

# ATFaceGAN: Single Face Image Restoration and Recognition from Atmospheric Turbulence

Chun Pong Lau , Hossein Souri and Rama Chellappa  
 Center for Automation Research , UMIACS  
 University of Maryland, College Park  
 {cplau,hsouri,rama}@umiacs.umd.edu

**Abstract**—Image degradation due to atmospheric turbulence is common while capturing images at long ranges. To mitigate the degradation due to turbulence which includes deformation and blur, we propose a generative single frame restoration algorithm which disentangles the blur and deformation due to turbulence and reconstructs a restored image. The disentanglement is achieved by decomposing the distortion due to turbulence into blur and deformation components using deblur generator and deformation correction generator respectively. Two paths of restoration are implemented to regularize the disentanglement and generate two restored images from one degraded image. A fusion function combines the features of the restored images to reconstruct a sharp image with rich details. Adversarial and perceptual losses are added to reconstruct a sharp image and suppress the artifacts respectively. Extensive experiments demonstrate the effectiveness of the proposed restoration algorithm, which achieves satisfactory performance in face restoration and face recognition.

**Index Terms**—Image Restoration; Face Recognition; Face Verification; Turbulence Mitigation; Generative Adversarial Networks

## I. INTRODUCTION

Capturing images at long ranges is always challenging as the degradation due to atmospheric turbulence is inevitable. Under the effects of the turbulent flow of air and changes in temperature, density of air particles, humidity and carbon dioxide level, the captured image is blurry and deformed due to variations in the refractive index [1], [2]. This will significantly degrade the quality of images and performances of many computer vision tasks such as object detection [3], recognition and tracking [4]. To suppress these effects, two classical approaches have been considered, one based on adaptive optics [5], [6] and the other based on image processing [7], [8], [9], [10], [11], [12], [13], [14], [15]. However, these methods require multiple image frames captured by a static imager. Mathematically, [9], [8], [13] the process of image degradation due to atmospheric turbulence can be represented as

$$\tilde{I}_k = D_k(H_k(I)) + n_k, \quad (1)$$

where  $\tilde{I}_k$  is the observed distorted images,  $I$  is the latent clear image,  $H_k$  is a space-invariant point spread function (PSF),  $D_k$  is the deformation operator, which is assumed to deform randomly and  $n_k$  is the sensor noise.

Recently, many learning-based face restoration algorithms such as face deblurring [16], [17], [18] and face superresolution [19], [20], [21] have been proposed. Moreover, the emer-

gence of Generative Adversarial Networks (GAN) has further improved the quality of reconstructed images. However, these methods have not tackled the problem of deformation, which greatly reduces the quality of the acquired images and the performance of many computer vision tasks.

Recently, [15] proposed a generative method to restore a clean image from multiple frames using a Wasserstein GAN [22] and a subsampled frames algorithm proposed by [13]. However, the method assumes a multi-frame setting with a static object. This assumption may not be practical in real life situations.

Motivated by the recent success of data-driven approach, we propose a generative single face image restoration algorithm, namely **Atmospheric Turbulence Face GAN (ATFaceGan)**, which reconstructs a clean face image with texture details preserved by simultaneously disentangling blur and deformation. We build two generators, namely, deblur function and deformation correction function to decompose the degradations in turbulence. Also, we propose a two path training approach to further disentangle the degradation and reconstruct two images. A fusion function is used to combine the information in the two restored images and reconstruct a sharp face image. Some sample restored images are shown in Fig. 1.

Our contributions are summarized below:

- 1) The proposed method tackles the atmospheric turbulence degradation problem with a single image input.
- 2) We propose a generative face restoration algorithm trained in an end-to-end manner, which tackles degradation due to both blur and deformation by building the deblur function and deformation correction function respectively.
- 3) We propose a two path training strategy to further disentangle the blur and deformation and improve the quality of the restored image.
- 4) We propose a fusion network to combine the latent features of the intermediate results and reconstruct one clean restored image.
- 5) Experiments demonstrate that the restored face image is satisfactory in both quantitative and visual assessment. Further, the restored face images yields improved recognition performance.



Fig. 1: Restored images of ATFaceGAN from the publicly available face dataset [23]. Rows 1 and 3 present the synthetic atmospheric turbulence degraded images. Rows 2 and 4 present the corresponding restored images obtained by the proposed algorithm.

## II. RELATED WORK

**Turbulence Degraded Image Restoration** Classical methods of restoring images degraded by turbulence generally include two approaches. One is "lucky imaging" [24], [25], which chooses a frame or a number of good frames in a turbulence degraded video and fuses the selected frames. Another one is the registration-fusion approach [8], [9], [26], [14], which first constructs a good reference image and aligns the distorted frames with the reference image using a non-rigid image registration algorithm. After alignment, the registered images are fused following which a restoration algorithm is applied to deblur the fused image to obtain the final restored image. Despite having satisfactory results, these methods assume *multi-frame inputs with static objects*. This assumption is violated easily in reality, for example, when pedestrians are moving in long range surveillance videos.

**Face Restoration** Due to recent successes of CNNs and GANs, several CNN-based face restoration algorithms have been proposed. [16] proposed a CNN with Residual Blocks to deblur face images. [17] proposed a multi-scale CNN that exploits global semantic priors and local structural constraints for face image deblurring as a generator and built a discriminator based on DCGAN [27]. [18] proposed an unsupervised method for domain-specific single image deblurring by disentangling the content information and blur information using the KL divergence constraint and improves the performance of face recognition. However, since *degradation due to turbulence contains motion blur, out-of-focus blur or compression arti-*

*facts*, these methods could not obtain satisfactory results.

## III. PROPOSED ALGORITHM

The proposed face image restoration algorithm is trainable in an end-to-end manner. Our goal is to reconstruct a sharp face image from the distorted face image and enhance the performance of face recognition systems.

### A. Problem Setting

Following the formulation of the degradation model discussed in [14], [9], [13], we assure the mathematical model in (1). This is the general setting for restoring the latent clean image from a sequence of turbulence-degraded image frames. However, we assume only one frame is available to reconstruct the latent clean image, a more challenging and practical problem than considered earlier. As a result, the subscript  $k$  is removed. Also, we notice that the "mixing" of deformation and blur in realistic turbulence face images is very fast and we could not be sure whether deformation precedes blur or blur precedes deformation. Therefore, we use a general turbulence function  $T$  to replace  $D \circ G$  in (1). Hence, our model becomes

$$\tilde{I} = T(I) + n. \quad (2)$$

Let  $\mathcal{I}_b = \{I_b := H(I) | I \in \mathcal{I}\}$  and  $\mathcal{I}_d = \{I_d := D(I) | I \in \mathcal{I}\}$  be the space of blurry and deformed face images respectively. Our goal is to construct a restoration function  $G$  to restore the distorted face images, i.e.  $G(\tilde{I}) = I$ . However, it is a highly ill-posed problem as we have very little prior information to

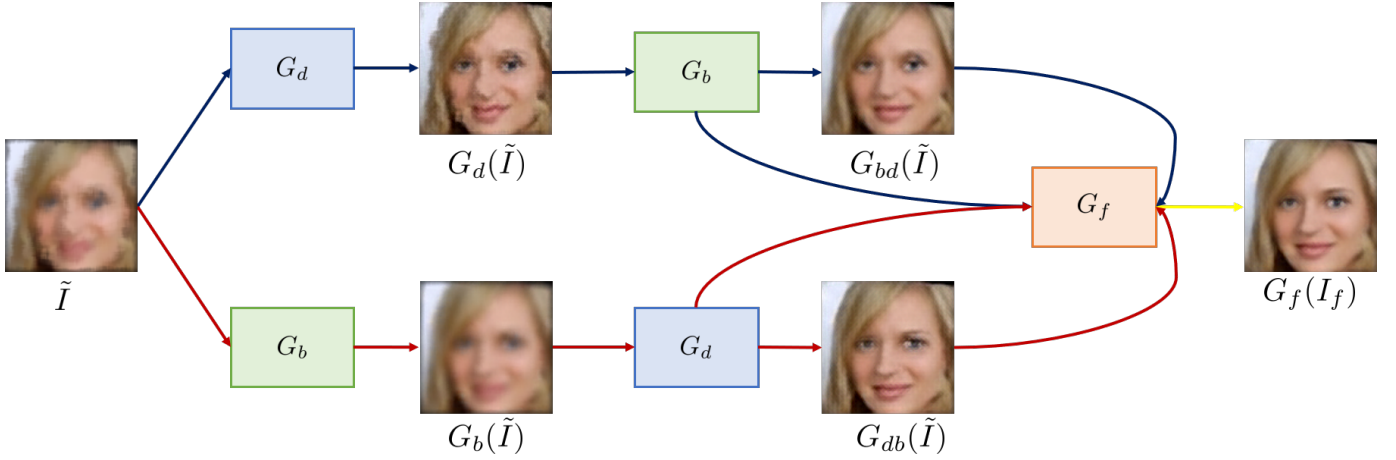


Fig. 2: Overview of the proposed restoration algorithm. Given a distorted image  $\tilde{I}$ , it is first passed through the deformation correction function  $G_d$  to get it aligned and deblurred with deblur function  $G_b$  and generate  $G_{bd}(\tilde{I})$ . It is then passed through the bottom branch (red), gets deblurred by  $G_d$  and then aligned by  $G_b$  to obtain  $G_{db}(\tilde{I})$ . Features of  $G_{bd}(\tilde{I})$  and  $G_{db}(\tilde{I})$  are concatenated and passed through the fusion function  $G_f$  to obtain the final restored image  $G_f(I_f)$ .

reconstruct  $\tilde{I}$ . Hence, a data-driven approach, in particular the Wasserstein GAN with gradient penalty, is applied to restore it. Moreover, blur and deformation are always combined in the turbulence-degraded face images. We hope to build a deblur function  $G_d$  and a deformation correction function  $G_b$  to remove the undesired blur and deformation, i.e.  $G_d(\tilde{I}) = I_d$  and  $G_b(\tilde{I}) = I_b$ . Therefore, we split the turbulence degradation due to blur and deformation in the training stage. In order to restore a general turbulence function  $T$  which contains both the blurring operator  $H$  and the deformation operator  $D$ , we propose a two path training approach, which tries to obtain more information to obtain a better result. Therefore, two restored images are obtained, i.e.  $G_d(G_b(\tilde{I}))$  and  $G_b(G_d(\tilde{I}))$ . A fusion network is implemented to improve the restoration results. Denote  $G_{bd} = G_b \circ G_d$ ,  $G_{db} = G_d \circ G_b$  and  $F(I)$  be the features of image  $I$ . Mathematically,

$$G(\tilde{I}) = G_f(I_f), \quad (3)$$

where  $G_f$  is a image fusion function and  $I_f$  is the feature pairs  $(F(G_{db}(\tilde{I})), F(G_{bd}(\tilde{I})))$ . The end-to-end architecture is illustrated in Fig. 2.

### B. Data Augmentation

In order to apply a data-driven method to restore a clean face image from distorted faces, sufficient amount of synthetic training data are needed. Therefore, the blur operator  $H$  and the deformation operator  $D$  are required to synthesize the distorted images. In this paper, we use the turbulence generation algorithm from [13], [14], [15] due to its efficiency in choosing different parameters to generate turbulence-degraded images with various severity.

We follow the procedure discussed in [13], [14], [15] to generate a random motion vector field to deform the face images.  $M$  points are selected in a face image  $I$ . For each point  $(x, y)$ , a  $N \times N$  patch  $P_{x,y}^N$  centered at  $(x, y)$  is

considered. A random motion vector field  $V_{x,y}$  is obtained in  $P_{x,y}^N$ . Mathematically,

$$V_{x,y} = \eta(G_\sigma * \mathcal{N}_1, G_\sigma * \mathcal{N}_2), \quad (4)$$

where  $G_\sigma$  is the Gaussian kernel with standard deviation  $\sigma$ ,  $\eta$  is the strength value,  $\mathcal{N}_1$  and  $\mathcal{N}_2$  are randomly selected from a Gaussian distribution. The overall motion vector field is generated after  $M$  iterations as follows:

$$V = \sum_{i=1}^M V_{(x,y)_i} \quad (5)$$

Then this motion vector field would be our deformation operator  $D$  defined as

$$D(I) = I \boxplus V, \quad (6)$$

where  $\boxplus$  is the warping operator. The blurring operator  $H$  is simply a Gaussian kernel. For more details, please see [13], [14], [15].

In order to construct the deblur function  $G_d$  and the deformation correction function  $G_b$ , we need to generate a blurry image  $I_b$ , a deformed image  $I_d$  and a distorted images  $\tilde{I}$  from each clean face image  $I$ . To generate  $I_b$ , Gaussian blurring filter with parameter  $\tau$  is applied on  $I$  to get  $I_b$ . To obtain  $I_d$ , the random motion vector field with strength  $\eta$  is applied on  $I$ .

### C. Network Architecture

A Wasserstein GAN with gradient penalty is applied to restore the distorted face images. The generator architecture is a CNN, similar to [28] used for image deblurring. It contains two strided convolution blocks with stride 12, six residual blocks [29] (ResBlocks) and two transposed convolution blocks. There are one convolution layer, instance normalization layer [30], ReLU activation [31] and a Dropout layer with  $p = 0.5$  in each ResBlock. A global skip connection

mentioned in [28] is also added. The deblur function  $G_d$  and deformation correction function  $G_b$  are included in both paths of this architecture. The fusion network  $G_f$  takes the concatenation of the features from face images  $G_{bd}\tilde{I}$  and  $G_{db}(\tilde{I})$  as inputs. The features are extracted after the activation function of the third ResBlock in  $G_b$  and  $G_d$ . The architecture of the fusion network  $G_f$  is exactly the latter half of the structure of  $G_b$  and  $G_d$ , which contains three ResBlock and two transposed convolution blocks. Since the input is the concatenation of two feature vectors, the number of channels in the ResBlocks of  $G_f$  is doubled. In order to keep the global skip connection, which has been shown to converge faster, pixel-wise average of  $G_{bd}\tilde{I}$  and  $G_{db}(\tilde{I})$  is added to  $G_f(I_f)$ . During training, three discriminators, namely  $D_b$ ,  $D_d$  and  $D_f$ , are designed.  $D_b$ ,  $D_d$  and  $D_f$  determine whether  $G_b(\tilde{I})$  and  $G_d(\tilde{I})$  and  $G_f(\tilde{I})$  are real or fake. The discriminators are Wasserstein GAN [22] with gradient penalty [32] (WGAN-GP). Their architectures are same as PatchGAN [33], [34]. All the convolutional layers except the last are followed by InstanceNorm layer and LeakyReLU [35].

#### D. Disentanglement of Blur and Deformation

In order to disentangle the turbulence distortion into blur and deformation, the deblur function  $G_d$  and the deformation correction function  $G_b$  are built. The content loss  $\mathcal{L}_{con}$  is defined as

$$\mathcal{L}_{con} = \|G_b(\tilde{I}) - I_b\|_1 + \|G_d(\tilde{I}) - I_d\|_1, \quad (7)$$

which is the sum of the  $L_1$  loss between aligned image  $G_b(\tilde{I})$  and  $I_b$  and the  $L_1$  loss between deblurred image  $G_d(\tilde{I})$  and  $I_d$ .

#### E. Two path training

The two path training strategy helps to disentangle the blur and deformation effects. One fixed order of restoration is needed if two path training is not implemented. For example, the distorted image is restored by  $G_b$  and followed by  $G_d$  according to (2). Then during the training phase,  $G_b$  is trained with the turbulence degraded images which are both blurry and deformed. In other words, the training images for  $G_b$  are implicitly assumed to be both blurry and deformed but not merely deformed. Therefore, if two path training is used, then  $G_b$  could learn from turbulence degraded images  $\tilde{I}$  and the deblurred images  $G_d(\tilde{I})$ .

Moreover, the search space of the optimization problem is larger because no implicit structure of degradation is assumed. As the turbulence function  $T$  only consists of blur and deformation but not the order of degradation, this gives more information ( $G_{bd}\tilde{I}$  and  $G_{db}(\tilde{I})$ ) to the network and improve the performance.

#### F. Fusion Loss

After both restored images  $G_{bd}(\tilde{I})$  and  $G_{db}(\tilde{I})$  are obtained, their features are fused together to obtain the final restored image. The fusion loss is defined as the  $L_1$  loss of the restored image and the real clean image  $I$ , i.e.

$$\mathcal{L}_f = \|G_f(I_f) - I\|_1. \quad (8)$$

#### G. Adversarial Loss

The Wasserstein-1 distance in WGAN has been shown to have good convergence property and is more stable in training given that the function is 1-Lipschitz. To enforce the 1-Lipschitz constraint, gradient penalty is applied. Then the discriminator and generator losses are defined as

$$\mathcal{L}_{Dis}^{\mathcal{I}_i} = \mathbb{E}_{\tilde{I} \sim \tilde{\mathcal{I}}} [D_i(G_i(\tilde{I}))] - \mathbb{E}_{I_i \sim \mathcal{I}_i} [D_i(I_i)] + \lambda_{WGAN} \cdot \mathbb{E}_{\hat{I}_i \sim \hat{\mathcal{I}}_i} [(\|\nabla_{\hat{I}_i} D_i(\hat{I}_i)\|_2 - 1)^2], \quad (9)$$

$$\mathcal{L}_{Dis}^{\mathcal{I}_f} = \mathbb{E}_{I_f \sim \mathcal{I}_f} [D_f(G_f(I_f))] - \mathbb{E}_{I \sim \mathcal{I}} [D_f(I)] + \lambda_{WGAN} \cdot \mathbb{E}_{\hat{I}_f \sim \hat{\mathcal{I}}_f} [(\|\nabla_{\hat{I}_f} D_f(\hat{I}_f)\|_2 - 1)^2], \quad (10)$$

$$\mathcal{L}_{Gen}^{\mathcal{I}_i} = -\mathbb{E}_{I_i \sim \mathcal{I}_i} [D_i(G_i(\tilde{I}))] \quad (11)$$

$$\mathcal{L}_{Gen}^{\mathcal{I}_f} = -\mathbb{E}_{I_f \sim \mathcal{I}_f} [D_f(G_f(I_f))] \quad (12)$$

where  $\hat{\mathcal{I}}_i$  is the distribution obtained by randomly interpolating between real images  $I_i$  and restored images  $G_i(\tilde{I})$ ,  $i \in \{b, d\}$  and  $j \in \{b, d, f\}$ . The adversarial loss is

$$\mathcal{L}_{adv} = \sum_{j \in \{b, d, f\}} \mathcal{L}_{Dis}^{\mathcal{I}_j} + \mathcal{L}_{Gen}^{\mathcal{I}_j} \quad (13)$$

#### H. Perceptual Loss

Using  $L_2$  loss or  $L_1$  loss merely as the content loss would lead to blurry artifacts and loss in texture details as these losses favor pixelwise averaging. On the other hand, Perceptual Loss, which is an  $L_2$  loss function between the feature maps of real image and generated image, has been demonstrated to be beneficial for image restoration tasks [17], [28], [18]. Therefore, perceptual loss is adopted, which includes

$$\mathcal{L}_p^{\mathcal{I}_i} = \|\phi_l(G_i(\tilde{I})) - \phi_l(I_i)\|_2^2, \quad i \in \{b, d\} \quad (14)$$

$$\mathcal{L}_p^{\mathcal{I}_f} = \|\phi_l(G_f(I_f)) - \phi_l(I)\|_2^2 \quad (15)$$

where  $\phi_l(\cdot)$  is the features of the  $l^{\text{th}}$  layer of a pretrained CNN. In this paper, the `conv3_3` layer of VGG-19 [36] network pretrained on ImageNet [37] is adopted. The total perceptual loss is

$$\mathcal{L}_p = \sum_{i \in \{b, d\}} \mathcal{L}_p^{\mathcal{I}_i} + \mathcal{L}_p^{\mathcal{I}_f} = \mathcal{L}_p^{b, \mathcal{I}_b} + \mathcal{L}_p^{\mathcal{I}_f} \quad (16)$$

The full loss function is a weighted sum of all the losses,

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_{con} \mathcal{L}_{con} + \lambda_f \mathcal{L}_f + \lambda_p^{b, d} \mathcal{L}_p^{b, \mathcal{I}_b} + \lambda_p^f \mathcal{L}_p^{\mathcal{I}_f} \quad (17)$$

The weights are empirically set for each loss to balance their importance.

#### I. Testing

At test time, only the generators are used. Given a turbulence distorted image  $\tilde{I}$ , the restored image is generated as follows:

$$I^r = G_f(F(G_{bd}(\tilde{I})), F(G_{db}(\tilde{I}))). \quad (18)$$



TABLE I: Ablation study tested with LFW dataset

Method	Degraded images	One generator	Decompose into two generators	Add two path training	Add fusion function
PSNR	24.17	25.99	25.90	26.16	<b>27.29</b>
SSIM	0.878	0.901	0.897	0.902	<b>0.924</b>

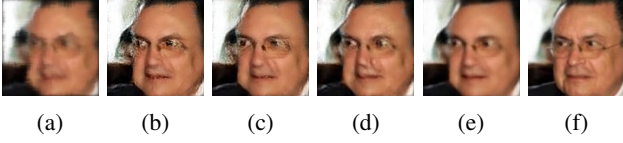


Fig. 3: Ablation study. (a) is the distorted image and (f) is the sharp image. (b) only contains one generator. (c) is split into  $G_d$  and  $G_b$ . (d) adds two path training and (e) adds fusion network

#### IV. EXPERIMENTS

Our algorithm is trained on [23] and evaluated on six face recognition datasets, including LFW [38], CFP [39], AgeDB [40], CALFW [41], CPLFW [42] and VGGFace2 [43].

##### A. Training details

The end-to-end design is implemented in Pytorch [44]. The training was performed on two GeForce RTX 2080 Ti GPU. In training, 10000 aligned face images are randomly picked, which are with resolution  $112 \times 112$  from [23] with the turbulence degradation algorithm in Sec(III-B) and a batch size of 16. During training, we use the Adam solver [45] with hyper-parameters  $\beta_1 = 0.9, \beta_2 = 0.999$  to perform five steps of update on discriminators and then one step on generators. The learning rate is initially set at 0.0001 for the first 30 epochs, then linear decay is applied for the next 20 epochs. For hyper-parameters in deformation operator  $D$ , we empirically set  $\eta = 0.13, N = 4, \sigma = 16$  and  $M = [1000, 3000, 7000, 10000]$ . For hyper-parameters in blurring operator  $H$ , the parameter  $\mu$  is set to be  $[1, 2, 3, 4]$ . For hyper-parameters in the loss function, we empirically set  $\lambda_{con} = \lambda_f = 1000, \lambda_p^{b,d} = 10$  and  $\lambda_p^f = 1$ . Note that various parameters in  $M$  and  $\tau$  are randomly picked to synthesize various strength of blur and deformation. The computation time of restoring a  $112 \times 112$  image is 0.031 seconds per image on average.

##### B. Testing details

In all the six testing dataset, all the pairs of the face images are degraded by the algorithm from [13]. PSNR and SSIM are used for evaluating the quality of the restored image. We use a pretrained face recognition network [46], which is trained as reported in [47], to test the face verification performance<sup>1</sup>.

##### C. Ablation study

In this section, the results of an ablation study preformed to analyze the effectiveness of each component or loss in the proposed algorithm are presented. Both quantitative and

qualitative results on face dataset in [23] are evaluated for the following four variants of our methods where each component is gradually added: 1) only one generator and one discriminator; 2) splitting the generators into two,  $G_b$  and  $G_d$ , and the restored image is  $G_b(G_d(\tilde{I}))$ ; 3) Applying two path training and the restored image is  $G_b(G_d(\tilde{I}))$  and 4) adding fusion network and fuse them by  $G_f$

We present the PSNR and SSIM for each variant in Table I and visual comparisons in Fig. 3. From Fig. 3, we observe that the resultant images with direct restoration, which only uses one generator, is not satisfactory. This is because turbulence degradation is a very ill-posed problem. There is a large gap between turbulence-degraded and clean image and one generator could not provide enough information to the network. By decomposing the network into two generators, the quantitative performance is similar to one generator but it is less noisy. This is because we have more information for the generators to learn as the intermediate results ( $G_d(\tilde{I})$ ) provides additional supervision to the final restored image. When we apply the two path training step and as both  $I_b$  and  $I_d$  are added to supervise the training, the results are good even groundtruth  $I$  is not used in the training. Adding the fusion network further improves the result as more information (features of  $G_{bd}(\tilde{I})$  and  $G_{ab}(\tilde{I})$ ) is given to the network and the information is combined by the fusion function  $G_f$ . Table I also justifies the result.

##### D. Qualitative and quantitative Evaluation

Since the proposed algorithm is the first single frame-based image restoration method with turbulence-degraded images, which involve blur and deformation, it is hard to compare with other methods. Therefore, we compare with some state-of-the-art image restoration methods including [28], [18], which could train with our turbulence-degraded image dataset. These two methods are the representative methods for applying GAN in deblurring in supervised and unsupervised ways respectively. For [28], we change the batch size from 1 to 16 and the number of training epoch to 100. For [18], we use the default setting.

The quantitative results are shown in Table II and the visual comparison are illustrated in Fig. 4. In Fig. 4, we have demonstrated three images: one from LFW, one from CFF and one from AGEDB. The top one is a frontal image with mild blur and mild deformation, the middle one is a frontal image with moderate blur and severe deformation and the bottom one is a non-frontal gray-scale face image with severe blur and mild deformation. For the top image, we can see that blur is suppressed in all three methods. [28] and [18] shows sharper visual result then ours. However, the result from [28] is noisy and that from [18] is deformed. The proposed method

<sup>1</sup>Please refer to the corresponding project page for the face verification policy: <https://github.com/ZhaoJ9014/face.evoLve.PyTorch>



Fig. 4: Visual performance comparison with state-of-the-art methods. (a) is the distorted image. (b) [28]. (c) [18]. (d) Ours. (e) Groundtruth.

TABLE II: Quantitative performance comparison with state-of-the-art methods

Dataset	Sharp	Distorted			[28]			[18]			Ours		
	Acc	PSNR	SSIM	Acc	PSNR	SSIM	Acc	PSNR	SSIM	Acc	PSNR	SSIM	Acc
AgeDB [40]	0.981	22.71	0.769	0.819	24.04	0.781	0.830	21.22	0.657	0.750	<b>25.24</b>	<b>0.838</b>	<b>0.835</b>
CALFW [41]	0.959	22.85	0.831	0.842	24.43	0.843	0.844	21.97	0.771	0.780	<b>25.78</b>	<b>0.890</b>	<b>0.857</b>
CFP_FF [39]	0.997	22.13	0.830	0.892	22.88	0.833	0.916	21.40	0.683	0.861	<b>24.37</b>	<b>0.889</b>	<b>0.922</b>
CFP_FP [39]	0.981	22.94	0.850	0.799	23.84	0.850	0.812	21.83	0.629	0.743	<b>25.20</b>	<b>0.901</b>	<b>0.815</b>
CPLFW [42]	0.925	24.21	0.875	0.787	26.11	0.882	0.797	22.68	0.787	0.732	<b>27.29</b>	<b>0.919</b>	<b>0.800</b>
LFW [38]	0.998	24.17	0.878	0.936	25.80	0.884	<b>0.951</b>	23.10	0.824	0.896	<b>27.29</b>	<b>0.924</b>	0.946
VggFace2 [43]	0.952	23.44	0.849	0.837	24.99	0.856	0.853	22.06	0.774	0.784	<b>26.16</b>	<b>0.896</b>	<b>0.854</b>

restores the image effectively. On the other hand, if both blur and deformation exist, [28] would induce more noise as shown in Fig. 4 (b) and [18] could not remove the deformation as shown in Fig. 4 (c). The proposed method suppresses both blur and deformation. Moreover, as our training set only consists of 10000 images, which include both colored and grey-scale images, the quantitative results generated by [18] are not good compared to [28] and the proposed method as the number of training sample is not large enough. The proposed method trained with a relatively small training set is effective in the presence of severe blur, deformation and pose. The PSNR and SSIM in Table II both demonstrate that the proposed method performs better than state-of-the-art methods.

For the face verification task, we note that [28] is slightly better than the proposed method in one out of seven experiments even though both the visual quality and quantitative results of the proposed method is better than [28]. Except LFW, the proposed method is more accurate than the other two methods. The verification accuracy of [28] is comparable with the proposed method. It is because [28] uses only perceptual loss as their content loss. As a result, the restored image from [28] is perceptually similar than the proposed method. Using the  $L_2$  distance from two feature output from  $\text{conv}_{3,3}$  layer of VGG-19 [36] network as a perceptual metric, namely  $d_{VGG}$ , we found that the  $d_{VGG}$  between restored image by [28] and the original clean image is 110.82 in LFW while the

$d_{VGG}$  between the restored image by the proposed method and original clean image is 118.55.

Atmospheric turbulence degradation severely harms the task of face verification as the verification accuracy is reduced by more than 10% on average. There could be a significant drop (as much as 20% for CFP\_FP) even though the face verification system is trained with [47], which consists of over 5 million images. Also, as the task of restoration from turbulence is very challenging, the restoration results from other state-of-the-art method do not yield satisfactory results even they are trained with our dataset. Moreover, the proposed method restores the turbulence degraded images effectively even with a relatively small dataset.

#### E. Performance of the disentangled representation

We try to disentangle the blur and deformation from atmospheric turbulence by training the deblur function  $G_d$  and the deformation correction function  $G_b$  with a commutative constraint. To see the performance of the disentanglement,  $G_b$  and  $G_d$  are tested. We try to use  $G_d$  to deblur the blurry image and  $G_b$  to correct the deformed images. Note that during the training,  $G_b$  is only fed with the distorted image  $\tilde{I}$  and the deformation corrected image  $G_d(\tilde{I})$  of the distorted image.

We test  $G_b$  and  $G_d$  with  $I_d$  and  $I_b$  respectively, where  $I$  are from the LFW dataset. The PSNR, SSIM and the accuracy of face verification are presented in Table III. The

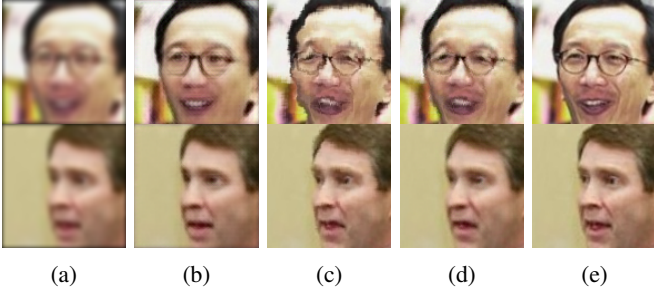


Fig. 5: Visual performance comparison of the deblur function  $G_d$  and deformation correction  $G_b$  with the LFW dataset. (a) Blurry image. (b) Restored image of (a) by  $G_d$ . (c) Deformed image. (d) Restored image of (c) by  $G_b$ . (e) Groundtruth.



Fig. 6: Visual performance comparison of original face image and the resultant image that passing through the pipeline with the original face image. Left: Original. Right: Ours.

visual performance is shown in Fig. 5. For the first row, the image is moderately blurred (Fig. 5 (a)) and severely deformed (Fig. 5 (c)). From Figs. 5 (b) and (d), we see that  $G_d$  and  $G_b$  successfully remove the blur and deformation from the image and preserve the features of the subject. On the other hand, note that the image in the second row is a profile face with moderate blur (Fig. 5 (a)) and mild distortion (Fig. 5 (c)). Still,  $G_d$  and  $G_b$  successfully restore the degraded images. Moreover, the PSNR, SSIM and face verification results confirm that  $G_d$  and  $G_b$  restore the images, preserve shape and semantic information and are robust to severity of blur, deformation and pose.

#### F. Robustness to clean images input

We tested the case of feeding the groundtruth clean image into the proposed method. We tested 5222 images from the IJB-C dataset [48]. The PSNR and SSIM of the resultant images after passing the original clean image to the pipeline are 33.89 and 0.980 respectively. The visual comparison result is shown in Fig. 6. Both qualitative and quantitative results show that the proposed method is robust to non-blurred images.

#### G. Failure cases

We have shown some failure cases in Fig. 7. The proposed method produces over-smoothed images if the turbulence

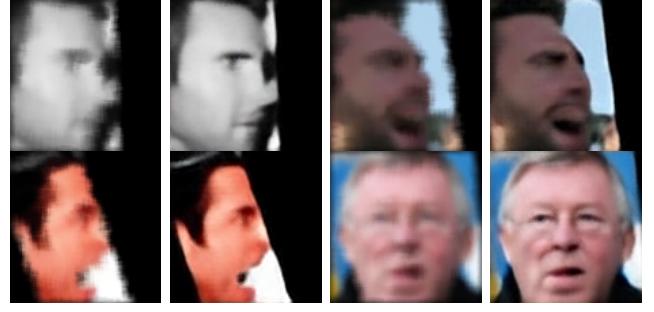


Fig. 7: Failure cases. Left: AT. Right: Restored.

TABLE III: PSNR, SSIM and face verification results for LFW dataset with  $G_b$  and  $G_d$ .

	$I_b$	$I_d$	$G_d(I_b)$	$G_b(I_d)$
PSNR	25.33	29.78	<b>28.72</b>	<b>29.93</b>
SSIM	0.895	0.958	<b>0.931</b>	<b>0.961</b>
Accuracy	0.793	0.649	<b>0.817</b>	<b>0.809</b>

degradation is very strong.

## V. CONCLUSION

In this paper, we proposed a single frame image restoration method ATFaceGAN, which is a generative algorithm to disentangle the turbulence distortion into blur and deformation and restores a sharp image. In order to disentangle turbulence, a deblur generator and a deformation correction generator are introduced. To further separate the blur and deformation, two paths are employed to produce two restored images. Finally, a fusion function combines the two restored images and generates one clean image. Ablation studies on each component demonstrate the effectiveness of different components. We have conducted extensive experiments on face restoration and face verification using the restored face images. Both quantitative and visual results show promising performance.

## ACKNOWLEDGMENT

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2019-022600002. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

## REFERENCES

- [1] R. Hufnagel and N. Stanley, "Modulation transfer function associated with image transmission through turbulent media," *JOSA*, vol. 54, no. 1, pp. 52–61, 1964.
- [2] M. C. Roggemann, B. M. Welsh, and B. R. Hunt, *Imaging through turbulence*. CRC press, 1996.
- [3] O. Oreifej, X. Li, and M. Shah, "Simultaneous video stabilization and moving object detection in turbulence," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 2, pp. 450–462, 2013.



- [4] E. Chen, O. Haik, and Y. Yitzhaky, "Detecting and tracking moving objects in long-distance imaging through turbulent medium," *Applied optics*, vol. 53, no. 6, pp. 1181–1190, 2014.
- [5] J. E. Pearson, "Atmospheric turbulence compensation using coherent optical adaptive techniques," *Applied optics*, vol. 15, no. 3, pp. 622–631, 1976.
- [6] R. K. Tyson, *Principles of adaptive optics*. CRC press, 2015.
- [7] M. H. Furhad, M. Tahtali, and A. Lambert, "Restoring atmospheric-turbulence-degraded images," *Applied optics*, vol. 55, no. 19, pp. 5082–5090, 2016.
- [8] M. Hirsch, S. Sra, B. Schölkopf, and S. Harmeling, "Efficient filter flow for space-variant multiframe blind deconvolution," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 607–614.
- [9] X. Zhu and P. Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 157–170, 2013.
- [10] M. Micheli, Y. Lou, S. Soatto, and A. L. Bertozzi, "A linear systems approach to imaging through turbulence," *Journal of mathematical imaging and vision*, vol. 48, no. 1, pp. 185–201, 2014.
- [11] E. Meinhardt-Llopis and M. Micheli, "Implementation of the centroid method for the correction of turbulence," *Image Processing On Line*, vol. 4, pp. 187–195, 2014.
- [12] Y. Lou, S. H. Kang, S. Soatto, and A. L. Bertozzi, "Video stabilization of atmospheric turbulence distortion," *Inverse Problems & Imaging*, vol. 7, no. 3, 2013.
- [13] C. P. Lau, Y. H. Lai, and L. M. Lui, "Variational models for joint subsampling and reconstruction of turbulence-degraded images," *Journal of Scientific Computing*, vol. 78, no. 3, pp. 1488–1525, 2019.
- [14] C. P. Lau, Y. H. Lai, and R. L. M. Lui, "Restoration of atmospheric turbulence-distorted images via rpca and quasiconformal maps," *Inverse Problems*, 2019.
- [15] W. H. Chak, C. P. Lau, and L. M. Lui, "Subsampled turbulence removal network," *arXiv preprint arXiv:1807.04418*, 2018.
- [16] G. G. Chrysos and S. Zafeiriou, "Deep face deblurring," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
- [17] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8260–8269.
- [18] B. Lu, J.-C. Chen, and R. Chellappa, "Unsupervised domain-specific deblurring via disentangled representations," *arXiv preprint arXiv:1903.01594*, 2019.
- [19] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "Fsnet: End-to-end learning face super-resolution with facial priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2492–2501.
- [20] X. Yu, B. Fernando, R. Hartley, and F. Porikli, "Super-resolving very low-resolution face images with supplementary attributes," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [21] X. Yu, B. Fernando, B. Ghanem, F. Porikli, and R. Hartley, "Face super-resolution guided by facial component heatmaps," in *The European Conference on Computer Vision (ECCV)*, September 2018.
- [22] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.
- [23] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, "Umdfaces: An annotated face dataset for training deep networks," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2017, pp. 464–473.
- [24] M. Aubailly, M. A. Vorontsov, G. W. Carhart, and M. T. Valley, "Automated video enhancement from a stream of atmospherically-distorted images: the lucky-region fusion approach," in *Proc. SPIE*, vol. 7463, 2009, p. 74630C.
- [25] M. A. Vorontsov and G. W. Carhart, "Anisoplanatic imaging through turbulent media: image recovery by local information fusion from a set of short-exposure images," *JOSA A*, vol. 18, no. 6, pp. 1312–1324, 2001.
- [26] Y. Xie, W. Zhang, D. Tao, W. Hu, Y. Qu, and H. Wang, "Removing turbulence effect via hybrid total variation and deformation-guided kernel regression," *IEEE Transactions on Image Processing*, vol. 25, no. 10, pp. 4943–4958, 2016.
- [27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [28] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8183–8192.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [30] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.
- [31] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
- [32] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems*, 2017, pp. 5767–5777.
- [33] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [34] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 702–716.
- [35] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.
- [36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [38] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*, 2008.
- [39] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–9.
- [40] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "Agedb: the first manually collected, in-the-wild age database," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 51–59.
- [41] T. Zheng, W. Deng, and J. Hu, "Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments," *arXiv preprint arXiv:1708.08197*, 2017.
- [42] T. Zheng and W. Deng, "Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments," *Beijing University of Posts and Telecommunications, Tech. Rep.*, pp. 18–01, 2018.
- [43] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 67–74.
- [44] 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," 2017.
- [45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [46] Y. Xu, Y. Cheng, J. Zhao, Z. Wang, L. Xiong, K. Jayashree, H. Tamura, T. Kagaya, S. Shen, S. Pranata *et al.*, "High performance large scale face recognition with multi-cognition softmax and feature retrieval," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1898–1906.
- [47] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European Conference on Computer Vision*. Springer, 2016, pp. 87–102.
- [48] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney *et al.*, "Iarpa janus benchmark-c: Face dataset and protocol," in *2018 International Conference on Biometrics (ICB)*. IEEE, 2018, pp. 158–165.