Visualization of Midline Diastema Fixing

Dr. Kumud Kundu[1], Aman Verma[2], Bulbul Tyagi[3], Prashant Vashisth[4]

[1]Asst Professor, Dept. of Computer Science & Engineering, Inderprastha Engineering College, Uttar Pradesh, India
[2,3,4]Dept. of Computer Science & Engineering, Inderprastha Engineering College, Uttar Pradesh, India

---***---

Abstract - *A lot has been achieved in healthcare industry but still there is some gap between the patient expectation and the healthcare treatment results. For Diastema treatment, like filling of gap between the front teeth, patient desires to have a view of after treatment results before actually undergoing treatment. To fill up this gap, this paper presents the outcomes of orthodontic treatments by utilizing image processing approaches and deep learning. Accurate results are required for facial recognition, age and gender detection to identify teeth area and a teeth mask. This work attempts to utilize image processing techniques to recognize the face area and consecutively region of interest within teeth area. Deep learning approach is utilized to identify correctly the age, gender of patient undergoing treatment and then identifying a teeth mask according to the obtained results.*

***Key Words*: Teeth masking; ResNet, feature classification; image recognition; Facial technology; signal processing; image classification**

1. INTRODUCTION

Diastema [1], often termed as over separation distance between two teeth specifically the gap between two front teeth. This over-separation spacing can occur anywhere in between two teeth. Though this over-separation between two front teeth happens to many people however orthodontic treatment of diastemas can eliminate the spaces. Specifically, for midline diastema, treatment is not typically required for medical reasons but if a person is not happy with the presence of their diastema, then both cosmetic as well as orthodontic treatments can come to rescue. There can be many possible treatments like dental bridging, dental implants or braces, it is possible to close or narrow the gap. Visualization of outcomes of treatment infuses confidence in patient towards the diastema treatment. In this paper we implemented the visualization of outcomes of the midline diastema treatment. This work attempts to utilize deep learning and image classification to recognize face and smile of a person according to the age and gender of the person undergoing treatment. Haar cascade classifier is utilized to detect face and smile across the images of dataset. Cropped images are further processed via ResNet for the estimation of age and gender. The results of ResNet facilitate in applying a perfect teeth mask. The purpose of this work is to develop an artificial intelligent model to help a person to visualize the outcomes of midline diastema. Second section of paper presents related literature survey while section three presents the formulated model based on deep learning pretrained model- ResNet. Section four presents the experimental results of the outcomes of orthodontic treatments. Conclusion and future work are presented in section five. OpenCV uses two types of classifiers, LBP (Local Binary Pattern) and Haar Cascades. We using the latter classifier.

1.1 Understanding Haar Cascades

A *Haar Cascade* is based on "Haar Wavelets" which Wikipedia defines as:

A sequence of rescaled "square-shaped" functions which together form a wavelet family or basis.

It is based on the Haar Wavelet technique to analyze pixels in the image into squares by function. This uses machine learning techniques to get a high degree of accuracy from what is called "training data". This uses "integral image" concepts to compute the "features" detected. Haar Cascades use the **Adaboost** learning algorithm which selects a small number of important features from a large set to give an efficient result of classifiers. Interpretation is aided by context, body gesture, voice, individual differences, and cultural factors as well as by facial configuration and timing. Computer facial expression analysis systems need to analyze the facial actions regardless of context, culture, gender, and so on.

1.2 Age and Gender Detection:

This model further uses age and gender detection using multi model and ResNet. The feature extraction part of the neural network uses the WideResNet architecture, short for Wide Residual Networks. It leverages the power of Convolutional Neural Networks (or ConvNets for short) to learn the features of the face. From less abstract features like edges and corners to more abstract features like eyes and mouth.

The unique thing about WideResNet architecture is that the author decreased the depth and increased the width of original residual networks so it trained several times faster. For the age prediction, the output of the model is a list of 101 values associated with age probabilities ranging from 0~100, and all the 101 values add up to 1 (or what we call softmax). The gender prediction is a binary classification task. The model outputs value between 0~1, where the higher the value, the more confidence the model think the face is a male.

2. LITERATURE SURVEY

A detailed study on the facial recognition is discussed which exposes the properties of dataset, facial recognition study classifier. Visual features of image are examined and some of the classifier techniques are discussed in which is helpful in the further inspection.

Humans gain a great deal of information through vision, and there are many problems which computers could solve if they were similarly capable. Unfortunately, capturing and storing images on computers is trivial, deriving meaningful information from them is not. An object detection algorithm takes a large image and should return all subregions of that image representing objects of a given class. It should be fast and robust in doing so; in the face detection application, the point is to save time by not running the face recognition algorithm all over the image.

Haar Classifier Cascades This approach to object detection was first proposed and implemented by Viola and Jones (Viola & Jones, 2001b). A cascade steps through several stages in turn; it stops and returns false if a stage returns false. If all stages return true, the cascade returns true. 2. A stage returns true if the sum of its feature outputs exceeds a chosen threshold. 3. A feature returns the sum of its rectangle outputs. 4. A rectangle returns the sum of pixel values in the image region bounded by that rectangle, multiplied by a chosen weight. Some details of Haar Classifier Cascade training or detection are implementation dependent. Sometimes such details will be described with reference to Viola and Jones for the original implementation (Viola & Jones, 2001b) or to Leinhardt and Maydt for their extensions (Lienhart & Maydt, 2002).

FEATURE EXTRACTION: Feature Extraction is next step of Pre-Processing we need to extract face features. Generally, two types of features are extracted namely Geometric based features and appearance-based features.

GEOMETRIC-BASED FEATURE EXTRACTION: In geometric-based method, features are extracted from some facial points like face, nose and eyes. Loss of useful information takes place using geometric-based techniques Swathi Kalam and Geetha Guttikonda [18] reported 95.6% accuracy rate after calculating the ratios of left eye to right eye distance, eye to nose, eye to lip and eye to chin distance for feature extraction. Syed Zulqarnain Gilani and Ajmal Mian [20] experimented with 3D Euclidean Distances and for that accuracy rate is 88.36%. Li et al. [19] classified the gender by utilizing not only the five facial features (nose, eyes, mouth, forehead, brows) but also external information like clothes and hair features. Ziyi Xu et. al. [15] used AAM to get 83 landmarks from the face images.

CLASSIFICATION: The last step of gender classification is classification in which the face is successfully classified as that of a male or female. In classification different classifiers are trained and tested by different extracted features. Different classifiers are combined to minimize the classification error rate. These classifiers are called as ensemble classifiers. Combining the outcomes of different classifiers is known as ensemble classifiers. As compare to ensemble classifiers, single classifier accuracy rate is less. E. Makinen et.al [2] used ensemble classifiers such as multilayer Neural Network, Support Vector machine, SVM with LBP features, discrete Adaboost with Haar like Features to increase the classification rate. Nazir et al [13] ensemble different classifiers (i.e. K-nearest neighbor, Mahalanobis distance, Linear Discriminant analysis and k-means).

3. PROPOSED SOLUTION FOR THE IDENTIFIED PROBLEM

HAAR CASCADE CLASSIFIER: We implemented our use case using the Haar Cascade classifier. Haar Cascade classifier is an effective object detection approach which was proposed by Paul Viola and Michael Jones in their paper, "**Rapid Object Detection using a Boosted Cascade of Simple Features**" in 2001.

A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. This modified feature set is called *2-rectangle feature*. Viola and Jones also defined 3-rectangle features and 4-rectangle features. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or absence) of certain characteristics in the image, such as edges or changes in texture. For example, a 2-rectangle feature can indicate where the border lies between a dark region and a light region.

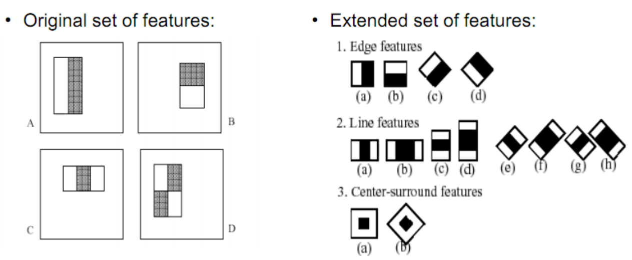

Fig 1: Haar features

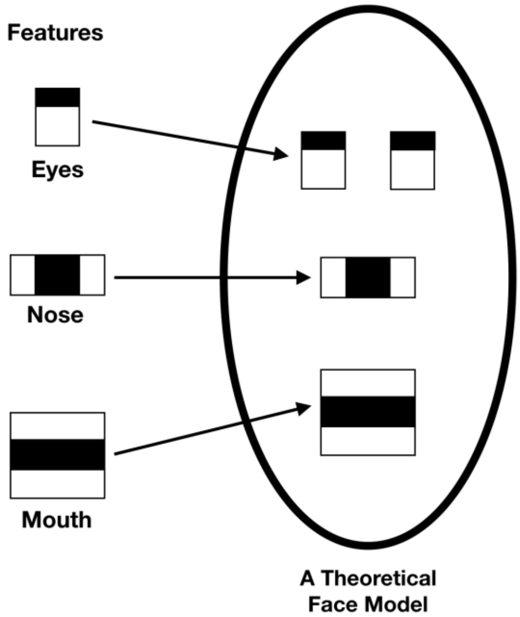

Fig 2: A theoretical face model.

Result obtained using Haar cascades:

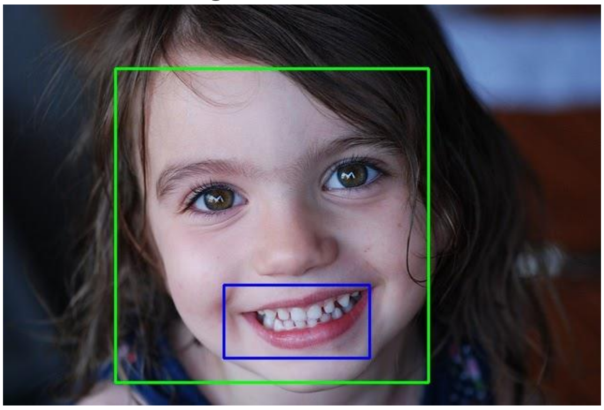

Fig 3: Implementation of Haar cascade classifier.

MULTI MODEL APPROACH: Models that are constructed within the bounds of a single paradigm are not sufficient for modeling all aspects of complex systems. Therefore, even though reasoning and simulation systems that utilize a single modeling paradigm are the current norm, we explore a multi model approach in this paper. A multi model approach is defined as one in which more than one model-each derived from a different perspective, and utilizing correspondingly distinct reasoning and simulation strategies-are employed. By describing four models which illustrate the use of different modeling techniques, we show how a multi model approach can enrich the modeling environment and make it correspond better with real world information. Our models come from many sources-Systems and Simulation literature for the modeling of natural phenomena and artificial devices, and Artificial Intelligence and Cognitive Science for the modeling of human intuition and expertise in reasoning.

Generalizing from these four models, we suggest that modeling complex systems may best be approached from an integrated architectural viewpoint which combines multiple modeling paradigms.

RESNET MODEL: ResNet, short for Residual Networks is a classic neural network used as a backbone for many computer visions tasks. This model was the winner of ImageNet challenge in 2015. The fundamental breakthrough with ResNet was it allowed us to train extremely deep neural networks with 150+layers successfully. Prior to ResNet training very deep neural networks was difficult due to the problem of vanishing gradients.

AlexNet, the winner of ImageNet 2012 and the model that apparently kick started the focus on deep learning had only 8 convolutional layers, the VGG network had 19 and Inception or GoogleNet had 22 layers and ResNet 152 had 152 layers. In this blog we will code a ResNet-50 that is a smaller version of ResNet 152 and frequently used as a starting point for transfer learning.

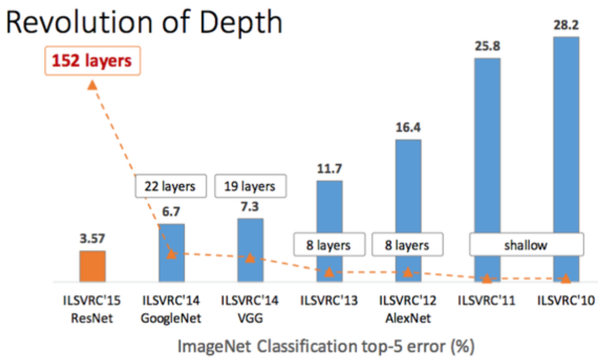

Fig 4: Revolution of Depth

However, increasing network depth does not work by simply stacking layers together. Deep networks are hard to train because of the notorious vanishing gradient problem — as the gradient is back-propagated to earlier layers, repeated multiplication may make the gradient extremely small. As a result, as the network goes deeper, its performance gets saturated or even starts degrading rapidly.

PREPARATION OF TRAINING DATASET: The dataset used for age and gender detection came from IMDB-WIKI – 500k+ face images with age and gender labels. Each image before feeding into the model we did the same preprocessing step as shown:

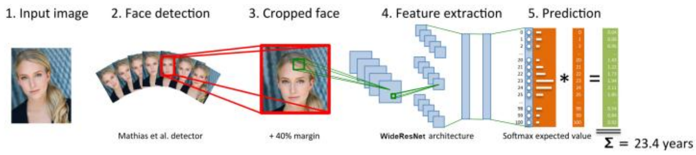

Fig 5: Steps for age prediction

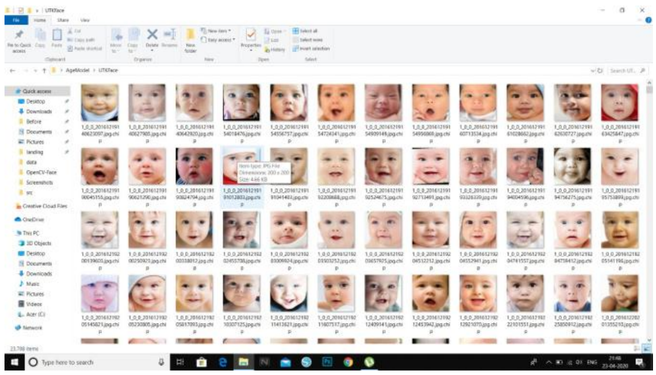

Fig 6: Dataset Used to train the age and gender model

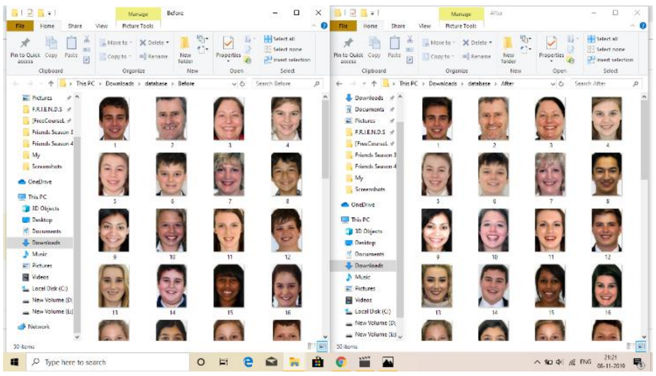

Fig 7: Dataset for teeth masking model

AGE AND GENDER DETECTION: We implemented the age and gender classification using multimodal system. We show that well-designed network architecture and properly tuned training hyperparameters, give better results. The experimental results on OIU-Audience dataset confirm that our model outperforms other studies on the same dataset, showing significant performance in terms of classification accuracy.

DATABASE OF TEETH USED FOR MASKING: A database of teeth according to the gender and age group is maintained for the masking purpose, Oval and triangular for female, Rectangular and square for male.

Fig 9: Snapshot of oval type teeth dataset

4. RESULTS

The age and gender detection model achieve classification accuracy values of 84.8% on age group and classification accuracy of 89.7% on gender.

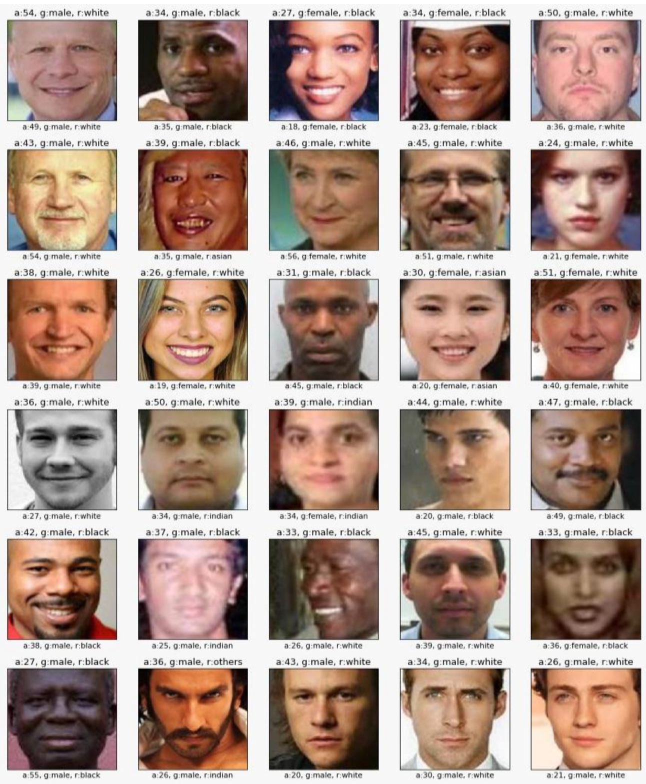

Fig 8: Results from Age and Gender detection model

Various results obtained after masking being applied to multiple images are as follows:

INPUT IMAGE **RESULT IMAGE**

INPUT IMAGE RESULT IMAGE

INPUT IMAGE RESULT IMAGE

5. CONCLUSION

Numerous researches and studies about face detection, Deep learning techniques used for age and gender detections are conducted. It is required in future to have a model like this with much more reliable, which has limitless possibilities in all fields. This project tried to use Haar cascades, ResNet, multi model and image processing to produce results and can be used in various fields in future with required modifications. Efficient results were obtained from the model and were reliable. In future, various other models can be developed using the same architecture.

REFERENCES

[1.] Y. Lecan, Y. Bengeo, and G. Hintin, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2.] J. Schmidhubar, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[3.] J. Ngiem, A. Khoslaa, M. Kam, J. Nim, H. Lei, and A. Y. Nig.

[4.] "Multimodal Deep Learning," *Proc. 28th Int. Conf. Mach. Learn.*, pp. 689–696, 2011.

[5.] Jae-Joon Hwang , Yun-Hoa Jung , Bong-Hae Cho, Min-Suk Heo, "An overview of deep learning in the field of dentistry", Journal of Information Science and Engineering March 2019.

[6.] Rameswari Poornima Janardanan and Rajasvaran Logeswaran, "Recent Image Processing Techniques in Forensic", March 2018.

[7.] Young-jun, "Machine Learning for Dental Image Analysis", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012.

[8.] Dalal, N., Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (Vol. 1, pp.886-893). San Diego, CA: Institute of Electrical and Electronics Engineers. Vivek and N. Dianna Radpour, "Using Deep Convolutional Networks for Gesture Recognition in American Sign Language", Department of Computer Science, Department of Linguistics, State University of New York at Buffalo, 2017.

[9.] Indira, A. P., Markande, A., David, M.P. (2012). Mandibular ramus: An indicator for sex determination-A digital radiographic study. *Journal of forensic dental sciences*, 4(2), 58.

[10.] Nomir, O. & Abdel-Mottaleb, M. Human Identification From Dental X-Ray Images Based on the Shape and Appearance of the Teeth. *IEEE Transactions on Information Forensics and Security* **2**, 188–197 (2007). 2. Zhou, J. & Abdel-Mottaleb, M. A content-based system for human identification based on bitewing dental X-ray image.