Single image dehazing algorithm based on pyramid multi-scale transposed convolutional network

Keping Wang, Yumeng Duan and Yi Yang

College of Electrical Engineering and Automation, Henan Polytechnic University, Henan Jiaozuo, People’s Republic of China

ABSTRACT
In this paper, a novel single image dehazing method based on pyramid multi-scale transposed convolutional network (MST-Net) is proposed. Conventional haze removal algorithms based on the atmospheric scattering model may lead to some problems such as incomplete dehazing and colour distortion due to the inaccurate parameter approximations or indirect image optimization and reconstruction. Therefore, we design a real end-to-end image dehazing network to directly learn the mapping relationship between hazy images and the corresponding clear images. In this network, the cascaded feature extraction blocks extract the diversified feature information of the input images by multi-channel concatenation structure, which can effectively fuse the local features of the first convolution layer into the semantic features of subsequent layers in the block. In order to reconstruct high-quality dehazed images relieving the colour distortion, we design a multi-scale transposed convolution block to gradually expand the resolution of the obtained feature maps, and introduce skip connections from the feature extraction module to supplement the detailed information of the feature map pyramid. Extensive experimental results demonstrate that the proposed method in this paper can remove the haze completely and achieve superior performance in subjective and objective evaluation over the other state-of-the-art methods.

ARTICLE HISTORY
Received 11 June 2020
Accepted 2 October 2020

KEYWORDS
Image dehazing; convolutional neural network; multi-scale transposed convolution; feature extraction; image reconstruction

1. Introduction
Haze is the most common atmospheric phenomenon which is influenced by the smoke, mist, dust and various suspended particles. Images captured by the devices often have been degraded by haze weather conditions. The degraded images are often subject to colour distortion, blurring and have poor contrast for visual performance, which may in turn seriously affect the performance of many computer vision tasks such as classification and detection (D, 2015). Thus, it is of great significance to overcome the destruction caused by haze and restore clear scenes.

In order to overcome the image degradation caused by the haze, a large number of image dehazing methods have been proposed. Since the atmospheric scattering model (ASM) (Narasimhan & Nayar, 2002) is introduced into the imaging mechanism to depict the physical degradation process of the hazy image explicitly, lots of works have been proposed based on this model.

In the ASM, since the two very important parameters \( t(x) \) and \( A \) are unknown, restoring the haze-free image from a hazy one is a severely ill-posed problem. Some approaches try to estimate the parameters and reconstruct the clear image by leveraging different priors of the hazy image such as dark channel prior (He et al., 2009), contrast colour-lines (Fattal, 2014) and haze-line prior (Berman et al., 2017). Recently, Convolutional Neural Networks (CNNs) highlight their performance in computer vision tasks and have been introduced into image dehazing as well. Instead of using handcrafted priors to estimate parameters of the ASM, CNNs are employed to calculate them automatically driven by a mass of haze/haze-free image pair data. However, the dehazing effects of these methods heavily rely on the suitable extraction of prior information and the accurate estimation of the ASM’s parameters. The inaccurate parameter estimation may result in haze remaining and colour distortion, especially in the cases of the scenes where the colours of objects are inherently similar to those of atmospheric lights.

Moreover, to simplify the mathematic model of the image degradation process, there exists a strong assumption in the ASM that the atmosphere is homogenous, which means particles in the atmospheric medium are the same and their distribution is uniform anywhere. Haze in the atmosphere influenced by the smoke, mist, dust and various suspended particles, essentially brings random, non-uniform noise in the images. It is difficult to
ensure its concentration and particle size are all the same in a large visual range. Furthermore, a large number of inhomogenous haze images are produced constantly. So image dehazing based on the ASM may fail in cases where the assumption of this model is broken.

Based on the above observations, we propose a novel image-to-image dehazing method with pyramid multi-scale transposed convolutional network (MST-Net), which directly learns the mapping between the input hazy images and the corresponding clear images. The feature extraction module of the MST-Net extracts and fuses the feature information of the input images through several feature extraction blocks composed of multi-channel concatenation structure. This structure transmits detailed features of the input layer to all subsequent layers by concatenation operations, which can supplement the detailed features lost in the process of convolution layers. In the image reconstruction module, we expand the resolution of the feature maps gradually to restore the high-pixel images by the multi-scale transposed convolution blocks composed of parallel transposed convolution layers. The skip connection is also introduced from the feature extraction blocks to increase the detailed information of the feature map pyramid. Finally, we use a transposed convolution layer to restore the dehazed images from the output feature maps of the multi-scale transposed convolution blocks.

This paper makes the following contributions:

- We propose a novel end-to-end pyramid multi-scale transposed convolutional network MST-Net for single image dehazing. MST-Net is composed of a feature extraction module and an image reconstruction module, which directly learns the mapping relationship between the hazy images and the corresponding clear images. It can remove haze more thoroughly and also has a strong advantage in terms of colour fidelity.
- A multi-channel concatenation structure is proposed, which transmits the input feature information to the subsequent layers through the concatenation operations in the feature extraction blocks. It effectively retains detailed information of various levels in the feature extraction process.
- We propose a multi-scale transposed convolution block, which consists of three parallel transposed convolutions with different filters and can more comprehensively restore the original features of the hazy images. In addition, we obtained feature map pyramids of different resolutions in the form of hierarchical expansion, which is conducive to the reconstruction of haze-free images.

2. Related works

Single image dehazing has drawn significant attention and a large number of dehazing methods has been proposed to handle this challenging ill-posed problem. Early researchers focus on image enhancement to improve the visual effects, mainly using traditional enhancement algorithms to increase the contrast and sharpness of hazy images. Classic algorithms include histogram-based (Zhu et al., 1999), Retinex (Ma et al., 2017) and contrast-based (Stark, 2000), etc. The image degradation mechanism has not been considered in this type of algorithms, hence the dehazing effect of these algorithms is limited. Later, many image dehazing algorithms were proposed with significant advancements due to the birth of the atmospheric scattering model (ASM). This model describes the imaging process of haze images in a mathematical model. The atmospheric scattering model is formulated as below:

\[ l(x) = J(x)t(x) + A(1 - t(x)) \]  \hspace{1cm} (1)

Where \( x \) is the location of each pixel within the image, \( l(x) \) is the observed hazy image, \( J(x) \) is the corresponding scene radiance, i.e. the ideal clean image to be recovered. \( A \) denotes the global atmospheric light, and \( t(x) \) is the medium transmission that can describe the portion of the light that is not scattered and reaches the imaging device. Among them, \( t(x) \) is related to the distance from the object to the imaging device and can be formulated as:

\[ t(x) = e^{-\beta d(x)}, 0 \leq t(x) \leq 1 \]  \hspace{1cm} (2)

Where \( \beta \) is the scattering coefficient of the atmosphere, which represents the ability of atmospheric scattering media to scatter visible light in different directions. When the atmosphere is evenly distributed, the atmospheric scattering coefficient \( \beta \) is a constant for the entire image at a certain moment. \( d(x) \) is the distance from the scene object to the image sensor.

Most of the algorithms based on ASM achieve the purpose of dehazing by estimating unknown parameters in the model. Fattal et al. (Fattal, 2008) use the prior information that the surface colour of the object is not related to the transmittance to estimate the transmission map of ASM, and adopt the Markov model to calculate the colour information. This method achieves the purpose of dehazing while loses the dehazing quality in areas where the haze is dense. He et al. use Dark Channel Prior (DCP) (He et al., 2009) to roughly estimate the transmission map of hazy images, and refine the rough transmission map through the soft matting algorithm. Then, the guided image filtering (He et al., 2010) is proposed to replace the soft matting due to its heavy computing. DCP is effective
for the most outdoor images, but usually fails to deal with the large white areas or sky areas in the image. In (Zhu et al., 2015), the Colour Attenuation Prior (CAP) is proposed, in which the parameters of the linear model are trained by the supervised learning, and then the scene depth and transmission are estimated in turn. Finally, the prior information and parameters are combined with the atmospheric scattering model to restore the haze-free images.

In recent years, deep learning has attracted much attention in computer vision due to its powerful nonlinear mapping and learning ability. It has been introduced into image dehazing as well. Convolutional Neural Networks (CNNs) is one of the typical algorithms of deep learning. Many methods (Cai et al., 2016; Deng et al., 2019; Dudhane & Murala, 2019; Liu et al., 2019; Ma et al., 2019; Park et al., 2020; Ren et al., 2016; Li, Miao, et al., 2019; Zhang & Patel, 2018) based on CNNs are proposed by people to estimate the parameters of the ASM. For example, DehazeNet proposed by Cai (Cai et al., 2016) et al. is the first time using CNN to directly estimate the medium transmission map from the haze image to achieve the purpose of image dehazing. Subsequently, a multi-scale convolutional neural network (MSCNN) (Ren et al., 2016) is proposed, which extracts the feature information in the image through a cascade of coarse-scale and fine-scale neural network. This can learn the mapping relation between the hazy image and its corresponding medium transmission map more effectively. Zhang (Zhang & Patel, 2018) et al. propose a densely connected pyramid dehazing network (DCPDN), in which the transmission map, atmospheric light, and dehazed image are estimated simultaneously. Finally, they use a joint discriminator based on generative adversarial network (GAN) (Goodfellow et al., 2014) to determine whether the obtained dehazed images and transmission images are real or not.

The deep learning-based methods may have some errors in estimating the transmission map and atmospheric light, and these errors would be superimposed and enlarged in subsequent calculations. Aiming at this problem, Li (Li et al., 2017) et al. proposes an AOD-Net to directly estimate the mapping relationship between hazy images and haze-free images. They transform the atmospheric scattering model and embed it into the dehazing network, and finally, combine Faster R-CNN (Ren et al., 2015) to quantitatively evaluate the effect of dehazing on the high-level visual task. In (Qin et al., 2019), an image-to-image network named FFA-Net is proposed to restore the haze-free images combined with the attention mechanism. This network uses the Feature Attention (FA) module combined channel attention and pixel attention to flexibly deal with the different features and pixels.

The aforementioned references based on the ASM have achieved significant progress in image dehazing. However, the dehazing effect of these methods would be affected in cases where the assumptions that the atmosphere is homogenous in ASM are broken, and their two-step optimizations instead of end-to-end methods to generate clear images may further cause performance degradation. Different from the above methods, our method directly constructs a novel network to capture the inherent relation between the hazy and haze-free images without estimating the intermediate parameters.

The rest of this paper is organized as follows. Section 3 presents the architecture of the proposed network, including the module of feature extraction and image reconstruction based on the multi-scale transposed convolution. Experimental results and analysis are provided in Section 4, and finally, we conclude our work in Section 5.

3. Proposed method

In this section, the proposed MST-Net is explained. As shown in Figure 1, the entire network is an image-to-image system, which can directly learn the mapping between the hazy images and the corresponding clear images through the cascaded feature extraction blocks and the multi-scale transposed convolution blocks. In the feature extraction, we propose a multi-channel concatenation structure to capture features at different scales, which fuses the information of the input feature maps and all subsequent layers through channels skip connections. Three cascaded multi-scale transposed convolution blocks are used in the image reconstruction module to expand the resolution of the feature maps. We fuse multiple feature maps from feature extraction blocks to make up for the lost information and recover the final clear images from the feature maps by a transposed convolution.

3.1. Feature extraction

Generally, the shallow layers of CNNs have strong ability to get geometric details, and they are suitable for extracting relatively simple features in local areas, such as edge features, texture features, etc. The feature maps obtained by the deep layers usually have a strong representation of the semantic information, which own high discrimination ability for different objects. Meanwhile, with the layers increasing, the generated feature maps tend to have low resolution and may lose some detailed geometric features gradually.

In order to retain more detailed information lost in feature extraction process, this paper proposes a
multi-channel concatenation structure which transmits
the input feature information to the following subse-
quently layers through multi-channel concatenation oper-
ations. Generally, the input feature maps in the network
tend to contain more local detail information, with the
layers increasing, the perception field of convolution layer
is expanded, and the more complex semantic information
of the image would be captured. The disadvantage is that
the above process may result in the loss of some detailed
information, such as edges, corners and straight lines, so
we introduce skip connections from the input layer to all
subsequent layers to supplement detailed information.

As shown in Figure 1, four cascaded feature extrac-
tion blocks are designed to extract the useful features for
image recovery. Each block is composed of multi-channel
concatenation structure, as shown in Figure 2. The conv1
is used as a bottleneck layer to reduce the number of
channels of the input feature maps, and the conv2 can
extract feature information and act as a down-sampling
layer to change the size of the feature maps, its convolu-
tion kernel size is $3 \times 3$. The conv3 is added to reduce the
number of channels to a smaller preset value to limit the
width of the network. With the deepening of the network,
more feature maps are needed to express the information
of the input images, so we add a branch from the previ-
ous layer and use conv4 to integrate the number of fea-
ture maps to an appropriate value. Besides, the network
performs dimension matching between different feature
maps by convolution layers. Compared with the densely
connected convolutional networks (Huang et al., 2017)
that connect all the different layers, the proposed net-
work is relatively lightweight and does not require addi-
tional dropout operations to reduce branches to avoid
overfitting.

Through experiments, it is found that the semantic
information of the feature maps obtained by the feature
extraction blocks composed of only four convolution lay-
ers is insufficient, and the dehazing effect of the restored
images is not complete. Therefore, the convolution lay-
ers conv5-conv8 are added to extend the network depth,
and its structure is the same as the previous network. The
deepened feature extraction block has higher nonlinear
complexity, and it can extract more complex and accurate
feature information.
Since high-pixel image restoration requires that the detailed information of the image is retained in the feature maps as much as possible, the excessive downsampling process may result in the loss of image details. Therefore, pooling layers are removed from the network. Each convolution layer in the feature extraction block is followed by Batch Normalization (BN), and we use the Rectified Linear Unit (ReLU) (Nair & Hinton, 2010) as the nonlinear activation function. ReLU can suppress noise while activating effective features, and it is simple to calculate and fast to converge.

3.2. Image reconstruction

The ultimate goal of MST-Net is to obtain the dehazed images of the same size as the input hazy images. Due to the existence of the downsampling layer in the feature extraction block, the size of the obtained feature maps is smaller, and a lot of features are lost, it is often difficult to obtain high-quality dehazed images by simple upsampling operations. To solve this problem, we propose the cascaded multi-scale transposed convolution blocks to improve the resolution of the feature maps gradually and obtain feature map pyramid. In addition, we introduce skip connections from the feature extraction blocks to compensate for the information loss of the feature map pyramid and finally recover the dehazed images using a transposed convolution.

3.2.1. Multi-scale transposed convolution block

The multi-scale transposed convolution block is designed to enlarge the size of the feature maps. Unlike traditional interpolation methods, the transposed convolution can learn different parameters to achieve a better upsampling effect. Inspired by Google Net (Szegedy et al., 2016), we use parallel transposed convolution operations with different filter to describe the feature information in the image. It is not susceptible to noise and has stronger robustness.

As shown in Figure 3, the multi-scale transposed convolution block contains three parallel transposed convolutions with stride 2, where the filter sizes are $2 \times 2$, $3 \times 3$, and $4 \times 4$.

The parallel method of multi-scale transposed convolution operations can capture the features of the image at different scales and restore the original features of the images more comprehensively. The multi-scale transposed convolution layer is formulated as:

$$V_i^s(T_s) = W_i^s *^{-1} T_s + B_i^s, (s = 1, 2, 3, i = 1, 2, 3)$$

Where $V_i^s$ is the $i$-th output feature maps of the transposed convolution block at the s-th layer, $W_i^s$ and $B_i^s$ represent the i-th filter and bias of transposed convolution block at the s-th layer, $*^{-1}$ denotes transposed convolution, $T_s$ is the input feature maps of the multi-scale transposed convolution block at the s-th layer.

High-resolution images reconstructed by transposed convolutions from low-resolution images are prone to checkerboard artifacts, which negatively affects visual effects. Therefore, we concatenate output feature maps of the multi-scale parallel transposed convolution to obtain the fused feature maps, and add a convolution layer with a kernel size of $1 \times 1$ after the fused feature maps. This improves the nonlinearity of the image reconstruction network and suppresses the checkerboard artifacts to some extent. It is defined as the following:

$$V_s = [V_1^s, V_2^s, V_3^s], (s = 1, 2, 3)$$

$$T_s = W_s * V_s + B_s, (s = 1, 2, 3)$$

$V_s$ is the fused feature maps of the s-th layer multi-scale transposed convolution, $T_s$ represents the output feature maps of the s-th layer multi-scale transposed convolution block, $W_s$ and $B_s$ is the convolution filter and bias added to suppress the checkerboard artifacts in the s-th layer block.

Figure 2. The feature extraction block (E-1) architecture. The architecture of other feature extraction blocks (E-2 ~ E-4) in the network is the same as E-1.
3.2.2. Image reconstruction method
Reconstructing high-resolution clear images from low-resolution feature maps is a challenging task because the obtained feature maps lose a lot of detailed information during the feature extraction process. In this paper, we expand the size of the feature maps hierarchically by cascading three multi-scale transposed convolution blocks and obtain feature map pyramid with different resolutions. This method reduces the loss of feature information caused by the direct up-sampling operation and is beneficial to the reconstruction of dehazed images.

In order to supplement the detailed features lost in the feature extraction process, we introduce skip connections from the feature extraction blocks so that the detailed features of the network directly participates in the image reconstruction process. As shown in Figure 1, feature maps of the same dimension are concatenated. This improves the quality of the reconstructed clear images and avoids gradient explosion or disappearance, making the network easier to train. Finally, we generate clear images with the same size as the input hazy images from the concatenated higher-dimensional feature maps by a transposition convolution layer.

3.3. Loss function
The smooth L1 loss is less sensitive to outliers. It has smaller gradient changes than the L2 loss during the initial training of the network and is more robust. When the value is close to 0, smooth L1 loss is smoother than L1 loss, which makes the network easier to converge in the subsequent training process. Therefore, we adopt smooth L1 loss as the loss function to calculate the error between the dehazed images and corresponding clear images. The function expression is:

$$L(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \begin{array}{ll} \frac{0.5}{2} (l_{gt} - F(l_{haze}))^2 & \text{if } |l_{gt} - F(l_{haze})| < 1 \\ |l_{gt} - F(l_{haze})| - 0.5 & \text{otherwise} \end{array} \right.$$  

(6)

Where $\Theta$ denotes the network parameters to be learned, $F$ represents the dehazing network proposed in this paper, $l_{gt}$ stands for the clear images used as the label, $l_{haze}$ stands for the input images, and $N$ is the number of image pixels.

4. Experimental results
To demonstrate the effectiveness of the proposed method, we conduct experiments on both synthetic and real-world haze image datasets. The synthetic dataset REAListic Single Image DEhazing (RESIDE) proposed by Li et al. (Li, Ren, et al., 2019) is applied to train and test our network. RESIDE is a large-scale benchmark which contains a large number of synthetic indoor and outdoor haze images with corresponding haze-free images. Because the indoor environment is limited by the illumination and scene depth, etc., the indoor synthetic hazy image is very different from the real-world haze images, which will interfere with the training of the network. Therefore, we train the...
network both on indoor and outdoor synthetic images to validate the dehazing effect of our method respectively. Moreover, to evaluate the generalization of the proposed method in this work, the experiments on several real-world haze images are carried out. All the results are compared with five state-of-the-art dehazing methods: Dark Channel Prior (DCP) (He et al., 2009), DehazeNet (Cai et al., 2016), MSCNN (Ren et al., 2016), AOD-Net (Li et al., 2017) and FFA-Net (Qin et al., 2019).

4.1. Results on RESIDE dataset

ITS (Indoor Training Set)-V2 is one of the subsets of the RESIDE-standard, which is selected as the indoor training data. There are 1399 clear images and 13,990 hazy images synthesized by corresponding clear images, in which each clear image generates 10 different hazy images. 10,000 indoor hazy images are selected randomly for training and the rest hazy images as a non-overlapping test dataset. Similarly, OTS (Outdoor Training Set) – ALPHA in the RESIDE-v0 is used as the outdoor training dataset. OTS-ALPHA contains 313,950 synthetic outdoor hazy images which are generated by 8,970 clear images, and each clear image corresponds to 35 hazy images. 14,000 outdoor haze images and 3000 outdoor hazy images are selected as non-overlapping training samples and test samples respectively.

All experiments were performed on a PC with NVIDIA GeForce GTX 1080Ti, and our network training is based on the PyTorch framework. The samples are resized to 512 x 512 as the input images of the MST-Net. The whole network is trained for 50 iterations and batch size of 2. In the training, we adopt Adadelta (Zeiler, 2012) as the optimization algorithm, which can adjust the learning rate adaptively.

4.1.1. Evaluations on indoor dataset

We quantitatively and qualitatively evaluate the performances of our proposed method with the other five dehazing methods. The visual results are shown in Figure 4. We observe that DCP (He et al., 2009) can effectively remove the haze in the images, however, the colour of the dehazed images tends to be darker and supersaturated visually. DehazeNet (Cai et al., 2016) and MSCNN (Ren et al., 2016) cannot remove the haze completely, which may be due to the estimates of transmission maps and global atmospheric light are not accurate enough. Although AOD-Net (Li et al., 2017) shows a good dehazing effect, there is still a small amount of haze remaining. It can be observed that the result of FFA-Net (Qin et al., 2019) retains the image details well but the dehazing effect is not very ideal. In order to observe the detail information of the indoor results, some regions in the experiment results are enlarged and shown in Figure 5. We can see that the dehazing effect of DehazeNet (Cai et al., 2016), AOD-Net (Li et al., 2017) and FFA-Net (Qin et al., 2019) is not complete, especially FFA-Net (Qin et al., 2019) still retains a lot of haze. The dehazing results of
DCP (He et al., 2009) and MSCNN (Ren et al., 2016) are better but there is slight colour distortion. In contrast, the method proposed in this paper can remove the haze more completely, and the obtained dehazed images have good colour fidelity and are closer to the clear images.

In order to demonstrate the method proposed in this paper quantitatively, we choose Structural Similarity (SSIM) (Wang et al., 2004) and Peak Signal-to-Noise Ratio (PSNR) (Huynhthu & Ghanbari, 2008) as evaluation indexes. SSIM and PSNR respectively reflect the similarity and completeness of the structural information of the images. The larger the value denotes the better the dehazing effect of the images. As shown in Table 1, the method proposed by us achieves the best results among all the methods.

| Methods                  | SSIM | PSNR  |
|--------------------------|------|-------|
| DCP (He et al., 2009)    | 0.72 | 15.78 |
| DehazeNet(Cai et al., 2016) | 0.79 | 19.05 |
| MSCNN (Ren et al., 2016) | 0.77 | 18.03 |
| Aod-Net (Li et al., 2017) | 0.76 | 17.97 |
| FFA-Net (Qin et al., 2019) | 0.72 | 14.92 |
| Ours                     | 0.88 | 25.80 |

4.1.2. Evaluations on outdoor dataset

Similar to the previous section, we select 3000 synthetic outdoor hazy images non-overlapping with the training images to evaluate these dehazing methods. From Figure 6, we can observe that the dehazing results of DCP
Figure 7. Dehazing comparison of the enlarged area on outdoor synthetic hazy images.

Table 2. The average results of SSIM and PSNR on the outdoor test samples.

| Methods                | SSIM  | PSNR |
|------------------------|-------|------|
| DCP (He et al., 2009)  | 0.83  | 17.52|
| DehazeNet (Cai et al., 2016) | 0.88   | 23.17|
| MSCNN (Ren et al., 2016) | 0.83  | 17.70|
| AOD-Net (Li et al., 2017) | 0.89  | 18.39|
| FFA-Net (Qin et al., 2019) | 0.82  | 18.32|
| Ours                   | 0.91  | 26.14|

(He et al., 2009) tend to severe colour distortion, especially for the sky regions, such as the images in the second and third rows. DehazeNet (Cai et al., 2016) and FFA-Net (Qin et al., 2019) retain the colour information of the input haze images greatly, however, in some dehazed images, there are some haze remaining. The dehazing results of MSCNN (Ren et al., 2016) and AOD-Net (Li et al., 2017) are impressive, but the disadvantage is there is colour distortion. For the dehazing effect and colour restoration of the outdoor synthetic hazy images, our method obtains the best visual result. Moreover, the quantitative results are shown in Table 2 also demonstrate that our results are better than the five other state-of-the-art methods. In order to observe the detail information of the results, we enlarge some regions of the dehazing results as shown in Figure 7.

The results show that DCP (He et al., 2009), MSCNN (Ren et al., 2016) and AOD-Net (Li et al., 2017) all have a certain degree of colour distortion, especially DCP (He et al., 2009), which shows a large degree of colour difference. DehazeNet (Cai et al., 2016), FFA-Net (Qin et al., 2019) and our method better retain the original colour of the input images.

4.2. Results on real-world haze images

Since haze in real-world haze images maybe not evenly distributed, removing haze in natural images becomes more challenging than in synthesized images. To demonstrate the dehazing ability of the proposed method in this paper, we conduct comparisons with five methods Dark Channel Prior (DCP) (He et al., 2009), DehazeNet (Cai et al., 2016), MSCNN (Ren et al., 2016), AOD-Net (Li et al., 2017) and FFA-Net (Qin et al., 2019) on several real-world haze images usually used for testing image dehazing.

The dehazed results are shown in Figure 8. It can be observed that DCP (He et al., 2009) has excellent dehazing results but suffers from colour distortion, especially in the sky region (see the first row, second column). FFA-Net (Qin et al., 2019) cannot completely remove the haze, such as the dolls in the third row, sixth column. In contrast, DehazeNet (Cai et al., 2016) and MSCNN (Ren et al., 2016) achieve a better dehazing effect but there is still a small amount of haze remaining on the road in the fourth image. AOD (Li et al., 2017) tends to produce darker dehazed images in some regions (notice the fifth row). Among them, some areas were enlarged and displayed in Figure 9. The results show that the dehazing effect of DehazeNet (Cai et al., 2016), MSCNN (Ren et al., 2016), AOD (Li et al., 2017) and FFA-Net (Qin et al., 2019) are not thorough enough, especially the blue license plate has obvious haze, DCP (He et al., 2009) and our method have better dehazing effect. Compared with the above five methods, the image dehazing results generated by our method are more complete and the colours are more realistic in visual.

5. Conclusion

In this paper, we propose a novel method based on pyramid multi-scale transposed convolutional network (MST-Net) for single image dehazing. Different from the traditional methods based on the atmospheric scattering model, the proposed network constructs the end-to-end mapping from the hazy images to the clear
Haze-free image reconstruction is an ill-posed problem, in order to obtain the high-pixel clear image, the network adopts multi-channel concatenation structure to extract diversified feature information and designs a multi-scale transposed convolution block to gradually expand the image resolution. We compare MST-Net with various state-of-the-art methods on synthetic and real-world haze images. The experimental results show that the images obtained by our network have excellent dehazing effects and colour fidelity visually.

**Disclosure statement**
No potential conflict of interest was reported by the author(s).

**Funding**
This work was supported by the [National Key Research and Development Project #1] under Grant [number 2018YFC0604500]; [Technology Project of Henan Province #2] under Grant [number 192102210100]; and [Key Scientific Research Projects of Higher Education Institutions in Henan Province #3] under Grant [number 19A413008].
References

Berman, D., Treibitz, T., & Avidan, S. (2017). *Air-light estimation using haze-line*. International conference on computational photography, Stanford, CA, USA, 12-14 May 2017.

Cai, B., Xu, X., Jia, K., Qing, C., & Tao, D. (2016). Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11), 5187–5198. https://doi.org/10.1109/TIP.2016.2598681

D. W. (2015). The latest research progress of image dehazing. *Acta Automatica Sinica*, 2015, 41(2), 221–239. https://doi.org/10.16383/j.aaas.2015.c131137

Deng, Z., Zhu, L., Hu, X., Fu, C., Xu, X., Zhang, Q., Qin, J., & Heng, P. (2019). Deep multi-model fusion for single-image dehazing. International conference on computer vision, Seoul, Korea (South), Korea (South), 27 Oct.-2 Nov. 2019.

Dudhane, A., & Murala, S. (2019). CDNet: Single image de-hazing. Workshop on applications of computer vision.

Fattal, R. (2008). *Single image dehazing*. International conference on computer graphics and interactive techniques, https://doi.org/10.1145/1399504.1360671.

Fattal, R. (2014). Dehazing using color-lines. *ACM Transactions on Graphics*, 34(1), 1–14. https://doi.org/10.1145/2651362

Goodfellow, I., Pougetabadjie, J., Mirza, M., Xu, B., Wardefarley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Neural information processing systems.

He, K., Sun, J., & Tang, X. (2009). Single image haze removal using dark channel prior. Computer vision and pattern recognition.

He, K., Sun, J., & Tang, X. (2010). *Guided image filtering*. European conference on computer vision, Heraklion, Crete, Greece, September 5-11, 2010.

Huang, G., Liu, Z., Der Maaten, L. V., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Computer vision and pattern recognition.

Huynhthu, Q., & Ghanbari, M. (2008). Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 44(13), 800–801. https://doi.org/10.1049/el:20080522

Li, Y., Miao, Q., Ouyang, W., Ma, Z., Fang, H., Dong, C., & Quan, Y. (2019). LAP-net: Level-aware progressive network for dehazing. International conference on computer vision, Seoul, Korea (South), Korea (South), 27 Oct.-2 Nov. 2019.

Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). AOD-net: All-in-one dehazing network. International conference on computer vision, Venice, Italy, 22-29 Oct. 2017.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., & Wang, Z. (2019). Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1), 492–505. https://doi.org/10.1109/TIP.2018.2867951

Liu, Y., Pan, J., Ren, J., & Su, Z. (2019). Learning deep priors for image dehazing. International conference on computer vision, Seoul (South), Korea (South), 27 Oct.-2 Nov. 2019.

Ma, J., Fan, X., Ni, J., Zhu, X., & Xiong, C. (2017). Multiscale retinex with color restoration image enhancement based on Gaussian filtering and guided filtering. *International Journal of Modern Physics B*, 31(16-19), 1744077. https://doi.org/10.1142/S0217979217440775

Ma, Q., Pei, T., Wang, X., Tang, Q., Zhou, Q., & Yu, T. (2019). *Single image dehazing using sky adaptive fusion*. International conference on image vision and computing, Xiamen, China, China, 5-7 July 2019.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. International conference on machine learning, Haifa, Israel, June 21-24, 2010.

Narasimhan, S. G., & Nayar, S. K. (2002). Vision and the atmosphere. *International Journal of Computer Vision*, 48(3), 233–254. https://doi.org/10.1023/A:1016328200723

Park, J., Han, D. K., & Ko, H. (2020). Fusion of heterogeneous adversarial networks for single image dehazing. *IEEE Transactions on Image Processing*, 29, 4721–4732. https://doi.org/10.1109/TIP.2020.2975986

Qin, X., Wang, Z., Bai, Y., Xie, X., & Jia, H. (2019). FFA-net: Feature fusion attention network for single image dehazing. *arXiv: Computer Vision and Pattern Recognition*. 28

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Neural information processing systems.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M. (2016). *Single image dehazing via multi-scale convolutional neural networks*. European conference on computer vision, Amsterdam, The Netherlands, October 11-14, 2016.

Stark, J. A. (2000). Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing*, 9(5), 889–896. https://doi.org/10.1109/83.841534

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. *IEEE Transactions on Image Processing*, 24(11), 3522–3533. https://doi.org/10.1109/TIP.2015.2446191