IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 October 2022, accepted 8 November 2022, date of publication 17 November 2022, date of current version 28 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3223101

RESEARCH ARTICLE

Unsupervised Denoising for Super-Resolution (UDSR) of Real-World Images

KALPESH PRAJAPATI[1], VISHAL CHUDASAMA[1], (Member, IEEE), HEENA PATEL[1], (Member, IEEE), ANJALI SARVAIYA[1], (Graduate Student Member, IEEE), KISHOR UPLA[1], (Member, IEEE), KIRAN RAJA[2], (Senior Member, IEEE), RAGHAVENDRA RAMACHANDRA[3], (Senior Member, IEEE), AND CHRISTOPH BUSCH[3], (Senior Member, IEEE)

[1]Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat 395007, India
[2]Department of Computer Science, Norwegian University of Science and Technology (NTNU), 2802 Gjøvik, Norway
[3]Department of Information Security and Communication Technology, Norwegian University of Science and Technology (NTNU), 2802 Gjøvik, Norway

Corresponding author: Kishor Upla (kishor_upla@gmail.com)

This work was supported in part by the Science and Engineering Research Board (SERB), a statutory body of the Department of Science and Technology (DST), Government of India, under Grant ECR/2017/003268, and in part by the Norwegian Biometric Laboratory (NBL), NTNU, Gjøvik, Norway.

ABSTRACT Single Image Super-Resolution (SISR) using Convolutional Neural Networks (CNNs) for many applications in *supervised* manner has resulted in significant improvement in state-of-the-art performance. Such supervised models achieve remarkable accuracy; albeit their poor generalization ability for real-world Low-Resolution (LR) images. Supervised training in many SR works involves synthetically generated LR images from its corresponding High-Resolution (HR) images. As the distribution of such LR observation is relatively different from that of real LR image, the supervised training in SISR task results in a degradation when applied on real-world data. SISR has been scaled to real-world data recently by posing the unsupervised problem into a supervised one through learning the distribution of noisy LR observation first, following which supervised training is performed to obtain the SR image. It therefore involves two steps where the accuracy of SR image relies on how closely the LR's distribution is learnt in the first step. In this work, we overcome such limitation by introducing unsupervised denoising network to transform real noisy LR image to clean image and then pre-trained SR network is utilised to increase the spatial resolution of cleaned LR image to generate SR image. Thus, instead of evaluating the denoised image in LR space to train the denoising network, we inspect the denoised image in SR space which allows to overcome the SR network's generalization problem. The proposed Unsupervised Denoising framework for Super-Resolution (*UDSR*) is validated on real-world datasets (NTIRE-2020 Real-World SR Challenge validation and testing dataset (Track-1)) by comparing it with many recent unsupervised SISR methods. The performance of denoising and SR networks is superior in terms of various perceptual indices such as Perceptual Index (PI) and Ma Score in addition to numerous non-references metrics.

INDEX TERMS Convolutional neural network, generative adversarial network, image enhancement, image restoration, single-image super-resolution, unsupervised learning.

I. INTRODUCTION

In many vision-driven applications such as surveillance and autonomous driving, the fidelity of system relies on the sensor's capability of capturing High-Resolution (HR) data. Such sensors capture precise details of the scene being observed which is preferred for detection and/or feature extraction for both machines as well as to humans. However, installation and use of such high resolution sensors is limited by many factors such as cost of production, sensor space requirements, and ease of manufacturing which prevent them in those applications. To tackle this problem, many works have proposed image *Super-Resolution (SR)* as an alternate approach to obtain HR image from given LR observations

The associate editor coordinating the review of this manuscript and approving it for publication was Felix Albu.

(one or many). Despite its advantages, image SR remains an open research problem due to its ill-posed nature, complexity and unavailability of proper quantitative assessments [1]. Mathematically, the relation between an LR and HR images can be expressed as,

$$I_{LR} = (I_{HR} \otimes k) \downarrow_s + n, \tag{1}$$

where, k, s and n denote blur kernel, down-scaling factor and noise, respectively. Further, ↓ indicates down-scaling operation.

Traditionally, the image SR problem was successfully solved by many theoretical approaches such as prediction-based methods [2], [3], [4], edge-based methods [5], [6], statistical methods [7], patch-based methods [8], and sparse representation methods [9], [10]. However, these approaches were limited to attain the accuracy upto certain extent until the use of deep learning in the field of computer vision. The recent advancements in the field of computational resources and abundant availability of datasets have allowed the increased use of neural networks (specifically Convolutional Neural Networks (CNNs)) based processing to numerous problems of computer vision including the task of Single Image Super-Resolution (SISR). By employing CNN for SISR, researchers have obtained better accuracy over the traditional methods; however, most of approaches are trained in *supervised* manner. The drawback of such supervised training is the unavailability of true Low-Resolution (LR) images and thus they are created by applying known degradation (usually bicubic downsampling) to the HR images [11], [12], [13], [14]. Such LR-HR pairs are employed in the training of CNN model which performs superior for synthetic test images. However, the same network tested on the real-world LR images performs poorly as a result of what is known as domain shifting problem in deep learning where the statistical distribution of synthetic LR image used for training is relatively different from that of true/natural LR observation [15], [16]. Further, many SR works in [17], [18], and [19] are built on the estimation of the kernel (i.e., k in Equation 1); however, such estimated kernel is often dissimilar from the real-world scenario affecting the quality of final SR image [20].

The above mentioned shortcoming associated to LR generation can be solved by utilizing true LR-HR pairs in supervised training. One such attempt was made by Cai et al. [21] in RealSR dataset, where the LR images have been acquired by changing camera's settings along with image registration technique to reduce the registration error between LR and HR data. Based on this dataset, an SR challenge had been organized in NTIRE 2019 [22] and numerous works have been published [23], [24], [25]. However, it is prominent that RealSR dataset is limited to certain upscaling factors and, gathering of such genuine photographs is inefficient and time-consuming, necessitating the use of specifically developed technology and related knowledge to reduce errors [26], [27]. To circumvent this limitation, Lugmayr et al. [16] introduced unsupervised training to the problem of SISR, where true LR-HR pair is not mandatory. In this direction, a few challenges such as AIM 2019 [28] and NTIRE 2020 [15], have been organised recently to explore the idea of unsupervised training for SISR to real-world data. Through such challenges, numerous works [23], [29], [30], [31], [32], [33] obtain state-of-the-art accuracy for real LR images; however, most of these works pose unsupervised training as supervised by learning the distribution of synthetic LR observation to that of *natural* LR data. A learned LR distribution is then used to clean the LR observation and later paired with its corresponding HR image for supervised training. Additionally, the idea of using adversarial samples for unsupervised training has also been introduced in USISResNet [33], where the noisy real LR image is transformed to clean SR without learning LR distribility explicitly. However, optimising a model using unsupervised learning to obtain SR via combining these individual steps (i.e., denoising and SR) is challenging. Therefore, Ahn et al. proposed a solution called SimUSR [34] that employs the BM3D denoising algorithm [35] followed by modified zero-shot learning method. Such idea obtains fair fidelity on distortion metrics; however, the perceptual outcome is inferior due to use of loss based on texture/finer details in denoising task with BM3D method [35] which is used to train zero-shot learning stage.

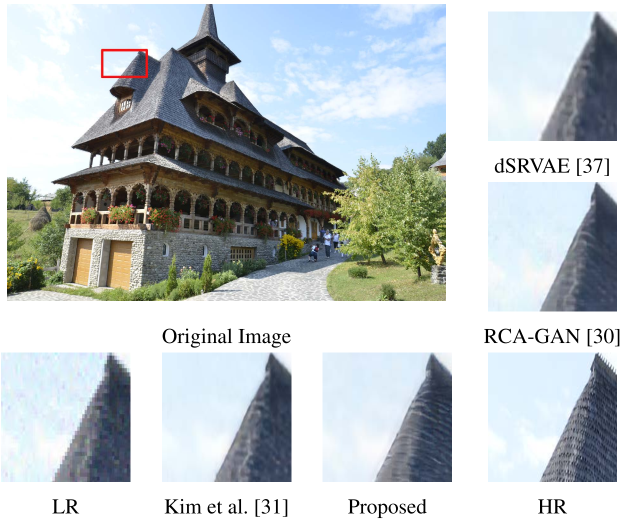


FIGURE 1. The performance of the proposed method (UDSR) compared with different existing methods for unsupervised super-resolution for the up-scaling factor of 4 on representative image of NTIRE-2020 Real-World SR challenge validation dataset.

Noting these constraints, we propose an unique SR design based on an unsupervised approach using a two-step process of denoising and super-resolution. The proposed unsupervised denoising for real-world SR tasks, referred as *UDSR*, not only improves the SR performance but also achieves state-of-the-art denoising performance. To demonstrate applicability, the proposed method is compared to many state-of-the-art unsupervised approaches such as dSRVAE [36], Kim et al. [31], and RCA-GAN [30] and their SR results as depicted in Fig. 1 to visually inspect various methods for upscaling factor ×4 on a single image of the NTIRE 2020 Real-world SR challenge Track-1 validation dataset [15]. The proposed model (i.e. *UDSR*) gains improvement in SR images with superior preservation of high-frequency details, as one can note from Fig. 1. Thus, the proposed approach suppresses noise of LR image in a robust and efficient manner and obtains superior SR result over the other existing state-of-the-art unsupervised techniques. Thus, the primary contributions of this work are summarised as follows:

- We propose a new unsupervised approach for SISR task by including denoising network for real-world LR data. Further, the proposed SR network is placed in between denoising and discriminator networks; thus, they are trained simultaneously to improve the quality of the denoising algorithm as well as the combined performance of both networks (denoising and super-resolution).
- Further, we introduce Residual Channel Split Block (RCSB) in the proposed SR network to act as a backbone layer which utilizes channel splitting concept with residual and densed operations. We assert that such a design of Residual Block (ResBlock) helps to extract meaningful details from the LR observation in optimized manner and the same is also validated experimentally.
- The optimality and robustness of the proposed denoising network to derive clean LR observation is validated quantitatively and quantitatively in an extensive ablation study. In addition to the SR performance, we have also evaluated the performance of the denoising network in subjective and quantitative manners.
- The proposed approach benefits from the incorporated Triplet loss that helps to find an embedding function mapping data with the same label to be close in embedding space and push the data of different classes far apart. Such a design further addresses the training challenges seen in GAN models. In other words, vanilla GAN distinguishes the SR from HR only, which is frequently unstable; the triplet loss, on the other hand, seeks to distinguish between SR with LR as well as with HR.
- The proposed method (i.e., UDSR) is extensively validated on the real-world data (i.e., NTIRE 2020 Real-world SR Challenge dataset [15]), where the noisy LR images are given (i.e., Track-1). In addition to the visual inspection of different SR results, the proposed method is also evaluated on reference-based scores (i.e., LPIPS, pair of PI-RMSE and Ma et al. score) and no-reference-based metrics (i.e., NIQE, BRISQUE and PIQE). The thorough experimental analysis shows that the proposed method trained in unsupervised manner outperforms to the recently proposed SR techniques for real-world data.
- Finally, the statistical evaluation of different quantitative measurements has also been performed in order to support the performance gain of the proposed method.

In the rest of the paper, Section II reviews the different prior SISR works on deep learning-based approaches for handling real-world SR problem. The details of the proposed method (i.e., UDSR) is discussed thoroughly in Section III followed by experimental justification in Section IV. Finally, the limitation and conclusion of work is reported in Section V and Section VI, respectively.

II. RELATED WORKS

Single Image Super-Resolution (SISR) is more frequently encountered challenge as Multiple Image Super-Resolution (MISR) [37]. Considering the scope of the work, we have categorised the existing literature of SISR in terms of traditional, supervised and unsupervised approaches. Since the proposed approach is based on unsupervised training, it is elaborated at length in comparison with other categories.

A. TRADITIONAL SR TECHNIQUES

Traditionally, numerous SR works were based on different mathematical formulations to improve the quality of SR images. For instance many works were based on reconstruction [38], [39], regularization [40], patch [41], [42] and also on self similarity methods [43], [44]. In these methods, an LR image formation is assumed and problem is posed as inverse process. Apart from these, many methods have been utilized interpolation techniques which is computationally efficient [5], [45], [46], [47], [48]; however, artifacts are observed in the SR results obtained using these methods. Importantly, the aforementioned traditional methods are limited to obtain the accuracy upto some extent. In the recent past, the use of deep learning in the field of computer vision has been swelled due to the availability of computing resources and abundant datasets. The emerging use of deep models to the various vision-driven tasks including to the problem of SISR provides higher accuracy level as compared to traditional methods.

B. SUPERVISED SR METHODS

Convolutional Neural Networks (CNNs) based SR methods train deep networks in a highly supervised manner using synthetically generated LR images from the given HR data. Since learning from LR to HR space is complex process, they typically obtain an upsampled LR image first and then improve the quality of SR image using deep neural network. Dong et al. [49] proposed a pioneer model referred as SRCNN, to learn an end-to-end mapping from interpolated LR images. Such pre-upsampling, on the other hand, has many drawbacks as most operations are conducted in high-dimensional space. Further, the time and space costs are substantially higher than with other frameworks [50]. Later, Dong et al. [14] advocated that most computations can be performed in low-dimensional space by applying post-upsampling to improve computational efficiency. Here, the computation and spatial complexity are greatly decreased as the feature extraction procedure, which has a high computational cost, only occurs in low-dimensional space and the resolution rises only at the end. As a result, such pipeline has become one of the most popular in many state-of-the-art works [12], [13], [51].

Further, He et al. [52] proposed ResNet model for learning residuals instead of a comprehensive mapping which has been frequently used by many recent SR models [13], [51], [53]. This can be classified into two categories: global [54], [55] and local [56], [57]. Additionally, by precisely modelling channel dependency, Hu et al. [58] propose a ''squeeze-and-excite'' block to improve learning ability. Combining such channel attention mechanism with SR, Zhang et al. [56] introduced RCAN model, which considerably increases the model's representation ability and hence, also the accuracy of SR results. To aid in the learning of feature correlations, Dai et al. [59] further propose a Second-Order Channel attention (SOCA) module. Moreover, in SR model [60], Zhang et al. substitute conventional convolution with dilated convolution to double the receptive field and obtain superior performance. Further, with the help of recent breakthroughs in lightweight CNNs, IDN [61], ChasNet [62] and CARN-M [63] recommended for to employ group convolution instead of conventional convolution. The group convolution greatly decreases the number of parameters and operations while sacrificing a small amount of performance. Further, Ignatov et al. [64] recently proposed depthwise separable convolution to improve the speed up of SR architecture. Recently, the transformers for natural language processing has further prompted researchers to adapt it for computer vision applications, including super-resolution [65], [66], [67], [68] where it is noted promising to improve accuracy. However, even in such frameworks, a SR model is generally trained using a supervised manner with LR images created by bicubic downsampling. This leaves the problem of realistic degradation HR images unresolved and such models tends to fail when tested on an actual LR image [15].

C. UNSUPERVISED SR TECHNIQUES

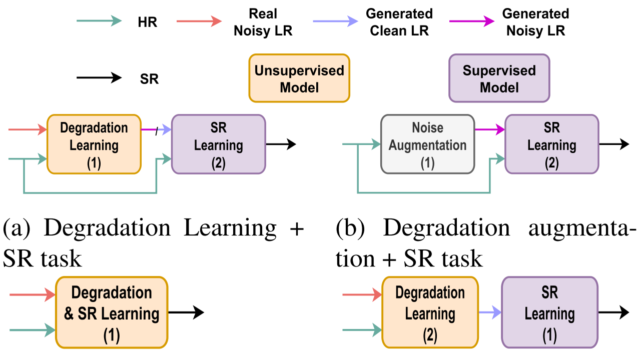
The supervised training in SISR does not translate well to real-world data as its distribution differs significantly from that of synthetically generated LR observations [15], [16], [28]. To tackle this problem, Cai et al. [21] created the RealSR dataset with real LR and HR data. On the basis of this dataset, the NTIRE-2019 challenge [22] was release, and numerous papers have been published [69], [70], [71]. However, creating a real LR-HR pair, on the other hand, is a time-consuming job that necessitates additional hardware and professional hands to avoid image acquisition problems [72], [73].

An alternative way to solve the above challenge is *unsupervised training* where original LH-HR pairs are not demanded. Here, unsupervised training is accomplished through adversarial learning by employing Generative Adversarial Network (GAN), Cycle GAN, Cycle in Cycle GAN and variational autoencoder within the networks. The primary task of such network is to learn unknown degradation of LR observation. However, such models often suffer from stability issues and thus, they are hard to get successful trained model. To accelerate the solution for unsupervised training, couple of challenges have been organised recently in ICCV-2019 (*called AIM 2019 Challenge [28]*) and CVPR-2020 (*called NTIRE 2020 Real-world SR Challenge [15]*.

Different strategies used to handle the problem of unsupervised SR in the existing literature are depicted in Fig. 2 for the convenience of the reader. Many works in [31], [32], [36], and [23] achieve the unsupervised training by splitting it into two sub-tasks as shown in Fig. 2(a). In first task, degradation of real-world LR observation is learned and later, the learnt distribution is used to create synthetic LR image [23], [32]. A few methods in [34], [36], and [31] generate clean image from real-LR image by employing adversarial training. In the following step, such generated/synthetic LR image are paired with its true HR image to train the SR network in supervised manner. However, the accuracy of the SR network is largely depends on the learning capability of unsupervised network and thus, error in the first step affects significantly the sequential task which deteriorates the overall performance. Additionally, a few techniques in [30], [74], and [75] employ random augmentation with multiple noise models to train the SR network (see Fig. 2(b)). Based on this concept, method in [76] combines several augmentation techniques. Similarly, authors in [77] proposed a second order degradation to model the realistic noise environment to train SR network. Such idea makes the SR network robust to the actual noisy LR to tackle the aforementioned problem with two sequential tasks. However, this strategy presumes a true noise model, which may differ from the real-world noise degradation. To mitigate this issue, Mou et al. [78] proposed a metric calculation strategy to handle un-quantified noise. Further, Prajapati et al. introduced USISResNet [33] and DUSGAN [29], a GAN-based approach that uses the capabilities of unsupervised learning with content loss estimated from a bicubic up-sampled picture to maintain the content of an LR image. Instead of sequential learning, they directly learn the noise degradation along with SR in a single step which is depicted in Fig. 2(c). However, the drawback of this idea is training of network which may result in poor preservation of the original content (i.e. less PSNR) and thus, it limits their applicability. In comparison to all of these existing methods, an unique method is been suggested in this paper that learns super-resolution first and then noise reduction with an SR network to generalise the entire framework's performance is illustrated in Fig. 2(d). Thus, the proposed *UDSR* framework not only works with super-resolution on real-LR images, but it also outperforms in terms of both quantitative and qualitative removal of noise.

III. PROPOSED METHOD

The framework of the proposed method is depicted in Fig. 3. To obtain superior perceptual quality of SR solutions in unsupervised manner, we adopt GAN models due to their ability to generate images with better visual quality. As mentioned earlier, it is designed for two specific tasks: Denoising and Super-Resolution (SR). Thus, the proposed approach consists of four different networks: Generator-DeNoise and Generator-SR, Discriminator (D) and Quality Assessment



(a) Degradation Learning + SR task

(b) Degradation augmentation + SR task

(c) Degradation Learning & SR at once

(d) SR + Degradation filtering task

FIGURE 2. Different strategies involved in the existing unsupervised SR approaches. Here, (1) and (2) indicate the order of two steps in the given strategy.

(QA) networks. The function of Generator-DeNoise network is to obtain clean LR image from the given noisy LR observation. The output of this is provided to the Generator-SR network which produces SR image from the available clean LR image. The use of Discriminator (D) network is to train generator networks in adversarial fashion. The QA network is helpful to improve the visual quality of SR image through loss function which is used to train Generator-DeNoise network. In Fig. 3, all above four networks are represented with different colors in rectangle cubical blocks fashion. The Generator-SR and QA networks are shown with blue colors which mean that these networks are pre-trained first and retained fixed in the framework. The SR network is pre-trained in adversarial manner with conventional supervised training. Later, the remaining networks (i.e., Generator-Denoise and Discriminator) are optimized in the adversarial fashion and hence, they are drawn with yellow color. They are trained in unsupervised manner to obtain denoised LR (i.e., clean LR) as well as SR images. The SR network trained prior using synthetically clean dataset cannot handle the noise presented in real-world environment. Moreover, we freeze the parameters of the SR network in the training of the generator network. Hence, in the proposed framework, to improve the quality of SR images using suggested losses, the generator network acts like denoising network which provides clean LR images from the real-world noisy images to the SR network.

Thus, as depicted in Fig. 3, a real-world noisy LR image is passed through the denoising network which acts as a generator in adversarial fashion and hence, cleaned LR image is obtained from this network. It is further fed to pre-trained Generator-SR network to obtain the super-resolved image. The available SR image is given to discriminator and QA networks for computing necessary losses and based on those losses, the Generator-DeNoise and discriminator networks are trained in unsupervised manner. It is necessary to mention here that the training of the Generator-DeNoise and discriminator networks is performed with unpaired LR-HR data (see in Fig. 3) of NTIRE-2020 Real-World SR Challenge dataset [15]. Thus, due to the unavailability of true LR-HR pairs in this dataset, the problem of the SR task is converted into unsupervised training which is hard to solve with traditional GAN loss. Hence, we introduce a modified GAN loss which is conceptualized from triplet loss [79]. In addition to the modified GAN loss (i.e., triplet loss), the proposed framework also includes color and Quality Assessment (QA) [33] losses to remove the noise of LR image and to improve the perceptual quality of the SR image. The design aspects of each of above network are elaborated in details in the following texts.

Generator-SR network: The proposed SR network is pre-trained initially in an adversarial manner using supervised training. It also facilitates the task of unsupervised SR by employing it after denoising network during inference time. The architecture of Generator-SR network is displayed in the Fig. 4(a). To make it simple, the functionality of whole SR network is divided in three modules as: Low-Level Feature (LLF) extraction, High-Level Feature (HLF) extraction and Image Reconstruction (IR) modules. The LLE extraction module consists of three parallel convolutional layers with different kernel size of 3×3, 5×5, and 7×7. Such dissimilar size of kernels assists the network to learn different features available at various reception fields in the LR image. They are equipped with 64 channels which are concatenated further to make them to 192 features maps. In the last layer, we pass these feature maps to a convolution layer with 3×3 kernel and 64 channels. Thus, the output of the LLF extraction module (I_{LLF}) can be represented mathematically as,

$$I_{LLF} = \mathcal{F}_{LLF}(I_{LR}), \tag{2}$$

where I_{LR} indicates LR image and $\mathcal{F}_{LLF}$ denotes the functionality of the LLF extraction module in the proposed SR network. rk in the proposed framework.

The low-level features are further passed through a deeper CNN module referred here as High-Level Feature (HLF) extraction module. This consists of sequence of Residual Channel Splitting Blocks (RCSBs) followed by Channel Attention Block (CAB). The design of RCSB is the backbone of SR network as it extracts vital features from the LR image. In addition to such sequential layers, one skip connection is also appended to improve the stability of the network [13], [51], [80]. The RCSB is designed using a sequential arrangement of Channel Splitting Blocks (CSBs) followed by Pixel Attention Blocks (PABs) connected in residual fashion as displayed in the Fig. 4(b). Inspired from [81], the architecture of the CSB comprises of channel splitting technique to perform two independent operations on two separate paths which are combined at end and multiplied to the skipped path. Inspired from the literature [1], [13], [51], [63], [82], the most common operations to extract features consist of residual and dense operations. Hence, in the proposed RCSB module, we employ them on different paths after the channel splitting (see Fig. 4(b)). Further, the attention based approaches improve the performance of the SR network by changing the features based on the statistics of the up-coming

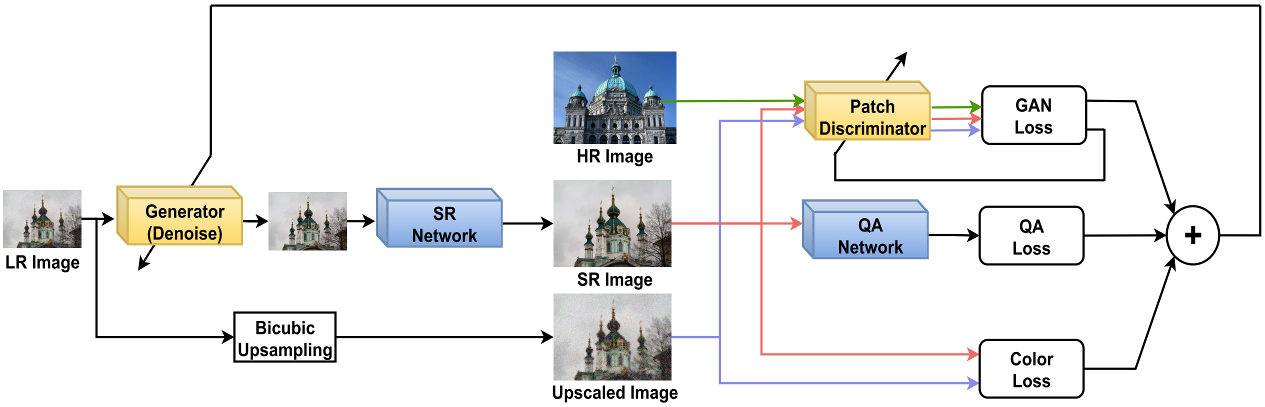


FIGURE 3. The framework of the proposed unsupervised denoiseing SR (*UDSR*) for Real-World image for the up-scaling factor of ×4. Here, blocks represented with blue color are pre-trained networks initially in supervised fashion and networks shown with yellow color indicate that they are trained later in unsupervised manner using unpaired LR-HR images.

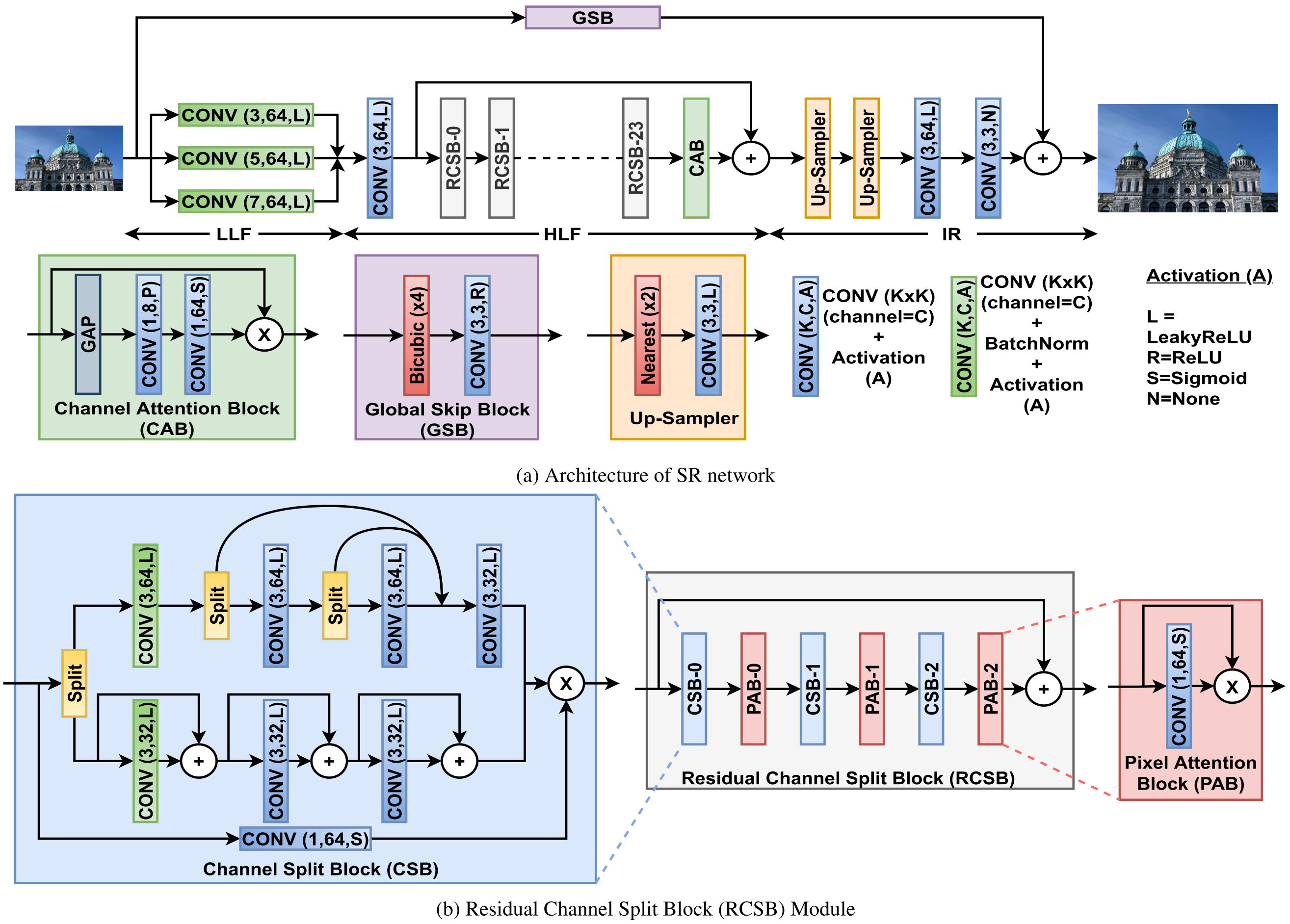


FIGURE 4. The architecture of the proposed SR network-*UDSR* to upsample the denoised LR image for the factor of ×4.

features. For instance, RCAN [56] introduces the channel attention block while PAN [83] introduced pixel attention scheme which boosts the performance significantly. Inspired from these, the proposed SR model utilizes channel attention scheme in CAB block and pixel attention scheme in PAB block which are illustrated in the Fig. 4(a) and Fig. 4(b), respectively. The concatenated features available from two different paths are further multiplied by constant available from residual connection which is inspired from [83]. Further, it is noteworthy that the first convolutional layer in RCSB uses batch normalization to improve the stability of the training process [84] (see Fig. 4(b)). By denoting the functionality of the HLF extraction module by $\mathcal{F}_{HLF}$, we can represent the high-level features I_{HLF} by following equation.

$$I_{HLF} = \mathcal{F}_{HLF}(I_{LLF}) + I_{LLF}. \tag{3}$$

In the last module (i.e., Image Reconstruction module (IR)), the high-level features are mapped to the required dimension of the SR image. First, the usage of each up-sampler layer resizes the features by ×2 which effectively increases the spatial dimension by ×4 (use of two such up-sampler blocks). Each up-sampler consists of nearest up-sampler followed by convolutional layer with kernel size of 3×3. However, the channels of such features are 64 which is finally reduced by 3 channels by using couple of convolutional layers at end in this module network. Additionally, we also employ the usage of Global Skip Block (GSB) to improve the stability of the proposed network. The resultant SR image (I_{SR}) can be formulated as,

$$I_{SR} = \mathcal{F}_{IR}(I_{HLF}) + \mathcal{F}_{GSB}(I_{LR}). \tag{4}$$

Here, $\mathcal{F}_{GSB}$ and $\mathcal{F}_{IR}$ indicates the functionality of the GSB and IR modules, respectively. As depicted in Fig. 3, the denoised SR image available at the end of SR network is given to discriminator due to following advantages.

- Usage of SR network before discriminator helps to enlarge the contents of denoised image which improves the overall quality of adversarial learning pipeline due to higher spatial resolution.
- Moreover, the intermediate usage of SR network in between Generator-DeNoise and Discriminator network helps to generalize the performance of the denoising network along with SR network which improves the task of unsupervised SR.

Generator-DeNoise Network: Employing a single generator network for the SR task in unsupervised manner for real world noisy image, is difficult which motivates us to deploy an another generator network (i.e., Generator-DeNoise) for cleaning the noisy LR observation before SR task. Such design of two networks allows to train the proposed denoising network along with pre-train SR network in unsupervised manner with adversarial learning. Hence, as mentioned earlier, by making the SR network fixed, the denoising network is trained in unsupervised setting to eliminate the unknown degradation available in real-world LR images. The architecture of the denoising network in the proposed framework is depicted in the Fig. 5 which is inspired from DnCNN [85] where residual has been learnt using sequence of 16 (i.e., $k = 16$) convolutional layers. However, instead of learning residual directly on 3 channels, the proposed network employs residuals at 64 feature maps which are subtracted from the input features. Moreover, the initial features are extracted using parallel usage of separate convolutional layers having different kernel sizes similar to the proposed SR network. Additionally, one can notice that the convolutional layers used in residual path utilized the batch normalization which effectively improve the denoise performance. The effectiveness of each parameter and/or setting utilized in the denoising network are studied separately and also demonstrated experimentally in ablation section later in the manuscript.

Quality Assessment (QA) Network: The QA network is used to evaluate the quality of SR image based on Mean Opinion Score (MOS) score provided at the training of the network. The architecture of this network used as proposed by Prajapati et al. [33]. It is used as a loss fucntion in order to train the two generators and its effectiveness in the proposed framework is also justified in ablation study.

Discriminator (D) network. Instead of traditional discriminator, the proposed framework utilizes patch discriminator [86] which examines the image into the small patch and discriminates each patch either as HR patch or fake/generated patch. Compared to a classical discriminator, this discriminator does not use fully connected layers to generate a single predictive value for the entire image. Instead, it generates localized predictive values to describe information for each patch. The architecture of the patch discriminator consists of 6 sequential convolutional layers with 64-128-256-512-512-1 channels. Moreover, each convolutional layer uses a 4×4 kernel as suggested in the original patch discriminator [86]. All convolutional layers except the last one follow with leaky ReLU while the last convolutional layer has a Sigmoid activation function to generate output in the range of 0 to 1.

A. LOSS FUNCTIONS

It is noteworthy to mention that the adversarial learning produces better SR results; however, it results in stability problem during training due to involvement of two different networks. Hence, the loss function plays a major role to train network when trained in an adversarial manner [87], [88], [89]. In this subsection, we discuss different loss functions employed in the proposed framework to train each module.

As discussed earlier, we train SR network prior to the adversarial training of denoise network in the framework. Since the aim of the work is to obtain high fidelity SR solutions for real-world degraded LR images, we use adversarial approach to train the SR network in supervised manner. To train the SR network, we use following combination of losses:

$$\mathcal{L}^{SR} = \lambda_1 \mathcal{L}_1(I_{SR}^{Synthetic}, I_{HR}) + \lambda_2 \mathcal{L}_{VGG}(I_{SR}^{Synthetic}, I_{HR}) + \lambda_3 \mathcal{L}_{Ra-GAN}(I_{SR}^{Synthetic}, I_{HR}). \tag{5}$$

Here, $I_{SR}^{Synthetic}$ denotes SR image generated on synthetically generated LR image which is represented by $I_{SR}^{Synthetic} = \mathcal{F}_{SR}(I_{LR}^{Synthetic})$, where $\mathcal{F}_{SR}$ denotes the function of SR network. It is worth mentioning here that during the training phase of SR network only, we generate the LR image ($I_{LR}^{Synthetic}$), using known down-sampling operation

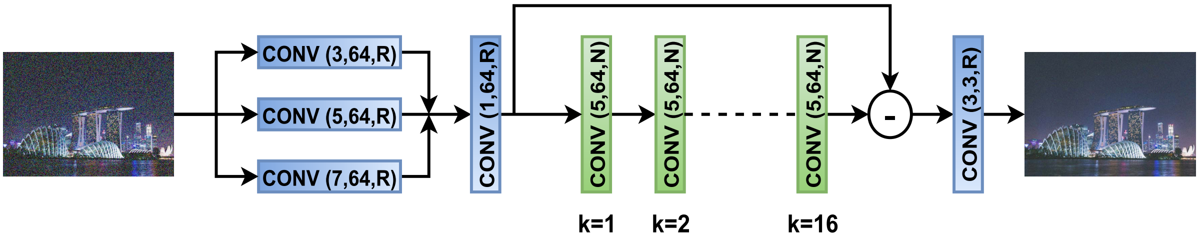


FIGURE 5. The architecture of the proposed denoising network (Generator-DeNoise).

(i.e., bicubic) from the available HR images (I_{HR}). $\mathcal{L}_1$ denotes L_1 loss between $I_{SR}^{Synthetic}$ and I_{HR}, while $\mathcal{L}_{VGG}$ represents VGG perceptual loss suggested by Ledig et al. [51]. In addition, we use relativistic GAN approach [90] for estimating adversarial loss which is denoted by $\mathcal{L}_{Ra-GAN}$. The values of λ_i are set empirically to 0.01, 0.1, and 0.05 for $i = 1, 2, 3$.

Further, the denoising network (Generator-DeNoise) is trained in an adversarial manner using discriminator network which are illustrated with yellow color in Fig. 3. This is accomplished by using unsupervised mode of training. In the proposed framework, we use a combination of three types of losses as:

$$\mathcal{L}_{Denoise} = \mathcal{L}_1(I_{SR}^{Real}, \mathcal{B}(I_{LR}^{Real})) + \mathcal{L}_{QA}(I_{SR}^{Real}) + \mathcal{L}_{LSGAN-G}(I_{SR}^{Real}). \quad (6)$$

The SR image in the learning is generated by denoising network following the SR network and same is represented as $I_{SR}^{Real} = \mathcal{F}_{SR}(\mathcal{F}_{Denoise}(I_{LR}^{Real}))$. $\mathcal{B}$ denotes the bicubic upsampling function. Further, inspired by Prajapati et al. [33], we incorporate QA loss i.e., L_{QA} to improve the perceptual quality of the SR image. Due to the stability issue in unsupervised training, here we employ Least-Square (LS) GAN loss (i.e., $\mathcal{L}_{LSGAN-G}$) [91] represented as,

$$\mathcal{L}_{LSGAN-G} = \sum^{N}(1 - \mathcal{D}(I_{SR}^{Real}))^2. \quad (7)$$

Here, N represents batch size during training iteration. The classical GAN framework which uses non-saturating loss (i.e. $-log(D(\cdot))$) or min-max loss (i.e. $log(1 - D(\cdot))$) saturates on either side and hence it creates problem of vanishing gradient. It can be resolved by least-square based loss which is not saturating on either side. Additionally, authors in [91] also claim that the LSGAN generates better quality of images as compared to classical GAN.

Additionally, to improve the quality of generated SR image and also to stabilize the training, the loss function of the discriminator has been modified in this manuscript. We incorporate the notion of triplet loss with few modifications to optimize the discriminator network. It can be written as

$$\mathcal{L}_{LSGAN-D} = \sum^{N}\left(\mathcal{D}(\mathcal{B}(I_{LR}^{Real})) + \mathcal{D}(I_{SR}^{Real}) + (1 - \mathcal{D}(I_{HR}))\right), \quad (8)$$

where, the bicubic upsampled image ($\mathcal{B}(I_{LR}^{Real})$), super-resolved image (I_{SR}^{Real}) and unpaired HR image (I_{HR}) are treated as negative, anchor and positive samples respectively. By inserting negative samples (i.e. bicubic up-sampled noisy LR image), we are providing more clue to the discriminator network which helps to stabilize the adversarial learning [79]. Thus, by comparing anchor with negative and positive samples, we maximize discriminator score for positive sample and minimizing the score for negative and anchor samples by optimizing the discriminator network. Further, the vanilla GAN suffers with the problem of mode collapse [92] which can also be solved by triplet loss based GAN framework [93], [94].

IV. EXPERIMENTAL ANALYSIS

We have conducted numerous experiments to validate the proposed method. These experiments have been performed on a system equipped with an Intel Xeon(R) Dual CPU with 128GB RAM and a dual NVIDIA Quadro P5000 GPU with 16GB memory. The PyTorch library is used to implement the proposed framework. To show the different details associated with experimental details, this section is categorized in the series subsections. The training dataset and augmentation strategies, as well as the hyper-parameters utilized to train the network, are discussed in Section IV-A. Further, the testing dataset, as well as the reference-based and no-reference-based image quality assessment criteria utilized here to compare the proposed network's performance are detailed in Section IV-B followed by ablation study in Section IV-C. Later, the performance of the proposed SR method is compared with other state-of-the-art methods quantitatively in Section IV-D and statistical analysis of the performance is presented in Section IV-E. Apart from the quantitative comparison, the qualitative comparison is also provided in the Section IV-F. Last, the computational complexity of the proposed method is compared with other existing methods and same is presented in Section IV-G.

A. TRAINING DETAILS AND HYPER-PARAMETER SETTINGS

The proposed method uses two independent training pipelines to address the unsupervised real-world SR problem. In the first stage, the SR network along with discriminator are trained on DIV2K [95] by creating an LR image synthetically using bicubic downsampling. This dataset consists of total 1000 images among which it is divided into 800-100-100 for training-validation-testing purposes. It is trained upto 2×10^5 number of iterations with a batch size of 16. Apart from the SR network, we also performed pre-training of the Quality Assessment (QA) network which has been trained to estimate better Mean Opinion Score (MOS) based on human rating. To train this network, we used KADID-10K [96] dataset which contains 10,125 images obtained using 81 unique pristine images with 25 types of degradation at 5 levels of each. The detailed description related to network architecture and training strategy are discussed in USISResNet [33].

After the completion of the first stage of training, the trained SR and QA networks are kept fixed to train the denoising network (i.e., Generator-DeNoise) further using an unsupervised approach in the second stage. Here, we use the NTIRE-2020 Real-world SR Challenge dataset [15] which consists of 2650 noisy images from DF2K dataset [97] where the noise is generated synthetically. Apart from the noisy images, it also consists the 800 clean DIV2K [95] images. It is important to that the dataset in this stage is not in paired which forces us to use unsupervised learning. Initially, both networks are initialized with a Kaiming initialization [98] (i.e., Denoising and SR networks). In the training of both networks, the random crop with 192×192 on HR images and with 48×48 on LR images along with random horizontal flipping and random rotation with 90° are used in augmentation process. An Adam optimizer with β_1 and β_2 having the value of 0.9 and 0.99 are set to train the SR and denoising networks. The denoising network is trained upto the 4×10^5 number of iterations. In both stages of training, we start the training with the learning rate of 1×10^{-4} which is decayed by half at every $(1/4)^{th}$ of the total iterations.

B. TESTING DETAILS

The potential of the proposed unsupervised SR approach is validated on two testing datasets. Along with the NTIRE-2020 Real-world SR Challenge Track-1 [15] validation dataset where HR images are available, we have also employed the testing dataset of that challenge (Track-1) in which original ground-truth is not unavailable. Further, the performance of the proposed denoising network is compared with many state-of-the-art methods such as BM3D [35],[1] DIP [99][2] and NAC [100].[3] Similarly, the unsupervised SR results are compared with ZSSR [101], SRMD [11], USISResNet [33],[4] SimUSR [34], dSRVAE [36],[5] Kim et al. [31],[6] SRResCGAN [32] [7] and RCA-GAN [30] methods.

Additionally, we compare the quantitative performance of the proposed method for real-world unsupervised SR using fidelity based measures such as Root Mean Squared Error (RMSE) and perceptual measures such as Ma et al. score [102] and Perceptual Index (PI) on NTIRE-2020 Real World-SR validation dataset [15] since they require ground-truth images. In addition, we also utilize Naturalness Image Quality Evaluator (NIQE) [103], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [104], and Perception Based Image Quality Evaluator (PIQE) [105] measures based on no-reference image quality assessment to compare the performance on NTIRE-2020 Real-World validation and Track-1 testing dataset. The deep learning based image quality measure termed LPIPS [106] is also estimated which has a better correlation with human perception. It is noteworthy to mention that there are always a trade-off between perception and distortion [107] and the same is also analyzed in the PI-RMSE plot in the results. Finally, we show the statistical analysis on the quantitative evaluation of the different methods along with the proposed method to judge the performance of all the above methods in statistical sense.

C. ABLATION STUDY

In this section, we discuss the importance of different hyperparameters settings of the proposed framework. This is categorized into three parts for better readability. First, the effectiveness of different modules in the SR network is discussed. In the second part, we show the ablation study on denoising network. At last, we discuss the effectiveness of loss function in the proposed whole framework (i.e., denoising and SR).

Numerous experiments to study the effect of the pixel and channel attention modules in the SR network have been conducted. Here, we have trained the SR network without using the above two modules by adopting training strategy as mentioned in the Section IV-A. To observe the effect that are comparable with other experiments, denoising model has been trained separately after pre-training of both SR networks as discussed earlier. The quantitative results obtained using these experiments is compared in terms of PSNR, SSIM and LPIPS values in Table 1. By observing this table, one can note that the SR performance is boosted by employing pixel and channel attention modules in the proposed SR network. Along with quantitative evaluation, the visual inspection of above cases are also depicted in Fig. 6. Here, both of these cases are shown in the Fig. 6(b-c) lack in preserving perceptual fidelity (i.e., visual details) as compared to the proposed method with pixel and channel attentions.

In the proposed denoising network, we employ 16 convolutional layers with 5×5 kernel size (see Fig. 5); to understand the effectiveness of the size of kernel in SR results, we have carried out first experiment with different size of kernel i.e., 3×3 and 7×7. Similarly, above experiment is extended by increasing the number of convolution layers to 24 instead of 16. Further, one can note that the intermediate layers of the proposed denoising network are designed without any activation function which is also tested by using ReLU activation function. The quantitative comparison of all above settings are depicted in Table 1 in terms of PSNR, SSIM and LPIPS.

[1] https://github.com/gfacciol/bm3d

[2] https://github.com/DmitryUlyanov/deep-image-prior

[3] https://github.com/csjunxu/Noisy-As-Clean-TIP2020

[4] https://github.com/kalpeshjp89/USISResNet

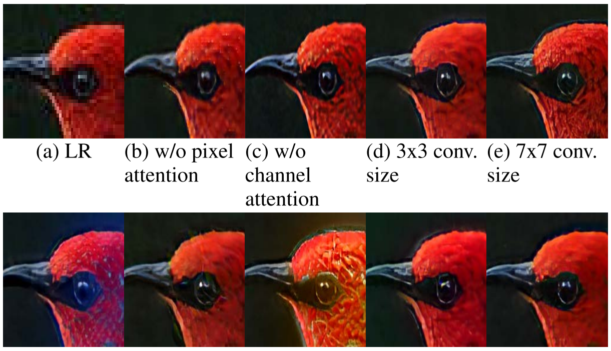
[5] https://github.com/Holmes-Alan/dSRVAE

[6] https://github.com/GT-KIM/unsupervised-super-resolution-domain-discriminator

[7] https://github.com/RaoUmer/SRResCGAN

TABLE 1. The quantitative comparison of different settings to validate the proposed framework design for Real-World SR problem.

Configuration	PSNR↑	SSIM↑	LPIPS↓
SR network configuration			
w/o pixel-attention	26.5617	0.7119	0.3342
w/o channel-attention	26.7629	0.7158	0.3296
Proposed	**26.7797**	**0.7195**	**0.3272**
Denoise (residual) network configuration			
3x3 convolutional size	26.7423	**0.7265**	0.3431
7x7 convolutional size	**26.9722**	0.7241	0.3317
24 convolutional layers	26.1457	0.7131	0.3488
ReLU activation function	26.6938	0.7169	0.3552
Proposed	26.7797	0.7195	**0.3272**
Loss function			
w/o QA loss	25.5744	0.6946	0.3446
w/ standard GAN loss	26.0918	0.7186	0.3532
Proposed	**26.7797**	**0.7195**	**0.3272**


(a) LR (b) w/o pixel attention (c) w/o channel attention (d) 3x3 conv. size (e) 7x7 conv. size

(f) 24 conv. blocks (g) ReLU function (h) w/o QA loss (i) w/ standard GAN loss (j) Proposed UDSR

FIGURE 6. The visual SR results for various cases of ablation study.

TABLE 2. The effectiveness of the SR network prior to discriminator in unsupervised denoising problem.

Configuration	PSNR↑	SSIM↑	LPIPS↓
w/o SR network	21.8404	0.3808	0.7160
Proposed	**26.7797**	**0.7195**	**0.3272**

TABLE 3. The comparison of the proposed de-noising network with other existing state-of-the-art methods on NTIRE-2020 Real World-SR validation dataset. The best two values in each measures are highlighted using red and blue colors respectively.

Method	NIQE↓	PIQE↓	BRISQUE↓	RMSE↓	Ma score↑	PI↓
BM3D [35]	4.3689	49.8367	33.8431	3.3743	8.3099	3.1230
DIP [99]	5.6317	70.5863	40.2249	15.4300	6.0035	5.4024
NAC [100]	4.7935	36.3319	31.9542	15.6760	7.7499	3.5218
Proposed	3.7670	31.5157	25.5600	5.1413	8.5872	2.5889

One can observe that the proposed denoising network with 16 convolutional layers and 5 × 5 earns better gain in different metrics than those configurations. From the visual comparison also, one can note that the proposed setting obtains better visual quality than that of other configurations displayed in Fig. 6(d-g). Additionally, the proposed framework consists of a modified triplet and LSGAN losses in addition to QA loss to improve the perceptual quality of SR results. The performance of the proposed method without QA loss and without LSGAN, with triplet loss is also analysed. The results depicted in Table 1 indicate that the proposed configuration achieves better measures over the different losses. The same have been compared visually in Fig. 6(h-j) with the proposed method where the efficacy of the proposed method can be easily observed.

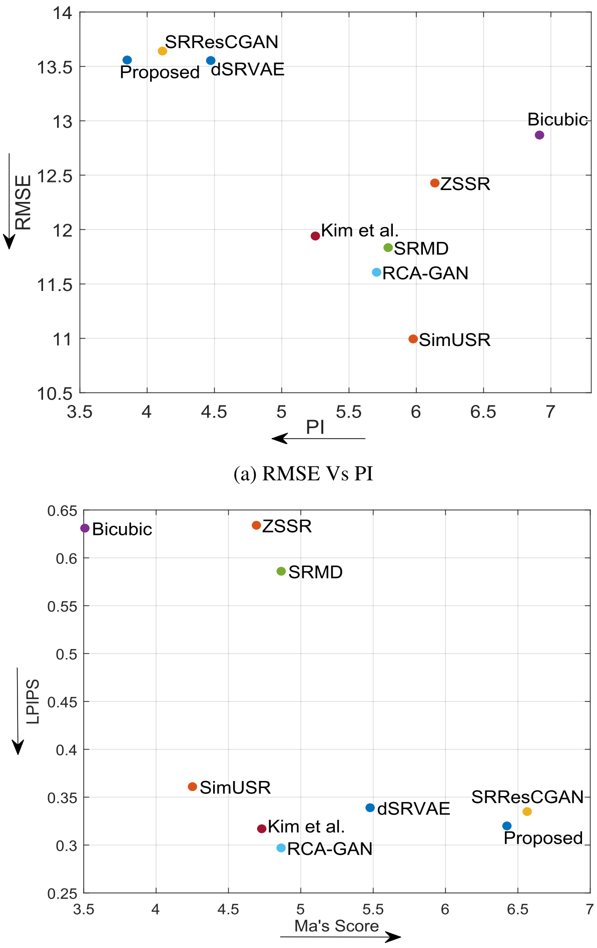
Here, our prime focus is to deal with real-world SR problem in unsupervised manner. To solve the problem, the proposed approach is employed with denoising network which is optimized by introducing SR network prior to the discriminator network as suggested in Fig. 3. Such insertion of the SR network in the optimization of the denoise network improves the performance of the denoise network along with SR output. To check the effectiveness of the proposed setting, we measure the outcomes obtained from the denoising network and the quantitative measurements are listed in the Table 2. It can be easily observed that the SR network in the proposed framework is helpful to improve the performance of denoising network and hence to obtain SR image too. This is due to estimation of loss functions in SR space instead of on LR output. Hence, the enhanced images obtained from SR network further useful to improve the performance of denoising network.

D. QUANTITATIVE EVALUATION

Here, we show the quantitative evluation of the proposed denoising model along with SR network in terms of various measures. In addition to the distortion metric (i.e., RMSE), we also incorporate the perceptual measures such as Perceptual Index (PI) and Ma et al. score where ground-truth images are required. Further, the performance of the proposed framework is also judged quantitatively by evaluting various non-reference metrics such as NIQE, BRISQUE and PIQE (to configure the case of unavailability of original image). In Table 3, we show the performance of the proposed denoising network (output is taken from denoising network) in terms of NIQE, BRISQUE, PIQE, RMSE, Ma et al. score and PI. The results here are measured on the degraded images from NTIRE-2020 Real-World SR validation data. It is worth mentioning that the lower value of NIQE, BRISQUE, PIQE, RMSE and PI suggests the better performance while the higher value in the case of Ma et al. score indicates superior quality of an image. By inspecting Table 3, one can note that the performance of NAC [100] method is inadequate to denoise the noisy images. In this NTIRE-2020 Real-World SR validation dataset, NAC [100] fails on 10 images out of total 100 images. Further, one can easily observe from Table 3 that the proposed method outperforms over the other state-of-the-art methods in terms of the NIQE, BRISQUE, PIQE and PI. However, it achieves second position for RMSE measure in which the BM3D [35] has superior performance.

TABLE 4. The quantitative comparison of the different existing unsupervised SISR methods along with the proposed method for upscaling factor ×4 on NTIRE Real-world SR challenge dataset.

Method	Validation Dataset			Testing Dataset (Track-1)		
	NIQE↓	BRISQUE↓	PIQE↓	NIQE↓	BRISQUE↓	PIQE↓
Bicubic	5.972	55.153	83.659	5.716	55.790	83.839
SRMD [11]	5.842	45.256	71.318	5.847	45.763	71.649
ZSSR [101]	6.384	47.924	77.062	6.478	48.713	77.722
SRResCGAN [32]	3.928	25.384	25.574	3.949	24.097	24.929
SimUSR [34]	4.424	46.928	81.271	4.453	45.723	81.633
RCAGAN [30]	5.472	33.511	42.563	5.519	32.671	42.082
Kim et al. [31]	4.262	38.847	46.719	4.333	38.422	47.041
dSRVAE [36]	3.606	32.606	39.138	3.650	32.708	38.391
Proposed	2.938	23.935	21.444	3.007	21.463	20.178



(a) RMSE Vs PI

(b) LPIPS Vs. Ma et al. score

FIGURE 7. The plots of perceptual measures obtained using the proposed and other existing unsupervised methods on NTIRE-2020 Real-world SR Challenge Track-1 validation dataset.

Additionally, the performance of the proposed unsupervised SR method is benchmarked with other recent state-of-the-art unsupervised SR methods in the Table 4. The comparison is carried out on NTIRE-2020 Real World SR validation and testing dataset (Track-1). Here, the no-reference based measures are used to gauge the performance of SR methods because of unavailability of reference images in the testing data. These measures suggest a quality of an image without considering the original content of LR image. From this table, we can observe the superiority of the proposed method with that of other methods. Further, there is always a trade-off between the perception and distortion metrics [107] and this is illustrated in the Fig. 7(a) where the performance of different techniques are plotted in terms of PI vs. RMSE plane [64][8] for the NTIRE 2020 Real-world SR Challenge Track-1 validation dataset for upscaling factor ×4. From this comparison, one can see the performance enhancement of the proposed method while comparing with other methods. Thus, the proposed method achieves better PI value than the other existing methods with the RMSE value similar with SRResCGAN [32] and dSRVAE [36] techniques. Additionally, the performance of different methods are also depicted on the plane of Ma et al. score vs. LPIPS value in Fig. 7(b). From both of these plots, we can observe that the proposed method is dominated by some extent over other methods considering the multiple performance metrics (i.e. PI-RMSE or Ma et al. score-LPIPS).

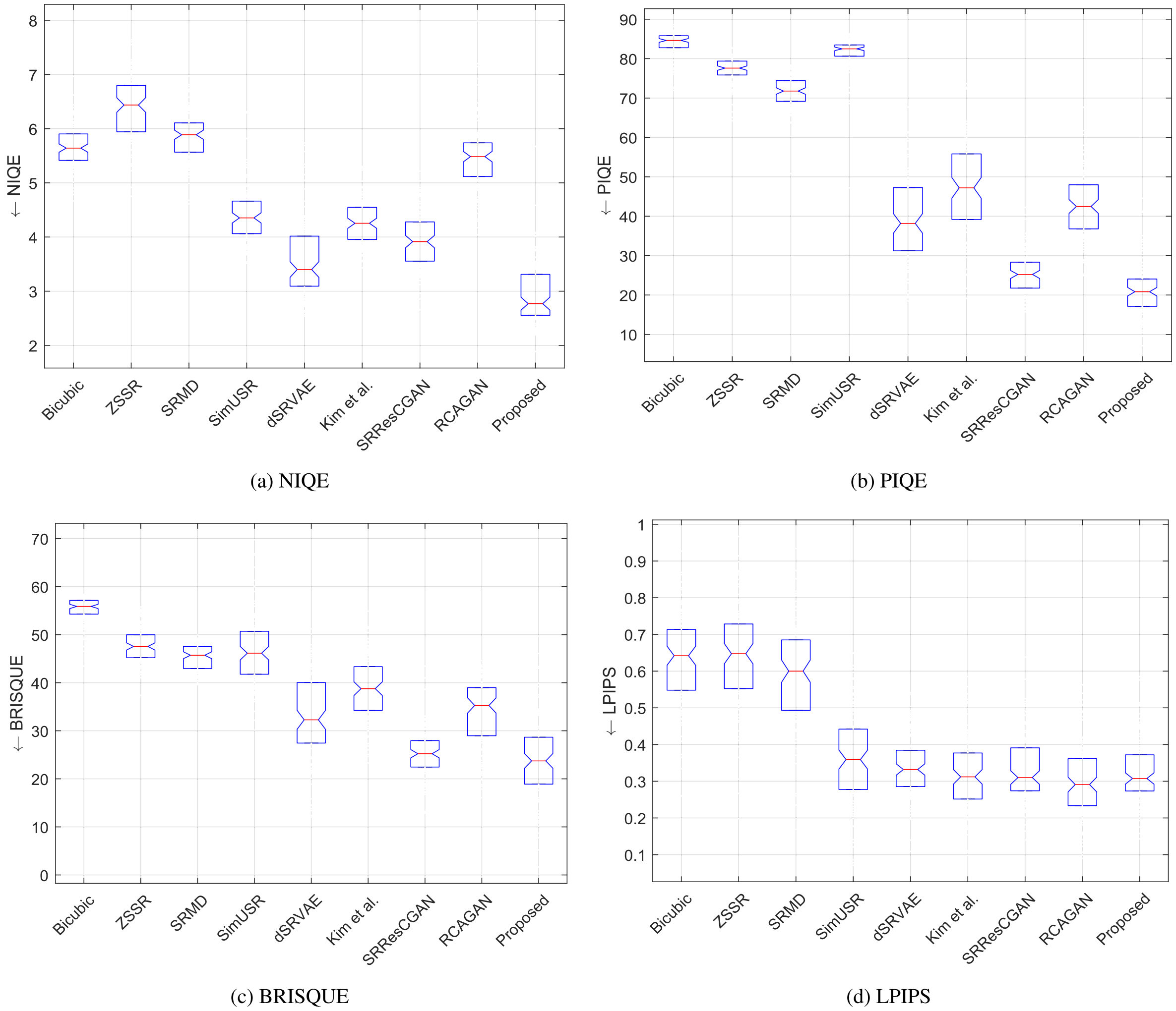
E. STATISTICAL ANALYSIS

The quantitative comparison of various method analyzed in last section are judged statistically here. We have carried out the Analysis of Variance (ANOVA) test, and the findings, computed at a 95 % Confidence Interval (CI), are depicted in Fig. 8 as a box-plot. From Fig. 8(a-c), one can observe that dSRVAE [36] and SRResCGAN [32] compete the other existing methods including SimUSR [34], Kim et al. [31] and RCA-GAN [30] in terms of NIQE, PIQE and BRISQUE respectively. Interestingly, it can be observed that zero-shot based method [101] (i.e., ZSSR) which learns degradation from LR image itself, does not work well here due to noise contain in LR image. Similarly, one can also observe that the generalization problem of CNN network by noticing the performance of SRMD [11] method which is trained on different types of noise patterns. However, in all of these cases the proposed method excels against other methods by achieving better quantitative performance. Further, one can note from Fig. 8(d) that the proposed method reveals slightly lower LPIPS value than RCA-GAN method; however, the variance of the LPIPS across the dataset for the proposed method is consistent over the RCA-GAN method. Moreover, as mentioned earlier that the performance of the proposed method in terms of other perceptual measures such as PI and Ma's score in addition to LPIPS (see Fig. 7) is better when compared to RCA-GAN and other existing state-of-the-art methods.

F. QUALITATIVE EVALUATION

Apart from the quantitative comparison, we discuss the subjective evaluation of the proposed method with respect to the other state-of-the-art methods in this subsection. As mentioned earlier, the insertion of SR network in unsupervised denoising algorithm improves the perceptual quality of the denoising framework which is depicted in the Fig. 9. Here,

[8] https://github.com/roimehrez/PIRM2018



(a) NIQE

(b) PIQE

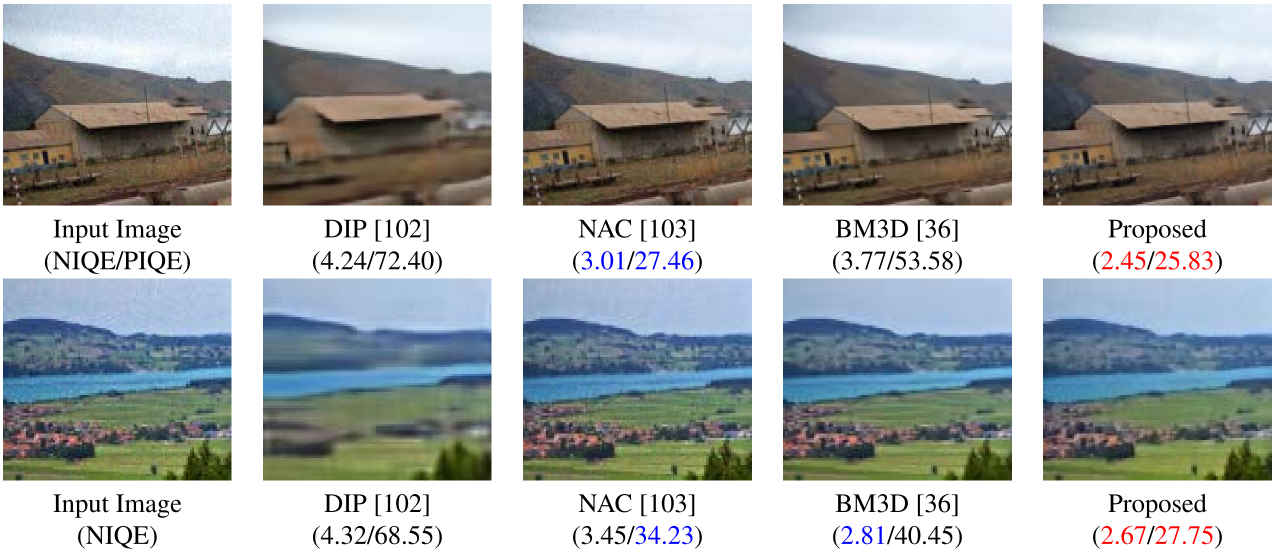
(c) BRISQUE

(d) LPIPS

FIGURE 8. **The statistical evaluation of the different SISR method with Analysis of Variance (ANOVA) test on NTIRE-2020 Real-World SR challenge Track-1 validation dataset [15].**

the proposed method is compared with the other denoising methods such as DIP [99], NAC [100] and BM3D [35] on NTIRE-2020 Real-World SR Challenge validation dataset. For fair comparison, the values of NIQE and PIQE of each result are also mentioned. By inspecting Fig. 9, one can easily observe that the DIP [99] method fails to preserve high frequency details and hence, yields blurry results. In contrast, NAC [100] extracts high frequency components successfully; however, it is not effective to eliminate noise completely and thus, noise is visible in the results (zoom these results for better visualization). Further, BM3D method [35] generates competing results with that of the proposed method; however, the visual performance of BM3D is slightly poorer than that of the proposed method. Additionally, as depicted in the Fig. 9, in terms of quantitative metrics (i.e. NIQE and PIQE), the performance of proposed method is superior than BM3D method.

Further, the visual comparison of the unsupervised real-world SR methods is depicted in the Fig. 10 and Fig. 11 for validation and testing datasets, respectively. In this comparison, the proposed method is competed with the other existing unsupervised real-world SR methods such as dSRVAE [36], Kim et al. [31], SRResCGAN [32], RCA-GAN [30] along with generalized method SRMD [11] and zero shot based ZSSR [101]. Fig. 10 shows comparison of above methods on the representative image of NTIRE-2020 Real-World SR validation dataset [15]. For the shake of illustration, no-reference based quality measurement NIQE and PIQE are also listed with each SR patch and the best two values are highlighted using red and blue colors, respectively. Here, tt can be observe that the ZSSR [101], dSRVAE [36], Kim et al. [31] and RCA-GAN [30] methods fail to generate texture from corresponding noisy LR image. Further, SRMD [11] estimates the texture up to some extent. Similarly, SRResCGAN can also generate better structure of wall; however, it is unable to eliminate noise in the SR image. Compared to all the existing methods, the SR result of the proposed method yields visually plausible data which have better capacity to generate texture



Input Image (NIQE/PIQE)

DIP [102] (4.24/72.40)

NAC [103] (3.01/27.46)

BM3D [36] (3.77/53.58)

Proposed (2.45/25.83)

Input Image (NIQE)

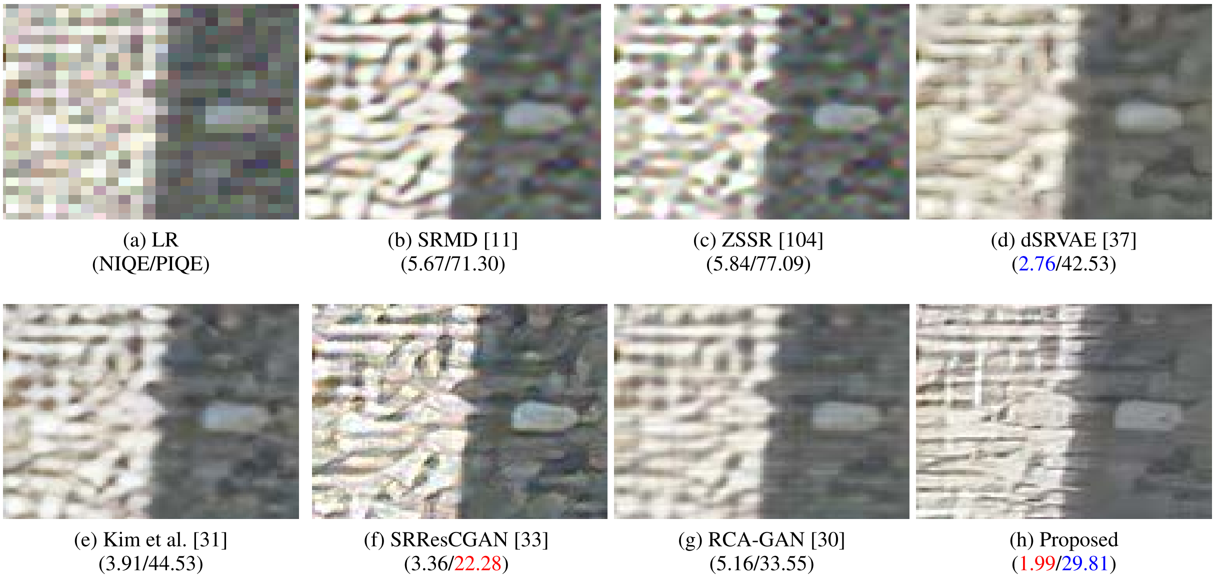
DIP [102] (4.32/68.55)

NAC [103] (3.45/34.23)

BM3D [36] (2.81/40.45)

Proposed (2.67/27.75)

FIGURE 9. **The visual comparison of the proposed denoising method with the other existing methods on NTIRE-2020 Real-World SR Challenge validation LR images [15]. The values of NIQE and PIQE are also mentioned alog side of each denoised patch. The highest and second highest are indicated with red and blue color texts, respectively.**



(a) LR (NIQE/PIQE)

(b) SRMD [11] (5.67/71.30)

(c) ZSSR [104] (5.84/77.09)

(d) dSRVAE [37] (2.76/42.53)

(e) Kim et al. [31] (3.91/44.53)

(f) SRResCGAN [33] (3.36/22.28)

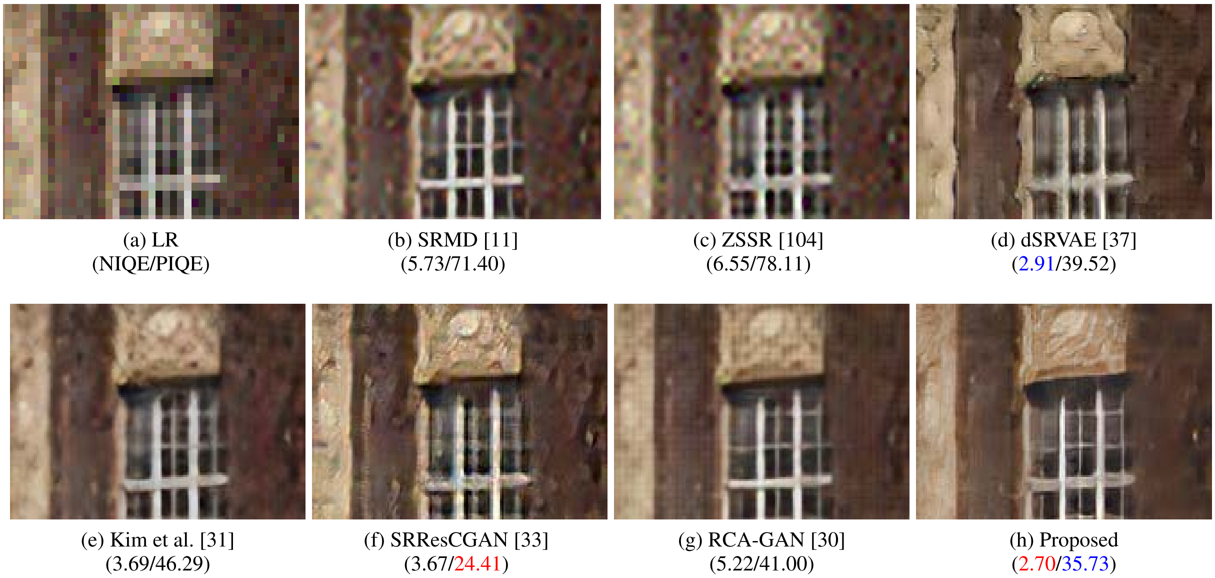
(g) RCA-GAN [30] (5.16/33.55)

(h) Proposed (1.99/29.81)

FIGURE 10. **The qualitative comparison of the SR images obtained using the proposed and other recent unsupervised SR methods on NTIRE-2020 Real-world SR Challenge Track-1 validation dataset [15]. The values of NIQE and PIQE are mentioned alogside of each SR patche and the highest two values are depicted red and blue color texts, respectively.**

with elimination of noise in robust manner. The visual result can also be validated by comparing the quantitative measurements which are listed with each method. The proposed method has highest value in terms of NIQE and second highest value in terms of PIQE which supports that the superior performance of the proposed method over the other methods for NTIRE-2020 validation dataset [15].

Similarly, the qualitative comparison of the proposed method with other state-of-the-art methods are depicted in the Fig. 11 where the SR results are taken from representative image from NTIRE-2020 Real-World SR testing Track 1 dataset [15]. It is worth to mention that the original images are not available in this dataset. Similar to earlier result, here one can notice that the performance of SRMD [11], ZSSR [101], Kim et al. [31], dSRVAE [36] and SRResCGAN [32] methods is poor as they are not preserving the proper shape of objects (see the window frame and vertical border between different colored wall). However, RCA-GAN [30] generates the SR image with better shapes of an objects; In contrast, the proposed method reconstructs SR image with proper shape of window and also looking plausible when compared to other methods. In addition, the quantitative measurements of



(a) LR
(NIQE/PIQE)

(b) SRMD [11]
(5.73/71.40)

(c) ZSSR [104]
(6.55/78.11)

(d) dSRVAE [37]
(2.91/39.52)

(e) Kim et al. [31]
(3.69/46.29)

(f) SRResCGAN [33]
(3.67/24.41)

(g) RCA-GAN [30]
(5.22/41.00)

(h) Proposed
(2.70/35.73)

FIGURE 11. The qualitative comparison of the SR images obtained using the proposed and other recent unsupervised SR methods on NTIRE-2020 Real-world SR Challenge Track-1 testing dataset [15]. The values of NIQE and PIQE are mentioned alogside of each SR patche and the highest two values are depicted red and blue color texts, respectively.

TABLE 5. The computational complexity in terms of the number of parameters and number of multiplication-addition operations of different unsupervised real-world SR models. The number of multiplication-addition operations is calculated for each model for SR image of size 512 × 512 resolution of an SR image.

Method	Trainable Parameters (SR network)	Multiply-Addition Operations
ZSSR [101]	0.22M	232.53G
SRMD [11]	1.48M	25.41G
SRResCGAN [32]	0.38M	1.5M
SimUSR [34]	15.59M	19.27G
RCA-GAN [30]	15.59M	19.27G
Kim et al. [31]	16.69M	22.89G
dSRVAE [36]	0.29M	78.22G
UDSR (Proposed)	9.3M	22.72G

this experiment are in support with that of visual assessment and thus, indicates the superiority of the proposed method when compared to other existing Real-World SR methods on NTIRE-2020 Real-World SR challenge testing Track-1 dataset [15].

G. COMPUTATIONAL COMPLEXITY

Additionally, Table 5 shows the number of trainable parameters along with multiplication-addition operations of the proposed and other real-world SR methods. The different SR methods such as SimUSR [34], Kim et al. [31], and RCA-GAN [30] have employed either RCAN [56] or ESRGAN [13] network for their SR network, which have almost 15M-16M number of trainable parameters. The proposed method consists of 9.3M parameters with SR network which is far less than of all of the other competitive methods. However, the proposed approach needs denoising network along with SR network to generate clean SR image from an LR observation. Adding the complexity of the denoising network (i.e. 1.6M parameters) in entire pipeline results in a total of 10.9M parameters and is marginally lower than other existing SR methods. Additionally, all methods have been compared based on multiply-addition operations for an SR network on 128×128 input LR image. Based on this measure, the proposed network shows less complexity than ZSSR [101], SRMD [11], Kim et al. [31] and dSRVAE [36] methods. The complexity of the proposed method is slightly higher than SimUSR [34] and RCA-GAN [30] based on this metric. Further, it is worth mentioning that computational complexity of SRResCGAN [32] method is better than that of the proposed method; however, the visual and quantitative evaluations indicate better performance compared to all of the above-mentioned existing state-of-the-art methods.

V. LIMITATIONS AND FUTURE SCOPE

Despite of achieving better quantitative and qualitative SR results, the proposed method has moderate distortion in terms of RMSE metric. It is always a trade-off between perceptual quality (which can be measured by LPIPS, Ma's score, NIQE, PIQE and BRISQUE) and distortion (which is mostly described by PSNR, SSIM and RMSE) for image restoration problem [107]. Further, the perceptual quality of the proposed method beats to other existing methods; however, same can be improved from that of visual SR results depicted in Fig. 10 and Fig. 11. Additionally, in this work, we have designed very lightweighted denoising network which also may be the reason for limited SR performance. In future work, it can also be extended to see the effect of complex denoising architecture.

Further, Aakerberg et al. [108] proposed semantic segmentation based annotation to solve the real-world SR problem. However, it demands specific kind of annotation (i.e., based on segmentation) which is not cost effective. Thus, one can study to utilize such other types of cheaper annotations which can improve the SR performance. Moreover, Xu et al. [109] observed the cross-domain limitation to broaden the super-resolution problem. Additionally, authors in [110] employed UNet based discriminator and obtained performance gain in the SR task. Such flavour of discriminator can also be investigated for real-world scenarios. In the same work, authors also introduce an uncertainty visualization about the genuineness of each pixel whether it is generated or natural. Such additional information might be useful in some critical application including medical imaging, biometric application etc., which is again an interesting domain to consider in future works.

VI. CONCLUSION

The requirement of the paired LR-HR dataset to train any deep network in supervised manner cannot be met in practical scenarios. However, such problem can be solved by training the network in unsupervised manner which has drawback of stability issue. In this manuscript, we have proposed a framework called *UDSR* using unsupervised way to handle the problem of Real-World Super-Resolution by inserting separate denoising network. The combination of SR network and denoising networks can be evaluated by measuring the performance of both networks and we found that it improves the performance of both. With respect to denoising network, the inclusion of SR network before to discriminator network improves the quality of the denoise network which has been evaluated quantitatively as well as qualitatively. However, because of stability issue, it is hard to train SR network that can perform both tasks simultaneously (i.e. denoising and SR) to handle real-world SR problem. In the proposed method, we train the SR network prior in supervised manner followed by optimizing denoising network in unsupervised way. In addition, we incorporate triplet loss based concept into LSGAN which stabilizes the training process more effectively which is also validated in the ablation study section of the manuscript. Based on vast amount of reference based and no-reference based measurements, one can conclude that the proposed method outperforms than the other state-of-the-art methods. Moreover, the SR performance of the proposed method is validated in subjective manner with the other existing methods.

REFERENCES

[1] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2020.

[2] P. Thévenaz, T. Blu, and M. Unser, "Interpolation revisited [medical images application]," *IEEE Trans. Med. Imag.*, vol. 19, no. 7, pp. 739–758, Jul. 2000.

[3] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.*, vol. 18, no. 8, pp. 1016–1022, 1979.

[4] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 3, pp. 231–239, May 1991.

[5] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, Aug. 2006.

[6] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 30, no. 2, p. 12, 2011.

[7] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definitio," *IEEE Trans. Image Process.*, vol. 3, no. 3, pp. 233–242, May 1994.

[8] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2004, pp. 1–8.

[9] H. Chavez-Roman and V. Ponomaryov, "Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1777–1781, Oct. 2014.

[10] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1127–1133, Jun. 2010.

[11] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3262–3271.

[12] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[13] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Computer Vision*, L. Leal-Taixé and S. Roth, Eds. Cham, Switzerland: Springer, 2019, pp. 63–79.

[14] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. ECCV*, Oct. 2016, pp. 391–407.

[15] A. Lugmayr et al., "NTIRE 2020 challenge on real-world image super-resolution: Methods and results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2020, pp. 2058–2076.

[16] A. Lugmayr, M. Danelljan, and R. Timofte, "Unsupervised learning for real-world super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3408–3416.

[17] J. Gu, H. Lu, W. Zuo, and C. Dong, "Blind super-resolution with iterative kernel correction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1604–1613.

[18] Z. Luo, Y. Huang, S. Li, L. Wang, and T. Tan, "Unfolding the alternating optimization for blind super resolution," 2020, *arXiv:2010.02631*.

[19] R. Zhou and S. Susstrunk, "Kernel modeling super-resolution on real low-resolution images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2433–2443.

[20] A. Liu, Y. Liu, J. Gu, Y. Qiao, and C. Dong, "Blind image super-resolution: A survey and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Aug. 30, 2022, doi: 10.1109/TPAMI.2022.3203009.

[21] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, "Toward real-world single image super-resolution: A new benchmark and a new model," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3086–3095.

[22] J. Cai, S. Gu, R. Timofte, and L. Zhang, "NTIRE 2019 challenge on real image super-resolution: Methods and results," in *Proc. IEEE Conf. CVPRW*, Jun. 2019, pp. 1–13.

[23] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang, "Real-world super-resolution via kernel estimation and noise injection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 466–467.

[24] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2020, pp. 492–511.

[25] R. Feng, J. Gu, Y. Qiao, and C. Dong, "Suppressing model overfitting for image super-resolution networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2019, pp. 1–10.

[26] C. Chen, Z. Xiong, X. Tian, Z.-J. Zha, and F. Wu, "Camera lens super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1652–1660.

[27] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to learn, learn to zoom," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3757–3765.

[28] A. Lugmayr et al., "AIM 2019 challenge on real-world image super-resolution: Methods and results," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, 2019, pp. 3575–3583.

[29] K. P. Prajapati, V. M. Chudasama, H. Patel, K. P. Upla, K. B. Raja, R. Ramachandra, and C. Busch, "Direct unsupervised super-resolution using generative adversarial network (DUS-GAN) for real-world data," *IEEE Trans. Image Process.*, vol. 30, pp. 8251–8264, 2021.

[30] J. Cai, Z. Meng, and C. Man Ho, "Residual channel attention generative adversarial network for image super-resolution and noise reduction," in *Proc. IEEE CVPRW*, Jun. 2020, pp. 1852–1861.

[31] G. Kim, J. Park, K. Lee, J. Lee, J. Min, B. Lee, D. K. Han, and H. Ko, "Unsupervised real-world super resolution with cycle generative adversarial network and domain discriminator," in *Proc. IEEE CVPRW*, Jun. 2020, pp. 1862–1871.

[32] R. M. Umer, G. L. Foresti, and C. Micheloni, "Deep generative adversarial residual convolutional networks for real-world super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1769–1777.

[33] K. Prajapati, V. Chudasama, H. Patel, K. Upla, R. Ramachandra, K. Raja, and C. Busch, "Unsupervised single image super-resolution network (USISResNet) for real-world data using generative adversarial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1904–1913.

[34] N. Ahn, J. Yoo, and K.-A. Sohn, "SimUSR: A simple but strong baseline for unsupervised image super-resolution," in *IEEE CVPRW*, Jun. 2020, pp. 1953–1961.

[35] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[36] Z.-S. Liu, W.-C. Siu, L.-W. Wang, C.-T. Li, M.-P. Cani, and Y.-L. Chan, "Unsupervised real image super-resolution via generative variational AutoEncoder," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1788–1797.

[37] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1. Greenwich, CT, USA: JAI Press, 1984, pp. 317–339.

[38] M. V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning-based method for image super-resolution from zoomed observations," *IEEE Trans. Syst., Man, B, Cybern.*, vol. 35, no. 3, pp. 527–537, Jun. 2005.

[39] H. Xu, G. Zhai, and X. Yang, "Single image super-resolution with detail enhancement based on local fractal analysis of gradient," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1740–1754, Oct. 2013.

[40] R. M. Bahy, G. I. Salama, and T. A. Mahmoud, "Adaptive regularization-based super resolution reconstruction technique for multi-focus low-resolution images," *Signal Process.*, vol. 103, pp. 155–167, Oct. 2014.

[41] G. Freedman and G. Fattal, "Image and video upscaling from local self-examples," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–10, Apr. 2010.

[42] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Proc. IEEE 12th ICCV*, Sep. 2009, pp. 349–356.

[43] C. Cruz, R. Mehta, V. Katkovnik, and K. O. Egiazarian, "Single image super-resolution based on Wiener filter in similarity domain," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1376–1389, Mar. 2018.

[44] M. Ebrahimi and E. R. Vrscay, "Solving the inverse problem of image zooming using 'self-examples,'" in *Proc. ICIAR*, 2007, pp. 117–130.

[45] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 7–1521, Oct. 2001.

[46] R. Franke, "Scattered data interpolation: Tests of some methods," *Math. Comput.*, vol. 38, no. 157, pp. 181–200, 1982.

[47] J. Allebach and P. W. Wong, "Edge-directed interpolation," in *Proc. 3rd IEEE Int. Conf. Image Process.*, vol. 3, Sep. 1996, pp. 707–710.

[48] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.

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

[50] W. Shi, J. Caballero, F. HuszÃİr, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. CVPR*, Jun. 2016, pp. 1874–1883.

[51] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4681–4690.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. CVPR*, Jun. 2016, pp. 770–778.

[53] R. Timofte, V. D. Smet, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. ACCV*, 2014, pp. 111–126.

[54] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. CVPR*, Jun. 2016, pp. 1637–1645.

[55] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. CVPR*, vol. 1, no. 4, Jul. 2017, pp. 3147–3155.

[56] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[57] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 517–532.

[58] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[59] T. Dai, J. Cai, Y. Zhang, S. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11057–11066.

[60] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2808–2817.

[61] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 723–731.

[62] K. Prajapati, V. Chudasama, H. Patel, A. Sarvaiya, K. Upla, K. Raja, R. Ramachandra, and C. Busch, "Channel split convolutional neural network (ChaSNet) for thermal image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 4363–4372.

[63] Y. Li, E. Agustsson, S. Gu, R. Timofte, and L. Van Gool, "CARN: Convolutional anchored regression network for fast and accurate single image super-resolution," in *Computer Vision* (Lecture Notes in Computer Science), vol. 11133. Springer, 2019, pp. 166–181.

[64] A. D. Ignatov et al., "PIRM challenge on perceptual image enhancement on smartphones: Report," 2018, *arXiv:1810.01641*.

[65] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, and T. Zeng, "Transformer for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2022, pp. 456–465.

[66] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5790–5799.

[67] Z. Lu, H. Liu, J. Li, and L. Zhang, "Efficient transformer for single image super-resolution," 2021, *arXiv:2108.11084*.

[68] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1833–1844.

[69] J. Xiao, H. Yong, and L. Zhang, "Degradation model learning for real-world single image super-resolution," in *Proc. ACCV*, 2020, pp. 1–17.

[70] M. S. Rad, T. Yu, C. C. Musat, H. K. Ekenel, B. Bozorgtabar, and J.-P. Thiran, "Benefitting from bicubically down-sampled images for learning real-world image super-resolution," 2020, *arXiv:2007.03053*.

[71] S. Son, J. Kim, W.-S. Lai, M.-H. Yang, and K. M. Lee, "Toward real-world super-resolution via adaptive downsampling models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8657–8670, Nov. 2021.

[72] Y. Guo, X. Wu, and X. Shu, "Data acquisition and preparation for dual-reference deep learning of image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 4393–4404, 2022, doi: 10.1109/TIP.2022.3184819.

[73] T. Wang, J. Xie, W. Sun, Q. Yan, and Q. Chen, "Dual-camera super-resolution with aligned attention modules," 2021, *arXiv:2109.01349*.

[74] H. Ren, A. Kheradmand, M. El-Khamy, S. Wang, D. Bai, and J. Lee, "Real-world super-resolution using generative adversarial networks," in *Proc. IEEE CVPRW*, Jun. 2020, pp. 1760–1768.

[75] A. Castillo, M. Escobar, J. C. P'erez, A. Romero, R. Timofte, L. V. Gool, and P. Arbel'aez, ''Generalized real-world super-resolution through adversarial robustness,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1855–1865.

[76] K. Zhang, J. Liang, L. V. Gool, and R. Timofte, ''Designing a practical degradation model for deep blind image super-resolution,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4771–4780.

[77] X. Wang, L. Xie, C. Dong, and Y. Shan, ''Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1905–1914.

[78] C. Mou, Y. Wu, X. Wang, C. Dong, J. Zhang, and Y. Shan, ''Metric learning based interactive modulation for real-world super-resolution,'' 2022, *arXiv:2205.05065*.

[79] M. Schultz and T. Joachims, ''Learning a distance metric from relative comparisons,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 1–8.

[80] J. Kim, J. K. Lee, and K. M. Lee, ''Accurate image super-resolution using very deep convolutional networks,'' in *Proc. IEEE CVPR*, Jun. 2016, pp. 1646–1654.

[81] X. Zhao, Y. Zhang, T. Zhang, and X. Zou, ''Channel splitting network for single mr image super-resolution,'' *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5649–5662, Jun. 2019.

[82] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in *Proc. IEEE CVPR*, Jul. 2017, pp. 4700–4708.

[83] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, ''Efficient image super-resolution using pixel attention,'' in *Proc. ECCV Workshops*, 2020, pp. 56–72.

[84] S. Ioffe and C. Szegedy, ''Batch normalization: Accelerating deep network training by reducing internal covariate shift,'' 2015, *arXiv:1502.03167*.

[85] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, ''Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising,'' *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[86] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, ''Image-to-image translation with conditional adversarial networks,'' in *Proc. IEEE Conf. CVPR*, Jul. 2017, pp. 1125–1134.

[87] Z. Wang, Q. She, and T. E. Ward, ''Generative adversarial networks: A survey and taxonomy,'' 2019, *arXiv:1906.01529*.

[88] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, ''Are GANs created equal? A large-scale study,'' in *Proc. NIPS*, 2018, pp. 1–10.

[89] M. Arjovsky and L. Bottou, ''Towards principled methods for training generative adversarial networks,'' 2017, *arXiv:1701.04862*.

[90] A. Jolicoeur-Martineau, ''The relativistic discriminator: A key element missing from standard GAN,'' 2018, *arXiv:1807.00734*.

[91] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley, ''Least squares generative adversarial networks,'' in *Proc. IEEE ICCV*, Oct. 2017, pp. 2794–2802.

[92] N. Kodali, J. Hays, J. D. Abernethy, and Z. Kira, ''On convergence and stability of GANs,'' 2018, *arXiv:1705.07215*.

[93] G. Cao, Y. Yang, J. Lei, C. Jin, Y. Liu, and M. Song, ''TripletGAN: Training generative model with triplet loss,'' 2017, *arXiv:1711.05084*.

[94] S. Yu, K. Zhang, C. Xiao, X. Bao, J. Z. Huang, and M. J. Li, ''Btgan: Training GAN with balanced triplet loss and two-branch architecture,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2021, pp. 1–8.

[95] E. Agustsson and R. Timofte, ''NTIRE 2017 challenge on single image super-resolution: Dataset and study,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1122–1131.

[96] H. Lin, V. Hosu, and D. Saupe, ''KADID-10k: A large-scale artificially distorted IQA database,'' in *Proc. 11th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2019, pp. 1–3.

[97] E. Agustsson and R. Timofte, ''NTIRE 2017 challenge on single image super-resolution: Dataset and study,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 114–125.

[98] K. He, X. Zhang, S. Ren, and J. Sun, ''Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,'' in *Proc. IEEE Conf. ICCV*, Dec. 2015, pp. 1026–1034.

[99] V. Lempitsky, A. Vedaldi, and D. Ulyanov, ''Deep image prior,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[100] J. Xu, Y. Huang, M.-M. Cheng, L. Liu, F. Zhu, Z. Xu, and L. Shao, ''Noisy-as-clean: Learning self-supervised denoising from corrupted image,'' *IEEE Trans. Image Process.*, vol. 29, pp. 9316–9329, 2020.

[101] A. Shocher, N. Cohen, and M. Irani, ''Zero-shot super-resolution using deep internal learning,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3118–3126.

[102] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, ''Learning a no-reference quality metric for single-image super-resolution,'' *Comput. Vis. Image Understand.*, vol. 158, pp. 1–16, May 2017.

[103] A. Mittal, R. Soundararajan, and A. C. Bovik, ''Making a 'completely blin' image quality analyzer,'' *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.

[104] A. Mittal, A. K. Moorthy, and A. C. Bovik, ''Blind/referenceless image spatial quality evaluator,'' in *Proc. Conf. Rec. 45th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Nov. 2011, pp. 723–727.

[105] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, ''Blind image quality evaluation using perception based features,'' in *Proc. 21st Nat. Conf. Commun. (NCC)*, Feb. 2015, pp. 1–6.

[106] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, ''The unreasonable effectiveness of deep features as a perceptual metric,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[107] Y. Blau and T. Michaeli, ''The perception-distortion tradeoff,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.

[108] A. Aakerberg, A. S. Johansen, K. Nasrollahi, and T. B. Moeslund, ''Semantic segmentation guided real-world super-resolution,'' in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 449–458.

[109] X. Xu, P. Wei, W. Chen, M. Mao, L. Lin, and G. Li, ''Dual adversarial adaptation for cross-device real-world image super-resolution,'' 2022, *arXiv:2205.03524*.

[110] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, ''DU-GAN: Generative adversarial networks with dual-domain U-Net-based discriminators for low-dose CT denoising,'' *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

KALPESH PRAJAPATI received the bachelor's degree in electronics and communication from Dharmsinh Desai University, Nadiad, India, and the master's degree in automatic control and robotics from the Maharaja Sayajirao University of Vadodara, India. He is currently pursuing the Ph.D. degree with the Sardar Vallabhbhai National Institute of Technology, Surat, India. His research interests include image enhancement, single-image super-resolution, image quality assessment, unsupervised learning, weakly supervised learning, and medical imaging.

VISHAL CHUDASAMA (Member, IEEE) received the bachelor's degree from Maharaja Sayajirao University, Vadodara, India, the master's degree in communication system from Dharmsinh Desai University, Nadiad, India, and the Ph.D. degree from the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India. His current research interests include image processing, deep learning with application to super-resolution, object detection and recognition, low-resolution face detection and recognition, medical imaging, and biometrics.

HEENA PATEL (Member, IEEE) received the Ph.D. degree from the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India. She is an AI Researcher at Logicwind, Surat. She has developed various real-time applications such as visiting card scanner, virtual try on jewelry, 2D floor plan analysis with 3D conversion, materials such as rods, stirrups, and flapping analysis in building construction, honeycomb detection at construction site, and object detection and tracking. Her research interests include image enhancement, domain translation, image super-resolution, thermal imaging, and computer vision applications using deep learning algorithms.

ANJALI SARVAIYA (Graduate Student Member, IEEE) received the B.E. degree in electronics and communications engineering and the M.E. degree in communication systems from Gujarat Technological University, India. She is currently pursuing the Ph.D. degree with the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India. Her research interests include signal/image processing, machine learning, image super-resolution using deep learning, and medical imaging.

KISHOR UPLA (Member, IEEE) received the Ph.D. degree from the Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, India. He is an Assistant Professor at the Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat, India. He has more than 15 years of academic and research experience from different technical universities across Gujarat, India. He worked as an ERCIM Postdoctoral Fellow with NTNU, Gjøvik, Norway. His research interests include signal and image processing, low-resolution face recognition, biometric, and multispectral and hyperspectral image analysis. He is a project partner in collaborative research work with NTNU, Norway, under ''INTPART–International Partnerships for Excellent Education and Research'' funded by the Research Council of Norway (RCN), Norway. He is also a member of the European Association for Biometrics (EAB).

KIRAN RAJA (Senior Member, IEEE) received the Ph.D. degree in computer science from the Norwegian University of Science and Technology, Norway, in 2016. He is a Faculty Member with the Department of Computer Science, Norwegian University of Science and Technology. His main research interests include statistical pattern recognition, image processing, and machine learning with applications to biometrics, security, and privacy protection. He was/is participating in EU projects SOTAMD, iMARS, and other national projects. He is a member of the European Association of Biometrics (EAB). He is the Chair of Academic Special Interest Group, EAB. He serves as a reviewer for number of journals and conferences. He is also a member of the editorial board for various journals.

RAGHAVENDRA RAMACHANDRA (Senior Member, IEEE) received the Ph.D. degree in computer science and technology from the University of Mysore, Mysore, India, the Institute Telecom, and Telecom Sudparis, Evry, France (carried out as a collaborative work), in 2010. He is currently appointed as a Full Professor with the Institute of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology, Gjøvik, Norway. He was a Researcher with the Istituto Italiano di Tecnologia, Genoa, Italy, where he worked with video surveillance and social signal processing. He has authored several papers. He also holds several patents in biometric presentation attack detection and morphing attack detection. His main research interests include deep learning, machine learning, data fusion schemes, and image/video processing, with applications to biometrics, multimodal biometric fusion, human behavior analysis, and crowd behavior analysis. He has received several best paper awards. He was/is also involved in various conference organizing and program committees. He was/is participating (as a PI/a Co-PI/a contributor) in several EU projects, IARPA USA, and other national projects. He is a reviewer for several international conferences and journals. He has served as the editor for ISO/IEC 24722 standards on multimodal biometrics and an active contributor for ISO/IEC SC 37 standards on biometrics. He is serving as an associate editor for various journals.

CHRISTOPH BUSCH (Senior Member, IEEE) is a member of the Norwegian University of Science and Technology, Norway. He holds a joint appointment with Hochschule Darmstadt (HDA), Germany. He has been Lectures biometric systems with DTU, Denmark, since 2007. On behalf of the German BSI, he has been the Coordinator for the project series BioIS, BioFace, BioFinger, BioKeyS Pilot-DB, KBEinweg, and NFIQ2.0. He was/is a partner of the EU projects 3D-Face, FIDELITY, TURBINE, SOTAMD, RESPECT, TReSPsS, and iMARS. He is also a Principal Investigator with the German National Research Center for Applied Cybersecurity (ATHENE). He is a Co-Founder of the European Association for Biometrics. He has coauthored more than 500 technical papers and has been a speaker at international conferences. Furthermore, he Chairs the TeleTrusT Biometrics Working Group and the German Standardization Body on Biometrics. He is a Convenor of WG3 in ISO/IEC JTC1 SC37. He is a member of the Editorial Board of the IET journal on *Biometrics*.