DOMAIN ADAPTATION IN IMAGE DEHAZING: EXPLORING THE USAGE OF IMAGES FROM VIRTUAL SCENARIOS

Angel D. Sappa[1,2], Patricia L. Suárez[1], Henry O. Velesaca[1], Darío Carpio[1]
[1]*Escuela Superior Politécnica del Litoral, ESPOL,*
Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863
[2]*Computer Vision Center, Campus UAB*
08193 Bellaterra, Barcelona, Spain

ABSTRACT

This work presents a novel domain adaptation strategy for deep learning-based approaches to solve the image dehazing problem. Firstly, a large set of synthetic images is generated by using a realistic 3D graphic simulator; these synthetic images contain different densities of haze, which are used for training the model that is later adapted to any real scenario. The adaptation process requires just a few images to fine-tune the model parameters. The proposed strategy allows overcoming the limitation of training a given model with few images. In other words, the proposed strategy implements the adaptation of a haze removal model trained with synthetic images to real scenarios. It should be noticed that it is quite difficult, if not impossible, to have large sets of pairs of real-world images (with and without haze) to train in a supervised way dehazing algorithms. Experimental results are provided showing the validity of the proposed domain adaptation strategy.

KEYWORDS

Domain adaptation; Synthetic hazed dataset; Dehazing.

1. INTRODUCTION

Images captured by sensors sensitive to the visible spectrum are affected by natural phenomena such as fog, dust, mist, rain, snow, etc. This considerably reduces their quality, so processes such as detection, segmentation and classification of objects, among others, are considerably affected, for which they will not be able to obtain results that meet the required objectives. The effect of haze on image quality is the result of random light scattering and therefore affects all pixels in the image. There are several methods to enhance images by removing haze, one of them is the estimation of the optical transmission in blurred scenes given a single input image (Fattal, 2008). In another approach, presented in (Wang et al., 2017), a fast algorithm for removing haze from a single image based on linear transformation is proposed. It assumes that there is a linear relationship in the minimum channel between the blurred image and the dehazed image. Ju et al. (2017) proposes an algorithm to address the inherent weaknesses of the atmospheric dispersion model to develop a way to smoothly remove haze using an adaptive method to adjust the transmission of the scene based on the characteristics of the scene.

In recent years, different convolutional neural networks (CNN) based approaches for haze removal have been proposed; these approaches achieve the best results in the haze removal problem. for example, Suárez et al. (2018) propose a stacked generative adversarial network (GAN) that has achieved outstanding results. There are other CCN based approaches for haze removal that relies on unpaired images for training a generative model (Anvari et al., 2020); it could be also mentioned the Cycle GAN based approach from (Engin et al., 2018). Despite the promising results of convolutional networks, it is evident that each of the approaches mentioned above needs a set of haze & clear image data for training, which is not very easy given the complexity of obtaining them. There are just a few datasets available on the literature with real haze and the corresponding dehazed real-world images (see Fig. 1).

Figure 1. Pairs of real-world clear/haze images from the dataset presented in (Lüthen et al., 2017).

Some of the techniques based on CNN have been trained with pairs of real-world images (clear) together with the corresponding synthetized haze images; an example of this strategy is the OTS RESIDE Beta dataset 2021 (Li et al. 2018), which has 2D synthetized haze images for each of the given real clear images. Unfortunately, haze images are obtained without depth information, hence obtained haze images do not follow real-world haze scenarios. In order to overcome the limitation of obtaining real-world images with haze, 3D models of different scenarios are required in order to simulate realistic haze image datasets. It should be mentioned that the usage of 3D virtual environments to generate dataset of synthetic images has already been considered for tackling different computer vision problems for instance object recognition (e.g., pedestrian (Fabbri et al., 2021), vehicles (Tang et al., 2019)), camera calibration (Charco et al., 2018; Charco et al., 2020; Charco et al., 2021), just to mention a few. In the current work, a similar strategy is followed to address the problem of image dehazing. The main contributions in this work are summarized below:

1) Datasets generation: a dataset with pairs of synthetic hazy/clear image has been generated using a realistic 3D scene simulator; multiple scenarios and different levels of haze are considered.
2) Training strategy: the dataset with synthetic haze/clear images is used to train the network and then transfer the learned knowledge to real environments where maybe just a few samples of real images are available.
3) Empirical analysis: results obtained from the proposed domain adaptation strategy (i.e., training with synthetic images and then fined tuned with real images) are analyzed for the model that solves the problem. Comparison with and without the proposed transfer learning are provided.

The manuscript is organized as follows. Section 2 briefly presents work related with the haze removal problem, as well as the basic concepts of domain adaptation strategies. Then, in Section 3, the proposed strategy is detailed. Experimental results with a set of synthetic and real-world images are presented in Section 4 together with a comparison and analysis of obtained results. Conclusions are given in Section 5.

2. RELATED WORKS

Image dehazing is a well-known method to remove haze, smoke, dust, or fog from single images without any prior knowledge about the camera, calibration data, scene content, 3D structure, or environmental conditions. A quite old but still very popular formulation of the problem is given in (McCartney, 1976), where haze effects are formulated using the atmospheric scattering model. During the last two decades several approaches have been proposed in the literature, starting with the use of global or local image information techniques, as well

as non-local pre-processing to remove haze from images (Berman et al., 2016). This technique is based on counting the different colors present in the images, assuming that the variety of the color range is sufficient to achieve a good representation of the images. Following the traditional techniques, Wang et al. (2017) proposes the estimation of the effective transmission using an atmospheric luminous value to guarantee the chromatic authenticity and the contrast of the images. A novel technique based on the fusion of the result of two energy functions is proposed by (Galdran et al., 2016). Recently, several convolutional neural networks (CNN) based approaches have been proposed achieving outstanding results compared to those obtained by traditional techniques. We can start by reviewing the approach presented by Wu et al. (2021), where the authors propose a contrastive regularization based on learning both information from blurred images and clear images as negative and positive samples, respectively. The technique applies an adaptive combination and the expansion of the receptive field to improve the learning of the network called AECR-Net. Continuing with the deep learning-based methods, in (Wang et al. 2021) a technique is presented that consists of training an adaptive model of haze concentration from end to end, whose layers include a pyramidal feature extractor and a multi-scaling attention mechanism to help the generalization of the trained model. Also, in Zhao et al. (2021), a two-stage dehaze model is proposed, the first in charge of processing a haze removal framework to restore visibility and the second stage in charge of enhancing reality with contradictory learning using unpaired clear and blurred images, ending with a perceptual fusion of the results.

As mentioned above, the haze removal task is quite challenging, not only because of the complexity of the problem, but also due to the lack of large datasets with pairs of clear and haze images to train models. This lack of images is because this is a natural phenomenon of distortion produced by nature, which depends on specific climatic conditions. There are a few datasets available on the literature with real haze and the corresponding dehazed real-world images; for instance, in (Lüthen et al., 2017) the authors acquire clear images in indoor scenarios and later produce haze, at different densities, and acquire a new set of images (see Fig. 1). The main problem with this dataset is on the limited number of pair of images. Another strategy followed for tackling the limitation on datasets, consists of using synthetized haze images, which are obtained by adding haze to real clear images (Li et al., 2018). This dataset is referred to as RESIDE beta dataset. The main problem with this dataset lies on the lack of realism on the added haze, since 3D information is missed the same density of haze is added all over the clear images.

The problem mentioned above, lack of large dataset for training deep learning based approaches, is a common problem in different images processing and computer vision approaches. This problem has been tackled by using images acquired in virtual 3D scenarios (e.g., (Fabbri et al., 2021; Tang et al., 2019; Charco et al., 2018). Images from these virtual scenarios, generally referred to as synthetic images, are used for training deep learning models. Then, trained models are adapted (e.g., fine-tuned) to work with real images. Some issues between feature space and layout could arise in real-world applications due to factors like lighting change, image quality and scale, just to name a few. To solve this, a process called Domain Adaptation can be used, where data labeled in one or more relevant source domains is applied to new tasks in a destination domain. Some superficial DA methods have been proposed to solve a domain switching problem between two domains. These can be divided into two classes: instance-based DA (Bruzzone et al., 2010; Chu et al., 2013) and function-based DA (Pan et al., 2011; Marzieh et al., 2018). Both classes of DA reduce the discrepancy between domains. The first selects certain pieces of data from the auxiliary datasets to compose the intermediate domains to train the deep network. These auxiliary datasets could be obtained by combining certain parts of the source and target data. For the second, the shared space is learned to match the distributions of both domains (i.e., source and destination). Some methods based on deep learning have recently been proposed, in which different DA strategies are integrated in the learning process (Wang et al., 2018). Depending on the relationship between the source and destination domains, knowledge transfer can be performed in a single or multi-step DA. In the first case, when both domains (source and destination) are directly related, that is, the feature space is the same, it is known as homogeneous DA, which is divided into three types, such as supervised DA with labeled data, semi- supervised DA with labeled and unlabeled data and unsupervised DA with unlabeled data; on the contrary, in the second case, when the feature spaces between the source and destination domains are not equivalent, it is known as heterogeneous DA. In the second case, that is, multi-step DA, an intermediate domain is used, which must be closely related to the source and destination domains.

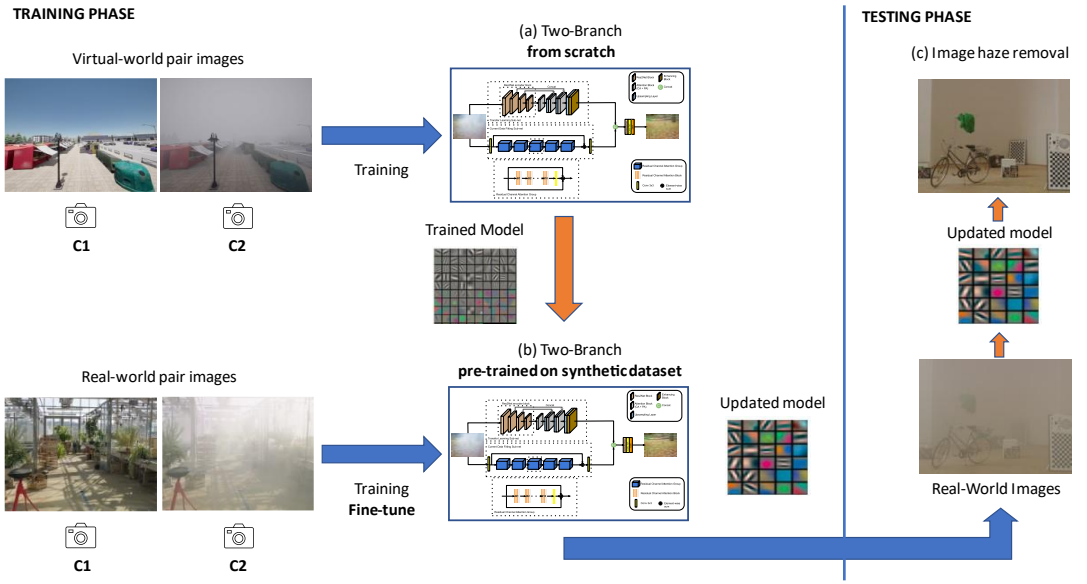

Figure 2. (a) Architecture trained using synthetic images. (b) Knowledge from the synthetic domain is adapted to be used with real-world images. (c) Updated weights after DA strategy used to evaluate image haze removal.

3. PROPOSED APPROACH

This section describes the strategy proposed for training haze removal approaches when few real pairs of (clear/haze) images are provided. This section stars providing details on the CNN architecture selected to carry out the haze removal. Additionally, the synthetic image dataset generation, which is acquired from virtual 3D outdoor environments, is presented. The CNN architecture is first trained with the synthetic dataset and then adapted to real world by means of a domain adaptation approach. It should be noted that the proposed strategy is valid not only for the network architecture used in this work, but it can be considered during the training of any model based on CNN. Figure 2 shows an illustration of the strategy proposed for training the CNN model in the current work (the Two-Branch (Yu et al., 2021) network has been selected).

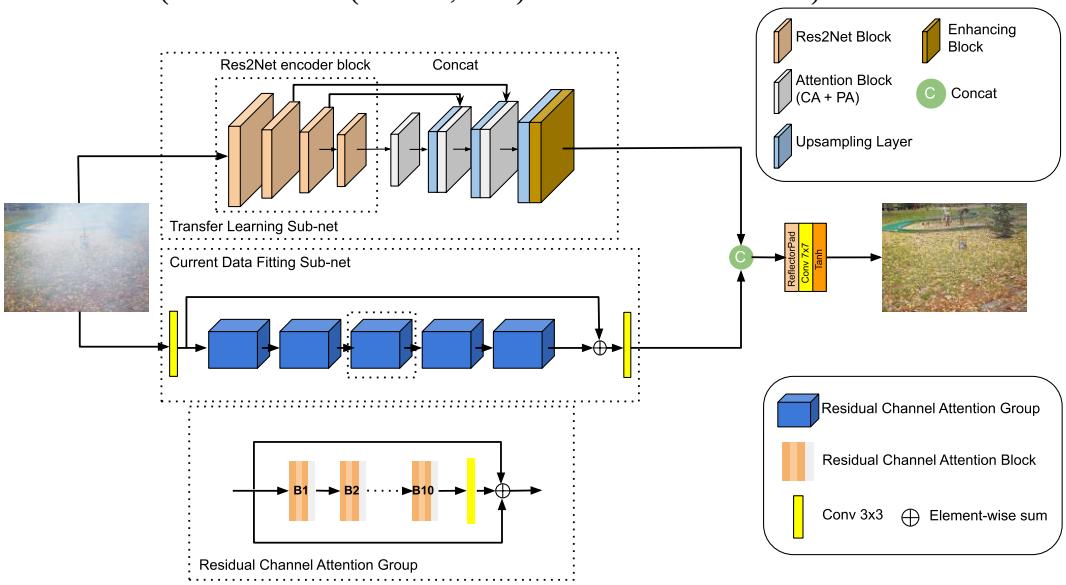

Figure 3. Dehaze model proposed by Yu et al. (2021)

3.1 Network architecture

This section presents details of the network architecture used as a benchmark to evaluate the proposed Domain Adaptation (DA) strategy. The architecture selected after analyzing a variety of state-of-the-art haze removal techniques has been the one proposed by Yu et al. (2021) since it has been designed with two branches, one that contains data-specific learning techniques (attention per channel blocks) and attention per pixel and the other branch that focuses on learning transfer. Because of the focused learning, the haze can be removed in a more adaptive way to its spatial representation. An illustration of the architecture is presented in Figure 3; the model consists of two learning branches, specializing in the removal of the haze present in the images, with a very limited dataset. The model performs a mapping of the most relevant characteristics and at the same time learns a domain transfer representation of another trained network, to facilitate the convergence of the model. In order to validate the results of this model, the authors have used real-world and synthetized haze images. It is worth mentioning that this model has also validated the removal of non-homogeneous haze, which allows us to evaluate this type of scenario in the future with our DA strategy.

Figure 4. Examples of pairs of images (clear and haze) captured from the different 3D scenarios proposed in the current work; these pairs of images are used during the training, testing and evaluation stages.

3.2 Synthetic dataset generation

This section presents the steps followed to generate the datasets with synthetic images, which are used to train the CNN architecture. The synthetic images are obtained from one of the virtual worlds configured by default in CARLA Simulator (Dosovitskiy et al., 2017). Among the different options available in the CARLA simulator is that of modifying the climate settings, this option was used to generate a climate with different intensities of haze in the virtual scenes. In addition, to obtain a dataset with greater variability of structures, shapes, colors and textures, different scenarios of the virtual worlds available by default in CARLA were explored. The different synthetic images used in this work are generated from the CARLA simulator, as detailed below. As a first task, the trajectory to be followed within the scenarios of the virtual world is defined.

Two virtual cameras are defined which move synchronously around the scene. The attributes of both cameras are defined as follows: width = 1600, height = 1200 and FOV = 100. The difference between both cameras is that camera #1 is used to obtain the (clear) ground truth images in RGB format while camera #2 adds a haze density percentage that varies from 20% to 60% in steps of 10%. For each camera and step 50 images are generated. The sequence is executed to generate 6 synthetic datasets with 250 image pairs (i.e., GT and haze image) for each one. At the same time that the configured cameras follow the predefined trajectory, their pose randomly varies at each instant of time (i.e., on the translation the X coordinate could take random values in between [−2, 2]m, the Y coordinate [− 2, 2]m, while the Z coordinate [−1, 1]m; regarding the orientations, the pitch angle could take random values in between [0, 7] degrees, yaw angle [−7, 7] degrees while roll angle [−7, 7] degrees). These values were defined according to the characteristics of each scene. Figure 4 shows examples of images of the synthetic dataset generated from 6 3D virtual scenarios and different haze intensity levels.

4. EXPERIMENTAL RESULTS

As described above, this work addresses the problem of haze removal in different real-world scenarios when the required number of images needed to train the model based on deep learning is not available. This problem of lack of data is addressed through a DA strategy. This section first presents details of results obtained when just a set with few pairs of real images is considered; then, experimental results with the proposed DA strategy are presented. In short, this strategy consists of first training with synthetic images, obtained from the CARLA simulator, and then transferring this learned knowledge to a real-world training process. The architecture evaluated in this work has been implemented in TensorFlow-Keras and trained with an NVIDIA Titan XP GPU and an Intel Core I9 3.3GHz CPU. As mentioned above, the Two Branch architecture is firstly trained from scratch with the dataset presented in (Lüthen et al., 2017), Figure 5 shows illustrations of results obtained by Two Branch network for qualitative evaluation; PSNR and SSIM are provided in Table 1 for a quantitative evaluation.

Figure 5. (*left*) Real clear image; (*middle*) Real haze image; (*rig*ht) Result from the Two Branch network trained from scratch with Lüthen et al. dataset.

As can be appreciated from Figure 5 and Table 1, having just a few sets of pairs of images for training the CNN architecture constrain the quality of results. In order to overcome this limitation, in this work a DA strategy is proposed. It consists of training the Two Branch network from scratch but with the acquired synthetic (clear/haze) images; then the resulting trained network is fine-tuned with images from the dataset presented in (Lüthen et al., 2017). Figure 6 shows the illustration of the results with the synthetic images (Fig. 6 (*bottom-left*)) as well as after the DA strategy (Fig. 6 (*bottom-right*)). It can be appreciated that the result from the proposed DA is considerably better than the one presented in Fig. 5(*right*); quantitative results presented in Table 1 confirm this improvement. In order to appreciate the need for fine tuning the model, result obtained by using the model just trained with synthetic images, no DA applied, are presented in Fig. 5 *(right)* and Table 1 *(4th row)*. Finally, in order to complement the analysis, the Two Branch network has been also trained from scratch but with the synthesized dataset RESIDE (Li et al., 2018). As can be seen in Table 1 (2*nd row*), the results do not overcome those obtained with the synthetic dataset (*4th row*). This is mainly caused by the lack of realism in the generation of the haze. Finally, the network trained with RESIDE is also used with the proposed DA strategy; in other words, the weights of the network are used as initialization to fine-tune

them with the dataset presented in (Lüthen et al., 2017), quantitative results are depicted in Table 1 (*3rd row*). The lack of realism in RESIDE dataset do not help to converge the training of the DA process.

Figure 6. (*top-left*) Real clear image; (*top-right*) Real haze image; (*bottom-left*) Result from the Two Branch trained with our synthetic dataset. (*bottom-right*) Result from the Two Branch trained with our synthetic dataset and fine-tuned with Lüthen et al. dataset.

The illustrations presented in Figure 6 and the results shown in Table I confirm that training with more data, even if it is synthetic, favors the convergence of the model and its corresponding generalization. It can be evidenced that the DA strategy applied in the training of few data of the real world, has helped to improve the convergence and avoid the excessive fit of the model when performing the haze removal. Undoubtedly, the features that are relevant to solve the problem have been transferred from the synthetic domain to the real world.

Table 1. Results from the Two Branch network trained with different strategies

Strategy	PSNR	SSIM
Trained from scratch with Lüthen et al.	18,303	0,749
Trained with RESIDE dataset and validated with Lüthen et al.	18,216	0,264
DA from RESIDE (trained with RESIDE images and fine-tuned with Lüthen et al.)	15,826	0,816
Trained with our synthetic images and validated with Lüthen et al.	21,624	0,867
Proposed DA (trained with our synthetic images and fine-tuned with Lüthen et al.)	**24,067**	**0,922**

5. CONCLUSION

This paper addresses the challenging problem of haze removal of real images with few labeled samples. This problem is addressed through a domain adaptation strategy, which avoids the need to have a large dataset of real-world images for training. The proposed DA strategy overcomes the approaches presented in the literature, which usually use synthetized images that do not represent the atmospheric haze model that is dependent on the distance and depth of the objects present in the scene. The proposed strategy consists of first training with synthetic images and then transferring this knowledge to real image scenarios. The knowledge transfer is performed with a DA strategy that fine-tuned the network parameters with few samples of real-world (clear/haze) images. As future work, we will consider explore DA strategy analyzing the model behavior results between different percentage of haze on the images.

REFERENCES

Anvari, Z., and Athitsos, V., 2020. Dehaze-GLCGAN: unpaired single image de-hazing via adversarial training. *arXiv preprint arXiv:2008.06632*.

Bruzzone, L., and Marconcini, M., 2010. Domain Adaptation Problems: A DASVM Classification Technique and a Circular Validation Strategy. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol 32, No. 5, pp. 770-787.

Charco, J. L., Vintimilla, B. X., and Sappa, A. D., 2018. Deep Learning Based Camera Pose Estimation in Multi-view Environment. *14th Int. Conf. on Signal-Image Technology & Internet-Based Systems*, 2018, pp. 224-228.

Charco, J. L., Sappa, A. D., Vintimilla, B. X., and Velesaca, H. O. 2020. Transfer Learning from Synthetic Data in the Camera Pose Estimation Problem. *15th Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pp. 498-505.

Charco, J. L., Sappa, A. D., Vintimilla, B. X., and Velesaca, H. O. 2021. Camera pose estimation in multi-view environments: From virtual scenarios to the real world. *Image and Vision Computing*, Vol. 110, pp. 1-12.

Chu, W., De la Torre, F., and Cohn, J. F., 2013. Selective Transfer Machine for Personalized Facial Action Unit Detection. *Proceedings of the Conference on Computer Vision and Pattern Recognition*. pp. 3515-3522.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V., 2017. CARLA: An Open Urban Driving Simulator. *Proceedings of the 1st Annual Conference on Robot Learning. pp. 1-16.*

Engin, D., Genç, A., and Kemal Ekenel, H., 2018. Cycle-dehaze: Enhanced cyclegan for single image dehazing. *Int. Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 825-833.

Fabbri, M., Brasó, G., Maugeri, G., Cetintas, O., Gasparini, R., Ošep, A., and Cucchiara, R., 2021. MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking?. *Int. Conf. on Computer Vision*, pp. 10849-10859.

Fattal, R., 2008. Single image dehazing. *ACM transactions on graphics*, Vol. 27, No. 3, pp. 1-9.

Galdran, A., Vazquez-Corral, J., Pardo, D., and Bertalmio, M., 2016. Fusion-based variational image dehazing. *IEEE Signal Processing Letters*, Vol. 24, No. 2, pp. 151-155.

Ju, M., Zhang, D., and Wang, X., 2017. Single image dehazing via an improved atmospheric scattering model. *The Visual Computer*, Vol. 33, No. 12, pp. 1613-1625.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., and Wang, Z. 2018. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, Vol. 28, No. 1, pp. 492-505.

Lüthen, J., Wörmann, J., Kleinsteuber, M., and Steurer, J., 2017. A RGB/NIR data set for evaluating dehazing algorithms. *Electronic Imaging*, Vol. 12, pp. 79-87.

McCartney, E., 1976. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons.*

Marzieh, G., and Mahdieh Soleymani, B., 2015, Unsupervised domain adaptation via representation learning and adaptive classifier learning, *Neurocomputing*, Vol. 165, pp. 300-311.

Pan, S. J., Tsang, I. W., Kwok, J. T., and Yang, Q., 2011, Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*. Vol 22, No 2, pp. 199-210.

Suárez, P. L., Sappa, A. D., Vintimilla, B. X., and Hammoud, R. I., 2018. Deep learning based single image dehazing. *Int. Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 1169-1176.

Tang, Z., Naphade, M., Birchfield, S., Tremblay, J., Hodge, W., Kumar, R., and Yang, X. 2019. Pamtri: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data. *Int. Conf. on Computer Vision*, pp. 211-220.

Wang, W., Yuan, X., Wu, X., and Liu, Y., 2017. Fast image dehazing method based on linear transformation. *IEEE Transactions on Multimedia*, Vol. 19, No. 6, pp. 1142-1155.

Wang, M. and Deng, W., 2018. Deep visual domain adaptation: A survey, *Neurocomputing*, Vol. 312, pp. 135-153.

Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., and Ma, L., 2021. Contrastive learning for compact single image dehazing. *Int. Conf. on Computer Vision and Pattern Recognition*, pp. 10551-10560.

Yu, Y., Liu, H., Fu, M., Chen, J., Wang, X., and Wang, K., 2021. A two-branch neural network for non-homogeneous dehazing via ensemble learning. *Int. Conf. on Computer Vision and Pattern Recognition*, pp. 193-202.

Zhao, S., Zhang, L., Shen, Y., and Zhou, Y., 2021. RefineDNet: a weakly supervised refinement framework for single image dehazing. *IEEE Transactions on Image Processing*, Vol. 30, pp. 3391-3404.