Gender Determination by Morphometry of Eyes Using Deep Learning

Group Members:

1. **Nabeel Ahmad Khan Jadoon**
2. **P.H. Amindu Dharmasena**
3. **Nutthapong Wongpromkal**

Affiliation:

Authors are affiliated with department of industrial system engineering, school of engineering and technology, technology Mechatronics and Machine Intelligence, Asian Institute of Technology, Thailand.

Abstract

Automatic gender classification from face images has been a popular topic among researchers for a decade. Feature extraction and classification methods are very important to create a successful automatic classification system. Due to the richness of face image datasets today, many successful machine learning, and deep learning methods have been implemented. It is very critical to extract accurate features from the datasets to achieve promising classification scores when traditional machine learning methods are used. However, deep learning models have been designed to extract the features automatically from the raw data directly. This also automatizes the feature extraction process besides classification. The hidden and unpredictable feature sets can be explored by the deep neural networks which can increase the classification performance compared to traditional machine learning methods. The motivation behind selecting this project derives from the regulations, and state-of-art technique used in just peculiar section of the face for classification method. The anthropometric analysis of the human face is an essential study for performing craniofacial plastic and reconstructive surgeries. Facial anthropometrics are affected by various factors such as age, gender, ethnicity, socioeconomic status, environment, and region. Plastic surgeons who undertake the repair and reconstruction of facial deformities find the anatomical dimensions of the facial structures useful for their surgeries. These dimensions are a result of the Physical or Facial appearance of an individual. Along with factors like culture, personality, ethnic background, age; eye appearance and symmetry contribute majorly to facial appearance or aesthetics. Our objective is to build a model to scan the image of an eye of a patient and find if the gender of the patient is male or female. The concept of multilayer deep learning model will be used to evaluate the training, testing and validation tests from eyes-rtte, 2021 dataset.

Keywords: Eye region, gender determination, facial data, morphology, classification with CNN

Section 1: Introduction

Human face detection has become a very important cog in our day-to-day activities. From unlocking our phone to crowd video surveillance to tracking humans by security services, human face detection is utilized in a wide field.

Today, artificial intelligence solves many problems related to human recognition. For instance, the face, fingerprint [1], speech, iris is some of the human data to be used by artificial intelligence models. Human gender recognition from face data is also a challenging topic today due to the rapid change of human face. The performance of the methods may easily decline by the change of the face like with/without moustache, beard or with aging. It can also be affected by the orientation of the head. In literature, many methods and models have been developed by researchers to solve those kinds of problems and provide accurate results [1].

Feature-based and Image-based approaches are the two main types of approaches utilized in face detection. An image-based approach finds the information using machine learning techniques and appropriate datasets. In feature-based approach, facial features like eyes, nose, mouth, eyebrows, skin color, etc. are extracted and statistical models are developed to obtain the relationship between the extracted features and the presence of a face. But a major problem faced in the feature-based approach is the corruption of the data due to illumination, noise, occlusion, feature boundary uncertainties due to shadows, etc. So, the algorithms implemented can end up being useless. Apart from these technical issues, there are plenty of cultural and religious beliefs which makes obtaining images required for face detection and related applications difficult. Especially, when it comes to gender detection, these socio-ethnic issues play a major role in determining the availability of images. Global implementation of these types of technologies gets hindered due to reasons like these.

One way to mitigate some of these issues is to focus on a few specific features of the face instead of utilizing the entire face. This reduces the complexity of the implemented algorithms and reduces the scrutiny that's been around these types of technologies and systems due to religious and social beliefs.

With these goals in our mind, we decided to construct a model which scans an eye of the subject and detects the gender of the person (male or female).

Section 2: Literature Review

Deep learning has gained much more interest in the application areas of automatic feature extraction, object recognition, classification, etc. Especially, more complex and deep convolutional neural networks perform high accuracy. By drawing inspiration from various areas, deep learning has been also applied for gender classification from face images

Comparative Model Research

The study in [2] creates a dataset which is composed of the facial images of Caucasian and Malaysians people and applies various Convolutional Neural Network (CNN) architectures to conduct a gender classification. It reports 88%, 85% and 49% accuracy by using VGG-16, ResNet-50 and MobileNet models respectively. The study conducted by [7] applies CNN and Local Recipient Areas Excessive Learning Machine (LRA-ELM) methods on Adience dataset [10] and achieves 80% and 87.13% for LRA-ELM and CNN, respectively. The study in [3], approximately uses 1500 face images where most of them were chosen from CASIA dataset [4]. It develops CNN and achieves 94.5% gender classification accuracy. The study conducted by [5] develops a deep learning method for classifying pedestrian gender. The pedestrian was segmented from the picture using a preprocessing step. Then, for classification, stacked auto encoders with a SoftMax classifier were utilized. It achieves accuracy rates of 82.9%, 81.8%, and 82.4% percent in the front, posterior, and mixed views, respectively on MIT dataset, and about 91.5% in the PETA dataset [15]. In [16] a comparative analysis is performed between the CNN models: Proposed CNN, AlexNet and VGG-16. 72.20%, 65.63% and 99.41% accuracies have been obtained respectively. Another study in [16] proposes a simple CNN to improve the performance of gender classification. Promising accuracy results have been obtained on the Adience dataset. A Resnet-101 CNN based method called Hyperface is proposed [13]. This method increases the gender recognition rate and speed.

The study in [14] reports the interchange of traditional CNN models with the PCANet model for gender categorization. In addition, by using PCANet, it is able to decrease the size of the network design in complicated CNN models. For gender categorization, this technique has an accuracy of 89.65%. In [17] researchers propose a CNN with reduced number of layers. By applying the method on a dataset composed of 1496 body images, it achieves 91.5% accuracy.

VGG network

Visual Geometry Group (VGG) is a standard deep Convolutional Neural Network (CNN) architecture which consists of multiple layers. This architecture is the very much used in object recognition models and is in fact a popular image recognition architecture [1],[2].

The model used in this project; VGG-19 [1] Has 19 convolutional layers. This has 16 Convolutional Layers and 3 Fully Connected Layers. The input for the model is an RGB image with predetermined pixel size which can vary according to the application. The image should be preprocessed before its fed into the model. The mean RGB value which is obtained considering the entire training set is deducted from each pixel of every image. In the convolutional layers, 3*3 kernels with a stride of 1 pixel were implemented to ensure that the entire image is covered, and spatial padding was implemented to conserve the resolution of the image. Max pooling was done over 2*2 size with the stride of 2 to reduce the computational load and reduce overfitting. After max pooling, Rectified Linear Unit (ReLU) is implemented to improve the non-linearity of the output [2] ReLU is preferred over activation functions like ‘Sigmoid’, “Leaky ReLU’, “Tanh” because ReLU doesn’t hinder the computational time while introducing the non-linearities [1], [3].

This is the generic architecture of the 16 Convolutional layers that are implemented in the model. Out of the 3 Fully Connected Layers, last layer is configured to have only 2 channels (male or female) which is incorporated with a “SoftMax” function. Altogether, the model has 16 Convolution layers, 3 Fully Connected Layers, 5 Max Pooling and 1 SoftMax layer [3].

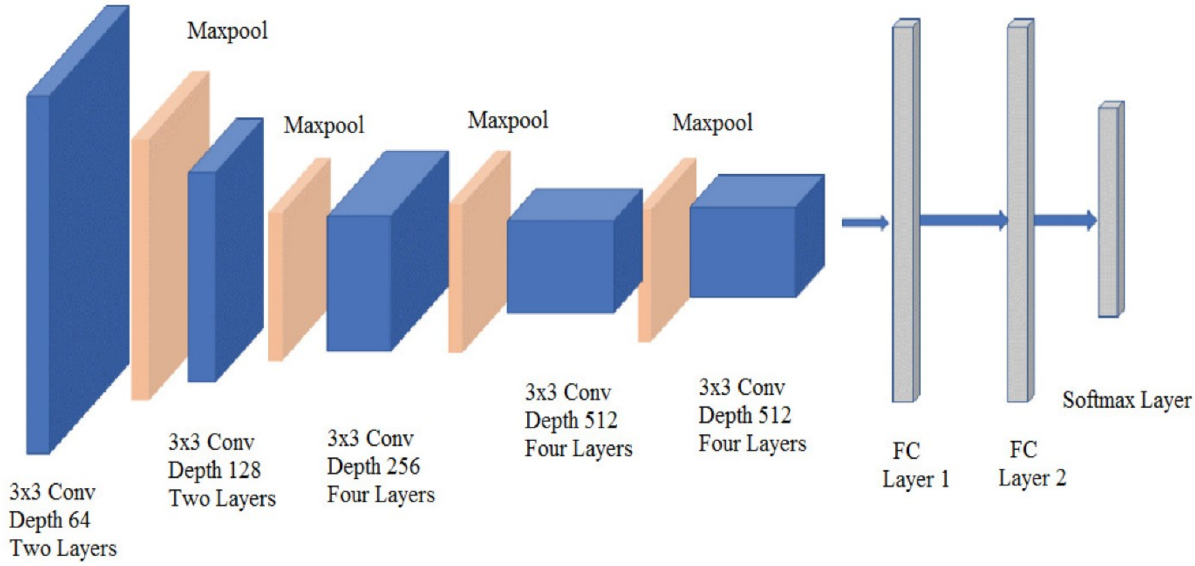


Figure 1: VGG 19 Architecture

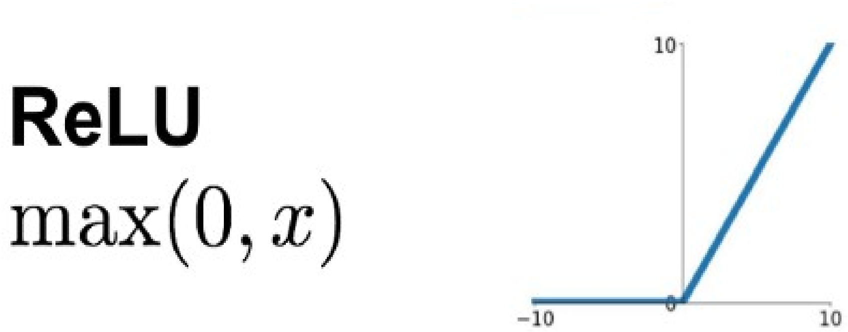


Figure 2: ReLU Activation Function

Datasets History:

This is a visual dataset that consists of about 15 million labeled high resolution images allocated to about 22,000 categories [4].

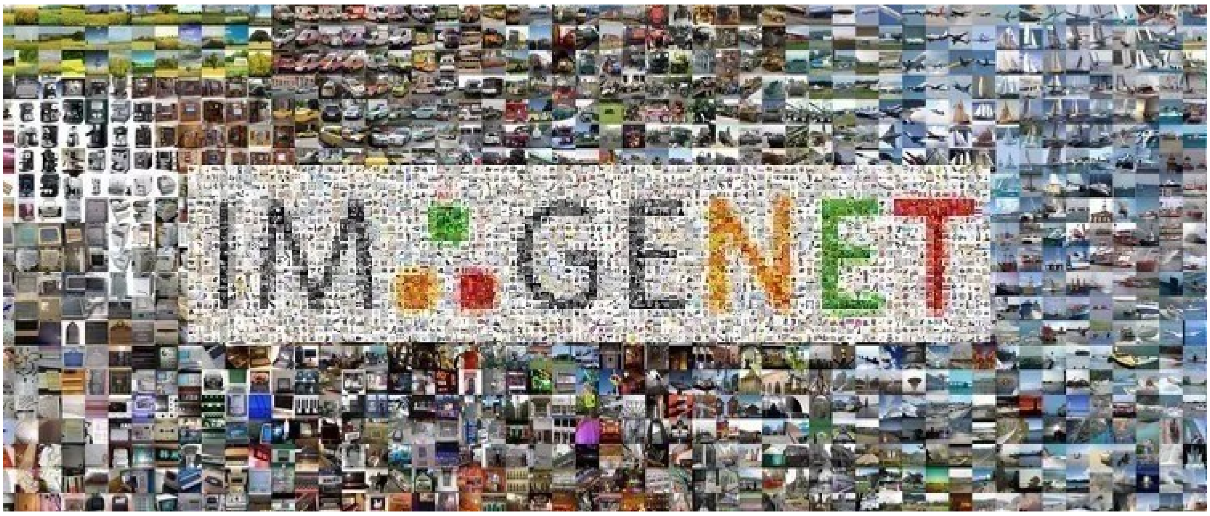


Figure 3: ImageNet Database

Since 2007, LFW dataset has been used for gender classification in literature. Similarly, Gallagher's dataset, FERET dataset which of recent advancement in color-feret-database 2021 [8][13][14].

Section 3: Methodology

A deep learning framework has been followed to effectively construct the methodology. Like other machine learning projects, the gathering of data was itself a difficult task. However, ImageNet data set was used, and processing was done using python script to focus on the peculiar section of the eye only. Our framework follows below strategy:

Workflow:

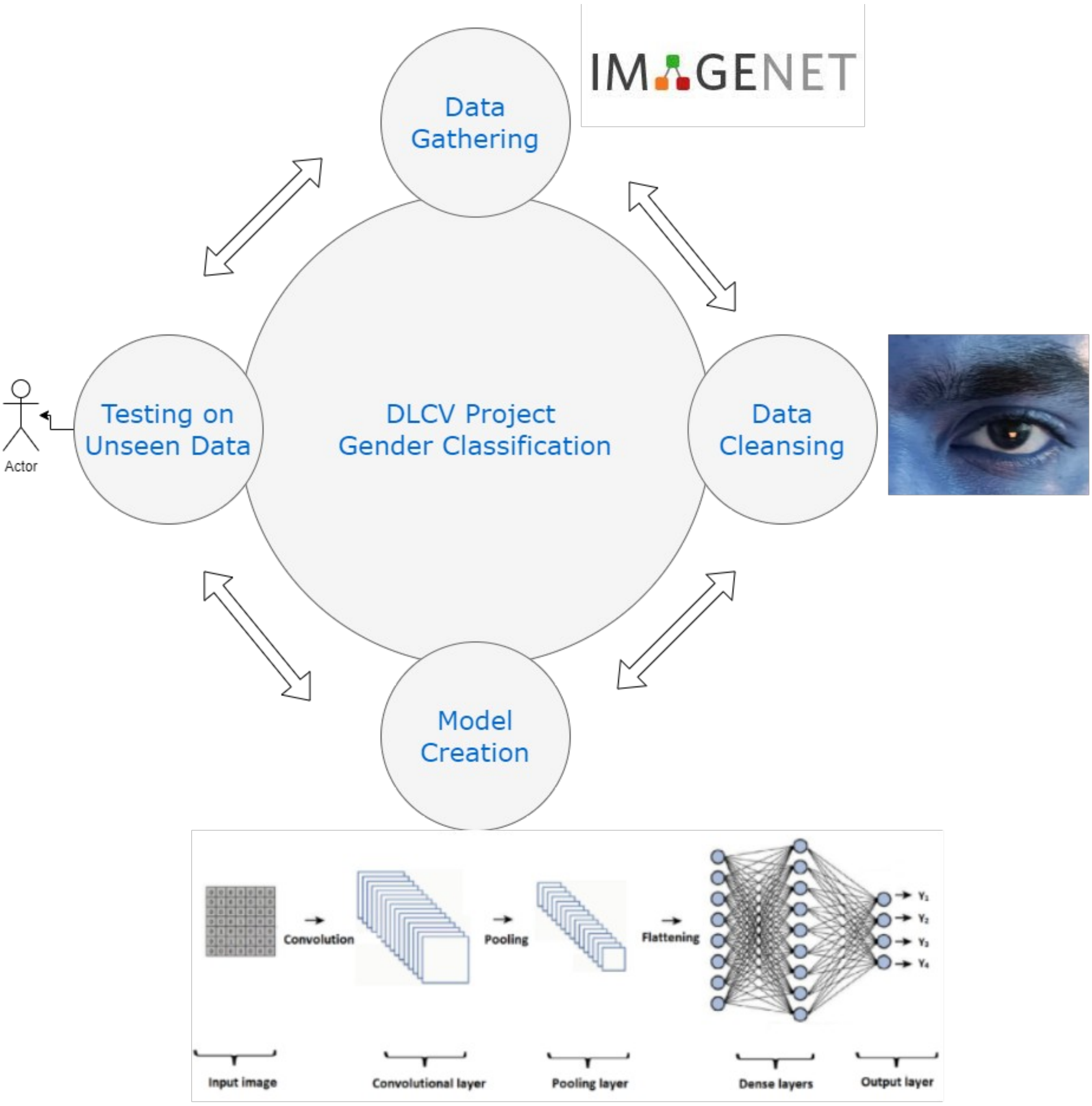


Figure 4: Custom methodology framework

The model selection was done using the benchmark for image classification along with tunning parameters [5].

Dataset:

Since our study conducts a comparative analysis between state-of-the-art CNN models only by using the eye images, we use the dataset named “Female and Male” which is referenced in [11]. This dataset includes only the eye images which are extracted from whole face images.

Note that it usually includes either complete or some part of eyebrows also. Moreover, it includes 5202 female and 6323 male eye images. Figure 1 and 2 show female and male eye image examples from this dataset.[11]

Figure 5: Selected male eye images from eyes-rtte & ImageNet Dataset [11]

Figure 6: Selected female eye images from eyes-rtte & ImageNet Dataset [11]

Size of Dataset:

For the final dataset, the size of the training and test data is shown as below

Gender DataSet DLCV Project		
Training Dataset	Male	Female
	5058	4162
Testing Dataset	2303	

Model Configuration:

We conjecture that despite the larger number of parameters and the greater depth of our nets compared to the nets required less epochs to converge due to:

- (a) -implicit regularization imposed by greater depth and smaller conv. filter sizes.
- (b)- pre-initialization of certain layers.

The initialization of the network weights is important since bad initialization can stall learning due to the instability of gradient in deep nets. To circumvent this problem, we began with training the configuration used in model summary of this DL project, shallow enough to be trained with random initialization. Then, when training deeper architectures, we initialized the first four convolutional layers and the last three fully connected layers with the layers of net A (the intermediate layers were initialized randomly). We did not decrease the learning rate for the pre-initialized layers, allowing them to change during learning. In this study, to remove the effect of different lighting conditions, each eye image in the dataset is normalized. The training is implemented by using TensorFlow on Python environment. The image sizes are adjusted as 55x55.

```
# view model layers
model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 55, 55, 128)       3584

 max_pooling2d (MaxPooling2D  (None, 27, 27, 128)      0
 )

 conv2d_1 (Conv2D)           (None, 25, 25, 64)        73792

 max_pooling2d_1 (MaxPooling  (None, 12, 12, 64)       0
 2D)

 conv2d_2 (Conv2D)           (None, 10, 10, 32)        18464

 max_pooling2d_2 (MaxPooling  (None, 5, 5, 32)         0
 2D)

 conv2d_3 (Conv2D)           (None, 3, 3, 16)          4624

 max_pooling2d_3 (MaxPooling  (None, 1, 1, 16)         0
 2D)

 flatten (Flatten)           (None, 16)                0

 dense (Dense)               (None, 256)               4352

 dense_1 (Dense)             (None, 256)               65792

 dense_2 (Dense)             (None, 2)                 514

=================================================================
Total params: 171,122
Trainable params: 171,122
Non-trainable params: 0
```

Table 1: Model Summary and Layers Information

The training parameters of our VGG19 model are given in below table

Parameter	Value
Optimizer	Adam
Loss	Sparse Categorical Cross-entropy
Shuffle	True
Number of epochs before/after fined tuned	25/6
Batch size	32

Table 2: Training Parameters of Vgg19

Testing scheme begins, given a trained ConvNet and an input image, it is classified in the following way. First, it is isotopically rescaled to a pre-defined smallest image side. Then, the network is applied densely over the rescaled test image in a way like [7]. Namely, the fully connected layers are first converted to convolutional layers. The resulting fully convolutional net is then applied to the whole (uncropped) image. The result is a class score map with the number of channels equal to the number of classes, and a variable spatial resolution, dependent on the input image size.

Finally, to obtain a fixed-size vector of class scores for the image, the class score map is spatially averaged (sum-pooled). Since the fully convolutional network is applied over the whole image, there is no need to sample multiple crops at test time, which is less efficient as it requires network re-computation for each crop. At the same time, using a large set of crops can lead to improved accuracy, as it results in a finer sampling of the input image compared to the fully convolutional net.

Region-of-Interest (ROI) Scheme:

Thus, multi-crop evaluation is complementary to dense evaluation due to different convolution boundary conditions: when applying a ConvNet to a crop, the convolved feature maps are padded with zeros, while in the case of dense evaluation the padding for the same crop naturally comes from the neighboring parts of an image (due to both the convolutions and spatial pooling), which substantially increases the overall network receptive field, so more context is captured. Region of interest was defined as scaling the whole image dataset uniformly and getting the ROI of eyes.

Section 4: Experimental results

Previous models have achieved the highest accuracy of 96% on NASNetLarge which of 80M parameters, now we have used this to benchmark up to 97% accuracy (with total 1.7M parameters). The comparison table is shown below. [7][8][17]

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
InceptionV3	92	0.779	0.937	23,851,784	159	42.25	6.86
InceptionResNetV2	215	0.803	0.953	55,873,736	572	130.19	10.02
Xception	88	0.790	0.945	22,910,480	126	109.42	8.06
NASNetLarge	343	0.825	0.960	88,949,818	-	344.51	19.96

Table 3: DL Models and accuracy comparison on gender classification

Figure 8: depicts the output images with the predicted gender classification. Female is denoted by 0 while Male is denoted by 1

Out of the 2305 images used for training, 1380 were predicted as 'male' while 925 were predicted as 'female'. An accuracy of 97% was achieved during training. Figure 9 and Figure 10 are the graphical representations of the model accuracy and loss during the training phase. We have used the tensor board call which offers smoothing function for curve fitting and tunning.

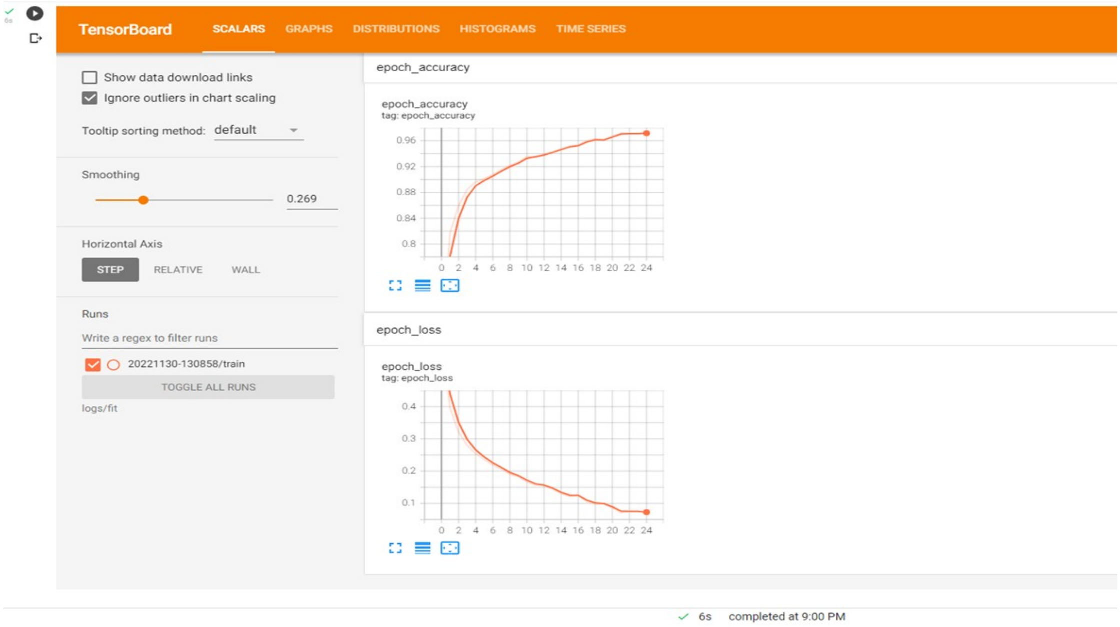
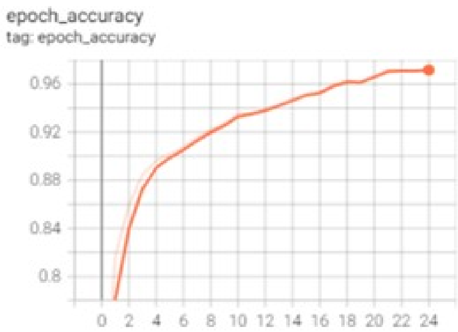


Figure 9: Tensor Board Callback



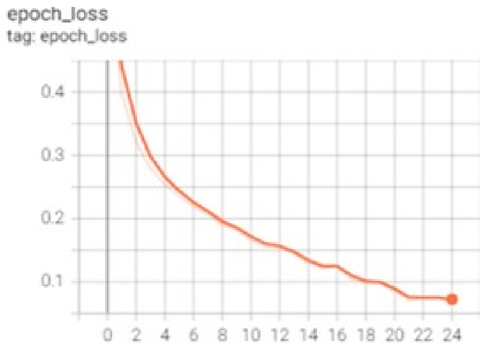


Figure 10-(a) Training Accuracy   10- (b) Training Loss

The following python code snippet shows the results for all the test images inside the python console along with file name and predicted label.

```
  elif (pred[j]==1):
    lt.append("male")
  print("File name is = "+test.filename[i], "&", "Pricted Label is = "+lt[j])
sub = pd.DataFrame(lt,columns=["label"])
print(test.filename[i],lt[j])
sub.to_csv("drive/My Drive/DLCVProject/submit2.csv",index=True)
#lt[:10]
```

```
73/73 [==============================] - 0s 4ms/step
File name is = Image_2305.jpg & Pricted Label is = male
File name is = Image_2305.jpg & Pricted Label is = male
File name is = Image_2305.jpg & Pricted Label is = female
File name is = Image_2305.jpg & Pricted Label is = male
File name is = Image_2305.jpg & Pricted Label is = male
File name is = Image_2305.jpg & Pricted Label is = male
File name is = Image_2305.jpg & Pricted Label is = female
File name is = Image_2305.jpg & Pricted Label is = male
File name is = Image_2305.jpg & Pricted Label is = female
File name is = Image_2305.jpg & Pricted Label is = female
File name is = Image_2305.jpg & Pricted Label is = male
```

Testing on our faces and checking results

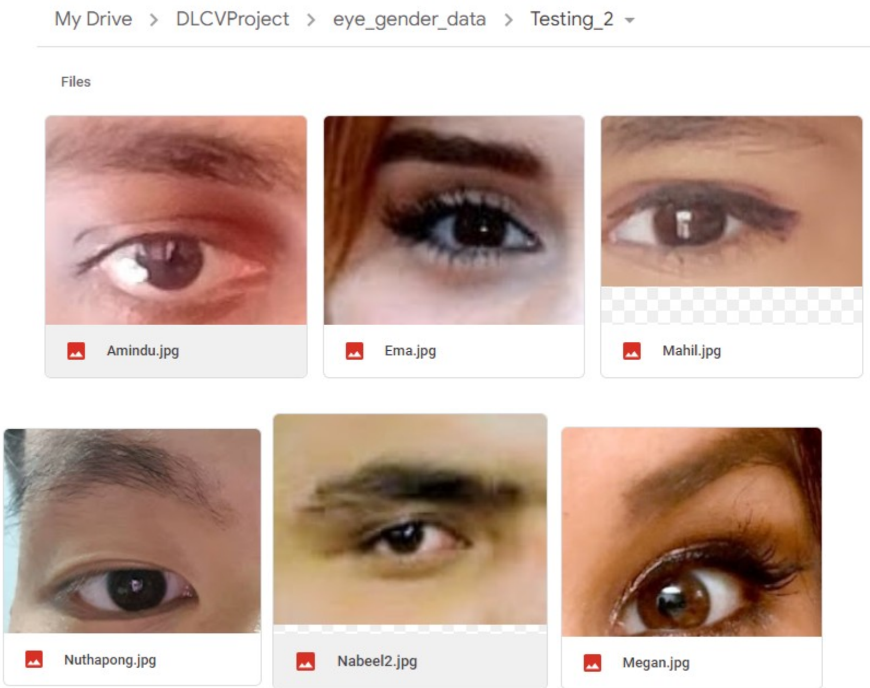
```
for j in range(test.shape[0]):
  if (pred[j]==0):
    lt.append("female")
  elif (pred[j]==1):
    lt.append("male")
  print("File name is = "+test.filename[j], "&", "Pricted Label is = "+lt[j])
sub = pd.DataFrame(lt,columns=["label"])
sub.to_csv("drive/My Drive/DLCVProject/submitTesting_2.csv",index=True)
#lt[:10]
```

```
1/1 [==============================] - 0s 18ms/step
File name is = Nabeel.jpg & Pricted Label is = male
File name is = Amindu.jpg & Pricted Label is = male
File name is = Ema.jpg & Pricted Label is = female
File name is = Mahil.jpg & Pricted Label is = female
File name is = Megan.jpg & Pricted Label is = female
File name is = Nuthapong.jpg & Pricted Label is = male
File name is = Nabeel2.jpg & Pricted Label is = male
```

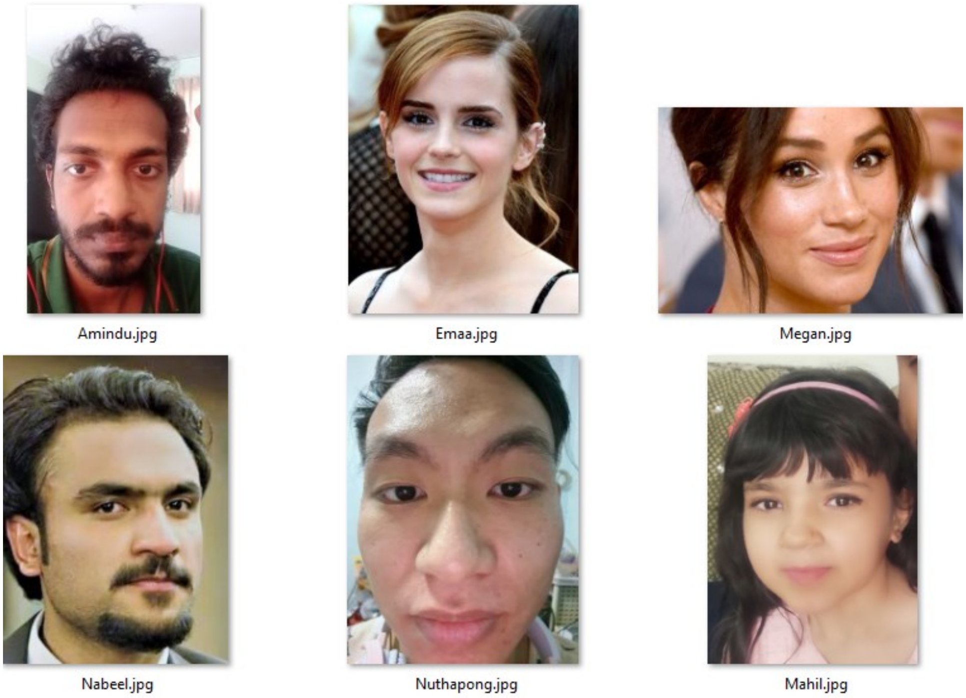
We have represented the output data in .csv format due to computational flexibility which allows one to match all the test data effectively by comparing the respective labels. Since it would be highly unprofessional to show the test images with their labels in programming console. Therefore, represented as shown below.

	label
0	male
1	male
2	female
3	female
4	female
5	male
6	male

Now, let's seen the test data which matches these labels from directory and clearly shows the identification of male or female as per images.



Now for comparison we have full-face images and it is clear that our algorithm gives correct predictions.



The results clearly demonstrate the correctness of our model and classification on the test data.

Section 5: Conclusion

This project study presents the classification performance of the state-of-the-art deep CNN model called Vgg19. During the training phase, we managed to obtain an accuracy of 97%. So, this is a clear indication that our approach is reasonable, which is detecting the gender utilizing the morphometry of the eyes. Our approach presents an alternative way of getting rid of face regulations that are imposed for face recognition, and classification. This approach classifies the gender with high accuracy.

Section 6: Future Recommendations

Since there are various classifications of genders for LGBQT etc., thus, for other categories new dataset is required along with better performing model on multiclass classification. Indeed, training a maximum number of epochs and parameters can lead to high accuracy which results in computationally expensive. Hence, further optimization can be used for this model along with convolution scheme to reduce the training time. Introducing multiclass classification, this project can be a precursor for gender equality labelling in the field of machine learning and can be recognized as the benchmark for different classes of genders.

References

[1] B. Gaudenz, “VGG Very Deep Convolutional Networks (VGGNet) – What you need to know,” *viso.ai*, 2022. https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/ (accessed Nov. 30, 2022).

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[3] A. Kaushik, “Understanding the VGG19 Architecture,” *OpenGenus Foundation*, 2022. https://iq.opengenus.org/vgg19-architecture/.

[4] P. Sharma, “7 Popular Image Classification Models in ImageNet Challenge (ILSVRC) Competition History,” *MLK making AI simple*, 2020. https://machinelearningknowledge.ai/popular-image-classification-models-in-imagenet-challenge-ilsvrc-competition-history/#:~:text=ImageNet is a visual Dataset that (accessed Dec. 01, 2022).

[5] Papers with code, meta-ai open benchmark platform, https://www. paperswithcode.com/sota/image-classification-on-imagenet.

[6] Karen Simonyan & Andrew Zisserman, Visual Geometry Group, Department of Engineering Science, University of Oxford, (ICLR 2015).

[7] Akbulut, Y., Şengür, A., & Ekici, S. (2017, September). Gender recognition from face 7images with deep learning. In 2017 International artificial intelligence and data processing symposium (IDAP), 1-4. IEEE.

[8] Color-feret-database. (12 October 2021). https://www.nist.gov/itl/products-and-services/

[9] Danisman, T., Bilasco, I. M., & Martinet, J. (2015). Boosting gender recognition performance with a fuzzy inference system. Expert Systems with Applications, 42(5), 2772-2784.

[10] Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. in IEEE Transactions on Information Forensics and Security, 9 (12), 2170-2179, doi: 10.1109/TIFS.2014.2359646.

[11] Eyes-rtte. (23 September 2021). https://www.kaggle.com/pavelbiz/eyes-rtte

[12] Gallagher, A. C., & Chen, T. (2008, June). Clothing segmentation for recognizing people. In 2008 IEEE conference on computer vision and pattern recognition, 1-8. IEEE.

[13] Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Univ. Massachusetts, Amherst, MA, USA.

[14] Janahiraman, T. V., & Subramaniam, P. (2019, October). Gender Classification Based on Asian Faces using Deep Learning. In 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) (84-89). IEEE.

[15] Keras applications. (20 October 2021). https://keras.io/api/applications/

[16] Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (34-42).

[17] Cimtay, Y. & Yilmaz, G.N. (2021). Gender classification from eye images by using pretrained convolutional neural networks. The Eurasia Proceedings of Science, T echnology, Engineering & Mathematics (EPSTEM), 14, 39-44.