SELF-SUPERVISED FRONTALIZATION AND ROTATION GAN WITH RANDOM SWAP FOR POSE-INVARIANT FACE RECOGNITION

Jiashu Liao[⋆] *Victor Sanchez*[⋆] *Tanaya Guha*[†]

[⋆] Department of Computer Science, University of Warwick, UK
[†] School of Computing Science, University of Glasgow, UK
jiashu.liao@warwick.ac.uk

ABSTRACT

The task of pose invariant face recognition (PIFR) has recently seen important improvements by introducing generative adversarial networks (GANs). These GAN-based models synthesize a frontal face image from an image in any pose to boost recognition performance. Most of these methods, however, require the ground-truth frontal face image during optimization as they rely on supervised or semi-supervised learning. In this work, we introduce the self-supervised Frontalization and Rotation GAN (FRGAN), which can synthesize a frontal face image from a non-frontal face image. For self-supervision, the synthesized image is rotated back to the original pose based on reconstruction and adversarial losses. To improve performance, the FRGAN uses the *Random Swap*, a parameter-free data augmentation strategy that swaps key facial regions between the input image and its reconstructed version to force the generator to synthesize more realistic images. Our qualitative and quantitative experiments on benchmark datasets confirm the strong performance of the FRGAN compared to the state-of-the-art (SOTA).

Index Terms— pose, GAN, face frontalization

1. INTRODUCTION

PIFR is a challenging task due to the self-occlusion of several regions of the face, which is caused by the camera's view angle. A common method to tackle the PIFR task is to use a 3D model [1] to obtain a morphable face, which can be rotated to reconstruct the self-occluded regions. Unfortunately, due to the inherent information loss, the rotated and reconstructed regions may lack the appropriate texture and depth.

Recently, several GAN-based models capable of synthesizing a frontal face image from a face image acquired in any other pose have been proposed to solve the aforementioned issues [2–5]. These models, however, are commonly constrained to a controlled environment during training or trained on very large datasets, such as the Multi Pose, Illumination, and Expressions (Multi-PIE) dataset [6]. These controlled environments and large datasets ensure the availability of a large collection of multi-view images from the same identity, including the ground truth frontal face images. Moreover, when trained under a supervised learning scenario, these GAN-based models may become constrained to a specific domain and lack robustness to be useful in other domains.

To overcome the aforementioned shortcomings, we propose the FRGAN, a self-supervised GAN for PIFR that can generate realistic frontal face images using an image captured in any other pose and without supervision from the corresponding frontal face image. To avoid supervised learning, the FRGAN rotates the synthesized face image back to the original pose and minimizes the reconstruction loss between this rotated-back image and the input image. To minimize the artifacts caused by rotation, the FRGAN also relies on an adversarial loss computed between the rotated-back image and the input image. Here, the generator is responsible for filling in the self-occluded regions with high-quality details, while the discriminator is responsible for distinguishing the input image from the rotated-back image, where the latter is computed based on the synthesized frontal face image. The FRGAN introduces also the *Random Swap*, which independently and randomly swaps the eye, nose, and mouth regions of the rotated-back and input images during optimization. Challenged by the new data distribution resulting from the *Random Swap*, the discriminator is forced to perform much better and consequently forces the generator to synthesize more realistic images.

2. RELATED WORK

This section reviews the most important works for face frontalization suitable for PIFR by grouping them according to the optimization strategy used.

Supervised: The Two-Pathway Generative Adversarial Network (TPGAN) [3] rotates a non-frontal face image to the frontal view in the 2D space. The TPGAN benefits from a multi-task learning strategy through the use of a global and a local generator. The global generator focuses on shapes and textures, while the local generator pays attention to the key facial areas, i.e., the eyes, nose, and mouth. The Couple-Agent Pose-Guided Generative Adversarial Network (CAPG-GAN) [5] uses a face heatmap to replace the local generator of the TPGAN. The CAPG-GAN concatenates the input face image with the corresponding heatmap to synthesize the cor-

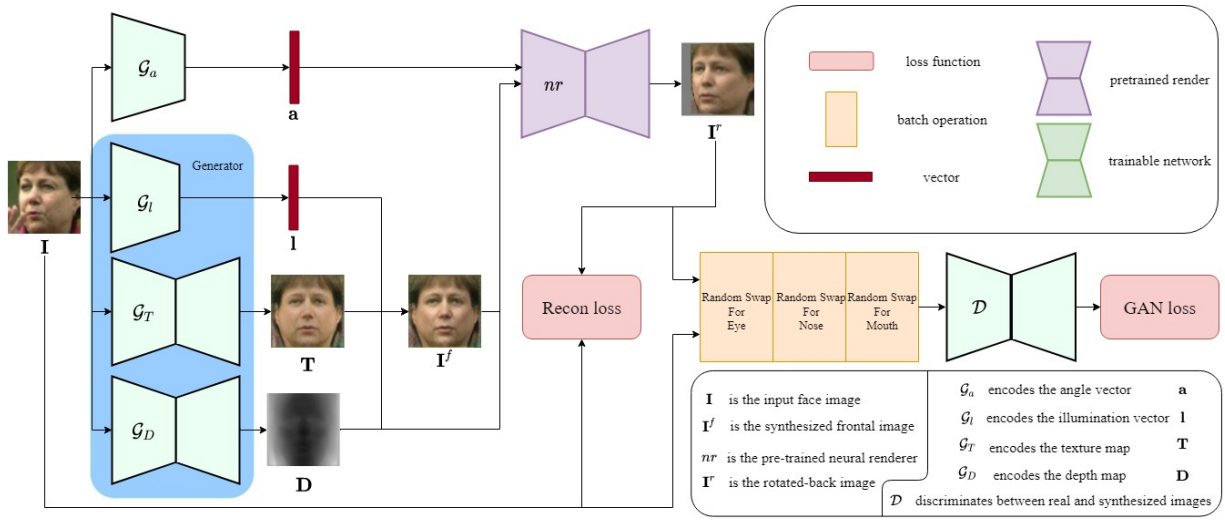

Fig. 1: Block diagram of the FRGAN. Three independent *Random Swap* operations are applied to $\mathbf{I}$ and $\mathbf{I}^r$ before forwarding them to the discriminator. .

responding frontal face image. The High Fidelity Pose Invariant Model (HF-PIM) [7] employs the off-the-shelf 3D-fitting network [8] to generate correspondence fields and extract texture maps to synthesize frontal face images with high quality. 3D-Aided Deep Pose-Invariant Face Recognition Model (3D-PIM) [9] uses a 3D Morphable Model to simulate the frontal view face. The simulated face is then refined by a global network and a local one.

Semi-supervised: Qian *et al.* [10] propose the Face Normalization Model (FNM), which takes two inputs: a non-frontal face image (from the CASIA-WebFace [11] dataset) and a frontal face image (from the Multi-PIE dataset). The FNM computes the reconstruction loss between the input frontal face image and the synthesized frontal face image to semi-supervise the rotation of the input non-frontal face image. The pixel-wise loss between the controlled face image from Multi-PIE and the corresponding frontal image is used as the supervision signal.

Self-supervised: The Face Frontalization GAN (FF-GAN) [12] uses two symmetry masks for visible and occluded regions to integrate the regression of the 3D Morphable Model (3DMM) [1] coefficients in a GAN. Han *et al.* [13] introduce a rotate-and-render (RR) pipeline by using a pre-trained 3D rendering model [14] with a generator with multiple residual blocks. It rotates any input face image to the frontal view and rotates this frontalized face image back to the input view to supervise the synthesis process. The Unsupervised Learning of Probably Symmetric Deformable 3D Objects (UNSUP3D) method [15] uses several sub-networks for decomposing the input image into several vectors and feature maps and subsequently synthesizing a frontal image. It then uses a pre-trained network [14] to rotate this synthesized image back to the input view to estimate the depth of the input image.

Algorithm 1: Random Swap

Input : $\mathbf{I}^1, \mathbf{I}^2, w, h$; Random Swap probability: p
Output: $\mathbf{I}^1_{rs}; \mathbf{I}^2_{rs}$

$p_S \leftarrow rand(0,1)$;
if $p_s > p$ **then**
  $\mathbf{I}^1_{rs} \leftarrow \mathbf{I}^1; \mathbf{I}^2_{rs} \leftarrow \mathbf{I}^2$;
else
  $w^{'} \leftarrow rand_i(0,w); h^{'} \leftarrow rand_i(0,h)$;
  $x \leftarrow rand_i(0, w-w^{'}); y \leftarrow rand_i(0, h-h^{'})$;
  $\mathbf{M}_s \leftarrow zero(\mathbf{M}_s[0{:}w, 0{:}h]); \mathbf{M}_u \leftarrow one(\mathbf{M}_u[0{:}w, 0{:}h])$;
  $\mathbf{M}_s \leftarrow one(\mathbf{M}_s[x{:}x+w^{'}, y{:}y+h^{'}])$;
  $\mathbf{M}_u \leftarrow zero(\mathbf{M}_u[x{:}x+w^{'}, y{:}y+h^{'}])$;
  $\mathbf{I}^1_{rs} \leftarrow \mathbf{M}_s(.*)\mathbf{I}^1 + \mathbf{M}_u(.*)\mathbf{I}^2$;
  $\mathbf{I}^2_{rs} \leftarrow \mathbf{M}_u(.*)\mathbf{I}^1 + \mathbf{M}_s(.*)\mathbf{I}^2$;

3. PROPOSED FRGAN

The FRGAN consists of three parts: a generator for face frontalization based on depth and rotation information, a pre-trained neural renderer for face rotation, and a discriminator (see Fig. 1). More specifically, our model decomposes the input face image, $\mathbf{I}$, of size $w \times h$, into four components by using four independent sub-networks: the angle vector $\mathbf{a}$, the illumination vector $\mathbf{l}$, the depth map $\mathbf{D}$, and the texture map $\mathbf{T}$. For vectors $\mathbf{a}$ and $\mathbf{l}$, the corresponding sub-networks comprise each a single encoder [16] to encode $\mathbf{I}$ into its pose and illumination, respectively. The other sub-networks are traditional generators [16]. The synthesized frontal face image, $\mathbf{I}^f$, is computed based on $\mathbf{T}$, $\mathbf{D}$, and $\mathbf{l}$. The model first computes $\mathbf{D}$ and $\mathbf{l}$ to obtain an appropriate shading for $\mathbf{T}$. It then multiplies the shading and $\mathbf{T}$ to get $\mathbf{I}^f$ with the appropriate depth and illumination. Note that our generator frontalizes the input face, as opposed to that in [13], which only re-paints the images rotated by a pre-trained model.

Guided by $\mathbf{a}$, a pre-trained 3D rendering model [14], hereinafter denoted by $nr(\cdot,\cdot,\cdot)$, rotates $\mathbf{I}^f$ and $\mathbf{D}$ back to input view to produce the reconstructed image, $\mathbf{I}^r$, here-

Fig. 2: Sample visual results on the LFW dataset. Each row represents one identity.

inafter called the rotated-back image. $\mathbf{I}^r$ is used to compute a reconstruction loss with respect to $\mathbf{I}$. The FRGAN uses a discriminator to distinguish between $\mathbf{I}^r$ and $\mathbf{I}$. To improve the generation of $\mathbf{I}^f$, the FRGAN uses the *Random Swap*, a new parameter-free data augmentation strategy inspired by the random erasing technique [17] and traditional data augmentation techniques, i.e., random flip and crop. Random erasing randomly erases a patch from the input image, which wrecks the image structure and may negatively affect the GAN's performance. Conversely, *Random Swap* randomly swaps small regions between images $\mathbf{I}$ and $\mathbf{I}^r$, as explained next.

Random Swap: In practice, the discriminator is likely to do very well and force the generator to fall into a plateau, which is a local optimization stage where the generator is unlikely to get updated any further [18]. One way to leave this plateau is by introducing a stochastic process such as a random walk [19]. *Random Swap* is based on this idea by swapping a random region of random size between the two images according to a certain probability. Algorithm 1 summarizes the pseudo-code of *Random Swap* for two images, $\mathbf{I}^1$ and $\mathbf{I}^2$, of size $w \times h$. Line 1 initializes the probability p_s, where $rand(a, b)$ returns a real random number in the interval $[a, b]$. Line 3 returns the output images $\mathbf{I}^1_{rs}$ and $\mathbf{I}^2_{rs}$ without applying any swap if the condition in Line 2 is satisfied. Otherwise, Line 5 first defines the size of the region to be swapped as $w^{'} \times h^{'}$, where $rand_i(a, b)$ returns a random integer in the interval (a, b). Line 6 then defines the (x, y) coordinates of the upper left corner of the region to be swapped. Lines 7 defines two binary masks of size $w \times h$: $\mathbf{M}_s$ containing only zeros and $\mathbf{M}_u$ containing only ones. Lines 8 and 9 set the values of $\mathbf{M}_s$ and $\mathbf{M}_u$ associated with the region to be swapped to ones and zeros, respectively. Lines 10 and 11 finally compute the output images $\mathbf{I}^1_{rs}$ and $\mathbf{I}^2_{rs}$ by using the binary masks, where $(.*)$ denotes element-wise multiplication.

It is important to note that in the FRGAN, the *Random Swap* is used during optimization to randomly swap three regions of fixed size and position depicting the eyes, nose, and mouth between $\mathbf{I}$ and $\mathbf{I}^r$. These regions are swapped independently with probabilities $p_{eye} = 0.5, p_{nose} = 0.5, p_{mouth} = 0.5$, respectively. Also note that the swapped regions are much smaller compared to the rest of the non-swapped regions, and these regions may not be swapped for all training images. This allows to truly introduce a stochastic process in the form of noise to the data distribution seen by the discriminator. After applying the *Random Swap*, the discriminator no longer depends on the features from any unique region. The generator then has a higher chance to leave the plateau and synthesize more realistic frontalized face images, $\mathbf{I}^f$.

Loss Function: The overall loss of the FRGAN is:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda_3 \mathcal{L}_{adv}. \tag{1}$$

The reconstruction loss, $\mathcal{L}_{recon}$, comprises two parts: the photo-geometric ($\mathcal{L}_{photo}$) and the VGG-perceptual loss ($\mathcal{L}_{vgg}$):

$$\mathcal{L}_{recon} = \lambda_1 \mathcal{L}_{photo} + \lambda_2 \mathcal{L}_{vgg}, \tag{2}$$

where $\lambda_1, \lambda_2, \lambda_3$ are hyper-parameters to control the effect of each loss. $\mathcal{L}_{photo} = ||nr(\mathbf{I}^f, \mathbf{D}, \mathbf{a}) - \mathbf{I}||_1$, and as defined before, $nr(\cdot, \cdot, \cdot)$ is the pre-trained 3D render model that produces $\mathbf{I}^r$. To attain a photo-realistic quality, the adversarial loss, $\mathcal{L}_{adv}$, is defined as [16]:

$$\begin{aligned} \min_{\theta_{\mathcal{G}}} \max_{\theta_{\mathcal{D}}} \mathcal{L}_{adv} = E_{\mathbf{I}^r_{rs} \sim (P(\mathbf{I}^r_{rs}))}[\log \mathcal{D}(\mathbf{I}_{rs}, \mathbf{I}^r_{rs})] \\ + E_{\mathbf{I}_{rs} \sim (P(\mathbf{I}_{rs}))}[\log(1 - \mathcal{D}(\mathbf{I}_{rs}, \mathbf{I}^r_{rs}))]. \end{aligned} \tag{3}$$

After applying the *Random Swap* on $\mathbf{I}$ and $\mathbf{I}^r$, we obtain $\mathbf{I}_{rs}$ and $\mathbf{I}^r_{rs}$, respectively. The discriminator, $\mathcal{D}(\cdot, \cdot)$, attempts to distinguish between $\mathbf{I}_{rs}$ and $\mathbf{I}^r_{rs}$. To synthesize more realistic images, $\mathbf{I}^f$ is then generated by maintaining the pose and visual information. Finally, to preserve the feature embeddings of the input image $\mathbf{I}$, we employ $\mathcal{L}_{vgg}$ as follows:

$$\mathcal{L}_{vgg} = ||vgg(\mathbf{I}^r) - vgg(\mathbf{I})||_1, \tag{4}$$

where $vgg(\cdot)$ gives the feature embeddings produced by a VGG-16 network pre-trained on ImageNet [20].

Fig. 3: Ablation study on the LFW dataset. 1^{st} row: input images, 2^{nd} row: images frontalized without the *Random Swap*. 3^{rd} row: images frontalized with the *Random Swap*

Table 1: ACC(%), AUC(%), FID(%), and FAR values on the LFW and IJB-A datasets.

Method	LFW			IJB-A	
	ACC↑	AUC↑	FID↓	@FAR=.01 ↑	@FAR=.001 ↑
3D-PIM [9]	-	-	-	**98.9 ±0.2**	**97.7 ±0.4**
FF-GAN [12]	96.42	99.45	83.81	85.2 ±1.0	66.3 ±3.3
CAPG-GAN [5]	99.37	99.90	-	-	-
TP-GAN [3]	96.13	99.42	-	-	-
FNM [10]	-	-	-	93.4 ±0.9	83.8 ±2.6
HF-PIM [7]	**99.41**	**99.92**	-	95.2 ±0.7	89.7 ±1.4
RR [13]	98.95	99.91	83.1	92.0 ±0.7	82.5 ±2.5
FRGAN W/O RS	98.20	99.53	85.8	91.7 ±1.0	82.0 ±2.8
FRGAN (ours)	99.40	**99.92**	**81.3**	93.7 ±0.9	86.1 ±1.2

- means not reported or non-applicable

4. EXPERIMENTAL EVALUATION

The FRGAN is implemented on the PyTorch [21] library with python 3.6. It is optimized on a single Nvidia GTX1080 Ti for 30 epochs with $\lambda_1 = 1$, $\lambda_2 = 0.5$, and $\lambda_3 = 1$. All images are of size 64×64. We use the Adam optimizer with a 0.0001 learning rate and 0.0005 weight decay. We use three datasets in the experiments. 1) CASIA-WebFace [11], which consists of 10,575 unique people with 494,414 images in total. This dataset is collected from the Internet as the second-largest public face dataset available. We use CASIA-WebFace as our training data, as done by other compared models. 2) LFW [22], which consists of 5,729 subjects with 3,223 images. LWF is the most commonly used dataset for unconstrained face recognition. We use LFW as the testing data. Compared to CASIA-WebFace, LFW is smaller and has cleaner data, which means fewer false labeled identities. Therefore, we use it for qualitative and quantitative evaluations. And 3) IJB-A [23], which consists of 5,712 images and 2,085 videos from 500 identities, with an average of 11.4 images and 4.2 videos per identity. We use the images from IJB-A for evaluation.

Figure 2 shows qualitative results on the LFW dataset after training on the CASIA-WebFace dataset. Note that the FRGAN synthesizes the background and facial features very well. TPGAN [3], FNM [10], and CAPG-GAN [5] are supervised 2D models. HF-PIM [7], FF-GAN [12], RR [13], and FRGAN are self-supervised 3D models. Note that without the involvement of a GAN, UNSUP3D [15] cannot generate photo-realistic images. Table 1 tabulates the 1:1 face verification accuracy (ACC), area-under-curve (AUC), and the Fréchet Inception Distance (FID) [24] on the LFW [22] dataset. Our face verification model is a ResNet-18 model borrowed from ArcFace, which is also used in [13, 15]. ResNet-18 has less parameters than LightCNN-29 used in [3, 5, 7, 10]. As a result, we report [3, 5, 7, 10] final values in Table 1. The FID measures the distance between the rotated-back image, $\mathbf{I}^r$, and the input image, $\mathbf{I}$. FAR is the false accept rate of 1:1 face verification tested on the IJB-A [23] dataset. A high @FAR=dr value denotes that two images belong to the same identity if their Euclidean distance is smaller than dr at FAR $= dr$. Note that the FRGAN outperforms RR [13] for the FID and achieves a very competitive performance for the other metrics.

Ablation Study: We remove the *Random Swap* from the FRGAN, whose results are reported in Table 1 as *FRGAN W/O RS*. As tabulated, the *Random Swap* improves the quantitative performance. Fig. 2 illustrates examples of frontalized images from the LFW dataset with and without the *Random Swap*. When *Random Swap* is used, the FRGAN can preserve more details located in the regions swapped, e.g., nasolabial folds, the shape of the lips, and the under-eye bags.

Complexity: The FRGAN has 37.1 million learnable parameters for the generator and 8.1 million parameters for the discriminator. RR [13], as an example, 225.1 million and 11.0 million parameters for the generator and discriminator. TPGAN [3] has 62.6 million parameters, while the CAPG-GAN [5] has 67.9 million parameters. The FRGAN attains a competitive quantitative and qualitative performance with less computational complexity in term of number of learnable parameters.

5. CONCLUSION

We proposed the FRGAN, a self-supervised GAN that can frontalize face images acquired in any other pose, where the self-supervision is based on rotating the synthesized frontal image back to the pose of the input face image. The FRGAN uses the novel *Random Swap* strategy to randomly and independently swap three key facial regions between the real and

reconstructed images before forwarding them to the discriminator. By swapping these key regions randomly, the generator is prevented from falling into a plateau and forced to synthesize more realistic images. Our evaluation results confirm the strong performance of our model compared to the SOTA, both quantitatively and qualitatively.

6. REFERENCES

[1] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *SIGGRAPH*, 1999.

[2] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, and J. Kim, "Rotating your face using multi-task deep neural network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[3] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis," in *International Conference on Computer Vision*, 2017.

[4] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning gan for pose-invariant face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[5] Y. Hu, X. Wu, B. Yu, R. He, and Z. Sun, "Pose-guided photorealistic face rotation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[6] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," in *IEEE International Conference on Automatic Face Gesture Recognition*, 2008.

[7] J. Cao, Y. Hu, H. Zhang, R. He, and Z. Sun, "Learning a high fidelity pose invariant model for high-resolution face frontalization," in *Conference on Neural Information Processing Systems (NeurIPS)*, 2018.

[8] X. Zhu, X. Liu, Z. Lei, and S. Z. Li, "Face alignment in full pose range: A 3d total solution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[9] Jian Zhao, Lin Xiong, Yu Cheng, Yi Cheng, Jianshu Li, Li Zhou, Yan Xu, Jayashree Karlekar, Sugiri Pranata, Shengmei Shen, Junliang Xing, Shuicheng Yan, and Jiashi Feng, "3d-aided deep pose-invariant face recognition," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. 7 2018, pp. 1184–1190, International Joint Conferences on Artificial Intelligence Organization.

[10] Y. Qian, W. Deng, and J. Hu, "Unsupervised face normalization with extreme pose and expression in the wild," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[11] D. Yi, Z. Lei, S. Liao, and S. Z Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[12] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Towards large-pose face frontalization in the wild," *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[13] H. Zhou, J. Liu, Z. Liu, Y. Liu, and X. Wang, "Rotate-and-render: Unsupervised photorealistic face rotation from single-view images," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[14] H. Kato, Y. Ushiku, and T. Harada, "Neural 3d mesh renderer," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[15] W. Shangzhe, R. Christian, and V. Andrea, "Unsupervised learning of probably symmetric deformable 3d objects from images in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[16] 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*. 2014.

[17] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

[18] T. Che, Y. Li, A. Paul Jacob, Y. Bengio, and W. Li, "Mode regularized generative adversarial networks," *International Conference on Learning Representations (ICLR)*, 2017.

[19] Gregory F. Lawler and V. Limic, "Random walk: A modern introduction," *Cambridge University Press, Cambridge Studies in Advanced Mathematics*, 2010.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, 2015.

[21] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," *Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

[22] G. B Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments,"

Tech. Rep., Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[23] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[24] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Conference on Neural Information Processing Systems (NeurIPS)*, 2017.