Supplementary Material for

RankGAN: A Maximum Margin Ranking GAN for Generating Faces

Felix Juefei-Xu*[1][0000−0002−0857−8611], Rahul Dey*[2][0000−0002−3594−5122], Vishnu Naresh Boddeti[2][0000−0002−8918−9385], and Marios Savvides[1]

[1] Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2] Michigan State University, East Lansing, MI 48824, USA

Abstract. This supplementary is for the work of RankGAN [3]. Due to space limit in the main paper, we are putting some of the face generation experimental results (in Section 1) and face completion experimental results (in Section 2) in the supplementary. These results demonstrate the effectiveness of the proposed RankGAN method. Further, we mention sample aware vs. sample agnostic RankGAN in Section 3. Finally, we discuss a connection between margin loss and f-divergence in Section 4.

Keywords: Generative adversarial networks · Maximum margin ranking · Face generation.

1 Face Generation Experiments

In Figure 1, we show face generation results from all the three stages of the RankGAN and the WGAN and LSGAN baselines. These images were generated by passing the input images through the encoder $\mathcal{E}$ and using the obtained latent vectors $\mathbf{z}$ as input to the generators. Since WGAN and LSGAN were trained without an encoder, image generation experiments don't preserve identity for them. One can observe that, the Stage-3 generated images look much more aesthetically appealing and realistic than both WGAN and LSGAN images.

2 Face Completion Experiments

2.1 Experimental Setup and Results

Database In addition to the CelebA dataset we used in the main paper, we also collect a single-sample dataset containing 50K frontal face images from 50K individuals, which we call the SSFF dataset. They are sourced from several frontal face datasets including the FRGC v2.0 dataset [9], the MPIE dataset [1],

* These authors contribute equally and should be considered co-first authors.

(a) Input faces. (b) Stage 1 generated faces, Open Set.

(c) Stage 2 generated faces, Open Set. (d) Stage 3 generated faces, Open Set.

(e) WGAN generated faces, Open Set. (f) LSGAN generated faces, Open Set.

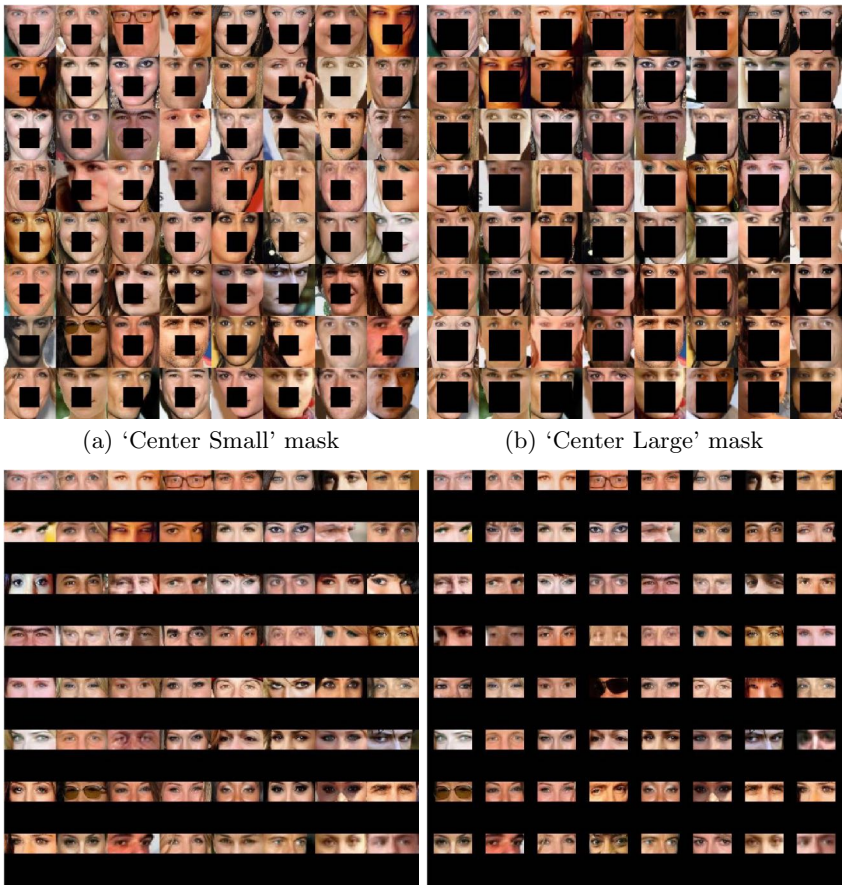
Fig. 1. Visualization of face generation results with RankGAN, Wasserstein GAN (WGAN) and Least Squares GAN (LSGAN). Latent vectors $\mathbf{z}$'s are obtained by passing the input faces through the encoder $\mathcal{E}$. Since WGAN and LSGAN have not been trained with an encoder, the face identities are not preserved.

the ND-Twin dataset [8], and mugshot dataset from Pinellas County Sheriff's Office (PCSO). Training and testing split is 9-1, with the image completion results being based only on the open-set. This dataset is single-sample, which means there is only one single image of a particular subject throughout the entire dataset. Images are aligned using two anchor points on the eyes, and cropped to 64×64.

Occlusion Masks We carried out face completion experiments on four types of facial masks, which are termed as: 'Center Small', 'Center Large', 'Periocular Small', and 'Periocular Large'. Examples can be seen from Figure 2.

Face Completion Results on Large Number of Iterations Some preliminary image completion progression visual results are shown in Figure 3 with

8000 iterations for optimizing for the $\hat{\mathbf{z}}$ using Stage-3 RankGAN. As can be seen, after 8000 iterations, the RankGAN is able to achieve decent image completion results. However, such an algorithm is still slow because it requires optimization over $\hat{\mathbf{z}}$ for any query image. Therefore, for the large-scale experiments to be carried out, we limit the algorithm to optimize for only 2000 iterations for the sake of time. Results can be improved if we allow further iterations.



(a) 'Center Small' mask (b) 'Center Large' mask

(c) 'Periocular Large' mask (d) 'Periocular Small' mask

Fig. 2. Four masks used in our experiments: 'Center Small', 'Center Large', 'Periocular Large', and 'Periocular Small' masks.

2.2 Discussions

As discussed above, we used *Face Completion* task to quantify the performance of our proposed approach. We used four different mask types to perform image

Fig. 3. Image completion visual results: progression of 8000 iterations for optimizing for the $\hat{\mathbf{z}}$.

completion and showed quantitatively that reconstruction improves with each stage of RankGAN. The four masks represent varying difficulties of image inpainting depending on the amount and type of visible image region. We use a large and a small square mask to occlude the central facial region. Another pair of large and small rectangular masks were used to make visible only the periocular region of the face image. In each of the above cases, we measured the performance of different stages of RankGAN using several metrics that have been detailed in Tables 2 to 8. In each case, we showed 48 best and worst examples of image inpainting from the open set in Figures 7 to 12. Note that, the SSFF dataset consists of real mugshot faces and hence for privacy concerns, we only report numerical metrics for the same.

From those figures, it can be seen that the most difficult cases of image completion occur when the faces are not front-facing. This can be related to the relative imbalance in the dataset between front-facing and non front-facing images, specifically the lack of the latter in the dataset. We generally observe better metrics in terms of Inception Score, PSNR, OpenFace NCS and PittPatt Score as we move up the stages. This shows that the higher stages are not just maintaining identity and visual appeal but also adding sharpness and details to the generated images.

Table 1. Data: CelebA, Mask: Center Large

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	2.3286	N/A	N/A	1.0000 (0.9965)	19.6092 (0.9109)
Stage-1	27.09	2.1524	22.76	0.7405	0.6726 (0.9724)	10.2502 (0.7134)
Stage-2	23.69	2.1949	21.87	0.7267	0.6771 (0.9573)	9.9718 (0.8214)
Stage-3	27.31	**2.2846**	**23.30**	**0.7493**	**0.6789 (0.9749)**	**10.4102 (0.7922)**
WGAN	**17.03**	2.2771	23.26	0.7362	0.5554 (0.9156)	8.1031 (0.7373)
LSGAN	23.93	2.2636	23.11	0.7361	0.6676 (0.9659)	10.1482 (0.7154)

Table 2. Data: CelebA, Mask: Center Small

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	2.2697	N/A	N/A	1.0000 (0.9916)	19.6108 (0.9135)
Stage-1	16.54	2.1331	26.64	0.8885	0.8246 (0.9913)	12.8730 (0.8201)
Stage-2	15.35	2.1898	25.68	0.8808	0.8331 (0.9946)	**13.0946 (0.8601)**
Stage-3	15.45	2.2903	26.65	**0.8888**	**0.8399 (0.9897)**	13.0318 (0.8260)
WGAN	**12.67**	2.2498	**26.69**	0.8834	0.7650 (0.9956)	11.3301 (0.8338)
LSGAN	15.30	**2.3088**	26.54	0.8833	0.8361 (0.9951)	12.8985 (0.8270)

Table 3. Data: CelebA, Mask: Periocular Large

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	2.257	N/A	N/A	1.0000 (0.9993)	19.6197 (0.9417)
Stage-1	60.96	1.732	15.02	0.4978	0.6793 (0.9643)	12.8902 (0.7345)
Stage-2	48.19	**1.839**	15.86	0.5550	**0.7080 (0.9812)**	13.4623 (0.8609)
Stage-3	65.83	1.818	15.80	**0.5629**	0.6892 (0.9788)	**13.5328 (0.8778)**
WGAN	**25.98**	1.756	**16.50**	0.5418	0.6591 (0.9510)	12.7374 (0.8543)
LSGAN	40.57	1.829	15.86	0.5216	0.6882 (0.9663)	12.8952 (0.7441)

Table 4. Data: CelebA, Mask: Periocular Small

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	2.2517	N/A	N/A	1.0000 (0.9992)	19.4195 (0.9273)
Stage-1	67.95	1.6861	13.97	0.4096	0.5810 (0.9506)	10.3975 (0.7319)
Stage-2	50.95	1.6034	14.47	0.4794	**0.6283 (0.9659)**	10.7539 (0.8314)
Stage-3	70.68	1.6394	14.54	**0.4884**	0.5928 (0.9507)	**11.1033 (0.7982)**
WGAN	**27.15**	**1.8060**	**14.69**	0.4649	0.5742 (0.9404)	10.0387 (0.7990)
LSGAN	42.59	1.7341	14.61	0.4396	0.5861 (0.9501)	10.4759 (0.7453)

Table 5. Data: SSFF, Mask: Center Small

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	1.845	N/A	N/A	1.0000 (0.9987)	19.1864 (0.9675)
Stage-1	44.57	1.766	**30.47**	**0.9087**	0.7447 (0.9868)	14.4907 (0.9859)
Stage-2	45.34	1.766	29.87	0.9061	0.7503 (0.9819)	**14.5053 (0.9929)**
Stage-3	**43.57**	**1.773**	30.07	0.9070	**0.7559 (0.9820)**	14.0614 (0.9924)

Table 6. Data: SSFF, Mask: Center Large

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	1.803	N/A	N/A	1.0000 (0.9973)	19.5385 (0.9802)
Stage-1	**81.01**	**1.883**	**26.29**	**0.7906**	0.5307 (0.9104)	8.1911 (0.9594)
Stage-2	82.02	1.840	26.26	0.7877	0.5360 (0.9214)	**8.2555 (0.9611)**
Stage-3	92.75	1.853	26.16	0.7835	**0.5394 (0.9115)**	8.2473 (0.9676)

Table 7. Data: SSFF, Mask: Periocular Large

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	1.803	N/A	N/A	1.0000 (0.9972)	19.8658 (0.9866)
Stage-1	**88.96**	1.717	**18.85**	**0.6041**	0.6561 (0.9571)	14.8129 (0.9860)
Stage-2	89.21	**1.721**	18.51	0.5988	0.6415 (0.9419)	14.7613 (0.9694)
Stage-3	105.85	1.660	18.74	0.5861	**0.6568 (0.9589)**	**14.9813 (0.9898)**

Table 8. Data: SSFF, Mask: Periocular Small

	FID	Inception	PSNR	SSIM	OpenFace (AUC)	PittPatt (AUC)
Original	N/A	1.803	N/A	N/A	1.0000 (0.9991)	19.1976 (0.9679)
Stage-1	**102.30**	1.634	17.15	**0.5218**	**0.5473 (0.9094)**	11.2008 (0.9673)
Stage-2	104.59	**1.665**	16.84	0.5177	0.5139 (0.8944)	11.2372 (0.9808)
Stage-3	118.66	1.636	**17.24**	0.5012	0.5412 (0.8965)	**11.4808 (0.9800)**

3 Sample Aware vs. Sample Agnostic RankGAN

In this section, we discuss further on the design principles regarding the Encoder $\mathcal{E}$. The RankGAN framework can be trained both with and without the encoder. We call the former method sample aware training and the latter, sample agnostic. In the case of the latter, as the stage progresses, the discriminator ranks generated samples from various stages while being sample agnostic, *i.e.*, the samples being ranked are arbitrary. This is still a viable solution for the ranker to provide meaningful feedback for the entire network. However, such sample agnostic ranking is 'fuzzy' and has been empirically observed to result in slower convergence of the ranking loss. This has been explored in [2]. The training pipeline for sample agnostic RankGAN is shown in Figure 4.

4 Connection Between Hinge Loss and f-Divergence

A connection can be established between the hinge loss we used for RankGAN, and f-divergence. In particular, minimizing the hinge loss is closely related to minimizing one particular instance of the f-divergence, which is the variational distance. The goal of the discussion is to establish a correspondence between the family of surrogate loss (convex upper bound on the 0-1 loss, such as hinge loss) functions and the family of f-divergences. It can be shown that any surrogate loss induces a corresponding f-divergence, and any f-divergence satisfying certain conditions corresponds to a family of surrogate loss functions. Readers are

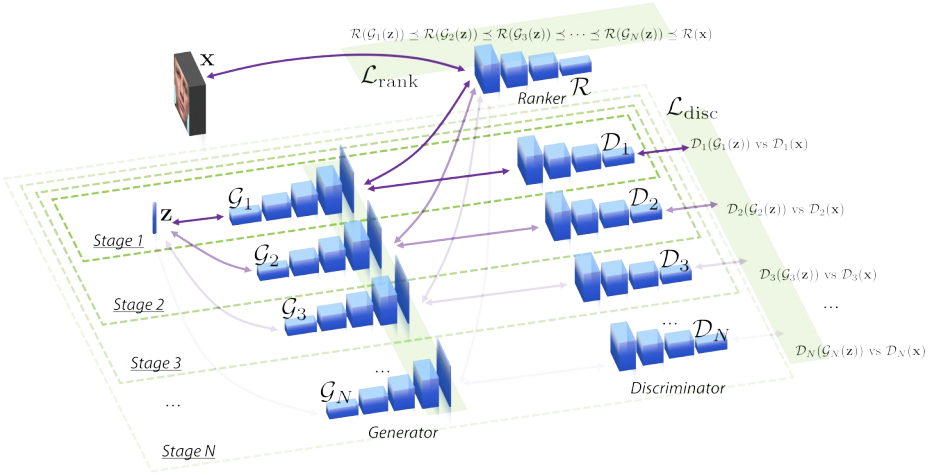


Fig. 4. Flowchart of the proposed RankGAN method without the Encoder $\mathcal{E}$.

encouraged to refer to the work of Ngugen *et al.* [7, 4, 6, 5] for such a connection in detail.

References

1. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing **28**(5), 807–813 (2010)
2. Juefei-Xu, F., Boddeti, V.N., Savvides, M.: Gang of GANs: Generative Adversarial Networks with Maximum Margin Ranking. arXiv preprint arXiv:1704.04865 (2017)
3. Juefei-Xu, F., Dey, R., Bodetti, V.N., Savvides, M.: RankGAN: A Maximum Margin Ranking GAN for Generating Faces. In: Proceedings of the Asian Conference on Computer Vision (ACCV) (December 2018)
4. Nguyen, X., Wainwright, M.J., Jordan, M.I.: Divergences, surrogate loss functions and experimental design. In: NIPS. pp. 1011–1018 (2005)
5. Nguyen, X., Wainwright, M.J., Jordan, M.I.: On Divergences, Surrogate Loss Functions, and Decentralized Detection. arXiv preprint math.ST/0510521 (2005)
6. Nguyen, X., Wainwright, M.J., Jordan, M.I.: On information divergence measures, surrogate loss functions and decentralized hypothesis testing. Department of EECS and Department of Statistics UC Berkeley, CA (2005)
7. Nguyen, X., Wainwright, M.J., Jordan, M.I.: On surrogate loss functions and f-divergences. The Annals of Statistics pp. 876–904 (2009)
8. Phillips, P.J., Flynn, P.J., Bowyer, K.W., Bruegge, R.W.V., Grother, P.J., Quinn, G.W., Pruitt, M.: Distinguishing identical twins by face recognition. In: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. pp. 185–192. IEEE (2011)
9. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on. vol. 1, pp. 947–954. IEEE (2005)

(a) Original faces.

(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.

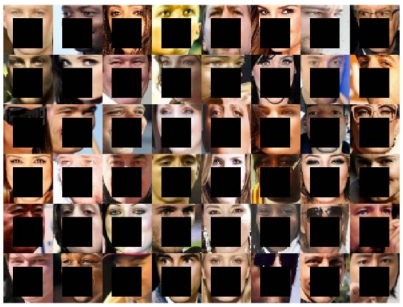
(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 5. Best completion results with RankGAN on CelebA, 'Center Large' mask.

(a) Original faces.



(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.

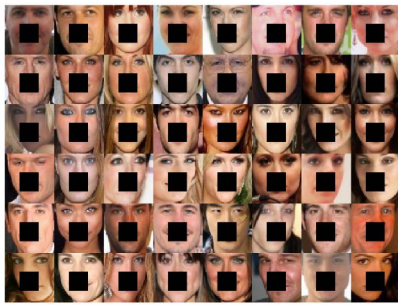
(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 6. Worst completion results with RankGAN on CelebA, 'Center Large' mask.

(a) Original faces.



(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.

(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 7. Best completion results with RankGAN on CelebA, 'Center Small' mask.

(a) Original faces.

(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.

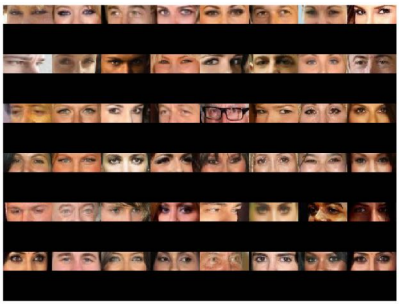
(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 8. Worst completion results with RankGAN on CelebA, 'Center Small' mask.

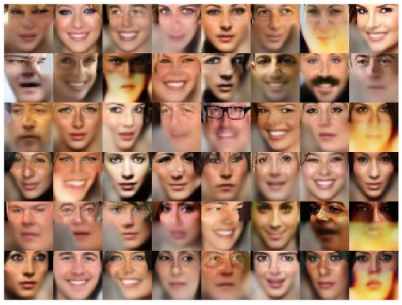
(a) Original faces.



(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.



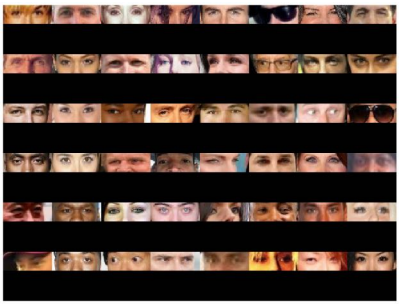
(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 9. Best completion results with RankGAN on CelebA, 'Periocular Large' mask.

(a) Original faces.



(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.

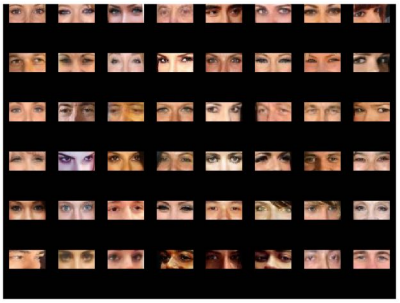
(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 10. Worst completion results with RankGAN on CelebA, 'Periocular Large' mask.

(a) Original faces.



(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.

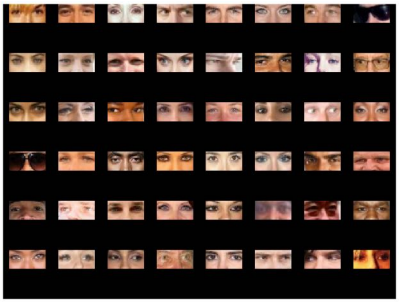
(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 11. Best completion results with RankGAN on CelebA, 'Periocular Small' mask.

(a) Original faces.



(b) Masked faces. ('Center Large')

(c) WGAN completion, Open Set.

(d) LSGAN completion, Open Set.

(e) Stage 1 completion, Open Set.

(f) Stage 2 completion, Open Set.

(g) Stage 3 completion, Open Set.

Fig. 12. Worst completion results with RankGAN on CelebA, 'Periocular Small' mask.