, we will elaborate on the working procedure of the model on the video field and engage the spatial relationship of the pixels.

Thanks

Thank you

References

[1] Alauddin M., Joarder A.H. (2004) Deafness in Bangladesh. In: Suzuki JI., Kobayashi T., Koga K. (eds) Hearing Impairment. Springer, Tokyo

[2] Hossain, Tonmoy & Shishir, Fairuz & Shah, Faisal. (2019). A Novel Approach to Classify Bangla Sign Digits using Capsule Network. 10.1109/ICCIT48885.2019.9038609.

[3] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

[4] Evtimova, K., & Drozdov, A. (2016). Understanding mutual information and its use in infogan.

[5] Gorijala, M., & Dukkipati, A. (2017). Image generation and editing with variational info generative AdversarialNetworks. arXiv preprint arXiv:1701.04568.

[6] M. Sanzidul Islam, S. Sultana Sharmin Mousumi, N. A. Jessan, A. Shahriar Azad Rabby and S. Akhter Hossain, “Ishara-Lipi: The First Complete MultipurposeOpen Access Dataset of Isolated Characters for Bangla Sign Language.” 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, 2018, pp. 1-4.