ImageSign: Sign Language Images to Text Generation

Sunday D. Ubur *

Abstract

In advancing research works for automating sign language using machine learning, and promoting communication access for sign language users and the general public, this work developed a Convolutional Neural Networks model, with the Sign Language MINST dataset obtained from Kaggle, to generate text from sign language dataset. The experiment result yielded 96% accuracy, and this result was further strengthened using a confusion matrix.

1. Introduction

This project is aimed at understanding image-to-text generation and will attempt to apply the techniques in setting the foundation for research work on sign language-to-text generation. Learning to communicate clearly in writing has never been easier for some traditional sign language users who mainly rely on visual clues to communicate with the world around them.[1] This study intends to extend further the work done[2] in an attempt to compare and understand existing text-to-image applications, and how further work may improve converting sign language-to-text using machine learning, and help sign language users improve communication mastery.

1.1. Aim of the Project:

- To detect sign language images and translate to text

- To get better results compared to past works using CNN with Relu, SoftMax, and Adam functions.

1.2. Scope of the Project:

- The scope of this project is limited to utilizing Sign MINST dataset to translate text to sign language and test the accuracy of the model using test dataset.

2. Related Works

Most research work already done in exploring sign language under machine learning is in translating sign language to speech, text-to-sign language, and video using avatars, hand gesture recognition, and camera. Very little work has been done in translating sign language to text. This work intends to use the popular available Sign Language MINST dataset to translate signs to text. Also, most research works in this domain have employed different machine learning algorithms without focusing on improving a particular algorithm that provides the most accurate and reliable result. There is a need to define a standard algorithm for future work in sign language advancement using machine learning.

In exploring an efficient approach to translating Indian Sign Language using Machine Learning,[3] an automated real-time system that translates English words to Indian Sign Language and vice versa was implemented using Neural Network Classifiers, and Google Speech Recognition API. Our work explores available dataset and does not need to employ hardware for dataset collection.

Using Support Vector Machine algorithm, Jiang, et al[4] developed a real-time vision-based static hand gesture recognition system for sign language. Hand signs data was fed to the system using a USB camera connected to a computer. Two other works that may seem related based on

*Sunday D. Ubur - Graduate Student, Department of Computer Science, College of Engineering, Virginia Tech, Blacksburg, VA, USA. uburs@vt.edu, uburs@github.io

the hardware used include an automatic interpretation using Artificial Neural Networks (ANN) and Leap Motion Controller[5] that employed an optical hand tracking hardware to feed the algorithm with data captured from a person signing words in real-time. Similarly,[6] also used hand gesture recognition and feature extraction and classified the real-time data using K-Nearest Neighbor (KNN) to implement a sign language learning system based on 2D image sampling.

There are other works that employed Convolutional Neural Network (CNN), such as Hypertuned deep CNN for sign language recognition[7] to recognize 24 alphabets obtained from an MNIST sign language database, and a real-time sign language recognition system using learning CNN[8] with the difference being the latter combined CNN with Tensorflow and Keras libraries in Python. Another model employed CNN and an android application to capture real-time sign language gestures for testing the model that has been trained using sign language dataset obtained from Kaggle. This work also used the sign language MINST dataset to further work in this domain. Our work is different by varying the hyperparameters and using a different number of epochs to obtain our result.

The most recent works in text-to-image generation, which inspired this work in translating sign language to text, used Generative Adversarial Networks (GANs). One such popular work generates images from text using transformers.[9] Relatedly, Text2Sign,[10] is another unique work, which in addition to applying GANs, used Neural Machine Translation (NMT) to produce sign language. GANs are known to be difficult to train due to the unstable nature of the training process and are sensitive to the choice of hyper-parameters,[11] and used the PHOENIX14T sign language translation dataset designed for weather forecast. Camgoz, et al[12] also used the PHOENIX14T dataset for their model of transformer-based joint end-to-end architecture for sign language recognition and translation.

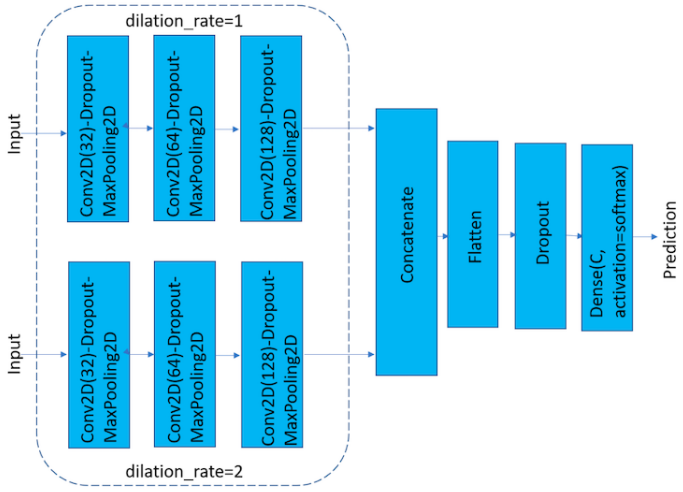
The work that is most related to what we are trying to do was a sign language translation that is bi-directional.[13] To translate text to sign language and sign-to-text, an online translation system that is commonly used to translate languages was used. Similarly, to translate sign language to text, data gloves and Microsoft Kinect sensor was used. The scope of this work, however, is limited to utilizing available datasets to translate sign language to text. This methodology choice is due to the limited timeframe allocated for this work as a semester course assignment, including having no access to the required hardware needed for capturing real-time sign language to generate dataset.

3. Dataset Description

This work explores Generative Adversarial Networks to create a CNN model using Sign Language MNIST dataset from Kaggle,[14] to generate alphabet texts from sign language. The dataset has 27455 train sets, and 7172 test sets of 785 columns respectively. We normalized the dataset, then allocated 20 percent for testing. This left us with 21964 samples for training, and 5491 samples for testing. Our labels consist of 0-25 alphabets (A-Z), of which letters J and Z were missing from the original dataset.

4. Our CNN Model

The model we used is a fully connected CNN layers with features like SoftMax, Relu, and Adam.

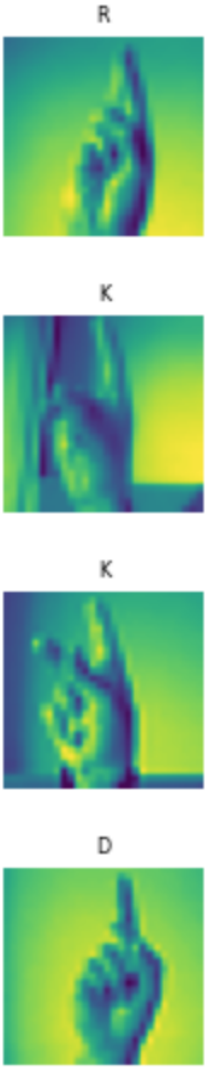


Our model used channels 32, 64, 128, and 512 to maximize and create compact images, and with regularized dropout rate 0f 0.25, dense: 25 softmax, and dataset reshaped into (28, 28, 1).

This gave us the following trainable parameters:

Layer (type)	Output Shape	Para
conv2d_15 (Conv2D)	(None, 26, 26, 32)	320
max_pooling2d_15 (MaxPooling2D)	(None, 13, 13, 32)	0
dropout_20 (Dropout)	(None, 13, 13, 32)	0
conv2d_16 (Conv2D)	(None, 11, 11, 64)	1849
max_pooling2d_16 (MaxPooling2D)	(None, 5, 5, 64)	0
dropout_21 (Dropout)	(None, 5, 5, 64)	0
conv2d_17 (Conv2D)	(None, 3, 3, 128)	7385
max_pooling2d_17 (MaxPooling2D)	(None, 1, 1, 128)	0
dropout_22 (Dropout)	(None, 1, 1, 128)	0
flatten_5 (Flatten)	(None, 128)	0
dense_10 (Dense)	(None, 512)	6604
dropout_23 (Dropout)	(None, 512)	0
dense_11 (Dense)	(None, 25)	1282

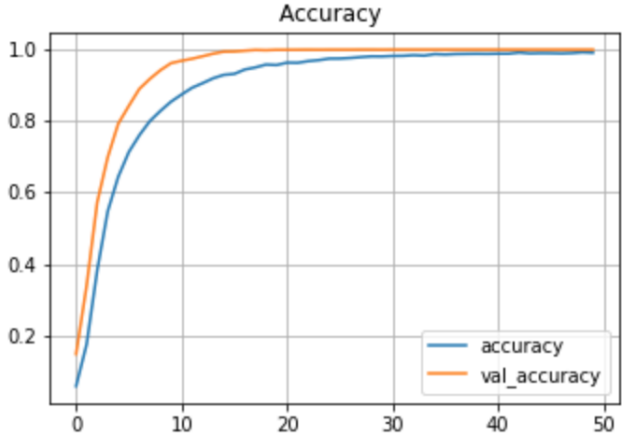
Total params: 171,545
Trainable params: 171,545
Non-trainable params: 0



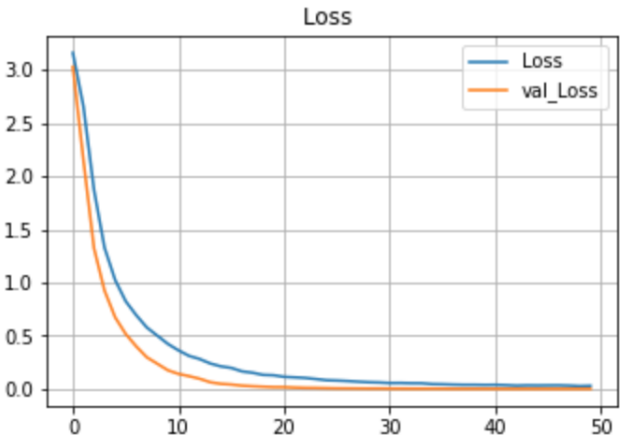
5. Results

Results obtained from the CNN have given an appreciable performance in the classification and generation of sign language images from the alphabet labels. The average accuracy score of the model is around 96%, which can be further improved by turning the hyperparameter values. Considering the importance of improving training with more data, if the training is done with more than 50 epochs, the accuracy can be higher than 96%, and a loss of approximately 4%.

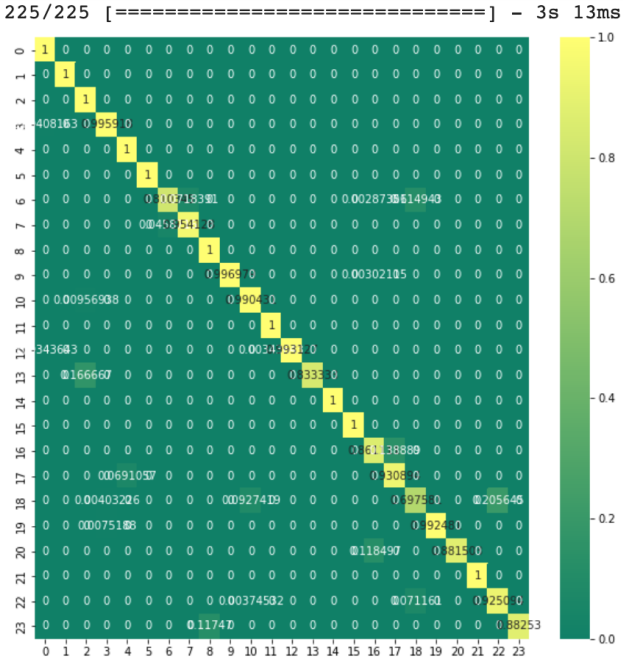
Analyzing the model, we obtained the following training and validation results:



Below are some predicted results obtained:



Finally, we implemented a normalized confusion matrix to explore some cases missed by the model, as shown below.



The result shows that our CNN model has given 100% accuracy in class label prediction for 10 classes, and the least predicted class is about 88% accuracy.

6. Conclusion, Challenge and Future Work

This work used the Convolutional Neural Networks algorithm to develop a model for translating sign language datasets to text. Our result is 96% accurate and the outcome shows that machine learning algorithms from deep learning will substantially accelerate the automation of sign language and improve communication between the hearing impaired and the general public.

The main challenge faced in this work is the lack of access to required hardware tools for capturing and translating real-time signed data and the limited time needed to perform advanced research in this domain. Another major challenge observed with CNN research with sign language is the limited availability of datasets in comparison to other research domains. Further study in this niche could be the production of a video-based sign language public dataset to support motion-based text-to-sign language generation.

Acknowledgement

This work acknowledges Dr. Ismini Lourentzou, Professor in the Advanced Machine Learning course, for inspiring us to undertake a wide range of research in the Machine Learning domain.

References

[1] J. G. Kyle, J. Kyle, B. Woll, G. Pullen, and F. Maddix, *Sign language: The study of deaf people and their language*. Cambridge university press, 1988.

[2] A. Borji, "Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2," *arXiv preprint arXiv:2210.00586*, 2022.

[3] S. Dhivyasri, K. H. KB, M. Akash, M. Sona, S. Divyapriya, and V. Krishnaveni, "An efficient approach for interpretation of indian sign language using machine learning," in *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pp. 130–133, IEEE, 2021.

[4] X. Jiang and W. Ahmad, "Hand gesture detection based real-time american sign language letters recognition using support vector machine," in *2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pp. 380–385, IEEE, 2019.

[5] J. Jenkins and S. Rashad, "An innovative method for automatic american sign language interpretation using machine learning and leap

motion controller," in *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0633–0638, IEEE, 2021.

[6] V. More, S. Sangamnerkar, V. Thakare, D. Mane, and R. Dolas, "Sign language recognition using image processing," *JournalNX*, pp. 85–87, 2021.

[7] A. Mannan, A. Abbasi, A. R. Javed, A. Ahsan, T. R. Gadekallu, and Q. Xin, "Hypertuned deep convolutional neural network for sign language recognition," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[8] M. Taskiran, M. Killioglu, and N. Kahraman, "A real-time system for recognition of american sign language by using deep learning," in *2018 41st international conference on telecommunications and signal processing (TSP)*, pp. 1–5, IEEE, 2018.

[9] M. Ding, W. Zheng, W. Hong, and J. Tang, "Cogview2: Faster and better text-to-image generation via hierarchical transformers," *arXiv preprint arXiv:2204.14217*, 2022.

[10] S. Stoll, N. C. Camgoz, S. Hadfield, and R. Bowden, "Text2sign: towards sign language production using neural machine translation and generative adversarial networks," *International Journal of Computer Vision*, vol. 128, no. 4, pp. 891–908, 2020.

[11] H. Ahn, T. Ha, Y. Choi, H. Yoo, and S. Oh, "Text2action: Generative adversarial synthesis from language to action," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5915–5920, IEEE, 2018.

[12] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10023–10033, 2020.

[13] T. Oliveira, P. Escudeiro, N. Escudeiro, E. Rocha, and F. M. Barbosa, "Automatic sign language translation to improve communication," in *2019 IEEE Global Engineering Education Conference (EDUCON)*, pp. 937–942, IEEE, 2019.

[14] "Sign language minst dataset. https://www.kaggle.com/datasets/datamunge/sign-language-mnist," Nov 20, 2022.