

Capstone Proposal

Shadman Ansari

AWS Machine Learning Engineer Nanodegree

American Sign Language hand gesture recognition using CNN Transfer learning

Domain Background:

Sign languages are an important means of communication for individuals with hearing impairments or speech disabilities. However, access to sign language education and resources can be limited, especially for people with special needs. This project aims to use computer vision technology to make sign language education more accessible and inclusive for people with special needs.

The project focuses on developing a Convolutional Neural Network (CNN) model for recognizing sign language gestures. The model will be trained using a pre-trained model as a base, which will be fine-tuned to better recognize the specific sign language being used. The use of a pre-trained model will minimize the amount of data required to train the model and make the process more efficient.

The developed model will be made available online for individuals with special needs to use for learning sign language. The user-friendly interface and real-time gesture recognition will provide an interactive and engaging learning experience, making sign language education more accessible and inclusive.

This project has the potential to improve the lives of individuals with special needs by empowering them to communicate effectively using sign language. By leveraging cutting-edge technology, this project aims to bridge the gap in access to sign language education and promote inclusivity for people with special needs.

Problem Statement:

Individuals who are specially enabled often face difficulties in accessing resources for learning sign languages, particularly American Sign Language (ASL) alphabets. This can pose challenges to their ability to communicate effectively and limit their full participation in society.

The aim of this project is to develop a computer vision model that can accurately recognize ASL alphabets, A through Z, to make sign language education more accessible to these individuals. The model will be based on a pre-trained ResNet architecture and fine-tuned to recognize the 29 classes of ASL alphabets. The challenge is to fine-tune the pre-trained model to accurately recognize the subtle differences between ASL alphabets and minimize the error rate in their recognition.

The goal is to create a model with high accuracy and minimal error in recognizing ASL alphabets, which can be made available online for individuals who are specially enabled to use for learning sign language. This model has the potential to play a crucial role in promoting inclusivity and access to sign language education for all.

Datasets and Inputs:

Data is collected from [Kaggle](https://www.kaggle.com/datasets/ariyadilshams/american-sign-language-alphabet). It has two folders – Train and Test, each folder has 29 classes – A through Z, Del, Space, nothing. Each training class has 2400 images, and each testing class has 600 images.



Solution:

Data Preprocessing:

- 1- Cleaning the image
- 2- Cropping the image size to (240, 240, 3) to feed them as input to the Convolutional Neural Network
- 3- Applied Normalization techniques as we have scaled pixel values in the range 0 -1

Benchmark Model:

A convolutional Neural Network (CNN) model will be used in this problem. CNN is a robust algorithm for images and video processing. It is currently one of the best algorithms for the automated processing images. Application involving object detection, image recognition, image segmentation, etc, are some of the tasks of CNN models

Convolutional Neural Network process data with a grid structure [6, 7]. CNN consists of three-layer: Convolutional Layer, Pooling layer, and Fully Connected Layer. The convolutional layer is the most essential and the primary layer in the CNN architecture. Next, to reduce the dimensionality of the feature map, the pooling layer comes into play, then outputs from the final pooling layer or the convolutional layer are fed as inputs to the full connected layer after flattening.

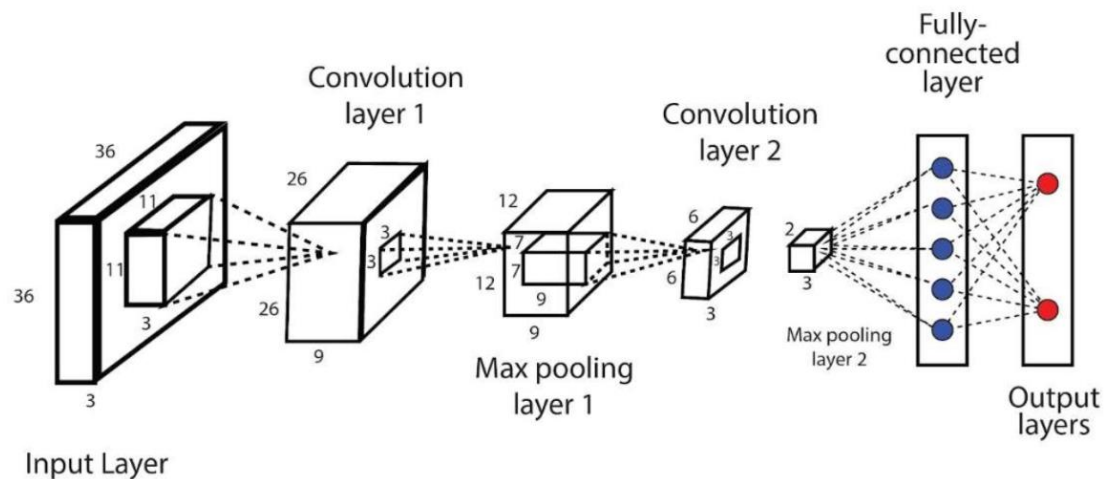
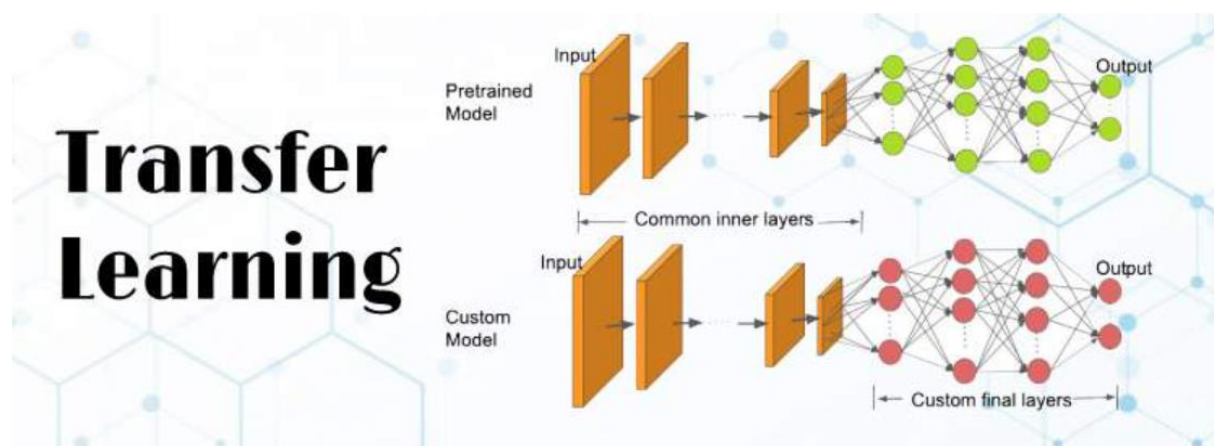


Fig. 3. CNN Architecture [3]

We have used Transfer learning using Resnet-50 Model, Transfer learning is a process to utilize a pre-trained model and only training the outer layer to save computational costs, and time.



Evaluation Metrics:

Average test loss

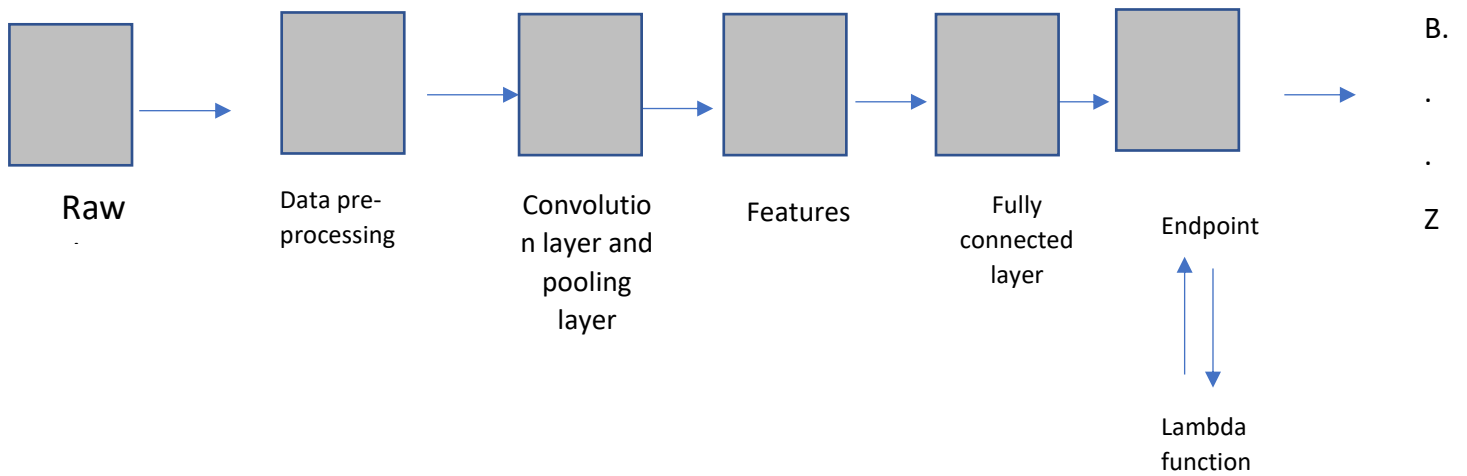
Loss is a value that represents the summation of errors in our model. **It measures how well (or bad) our model is doing.** If the errors are high, the loss will be high, which means that the model does not do a good job. Otherwise, the lower it is, the better our model works.

Project Design:

A Convolution Neural Network (CNN) model will be trained using Resnet-50 Pretrained model, hyperparameter tuning job will be trained to determine hyperparameters. The Best hyperparameters will be used to train an estimator and deploy the model.

Then A lambda function will be used to invoke the endpoint and print the output.

Proposed model:



References

- [1] Accuracy, Precision, and Recall in Deep Learning | Paperspace Blog. (2022). Retrieved 2 July 2022, from <https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/#:~:text=Accuracy%20is%20a%20metric%20that,the%20total%20number%20of%20predictions>
- [2] AUC-ROC Curve in Machine Learning Clearly Explained - Analytics Vidhya. (2022). Retrieved 1 July 2022, from <https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/>