Thermal Infrared Single Image Dehazing and Blind Image Quality Assessment

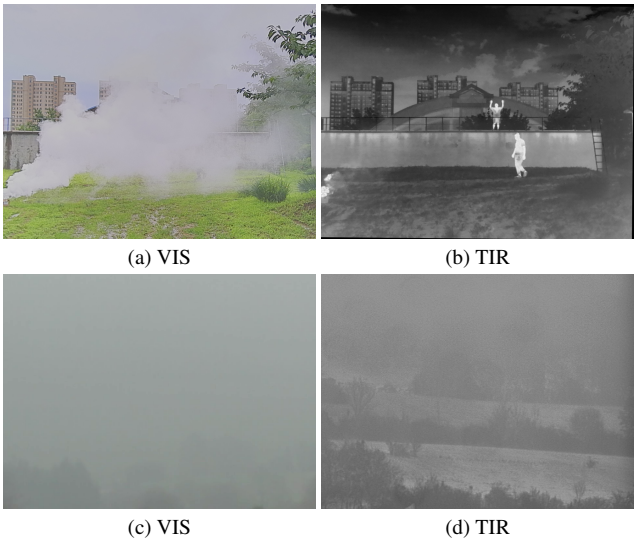
Fabian Erlenbusch[1,2], Constanze Merkt[1,3], Bernardo de Oliveira[1], Alexander Gatter[4], Friedhelm Schwenker[3], Ulrich Klauck[2], Michael Teutsch[1]

[1] Hensoldt Optronics GmbH, Germany, [2] Aalen University of Applied Sciences, Germany, [3] Ulm University, Germany, [4] Rheinmetall Electronics GmbH, Germany

fabian.erlenbusch@hs-aalen.de, {constanze.merkt, michael.teutsch}@hensoldt.net

Abstract

Image dehazing is a method to reduce the effects of haze, dust, or fog in images in order to provide a clear view of the observed scene. A large variety of traditional and machine learning-based approaches exists in the literature. However, these approaches mostly consider color images in the visual-optical spectrum. Apparently, the thermal infrared spectrum is much less affected by haze due to its longer wavelength. But atmospheric perturbations during long-range observation can cause image quality degradation in the thermal infrared (TIR) spectrum as well. In this paper, we propose a method to generate synthetic haze for TIR images. Then, we analyze the already existing blind image quality assessment measure Fog Aware Density Evaluator (FADE) for its applicability to the TIR spectrum. We further provide a comprehensive overview of the current state-of-the-art in image dehazing and empirically show that many approaches originally designed for visual-optical images perform surprisingly well when applied to the TIR spectrum. This is shown in experiments conducted on the recently published M3FD dataset.



(a) VIS (b) TIR

(c) VIS (d) TIR

Figure 1. Haze visibility is shown across two spectra. Smaller aerosol particles originating from smoke are clearly visible in the VIS image, but mitigated by a TIR camera (a,b) [28]. Heavy fog with larger particles affects both spectra (c,d).

1. Introduction

Haze and related atmospheric perturbations such as smoke, dust, fog, mist, rain, or snow are natural phenomena, in which (hydrated) aerosol particles of different sizes are suspended in the air [7, 19]. Light beams traveling through the air are absorbed, causing attenuation, and scattered, causing diffusion (also called airlight). This results in reduced contrast, fainted surfaces, color shift, opalescence, obscured clarity, and intensity blur, diminishing the visibility and the object perception capability [41, 47]. The performance of computer vision algorithms such as object detection, tracking, and segmentation is usually affected by the presence of haze [26, 42]. Hence, image dehazing algorithms can be used to provide a clear view of the scene supporting both machine and human. However, this task is challenging since the modeling of haze is dependent on the often unknown scene depth, the density of the atmospheric particles, the particle size, and the wavelength [33, 40]. Furthermore, the introduction of artifacts by the dehazing process such as color distortion, over-enhancement, halo, or blur must be avoided [1].

The visible haze effect of such aerosol particles is different dependent on the captured spectrum: shorter wavelengths in the Visual-optical (VIS) or Near-Infrared spectrum are more affected than longer wavelengths such as Thermal Infrared (TIR) [19]. An example is given in Fig. 1. There can be a range difference of up to 500 m in visibility between VIS and long-wave TIR [39]. However, rather

large particles like water droplets or ice crystals do affect the TIR spectrum. This leads to a degradation in visual image quality [22] and computer vision algorithm performance [23] that can be compared to the ones in the VIS spectrum. Hence, dehazing algorithms suitable for single-channel TIR images are needed, especially because most existing literature focuses on VIS images [52].

A large variety of traditional and machine learning-based approaches is discussed in this literature with deep neural networks as the current state-of-the-art technique [25, 26, 48]. Nowadays, experiments are usually conducted utilizing public datasets that are affected by artificially generated synthetic haze to provide ground-truth (GT) for supervised learning [26, 45]. Quantitative results are calculated using (1) full-reference Image Quality Assessment (IQA) measures such as Peak Signal-to-Noise Ratio (PSNR) or Structural Similarity (SSIM), as well as (2) no-reference *blind* IQA measures such as Natural Image Quality Evaluator (NIQE) [36] or Fog Aware Density Evaluator (FADE) [6].

Inspired by this evaluation approach, the main goal of this work is to investigate the applicability and the performance of traditional and deep learning-based dehazing algorithms assessed with TIR images. The recently published multi-spectral M3FD dataset [28] is used to create images with synthetic haze in both the VIS and the TIR spectrum for systematic comparison. This dataset is used to empirically show that FADE generalizes well on the TIR spectrum and thus can be applied as a blind IQA measure. To conduct the experiments, we leverage knowledge taken from the VIS domain by starting with color images that we convert to single-channel, monochromatic VIS images as an intermediate step, and then transfer the gathered information to the final single-channel TIR images. After that, we utilize traditional and deep learning-based dehazing techniques proven in the VIS spectrum to provide an extensive study on their performance when applied to TIR images. Additionally, a test with a deep learning-based object detector was performed to compare its behavior on hazy and haze-free TIR images.

Our contributions are (1) the generation of synthetic haze in the TIR spectrum based on depth images calculated in the VIS spectrum using the Multi-scenario Multi-Modality Benchmark to Fuse TIR and VIS for Object Detection (M3FD) [28], (2) the experimental analysis of the applicability of the blind IQA measure FADE to TIR images, and (3) a comprehensive analysis of traditional and machine learning-based methods applied to TIR images.

The paper is organized as follows: related work is presented in Section 2. Our proposed method for the generation of synthetic haze in IR images is presented in Section 3. The applicability of the blind IQA measure FADE to TIR images is analyzed in Section 4. Experimental results are described in Section 5. We conclude in Section 6.

2. Related Work

Dehazing: the Atmospheric Scattering Model (ASM) [32, 33] is the key foundation for most dehazing approaches, describing a hazy image $I(x)$ as $I(x) = J(x)t(x) + A(1 - t(x))$ with $J(x)$ being the haze-free image, $t(x) = e^{-\beta d(x)}$ being the transmission matrix with the atmosphere scattering coefficient β, and A being the airlight. Many dehazing methods exploit ASM by estimating $t(x)$, A, and $J(x)$, using prior- [4, 14, 50, 51] or learning-based approaches [5, 25, 29, 48], being the last the most prominent currently. Some end-to-end learning-based methods incorporate ASM, directly minimizing the reconstruction error on the haze-free image [25]. Alternatively, image enhancement methods are also employed [40, 41] especially in the gray-valued TIR spectrum [52].

Surveys: many surveys [1, 13, 26, 41, 47] investigate dehazing in the VIS spectrum from various aspects such as the dehazing algorithms, relevant datasets, or suitable evaluation measures. Some authors propose modified versions of prior-based algorithms for TIR spectrum [9, 55, 59] without properly comparing the performance to other dehazing techniques. Al Mansoori *et al*. [31] is the only identified survey, which investigated both prior-based and learning-based dehazing algorithms in the TIR spectrum. However, important information is missing such as fine-tuning or transfer-learning of the deep learning techniques to the TIR spectrum. Furthermore, just private data was used for the experiments. Therefore this work investigates the performance of various promising dehazing techniques in the TIR spectrum including image enhancement, prior-based dehazing, and learning-based dehazing.

Creation of hazy datasets: multiple authors [13, 26, 45, 46] provide relevant information on the creation of synthetic hazy images from non-hazy real images in the VIS spectrum. However, nearly no literature exists, in which the generation of synthetic haze in the TIR spectrum is discussed. Yan *et al*. [55] generate synthetic haze in the TIR spectrum using pix2pix image translation [15]. This approach can be trained to map scene appearance in the VIS spectrum to images in the TIR spectrum. In the past, this method was used to generate large datasets for learning single object tracking in the TIR spectrum [58] or multi-modal person re-identification [21]. Following a similar approach, Yan. *et al*. [55] map hazy VIS images to hazy TIR images, but the results are not convincing as haze severely disturbs the already hard problem of VIS-TIR image translation. Therefore this work proposes an improved approach for creating the synthetic haze in the TIR spectrum by leveraging the ASM similar to the proposed approach of Li *et al*. [26] in the VIS spectrum. The lack of training data in the TIR spectrum may affect supervised deep learning for dehazing, but transfer learning can provide a solution to this problem by leveraging the prior training on VIS images and fine-tuning

the model on a smaller volume of TIR images [11].

Blind IQA: early successful approaches for blind IQA extract image features using Natural Scene Statistics (NSS) [37]. The basic idea is to model scene statistics of visually appealing images and learn to distinguish between distorted and good-quality images. Different variants of features and related machine-learning approaches exist in the literature. Among the most prominent ones are: (1) NIQE [36], (2) BLind Image Integrity Notator using DCT Statistics (BLIINDS-II) [44], (3) Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [35], (4) Perception based Image QUality Evaluator (PIQUE) [38], or (5) Blockiness and Luminance Change (BALC) [57]. They all do not measure haze specifically but image quality in general. Deep learning-based methods came up recently that also measure rather a generic image quality [49, 56]. However, only very little literature is available for blind IQA in dehazing [6, 17, 34] and this related work considers the VIS spectrum only. Some authors measure rather generic image distortions such as color shift, contrast, or sharpness and utilize them to determine the level of haze [17, 34]. Such image properties cannot be transferred from the VIS to the TIR spectrum easily. However, there is one approach called FADE that is based on NSS [6]. Since NSS features do generalize quite well across the spectra [12], we utilize FADE for our studies in this paper.

3. Dataset Preparation

In this section, we first describe the datasets used in this paper and then we present the proposed approach to generate synthetic haze in the TIR spectrum.

3.1. Description of the Datasets

The datasets chosen to assess the dehazing performance of the tested methods are: (1) O-HAZE [2], (2) NH-HAZE [2], (3) the Synthetic Objective Testing Set of the Realistic Single Image Dehazing (RESIDE SOTS) [26], and (4) M3FD [28]. Table 1 shows their default characteristics. The first three were selected since they contain images from the VIS spectrum in several different scenarios with real or synthetic haze in different heaviness.

Only the multi-spectral M3FD dataset has TIR images, which are pixel-precisely aligned to the VIS images. However, M3FD is no dehazing dataset. Hence, we add synthetic haze as described in Section 3.2, which is common practice in the field [26, 45]. In the context of O-HAZE and NH-HAZE, the haze was created by a haze-generation machine [3]. For RESIDE, the haze was created using ASM with random values in predefined ranges for A and β, after calculating the depth field using a monocular depth estimator [26]. Due to limited memory, O-HAZE images were downscaled to 25% of their original size. Single-channel TIR images are represented as three-channel arrays to directly feed them to deep neural networks pre-trained for RGB images. As an additional test, all datasets in the VIS spectrum were converted to three-channel gray-scale images to investigate the impact of monochromaticity on the dehazing methods.

		Original datasets			
		I	II	III	IV
Number of images		45	55	500	4,200
Spectrum	VIS	✓	✓	✓	✓
	TIR				✓
Haze	No haze				✓
	Real	✓	✓		
	Synthetic			✓	
	Homogeneous	✓			
	Non-homogeneous		✓	✓	

Table 1. The default characteristics of the datasets employed in the experiments of this work: (I) O-HAZE, (II) NH-HAZE, (III) RESIDE SOTS, and (IV) M3FD.

Class	Train	Validation	Test
People	9,216	1,192	1,069
Car	14,467	1,939	1,890
Other Vehicle	1,382	153	173

Table 2. Number of class instances per split in M3FD.

Regarding the object detection task mentioned in Section 1, M3FD was split: (1) $0.9/0.1$ (3780 / 420 images) for DehazeFormer [48] and All-in-One Dehazing Network (AOD-Net) [25] that we use in Section 5; and (2) $0.8/0.1/0.1$ (3360 / 420 / 420 images) for the task of object recognition, being the testing sets the same and the training set in (1) equal to the training set plus the validation set in (2). The dataset classes were organized into three classes: (1) 'person', (2) 'car', and (3) 'other vehicles' ('bus' and 'truck' classes). The classes 'motorcycle' and 'lamp' were not numerous enough for being employed in this work. Table 2 shows the resulting number of instances of each class. Note that the number of images containing a certain class is quite balanced, at least for the classes 'person' and 'car'.

3.2. Adding Synthetic Haze to Real TIR Images

The simplest approach for creating synthetic hazy images from non-hazy images in the VIS spectrum is by employing ASM. Therefore, we need the image and its related depth map. The first step is to estimate the depth $d(x)$ regarding the input image $J(x)$ using a monocular depth estimator $E(x)$. β is set accordingly to the chosen strength of scattering. Similar to β, A needs to be set to the desired color and intensity. The hazy image $I(x)$ and $t(x)$ can be obtained using ASM (see Section 2) [26].

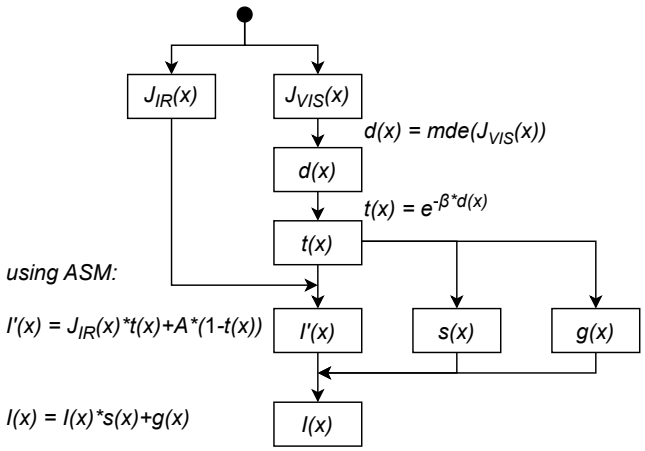


Figure 2. Proposed approach for creating synthetic hazy images for the TIR spectrum from non-hazy images. The monocular depth estimator (mde) is adopted from [61].

The novel approach of this work for adding synthetic haze to real TIR images requires only a few additional steps. An overview is given in Fig. 2. We start with the haze-free aligned images J_{VIS} and J_{TIR}. The depth map from the VIS image is utilized. We use the state-of-the-art method DIFFNet [61] for monocular depth estimation (mde) in the VIS spectrum. $t(x)$ can then be obtained by setting β to the desired strength of scattering. After selecting A, the hazy image $I_{\text{TIR}}(x)$ can be computed using ASM and the input TIR image J_{TIR}. Adding noise ensures that the generation is more realistic since the presence of haze usually results in a lower signal-to-noise ratio and thus noise becomes more apparent. We do not consider fixed-pattern noise here since this is usually handled by Non-Uniformity Correction (NUC) [30]. Inspired by several discussion on modeling noise in the TIR spectrum [16, 20], we propose a combined use of speckle noise $s(x)$ and Gaussian noise $g(x)$. While the additive noise is subject to a Gaussian distribution with zero mean and a variance of 0.025, the multiplicative speckle noise is subject to a Rayleigh distribution [27].The speckle noise is then injected multiplicatively and the Gaussian noise is injected additively to the hazy image as seen in Fig. 2.

To avoid halo effects that appear in the images after adding synthetic haze, the Gaussian blur was applied as a post-processing step. For VIS images, β was chosen between [0.6, 1.8] using a uniform distribution and A between [0.7, 1.0] also in an uniform distribution. Figure 3 shows some created samples using the described approach and parameters. For TIR, β was chosen between [0.6, 1.8] using a uniform distribution and A from a uniform distribution whose lower bound was 0.1765, derived from the best results for dehazing TIR images in the work of Fang et al. [9], while the upper bound was 0.588, derived from the most

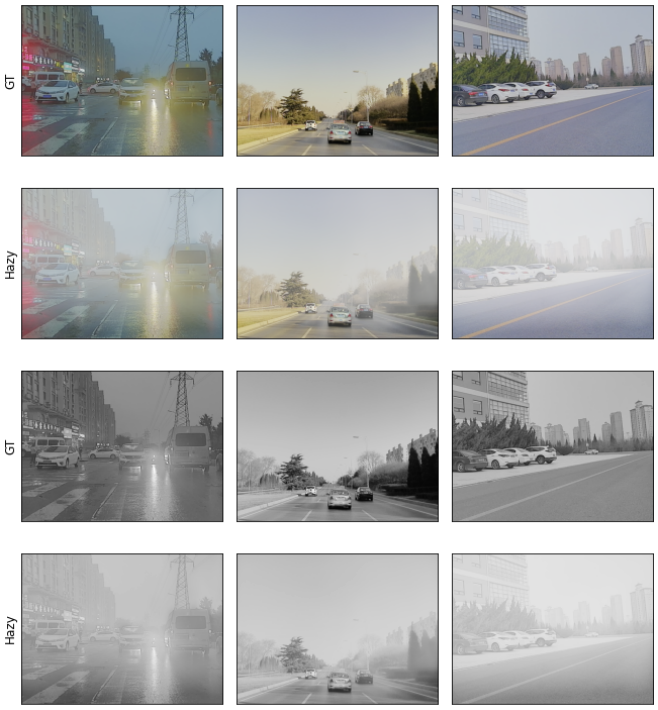


Figure 3. Samples from M3FD with respective synthetic haze in the VIS spectrum (colored/monochrome).

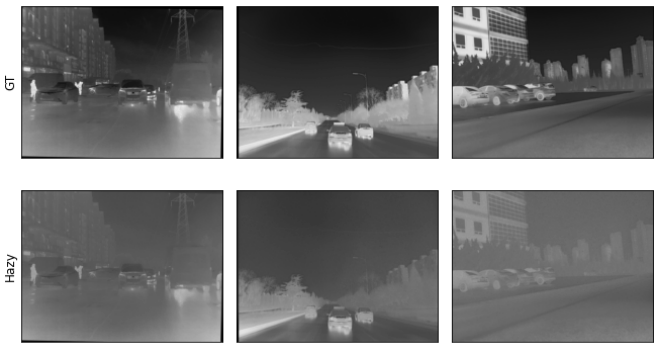


Figure 4. Samples from M3FD with respective synthetic haze in the TIR spectrum.

common value of the histogram of a hazy TIR image in the work of Zheng et al. [59]. Figure 4 illustrates a few samples using the described approach and parameters for the TIR spectrum.

4. Blind Image Quality Assessment

To the best of our knowledge, the only blind IQA measure in the literature specifically designed for dehazing is FADE [6]. It is based on NSS and was originally introduced for VIS imagery. Since then, several authors adopted it for quantitative evaluations comparing state-of-the-art approaches [1, 10, 24]. It has never been tested on TIR im-

agery, yet. Instead, existing literature for TIR image dehazing relies on full-reference IQA measures such as PSNR or SSIM [31] just like most of the related work in the VIS spectrum [25, 26, 40, 48]. We now analyze the FADE measure for its applicability to both the VIS and the TIR spectrum.

This analysis is conducted as follows: we use the four public datasets O-HAZE [2], NH-HAZE [3], RESIDE [26], and M3FD [28]. The three pure VIS datasets O-HAZE, NH-HAZE, and RESIDE come with GT for dehazing, *i.e.*, image pairs are provided with a clear image I_{gt} and a hazy image I_{hazy} for each scene. We generate such image pairs for M3FD using the approach described in Section 3.2 to create synthetic haze. A higher FADE value indicates a higher level of haze. Hence, FADE follows the principle *the lower the better*. We can assume that if the measure works as expected, a hazy image should produce a higher FADE value than a clear image. To quantify this, we calculate the difference in the FADE values for given image pairs:

$$\Delta\text{FADE}(I_{hazy}, I_{gt}) = \text{FADE}(I_{hazy}) - \text{FADE}(I_{gt}) \quad (1)$$

According to the assumption, we can expect that a value of ΔFADE larger than zero indicates a correct behavior of FADE as a blind IQA measure for dehazing. Within each dataset, we calculate ΔFADE for each image pair and average the results. ΔFADE follows the principle *the higher the better*. The results for the considered datasets are shown in Fig. 5 top left. We split M3FD in VIS and TIR. Furthermore, we split M3FD TIR in *with* and *without* artificially injected noisy since NSS-based IQA measures are quite sensitive to noise and thus the experiment may be biased towards the artificial noise. Hence, we receive six different datasets in the plot. The colored dots represent the mean value and the vertical lines represent the standard deviation within each dataset. The averaged ΔFADE values for all datasets are consistently above zero. The artificial noise affects the FADE measure quite strongly, but still the ΔFADE value is clearly above zero. So, it seems that FADE works well as a blind IQA measure for dehazing. Please note that since FADE is based on NSS [6] and the synthetic haze generation is based on ASM [45], *i.e.*, different mathematical foundations, we do not expect any undesirable correlation or stochastic dependence in our experiment here.

The plots in Fig. 5 show that this desired behavior is achieved by FADE only. We compare FADE with other blind IQA measures not specifically designed for dehazing: NIQE [36], PIQUE [38], BRISQUE [35], BLIINDS-II [44], and BALC [57]. Since these measures follow the principle *the lower the better*, we calculate their Δ-variants analogously to FADE. The plots show that ΔFADE is the only measure consistently above zero. Hence, we conclude that FADE can be a suitable blind IQA measure for dehazing in both the VIS and the TIR spectrum. Note that we introduced ΔFADE only to analyze the usefulness of FADE.

5. Experiments and Results

5.1. Algorithms

Six dehazing algorithms were selected: (1) Contrast Limited Adaptive Histogram Equalization (CLAHE) [62], (2) Multi-Scale Retinex with Chromaticity Preservation (MSRCP) [43], (3) Dark Channel Prior (DCP) [14], (4) Fast Visibility Restoration (FVR) [51], (5) AOD-Net [25], and (6) DehazeFormer [48]. CLAHE and MSRCP are prominent image enhancement approaches already employed for dehazing [52, 54]; DCP and FVR present well-known prior-based dehazing techniques; AOD-Net is a notorious light-weighted, real-time learning-based technique [25, 60]; and DehazeFormer, specifically the DehazeFormer-b, recently outperformed most previous state-of-the-art methods [48].

5.2. Experimental Setup

A grid search was implemented to optimize the parameters of CLAHE, DCP, and FVR for each dataset and spectrum, using SSIM as optimization criterion. For CLAHE, a tile size of 8 by 8 was set. The clip limit was chosen by the grid search, being limited in the range of $[1, 10]$, with a step size of 1. CLAHE was also applied channel-wise. For DCP, a patch size of 15 and a restriction of the lower bound of 0.15 was chosen, while ω was limited in the range of $[0.3, 0.9]$, with a step size of 0.1. For FVR, the window size s_v was set to 61, while p was limited in the range of $[0.3, 0.9]$, with a step size of 0.1.

The deep learning models were fine-tuned on the TIR images of the M3FD dataset with added synthetic haze. For both models, we utilize the pre-trained weights for dehazing in the VIS spectrum. The 0.9/0.1 train-validation-test-split was used as mentioned in Section 3.2. The default training strategy of the AOD-Net implementation was used to fine-tune the AOD-Net. Adam was used as an optimizer with a learning rate of 0.0001. A batch size of 8 was used to train the model for 30 epochs. Our training strategy for DehazeFormer-b strictly follows the authors' training on the outdoor RESIDE training dataset [48].

The Yolov5m model [18] was used for the M3FD TIR dataset in the object detection task. The pre-trained weights from the COCO dataset were taken as a baseline. The model was fine-tuned on M3FD using the non-hazy GT TIR images with the default training strategy and 0.8/0.1/0.1 split.

We use the full-reference IQA measures PSNR and SSIM [53], as well as the blind IQA measure FADE [6] to evaluate the dehazing performance. Additionally, the task-driven metric mean Average Precision (mAP) over all classes [8] was employed for the object detection analysis.

5.3. Experimental Results

The experiments are conducted in three steps: (I) dehazing color and gray-scale VIS images as a reference, (II)

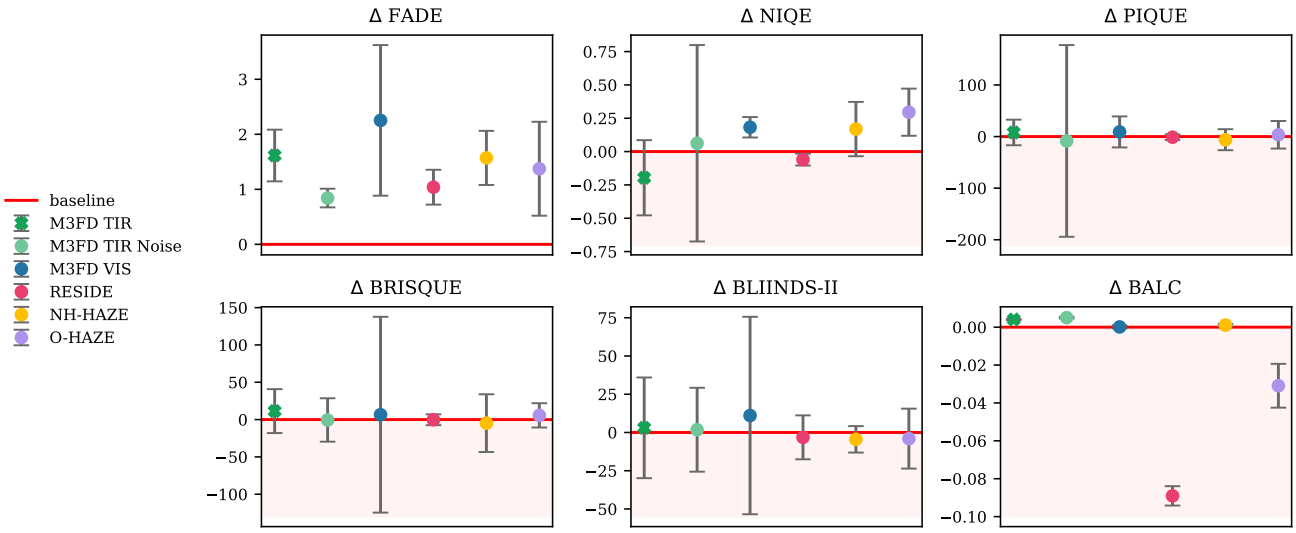


Figure 5. ΔFADE as calculated by Eq. 1 is the only blind IQA measure consistently above zero for all considered public datasets. This indicates that FADE [6] can be a suitable blind IQA measure for dehazing in both the VIS and the TIR spectrum.

dehazing TIR images, and (III) improvement in object detection by using dehazed TIR images as input. SSIM and PSNR are always measured in reference to the clear non-hazy GT image. In addition, we use FADE and mAP as described in Section 5.2.

5.3.1 Dehazing VIS Images

Table 3 shows the dehazing performance for each evaluated approach, each dataset, each IQA measure, and both types of input images, *i.e.* color or gray-scale VIS. Most dehazing methods achieve an improvement compared to the original hazy images according to all three measures. In general, the prior-based methods perform better than the learning-based ones. The results for the gray-value images are comparable and seem to highly correlate with the results for the color images. From these results, it can be deduced that most methods do not seem to be affected by monochromaticity. This is interesting since DCP uses a prior based on multi-channel color images. We expect this to be the result of the duality between DCP and Retinex [10] with the differences in performance originating from different parameterization. FADE generally seems to correlate with image properties such as brightness, saturation and contrast, which could explain the superior performance of MSRCP and AOD-Net.

The results on NH-HAZE show worse performance in both SSIM and PSNR compared to O-HAZE. This is probably due to NH-HAZE having a more dense and less homogeneous haze compared to O-HAZE. Simple image enhancement techniques like CLAHE, FVR and MSRCP seem to be superior in such hazy environments. The performance of the learning-based techniques could be explained by the fact that both O-HAZE and NH-HAZE are not very ASM-*conform*, meaning that haze strength and distribution are not as dependent on the scene depth.

Differently to the experiments on the O-HAZE and NH-HAZE datasets, various dehazing methods are able to almost fully restore the clear images from the hazy images in the RESIDE SOTS dataset. This is most likely the case because the RESIDE SOTS dataset is more ASM-conform concerning both the dependence on haze strength and the distribution of the scene depth. DehazeFormer was trained on and optimized for this dataset, which is expected to be the reason for its outstanding performance. The results on M3FD seem to be somewhat similar to those on RESIDE SOTS, arguably because the haze in both datasets was synthetically created using a similar technique. DehazeFormer is not able to significantly outperform the other techniques, scoring the third-best performance after DCP and FVR, presumably because it was not fine-tuned for this specific dataset. Again, no negative effect of monochromaticity on the various techniques can be observed.

Even though FADE was designed for VIS imagery [6], we cannot see a clear correlation between FADE and the full-reference IQA measures SSIM and PSNR in this experiment. Furthermore, Table 3 shows that FADE does not seem to generalize across datasets.

5.3.2 Dehazing TIR Images

Table 4 shows the dehazing performance on M3FD TIR. Most methods achieve an improvement in image quality. It can also be observed that fine-tuning the deep learning models is highly necessary. Specifically notable is that the

Method	SSIM ↑		PSNR ↑		FADE ↓	
	color	gray	color	gray	color	gray
O-HAZE						
None	0.576	0.591	13.77	14.63	1.919	2.573
CLAHE	0.677	0.697	14.11	14.90	1.291	1.860
MSRCP	0.605	0.618	12.02	12.48	**0.518**	**0.769**
DCP	0.685	**0.718**	16.94	**18.74**	0.681	1.142
FVR	**0.697**	0.717	**17.09**	18.02	0.621	1.104
AOD	0.554	0.603	14.98	16.01	0.535	0.938
DF	0.640	0.689	16.12	18.49	0.762	0.789
NH-HAZE						
None	0.432	0.445	11.40	2.020	1.844	2.403
CLAHE	**0.552**	**0.571**	11.89	12.39	1.187	1.671
MSRCP	0.511	0.524	10.62	11.03	0.650	0.904
DCP	0.493	0.514	13.01	13.75	0.636	1.068
FVR	0.519	0.532	**13.17**	**13.76**	0.549	0.993
AOD	0.409	0.434	11.85	12.40	**0.496**	**0.836**
DF	0.465	0.498	12.11	13.20	1.237	1.250
RESIDE SOTS						
None	0.813	0.824	15.92	16.03	1.762	2.025
CLAHE	0.842	0.856	16.44	16.66	1.121	1.401
MSRCP	0.785	0.800	14.86	15.09	**0.644**	0.803
DCP	0.919	0.930	21.63	20.72	0.746	0.976
FVR	0.912	0.927	21.51	21.48	0.810	1.058
AOD	0.892	0.908	19.55	19.66	0.659	**0.770**
DF	**0.990**	**0.988**	**38.58**	**33.80**	0.732	0.811
M3FD VIS						
None	0.851	0.854	15.75	15.94	3.190	3.335
CLAHE	0.899	0.902	16.75	17.03	2.103	2.400
MSRCP	0.799	0.800	14.20	14.48	**1.196**	**1.379**
DCP	**0.936**	**0.937**	19.96	19.75	1.587	1.989
FVR	0.931	0.930	19.33	19.22	1.524	1.854
AOD	0.893	0.894	17.51	17.42	1.314	1.482
DF	0.928	0.926	**20.19**	**20.05**	1.923	2.067

Table 3. Dehazing performance on each considered dataset. *None* indicates that no dehazing was performed. AOD stands for AOD-Net and DF stands for DehazeFormer. AOD-Net and DehazeFormer were originally trained with VIS data and we did not conduct any fine-tuning here.

fine-tuned DehazeFormer outperforms all other metrics by a large margin. Figure 6 shows that while most methods seem to improve the contrast slightly, the fine-tuned DehazeFormer improves the image and especially the sky region the most. FADE reacts in an unexpected way: according to Table 4, AOD-Net† performs best, although the qualitative evaluation in Fig. 6 definitely contradicts this observation. This behavior could be explained by its focus on both image brightness and contrast. As seen in Fig. 5, the artificially injected noise in the TIR images leads to better FADE values. This could be explained by noise introducing pseudo-contrast to the image. The good FADE value of the noise-amplifying MSRCP approach further confirms

Method	SSIM ↑	PSNR ↑	FADE ↓
None	0.839	19.01	2.448
CLAHE	0.839	19.23	1.877
MSRCP	0.609	13.82	1.050
DCP	0.861	19.01	1.931
FVR	0.855	19.27	1.713
AOD-Net†	0.601	12.97	**0.857**
AOD-Net*	0.848	19.65	1.836
DehazeFormer†	0.755	16.14	1.042
DehazeFormer*	**0.945**	**28.04**	1.778

Table 4. Average performance on M3FD TIR. *None* indicates that no dehazing was performed. The learning-based method with a † has the default training and with * has been trained with TIR data.

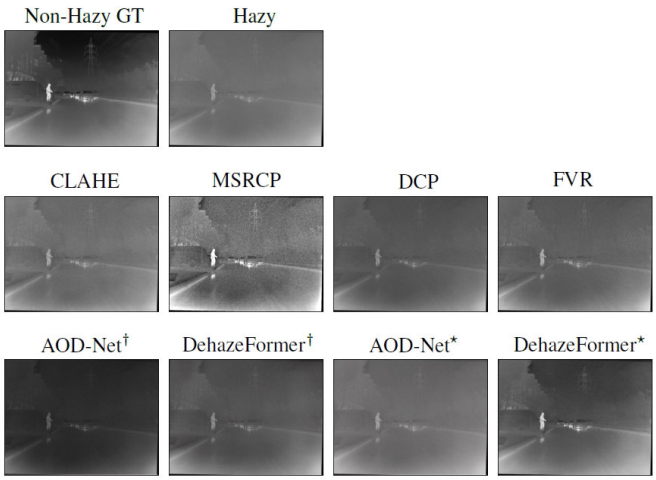


Figure 6. Image samples taken from M3FD TIR after the use of dehazing techniques. The learning-based method with a † has the default training and with * has been trained with TIR data.

this assumption. The adaption of FADE and the related NSS to the TIR spectrum [12] should be investigated.

In addition to the presented experiments, the images from M3FD were grouped according to the haze levels that were synthetically added and individual tests were performed for each haze level. Table 5 shows the best-performing techniques on M3FD TIR for each metric and type of input image for different levels of haze. We can observe that the haze level negatively affects the performance of all dehazing techniques. FADE again reacts in an unexpected way to the introduction of artificial noise: heavier haze improves the FADE values, which could be the result of pseudo image contrast introduced by noise. Most notable is the excellent and robust performance of the fine-tuned DehazeFormer even for heavy haze. However, it should be noted that fine-tuning was performed within the dataset.

Figure 6 shows such a sample with a heavy haze level. In the hazy image, it is hard to identify any objects except for

Method	SSIM ↑		PSNR ↑		FADE ↓	
	Light	Heavy	Light	Heavy	Light	Heavy
None	0.895	0.784	21.29	17.09	2.561	2.257
CLAHE	0.903	0.770	21.42	17.32	1.992	1.673
MSRCP	0.706	0.502	13.78	14.0	1.296	0.781
DCP	0.920	0.803	20.69	17.57	2.053	1.743
FVR	0.922	0.785	21.29	17.72	1.812	1.528
AOD†	0.598	0.592	13.08	12.86	**0.942**	**0.725**
AOD*	0.896	0.794	22.26	17.42	1.871	1.713
DF†	0.788	0.723	17.11	15.4	1.148	0.937
DF*	**0.972**	**0.919**	**29.83**	**26.8**	1.745	1.756

Table 5. Average performance on M3FD TIR with different haze levels. *None* indicates that no dehazing was performed. AOD stands for AOD-Net and DF stands for DehazeFormer. Each deep learning-based method with a † has been trained with VIS data and with a * with TIR data.

one person. Most techniques are only marginally able to remove the haze. MSRCP performs well in qualitative terms, while the fine-tuned DehazeFormer performs well in both qualitative and quantitative terms. The MSRCP improves the perceptual quality of the image by improving the variation of lightness in local image regions, but it also amplifies the image noise. The fine-tuned DehazeFormer model is the technique that restores the image the best probably due to its strong Transformer backbone.

5.3.3 Dehazed TIR Images for Object Detection

Object detection performance can be used as an indicator of dehazing quality [26]. The last experiment considers using dehazing techniques as a pre-processing step in an object detection task with the pre-trained YOLOv5 object detector fine-tuned on the non-hazy M3FD TIR dataset. The object detector confirms the observations made in the previous experiments, as there is a correlation between dehazing quality and object detection performance. However, while MSRCP improved the perceptual quality, the object detector performs poor on the related dehazed images probably due to the amplified image noise. Table 6 shows the results of the task-driven evaluation for all classes. Considering mAP@.50 and mAP@0.50-.95, the DehazeFormer outperforms all other techniques. The fine-tuned DehazeFormer also outperforms the other methods on each individual class, namely 'people', 'car' and 'other vehicles'. It is able to nearly close the gap to the non-hazy GT.

Figure 7 shows the distance from both the upper (*i.e.*, object detection on non-hazy images) and the lower baseline (*i.e.*, object detection on hazy images) of the evaluation. Dehazing methods between the two lines improve the object detection performance. Most methods do so. An exception is MSRCP: the change in image appearance or the noise amplification seem to disturb the object detection.

	Input	mAP@.5 ↑	mAP@.5-.95 ↑
	Hazy	0.838	0.583
	Non-Hazy GT	**0.915**	**0.645**
After dehazing	DehazeFormer*	**0.902**	**0.630**
	DehazeFormer†	0.866	0.601
	AOD-Net*	0.849	0.593
	AOD-Net†	0.828	0.557
	FVR	0.859	0.597
	DCP	0.860	0.599
	CLAHE	0.853	0.594
	MSRCP	0.813	0.554

Table 6. Object detection performance of YOLOv5 given by mAP using dehazing techniques on M3FD for both clear images and after dehazing. † indicates training on VIS and * on TIR data.

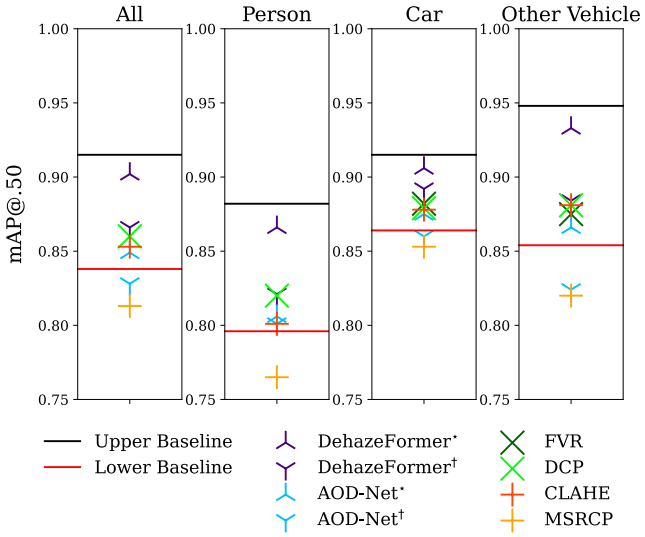


Figure 7. Task-driven evaluation using dehazing as pre-processing step on M3FD for object detection (mAP@.50). The lower and upper baseline are produced by hazy images and the non-hazy GT.

6. Conclusion

We presented an approach to generate synthetic haze in TIR images using aligned VIS-TIR image pairs as provided by the M3FD dataset. We then demonstrated that the blind IQA measure FADE originating from the VIS spectrum has some potential to be applied to TIR images. In a survey-like evaluation, we showed that the Transformer-based method DehazeFormer performs best on the M3FD dataset in dehazing TIR images. The margin is large as shown with two full-reference IQA measures SSIM and PSNR. Furthermore, TIR images dehazed with DehazeFormer provide a good input for object detection using YOLOv5 according to the mAP. FADE is applicable to the TIR spectrum, but some results are unexpected and need further analysis.

References

[1] Subhash Chand Agrawal and Anand Singh Jalal. A comprehensive review on analysis and implementation of recent image dehazing methods. *Archives of Computational Methods in Engineering*, 2022. 1, 2, 4

[2] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe de Vleeschouwer. O-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Outdoor Images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 867–8678, 2018. 3, 5

[3] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1798–1805, 2020. 3, 5

[4] Dana Berman, Tali Treibitz, and Shai Avidan. Single Image Dehazing Using Haze-Lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(3):720–734, 2020. 2

[5] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 2

[6] Lark Kwon Choi, Jaehee You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11):3888–3901, 2015. 2, 3, 4, 5, 6

[7] Thierry Elias, Martial Haeffelin, Philippe Drobinski, Laurent Gomes, Jerome Rangognio, Thierry Bergot, Patrick Chazette, Jean-Christophe Raut, and Michele Colomb. Particulate contribution to extinction of visible radiation: Pollution, haze, and fog. *Elsevier Atmospheric Research*, 92:443–454, 2009. 1

[8] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision (IJCV)*, 88(9):303–338, 2010. 5

[9] Tao Fang, Zhiguo Cao, and Ruicheng Yan. A unified dehazing approach for infrared images. In Xinyu Zhang and Jianguo Liu, editors, *MIPPR 2013: Multispectral Image Acquisition, Processing, and Analysis*, SPIE Proceedings, page 89170V. SPIE, 2013. 2, 4

[10] Adrian Galdran, Aitor Alvarez-Gila, Alessandro Bria, Javier Vazquez-Corral, and Marcelo Bertalmio. On the duality between retinex and image dehazing. In *IEEE CVPR*, 2018. 4, 6

[11] Yona Falinie A. Gaus, Neelanjan Bhowmik, Brian K. S. Isaac-Medina, and Toby P. Breckon. Visible to infrared transfer learning as a paradigm for accessible real-time object detection and classification in infrared imagery. In *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies IV*, volume 11542, pages 13–27, 2020. 3

[12] Todd Richard Goodall, Alan Conrad Bovik, and Nicholas G. Paulter. Tasking on Natural Statistics of Infrared Images. *IEEE Transactions on Image Processing*, 25(1):65–79, 2016. 3, 7

[13] Jie Gui, Xiaofeng Cong, Yuan Cao, Wenqi Ren, Jun Zhang, Jing Zhang, Jiuxin Cao, and Dacheng Tao. A comprehensive survey and taxonomy on single image dehazing based on deep learning. *ACM Computing Surveys*, dec 2022. 2

[14] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. 2, 5

[15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017. 2

[16] Wilhelm Isoz, Thomas Svensson, and Ingmar Renhorn. Nonuniformity correction of infrared focal plane arrays. In *Proceedings of SPIE Vol. 5783*, 2005. 4

[17] Bin Ji, Yunyun Ji, Han Gao, Xuedong Hu, and Feng Ding. No-reference image quality assessment for dehazed images. *Journal of Electronic Imaging*, 31(1):013013, 2022. 3

[18] Glenn Jocher. yolov5, 2022. Accessed on 03.11.2022. 5

[19] Stephen J. Fenley John N. Sanders-Reed. Visibility in degraded visual environments (dve). In *Proceedings of SPIE*, volume 10642, 2018. 1

[20] Howard V. Kennedy. Modeling noise in thermal imaging systems. In *Proceedings of SPIE Vol. 1969*, 1993. 4

[21] Vladimir V. Kniaz, Vladimir A. Knyaz, Jiří Hladůvka, Walter G. Kropatsch, and Vladimir Mizginov. Thermalgan: Multimodal color-to-thermal image translation for person re-identification in multispectral dataset. In Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision – ECCV 2018 Workshops*, pages 606–624. Springer International Publishing, 2019. 2

[22] M. Krišto and M. Ivašić-Kos. Thermal imaging dataset for person detection. In *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1126–1131, 2019. 2

[23] Mate Krišto, Marina Ivašić-Kos, and Miran Pobar. Thermal object detection in difficult weather conditions using yolo. *IEEE Access*, 8:125459–125476, 2020. 2

[24] Rahul Kumar, Brajesh Kumar Kaushik, and R. Balasubramanian. Multispectral transmission map fusion method and architecture for image dehazing. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 27(11):2693–2697, 2019. 4

[25] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-One Dehazing Network. In *IEEE International Conference on Computer Vision (ICCV)*, 2017. 2, 3, 5

[26] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 1, 2, 3, 5, 8

[27] Haiyan Li, Jun Wu, Aimin Miao, Pengfei Yu, Jianhua Chen, and Yufeng Zhang. Rayleigh-maximum-likelihood bilateral filter for ultrasound image enhancement. *Biomedical engineering online*, 16(1):46, 2017. 4

[28] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *IEEE/CVF International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2, 3, 5

[29] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. In *IEEE International Conference on Computer Vision (ICCV)*, 2019. 2

[30] Baolin Lv, Shoufeng Tong, Qiaoyuan Liu, and Haijiang Sun. Statistical Scene-Based Non-Uniformity Correction Method with Interframe Registration. *Sensors*, 19(24), 2019. 4

[31] Abdulla Al Mansoori, Issacniwas Swamidoss, Abdulrahman Almarzooqi, and Slim Sayadi. An investigation of various dehazing algorithms used on thermal infrared imagery for maritime surveillance systems. In *Target and Background Signatures VII*, volume 11865, page 118650M. International Society for Optics and Photonics, SPIE, 2021. 2, 5

[32] E. J. McCartney and Freeman F. Hall. Optics of the atmosphere: Scattering by molecules and particles. *Physics Today*, 30(5):76–77, 1977. 2

[33] Răzvan-Cătălin Miclea, Vlad-Ilie Ungureanu, Florin-Daniel Sandru, and Ioan Silea. Visibility enhancement and fog detection: Solutions presented in recent scientific papers with potential for application to mobile systems. *Sensors*, 21(10), 2021. 1, 2

[34] Xiongkuo Min, Guangtao Zhai, Ke Gu, Yucheng Zhu, Jiantao Zhou, Guodong Guo, Xiaokang Yang, Xinping Guan, and Wenjun Zhang. Quality evaluation of image dehazing methods using synthetic hazy images. *IEEE Transactions on Multimedia*, 21(9):2319–2333, 2019. 3

[35] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. 3, 5

[36] Anish Mittal, Rajiv Soundararajan, and Alan Conrad Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. 2, 3, 5

[37] Anush Krishna Moorthy and Alan Conrad Bovik. Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality. *IEEE Transactions on Image Processing*, 20(12):3350–3364, 2011. 3

[38] Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, Sumohana S. Channappayya, and Swarup S. Medasani. Blind image quality evaluation using perception based features. In *Twenty First National Conference on Communications (NCC)*, 2015. 3, 5

[39] Roberto Nebuloni. Empirical relationships between extinction coefficient and visibility in fog. *SPIE Applied Optics*, 44(18):3795–3804, 2005. 1

[40] Anil Singh Parihar, Yash Kumar Gupta, Yash Singodia, Vibhu Singh, and Kavinder Singh. A comparative study of image dehazing algorithms. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 766–771. IEEE, 2020. 1, 2, 5

[41] M. Pavethra and M. Umadevi. Deep Learning approaches for Image Dehazing. In *6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, 2021. 1, 2

[42] Yanting Pei, Yaping Huang, Qi Zou, Yuhang Lu, and Song Wang. Does Haze Removal Help CNN-based Image Classification? In *European Conference on Computer Vision (ECCV)*, 2018. 1

[43] Ana Belén Petro, Catalina Sbert, and Jean-Michel Morel. Multiscale retinex. *Image Processing On Line*, 4:71–88, 2014. 5

[44] Michele A. Saad, Alan C. Bovik, and Christophe Charrier. Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain. *IEEE Transactions on Image Processing*, 21(8):3339–3352, 2012. 3, 5

[45] Christos Sakaridis, Dengxin Dai, and Luc van Gool. Semantic Foggy Scene Understanding with Synthetic Data. *International Journal of Computer Vision (IJCV)*, 126(9):973–992, 2018. 2, 3, 5

[46] Aditi Sarker, Morium Akter, and Mohammad Shorif Uddin. Simulation of hazy image and validation of haze removal technique. *Journal of Computer and Communications*, 07(02):62–72, 2019. 2

[47] Sweta Shaw, Rajarshi Gupta, and Somshubhra Roy. A review on different image de-hazing methods. In *Advances in Intelligent Systems and Computing*, pages 533–540. Springer Singapore, July 2019. 1, 2

[48] Yuda Song, Zhuqing He, Hui Qian, and Xin Du. Vision Transformers for Single Image Dehazing. *IEEE Transactions on Image Processing*, 32:1927–1941, 2023. 2, 3, 5

[49] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly Assess Image Quality in the Wild Guided by a Self-Adaptive Hyper Network. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[50] Song Sun and Xinhua Guo. Image Enhancement Using Bright Channel Prior. In *International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII)*, 2016. 2

[51] Jean-Philippe Tarel and Nicolas Hautière. Fast visibility restoration from a single color or gray level image. In *IEEE International Conference on Computer Vision (ICCV)*, 2009. 2, 5

[52] Michael Teutsch, Angel D. Sappa, and Riad I. Hammoud. Computer vision in the infrared spectrum: Challenges and approaches. *Synthesis Lectures on Computer Vision*, 10(2):1–138, 2021. 2, 5

[53] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[54] Garima Yadav, Saurabh Maheshwari, and Anjali Agarwal. Contrast limited adaptive histogram equalization based enhancement for real time video system. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2392–2397. IEEE, 2014. 5

[55] Siyu Yan, Jingwen Zhu, Kai Yun, Yuexing Wang, and Chuangang Xu. An infrared image dehazing method based

on modified dark channel prior. In Wei Wei and Yang Yue, editors, *International Conference on Biometrics, Microelectronic Sensors, and Artificial Intelligence (BMSAI)*, volume 12252, page 122520N. SPIE, 2022. 2

[56] Xiaohan Yang, Fan Li, and Hantao Liu. A Survey of DNN Methods for Blind Image Quality Assessment. *IEEE Access*, 7:123788–123806, 2019. 3

[57] Yibing Zhan and Rong Zhang. No-Reference JPEG Image Quality Assessment Based on Blockiness and Luminance Change. *IEEE Signal Processing Letters*, 24(6):760–764, 2017. 3, 5

[58] Lichao Zhang, Abel Gonzalez-Garcia, Joost van de Weijer, Martin Danelljan, and Fahad Shahbaz Khan. Synthetic data generation for end-to-end thermal infrared tracking. *IEEE Transactions on Image Processing*, 28(4):1837–1850, 2019. 2

[59] Lintao Zheng, Hengliang Shi, and Ming Gu. Infrared traffic image enhancement algorithm based on dark channel prior and gamma correction. *Modern Physics Letters B*, 31(19-21):1740044, 2017. 2, 4

[60] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-High-Definition Image Dehazing via Multi-Guided Bilateral Learning. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 5

[61] Hang Zhou, David Greenwood, and Sarah Taylor. Self-Supervised Monocular Depth Estimation with Internal Feature Fusion. In *British Machine Vision Conference (BMVC)*, 2021. 4

[62] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In Paul S. Heckbert, editor, *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994. 5