Detection of diseases from facial features using Convolutional Neural Network

Ayush Kumar - 2017B5A70761P
Project Report

December 02, 2020

Abstract

This project report is based on the Design Oriented Project that I am doing under Prof. Mukesh Kumar Rohil in semester 1 of the academic year 2020-2021. This is based on detection of rare diseases using CNN by analysing different facial features.

1 Introduction

Convolutional Neural Networks(CNNs)[1] have widely been used in visual recognition tasks and has achieved remarkable accuracy in such tasks. Image classification is one such task and the rapid growth of data sizes and computing capability has enabled its use in various domains. The architectures developed had the performance similar to that of humans and the performance has improved ever since.

Technologies in facial analysis have been used to detect the distinct facial features to detect visually observable diseases. One such category of it is rare genetic diseases. These rare diseases affects almost 6-8% of the global population and and out of those almost 40% are genetic. According to the study[2], because of the heterogeneous genetic diverse population of India, rare genetic diseases are one of the major public health concerns. Though these diseases are rare but due to the large population even the rarest of the rare disease can essentially translate to a huge disease burden.

An early detection of such diseases can help people figure out a way for proper prevention and further medical assistance. Since most of the syndromic genetic conditions have distinct facial phenotype, thus it can be used to distinguish them. CNNs can be used as a diagnostic aid as it can be used to classify distinct cranio-facial characteristics by analysing non medical images of people's faces. If not done, then the medical workup to check for the diseases can be quite tedious and costly because of the rarity of the diseases.

Some of the work in this field have focused on binary classification model i.e. classifying affected individuals of one syndrome from the non- affected ones. The task is challenging because of limited availability of dataset and very subtle difference in facial features corresponding to various diseases. Our proposed model is basically a multi-class model which will assign probability to different diseases after deep analysis of different parts of the face individually.

2 Related work

As discussed above, various models have been developed to distinguish affected vs non-affected for a certain disease. [3] is based on the same procedure. After dimension reduction is performed on the dataset using principal component analysis(PCA), k-nearest neighbour(kNN) and Support Vector Machine(SVM) is used for classification of Down syndrome.

In [4] published in 2016, cerrolaza et al. used 2D extension of Linear Discriminant Analysis to extract features to classify positive dysmorphic syndromes from negative ones. Dysmorphic syndromes accounts for almost half of genetic disorders. This model gave an accuracy of 0.95 on a test-set of 145 images which included 15 different genetic disorders. However, this model just classified the image in a binary manner and didn't give probabilistic chances corresponding to different disorders.

Deep Learning Frame-Work for Recognizing Developmental Disorders [5]: This paper was published in 2017 and it proposed a novel framework to determine the developmental disorder. This framework relies on DCNN for global and local feature extraction. This uses facial features to determine from a spectrum of disorders such as down syndrome, fetal alcohol syndrome etc. Preprocessing of the input image was done using [6]. The model was fine tuned on a pre-trained AlexNet model[7].
Analysis of the result were presented in 4 scenarios. In scenario 1 the model achieved an accuracy of 98.80 %. In scenario 2, high Mean Average Precision and Mean Average Recall were achieved for all disorders. In scenario 3, to identify a disorder from a given pool of disorders, the model outperformed manual classification. In scenario 4, model's highest sensitivity i.e. had highest MAR and MAP for age group of 0-6 years.

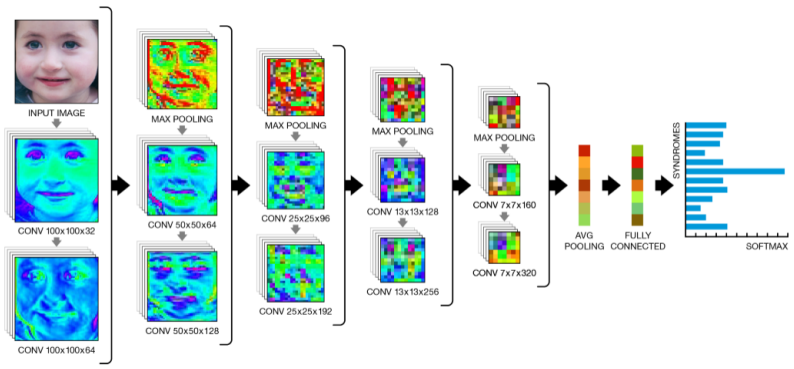

Figure 1: Architecture of DeepGestalt [8]. The network consists of ten convolutional layers and all but the last one are followed by batch normalization and ReLU. After each pair of convolutional layers, a pooling layer is applied (max pooling after the first four pairs and average pooling after the fifth pair). This is then followed by a fully connected layer with dropout (0.5) and a softmax layer.

DeepGestalt-Identifying rare genetic syndromes using deep learning [8]: This is by far the most successful framework for detection of multiple genetic syndrome using Computer Vision and deep learning and has been developed by FDNA for Face2Gene application. This model is trained on a community driven dataset of over 26000 patient cases and has achieved a remarkable 91% top-10 accuracy in identifying over 215 genetic syndromes.

Since the model is operating on uncontrolled real world images, so preprocessing of the image is necessary. This model uses DCNN cascade based face detection proposed in [9]. Then using 130 facial-landmarks, the image is geometrically normalised. Then the aligned image is passed through a region generator which defines multiple regions of interest. The backbone architecture[10] used has been illustrated in Figure 1. Model is trained separately for each crop and combined to form a robust model. For facial recognition task the model was trained on CASIA web-face dataset [10] and to fine tune the network a proprietary dataset was used.
DeepGestalt is a multi-class model and uses aggregation of various facial regions and thus achieved top-5 accuracy of 83.7% and top-10 accuracy of 91.0%. Even the standard deviation corresponding to top-n accuracy showed that model is generalizing well to specific problems and outperformed clinical experts in three separate experiments.

3 Proposed Approach

Our goal is to make an end-to-end model which takes an image of the face as input provides result as probabilities against various diseases for the given face.

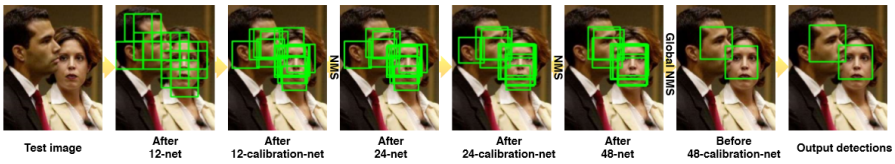

Figure 2: Test pipeline for [9]. As we go from left to right the resolution increases and detection window gets reduced

Preprocessing of the Image: The image that we are getting as input has been taken in real world in an uncontrolled condition. There can be large visual variation between the images because of pose, face size, brightness level, expressions etc. and thus a robust model is needed with powerful discriminative capability as well as computationally high performance. Thus, our model adopts the method proposed in [9], which is a cascade architecture built on CNN.
[9] uses multiple resolutions and thus is able to reject most of the background

in the quick initial low resolution stages of the processing. A CNN based calibration of bounding boxes is done after each detection stage. This model uses 6 CNNs, 3 CNN namely 12-net, 24-net and 48-net for face vs non face binary classification and rest 3 namely 12-calibration-net, 24-calibration-net and 48-calibration-net for bounding box calibration. The overall test pipeline for this approach has been given in Figure 2.

Architecture for syndrome classification: To learn the baseline facial recognition Inception-Resnet-v1[11] is used as backbone architecture. The whole architecture is illustrated in Figure 3. To train the model for facial recognition task, the model is trained on CASIA web-face dataset [10] which contains 453,453 images over 10,575 identities. After the preprocessing of the image, the image is segmented into different facial regions and model is then fine tuned for each facial feature. The syndromic images used to fine-tune the model has been collected from different sources.

A softmax classifier is used on each facial region to make separate prediction for each region and then results are averaged out to make a robust multi-class prediction. The evaluation is done based on top-n accuracy. Top-N accuracy basically measures how often the predicted class falls in top N values of softmax distribution. The results are also evaluated based on confusion matrix to visulaize in what ways the model is making an inaccuracy in classifying the image.

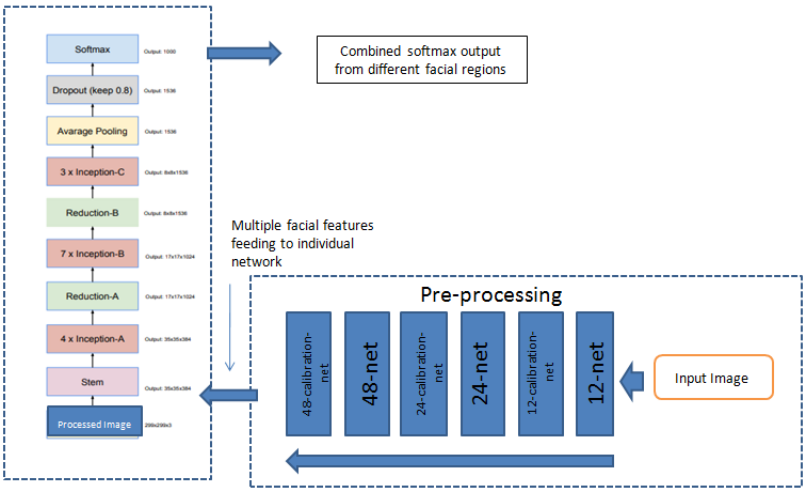

Figure 3: Proposed architecture with the backbone of Inception-Resnt-v1 [11]. The pre-processing has been done similar to that of [9]. The softmax probabilities calculated on different regions has been averaged in the final probability.

4 Experiments & Results

Since we are using Inception-Resnet-v2 architecture, each facial feature should be resized to a resolution of 299*299.
Dataset: The architecture is pre-trained on CASIA web-face dataset and fine tuning has been done using customised datset collected from various sources.
Performance Metrics: Top-N accuracy will be used to analyse the results. Also, results will be evaluated using confusion matrix.

A custom dataset was created to test the architecture. The dataset consist of the following diseases:

a) Down Syndrome [12]: It is a is a genetic disorder caused by the presence of all or part of a third copy of chromosome 21. It is usually associated with physical growth delays, mild to moderate intellectual disability. Down syndrome is one of the most common chromosome abnormalities in humans. It occurs in about 1 in 1,000 babies born each year. In 2015, Down syndrome was present in 5.4 million individuals globally and resulted in 27,000 deaths, down from 43,000 deaths in 1990.

People with Down syndrome may have some or all of these physical characteristics: a small chin, slanted eyes, poor muscle tone, a flat nasal bridge, a single crease of the palm, and a protruding tongue due to a small mouth and relatively large tongue. These airway changes lead to obstructive sleep apnea in around half of those with Down syndrome. Other common features include: a flat and wide face, a short neck, excessive joint flexibility, extra space between big toe and second toe, abnormal patterns on the fingertips and short fingers.

b) Fetal Alcohol Syndrome [13]: Fetal alcohol syndrome is a condition in a child that results from alcohol exposure during the mother's pregnancy. Fetal alcohol syndrome causes brain damage and growth problems.Surveys from the United States found that about 10% of pregnant women drank alcohol in the past month, and 20% to 30% drank at some point during the pregnancy.

The three FAS facial features are: (i) A smooth philtrum: The divot or groove between the nose and upper lip flattens with increased prenatal alcohol exposure. (ii) Thin vermilion: The upper lip thins with increased prenatal alcohol exposure. (iii) Small palpebral fissures: Eye width decreases with increased prenatal alcohol exposure.

c) Cerebral palsy [14]: It is a group of permanent movement disorders that appear in early childhood. Signs and symptoms vary among people and over time. Often, symptoms include poor coordination, stiff muscles, weak muscles, and tremors. There may be problems with sensation, vision, hearing, swallowing, and speaking. Cerebral palsy is the most common movement disorder in children. It occurs in about 2.1 per 1,000 live births.

It causes paralysis or severe weakness of the facial muscles on one side of the face. It is believed to be due to a swelling of the nerve that controls the muscles of the face.

The training data consists of almost 20-25 images of each disease and 150 images of face with no genetic condition. I customised the face recognition algorithm to output the boundary of the face identified in the picture and there

are options for both GPU enabled and non-enabled devices. The original architecture has an accuracy of 99.38% on the Labeled Faces in the Wild benchmark.

The original architecture is using InceptionResNetV2 and I have also coded to run the same dataset on VGG16 also to compare the results. The model uses loss as categorical crossentropy, optimizer as Adam optimizer and metrics as accuracy. However manually confusion matrix is also calculated. All the codes are available on Github, however due to ethical issues the dataset is not public.

The model was trained on Google Colab. The last layer of both VGG and Inception Resnet was removed and replaced with softmax classifier corresponding to the number of class of training dataset. After making the hidden layers non-trainable, the Inception Resnet architecture has 54,729,956 total parameters with 393,220 trainable parameters. The VGG 16 architecture has 14,815,044 total parameters and 100,356 trainable parameters. The model achieved a loss of 0.1617 remarkable accuracy of 0.9575.

Figure 4: Results of the pre-processing of the image. The code successfully outputs the bounding box dimensions in which the face has been identified in the image.

```
Epoch 1/5
tcmalloc: large alloc 3699564544 bytes == 0x3ff34000 @  0x7f41b87521e7 0x7f418fd31e95 0x7f418fd48a85 0x7f4195fa16ca 0x7f419644ebb4 0x7f41
1/7 [===>..........................] - ETA: 0s - loss: 1.6976 - accuracy: 0.2500tcmalloc: large alloc 3699564544 bytes == 0x3ff34000 @  0
2/7 [=======>......................] - ETA: 40s - loss: 1.6515 - accuracy: 0.5000tcmalloc: large alloc 3699564544 bytes == 0x3ff34000 @
3/7 [===========>..................] - ETA: 44s - loss: 2.0858 - accuracy: 0.5521tcmalloc: large alloc 3699564544 bytes == 0x4a498000 @
4/7 [===============>..............] - ETA: 37s - loss: 2.1284 - accuracy: 0.5859tcmalloc: large alloc 3699564544 bytes == 0x4a498000 @
5/7 [===================>..........] - ETA: 26s - loss: 1.9936 - accuracy: 0.6125tcmalloc: large alloc 3699564544 bytes == 0x3ff34000 @
6/7 [=======================>......] - ETA: 13s - loss: 1.8563 - accuracy: 0.6250tcmalloc: large alloc 2312298496 bytes == 0x2c6d4000 @
7/7 [==============================] - 92s 13s/step - loss: 1.7830 - accuracy: 0.6132
Epoch 2/5
tcmalloc: large alloc 3699564544 bytes == 0x4a498000 @  0x7f41b87521e7 0x7f418fd31e95 0x7f418fd48a85 0x7f4195fa16ca 0x7f419644ebb4 0x7f41
1/7 [===>..........................] - ETA: 0s - loss: 0.5971 - accuracy: 0.8125tcmalloc: large alloc 3699564544 bytes == 0x4a498000 @  0
2/7 [=======>......................] - ETA: 40s - loss: 0.6844 - accuracy: 0.7969tcmalloc: large alloc 3699564544 bytes == 0x3ff34000 @
3/7 [===========>..................] - ETA: 43s - loss: 0.8143 - accuracy: 0.7708tcmalloc: large alloc 3699564544 bytes == 0x3ecd4000 @
7/7 [==============================] - 91s 13s/step - loss: 0.8044 - accuracy: 0.7736
Epoch 3/5
7/7 [==============================] - 96s 14s/step - loss: 0.4466 - accuracy: 0.8396
Epoch 4/5
7/7 [==============================] - 91s 13s/step - loss: 0.4316 - accuracy: 0.8726
Epoch 5/5
7/7 [==============================] - 91s 13s/step - loss: 0.1617 - accuracy: 0.9575
```

Figure 5: Results of the final processing of the dataset. An accuracy of 0.95 was obtained after 5 epochs of training.

5 Conclusion and Future work

This model has been developed keeping in mind previous deep learning models has provided the accuracy of diagnosis higher than the authorised clinicians in some study and thus these models can help as a primary diagnostic tool for multiple rare diseases which shows some facial feature. One of the major difficulty faced was collection of dataset as most of the dataset were not public considering ethical reasons. With addition of more data in training and testing a more proper and elaborate result can be given and can help build an even powerful model.

This project has lot of future scopes by testing the data on different architectures and adding additional input such as age and other parameters. However this will require an extensive work in formation on dataset. With model becoming more powerful and generalised it can find commercial use in medical field as well.

References

[1] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[2] Sridhar Sivasubbu and Vinod Scaria. Genomics of rare genetic diseases—experiences from india. *Human genomics*, 13(1):1–18, 2019.

[3] Şafak Saraydemir, Necmi Taşpınar, Osman Eroğul, Hülya Kayserili, and Nuriye Dinçkan. Down syndrome diagnosis based on gabor wavelet transform. *Journal of medical systems*, 36(5):3205–3213, 2012.

[4] Juan J Cerrolaza, Antonio R Porras, Awais Mansoor, Qian Zhao, Marshall Summar, and Marius George Linguraru. Identification of dysmorphic syndromes using landmark-specific local texture descriptors. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1080–1083. IEEE, 2016.

[5] Pushkar Shukla, Tanu Gupta, Aradhya Saini, Priyanka Singh, and Raman Balasubramanian. A deep learning frame-work for recognizing developmental disorders. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 705–714. IEEE, 2017.

[6] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2879–2886. IEEE, 2012.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[8] Yaron Gurovich, Yair Hanani, Omri Bar, Nicole Fleischer, Dekel Gelbman, Lina Basel-Salmon, Peter Krawitz, Susanne B Kamphausen, Martin

Zenker, Lynne M Bird, et al. Deepgestalt-identifying rare genetic syndromes using deep learning. *arXiv preprint arXiv:1801.07637*, 2018.

[9] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. A convolutional neural network cascade for face detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5325–5334, 2015.

[10] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[11] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*, 2016.

[12] Down syndrome, Nov 2020.

[13] Fetal alcohol spectrum disorder, Dec 2020.

[14] Cerebral palsy, Dec 2020.