

Pairwise-GAN: Pose-based View Synthesis through Pair-Wise Training

Xuyang Shen, Jo Plested, Yue Yao, and Tom Gedeon

Research School of Computer Science,
Australian National University
first.second@anu.edu.au

Abstract. Three-dimensional face reconstruction is one of the popular applications in computer vision. However, even state-of-the-art models still require frontal face as inputs, which restricts its usage scenarios in the wild. A similar dilemma also happens in face recognition. New research designed to recover the frontal face from a single side-pose facial image has emerged. The state-of-the-art in this area is the Face-Transformation generative adversarial network, which is based on the CycleGAN. This inspired our research which explores the performance of two models from pixel transformation in frontal facial synthesis, Pix2Pix and CycleGAN. We conducted the experiments on five different loss functions on Pix2Pix to improve its performance, then followed by proposing a new network Pairwise-GAN in frontal facial synthesis. Pairwise-GAN uses two parallel U-Nets as the generator and PatchGAN as the discriminator. The detailed hyper-parameters are also discussed. Based on the quantitative measurement by face similarity comparison, our results showed that Pix2Pix with L1 loss, gradient difference loss, and identity loss results in 2.72% of improvement at average similarity compared to the default Pix2Pix model. Additionally, the performance of Pairwise-GAN is 5.4% better than the CycleGAN and 9.1% than the Pix2Pix at average similarity.

Keywords: Face Frontalization · Novel View Synthesis · Image Translation

1 Introduction

Three-dimensional face reconstruction from a single image is one of the popular topics in computer vision over the past twenty years. Although end-to-end learning methods achieve the state-of-the-art in three-dimensional facial reconstruction [1,2], they still cannot infer the missing facial information if the input facial image is a side-pose at more than 30 degrees. This issue not only happens on three-dimensional facial reconstruction tasks but also exists on other facial applications in computer vision. Face recognition is a widely used authentication and detection technique that currently requires front face images. However, this condition is challenging to fulfill in real life, especially for video surveillance which captures the object under any situation.

Inferring the missing facial information, known as frontal facial synthesis, is a good alternative for these applications to solve the issue raised by the side-pose facial image. From 2013, an increasing number of deep learning methods have been proposed

to improve performance on frontal face synthesis problems [3,4,5,6]. With the publication of Two Pathways Generative Adversarial Network (TP-GAN) [7] and complete representations generative adversarial network (CR-GAN) [8], the focus of solving face synthesis problems was shifted into models based on GAN and CGAN [7,8,9,10].

The current state-of-the-art frontal facial synthesis is Pose-Invariant Generator Adversarial Network (FT-GAN) proposed by [9]. FT-GAN combines Cycle-Consistent Adversarial Networks (CycleGAN), considered state-of-the-art in pixel-to-pixel transformation [21,29,30,31], with key point alignment to generate frontal facial synthesis more realistically. Their results indicated that FT-GAN is 10% and 4% better than CR-GAN and CycleGAN respectively, when the input side-pose image is between 60 and 75 degrees.

With these motivations, we begin our work by exploring two models, Pix2Pix and CycleGAN, from style transformation into frontal facial synthesis. Our research aimed to evaluate the performance of these two models under default hyper-parameters using the Color FERET Database [11]. We conducted our experiments in two stages:

1. We applied different loss functions to the Pix2Pix model and evaluated their performance.
2. We proposed a new CGAN network named Pairwise-GAN, based on the network architecture of Pix2Pix and CycleGAN, and compared its performance with state-of-the-art models.

The main contributions of this work are listed below:

- Five different loss functions were analyzed to improve the performance of the Pix2Pix in frontal facial synthesis. The best loss results achieved an 2.72% improvement compared to Pix2Pix, which is close to CycleGAN’s performance.
- Pairwise-GAN which targets frontal facial synthesis is proposed as a new network architecture of CGAN. Pairwise-GAN reaches 44.3 average similarity and 74.22 maximum similarity between synthesis and ground truth. It gains 5.4% better results than the CycleGAN and 9.1% than the Pix2Pix. Compared to a 4% improvement over CycleGAN by the FT-GAN [9].
- A new quantitative measurement on face frontalization is introduced to evaluate the similarity between the generated frontal face and the ground-truth image.

2 Methods

2.1 Dataset and Pre-processing

We selected the Color FERET Database from NIST [11] since it is free to access for research purposes. It contains 11,338 facial images from 994 different subjects. Images were collected at 14 different times between December 1993 to August 1996 [11], so the focus length and position of the camera are different each time. This results in various facial sizes. To minimize the influence of this noise to GAN training, we used MTCNN [12] to crop out the face for each image.

Another source of variability of the dataset is the angle of side-images. It is commonly agreed that synthesizing the frontal image based on the side-image at a small

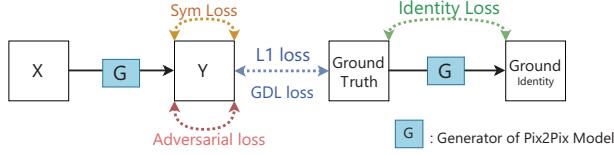


Fig. 1: Overview of loss functions used in "exploring loss function" experiment. X and Y in the figure are the input image (side-pose facial image) and the generated frontal image. The ground identity is generated by applying the generator to the ground truth image

angle is more accessible than the large-angle since the former one can provide more facial features. Therefore, we added another constraint to frontal facial generation by requiring the angle of side-pose images to be at least 60 degrees. We also restricted the maximum angle of fewer than 90 degrees as models gain fewer facial information after this degree.

Train set and Test set After pre-processing, there are 3,135 images (2,090 pairs) in the training set and 369 (246 pairs) in the test set, where one pair contains one frontal image and one side-pose image. The size of input images in experiments is further resized into 256 by 256 pixel, due to the limitation of computation. The reason we chose 256 is that elaborate facial features can be reflected in the generated image from 256 to 512 pixels, such as hairstyle or detailed information on eyes [14].

2.2 Synthesis Loss Functions

Experimental setup Based on the train set and test set, we performed our first stage of experiments, which explored five different loss functions on Pix2Pix [23]. The overview of four different loss functions is shown in Fig. 1. This model was chosen because Pix2Pix and CycleGAN are two state-of-art models in style transformation where frontal facial synthesis also belongs. Additionally, the Pix2Pix is more applicable to experiment with different loss functions compared to the CycleGAN.

The generator in Pix2Pix is a modified U-Net [15,23] which contains eight blocks in the encoder and eight blocks in the decoder. Each block in the encoder is constructed by one convolution layer, one batch normalization layer, and Leaky ReLU as the activation function. Each block in the decoder consists of one transposed convolution layer, one batch normalization layer, and ReLU as the activation function. Additionally, the dropout is also applied to the first three blocks of the decoder.

Adversarial loss aims to optimize the generator, which achieves a minimum Kullback-Leibler divergence (KL-divergence) between the generated data and ground truth data [19]. Theoretically, the generator tries to get a higher score from the discriminator. In other words, the generated image tends to be a sharp frontal facial image whose facial feature is closer to the input person. However, in practice, it commonly appears that the

discriminator in the conditional adversarial network only focuses on whether the generated image is sharp and contains basic facial features or not, ignoring the input of the domain. Based on this, we continued to explore other loss functions.

$$\text{Adversarial Loss} = E_{x \sim P_x(X)}[\log(D(G(x)))] \quad (1)$$

where $G(x)$ refers to generated data (fake image) by the generator; D refers to the discriminator in Pix2Pix; G refers to the generator in Pix2Pix mode.

L1 Loss (Mean Absolute Error) is employed to facilitate content consistency between the generated frontal image and the ground-truth frontal image. L1 loss also accelerates optimization in CGAN [7].

$$L1 \text{ Loss} = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H |GT_{i,j} - Y_{i,j}| \quad (2)$$

where GT refers to the ground truth image; Y refers to the generated image; W, H is the width and height of the images; i, j is the x coordinate and y coordinate of each pixel.

Gradient Difference Loss (GDL) aims to penalize the differences of image gradient predictions directly, where gradients are the differences between neighboring pixel values [20]. Since the facial images usually have continuous value within the neighboring, this formula can strengthen this relationship in the generated image, making it more realistic.

$$\begin{aligned} GDL \text{ Loss} = & \sum_{i,j} (|Y_{i,j} - Y_{i-1,j}| - |GT_{i,j} - GT_{i-1,j}|)^\alpha + \\ & ||Y_{i,j-1} - Y_{i,j}| - |GT_{i,j-1} - GT_{i,j}|)^\alpha) \end{aligned} \quad (3)$$

where Y, GT represents the generated image and ground truth image respectively; i, j is the x coordinate and y coordinate of that pixel; α can be any integer greater or equal to 1, we tested $\alpha = 1$ and $\alpha = 2$ in our experiments.

Symmetry Loss Symmetry is one of the features in human faces. Based on this phenomenon, [7] proposed the symmetry loss, which encourages symmetrical structure generated by the generator. Additionally, this formula behaves more robustly in the Laplacian image space as it can overcome the illumination difference between two sides of faces. However, we only apply it to the original pixel image, because the illumination difference is minor in Color FERET dataset.

$$Symmetry \text{ Loss} = \frac{1}{W/2 \times H} \sum_{i=1}^{W/2} \sum_{j=1}^H |Y_{i,j} - Y_{W-(i-1),j}| \quad (4)$$

where Y refers to the generated image; W, H is the width and height of the images; i, j is the x coordinate and y coordinate of that pixel.

Identity Loss first appeared in the implementation of CycleGAN [21] and was discussed in Cycada [22]. One intention of this function is to regularize the generator, which should not map the input image into a different domain image if it is already in the target domain. Our experiment found that this loss function helps preserve facial identity information from different people.

$$\text{Identity Loss} = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H | GT_{i,j} - GI_{i,j} | \quad (5)$$

where GT refers the ground truth image; GI refers to the ground identity image; W, H is the width and height of the images; i, j is the x coordinate and y coordinate of that pixel.

2.3 Network Architecture of Pairwise-GAN

During the experiments, we discovered that Pix2Pix and CycleGAN have outstanding achievements in general pixel-to-pixel transformation but gain ordinary performance on frontal facial synthesis. Notably, it was hard to balance the ratio of adversarial loss and other types of losses. If the proportion of adversarial loss is adjusted higher than the other, the generator can easily deceive the discriminator by generating a realistic image unrelated to the input domain. Alternatively, parts of the generated image are close to the original person but the entire image is not a real face, for example, the face may contain three eyes. As a result of further study of the network architecture, we developed Pairwise-GAN which is based on the network architecture of Pix2Pix and CycleGAN.

Generator The generator of Pairwise-GAN is based on U-Net [15] which is also used in the generator of Pix2Pix and CycleGAN [21,23]. Unlike usual conditional adversarial networks and [7,27,28], Pairwise-GAN has two independent auto-encoders and weight sharing of the first two-layers of the decoder (Fig. 2). The layered architecture of U-Nets contains eight blocks in the encoder, eight blocks in the decoder, and skip connections between the encoder and decoder. Specifically, the detail layers used in each block of encoder and decoder are the same as Pix2Pix, except for the normalization layer. In Pairwise-GAN, we use instance normalization instead of batch normalization to boost the performance of generator [24,25,26].

When training, Pairwise-GAN requires two side-pose images at the same angles from the same person but in a different direction as inputs (X_{left}, X_{right}). Left generator G_1 takes the side-pose image from the left direction, while another side-pose image will be passed into generator G_2 . After different encoders extract facial features, two frontal facial images are generated through two decoders. As these two generators are independent, only one side-pose image is required to be input for prediction. This approach addresses affordability and the difficulties in pixel-to-pixel transformation inherent in the complexities of human faces.

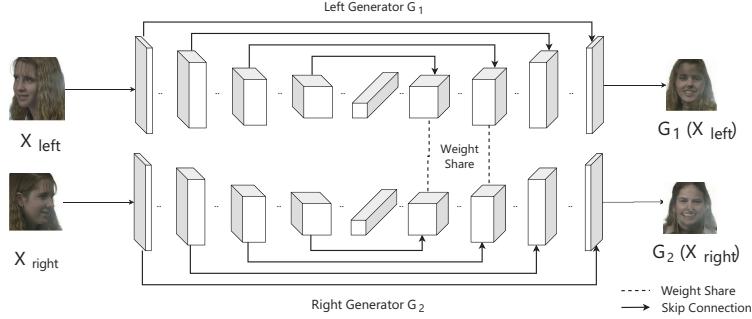


Fig. 2: Generator Architecture of Pairwise-GAN. X_{left} left-side facial input; X_{right} right-side facial input. Providing both side faces is a common attribute for most facial databases, such as Multi-PIE Face Database, CAS-PEAL Face Database, and Color FERET Database

Discriminator A standard discriminator in a GAN maps the input image into a scalar value which classifies whether it is real or fake. The output scalar value is a weighted sum of the whole data field, which cannot reflect the characteristics of the local feature. To address this lack of precision, we used the patch-based discriminator of GANs (PatchGAN) as the discriminator for Pairwise-GAN as proposed in Pix2Pix [23]. PatchGAN maps the input image (256 by 256 pixel) to an N-by-N matrix of outputs X , where each $X_{i,j}$ represents whether the patch (i, j) is real or fake. Additionally, the patch (i, j) is the convolution result of one receptive field which the discriminator is sensitive to, for example, the ear, tooth, and eyes if the input image is a human face (Fig. 3).

The operation described above is mathematically equivalent to cropping the input data into multiple overlapping patches, respectively discriminating the difference by the classifier (discriminator), and averaging the results. The result outlined in [23,21], indicated that PatchGAN extracted the local characterization of the images. This is conducive to generating the images in a high resolution.

From characteristics of patches, the size of the receptive field can be adjusted in different tasks. Specifically, a broader scope of receptive field focuses on the relationship between objects in a large area. However, it also consumes more expensive computation power. From the experimental results offered by Pix2Pix [23], the receptive field with 16-by-16 size already achieves a sharp output. In contrast, results from a smaller receptive field (1-by-1) lose spatial statistics, and a larger receptive field (70-by-70) can force results with more colorfulness. As a result, under our computation power, we set the size of the receptive field as 70-by-70, which results in the output matrix being 30-by-30.

Pair Loss To reduce the difference of generated images between the left generator and right generator, we examined the weight sharing between two encoders and added the pair loss into the loss function. This loss was expected to penalize the network parameters if there exists a major difference between generated images by the left generator

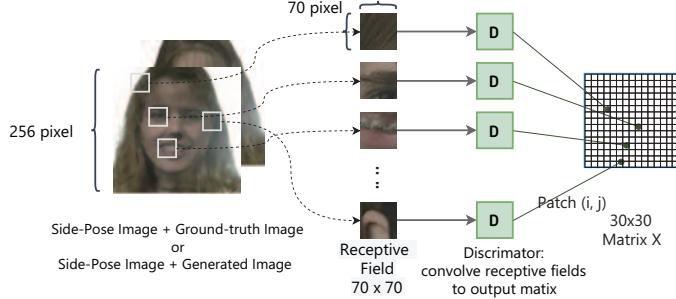


Fig. 3: Intuitive view of the discriminator in Pairwise-GAN (Patch-based discriminator of GANs [23])

and right generator.

$$\text{Pair Loss} = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H |YL_{i,j} - YR_{i,j}| \quad (6)$$

where YL and YR refer to the image generated by the left generator and right generator respectively; W, H is the width and height of the images; i, j is the x coordinate and y coordinate of that pixel.

3 Results and Evaluation

To measure the results quantitatively, we employed the facial similarity, structural similarity (SSIM), Frechet Inception Distance (FID), and Peak Signal-to-Noise Ratio (PSNR) by comparing the frontal facial synthesis and ground truths. Notably, The result of facial similarity (confidence) is provided by a commercial face comparison API (Face++) based on face recognition.

3.1 Loss Function Analysis

In this experiment, a Pix2Pix with various loss functions was trained with 125 epochs, with one epoch taking 2,090 pairs of images. Additionally, both generator and discriminator were optimized with Adam and a learning rate of 0.0002, following [15].

Based on the default ratio of adversarial loss and L1 loss in Pix2Pix (1:120), we first adjusted the proportion of the L1 loss. With a low penalty coming from L1 loss, Pix2Pix tends to generate the frontal facial less similar to the original domain. If the L1 loss is removed from the loss penalty, the average similarity drops down 40% compared to the default one (Table 1). However, although a large L1 penalty results in a highly similar image, the blurring issue of synthetic frontal images becomes more serious. For instance, we found that the shape of the nose is difficult to determine in the generated image by 120 weights of L1 loss, compared to the adversarial loss one.

Table 1: Compare average similarity between ground truth and generated image on different loss configurations among 40 test images. Higher similarity is better

Hyper-parameters	Adv	L1	GDL	Sym	Id	Average Similarity
Pix2Pix	1	0	0	0	0	25.68
Pix2Pix	1	40	0	0	0	33.11
Pix2Pix	1	120	20	0	0	31.17
Pix2Pix	20	3	0.5	0	0	34.89
Pix2Pix	20	3	0.1	0.05	0	31.07
Pix2Pix	20	3	0.1	0	5	37.92
<i>Pix2Pix Default</i>	<i>1</i>	<i>120</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>35.2</i>
<i>CycleGAN Default</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>5</i>	<i>38.7</i>

After applying the GDL (gradient difference loss) into Pix2Pix and keeping the original ratio, the average similarity decreased. To further analyze the influence of the GDL penalty to model training, the proportion of L1 loss is decreased below the adversarial loss; the ratio between L1 and GDL remains untouched. We noticed that the average similarity grows to 35. However, the obscured side of the input image is blurred in the synthesis, which we considered a drawback of the GDL penalty.

The behavior of symmetry loss resulted in a negative influence on the performance of Pix2Pix as the average similarity decreases into 31. It was hard to converge during the training progress. As a result, we abandoned the this loss in later experiments.

We explored identity loss which aims to regularize the generator. Compared to the L1 penalty and GDL penalty, identity loss leads to a significant improvement in the frontal facial of Pix2Pix, which the synthetic frontal face is most natural and sharp compared to others, notably, hairstyle and glasses were recovered in high quality. As a result, identity penalty helped to preserve facial identity information, however, adding identity loss to GAN learning doubled the training time required in each epoch to calculate the ground identity image through the generator.

3.2 Pairwise-GAN Analysis

We trained each of the experiments of Pairwise-GAN for 250 epochs, with one epoch taking 1,045 images and used the same data distribution as previous experiments. We set the batch size to 1 and the learning rate to 0.0002. Additionally, Adam was selected as the optimizer for both generator and discriminator.

We expected the network architecture with weight sharing gains to achieve a better performance than using the loss penalty as the former is closer to a coercive specification. However, the results indicated another option which the pair loss helps to improve the average performance, and weight sharing contributes to refresh the peak performance (Table 2). If both weight sharing and pair loss are employed in the Pairwise-GAN, it decreased both peak and average performance, which results from the network is hard to converge.

Comparing with State-of-the-Art To further analyze the performance of Pairwise-GAN, we included comparisons with current state-of-the-art models. FT-GAN, the cur-

Table 2: Results of different configurations on Pairwise-GAN (ablation study) and comparisons on Pairwise-GAN with Pix2Pix, CycleGAN, and CR-GAN

Hyper-parameters						Similarity Measure			Other Measure		
Adv	L1	GDL	Id	Pair	Weight Share	Avg	Max	Min	SSIM	FID	PSNR
10	0	0	5	10	Disable	44.23	79.14	22.02	0.576	102.141	13.968
<i>10</i>	<i>3</i>	<i>0</i>	<i>5</i>	<i>10</i>	<i>Disable</i>	<i>44.30</i>	<i>74.22</i>	<i>23.56</i>	<i>0.600</i>	<i>91.517</i>	<i>14.517</i>
10	3	0	5	0	Enable	42.34	81.83	17.16	0.550	90.531	14.199
10	0	0	10-5	10-2	Enable	42.79	70.61	16.43	0.515	93.433	13.987
10	0	0	0	10-2	Enable	45.16	71.95	20.07	0.505	119.211	13.209
<i>CR-GAN Default [8]</i>						<i>30.08</i>	<i>47.16</i>	<i>12.9</i>	<i>0.362</i>	<i>171.28</i>	<i>10.916</i>
<i>Pix2Pix Default [23]</i>						<i>35.2</i>	<i>62.12</i>	<i>0</i>	<i>0.502</i>	<i>118.791</i>	<i>13.775</i>
<i>CycleGAN Default [21]</i>						<i>38.9</i>	<i>70.22</i>	<i>14.64</i>	<i>0.493</i>	<i>99.777</i>	<i>13.274</i>
<i>Groud Truth</i>						<i>100</i>	<i>100</i>	<i>100</i>	<i>1</i>	<i>0.0</i>	<i>+Inf</i>

rent state-of-the-art model in frontal facial generation used CMU Multi-PIE database [13], and did not release the codes [9]. Therefore, we compared our model with CycleGAN since FT-GAN is built on CycleGAN and within 4 % better than it. Additionally, we also the CR-GAN in the comparison.

Table 2 shows Pairwise-GAN reaches 44.3 average similarity and 74.22 maximum similarity between synthesis and ground truth. In other words, the performance of Pairwise-GAN is around 5.4% better than the CycleGAN and 9.1% than the Pix2Pix, measured by the average similarity of synthesis and ground truth among 40 test images. It also obtains the best evaluation results from SSIM, FID, and PSNR. From the synthetic frontal images, we note that CycleGAN only generated the main facial features, and stacks onto the original side-pose image. On Pairwise-GAN, the field recovered is broader as it focuses on the facial features as well as other elements, such as hairstyle and neck. Apart from that, Pairwise-GAN also requires less computation power in both training and prediction as it only consumes 224 seconds compared to CycleGAN requiring 315 seconds. The prediction can achieve 34 fps by using GPU, RTX 2060S.

4 Conclusion and Future Work

In this paper, we extend current work on synthesizing front face images based on side-pose facial images. Initially, we focused on the performance of Pix2Pix, which tests and analyses five different loss functions, including adversarial loss, L1 loss, gradient difference loss, symmetry loss, and identity loss. Since the improvement based on different loss penalty was minor, we continued our work to propose Pairwise-GAN. Through the analysis of experimental results on different loss functions, we concluded that L1 loss, GDL loss, and identity loss help alleviate the common issue existing in CGAN, where the output face is related to the input face. The ratio of these five-loss, achieving the highest score in our experiment is 20:3:0.1:0:5 (adversarial:L1:GDL:symmetry:identity).

In our second experiment, we proved that Pairwise-GAN generated better frontal face results than the CycleGAN in Color FERET database by using less computation resources; specifically, it improves around 5.4% on average compared to CycleGAN

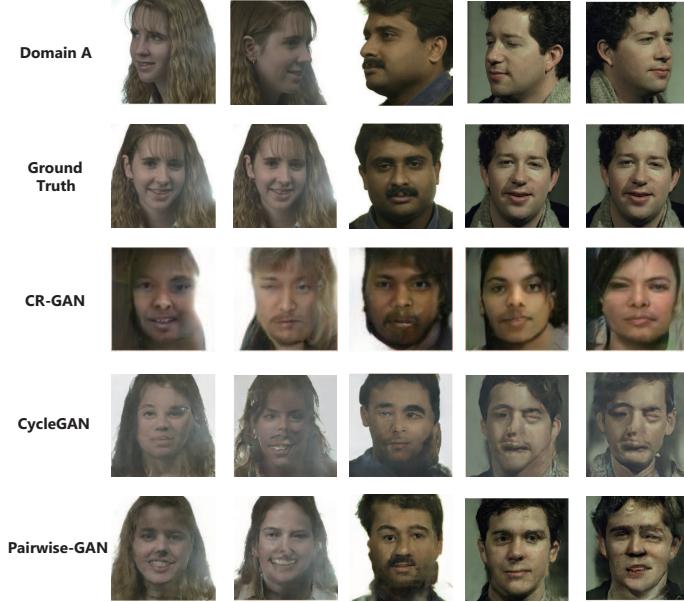


Fig. 4: Qualitative results of Color FERET Database. Comparison with CR-GAN, CycleGAN and Ground Truth

in generated quality. We also explored various configurations of Pairwise-GAN, which demonstrated that either, but not both, pair loss (soft penalty) or weight sharing (coercive specification) positively contribute to improvements. Further research on the network architecture is required to continue minimizing the difference of generated images by the two generators of Pairwise-GAN.

Acknowledgements

We thank Dawn Olley and Alasdair Tran for their invaluable editing advice.

References

1. Jackson, A.S., Bulat, A., Argyriou, V., Tzimiropoulos, G.: Large pose 3D face reconstruction from a single image via direct volumetric CNN regression. In: IEEE International Conference on Computer Vision, pp. 1031-1039 (2017)
2. Feng, Y., Wu, F., Shao, X., Wang, Y., Zhou, X.: Joint 3d face reconstruction and dense alignment with position map regression network. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 534-551 (2018)
3. Zhang, Y., Shao, M., Wong, E.K., Fu, Y.: Random faces guided sparse many-to-one encoder for pose-invariant face recognition. In: IEEE International Conference on Computer Vision, pp. 2416-2423 (2013)

4. Kan, M., Shan, S., Chang, H., Chen, X.: Stacked progressive auto-encoders (spae) for face recognition across poses. In: IEEE conference on computer vision and pattern recognition, pp. 1883-1890 (2014)
5. Hassner, T., Harel, S., Paz, E., Enbar, R.: Effective face frontalization in unconstrained images. In: IEEE conference on computer vision and pattern recognition, pp. 4295-4304 (2015)
6. Yin, X., Yu, X., Sohn, K., Liu, X., Chandraker, M.: Towards large-pose face frontalization in the wild. In : IEEE international conference on computer vision, pp. 3990-3999 (2017)
7. Huang, R., Zhang, S., Li, T., He, R.: Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In: IEEE International Conference on Computer Vision, pp. 2439-2448 (2017)
8. Tian, Y., Peng, X., Zhao, L., Zhang, S., Metaxas, D.N.: CR-GAN: learning complete representations for multi-view generation. arXiv preprint arXiv:1806.11191 (2018)
9. Zhuang, W., Chen, L., Hong, C., Liang, Y., Wu, K.: FT-GAN: Face Transformation with Key Points Alignment for Pose-Invariant Face Recognition. Electronics, vol. 8, pp. 807 (2019)
10. Yin, Y., Jiang, S., Robinson, J.P., Fu, Y.: Dual-attention GAN for large-pose face frontalization. arXiv preprint arXiv:2002.07227 (2020)
11. color FERET Database, <https://www.nist.gov/itl/products-and-services/color-feret-database>
Last accessed 1 June 2020
12. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, **23**(10), pp.1499-1503 (2016)
13. The CMU Multi-PIE Face Database, <http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie>
Last accessed 1 June 2020
14. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: IEEE conference on computer vision and pattern recognition, pp. 4401-4410 (2019)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234-241 (2015)
16. Face Comparing by Face++, <https://www.faceplusplus.com/face-comparing/>
Last accessed 1 June 2020
17. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, **13**(4), pp.600-612 (2004)
18. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in neural information processing systems, pp. 6626-6637 (2017)
19. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems, pp. 2672-2680 (2014)
20. Mathieu, M., Courrie, C., LeCun, Y.: Deep multi-scale video prediction beyond mean square error. arXiv preprint arXiv:1511.05440 (2015)
21. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE international conference on computer vision, pp. 2223-2232 (2017)
22. Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T.: Cycada: Cycle-consistent adversarial domain adaptation. In: International conference on machine learning, pp. 1989-1998 (2018)
23. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: IEEE conference on computer vision and pattern recognition, pp. 1125-1134 (2017)

24. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
25. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
26. Wu, Y., He, K.: Group normalization. In: Proceedings of the European conference on computer vision (ECCV), pp. 3-19 (2018)
27. Liu, M.Y., Tuzel, O.: Coupled generative adversarial networks. In: Advances in neural information processing systems, pp. 469-477 (2016)
28. Anoosheh, A., Agustsson, E., Timofte, R., Van Gool, L.: Combogan: Unrestrained scalability for image domain translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 783-790 (2018)
29. Yao, Y., Zheng, L., Yang, X., Naphade, M., Gedeon, T.: Simulating content consistent vehicle datasets with attribute descent. arXiv preprint arXiv:1912.08855 (2019)
30. Yao, Y., Plested, J., Gedeon, T.: Information-preserving Feature Filter for Short-term EEG signals. Neurocomputing (2020)
31. Tewari, A., Fried, O., Thies, J., Sitzmann, V., Lombardi, S., Sunkavalli, K., Martin-Brualla, R., Simon, T., Saragih, J., Niener, M., Pandey, R.: State of the Art on Neural Rendering. arXiv preprint arXiv:2004.03805 (2020)