

LEARNING TO RESTORE IMAGES DEGRADED BY ATMOSPHERIC TURBULENCE USING UNCERTAINTY

Rajeev Yasarla and Vishal M. Patel

Department of Electrical & Computer Engineering
Johns Hopkins University
{ryasar1, vpatel36}@jhu.edu

ABSTRACT

Atmospheric turbulence can significantly degrade the quality of images acquired by long-range imaging systems by causing spatially and temporally random fluctuations in the index of refraction of the atmosphere. Variations in the refractive index causes the captured images to be geometrically distorted and blurry. Hence, it is important to compensate for the visual degradation in images caused by atmospheric turbulence. In this paper, we propose a deep learning-based approach for restoring a single image degraded by atmospheric turbulence. We make use of the epistemic uncertainty based on Monte Carlo dropouts to capture regions in the image where the network is having hard time restoring. The estimated uncertainty maps are then used to guide the network to obtain the restored image. Extensive experiments are conducted on synthetic and real images to show the significance of the proposed work. Code is available at : <https://github.com/rajeevyasarla/AT-Net>

Index Terms— Atmospheric turbulence degradation, image restoration, deep learning.

1. INTRODUCTION

Atmospheric turbulence can significantly degrade the quality of images acquired by long-range visible and thermal imaging systems by causing spatially and temporally random fluctuations in the index of refraction of the atmosphere. Variations in the refractive index causes the captured images to be geometrically distorted and blurry [1, 2]. Hence, it is very important to compensate for the visual degradation in images caused by atmospheric turbulence. Adaptive optics-based techniques can be used to compensate for turbulence effects in images. However, they require very large, complex, fragile, and expensive hardware. On the other hand, image processing-based approaches are cheap and effective. In this paper, we pose the turbulence degraded image restoration problem as a nonlinear regression problem, where the optimal parameters are learned from synthetically generated data. As a function approximator, we propose to use Deep Convolutional Neural Networks (DCNNs).

This research was supported by NSF CAREER award 2045489.



Fig. 1. Sample image restoration results comparing TDRN with the proposed method, AT-Net.

Under the assumption that the scene and the imaging sensor are both static and that observed motions are due to the air turbulence alone, the image degradation due to atmospheric turbulence can be mathematically formulated as follows [3, 4, 5, 6, 7]

$$\mathbf{y}_k = \mathbf{G}_k(\mathbf{H}_k(\mathbf{x})) + \epsilon_k, \quad k = 1, \dots, M, \quad (1)$$

where \mathbf{x} denotes the ideal image, \mathbf{y}_k is the k -th observed frame, \mathbf{G}_k and \mathbf{H}_k represent the deformation operator and air turbulence-caused blurring operator, respectively and ϵ_k denotes additive noise. The deformation operator is assumed to deform randomly and \mathbf{H}_k correspond to a space-invariant diffraction-limited point spread function (PSF). As can be seen from (1), atmospheric turbulence has two main degradations on the observed images: geometric distortion and blur. Various optics-based [8, 9, 1] and image processing-based [10, 11, 4, 12, 13, 5, 14, 15, 7, 16, 17] turbulence removal algorithms have been proposed in the literature. In general, most image processing methods follow a similar pipeline: lucky region fusion or non-rigid image registration, and blind deconvolution.

In many applications of long-range imaging, such as surveillance, we are faced with a scenario where we have to restore a single image degraded by atmospheric turbulence. In this case, we have the following observation model,

$$\mathbf{y} = \mathbf{G}(\mathbf{H}(\mathbf{x})) + \epsilon, \quad (2)$$

where the subscript k has been removed from (1). This is an extremely ill-posed problem as we have to overcome the effects of both blur and geometric distortion from a single im-

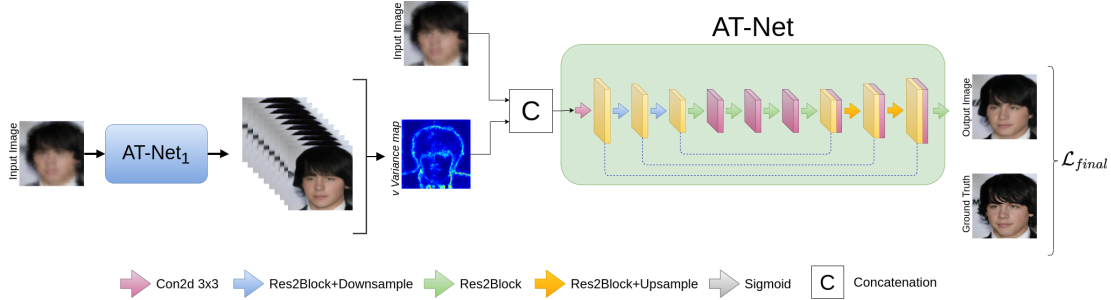


Fig. 2. An overview of the proposed AT-Net atmospheric turbulence degraded image restoration method.

age. Our prior work [18] was one of the first deep learning-based attempts in the literature to recover a single image degraded by atmospheric turbulence. In [18] we first estimate the prior information regarding the amount of geometric distortion and blur at each pixel using two separate networks based on Monte Carlo dropouts [19]. The estimated priors are then used by a restoration network called, Turbulence Distortion Removal Network (TDRN), to restore an image. Although this framework is capable of alleviating blur and geometric distortion caused by atmospheric turbulence, it requires the design and optimization of three separate networks – image deblurring network to estimate the blur prior, geometric distortion removal network to estimate the distortion prior and TDRN to restore an image.

Rather than dealing with geometric distortion and blurry degradation separately, in this paper we improve our work in [18] by developing an approach that estimates only a single prior (in terms of uncertainty) corresponding to a combination of both blur and geometric distortions in turbulence degraded images. The estimated uncertainty maps are then used to guide the network to obtain the restored image. Extensive experiments are conducted on synthetic and real images to show the significance of the proposed work. Figure 1 presents a comparison between TDRN and the improved version of TDRN proposed in this paper, which we call AT-Net (Atmospheric Turbulence distortion removal Network). As can be seen from this figure, the new approach is able to recover details in the image better than [18].

2. PROPOSED METHOD

Figure 2 gives an overview of the proposed restoration network. Given pairs of clean and atmospheric turbulence distorted images, $\{\mathbf{y}_k, \mathbf{x}_k\}_{k=1}^N$, we train two networks similar to AT-Net for restoration. We refer to them as AT-Net₁ and AT-Net. By applying Monte Carlo dropout in every layer of AT-Net₁, we can formulate the epistemic uncertainty [19] and use the corresponding variance as a prior information. For example, given a turbulence distorted image \mathbf{y} , we pass it as an input to AT-Net₁ S times and obtain a set of outputs $\{\mathbf{p}_i\}_{i=1}^S$, where i corresponds to the i th instance of dropout

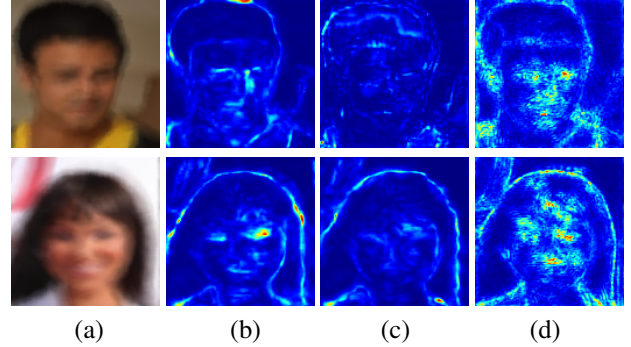


Fig. 3. Image priors in terms of uncertainty maps on sample turbulence distorted images. (a) Turbulence degraded images. (b) Blur prior [18]. (c) Geometric distortion prior [18]. (d) Distortion map estimated using the proposed method.

with corresponding parameters θ^i and $\mathbf{p}_i = \text{AT-Net}_1(\mathbf{y}; \theta^i)$. We define the distortion prior \mathbf{d} as the variance of the outputs $\{\mathbf{p}_i\}_{i=1}^S$, i.e. $\mathbf{d} = \text{variance}(\{\mathbf{p}_i\}_{i=1}^S)$. As explained in [19] this variance is defined as the model uncertainty. However, in our case we use it as a measure of the ability or competence of the network in addressing image restoration. Hence a high variance value at a pixel location in \mathbf{d} means that the restoration network is not able to reconstruct the underlying clean image properly at that pixel in the output image. Figure 3 compares the variance maps estimated by the proposed method with blur and geometric distortion variance maps estimated by our previous work [18]. We can clearly see that the combined prior closely resembles the degradations that are present in the input image compared to the blur and geometric deformation priors.

AT-Net₁. The goal of the AT-Net₁ network is to estimate the uncertainty maps corresponding to the input degraded images. To this end, AT-Net₁ is trained to restore degraded images. In particular, AT-Net₁ is constructed using the UNet [20] architecture with Res2Block as the basic building block [21]. AT-Net₁ consists of the following layers, Res2Block(3,64)-Downsample-Res2Block(64,64)-Downsample-Res2Block(64,64)-Res2Block(64,64)-Res2Block(64,64)-Res2Block(64,64)-Res2Block(64,64)

-Upsample-Res2Block(64,64)-Upsample-Res2Block(64,16)-Res2Block(16,3),
 where Downsample means average pooling layer and Res2Block(m, n) denotes Res2Block with m input channels and n output channels.

AT-Net. Given turbulence degraded image \mathbf{y} , turbulence variance map \mathbf{d} is first computed using AT-Net₁. Then, \mathbf{y} and \mathbf{d} are used as inputs to AT-Net to obtain the restored image $\hat{\mathbf{x}}$, *i.e.* $\hat{\mathbf{x}} = \text{AT-Net}(\mathbf{x}, \mathbf{d})$. The AT-Net network architecture consists of the following layers, Conv2d $3 \times 3(5,16)$ -Res2Block(16,64)- Downsample Res2Block(64,64)-Downsample-Res2Block(64,64) -Res2Block(64,64)-Res2Block(64,64) -Res2Block(64,64) -Res2Block(64,64)-Upsample-Res2Block(64,64) -Upsample-Res2Block(64,3),
 where Conv2d $3 \times 3(m, n)$ denotes a 3×3 convolutional layer with m input channels and n output channels.

The loss function we use to train both AT-Net₁ and AT-Net networks is a combination of the perceptual loss and the \mathcal{L}_1 loss

$$\mathcal{L} = \mathcal{L}_1 + \lambda_p \mathcal{L}_p, \quad (3)$$

where $\mathcal{L}_1 = \|\hat{\mathbf{x}} - \mathbf{x}\|_1$ and $\mathcal{L}_p = \|F(\hat{\mathbf{x}}) - F(\mathbf{x})\|_2^2$. Here, $F(\cdot)$ denotes the features from layer *pool3* of a pretrained VGG-Face network [22]. In our experiments, we set λ_p equal to 0.002.

3. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed AT-Net method on both synthetic and real-world datasets [18]. The performance of different methods on the synthetic data is evaluated in terms of Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) and d_{VGG} . Here, d_{VGG} is defined as feature distance between the restored image $\hat{\mathbf{x}}$ and the ground truth clean image \mathbf{x} . We use the outputs of *pool5* layer from the VGG-Face [22] to compute d_{VGG} . Furthermore, in order to show the significance of different face restoration methods, we perform face recognition on the restored images using ArcFace [23]. The performance of the proposed AT-Net method is compared against the following recent state-of-the-art face image restoration methods [24, 25, 26] and generic single image deblurring methods [27, 28]. Note that we re-train these methods using the same turbulence degraded images that are used to train our network. While re-training the networks, we followed the training procedure including the parameter selection mentioned in those respective papers.

Training Details. We collected 2,000 clean Face images for training AT-Net from the Helen dataset [29]. In addition, 25,000 images were randomly selected from the CelebA dataset [30]. Atmospheric turbulence degraded images are generated using the turbulence synthesis approach presented in [15, 18]. As a result, we generated 1.5 million pairs (*i.e.*

$\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^{1.5 \times 10^6}$) of turbulence distorted images and corresponding clean images for training the AT-Net₁ and AT-Net networks. Given \mathbf{y} , we estimate \mathbf{d} (turbulence variance-map) using AT-Net₁ where we set $S = 10$. Adam optimizer with learning rate of 2×10^{-4} and a batch-size of 10 is used for training both networks. We train AT-Net₁ and AT-Net for 2×10^5 and 1.5×10^6 iterations, respectively.

Testing. The synthetic test set contains 24,000 images from Helen and 24,000 images from CelebA as test images (*i.e.* in total 48,000 test images denoted by \mathcal{D}_{test}). In addition, we use 600 real-world turbulence distorted images collected by the US Army in a variety of different atmospheric conditions as another test set.

Table 1. Quantitative results in terms of PSNR, SSIM, and d_{VGG} on the synthetic datasets. PSNR/SSIM higher the better, and d_{VGG} lower the better

Deturbulence Method	CelebA			Helen		
	PSNR	SSIM	d_{VGG}	PSNR	SSIM	d_{VGG}
Turbulence-distorted	22.43	0.731	5.13	22.35	0.667	6.11
Pix2Pix [31](CVPR 2017)	22.51	0.738	5.28	22.62	0.671	5.83
Pan et al. [24](ECCV 2014)	20.73	0.679	6.04	20.01	0.627	7.28
Shen et al. [25](CVPR 2018)	23.08	0.745	4.72	23.01	0.681	5.14
Yasarla et al. [26](TIP 2020)	24.06	0.768	4.11	23.81	0.702	4.49
Kupyn et al. [27](CVPR 2018)	23.54	0.748	4.51	23.28	0.693	4.98
Zhang et al. [28](CVPR 2019)	24.16	0.770	3.94	23.95	0.709	4.36
TDRN[18] (trained with $\mathcal{L}_1, \mathcal{L}_p$)	25.04	0.802	3.57	24.47	0.725	4.17
AT-Net (ours)	25.31	0.810	3.09	24.95	0.750	3.80

Table 2. Top-1, Top-3 and Top-5 face recognition accuracies on a real-world dataset.

Method	Top-1	Top-3	Top-5
Turbulence-distorted	38.10	49.32	56.13
Pix2Pix [31](CVPR 2017)	37.82	50.76	57.41
Pan et al. [24] (ECCV 2014)	35.67	45.18	50.79
Shen et al. [25](CVPR 2018)	39.91	52.21	58.17
Yasarla et al. [26](TIP 2020)	42.31	57.72	64.40
Kupyn et al. [27](CVPR 2018)	40.72	54.86	62.28
Zhang et al. [28](CVPR 2019)	44.76	60.96	69.75
TDRN[18] (trained with $\mathcal{L}_1, \mathcal{L}_p$)	47.16	62.91	72.87
AT-Net (ours)	48.38	63.96	73.83

Results on Synthetic Data. Results corresponding to different methods on synthetic data are shown in Table 1 and Fig 4 (Rows 3 & 4). Higher PSNR/SSIM and lower d_{VGG} correspond to better quality of the reconstructed images. As can be seen from Fig 4 (Rows 3 & 4) and Table 1, AT-Net outperforms the state-of-the-art face image restoration methods. In particular, generic deblurring methods [27, 28] are not able to perform well due to the lack of prior information. On the other hand, methods that make use of some prior information about the face [24, 25, 26] are unable to perform better because of improper prior estimation from the input images. AT-Net outperforms the state-of-the-art methods including our prior method [18], which clearly demonstrates the effectiveness of the proposed method.

Results on Real-world Data. We also evaluate the performance of different methods on several real-world turbulence



Fig. 4. Image restoration results on sample real-world (rows 1-2) and synthetic (rows 3-4) turbulence distorted images. Compared to the other methods, the proposed AT-Net method produces sharp and clean face images.

distorted images collected by the US Army, published in our previous work [18]. Fig 4 (Rows 1 & 2) illustrate the qualitative performance of different methods on two sample real-world turbulence distorted face images from this dataset. As can be seen from this figure, state-of-the-art restoration methods produce artifacts and blurry outputs especially around the mouth, eyes, and nose regions of the face. On the other hand, AT-Net is able to recover details of the face better and significantly improves the visual quality.

more than 1% improvement in Rank 1, 3 and 5 accuracy compared to our previous method [18].

Results on Real-world Non-face Images. Finally, we conducted experiments on non-face real-world atmospheric-turbulence degraded images to show that our method can be used to restore non-face images as well. As can be seen from Fig 5, AT-Net is able to restore better quality images as compared to our previous work [18].

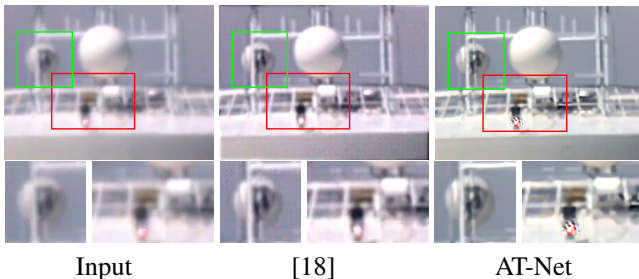


Fig. 5. Image restoration results on sample real-world turbulence distorted non-face images

Additionally, we perform face recognition task on real-world turbulence distorted images using ArcFace [23]. The most similar faces (i.e. Top-K nearest matches) from the restored image are selected from the gallery set (set of different clean images corresponding to 100 identities) to check whether they belong to the same identity or not. Results corresponding to this experiment are shown in Table 2. As can be seen from this table, AT-Net is able to restore real-world images better and preserves identity in the restored images better than the other methods. In particular, AT-Net gives

4. CONCLUSION

We proposed AT-Net to address the atmospheric turbulence distortion removal problem from a single image. This work can be viewed as an extension of our previous approach [18]. Rather than having two separate networks for estimating blur and geometric deformation uncertainties, we only use a single network to estimate the overall distortion introduced by turbulence. As a result, the proposed method reduces the complexity and at the same time is able to provide on par or better results compared to previous state-of-the-art methods. We conducted extensive experiments using synthetic and real data to show that AT-Net network mitigates the geometric deformation and blur introduced by turbulence, and restores better quality images.

5. REFERENCES

[1] Michael C. Roggemann and Byron M. Welsh, *Imaging Through Turbulence*, CRC Press, 1996.

- [2] Valerian Ilich Tatarski, *Wave propagation in a turbulent medium*, Courier Dover Publications, 2016.
- [3] D. H. Frakes, J. W. Monaco, and M. J. T. Smith, "Suppression of atmospheric turbulence in video using an adaptive control grid interpolation approach," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, 2001, vol. 3, pp. 1881–1884 vol.3.
- [4] Md Hasan Furhad, Murat Tahtali, and Andrew Lambert, "Restoring atmospheric-turbulence-degraded images," *Applied optics*, vol. 55, no. 19, pp. 5082–5090, 2016.
- [5] Xiang Zhu and Peyman Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 157–170, 2012.
- [6] Michael Hirsch, Suvrit Sra, Bernhard Schölkopf, and Stefan Harmeling, "Efficient filter flow for space-variant multiframe blind deconvolution," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 607–614.
- [7] 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.
- [8] James E Pearson, "Atmospheric turbulence compensation using coherent optical adaptive techniques," *Applied optics*, vol. 15, no. 3, pp. 622–631, 1976.
- [9] Robert K Tyson, *Principles of adaptive optics*, CRC press, 2015.
- [10] Samy Metari and François Deschenes, "A new convolution kernel for atmospheric point spread function applied to computer vision," in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.
- [11] Masao Shimizu, Shin Yoshimura, Masayuki Tanaka, and Masatoshi Okutomi, "Super-resolution from image sequence under influence of hot-air optical turbulence," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. Cite-seer, 2008, pp. 1–8.
- [12] Enric Meinhardt-Llopis and Mario Micheli, "Implementation of the centroid method for the correction of turbulence," *Image Processing On Line*, vol. 4, pp. 187–195, 2014.
- [13] Mario Micheli, Yifei Lou, Stefano Soatto, and Andrea L Bertozzi, "A linear systems approach to imaging through turbulence," *Journal of mathematical imaging and vision*, vol. 48, no. 1, pp. 185–201, 2014.
- [14] Chun Pong Lau, Yu Hin Lai, and Lok Ming Lui, "Restoration of atmospheric turbulence-distorted images via rpca and quasi-conformal maps," *Inverse Problems*, vol. 35, no. 7, pp. 074002, 2019.
- [15] Wai Ho Chak, Chun Pong Lau, and Lok Ming Lui, "Subsampled turbulence removal network," *arXiv preprint arXiv:1807.04418*, 2018.
- [16] D. Li, R. M. Mersereau, and S. Simske, "Atmospheric turbulence-degraded image restoration using principal components analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 3, pp. 340–344, 2007.
- [17] Chun Pong Lau, Hossein Souri, and Rama Chellappa, "Atfacegan: Single face image restoration and recognition from atmospheric turbulence," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 2020, pp. 32–39.
- [18] 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.
- [19] Alex Kendall and Yarin Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Advances in neural information processing systems*, 2017, pp. 5574–5584.
- [20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [21] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr, "Res2net: A new multi-scale backbone architecture," *arXiv preprint arXiv:1904.01169*, 2019.
- [22] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al., "Deep face recognition.," in *bmvc*, 2015, vol. 1, p. 6.
- [23] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.
- [24] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang, "Deblurring face images with exemplars," in *European conference on computer vision*. Springer, 2014, pp. 47–62.
- [25] Ziyi Shen, Wei-Sheng Lai, Tingfa Xu, Jan Kautz, and Ming-Hsuan Yang, "Deep semantic face deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8260–8269.
- [26] Rajeev Yasarla, Federico Perazzi, and Vishal M Patel, "Deblurring face images using uncertainty guided multi-stream semantic networks," *arXiv preprint arXiv:1907.13106*, 2019.
- [27] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří 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.
- [28] Kai Zhang, Wangmeng Zuo, and Lei Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [29] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang, "Interactive facial feature localization," in *European conference on computer vision*. Springer, 2012, pp. 679–692.
- [30] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3730–3738.
- [31] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei 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.