Sign Language Detection Using Artificial Intelligence

Mithil Salunkhe
10B

Parth Kapuskari
10B

Ronit Subhedar
10B

September 1, 2022

1 Introduction

1.1 Need for Sign Language Detection Using Artificial Intelligence

More than 70 million people in the world suffer from complete hearing loss. The Only way for these people to communicate in there day-to-day life is by using sign language. Unfortunately, Only 1 percent of the world population knows how to communicate with sign language. Many people simply do not have the time to learn how to communicate with sign language. This makes it difficult for deaf people to communicate, making the goal of inclusivity hard to achieve.

1.2 A novel way to solve the problem

Artificial Intelligence has made huge strides in recent years. From Cancer Detection to Generating Art, Artificial Intelligence is perhaps our best way to save time of people and achieve the goal of inclusivity. We Propose a novel end-to-end solution for making this a reality which involves the following steps *

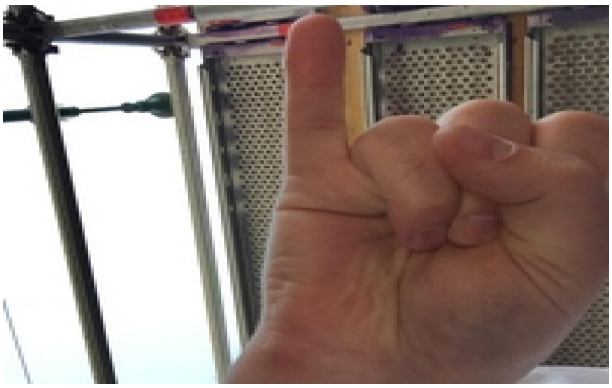
- Capturing the input from the Camera †
- Segmentation of the hand from the background ‡
- Classification of the hand gesture §

*Further Description Of the steps is given in Approach
†Capturing the input from the Camera
‡Segmentation of the hand from the background
§Classification of the hand

2 Approach

2.1 Capturing the input from the Camera

We First get the image from a camera , which is then resized and done augmentations on. We use a package called opencv to do this, helping us capture live camera feed in frames which are then feed to a segmentation model.

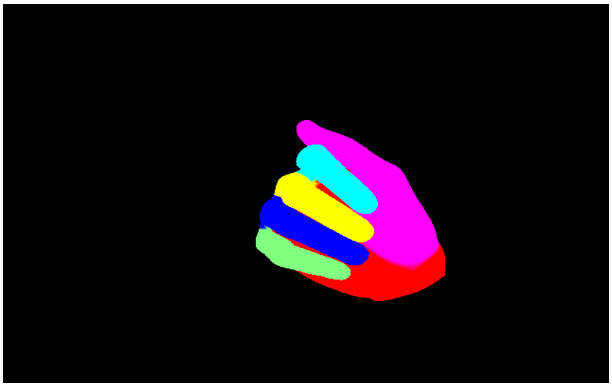


(1) Example of a raw image which is captured in the camera

2.2 Segmentation of the hand from the background

Segmentation is the process in the which objects are seperated from the background, which in our case here is being used to improve accuracy of the model and for generalizing. Segmentation here is done using an Unet model.The model consists of a contracting path to capture context and a symmetric expanding path that enables precise localization ie an encoder decoder architecture. We Get the model from segmentation-models-pytorch, a library consisting of many unet like models and backbones. The removal of the background helps the model not over fit , helping the model to work irrespective of the background.

(1) Example Image



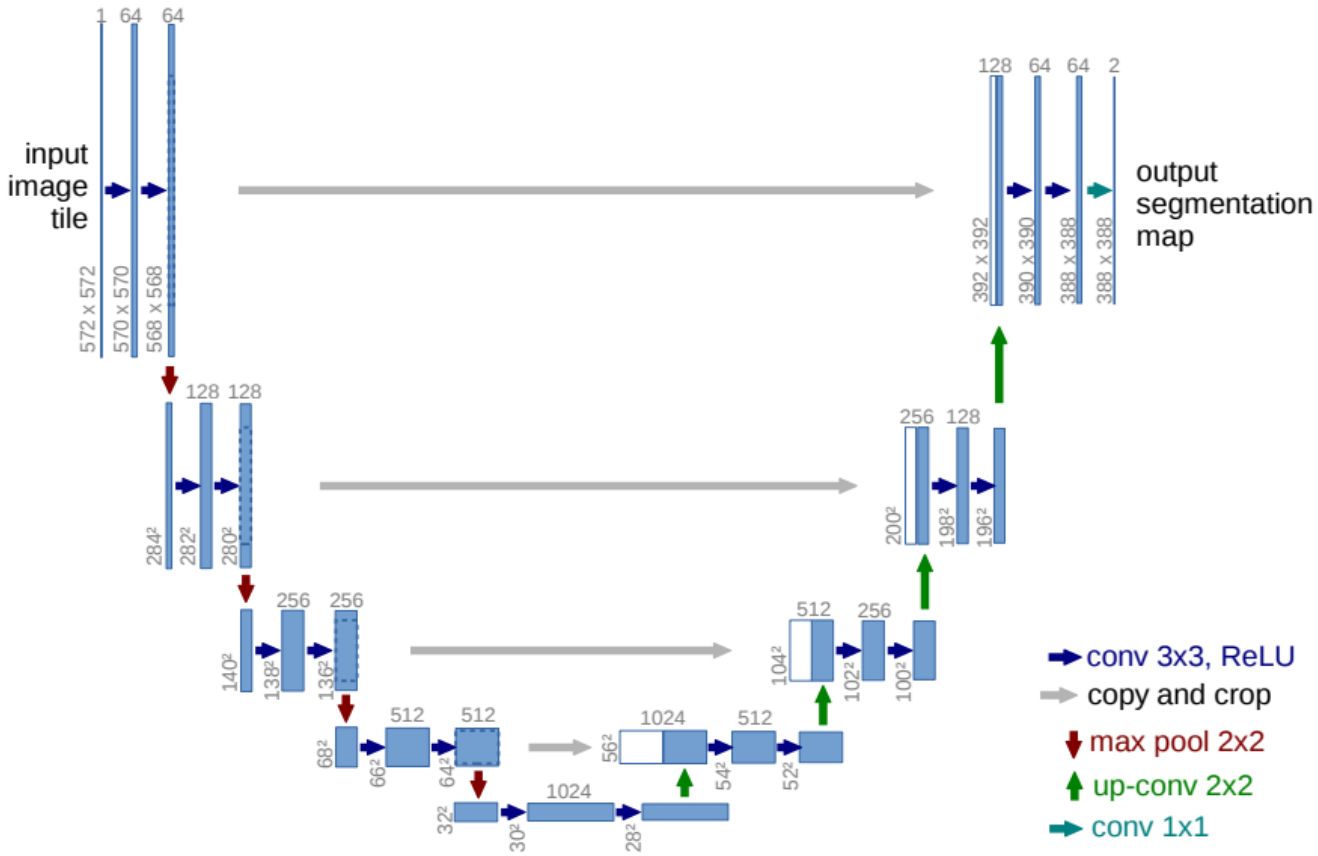
(2) Example of a segmentation mask generated from an image

2.3 Classification of the hand

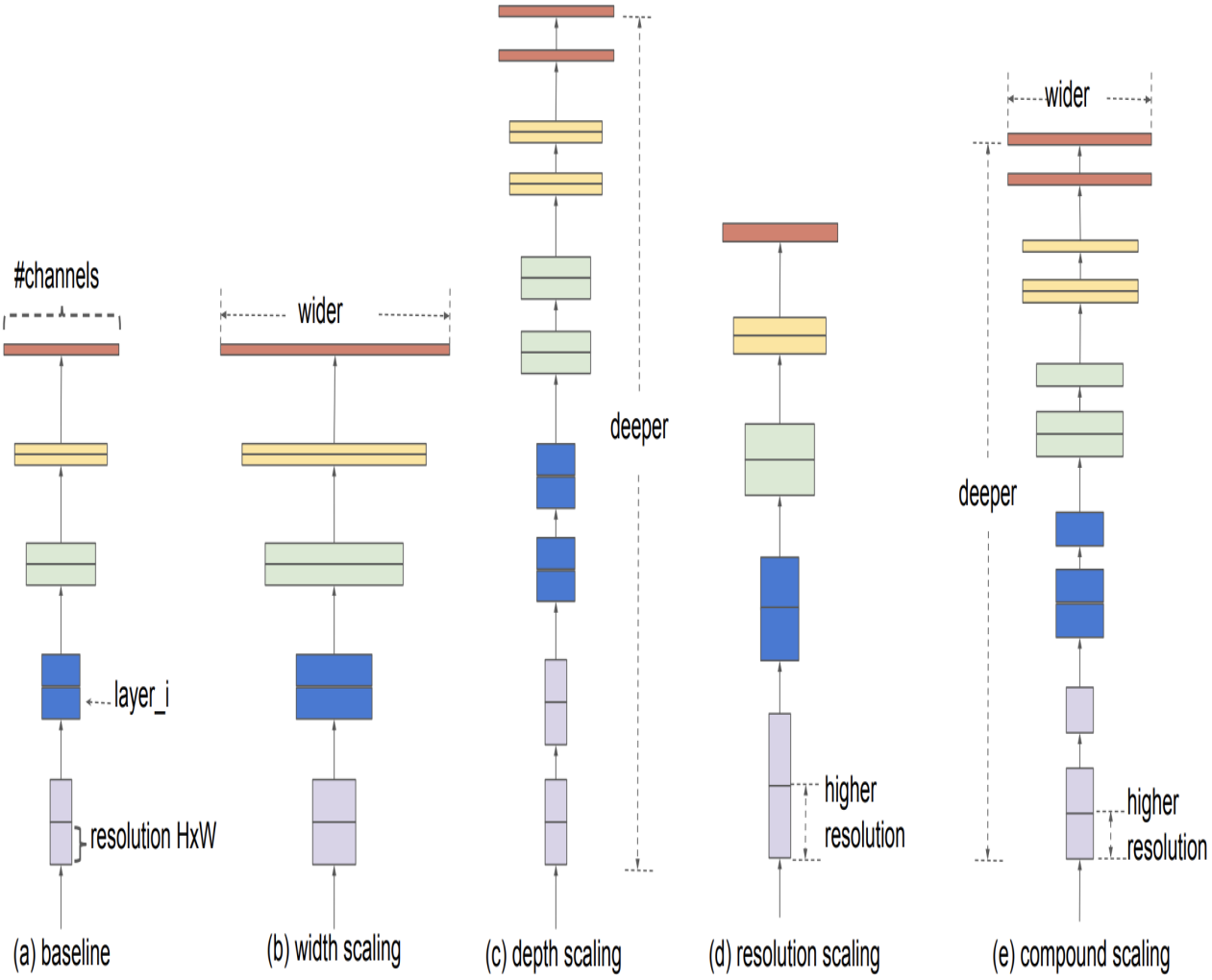
Classification is the main part of our end to end sign language detection pipeline.It is a non-trivial problem in this context,because of 29 classes present in the dataset.Our task is made easier by the additional step of Segmentation, which helps reduce the chance of model overfitting to the background.But the problems of noisy data and different hand shades is hard one to

solve.To combat we use data augmentations like Random Contrast And Random Brightness.We also use a Convolution Neural Network called efficientnet b0, which gives us both, fast inference speed and highly accuracy and confident result.EfficientNet B0,A recent state-of-the-art model,achieves much better accuracy and efficiency than previous ConvNets. CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns to optimize the filters (or kernels) through automated learning, whereas in traditional algorithms these filters are hand-engineered. This independence from prior knowledge and human intervention in feature extraction is a major advantage.

We achieve 97 percent accuracy on our testing dataset which is considered excellent accuracy for a real word application.



(1) Unet architecture



(1) efficientnet b0 architecture

Important libraries we used in our project are.

3 Appendix

3.1 Libraries used

- Pytorch - It is the most used package in our pipeline. Pytorch is the leading deep learning library used by researchers around the world. Pytorch is selected for its ease of scalability,fast inference time and for its more pythonic approach making it easy to integrate others useful library in our pipeline.

- Segmentation Models Pytorch And TIMM - Both of these library are where we get our models from. Segmentation Models Pytorch is from where we get our Segmen-

tation Models from while TIMM is where we get our classifications models from.

- Opencv - Opencv is a computer vision library which helps us load images for training and helps us get our camera feed

- Miscellaneous - sklearn , albumentations , glob

Made Using LaTeX

3.2 Dataset used

Credits to the dataset we used to train our models.

- *https://www.kaggle.com/datasets/unfrien*
  hand-segmentation : Segmentation Dataset

- *https://www.kaggle.com/datasets/debashis*
  sign-language-aplhabet-dataset : Classification Dataset