



Learning Representations for Masked Facial Recovery

Zaigham A. Randhawa, Shivang Patel, Donald A. Adjeroh,
and Gianfranco Doretto

West Virginia University, Morgantown, WV 26506, USA
{zar00002,sap00008,daadjero,h,gidoretto}@mix.wvu.edu

Abstract. The pandemic of these very recent years has led to a dramatic increase in people wearing protective masks in public venues. This poses obvious challenges to the pervasive use of face recognition technology that now is suffering a decline in performance. One way to address the problem is to revert to face recovery methods as a preprocessing step. Current approaches to face reconstruction and manipulation leverage the ability to model the face manifold, but tend to be generic. We introduce a method that is specific for the recovery of the face image from an image of the same individual wearing a mask. We do so by designing a specialized GAN inversion method, based on an appropriate set of losses for learning an unmasking encoder. With extensive experiments, we show that the approach is effective at unmasking face images. In addition, we also show that the identity information is preserved sufficiently well to improve face verification performance based on several face recognition benchmark datasets.

Keywords: Face unmasking · GAN inversion · Face verification

1 Introduction

Face recognition in unconstrained environments is still a challenging problem, despite the impressive progress of recent approaches based on deep learning [8, 43]. A major factor affecting performance is the presence of occluded parts of the face. Although face recognition under occlusions is not a new problem [53], its relevance has been refreshed in light of the COVID-19 pandemic, which has led to a dramatic increase of people wearing protective masks of various kinds in public venues. This new *status quo* is posing challenges to the pervasive use of face recognition technology, leading to government institutions initiating studies to better evaluate the effects of face masks on current approaches [36].

There is more than one way to mitigate the loss of performance of face matchers dealing with face images wearing masks [53], and one of them is to attempt to reconstruct the face appearance on the occluded region. The main advantage of this approach is that it can be used to potentially improve the performance of any face matcher.

Recent approaches for face reconstruction and manipulation based on deep learning [11, 40, 41] leverage the extraordinary generative power of these methods in capturing the statistics of the face manifold [24]. In this work we plan to harness that capability even further. Differently than previous approaches, which aim at generic face manipulations, we develop a method that is specifically focussed on unmasking images of faces wearing masks. Our method does not involve the detection or segmentation of face masks, and can be used as a preprocessing step to unmask a face image, which can then be fed to a face matcher.

We frame the problem as a special instance of a GAN inversion [7, 42], where the GAN network is a StyleGAN2 architecture [24]. We do so by designing a set of losses and a training procedure for learning an encoder network that maps the input image of a face wearing a mask onto an appropriate code space of faces not wearing masks. This is meant to be the input space of the generator network that will then reproduce the face image without mask.

Ultimately, the challenge is to generate face images that preserve the identity of the input in order to improve face recognition performance. This is why we test our approach with several face recognition datasets. In particular, we show that it can produce compelling face reconstructions with competitive image quality metrics. In addition, we evaluate extensively how our method works for improving face verification under several face masking conditions.

2 Relevant Works

Image Recovery Under Occlusion and Recognition. Our work can be considered as a type of occlusion recovery which could be used for face recognition [53]. A lot of works treat occlusions as noise and compress the occluded images/faces down to a lower resolution or latent space. This helps to filter out the noise and reconstruct the images back at a higher resolution. Some of these approaches employ more traditional methods like sparse representations [12, 22, 22, 25, 28, 32, 45, 56] and PCA [9, 32, 38], while others rely on neural networks [13, 17, 26, 50] to accomplish this task. Our approach is more similar to the latter. Additionally, some of these methods are occlusion aware and rely on occlusion segmentation or contours to help with the image recovery process [2, 5, 10, 47, 51, 52]. Therefore, occlusion map prediction is part of their model. On the other hand, we do not require any kind of occlusion information. Additionally, some works try to make occlusion neutral feature-extractors/encoders [6, 30] or train the face matching networks to adapt to occlusions [46]. In that sense, our method is not occlusion neutral, and although we do employ ArcFace [8] and FaceNet [43] to help with the training and facial verification tasks, at no point do we train these matchers.

GAN Inversion. A lot of approaches used local discriminators and global discriminators and trained GANs from scratch to reproduce faces/images free of occlusions [14, 21, 27]. Some others used variations of cyclic losses for image/facial

deocclusion [19, 29]. We decided to use a pretrained StyleGAN2 [24] as our generator. As stated in [1, 57], real life face reconstruction via StyleGAN based on the original \mathcal{W} space is a very hard task. Some approaches have extended the \mathcal{W} space to new ones, named $\mathcal{W}+$ [41, 49], \mathcal{W}^* [44], p [57] etc, while others have trained the decoders with various losses to achieve exact facial GAN inversion [7, 16, 42]. We decided to build our approach based on the $\mathcal{W}+$ space via the pSp model [41] and keep the generator fixed. Additionally, related to us, a couple of works employ ℓ -norm losses in the latent space of their StyleGAN architectures [16, 37, 40]. However, none of them use the latent space loss for image recovery or inpainting with StyleGAN.

We also report that [27, 39, 49] employed face parser losses in the output space while training for GAN inversion/facial reconstruction). However, our work does not use such a loss. Perceptual losses like LPIPS were also used by us and other works [33, 49, 51]. ID losses based on Facenet, ArcFace, LightCNN etc. were also used while training for GAN inversion/face unmasking [14, 33, 37, 49, 51, 55]. Just like [27, 29, 33, 34, 49], we also employed an ℓ -norm loss on the output image space to help with image reconstruction.

3 Method

Given an image M of a face wearing a mask, we are interested in developing an approach for face unmasking, which is the task of mapping M onto a new image U , depicting the same person in M , only without the mask. We assume that the unmasking process can be modeled by the relationship $U = g \circ f(M)$, where f maps M onto a representation w , and g generates U from the representation. We do not make assumptions about the specific type of face mask, nor do we require a mask detection or segmentation process to be involved in the unmasking task. We do however, require the face in M to be aligned in terms of 2D position, 2D orientation and scale with the nominal alignment of the dataset used for training the model $g \circ f$.

3.1 Baseline Model

In the case when the face in the image M was not wearing a mask, since no mask needs to be removed, we would expect this condition to be true: $U = M$. Also, let us indicate with $T \doteq U = M$ the image of the face without mask. Therefore, the model $g \circ f_0$ should behave like a face autoencoder, where the encoder in this particular case is indicated with f_0 . While there are several implementations of face autoencoders [11, 40, 41], since we are ultimately interested in evaluating how the approach would improve the performance of face recognition, we want one that executes face autoencodings that are photorealistic, and that can maintain face identity. The state-of-the art in that category is the pSp model [41], where the generator g is a StyleGAN2 network [24], and the encoder f is based on a feature pyramid model built on top of a ResNet backbone, and followed by a mapping to a set of 18 *styles*. The styles capture different levels of image

detail, roughly divided in three groups, coarse, medium, and fine. Every style is a 512-dimensional vector. The collection of the 18 style vectors constitutes the representation w , which is an element of the space referred to as $\mathcal{W}+$ in [41].

The training of the pSp model is approached as a “GAN inversion” task, meaning that the generator network is trained offline (i.e., StyleGAN2), and is kept locked while only the encoder f_0 is being trained, with the task of “learning to invert” the operation of the generator. This approach is mainly due to the success of StyleGAN2 in modelling the face space, and also due to the difficulty in designing and training such kind of models.

In order to train the encoder f_0 , the pSp model combines a number of losses. The first one is a reconstruction loss based on the ℓ_2 -norm

$$\mathcal{L}_R(T) = \|T - g \circ f_0(T)\|_2. \quad (1)$$

The second aims at maintaining the perceptual similarity between input and reconstructions, and is based on the LPIPS metric $P(\cdot)$ [54]

$$\mathcal{L}_{LPIPS}(T) = \|P(T) - P(g \circ f_0(T))\|_2. \quad (2)$$

In order to preserve the face identity of the input in the reconstructions, an identity preserving loss is used to maximize the cosine similarity between the normalized ArcFace [8] representations $AF(\cdot)$ of the input image and the reconstruction

$$\mathcal{L}_{ID}(T) = 1 - AF(T) \cdot AF(g \circ f_0(T)). \quad (3)$$

The encoder f_0 , which we refer to as the *baseline encoder*, is then trained by minimizing this loss, which is written on a per-image basis as

$$\mathcal{L}_0(T) = \mathcal{L}_R(T) + \alpha \mathcal{L}_{LPIPS}(T) + \beta \mathcal{L}_{ID}(T), \quad (4)$$

where α , and β are hyperparameters striking a balance between the loss terms.

3.2 Unmasking Model

Given an image M with a masked face, we can still make the assumption that in our original model $g \circ f$, g is a face image generator, modeled with StyleGAN2, and that we keep it fixed. Therefore, training the encoder f becomes a specialized GAN inversion problem. If T is an image of a face not wearing a mask, we make the assumption that T is identical to M , except for the area of M corresponding to the pixels on the face mask.

To train the encoder f we combine several losses, most of which are a modification of those used to train the baseline model f_0 . See Fig. 1. Specifically, we require the autoencoding of M (i.e., U) to be close to T in the ℓ_2 -norm sense

$$\mathcal{L}_R(T, M) = \|T - g \circ f(M)\|_2. \quad (5)$$

We also want the autoencoding of M to be perceptually similar to T according to the LPIPS metric $P(\cdot)$ by minimizing

$$\mathcal{L}_{LPIPS}(T, M) = \|P(T) - P(g \circ f(M))\|_2. \quad (6)$$

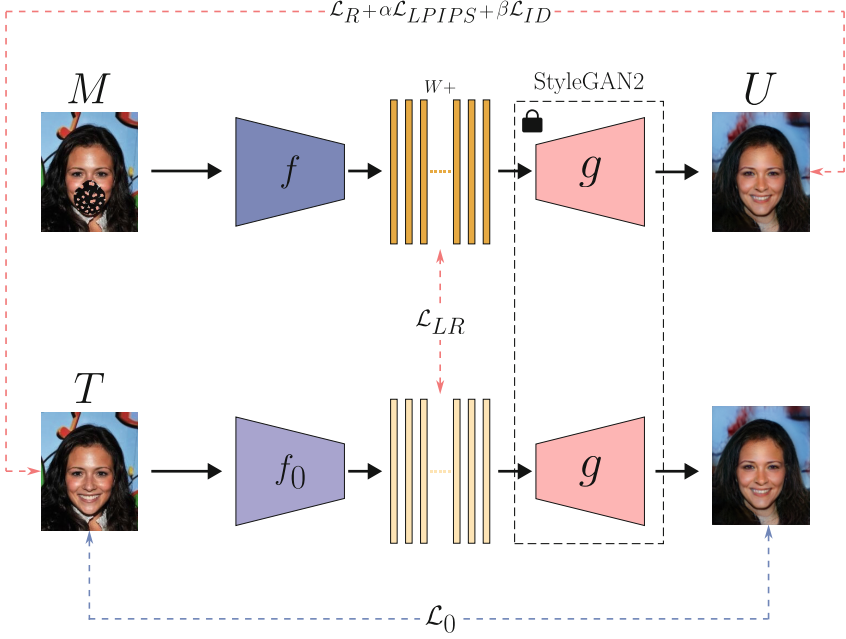


Fig. 1. Unmasking architecture. Overview of the architecture and the losses used to train the baseline encoder f_0 , and the unmasking encoder f . The generator g is kept fixed at all times. T is a face image. M is the same face image wearing a mask. U is the autoencoded version of M .

In addition, the identity of the autoencoding of M should be as close to the identity of T as possible, and for that we maximise the similarity between the respective normalized ArcFace representations $AF(\cdot)$ by minimizing

$$\mathcal{L}_{ID}(T, M) = 1 - AF(T) \cdot AF(g \circ f(M)). \quad (7)$$

We also observe that ideally, the baseline model should be such that $T = g \circ f_0(T)$. Therefore, we would want as much as possible that $g \circ f_0(T) = g \circ f(M)$, but this could be achieved by simply having $f_0(T) = f(M)$. So, we encourage that with the loss

$$\mathcal{L}_{LR}(T, M) = \|f_0(T) - f(M)\|_2, \quad (8)$$

which we name *latent reconstruction loss*. Finally, the encoder f , which we refer to as the *unmasking encoder*, is trained by minimizing this combined *unmasking loss*, which is written on a per-image basis as

$$\mathcal{L}_{UM}(T, M) = \mathcal{L}_R(T, M) + \alpha \mathcal{L}_{LPIPS}(T, M) + \beta \mathcal{L}_{ID}(T, M) + \gamma \mathcal{L}_{LR}(T, M), \quad (9)$$

where α , β , and γ are hyperparameters striking a balance between the loss terms.

Table 1. Datasets. Quantitative summary of the datasets used.

Name	Original # of images	Total # of masked images	# of train, test IDs	# of train, test images
FFHQ	70000	69794	—, —	55811, 13593
CelebA	202599	196999	8141, 2036	157597, 39402
LFW	13233	13168	4754, 1144	10794, 2374
RMFRD	2118 M + 90468	806	218, 64	597, 209

3.3 Datasets

Because of the pandemic, there is a number of datasets and tools to add masks to face images. For instance, Masked-FaceNet is a dataset with faces from the FFHQ dataset wearing masks correctly and incorrectly [4], RMFD is a collection of real world masked faces, including also face images with same identity not wearing masks [48], and so is DS-IMF [35]. The MAFA dataset [15] has a lot of real world masked images, but neither with identification information nor with corresponding identities without wearing masks. In our experiments we used the FFHQ dataset [23], CelebA [31], and LFW [20], and we used the MasktheFace toolkit [3] to create the pairs of face images (T, M) , where M is a version of T with a synthetic mask added. Note that the MasktheFace toolkit failed to mask faces of certain images and those were not included in further training and testing of the models. In addition, we used a subsection of RFRD called RMFRD, which has real life masks only. Table 1 gives details on the size of the datasets used.

3.4 Implementation Details

For training our approach, we assume that a StyleGAN2 generator model g is given to us and is kept locked. Then, we train the baseline encoder f_0 with the loss (4). Subsequently, we use f_0 to initialize the unmasking encoder f , and we train it with the loss (9). Also, we conduct the experiments by first learning the model for the dataset with higher resolution, and then we use the baseline model to initialize the baseline model of the dataset with the immediate smaller resolution. So, we start from the FFHQ dataset [23], then we process CelebA [31], then LFW [20], and finally RMFRD [48].

The generator g based on StyleGAN2 allows to generate images at 1024×1024 resolution with an architecture based on 18 layers. It is also possible to use only the first 14 layers of StyleGAN2 and work with a model that generates images at 256×256 resolution [41], which is also much faster to train. We verified, as reported in Sect. 4, that working with the smaller network does not affect face verification results significantly, since the images are downscaled before feeding them to the face matcher. Therefore, unless otherwise stated, we always use the model with the smaller 14 layers generator architecture.

The approach we use assumes that the input images have faces that are sufficiently aligned. This is the case for FFHQ and CelebA. For LFW, we used the deep funneled images [20], which correct for the orientation of the faces and properly align them. We also crop a 150×150 region out of the original 250×250 images to leave out image areas containing significant background clutter that were making the training difficult to converge. Note also that the cropped images were then resized up to 256×256 prior to be used.

Additionally, RMFRD has very low quality images and a very variable resolution, and it is in general a very challenging dataset. Because of this, we used OpenCV to only keep faces where we could detect both eyes, and we rotated the faces to make the eyes horizontal, and resize them to 256×256 . Moreover, this dataset has faces wearing real masks as M images, and there are no identical images with faces without masks as T images. Therefore, the training is approached in two phases. First, we used the unmasked faces in the training set to train the baseline and the unmasking models just like we did for the other datasets. Second, we fine-tune the model with the images with real masks as follows. The losses (5) and (6) are computed based only on the periorbital region of the face because it is visible, which is identified automatically from the position of the eyes. The loss (7), instead, uses as T , an image with the same identity and that is not wearing a mask. Finally, in the loss (8) T is replaced with an estimate \hat{T} of a face image without mask, generated by g to have the same periorbital region of the masked face in the M image according to the ℓ_2 -norm.

4 Experimental Results

We evaluate our unmasking model extensively, by providing results pertaining the unmasking of face images wearing masks, the image quality metrics of those images, and we evaluate to what extent the unmasking process might help improving the performance of a face matcher. We use four datasets, FFHQ [23], CelebA [31], and LFW [20], and RMFRD [48], and two face matchers, ArcFace [8] and FaceNet [43].

Face Verification Notation. The face verification experiments are conducted with different unmasking settings. The notation used to indicate these settings is defined as follows. MM indicates when both the probe and the gallery face images are masked. MT indicates when the probe image is masked and gallery images are not. UU indicates that both probe and gallery images were originally masked but they were both unmasked by our approach before verification. UT indicates that the probe was originally masked but was unmasked by our approach before verification and the gallery images were not masked. TT indicates when the probe and the gallery images were not masked. This is expected to provide the upper bound results.

18 Layers vs 14 Layers Architecture. We compared the two generator architectures based on 14 and 18 layers in terms of face verification performance as well as image quality of the unmasked images. A key difference between the



Fig. 2. Face unmasking. Qualitative face unmasking results on various datasets. For each row we have original face image (T), masked image (M), and unmasked image (U).

architectures is that the 14 layers provides 256×256 images and the 18 layers provides 1024×1024 images. Table 2 shows the results on CelebA. The metrics used are area under curve (AUC) for face verification, and peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM) for image quality [18]. The main conclusion is that the difference between architectures is not significant. For AUC purposes, this is not surprising because the images are downsampled to 112×112 for the face matchers. Also, the pSp framework only accepts input at a resolution of 256×256 . Therefore when calculating SSIM and PSNR, the 1024×1024 images are downsized to 256×256 for the model input and the resulting output quality is very similar to that of the 14 layers model. The 18 layer model has slightly better facial verification AUCs because it has more expressive power. On the other hand, the 14 layer model has marginally better SSIM and PSNR values because it has to produce images at a lower scale. Therefore, in the remaining experiments we used the 14 layered architecture, which also allows for a faster training time.

Face Unmasking. We conducted face unmasking experiments with the four datasets FFHQ, CelebA, LFW, and RMFRD. Qualitative unmasking results can be seen in Fig. 2. The unmasked images have resolution 256×256 . For FFHQ the mask was simply black, for CelebA and LFW we show the case of different

Table 2. Architecture depth. Effects of the generator network depth on face verification in terms of AUC, and unmasking image quality in terms of PSNR and SSIM. The dataset used is CelebA.

Architecture	UU	UT	PSNR	SSIM
14-layer	0.9590	0.9422	19.00	0.75891
18-layer	0.9596	0.9440	18.51	0.75568

Table 3. Face verification. AUC, SSIM, and PSNR values of an ablation study of the unmasking loss for different face verification settings. ArcFace was used as face matcher.

Dataset	\mathcal{L}_R	\mathcal{L}_{LIPS}	\mathcal{L}_{ID}	\mathcal{L}_{LR}	UU		TT	SSIM	PSNR
					MM	MT			
FFHQ	✓	✓	✓	✓				0.69450	17.92
CelebA	✓				0.750 0.825	0.750 0.824	0.891		
CelebA				✓	0.833 0.825	0.840 0.824	0.891	0.72905	18.30
CelebA	✓	✓	✓		0.952 0.931	0.964 0.940	0.984		
CelebA		✓	✓	✓	0.947 0.931	0.962 0.940	0.984		
CelebA	✓	✓	✓	✓	0.959 0.931	0.971 0.940	0.984	0.75891	19.00
LFW	✓	✓	✓		0.944 0.952	0.958 0.958	0.990		
LFW	✓	✓	✓	✓	0.957 0.952	0.968 0.958	0.990	0.67737	17.04
RMFRD	✓	✓	✓	✓		0.609 0.602			

shape and color masks added, and the masks in RMFRD are real masks worn by the subjects.

Ablation of the Unmasking Loss. In Table 3 we report an ablation study where in the unmasking loss (9) we include only the components indicated. We do so for a face verification experiment using the CelebA, LFW, and RMFRD datasets. The AUC values highlight the contribution coming from using only \mathcal{L}_R , only \mathcal{L}_{LR} , and how much performance deteriorates when each of them is removed from the full model.

The first two experiments in Table 3 concentrate on the cases when the model is only trained with either \mathcal{L}_R or \mathcal{L}_{LR} . The facial verification results for these two models only use 3000 images from the test dataset (the rest of the models use the entirety of the test dataset). Please note that \mathcal{L}_R is not enough to learn a model where the UU case is better than the MM, or the UT case is better than the MT, highlighting the fact that adding the \mathcal{L}_{LIPS} and \mathcal{L}_{ID} is important for \mathcal{L}_R . For CelebA, a model with only \mathcal{L}_{LR} allows the UU and UT cases to outperform the MM and MT cases respectively, but adding the rest of the losses

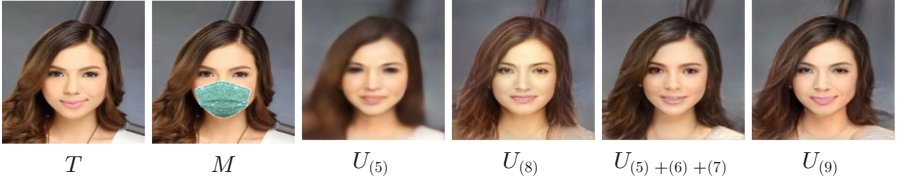


Fig. 3. Ablation of face unmasking. Left two images: original face image (T) and same image with mask (M). The right four images are unmasked versions of M obtained with different models.

further increases the facial verification AUCs by a large margin. Additionally, for LFW, which is a lower resolution dataset, removing \mathcal{L}_{LR} causes UU and UT to not outperform MM and MT respectively, further highlighting its importance. The final takeaway is that the full model allows for UU and UT to surpass MM and MT, respectively, and allows to approach the upper bound set by the TT scenario.

Also, Table 3 summarizes the results on image quality metrics such as SSIM and PSNR. LFW has lower metrics than the other datasets because it has the lowest quality images, especially due to the aforementioned cropping of the T , and M images from a size of 250×250 to 150×150 followed by an upsampling to 256×256 . Additionally, FFHQ also performs worse than CelebA because the comparison between U and T is done at a resolution of 256×256 , for which the T has to be downsampled, which detrimentally affects the SSIM and PSNR metrics. CelebA images, instead, are three times more than FFHQ for training, and are upsampled from 178×218 to 256×256 , suffering the least amount of distortion. Finally, the CelebA model with only \mathcal{L}_{LR} performs worse than the full model in terms of both SSIM and PSNR. This is because the latter has more constraints to satisfy, which gives better GAN-inversion results. More qualitative results can be seen in Fig. 3, which follows similar trends.

Figure 3, instead, shows a qualitative ablation of the unmasking results, obtained by progressively adding more components in the unmasking loss (9). From the left, we have the original face image (T), and the version wearing the mask (M), followed by an unmasked face (U) with a model trained only with \mathcal{L}_R , which is rather blurry, despite the fact that the generator is a StyleGAN2 network that produces sharp face images. Then, fourth from the left, the unmasked face was obtained with a model trained only with \mathcal{L}_{LR} , which is relatively sharp, but the identity drift is noticeable. Second from the right, the unmasked face was obtained with a model without only \mathcal{L}_{LR} , whereas the last image was unmasked by the full model.

Face Verification with FaceNet. In Table 4 we report face verification results on CelebA and LFW based on FaceNet [43] as the face matcher. The train/test split used are the same as those used in the complete \mathcal{L}_{UM} models in Table 3. Note that the results establish the same relationships between the various settings as those deductible from Table 3. This is relevant because now we have

trained the unmasking models with one face matcher (i.e., ArcFace), while we have tested them with another one (i.e., FaceNet), confirming that even the previous results were not subject to strong biases, since models were trained and tested with the same face matcher, since ArcFace is used in the loss (7).

Table 4. Face verification. AUC face verification results on CelebA and LFW. FaceNet was used as face matcher during testing.

Dataset	\mathcal{L}_R	\mathcal{L}_{LPIPS}	\mathcal{L}_{ID}	\mathcal{L}_{LR}	UU	UT
					MM	MT
CelebA	✓	✓	✓	✓	0.868	0.867
					0.832	0.814
LFW	✓	✓	✓	✓	0.9833	0.9845
					0.9834	0.9844

5 Conclusions

In this work we have proposed a method for unmasking the face image of a subject wearing a mask. We formulate the problem as a GAN inversion, because we leverage the generative modeling of the face manifold of current methods. We designed a set of losses to learn an unmasking encoder that enables mapping the input image onto a new face image. Our set of experiments show that the unmasking process recovers compelling face images, with competitive image quality metrics. In addition, by testing the unmasking process with two face matchers, our set of results on face verification confirms that the identity is preserved sufficiently well to provide a consistent significant improvement on three commonly used face recognition benchmarks.

Acknowledgements. This material is based upon work supported in part by the Center for Identification Technology Research and the National Science Foundation under Grants No. 1650474 and No. 1920920.

References

1. Abdal, R., Qin, Y., Wonka, P.: Image2StyleGAN: how to embed images into the styleGAN latent space? In: IEEE ICCV, pp. 4432–4441 (2019)
2. Abdal, R., Zhu, P., Mitra, N.J., Wonka, P.: Labels4Free: unsupervised segmentation using StyleGAN. In: IEEE ICCV, pp. 13970–13979 (2021)
3. Anwar, A., Raychowdhury, A.: Masked face recognition for secure authentication. arXiv preprint [arXiv:2008.11104](https://arxiv.org/abs/2008.11104) (2020)
4. Cabani, A., Hammoudi, K., Benhabiles, H., Melkemi, M.: MaskedFace-Net-a dataset of correctly/incorrectly masked face images in the context of COVID-19. Smart Health **19**, 100144 (2021)
5. Chen, Y.A., Chen, W.C., Wei, C.P., Wang, Y.C.F.: Occlusion-aware face inpainting via generative adversarial networks. In: IEEE ICIP, pp. 1202–1206. IEEE (2017)

6. Cheng, L., Wang, J., Gong, Y., Hou, Q.: Robust deep auto-encoder for occluded face recognition. In: ACM International Conference on Multimedia, pp. 1099–1102 (2015)
7. Daras, G., Dean, J., Jalal, A., Dimakis, A.G.: Intermediate layer optimization for inverse problems using deep generative models. arXiv preprint [arXiv:2102.07364](https://arxiv.org/abs/2102.07364) (2021)
8. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: IEEE CVPR, pp. 4690–4699 (2019)
9. Deng, Y., Dai, Q., Zhang, Z.: Graph Laplace for occluded face completion and recognition. IEEE TIP **20**(8), 2329–2338 (2011)
10. Din, N.U., Javed, K., Bae, S., Yi, J.: A novel GAN-based network for unmasking of masked face. IEEE Access **8**, 44276–44287 (2020)
11. Donahue, J., Krähenbühl, P., Darrell, T.: Adversarial feature learning. arXiv preprint [arXiv:1605.09782](https://arxiv.org/abs/1605.09782) (2016)
12. Fidler, S., Skocaj, D., Leonardis, A.: Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. IEEE PAMI **28**(3), 337–350 (2006)
13. Gao, R., Grauman, K.: From one-trick ponies to all-rounders: on-demand learning for image restoration. arXiv preprint [arXiv:1612.01380](https://arxiv.org/abs/1612.01380) (2016)
14. Ge, S., Li, C., Zhao, S., Zeng, D.: Occluded face recognition in the wild by identity-diversity inpainting. IEEE T-CSVT **30**(10), 3387–3397 (2020)
15. Ge, S., Li, J., Ye, Q., Luo, Z.: Detecting masked faces in the wild with LLE-CNNs. In: IEEE CVPR, pp. 2682–2690 (2017)
16. Ghosh, P., Zietlow, D., Black, M.J., Davis, L.S., Hu, X.: InvGAN: invertable GANs. arXiv preprint [arXiv:2112.04598](https://arxiv.org/abs/2112.04598) (2021)
17. He, Z., Zuo, W., Kan, M., Shan, S., Chen, X.: AttGan: facial attribute editing by only changing what you want. IEEE TIP **28**(11), 5464–5478 (2019)
18. Horé, A., Ziou, D.: Image quality metrics: PSNR vs. SSIM. In: ICPR, pp. 2366–2369 (2010)
19. Hu, B., Zheng, Z., Liu, P., Yang, W., Ren, M.: Unsupervised eyeglasses removal in the wild. IEEE Trans. Cybern. **51**(9), 4373–4385 (2020)
20. Huang, G.B., Mattar, M., Lee, H., Learned-Miller, E.: Learning to align from scratch. In: NIPS (2012)
21. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Trans. Graph. (ToG) **36**(4), 1–14 (2017)
22. Iliadis, M., Wang, H., Molina, R., Katsaggelos, A.K.: Robust and low-rank representation for fast face identification with occlusions. IEEE TIP **26**(5), 2203–2218 (2017)
23. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: IEEE CVPR, pp. 4401–4410 (2019)
24. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: IEEE CVPR, pp. 8110–8119 (2020)
25. Leonardis, A., Bischof, H.: Robust recognition using eigenimages. Comput. Vis. Image Underst. **78**(1), 99–118 (2000)
26. Li, C., Ge, S., Zhang, D., Li, J.: Look through masks: towards masked face recognition with de-occlusion distillation. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 3016–3024 (2020)
27. Li, Y., Liu, S., Yang, J., Yang, M.H.: Generative face completion. In: IEEE CVPR, pp. 3911–3919 (2017)

28. Li, Y., Feng, J.: Reconstruction based face occlusion elimination for recognition. *Neurocomputing* **101**, 68–72 (2013)
29. Li, Z., Hu, Y., He, R., Sun, Z.: Learning disentangling and fusing networks for face completion under structured occlusions. *Pattern Recogn.* **99**, 107073 (2020)
30. Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. In: *ECCV 2018*, pp. 85–100 (2018)
31. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: *Proceedings of the IEEE ICCV*, pp. 3730–3738 (2015)
32. Luan, X., Fang, B., Liu, L., Yang, W., Qian, J.: Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion. *Pattern Recogn.* **47**(2), 495–508 (2014)
33. Ma, X., Zhou, X., Huang, H., Jia, G., Chai, Z., Wei, X.: Contrastive attention network with dense field estimation for face completion. *Pattern Recogn.* **124**, 108465 (2022)
34. Menon, S., Damian, A., Hu, S., Ravi, N., Rudin, C.: Pulse: self-supervised photo upsampling via latent space exploration of generative models. In: *IEEE CVPR*, pp. 2437–2445 (2020)
35. Mishra, S., Majumdar, P., Dosi, M., Vatsa, M., Singh, R.: Dual sensor Indian masked face dataset. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pp. 1–8. IEEE (2021)
36. Ngan, M., Grother, P., Hanaoka, K.: Ongoing face recognition vendor test (FRVT) part 6a: face recognition accuracy with masks using pre-COVID-19 algorithms. Technical report, NIST (2020)
37. Nitzan, Y., Bermano, A., Li, Y., Cohen-Or, D.: Face identity disentanglement via latent space mapping. *arXiv preprint [arXiv:2005.07728](https://arxiv.org/abs/2005.07728)* (2020)
38. Park, J.S., Oh, Y.H., Ahn, S.C., Lee, S.W.: Glasses removal from facial image using recursive error compensation. *IEEE PAMI* **27**(5), 805–811 (2005)
39. Pernuš, M., Štruc, V., Dobrišek, S.: High resolution face editing with masked GAN latent code optimization. *arXiv preprint [arXiv:2103.11135](https://arxiv.org/abs/2103.11135)* (2021)
40. Pidhorskyi, S., Adjeroh, D.A., Doretto, G.: Adversarial latent autoencoders. In: *IEEE CVPR*, pp. 14104–14113 (2020)
41. Richardson, E., et al.: Encoding in style: a StyleGAN encoder for image-to-image translation. In: *IEEE CVPR*, pp. 2287–2296 (2021)
42. Roich, D., Mokady, R., Bermano, A.H., Cohen-Or, D.: Pivotal tuning for latent-based editing of real images. *arXiv preprint [arXiv:2106.05744](https://arxiv.org/abs/2106.05744)* (2021)
43. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: *IEEE CVPR*, pp. 815–823 (2015)
44. Shukor, M., Yao, X., Damodaran, B.B., Hellier, P.: Semantic and geometric unfolding of StyleGAN latent space. *arXiv preprint [arXiv:2107.04481](https://arxiv.org/abs/2107.04481)* (2021)
45. Su, Y., Yang, Y., Guo, Z., Yang, W.: Face recognition with occlusion. In: *IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 670–674. IEEE (2015)
46. Trigueros, D.S., Meng, L., Hartnett, M.: Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss. *Image Vis. Comput.* **79**, 99–108 (2018)
47. Wang, M., Hu, Z., Sun, Z., Zhao, S., Sun, M.: Varying face occlusion detection and iterative recovery for face recognition. *J. Electron. Imaging* **26**(3), 033009 (2017)
48. Wang, Z., et al.: Masked face recognition dataset and application. *arXiv preprint [arXiv:2003.09093](https://arxiv.org/abs/2003.09093)* (2020)
49. Wei, T., et al.: A simple baseline for StyleGAN inversion. *arXiv preprint [arXiv:2104.07661](https://arxiv.org/abs/2104.07661)* (2021)

50. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: *Advances in Neural Information Processing Systems*, vol. 25 (2012)
51. Xinyi, Z., Runqing, J., Tianxiang, H., Hao, Y.: Identity preserving face completion with landmark based generative adversarial network (2021)
52. Xiong, W., et al.: Foreground-aware image inpainting. In: *IEEE CVPR*, pp. 5840–5848 (2019)
53. Zeng, D., Veldhuis, R., Spreeuwiers, L.: A survey of face recognition techniques under occlusion. *IET Biometrics* **10**(6), 581–606 (2021)
54. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: *IEEE CVPR*, pp. 586–595 (2018)
55. Zhao, F., Feng, J., Zhao, J., Yang, W., Yan, S.: Robust LSTM-autoencoders for face de-occlusion in the wild. *IEEE TIP* **27**, 778–790 (2016)
56. Zhao, S., Hu, Z.P.: A modular weighted sparse representation based on fisher discriminant and sparse residual for face recognition with occlusion. *Inf. Proc. Lett.* **115**(9), 677–683 (2015)
57. Zhu, P., Abdal, R., Qin, Y., Femiani, J., Wonka, P.: Improved StyleGAN embedding: where are the good Latents? arXiv preprint [arXiv:2012.09036](https://arxiv.org/abs/2012.09036) (2020)