THERMAL TO

Dept. of E

2022 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-9620-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIP46576.2022.9897975

A

In many practical applicati
metrics and surveillance, t
used to capture images in lo
ever, such imaging system
lence, which introduces sev
captured images. Such an i
ing and significantly decrea
paper, we first investigate th
method on real-world therm
method is then proposed w
ages into visible-spectrum i
based on a pre-trained Sty
existing two-steps methods
thermal to visible image tra
be effective in terms of bot
results and face verification
knowledge, this is the first v
to visible image translation

Index Terms— Atmos
visible image translation, D

1. IN

In many applications of lo
we are faced with a scenari
tity of a person appearing in
spheric turbulence. One wa
methods that can remove the
ever, restoring images degr
cult since it causes images
Such a scenario becomes m
tured from thermal modality
time conditions. In this wor
to synthesize visible images
are degraded by atmospheri

Existing efforts on therm
be divided into two main ca
tion [1, 2, 3] and *recogniti*
methods aim to find a comm
tive information, where ima
ing to the same person can b
measurements. Classical ap
ilarities [1], partial least squ
tors such as SIFT and HOG
ods have shown to be effec
tasks, motivating more rece
formance. Common techn

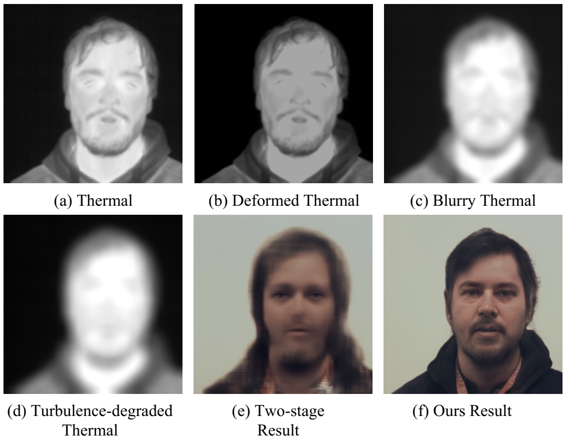


Fig. 1: Visualization of the degraded thermal image under atmospheric turbulence and its corresponding reference image. Most high-frequency details are missing in the thermal image under turbulence compared with the deformed only and the blur only images.

to reduce the domain gap between thermal and visible modality or leveraging convolutional neural networks (CNNs) to extract domain-invariant features [6, 7, 8, 9].

Recently, *recognition by synthesis* [10, 11, 12, 13, 14] has been used to address the problem of heterogeneous face recognition since any off-the-shelf face recognition method can be seamlessly applied on the translated visible images. Riggan et al. [15] synthesized images by leveraging both global and local regions, resulting in better discriminative quality. Later, Mallat et al. [10] introduced cascaded networks to gradually refine the generated images. Several recent methods leverage Generative Adversarial Networks (GANs) to further improve the perceptual quality of the synthesized images. Specifically, Zhang et al. [11] proposed GAN-VFS that learns to jointly optimize visible feature estimation and facial reconstruction. Di et al. [13] and Immidisetti et al. [14] adopt mulltiple self-attention modules into their GANs to allow long-range correlation modeling, which further enhance the synthesis quality. Our work falls into the later category, which learns to synthesis visible images from thermal modality.

The visual quality of imaging through turbulence suffers distortion from both the blur and deformation operations in the pixel space. The physical model corresponding to turbulence degradation has been established in [16, 17, 18, 19, 20]. It has been reformulated and simplified in the turbulence mitigation works [21, 22, 23, 24] as

$$y = T(H(x)) + \xi, \tag{1}$$

where y is the observed turbulence degraded image, x is the clean

This work was supported b

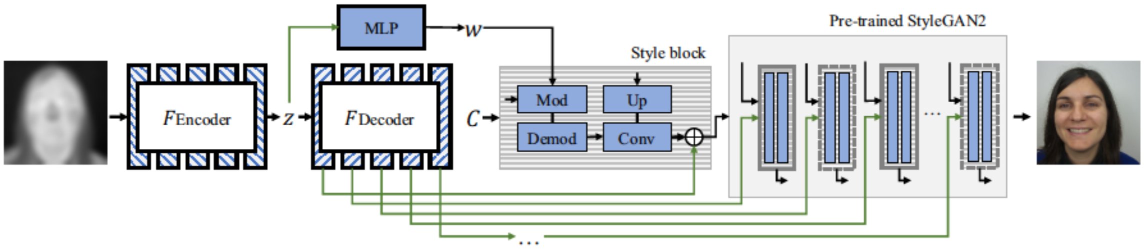


Fig. 2: An overview of the proposed end-to-end network for thermal to visible image synthesis under atmospheric turbulence. The network ensembles a pre-trained StyleGAN2 and learns to project the thermal images into the style-space of a pre-trained StyleGAN2. Resulting from the rich generative priors of StyleGAN2, the network can produce sharper visible images in an end-to-end manner without processing turbulence or thermal images separately.

image, ξ is the additive noise, T is the turbulence degradation operator, and H corresponds to visible to thermal operator. Practical ways for simulating turbulence-like images include physics-based parameter simulation, e.g., Chimitt et al. [25] and Mao et al. [26], and visual effects simulation, e.g., Lau et al. [22] and Yasarla et al. [23]. Though large differences exist in these two ways of simulating turbulence, their results are similar in the sharpness and color bias. Hence, most mitigation methods make use of them without any specific configurations. In this work, we simulate turbulence degraded thermal images according to Mei and Patel [27], which combines multiple random blur and noise with Elastic deformation augmentation. We empirically find that the parameters of simulation following the 300 meters long-range distance configuration leads to the best simulation effects in the thermal images.

Turbulence mitigation is another emerging topic beyond simulation. One straightforward way is build upon two-step mitigation and has been widely applied in recent deep-learning based methods, e.g., TDRN [23], ATFaceGAN [22]. Such methods tend to process the turbulence degraded image with the deformation correction module and deblurring module, and the two results are then fused to obtain the final result. However, recent research [27] shows that learning mitigation in an end-to-end fashion avoids error propagation which often happens in the two-step process. Inspired by [27], in this work, we propose a novel network for thermal to visible image reconstruction under turbulence. The network ensembles a pre-trained GAN and utilizes the GAN prior for learning the reconstruction. Benefiting from the generative priors, we find that such a network is able to simultaneously restore and translate a turbulence degraded thermal image into a high-quality visible image.

2. PROPOSED METHOD

The observation model we follow is $\tilde{I} = T(H(I)) + \xi$ that is mentioned before. Given a thermal image $\tilde{I}$ captured under atmospheric turbulence, we propose to learn to reconstruct a visible image I using a deep neural network $G_\theta(\cdot)$ and optimize its parameters θ according to an objective with the ground truth I. The goal of the network G is to simultaneously restore and translate a thermal image into a visible image. Specifically, the objective includes the adversarial loss of GAN which is combined with the pixel-wise loss, perceptual loss in the pre-trained VGG19 network $\phi(\cdot)$, and identity preserving loss defined in the pre-trained face recognition network $\eta(\cdot)$ as

$$\begin{aligned}\mathcal{L}(G) = &-\lambda_{adv}\mathbb{E}_{G(\tilde{I})}\,\text{softplus}(D(G(\tilde{I}))) \\ &+ \|I - G(\tilde{I})|\ + \lambda_{per}\|\phi(I) - \phi(G(\tilde{I}))\| \\ &+ \lambda_{id}\|\eta(I) - \eta(G(\tilde{I}))\|,\end{aligned} \tag{2}$$

where the λ_{adv}, λ_{per} and λ_{id} are the weights of the adversarial loss, perceptual loss, and identity preserving loss, respectively. The overall architecture in illustrated in Figure 2.

The proposed project module is build upon an encoder-decoder network which takes thermal images $\tilde{I}$ as input and outputs a latent code z and a set of modulation features. In what follows, we provide details of our network.

2.1. Thermal-Turbulence Projection Module

To find the latent code z and modulation features that correspond to the clear images, the projection module should be capable enough to capture both the identity information and local structures. However, extracting local information from thermal images is pretty challenging since most of the details are distorted. We propose to utilize a multi-scale encoder-decoder network for extracting features from $\tilde{I}$, where the number of encoder and decoder layers and their resolutions of the output follows the configuration of StyleGAN2 [28]. Denoting the first feature extraction layer of the encoder part as $\mathbf{E}_0(*)$, we have the shallow feature F_0 as $F_0 = \mathbf{E}_0(\tilde{I})$. In particular, n number of encoder layers with a pooling operation of scale 1/2 are used to extract multi-scale features and preserve the details as follows

$$F_i = \mathbf{E}_i(\text{Pooling}(F_{i-1})), i \in \{1, 2, \dots, n\}. \tag{3}$$

The final output of F_n is then taken as the predicted latent code z for projection. In order to preserve the details of the reconstructed images at different scales, the extracted features $\{F_1, F_2, \dots, F_n\}$ are then processed by n decoder layers $\mathbf{D}_i(\cdot)$ as

$$\bar{F}_i = \mathbf{D}_i(\text{Deconvolution}(F_{i-1})) + F_i, i \in \{1, 2, \dots, n\}. \tag{4}$$

These multi-scale features $\{\bar{F}_1, \bar{F}_2, \dots, \bar{F}_n\}$ are then applied as the feature modulation parameters for gradually correcting the style features of a pre-trained StyleGAN2 at its generation process. Based on the aforementioned two types of encoder and decoder layers, the proposed projection module can learn to project the thermal images under turbulence into the natural image space encoded by StyleGAN2. Here we use green lines in Figure 2 to denote the connections between the projection module and the pre-trained StyleGAN2 for clarification. Note that at the learning process, only the parameters of the projection module are updated according to the gradients, while the parameters of the pre-trained StyleGAN2 are fixed.

2.2. Image Reconstruction Module

As mentioned in Section 2.1, the parameters of the pre-trained StyleGAN2 are fixed during the entire learning procedure, and thus its output always fits the natural image statistics given an arbitrary latent code. Such property significantly simplifies the reconstruction learning since the latent space is limited to the manifold corresponding to natural images only. However, since the generation process of StyleGAN2, i.e., mapping random latent code to natural images is stochastic, ensuring the identity consistency of reconstructed images can be difficult. To overcome this issue, we leverage multi-scale features $\{\bar{F}_1, \bar{F}_2, \ldots, \bar{F}_n\}$ produced by the decoder to modulate the features of StyleGAN2 at generation. In particular, for each output layers $\mathbf{L}_i(\cdot)$ of StyleGAN2 at each resolution, the original procedure takes features $\hat{F}_{i-1}$ extracted from noise and generates the features in the next level as $\hat{F}_i = \mathbf{L}_i(\hat{F}_{i-1}, \epsilon)$, where ϵ is the noise corresponding to normal distribution. In contrast, our modified version modulates the generation process as

$$\begin{aligned} \bar{F}_i^{\text{mean}}, \bar{F}_i^{\text{std}} &= \text{Split}(\bar{F}_i), \\ \hat{F}_i &= (\mathbf{L}_i(\hat{F}_{i-1}, \epsilon) + \bar{F}_i^{\text{mean}}) * \bar{F}_i^{\text{std}}, \end{aligned} \tag{5}$$

where $\text{Split}(\cdot)$ divides the decoded feature $\bar{F}_i$ into two modulate parameters $\bar{F}_i^{\text{mean}}$ and $\bar{F}_i^{\text{std}}$ at the channel dimension. We empirically find that such modulation is able to correct the features during generation, and it helps to preserve details at reconstruction. Following such a modulation process, the final output of the pre-trained StyleGAN2 can preserve both the identity-related details and the natural image statistics.

3. EXPERIMENTS

In this section, we conduct experiments to evaluate our approach against existing state-of-the-art approaches. We select two commonly used thermal-visible datasets: the VIS-TH [29] dataset and the ARL-VTF [30] dataset. In the following, we first briefly introduce them. Then we describe the evaluation metrics, and training and implementation details. Finally, we present both quantitative and qualitative results to showcase the superiority of our method.

3.1. Dataset and Evaluation Metrics

VIS-TH is a challenging Visible-Thermal dataset which is captured in the Long Wave Infrared (LWIR) modality. It contains data from 50 subjects. Images from each subject contain variations in expression, pose and illumination conditions. The paired thermal and visible images are captured by a dual-sensor camera and thus are well-aligned. We randomly select data from 35 subjects for training, data from 5 subjects for validation. The remaining 10 subjects are used for testing.

ARL-VTF is a popular dataset for thermal-to-visible face verification, consisting of data from 220 identities. Images for each subject vary only in expressions. Annotations are provided for alignment. We construct the training set by randomly selecting 160 subjects. We also randomly selects 40 subjects for evaluation and use the rest 20 subjects for testing. This results in 3,200 training pairs, 400 validation pairs and 985 testing pairs. We apply a simple color adjustment to mitigate overexposure over the VIS modality.

3.2. Evaluation Metrics.

To best demonstrate the effectiveness of our approach, we report results with both face verification metrics and image quality measurements. Following [31], we report Rank-1 accuracy, Verification Rate (VR) @ False Accept Rate (FAR)=1% and VR@FAR=0.1% for evaluating face recognition. We create the gallery set by selecting one visible image for each subject and use all thermal images as the probe set. For image quality, perceptual metrics LPIPS [32], NIQE [33], identity metric Deg (cosine distance between LightCNN [34] features), and pixel-wise PSNR and SSIM [35] are reported for comparison.

3.3. Implementation and Training Details

For reconstructing faces, we leverage the pre-trained StyleGAN2 [28]. The projection module for encoding styles and modulation features contains 7 downsample layers and 7 upsample layers. At the lowest level, features have a spatial dimension of 4×4. The size of all convolution filters is set to 3×3. During training, each input batch contains 4 thermal images. We set $\lambda_{adv} = 1$; $\lambda_{per} = \lambda_{id} = 10$. To optimize the parameters, we adopt the Adam [36] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1e - 8$. The initial learning rate is set equal to 2e-3 and reduces to a half after 140K iterations. The training stops at 150K iterations. We implement the proposed model using PyTorch on Nvidia RTX8000 GPUs.

3.4. Turbulence Data Simulation

The simulation method is inspired by Mei and Patel [27], which is originally proposed for turbulence simulation on visible-spectrum images. We experimentally find that such a simulation method is also suitable for thermal images. Specifically, we applied the random blur and elastic deformation on the thermal images, where *isotropic* and *anisotropic* Gaussian kernels are used with a fixed blur kernel size 11 and sampled blur σ from $[1, 11]$. For the elastic deformation, we empirically choose the parameters α and β from the uniform sampling of $[41, 51]$ and $[11, 21]$. Note that the testing sets used in all evaluations are simulated with the same parameter settings for a fair comparison.

Table 1: Image quality results on the **VIS-TH** dataset.

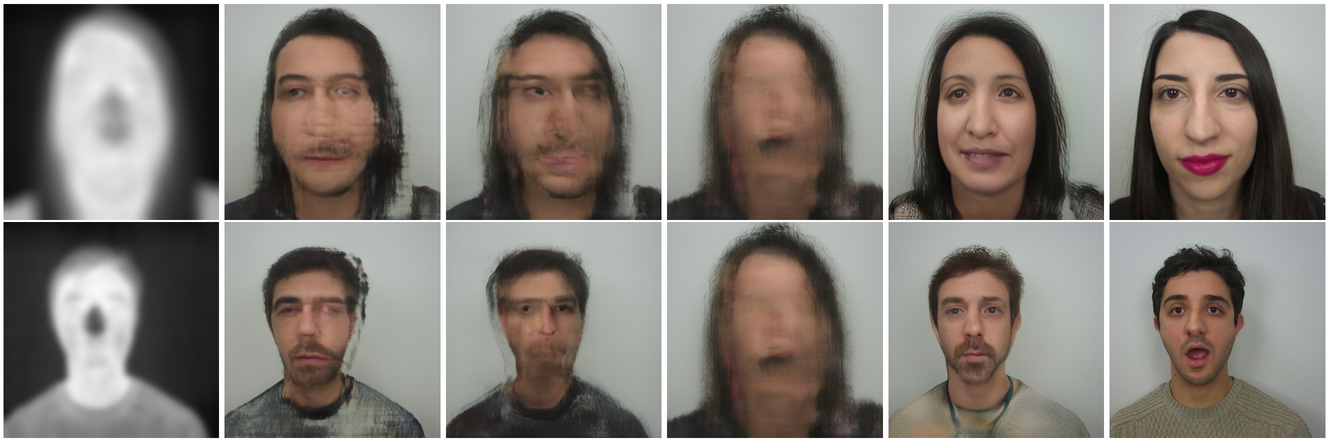
Methods	LPIPS↓	NIQE↓	Deg.↑	PSNR↑	SSIM↑
TH+TB	0.7162	16.547	32.21	6.59	0.3842
One-Stage [37]	0.3355	6.532	50.76	**17.64**	**0.7208**
Two-Stage	0.3740	5.967	50.04	15.92	0.6941
TH Only	0.4243	6.445	40.78	15.59	0.6819
Ours	**0.3127**	**5.547**	**51.68**	16.91	0.6836

Table 2: Verification results on the **VIS-TH** dataset.

Method	Rank-1	VR@FAR=1%	VR@FAR=0.1%
LightCNN [34]	18.10	0.48	0
One-Stage [37]	41.43	7.14	2.86
Two-Stage	32.38	4.76	0
Direct	12.38	0	0
Ours	**48.10**	**10.95**	**3.33**

3.5. Results on VIS-TH Dataset

We evaluate our method (denoted as LRTT) on the VIS-TH dataset and compare it with previous state-of-the art face hallucination approach HifaceGAN [37] trained on the same simulated dataset (denote as *One-Stage*). We further introduce a *Two-Stage* strategy,



Turbulence-degraded Thermal

One-Stage

(c) Two-Stage

Thermal Only

(e) Ours

(f) Reference

Fig. 3: Visualization results of compared methods on the thermal images with simulated turbulence effects.

Table 3: Image quality results on the **ARL-VTF** dataset.

Methods	LPIPS↓	NIQE↓	Deg.↑	PSNR↑	SSIM↑
TH+TB	0.7111	17.52	26.98	5.85	0.3365
One-Stage [37]	0.3963	10.58	50.71	17.88	0.7711
Two-Stage	0.2819	8.564	56.35	18.31	0.7848
Direct	0.3786	8.773	42.86	17.56	0.7615
Ours	**0.2185**	**6.093**	**61.99**	**19.06**	**0.7586**

Table 4: Verification results on the **ARL-VTF** dataset.

Method	Rank-1	VR@FAR=1%	VR@FAR=0.1%
LightCNN [34]	5.69	5.38	0.05
One-Stage [37]	19.90	11.47	4.47
Two-Stage	29.54	22.34	10.46
Direct	15.84	7.51	0.29
Ours	**46.40**	**25.58**	**10.96**

which first reconstructs a clear thermal image via the state-of-the-art turbulence removal approach TDRN [23] and then translates the thermal image to visible domain using HifaceGAN. To showcase the difficulty induced by the turbulence, we also report results of directly translating the degraded thermal image to visible domain using HifaceGAN (denote as *Direct*).

Image Quality Results. Visual results are shown in Figure 3. From this figure, one can see that our approach can synthesize the most clear and accurate faces. In contrast, *Direct* method reconstructs faces with severe artifacts and distortions. Training with the simulation apparently improves the synthesis quality, but the results from *One-Stage* are still very blurry. Moreover, the *Two-Stage* baseline failed to generate high quality faces and yields the worse results compared to *One Stage*, due to the error accumulation as discussed in Section 1. We report the quantitative results in Table 1. Our approach achieves the highest performances in all metrics. It is worth noting that our method also achieves the best Deg. score, which indicates our approach can better preserve the identity information, which is crucial for accurate face verification.

Face Verification Results. In Table 2, we report face verification results. When comparing with HifaceGAN and *Two-Stage*, our method achieves the best performance under all verification metrics. It significantly improves the rank-1 accuracy of the visible domain recognizer LightCNN [34] to 48.10%. This demonstrate its effectiveness in generating high fidelity faces. In contrast, due to the very low synthesis quality, other strategies even reduce the performance of the LightCNN.

3.6. Results on the ARL-VTF Dataset

To further validate the effectiveness of our approach, we conduct experiments on the ARL-VTF dataset. Visual results are shown in Figure 3. ARL-VTF is an easier dataset as it contains more data with variations only in expressions. Therefore, baseline methods can generate faces with reasonable quality. However, although they can produce a rough outline and major facial components, they failed to recover the detailed facial structures and the output images still contain many noticeable artifacts. Our approach can accurately recover the detailed face structures and achieves the highest synthesis quality. Quantitative results are reported in Table 3. Our method performs the best in almost all image quality metrics. The superiority of our approach in thermal-visible face synthesis can further benefit the face verification accuracy. As shown in Table 4, one can see that all methods improve the face recognition accuracy based on LightCNN by a large margin. This is mainly because all baselines can produce the face outline. Since LightCNN is a very powerful visible domain face classifier, this can already improve LightCNN to reach a reasonable performance. However, benefiting from more accurate facial details, our method still yields the best performances in all verification metrics.

4. CONCLUSION

We presented a novel GAN inversion network for end-to-end thermal to visible image translation, where the input suffers from atmospheric turbulence. Compared with the recent thermal image translation approach, two-step turbulence mitigation approach and thermal to visible translation procedure, our method outperforms the other approaches in both visual quality and identity consistency. Though the evaluation is conducted on the synthetically generated turbulence degraded thermal images, we point out that both the network backbone and data augmentation are thoroughly investigated in real-world cases separately. Therefore, we believe the proposed method is a new strong baseline for the similar thermal-spectrum translation tasks affected by atmospheric turbulence.

5. REFERENCES

[1] Brendan F Klare and Anil K Jain, “Heterogeneous face recognition using kernel prototype similarities,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1410–1422, 2012.

[2] Benjamin S Riggan, Nathaniel J Short, and Shuowen Hu, “Optimal feature learning and discriminative framework for polarimetric thermal to visible face recognition,” in *IEEE winter conference on applications of computer vision*. IEEE, 2016, pp. 1–7.

[3] Dihong Gong, Zhifeng Li, Weilin Huang, Xuelong Li, and Dacheng Tao, “Heterogeneous face recognition: A common encoding feature discriminant approach,” *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2079–2089, 2017.

[4] Jonghyun Choi, Shuowen Hu, S Susan Young, and Larry S Davis, “Thermal to visible face recognition,” in *Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX*. International Society for Optics and Photonics, 2012, vol. 8371, p. 83711L.

[5] Shreyas Saxena and Jakob Verbeek, “Heterogeneous face recognition with cnns,” in *European Conference on Computer Vision*. Springer, 2016, pp. 483–491.

[6] Cedric Nimpa Fondje, Shuowen Hu, Nathaniel J Short, and Benjamin S Riggan, “Cross-domain identification for thermal-to-visible face recognition,” in *IEEE International Joint Conference on Biometrics*, 2020, pp. 1–9.

[7] Ran He, Xiang Wu, Zhenan Sun, and Tieniu Tan, “Learning invariant deep representation for nir-vis face recognition,” in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[8] Ran He, Xiang Wu, Zhenan Sun, and Tieniu Tan, “Wasserstein cnn: Learning invariant features for nir-vis face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1761–1773, 2018.

[9] Seyed Mehdi Iranmanesh, Ali Dabouei, Hadi Kazemi, and Nasser M Nasrabadi, “Deep cross polarimetric thermal-to-visible face recognition,” in *IEEE International Conference on Biometrics*, 2018, pp. 166–173.

[10] Khawla Mallat, Naser Damer, Fadi Boutros, Arjan Kuijper, and Jean-Luc Dugelay, “Cross-spectrum thermal to visible face recognition based on cascaded image synthesis,” in *IEEE International Conference on Biometrics*, 2019, pp. 1–8.

[11] He Zhang, Benjamin S Riggan, Shuowen Hu, Nathaniel J Short, and Vishal M Patel, “Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks,” *International Journal of Computer Vision*, vol. 127, no. 6, pp. 845–862, 2019.

[12] Xing Di, He Zhang, and Vishal M Patel, “Polarimetric thermal to visible face verification via attribute preserved synthesis,” in *IEEE International Conference on Biometrics Theory, Applications and Systems*, 2018, pp. 1–10.

[13] Xing Di, Benjamin S Riggan, Shuowen Hu, Nathaniel J Short, and Vishal M Patel, “Polarimetric thermal to visible face verification via self-attention guided synthesis,” in *IEEE International Conference on Biometrics*, 2019, pp. 1–8.

[14] Rakhil Immidisetti, Shuowen Hu, and Vishal M. Patel, “Simultaneous face hallucination and translation for thermal to visible face verification using axial-gan,” in *IEEE International Joint Conference on Biometrics*, 2021, pp. 1–8.

[15] Benjamin S Riggan, Nathaniel J Short, and Shuowen Hu, “Thermal to visible synthesis of face images using multiple regions,” in *IEEE Winter Conference on Applications of Computer Vision*, 2018, pp. 30–38.

[16] Andrei Nikolaevich Kolmogorov, “The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers,” *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 434, no. 1890, pp. 9–13, 1991.

[17] Valerian Ilich Tatarski, *Wave propagation in a turbulent medium*, Courier Dover Publications, 2016.

[18] David L Fried, “Statistics of a geometric representation of wavefront distortion,” *JoSA*, vol. 55, no. 11, pp. 1427–1435, 1965.

[19] David L Fried, “Optical resolution through a randomly inhomogeneous medium for very long and very short exposures,” *JOSA*, vol. 56, no. 10, pp. 1372–1379, 1966.

[20] David L Fried, “Probability of getting a lucky short-exposure image through turbulence,” *JOSA*, vol. 68, no. 12, pp. 1651–1658, 1978.

[21] Rajeev Yasarla and Vishal M. Patel, “Learning to Restore a Single Face Image Degraded by Atmospheric Turbulence using CNNs,” *arXiv preprint arXiv:2007.08404*, 2020.

[22] Chun Pong Lau, Carlos D. Castillo, and Rama Chellappa, “ATFace-GAN: Single Face Semantic Aware Image Restoration and Recognition From Atmospheric Turbulence,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 2, pp. 240–251, 2021.

[23] Rajeev Yasarla and Vishal M. Patel, “Learning to Restore Images Degraded by Atmospheric Turbulence Using Uncertainty,” in *IEEE International Conference on Image Processing*, 2021, pp. 1694–1698.

[24] Nithin Gopalakrishnan Nair and Vishal M. Patel, “Confidence Guided Network For Atmospheric Turbulence Mitigation,” in *2021 IEEE International Conference on Image Processing*, 2021, pp. 1359–1363.

[25] Nicholas Chimitt and Stanley H Chan, “Simulating anisoplanatic turbulence by sampling intermodal and spatially correlated zernike coefficients,” *Optical Engineering*, vol. 59, no. 8, pp. 083101, 2020.

[26] Zhiyuan Mao, Nicholas Chimitt, and Stanley H Chan, “Accelerating atmospheric turbulence simulation via learned phase-to-space transform,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14759–14768.

[27] Kangfu Mei and Vishal M Patel, “Ltt-gan: Looking through turbulence by inverting gans,” *arXiv preprint arXiv:2112.02379*, 2021.

[28] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila, “Analyzing and improving the image quality of stylegan,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.

[29] Khawla Mallat and Jean-Luc Dugelay, “A benchmark database of visible and thermal paired face images across multiple variations,” in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2018, pp. 1–5.

[30] Domenick Poster, Matthew Thielke, Robert Nguyen, Srinivasan Rajaraman, Xing Di, Cedric Nimpa Fondje, Vishal M Patel, Nathaniel J Short, Benjamin S Riggan, Nasser M Nasrabadi, et al., “A large-scale, time-synchronized visible and thermal face dataset,” in *IEEE Winter Conference on Applications of Computer Vision*, 2021, pp. 1559–1568.

[31] Boyan Duan, Chaoyou Fu, Yi Li, Xingguang Song, and Ran He, “Cross-spectral face hallucination via disentangling independent factors,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7930–7938.

[32] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.

[33] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[34] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan, “A light cnn for deep face representation with noisy labels,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2884–2896, 2018.

[35] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[36] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” in *International Conference on Learning Representations*, 2015.

[37] Lingbo Yang, Shanshe Wang, Siwei Ma, Wen Gao, Chang Liu, Pan Wang, and Peiran Ren, “Hifacegan: Face renovation via collaborative suppression and replenishment,” in *ACM International Conference on Multimedia*, 2020, pp. 1551–1560.