



High-resolution image de-raining using conditional GAN with sub-pixel upscaling

Prasen Kumar Sharma¹ · Sathisha Basavaraju¹ · Arijit Sur¹

Received: 8 April 2020 / Revised: 3 August 2020 / Accepted: 13 August 2020 /

Published online: 6 September 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

High-quality image de-raining is a challenging task that has been given considerable importance in recent times. To begin with, this problem is modeled as an image decomposition task where a rainy image is decomposed into the rain-free background and the associated rain streak map. Most of the existing methods have been successful in removing the rain streaks but fails to restore the image quality, which is degraded due to noise removal. This paper proposes a novel architecture called High-Resolution Image De-Raining using Conditional Generative Adversarial Networks (HRID-GAN) to generate a de-rained image with minimal artifacts and better visual quality. Extensive experiments on publicly available synthetic as well as real-world datasets show a substantial improvement over the state-of-the-art methods SPANet (Wang et al. 2019) by $\sim 2.43\%$ in PSNR and, DID-MDN (Zhang and Patel 2018) by $\sim 2.43\%$, $\sim 10.12\%$ and ID-CGAN (Zhang et al. 2017) by $\sim 11.80\%$, $\sim 34.70\%$ in SSIM and PSNR respectively.

Keywords Image restoration · Deep learning · Conditional GAN

1 Introduction

Rain-streaks, in the real-world scenario, may induce problems in many images and video-based applications such as UAV-based tracking, surveillance, autonomous car driving, etc. The rain streaks near camera lens appear to be raindrops and far from camera lens appears to be like haze veil, [39] thereby making it visually degraded. So far, the proposed solutions for the image de-raining, can be categorized based on (1) Supervised models with labelled-constrained, (2) Semi-supervision based learning paradigms, and (3) Unsupervised models with self-supervised constrained. Further, they span from traditional prior-based methods to deep learning-based models. However, a majority of the state-of-the-art (SOA) methods for image de-raining use supervised learning with labelled-constrained models, trained on synthetically generated datasets (Fig. 1).

✉ Prasen Kumar Sharma
kumar176101005@iitg.ac.in

¹ Indian Institute of Technology Guwahati, Guwahati, India

Supervised models with labelled-constrained require paired dataset when training, which is uncertain in the case of *real-world rainy images*. And, the manually generated rainy images may not cover a variety of rain information such as rain-streaks with various rain-density, direction, and scene illumination. Therefore, the supervised models, in general fail to remove the rain streaks from real-world samples. In last few years, a paradigm shift has been observed in the case of image de-raining from supervised to **Semi-supervision based learning** where the researchers have used both paired and unpaired training samples. The limitations of supervised learning can be reduced with the use of unpaired dataset (*especially by using real-world rainy images*). However, such models are highly resource exhaustive. Recent image de-raining methods are based on **Unsupervised learning with**

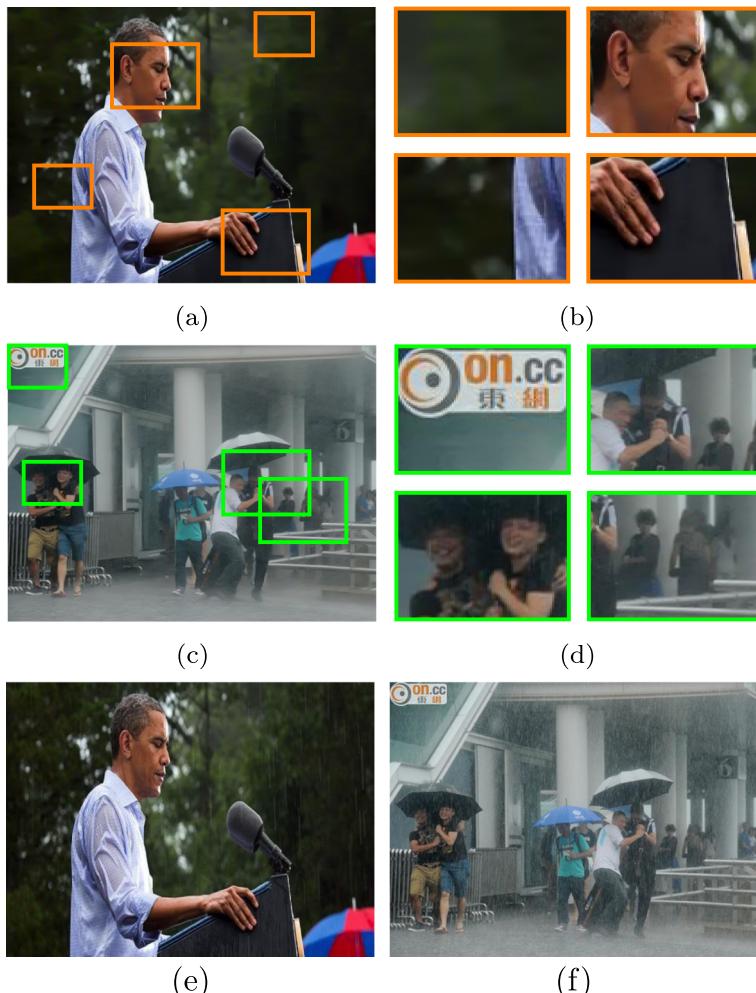


Fig. 1 Sample results of the proposed HRID-GAN method for single image de-raining on two real-world rainy images (e) & (f). (a) & (c) are de-rained images of (e) & (f) respectively

self-supervised constrained. The objective of this approach is to learn the intrinsic characteristics of the rain streaks and overcome the problems raised by using labelled or under constrained models.

Given a rainy image $\mathbf{R} \in [0, 255]^{M \times N}$ with rain streaks denoted by $\mathbf{M}_R \in [0, 255]^{M \times N}$ and clean background by $\mathbf{B} \in [0, 255]^{M \times N}$, single image de-raining becomes an image decomposition problem to recover the \mathbf{B} from \mathbf{R} by using a traditional equation given below as

$$\mathbf{B} = \mathbf{R} - \mathbf{M}_R \quad (1)$$

In this paper, we attempt to generate the de-rained images with better visual quality using Generative Adversarial Networks with efficient sub-pixel convolution [40] instead of conventional transpose convolution method.

2 Related work

In recent literature, there have been many methods which handled this image de-raining problem. For example, Fu et al. [5, 6] devised a network based on negative residual learning using Deep Residual Network (ResNet) [13] which can be added to rainy images to get their de-rained versions. Shen et al. [39] proposed a deep CNN that takes Haar [11] wavelets coefficients along with the Dark channel [12] prior as input to the proposed model for image de-raining. In [24, 25, 55], the proposed works considered the apriori image processing domain knowledge such as (a) centralized sparse representation, (b) predicted rain direction, (c) rain streak layer and (d) Gaussian mixture models, to estimate the rain-free background image from the corresponding rainy image. Zhang et al. [51] learns the sparsity and low-rank representation based convolutional kernels to determine the clear image and rain streaks. The work in [3, 14, 18, 33, 50] proposed a method where bilateral filtering is used to split the frequency parts (low and high) of the rainy image and the high-frequency part is further split into rainy textures and non-rain geometric details. It is achieved using structured dictionary learning, the histogram of oriented gradients (HOG), eigen colors and depth of field.

Chang et al. [2] extracts the periodic noise which follows line patterns such as rain streaks, stripes, fences etc. Chen et al. [4] introduced a model based on the Conditional Generative Adversarial Network [7] to estimate the high quality rain free image. Gu et al.

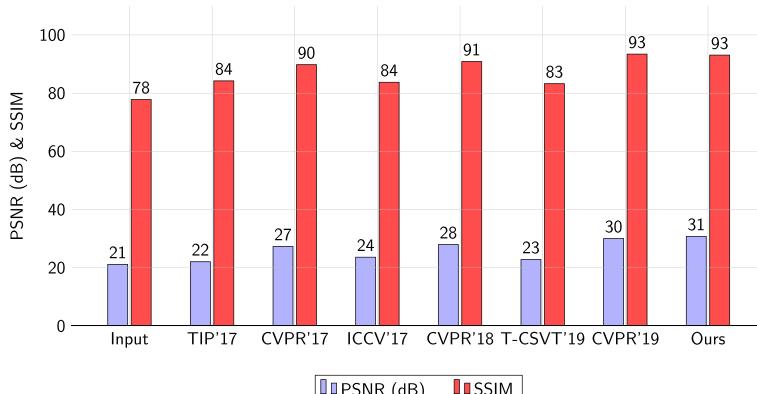


Fig. 2 Graphical comparison with the existing state-of-the-art schemes in terms of PSNR (dB) and SSIM

[9] decompose the rainy image into two layers, one which portrays large-scale structures is approximated by sparse analysis representation and the other by synthesis sparse representation which exhibits the finer textures in the image for the rain streaks removal. Yang et al. [48] proposed a joint rain-streak detection and removal method which uses a binary map. If a value, referencing a pixel in an image, in the binary map is ‘1’, it means that corresponding pixel has been affected by the rain-streaks. They simulate the heavy rain by modeling the appearance of rain streak accumulation, various shapes, and directions. Ren et al. [35] divided the rain streaks into sparse and dense using a matrix decomposition framework. Yeh et al. [49] proposed a method of decomposition which retrieves the high and low-frequency components of a rainy image by using the Gaussian filter. To remove the rain streaks from the high-frequency part, a Canny edge detection algorithm has been used. The rain streaks present in the low-frequency part are removed by using a non-negative matrix factorization technique. Wang et al. [44] observed that the high-frequency component of the rainy image consists of most of the rain streak part. The rain-free details from the high-frequency component are obtained using dictionary-based learning method. Zhang et al. [53] proposed a model that utilizes a conditional GAN framework [7] with baseline model built upon an encoder-decoder [41] framework and perceptual loss function [17]. Zhang et al. [52] proposed a method that first classify the rain-streak map into *light*, *medium* and *heavy* rain-streaks and then used a Densenet [15] based model to estimate the rain-streak map. Shen et al. [39] proposed a model which is based on the Haar Wavelets used to jointly remove the haze and rain streak from an image using a deep convolutional neural network. However, Sharma et al. [36] proposed a conditional GAN [7] based framework that utilizes both spatial and frequency domain cues to predict the rain-streak map from a rainy image.

2.1 Our contributions

It has been observed from the above mentioned existing works that a majority of them have been able to successfully de-rain a rainy image Fig. 2, but most of them suffer from the problem of over-coloring in the de-rained images. In other words, most of the existing methods fail to reconstruct the original perceptual quality of the clean image. This limitation motivates us to enhance the perceptual quality of the rain-free image and, therefore the contributions made by this paper can be summarised as follows:

1. U-Net [41] based architecture has been very successful in the case of image de-noising and reconstruction tasks [21, 36, 54], due to its ability to preserve important features for the reconstruction of images and discard the irrelevant and noisy components. Therefore, we propose an image de-raining model based on the U-Net framework.
2. We propose to use the efficient sub-pixel convolution [40] instead of conventional transpose convolution to avoid the checkerboard artifacts in the generated de-rained images.
3. To further improve the quality of the de-rained image generated by the encoder-decoder network, a deep residual network [13] has been used.
4. The cGAN based adversarial training has been incorporated in order to achieve better de-rained images.
5. An ablation study has been given at the end of the paper to demonstrate the effects of certain modules in the network with detailed comparisons.

The rest of this manuscript is organized as follows: Section 3 describes the proposed method for single image de-raining. Section 4 shows the experiments conducted and results

obtained, with qualitative and quantitative comparison between proposed and existing methods. The paper is concluded in Section 6.

2.2 Conditional GANs

Goodfellow et al. [7] designed a generative modelling framework called “Generative Adversarial Network” which comprises of two following sub-models : Generator (**G**) and Discriminator (**D**). For a given pair of rainy and clean images **R**, **B** respectively, the networks **G** and **D** play a zero-sum 2 player minimax game until Nash equilibrium [31] is achieved based on the (2)

$$\min_{\mathbf{G}} \max_{\mathbf{D}} \quad \mathbb{E}_{\mathbf{R} \sim p_{data}(\mathbf{R})} [\log(1 - \mathbf{D}(\mathbf{R}, \mathbf{G}(\mathbf{R})))] + \\ \mathbb{E}_{\mathbf{R} \sim p_{data}(\mathbf{R}, \mathbf{B})} [\log(\mathbf{D}(\mathbf{R}, \mathbf{B}))] \quad (2)$$

where **D** is trained in such a way that the probability of precisely classifying the generated de-rained images between real and fake ones can be maximized whereas **G** is trained to estimate more natural de-rained images similar to ground truth images such that **D** can be fooled.

Following [54], for this de-noising problem, the proposed generator model, conditioned on rainy image **R**, tries to generate the de-noised image similar to its clean counterpart so that it is difficult for discriminator to differentiate between real clean and de-noised (generated) clean image. Wang et al. [42] have used the conditional GAN to jointly detect and remove the shadows from single images. Qian et al. [34] have used attention map based CGAN to remove rain drops from single images. They have used residual and LSTM [8] blocks to predict the attention maps in the generator network. Kupyn et al. [22] have proposed a method for blind motion de-blurring from single images based on conditional GAN. Li et al. [23], and Sharma et al. [21] proposed CGAN based de-hazing method from single images.

2.3 Efficient sub-pixel convolution

During the process of downsampling of an image, the most prominent features remain in the compressed form, whereas the noise is discarded. To reconstruct the same image, from the downsampled features, the most common method used is transposed convolution, which is also popularly known as deconvolution. Even though bicubic interpolation is a special case of deconvolution [40], it has been observed that transpose convolution often induces checkerboard artifacts in the generated images [32].

The other way to upscale an image is to perform convolution operation with a fractional stride of $\frac{1}{r}$ based on the below equation.

$$\mathbf{I}^{HR} = f^L(\mathbf{I}^{LR}) = \mathcal{PS}(\mathbf{W}_L * f^{L-1}(\mathbf{I}^{LR}) + \mathbf{b}_L) \quad (3)$$

where \mathbf{I}^{LR} is the lower resolution (downsampled) image, \mathbf{I}^{HR} is the generated higher resolution image (upsampled) and \mathcal{PS} is the pixel-shuffle operation based on the following equation.

$$\mathcal{PS}(\mathbf{T})_{x,y,c} = \mathbf{T}_{\lfloor x/r \rfloor, \lfloor y/r \rfloor, c.r.mod(y,r)+c.mod(x,r)} \quad (4)$$

3 Proposed approach

3.1 Baseline generator model

The proposed baseline generator model consists of an encoder-decoder network where the decoder network comprises of conventional transpose convolution layers for up-sampling, followed by a deep residual network of 10 layers. The encoder consists of four encoding units (**E-Units**) as shown in Fig. 3, whereas the decoder consists of four conventional deconvolution¹ layers with symmetric skip connection to avoid the loss of image details when the network goes deeper [30] and helps deconvolution to recover a better clean image [38]. The encoder network can be summarised as follows:

- **E-Unit A** comprises of 2 convolution layers with kernel size 5×5 , number of kernels 64 each layer, spatial stride of 1×1 with activation function as ReLU at each layer. Batch Normalization [16] (BN) is applied only in second convolution layer.
- **E-Unit B** comprises of 2 convolution layers with kernel size 4×4 , number of filters 128 each layer, spatial stride of 2×2 in the first layer and 1×1 in second.
- **E-Unit C** comprises of 2 convolution layers with kernel size 3×3 , number of filters 256 each layer, spatial stride of 2×2 in the first layer and 1×1 in second.
- **E-Unit D** comprises of 2 convolution layers with kernel size 2×2 , number of filters 512 each layer, spatial stride of 2×2 in the first layer and 1×1 in second.

Each convolution layer of **E-Units (B:D)** consists of ReLU as activation function followed by Batch Normalization (BN) for faster convergence [16].

3.2 Generator with efficient sub-pixel convolution

The baseline generator model consists of conventional deconvolution operation which is quite computationally expensive. Several upsampling methods are available that have shown remarkable success in the recent times [27, 29]. Lu et al. [27] proposed Non-Convex JBU (NCJBU) by extending the well-known Joint Bilateral Upsampling (JBU) [20] with a novel non-convex optimization framework for guided depth-map upsampling. In the proposed model, the conventional transpose convolution operations are replaced with the efficient sub-pixel convolution blocks (**D-Units**) for upsampling, as shown in Fig. 4. The overall architecture of the generator model remains the same as the baseline generator as discussed in the previous Section 3.1.

3.3 Discriminator

Following Zhang et al. [53], we have employed the adversarial training for single image de-raining problem. While the objective of the proposed generator model is to estimate the de-rained image from the rainy image, the proposed discriminator model is trained to differentiate whether the estimated de-rained image is real or fake. In other words, the feedback from the proposed discriminator is used to train the proposed generator more efficiently. Following [36], an overview of the architecture of the proposed discriminator model, which consists of 5 convolution layers, is shown in Fig. 5. Each layer comprises of $\mathbf{k}/2$, \mathbf{k} , \mathbf{k}^2 , \mathbf{k}^3 and $\mathbf{k}/2$ convolution filters respectively where $\mathbf{k} = 2$. ReLU activation function has been

¹Transpose convolution is also referred as Deconvolution

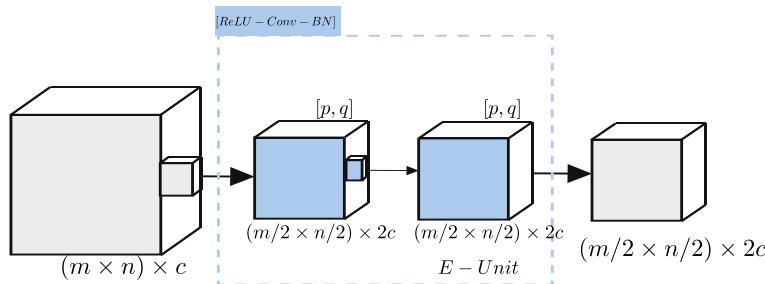


Fig. 3 An overview of single encoder unit (E-Unit) which consists of two convolution layers. An E-Unit outputs a tensor of spatial dimension downsampled (/2) with upsampled ($\times 2$) channel dimensions compared to the input given

used to induce the non-linearity in the model. A fully-connected neural network layer with 128 neurons, followed by a Sigmoid layer has been used to predict the logits.

3.4 Cost function

With the evolution of deep learning-based single image restoration methods, several regression, image reconstruction and object detection cost functions have been introduced, such as [26, 28]. To formally define the cost function used for the optimization of the proposed model, let ψ_G be the proposed non-linear generator model which outputs the de-rained image when a rainy image $x \in [0, 1]^{h \times w \times c}$ is given as input with w, h, c as width, height and channels of rainy image respectively. To retain the structural information, the mean-squared error (MSE) is mostly used error function in denoising algorithms and can be defined as

$$\mathcal{L}_{MSE} = \frac{1}{w.h.c} \sum_{i=1, j=1, k=1}^{w, h, c} ||\psi_G(x)^{i, j, k} - y^{i, j, k}||_2^2 \quad (5)$$

The adversarial loss to train the proposed generator model is defined as

$$\mathcal{L}_{Adv} = -\frac{1}{N} \sum_{i=1}^N \log D(x_i) \quad (6)$$

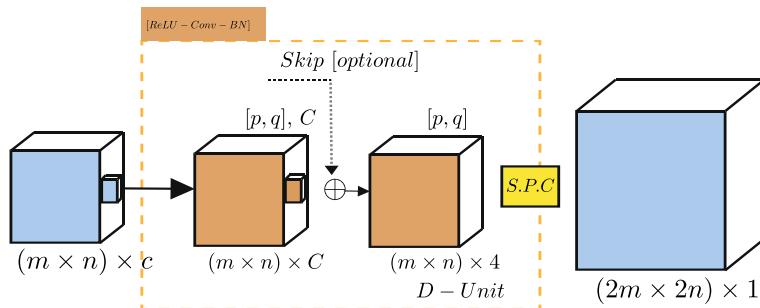


Fig. 4 An overview of single decoder unit (D-Unit) which consists of two convolution layers followed by an efficient sub-pixel re-arrangement(S.P.C) block. A D-Unit outputs a tensor of spatial dimension upsampled ($\times 2$) compared to the input given

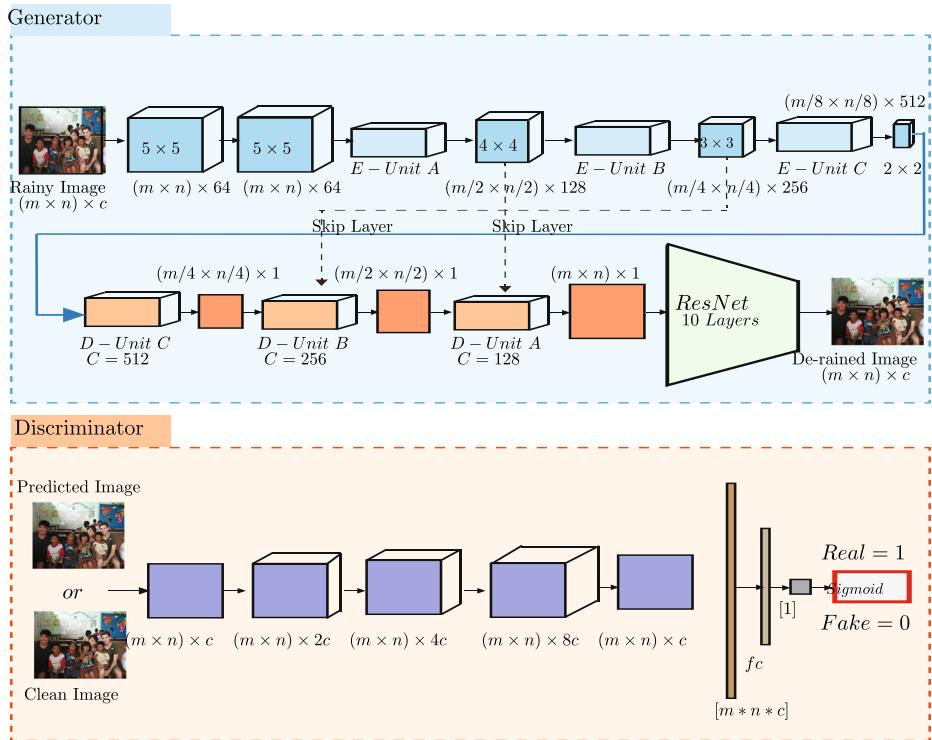


Fig. 5 An overview of the architecture of the proposed framework for rain streak removal from the single image

where N is the number of generated de-rained images from generator. Therefore, the total generator loss is the linear combination of mean-squared error and entropy losses and can be written as

$$\mathcal{L}_G = \lambda_M \cdot \mathcal{L}_{mse} + \lambda_A \cdot \mathcal{L}_{Adv} \quad (7)$$

where λ_M and λ_A are pre-defined weights for the cost functions defined above. The objective of the proposed method is to generate the de-rained image given a rainy image as an input. The proposed generator tries to generate the de-rained image such that it is difficult for the discriminator to decide whether generated de-rained image is real clean image or fake. The proposed network tries to minimize the generator cost function \mathcal{L}_G .

4 Experiments and results

The synthetic datasets of rainy and clean images given by the authors of [52] have been used for training and testing. Zhang et al.. [52] have included the medium rain density images in addition to images with light and heavy rain densities. We have augmented the rainy and clean images from our selected training dataset into the disjoint patches of size 128×128 , thus creating a total of 1,92,000 patches. Our selected test dataset consists of 1201 rainy images with ground truth. The proposed method, along with baseline configurations, have also been tested on real-world rainy images, as shown in Fig. 6. However, the ground truth

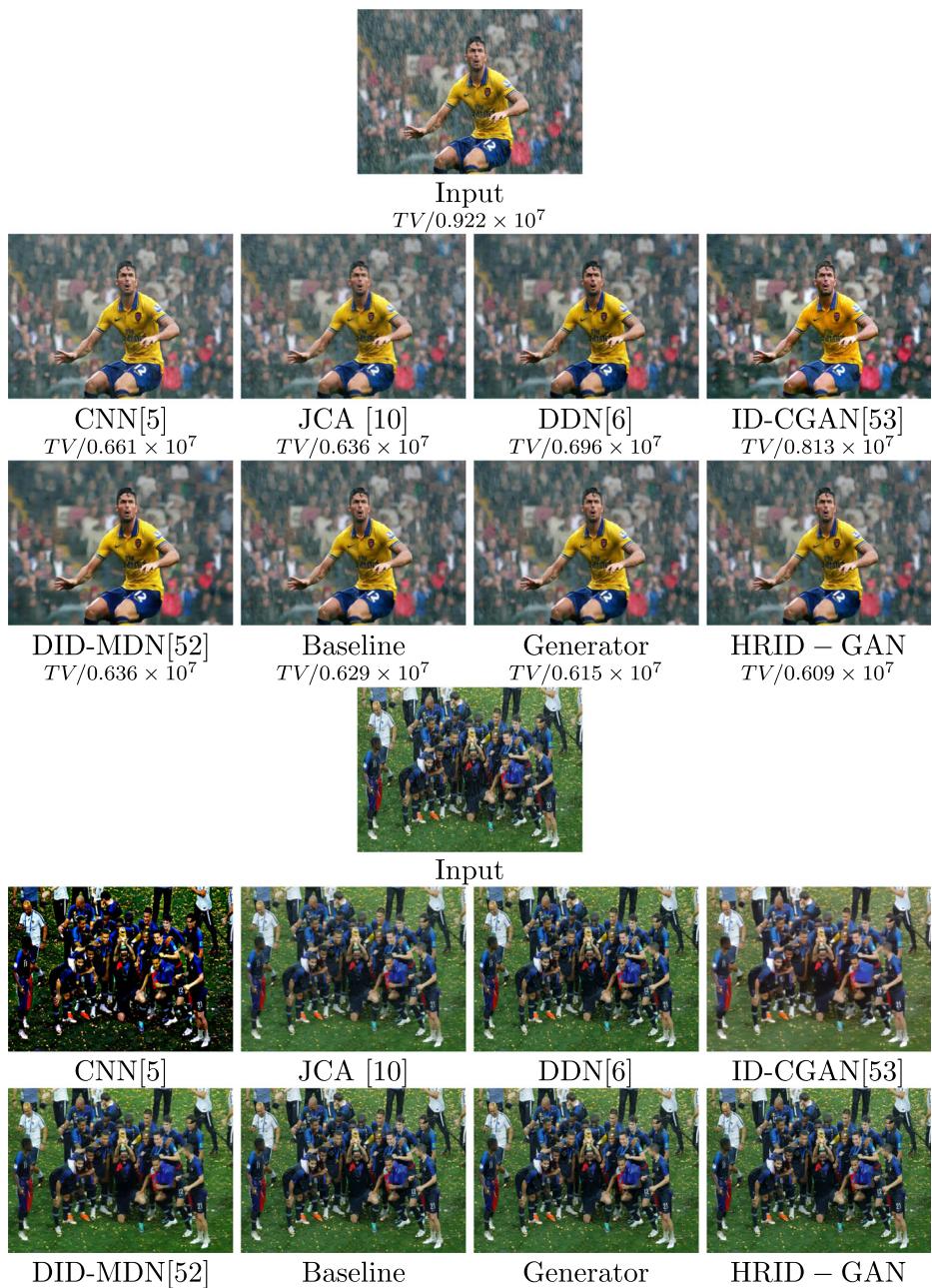


Fig. 6 Sample results on real-world rainy images in terms of Total Variation (TV)

of such images does not exist in order to compare with. To give a quantitative comparison of such images, we have used the following evaluation metrics.

4.1 Quality measures

The proposed scheme along with the existing state-of-the-art methods have been evaluated on the following image quality evaluation metrics: Universal Quality Index (UQI) [45], Structural Similarity Index (SSIM) [46] for measuring the similarity between de-rained and ground truth images, Peak Signal-to-Noise Ratio (PSNR), Visual Information Fidelity (VIF) [37], Multi-scale Structural Similarity (MS-SSIM) [47], Mean-squared Error (MSE) and Total Variation (TV) to calculate amount of noise present in the image after de-noising. *Color images are given as input to measure all quality metrics for every compared paper to be fair instead of luminance only as done in [10, 53].*

4.2 Model parameters

In this sub-section, the dataset and parameters which are being used during training and testing of the proposed framework are discussed briefly. The publicly available synthetic dataset² [52] is used for training and testing where training set consists of 12000 images of size 512×512 . For augmentation, we have cropped the training images into the disjoint patches of shape 128×128 , resulting in a total of 192,000 patches in the training set. For evaluation, the synthetic test set comprises of 1.2K images of shape 512×512 . To test the generality of the proposed scheme, we have also evaluated the proposed model on the real-world rainy images that do not have the ground truth clean images. The proposed model is built upon Tensorflow [1] framework and is trained on Nvidia-GTX 1080 GPU for 50 epochs. The learning rate (lr) is initially set to 0.01 and reduced by $\times 0.1$ after every 15–20 epochs. The batch size of 20 and Adam [19] optimization algorithm have been used when training. The weights are set as follows: $\lambda_E = 1$ and $\lambda_{adv} = 0.01$.

4.3 Comparison configurations

The proposed method has been compared with the following baseline configurations :

1. **Baseline** : We propose a baseline model described in Section 3.1.
2. **Generator** : The proposed generator with efficient sub-pixel convolution model as described in Section 3.2 with λ_{adv} set to zero.

The proposed method is compared with the following existing methods :

1. **DDN** : A deep residual networks based architecture that requires prior image processing domain knowledge [6].
2. **DID-MDN** : A rain streak density-aware method based on densely connected CNN and a classifier [52] is the existing state-of-the-art method for single image de-raining problem.
3. **ID-CGAN** : A conditional GAN based framework which has an underlying model of encoder-decoder network [41] with perceptual loss function for training.
4. **JCA** : A layer separation method based on convolution analysis jointly with synthesis sparse representation [10] used for rain streak removal from single images.

²<https://github.com/hezhangsprinter/DIDMDN>

Table 1 Quantitative comparison with existing methods on test dataset in terms of SSIM and PSNR. Best and second best results are highlighted in bold and underlined fonts, respectively

-	Methods	SSIM	PSNR
TIP'17	Input	0.7781	21.15
CVPR'17	CNN [5]	0.8422	22.07
ICCV'17	DDN [6]	0.8978	27.33
T-CSVT'19	JCA [10]	0.8374	23.63
CVPR'18	ID-CGAN [53]	0.8325	22.85
CVPR'19	DID-MDN [52]	0.9087	27.95
Ours	Baseline	0.9269	30.07
Ours	Generator(G)	0.9297	<u>30.62</u>
Ours	Proposed	<u>0.9308</u>	30.78

5. **CNN** : Clearing the skies : A deep network architecture for single image rain removal [5] is based on the deep residual network similar to [6].

4.4 Quantitative results

Quantitative comparison results with state-of-the-art methods [5, 6, 10, 52, 53] and baseline configurations are given in Tables 1 and 2. The proposed baseline generator model with conventional transpose convolution for up-scaling operation has an impressive improvement over recent state-of-the-art methods. The generator model (G) where transpose convolution operations have been replaced with efficient sub-pixel convolution operations, has a further improvement over the baseline model. The generator model with efficient sub-pixel convolutions (G) as shown in Table 2 is only trained on mean-squared error as a loss function. The inclusion of discriminator and adversarial loss, denoted as (G+D), along with mean-squared error on the proposed model (G) has further improvement overall recent methods as well

Table 2 Quantitative comparison with existing methods on test dataset. Best and second best results are highlighted in Bold and Underlined fonts, respectively

-	Methods	SSIM	PSNR	VIF	MS-SSIM	TV-Error	UQI	MSE
TIP'17	Input	0.7781	21.15	0.3734	0.7334	1.55	0.8636	0.766
CVPR'17	CNN [5]	0.8422	22.07	0.4082	0.8384	1.25	0.8650	0.708
ICCV'17	DDN [6]	0.8978	27.33	0.4246	0.8650	1.14	0.9526	0.124
T-CSVT'19	JCA [10]	0.8374	23.63	0.3867	0.8145	1.05	0.8865	0.520
CVPR'18	ID-CGAN [53]	0.8325	22.85	0.5177	0.9007	1.19	0.8922	0.513
Ours	Baseline	0.9269	30.07	0.4741	0.9058	0.94	<u>0.9677</u>	0.080
Ours	Generator(G)	<u>0.9297</u>	<u>30.62</u>	0.4833	<u>0.9075</u>	1.01	0.9674	<u>0.072</u>
Ours	Proposed	0.9308	30.78	<u>0.4851</u>	0.9090	<u>1.01</u>	0.9682	<u>0.070</u>

† TV-E is $\times 10^7$. ‡ MSE is $\times 10^{-3}$

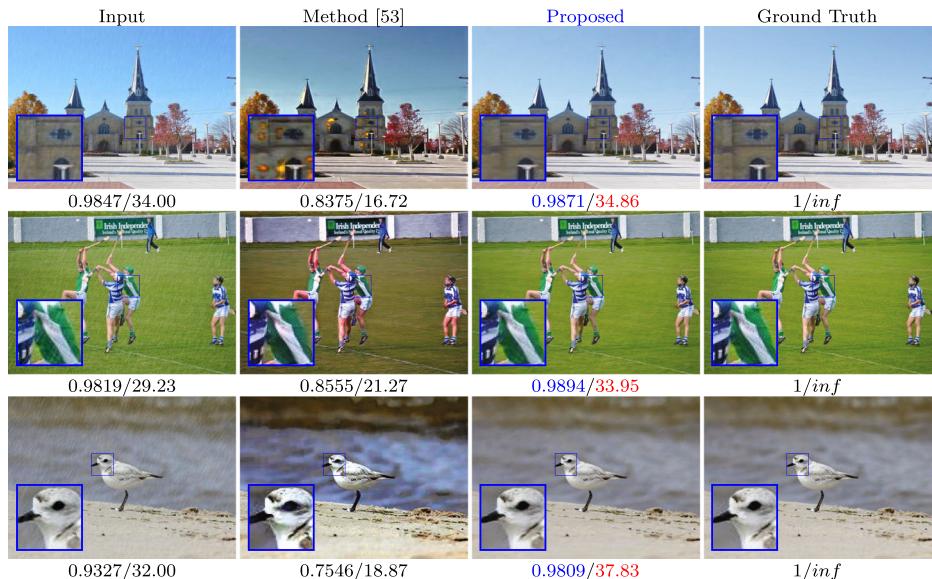


Fig. 7 Qualitative comparison with ID-CGAN [53] on synthesized test images in terms of SSIM/PSNR

as baseline configurations. The visual comparison on synthetic as well as real-world rainy images are given as follows.

4.5 Qualitative results

It can be observed from Figs. 7, 6, 9 and 10 that the visual quality of the de-rained images estimated by using the proposed model is far better than the existing state-of-the-art methods. It is observed that the de-rained images generated by the existing methods [6], [10] and [5] still contains the rain-streaks. The methods [52] and [53] estimates the rain-streak free images in RGB domain. It can be seen that such methods due to high correlation among color channels suffer from over-coloring (color saturation) and white round artifacts (maybe due to the use of perceptual loss in [53]) respectively. However, the de-rained images generated by the proposed scheme do not suffer from these artifacts and degradations. The visual quality of the de-rained images may have been improved by incorporating the adversarial training in the proposed model, unlike in [5] and [6]. Following [53], learning the difference between real and fake de-rained images via adversarial training has been beneficial for single image rain-streak removal. However, the proposed model does not include the perceptual loss function [17] due to the possibility of white round artifacts. The proposed model has made a relaxation on highly correlated color spaces, i.e., unlike [52] and [53], instead of estimating the de-rained images in colors, such as RGB or YCbCr, it estimates only the Y channel of the de-rained images. The relaxation of chrominance components Cb, Cr has been beneficial, and an improvement of $\sim 2\%$ in SSIM and $\sim 10\%$ in PSNR has been recorded over existing scheme [52]. The proposed method has also been evaluated on real-world rainy images as shown in Figs. 6, 8, 11 and based on the total variation error (TV), it can be observed that the visual quality of the de-rained images generated by the proposed model is better than the existing methods (Figs. 8, 9, 10 and 11).



Fig. 8 Sample results on real-world rainy images

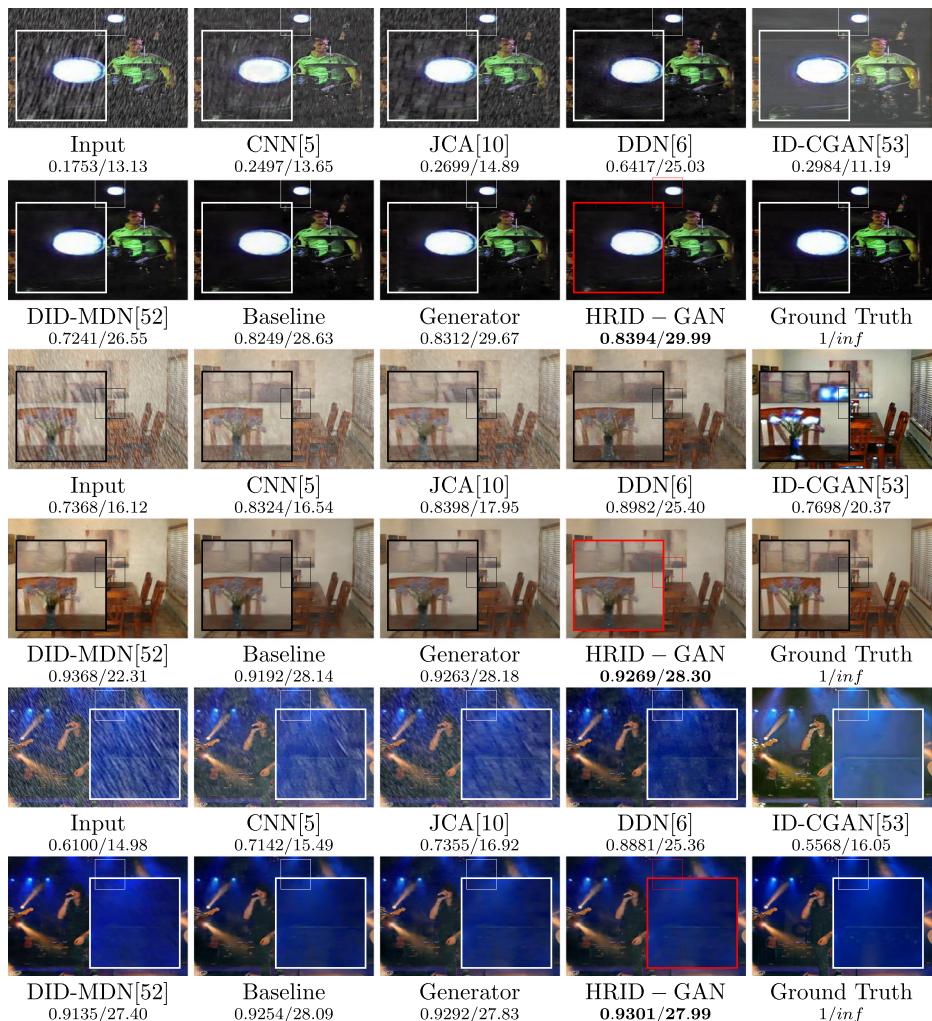


Fig. 9 Sample results on synthetic rainy images compared with the existing schemes in terms of SSIM/PSNR

5 Discussion

In this work, a cGAN based image de-raining scheme has been proposed, and it has been experimentally justified that proposed scheme outperformed the state-of-the-art schemes. Although proposed work has employed cGAN architecture, it is significantly different from cGAN based scheme [53], which has been proposed in the recent literature. The major differences are pointed out as follows:

1. The model used in [53] is trained in RGB color space, whereas the proposed model in this paper has utilized the decorrelated YCbCr color space.
2. The model used in [53] has incorporated the perceptual loss [17] unlike the proposed model in this paper.

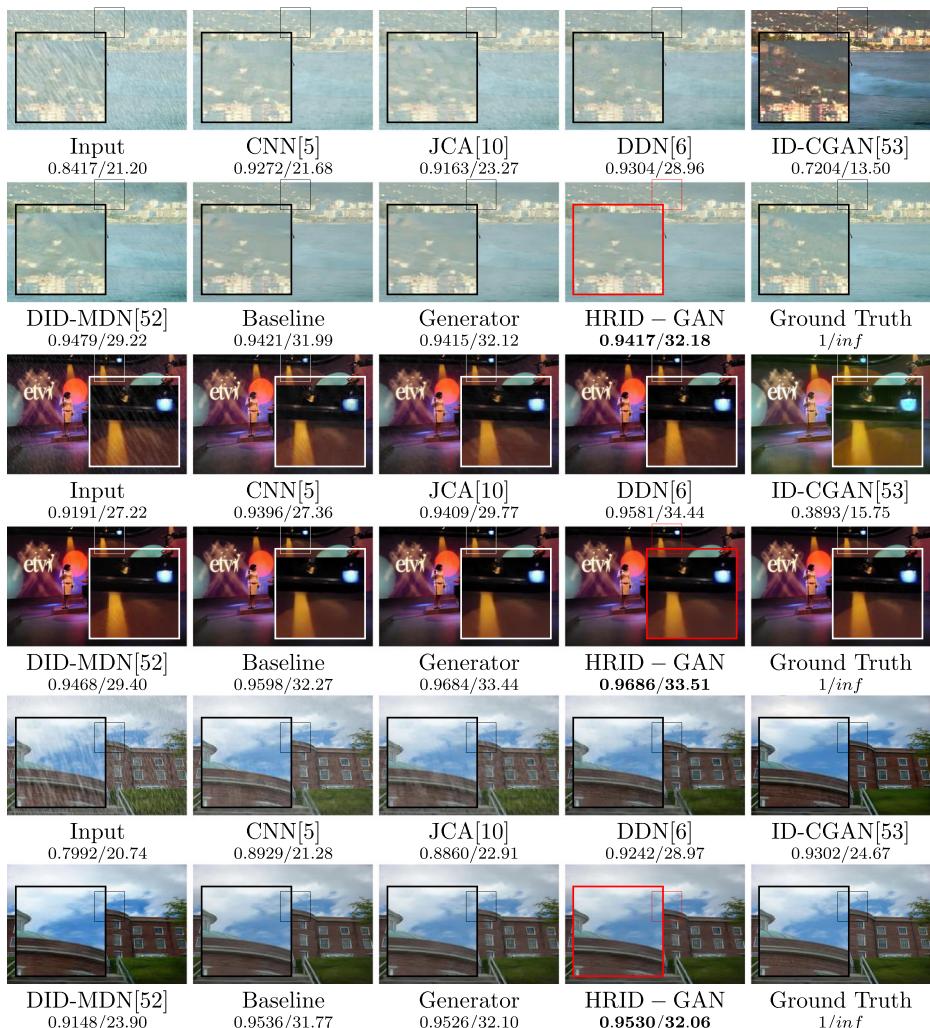


Fig. 10 Sample results on synthetic rainy images compared with the existing schemes in terms of SSIM/PSNR

3. The proposed model in this paper has incorporated a 10 layer ResNet [13] unlike in [53].

However, the improvement reported by the proposed baseline model in this paper compared to [53] is substantially huge in terms of both SSIM and PSNR, as shown in Table 2. This is mainly due to the fact that inclusion of perceptual loss in [53] have added the white round artifacts in the generated de-rained images when operating in RGB color space (as reported in [53]) whereas this paper has retained the color information in terms of chrominance channels and has not utilized the benefits of perceptual loss on generated de-rained (grayscale) images. In addition to this, estimating the RGB values of the pixels in the predicted de-rained images as done in [52, 53] may be difficult for a network when compared to only predicting the grayscale values of the de-rained images. The proposed model has



Fig. 11 Sample results on real-world rainy images

shown substantial improvement over the work reported in [6] which has utilized the ResNet [13] architecture and estimates the negative residual of the rain-streaks using the result of high-pass filter over the given input as a prior to the network. However, in this work, no such prior knowledge is required. Moreover, in addition to the residual network, this paper also incorporates the encoder-decoder framework along with cGAN, which makes it more effective over [6].

While other methods have been successful in removing the rain-streaks up to some extent, most of the existing methods failed to consider the spatial resolution of the image. It is observed that after removing the rain-streaks, the obtained results suffer from the problems of over-smoothness and blurriness. To avoid these issues, unlike existing methods, this paper has incorporated a sub-pixel convolution [40] to enhance the spatial resolution of the image. The use of efficient sub-pixel convolution instead of traditional deconvolution

for up-sampling of the features in the proposed generator model (G) compared to proposed baseline model initially has found to be beneficial, and an improvement of ≈ 0.5 dB in PSNR has been observed as shown in Table 2. This can be because efficient sub-pixel convolution might have improved the spatial resolution of the de-rained images by the proposed architecture which in turn resulted in better PSNR values and a slight improvement in SSIM and other evaluation metrics. The adversarial loss, in addition to the MSE on the proposed generator, has been useful in further improving the results.

6 Conclusion

In this paper, a conditional GAN based framework is proposed for a single image rain-streak removal task. The proposed scheme employed the computational and visual efficiency of efficient sub-pixel convolution over conventional transpose convolution for up-scaling and produced results are better than existing state-of-the-art methods. It is shown how conventional deconvolution can be replaced by sub-pixel convolution for image de-noising tasks, and noticeable improvement can be achieved if adversarial training is used in addition to traditional loss functions. The spatial resolution of the de-rained image has been achieved by the efficient sub-pixel convolution for upscaling as generated results have better resolution than the results obtained by the existing methods. Unlike existing state-of-the-art methods, the de-rained images generated by the proposed scheme do not contain white rounds artifacts and blurriness, which ensures applicability of the proposed method on single image de-raining tasks.

Acknowledgments Authors would like to thank the anonymous reviewers for their insightful comments and suggestions. Authors would also like to acknowledge the funding agency, Ministry of Human Resource Development, Government of India.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Abadi M, Agarwal A (2015) TensorFlow: Large-scale machine learning on heterogeneous systems. <http://tensorflow.org/>. Software available from tensorflow.org
2. Chang Y, Yan L, Zhong S (2017) Transformed low-rank model for line pattern noise removal. In: 2017 IEEE International conference on computer vision (ICCV), pp 1735–1743. <https://doi.org/10.1109/ICCV.2017.191>
3. Chen DY, Chen CC, Kang LW (2014) Visual depth guided color image rain streaks removal using sparse coding. IEEE Trans Circ Syst Vid Technol 24(8):1430–1455. <https://doi.org/10.1109/TCSVT.2014.2308627>
4. Chen Q, Yi X, Ni B, Shen Z, Yang X (2017) Rain removal via residual generation cascading. In: 2017 IEEE Visual communications and image processing (VCIP), pp 1–4. <https://doi.org/10.1109/VCIP.2017.8305092>
5. Fu X, Huang J, Ding X, Liao Y, Paisley J (2017) Clearing the skies: a deep network architecture for single-image rain removal. IEEE Trans Image Process 26(6):2944–2956. <https://doi.org/10.1109/TIP.2017.2691802>
6. Fu X, Huang J, Zeng D, Huang Y, Ding X, Paisley J (2017) Removing rain from single images via a deep detail network. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1715–1723. <https://doi.org/10.1109/CVPR.2017.186>

7. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds) Advances in Neural Information Processing Systems 27, Curran Associates, Inc., pp 2672–2680. <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>
8. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2017) Lstm: a search space odyssey. IEEE Trans Neural Netw Learn Syst 28(10):2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
9. Gu S, Meng D, Zuo W, Zhang L (2017) Joint convolutional analysis and synthesis sparse representation for single image layer separation. In: 2017 IEEE International conference on computer vision (ICCV), pp 1717–1725. <https://doi.org/10.1109/ICCV.2017.189>
10. Gu S, Meng D, Zuo W, Zhang L (2017) Joint convolutional analysis and synthesis sparse representation for single image layer separation. In: 2017 IEEE International conference on computer vision (ICCV), pp 1717–1725. <https://doi.org/10.1109/ICCV.2017.189>
11. Haar A (1910) Zur theorie der orthogonalen funktionensysteme. Math Ann 69(3):331–371. <https://doi.org/10.1007/BF01456326>
12. He K, Sun J, Tang X (2011) Single image haze removal using dark channel prior. IEEE Trans Patt Anal Mach Intel 33(12):2341–2353. <https://doi.org/10.1109/TPAMI.2010.168>
13. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: The IEEE conference on computer vision and pattern recognition (CVPR)
14. Huang DA, Kang LW, Wang YCF, Lin CW (2014) Self-learning based image decomposition with applications to single image denoising. IEEE Trans Multimedia 16(1):83–93. <https://doi.org/10.1109/TMM.2013.2284759>
15. Huang G, Liu Z, Weinberger KQ (2016) Densely connected convolutional networks. arXiv:[abs/1608.06993](https://arxiv.org/abs/1608.06993)
16. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Bach F, Blei D (eds) Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research, vol 37. PMLR, Lille, pp 448–456. <http://proceedings.mlr.press/v37/ioffe15.html>
17. Johnson J, Alahi A, Li F (2016) Perceptual losses for real-time style transfer and super-resolution. arXiv:[abs/1603.08155](https://arxiv.org/abs/1603.08155)
18. Kang LW, Lin CW, Fu YH (2012) Automatic single-image-based rain streaks removal via image decomposition. IEEE Trans Image Process 21(4):1742–1755. <https://doi.org/10.1109/TIP.2011.2179057>
19. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: International conference on learning representations (ICLR)
20. Kopf J, Cohen MF, Lischinski D, Uyttendaele M (2007) Joint bilateral upsampling. ACM Trans Graph 26(3):96–es. <https://doi.org/10.1145/1276377.1276497>
21. Kumar Sharma P, Jain P, Sur A (2020) Scale-aware conditional generative adversarial network for image dehazing. In: 2020 IEEE Winter conference on applications of computer vision (WACV), pp 2344–2354
22. Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J (2018) Deblurgan: Blind motion deblurring using conditional adversarial networks. In: The IEEE conference on computer vision and pattern recognition (CVPR)
23. Li R, Pan J, Li Z, Tang J (2018) Single image dehazing via conditional generative adversarial network. In: The IEEE conference on computer vision and pattern recognition (CVPR)
24. Li Y, Tan RT, Guo X, Lu J, Brown MS (2016) Rain streak removal using layer priors. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR), pp 2736–2744. <https://doi.org/10.1109/CVPR.2016.299>
25. Li Y, Tan RT, Guo X, Lu J, Brown MS (2017) Single image rain streak decomposition using layer priors. IEEE Trans Image Process 26(8):3874–3885. <https://doi.org/10.1109/TIP.2017.2708841>
26. Lin T, Goyal P, Girshick R, He K, Dollár P (2017) Focal loss for dense object detection. In: 2017 IEEE International conference on computer vision (ICCV), pp 2999–3007
27. Lu X, Guo Y, Liu N, Wan L, Fang T (2018) Non-convex joint bilateral guided depth upsampling. Multimed Tools Appl 77(12):15521–15544. <https://doi.org/10.1007/s11042-017-5131-x>
28. Lu X, Ma C, Ni B, Yang X, Reid I, Yang MH (2018) Deep regression tracking with shrinkage loss. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (eds) Computer vision – ECCV 2018. Springer International Publishing, Cham, pp 369–386
29. Lu X, Wang W, Ma C, Shen J, Shao L, Porikli F (2019) See more, know more: Unsupervised video object segmentation with co-attention siamese networks. In: The IEEE conference on computer vision and pattern recognition (CVPR)
30. Mao X, Shen C, Yang YB (2016) Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I,

- Garnett R (eds) Advances in Neural Information Processing Systems 29, pp 2802–2810. Curran Associates, Inc. <http://papers.nips.cc/paper/6172-image-restoration-using-very-deep-convolutional-encoder-decoder-networks-with-symmetric-skip-connections.pdf>
- 31. Nash JF (1950) Equilibrium points in n-person games. Proceedings of the National Academy of Sciences 36(1):48–49. <https://doi.org/10.1073/pnas.36.1.48>
 - 32. Odena A, Dumoulin V, Olah C (2016) Deconvolution and checkerboard artifacts. Distill. <http://distill.pub/2016/deconv-checkerboard/>
 - 33. Park K, Yu S, Jeong J (2018) A contrast restoration method for effective single image rain removal algorithm. In: 2018 International workshop on advanced image technology (IWAIT), pp 1–4. <https://doi.org/10.1109/IWAIT.2018.8369644>
 - 34. Qian R, Tan RT, Yang W, Su J, Liu J (2018) Attentive generative adversarial network for raindrop removal from a single image. In: The IEEE conference on computer vision and pattern recognition (CVPR)
 - 35. Ren W, Tian J, Han Z, Chan A, Tang Y (2017) Video desnowing and deraining based on matrix decomposition. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 2838–2847. <https://doi.org/10.1109/CVPR.2017.303>
 - 36. Sharma PK, Jain P, Sur A (2019) Dual-domain single image de-raining using conditional generative adversarial network. In: 2019 IEEE International conference on image processing (ICIP), pp 2796–2800. <https://doi.org/10.1109/ICIP.2019.8803353>
 - 37. Sheikh HR, Bovik AC (2006) Image information and visual quality. IEEE Trans Image Process 15(2):430–444. <https://doi.org/10.1109/TIP.2005.859378>
 - 38. Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Patt Anal Mach Intel 39(4):640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
 - 39. Shen L, Yue Z, Chen Q, Feng F, Ma J (2018) Deep joint rain and haze removal from single images. arXiv:[abs/1801.06769](https://arxiv.org/abs/1801.06769)
 - 40. Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z (2016) Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: The IEEE conference on computer vision and pattern recognition (CVPR)
 - 41. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds) Advances in Neural Information Processing Systems 27, Curran Associates, Inc., pp 3104–3112. <http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf>
 - 42. Wang J, Li X, Yang J (2018) Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: The IEEE conference on computer vision and pattern recognition (CVPR)
 - 43. Wang T, Yang X, Xu K, Chen S, Zhang Q, Lau RW (2019) Spatial attentive single-image deraining with a high quality real rain dataset. In: The IEEE conference on computer vision and pattern recognition (CVPR)
 - 44. Wang Y, Chen C, Zhu S, Zeng B (2016) A framework of single-image deraining method based on analysis of rain characteristics. In: 2016 IEEE International conference on image processing (ICIP), pp 4087–4091
 - 45. Wang Z, Bovik AC (2002) A universal image quality index. IEEE Signal Process Lett 9(3):81–84. <https://doi.org/10.1109/97.995823>
 - 46. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612. <https://doi.org/10.1109/TIP.2003.819861>
 - 47. Wang Z, Simoncelli EP, Bovik AC (2003) Multiscale structural similarity for image quality assessment. In: The thirty-seventh asilomar conference on signals, systems computers, 2003, vol 2, pp 1398–1402. <https://doi.org/10.1109/ACSSC.2003.1292216>
 - 48. Yang W, Tan RT, Feng J, Liu J, Guo Z, Yan S (2017) Deep joint rain detection and removal from a single image. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1685–1694. <https://doi.org/10.1109/CVPR.2017.183>
 - 49. Yeh CH, Liu PH, Yu CE, Lin CY (2015) Single image rain removal based on part-based model. In: 2015 IEEE International conference on consumer electronics - taiwan, pp 462–463. <https://doi.org/10.1109/ICCE-TW.2015.7216999>
 - 50. Yu S, Ou W, You X, Mou Y, Jiang X, Tang Y (2015) Single image rain streaks removal based on self-learning and structured sparse representation. In: 2015 IEEE China Summit and international conference on signal and information processing (chinaSIP), pp 215–219. <https://doi.org/10.1109/ChinaSIP.2015.7230394>

51. Zhang H, Patel VM (2017) Convolutional sparse and low-rank coding-based rain streak removal. In: 2017 IEEE Winter conference on applications of computer vision (WACV), pp 1259–1267. <https://doi.org/10.1109/WACV.2017.145>
52. Zhang H, Patel VM (2018) Density-aware single image de-raining using a multi-stream dense network. In: The IEEE conference on computer vision and pattern recognition (CVPR)
53. Zhang H, Sindagi V, Patel VM (2017) Image de-raining using a conditional generative adversarial network. arXiv:[abs/1701.05957](https://arxiv.org/abs/1701.05957)
54. Zhang H, Sindagi V, Patel VM (2019) Image de-raining using a conditional generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology, pp 1–1
55. Zhu L, Fu CW, Lischinski D, Heng PA (2017) Joint bi-layer optimization for single-image rain streak removal. In: 2017 IEEE International conference on computer vision (ICCV), pp 2545–2553. <https://doi.org/10.1109/ICCV.2017.276>

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