iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** II **Month of publication:** February 2022

DOI: https://doi.org/10.22214/ijraset.2022.40442

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

Recognition of Hand Movements for Specially-Abled

Abhishek Upmanyu[1], Arundhati Singh[2], Siddharth Sharma[3]
[1, 2, 3]*Department of Computer Science, Amity School of Engineering and Technology, Noida*

Abstract: ***The procedure through a user can make gestures to provide or deliver information is called Recognizing gestures. In daily life our physical gestures have proved to be an incredible tool of communication. An entire set of physical gesture man contribute to a whole language like that of the sign language. The have the ability to smoothly dispatch information, facts, emotions and feelings. Rightly so a combination of physical behaviour and emotional expression. Broadly speaking, hand gesture can be broken down into two categories: static gestures and dynamic gestures. For the initial type, the gesture the hand makes denotes a sign. For the type that follows which is dynamic gestures, there is a sequence of actions that takes shape of gestures, taking place which communicates certain message. Recognizing gestures can predict the intent of the user via a model without human interference. Recent times has seen tremendous amount of work done on designing, building, and upgrading existing Neural Network model into a different form of structure which carries the original purpose of classification but does so with even more promising accuracy and precision. The EfficientNet Model is one that has recently resurfaced. A model constructed by scaling the parameters of an existing Convolutional Neural Network. In this research, we suggest employing an EfficientNet Convolutional Neural Network to identify static hand motions. The process would involve extracting static hand gestures, converting them to gray scale to lessen the computational cost. Segregating the dataset obtained into training, testing, and validating datasets. Later in the upcoming Major Project, a CNN will be built with EfficientNet, and K-fold cross validation will be used for training and testing. We will later in the upcoming Major Project, compare our findings to earlier research and work done on the architecture of Neural Networks.***

Keywords: Hand Gestures, CNN, EfficientNet, ResNet, disability

I. INTRODUCTION

Humans have no major difficulty in identifying and inferring some kind of meaning from gestures. This only is possible due to the spectacular union we have inbuilt of vision and interaction of synapses in our brain, that are formed and strengthened as the brain develops over the years.

To make a computer duplicate this skill, various challenges must be overcome, such as separating portions of interest in photos from the backdrop and carefully selecting the most appropriate image capturing technology and strategy to address the problem. Computing has evolved and with it the access to newer technologies has been eased for us. Devices like Kinect and Leap Motion are innovation of input device technologies for capturing human gestures. These devices have also successfully found uses in areas like that of Robotics, Computer Graphics, Augmented Reality etc.

As was mentioned, the two categories that gestures fall in, likewise, Gesture Recognition methodologies too usually are decomposed into two categories: Static or Dynamic.

Static Gestures need only the processing of one input image to be fed into the classifier. On the contrary, dynamic gestures process sequence of actions that contribute to a gesture and one that is far more complex in terms of gesture recognition methodologies and computational cost.

In this article, we suggest employing a CNN created with EfficientNet to work with static photos of hand motions.

II. OBJECTIVE AND SCOPE

Gesture recognition is the technique of using a user's motions to communicate information or control a device. Physical gestures are a strong form of communication in everyday life. As in sign languages, a set of physical gestures can make up an entire language. They can effectively transmit a wide range of facts and emotions. This research offers the modest notion that gesture-based input is an extremely useful way for conveying information or controlling devices by recognizing distinct human gestures.

The goal of this paper is to use the EfficientNet Deep Learning model to detect embedded hand gesture patterns in images. EfficientNet is a convolutional neural network design and scaling method that uses a compound coefficient to scale all depth/width/resolution dimensions evenly.

©IJRASET: All Rights are Reserved

The EfficientNet scaling method consistently increases network breadth, depth, and resolution with a set of pre-set scaling coefficients, unlike standard practice, which adjusts these factors erratically. If we want to employ 2N times more computational resources, for example, we may simply increase the network depth by N, the breadth by N, and the picture size by N, where are constant coefficients obtained by a small grid search on the initial small model. EfficientNet equally scales network breadth and depth using a compound coefficient.

If the input image is larger, the network requires more layers to increase the receptive field and more channels to detect more fine-grained patterns on the larger image, according to the compound scaling Approach.

The decision to adopt EfficientNet for this project was based solely on its accuracy as well as complex efficiency as compared to the other CNN models. We can scale up the CNN in several dimensions. If you only scale CNNs in one way (for example, depth only), the gains will quickly deteriorate in comparison to the computing increase required.

The most common way to scale up CNNs is to add more layers or make them deeper. ResNet18 has 18 layers, and the more layers the CNN has, the more "power" it has. EfficientNet Scales up at this point. This was not a trivial improvement, but rather a significant increase in the accuracy of ResNet34, ResNet152, and other similar networks. most contemporary CNNs.

III.LITERATURE STUDY

For the purpose of literature study, there was a need of reviewing previously done work or work related to the field of image classification models and Gesture recognition of static images. Regarding the databases, the following were referred to find literature of relevance.

IEEE (https://explore.ieee.com/)
SpringerLink (https://link.springer.com/)
ScienceDirect (https://www.sciencedirect.com/)
Hindawi (https://www.hindawi.com/)

For the purpose of reviewing literature, we opted a systematic plan of reading them. We went thoroughly through the abstract and probed the keywords to grasp the context of the research papers on the surface. An understanding of the essence of the research was gradually build which aided us in the procedure of inclusion and exclusion. In cases, where the abstract was not able to provide us enough clarity, we further studied the introduction and the conclusion for the purpose of filtering papers not relevant for our study.

DATABASES	CHOSEN
IEEE	5
SpringerLink	7
ScienceDirect	3
Hindawi	4

A. Analysis and Discussion

1) The Paper based on Image classification Algorithm Based on Deep Learning Kernel Function, aimed to bring the solution to the problem of complex function approximation and poor classifier effect. It adopted the process of image classification methods which were based on traditional colors, textures, local features, the SIFT method, which looks up for the position scale and rotation invariant on different spatial scales. But it faced the problem of dimensionality disaster and concerning the efficiency of the computational power which was low. Krizhevsky presented the AlexNet model in 2012 which was later optimized to CNN on inclusion of a deeper structure, ReLU activation function, adopting the dropout method to reduce generalization. It reduced Image Classification Top Error rate from 25.8% to 16.4%. VGG Model reduced it to 7.3%. It used more convolutional layers. For this paper being discussed, there have been good application results but problems such as excessive gradient propagation path and overfitting have been seen. A concept of using deep learning model based on stack sparse coding was introduced. A sparse autoencoder was stacked to minimize the value taken between the 'input signal' and the 'output signal'. It trains one layer in the network each time. This encoder abbreviated as SSAE, has the function of feature extraction only. To achieve data classification, a classifier was added to the last layer of the network. The output of SSAE was used as input to the classifier. Doing this kept the parameters of the layers that were trained, unchanged. The SSAE model becomes simpler and easier to implement. The inclusion of sparse representation in the basic network produces better generalization ability and classification accuracy.

2) The second paper based on "COVID-19 Image Classification using deep features and fractional-order marine predators' algorithm", combined strengths of CNN (using Inception architecture) to extract features and a swarm-based feature selection algorithm to select the most relevant features. It achieved 98.7% 98.2% and 96.6% 99% of classification accuracy and F-score on dataset 1 and dataset 2 of Covid-19 X-Ray datasets. It implemented high performance and saw reduction in computational complexity. Considering how some of the extracted features by CNN might not be sufficient, therefore the use of the marine algorithm was tasked to remove the irrelevant features. The authors combined deep features, an architecture of neural network plus a feature selection algorithm.

3) The third paper based on "Automatically Designing CNN architecture using Genetic Algorithm for Image Classification", unlike the traditional CNN, pooling layers, occasionally fully connected layers, proposed an encoding strategy where they designed a new building block by directly using the skip connection to replace the convolutional layers when forming a CNN. Also, the fully connected layers in the end were discarded.

4) This paper proposed the concept of EfficientNet: Rethinking model scaling for CNN. The authors described how model scaling and identifying that properly keeping the balance amongst the network's depth, width and resolution can pave way for a better performance. The mathematical thought was to keep all dimensions of a Neural Network balanced, by simply scaling each of the dimensions by a constant rate to improve accuracy and efficiency. Unlike arbitrary scaling, their method uses a set of preset scaling coefficients to scale the dimensions of networks equally. Scaling up any dimension of a network enhances accuracy, but the accuracy benefit reduces for larger models, according to the findings. Thus, a proposal of this new compound scaling methods paved way for achieving better results. They used a compound coefficient phi to uniformly scale the network dimension.

Depth: $d= \alpha^{\Phi}$

Width: $w= \beta^{\Phi}$

Resolution: $r= \gamma^{\Phi}$

Alpha, Beta, and Gamma are constants that may be found using a grid search, whereas phi is a numerical value coefficient which must be explicitly specified. The authors applied the scaling method to widely used MobileNets and ResNets architecture to look for improvements. In summary, their compound scaling strategy increases the accuracy of these models when compared to single-dimension scaling methods.

5) The final paper of relevance to our study was centered on designing an automated Medical Diagnosis of COVID-19 though EfficientNet. Their work was built upon the previous paper we discussed. The authors build a CNN and on its principles of EfficientNet were applied to scale the dimensions uniformly. EfficientNet B4 model was used for transfer learning. In terms of binary and multi-class classification, the suggested CNN model utilizing EfficientNet has an 'average recall value' of 99.63 percent and 96.69 percent, respectively.

IV. RESEARCH METHODOLOGY

This section of the report describes the methodology along with materials that are planned to be used.
Section 4.1 presents the dataset of hand gestures images to be used to test, train, and validate the model. The CNN method to be used is discussed in Section 4.2

A. Hand Gesture Dataset

The samples involved to train and test the method to be adopted have been collected from public datasets.
It is important to make sure that the number of samples that are to be analyzed for each class/category of images in the dataset are equal in number. The dataset consists of 10 different hand gestures captured of 10 different subject (5 men and 5 women). It is structured in different folders. Each folder containing images from one subject. In total there are 20000 images, 2000 images per directory/folder belonging to one subject.

B. CNN

We have planned on constructing and using EfficientNet B4 model. Also, for minimizing overfitting by reducing the total number of parameters, a global_average_pool2d layer will be used. We will have a sequence of most likely 3 inner dense layers with ReLU activation functions being the most widely used activation function for classification.
Along with this, to ensure generalization to some extent, dropout layers will be included in the structure. Some percentage (approximately 30%) of dropout rate will be introduced.

In the end, to finish of the structure, one final output dense layer will be attached. Containing 10 output slots/units, each for every class/category combined with SoftMax activation function, added to create the classification system.
All software and libraries that will be made use of are open source. For a functioning model, Google Colab will be used on the GPU runtime, a cost-free software.
EfficientNet models are pre-trained scaled CNN models. Because the foundation of the EfficientNet models are highly effective compound scaling methods, this approach allows scaling up a baseline Convolutional Neural Network and maintains the model's efficiency over pre-existing CNNs (AlexNet, ImageNet, GoogleNet, MobileNet).
EfficientNet has 8 versions of itself ranging from EfficientNetB0 to EfficienyNetB7. Each differing on the grounds of number of parameters from 5.3 million to 66 million. We will make use of either B3 or B4 depending upon the requirement of parameters for our dataset.

C. Validation and Experimental Setup

The model will be validated using K-fold Cross Validation method. Also, a separate Database of samples that have not been used during the training phase to validate the performance of the model will be used.
For keeping track of metrics, its precision, recall and F1-score will be computed concerning each class/category. The average values for all folds will be calculated and displayed.

V. FUTURE SCOPE

Our focus of this Project is to utilize the Convolutional Deep Learning Efficient Net model to detect embedded patterns of hand gestures in an image. Convolutional Neural Network is a specific type of a Neural Network involving layers of convolutions to extract local features from an image. Moving to EfficientNet, scaling a neural network model, and discovering that properly balancing a network's depth, width, and resolution can result in improved performance. Using a compound coefficient and a constant rate, this equally scales all dimensions depth, width, and resolution.
In this segment, we will observe the future scope or aspects of our project. As discussed earlier about the methodology, we can implement the model using mentioned dataset and train our model accordingly. Then with the use of appropriate hardware, we want to switch from static hand gestures to dynamic hand gestures. Acquirement of hardware comes under our one of the limitations. But in the foreseeable future, we wish to take our project to physical environment and test it with real-time subjects and train our model accordingly.

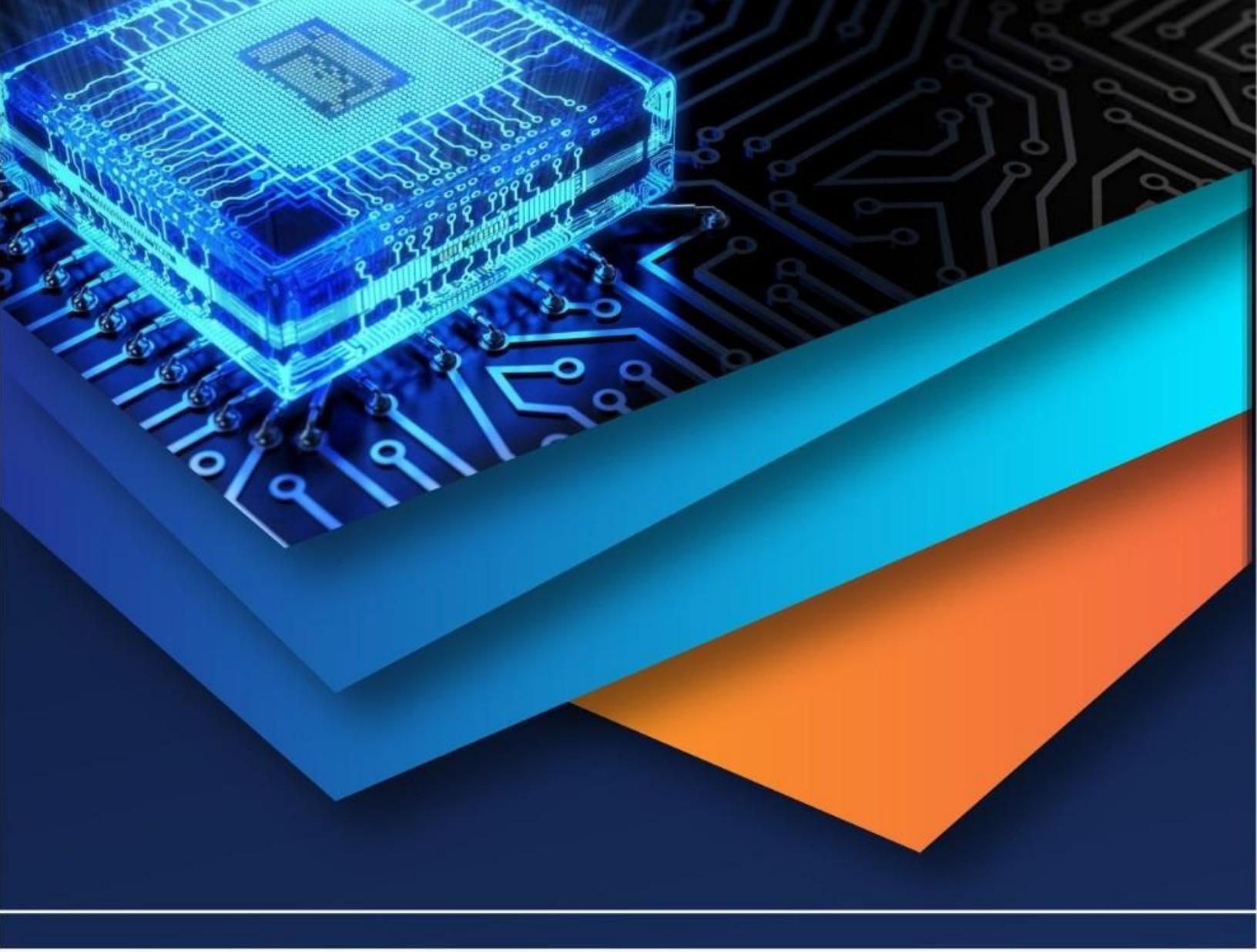
VI. LIMITATIONS AND CONCLUSION

As mentioned in the earlier sections about static profile photos which are being used in the prototype implementation of our project. The images of these subjects are being pulled from LeapGestRecog Dataset. The database consists of 10 different hand-gestures performed different subjects. In other words, there is set of pre-clicked images which are stored in a dataset. Using this dataset, we can link and train our model to work efficiently for this dataset. For a real time, project which should run successfully in current environment, we cannot rely on pre-clicked set of images as it would defy the very purpose of our study along with the idea of feasibility for the users. We will look upon some limitations which hinders the implementation of real-world application in this segment.
In this real-world scenario where our project is at crossroads of physical implementation. To understand the limitations pertaining to our project, we need to understand the real actual barriers that come along with it. First barrier to physical implementation is hardware, second is lack of funding. Without the resolution of these limitations, the physical working in current environment is not possible. We can continue to link our programming to LeapGestRecog dataset and train our model as much as we would like, but it would not work as efficiently and accurately as it should. And on account of the absence of real time data, if it is implemented without training our model to the real-world images, it may direct the users in wrong way which may in turn be harmful. In order to be redirected after being astray, we need to overcome these limitations and make world a better place for the specially-abled.
For our project to be successfully implemented, we would require certain hardware which are hard to come by. This is the first limitation to our project. The software implementation of our project which includes programming can be easily handled on our end. The programming can be carried out on our very own laptops. With the evolving technology, we can use Kaggle, Google Colab, and TensorFlow to carry out the initial phases of development. The project's programming needs to be linked to the hardware so that it can make a real difference and help the user in his/her day-to-day errands. Hardware may include sensors, high quality cameras, digital outlet for user interface and miscellaneous gear. Due to lack of accessibility of hardware, the prospects of the physical implementation of any project are dim.

In addition to accessibility of hardware, comes funding. Top of the end gears come with a heavy pricing. For efficient and accurate working of our prototype, it is necessary to gain hands on the top of the notch hardware that the current technology industry has to offer. If any of the gear is not compatible with one another, then we would fail in the initial stages rendering us of futile efforts. We would need a high pixel camera to be instilled with all the gears for our prototype to work in a well-structured and logical manner. This is the second limitation to our project. If we think of an alternative to this, we can ask for help to universities having access to the required gear. But that would be limited access to resources which would also hinder in our ongoing course. Having our own resources would have a huge impact as we could get to work on our own pace.

REFERENCES

[1] https://ieeexplore.ieee.org/abstract/document/6343787/authors#authors
[2] https://arxiv.org/pdf/1901.10323.pdf
[3] https://www.researchgate.net/publication/284626785_Hand_Gesture_Recognition_A_Literature_Review
[4] https://www.ijsr.net/archive/v3i8/MDcwODE0MDE=.pdf
[5] https://www.researchgate.net/profile/Mehdi-Habibzadeh/publication/324558063_Automatic_white_blood_cell_classification_using_pre-trained_deep_learning_models_ResNet_and_Inception/links/5c5c66f892851c48a9c173a7/Automatic-white-blood-cell-classification-using-pre-trained-deep-learning-models-ResNet-and-Inception.pdf
[6] https://arxiv.org/pdf/2019.14395.pdf
[7] https://ieeexplore.ieee.org/document/8124506
[8] https://www.hindawi.com/journals/sp/2020/7607612/
[9] https://www.nature.com/articles/s41598-020-71294-2
[10] https://ieeexplore.ieee.org/abstract/document/9075201?casa_token=uxq_83ztStsAAAAA:AFlSRr0C5Moxd5rewSPT4cYQUvtPHSNWBRL4jZaCF6FcE7xutQKeYyuZSUndAeCL3hjrjVg
[11] https://ieeexplore.ieee.org/abstract/document/9210789?casa_token=l7eee_7XiBkAAAAA:lh8DgrHQhcDMUYfA3FUXSS90KLxPsQwrU674ZJniXwhZ4t1B6010qaqkx4jkBHso2PSnkYkR
[12] http://proceedings.mlr.press/v97/tan19a/tan19a.pdf
[13] https://www.sciencedirect.com/science/article/pii/S1568494620306293



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)