Vol.2022-CG-188 No.1
Vol.2022-DCC-32 No.1
Vol.2022-CVIM-231 No.1
2022/11/18

3D Face Reconstruction-based Augmentation for Gaze and Head Redirection

Jiawei Qin[1,2,a)] Xueting Wang[2,b)]

Abstract: Gaze and head redirection is to change the gaze and head orientation of a full-face image to a target direction. Instead of only changing face direction in existing generation tasks, controllable both gaze and head redirection containing more complicated information is also critical for creating expressive face image. However, the redirection precision of existing gaze and head redirection models critically degrades once the target direction goes out of the model's capable range, which is limited by the angle range of the training data. In this paper we propose to use monocular 3D face reconstruction as data augmentation to extend the redirection range of the existing limited real data. The augmentation can largely extend the head pose range by rotation while preserving the original gaze information of real data. Consequently, the range of the head pose and gaze can both be extended. Experiments show that the proposed data augmentation significantly improved the redirection performance especially when redirecting to a relatively large angle while keeping the image quality.

Keywords: Gaze and head redirection, data augmentation, 3D face reconstruction.

1. Introduction

In recent years, gaze has gained more and more attention in the computer vision community due to its significance in human behavior. Although estimating gaze direction has been making great progress from model-based to the current appearance-based approaches using deep neural network [6], [9], [33], redirecting gaze and head of a given image becomes an even more important topic due to the growing demand for face synthesis and face editing in applications such as avatar in online advertising. Previous works redirected the gaze based on image-warping [5] or combining graphics [28]. Generative adversarial network (GAN) is also introduced to achieve better photo-reality [11]. While these methods can only redirect gaze, learning-based face synthesis has made great progress in simultaneously redirecting head and gaze [35].

Learning-based face synthesis requires gaze and head pose labels during training. It is believed that the learning-based generative model cannot precisely redirect to a large angle that is out of the range of the training dataset, and therefore collecting comprehensive training data is the main challenge in gaze redirection. Specifically, while head pose labels are relatively easy to get, sophisticated devices are required to obtain accurate gaze labels. Previous work has made great efforts to collect accurate gaze from in-the-wild settings [18], [34]. Although they obtained a diverse environment and large subject scale, the gaze is naturally limited by the screen size. On the other hand, extreme direction data can be collected in lab-controlled settings [31], but the high cost makes it difficult to reach a large subject scale. To address the above challenges, we propose to create synthetic data as data augmentation to extend the angle range of existing limited real data.

Since gaze is subtle information in the face image, the augmentation data must contain an accurate gaze label, which is defined as a 3D vector in physical space. The GAN-based face generation approach cannot be directly applied since it manipulates head pose in latent space instead of real physical space [13], [17], [25], [26], [27]. Besides, the synthesized image is also not reliable and few GAN-based works have considered gaze feature manipulation. On the other hand, 3D face reconstruction has been used to enhance face recognition [36] and gaze estimation [21] because its sampled texture preserves the original gaze feature. The 3D face can be rotated in 3D space without estimation error, and therefore we decide to use this approach to augment the learning-based head and gaze redirection method [14], [35].

In conclusion, we adopt a state-of-the-art head and gaze redirection method, and create augmentation data to extend its capable redirection angle range. Our contribution can be summarized as

- We hypothesized that the angle range in training data hugely influenced the model's redirection range and proved it by experimental analysis.
- We create augmentation data with a much larger angle range by 3D face reconstruction.
- The experimental results proved that the 3D face reconstruction-based augmentation is effective for extending the angle range of existing limited datasets for face and gaze redirection.

[1] The University of Tokyo, Meguro, Tokyo 153–0041, Japan
[2] CyberAgent, Inc., Shibuya, Tokyo 150–6121, Japan
[a)] jiawei_qin@cyberagent.co.jp
[b)] wang_xueting@cyberagent.co.jp

In this paper, we will introduce the dataset settings and experimental settings in section 3 used in the following hypothesis validation and augmentation effect evaluation. In Section 4, we first analyze the influence of training data's angle range using an existing large-range real dataset. In Section 5, we introduce the augmentation method and create the augmentation data from narrow real data for redirection tasks to evaluate the effect of augmentation.

2. Related Works

2.1 GAN-based Face Editing

Generative adversarial networks (GAN) have been widely used in face editing to generate images with high photoreality [15], [16], [17], [25], [26]. There exist high-quality training datasets such as FFHQ [17], CelebA-HQ [16] allowing the model to edit face expression, age, glasses, pose, etc. Many previous works trained a GAN that can change the pose of a given image [13], [17], [25], [26], [27]. Most of the previous works were designed to manipulate the head pose in the latent code space for generating an image of different head poses. However, head and gaze redirection is quite a different task that is determined to redirect images based on physical direction, which requires labels during training. Therefore, the GAN-based method is difficult to apply directly on head and gaze redirection. In addition, the gaze is much more subtle information so the existing GAN-based method cannot handle it.

2.2 3D Face Reconstruction

The rapidly progressing monocular 3D face reconstruction led to the improvement of other face-related tasks such as face recognition [19], [29], [36]. 3DM Morphable Model (3DMM) [2] is a parameterized model that has been widely used in 3D face reconstruction [1], [3], [4], [22]. Although 3DMM is becoming robust in reconstructing the face shape, the predicted texture cannot preserve the authentic information of the input 2D image, especially for the subtle gaze appearance. Another line of work directly sampled RGB values from the input image to faithfully keep the original appearance. Zhou *et al*. rotated the 3D reconstructed face to create more pose-diverse training data for face recognition [36], and Qin *et al*. rotated the 3D reconstructed face to extend the gaze range for more pose-robust gaze estimation [21]. The proposed projective-matching [21] process transforms the pixel-space 3D face to the camera coordinate system such that the 3D face can be rotated and translated inside 3D physical space. The 3D reconstructed face can be arbitrarily manipulated and rendered to new images, and it does not require any extra training process.

2.3 Gaze and Head Redirection

Gaze direction is defined as the 3D vector that starts from the center of the eye or face to the gaze target in the camera coordinate system. Gaze redirection started from the eye-only image by deep warping [5] method to GAN [11]. Although there exists full-face gaze redirection using the 3D eyeball model [28], the model-based approaches cannot redirect head or handle glasses, although they do not need training data.

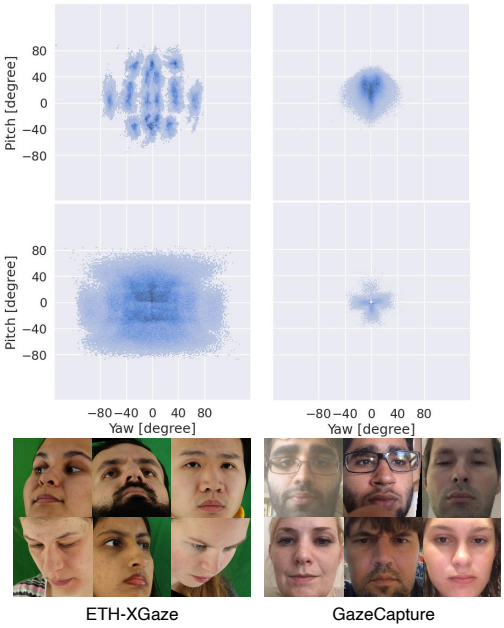

Fig. 1 The distribution and samples of existing real datasets. The top row is the head pose distribution, and the middle row is the gaze distribution. The bottom row is the image samples of the datasets.

Currently, learning-based full-face gaze and head redirection [14], [35] becomes a more significant topic, and head pose and gaze direction labels are needed for training the redirection model. Park *et al*. [20] proposed an encoder-decoder model that disentangles the appearance, head, and gaze features and transforms the source embeddings to target by rotation. During training, the paired images are used such that the model learns the transformation supervised by the head pose and gaze label. ST-ED [35] focused on improving the image quality performance and took into consideration extraneous factors such as illumination and hue by introducing an unsupervised self-learning pipeline. Consequently, the model enables redirecting a source image to a target image or any target angle.

The related works are summarized in Table 1. Since our task is to redirect the head and gaze based on the physical direction, GAN-based methods are not suitable. ST-ED is a state-of-the-art model in head and gaze redirection, but its ability to redirect to a very large angle is not verified. There is a lack of training data with a large subject scale and angle range, and collecting such data is not trivial. On the other hand, 3D face reconstruction can create an arbitrary amount of face-rotated images with accurate angles without extra training. Therefore, we decide to use ST-ED as the baseline synthesis model but use 3D face reconstruction to render images as augmentation data to extend the limited angle range of real data, such that the ST-ED model can be more robust in large angles.

3. Datasets and Experimental Settings

In this section, we will introduce the used datasets and experimental settings used for the following analysis.

Table 1 Comparison of related works of .

Methods	Pros	Cons
GAN-based redirections	High image quality	Cannot redirect based on physical target angle
3D Face Reconstruction	Accurate physical redirection angle	Cannot redirect gaze Requires frontal source image
ST-ED (Deep generative model)	Redirects head and gaze separately Redirects from any source image to physical target angle	Requires labeled training data Performance relies on training data

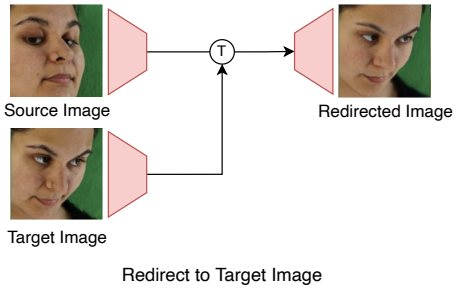

Fig. 2 The illustration of the experiment for evaluating the redirection performance on large target angle.

Table 2 The estimation errors (degree) of the estimators for each dataset.

Average Error	ETH-XGaze	GazeCapture
Head Pose	0.61	0.88
Gaze	0.83	1.37

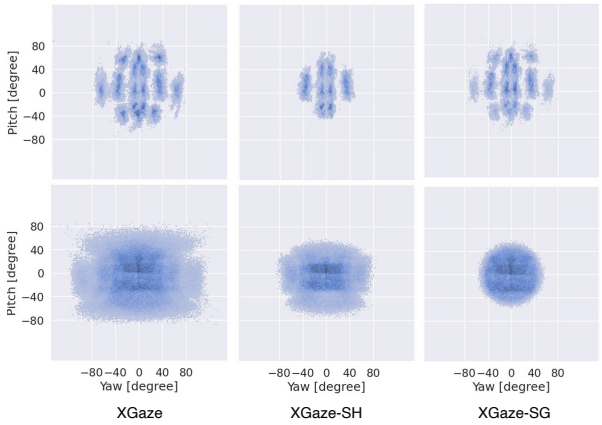

Fig. 3 The distribution and samples of original XGaze and created subsets for training the baseline model. The top row is the head pose distribution, and the bottom row is the gaze distribution.

3.1 Datasets

- **GazeCapture** [18] is collected through crowdsourcing when subjects are using tablet, resulting in a very large subject scale with more than 1,400 subjects. It totally contains more than 1.6 million images.
- **ETH-XGaze** [31] put 18 synchronous cameras under different view angles to take pictures, which achieves a very large angle range. It contains 110 subjects and each subject has around 600 frames, and therefore it has more than 1 million images in total.

The head pose and gaze distribution as well as some samples of the datasets are shown in Fig. 1. We can find that the angle range of ETH-XGaze is relatively large while GazeCapture is much smaller because it requires the subject to fix their gaze at some points in the screen, which fundamentally limits the range of the gaze and head direction. On the other hand, crowdsourcing makes it easier to collect a dataset with a huge subject scale, but the high cost of a lab-control setting is difficult.

3.2 Experimental Settings

We introduce the common experimental settings used in the following hypothesis verification and augmentation evaluation.

For evaluation, previous work evaluates image-to-image redirection using ground-truth paired images [5], [11], [14], [35] as shown in Fig. 2.

To obtain the head pose and gaze direction of the generated image, we use a ResNet-18 [10] estimator. The estimator is trained on the corresponding dataset when evaluating the model on each dataset such that it can predict accurate head pose and gaze. The average estimation errors for each dataset are shown in Table 2, indicating that the estimator is reliable.

For pre-processing, we adopt the data normalization [32], which is commonly used in gaze-related tasks [31], [35]. For the training hyperparameters, we follow the same setting as the original ST-ED [35].

3.3 Evaluation Metrics

Redirection Error. We compute the angular error between the ground-truth target direction and the estimated direction of the generated image. Both the head redirection error and the gaze redirection error are evaluated.

LPIPS. We adopt LPIPS [11], [30], [35] to evaluate the general similarity of the image between the generated image and the target image.

FID. Fréchet Inception Distance [12], [24] has been used to evaluate the visual similarity of two groups of images that is close to human perception.

Identity Distance. In addition, we also compute the identity feature distance between target image and generated image using FaceNet [23] as an evaluation of the identity preservation ability.

4. Analysis of Angle Range in Training Data

Hypothesis: the performance of the redirection model will drop significantly once the target direction is larger than the training data.

In this section, we verify this hypothesis by narrowing down an existing large-range dataset to subsets and comparing their influence on the model as training data. We use the ST-ED model as the baseline model for redirecting head and gaze.

4.1 Narrow Angle Range Creation

Since XGaze already contains a very wide range of head pose and gaze, we create two subsets from it with a smaller range of head pose (XGaze-SH) and small gaze (XGaze-SG). We train the ST-ED model using these three training sets and compare their

Table 3 The evaluation on XGaze (frame 200 to 400) of models trained by XGaze and its narrow subsets.

	Head Error (↓)	Gaze Error (↓)	LPIPS (↓)	FID (↓)	Identity (↓)
XGaze (Full Range)	**2.007**	**5.097**	**0.073**	**16.810**	**0.982**
XGaze-SH †	5.749	7.794	0.138	25.149	1.398
XGaze-SG *	2.204	9.679	0.077	17.238	0.989

†: SH is short for small head; *: SG is short for small gaze

Table 4 The evaluation on XGaze Test of models trained by XGaze and its narrow subsets.

	Head Error (↓)	Gaze Error (↓)	LPIPS (↓)	FID (↓)	Identity (↓)
XGaze (Full Range)	**3.098**	**7.968**	**0.148**	**48.05**	2.120
XGaze-SH †	5.873	10.218	0.183	56.56	2.225
XGaze-SG *	3.292	11.800	0.158	49.47	**2.097**

†: SH is short for small head; *: SG is short for small gaze

performance with the same testing sets.

The distributions of the three training datasets are shown in Fig. 3. XGaze uses 18 synchronous cameras. For the subsets, XGaze-SH only contains the samples from the 10 relatively smaller cameras, which is shown in the second column of Fig. 3. XGaze-SG is sorted based on the $\ell_2 - norm$ of gaze and chosen from the smaller portion such that it has the same number of samples as XGaze-SH. The distribution is shown in the third column of Fig. 3.

Since the official ETH-XGaze only has public gaze label for 80 subjects, we split them into 75 for training and 5 for testing. We denote the first 75 subjects as **XGaze** and the last 5 subjects as **XGaze Test**.

If not specifically mentioned, only the first 200 frames are used for training because we also evaluate the model on seen subjects. Therefore, for training, XGaze, XGaze-SH, and XGaze-SG have 270,000, 150,000, and 150,000 samples, respectively.

4.2 Experiments and Result Analysis

After creating the training datasets, we conduct the experiment and compare the influence of different angle ranges of the training data.

We evaluate the redirection error, LPIPS, FID, and identity distance. We choose the 200th- to 400-th frames of XGaze and the first 200 frames of XGaze Test for testing, which have 150,000 and 10,000 samples, respectively. We use XGaze (frame 200 to 400) and XGaze Test as testing data. During testing, each image is randomly redirected to another image.

The results of the two test sets are shown in Table 3 and Table 4, respectively. Intuitively, the full-range XGaze shows the best performance on all evaluation metrics. While the XGaze-SH shows a huge performance drop on all metrics, especially the head redirection error due to its limited range of head pose, the XGaze-SG only drops the performance on gaze redirection error due to its limited range of gaze and shows a comparable performance on other metrics as the full-range XGaze. The reason for the difference is that the other metrics are all influenced by incorrect head redirection, as shown on the left side of Fig. 4.

It is also noticed that the identity distance becomes much larger when redirecting on unseen subjects. Visualized examples can also be seen on the right side of Fig. 4, where the redirected image of unseen subjects becomes a different person, which reveals an inevitable drawback of the current ST-ED model.

In conclusion, narrow training data will cause a performance drop on redirection. In particular, the limitation of the head pose will have a greater effect on the overall image and will degrade all the evaluation metrics.

Discussions The experimental results confirm that the hypothesis is valid. The error is especially large when redirecting to a large target angle out of its training data's angle range. The model can only redirect in the angle range of the training data, indicating that it is important to use training data that contain a large enough angle range.

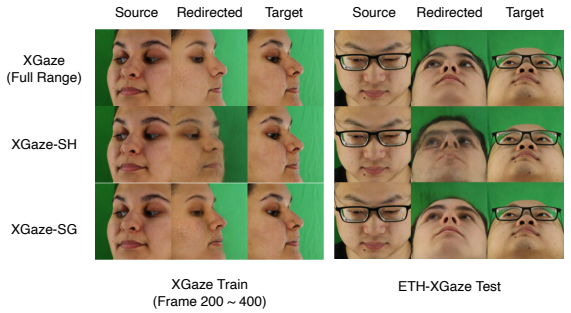

Fig. 4 Examples of image-to-image redirection. Each row corresponds to each training data setting. Left side is the seen subjects from XGaze Train, and right side is the unseen subjects from XGaze Test.

5. Angle Range Augmentation

In this section, we first create augmentation data for GazeCapture, which has limited angle range, and then perform evaluation experiments as explained in Section 3.2.

5.1 Augmentation Method

We follow an off-the-shelf 3D face reconstruction pipeline 3DDFA [7], [8], [37] to do reconstruction and apply the projective-matching [21] to finally obtain a 3D face. Given a target head direction or gaze direction, we can compute a rotation matrix, rotate the face, and render it to a new image.

To create augmentation data, we formulate head-based sampling and gaze-based sampling. For the head-based sampling, we computed a rotation matrix based on the origin and the sampled target head pose and rotated the 3D face using the rotation matrix. Similarly for the gaze-based sampling. In detail, we sample the target direction from a circle-shaped uniform distribution, and the circle radius is chosen to be 40 degrees. Consequently, the distributions of the original GazeCapture dataset and its two augmentation datasets are shown in Fig. 6. Notice that when forcing the head pose to be uniform distribution, the gaze distribution may have a larger range and vice versa.

We follow the split of ST-ED [35] to create the GazeCapture Train and GazeCapture Test. If not specifically mentioned, we abbreviate GazeCapture Train as GazeCapture later. Specifically, GazeCapture contains 1177 subjects with 1,379,083 images, and GazeCapture Test contains 139 subjects with 191,842 images.

For augmentation data, we filter out subjects with less than 30 samples in the original GazeCapture, and we randomly sample 30 images from the remained 861 subjects. Each source image will be augmented to 10 new images, and some augmentation data ex-

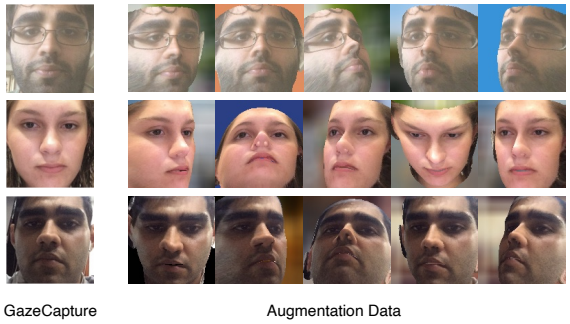

Fig. 5 The samples of original GazeCapture (left) and its augmentation (right).

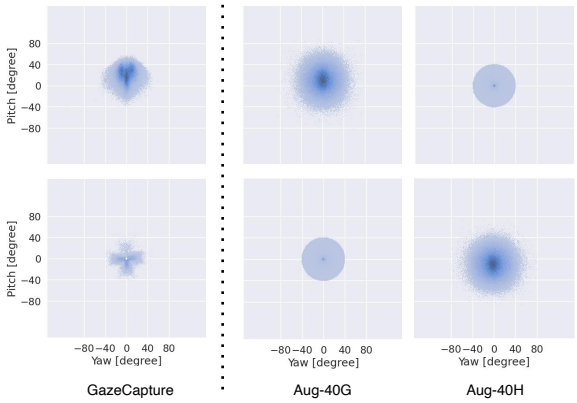

Fig. 6 The label distribution of GazeCapture (left) and the augmentation datasets (right). The top row is the head pose distribution and the bottom row is the gaze distribution.

amples are shown in Fig. 5. As a result, the augmentation dataset contains 257,470 samples.

5.2 Experiments and Result Analysis

After creating the augmentation data, we conduct the two experiments to compare their influence on the model.

We use GazeCapture Test and XGaze Test as the testing data. During testing, each image is randomly redirected to another image. Redirection error, LPIPS, FID, and identity distance are evaluated between target and redirected images. The results for the two test sets are shown in Table 5 and Table 6, respectively.

Since the GazeCapture Test has a limited label range similar to the GazeCapture Train, the head redirection error and gaze redirection error are already at a small level. Therefore, adding augmentation (target on the large range) data does not show an obvious influence but only fluctuation in a very small range for this test set. For the LPIPS, FID, and identity distance, the results in Table 5 are generally worse than that in Table 3. The reason is that the model is tested on unseen subjects from GazeCapture Test. The model cannot preserve the appearance of unseen subjects after redirection, resulting in a larger difference between the target image and the redirected image. Examples can be observed in the left side of Fig. 7, where the redirected image has a similar but not perfectly identical identity. Therefore, in addition to the face identity distance, LPIPS and FID are all at a higher level.

Since XGaze Test has a much larger direction range than GazeCapture, it is much more difficult for the model to perform well, as shown in the first row of Table 6. From the second and third

Table 5 The evaluation of augmentation on GazeCapture Test.

	Head Error (↓)	Gaze Error (↓)	LPIPS (↓)	FID (↓)	Identity (↓)
GazeCapture	1.179	3.960	0.201	**39.64**	1.923
GC † + Aug-40G	1.226	3.829	0.192	41.24	**1.867**
GC + Aug-40H	**1.171**	**3.789**	**0.190**	45.52	1.873

†: GC is short for GazeCapture.

Table 6 The evaluation of augmentation on the XGaze Test.

	Head Error (↓)	Gaze Error (↓)	LPIPS (↓)	FID (↓)	Identity (↓)
GazeCapture	22.265	32.837	0.383	145.68	2.383
GC + Aug-40G	**8.335**	21.222	**0.296**	**121.86**	**2.288**
GC + Aug-40H	11.499	**20.399**	0.322	133.70	**2.288**

†:GC is short for GazeCapture.

Fig. 7 Examples of image-to-image redirection. Each row corresponds to the training data setting. Left side is the unseen subject from GazeCapture Test, and right side is the unseen subject from XGaze Test.

row of Table 6, the augmentation showed significant improvement in the head redirection error and gaze redirection error. Note that though the augmentation data improved the performance on LPIPS, FID, and identity distance, it may be due to better head pose redirection. As can be seen on the right side of Fig. 7, the metrics between the redirected image and the target image are reduced due to a more similar head pose. All of the metrics are at a very high level, meaning that the ST-ED model cannot generalize the performance well on unseen subjects and the augmentation has a limited effect on image quality and identity preservation.

Discussions Since the data augmentation focuses on angle extension, it can effectively reduce the redirection angle error. However, it is difficult to directly reduce the visual distance to the target image, especially for the preservation of identity.

6. Conclusion

In this paper, we analyzed the angle range of training data for a face image synthesis model that redirects head pose and gaze. The redirection performance hugely relies on the training data. To tackle the limited angle range of training data, we proposed to use 3D face reconstruction to create training data with a larger angle range. We conducted thorough experiments and showed that data augmentation can effectively reduce redirection errors. Besides, we also found that subject identity preservation remains a challenge, which will be future work, along with the demand for higher photo-reality.

References

[1] *Towards Metrical Reconstruction of Human Faces* (2022).
[2] Blanz, V. and Vetter, T.: A morphable model for the synthesis of 3D faces, *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 187–194 (1999).
[3] Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y. and Tong, X.: Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set, *Proc. CVPRW* (2019).

[4] Feng, Y., Feng, H., Black, M. J. and Bolkart, T.: Learning an Animatable Detailed 3D Face Model from In-The-Wild Images, Vol. 40, No. 8, (online), available from ⟨https://doi.org/10.1145/3450626.3459936⟩ (2021).

[5] Ganin, Y., Kononenko, D., Sungatullina, D. and Lempitsky, V.: Deepwarp: Photorealistic image resynthesis for gaze manipulation, *Proc. ECCV*, Springer, pp. 311–326 (2016).

[6] Ghosh, S., Dhall, A., Hayat, M., Knibbe, J. and Ji, Q.: Automatic gaze analysis: A survey of deep learning based approaches, *arXiv preprint arXiv:2108.05479* (2021).

[7] Guo, J., Zhu, X. and Lei, Z.: 3DDFA, https://github.com/cleardusk/3DDFA (2018).

[8] Guo, J., Zhu, X., Yang, Y., Yang, F., Lei, Z. and Li, S. Z.: Towards Fast, Accurate and Stable 3D Dense Face Alignment, *Proceedings of the European Conference on Computer Vision (ECCV)* (2020).

[9] Hansen, D. W. and Ji, Q.: In the Eye of the Beholder: A Survey of Models for Eyes and Gaze, Vol. 32, No. 3, pp. 478–500 (2010).

[10] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *Proc. CVPR*, (online), DOI: 10.1109/CVPR.2016.90 (2016).

[11] He, Z., Spurr, A., Zhang, X. and Hilliges, O.: Photo-realistic monocular gaze redirection using generative adversarial networks, *Proc. ICCV*, pp. 6932–6941 (2019).

[12] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. and Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium, *Advances in neural information processing systems*, Vol. 30 (2017).

[13] Hu, Y., Wu, X., Yu, B., He, R. and Sun, Z.: Pose-guided photorealistic face rotation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8398–8406 (2018).

[14] Jindal, S. and Wang, X. E.: CUDA-GHR: Controllable Unsupervised Domain Adaptation for Gaze and Head Redirection, *arXiv preprint arXiv:2106.10852* (2021).

[15] Kammoun, A., Slama, R., Tabia, H., Ouni, T. and Abid, M.: Generative Adversarial Networks for face generation: A survey, *ACM Computing Surveys (CSUR)* (2022).

[16] Karras, T., Aila, T., Laine, S. and Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation, *International Conference on Learning Representations*, (online), available from ⟨https://openreview.net/forum?id=Hk99zCeAb⟩ (2018).

[17] Karras, T., Laine, S. and Aila, T.: A Style-Based Generator Architecture for Generative Adversarial Networks, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).

[18] Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W. and Torralba, A.: Eye Tracking for Everyone, *Proc. CVPR*, (online), DOI: 10.1109/CVPR.2016.239 (2016).

[19] Masi, I., Hassner, T., Tran, A. T. and Medioni, G.: Rapid Synthesis of Massive Face Sets for Improved Face Recognition, *Proc. FG*, (online), DOI: 10.1109/FG.2017.76 (2017).

[20] Park, S., Mello, S. D., Molchanov, P., Iqbal, U., Hilliges, O. and Kautz, J.: Few-Shot Adaptive Gaze Estimation, *Proc. ICCV* (2019).

[21] Qin, J., Shimoyama, T. and Sugano, Y.: Learning-by-Novel-View-Synthesis for Full-Face Appearance-Based 3D Gaze Estimation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 4981–4991 (2022).

[22] Sanyal, S., Bolkart, T., Feng, H. and Black, M.: Learning to Regress 3D Face Shape and Expression from an Image without 3D Supervision, *Proc. CVPR* (2019).

[23] Schroff, F., Kalenichenko, D. and Philbin, J.: Facenet: A unified embedding for face recognition and clustering, *Proc. CVPR*, pp. 815–823 (2015).

[24] Seitzer, M.: pytorch-fid: FID Score for PyTorch, https://github.com/mseitzer/pytorch-fid (2020). Version 0.2.1.

[25] Shen, Y., Gu, J., Tang, X. and Zhou, B.: Interpreting the Latent Space of GANs for Semantic Face Editing, *CVPR* (2020).

[26] Shen, Y., Yang, C., Tang, X. and Zhou, B.: InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs, *TPAMI* (2020).

[27] Tripathy, S., Kannala, J. and Rahtu, E.: Icface: Interpretable and controllable face reenactment using gans, *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 3385–3394 (2020).

[28] Wood, E., Baltrušaitis, T., Morency, L.-P., Robinson, P. and Bulling, A.: Gazedirector: Fully articulated eye gaze redirection in video, *Computer Graphics Forum*, Vol. 37, No. 2, Wiley Online Library, pp. 217–225 (2018).

[29] Yuxiao Hu, Dalong Jiang, Shuicheng Yan, Lei Zhang and Hongjiang zhang: Automatic 3D reconstruction for face recognition, *Proc. FG*, (online), DOI: 10.1109/AFGR.2004.1301639 (2004).

[30] Zhang, R., Isola, P., Efros, A. A., Shechtman, E. and Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, *Proc. CVPR* (2018).

[31] Zhang, X., Park, S., Beeler, T., Bradley, D., Tang, S. and Hilliges, O.: ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head Pose and Gaze Variation, *Proc. ECCV* (2020).

[32] Zhang, X., Sugano, Y. and Bulling, A.: Revisiting Data Normalization for Appearance-Based Gaze Estimation, *Proc. ETRA* (2018).

[33] Zhang, X., Sugano, Y., Fritz, M. and Bulling, A.: Appearance-Based Gaze Estimation in the Wild, *Proc. CVPR* (2015).

[34] Zhang, X., Sugano, Y., Fritz, M. and Bulling, A.: It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation, *Proc. CVPRW* (2017).

[35] Zheng, Y., Park, S., Zhang, X., Mello, S. D. and Hilliges, O.: Self-Learning Transformations for Improving Gaze and Head Redirection, *Proc. NeurIPS* (2020).

[36] Zhou, H., Liu, J., Liu, Z., Liu, Y. and Wang, X.: Rotate-and-Render: Unsupervised Photorealistic Face Rotation from Single-View Images, *Proc. CVPR* (2020).

[37] Zhu, X., Liu, X., Lei, Z. and Li, S. Z.: Face alignment in full pose range: A 3d total solution, *IEEE transactions on pattern analysis and machine intelligence* (2017).