International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Jul-2022)

https://dx.doi.org/10.12785/ijcds/120111

Video Gaze Redirection Using Generative Adversarial Network (GAN)

Sikha Sunil[1], Sneha Johnson[2], Treasa Mani[3], Vishak J Nair[4] and Anjusree V.K[5]

[1,2,3,4,5]*Department of Computer Science and Engineering, Rajagiri School Of Engineering and Technology, Kochi, Kerala,682030, India*

Received 3 Jun. 2021, Revised 9 Mar. 2022, Accepted 15 Jun. 2022, Published 1 Jul. 2022

Abstract: Gaze correction is a type of video re-synthesis problem that trains to redirect a person's eye gaze into camera by manipulating the eye area. It has many applications like video conferencing, movies, games and has a great future in medical fields such as to experiment with people having autism. Existing methods are incapable of gaze redirection of video using GAN. The suggested approach is based on the in-painting model to read from the face and fill the missed eye regions with new contents, reflecting corrected eye gaze in this paper. Both gaze estimation as well as gaze redirection have been implemented . The Hourglass model of CNN was used for gaze estimation and the Generative Adversarial Network(GAN) for video gaze redirection, in which two neural networks compete in a game to learn and produce new data with the same statistics as the training set. In addition, various losses were estimated such as discriminator, generator loss and perceptual loss in order to determine the accuracy of our model and evaluate the performance by adversarial divergence, reconstruction error and image quality measures. We demonstrate that the proposed method outperforms in terms of quality of the image and redirection precision in comprehensive tests.

Keywords: Gaze estimation, Gaze redirection, GAN, CNN, Computer Vision

1. Introduction

In the twenty-first century, Technology has been an inevitable part of life and therefore the increasing demands for innovation to satisfy the deeds of the world lead to additional developments. The adequate development that has got to be dropped at the gaze is showing associate inclination in recent times even so solely few. Most of the Gaze Project uses a gaze-tracking software system to know how folks understand visual media though most of the connected projects have been used for eye gaze estimation, eye-gaze input, vision and communication; all the mentioned square measures enforced victimization of deep learning [1]. In this paper we've handled the improvisation that has been created to the gaze with regards to estimation and redirection.

In Every Social communication of the folks the Gaze estimation [2] and Gaze redirection [3] strategies are used as the core implementation of our project. The Eye-gaze technology [4][5] is one amongst the rising technologies that brought major upheavals within the Gaze projects.

The developments within the field of gaze were concerning studies relating to eye gaze and connected contents from wherever it really began. Additional works during this field created a scope for transformation. So far there have been studies and projects concerning the gaze estimation and redirection at intervals within the constraint technologies. Gaze and its application play an important role in cognitive science and non-verbal communication like emotions, expressions and attention. Gaze Estimation is a work to predict wherever an individual is staring at a given position. The task contains 2 directions: 3D gaze vector [5][6] and 2-D gaze position estimation. The Former is to predict the gaze vector that is typically employed in automotive safety. Later on a 2D screen it is to predict the horizontal and vertical coordinates which permits utilizing cursor to regulate gaze points for Human-Machine interaction. Also all the works thus far have been introduced within the video conferencing [7], films, games and relevant things dealt by humans to supply the photo-realistic pictures [2]. This could solely be achieved by generation of the training data for gaze estimation thereby implementing the gaze redirecting without twisting the area that surrounds the eye.

It is found out that developments within the gaze redirection with additional progress in innovations, introductions and plenty of projects have been developed with the present technologies[8]. The main and novel concept is to boost the adaptability by producing additional training samples by combining redirected eye images from existing used samples [3]. The current developments focus on the gaze redirection of video that is primarily implemented using the Gonvolutional Neural Network(CNN). There were additional works that happened identical, however none has been

E-mail address: sikhasunil007@gmail.com, snehapanamkattu@gmail.com, treasamani@gmail.com, vishak4u.vj@gmail.com, anjusreevk@rajagiritech.edu.in

developed through the means of Generative Adversarial Network(GAN) [9][10]. GAN is essentially a new scheme for estimating generative models via an adversarial operation, within which we tend to simultaneously train 2 models: the data distribution captured by a Generative Model G, and a sample derived from the training data instead of G which estimates the probability by Discriminate model D.

In this work we focused on implementing a completely unique plan of human-machine interaction [5]. Work on Gaze chiefly embraces detection, estimation and redirection. Initially live video [10] is being captured and kept stored for the processing. The process video can bear the estimation section wherever the face region is estimated [11] as a whole and thus the input vector of the corresponding gaze direction is identified and produced. Additional progress within the work results in the video redirection where the gaze is redirected to a specific angle and output is created from the estimated input.

2. Related Work

A. Real Time Implementation

To provide the real time inputs, frames are grabbed from a webcam. Use dlib [8] for detecting the face and a 5-point facial landmark. Although in-depth face recognition methods like VGG face applied to the VGG network [4], deep residual networks (ResNet) [12] have high efficiency, the models are compound and the speed of vision is slow. To see real-time face recognition in video reading and dlib [11][13][14] can be used. The dlib environment is configured and it can be used as a face recognition tool.

B. Gaze Estimation

There is a wide collection of works in Gaze-estimation methods. Brief analyses of some of the methods are specified below.

Feature-Based Gaze Estimation: It works with geometric concepts to make feature vectors that connect the shape of the eye and many other details, such as head pose and measuring direction of viewing [5]. It downgrades a set of features and feeds these into the simple, standard Machine Learning models to bring back the viewing rate. A common approach is the PC-EC vector (Pupil Center- Eye Corner vector) [14]. It can restore the traditional corneal expressions which are used for eye tracking.

Model Based Estimation: It attempts to match the well-known 3D model to the eye image [15] by reducing the appropriate energy. In the 3D-model based gaze-estimation [15][6] methods during user calibration procedures, the radius and center of the eyeball, the angular offset between optical axes and visual axes are determined. With relative to the facial landmark the eyeball center can be determined Conventional feature-based and model-based gaze-estimation methods have been shown to be effective in lighting and control of cameras. Due to the complexity of modeling features like visual elements and lighting changes, such methods outdo by appearance based methods. So we used a novel learning based method for detecting the eye landmark that makes former methods competitive with the latest appearance based methods.

C. Gaze Redirection

Gaze redirect approaches can break down mainly into novel view and monocular view synthesis.

Novel-view synthesis methods [7] re-render the captured user's face so it seems to be looking at the webcam. It provides a scene which contains the face of the subject from the input gaze view to get gazing at the webcam. These require a depth view of the face, and also need to integrate a new image of the subject with a redirected gaze by performing 3D transformation.For that it requires dedicated hardware and also needs to change the whole scene that limits their applicability.

Monocular-gaze synthesis [16] targets to change the gaze within the eye region. There are wraping based methods [2] which use deep neural networks or random forests but it cannot be used in the case where eyes are occluded, because without generating new pixels they only change pixels with existing pixels from the input image. Euclidean distance is used as an error metric but it does not precisely reflect the perceptual difference between images. There are approaches based on 3D modeling [3] .The 3D model is used to balance the shape and texture of the eye patch, and join eyeballs that are placed at the top of the input image.

D. Generative Adversarial Network

Generative Adversarial Networks [9] produces model distribution which mimics a given target distribution. In many applications like video prediction and generation [17] , image super resolution [18] , image in-painting and also in classical computer vision problems [19] this method has been adopted. GAN-based approaches like image-to-image translation [10] resulted in impressive outputs.

Our work connects to the works of gaze redirection in the image using GAN [20][21]. The work [22] generates the photo realistic image, without deforming the look of the eye and maintaining the desired gaze direction.

It ensures the consistency and perceptual similarity of synthesized images to the real images. Another work [22] proposes the image inpainting model, that uses a self-guided pretrained model to obtain angle invariant features and creates a self-guided generative adversarial network learning module to stabilize the adversarial training and to increase the quality of in-painted results.

3. Implementation

A. Proposed Approach

The overall workflow of our project model is as shown in Figure 1. The live video is captured from the respective webcam and passed to the eye landmark detection model followed by gaze estimation to estimate the gaze. The frames of the captured live video are extracted and stored

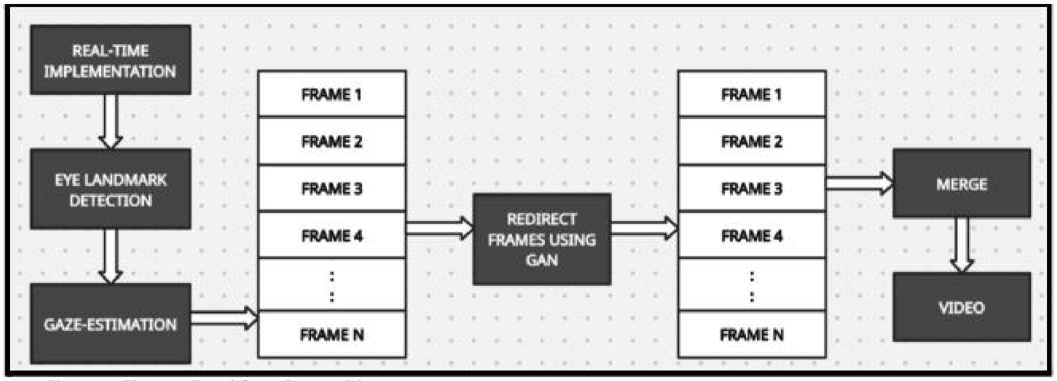


Figure 1. The overall workflow of our model

in a folder. These outputs from the gaze estimation along with the frames are passed for gaze redirection where respective frames are redirected. The redirected frames are then combined together to obtain the output of the redirected video.

B. *Real-time Implementation*

Initially a video is taken as an input. The process starts by extracting the face and eyes from the video. Here we use a real-time implementation [15][6] for input. We provided access to the webcam and collected the live video as frames of images 1280 × 720 RGB and a 5 point facial landmarks and face are detected using dlib. Using the detected eye corners two eyes are segmented and given as input to the gaze estimation model. Each captured frame and estimated gaze are stored and given as input for the gaze redirection.

C. *Gaze Estimation*

Gaze directions have a vital role in a person's cognitive science and for social communication. So that estimation would have a wide application for future innovations. The proposed method can provide accurate gaze estimation for different angular poses of eyes [12], besides it provides the head, eyes and all eye landmark information into the proposed gaze estimation framework. It uses a network that is pre-trained to encode the ocular information such as landmark detection of eyeball and its center, also endpoints of eyes. More than that it helps to find direction of the eye by using the coordinate points generated by eye position.

D. *Hourglass Model*

This is the architecture we used for eye landmark detection [23]. It is a type of Convolutional Neural Network which holds encoder and decoder networks. The Hourglass model uses different convolutional layers to break down and reconstruct the image. Here we load a video for transforming and later it encodes and decodes input to the feature matrix. On decoding we meld a feature matrix with an earlier layer of spatial information. Encoder will extract the features of the object by discarding all background pixels. Pixels are not involved in feature extraction [6], so they detach all details of object location. By integrating the feature matrix and spatial information to get a knowledge of input images or videos. In this model we use different types of layers namely max-pooling layers, convolutional layers, residual layers, bottleneck layers and unsampling layers.

Hourglass modules have feature maps that are Down-scaled via pooling operation, then Upscaled using bilinear interpolation. Here we use Hourglass architecture to detect landmarks of an eye image. About 64 refined feature maps are combined by the 1×1 convolutional layer to bring out 18 heatmaps where each represents a particular eye region landmark location. We use 3 hourglass modules with the residual layers on each stage to get an ideal eye landmark. Hourglass model contains 2 sections i.e. encoder and decoder. In between this encoder and decoder there are 3 bottleneck layers. It also performs element-wise addition with the output of decoder and inputs of network, that helps for loss calculation and passes on to the next Hourglass.

E. *GUI Interface*

The GUI is a Graphical User Interface used as a visual interactive component on the software system. Here we use the GUI to connect Gaze-Estimation and Gaze-Redirection. We use 'tkinter' for the implementation of the GUI. It is a standard Python GUI library. It provides an easy and fast way to make GUI applications and also a powerful object-oriented interface to the GUI toolkit.

F. *Gaze Redirection*

The frames extracted and the gaze position outputs from the gaze estimation method are given to the video gaze redirection method. Here each frame is given for the gaze correction using GAN. The architecture of our gaze redirector is as shown in Figure 2.

Our Gaze redirection model aims to manipulate the eye region and redirect the eye gaze into a single direction. It is based upon the inpainting model, studied from the facial image to fill in the deleted eye areas with a new eye region that describes the correct eye gaze. For that we use the fully convolutional network as the basis of our model with an image inpainting model.

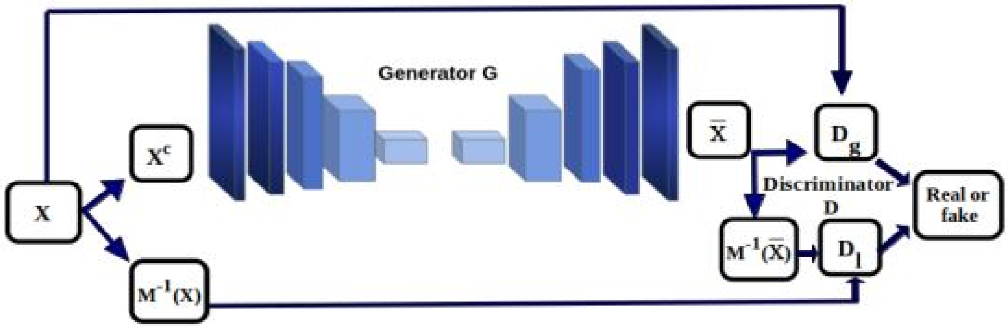


Figure 2. Architecture of our GAN model for the gaze redirector

Our gaze-redirection model consists of two parts, one generator and the other is a discriminator. Generator which aims in inpainting the network. We have a global discriminator, where the input to this network is the entire face, and we also have a local discriminator, where the input to this network is only the local eye-region i.e, to perform the adversarial learning with a generator.

1) **Generator G:** Generator is trained with two domains of images X and Y, each with dimension 256 x 256. Where X contains the images looking straight to the camera and Y looking somewhere else. Then a mask function is given to both the domains, where it crops out the eye region and results in X^c (masked image of X) and Y^c(masked image of Y). Then it trains the generator in such a way that it generates X from X^c and giving Y^c as input to the generator will output an element subset of X. Generator G is designed as an auto-encoder which uses fully connected networks, to fill the removed eye region with the redirected eye gaze.
2) **Discriminator D:** There are two discriminators D_g and D_1, both are convolutional networks. The main aim of the discriminator is to avoid the reconstruction loss of the output image from the generator. It performs adversarial loss to increase visual quality of inpainted output images. Where the global discriminator takes entire image X and the generated an image $\overline{X}$ as input to check the generated image is coherent to the global region. And the local discriminator takes the local eye region, which is the masked image M^{-1} and masked image of the generated image $M^{-1}(\overline{X})$ as input to ensure realistic output.
The output from both these discriminators are connected together and given as the input to the sigmoid function to determine whether the image is fake or real. This results in the probability of the output being real.

4. Experimental Methods

A. Datasets

The Proposed model has been generated by several training and testing through the dataset. Though the MPII dataset provides sufficient data for the evaluation; we have used another dataset which was found to be more feasible later. Despite the complexity in using the GAN further leads to the adoption of the NewGazeDataset where the implementation was processed. NewGazeDataset provides the required images for both testing and training which helps for eye landmarks for redirection. Out of the total images within the dataset; around 5000 used for the testing purposes and 25000 images for training.

But considering fact of redirection of video can output the frames inpainting different eyes in each frame or loss of originality in biological eye regeneration, we added our own dataset containing user images to the existing dataset (Figure 3). The dataset is obtained in the same manner as that of NewGazeDataset. Implementing the same functionality for every user will be discussed in section 6.

B. Experiments

The evaluation for the model has been implemented in such a way that as soon as the live video is captured and recorded as input; it is divided into a set of frames where each frame acts as a point of instance of the video. The live video is recorded in the webcam having 60 fps. The Input video obtained using OpenCV and DLib. Both these help in the generation of facial key points detectors which can detect eyes in real time. Moreover predictions in real time can be obtained through the Dlib. It contains a pretrained network from which it was able to detect eye landmarks. Captured the frames in aspect ratio 128x720 and given to gaze estimation, where the faces of images hourglass model and estimate the eye position and gaze direction. The frames captured earlier are stored as a cropped size of 256x256 in the image file with a corresponding frame name. Each frame is identified through this frame name and is used for the manipulation of eye redirection one after the other. The redirected frames are then joined together to output redirected video.

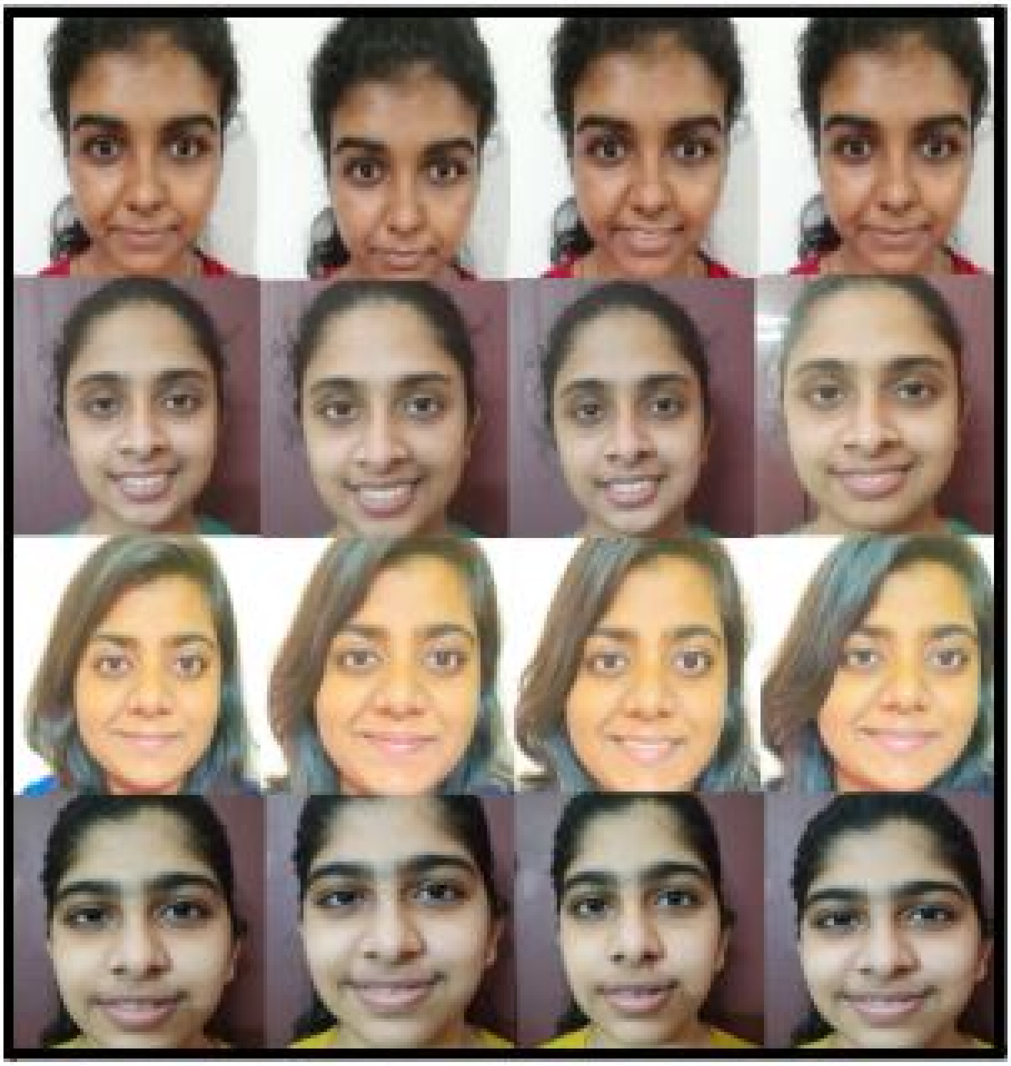


Figure 3. Some pictures from the added dataset

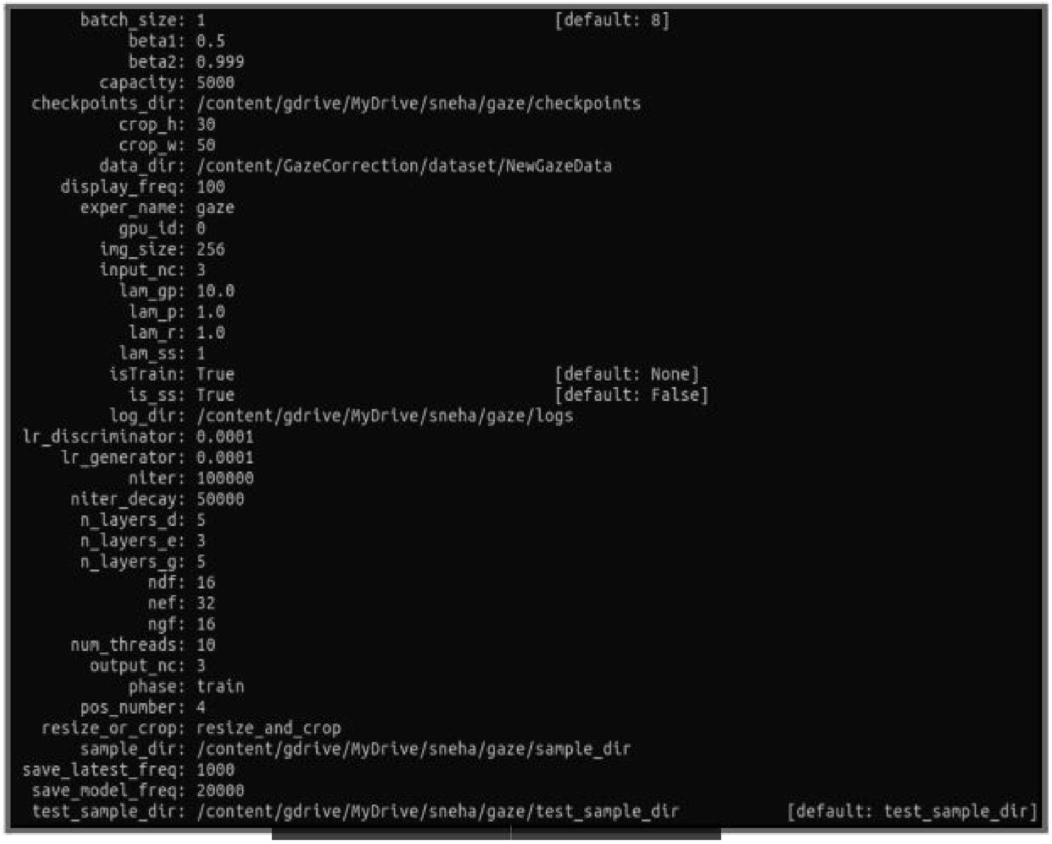
C. Model

For the model the requirements are a gpu system with python version 3.5 , tensorflow of version 1.14, scipy version 1.2 , numpy version 1.18.1, dlib and opencv-python.

The detailed view of the GAN model is shown in Figure 4. Scales the image size to 256x256. Crop the eye region with the size of width 50 and height 30. The initial learning rate for adam in discriminator and generator is taken as 0.0001 . Exponential decay rates for first and second moment estimation, ie beta1 and beta2 for adam, are taken as 0.5 and 0.999 respectively. Weight for the gradient penalty of GAN is taken as 10 and the weight for perceptual loss of generator as 1.0 and the weight for the reconstruction loss for generator as 1.0. Number of encoder filters in the first deconvolutional layer is taken as 32 . Whereas the number of generator filters in the first convolutional layer is 16 and also the number of discriminator filters in the first convolutional layer is 16. Kernel size of 7x7 with stride size of 1x1 and kernel size of 4x4 with stride size of 2x2 are used for the generator, and for the discriminator kernel size of 5x5 with stride size of 2x2. Number of layers of sparse autoencoder is taken as 3, whereas number of layers of layers of generator as 5 and also number of layers for discriminator as 5. Figure 4 describes the model with a vast view.

D. Training Details

Using this proposed model; we tried to demonstrate live gaze redirection using the GAN method. Basically the implementation consists of testing and training of the model from which the images in the MPII datasets used for gaze-estimation and NewGazeDataset for gaze-redirection. While obtaining the outputs as frames there were errors related



```
       batch_size: 1                                    [default: 8]
            beta1: 0.5
            beta2: 0.999
         capacity: 5000
  checkpoints_dir: /content/gdrive/MyDrive/sneha/gaze/checkpoints
           crop_h: 30
           crop_w: 50
         data_dir: /content/GazeCorrection/dataset/NewGazeData
     display_freq: 100
       exper_name: gaze
           gpu_id: 0
         img_size: 256
         input_nc: 3
           lam_gp: 10.0
            lam_p: 1.0
            lam_r: 1.0
           lam_ss: 1
          isTrain: True                                 [default: None]
            is_ss: True                                 [default: False]
          log_dir: /content/gdrive/MyDrive/sneha/gaze/logs
lr_discriminator: 0.0001
     lr_generator: 0.0001
            niter: 100000
      niter_decay: 50000
       n_layers_d: 5
       n_layers_e: 3
       n_layers_g: 5
              ndf: 16
              nef: 32
              ngf: 16
      num_threads: 10
        output_nc: 3
            phase: train
       pos_number: 4
   resize_or_crop: resize_and_crop
       sample_dir: /content/gdrive/MyDrive/sneha/gaze/sample_dir
 save_latest_freq: 1000
  save_model_freq: 20000
  test_sample_dir: /content/gdrive/MyDrive/sneha/gaze/test_sample_dir        [default: test_sample_dir]
```

Figure 4. Detailed structure of the model.

to the accuracy from which the biological eye regeneration causes errors resulting in low accurate outputs. Considering it as a collective frames; each error in the frames stood as a severe implementation fault which is supposed to be obtained as a video. Thus frame manipulation for the intended output was dealt through the generation of a personalised dataset of the user where more accurate and live redirected video is obtained as output. Though obtaining an output through such a method can be considered as a complex procedure; we have made an implementation algorithm for maintaining the accuracy and regeneration of the redirected output in such a way that a new dataset is added to the existing NewGazeDataset. By doing so; the accuracy is maintained and the loss can be minimised to an extend; producing a well defined redirected output.

5. Performance Estimation

Twenty four methods are there [21], for estimating and evaluating the accuracy of GAN models [24]. Among them we used three methods to evaluate the accuracy of output from our model.

To calculate the performance of our model we adopted three measures [21] adversarial divergence [25], reconstruction error [26] and image quality measures [27].

A. Adversarial Divergence

Computed the Kullback–leibler divergence [25] between the generated data distribution $P_g(y \mid x)$ and real data distributions $P_r(y \mid x)$. The low value for this matrix is zero, so $P_g(y \mid x) = P_r(y \mid x)$. Which means, less the value of adversarial divergence and that is closer to the two distributions And we obtained divergence value to zero in the epoch 1500. That is the difference between the generated data distribution and real data distribution is zero, which shows our model produces high accurate output. Adversarial divergences obtained in each 500 iterations are shown in Table 1.

B. Reconstruction Error

Measures the reconstruction loss [26] between the closest generated image and test image. We use L1 distance between output image X and ground truth image $\overline{X}$. As shown in graph Figure 5 we obtained the reconstruction loss to be less than 0.1 from epoch 1500 and further iterations tends the value to zero. That is in 1500 epochs our model is able to output with reconstruction error $< 0.08 \sim 0$. Which shows our model is able to generate the output with least reconstruction loss.

TABLE I. ADVERSARIAL DIVERGENCE IN EACH 500 ITERATIONS

Divergence	0	500	1000	1500
Left eye	0.0383	0.0003	0.0002	0
Right eye	0.0520	0.0009	0.0002	0

TABLE II. GENERATOR AND DISCRIMINATOR LOSS IN EACH 500 ITERATIONS

Loss	1000	2000	3000	4000	5000	6000
Discriminator	1.513	1.429	1.389	1.329	1.28	1.18
Generator	0.940	0.927	0.787	0.786	0.78	0.74

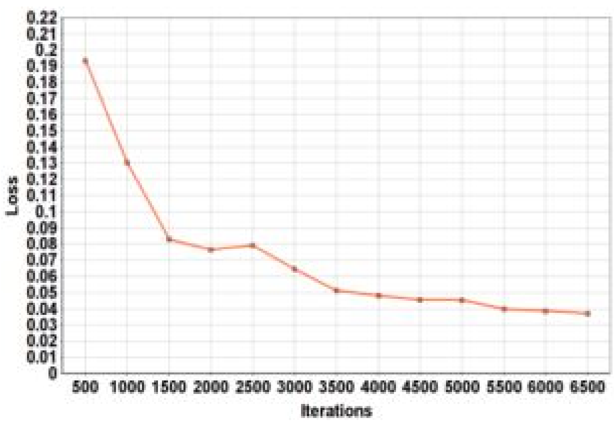


Figure 5. Graph which shows the reconstruction loss of our model with respect to the iterations

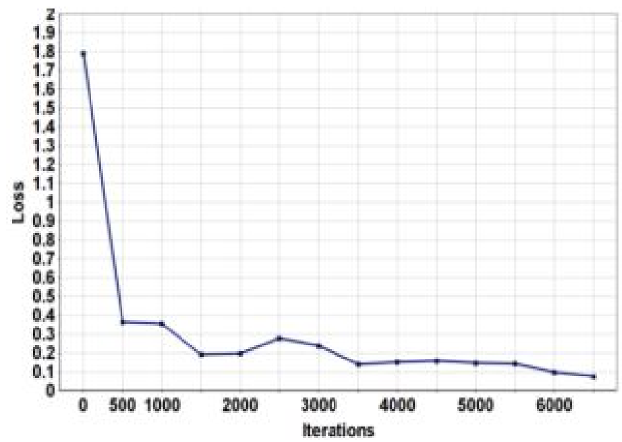


Figure 6. Graph which shows the perceptual loss of our model with respect to the iterations.

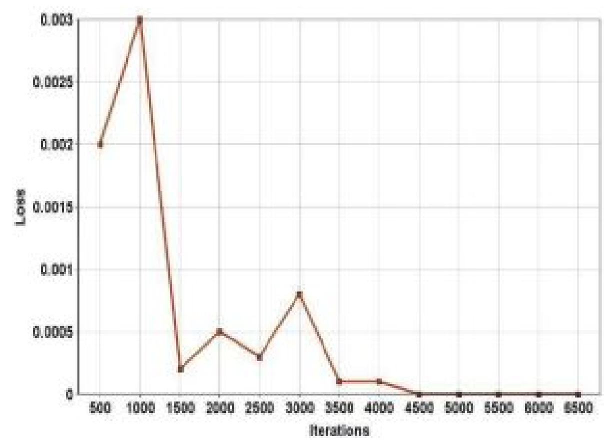


Figure 7. Graph showing the real class loss with respect to iterations

C. Image Quality Measure

We used perceptual loss [27] between the real image and generated image to measure the image quality. Where it compares the high level differences like pixel variations, style and content discrepancies or any other minor misalignment between the generated and real images. As shown in graph Figure 6, we obtained the perceptual loss that tends the value to zero. Hence it shows that our model generate output with high quality and least perceptual loss.

Other loss functions we obtained for our model are generator loss, discriminator loss (Table 2), real class loss and fake class loss for each iteration (Figure 7, Figure 8). And it clearly shows that our model outputs very least losses.

To the best of our knowledge this is the first approach towards the video gaze redirection using generative adversarial network. There are models for video gaze redirection using other than GAN, as mentioned in section 2 [3][4]. But GAN models learn the distribution much quicker and produce more accurate output than other methods. That is, GAN models outperforms other methods [22][20][27].

6. Results

Figure 8 shows the output obtained when some of the consecutive frames were given as the input to the gaze redirection using GAN. First column of the Figure 9 shows input frames, whereas the second column is a masked image of the input image which is then given for the inpainting done by the trained generator and the third

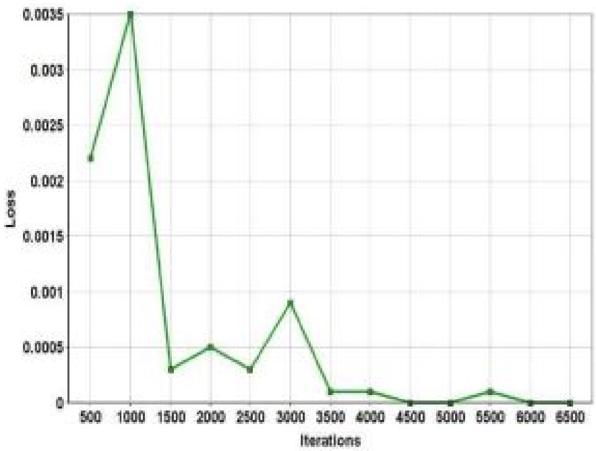


Figure 8. Graph showing the fake class loss with respect to iterations

column represents the redirected output image respectively.

These redirected output frames are then combined together to form the video. This is the output of our model. A screenshot from our output video is shown in Figure 10. Where the left half side of the video shows the input and whereas the right side is the redirected output.

7. Future Works

From this work, we explicitly learn landmark-based features and molecular gaze redirection. However, more research in future is needed to completely understand the best methods to store and utilize calibration samples while improving the gaze-estimation and redirection accuracy. It should have a great future with the project that comes up with gaze related work such as video conferencing, video callings etc. The hope is, our work will enable us to implement the commonly used cameras on devices such as tablets, smart phones and laptops as inbuilt software. It should have a great future in medical fields such as to experiment with people having autism.

Establishing our model as a live video gaze redirection software in future, which enables to captures the user's images while signing up and add those images to the existing dataset and train the model in due time helps our project model to be useful to any users who sign up to output their redirected video with more accuracy.

8. Conclusion

In this paper, we perform a live video redirection using a generative adversarial network. We used a personalized gaze estimation method using pre trained eye gaze model with capsule layers, the hourglass model of CNN and a gaze redirection method leveraging GAN. The proposed method can bring about redirected video in real time implementation. To the best of our knowledge our model is the first approach to the video gaze redirection using GAN. The extensive

Figure 9. Some of the redirected frame outputs of consecutive frames extracted from the video.

Figure 10. .Screenshot from output video. Left side of the video shows the input and the right side shows the output

evaluation measure shows that our model outperforms the other methods of video redirection in terms of accuracy of both redirection precision and image quality. Establishing this model in future can help in video conferencing and many other applications.

Acknowledgment

The authors are thankful to the Department of Computer Science and Engineering of Rajagiri School of Engineering and Technology for the guidance and support to accomplish the project

References

[1] P. A. Punde, M. E. Jadhav, and R. R. Manza, "A study of eye tracking technology and its applications," in *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, 2017, pp. 86–90.

[2] Y. Ganin, D. Kononenko, D. Sungatullina, and V. Lempitsky, "Deepwarp: Photorealistic image resynthesis for gaze manipulation," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 311–326.

[3] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and A. Bulling, "Gazedirector: Fully articulated eye gaze redirection in video," *Computer Graphics Forum*, vol. 37, no. 2, pp. 217–225, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13355

[4] H. Zhao, J. Liu, and W. Wang, "Research on human behavior recognition based on video key frame," in *The 2nd International Conference on Computing and Data Science*, ser. CONF-CDS 2021. New York, NY, USA: Association for Computing Machinery, 2021.

[5] M. X. Huang, T. C. Kwok, G. Ngai, H. V. Leong, and S. C. Chan, "Building a self-learning eye gaze model from user interaction data," in *Proceedings of the 22nd ACM International Conference on Multimedia*, ser. MM '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1017–1020.

[6] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and A. Bulling, "A 3d morphable eye region model for gaze estimation," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 297–313.

[7] C. Kuster, T. Popa, J. Bazin, C. Gotsman, and M. Gross, "Gaze correction for home video conferencing," *ACM Transactions on Graphics*, vol. 31, no. 6, Nov. 2012.

[8] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, p. 1755–1758, dec 2009.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014.

[10] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014. [Online]. Available: http://arxiv.org/abs/1411.1784

[11] M. Xu, D. Chen, and G. Zhou, "Real-time face recognition based on dlib," in *Innovative Computing*, C.-T. Yang, Y. Pei, and J.-W. Chang, Eds. Singapore: Springer Singapore, 2020, pp. 1451–1459.

[12] S. Tathe, A. Narote, and S. Narote, "Face recognition and tracking in videos," *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, pp. 1238–1244, 07 2017.

[13] S. Praveenkumar, K. J. Varma, Subramanya, and A. VenkataSaiHarish, "A multiple face recognition system with dlib's resnet network using deep metric learning," 2020.

[14] L. Sesma, A. Villanueva, and R. Cabeza, "Evaluation of pupil center-eye corner vector for gaze estimation using a web cam," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 217–220.

[15] K. Wang and Q. Ji, "Real time eye gaze tracking with 3d deformable eye-face model," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1003–1011.

[16] L. Wolf, Z. Freund, and S. Avidan, "An eye for an eye: A single camera gaze-replacement method," 07 2010, pp. 817 – 824.

[17] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8798–8807.

[18] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[19] X. Wang, A. Shrivastava, and A. Gupta, "A-fast-rcnn: Hard positive generation via adversary for object detection," *CoRR*, vol. abs/1704.03414, 2017. [Online]. Available: http://arxiv.org/abs/1704.03414

[20] J. Zhang, M. Sun, J. Chen, H. Tang, Y. Yan, X. Qin, and N. Sebe, "Gazecorrection: Self-guided eye manipulation in the wild using self-supervised generative adversarial networks," *CoRR*, vol. abs/1906.00805, 2019. [Online]. Available: http://arxiv.org/abs/1906.00805

[21] A. Borji, "Pros and cons of GAN evaluation measures," *CoRR*, vol. abs/1802.03446, 2018. [Online]. Available: http://arxiv.org/abs/1802.03446

[22] Z. He, A. Spurr, X. Zhang, and O. Hilliges, "Photo-realistic monocular gaze redirection using generative adversarial networks," *CoRR*, vol. abs/1903.12530, 2019. [Online]. Available: http://arxiv.org/abs/1903.12530

[23] E. Wood and A. Bulling, "Eyetab: Model-based gaze estimation on unmodified tablet computers," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 207–210.

[24] S. Barua, S. M. Erfani, and J. Bailey, "FCC-GAN: A fully connected and convolutional net architecture for gans," *CoRR*, vol. abs/1905.02417, 2019. [Online]. Available: http://arxiv.org/abs/1905.02417

[25] J. Yang, A. Kannan, D. Batra, and D. Parikh, "LR-GAN: layered recursive generative adversarial networks for image generation," *CoRR*, vol. abs/1703.01560, 2017. [Online]. Available: http://arxiv.org/abs/1703.01560

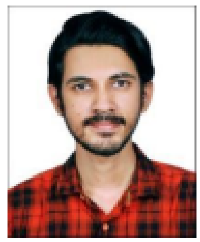
[26] S. Xiang and H. Li, "On the effects of batch and weight normalization in generative adversarial networks," 2017.

[27] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *CoRR*, vol. abs/1603.08155, 2016. [Online]. Available: http://arxiv.org/abs/1603.08155

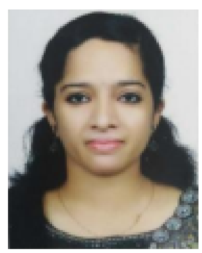
Sikha Sunil is a B.Tech (2017-21) graduate in computer science and engineering. She completed her Btech degree from Rajagiri School of Engineering and Technology under Kerala Technological University (KTU). Her project topics of interests are soft computing, web development, image processing and information security.

Sneha Johnson is a B.Tech (2017-21) graduate in computer science and engineering. She completed her Btech degree from Rajagiri School of Engineering and Technology under Kerala Technological University (KTU). Her project topics of interests are soft computing, web development, machine learning and information security

Treasa Mani is a B.Tech (2017-21) graduate in computer science and engineering. She completed her Btech degree from Rajagiri School of Engineering and Technology under Kerala Technological University (KTU). Her project topics of interests are soft computing, web development, machine learning and information security Rajagiri School of Engineering and Technology under Kerala Technological University (KTU). Her project topics of interests are soft computing, web development, machine learning and information security



Vishak J Nair is a B.Tech (2017-21) graduate in computer science and engineering. He completed his Btech degree from Rajagiri School of Engineering and Technology under Kerala Technological University (KTU). His project topics of interests are soft computing, web development, image processing, and information security.



Anjusree V.K is an Assistant professor of the department of computer science and engineering at Rajagiri School of Engineering and Technology. She did her B.Tech in computer science and engineering from Sree Buddha College of Engineering in 2013 and M.Tech in computer science and engineering from the same institution in 2015. She has been working in Rajagiri School of Engineering since 2017. Before joining Rajagiri she worked in Sree Narayana Gurukulam College of Engineering, Kolencherry. She is an active member in IEEE. She has published technical papers in international and national conferences.