IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 7, 2021, accepted May 17, 2021, date of publication May 20, 2021, date of current version June 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082207

A Framework for Objective Evaluation of Single Image De-Hazing Techniques

ALESSANDRO ARTUSI[1] **AND KONSTANTINOS A. RAFTOPOULOS**[1,2,3]
[1]DeepCamera MRG Laboratory, CYENS–Centre of Excellence, 1016 Nicosia, Cyprus
[2]Hellenic Military Academy, 16673 Attica, Greece
[3]Department of Civil Engineering, University of West Attica, 12244 Attica, Greece

Corresponding author: Alessandro Artusi (artusialessandro4@gmail.com)

The work of Alessandro Artusi was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under Grant 739578, and in part by the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

ABSTRACT Real-world environment, where images are acquired with digital camera, may be subject to sever climatic conditions such as haze that may drastically reduce the quality performance of sophisticated computer vision algorithms used for various tasks, e.g., tracking, detection, classification etc. Even though several single image de-hazing techniques have been recently proposed with many deep-learning approaches among them, a general statistical framework that would permit an objective performance evaluation has not been independently introduced yet. In this manuscript, certain performance metrics that emphasize different aspects of image quality, output ranges and polarity, are identified and combined into a single performance indicator derived in an unbiased manner. A general methodology is thus introduced, as a framework for objective performance evaluation of current and future dehazing tasks, through an extensive comparison of 15 single image de-hazing techniques over a vast range of image data sets. The proposed unified framework shows several advantages in evaluating diverse and perceptually meaningful image features but also in elucidating future directions for improvement in image dehazing tasks.

INDEX TERMS Haze, single image de-hazing, deep-learning, generative adversarial network (GAN), convolutional neural network, bench-marking, survey, computational-performance, computer vision, image processing.

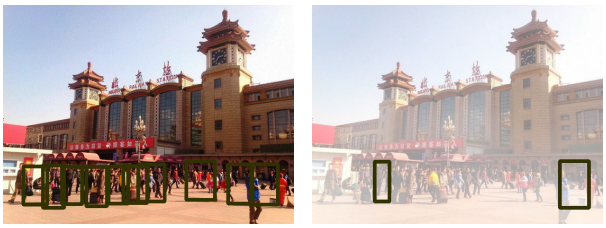
I. INTRODUCTION

Image capturing is the first step in an imaging pipeline and plays an important role in providing input with an acceptable level of quality that will not compromise the performance of existing computer vision and image processing tasks, i.e., classification, object detection, details extraction etc. The quality of the captured image may be degraded not only by noise generated by the device sensor but also by environmental conditions such as fog, rain drops, haze, illuminations conditions etc. affecting visibility of objects, details etc. and drastically compromising the recognition capability of computer vision and image processing tasks.

Atmospheric haze in particular, is generated by particles that attenuate light passing through them. These particles cause light absorption and scattering, so as only a certain percentage of the reflected light reaches the camera sensor. This results in an acquired image of reduced visibility and color shift when compared to an image acquired under standard sunlight environmental conditions and it drastically reduces the performance of state-of-the-art object detection algorithms as shown in Figure 1. De-hazing techniques for single images exist. They provide an attempt to enhance the hazy image by solving the so called light scattering model, as described in Section II. They can be divided in two main categories. Methods that use image prior features to constraint the solution of the scattering model and methods that use a black-box approach where deep-learning is used to limit the constraints required by the prior features approaches (see Section II).

While methods in the first category are more or less understood in a probabilistic context, the methods in the second category exhibit a well known dependency to the data sets used for their training. Even though the performance of the second group is typically measured on unseen images with cross validation techniques, a hidden bias to certain choices, data sets or parameters is generally difficult to quantify. The contribution in this paper is in introducing a *framework* for objective evaluation using diverse performance metrics *unified* under a statistical context. Existing benchmark data sets

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino.



(a) free-haze image (b) haze image

FIGURE 1. Images acquired under different environmental conditions - (a) without haze, while (b) with haze. The image in (b) clearly shows reduced visibility where objects and details are less visible, and color shift where colors in various areas are desaturated, when compared to the ground truth (a). Applying state-of-the-art object detection algorithm only few persons in the haze image are detected (b). Images from the RESIDE 2018 data set [3].

with different characteristics are integrated and by means of the proposed statistical framework, a comprehensive performance evaluation is performed, between the most popular de-hazing methods at a scale not seen before for this particular problem. For methods based on deep learning in particular, fusing testing data sets affects the hidden bias in training / cross validation choices. Indeed, different data sets with types of images ranging from indoor to outdoor settings, low to high resolutions, synthetic to real haze, as well as different haze levels are used to check the level of dependency of deep learning methods to their original training sets.

Five objective metrics are used, providing a mixture of orthogonal criteria in performance evaluation, namely, detection of structural changes, color distortion and identifying visibility changes. Two reference-less metrics predict losses in naturalness [1] thus estimating the quality of an image in a way that is consistent with human perception [2].

The proposed framework addresses the unification problem of these metrics by introducing a general statistical framework, one that provides a unique performance indicator, derived in an unbiased manner. The generality of the proposed framework makes it possible to incorporate current and future single image de-hazing techniques into a uniform objective evaluation.

The rest of the paper is structured as follows: the single image de-hazing problem formulation and related-work are described in Section II. A total of $78K$ images with different characteristics, i.e., indoor, outdoor, synthetic and real haze, as well as dense haze, different haze levels, synthetic and real images, are used and described in Section III. Section IV provides the methodology used for the analysis of the results, including quality evaluation as well as analysis of the computational performance. Final remarks are discussed in Section VI.

II. SINGLE IMAGE DE-HAZING, PROBLEM FORMULATION AND CURRENT METHODS

A simple way to model the scattering of light into a medium is the so called classical atmospheric scattering model:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

where I is the hazy image, J is the recovered haze free radiance scene, while A and t are the global atmospheric light and the medium transmission coefficients respectively. The medium transmission coefficient describes the portion of the light that is not scattered and reaches the camera. The first term in eq. 1, describes the scene radiance decay in the medium and is called the direct attenuation term. The remaining term is the airlight and it is responsible for the color shift in the image due to the scattered light in the medium [4]. In the case of homogeneous atmosphere, the transmission t is expressed as an exponential function of the scattering coefficient of the atmosphere β and the scene depth d:

$$t(x) = \exp^{\beta(x)d(x)}. \tag{2}$$

The above scattering model is typically used to synthesize haze images starting from a free-haze image, and it is used as the base model for the majority of single image de-hazing techniques. Based on the physical model described in eq. 1, the haze free image J is recovered by estimating, first the transmission term (t), then the coefficient A. This is achieved either by making use of an image prior and imposing specific constraints or by using data-driven approaches, e.g., deep-learning. Eq. 1 describes an ill-posed problem due to the fact that the number of unknown variables is larger then the number of equations. Moreover, the so called airlight-albedo ambiguity where the albedo term is not constant over all pixels of the haze image, gives rise to a large number of undetermined degrees of freedom as noted by Fattal [5]. A solution to this problem would be to impose constraints either on image properties or on the computation of the coefficient terms t and A in eq. 1. For example, Fattal [5], [6] imposed constraints on the estimation of the albedo term making it easy to estimate the depth term (d). Other constrains can be imposed in the formulations of the light and object chromaticity as well as on the atmospheric light (A) [7]. Dark channel prior (DCP) [4] can be used to constraint the computation of the transmission term (t). White balancing the input image may assume that the haze can be represented with perfect withe and constraint the so called atmospheric veil, as defined in [8].

Tang *et al.* [9] extended the DCP concept considering other haze image features as constraints, e.g., local maximum contrast, saturation and hue disparity. Chen *et al.* [10] use DCP to recover the transmission term t and Total Generalized Variation (TGV) for refining, when DCP is not precisely estimating it. Scene transmission term t, can also be estimated imposing boundary constraints on the radiance cube combined with a weighted $L1$ norm based contextual regularization, as proposed in [11]. A global image prior, sensitive to dependencies created from the use of a wrong $||A||$, is proposed by Sulami *et al.* [12] introduced a simple procedure for recovering $||A||$ by minimizing these dependencies. Zhu *et al.* [13] observed that the difference between brightness and saturation can approximately represent the concentration of haze. Using this color attenuation prior, they derive the depth map (d) by proposing a linear model and

learning the parameters of the model with a supervised learning method. Berman *et al.* [14] proposed a new non-local prior to recover the depth map (d). The key concept is to assume that colors of a haze-free image are well approximated by a few hundred distinct colors and each one can be represented as a line in the *RGB* space (haze-line). Alternatively, to solve the albedo and depth ambiguity, the haze input image can be modeled by two statistically independent terms, scene albedo and depth. The dependency between these two terms and the input haze image can be modeled using a probabilistic model based on factorial Markov random field [15], [16]. Fattal [6] used a Markov random field for producing a complete and regularized transmission term (t), given noisy and scattered estimates.

Imposing specific constraints to haze removal, as shown in the above prior-based techniques, has the advantage of effectively formulating an ill-posed problem while simultaneously reducing computational complexity. However, these constraints may be easily violated when treating real-world images. Moreover, input parameters that could work for a large variety of images may be difficult to estimate. This makes the proposed solution often unpractical, especially when a realistic reconstruction of the image is required. In these cases, artifacts are often visible as shown in Figure2.

Fusion based techniques can be used to overcome the need of estimating the depth term (d), thus avoiding its costly refinement processes, and so estimate directly the free-hazy image. To achieve this, one can start from the haze image, derive a number of enhanced images, i.e., through gamma correction [17]–[19], contrast enhancement [20] or by extracting statistical features [21] and then fuse or merge them into a haze-free image.

Multi-scale blending may also be used to improve the overall quality of the final haze-free image [17], [20], [21].

Avoiding the estimation of the depth term, makes fusion-based techniques a good alternative to the prior-based methods. However, these approaches share similar drawbacks with prior-based methods. Indeed, the enhancement of the derived-from-the-hazed input images, can be seen as imposing constraints that are difficult to estimate in a large variety of images. Moreover, multi-scale blending increases the computational complexity.

Imposing constraints (image priors) in the estimation of model coefficients described in eq. 1, provide solutions that fail when these priors are violated. Unfortunately this is usually the case in real-world images. To overcome this problem, deep-learning approaches have been considered to either learn the mapping between the transmission term (t) and its haze input image [22]–[25], or directly generate the haze-free image [26]–[40]. However, as will be shown hereinafter, deep-learning approaches suffer from increased dependency to their training sets and thus their generalization capability may be questioned across diverse datasets.

Previous bench-marking on single image de-hazing algorithms exist. Ancuti *et al.* [41] analyzed the performance of 6 existing single image de-hazing techniques using a small number of objective metrics, e.g., SSIM [42] and the CIE color difference formula $\Delta E2000$ [43]. The data set used is the D-haze [44], which provides a small number of images that are both indoor and outdoor where haze is produced synthetically. Li *et al.* [45] bench-marking, compares 11 de-hazing techniques of which 2 are deep-learning approaches. Recently, Li *et al.* [3] provided a bench-marking on a large data set, RESIDE 2018 [3], comparing 9 de-hazing techniques, of which 3 are deep-learning approaches. They used 4 objective metrics of which 2 are no-reference metrics. They were also concerned whether de-hazing can help task-oriented computer vision approaches, i.e., object detection, in the sense of improving the performance of object detection in the presence of haze. Such a metric was incorporated in their bench-marking. We share the same concern in this work. A similar metric has been incorporated in the proposed framework.

FIGURE 2. (top) - Example result from Kratz *et al.* [15] and Fattal [5], clear color shift artifacts are visible. (bottom) - Quality comparison among the DCP [4] technique and its fast version *DCP_F*. Visible halos artifacts, around objects, are produced with the fast version.

III. IMAGES DATA SETS

To demonstrate the proposed statistical framework, a sub-set of 34 state-of-the-art de-hazing techniques, shown in Table 1, was selected. The selection was based on the availability of their original code and in the case of deep-learning based approaches, the availability of their trained model. For some of these methods the code/trained network model is either not provided or not accessible [6], [7], [20], [24], [25], [28], [30]–[35], [37], [38], [40], [46], thus these methods were excluded, to avoid own implementations not intended by the authors. This leaves a subset of 17 techniques of which 6 are based on deep-learning techniques. Preliminary testing, among these 17 techniques, has shown that Fattal [5] and Kratz and Nishino [15] methods have difficulties to find suitable input parameters for all type of images used. Figure 2 shows several artifacts in the de-hazed images, i.e., color shift, over saturation etc. as also confirmed in [45]. As overcoming this problem was not trivial to us, nor a method was suggested in the paper, these two techniques were also excluded.

(a) D-hazy [41] (b) NTIRE 2018-2019 [44], [46] (c) RESIDE 2018 [3]

FIGURE 3. Examples of images belonging to the data sets used in this bench-marking: frame color indicate free-haze (red) and haze (blue) image.

TABLE 1. Original pool of de-hazing techniques where 15 techniques have been selected. Red text indicates methods that in preliminary tests have provided poor results due to the difficulties of finding common parameters setting for all input images.

Method	Acronymum	Code Availability
Fattal [5]	FAT	YES
Kratz et al. [15]	KRATZ	YES
He et al. [4]	DCP/DCP_F	YES
Tarrel et al. [8]	TFV	YES
Meng et al. [11]	BCCR	YES
Sulami et al. [12]	ATML	YES
Choi [21]	DEF	YES
Zu et al. [13]	CAP	YES
Bermann et al. [14]	NLD	YES
Chen et al. [10]	GRM	YES
Galdran [17]	AMEF	YES
Fattal [6]	-	NO
Tan [7]	-	NO
Ancuti et al. [20]	-	NO
Wu et al. [47]	-	NO
Zheng et al. [18]	-	NO
Zhu et al. [19]	-	NO
Ren et al. [22]	M_NN	YES
Cai et al. [23]	DE_Z	YES
Li et al. [26]	AOD	YES
Engin et al. [29]	CY_D	YES
Qin et al. [36]	FFA_Net	YES
Sourya et al. [39]	DMPH	YES
Shao et al. [48]	-	NO
Ren et al. [31]	-	NO
Dong et al. [37]	-	NO
Zhang et al. [24]	-	NO
Dong et al. [38]	-	NO
Yang et al. [30]	-	NO
Swami et al. [28]	-	NO
Dudhane et al. [32]	-	NO
Ren et al. [33]	-	NO
Li et al. [34]	-	NO
Li et al. [35]	-	NO
Pang et al. [25]	-	NO
Shen et al. [40]	-	NO

Concerning DCP, a fast solution is also available in its paper. Both solutions were preliminary tested. Results in Figure 2, show visible halos artifacts around objects when the fast solution is used. Based on this result, only the results of the original solution [4] is provided here. For the computational evaluation, however, see Section V-C, the fast solution is also included. This is reducing to 15 the number of de-hazing techniques tested here. For all techniques the original code was used, with default input parameters.

TABLE 2. Characteristics of data sets used in the proposed framework.

	Data sets		
Type	RESIDE β [3]	D-Hazy [41]	NTIRE [44], [46]
Indoor	YES	YES	YES
Outdoor	YES	NO	YES
Haze-level	YES	NO	NO
Task-driven	YES	NO	NO
De-Haze	NO	NO	YES [46]

To present the proposed framework for objective evaluation, the 15 aforementioned de-hazing techniques have been used, together with a vast range of images from several existing benchmark data sets. To the best of our knowledge, this makes the largest data set used in the literature so far in de-hazing activities, constituted of approximately 78K images.

The data sets incorporated in the proposed framework were carefully chosen to represent a wide range of real life conditions, capturing scenarios, diverse image features and in the case of RTTS, high quality annotations necessary for testing object recognition performance after haze removal. In Tables 2 and 3 the main characteristics of these data sets are shown and some of these images can be inspected in Figure 3. They comprise indoor and outdoor images at different resolutions, images with different degrees of haze, including dense haze but also annotated images needed for the task-driven evaluation as will be discussed in Section V-D.

A. D-HAZY DATA SET

The D-haze data set [41] provides two sets of images. Two existing data sets, Middleburry [49] and NYU-Depth V2 (NYU) [50], provide the depth map. Real-world ground truth images with synthesize haze, employing Koschmieder's

TABLE 3. Data sets used in our bench-marking.

Type	# Images	W x H
Midlleburry	23	2,880 x 1,988
NYU	1,448	640 x 480
I-haze	30	4,657 x 2,833
O-haze	40	2,966 x 3,202
De-haze [46]	55	1,600 x 1,200
OTS (x 35 haze levels)	2,061	550x413
RTTS	4,322	-

physical model [51] of light transmission in hazy scenes have been augmented under the assumption that atmospheric intensity and haze density are uniform. Table 3 shows the number of images belonging to the two subsets and their resolution.

B. NTIRE DATA SET

The NTIRE data set series includes three collections of data, developed between 2018 and 2020 [44]. At the time of this writing the 2020 version was not available for testing. NTIRE data sets are the only ones that provide hazy images of very high resolution and this is the reason they were selected here. Hazy images are generated using real haze produced by a professional haze machine (see Table 3). All NTIRE data sets consist of indoor (I-haze) and outdoor (O-haze) images [44] as well as dense uniform haze (De-haze) [48] images. Each scene includes a MacBeth color checker, allowing more precise image color calibration and better assessment of performance. Moreover, since images are captured in a controlled environment, both haze-free and hazy images are captured under the same illumination conditions.

C. RESIDE β DATA SET

D-hazy and NTIRE data sets provide mainly indoor images. In order to enrich with further content including indoor, outdoor and annotated images the RESIDE data set was also selected together with the Outdoor Training Set (OTS) and the Task-driven Testing Set (RTTS) (see Table 3). OTS extends the number of outdoor images with different haze levels, e.g., 35 haze levels for a total of 72K images. RTTS provides a large annotated data set, necessary to address whether de-hazing is capable of increasing the performance of existing computer vision algorithms. In RESIDE dataset the haze has been synthetically added using the atmospheric model described in eq. 1 with atmospheric light (A) and atmospheric scattering coefficient (β) varying in the range of [0.8, 1.0] and [0.04, 0.2] with a step of 0.5 and a variable step of 0.02 and 0.04 respectively.

IV. EXPERIMENTS AND ANALYSIS

The proposed framework is evaluated in the next section through experimental results that compare the performance of single image de-hazing techniques. The aim is to introduce a unifying benchmark that will incorporate many diverse data sets with different aspects of image quality. The merit of such a framework is demonstrated in the conceptual results regarding dependency to training sets, computational constraints but also in understanding whether haze removal is sufficient for improving the performance of computer vision task-driven applications.

In Section V, a qualitative comparison using a total of 5 objective metrics is described. In Section V-C, various tests on the computational performance are presented. Finally, in Section V-D a bench-marking on a task-driven computer vision application, related to an object detection task is reported.

V. PERFORMANCE EVALUATION

In the proposed framework, five objective metrics have been used. Among them, Peak to Signal Noise Ration (PSNR), measures the noise ratio in the image signal and Structural Similarity Index (SSIM) [42] measures structural changes within the image.

Despite the fact that these objective metrics provide a useful indication of image quality, they have limited use in predicting the probability that a specific change in an image is visible or not to an average observer. Thus, the so called HDR-VDP metric [52] is also used, which through implementation of important aspects of the human visual system is capable to predict the visibility changes as well as the quality degradation with respect to the reference image. Moreover, two non-reference objective metrics, BRISQUE [1] and IQVG [2] are also used, where the quality of an image is estimated automatically, in a way that is consistent with human perception and without any prior knowledge of the reference image. These will help to understand whether the evaluation is consistent across different types of objective metrics.

The higher the value of the PSNR, SSIM, HDR-VDP and IQVG metrics, the higher the image quality, whereas the inverse holds for the values of the BRISQUE objective metric.

This certain combination of metrics introduced by the proposed framework were chosen to emphasize different aspects of image quality and allow for an objective evaluation. Their combined effect is assessed in the context of a statistical treatment that *objectifies* the results, leading to a single performance indicator, derived in an unbiased manner from all five metrics. By converting raw metric scores into z-scores and rank methods according to their mean z-score from all metrics, one achieves a *standardization of values* that aids the comparison between metrics of different scales or widely different ranges measured on the same scale. The *standardization process* used hereinafter is now described in distinct steps:

Let i indexing methods, j indexing metrics:

1) Calculate the mean value for each metric and denote μ_j
2) Calculate the standard deviation for each metric and denote it σ_j
3) Calculate the distance of every raw score from the respective mean, measured in standard deviations e.g.

TABLE 4. Standardized values (z-scores) of prior based methods comparative performance with overall performance indicators. Each entry is the number of standard deviations by which the method's score was above or below the mean score of all methods for that metric. BRISQUE values have been negated to be consistent with other metrics where positive entries show better performance than negative entries. Rows labeled MEAN hold the means of the respective columns. These numbers serve as performance indicators for the various methods. Outdoor and indoor environments are treated separately. An overall score, as the mean of all indoor and outdoor scores from all data sets, is calculated in Table 5. Green to red color scale means best to worst in the row.

PRIOR BASED METHODS - STANDARDIZED PERFORMANCE			AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
NTIRE 2019	INDOOR	PSNR	1.08	-2.25	-0.59	0.58	0.60	0.28	0.09	0.66	-0.44
		SSIM	0.80	-2.38	-0.51	0.46	0.34	0.46	-0.40	0.58	0.65
		HDR	-0.30	1.94	1.09	0.48	-0.49	-0.56	-0.16	-0.78	-1.23
		IQVG	0.08	-1.96	1.43	-0.22	1.04	0.56	-0.12	-0.02	-0.80
		BRISQUE	0.86	-0.62	-0.33	1.19	-0.07	-0.07	-2.14	0.75	0.42
		MEAN	0.50	-1.05	0.22	0.50	0.29	0.13	-0.55	0.24	-0.28
			AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
	OUTDOOR	PSNR	1.08	-0.89	0.92	1.10	-0.62	0.67	0.15	-0.84	-1.56
		SSIM	0.57	-0.99	0.19	1.68	0.05	-1.06	-0.85	1.18	-0.77
		HDR	0.00	2.18	0.43	0.06	-0.82	0.52	-0.41	-0.89	-1.07
		IQVG	-0.52	-1.13	0.34	-0.52	0.34	0.16	2.30	-0.21	-0.76
		BRISQUE	0.11	-0.69	0.68	1.11	0.07	0.52	-2.12	0.86	-0.54
		MEAN	0.25	-0.31	0.51	0.69	-0.20	0.16	-0.18	0.02	-0.94
			AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
NTIRE 2018	INDOOR	PSNR	0.65	-2.05	-0.74	0.16	0.77	0.76	-0.86	0.54	0.77
		SSIM	0.84	-2.16	-0.22	0.34	0.67	0.45	-1.05	0.40	0.73
		HDR	-0.17	1.83	1.17	-0.43	-0.24	-0.28	0.46	-1.19	-1.14
		IQVG	0.72	-2.14	-0.06	0.20	0.72	0.72	-1.10	0.20	0.72
		BRISQUE	0.96	-1.86	-0.60	0.51	0.37	0.56	-1.26	0.87	0.44
		MEAN	0.60	-1.28	-0.09	0.16	0.46	0.44	-0.76	0.16	0.30
			AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
	OUTDOOR	PSNR	0.77	-2.37	-0.08	1.06	0.63	0.05	0.28	-0.43	0.10
		SSIM	0.78	-2.37	0.05	0.86	0.38	-0.59	-0.03	0.70	0.22
		HDR	-0.23	2.07	1.03	-0.26	-0.63	-0.09	0.14	-1.11	-0.93
		IQVG	-0.06	-2.33	0.50	-0.06	1.07	-0.06	1.07	-0.06	-0.06
		BRISQUE	0.26	-2.18	0.63	0.56	0.85	0.70	0.54	-0.52	-0.85
		MEAN	0.30	-1.43	0.43	0.43	0.46	0.00	0.40	-0.28	-0.31

			AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
OTS	OUTDOOR	PSNR	0.30	-2.25	0.00	-0.11	1.40	0.27	0.70	0.21	-0.52
		SSIM	0.52	-2.61	0.23	0.09	0.71	0.42	0.23	0.28	0.13
		HDR	0.07	1.36	1.00	-1.31	0.46	0.05	0.53	-0.54	-1.62
		IQVG	0.45	-2.57	0.28	0.11	0.45	0.62	-0.22	0.62	0.28
		BRISQUE	0.54	-2.60	0.31	0.19	0.42	0.43	-0.07	0.65	0.12
		MEAN	0.37	-1.73	0.36	-0.21	0.69	0.36	0.23	0.24	-0.32
			AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
D-HAZE (INDOOR)	NYU	PSNR	0.68	-1.91	1.47	-1.09	0.09	0.19	0.69	0.06	-0.17
		SSIM	0.18	-2.46	0.90	-0.06	0.58	-0.06	0.82	-0.14	0.26
		HDR	-0.04	2.49	0.19	-0.66	-0.55	0.19	-0.25	-0.73	-0.63
		IQVG	0.58	0.00	0.00	0.00	-0.58	1.15	-2.31	0.58	0.58
		BRISQUE	1.26	-1.99	0.06	0.62	0.09	0.47	-1.09	-0.18	0.76
		MEAN	0.53	-0.78	0.52	-0.24	-0.07	0.39	-0.43	-0.08	0.16
			AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
	Middlesburry	PSNR	0.72	-2.41	0.93	0.47	0.18	-0.24	0.67	-0.19	-0.13
		SSIM	0.43	-2.59	0.60	0.20	0.49	0.03	0.60	-0.08	0.32
		HDR	-0.30	2.37	0.26	-0.49	-0.66	0.60	-0.27	-0.74	-0.77
		IQVG	-0.45	1.98	-0.45	0.36	-0.45	1.17	-1.26	-0.45	-0.45
		BRISQUE	0.44	-2.02	0.00	0.34	0.57	1.17	-1.27	0.25	0.51
		MEAN	0.17	-0.53	0.27	0.18	0.03	0.55	-0.31	-0.24	-0.11

TABLE 5. Ranking of 9 prior-based dehazing techniques based on their combined outdoors-indoors overall perforce across all tested data. Each column hold the means of Table 4. AMEF exhibits the best mean performance across all data sets. Standard deviations measure the method's sensitivity to the data sets. The MEAN STD is calculated to measure the overall degree of dependency of prior methods' performance to different data sets and will be compared to the MEAN STD of deep learning based methods.

PRIOR BASED METHODS – ALLOVER PERFORMANCE COMPARISON	AMEF	ATML	DCP	BCCR	CAP	DEF	GRM	NLD	TFV
NTIRE19-IN	0.50	-1.05	0.22	0.50	0.29	0.13	-0.55	0.24	-0.28
NTIRE19-OUT	0.25	-0.31	0.51	0.69	-0.20	0.16	-0.18	0.02	-0.94
NTIRE18-IN	0.60	-1.28	-0.09	0.16	0.46	0.44	-0.76	0.16	0.30
NTIRE18-OUT	0.30	-1.43	0.43	0.43	0.46	0.00	0.40	-0.28	-0.31
OTS-OUT	0.37	-1.73	0.36	-0.21	0.69	0.36	0.23	0.24	-0.32
D_H-NYU-IN	0.53	-0.78	0.52	-0.24	-0.07	0.39	-0.43	-0.08	0.16
D_H-MIDD-IN	0.17	-0.53	0.27	0.18	0.03	0.55	-0.31	-0.24	-0.11
MEAN	0.39	-1.02	0.32	0.22	0.24	0.29	-0.23	0.01	-0.21
STD	0.16	0.51	0.21	0.35	0.33	0.19	0.42	0.22	0.40
MEAN STD								0.31	

for the raw score of method i with metric j, denoted by r_{ij}, the respective standardized score is denoted by s_{ij} and equals $\frac{r_{ij}-\mu_j}{\sigma_j}$

4) For each method i, calculate the average of its standardized scores $s_{ij}, j \in metrics$ and use this single number as a performance indicator for method i

For the BRISQUE metric, the calculation in 3 is negated: $s_{ij} = \frac{\mu_j - r_{ij}}{\sigma_j}$ to be consistent with the other metrics, in that positive numbers indicate better performance. Standardization results for prior based and deep learning based methods are shown in Tables 4-5 and 6-7 respectively.

A. INDOOR VS. OUTDOOR ENVIRONMENT

The environment where the image is taken may influence an algorithm's performance. To verify this, images described in Section III, have been further divided into two groups representing two distinct types of environments, indoors and outdoors. Based on the normalization strategy of the proposed framework, standardized scores with combined metric performance indicators are shown, for prior based and deep learning de-hazing techniques respectively. An *outdoor-indoor overall combined ranking* is provided in Tables 5 and 7 for prior-based and deep learning techniques respectively.

Among prior-based methods, AMEF (Artificial Multiple-Exposure Image Fusion) [17] provides an overall better quality performance, considering both types of environments. AMEF performs haze removal by fusing artificially under-exposed images. Initially, the original hazy image is artificially under-exposed via a sequence of gamma-correction operations. The resulting set of multiply-exposed images is merged into a haze-free image

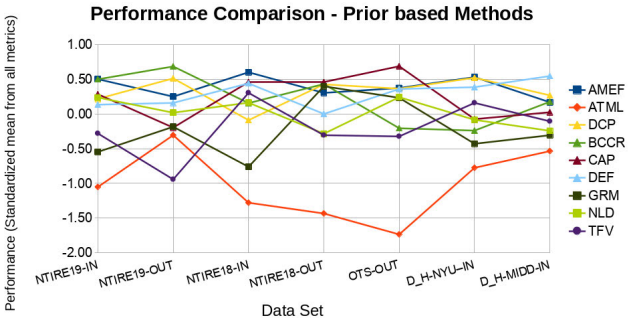


FIGURE 4. Performance graph of 9 prior-based dehazing techniques based on their mean score from 5 objective metrics. The data for each method are the columns in Table 5.

through a multi-scale Laplacian blending scheme. Even though AMEF is the winner among prior based methods, its performance degrades when only the outdoor dataset is considered. There, DCP is the combined (from all metrics) winner among the prior-based methods, but it is never the winner for a single metric. This reveals the usefulness of the proposed framework, as conclusions can be drawn from a global perspective that unifies different aspects of performance. As one also observes, DCP gives consistently good results (above average) for outdoors environment, in relation to all other metrics.

Also in relation to the different metrics, ATML is performing very well when HDR-VDP is used but its performance degrades in all other metrics, in both indoor and outdoor environments. This clearly indicates that relying on the results of a single metric may lead to wrong interpretations. One may also notice differences in performance when a data set belonging to the same type of environment, either indoor or outdoor is used. For example, BCCR performs considerably well, similar to AMEF, for the indoor environment of NTIRE2019 data set. However, its performance degrades, in the same environment (indoors), when the NTIRE2018 data set is used. This may be due to the different type of haze provided by these two data sets, e.g., dense vs. no-dense haze. This can also be easily deduced from figure 4.

In the case of deep-learning, one may notice higher variability in performance among different data sets and objective metrics when compared to prior based methods, as shown in Table 6. The overall winner method for deep learning based methods is AOD (All-in-One Dehazing Network). It is a convolutional neural network designed on a re-formulated atmospheric scattering model. Instead of estimating the transmission matrix and the atmospheric light separately, AOD directly generates the clean image through a light-weight CNN that optimizes a combined, (transmission matrix/atmospheric light) parameter named K. AOD is composed of two parts: a K-estimation module that uses five convolutional layers to estimate the combined parameter K, followed by a clean image generation module that

TABLE 6. Standardized values (z-scores) of deep learning based methods comparative performance with overall performance indicators. Each entry is the number of standard deviations by which the method's score was above or below the mean score of all methods for that metric. BRISQUE values have been negated to be consistent with other metrics where positive entries show better performance than negative entries. Rows labeled MEAN hold the means of the respective columns. These numbers serve as performance indicators for the various methods. Outdoor and indoor environments are treated separately. An overall score among the deep learning methods, as the mean of all indoor and outdoor scores from all data sets, is calculated in Table 7. Green to red color scale means best to worst in the row.

DEEP LEARNING METHODS –STANDARDIZED PERFORMANCE

			M_NN	AOD	DE_Z	CY_D	FFA	DMPH
NTIRE 2019	INDOOR	PSNR	0.60	0.54	0.40	0.58	-1.95	-0.18
		SSIM	0.05	-0.13	-0.47	1.35	-1.55	0.75
		HDR	0.46	0.09	-0.81	-0.19	1.64	-1.18
		IQVG	0.84	0.98	0.84	-1.19	-1.02	-0.43
		BRISQUE	0.75	0.71	0.10	-0.89	-1.54	0.87
		MEAN	**0.54**	**0.44**	**0.01**	**-0.07**	**-0.88**	**-0.03**
			M_NN	AOD	DE_Z	CY_D	FFA	DMPH
	OUTDOOR	PSNR	-0.41	0.28	-1.09	1.56	-0.90	0.56
		SSIM	-0.30	-0.80	-0.62	1.39	-0.81	1.14
		HDR	-0.58	0.69	-1.15	1.51	-0.71	0.23
		IQVG	0.71	1.07	0.82	-0.72	-1.37	-0.51
		BRISQUE	0.77	0.66	0.48	-0.15	-1.92	0.15
		MEAN	**0.04**	**0.38**	**-0.31**	**0.72**	**-1.14**	**0.31**
			M_NN	AOD	DE_Z	CY_D	FFA	DMPH
NTIRE 2018	INDOOR	PSNR	-0.06	-1.03	0.30	1.78	-0.79	-0.20
		SSIM	0.36	-0.51	0.58	1.45	-0.51	-1.38
		HDR	0.34	1.55	0.39	-0.16	-0.95	-1.18
		IQVG	-0.18	0.36	0.09	-1.82	1.18	0.36
		BRISQUE	-0.21	0.29	-0.06	-1.63	0.11	1.49
		MEAN	**0.05**	**0.13**	**0.26**	**-0.07**	**-0.19**	**-0.18**
			M_NN	AOD	DE_Z	CY_D	FFA	DMPH
	OUTDOOR	PSNR	-0.04	-0.45	-0.56	1.99	-0.69	-0.24
		SSIM	0.00	-1.21	-0.50	1.57	-0.57	0.71
		HDR	-0.22	0.89	-0.82	1.51	-1.03	-0.33
		IQVG	-0.14	-1.00	0.71	-1.00	1.57	-0.14
		BRISQUE	0.49	1.33	0.17	-1.58	0.22	-0.63
		MEAN	**0.02**	**-0.09**	**-0.20**	**0.50**	**-0.10**	**-0.13**

			M_NN	AOD	DE_Z	CY_D	FFA	DMPH
OTS	OUTDOOR	PSNR	-0.02	0.39	0.20	-1.08	1.58	-1.07
		SSIM	0.32	0.62	0.53	-1.70	0.92	-0.69
		HDR	-0.32	0.15	0.40	-0.39	1.59	-1.43
		IQVG	0.52	0.30	0.07	-1.93	0.96	0.07
		BRISQUE	0.27	0.83	-0.09	-1.94	0.54	0.39
		MEAN	**0.16**	**0.46**	**0.23**	**-1.41**	**1.12**	**-0.55**
			M_NN	AOD	DE_Z	CY_D	FFA	DMPH
D-HAZE (INDOOR)	NYU	PSNR	-0.35	-0.14	-0.04	-0.04	1.82	-1.25
		SSIM	0.21	0.39	0.45	-1.28	1.32	-1.10
		HDR	-0.30	-0.24	-0.13	0.63	1.51	-1.47
		IQVG	0.57	0.70	0.57	-1.96	0.06	0.06
		BRISQUE	0.24	0.69	0.27	-2.01	0.47	0.34
		MEAN	**0.07**	**0.28**	**0.23**	**-0.93**	**1.04**	**-0.68**
			M_NN	AOD	DE_Z	CY_D	FFA	DMPH
	MIDDLSBURRY	PSNR	-0.26	0.23	0.16	0.24	1.36	-1.72
		SSIM	0.12	0.69	0.53	-1.27	1.10	-1.18
		HDR	-0.29	0.02	0.02	0.40	1.46	-1.62
		IQVG	0.79	0.92	0.79	-1.52	-0.75	-0.23
		BRISQUE	0.64	0.61	0.56	-1.95	0.28	-0.14
		MEAN	**0.20**	**0.49**	**0.41**	**-0.82**	**0.69**	**-0.98**

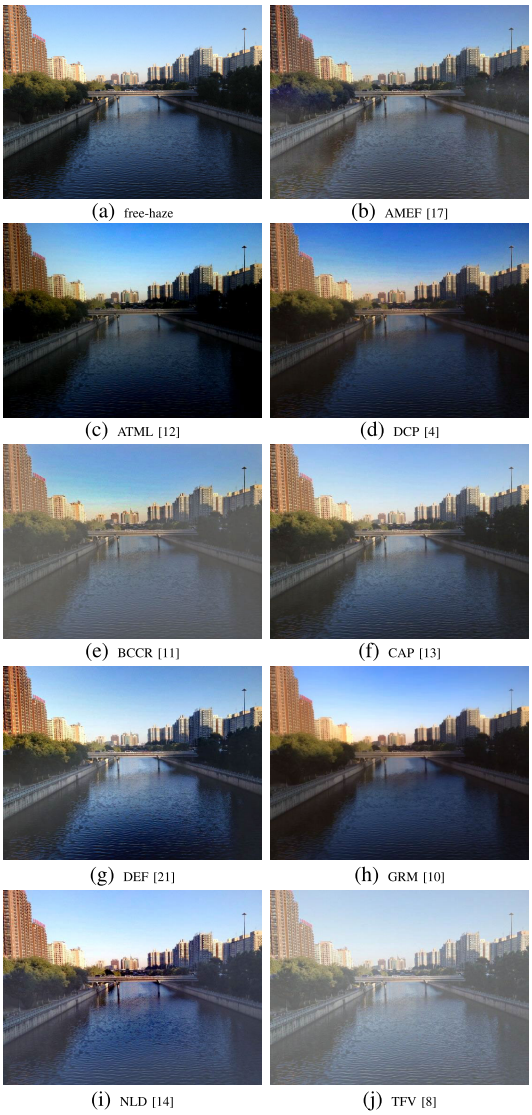


FIGURE 5. Results comparison - Image selected from the RESIDE data set processed with 9 image prior based de-hazing techniques and compared to the free-haze (a).

consists of an element-wise multiplication layer and several element-wise addition layers to generate the recovery image via the modified atmospheric scattering model involving the optimized parameter K. The K-estimation module is the critical component of AOD, being responsible for estimating the depth and relative haze level. AOD even though the winner among deep learning methods, shows significant variability among different metrics, data sets and environments. It is worth noticing that the performance of certain deep-learning methods is severely affected by the data set used e.g. CY_D and *FFA* in Figure 8.

Examples of dehazing are shown in Figures 5 and 7 for image prior and deep-learning based techniques respectively. One observes how haze may not be completely removed (TFV, AOD), introducing clear artifacts (CY_D, BCCR), contrast distortion (ATML, GRM, DCP) and washing out details as in NLD.

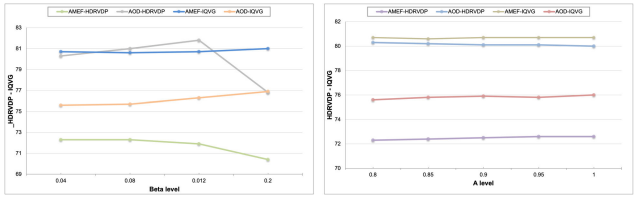


FIGURE 6. Objective evaluation varying - left: β levels keeping A fixed to the value of 0.8; - right: varying A levels keeping β fixed to the value of 0.04, for both metrics HDR-VDP and IQVG.

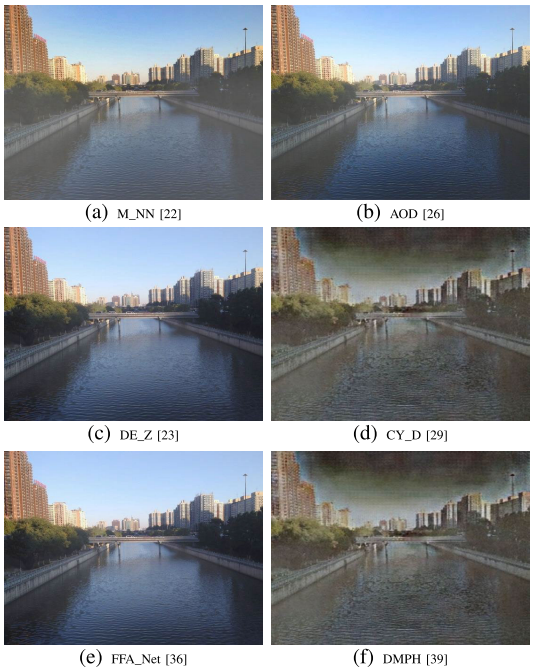


FIGURE 7. Results comparisons - Image from the RESIDE data set and processed with the 6 deep-learning based de-hazing techniques. The original free-haze image is shown in Figure 5.

B. VARYING β AND A

Atmospheric light A and the scattering coefficient of the atmosphere β are playing a role in generating a color shift and adding haze into the free-haze image. RESIDE 2018 data set provides a set of images where 35 levels of haze distortion are added to the free-haze image, e.g., 5 levels of A and 7 levels of β.

Firstly, how the quality of the de-hazed image may be influenced by β was investigated, using the levels of $[0.04, 0.06, 0.12, 0.2]$ where A has been fixed to 0.8. Then, we investigated how the quality of the de-hazed image may be influenced by the parameter A in the range of $[0.8, 1.0]$ with a step of 0.5, where β is fixed to the value of 0.04. This means testing all de-hazing methods used here, on 9 different data sets constituted of $2,061$ images each, in a total of $18,549$ images under all 6 objective metrics. Due to the high computational time required to perform this test, analysis was restricted to the 2 best performing techniques (AMEF [17] and AOD [26]), belonging to the prior based and deep learning based de-hazing techniques respectively, using 2 objective metrics; HDR-VDP for full reference

TABLE 7. Ranking of 6 deep learning dehazing techniques based on their combined outdoors-indoors overall perforce across all tested data. Each column hold the means of Table 4. AOD exhibits the best mean performance across all data sets. Standard deviations measure each method's sensitivity to different data sets. The MEAN STD is calculated to measure the overall degree of dependency of deep learning methods' performance to different data sets and will be compared to the MEAN STD of prior based methods.

DEEP LEARNING METHODS - PERFORMANCE COMPARISON						
	M_NN	AOD	DE_Z	CY_D	FFA	DMPH
NTIRE19-IN	0.54	0.44	0.01	-0.07	-0.88	-0.03
NTIRE19-OUT	0.04	0.38	-0.31	0.72	-1.14	0.31
NTIRE18-IN	0.05	0.13	0.26	-0.07	-0.19	-0.18
NTIRE18-OUT	0.02	-0.09	-0.20	0.50	-0.10	-0.13
OTS-OUT	0.16	0.46	0.23	-1.41	1.12	-0.55
D_H-NYU-IN	0.07	0.28	0.23	-0.93	1.04	-0.68
D_H-MIDD-IN	0.20	0.49	0.41	-0.82	0.69	-0.98
MEAN	0.15	0.30	0.09	-0.30	0.08	-0.32
STD	0.18	0.21	0.27	0.78	0.90	0.44
MEAN STD						0.46

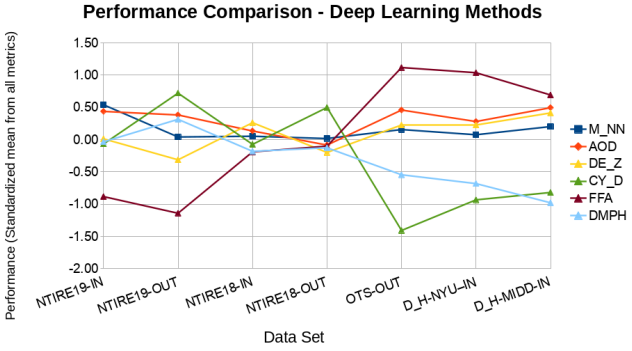


FIGURE 8. Performance graph of 6 deep learning dehazing techniques based on their mean score from 5 objective metrics. The data for each method are the columns in Table 7.

and IQVG for no-reference. The plots for HDR-VDP and IQVG objective metrics are shown in Figure 6 (left) when varying β.

One may notice here an interesting result. De-hazing techniques reproduce a suitable output image independently from the β value, the haze level applied to the free-haze image. Differences of just two points in the quality score, predicted by the two metrics, is not perceivable by the human visual system. The plots for varying A are shown in Figure 6 (right) as above. It may also be noticed that tested de-hazing techniques are capable of maintaining a suitable de-hazed image quality, independently from the level of color shift introduced by parameter A.

C. TIME ANALYSIS

To assess the efficiency of the proposed framework, computational times of the tested methods have been registered.

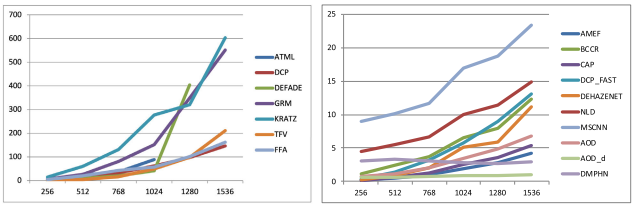


FIGURE 9. CPU time, in seconds, of each de-hazing technique for predicting the free haze image, varying the image resolution from 256 × 256 to 1536 × 1536 pixels.

Due to memory management issues, the size of the input haze image varied from 256 × 256 to 1536 × 1536 pixels. *CY_D* was excluded from this test, due to its hard requirement for an input image of 256 × 256 pixels. All experiments were performed on a Linux machine equipped with an Intel CPU 4 Core i.5 – 7500 (2.40 GHz) with 16 Gb of memory and Intel GPU HD Graphics 630(Kaby Lake GT2).

Based on their computational performance, de-hazing techniques were separated in two broad classes, one with the techniques having very high computational cost shown in Figure 9 (left) and the other class with the techniques that were faster, shown in Figure 9 (right). Some of the slow techniques have issues related to memory management when image resolution goes above 1024 × 1024 for ATML and 1280 × 1280 for DEFADE. Large improvement in computational cost is achieved by the techniques shown in Figure 9 (right), where DCP_F, AMEF, CAP and AOD are the ones with the fastest performance. However, there were cases where DCP_F produced halos artifacts around objects, as shown in Figure 2.

Time analysis must be assessed in relation to the computational complexity of the techniques used. For simplicity and without loss of generality, an input matrix of dimensions $n \times n$ is considered. Prior-based and fusion-based approaches may vary in their computational complexity, from $\mathcal{O}(n)$, when a simple subtraction or addition on the input image is required, to that of $\mathcal{O}(n^3 logn)$ when filtering and/or pre-processing of the input image are necessary, i.e., gamma-correction, Gaussian pyramid or other type of filtering. For deep-learning approaches, the computational complexity is considered for the inference process (forward propagation). Let n be the dimension of the input vector. Without loss of generality, one can consider each convolutional layer to correspond to a matrix multiplication, with matrix dimensions $n \times n$ and an activation function that follows. These are the two main operations performed at each layer. A matrix multiplication has a complexity of $\mathcal{O}(n^3)$, while for the activation function the complexity is of $\mathcal{O}(n)$. Considering n numbers of layers, having the same number of weights and parameters, the total computational complexity for a deep-learning approach under the above assumptions is $\mathcal{O}(n^4)$.

D. TASK-DRIVEN EVALUATION

The proposed framework incorporates a task driven evaluation using the annotated data set available in RESIDE

TABLE 8. Task-driven results for 5 object detection techniques on the original hazed images of the *RTTS* data set and on the de-hazed images obtained with the AMEF, DCP, M_NN and AOD de-hazing techniques. The results are the *Mean Average Precision* (*mAP*) among all objects.

Method	Haze	AMEF [17]	DCP [4]	M_NN [22]	AOD [26]
SSD_{mob_v1}	14.61	15.46	15.24	14.15	14.44
SSD_{mob_v2}	15.23	16.15	15.90	14.79	15.28
SSD_{inc_v2}	17.02	18.12	17.64	16.55	16.79
$FRCNN_{inc\_v2}$	27.50	27.76	27.09	27.32	26.75
$FRCNN_{res\_101}$	31.58	30.80	31.37	31.08	30.43

2018 as was also discussed in Section III. This annotated part is called RTTS and includes 4, 322 images, of which 35 images have been removed for memory management issues related to some of the tested de-hazing techniques, (e.g., DCP), reducing the total number of images to 4287. To reduce the required computational cost only the two best techniques for each category were tested, e.g., AMEF and DCP, for prior based techniques, *M_NN* and AOD for deep-learning based techniques.

Traditional object detection was performed both on the original hazed RTTS image, as well as the resulting from each of the four best methods de-hazed image and the results were compared to discover whether any method of dehazing helped recognition. Tensorflow Object Detection API [53], was used. For object detection, the methods used were: SSD [54], FRCCN [55] as feature extractors and mobilnet v1 and v2 [56], inception v2 [57] and *resenet*_101 [58] deep-learning networks. All object detection methods have been trained on the COCO2012 data set [59]. The dataset used for object detection consists of 5 different annotated objects, person, bicycle, car, bus and motorbike; and the metric used for evaluation was the one used in the PASCAL Visual Object Classes Challenge 2012 (VOC2012) [60]. The results are reported in Table 8 as Mean Average Precision (mAP) among all objects. Column labeled *Haze* holds the results of detection on the corresponding haze images before dehazing.

AMEF provides an overall improvement in all cases except FRCCN *resenet*_101, while DCP provides either limited or equivalent improvement when compared to the performance on the original RTTS haze data set (ground truth). *M_NN* and AOD are not improving the performance. This may due to the fact that state-of-the-art de-hazing techniques are not always capable to completely remove haze from the input image, as shown in Figure 10. Often these techniques may introduce other types of artifacts that may mislead certain computer vision applications. Future de-hazing techniques therefor need to be designed with a specific application in mind. A general-purpose de-hazing technique is very difficult to design due to two major factors. First, imposed design constraints may not be generalized on various applications as a general-purpose technique will require. Second, design imposed constraints on specific computer vision tasks may not be able to take advantage of the improved quality of the de-hazed image.

(a) DCP [4] (b) AMEF [17] (c) AOD [26] (d) M_NN [22]

FIGURE 10. Task-driven comparison: either haze is not removed completely or blocking artifacts are injected in the de-hazed image.

E. DEPENDENCY TO TRAINING SETS

The choice of training/testing data sets affect the performance of all methods as is demonstrated in the experiments. However, this dependency is stronger in the deep learning based approaches as was initially suspected. To measure this dependency, the mean standard deviation of the standardized scores of Tables 4 and 6 has been calculated for prior based and deep learning methods respectively. As is shown in the respective Tables 5 and 7, the mean standard deviation of deep learning methods is greater than that of the prior based methods. This proves that the performance of deep learning methods is more dependent to the data sets used for training. This result is not surprising since this dependency is well known for all Neural Network architectures and it emerges as an advantage of the prior based methods in dehazing tasks. As an example of such a behavior, one observes in Table 7 and the resulting Figure 8, FFA and CY_D exhibiting high variability in performance, depending on the data set used for testing. On the other hand, according to the results reported in Table 5 and the resulting Figure 4, prior based methods seem more consistent in performance across different data sets.

VI. CONCLUSION

A general framework for an objective evaluation of dehazing techniques has been introduced in this paper. It consists of a selection/fusion of several data-sets and performance metrics in a *unifying* context of statistical origin.

The framework was validated by means of a collection of state of the art dehazing techniques. Several advantages of the proposed framework were revealed in identifying diverse performance characteristics and *objectifying* global evaluation results.

The framework can be used to evaluate current and future dehazing techniques in an unbiased manner, on different aspects of performance and image capturing conditions.

REFERENCES

[1] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[2] Z. Gu, L. Zhang, and H. Li, "Learning a blind image quality index based on visual saliency guided sampling and Gabor filtering," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 186–190.

[3] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 492–505, Jan. 2019.

[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[5] R. Fattal, "Single image dehazing," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 72:1–72:9, Aug. 2008.

[6] R. Fattal, "Dehazing using color-lines," *ACM Trans. Graph.*, vol. 34, no. 1, pp. 13:1–13:14, Dec. 2014.

[7] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[8] J.-P. Tarel and N. Hautière, "Fast visibility restoration from a single color or gray level image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2201–2208.

[9] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Washington, DC, USA: IEEE Computer Society, Jun. 2014, pp. 2995–3002.

[10] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 9906, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46475-6_36.

[11] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*. Washington, DC, USA: IEEE Computer Society, Dec. 2013, pp. 617–624.

[12] M. Sulami, I. Glatzer, R. Fattal, and M. Werman, "Automatic recovery of the atmospheric light in hazy images," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2014, pp. 1–11.

[13] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.

[14] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.

[15] L. Kratz and K. Nishino, "Factorizing scene albedo and depth from a single foggy image," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep. 2009, pp. 1701–1708.

[16] K. Nishino, L. Kratz, and S. Lombardi, "Bayesian defogging," *Int. J. Comput. Vis.*, vol. 98, no. 3, pp. 263–278, Jul. 2012.

[17] A. Galdran, "Image dehazing by artificial multiple-exposure image fusion," *Signal Process.*, vol. 149, pp. 135–147, Aug. 2018.

[18] M. Zheng, G. Qi, Z. Zhu, Y. Li, H. Wei, and Y. Liu, "Image dehazing by an artificial image fusion method based on adaptive structure decomposition," *IEEE Sensors J.*, vol. 20, no. 14, pp. 8062–8072, Jul. 2020.

[19] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–23, 2021.

[20] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3271–3282, Aug. 2013.

[21] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3888–3901, Nov. 2015.

[22] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 9906, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46475-6_10.

[23] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.

[24] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3194–3203.

[25] Y. Pang, J. Nie, J. Xie, J. Han, and X. Li, "BidNet: Binocular image dehazing without explicit disparity estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5931–5940.

[26] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4780–4788.

[27] H. Zhang, V. Sindagi, and V. M. Patel, "Multi-scale single image dehazing using perceptual pyramid deep network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1015–1024.

[28] K. Swami and S. Kumar Das, "CANDY: Conditional adversarial networks based fully end-to-end system for single image haze removal," 2018, *arXiv:1801.02892*. [Online]. Available: http://arxiv.org/abs/1801.02892

[29] D. Engin, A. Genc, and H. K. Ekenel, "Cycle-dehaze: Enhanced CycleGAN for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 825–833.

[30] X. Yang, Z. Xu, and J. Luo, "Towards perceptual image dehazing by physics-based disentanglement and adversarial training," in *Proc. AAAI*, 2018, pp. 7485–7492.

[31] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3253–3261.

[32] A. Dudhane, H. S. Aulakh, and S. Murala, "RI-GAN: An end-to-end network for single image haze removal," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2019, pp. 2014–2023, doi: 10.1109/CVPRW.2019.00253.

[33] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *Int. J. Comput. Vis.*, vol. 128, no. 1, pp. 240–259, Jan. 2020.

[34] R. Li, X. Zhang, S. You, and Y. Li, "Learning to dehaze from realistic scene with a fast physics based dehazing network," 2020.

[35] B. Li, Y. Gou, S. Gu, J. Z. Liu, J. T. Zhou, and X. Peng, "You only look yourself: Unsupervised and untrained single image dehazing neural network," *CoRR*, vol. abs/2006.16829, 2020. [Online]. Available: https://arxiv.org/abs/2006.16829

[36] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proc. 34th AAAI Conf. Artif. Intell. (AAAI), 32nd Innov. Appl. Artif. Intell. Conf. (IAAI), 10th Symp. Educ. Adv. Artif. Intell. (EAAI)*, New York, NY, USA: AAAI Press, Feb. 2020, pp. 11908–11915. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/6865

[37] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing," in *Proc. 34th AAAI Conf. Artif. Intell. (AAAI), 32nd Innov. Appl. Artif. Intell. Conf. (IAAI), 10th Symp. Educ. Adv. Artif. Intell. (EAAI)*, New York, NY, USA: AAAI Press, Feb. 2020, pp. 10729–10736. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/6701

[38] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2157–2167.

[39] S. D. Das and S. Dutta, "Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 1994–2001, doi: 10.1109/CVPRW50498.2020.00249.

[40] J. Shen, Z. Li, L. Yu, G.-S. Xia, and W. Yang, "Implicit Euler ODE networks for single-image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 877–886.

[41] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, "D-HAZY: A dataset to evaluate quantitatively dehazing algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2226–2230.

[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[43] M. Luo, "The cie 2000 colour difference formula: Ciede2000," *Color Res. Appl.*, vol. 4421, pp. 340–350, Jun. 2002.

[44] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," 2018, *arXiv:1804.05091*. [Online]. Available: http://arxiv.org/abs/1804.05091

[45] Y. Li, S. You, M. S. Brown, and R. T. Tan, "Haze visibility enhancement: A survey and quantitative benchmarking," *Comput. Vis. Image Understand.*, vol. 165, pp. 1–16, Dec. 2017.

[46] Q. Wu, J. Zhang, W. Ren, W. Zuo, and X. Cao, "Accurate transmission estimation for removing haze and noise from a single image," *IEEE Trans. Image Process.*, vol. 29, pp. 2583–2597, 2020.
[47] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2805–2814.
[48] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 1014–1018, doi: 10.1109/ICIP.2019.8803046.
[49] (2003-2014). *Middleburry Dataset*. [Online]. Available: http://vision.middlebury.edu/stereo/data/
[50] P. K. N. Silberman, D. Hoiem, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. ECCV*, 2012, pp. 746–760.
[51] H. Israël and F. Kasten, "KOSCHMIEDERs theorie der horizontalen sichtweite," in *Die Sichtweite im Nebel und die Möglichkeiten Ihrer Künstlichen Beeinflussung* (Forschungsberichte des Wirtschafts- und Verkehrsministeriums Nordrhein-Westfalen), vol. 640. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften, 1959, doi: 10.1007/978-3-663-04661-5_2.
[52] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 40:1–40:14, Jul. 2011.
[53] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3296–3297.
[54] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46448-0_2.
[55] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2015, pp. 91–99.
[56] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861
[57] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[59] *Coco Dataset*. Accessed: Oct. 15, 2019. [Online]. Available: http://cocodataset.org/home
[60] *Pascal 2012 Competition*. Accessed: Oct. 15, 2019. [Online]. Available: http://host.robots.ox.ac.uk/pascal/VOC/

ALESSANDRO ARTUSI received the Ph.D. degree in computer science from the Vienna University of Technology, in 2004. He is currently the Managing Director of the DeepCamera Group, CYENS, Cyprus. Recently he has joined, representing CYENS as one of the funding members, and the Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI), a not-for-profit standards organization established in Geneva. His research interests include visual perception, image/video processing, computer graphics, high dynamic range technology, objective/subjective imaging/video evaluation, deep-learning, computer vision, and color science with particular focus to deploy the next generation of imaging/video pipeline. He has been active member of standardization committees, such as JPEG and MPEG. He has been a member of the IST/037 coding of picture, audio, multimedia, and hypermedia information, and the British Standard Institute (BSI). He has been one of the editors of the JPEG-XT standard for encoding High Dynamic Range content.

KONSTANTINOS A. RAFTOPOULOS received the Ptychion in mathematics from the National and Kapodistrian University of Athens (UOA), in 1994, the M.Sc. degree in computer science from the University of California at Los Angeles (UCLA), in 1996, the Ph.D. degree in computer science from the National Technical University of Athens (NTUA), and the E.D. degree from UCLA, in 2002. From 2004 to 2012, he was leading a consortium with SIEMENS for manufacturing state of the art mechatronic systems for the packaging industry. Since 2012, he has been collaborated with several startups in France, U.S., U.K., Singapore, and Greece, as an expert in machine vision and machine learning applications. He is currently an Active Researcher with tier 1 publications in the fields of computer vision, machine learning, pattern recognition, and data science.