




Single Image Desmoking via Attentive Generative Adversarial Network for Smoke Detection Process

Yufeng Huang * and Xiang Chen*, College of Electronic and Information Engineering, Shenyang Aerospace University, Daoyi South Avenue 37, Shenyang 110136 Liaoning, China

Lei Xu, Shenyang Fire Research Institute of MEM, Wenda Road 218-20, Shenyang 110136 Liaoning, China

Kaiyuan Li, School of Safety Science and Emergency Management, Wuhan University of Technology, Luoshi Road 122, Wuhan 430070 Hubei, China

Received: 4 June 2020/**Accepted:** 15 January 2021/**Published online:** 30 January 2021

Abstract. Smoke removal has been a challenging problem which refers to the process of retrieving clear images from the smoky ones. It has an extensive demand while the performances of existing image desmoking methods have limitations owing to the homogeneous medium based model, hand-crafted features, and insufficient datasets. Unlike the existing desmoking methods using an atmospheric scattering model, the proposed method utilizes a general deblurring model. Inspired by Generative Adversarial Network (GAN), we propose an end-to-end attentive DesmokeGAN which implements the visual attention into the generative network to effectively learn the smoke features and their surroundings. The architecture of critic network is identical to PatchGAN by adding multi-component loss function to the image patch. Due to the scarce of various and quality smoke datasets, a graphics rendering engine is used to synthesize the smoky images. The quantitative and qualitative results show that the designed framework performs better than the recent state-of-the-art desmoking approaches on both synthetic and real images in indoor and outdoor scenes, especially for the thick smoky images.

Keywords: Desmoking, Generative adversarial network (GAN), Attentive mechanism, Smoke synthesis

1. Introduction

Due to the existence of smoke in fires, the images taken under fire conditions are certainly subject to blurring, color distortion and other visible quality degradations in the background scene. The smoky image may seriously affect the subsequent tasks such as safety monitoring system, occupant evacuation and fire-fighting processes. The exiting smoke detection technologies almost focus on the

*Correspondence should be addressed to: Yufeng Huang, E-mail: yufengh_sau@sina.com; Xiang Chen, E-mail: xiangchen_sau@126.com



early detection for the thin smoke environment, while the thick smoke shielding in the fire scenario is more important for the rescuing process. Therefore, image desmoking, as a preprocessing step, is used to retrieve the clear images from the smoky ones. Single image desmoking is a fundamental image processing technique and it has attracted increasing attention in the computer vision community during recent years.

Notably, less attention has been paid on image desmoking as it is frequently misbelieved as a similar topic of dehazing which has been well studied for decades. The tractional approaches are mainly divided into mathematical and physical models. The mathematical models usually target at improving the pixel intensity, including the contrast enhancement [1], histogram analysis [2], homomorphic filtering [3] and Retinex [4]. The physical model usually considers the image processing as an inverse and ill-posed matter, solves it by optimizing its corrupted image using atmospheric light and scene transmission map. The commonly used models include the atmospheric scattering models such as a probabilistic graphical atmospheric light model [5–7], dark channel prior (DCP) [8–10] and reduced formation model [11]. Although these methods are generally simple and fast to implement, there are several limitations. The mathematical models need to adjust too many threshold parameters to accommodate different conditions, while the physical methods use the model to estimate the transmission map, so that it may fail in the case of unsuitable modeling.

Recently, learning-based methods start to utilize Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) for image dehazing and desmoking. Based on the atmosphere scattering model, the CNN based methods [12–15] have mainly focused on the regressing of transmission map and clear images using multiple models and features. The GAN based methods have employed the generator and discriminator to recover the corrupted image to a clear one. For instance, the conditional GAN [16, 17] have been adopted to remove haze from an image, where the clear image is estimated by certain conditional model and further optimized using the multi-task methods [18]. Engin [19] improves the Cycle-GAN networks by combining perceptual losses and cycle-consistency, and gets visually better dehazing image. Disentangled Dehazing Network (DDN) [20] is proposed to estimate transmission map, the scene radiance and atmosphere light by utilizing three generators simultaneously. Using the fusion coding of contours and colors, Tan [21] exploits an end-to-end model via simulating visual perception with depth decoding. The previous deep learning studies have been focusing on haze removal and seldom considered the special features of smoke itself. Therefore, a more adaptive desmoking approach, which should take into account the smoke features under various conditions, should be introduced.

This paper conducted a study targeting specifically at the problems of desmoking images, such as halo, unnatural colors and blurring. In order to get enough dataset for deep learning of desmoking, a graphics rendering engine, Blender, was used to synthesize the smoky image mimicking the smoky environments. We developed an end-to-end attentive desmoking method using a single neural network that contained a conditional GAN with gradient penalty, which did not use the prior model or any post-processing technique. Furthermore, the architecture

of the current network was identical to PatchGAN by adding the attentive weight information and multi-component loss. In summary, the contributions of the current work include:

① The smoky image dataset is synthesized via rendering the realistic smoke spread process. The datasets contained a set of diverse conditions and densities of smoke in both indoor and outdoor settings without the need of any manual labeling. It offers an effective and cheap way to generate continuous and realistic smoky image pairs.

② An attentive end-to-end DesmokeGAN is proposed to learn the effective features of smoky images for the estimation of transmission map. Considering the blurring and corruption caused by smoke, the network describes a kernel-free blind smoke blurring and the additive noise instead of using the atmospheric scattering model. It might provide a new idea to deal with the medium with nonuniform distribution and nonhomogeneous density in the entire scene.

③ A novel DesmokeGAN framework is employed. For the generator, the attentive mechanism is added according to the DCP information, since it helps the network to identify the smoke removal region. For the discriminator, a PatchGAN classifier is used with weighting the adversarial and perceptual Loss. The network could reduce the presence of artefacts and preserve the structure significantly.

2. Related Work

A variety of approaches have been proposed to overcome the degradation caused by smoke. In this section, we briefly reviewed the most related single image desmoking technology and GAN used in the paper.

2.1. The Dark Channel Prior (DCP)

The DCP proposed by He et al. [8] was based on the finding that the haze-free regions-pixels should have at least one color channel intensity value close to zero. Prior to adding the histogram equalization process, the DCP dehazing method was modified for the desmoking purpose [22]. The DCP was further refined using the optimized Gaussian Markov Random Field model to recover the image [23]. Pei [9] improved the DCP by estimating the transmission map for each color channel. Though there are various methods aiming at enhancing the contrast and color, the DCP methods suffered from falsely detecting the object and hardly preserved original color due to its own limitation.

2.2. Attention Mechanism

Attention helps visual system to focus on salient parts, which plays an important role to capture visual structure better [24]. Recently, several attempts have been put [25–30] to incorporate attention processing to improve the networks performance of image process. Hu et al. [26] introduced a channel-wise attention module to exploit the inter-relationship. Using the attention mechanism, the machine translation models [27, 28] could achieve a better result in image generation. The

methods [25] designed a Residual Attention Network, which refined the feature maps using an encoder-decoder style attention module. In [29], attention was formalized as a non-local operation to calculate the spatial dependencies in the video process. Zhang et al. [30] employed the residual channel attention network to optimize the image super-resolution. Inspired by the work, the feature attention module [31] is designed to recover the haze image using CNN network. In spite of these image progresses, the attention mechanism has seldom yet been explored in image desmoking.

2.3. Generative Adversarial Networks

The idea of GAN, introduced by Goodfellow [32], forms a two-player minimax game. It consists of two competing models: the generator G and discriminator D . The generator G learns to generate artificial samples and uses it to fool the discriminator. The discriminator D distinguishes the real data from the sample generated by the generator. The goal of capturing the real data distribution can be achieved as the generating convincing sample can not be identified from the real one. The minimax game between the G and D is formulated as the following function:

$$\min_G \max_D \int_{z \sim p_r} E [\log D(x)] + \int_{\bar{x} \sim p_\theta} E [\log(1 - D(\bar{x}))]; \text{ as, } \bar{x} = G(z), z \sim {}^g P(Z) \quad (1)$$

where D is the set of 1-Lipschitz functions. The value here approximates $K \cdot W(P_r, P_\theta)$ as $W(P_r, P_\theta)$, where W is the Wasserstein distance and K is the Lipschitz constant. To enforce K in the GAN, the weight clipping is set to $[-c, c]$ and the gradient penalty can be improved as [33]:

$$\lambda \int_{\bar{x} \sim p_{\bar{x}}} E \left[(\|\nabla_{\bar{x}} D(\bar{x})\|_2 - 1)^2 \right] \quad (2)$$

Furthermore, the conditional GAN [34–36] has been extensively applied to image translation as *pix2pix*. They receive the observed image or the label as inputs and markovian discriminator to the latent code. The game function with the value $V(D, G)$ can be written as:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(x)))] \quad (3)$$

2.4. Deep Learning in Desmoking

With the recent breakthrough of deep learning, there have been some techniques focusing on desmoking. Bolkar [37] might be the first person who applies deep learning desmoking approach to surgery videos. The De-Haze and Smoke GAN(DHSGAN) [38], an end-to-end network, proved that deep learning could help dealing with the smoky images. Sidorov [39] solved the surgery smoky problem by adding perceptual quality metric in GAN's loss function. The De-smoke

Generative cooperative network [40] was established with a novel scheme that regarded tasks of smoke detection and removal as two separate ones. The Synthetic dataset was also one important part in deep learning because of the lack of real smoky data. In [34], the data was generated by adding Perlin noise and using for fine tuning AOD-Net [13]. Later, the synthetic data could be made by Blender (<https://www.blender.org/>) for the Joint Surgical desmoking images [40]. Nevertheless, there are still a lot of work that needs to be done in network structures and the smoky dataset.

3. Proposed Method

The goal of this paper is to remove the smoke, while maximally keeping the original structure and color in the image. For the deep learning methods, large amounts of high-quality and easy-access synthetic data may play an important role in the training processing as well as the network model and architecture. Then based on human visual mechanism, the feature attention methods will be proposed, which can accelerate and improve the process of desmoking. We introduced the proposed pipeline in detail, including the smoke synthesis, the based model and overall network architecture. These components were successively described in detail as below.

3.1. Smoke Synthesis

Taking the huge and actual number of datasets for training networks is a very expensive and time-consuming work, particularly as the smoke datasets must require certain combustion conditions which affect the natural environments. It is not easy to get thousands of different scenes of image pairs (with and without smoke). For numerous image pairs, acquiring density masks and labeling manually seems impossible for real datasets. Synthetic datasets can provide detailed ground-truth scenes, as well as easy scalable alternatives to annotate images manually. To deal with the current practical issue, the need of synthetic smoky datasets has considerably increased.

As known, the traditional haze model [41, 42] and a Perlin noise function [37] cannot address the special characteristics of smoke [9]. To get better and more realistic synthetic smoke images, we employed an open source 3D graphics engine to generate the smoke images for training. The 3D graphics engine can render the smoke based on the physical smoke movements models and set depth information and colors value in RGB channel separately, which is offered by the Blender. Here, we introduced the synthetic process in details. The primary smoke $I_{smoke'}$ can be defined as:

$$I_{smoke'}(x, y) = Blender(D_{rand}, I_{rand}, P_{rand}) \quad (4)$$

where D_{rand} is the density, I_{rand} is the intensity and P_{rand} is the position of smoke generation. The density D_{rand} means that the smoky solid particles unevenly diffused at the certain volume. The intensity I_{rand} represents the smoky solid particles

transferred at certain degree. The Position P_{rand} is the general smoky starting position in the picture area. As the images are generally in color type, the luminosity function of RGB channels can be formulated as:

$$I_{smoke}(x, y) = \left(0.3 * I_{smoke'}(x, y)^R\right) + \left(0.59 * I_{smoke'}(x, y)^G\right) + \left(0.11 * I_{smoke'}(x, y)^B\right) \quad (5)$$

The smoky image can be made by overlaying the smoke image with different densities, intensities and locations on the smoke-free image.

$$I_{smoked-image}(x, y) = I_{smoke-free}(x, y) + I_{smoke}(x, y) \quad (6)$$

The random of rendered process can avoid the over-fitting of the network and can generate enough synthetic smoke images for training, as shown in Fig. 1. In the synthetic process, the synthetic scene has been generated with the ground truth image, which adds the smoke mask with various locations and smoke levels using the 3D graphics engine. Generally, the generated locations are roughly divided into four positions: top, bottom, left and right. The smoke densities are graded

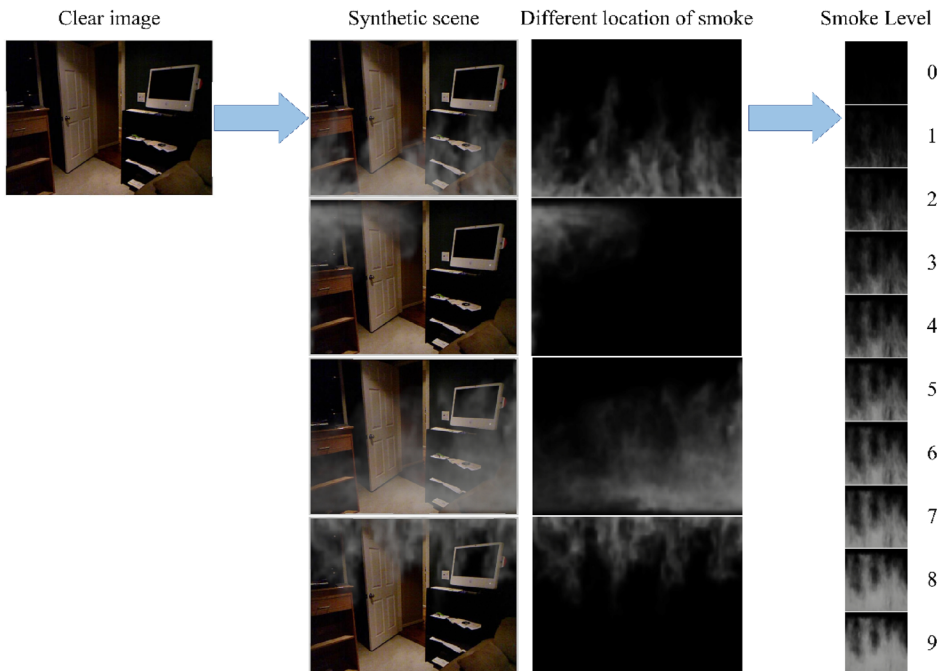


Fig. 1. Left: the ground truth image. Middle: smoke rendered images and the smoke masks of different location. Right: 10 levels of a smoke mask.

into 10 degrees ranging from 0 to 9, where 0 is defined as no smoke, 9 is defined as the max smoke density in Fig. 1.

3.2. Network Architecture and Loss Function

3.2.1. The physical model The goal of removing smoke is to produce a clear image with only one input of the smoky image while maximally keeping the original feature of the smoke-free image. The atmospheric scattering model is strongly based on the assumption that the medium is homogeneous and the light will follow the spreading law in atmospheric. However, the smoke density has its local property, that means, it may alternate from one area to another and is heterogeneously distributed in one scene. Moreover, the light conditions are also complex in most fire conditions, especially the indoor scene. As the atmospheric scattering model does not lend itself to fire scenario, we need to search for other proper ways to deal with the smoke removing issue. The smoky images exit structure blur and color shifts along with the smoke. Based on the characters of smoke, the common formulation of non-uniform smoke model is adopted as follows:

$$I_{smoke-image}(x, y) = k(M) * I_{smoke-free}(x, y) + N \quad (7)$$

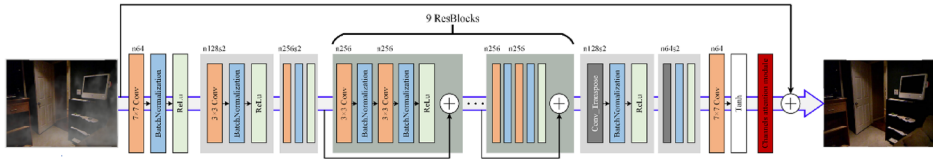
where $I_{smoke-image}$ is the observed smoky image, $k(M)$ the unknown blur kernel determined by smoke transmission influenced by the kinds of smoke, and $I_{smoke-free}$ the clear image. The $*$ denotes the convolution and N is an additive noise reference with the smoke thickness.

3.3. Network Architecture

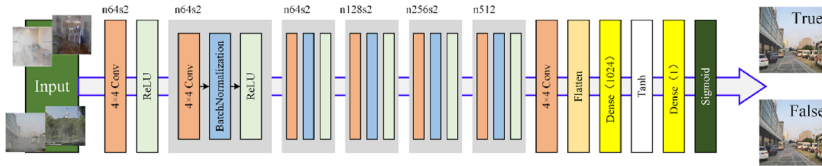
As a baseline for the implementation of the desmoking approach, we use the conditional GAN that is generally similar to the one developed by Isola [34]. This network structure consists of the generator G and discriminator D and the goal is to learn a generator which recovers the clear images properly.

The generator G is shown in Fig. 2 and it consists of two stride convolution blocks, nine residual blocks [43] with 3*3 kernels and two transposed convolution blocks. Similar to [44], we used batch-normalization [45] followed by ReLU with $\alpha = 0.2$ activation function in each residual block, as shown in Fig. 2a. For the generator, the resnet-based architecture with global skip connection is used, which should enhance the train ratio and reserve the original features of input sample. For the convolution layer in the ResBlock, the dropout processing has been imported with a probability value of 0.5.

The discriminator D is employed to figure out the discrimination between the target and generated images. Here, the conditional GAN's discriminator [34] performs patch-wise comparison of the desmoky image and clear image. The network consists of four convolutional layers with batch-normalization and ReLU with a parameter $\alpha = 0.2$ activation, as shown in Fig. 2b. The discriminator network also introduced a Dense net with an activation processing of tanh and sigmoid in sequence. During the test phase, the critic network gained a valuable performance, which contains the Wasserstein distance [46] and gradient penalty [33].



(a) Generator Architecture



(b) Discriminator Architecture

Fig. 2. The pipeline architecture of Generator a and Discriminator b network with corresponding number of feature maps(n) and strides(s) demonstrated in each convolution layer.

3.4. Loss Function

We calculated the loss functions as a sum of adversarial loss and perceptual loss. The adversarial loss focuses on restoring the texture details and the deeper layers denote the features of a higher abstraction [47, 48], while the perceptual loss restores general content. The total loss is given by:

$$Loss_{total} = W_{gan} * L_{adv} + W_{vgg} * L_{vgg} \quad (8)$$

Adversarial Loss: In [34, 49, 50], vanilla GAN objective related to conditional GAN has been used as the main loss function. To improve the numerical stability and effectiveness, the alternative least squares GAN has been used to generate higher quality results. The WGAN-GP is the discrimination function, which gets robust choosing from the generator. The adversarial loss can be calculated as:

$$L_{adv} = \sum_{n=1}^N -D_{\theta_D}(G_{\theta_G}(I^B)) \quad (9)$$

Perceptual Loss: As stated in [50, 51], the perceptual loss is a simple content loss added the feature reconstruction perceptual loss. The generated and target images are gone through the VGG-19 [52] conv 3*3 feature maps. The perceptual loss of these two images is calculated after passing through the pool layer. At test time, only the generator is kept.

$$L_{vgg} = \frac{1}{CWH} \sum_{c=1}^C \sum_{w=1}^W \sum_{h=1}^H \left\| V(G(x, z)^{c,w,h}) - V(y^{c,w,h}) \right\|_2^2 \quad (10)$$

where, C , W , H is respective the output's channels, width and height.

3.5. Channel Attention Module(CAM)

Reference to DCP [8], the proposed CAM follows that different channel features have totally different weighted information. The channel-wise global spatial information has been transformed into a channel descriptor using pooling process, with the equation as follow:

$$g_c = H_p(F_c) = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W X_c(h, w) \quad (11)$$

where, $X_c(h, w)$ is the value of c -th channel X_c at the location (h, w) and H_p is the pooling function. Then, the feature image shape changes from $C \times H \times W$ to $C \times 1 \times 1$. To obtain the weights of various channels, the Sigmoid and ReLU activation functions are selected following with two convolution layers as shown in Fig. 3.

$$CA_c = \sigma(\text{conv}(\delta(\text{conv}(g_c)))) \quad (12)$$

where σ is the sigmoid function and δ is the ReLU function. Thus, the CAM is listed under a multiplication cross process with the input F_c and weights information CA_c .

4. Experimental Results

This section briefly introduces the experimental datasets, training details and evaluation metrics. Then, we effectively and quantitatively verified our method against several state-of-the-art algorithms on synthetic and real-world smoky images. In addition, comparison of the density limit test shows the superiority in density smoky images.

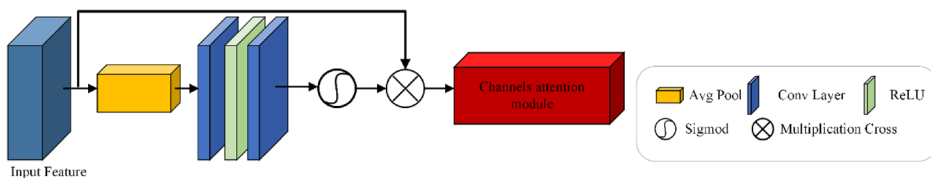


Fig. 3. The Channel Attention module.

4.1. Datasets

Similar to the recent deep learning methods, the desmoke methods require a huge number of datasets for training and testing. However, since no public dataset exists with a full set of smoky and clean images, we had to create our own dataset which contained synthetic and real image pairs for indoor and outdoor scenes.

To obtain the synthetic data, the clear background images from publicly indoor datasets NYU-Depth [53] and outdoor datasets RESIDE-OTS [54] are obtained. Then, the graphics rendering tool blender is used to generate the smoked data. It is confirmed that the smoke of different densities, intensities and positions were added to generate a diverse smoky dataset. In general, about 6000 synthetic image pairs are produced and an example is shown in Fig. 4. To ensure the effectiveness of the method, the real-world data are captured using a Sony A6000 camera while some data are obtained from the internet. While acquiring the real-world data, it is ensured that the images gathered are varied in terms of density, intensity and position of smoke. In total, about 2400 pieces of real-world smoky image are obtained with various background scenes and smoke. For the experimental convenience, the training and verification data are uniformly cropped into 640*480 pixels.

4.2. Training Details

The detailed structure and parameter setting of proposed model are given in Fig. 1 using TensorFlow [55] framework. In the training process, for optimization, we followed the method of WGAN [46] and carried the update ratio 5 on generator G and one on discriminator D. Using Adam optimization method [56] as a solver, its parameters can be set as followed: the momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, the learning rate 1×10^{-4} and batch size 4. Empirically, the weights values are set as $W_{gan} = 1$, $W_{vgg} = 100$. In this study, 90% of the synthetic image pairs (5400) are served as the training datasets randomly and the left 10% synthetic image (600) and real images (2400) pairs are used for evaluation. The training time is approximately 10 h for 100 epochs on a workstation with a NVIDIA Tesla V100 GPU (16 G).



Fig. 4. An example of synthetic smoke image. A smoky image b can be served as the superposition of a clear background image a and a smoke layer image c.

4.3. Evaluation Metrics

A smoke removal method's performance can be evaluated on several criteria and among them two of the most commonly used criteria are Peak Signal-to-Noise Ratio (PSNR in dB) and the Structural Similarity Index Measure (SSIM) [57]. The higher value of PSNR indicates better performance to remove smoke from the smoky image. The greater SSIM score nearest to 1 means that the two different images are more similar to each other. However, single criterion may not present the fully properties, so that a score metric with a weighted sum them two is mentioned [16] as:

$$Score = W_{PSNR} * PSNR + W_{SSIM} * SSIM \quad (13)$$

The PSNR weight (W_{PSNR}) and the SSIM weight (W_{SSIM}) are set to 0.05 and 1 separately. The higher score means good properties in visually pleasing and a good smoke removal quality. By using these criteria, the performances of models are evaluated.

4.4. Quantitative and Qualitative Comparison on Synthetic Images

In this section, we revealed the effective performance of our method by conducting a mass of experiments on synthetic and actual datasets. The proposed method was compared to four recent state desmoking methods: MSR [4], DCP [8], AOD-Net [13], and DHSGAN [38]. All the methods used the source codes and default parameters specified published in the literature. As the availability of ground truth in synthetic data, the results are evaluated using PSNR, SSIM and weight score under the indoor and outdoor synthetic smoke image. Table 1 shows the average evaluation criteria of each pairs of smoke-free and desmoking images. From the table, the proposed method obtained the highest value of PSNR over 20 in both indoor and outdoor image. The average SSIM results were also compared to the other related methods and even surpassing most of them, the higher indoor results might reveal that our method did an even better job in the indoor environment. The most notably increasing score noted that our approach could properly

Table 1
Quantitative comparisons on the indoor and outdoor synthetic datasets of different methods

Methods	Indoor			Outdoor		
	PSNR	SSIM	Score	PSNR	SSIM	Score
MSR	13.51	0.6510	0.6835	14.89	0.6414	1.3859
DCP	15.32	0.7284	1.4944	16.75	0.7955	1.6330
AOD-Net	17.86	0.7732	1.6662	16.69	0.7600	1.5945
DHSGAN	19.13	0.8316	1.7881	18.84	0.8080	1.7500
Ours	26.05	0.9098	2.2123	24.35	0.8794	2.0969

Bold values indicate best result of desmoking and used to emphasis the highest value

remove the smoke and restore the image which is more similar to the ground truth one.

Some corresponding pictures directly show visually difference in indoor and outdoor, as particularly seen in Fig. 5. It can be clearly seen that the proposed method outperforms all the other related approaches and produces significant visibility improvement even in cases of dark light and thick smoke. The MSR and DCP failed to remove the smoke especially in the bright scene because these methods often involve parametric models and are not robust to various conditions. The deep learning methods performed better in dealing with the smoke, but there is still some smoke left in AOD-Net and DHSGAN, particularly in bright condition. Another benefit of our method is being good at preserving of structure and color information similar to the ground truth.



Fig. 5. Qualitative results of synthetic smoky images using several state-of-the-art desmoking methods and our proposed method in indoor and outdoor scene.

4.5. Qualitative Comparison on Real-World Images

To evaluate the effectiveness of the proposed method, we conduct a comparison on the real-world smoky images. The original smoky image, along with the desmoked counterparts are illustrated in Fig. 6. Notably, the output of proposed method looks better in visual and natural restored. In contrast, there are some limitations in other methods. For instance, there are still lots of smoke left in the MSR, Retinex and DCP's results indicating that these classical methods may not suitable for smoke remove although they do well in removing haze. AOD-Net can deal with a certain amount of smoke, but the color shift gloomy in both the smoky and no smoky areas. The DHSGAN seems fail to remove smoke due to being not robust enough to various smoke. In general, the proposed method can successfully remove majority of smoke while lead to less color distortion.

4.6. Qualitative Comparison on Real Smoke Images

To verify the practical use in real fire scenario, we further compare the proposed method against other algorithms on the real smoky images with relatively heavy smoke. Figure 7 demonstrates three real smoke samples of which the original images have non-uniform smoke spatially varying in the images. Even though all the methods remove smoke successfully for the synthetic datasets and real smoke scene with light smoke, there are some smoke left in these cases, especially in the heavy smoke conditions. Through the following color sample, the outputs have

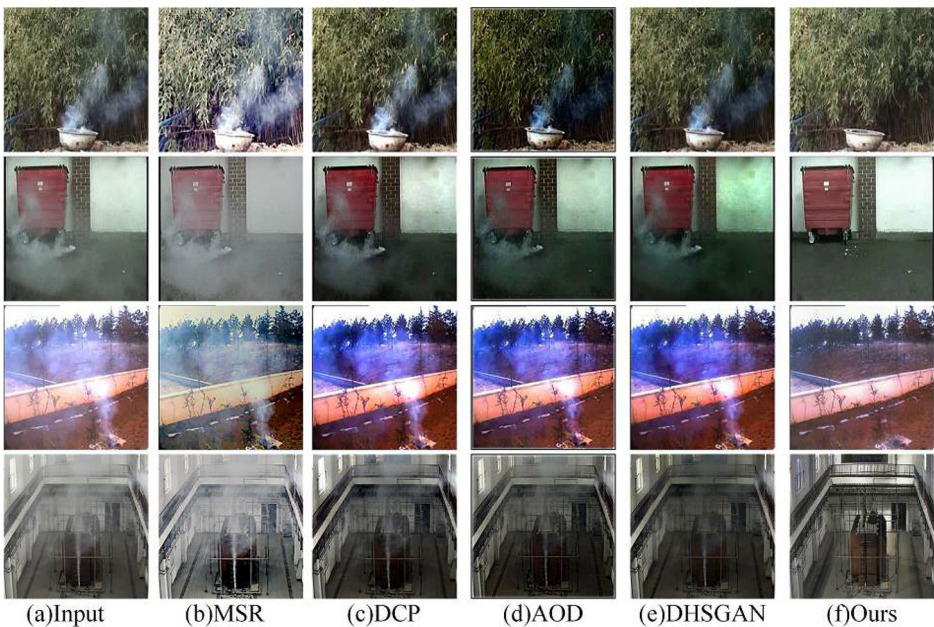


Fig. 6. Qualitative results of real light smoky images using several state-of-the-art desmoking methods and our proposed method.

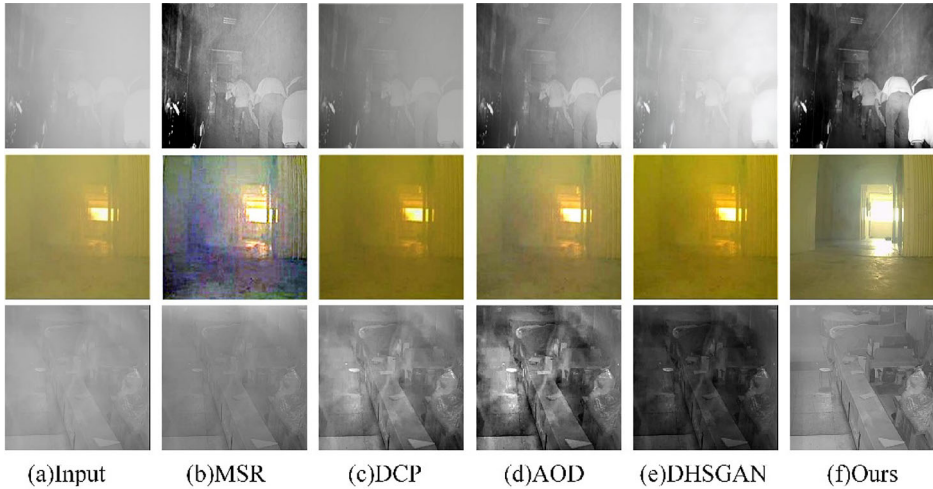


Fig. 7. Qualitative results of real heavy smoky images using several comparing desmoking methods and our designed method.

some limitations to restore the color, as the result of DCP and DHSGAN seem too dark and MSR have the unusual color shifting. Apparently, the proposed methods efficiently remove most of the smoke even in the non-uniform heavy smoke, as well as restore the color to some extent. In summary, the current algorithm achieved a satisfying result than the others with the real-world smokes which are thicker than those in Fig. 5 and 6.

4.7. Smoke Removal Limit Test

The structural loss and color fade, usually irreversible, largely depend on the degree of smoke densities. To further verify the ability of the above methods to recover the image under different degrees of smoke density, we carried out a performance research of desmoking under 10-degree smoke densities as mentioned in Smoke Synthesis section. We picked 60 images and its nearly 10-degree smoke densities image from the testing synthetic image at random, as well as the similar smoke densities of real-world image in two scenes: color and grey.

As shown in Fig. 8, the rendered smoke mask and an image with 10-degree smoke densities are in the first two rows and the desmoking images from related previous methods are shown in the following rows. Most of the previous methods can only efficiently remove smoke to a certain degree and are not robust enough to various degrees. With increasing smoke density, the previous methods' results contain more smoke residue and blurring. The deep learning methods give better performance than the traditional methods in most densities. The output results also have certain limitations to recover the proper colors as the MSR's results seem too bright while AOD-Net's results are too dark to recognize. These defects may lead to color shift in the no smoke areas and fail to recover the correct color for smoke covered place. The under or over-saturated exist as well as loss the

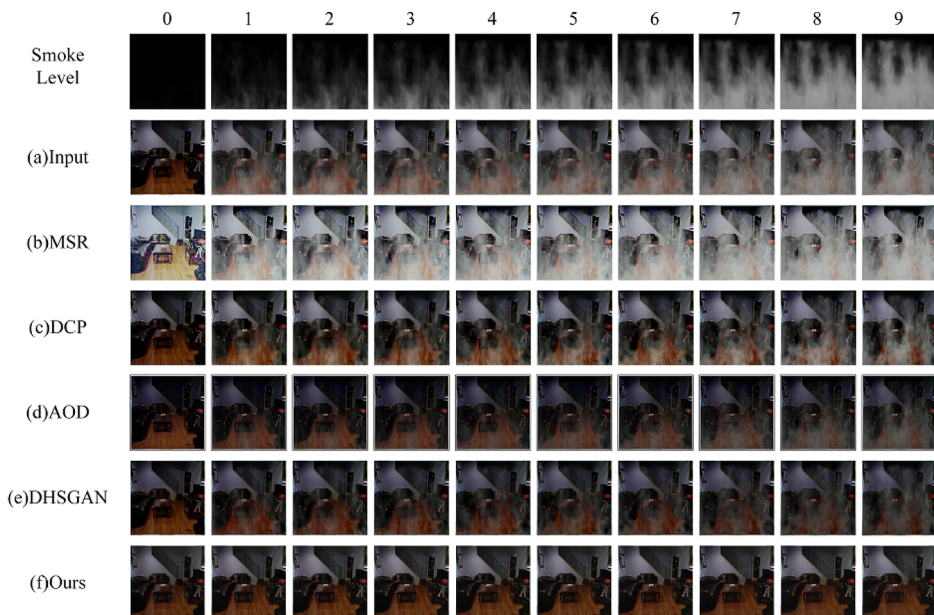


Fig. 8. Quantitative results on the indoor synthetic testing images for the smoke removal density test.

detail of the picture. Our approach has removed almost all the smoke and restored the clear images with fine structure and very slight color change even in the very thick smoke. Clearly, the proposed model performed well on the synthetic dataset with better robustness and reducing property.

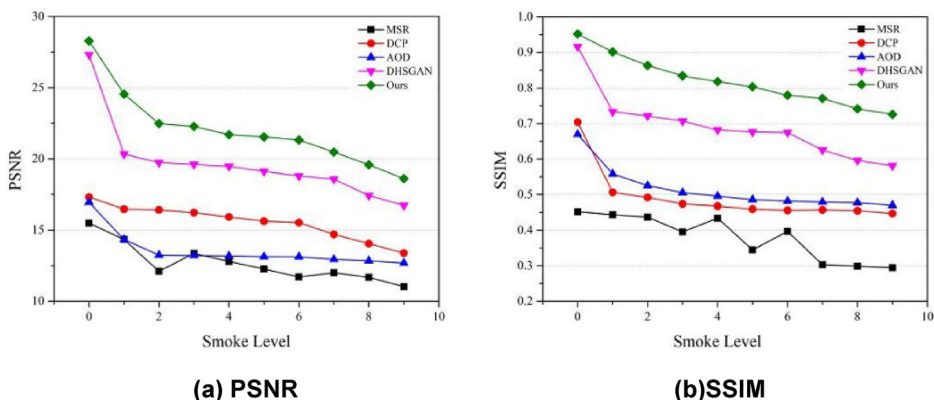


Fig. 9. The quantitative results of the smoke removal limit test. PSNR and SSIM results for our method and other comparison approaches under 10 different smoke degrees.

The quantitative results of different methods are shown in the Fig. 9, which indicates that the curves of SSIM and PSNR between image pairs of desmoke image and smoke-free image for our results and other methods under 10 different smoke levels. From thin to thick smoke, all the curves tend to decrease in PSNR and SSIM, which means that it is harder to deal with the high densities of smoke than the thin one. The curves are different from each other even at the beginning: 0 degree with no smoke, and it reveals that the methods may result in the color shift and structural change in the process even without the existence of smoke. The proposed model produces the highest SSIM and PSNR value for all 10 smoke levels, which are significantly superior to other previous methods. From the 0 to 9 degrees, the trends of SSIM and PSNR are stable at a high level, which indicates that our methods can robustly recover the correct structure and color under thin smoke as well as very thick smoke.

To further demonstrate the effectiveness of the proposed method, we test the related techniques on the real smoke scenes and the results of the two real scenes are shown in Fig. 10. Figure 10a illustrates the grayscale smoke removed results of the surveillance video compared with prior methods. Our method could almost remove all the smoke in different smoke levels, while the other methods still have some smoke left and it becomes worse with increasing smoke density. Due to the influence of the darker areas in the picture, the other deep learning-based results seem too dark to show the structure and detailed information, especially in the heavy smoke. Similar results can be seen in color scene in Fig. 10b, the proposed model can deal with the different levels of smoke of which the distribution concentrates in the upper position and effect less in the smoke-free area below. Other methods suffer from color shift too much and some of them (DHSGAN and AOD-Net) is too gloomy to identify the structure in smoke and smoke-free area.

5. Conclusion

In this paper, a novel GAN-based approach is proposed for smoke removal task in a single smoke image. Inspired by smoke relevant features and blurring, the learning model is reconstructed to estimate the features and transmission map. Meanwhile, the proposed method has improved the generator network using the feature attentive mechanism on the image patch. In addition, an innovative approach is used to create the realistic synthetic datasets using a 3D graphics engine. A new benchmark and evaluation protocol are employed to evaluate the synthetic and real datasets in indoor and outdoor scenes. Compare to the other methods, the proposed methods produce the better desmoking results in the synthetic and real datasets. The smoke removal limit test showed that the proposed method performs better with stronger robustness in even thick smoke.

The database, model algorithm and engineering application are the mainly challenges in the future. Therefore, our future work will focus on the interpretability of Deep Learning algorithm, which should be supervised by the optical ruler of smoke, such as the Lambert–Beer law and the Mie scattering. Moreover, the real-world smoke driven by the buoyancy plume, which is the key factor of the smoky

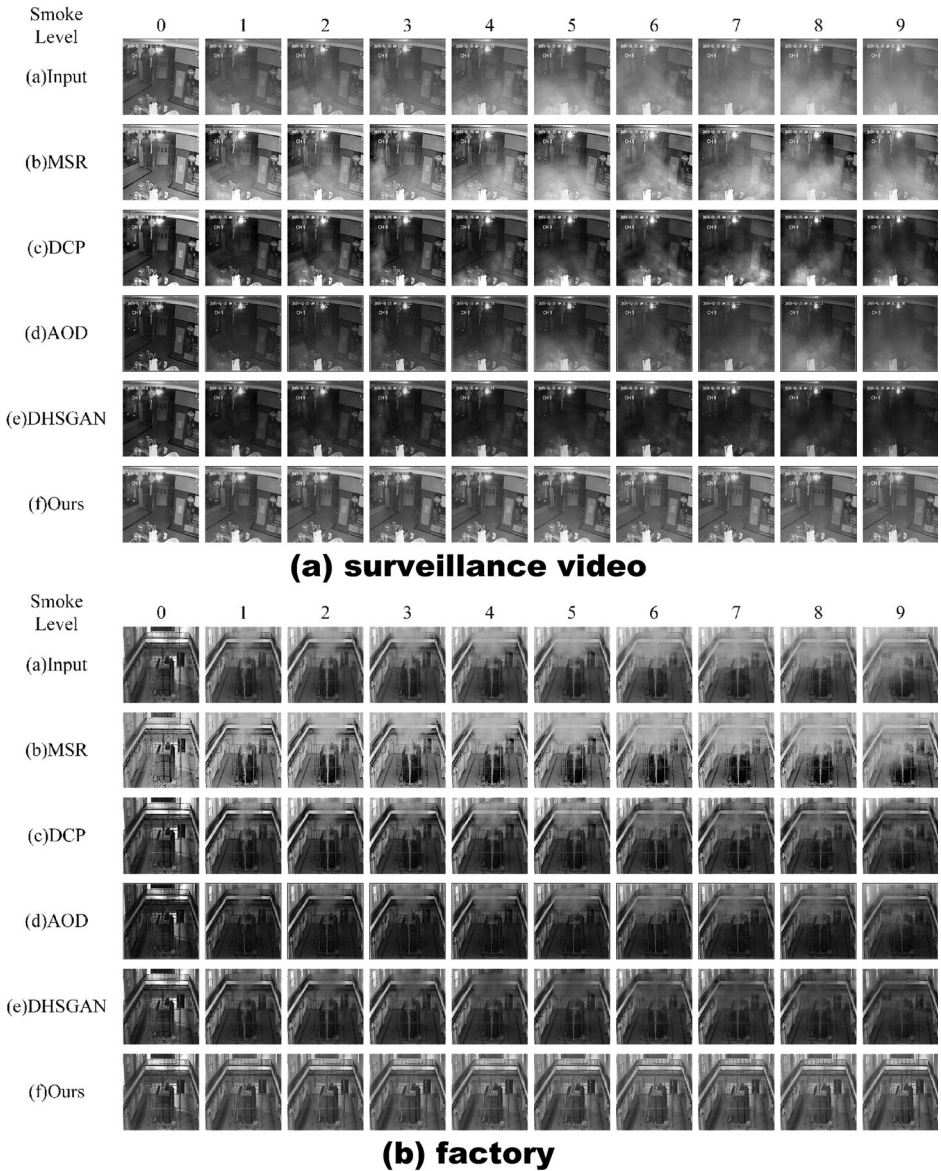


Fig. 10. Comparison of our smoke removal methods under 10 different smoke levels in two real smoke scenes (surveillance video and factory).

image synthetic process, should be also considered in the future researches. Once the database and algorithm model restrict are overcome, the research may achieve great improvements in the fire technology field with smart intelligence. Further development may also include gathering more realistic way to simulate the smoke,

modification of the code for better color restoration and application of real-time desmoking.

References

1. Hasikin K, Mat Isa NA (2015) Adaptive fuzzy intensity measure enhancement technique for non-uniform illumination and low-contrast images. *SIViP* 9(6):1419–1442
2. Rahman S et al (2016) An adaptive gamma correction for image enhancement. *Eurasip J Image Video Process* 1:35
3. Seow M-J, Asari VK (2006) Ratio rule and homomorphic filter for enhancement of digital colour image. *Neurocomputing* 69(7):954–958
4. Li S et al (2019) Real-time smoke removal for the surveillance images under fire scenario. *SIViP* 13(5):1037–1043
5. Fattal R, (2008) Single image dehazing. In international conference on computer graphics and interactive techniques
6. Baid A, et al., (2017) Joint desmoking, specular removal, and denoising of laparoscopy images via graphical models and Bayesian inference. In international symposium on biomedical imaging. Melbourne, VIC
7. Kotwal A, Bhalodia R and Awate SP (2016) Joint desmoking and denoising of laparoscopy images. In international symposium on biomedical imaging. Melbourne, VIC
8. He K, Sun J, and Tang X (2009) Single image haze removal using dark channel prior. In: *Computer vision and pattern recognition*, pp 1956–1963
9. Pei S, Chen Y, Chiu Y (2019) Single image desmoking using haze image model and human visual system. *J Electron Imaging* 28(04):043007
10. Meng G et al. (2013) Efficient image dehazing with boundary constraint and contextual regularization. In international conference on computer vision
11. Sulami M et al. (2014) Automatic recovery of the atmospheric light in hazy images. In *IEEE international conference on computational photography* Santa Clara, CA
12. Zhu Q, Mai J, Shao L (2015) A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans Image Process* 24(11):3522–3533
13. Li B et al. (2017) AOD-Net: all-in-one dehazing network. In international conference on computer vision
14. Chen D, et al. (2019) Gated context aggregation network for image dehazing and deraining. In *IEEE winter conference on applications of computer vision*
15. He L, Bai J, Ru L (2019) Haze removal using aggregated resolution convolution network. *IEEE Access* 7:123698–123709
16. Raj NB and Venkateswaran N (2018) Single image haze removal using a generative adversarial network *arXiv: computer vision and pattern recognition*
17. Li R et al. (2018) Single image dehazing via conditional generative adversarial network. In *computer vision and pattern recognition*
18. Zhang H, Sindagi VA and Patel VM (2017) Joint transmission map estimation and dehazing using deep networks. *arXiv: computer vision and pattern recognition*
19. Engin D, Genc A, and Ekenel HK (2018) Cycle-dehaze: enhanced cycleGAN for single image dehazing. In *computer vision and pattern recognition*
20. Yang X, Xu Z, and Luo J (2018) Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *national conference on artificial intelligence*
21. Tan M et al (2019) Image-dehazing method based on the fusion coding of contours and colors. *IEEE Access* 7:147857–147871

22. Tchaka K, Pawar V and Stoyanov D (2017) Chromaticity based smoke removal in endoscopic images. In Proceedings of SPIE. Orlando, Florida, US
23. Nishino K, Kratz L, Lombardi S (2012) Bayesian Defogging. *Int J Comput Vision* 98(3):263–278
24. Larochelle, H. and Hinton GE (2010) Learning to combine foveal glimpses with a third-order Boltzmann machine. In neural information processing systems
25. Wang F et al. (2017) Residual attention network for image classification. In computer vision and pattern recognition
26. Hu J et al (2020) Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* 42(8):2011–2023
27. Vaswani A et al. (2017) Attention is all you need. In neural information processing systems
28. Parmar N et al. (2018) Image transformer. arXiv: computer vision and pattern recognition
29. Wang X et al. (2018) Non-local neural networks. In computer vision and pattern recognition
30. Zhang Y et al. (2018) Image super-resolution using very deep residual channel attention networks. In European conference on computer vision
31. Qin X et al. (2019) FFA-Net: feature fusion attention network for single image dehazing. arXiv: computer vision and pattern recognition
32. Goodfellow I et al. (2014) Generative adversarial nets. In advances in neural information processing systems
33. Gulrajani I, et al. (2017) Improved training of wasserstein GANs. arXiv: Learning
34. Isola, P., et al., (2017) Image-to-Image translation with conditional adversarial networks. In computer vision and pattern recognition. Honolulu, HI
35. Ronneberger O, Fischer P, and Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In medical image computing and computer assisted intervention
36. Mirza M and Osindero S, (2014) Conditional generative adversarial nets. arXiv: learning
37. Bolkar S et al. (2018) Deep smoke removal from minimally invasive surgery videos. In proceedings of IEEE international conference on image processing
38. Malav R et al. (2018) DHSGAN: An end to end dehazing network for fog and smoke. In: Jawahar C, Li H, Mori G, Schindler K (Eds) Asian conference on computer vision
39. Sidorov O, Wang C, and Cheikh FA (2019) Generative smoke removal. arXiv: computer vision and pattern recognition
40. Chen L et al (2019) De-smokeGCN: generative cooperative networks for joint surgical smoke detection and removal. *IEEE Trans Med Imaging*. 39(5):1–1
41. Cai B et al (2016) DehazeNet: an end-to-end system for single image haze removal. *IEEE Trans Image Process* 25(11):5187–5198
42. Tang K, Yang J and Wang J, (2014) Investigating haze-relevant features in a learning framework for image dehazing. In computer vision and pattern recognition
43. He K et al. (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition
44. Xu, B., et al., (2015) Empirical evaluation of rectified activations in convolutional network. arXiv: learning
45. Ioffe S and Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv: learning
46. Arjovsky M, Chintala S and Bottou L (2017) Wasserstein gan. ArXiv e-prints

47. Mahapatra D et al. (2017) Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis. In international conference on medical image computing and computer-assisted intervention. Springer
48. Zeiler MD and Fergus R (2014) Visualizing and understanding convolutional networks. in European conference on computer vision. Springer
49. Nah S, Hyun Kim T and Mu Lee K (2017) Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition
50. Ledig C et al. (2017) Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition
51. Zhang H, Sindagi VA, and Patel VM (2017) Image de-raining using a conditional generative adversarial network. arXiv: computer vision and pattern recognition
52. Simonyan K and Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv: computer vision and pattern recognition
53. Silberman N et al. (2012) Indoor segmentation and support inference from RGBD images. In european conference on computer vision
54. Li B et al (2019) Benchmarking Single-Image Dehazing and Beyond. IEEE Trans Image Process 28(1):492–505
55. Abadi M et al. (2015) Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv: distributed, parallel, and cluster computing
56. Kingma DP and Ba J (2014) Adam: a method for stochastic optimization. arXiv: learning
57. Wang Z et al (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.