This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2025.3528544

Image Registration and Deep NeuroFuzzy Networks for Mitigating Atmospheric Turbulence Effects in Consumer-based Optical Imaging

Sunder Ali Khowaja, *Senior Member, IEEE*, Usman Ali, Kapal Dev, *Senior Member, IEEE*, and Ik Hyun Lee, *Member, IEEE* (Correspnding Author)

***Abstract*—Consumer-based optical imaging systems are characterized as big data processing systems, which are drastically affected by atmospheric turbulences that add geometric distortions and blur effect to the images when used in outdoor condition. Physics-grounded simulators have been proposed recently to generate synthetic data but the generalization to the real-world turbulent images is not so good. In this paper, we combine the characteristics of image registration, deep neurofuzzy methods, and channel-attention based discriminative learning strategy to propose image registration, neurofuzzy based denoising, and deblurring network (RND2Net). The RND2Net is designed on a principle that it does not require turbulent image pairs (ground truth images) to train the network, which closely resembles the real-world situation used as consumer devices. The registration module focuses on the region-based fusion techniques while the denoising and deblurring module incorporates deep neurofuzzy network along with dense residual blocks and channel attention mechanism to train the network. The RND2Net is also designed to reduce the noise and blur effect from images, while generalizing on the down-stream tasks, such as text recognition. Experimental results show that the RND2Net yields better performance quantitatively as qualitatively on synthetic and real-world datasets in comparison to existing state-of-the-art methods.**

***Index Terms*—Deep NeuroFuzzy Networks, Big Data Processing, Image Restoration, Atmospheric Turbulence, Channel Attention, Image Registration**

I. Introduction

There are many causes for the visual quality degradation of images acquired from consumer devices. Concerning the long-distance imaging, atmospheric turbulence is one of the most prevalent reasons for quality degradation. On very hot days, the phenomenon can usually be observed where the stationary objects seem to fluctuate or move that are near ground level. The same phenomenon can be observed with hot chimneys, looking through hot mediums such as roads in deserted areas on a hot summer, looking past the objects near jet engines, and so forth. Long-range objects can also exhibit turbulence characteristics at night such as flickering of stars or long-range objects seen through high-beam lights.

Dr. Sunder Ali Khowaja is with School of Computing, Dublin City University and ADAPT Centre, Dublin, Ireland. sunderali.khowaja@dcu.ie

Dr. Usman is with Department of Computer Science and Engineering, Sejong University, South Korea. usman.ali@sejong.ac.kr

Kapal Dev is with Department of Computer Science, Munster Technological University, Cork, Ireland. kapal.dev@mtu.ie

Dr. Ik Hyun Lee is with Department of Mechatronics Engineering, Tech University of Korea, Siheung 15073, Republic of Korea, and also with IKLAB Inc., Seoul 08513, Republic of Korea. ihlee@tukorea.ac.kr

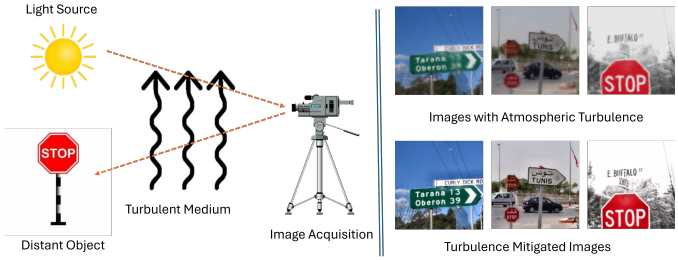

Fig. 1. Distortion formation concerning atmospheric turbulence in vision systems along with some examples.

The phenomenon is observed because of the large variations in temperature or density particles, thus causing rapid changes in refractive index. An example is illustrated in Figure 1. Subsequently, it can be said that when the optical transmission path faces random fluctuations in terms of refractive index, atmospheric turbulence is observed [2]. The most common issue with atmospheric turbulence is the temporal oscillation effect that causes space- and time-blur, and geometric distortions commonly known as motion blur. Depending on the intensity of atmospheric turbulence, the object might seem invisible or blend in with the background, even if the camera is dead static. The said effect creates drastic issues for object recognition, text recognition, semantic segmentation, and image restoration tasks. Some examples of atmospheric turbulence images are shown in Figure 1.

In general, the literature suggests that the solution to the distortion and blur caused by atmospheric turbulence is based on hardware- and image processing-based solutions. The hardware-based solutions are considered to be hard and expensive considering the implementation in comparison to image processing-based solutions [3], [58]. Existing approaches based on image processing techniques take into account a set of turbulent images to construct a restored image. However, the studies make a basic assumption that the object and camera are static. The reason for considering a set of frames instead of a single image is that it cannot model the time-varying geometric and blur distortions or embedded air turbulence. Many of the existing approaches are based on the "lucky frame" selection that selects high quality frames from the degraded sequences to perform frame-level fusion for restoring the image [4]. Such studies assume that some high-quality images might be available even with a short-exposure video.

Such an assumption of finding lucky frames usually seems less practical in real-world situations. Researchers addressed this problem by selecting good regions from different frames instead of considering the whole frame. The selection of good region method performs better than the "lucky frame" and to some extent exhibits real-world situation, however, in order to find good regions, the method needs a large volume of distorted frames. Furthermore, the approach cannot handle point spread function diffraction (PSFD) which is a different kind of blur in comparison to the spatial blur [5], [59]. In order to alleviate the issues associated with good regions, studies moved to image registration-based methods. These methods build an average frame from the available frames and then perform a registration method to align all other frames onto the average frame. Usually the studies considered B-spline registration method to perform the aforementioned task. The study in [6] used the B-spline basis function for image registration to construct an average frame followed by a Bayesian reconstruction framework that uses optimization function to generate a high-quality restored image. Such method lacks real-world applicability as it assumes the blur to be time and space invariant. The authors in [7] extended their work by adding a fusion module to generate near-diffraction-limited blur image. The restored image is then generated using blind deconvolution operation. The method was computationally exhaustive and was not suitable for high-resolution. The non-rigid registration method as mentioned above can miss fine details from the images that ought to be fused, which leads to the registration error [3].

Over the years, researchers have also proposed the use of deep learning methods to perform the restoration of images. However, most of the deep learning methods focus on solving deconvolution or image denoising problems. For atmospheric turbulence, deep learning approaches are still in their infancy. The study in [8] assumed that a Gaussian distribution can be used to model the spatial displacement between turbulent images. Based on the assumption, they used a standard Gaussian denoiser to restore the turbulent images. Subsequently, the study [9] exploited the UNet architecture for image restoration, which was initially proposed for medical image segmentation [10]. A volume of 50 concatenated frames was provided to the network architecture as an input for generating a single image output. However, researchers suggest that such methods cannot be used for moving objects as it requires long length inputs. The study in [55] explored several network architectures initially proposed for deblurring, denoising and deconvolution approaches for mitigating atmospheric turbulence mitigation. More recent methods explore the use of generative adversarial networks (GANs) and transformers to perform the image restoration [11], [60]. Most of the studies focus on designing a deep learning method that is based on the Gaussian distribution displacement of frames.

Recently, neurofuzzy systems have been proposed due to the characteristics that correspond to easy implementation and robustness to uncertainties [12]. Many studies have promoted the integration of neurofuzzy and deep learning systems due to their improved performances [13]. The aforementioned methods were proposed to improve the ability of a DNN to extract better representations. However, in doing so, the systems were experiencing additional computational burden and higher execution times.

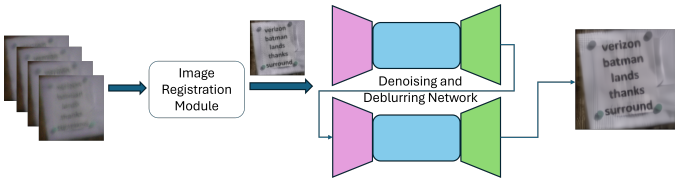

Fig. 2. Architectural workflow of the proposed RD2Net for atmospheric turbulence mitigation from images.

To the best of our knowledge, the fusion of registration methods and the deep neurofuzzy-based approaches have not been proposed in the context of atmospheric turbulence mitigation in images. To this end, we propose a non-rigid registration, and neurofuzzy network for image denoising followed by a deblurring module that leverages the channel attention and dense residual blocks (RND2Net) to solve the atmospheric turbulence mitigation challenge. The Registration module is based on discrete wavelet transform (DWT) and is responsible for generating an approximate registered image based on priority, boundary, and similarity maps. The priority and boundary maps are computed from overlapping regions, whereas the similarity maps are generated using region of interest (ROI) size, intensity similarity, and sharpness. We also propose a deep neurofuzzy network for image denoising and channel attention with dense residual blocks-based deblurring modules for preserving the fine details and generating high-quality restored image. The proposed method secured second position in UG2+ Atmospheric Turbulence Mitigation associated with IEEE Computer Vision and Pattern Recognition Challenges in 2022 and secured third position in UG2+ Coded Target Restoration challenge associated with IEEE Computer Vision and Pattern Recognition Challenges in 2023. The method was also on top of the board for UG2+ Atmospheric Turbulence Mitigation in the validation phase in 2023. The contributions of this work are summarized as follows:

- The fusion of non-rigid registration, neurofuzzy network with denoising and deblurring modules is proposed.
- A DWT based registration module is proposed to perform the fusion of available frames and a selected frame with identified region of interest.
- A neurofuzzy network for image denoising is proposed for effective restoration.
- A deblurring network with dense residual blocks and channel attention is proposed to generate a high-quality turbulence free image.
- A comparative analysis with state-of-the-art techniques is performed to prove the efficacy of the proposed approach.

The rest of the paper is structured as follows: Section II consolidates a brief literature review of existing works. Section III provides details for the proposed RND2Net. Section IV presents the experimental setting and results. Section V presents the concluding remarks, limitations, and future works corresponding to the proposed work.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2025.3528544

II. Proposed Methodology

The proposed work leverages the characteristics of both the registration as well as data-driven approaches to ensure the optimal mitigation of atmospheric turbulence. This section provides details concerning the architectural modules for mitigating the atmospheric turbulence from images. The proposed workflow of RND2Net is shown in Figure 2. The image registration module takes a series of atmospheric turbulent images, performs the registration, and then fuses the registered frame with the rest of the images. The resultant image will then undergo neurofuzzy denoising and deblurring network that enhances the image for a specific application. The details for each of the modules of RND2Net are given in the subsequent subsections.

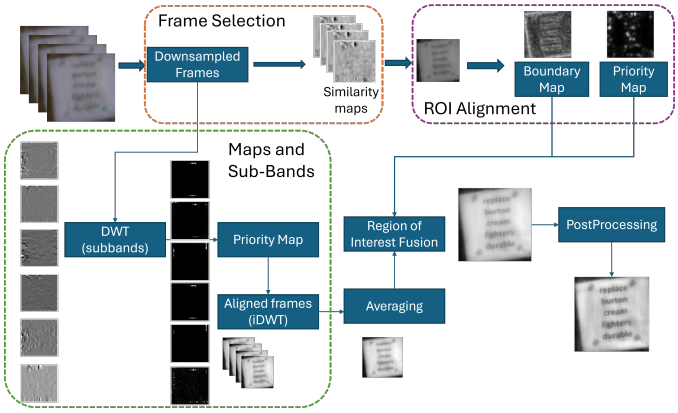

Fig. 3. Proposed block diagram for Image Registration Module.

A. Image Registration Module

The proposed registration module first selects the most informative frames and aligns the region of interests (ROIs) using morphological image processing, which computes the similarity maps from the selected frames. The ROI aligned image then undergoes the computation of boundary and priority maps that extract meaningful information. Similarly, another stream computes the discrete wavelet transform (DWT) on the downsampled frames on varying phases (subbands). A priority map from the DWT maps is constructed followed by the inverse DWT (iDWT) process. The frames constructed from iDWT are then averaged to construct a single image. The outputs from DWT/iDWT, boundary and priority maps are then fused to generate a registered image. The expanded version of the image registration module is shown in Figure 3. The steps performed in the Image Registration module are covered in this section.

1) Frame Selection: It has been suggested in the literature [18] that using all the available frames to perform image registration, especially for removing atmospheric turbulence, would cause more disruption in the form of blurriness. In this regard, a frame selection process has been a standard to perform before moving to further operations. The frame selection process carefully selects frames based on intensity and sharpness, using a highpass image filter to compute a sharpness metric. The similarity between frames is then evaluated in two parts: first, by comparing the magnitudes of the filtered coefficients; second, by calculating the mean square error (MSE) between each frame and an average intensity image to assess their similarity in terms of intensity. The selection of the optimal number of frames, which balances both sharpness and intensity similarity, is treated as a hyperparameter that needs to be fine-tuned. Therefore, we shall discuss the impact of variance on the selection of number of frames in the experimental section.

2) ROI Alignment: Generally, one of the effects that atmospheric turbulence causes to an image is the misalignment of region of interests. Image registration methods work quite well if the distorted objects in the successive frames are displaced with a smaller magnitude. For large displacements, conventional image registration methods do not work well. Similar to the conventional registration approaches, feature matching methods also are not suitable for large displacements in the distorted objects as the strong gradients are spatially distorted in a random manner within each frame. Furthermore, as the distortions are quite high, manual marking of the objects is considered to improve the ROI alignment. In real-world situations, it's not possible to mark the objects manually, therefore, in this regard, we leverage morphological image processing but instead of marking the objects manually. The morphological image processing performs dilation operation to expand the region of interests and then used connected component labeling to isolate and identify the ROI. The properties of ROI are then calculated, which then can be used for further alignment process. We consider the first frame to be the reference one. We apply the thresholding method proposed in [26] for the transformation of the frames to binary maps. The threshold for the method is selected from the histogram of the reference image. The mask of the reference frame is used to compute the center position. Each subsequent selected frame is then aligned in reference to the center position of the reference frame, accordingly.

3) Maps and Sub-bands: This step in the proposed image registration module is the most important one. Let us consider the ROI aligned image as $RI_u^{l,\vartheta}$, where ϑ represents the DWT sub-bands varying from $(1, ..., 6)$ and $u \in U$ of level l. The aligned image RI is divided into equal regions denoted by ς_u. A multiresolution priority map is then formulated as shown in equation 1.

$$PMap_u = \{pmap_{u,r_{u,1}}, pmap_{u,r_{u,2}}, ..., pmap_{u,r_{u,\varsigma_u}}\} \quad (1)$$

where r_{u,ς_u} represent the regions in each image u. For instance, if the regions $r_{1,1}$ and $r_{2,1}$ from images 1 and 2 overlap, the region will be splitted into two having the same priority. Considering the aforementioned example, the fusion of priority maps is performed from the DWT subbands such that $RI_{DWT} = RI_1 \cup RI_2 \cup ... RI_U$. The fusion in the existing studies with DWT is either performed with averaging of the regions [27] or using maximum priority [28]. In this study, we adopt the averaging-based fusion technique as proposed in [27], to obtain the desired wavelet coefficients. We then apply a lowpass filter on the DWT coefficients and perform inverse transform on the coefficients to obtain the aligned frames. All the frames are then averaged to get a registered image using

DWT coefficients as shown in Figure 3.

In the other stream, we apply highpass filter to each region r_u on the ROI aligned image RI. The normalization area of each region and the highpass coefficients are denoted by $|r_u|, h(x, y)$, accordingly. The following formulation is then used to obtain the priority map in spatial domain, shown in equation 2.

$$PMap(r_u) = \frac{1}{|r_u|} \sum_{\forall (x,y) \in r_u} |h_u(x, y)| \tag{2}$$

The boundary map is computed to find the temporal consistency and maximum sharpness. We use watershed multiscale segmentation approach along with dilation operation to construct the boundary map. The map also undergoes averaging filter to avoid discontinuities among the neighboring areas, accordingly.

In order to perform a fusion from the DWT, boundary and priority map, we compute the DWT coefficients on each level from a combined averaged image such that $avg(RI_{DWT} + PMap + BMap)$. The DWT coefficients for the fusion are denoted by $D_f^{l,\vartheta}$. We further reduce the distortion from the fused image by adjusting the phase of DWT coefficients to unit average vector. The study in [29] suggested that the turbulence causes pixel deviation, which is representative of the deviation from direction, velocity and wind. In this regard, quasi-periodic motion with zero-mean, i.e. average phase, can be used to reduce the aforementioned deviations. We compute the average phase from the selected frames and use them to improve the fused image as shown in equation 3.

$$D_f^{\sim l,\vartheta} = \frac{\sum_{u=1}^{U} D_f^{l,\vartheta}(x, y)}{|\sum_{u=1}^{U} D_f^{l,\vartheta}(x, y)|} |D_f^{l,\vartheta}(x, y)| \tag{3}$$

We then apply inverse DWT to the coefficients for obtaining the fused registered image RF, accordingly.

4) Post Processing: Existing studies have shown that the atmospheric turbulent images exhibit effects of fog, haze, and poor contrast. The suggestion is to perform a postprocessing step in order to improve the quality of the image. Several techniques are proposed in the literature [18], [19], [30], [31] to perform image enhancement operation, however, in this study, we adopt histogram clipping and top-hat transformation as proposed in [31] to balance out image contrast, thus, by extension, improving the quality of the registered fusion image as shown in Figure 3.

B. Denoising and Deblurring Network

This subsection presents the details regarding the network architecture for our denoising and deblurring module. The overall proposed network architecture for denoising and deblurring module is shown in Figure 4 and 5. The proposed architecture has been inspired by the study [32], however, there are several differences in comparison. For instance, the proposed architecture leverages deep neurofuzzy network and dense residual blocks (DRBs) for dealing with denoising and deblurring of images, respectively. The restoration process is further improved by adding a channel attention mechanism. The proposed architecture operates on a sub-pixel mechanism,

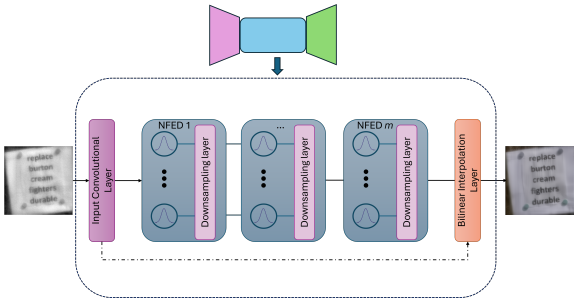

Fig. 4. Proposed Neurofuzzy Denoising Architecture for image restoration.

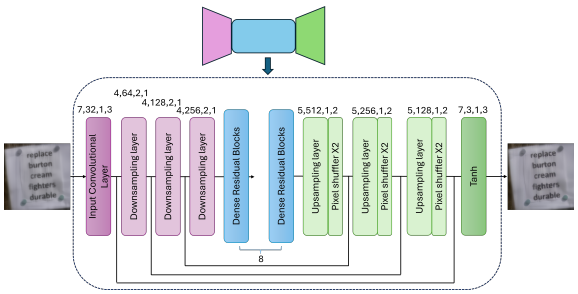

Fig. 5. Proposed Deblurring Architecture for image restoration. We use the convention (kernel size, number of feature maps, stride, padding) to represent the layer parameters, respectively.

which helps the network to learn fine details, thus, helping it to remove geometric distortions as well. Lastly, the loss function for the proposed architecture caters to the needs of the application at hand. The subsections below present further details concerning the network architecture shown in Figure 4 and Figure 5.

1) NeuroFuzzy Autoencoders: For the image denoising part, we propose the use of neurofuzzy based denoising with an architecture that follows encoder-decoder style. Specifically, the architecture involves an input, an output, and a fuzzy layer. In the proposed neurofuzzy encoder decoder (NFED), neurons are blurred using the Gaussian membership function. The formulation for the encoder is shown below:

$$\bar{\mathbf{H}}^1(t) = exp^{-\psi^1}(t) \tag{4}$$

where

$$\psi^1(t) = -\frac{(\bar{b}^1 - c^1(t) + \bar{W}^1(t)SI)^2}{2\sigma^1(t)^2} \tag{5}$$

The notation $\bar{\mathbf{H}}^1(t)$ refers to the output from the NFED encoder part, SI is a sample noisy training image that is considered for specifically training the denoising and deblurring

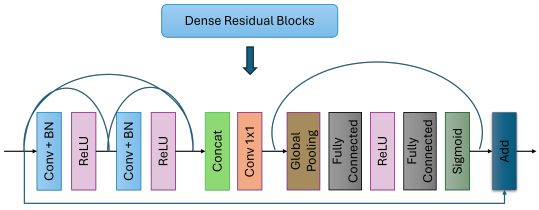

Fig. 6. Proposed Architecture for Dense Residual Blocks in Deblurring Module

network, however, in the working stream, it will be replaced by RF, which is the output from the registration block. From the registration block, $\bar{W}^1$ corresponds to the encoder's weight matrix, c^1 is the center vector, $\bar{b}^1$ is the bias vector, and σ^1 is the width associated with the Gaussian membership function. The membership function can also be considered as the transition behavior between the encoder and decoder. On the other hand, the fuzzy membership will be considered as an input to the decoder, which is realized using the following formulation.

$$\Omega^1(t) = \frac{1}{exp^{\hat{b}^{-1}(t) - \hat{W}^1(t)\bar{\mathbf{H}}^1(t)} + 1} \tag{6}$$

where $\Omega^1(t)$ is the output vector from NFED decoder, $\hat{W}^1(t)$ is the weight matrix of the decoder, and $\hat{b}^{-1}(t)$ is the bias vector. We optimize the loss function of the NFED as shown in the formulation below.

$$Loss_{NFED}(t) = \frac{1}{2}\|GT - \Omega^1(t)\|^2 \tag{7}$$

For addressing the slow training process, we employ the compact training strategy as proposed in the study [12]. The output from the NFED is represented by FD.

2) Dense Residual Blocks: Studies focusing on image restoration have shown that residual blocks have been efficient for the task as they propagate information effectively through residual connections. The study [32] leverages residual blocks for removing water turbulence from the images. As our architecture is inspired from the study [32], we replaced the residual blocks with the dense residual blocks [33] and instance normalization with the group normalization [34], as it exhibits better accuracy, training stability, and support for different batch sizes. We also add batch normalization layer with the dense residual block in order to further stabilize the training process. Although the dense residual blocks are capable of propagating information to the deeper layers, they do not consider channel relationships among the convolutional features. In this regard, we integrate channel attention layers [35] into DRBs. The goal of the channel attention is to extract better representations and improve the interdependencies between the channel and convolutional features. Our intuition regarding the dense residual blocks for improving the image restoration process is compliant with some of the recent works [36], [37]. The architecture for dense residual blocks is shown in Figure 6.

3) Network Architecture: The proposed network architectures for denoising and deblurring network are shown in Figure 4 and 5. The denoising network is trained to remove noise as well as geometric distortions, which further improves the quality of image in terms of atmospheric turbulence. However, the denoising network removes high frequency details, thus, adds blurry effect in the process. In this regard, a deblurring network is also used that minimizes the blur effect while retaining perceptual details. We discuss the components of the proposed network architecture below.
The downsampling layer comprises of convolutional, Rectified Linear Units (ReLU), and group normalization layers. The size of filters, depth, stride, and padding for each convolutional layer is shown in Figure 5. The dense residual blocks take into account the output from downsampling layers and then upsampled using the subsequent layers. The structure of the upsampling layer follows a similar convention to that of downsampling layers. The upsampling layer performs the deconvolution operation, thus decreasing the depth to 32 channels. The output convolutional layer provides us with the deblurred image depending on the employed network stream. The bilinear upsampling layer in the proposed architecture helps to identify the smaller movements at pixel level by increasing the field size. For the deblurring network, we add skip connections [10] as they are suggested to be efficient for restoring the details lost during the downsampling process. Furthermore, in the deblurring network we use the bilinear layers with the sub-pixel convolution layer as suggested in [38].

4) Objective Function: The objective of the deblurring module is to minimize the loss between the blurry image as an input and ground truth image at output. This loss is optimized in a supervised manner on some datasets that are publicly available but do not represent atmospheric turbulence effect. The aim of this method is to cope with atmospheric turbulence effect in an image without considering turbulence data. The presumption is beneficial as it caters to real-world situations where the availability of atmospheric turbulence datasets is another research problem. We train both the deblurring networks to optimize learnable parameters θ, respectively. Two losses are mostly associated with the proposed network architecture. The loss that minimizes generator output with the ground truth $Loss_{gn}$, and the loss that minimizes deblurring output in association with denoised image $Loss_{db}$. The $Loss_{gn}$ is composed of adversarial, perceptual, and content losses. To comply with the natural image manifold, adversarial loss was used in this study. We used DCGAN [39] architecture as our discriminator in this study. The DCGAN in the proposed work is trained using least square loss [40] rather than the cross-entropy loss, as it yields better stability during the training process and produces better output image. The formulation for the loss function is shown in equation 8.

$$\begin{aligned} Loss_{gen} = & \\ & \rho_1 \sum_{x=1}^{j_{4,3}} \sum_{y=1}^{k_{4,3}} |fmap_{4,3}(FD)_{x,y} - fmap_{4,3}(\theta(\hat{FD}))_{x,y}| \\ & + \sum_{p=1}^{M} \sum q = 1^N |FD_{p,q} - \theta(\hat{FD}_{p,q})| \\ & + \rho_2 [dis_\alpha(\theta(\hat{FD}) - 1)]^2 \end{aligned} \tag{8}$$

$$Loss_{dis} = [dis_\alpha(\theta(\hat{FD}))^2 + (dis_\alpha(FD) - 1)^2] \tag{9}$$

The notation α represents the trainable parameter, dis represents the discriminator, $fmap$ refers to the feature map, the notation $4, 3$ refers to the Conv4 _3 in the VGG16 architecture, its weight and height is represented by j, k and the parameters ρ_1 and ρ_2 are empirically selected for the optimal performance.

Subsequently, the $Loss_{db}$ is then trained for minimizing content and perceptual loss [41]. The former loss computes L1 distance in spatial domain from the reconstructed image and the original one, while the latter considers the perceptual similarity using the feature maps. The total loss for the proposed network is shown in equation 10.

$$Loss_{db} = \sum_{x=1}^{j_{4,3}} \sum_{y=1}^{k_{4,3}} |fmap_{4,3}(FD)_{x,y} - fmap_{4,3}(\hat{FD})_{x,y}| \sum_{p=1}^{M} \sum_{q=1}^{N} |FD_{p,q} - \hat{FD}_{p,q}| \quad (10)$$

The restored image is represented by $\hat{FD}$, respectively.

III. Experimental Setup and Analysis

A. Datasets

There are a few methods available in the literature for rendering atmospheric turbulent images such as physics-based simulators or numerical simulations. However, both of the image generation methods exhibiting atmospheric turbulence require high computational time as well as power [2]. Furthermore, getting atmospheric turbulent images from simulator does not represent a real-world situation. In this regard, we have not used any training images that represent atmospheric turbulence, which is the one of the main motivations of this study. However, we do use Berkeley segmentation dataset [42], ImageNet validation dataset [43], and waterloo exploration dataset [44] to train our denoising and deblurring module. We add noise and blur to the images using the protocol proposed in study [19].

For the evaluation, we conduct our experiments on two types of datasets, the synthetic and real datasets. For the synthetic dataset, we leverage the simulator proposed in [9] to create distorted sequences from videos. Three levels were chosen for distorting the videos, varying from strong turbulence to weak turbulence. The strong turbulence used a value of 4.5, the medium turbulence used a value of 3.0, and the value for weak turbulence was selected to be 1.5.

For the real dataset, we conduct our experiments on the static as well dynamic scenes. For the static scenes, we present our results using the text dataset [45], while for the dynamic scenes we use the dataset proposed in [48]. All of the results concerning atmospheric turbulence mitigation are evaluated using Peak Signal Noise Ratio (PSNR) and structural similarity (SSIM), accordingly.

B. Implementation and Training Strategy

We train the denoising and deblurring network on the aforementioned datasets discussed in the previous section. Each of the images in the training dataset was resized, cropped, and flipped before adding noise and blur effects. We trained the proposed denoising and deblurring network in an end-to-end fashion. For the denoising network, the number of NFED were set to 7, the number of neurons in NFED is set to 32, and the value of corrupt probability was set to 0.3, respectively. We used the batch size of 16, with a learning rate of $1e^(-5)$ for the first 8 epochs and $4e^(-5)$ for the next 5 epochs. ADAM optimizer with default values was used to optimize the network. The hyperparameters ρ_1 and ρ_2 were empirically determined to be 0.7 and 0.5. The generator and discriminator were updated only once for each iteration. The network was trained on NVIDIA RTX 3060Ti GPU.

C. Experimental Results

This section provides quantitative and qualitative comparisons with existing and state-of-the-art works on synthetic as well as real datasets. We first present the results on the synthetic dataset. For the comparative analysis we add the comparison of the proposed work with CLEAR [46], NERT [24], NDIR [47], TSR-WGAN [48], TurbNet [45], SwinIR [49], Restormer [50], Uformer [51], MPRNet [52], TDRN [53], and PiRN [25]. All the results have been assessed using PSNR and SSIM metrics. The results for the mitigation of atmospheric turbulence from synthetic images are presented in Table 1. The results clearly indicate the superiority of the transformer-based architectures for the mitigation of atmospheric turbulence on images as SwinIR [49], Uformer [51], Restormer [50], TurbNet [45], PiRN [25], and the proposed method used transformer-based backbones for either learning or extracting representative features. However, the proposed method outperforms all other existing works across all the turbulence levels using the synthetic datasets, respectively. The results indicate that using the fusion of registration and generative methods can help mitigating the atmospheric turbulence effect in an optimal manner.

For the static scene dataset, we adopt the idea of performing the evaluation using downstream task, which is the text recognition from atmospheric turbulent images. We perform both the quantitative and qualitative evaluation on the text dataset. The quantitative evaluation evaluates the downstream task of text recognition from restored images using average detected longest common subsequence (AD-LCS), and average word detection ratio (AWDR). The aforementioned metrics were introduced in [45] for the task of optical character recognition. It should be noted that the proposed work has secured 2nd position in Ug2+ atmospheric turbulence mitigation challenge held in conjunction with CVPR 2022-2023. We compare our methods with PiRN [25], TurbNet [45], and SwinIR [49] for the quantitative comparison, while we use TurbNet and PiRN for the qualitative comparison, respectively. The quantitative results on the text dataset are presented in Table 2. It can be realized that the proposed method secures around 2.438 and 0.171 performance gains over the raw input affected by atmospheric turbulence in terms of AD-LCS and AWDR, respectively. Furthermore, the proposed method is able to achieve better results than the existing state-of-the-art approaches. The improvement can further be demonstrated by the visual results shown in Figure 7. The visual results show a drastic improvement when it comes to the blurriness and sharpness, which are highly degraded by the effects of atmospheric turbulence. In addition, the text is much clearer, which helps in improving the classification and recognition tasks, accordingly.

TABLE I
COMPARATIVE ANALYSIS ON SYNTHETIC DATASET USING THE PROPOSED WORK AND STATE-OF-THE-ART WORKS.

Method	Weak		Medium		Strong	
	PSNR	SSIM	PSNR	SSIM	PNSR	SSIM
CLEAR [46]	20.164	0.704	19.341	0.611	17.715	0.488
TSR-WGAN [48]	20.428	0.600	18.811	0.457	17.198	0.347
NDIR [47]	21.366	0.716	19.606	0.603	18.812	0.544
SwinIR [49]	28.016	0.742	25.958	0.689	23.868	0.638
MPRNet [52]	26.943	0.720	24.805	0.669	23.010	0.618
TDRN [53]	26.597	0.699	23.398	0.662	21.592	0.601
Uformer [51]	27.756	0.719	25.294	0.690	23.498	0.629
Restormer [50]	27.897	0.731	25.702	0.692	23.866	0.650
TurbNet [45]	27.995	0.740	26.003	0.741	23.894	0.651
NeRT [24]	22.109	0.766	20.576	0.659	19.311	0.567
PiRN [25]	29.150	0.766	26.860	0.747	25.070	0.665
RND2Net	**30.128**	**0.804**	**28.236**	**0.762**	**26.140**	**0.718**

TABLE II
COMPARATIVE ANALYSIS ON THE TURBULENCE TEXT DATASET USING THE PROPOSED WORK AND STATE-OF-THE-ART METHODS.

Methods	AD-LCS	AWDR
Raw Input	5.076	0.623
TurbNet [45]	7.314	0.758
SwinIR [49]	7.002	0.740
PiRN [25]	7.498	0.776
RND2Net	7.514	0.794

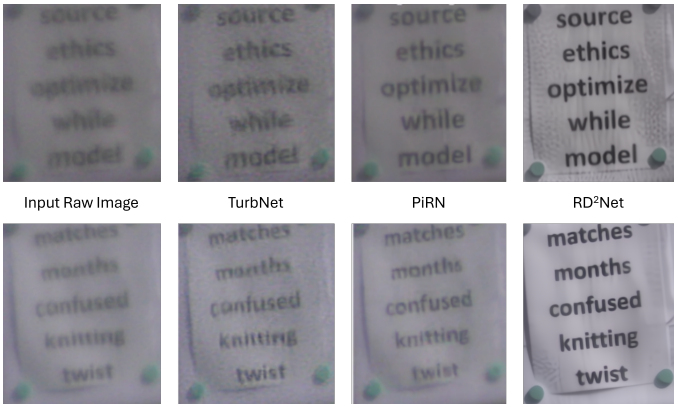

Fig. 7. Qualitative comparison of RND2Net pm real-world Turbulence Text dataset [9] with the existing state-of-the-art methods.

As suggested earlier, we also evaluate the performance of the proposed method on dynamic scene videos [48] and perform a comparative analysis with existing SOTA methods such as TurbNet [45], TSR-WGAN [48], VRT [54], and TMT [55]. The dynamic scenes dataset comprise of videos instead of static images, therefore, we use perceptual quality learned perceptual image patch similarity (LPIPS) [56], and complex-wavelet SSIM (CW-SSIM) [57] along with SSIM and PSNR to evaluate the performance. The use of such evaluation metrics are in compliance with existing works employing the dynamic scenes dataset. For the reference, lower LPIPS represents better frame quality and higher CW-SSIM depict better performance. We present the quantitative results on dynamic scene dataset in Table 3. For VRT [54], we kept the number of frames to 12 as it requires large GPU memory space to deploy if more frames are used. Considering the computational power and memory constraint, we limit VRT to 12 frames. Considering the result achieved, increasing the number of frames won't make much difference in the result. This observation and the idea to limit the frames to 12, is compliant with the existing studies [55] that use VRT for comparative analysis. It can be observed that the proposed method outperforms all the existing SOTA methods across all the evaluation metrics. The results also highlight the generalization power of the RND2Net on static and dynamic scenes, which is essential for a real-world implementation.

TABLE III
COMPARATIVE ANALYSIS WITH STATE-OF-THE-ART METHODS ON DYNAMIC SCENES DATASET

Method	Frames	LPIPS	CW-SSIM	SSIM	PSNR
TSR-WGAN [48]	15	0.2606	0.8596	0.7957	26.33
VRT [54]	12	0.2485	0.8691	0.8300	27.61
TurbNet [45]	1	0.4445	0.8072	0.7149	24.22
TMT [55]	12	0.2475	0.8705	0.8318	27.88
TMT [55]	20	0.2412	0.8741	0.8352	28.01
RNDReg [55]	15	0.2554	0.8604	0.8298	26.22
RNDDN [55]	15	0.2462	0.8688	0.8304	27.54
RNDDB [55]	15	0.2458	0.8690	0.8306	27.59
RND2Net	15	0.2306	0.8763	0.8368	28.54
RND2Net	20	0.2259	0.8854	0.8446	28.86

IV. ABLATION STUDY

It is important to highlight the performance of the proposed method with respect to sensitive parameters and multiple modules. In this regard, we perform an ablation study to show the contribution of each of the modules in the performance improvement concerning atmospheric turbulence mitigation. Furthermore, we also evaluate the sensitivity of the proposed method to number of frames being used that can affect the quality of restoration. We perform these experiment on the dynamic scenes dataset, respectively.

A. Number of Input frames

Image restoration under turbulence, especially with videos, require image sequences to estimate the degradation field from large areas of spatiotemporal domain. It also estimates if the sequences have high temporal correlation. The same degradation property across the frames represents a temporal correlation of 1. Therefore, it is necessary to understand the

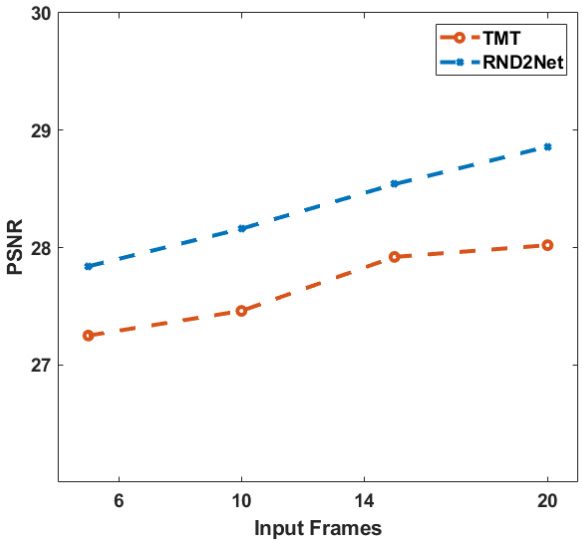

Fig. 8. Ablation results on the performance of Atmospheric Turbulence Restoration with varying number of frames

effect of number of frames on the restoration quality. In this regard, we trained RND2Net on the synthetic dataset using 5, 10, 15, and 20 input frames. We report the ablation study results in relation to the input frames in Figure 8. The results indicate the higher number of frames yields better results. The results also infer that the proposed method has less computational complexity in comparison to the method such as VRT, which is hard to train on more than 12 frames. Based on the results, we used the number of frames as 15 and 20 in our previous experiments. We also compare our results with that of TMT, which uses similar number of frames in their experiments.

B. Impact of Denoising and Deblurring Module

The proposed RND2Net introduces a neurofuzzy network for image denoising and dense residual block-based image deblurring modules. We perform an ablation study to show that both these modules contribute to the performance for mitigating atmospheric turbulence in images. We report our results on dynamic scenes dataset using 15 frames in Table 3. First we evaluate the RND2Net without both the denoising and deblurring modules annotated as RNDReg, then we only use the denoising module without deblurring, annotated as RNDDN, subsequently we only use the deblurring module without denoising annotated as RNDDB, and the final results are shown using the combination of Image Registration, Denoising and Deblurring modules annotated as RND2Net. It can be realized that the image registration module alone achieves better results than TurbNet and TSR-WGAN, while when implemented either of the modules, i.e. denoising or deblurring, achieves better performance than VRT. The modules when combined with image registration module can achieved state-of-the-art results by outperforming TMT method as well.

V. Conclusion

In this paper, we propose the registration, denoising, and deblurring network (RD2Net) to mitigate the atmospheric turbulence effect on the images. The RD2Net combines the characteristics of image registration methods with generative and discriminative learning strategies while integrated transformer networks. The proposed method improves the physical attributes of the image that are degraded during the atmospheric turbulence while improving the downstream tasks that might vary from prediction to the classification tasks. One of the highlighting characteristics of the proposed method is that it is not trained on the atmospheric turbulent images to begin with, which makes the proposed method a plug-and-play method to restore the turbulent images as well as for the restoration of images affected by blurring and noise. This is an extension to the method that has won CVPR UG2+ competition for atmospheric turbulence mitigation from 2022-2023 as a runner-up.

In the proposed work, we use separate modules for denoising and deblurring to mitigate the atmospheric turbulence, which might yield a little bit more computational complexity in comparison to the embedded approach. In this regard, as our future work, we intend to make an embedded end-to-end trainable network to perform the mitigation task. We would also like to extend this work to add more plug-and-play modules such as image superresolution to make it a generalized image restoration method that can be used to denoise, deblur, and restore geometric distortions from the image. Furthermore, we intend to transform the method into a viable product for its immediate use on the browser or a mobile application.

References

[1] Lau, C.P., Lai, Y.H. and Lui, L.M., 2019. Restoration of atmospheric turbulence-distorted images via RPCA and quasiconformal maps. Inverse Problems, 35(7), p.074002.

[2] Nair, N.G., Mei, K. and Patel, V.M., 2023. At-ddpm: Restoring faces degraded by atmospheric turbulence using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3434-3443).

[3] Fazlali, H., Shirani, S., BradforSd, M. and Kirubarajan, T., 2022. Atmospheric turbulence removal in long-range imaging using a data-driven-based approach. International Journal of Computer Vision, 130(4), pp.1031-1049.

[4] Zhang, S., Rao, P. and Chen, X., 2023. Blind turbulent image deblurring through dual patch-wise pixels prior. Optical Engineering, 62(3), pp.033104-033104.

[5] Zhu, X. and Milanfar, P., 2012. Removing atmospheric turbulence via space-invariant deconvolution. IEEE transactions on pattern analysis and machine intelligence, 35(1), pp.157-170.

[6] Shimizu, M., Yoshimura, S., Tanaka, M. and Okutomi, M., 2008, June. Super-resolution from image sequence under influence of hot-air optical turbulence. In 2008 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-8). IEEE.

[7] Zhu, X. and Milanfar, P., 2010, January. Image reconstruction from videos distorted by atmospheric turbulence. In Visual Information Processing and Communication (Vol. 7543, pp. 228-235). SPIE.

[8] Gao, J., Anantrasirichai, N. and Bull, D., 2019. Atmospheric turbulence removal using convolutional neural network. arXiv preprint arXiv:1912.11350.

[9] Mao, Z., Chimitt, N. and Chan, S.H., 2021. Accelerating atmospheric turbulence simulation via learned phase-to-space transform. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 14759-14768).

[10] Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

[11] Chak, W.H., Lau, C.P. and Lui, L.M., 2018. Subsampled turbulence removal network. arXiv preprint arXiv:1807.04418.

[12] Han, H.G., Zhang, H.J. and Qiao, J.F., 2020. Robust deep neural network using fuzzy denoising autoencoder. International Journal of Fuzzy Systems, 22(4), pp.1356-1375.

[13] Talpur, N., Abdulkadir, S.J., Alhussian, H., Hasan, M.H., Aziz, N. and Bamhdi, A., 2023. Deep Neuro-Fuzzy System application trends, challenges, and future perspectives: A systematic survey. Artificial intelligence review, 56(2), pp.865-913.

[14] Estrada, D., Hou, W., Matt, S. and Ouyang, B., 2023. Multi-frame image fusion using a machine learning-based weight mask predictor for turbulence-induced image degradation. Journal of Applied Remote Sensing, 17(1), pp.016514-016514.

[15] Zhang, C., Xue, B., Zhou, F. and Xiong, W., 2018. Removing atmospheric turbulence effects in unified complex steerable pyramid framework. IEEE Access, 6, pp.75855-75867.

[16] Xie, Y., Zhang, W., Tao, D., Hu, W., Qu, Y. and Wang, H., 2016. Removing turbulence effect via hybrid total variation and deformation-guided kernel regression. IEEE Transactions on Image Processing, 25(10), pp.4943-4958.

[17] Nawreen, F., Halder, K.K., Tahtali, M. and Anavatti, S.G., 2023. Stabilization of turbulence-degraded video using patch-based reference frame. Optics Continuum, 2(12), pp.2484-2499.

[18] Anantrasirichai, N., Achim, A., Kingsbury, N.G. and Bull, D.R., 2013. Atmospheric turbulence mitigation using complex wavelet-based fusion. IEEE Transactions on Image Processing, 22(6), pp.2398-2408.

[19] Khowaja, S.A., Yahya, B.N. and Lee, S.L., 2021. Cascaded and Recursive ConvNets (CRCNN): An effective and flexible approach for image denoising. Signal Processing: Image Communication, 99, p.116420.

[20] Li, Y., Zhang, Y., Timofte, R., Van Gool, L., Tu, Z., Du, K., Wang, H., Chen, H., Li, W., Wang, X. and Hu, J., 2023. NTIRE 2023 challenge on image denoising: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 1904-1920).

[21] Lau, C.P., Souri, H. and Chellappa, R., 2020, November. Atfacegan: Single face image restoration and recognition from atmospheric turbulence. In 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (pp. 32-39). IEEE.

[22] Anantrasirichai, N., 2023. Atmospheric turbulence removal with complex-valued convolutional neural network. Pattern Recognition Letters, 171, pp.69-75.

[23] Rai, S.N. and Jawahar, C.V., 2022. Removing atmospheric turbulence via deep adversarial learning. IEEE Transactions on Image Processing, 31, pp.2633-2646.

[24] Jiang, W., Boominathan, V. and Veeraraghavan, A., 2023. Nert: Implicit neural representations for unsupervised atmospheric turbulence mitigation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4235-4242).

[25] Jaiswal, A., Zhang, X., Chan, S.H. and Wang, Z., 2023. Physics-driven turbulence image restoration with stochastic refinement. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 12170-12181).

[26] Khowaja, S.A. and Lee, S.L., 2020. Semantic image networks for human action recognition. International Journal of Computer Vision, 128(2), pp.393-419.

[27] Wan, T., Canagarajah, N. and Achim, A., 2009. Segmentation-driven image fusion based on alpha-stable modeling of wavelet coefficients. IEEE Transactions on Multimedia, 11(4), pp.624-633.

[28] Lewis, J.J., O'callaghan, R.J., Nikolov, S.G., Bull, D.R. and Canagarajah, C.N., 2004, June. Region-based image fusion using complex wavelets. In Proc. 7th International Conference on Information Fusion (Vol. 1, pp. 555-562).

[29] Li, D., 2009. Suppressing atmospheric turbulent motion in video through trajectory smoothing. Signal Processing, 89(4), pp.649-655.

[30] Ling, P., Chen, H., Tan, X., Jin, Y. and Chen, E., 2023. Single image dehazing using saturation line prior. IEEE Transactions on Image Processing.

[31] Khowaja, S.A., Khuwaja, P. and Ismaili, I.A., 2019. A framework for retinal vessel segmentation from fundus images using hybrid feature set and hierarchical classification. Signal, Image and Video Processing, 13, pp.379-387.

[32] Li, Z., Murez, Z., Kriegman, D., Ramamoorthi, R. and Chandraker, M., 2018, March. Learning to see through turbulent water. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 512-520). IEEE.

[33] Zhang, Y., Tian, Y., Kong, Y., Zhong, B. and Fu, Y., 2018. Residual dense network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2472-2481).

[34] Wu, Y. and He, K., 2018. Group normalization. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).

[35] Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E., 2020. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 42(08), pp.2011-2023.

[36] Wu, W., Liu, S., Xia, Y., and Zhang, Y., 2024. Dual Residual Attention Network for Image Denoising. Pattern Recognition, 149, pp. 110291.

[37] Zhang B., Sun J., Sun F., Wang F., Zhu B., 2024. Image Deblurring Method based on Self-Attention and Residual Wavelet Transform. Expert Systems with Applications, 244, pp. 123005.

[38] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).

[39] Radford, A., Metz, L. and Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

[40] Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z. and Smolley, S.P., 2018. On the Effectiveness of Least Squares Generative Adversarial Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(12), pp.2947-2960.

[41] Khowaja, S.A., Mujtaba, G., Yoon, J. and Lee, I.H., 2023. Face-PAST: Facial Pose Awareness and Style Transfer Networks. arXiv preprint arXiv:2307.09020.

[42] Arbelaez, P., Maire, M., Fowlkes, C. and Malik, J., 2010. Contour detection and hierarchical image segmentation. IEEE transactions on pattern analysis and machine intelligence, 33(5), pp.898-916.

[43] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

[44] Ma, K., Duanmu, Z., Wu, Q., Wang, Z., Yong, H., Li, H. and Zhang, L., 2016. Waterloo exploration database: New challenges for image quality assessment models. IEEE Transactions on Image Processing, 26(2), pp.1004-1016.

[45] Mao, Z., Jaiswal, A., Wang, Z. and Chan, S.H., 2022, October. Single frame atmospheric turbulence mitigation: A benchmark study and a new physics-inspired transformer model. In European Conference on Computer Vision (pp. 430-446). Cham: Springer Nature Switzerland.

[46] Anantrasirichai, N., Achim, A. and Bull, D., 2018, October. Atmospheric turbulence mitigation for sequences with moving objects using recursive image fusion. In 2018 25th IEEE international conference on image processing (ICIP) (pp. 2895-2899). IEEE.

[47] Li, N., Thapa, S., Whyte, C., Reed, A.W., Jayasuriya, S. and Ye, J., 2021. Unsupervised non-rigid image distortion removal via grid deformation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 2522-2532).

[48] Jin, D., Chen, Y., Lu, Y., Chen, J., Wang, P., Liu, Z., Guo, S. and Bai, X., 2021. Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning. Nature Machine Intelligence, 3(10), pp.876-884.

[49] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L. and Timofte, R., 2021. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1833-1844).

[50] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S. and Yang, M.H., 2022. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5728-5739).

[51] Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J. and Li, H., 2022. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 17683-17693).

[52] Mehri, A., Ardakani, P.B. and Sappa, A.D., 2021. MPRNet: Multi-path residual network for lightweight image super resolution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 2704-2713).

[53] Yasarla, R. and Patel, V.M., 2021, September. Learning to restore images degraded by atmospheric turbulence using uncertainty. In 2021 IEEE International Conference on Image Processing (ICIP) (pp. 1694-1698). IEEE.

[54] J. Liang et al., "VRT: A Video Restoration Transformer," in IEEE Transactions on Image Processing, vol. 33, pp. 2171-2182, 2024, doi: 10.1109/TIP.2024.3372454.

[55] X. Zhang, Z. Mao, N. Chimitt and S. H. Chan, "Imaging Through the Atmosphere Using Turbulence Mitigation Transformer," in IEEE Transactions on Computational Imaging, vol. 10, pp. 115-128, 2024, doi: 10.1109/TCI.2024.3354421.

[56] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 586-595, doi: 10.1109/CVPR.2018.00068.

[57] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik and M. K. Markey, "Complex Wavelet Structural Similarity: A New Image Similarity Index," in IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2385-2401, Nov. 2009, doi: 10.1109/TIP.2009.2025923.

[58] Z. Ji, H. Zheng, Z. Zhang, Q. Ye, Y. Zhao and M. Xu, "Multi-Scale Interaction Network for Low-Light Stereo Image Enhancement," in IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 3626-3634, Feb. 2024, doi: 10.1109/TCE.2023.3280229.

[59] C. -H. Son and H. -M. Park, "A pair of noisy/blurry patches-based PSF estimation and channel-dependent deblurring," in IEEE Transactions on Consumer Electronics, vol. 57, no. 4, pp. 1791-1799, November 2011, doi: 10.1109/TCE.2011.6131155.

[60] N. Yuan, J. Li and B. Sun, "Consumer Camera Demosaicking and Denoising With a Collaborative Attention Fusion Network," in IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 509-521, Feb. 2024, doi: 10.1109/TCE.2023.3342035.