Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.026674
Article

Tech Science Press

Artificially Generated Facial Images for Gender Classification Using Deep Learning

Valliappan Raman[1], Khaled ELKarazle[2,*] and Patrick Then[2]

[1]Coimbatore Institute of Technology, Coimbatore, 641014, India
[2]Swinburne University of Technology, Kuching, 95530, Malaysia
*Corresponding Author: Khaled ELKarazle. Email: kelkaeazle@swinburne.edu.my

Received: 01 January 2022; Accepted: 22 February 2022

Abstract: Given the current expansion of the computer vision field, several applications that rely on extracting biometric information like facial gender for access control, security or marketing purposes are becoming more common. A typical gender classifier requires many training samples to learn as many distinguishable features as possible. However, collecting facial images from individuals is usually a sensitive task, and it might violate either an individual's privacy or a specific data privacy law. In order to bridge the gap between privacy and the need for many facial images for deep learning training, an artificially generated dataset of facial images is proposed. We acquire a pre-trained Style-Generative Adversarial Networks (StyleGAN) generator and use it to create a dataset of facial images. We label the images according to the observed gender using a set of criteria that differentiate the facial features of males and females apart. We use this manually-labelled dataset to train three facial gender classifiers, a custom-designed network, and two pre-trained networks based on the Visual Geometry Group designs (VGG16) and (VGG19). We cross-validate these three classifiers on two separate datasets containing labelled images of actual subjects. For testing, we use the UTKFace and the Kaggle gender dataset. Our experimental results suggest that using a set of artificial images for training produces a comparable performance with accuracies similar to existing state-of-the-art methods, which uses actual images of individuals. The average classification accuracy of each classifier is between 94% and 95%, which is similar to existing proposed methods.

Keywords: Facial recognition; data collection; facial images; generative adversarial networks; facial gender estimation

1 Introduction

Deep learning has become the predominant solution for most facial analysis problems. The preference for deep learning models in the computer vision field is due to the robustness of architectures such as convolutional neural networks or residual connections, which effectively extracts all the necessary facial features without manually engineering the features using several filters [1,2]. Gender estimation is the idea of training a machine learning model on thousands of labelled facial images to produce a function that maps an input image X to a corresponding label Y.

After training a model for several epochs, it is then evaluated either on a test portion of the same dataset or a different dataset entirely. The accuracy is recorded, and the model hyperparameters and design are modified accordingly to improve the performance. The training and testing processes are summarised in Figs. 1 and 2.

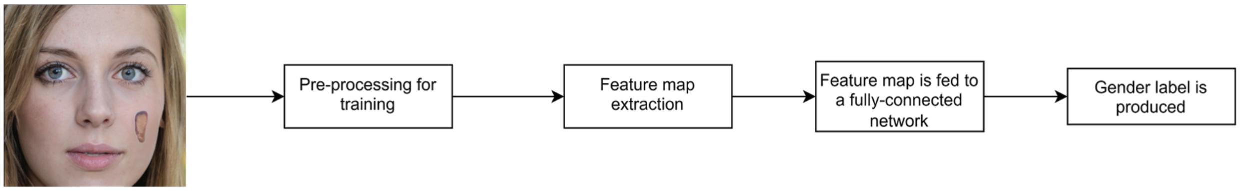
Pre-processing for training
Feature map extraction
Feature map is fed to a fully-connected network
Gender label is produced

Figure 1: A typical training process of a gender classifier

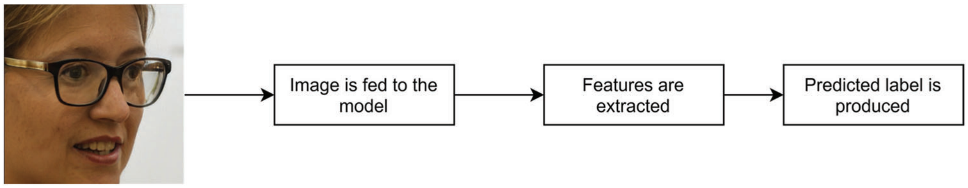

Figure 2: The testing process of a typical gender estimation model

Despite the numerous applications that can benefit from automatic facial gender classification, acquiring enough training images remains a significant challenge for several reasons. Firstly, collecting photos of individuals comes with privacy risks as facial images are sensitive data; therefore, several precautions should be taken before building a training dataset for a facial analysis model [3]. Secondly, capturing images to construct a dataset of a few thousand samples is a time-consuming process.

To solve the issues mentioned above, we construct a dataset of images that have been artificially generated using Style-Generative Adversarial Networks (StyleGAN) [4]. We then label the images based on the observed gender of the subject in the image using a set of features as a guideline to decide whether the subject is a male or a female. Next, we train a custom-built Convolutional Neural Network (CNN), a pre-trained Visual Geometry Group (VGG16) and a (VGG19) that differentiate gender classes from a given facial image. We choose to train three different models to validate the usage of artificially generated images to solve a facial analysis problem. We test the models on two different datasets, the Kaggle gender dataset [5] and the UTKFace dataset [6].

We present the accuracy of our classifiers and compare the performance with existing state-of-the-art gender classification models to understand how is the performance affected when artificial images are used for training. To the best of our knowledge, this is the first study to explore the idea of training a facial analysis task on an artificial dataset.

There have been several motivations behind our research. Firstly, we are keen to explore a different mechanism to acquire a labelled dataset without collecting images of actual subjects. We are interested in this process to minimise the potential high privacy risk when collecting many facial images to train deep learning models. Several studies, such as [7,8], have highlighted that the process of building any facial analysis system, whether these models estimate gender, age or emotions, comes with several privacy issues.

Some of these well-known issues include the absence of permission to use an individual's facial photo or regulations that may prohibit the usage of actual facial images for machine learning research. These issues have significantly inspired this study to be conducted.

Additionally, as it takes a long time to acquire a significant number of facial images by either manually capturing them or applying for access to an existing database for research purposes, we are keen to find an efficient alternative to the existing process.

Finally, due to the lack of research into utilising StyleGAN-produced images as training samples, especially for a task such as facial gender estimation, we developed an interest to discover the advantages and disadvantages of training a deep learning model entirely on an artificial dataset. This manuscript is organised into eight sections.

The first section is the introduction which includes the definition of facial gender estimation, a brief overview of the typical way of building a gender estimation model and a brief introduction to our method. The second section presents the motivation and organisation of the manuscript. In section three, we present our contribution to the body of the literature. In section four, we present several existing methods to estimate facial gender. Section five thoroughly presents our method of building the dataset and training our classifier. Section six presents the accuracies we obtained and compares them to several state-of-the-art methods. Section seven presents a discussion on our results, several abnormal encounters we noticed during training and testing, the advantages and disadvantages of using our dataset, and potential suggestions to improve the classification performance when a model trains on our dataset. Finally, in section eight, we conclude the study and present our future works. The scientific contribution of this study is mainly creating the StyleGAN-generated dataset rather than building a gender classifier. However, as we attempt to test the generated dataset, part of the contribution is training a gender classifier on the dataset. The main distinction between this study and similar ones is that instead of optimising or introducing a gender classifier, we investigate alternative ways to gather training samples without collecting actual facial images. The main contributions of this study are as follows:

- We generate many facial images using StyleGAN to reduce the need for actual facial images. We then label the generated images based on the observed facial gender.
- We train a custom-built binary classifier on the artificial dataset to classify genders from a given facial image.
- We modify two pre-trained networks and train them on the StyleGAN-generated dataset.

2 Related Work

There have been numerous algorithms and models proposed to solve the problem of gender estimation. However, the main flow in which the images are pre-processed, the feature map is extracted, and gender is estimated remain the same. In one study conducted by [9], the authors introduced a gender estimation model based on several steps. Firstly, the authors used the Viola-Jones face detection framework to locate a face in a given image. The authors used this framework instead of building their face detection algorithm because this framework has proven effective in various face analysis tasks [10–12].

Once a face is detected, it is extracted from the image by cropping the bounding box around it. This step is necessary for eliminating unwanted background noises that might interfere with training or testing. The authors then extract the feature map using a local binary pattern texture operator with their related patch LBP codes. The texture operator constructs the feature maps using the extracted features and shapes from a given image. The next step in this method is training a model on the extracted gender facial features. The authors chose a modified version of the linear Support Vector Machine (SVM) annotated as dropout-SVM. The distinct difference between a normal SVM and the proposed one is that the modified model has a dropout component to reduce overfitting as much as possible. The authors reported an accuracy of 88.6% when testing the model on the Gallagher benchmark dataset [13]. Despite reporting classification accuracy of over 80%, the main drawback of using this method lies in the feature extraction stage. Manual feature extraction can be less effective than deep learning methods, and that is because certain facial features might not be captured by a texture operator such as Local Binary Patterns (LBP).

In a different study carried out by [14], the authors trained a convolutional neural network from scratch to classify facial gender. The authors used a challenging dataset denoted as Adience, which they introduced. The dataset consists of more than 20,000 facial images taken in unrestrained conditions on different devices, and it has been used in several studies. The classifier consists of three convolutional layers; each is activated using the Rectified Linear Activation Function (ReLU). Each convolutional layer is followed by a max-pooling layer with a pool size of 3 × 3 and a local response normalisation layer.

Training of the classifier begins by firstly resizing the facial images to 256 × 256. A 227 × 227 crop is then fed to the input layer of the classifier. The classifier is comparatively shallow as it consists of 3 convolutional layers and two fully connected layers. The first layer consists of 96 filters of size 7 × 7. The second layer consists of 256 filters with a kernel size of 5 × 5. The third and last convolutional layer consists of 384 filters with a kernel size of 3 × 3. The fully connected layers consist of 512 neurons each, and they are activated using the ReLU activation function. A single dropout layer follows both fully connected layers to overcome overfitting with a rate of 50%. The output layer consists of two classes, each mapping to a gender class. The authors stated that this network design was decided on after several experiments. The authors have reported a classification accuracy of 86.8% on the Adience dataset. The low resolution of several samples has been the main limitation of this study due to the disappearance of essential facial features in blurred images.

Another study [15] approached the problem of gender estimation differently. The authors employed a pre-trained Visual Geometry Group Face (VGG-Face) [16] network, which has been initially trained to recognise 10000 celebrities from their facial images. The VGG-Face network takes an input RGB facial image of input size 224 × 224. The network comprises 33 convolutional layers, where each layer is followed by a max-pooling layer and activated using the ReLU function. The fully connected portion of the network consists of two layers; each layer consists of 4096 neurons, and each layer is activated using the ReLU function. To fit the nature of the gender estimation task, the authors modified the pre-trained network by adding an output softmax layer with two outputs to represent the gender classes. In addition, the authors attempted to reduce overfitting by introducing a dropout layer with a rate of 50% after each layer. The authors used a large-scale dataset denoted as the Celebrity face [17] for training and testing. The dataset consists of 100,000 facial images of celebrities; however, only 200 images out of this dataset was used in this experiment. The training dataset is evenly distributed among both classes, where each class consists of 100 samples. Of the 200 samples, only 160 were used for training, while the remaining 40 were used for testing. The authors reported an accuracy of 95%; however, the main issue with this study is the low number of testing samples. We theorise that the testing portion has lacked challenging samples (e.g., different illuminations or non-frontal faces) that, if included, may decrease the presented accuracy.

The authors in [18] have introduced a CNN trained from scratch to classify facial gender. The proposed network consists of four convolutional layers; each layer is activated using the ReLU activation function followed by a max-pooling layer. The network's input layer consists of 16 filters, and it takes an RGB facial image of size 224 × 224. The first hidden layer consists of 32 filters, and the second hidden layer consists of 64 filters. The last convolutional layer consists of 128 filters. The authors maintained a kernel size of 3 × 3 throughout the convolutional portion of the network. The fully connected portion of the network contains two layers, with each layer containing 4096 neurons and activated using the ReLU activation function. The final output softmax layer then maps the features to two gender classes, male or female.

A dataset of facial images was acquired for training from Kaggle [19], consisting of 19,000 facial images taken in uncontrolled conditions. Out of the total number of images, 4000 samples were used for training while 1000 were used for testing. During the testing phase, the authors compared the performance of their model with a pre-trained AlexNet [20] network. The proposed model produced an accuracy of 92.40%, while the AlexNet model produced a 90.50% accuracy. One of the observations reported by the authors is the reduction in the classification accuracy after training both models for more than 20 epochs. The reduction in accuracy could indicate possible overfitting since there was no mention of any additional steps to help the network generalise to new samples.

Moreover, the authors did not mention a pre-processing step to reduce the background noises that might affect the training process. In conclusion, current work focuses mainly on introducing a new algorithm or optimising an existing deep learning network such as AlextNet or the VGG network to estimate gender from a given facial image, with little to no studies on looking into the training samples. One of the significant gaps in existing work is that all gender estimation models depend heavily on facial images of actual individuals to learn the patterns of a face. This reliance on facial images for training can cause potential privacy or copyright issues. The proposed method could help minimise the need for actual facial images to train facial gender classifiers.

3 Method

The proposed method consists of two components. The first component is a pre-trained StyleGAN network that generates realistic-looking unlabelled facial images sized 1024 × 1024 pixels. The generated dataset is then labelled based on the observed gender using a set of criteria such as the shape of the jawline or the presence of facial hair. More details on the labelling part are discussed in Section 3.1. The second module is the gender classifier we train on the constructed dataset. The primary gender classifier is built and trained from scratch; however, we fine-tune and employ pre-trained VGG16 and VGG19 networks for further experiments. In Fig. 3, we provide a summarised overview of our method:

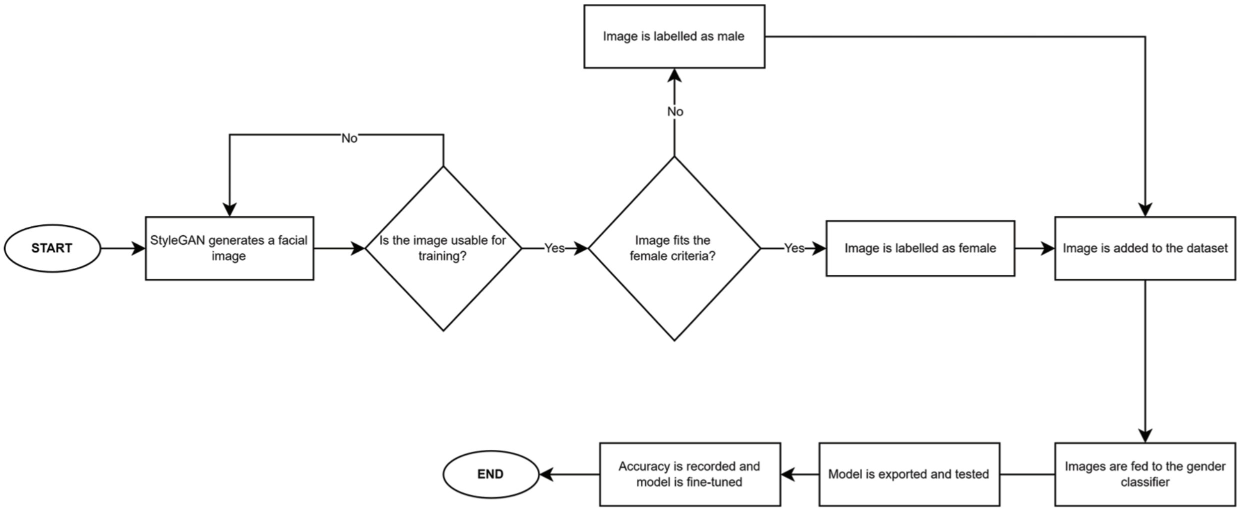
Image is labelled as male
No
No
START
StyleGAN generates a facial image
Is the image usable for training?
Yes
Image fits the female criteria?
Yes
Image is labelled as female
Image is added to the dataset
END
Accuracy is recorded and model is fine-tuned
Model is exported and tested
Images are fed to the gender classifier

Figure 3: A formalised overview of the proposed method

3.1 Dataset Formulation

We acquire the pre-trained StyleGAN network from the authors official GitHub page [21]. The StyleGAN network has already been trained on a large-scale facial images dataset denoted as Flickr-Faces-HQ (FFHQ). This dataset consists of 70,000 facial images crawled from Flickr, and only images with permissive licenses were scrapped. The images are of high quality, sized 1024 × 1024. The images are automatically aligned and cropped using the dlib library before being fed to the StyleGAN network for training.

The StyleGAN network plays a min-max game in which the generator consistently attempts to fool the discriminator by generating realistic-looking images. In contrast, the discriminator learns the features of the original images, takes the generator's image and produces a value that corresponds to the probability of whether the image is actual or synthesised. This process is mathematically represented as follows:

$$L(G) = \max[\log(D(x)) + \log(1 - D(G(z)))] \tag{1}$$

$$L(D) = \max[\log(D(x)) + \log(1 - D(G(z)))] \tag{2}$$

In Eqs. (1) and (2), the loss of the generator and discriminator is computed, respectively. D usually represents the discriminator, G represents the generator, and z is the latent vector that has been randomly generated and fed to G. The following formula is then produced when the equations above are combined to train the whole system.

$$\min\max L(D,\ G) = \min\max(E_{x\sim P_{data}(x)}[logD(x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z)))]) \tag{3}$$

Eq. (3) combines both losses obtained in Eqs. (1) and (2) in which $P_{data}(x)$ is the distribution of the actual data and $P_z(z)$ is the distribution of the noise input. By taking the derivative of Eq. (3), we can formulate the global optimality using Eq. (4), in which $P_g(x)$ is the distribution of the generated samples. This process is illustrated in Fig. 4.

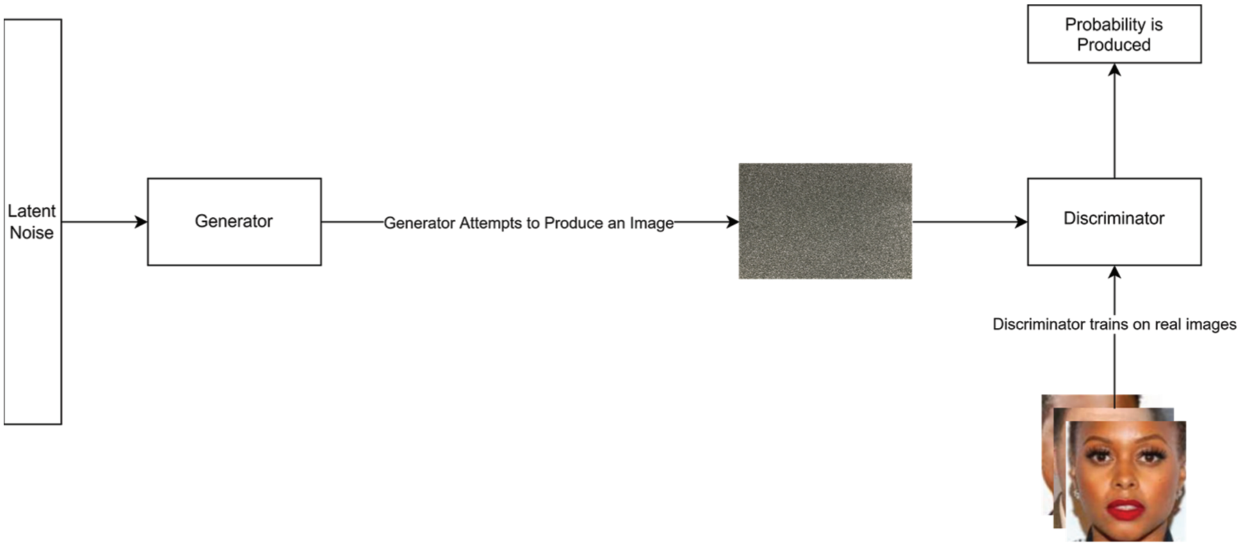
Latent Noise
Generator
Generator Attempts to Produce an Image
Discriminator
Probability is Produced
Discriminator trains on real images

Figure 4: A high-level overview of training a GAN network to produce facial images

$$-\frac{P_{data}(x)}{D(x)}+\frac{P_g(x)}{1-D(x)}=0$$

$$D_G^*(x)=\frac{P_{data}(x)}{P_{data(x)}+P_{g(x)}} \tag{4}$$

The network design of the StyleGAN generator is similar to that of Progressive Generative Adversarial Networks (ProGAN) [22], in which the input layer is a 4 × 4 convolutional layer, and it increases by a multiple of 4 until 1024 × 1024. Despite following the design of a ProGAN, there are a few distinct differences in the design of StyleGAN's generator.

The first difference is the addition of eight fully-connected layers of size 512 neurons, which takes the latent vector z as an input. The second difference is the addition of an intermediate latent space denoted as W. The W latent space takes the processed output of the last fully connected layer in the mapping network and concatenate it to the generator network. This latent space controls the generator's output using Adaptive Instance Normalisation (AdaIN) at every convolutional layer. The adaptive instance normalisation is defined in Eq. (5):

$$\mathrm{AdaIN}(\mathrm{x_i},\ \mathrm{y})=\mathrm{y_{s,i}}\ \frac{\mathrm{x_i}-\mu(\mathrm{x_i})}{\sigma(\mathrm{x_i})}+\mathrm{y_{b,i}} \tag{5}$$

where x represents the feature map, y is a style from which a biased scalar component denoted as $\mathrm{y_{s,i}}$ is obtained, μ is the mean, σ is the standard deviation and $\mathrm{y_{b,i}}$ is the bias parameter. We use the pre-trained StyleGAN generator above to produce more than 10,000 unlabelled facial images. After generating the facial images, we eliminate several unsuitable samples from the generated dataset. We define "unsuitable" samples as either facial images of toddlers where it is impossible to identify their facial gender or corrupted images where the generated face is entirely distorted. This cleaning process leaves us with just 11,009 images in total.

The next step is labelling each image in our dataset based on the observed gender. Since the StyleGAN-generated images are random and do not consider gender, age or ethnicity, we use a set of guidelines similar to those mentioned in medical publications like [23–25] to determine whether a subject is female or male based on a set of features. The features for both genders are detailed in Tab. 1.

Table 1: A set of requirements used to check whether a face belongs to a female or male subject

Gender	Features
Males	More prominent chin and jawline
	Facial hair existence
	Much angular face with stronger bones
Females	Less prominent jawline and chin
	Absence of facial hair
	Much more rounded face

Each image is judged by three different individuals based on the features in Tab. 1. Images that were too distorted or images of toddlers were excluded from the dataset as it was impossible to extract any gender information. Moreover, since gender-neutral subjects are hard to identify from the generated images, we treat the problem as a binary classification task, and this issue significantly limits the proposed method. In addition to using the features in Tab. 1, we consider subjects' overall appearance before assigning a class to an image. We provide a few samples of the generated images in Fig. 5. Before feeding the dataset to a gender classifier, the next and final step is pre-processing our images by cropping and rotating a detected face in a given image. We carry out this operation using dlib and OpenCV.

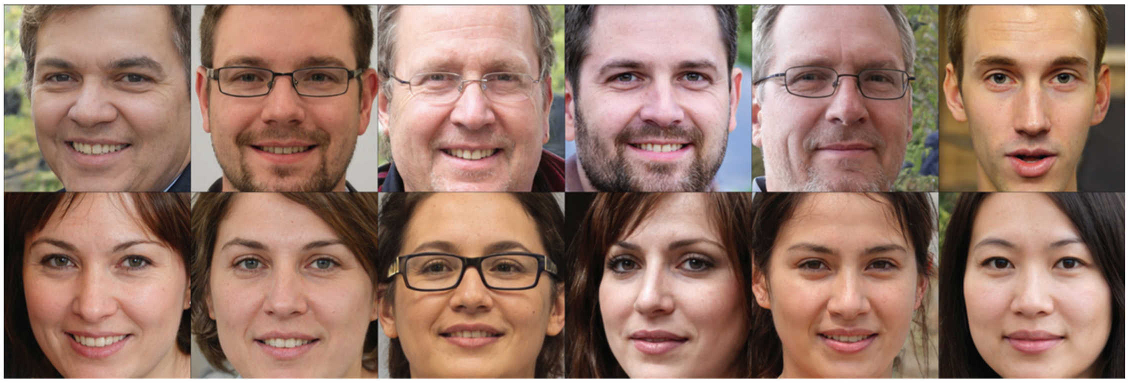

Figure 5: Randomly selected samples of the generated images. The top row is assigned the "males" class, while the bottom row is assigned the "female" class

3.2 Training Algorithm

We propose a lightweight binary classifier that consists of two hidden layers and one fully connected layer. We consider X to represent a set of training samples, where $X = \{x_1, x_2, \ldots x_i\}$ and Y is a set of matching labels where $Y = \{y_1, y_2, \ldots y_i\}$, our objective is to define a function in which an image x is mapped to a label y. This function can be represented as $\hat{y} = f(x_i)$, in which $\hat{y}$ is the predicted gender label of an image. The first portion of our network extracts the feature map using several convolutional and pooling layers. Each convolutional layer is activated using the Rectified Linear Unit (ReLU) function. The second portion of our model is a fully connected network with a single hidden layer that maps to a sigmoid output layer with two classes as defined in Eq. (6). The variable x in Eq. (6) represents the extracted feature map processed using the sigmoid activation function.

$$\mathrm{f(x)} = \frac{1}{1 + e^{-x}} \tag{6}$$

Since our classifier produces a binary output (e.g., female or male), we use the binary cross-entropy [26] as our loss function. This loss function is defined in Eq. (7):

$$l = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p(y_i)) + (1 - y_i)(\log(1 - p(y_i)) \tag{7}$$

where we denote y as the gender label, $p(y_i)$ is the probability of an image being one of a class. While $\log(1 - p(y_i))$ represents the probability of an image belonging to the opposite class for N number of samples. The images are resized to 64 × 64 to reduce the training time and complexity before feeding them to the network. The resizing step is essential because feeding the network a 1024 × 1024 image will increase the computational time and consume more resources. This input layer consists of 64 filters with a kernel size of 3 × 3, followed by a max-pooling layer with a pool size of 2 × 2. The subsequent hidden layers are defined as follows:

1) The first hidden layer comprises 64 filters, each with a kernel size of 3 × 3, followed by a max-pooling layer with a pool size of 2 × 2. A single dropout layer follows with a rate of 0.5.
2) The second hidden layer comprises 128 filters, each with a kernel size of 3 × 3, followed by a max-pooling layer with a pool size similar to the previous layer. A dropout layer is then added with a rate of 0.5.

The feature map is then flattened and fed to a dense network that consists of 256 neurons activated using the ReLU function and followed by a dropout layer of rate 0.5. The output layer consists of two neurons, each mapping to one class. The network architecture is presented in Fig. 6.

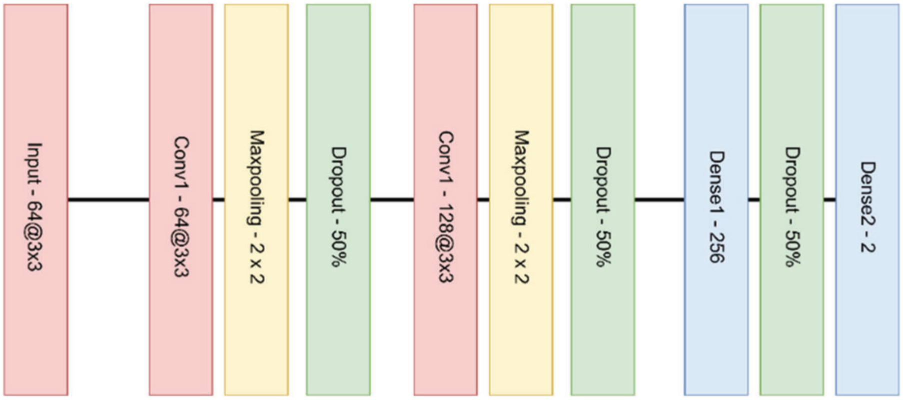

Figure 6: An overview design of our classifier

We train our classifier for 100 epochs with the early stopping mechanism to prevent overfitting. We use the adam optimiser and set the learning rate to 0.001 with a decay rate of 0.00001 and a batch size of 64. The early stopping stops the training process on the 62nd epoch and saves the best weight for further evaluation.

4 Experiments and Evaluation

This section presents the accuracy of our gender classifiers, the breakdown of the training dataset, and the breakdown of the testing datasets. We modified the VGG16 and VGG19 networks by replacing the last layer with a sigmoid layer with two outputs and the input layer to accept images sized 64 × 64 pixels.

4.1 Datasets

Our artificially generated dataset consists of 11,009 pre-processed facial images. We divide the dataset into 80% training and 20% testing. Out of the total number of images, 4841 are female subjects while 6168 are male subjects. Tab. 2 summarises the distribution of the samples.

Table 2: Overview of the distribution of our dataset

Number of images	11,009
Number of training samples	8807
Number of testing samples	2202
Number of male subjects	4841
Number of female subjects	6168
Training-to-testing ratio	80:20

We use two additional datasets to test our implementation. The details of these datasets are as below:

- **UTKFace dataset**: UTKFace is a large-scale facial image dataset consisting of more than 20,000 samples of subjects between 0 and 116 years old. In addition, the samples are all labelled with age, gender and ethnicity. We use 16,718 images from the total samples and use the gender label as the ground-truth label from each image.
- **Kaggle gender dataset**: This dataset is available for use on Kaggle, and it consists of more than 23,000 training samples and 11,000 validation samples. The dataset is a collection of celebrities facial images that have been rotated and processed. The images are taken in various conditions and illuminations. We use all 11,000 images from the validation sub-set to test our classifier.

4.2 Experimental Results and Analysis

We obtain the accuracy of our model by dividing the number of correctly classified samples over the number of testing samples. This testing metric is mathematically defined in Eq. (8).

$$\text{Accuracy} = \frac{n}{N} \ \text{ x } \ 100 \tag{8}$$

where n is the number of correctly classified images, and N is the total number of testing samples. Tab. 3 presents the accuracy of our classifier when tested on both the cross-validation datasets and the test portion of our generated dataset. In addition to our custom-built model, we present the accuracies of our pre-trained models in Tabs. 4 and 5. Our training and validation accuracies and losses are presented in Figs. 7–9.

Table 3: The testing accuracy of our model

Portion	Misclassified	Correctly classified	Accuracy
Generated dataset	115	2087	94.78%
UTKFace	1821	14897	89.10%
Kaggle gender dataset	993	10656	91.47%

Table 4: The testing accuracy of the VGG16 model

Portion	Misclassified	Correctly classified	Accuracy
Generated dataset	102	2100	95.36%
UTKFace	1131	15587	93.23%
Kaggle gender dataset	697	10952	94.01%

Table 5: The testing accuracy of the VGG19 model

Portion	Misclassified	Correctly classified	Accuracy
Generated dataset	103	2099	95.32%
UTKFace	1821	15057	90.06%
Kaggle gender dataset	761	10888	93.46%

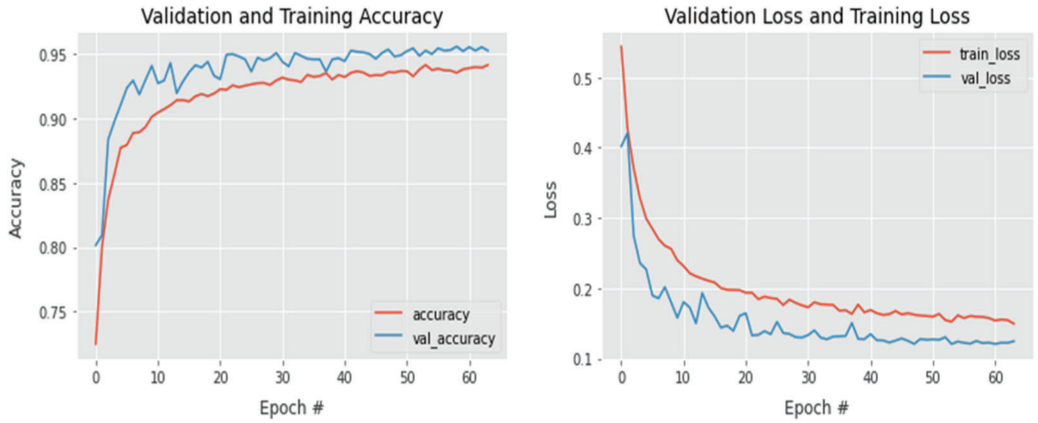

Figure 7: Training and validation accuracies (left) and loss (right) of our custom-built classifier

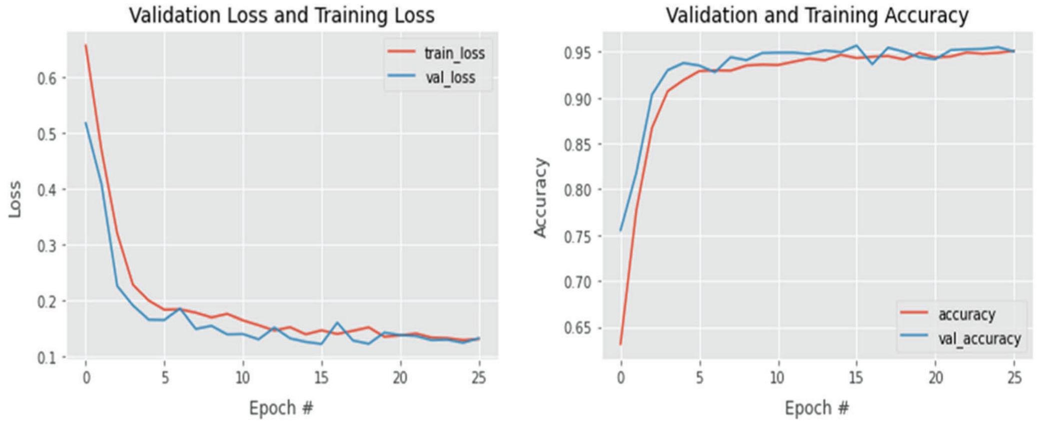

Figure 8: Training and validation accuracies (left) and loss (right) of the VGG19 network

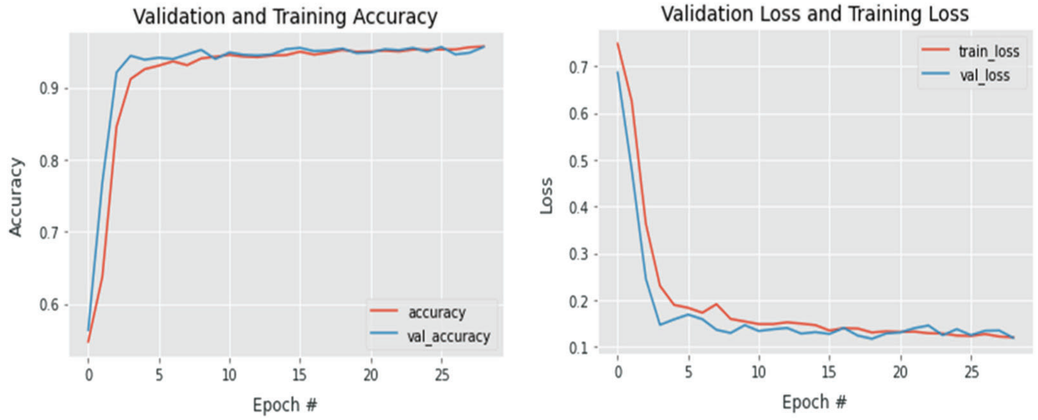

Figure 9: Training and validation accuracies (left) and loss (right) of the VGG16 network

Furthermore, a comparison of accuracies is presented in Tab. 6 to demonstrate how our method performs against existing state-of-the-art models. Moreover, in Tab. 7, we present our method's time complexity, which includes the time needed to generate one image and the required time for each classifier to process one image.

Table 6: The testing accuracy of the VGG16 model

Method	Accuracy
Tilki et al. [18]	92.40%
Alsellami et al. [27]	97.84%
Bekhet et al. [28]	89.00%
Mittal et al. [29]	93.68%
Ours (Custom model)	94.78%
Ours (VGG16)	95.37%
Ours (VGG19)	95.96%

Table 7: The time complexity of our proposed method

Operation (per image)	Average time (s)
Facial image generation	0.264
Custom model classification speed	0.042
VGG16 classification speed	0.046
VGG19 classification speed	0.051

4.3 Discussion

This section discusses and interprets the results obtained throughout our experiments to better understand the factors that influence our accuracy and the potential aspects that might impact future experiments. In addition, we present several unusual encounters that emerged during training and testing.

Although data collection and data quality are vital processes to any deep learning problem, especially facial recognition, most current work focuses on the model aspect of the problem rather than the dataset. The main difference between our approach and similar works is that more effort is given to the dataset aspect of the problem rather than optimising the model. According to the results presented in Tabs. 3–5, we can verify that training a model on a set of artificially generated images could produce nearly comparable accuracies to existing state-of-the-art gender estimation methods. The produced accuracies and the performance of each model verify that the labelling process was almost accurate but is still not the best labelling mechanism. An alternative solution would be crowdsourced labelling in which more than three observers judge the generated samples and cross-validators to ensure the accuracy of the labels.

During training, we find that fixing the number of epochs to a constant value overtrains all three networks, increasing the validation loss and decreasing each network's generalisation ability despite adding dropout layers throughout the network. We use early stopping as an additional preventive measure to decrease overfitting, which stops training on the 62^{nd} epoch. We suspect that the accuracy deteriorates without early stopping because the number of training samples is small; therefore, adding more samples might eliminate the need for early stopping and further enhance the model's accuracy. Moreover, we find that using image augmentation as a substitution to adding more samples improves our custom-build model's performance. The accuracy of our custom model improved by 1% when we introduced an augmentation layer before training. However, our findings prove that using a modified pre-trained model such as VGG16 can outperform a custom-built model, especially if the model has been trained on a similar facial analysis task. In addition to the improved accuracy, adding early stopping to our pre-trained models decreased the training time significantly, and each VGG network took around 23 epochs to train.

In Tab. 7, we present the time taken for the facial images generator to produce a single image and the time needed for the classifier to classify one input image. The presented results show that the image generation process is efficient since all the 11,009 images were generated in just 48 min at a rate of 0.264 s per image. In addition, each classifier took an average of 0.046 s to classify a single image with a more complex model like the VGG19 requiring slightly more time.

There are three main limitations to our method. Firstly, our training images are all of the frontal faces, which means that none of our classifiers is reliable enough to be used in real-life situations where facial images are taken in different conditions using different devices. Different conditions may include various illuminations, poses, or non-frontal images. To tackle this issue, we can either combine our generated dataset with challenging samples from other datasets such as Adience or MORPH datasets or modify the architecture of StyleGAN to generate challenging samples which simulate facial images captured in real-life scenarios. The issue of working with unrestrained images can be demonstrated in Tab. 3, in which we present the accuracy of our model attempting to classify more than 16,000 images. Out of the total number of testing samples, we see that our model could not classify 1821 samples. Most of the misclassified samples are either non-frontal facial images or images of low resolution.

The second limitation is that the generated samples are limited in diversity. Since the generated images are unlabelled, we would not extract information on the age or ethnicity distributions; however, from a high-level observation, we notice that the images are not ethnically diverse and slightly biased towards a single ethnic group.

The third limitation is the labelling process, which judges an image based on the observed gender. Our classifier is a binary model, and we assume that the StyleGAN-generated images are either males or females. This assumption significantly limits the model's ability to perform on gender-neutral subjects or individuals who neither identify as males nor females. There are no datasets or methods to build non-binary gender classifiers, making this issue an exciting area of research.

Given the presented work and results, we see a significant potential of using artificially generated images as training samples for various tasks such as age estimation, facial recognition or object detection. This mechanism can help with two issues. Firstly, artificially generated samples can be considered additional samples to be upscale an existing small-scale dataset. Secondly, we can potentially reduce several privacy issues with generated facial images since all training samples are from a GAN network and not collected from actual individuals.

5 Conclusion and Future Works

In this study, we propose a dataset of facia images that have been created from only artificially generated images using StyleGAN. We generate 11,000+ facial images then label them according to the observed gender. We then pre-process the images by locating and extracting the face in a given image to ease the process of training. The images are then used as a training dataset for a gender classifier to distinguish females and males apart. We present high accuracies that are almost comparable to existing state-of-the-art methods. These accuracies open the door for more research into potentially using StyleGAN-generated samples are training samples to minimise the need for actual facial images as much as possible, reducing probable privacy risks.

For future work, we are interested in finding a way to control the output of StyleGAN networks based on age, gender and ethnicity to produce samples based on a set of pre-defined conditions. We propose additional research into finding novel ways to generate images conditionally based on ethnicity, age or pose for the presented limitations. This proposal would potentially overcome the first and second limitations. As for the third limitation, we propose a crowdsourcing mechanism to get more opinions on the label of the generated images. Having more views on the gender of the subject in the image could enhance the accuracy of the ground truth label of the image.

Acknowledgement: The authors would like to thank Swinburne University of Technology (Sarawak Campus) for providing the necessary resources to carry out this study.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. K. Dubey and V. Jain, “A review of face recognition methods using deep learning network,” *Journal of Information and Optimization Sciences*, vol. 40, no. 2, pp. 547–558, 2019.

[2] N. S. Singh, N. Shanker, S. Hariharan and M. Gupta, “Facial recognition using deep learning,” in *Advances in Data Sciences, Security and Applications*, Springer, Singapore, pp. 375–382, 2020.

[3] P. Bery, “Ethical aspects of facial recognition systems in public places,” *Journal of Information, Communication and Ethics in Society*, vol. 2, no. 2, pp. 97–109, 2004.

[4] T. Karras, S. Laine and T. Aila “A Style-based generator architecture for generative adversarial networks,” in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, California, USA, pp. 4401–4410, 2019.

[5] KAGGLE, https://www.kaggle.com/cashutosh/gender-classification-dataset, 2019.

[6] Z. Zhang, H. Qi and Y. Song, “Age progression/regression by conditional adversarial autoencoder,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Hawai, USA, pp. 4352–4360, 2017.

[7] A. loannau, I. Tussyadiah and G. Miller, “That's private! understanding travelers’ privacy concerns and online data disclosure,” *Journal of Travel Research*, vol. 60, no. 7, pp. 1510–1526, 2021.

[8] Q. Bu, “The global governance on automated facial recognition (AFR): Ethical and legal opportunities and privacy challenges,” *International Cybersecurity Law Review*, vol. 2, no. 1, pp. 13–45, 2021.

[9] E. Eidinger, R. Enbar and T. Hassner, “Age and gender estimation of unfiltered faces,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[10] K. Vikram and S. Padmavathi, “Facial parts detection using Viola Jones algorithm,” in *4th Int. Conf. on Advanced Computing and Communication Systems (ICACCS)*, New Jersey, USA, pp. 1–4, 2017.

[11] M. K. Dabhi and B. K. Pancholi, “Face detection system based on viola-jones algorithm,” *International Journal of Science and Research*, vol. 5, no. 4, pp. 62–64, 2016.

[12] N. T. Deshpande and S. Ravishankar, “Face detection and recognition using viola-jones algorithm and fusion of PCA and ANN,” *Advances in Computational Sciences and Technology*, vol. 10, no. 5, pp. 1173–1189, 2017.

[13] A. C. Gallagher and T. Chen, “Clothing cosegmentation for recognising people,” in *IEEE Conf. on Computer Vision and Pattern Recognition*, Alaska, USA, pp. 1–8, 2008.

[14] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, USA, pp. 34–42, 2015.

[15] A. Dhomne, R. Kumar and V. Bahan, “Gender recognition through face using deep learning,” *Procedia Computer Science*, vol. 132, no. 2, pp. 2–10, 2018.

[16] M. Nakada and H. Wang, “AcFR: Active face recognition using convolutional neural networks,” in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Hawaii, USA, pp. 35–40, 2017.

[17] Z. Liu, P. Luo, X. Wang and X. Tang, “Deep learning face attributes in the wild,” in *Proc. of the IEEE Int. Conf. on Computer Vision*, Santiago, Chile, pp. 3730–3738, 2015.

[18] S. Tilki, H. B. Dogru, A. A. Hamdeed, A. Jamil and J. Rasheed, “Gender classification using deep learning techniques,” *Manchester Journal of Artificial Intelligence & Applied Sciences*, vol. 2, no. 1, pp. 126–131, 2021.

[19] KAGGLE, 2018, https://www.kaggle.com/ttungl/adience-benchmark?gender-and-age classification.

[20] A. Krizhevsky, I. Sutskever and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 2, no. 5, pp. 1097–1105, 2012.

[21] Tero Karras (2018), StyleGAN. https://github.com/NVlabs/stylegan.

[22] T. Karras, T. Aila, S. Laine and J. Lehtinen, “Progressive growing of GANs for improved quality, stability and variation,” arXiv preprint, arXvi:1710.10196, 2017

[23] V. Bruce, A. M. Burton, E. Hanna and P. Healy, “Sex discrimination: How Do We tell the difference between male and female faces?,” *Perception*, vol. 22, no. 2, pp. 131–152, 1993.

[24] E. Brown and D. Perret, “What gives a face Its gender?,” *Perception*, vol. 22, no. 7, pp. 829–840, 1993.

[25] A. M. Burton, V. Bruce and N. Dench, “What's the different between Men and women?,” *Perception*, vol. 22, no. 2, pp. 153–176, 1993.

[26] Z. Zhang and M. Sabuncu, “Generalised cross entropy loss for training deep neural network with noisy labels,” in *32nd Conf. on Neural Information Processing Systems (NeurIPS)*, Montreal, Canada, pp. 1–11, 2018.

[27] B. Alsellami and P. Deshmukh, “Gender recognition using deep learning convolutional neural network,” in *Int. Conf. on Innovative Computing and Communication*, New Delhi, India, pp. 355–364, 2022.

[28] S. Bekhet, A. Alghamdi and I. Eddin, “Gender recognition from unconstrained selfie images: A convolutional neural network approach,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 2066–2078, 2022.

[29] S. Mittal and S. Dana, “Gender recognition from facial images using hybrid classical-quantum neural network,” in *Students Conf. on Engineering and Systems*, Allahabad, India, pp. 1–6, 2020.