IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 4, 2019, accepted August 16, 2019, date of publication August 30, 2019, date of current version September 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938218

Haze Removal Using Aggregated Resolution Convolution Network

LINYUAN HE[1,2], **JUNQIANG BAI**[1], **AND LE RU**[2]

[1]Unbanned system Research Institute, Northwestern Polytechnical University, Xi'an 710072, China
[2]Department of Aeronautical Engineering, Air Force Engineering University, Xi'an 710038, China

Corresponding author: Linyuan He (hal1983@163.com)

This work was supported in part by the National Natural Science Foundation of China, under Grant 61701524, and in part by the China Postdoctoral Science Foundation under Grant 2019M653742.

ABSTRACT The haze removal technique refers to the process of reconstructing haze-free images from scenes of inclement weather conditions. This task has an extensive demand in practical applications. At present, models based on deep convolution neural networks have made significant progress in the haze removal field, greatly outperforming the traditional prior and constraint methods. However, the current CNNs methods, which involve only a single input image, do not provide sufficient features to determine the optimal transmission maps for haze removal; therefore, we propose and design an aggregated resolution convolution network (ARCN) that uses multiple inputs and aggregates features from a CNN model and the adversarial loss algorithm. Experiments comparing the visual results of our network with those of several previous methods reveal substantial improvements.

INDEX TERMS Haze removal, single image dehazing, deep convolutional neural network.

I. INTRODUCTION

Due to scattering by suspended particles and atmospheric light from the scene, captured images are often accompanied by low contrast and shifted luminance characteristics, which seriously affects subsequent tasks such as automatic driving, smart cities and other technologies [1], [2]. Dehazing is a type of image processing technique for restoring haze-free scenes as much as possible.

At present, most approaches have been developed around the atmospheric scattering model, which is described as:

$$I^c(x) = J^c(x)t(x) + A^c\left(1 - t(x)\right) \tag{1}$$

where A^c is the atmospheric light, $c \in \{r, g, b\}$, $I^c(x) \in \mathbf{R}^C$ is the degraded hazy image, $J^c(x)$ indicates the underlying haze free image and $t(x)$ denotes the media transmission correlated with the scene depth $t(x) = e^{-\beta d(x)}$. Due to the importance of $t(x)$, an optimal transmission map is considered the essential prerequisite to solve the dehazing problem [3]–[6]. Currently, dehazing methods are mainly divided into two categories: artificial calibration and database learning methods. Artificial calibration methods use a large number of statistical priors to construct a variety of filters and models used to recover haze-free images. For example, For example, Tan [7] maximized the neighbourhood contrast to compute the transmission of each pixel. Under the Bayesian probability model, Nishino *et al.* [8] employed statistical features to eliminate the influence of fog and haze. Fattal [9] assumed that neighbourhood chromaticity and transmission were uncorrelated and used the statistical colour property for haze removal. In addition, colour line prior [10] has also been utilized. This approach combined an augmented Markov random field model to obtain the optimal transmission map. More importantly, the dark channel prior (DCP) [11] is assumed to be the most successful prior statistical knowledge and it is widely used in various haze removal algorithms. Additionally, Tarel and Hautière [12], Yu *et al.* [13], and He *et al.* [14] employed a median filter, a bilateral filter and a guided filter instead of soft matting. Meanwhile, some statistical constraints still exist in other approaches, such as the colour attenuation prior [15] and non-local patch prior [16]. Clearly, artificial calibration has received extensive attention and has achieved better results. However, prior based methods are not suitable for all images; consequently scholars have increasingly focused on to how to improve the universality of dehazing approaches by learning from a database. For instance, Caraffa and Tarel [17] trained a local dictionary from the FRIDA database and used it to converge the final transmission map. By revealing that a synthetic database can be highly similar to real scene features, Tang *et al.* [18] improved the accuracy of the transmission map by combining multiple colour features

The associate editor coordinating the review of this article and approving it for publication was Andrea F. Abate.

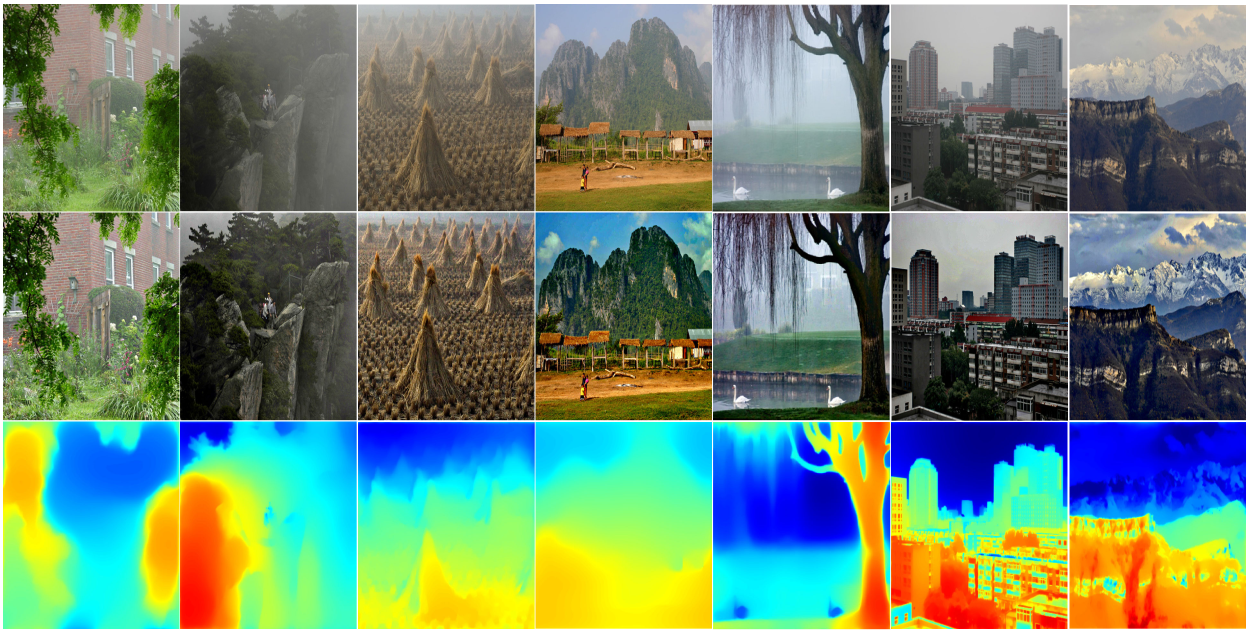

FIGURE 1. The haze-free images and depth maps restored by aggregated resolution convolution network.

with random forests. Moreover, a new deep learning-based method [19] via joint estimation clear image detail and transmission map has been proposed, which could use a global regularization method to eliminate the halos and artifacts. To sum up, applications of the above methods have inspired researchers to explore database learning-based algorithms. In particular, after compiling haze-free images into a database, Cai *et al.* [20] proposed a specific CNN to recover the transmission map. Based on the NYU database, Ren *et al.* [21] trained a multi-scale CNN to reconstruct the transmission that was able to approximate the ground truth. Recently, Li *et al.* [22] and Zhang and Patel [23] have also employed a K-estimation module and a densely connected pyramid dehazing network to enable dehazing between the transmission map and the atmospheric light. All these works indicate that CNNs have gradually played an increasingly important role in haze removal tasks. Nevertheless, the current dehazing networks consider only one input image; consequently, they are unable to capture sufficient features to indicate the optimal transmission map. In contrast, the traditional fusion methods [24] have demonstrated that using multiple input images can provide more robust feature representations. More importantly, the final feature map in [20]–[22] was computed by a series of multi-scale convolution layers. This approach can neglect the opportunity to acquire edge and circle information, which is indispensable for transmission estimation. Motivated by these observations, we argue that acquiring feature at different levels from multiple input images and aggregating different features would be more conductive for the dehazing task. Hence, the crux of the matter becomes how to take advantage of feature extraction and construct a dehazing network.

In this article, we propose a new aggregated resolution convolution network (ARCN) that capitalizes on different levels of input images to fuse and reconstruct transmission maps. Different from the previous CNN based approaches, which depend only on the multi-scale feature extraction output to construct a feature map, our network can efficiently search for a more accurate final map by using a hierarchical progression strategy and aggregated features. This model combines both high- and low-level features to generate sharp, detailed depth predictions.

1) Unlike previous dehazing networks, which concentrate on extracting deep features from a single haze image, we propose a novel input strategy that can aggregate the feature maps from multi-scale input images into a single input image. The advantage of this approach is that it can highlight the features of the input image without relying on an additional external reference image.
2) Residual and dense blocks are typical techniques used in deep networks. Here, we combine them effectively according to the dehazing task. Depth estimation is closely related to edge jumps; thus, we can use multiple dense blocks to extract features at different deep levels. To correlate these feature maps, local residuals are used to connect adjacent dense blocks to effectively aggregate features.
3) In view of the importance of edges in haze removal, we not only add TV regularization terms to estimate transmission but also apply style loss to constrain the entire image restoration. The experiments show that this approach is superior to traditional methods for restoring images.

The rest of paper is arranged as that: the typical CNN based dehazing approaches are carried out in Section II. In addition, ba The rest of paper is arranged as follows. The typical CNN based dehazing approaches are described in Section II. Then, based on the analyses of the traditional methods, we use the properties of multiple inputs to fuse more features in our network. Moreover, we construct a residual unit to extract depth-dependent features. Finally, we achieve better results using the edge preservation loss function. The experiments presented in Section IV shows that the proposed method performs equivalently or better than do other advanced dehazing methods. Finally, we provide conclusions and suggest future research in Section V.

II. RELATED WORKS

Numerous CNN based dehazing approaches have been proposed in recent years. First, we analyse several representative methods and then concentrate on their network structures, which can help us to determine the advantages of the underlying architectures.

After CNN were successfully applied in the object recognition field [25]–[27], scholars began to research CNN-based approaches for the underlying vision tasks [28]–[31]. In contrast to artificial calibration methods that use different statistical assumptions and models to estimate the transmission map, an end-to-end CNN can obtain the optimal transmission map through its cascading convolutional layers. For example, DehazeNet [20] was the first CNN based dehazing network. It includes four main operations: feature extraction, multi-scale mapping, finding the local extremum and non-linear regression. Each layer was carefully designed to reflect and represent the previously successful assumptions and priors. In other words, DehazeNet succeeded in reproducing the artificial calibration method and demonstrated the feasibility of using a network approach. The multi-scale neural network [21] is composed of two main modules: a coarse-scale network and a fine-scale network. Note that both networks employed a single hazy image as input. Moreover, the output of the course-scale network can be considered as a useful supplement to the fine-scale network. The resulting predicted transmission map include more depth jumps. AOD-Net [22] proposed a new architecture under a rewritten physical model. For estimation, it employed a combined K-module combine that could combine the transmission map and atmospheric light through a linear transformation. This approached proved beneficial by transforming the problem of two variables into one variable through a single convolutional network. Furthermore, the image dehazing approach using a deep fully convolutional network in [33] proposed a deep lightweight residual model that employed residual learning to directly project the given hazy image onto both a hazy image and the corresponding haze-free image.

To sum up, the current CNN-based dehazing methods have informally shown great promise. Their primary goal is to learn diverse characteristics from a convolutional network, and the models include the following similar behaviours:

1) All the networks take a single image as input data to estimate the transmission map. Although multi-scale convolutional layers are used to extract more details, because the input feature dimension has not changed, the recovered depth information is not sufficiently obvious. In contrast, traditional multiple-input fusion strategies [24], [31] provide more information for restoring depth maps, which motivated us to address the input feature dimension problem by increasing the amount of input data.
2) In CNN based dehazing methods, the input image needs to pass through all the network layers to achieve the final output prediction. With deep network designs, a long-range memory is required to avoid the gradient explosion problem. To effectively train our network, we employ residual learning to produce skip connections. We found that our network architecture directly propagates the gradient information, which is conducive to constructing a better depth map.
3) At present, the use of Euclidean loss is the standard configuration in most CNN methods. However, this loss type does not apply to depth maps due to the relationship between depth and edge information. Although adversarial and Euclidean loss was employed in the jointly learned network in [22], all the attention was still concentrated on global differences. Based on this observation, we need a gradient regularization term to compensate for outliers or errors. Therefore, we employ a gradient constraint to reduce the error probability.

III. AGGREGATED RESOLUTION CONVOLUTION NETWORK

In this section, we elaborate and explain the proposed structure. As illustrated in Fig. 2, our architecture includes multiple inputs and a connected generator and discriminator. The task of the generator is to generate a transmission map through the aggregated networks using multiple input scales such that that the discriminator considers the generated t-map to be the indistinguishable from the distribution of the ground truth. This process makes it possible to generate visually complete and statistically consistent transmission maps from given hazy images. The task of the discriminator is to determine the authenticity of estimated t-maps and dehazed images.

Unlike most of the current CNNs dehazing approaches, in which the output depend on only a single input image that does not provide adequate feature dimensions, our framework is built on a strategy of fusing multi-resolution inputs. This approach enables us to identify abundant details that they can be mapped and used to reconstruct the depth map. Moreover, to extract more edge jumps and better reflect the transmission features, we adopt feature aggregation and an edge preserving

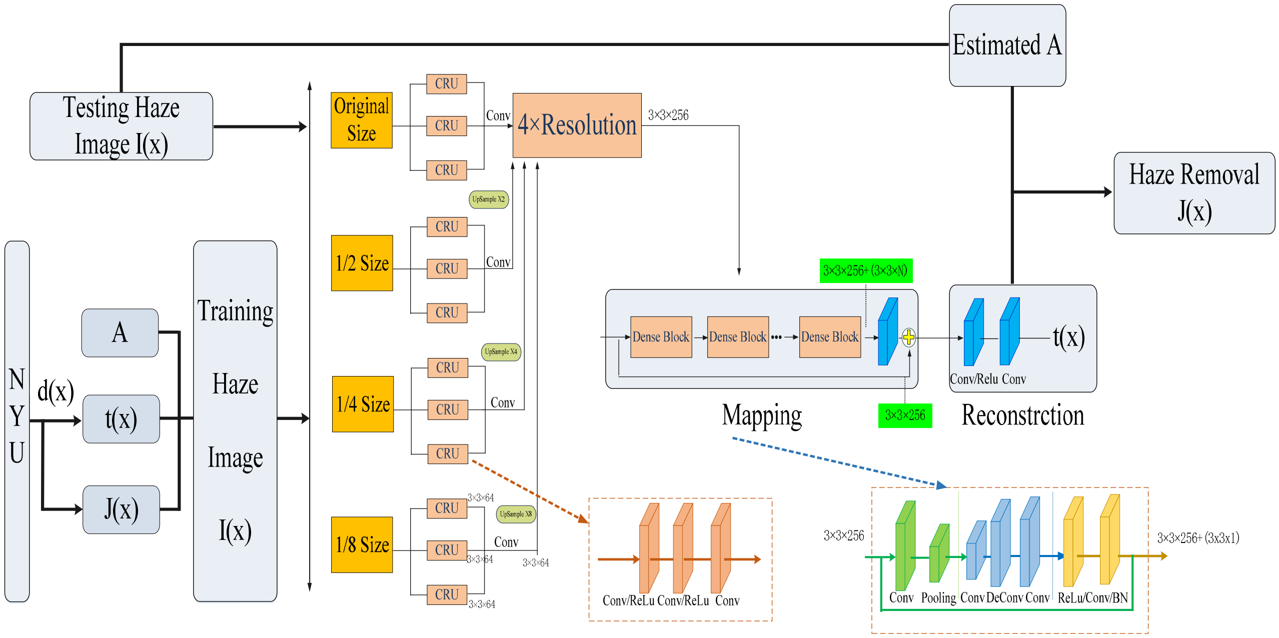

FIGURE 2. The individual components of our aggregated resolution convolution network.

loss function. These two features are described in detail in Sec. C.

A. THE GENERATOR FOR ESTIMATED TRANSMISSION MAP

Unlike the multiple-input approach used in the gated fusion network, we do not add additional input images. Instead, while guaranteeing the input of the original information, we increase the information dimensions by sampling the original image to form multiple inputs. Although this approach increases the model's memory consumption, it also enables the network to find the optimal feature set without requiring additional information to successfully aggregate the transmission map. Moreover, compared with other image dehazing methods based on convolutional networks, our generator achieves better performance and requires fewer layers by aggregating features at the level of each layer. This architecture both makes training more fluent and reduces the testing time. Our generator model consists of three parts: an input unit, a feature mapping unit and a reconstruction unit.

1) FUSION OF LAYERED PYRAMID INPUTS

As noted previously, when using a single input image, insufficient features may be extracted to generate the depth map. In contrast, multiple inputs can transmit more features simultaneously and merge them to produce the final prediction. However, the pooling operation requires output size reduction and is the biggest disadvantage of using multi-CNNs directly to solve the haze problem, because the reduction causes the loss of a great many details and produces inaccurate results. Considering the above aspects, our method employs a combination of multi-resolution and multi-stage accumulation strategies. First, the multi-resolution approach helps the model extract more distinct features. Second, multi-stage accumulation allows the model to focus on the most important details so that they are not discarded during the pooling operation. Finally, our model fuses multi-resolution features at different scales to generate an abundant feature map.

As illustrated in Fig. 2, we decompose a single image into multi-resolution images using a down-sampling strategy. The images we used for training come from the NYU data set, and they become 40 × 30 pixels in size after four down-samplings—too small to accurately reflect the characteristics of the entire image. Therefore, we employ three down-sampling results and fuse those with the original image. Moreover, to simplify the calculations, we use one collective residual unit (CRU) in each layer that extracts image features and fuses them internally to maximize the image feature aggregation, which is consistent with the conception of Ancuti [23]. Although all four components have the same internal structure, their parameters are not exactly the same due to the details of their respective resolutions.

Constrained by the need to guarantee an invariant size for the input images, we fuse input images of different resolutions with the original image through a relevant up-sampling strategy. As described in Fig. 2, we first extract the features of each resolution using the CRU units. In addition, similar to the last layer of SRCNN, we employ a convolution layer to aggregate the three channels of features. As a result, we obtain abundant feature sets after the up-sampling operation. Here, we set the channel numbers to d = 64 and adopt 3 × 3 filters to maintain performance in our approach, which convolves over the input (3 × 64 = 192) and outputs 1 × 64 = 64 on both sides of the convolution layer. Finally, there are 4 different resolutions ($4 \times 64 = 256$) in the feature activations. This provides sufficient elementary features to trigger the activation maps.

2) FEATURE MAPPING AND RECONSTRUCTION

According to the theory of CNN-based dehazing methods, algorithms are typically designed to find a mapping function $F(\tilde{I}) = t$ to predict the transmission map and the latent dehazed image. However, deep network architectures require a long-range memory to avoid the gradient explosion problem. Therefore, we adopt a dense block for mapping and reconstruction. Unlike the typical meaning, in this case, dense refers to the dense module, which greatly reduces the amount of computation, not to dense connections [23]. More importantly, because edge jumps are the main features that represent depth information, mapping high-frequency information is critical to successful haze removal. Therefore, the principle of residual learning has been globally accepted for training a residual mapping. In Fig. 2 we have: (a) *Conv+Pooling*: there is one module that consist of groups of convolution and pooling layers. This architecture captures the stable features used in depth mapping at different scales, which helps us in searching for edge jump information. (b) *Conv+Deconv+Conv*: In order to obtain consistent feature mapping with input image, we need to up-sample the pooled images to restore the sizes of feature maps. Following the suggestion of Wang *et al.* [34], we adopt a 16×16 de-convolution kernel to enhance the quality of the results according to the transmission map. Moreover, two convolutional layers are deployed, one on each side of the de-convolutional layer; these effectively reduce the computational complexity. c) *Relu+Conv+BN:* In this module, three layers are composed for mapping. The most different is that we have adjusted the order of Relu and Conv. First, a non-linear function is used to adapt to the edge jumps of depth transmission. In addition, a convolution is employed to acquire the relevant features. The BN makes the feature distribution consistent with the statistical distribution of the transmission map. Finally, we need to reduce the $(256 + \mathrm{N})$ feature dimensions to 256 dimensions through the convolutional layer to cooperate with the global residual for reconstruction.

The reconstruction unit contains two convolutional layers. One has a non-linear stretching function. The main purpose of this first convolution is to convert the 256 dimension parameters to a 1-dimensional vector matching that of the original image, and the last convolution is used to acquire a robust transmission map.

B. THE DISCRIMINATOR FOR HAZE-FREE IMAGE

Because there are two ground truths in the haze removal task, the discriminator consists of two steps. First, we need to distinguish the quality of the generated transmission map. If the quality is high, we can then judge whether the estimated image is a dehazed image. In this way, we can better estimate the two variables in the physical model. Because the discriminator is a classification network, we adopt the same discriminator network as was used in [37] to simplify the classification task. We deduce the second step as follows:

For the classical model in Eq. (1), we know that $\mathbf{A}$ is also an important variable; thus μ, we need to estimate its value to restore a haze-free image in the testing step. Typically, $\mathbf{A}$ is considered to be the pixel with the highest intensity in the haze image; therefore, many methods employ the brightest pixel in the corresponding degraded image as $\mathbf{A}$. We follow this strategy when searching for atmospheric light. However, unlike these methods, which directly determine the maximum value in all three channels, we first apply a white balance and then look only for the brightest pixel and adopt that as $\mathbf{A}$. This practice transforms atmospheric light into pure white light, which avoids the problem of colour perturbations.

After estimating $\mathbf{A}$ and $\mathbf{t}$, using the atmospheric scattering model, the dehazed image $\mathbf{J}$ can be restored. Consequently, the final result $J(x)$ is restored as follows:

$$\mathbf{J} = \frac{\mathbf{I\text{-}A}}{\mathbf{t}} + \mathbf{A} \tag{2}$$

C. COST FUNCTION

Currently, the goal of the output of the existing CNN based methods is to approach the real transmission map as closely as possible; therefore, they apply the Euclidean loss to deduce the final result. The inference is performed for macroscopic perspective, such as recognition and tracking tasks. However, it is not appropriate for a transmission map, in which local smoothness and piecewise discontinuity appear in most regions. In contrast, traditional variation methods have been demonstrated to recover these characteristics, which rely on an additional gradient regularization term. Therefore, we designed a restricted loss function for the dehazing task that considers the perceptual loss as weighted TV and style losses.

1) GRADIENT LOSS

Given one output transmission map $\mathbf{t}$ and a ground truth $\mathbf{t}$, we minimize a loss function defined with the TV constraint as follows:

$$L_t = \left(\frac{1}{2}\left\|\hat{\mathbf{t}} - \mathbf{t}\right\|_2^2 + \lambda\left\|\nabla\hat{\mathbf{t}}\right\|_1\right) \tag{3}$$

Nevertheless, our experiments clearly showed that the edge jumps cannot be preserved well under the TV regularization. Unlike the previous locally based regularization methods, TV regularization performs on a global level and is thus less sensitive to local textures. To solve this problem, we propose the following new regularization constraint for transmission map restoration:

$$\mathrm{E}(\hat{\mathbf{t}}) = \mu\left(\left\|\nabla\hat{\mathbf{t}} - \nabla\mathbf{t}\right\|_2^2\right) \tag{4}$$

where μ is an experimental regulation parameter. When $\hat{\mathbf{t}}$ approaches the ground truth, the estimated $\hat{\mathbf{t}}$ will include the

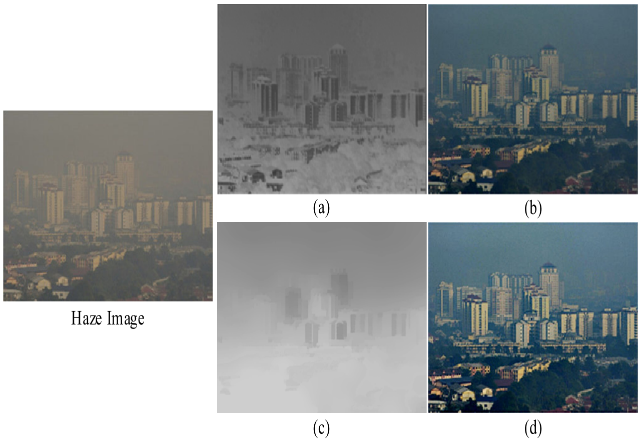

FIGURE 3. Estimated results of our method under different constraints. Figures (a) and (b) are the results by Formula 3, Figures (c) and (d) are the results by Formula 5.

correct and sharpened edges, making the value of Formula 4 is small. In contrast, a fuzzy or defective $\hat{\mathbf{t}}$ produces incorrect edges, which lead to larger values of Formula 5. Therefore, minimizing the above formula yields the correct sharp images. Based on the above considerations, we combined Formula 3 and 4 to construct the final loss function:

$$L_t = \left(\frac{1}{2}\left\|\hat{\mathbf{t}} - \mathbf{t}\right\|_2^2 + \lambda\left\|\nabla\hat{\mathbf{t}}\right\|_1 + \mu\left(\left\|\nabla\hat{\mathbf{t}} - \nabla\mathbf{t}\right\|_2^2\right)\right) \quad (5)$$

Compared with the conventional total variation, which encourages piecewise constant images and often suffers from undesirable artefacts, Eq. (5) prefers piecewise smooth images. This smoothness is a desirable property in depth estimation because an image may have a slanted plane whose transmission varies smoothly along with the change of depth.

2) STYLE LOSS

The task of haze removal consists of two important steps. After refining the transmission map, the next task is to recovery the scene radiance $\mathbf{J}$. One observed phenomenon is that artefacts appear in certain areas of the restored images. However, these visual artefacts are usually invisible in the input image. It is precisely the newly restored edges that cause the artefact effect. Based on this idea, we propose a novel loss function to constrain the edges in the dehazed image. We were motivated to minimize the residual of the gradients between the input and output images under the sparse-inducing norm. Thus, our optimization problem becomes

$$L_J = \left(\frac{1}{2}\left\|\hat{\mathbf{J}} - \mathbf{J}\right\|_2^2 + 0.1\left(\left\|\nabla\hat{\mathbf{J}} - \nabla\mathbf{I}\right\|_2^2\right)\right) \quad (6)$$

Fig. 3 shows the estimated results of our method under the constraints in Formulas 3 and 5, respectively. Under the Formula 3 constraint, we can clearly see that a more accurate transmission map is obtained within a certain gradient scope. However, when the same object has too many details, this piecewise smoothing strategy causes the transmission map to be too slight. It leads to dark colours in the restored image. In contrast, Formula 5 makes the transmission map smoother, and the estimation is more consistent with the image depth. Therefore, the restored image looks more natural than when using Formula 3. Fig. 4 shows the results of gradient loss achieved when using only Formula 5 and when integrating gradient loss and style loss with both Formulas 5 and 6. It can be seen that ensuring only the correctness of the transmission map, the restored image presents colour disturbances, block effects and some colour-darkened areas (e.g., the sky, the clock tower and others) in figures (a) and (c). In contrast, the dehazing results appear more natural when multiple loss functions are fused, such as figures (b) and (d).

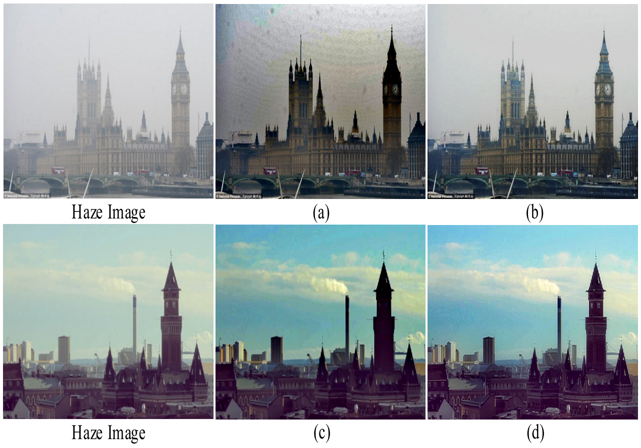

FIGURE 4. Estimated results of our method under different constraints. Figures (a) and (c) are the results by Formula 5, Figures (b) and (d) are the results by Formulas 5 and 6.

IV. EXPERIMENTS

To verify the rationality and effectiveness of our architecture, we performed comprehensive experiments on three synthetic datasets and a large number of natural hazy images. There are four main procedures. The first sub-section describes the datasets used for training and testing in our experiments. The second sub-section discusses experimental details and reports the parameter settings. The third sub-section discusses the a comparison results on some challenging hazy images—both natural and synthetic scenes. The last sub-section presents a quantitative evaluation of our results.

A. DATASETS

Because no public dataset exists that includes a full set of hazy and clear images along with their transmission maps, no ground truth is available to serve as a definitive reference value. Consequently, we synthesized training datasets by following the process reported by [20]–[23]. First, we randomly selected 1000 NYU images are as training samples $J^c(x)$ and $d(x)$. In addition, we adopted and applied random atmospheric light $A^c = [a, a, a] \in [0.5, 1.0]$ and scattering coefficients $\beta = [0.3, 1.8]$ to generate correlated haze images $I^c(x)$ and transmission maps $t(x)$. Moreover, in order to weaken the influence of rotation on feature extraction, we applied four different rotations and flipping techniques, which image employed from the dataset. As a result, we obtained 16,000 synthetic images to train the optimal model parameters. To verify the rationality and validity of our model, we also created a validation dataset which contains 50 real hazy images and 50 synthetic hazy images. The proposed model was trained using the MatConvNet package on a workstation equipped with an Intel i7 2.8 GHz CPU and two GTX1080Ti GPUs. Training requires approximately 100 epochs to converge.

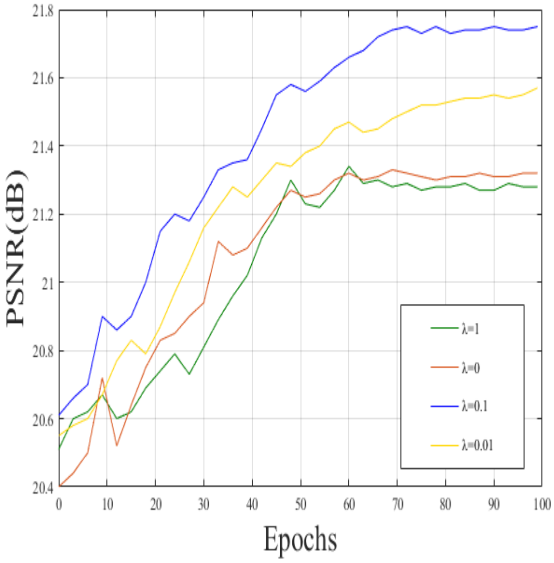

FIGURE 5. The PSNR values on the validation set during training with different value of λ.

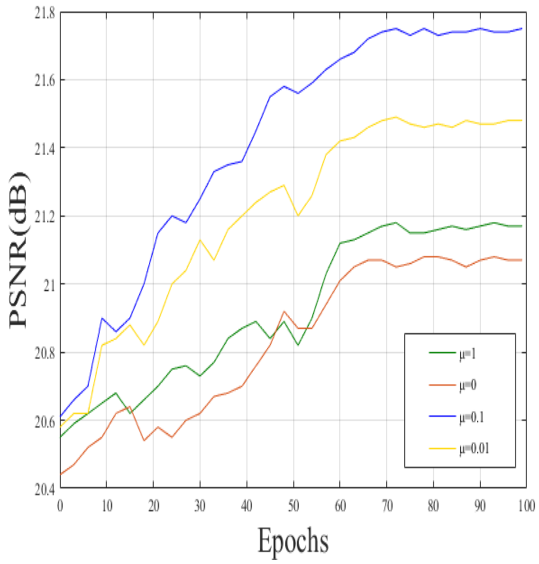

FIGURE 6. The PSNR values on the validation set during training with different value of λ.

TABLE 1. Component evaluation with different.

Method	Accuary		Error		
	δ<1.5	$\delta<1.5^2$	Rel	$\log_{10}$	Rms
No multi-inputs	82.6	92.6	0.124	0.051	0.537
No Feature Aggregation	84.4	94.2	0.112	0.046	0.505
No Gradient Loss	85.3	95.9	0.116	0.049	0.521
Our	86.4	97.2	0.107	0.043	0.497

B. PARAMETER SETTINGS

We performed a series of special experiments on our test dataset to investigate each parameter of the loss function. In addition, we set the initial learning rate to 0.1 and reduced it by a factor of 10 every 10 epochs. Updating stops after the learning rate reaches 0.1×10^{-10}. Moreover, we use batch processing during train to speed up the training operation. In these special experiments, we first fix $\mu = 0.1$ to verify the property of $\lambda = 1, 0, 0.1, 0.01$. The PSNR value of $\lambda = 0.1$ is better than others, as illustrated in Fig. 5, and the restored images also confirmed these results. When the parameter $\lambda = 1, 0$, the dehazed images may still have large or small amounts of remaining fog and haze. However, when $\lambda = 0.01$, the process removes most of the fog, but part of the texture also disappears. Compared with those two values, $\lambda = 0.1$ produces better results, although it introduces some noise in the restored image. Meanwhile, Fig. 6 also confirms our inference, because a higher PSNR value can be obtained by fixing $\lambda = 0.1$ and setting $\mu = 0.1$. Therefore, in the subsequent experiments, we set $\mu = 0.1$ and $\lambda_1 = 0.1$. This approach not only increases the dehazing level but also makes good use of the edge information. Moreover, as the most important three aspects of our network—multiple inputs, multi-scale residual units and regularization loss function—affect the final estimation, we adopted typical error metrics for the quantitative evaluation. Therefore, we also selected another 50 RGB-D image pairs and set $\beta = 0.8$ to establish some test data as internal sets for testing each step in our method. Examples of the results are shown in Fig. 7 and the averaged objective indexes are listed in Table 1. Here, we choose:

$$\begin{aligned}
\text{Threshold:}\quad & \max(\frac{\hat{t}}{t}, \frac{t}{\hat{t}}) = \delta < thr \\
\text{Mean relative error (Rel):}\quad & \frac{1}{|\Omega|}\sum\nolimits_{t\in\Omega} |\hat{t}-t|/t \\
\text{Mean log10 error (}\log_{10}\text{):}\quad & \frac{1}{|\Omega|}\sum\nolimits_{t\in\Omega} |\log_{10}\hat{t} \\
& -\log_{10} t| \\
\text{Root mean squared error (Rms):}\quad & \sqrt{\frac{1}{|\Omega|}\sum\nolimits_{t\in\Omega} |\hat{t}-t|^2}
\end{aligned}$$

where $\hat{t}$ and t denote the estimated transmission map and the ground truth, respectively, and Ω represents all the pixels in the images. As shown in Table 1, using every contribution provides better results than using none of them.

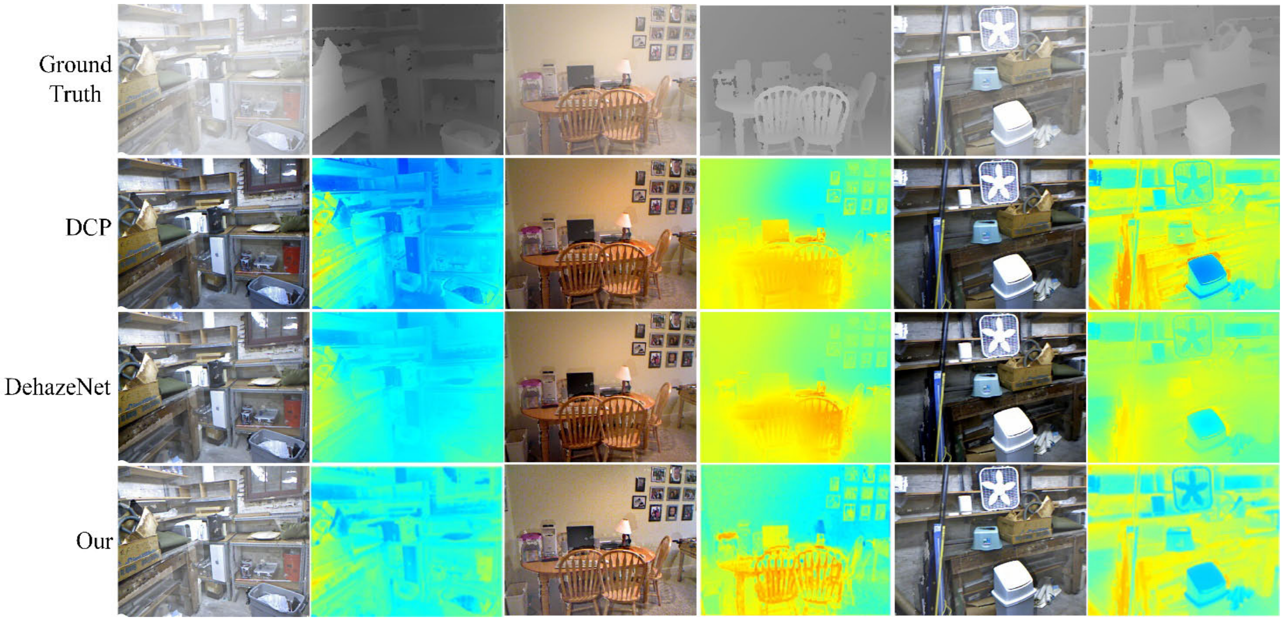

FIGURE 7. Comparison classical approaches for synthetic NYU datasets.

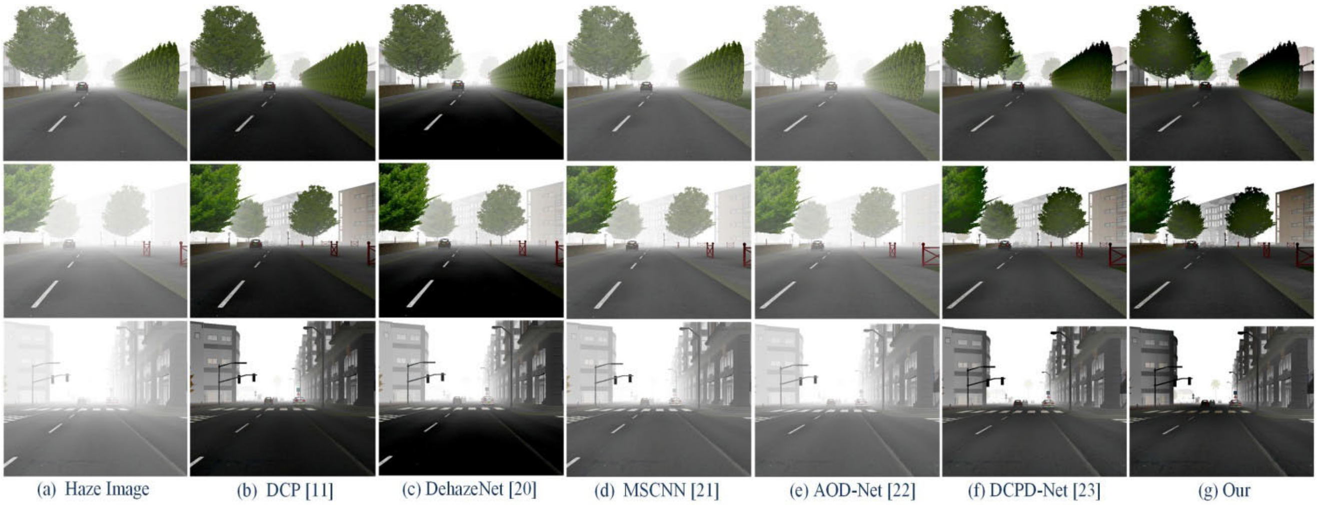

FIGURE 8. Visual qualitative comparisons for synthetic road images dehazing.

C. EVALUATION ON BENCHMARK DATA

Because dehazing is very important for automated driving in bad weather, we first present a group of synthetic dehazed road images for illustrative purposes. These artificial images are synthetically transformed into hazy images as shown in Fig. 8(a). As one of the most popular haze removal methods, He *et al.* [11] utilized the DCP to assume that the minimum value of each patch approximates zero; thus, the transmission map can easily be calculated. However, this strategy leads to an over-estimated transmission map, resulting in a darker restored image as shown in Fig. 8(b). A convolutional network was used in DehazeNet [20], but most of the layers are specially organized to accommodate prior constraints, such as DCP and colour attenuation. In contrast, due to its end-to-end architecture, MSCNN [21] achieves more natural results, but residual haze still exists in the dehazed image. A similar situation prevails in the results of AOD-Net [22]. The DCPD-Net model is superior to the previous models due to its dense connections, which result in more features for estimating the transmission map. However, this connection is not guided by edge information; consequently, distant haze is still not completely removed. Compared with DCPD-Net, our results (Fig. 8(g)) not only look more natural but also show richer edge features. This result occurs because our ARCN architecture uses multiple inputs and feature aggregation, allowing it to obtain more robust features for estimating the optimal transmission map.

Furthermore, we synthetically transformed some stereo images into hazy images to test the accuracy of colour reduction. Fig. 9 shows the results of various conventional

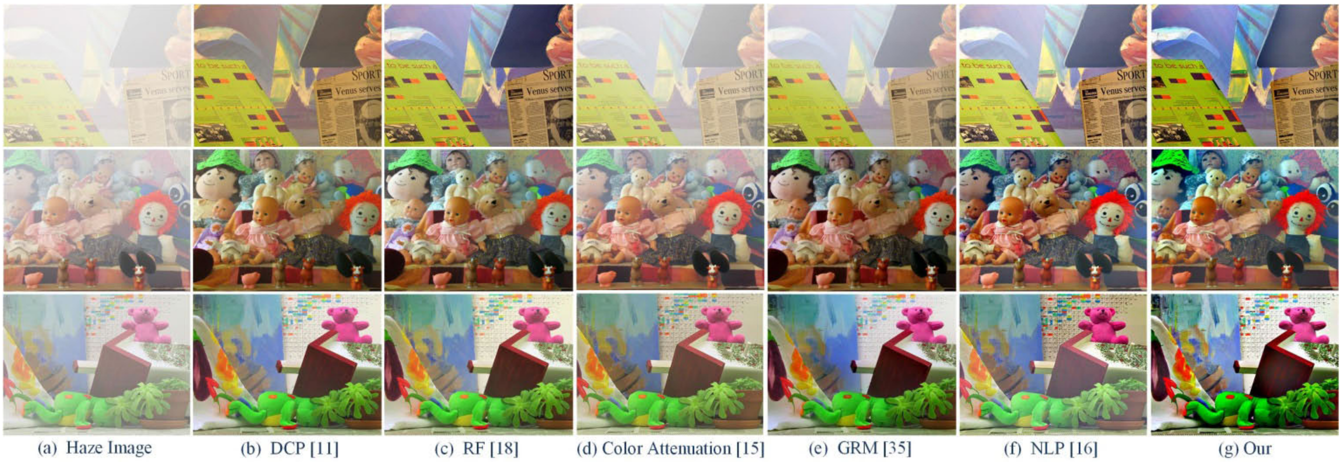
(a) Haze Image (b) DCP [11] (c) RF [18] (d) Color Attenuation [15] (e) GRM [35] (f) NLP [16] (g) Our

FIGURE 9. Visual qualitative comparisons for synthetic stereo images dehazing.

TABLE 2. Quantitative comparison shown in Fig. 8.

Approach	*e*/*ϑ*/*r*		
	Fig 8(1)	Fig 8(2)	Fig 8(3)
He [11]	1.10/0.01/1.19	4.99/0.00/1.89	4.07/0.01/2.10
Cai [20]	0.94/20.21/1.28	4.04/19.55/1.92	3.73/18.58/1.91
Ren [21]	0.01/0.00/1.04	1.35/0.00/1.70	1.09/0.00/2.01
Li [22]	0.25/0.00/2.21	0.85/0.00/4.10	0.51/0.00/4.69
Zhang [23]	0.37/0.33/1.75	2.91/0.00/4.00	2.36/0.13/4.59
Our [11]	1.03/0.02/2.32	4.72/0.02/4.14	3.94/0.04/4.82

approaches for restoring the hazy images. Typically, these methods rely on a certain type of scene or on statistical features to determine the constraint limits. However, no prior applies equally to all hazy images. This is the main reason why using a single prior produces colour shifts or ambiguous edges on different test images. Compared with prior results in Fig. 9, our ARCN approach, which uses a GAN strategy, is better than the traditional approaches at both colour restoration and at maintaining authenticity. Most of the details and edges are precisely captured and restored. The principal reasons for these results are the style loss and edge loss functions used in the training process. For example, the characters in the newspaper and the edges of each doll are recovered more accurately. Moreover, many of the other methods contrast do not fully remove the haze.

Like most successful dehazing methods, we adopt mean square error and structural similarity to evaluate the results of our approach and those of some typical haze removal approaches, especially CNN based approaches. To ensure a more objective and impartial evaluation, we employed two different objective evaluation methods. Because the purpose of dehazing is to restore the original appearance of the test image, following Hautière *et al.* [38], we use three metrics to measure the changes in edges and contrast before and after dehazing. In Table 2, the metric *e* represents the rate of new visible edges in the dehazed image compared to the hazy image, while the metric *ϑ* denotes the percentage of pixels that become black or white following the dehazing operation. Higher positive *e* values and values of *ϑ* closer to zero imply better performance. In addition, the metric *r* denotes the mean ratio of the gradient norms before and after dehazing. High *r* values represent better restoration of the local contrast, whereas low *r* values suggest fewer spurious edges and artefacts.

TABLE 3. The average results of four evaluation index methods SOTS HSTS.

Methods	SOTS	HSTS
DCP [11]	19.79/0.848	15.96/0.877
CAP [15]	19.05/0.836	21.54/0.867
NLP [16]	17.29/0.749	17.62/0.792
DehazeNet [20]	21.14/0.847	24.49/0.915
MsCNN [21]	17.11/0.805	18.29/0.841
AOD-Net [22]	19.38/0.849	21.58/0.922
DCPD-Net [23]	22.41/0.892	23.13/0.887
Our	22.55/0.906	23.39/0.938

Next, we present a quantitative evaluation of the dehazed outputs in Fig. 8. One obvious conclusion is that the method in Cai *et al.* [20] produces more black after restoration; therefore is achieves the highest *ϑ* score. We sequenced the other five algorithms in decreasing order with respect to the increase in new visible edges; the resulting sequence was He *et al.* [11], Ours, Ren *et al.* [21], Li *et al.* [22], and Zhang and Patel [23]. This result demonstrates that our method can generate more edges. Due to the importance of contrast in dehazing task, the index *r* is utilized to evaluate whether the new edges improve visibility. Based on this metric, our approach is superior to all the other methods. This implies that the gradient produced by our method is more effective but that a further increase would probably be too strong.

We also employed SOTS and HSTS test sets for assessment. In Table 3, the comparative results confirm that in most cases, our approach restores more accurate transmission maps and vivid haze-free images than do the other algorithms. Depending on the neural network and the availability

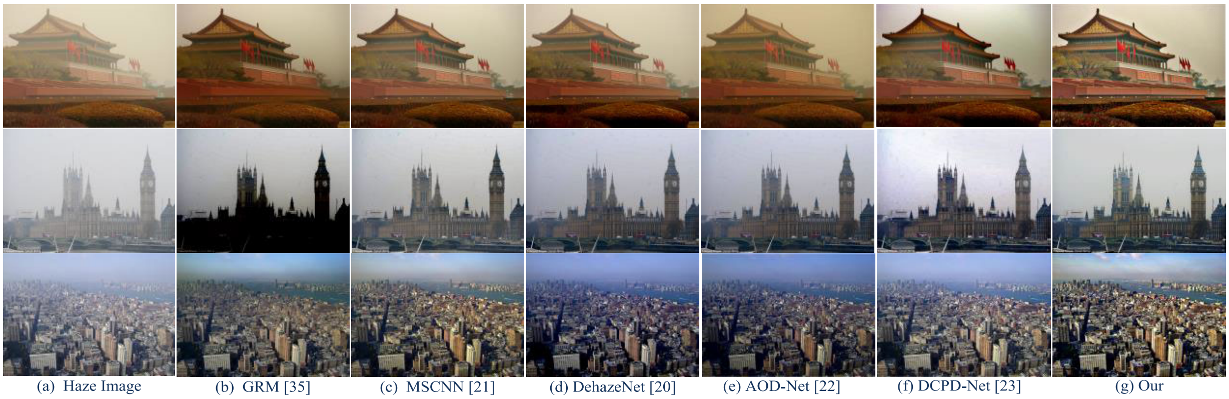

FIGURE 10. Visual qualitative comparisons for real prospect images dehazing.

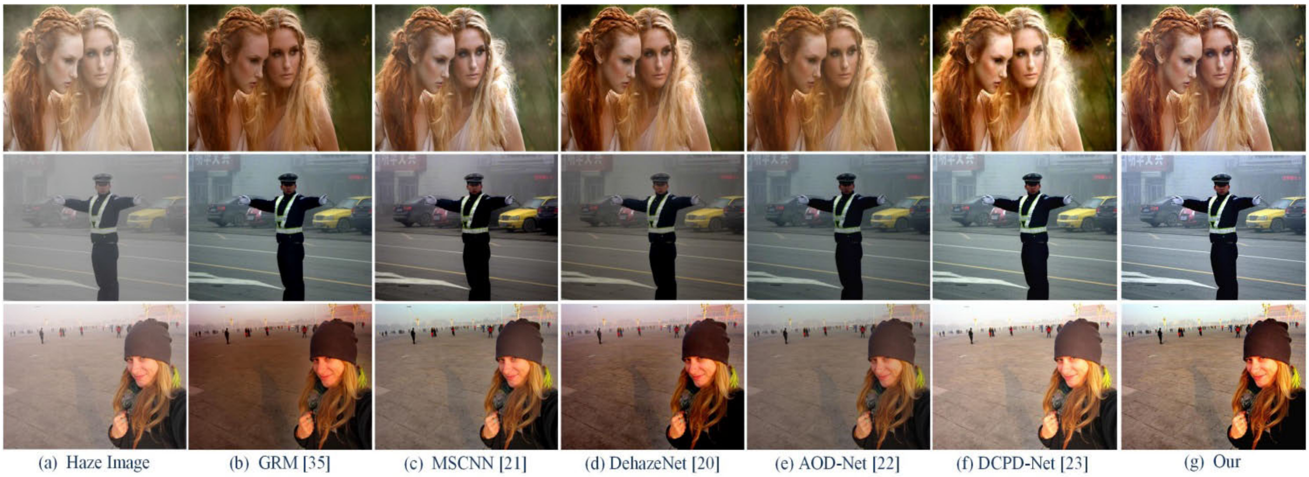

FIGURE 11. Visual qualitative comparisons for real close-range images dehazing.

of sufficient data, more details can usually be obtained compared with the traditional methods, such as DCP, CAP and NLP. For instance, prior-based methods always utilize statistical features to directly minimize the energy. However, in some areas of the dehazed images, this approach leads to over-saturation. In contrast, we can attribute our high scores to layer-by-layer feature extraction, which effectively prevents blackening of the restored image.

The next comparison concentrates on data-driven approaches. Edges are a key variable in haze removal; consequently, many other CNN-based methods pay attention to learning the edges. Nevertheless, they obtain rough misplaced edges in the transmission map, which subsequently results in inaccurate dehazed images. In contrast, our method achieved the highest performance through feature aggregation and the specific edge loss function.

D. EVALUATION ON REAL-WORLD HAZE IMAGES

In the comparison and analysis of the synthetic image results, our ARCN architecture resulted in better representations than those of previous approaches. To further illustrate the expressiveness of our model on real-world images, we conducted a further comparison and evaluation with the typical methods. Because of the diversity of real scenes, we divide the dataset scenes into three types: prospect images, close-up images and challenging images. Notably, regardless of the natural scene type, our method results in clear and natural restored images, which benefits from its ability to combine different features as well as the ability to highlight the edge information. In contrast, the results of the other methods are not as good as those of our method. For instance, GRM [18] causes artefacts in the sky region and black areas appear in the building. In addition, colour distortions and shadows are present in the dehazed images. Similar phenomena also exist in DehazeNet [20], which also adopts multiple prior conditions. Although MSCNN [21] employed multi-scale information to correct the above mistakes, some areas still include residual haze and fog, such as the restored image in Fig. 10(c). By adopting the K modules of transmission for hazy scenes, AOD-Net [22] effectively improved the dehazing results. Nevertheless, the colours of the restored image appear gloomy due to several pattern modules. Fig. 10(e) and Fig. 11(e) show that the restored images appear hazy in distant areas. DCPD-Net [22] performed better than did

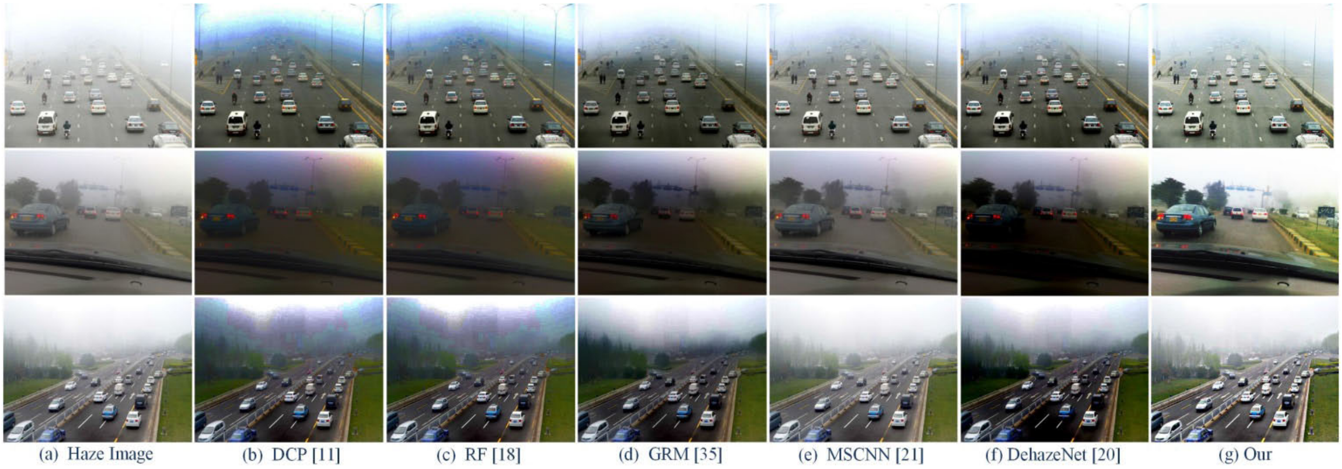

FIGURE 12. Visual qualitative comparison for real challenged images dehazing.

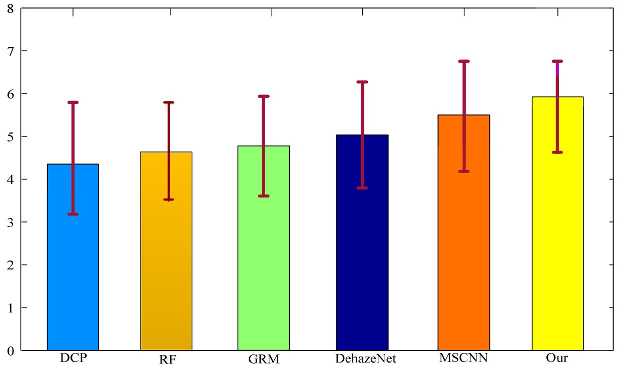

FIGURE 13. MOS for different approaches.

the other models, but compared to our model, a residual haze always exists in the restored images, as illustrated in Fig. 10(f) and Fig. 11(f).

Finally, we also compared our method with some methods commonly used in practice to verify our results, as shown in Fig. 12. Obviously, the traditional methods tend to produce artefacts in sky areas, but our method seems to recover the natural images. These results also demonstrate the importance of reconstructing edges in line with the input edges.

Because challenging hazy images do not have a relevant ground truth, we cannot use SSIM or MSE as indicators to evaluate the dehazing quality. Therefore, we employ the mean opinion score (MOS) [36] to compare the mean and standard deviation of each approach, as illustrated in Fig. 13. Although MOS is only a rough subjective criterion, it still reveals some clues and observations. First, we note that each algorithm has a certain error bar, which implies that no approach is competent at all dehazing tasks. Compared with other methods, the error bar of the proposed method is smaller, as shown in Fig. 13. Moreover, our average score is the highest, which indicates that a larger percentage of our method's dehazed images performed well than did those of the other methods. More importantly, the MOS indicators are largely consistent with the objective comparison in Table 2, which further confirms the correctness of the proposed method when training the transmission map reconstruction using an indoor image library (NYU).

E. FAILURE CASES

As described above, we rely on training synthetic indoor dataset to obtain a transmission map and our method achieves better results than those of other methods. However, indoor depth is not truly equivalent to the depth of outdoor scenes. Consequently, our method does not restore some of the foreground images very well, as shown in Fig. 8. In addition, because we considered only the depth information factors and ignored the characteristics of real scenes, the restored images include some artefacts in the sky region, as illustrated in Fig. 12. In future work, we will continue to strive to improve and perfect our network model to address these problems.

V. CONCLUSION

In this paper, we provide an Aggregated Resolution Convolution Network to address the image dehazing problem. In order to achieve enough features from one single haze image, we design a novel multi inputs frame to extract different level features according to the previous successful method. Moreover, for improving accuracy and reducing disturbance, we propose to train and verify our ARCN network with a relative edge and style Loss function that its main reasonability is to ensure the accuracy of the gradient while our network can better achieve the feature extraction. Extensive experiments have been demonstrated that our algorithm restores better than many other classical methods on a massive number of synthetic and real-world scenes.

ACKNOWLEDGMENT

The authors would like to thanks the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] M. Negru, S. Nedevschi, and R. I. Peter, ‘‘Exponential contrast restoration in fog conditions for driving assistance,’’ *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2257–2268, Aug. 2015.

[2] S.-C. Huang, B.-H. Chen, and W.-J. Wang, ‘‘Visibility restoration of single hazy images captured in real-world weather conditions,’’ *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1814–1824, Oct. 2014.

[3] P. Carr and R. Hartley, ‘‘Improved single image dehazing using geometry,’’ in *Proc. Digit. Image Comput. Techn. Appl.*, Dec. 2009, pp. 103–110.

[4] B. Xie, F. Guo, and Z. Cai, ‘‘Universal strategy for surveillance video defogging,’’ *Proc. SPIE*, vol. 51, no. 10, May 2012, Art. no. 101703.

[5] 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)*, Dec. 2013, pp. 617–624.

[6] J. Park, K. Kim, S. Lee, C. S. Won, and S.-W. Jung, ‘‘Text-aware image dehazing using stroke width transform,’’ in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2231–2235.

[7] R. Tan, ‘‘Visibility in bad weather from a single image,’’ in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

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

[9] R. Fattal, ‘‘Single image dehazing,’’ *ACM Trans. Graph*, vol. 27, no. 3, pp. 1–9, Aug. 2008.

[10] R. Fattal, ‘‘Dehazing using Color line,’’ *ACM Trans. Graph*, vol. 34, no. 1, pp. 256–269, Dec. 2014.

[11] 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.

[12] J. P. Tarel and N. Hautière, ‘‘Fast visibility restoration from a single color or gray level image,’’ in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Sep. 2009, vol. 30, no. 2, pp. 2201–2208.

[13] J. Yu, C. Xiao, and D. Li, ‘‘Physics-based fast single image fog removal,’’ in *Proc. IEEE Int. Conf. Signal Process. (ICSP)*, Oct. 2010, pp. 1048–1052.

[14] K. He, J. Sun, and X. Tang, ‘‘Guided image filtering,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Oct. 2013.

[15] 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.

[16] D. Berman, T. Treibitz, and S. Avidan, ‘‘Non-local image dehazing,’’ in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1674–1682.

[17] L. Caraffa and J.-P. Tarel, ‘‘Combining stereo and atmospheric veil depth cues for 3d reconstruction,’’ *IPSJ Trans. Comput. Vis. Appl.*, pp. 1–11, 2014.

[18] K. Tang, J. Yang, and J. Wang, ‘‘Investigating haze-relevant features in a learning framework for image dehazing,’’ in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2995–3002.

[19] S. Zhang, F. He, W. Ren, and J. Yao, ‘‘Joint learning of image detail and transmission map for single image dehazing,’’ *Vis. Comput.*, Nov. 2018, pp. 1–12.

[20] 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.

[21] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, ‘‘Single image dehazing via multi-scale convolutional neural networks,’’ in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2016, pp. 154–169.

[22] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, ‘‘An all-in-one network for dehazing and beyond,’’ in *Proc. ICCV*, 2017, pp. 4770–4778.

[23] H. Zhang and M. Vishal Patel, ‘‘Densely connected pyramid dehazing network,’’ in *Proc. CVPR* Jun. 2018, pp. 3194–3203.

[24] Y. Wang and C. Fan, ‘‘Single image defogging by multiscale depth fusion,’’ *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4826–4837, Nov. 2014.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘‘Imagenet classification with deep convolutional neural networks,’’ in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[26] C. Ding and D. Tao, ‘‘Robust face recognition via multimodal deep face representation,’’ *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2049–2058, Nov. 2015.

[27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ‘‘Rich feature hierarchies for accurate object detection and semantic segmentation,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[28] C. Dong, C. C. Loy, K. He, and X. Tang, ‘‘Image super-resolution using deep convolutional networks,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[29] J. Kim, J. K. Lee, and K. M. Lee, ‘‘Accurate image super-resolution using very deep convolutional networks,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[30] K. Kim, J. K. Lee, and K. M. Lee, ‘‘Deeply-recursive convolutional network for image super-resolution,’’ in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.

[31] D. Zhao, L. Xu, Y. Yan, J. Chen, and L.-Y. Duan, ‘‘Multi-scale Optimal Fusion model for single image dehazing,’’ *Signal Process., Image Commun.*, vol. 74, pp. 253–265, May 2019.

[32] D. Liu, Z. Wang, N. Nasrabadi, and T. Huang, ‘‘Learning a mixture of deep networks for single image super-resolution,’’ Jan. 2017, *arXiv:1701.00823*. [Online]. Available: https://arxiv.org/abs/1701.00823

[33] J. Ding, Z. Yan, X. Wei, and X. Li, ‘‘Light-weight residual learning for single image dehazing,’’ *Proc. SPIE*, vol. 28, no. 3, May 2019, Art. no. 033013.

[34] Y. Wang, L. Wang, H. Wang, and P. Li, ‘‘End-to-end image super-resolution via deep and shallow convolutional networks,’’ Jul. 2016, *arXiv:1607.07680*. [Online]. Available: https://arxiv.org/abs/1607.07680

[35] C. Chen, M. N. Do, and J. Wang, ‘‘Robust image and video dehazing with visual artifact suppression via gradient residual minimization,’’ in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2016, pp. 579–591.

[36] K. Ma, W. Liu, and Z. Wang, ‘‘Perceptual Evaluation of single image dehazing algorithms,’’ in *Proc. Int. Conf. Image Anal. Process.*, Sep. 2015, pp. 3600–3604.

[37] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, ‘‘Photo-realistic single image super-resolution using a generative adversarial network,’’ in *Proc. CVPR*, Jul. 2017, pp. 4681–4690.

[38] N. Hautière, J.-P. Tarel, D. Aubert, and É. Dumont, ‘‘Blind contrast enhancement assessment by gradient ratioing at visible edges,’’ *Image Anal. Stereol.*, vol. 27, no. 2, pp. 87–95, 2008.