IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 January 2023, accepted 23 February 2023, date of publication 2 March 2023, date of current version 8 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3251396

RESEARCH ARTICLE

Deep Learning-Based Single-Image Super-Resolution: A Comprehensive Review

KARANSINGH CHAUHAN[1], SHAIL NIMISH PATEL[2], MALARAM KUMHAR[3], JITENDRA BHATIA[3], SUDEEP TANWAR[3], (Senior Member, IEEE), INNOCENT EWEAN DAVIDSON[4], (Senior Member, IEEE), THOKOZILE F. MAZIBUKO[5], AND RAVI SHARMA[6]

[1]Department of Computer Science, Dalhousie University, Halifax, NS B3H 4R2, Canada
[2]Department of Computer Science and Engineering, Ahmedabad University, Ahmedabad, Gujarat 380009, India
[3]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat 382481, India
[4]Department of Electrical, Electronic and Computer Engineering, Cape Peninsula University of Technology, Bellville 7535, South Africa
[5]Department of Electrical Power Engineering, Durban University of Technology, Durban 4000, South Africa
[6]Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun 248001, India

Corresponding authors: Jitendra Bhatia (jitendra.bhatia@nirmauni.ac.in), Sudeep Tanwar (sudeep.tanwar@nirmauni.ac.in), and Thokozile F. Mazibuko (ThokozileM1@dut.ac.za)

ABSTRACT High-fidelity information, such as 4K quality videos and photographs, is increasing as high-speed internet access becomes more widespread and less expensive. Even though camera sensors' performance is constantly improving, artificially enhanced photos and videos created by intelligent image processing algorithms have significantly improved image fidelity in recent years. Single image super-resolution is a class of algorithms that produces a high-resolution image from a given low-resolution image. Since the advent of deep learning a decade ago, this field has made significant strides. This paper presents a comprehensive review of the deep learning assisted single image super-resolution domain including generative adversarial network (GAN) models that discusses the prominent architectures, models used, and their merits and demerits. The reason behind covering the GAN models is that it is been known to perform better than the conventional deep learning methods given the resources and the time. For real-world applications with noise and other issues that can cause low-fidelity super resolution (SR) images, we examine another solution based on GAN model. This GAN model-based technique popularly known as blind super resolution is more resilient. We examined the various super-resolution techniques by varying image scaling factors (i.e., 2x, 3x, 4x) to measure PSNR and SSIM metrics for the different datasets. PSNR across the different datasets covered in the experimental section shows an average of 14-17 % decrease in the score as we move up the image resolution scale from 2x to 4x. This is observed across all the datasets and for every model mentioned in the experimental section of the paper. The results also show that blind super-resolution outperforms the conventional deep learning methods and the more complex GAN models. GAN models are complex and preferred when the upscale factor is high, while residual and dense models are recommended for smaller upscaling factors. This paper also discusses the applications of image super-resolution, and finally, the paper is concluded with challenges and future directions.

INDEX TERMS Image super-resolution, deep learning, convolutional neural network, generative adversarial network.

I. INTRODUCTION

Single Image Super-Resolution is an image processing algorithm that focuses on recovering a high-resolution image from a single low-resolution image. Various techniques that have been developed in the past have produced notable outcomes. In this section, a few of those techniques are covered. These techniques include regularisation techniques, neighbourhood embedding-based algorithms, and the use of similar patches that are redundant in low-resolution images

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

in their counterparts with higher resolution. Different methods can achieve super-resolution of images. One approach involves fixed mapping methods in which a fixed relation is obtained from some a priori information. Few other approaches are leaned towards machine learning (ML) based methods that use a dictionary and aim at learning the mapping from low-resolution to high-resolution images. There also exist some low-complexity algorithms that aim at better results without utilizing a lot of resources. A neighborhood-embedding-based set of algorithms exists in which the high-resolution counterpart of the image is learned from the apriori HR patches, which are stored in the dictionary. These patches can be combined in the same way they correspond to low-resolution images. This method is made better by using euclidean distance to perform a neighbour search, followed by optimizing a least squares problem. However, these processes are not optimal and fail at the task when the scale of the image in question is comparatively higher [1], [2], [3]. Thus deep learning-based techniques are evolved.

Further recently, there has been a paradigm shift from the traditional algorithms to deep-learning-based approaches, beginning with the first CNN-based algorithm proposed by Dong et al. [4] to the recently proposed Generative Adversarial Network (GAN) based algorithms. Recently, some methods of GAN-based or learning super-resolution space can generate simulated textures but do not promise the accuracy of the textures which have low quantitative performance. Rethinking both, there is a model that learns the distribution of underlying high-frequency details in a discrete form and proposing a two-stage pipeline i.e., divergence stage to convergence stage [5].

Deep learning methods have been the dominant performer in image processing in recent years. Image segmentation has largely benefited from the introduction of deep learning in terms of accuracy, label classes, and inference time [6], [7]. Gatys et al. [8] used convolutional neural networks in their work on neural style transfer to fuse the artistic qualities of one image into the other. Xie et al. [9] have proposed a neural network to reconstruct images with missing patches, such as old torn images. Apart from this, the machine learning pipeline has seen a rise in automation [10]. Similarly, in the domain of Single Image Super-Resolution, the use of deep neural networks has shown a phenomenal increase in the perceptual quality of the generated image. Dong et al. have already shown in their work that the previous traditional state-of-the-art methods, such as sparse representation, are a special case of their three-layer convolutional neural network. Their work also showed how traditional methods are sub-optimal for image super-resolution [11].

Image super-resolution attempts to extract the visual features from a given image to create a higher-resolution version of the same image. These features include low-complexity ones such as outlines, shapes, and shades to highly abstract the features like textures and luminance. By taking these features and extrapolating the details across a higher resolution, Image super-resolution aims to generate images with better resolution without needing any change in the capturing hardware. This intrinsic property of learning to extrapolate features motivates us to adopt deep learning as a perfect match for such a computational problem. It is possible to represent higher-order abstract features using the deep network.

It is imperative to frame the super-resolution process as a mathematical model to understand it better. For remaining of the literature, we will be defining ILR as a low resolution image with Width W, Height H, and channels C (For a normal RGB image, $C = 3$). We also define I_{HR}, which is the high-resolution counterpart of I_{LR} and has an upscaling factor of S. Since IHR has a higher resolution, it has a width of $S*W$, Height $S * H$, and channels C. In practice, to create a dataset of I_{HR} and I_{LR} pairs, we downgrade a given image to generate I_{LR} whereas the original image is taken as the I_{HR}. The degradation process to generate I_{LR} is performed by downsizing the I_{HR} by a scale of S and optionally introducing noise in the I_{LR}. Mathematically, we can define the degradation process ϑ as follow.

$$I_{LR} = \vartheta_S(I_{HR}) \tag{1}$$

considering the Eq. 1 we can frame the super-resolution process as its inverse, ϑ^{-1}. Therefore,

$$I_{SR} = \vartheta_S^{-1}(I_{LR}) \tag{2}$$

where, ISR is the super-resolution version of the I_{LR} and has the exact dimensions as that of I_{HR}. In an ideal scenario, I_{SR} should be the same as I_{HR}, i.e., the inverse operation inverse perfectly maps the I_{LR} to I_{HR}. However, this is not true in practice because ϑ is a one-to-many function, i.e., the same I_{LR} can be generated by degrading multiple different I_{HR}. This is because when we downsize a high-resolution image, we are mapping multiple pixel values to a single pixel value. For brevity's sake, consider the mapping operation an average function that takes multiple pixel values from I_{HR} and maps them to a single pixel value in I_{LR}. Since multiple input values can correspond to the same averaged output value, the mapping operation is not an injective function. Due to this, a single low-resolution image can be generated from multiple high-resolution images with minor variations amongst them. Therefore, Eq. 2 is a one-to-many function, i.e., it is surjective. In other words, single-image super-resolution is an under-determined inverse problem, of which the solution is not unique. Figure 1 shows this problem with an example.

This multi-solution problem is usually addressed by constraining the solution space by extracting prior information. The preliminary information in the case of a deep learning-based super-resolution model usually refers to the original high-resolution image used to generate its low-resolution counterpart. These high-resolution images are taken as the target output of the convolutional networks that learns the mapping operation between the I_{LR} and I_{HR} through back-propagation. Most deep-learning models are exemplar-based, where the low-resolution images and their corresponding high-resolution images are given during training. Some models exploit self-similarity within the image to find low

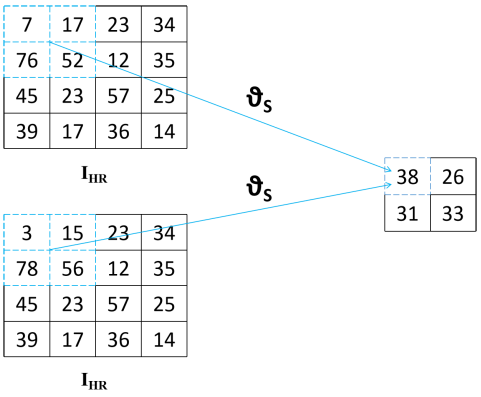

FIGURE 1. Multiple I_{HR} generating the same I_{LR} assuming average mapping operator for downscaling.

and high-resolution patch pairs and thus are self-supervised [12], [13]. However, the success of such models is limited; hence, we will not discuss them in depth in this paper. The primary motivation behind the review is the increasing research focus on the domain that demands an extensive review of the ongoing research work to provide the right direction to the community. In addition, due to the rising use of high-resolution content on the web, the applicability of image super-resolution is expanding rapidly. While there have been some reviews [14], [15], [16], [17], [18] in the past on the same, there was a distinct lack in exploring newer models mainly GANS and Blind SR which have a more meaningful impact on real life applications. The main contributions of this review are:

1) We present the single-image super-resolution approaches, specifically deep learning.
2) We provide a study of the accuracy of various algorithms and the evaluation criteria.
3) We cover a in depth review of GAN models. Compare them with conventional deep learning models and explore how they are able to perform better.
4) We also cover a section on blind super resolution technique which more closely resembles the real life images that contains various noises and unpredictable down sampling kernels.
5) Finally, We discuss the current challenges and suggested future directions.

We also discuss how deep learning can be utilized to address the issue of image super-resolution after it has been mathematically defined. The remainder of the paper is organized into sections resembling how most supervised deep-learning models are laid out. Section II discusses the related work as per taxonomy depicted in Figure 2. The detailed challenges and research directions are discussed in section III based on the review of the current research. Section IV discusses the applications of image super-resolution. Finally, section V concludes the review with the various outcomes explored in this paper. All the acronyms used in this article are listed in the Appendix A along with their meaning.

II. RELATED WORK

Different upscaling methods, architectures, loss functions, and performance metrics are the focus of this study on image super-resolution. The detailed taxonomy of this related work is shown in Figure 2.

A. UPSCALING METHODS

Upscaling the input image to generate the super-resolution is the core component of image super-resolution. This section will discuss the upscaling mechanism used by various proposed works. Predominantly, either the low-resolution image can be upscaled to the desired dimensions before feeding it to the deep-learning model, or upscaling can be performed as a final layer of the deep-learning model. Section II-A1 belongs to the former whereas section II-A2 and section II-A3 fall under the latter category.

1) BICUBIC INTERPOLATION

Bicubic interpolation performs interpolation on the original pixel values to generate the I_{SR}. It is the successor to techniques like the nearest neighbor and bilinear interpolation, which were used previously. As it performs a cubic spline interpolation on both the image axis pixels in the enlarged image, it is denoted as bicubic interpolation.

Bicubic interpolation upscaling method was employed by a number of models as a preprocessing step on the I_{LR} prior to passing it through the network. The model has to learn the filters that fill in the missing details in the interpolated I_{LR} to produce its upscaled I_{SR} as shown in figure 3. The models discussing the usage of Bicubic Interpolation techniques for the task of enhancing the resolution of the image are discussed. Bicubic interpolation is fast, but since it performs a polynomial (cubic) fit on the pixels, it does not generate meaningful contextual details in the image. Apart from this, since the model now has to perform convolutions on a bigger image, the overall execution time of the model increases, which gets more prominent as the models get deeper. This makes the model heavy and computationally intensive, rendering it useless for production usage or video super-resolution.

Wang et al. [19] presented a sparse coding-based network to enhance the resolution of the image taken as an input. The advantage of this technique is its ability to learn a more complex regression function and thus can not be converted into an equivalent sparse coding model. On top of this, the network discussed in this work is also a CNN model used for patch extraction and reconstruction to give the best possible resolution. The CNN uses a LISTA sub-network specifically designed to enforce the sparse representation prior. This results in a lower training speed and performs better than a vanilla CNN. Kim et al. [20] proposed a very deep convolutional network inspired by a VGG-net image model used for ImageNet Classification. An increase in the network depth shows an improvement in the accuracy by

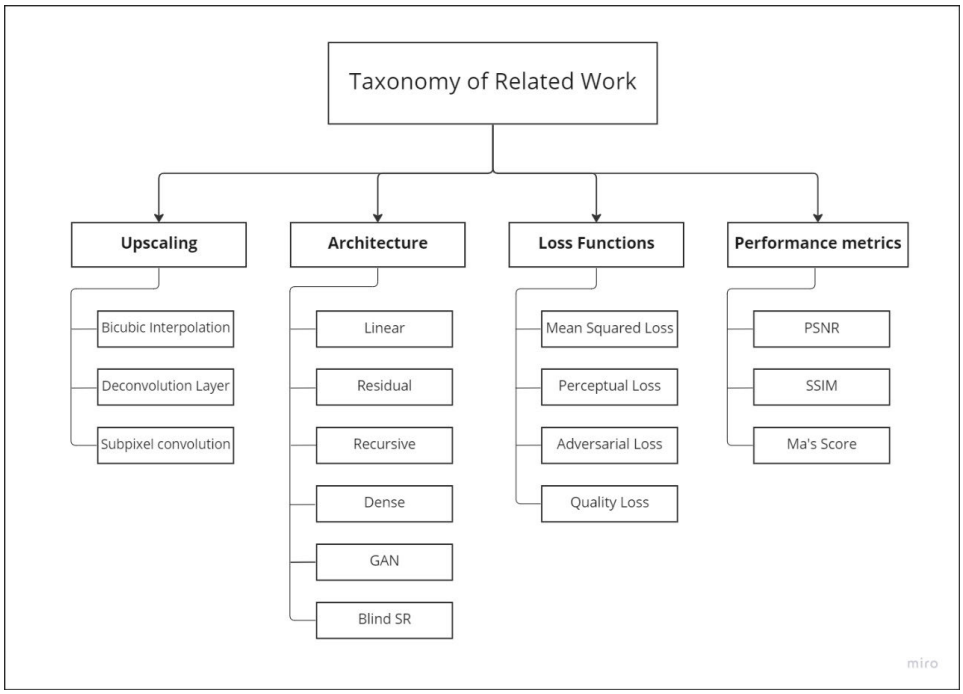

FIGURE 2. Taxonomy of related work.

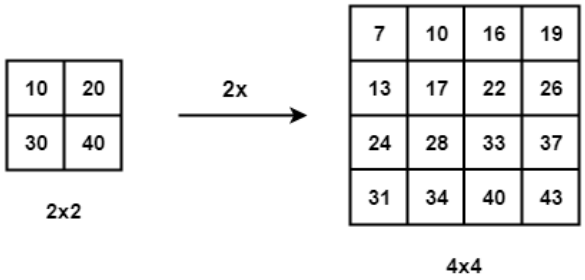

FIGURE 3. Bicubic interpolation.

using 20 weight layers. Only the residuals are learned during the training of this model, and the learning rates are very high compared to the SRCNN model. Tai et al. [21] discuss a memory-based approach for Image Restoration. The benefit of this network is that they explicitly propose a deep memory network (MemNet) that introduces a memory block to mine persistent memory through an adaptive learning process. The features are first extracted from the low-quality image. Then following the residual architecture, several memory blocks are stacked in a densely connected structure to solve the image restoration task.

2) DECONVOLUTION LAYER

The deconvolution layer is essentially a backward pass of the convolution layer, i.e., it produces the input which generates a given feature map [22]. In a deep-learning model, deconvolution layers can be positioned at the end to upscale the I_{LR} without increasing the convolutions needed to be performed, as discussed in bicubic interpolation. Deconvolution uses learnable filters, which are computed during backpropagation. This gives superior performance over the bicubic interpolation, which is rigid. Due to this, better performance can be seen in multiple models [23], [24], [25] which adopted it. Dong et al. [4] also updated their SRCNN model to benefit from the transpose convolution layer for faster execution in their work [26]. While deconvolution has some superiority over bicubic interpolation, it generates checkerboard artifacts in the resulting image. Checkerboard artifacts are netted block-like anomalies in images frequently seen in image super-resolution, as shown in figure 5. This is caused when the strides are not a factor of the filter size, causing overlapping in the output as shown in figure 4.

3) SUB-PIXEL CONVOLUTION LAYER

Transpose convolutions boosts the network effectiveness, but the checkerboard artifacts reduces the output fidelity. Shi et al. [29] developed a new technique for upscaling that addressed this issue by generating the I_{SR} by randomly rearranging the network's output features. They combined their method into a layer known as the sub-pixel convolution, also known as the pixel shuffler [30]. As can be seen in the figure 6, the layer produces S^2 feature maps by using a convolutional layer which are then reshuffled into the dimensions $SH * SW * C$ which is the required I_{SR}. Since no overlapping of the outputs occurs in the shuffle operation, the blocky checkerboard artifacts are no longer present. Many subsequent models utilized this approach for upscaling the feature maps into I_{SR}. For example, Shi, et al. [29] discuss about the

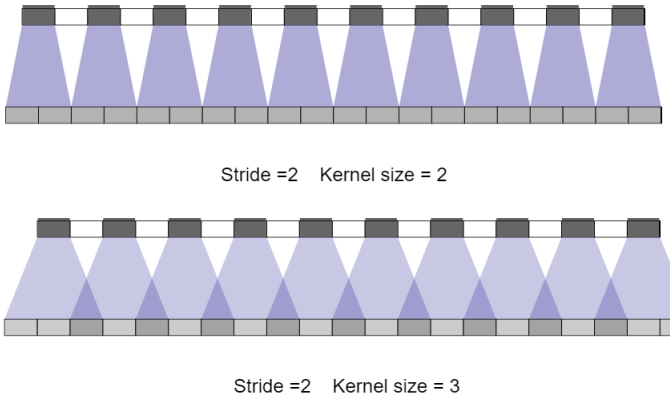

FIGURE 4. Checkerboard artifacts due to the stride=2 not being a multiple of kernel size=3 [27].

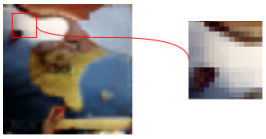

FIGURE 5. An example of checkerboard artifacts from dumoulin, et al. [28].

task of increasing the resolution of the image from LR to HR is performed at the very end of the network and super-resolve the HR data from LR feature maps. Subpixel convolution is used to achieve this at the last layer of the CNN. This HR image is then improved by deriving information from the feature maps which were learnt from the LR images in the previous layers. Zafeirouli et al. [31] proposed an efficient, lightweight framework that is built upon the recursive architecture. The structure is based on progressive reconstruction that strengthens the information flow by taking advantage of dense and residual connections. A coordinate convolution layer exploits the coordinate information allowing the network to learn the translation dependency required by the SR task. A sub-pixel convolution layer is put in place to convert the image from an LR to HR.

Zhang et al. [32] proposed a model in which each layer has direct access to the original LR input leading to deep supervision. This model uses a GRL approach to achieve image super-resolution. The upscaling technique in the second last layer comprises subpixel convolution layers. Shamsolmoali et al. [33] introduce a new approach to achieve the complex task of image super-resolution. The proposed approach is based on a progressive dilated convolution network for image super-resolution. A subpixel convolution layer is put in place to convert the LR to HR. The model also builds upon an efficient feature extraction and a feature reuse mechanism. The dense connections in place facilitate the reuse required to improve the model's performance.

B. NETWORK ARCHITECTURES

Network architecture decides how the feature maps interact with each other and hence plays a critical role in image processing as a whole. Multiple approaches have been proposed during the past half a decade which can be categorized based on some common strategy or principle followed. This section discusses the broad categories of these network architectures.

1) LINEAR NETWORKS

This type of architecture represents a rudimentary network model in which convolutional neural networks are stacked on top of one another. With their work in SRCNN [4] as shown in figure 7, Dong et al. [4] presented the pioneering work of deep-learning assisted image super-resolution. It proposes a three-layer convolutional neural network architecture that establishes an end-to-end mapping between low-resolution and high-resolution images.

There are other works such as SCN [19] and ESPCN [29] which are built upon Linear architecture. Wang et al. [19] improved the conventional SRCNN [4], where they have argued that human expertise in the field of super-resolution can be leveraged to improve the performance of neural networks. They improved the performance of the SRCNN network by initializing the parameters of the network based on their expertise. Their work introduces a Learned Iterative Shrinkage and Thresholding Algorithm (LISTA) network in addition to the neural network described in SRCNN. LISTA mimics the various stages of Sparse-coding and is used to enforce sparse representation within the network. Due to the simplicity of such networks, a deeper model is deemed impractical due to vanishing/exploding gradients. Apart from this, the model's convergence time also increases, making it hard to train intense models.

2) RESIDUAL NETWORKS

He et al. [34] introduced their work based on deep residual learning for image recognition. The deep learning community has widely adopted residual learning for creating deep models that can be trained easily. Residual learning consists of learning the difference between the input and the desired output rather than learning the desired output directly. In the

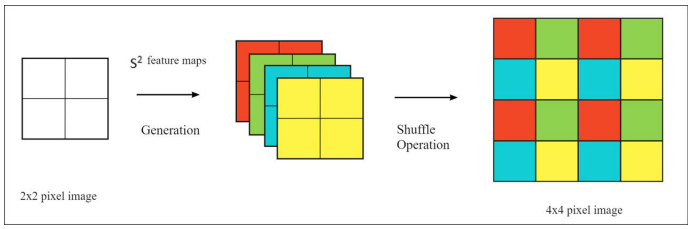

FIGURE 6. Working of sub-pixel convolution for an upscaling factor of 2 (s=2).

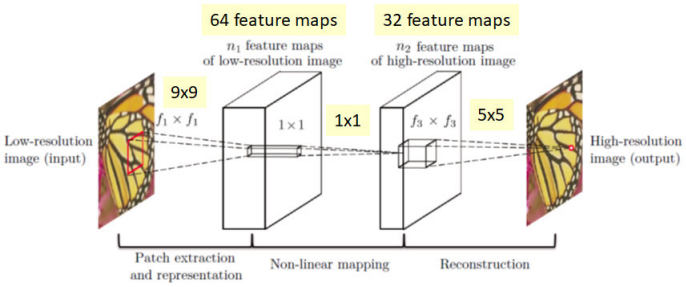

FIGURE 7. SRCNN architecture [4].

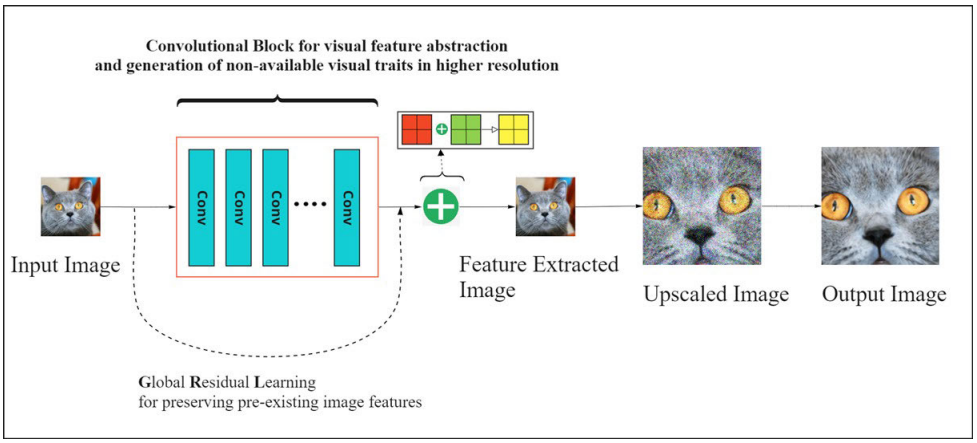

FIGURE 8. Global residual learning in VDSR [20].

context of image super-resolution, it consists of learning just the features in I_{HR} and combining those with the already available features from the input I_{LR}. Therefore, the residual between LR and HR can be represented as follows:

$$Residual = \{I_{HR} - I_{LR}\} \tag{3}$$

We relieve the network from learning the redundant information already in the I_{LR} by using residual learning. Since the network no longer has to carry the current information of the low-resolution image, which remains unchanged mainly during various convolutions, the network can focus on learning the residual details as denoted in Eq. 3. Residual learning has two primary types:

- Global Residual Learning (GRL) GRL performs an element-wise addition between the input I_{LR} and the feature maps of the final layer of the model. This provides a direct path for the existing I_{LR} features to be available at the end of the network without having to go through the actual model and perform futile computations. An example of this is Very Deep Super Resolution (VDSR) as shown in figure 8.
- Local Residual Learning (LRL) LRL performs element-wise addition of the feature maps between the various layers of the model. It differs from GRL in that the LRL connections exist within the various layers of the network, whereas GRL originates from the input image and ends in the final layer of the network. As a result, earlier feature maps of the model have higher activations than the latter part of the model. LRL keeps these activations alive by adding them to the latter feature maps. An example of this is shown in figure 9. Note that unlike GRL in figure 8, the skip connection

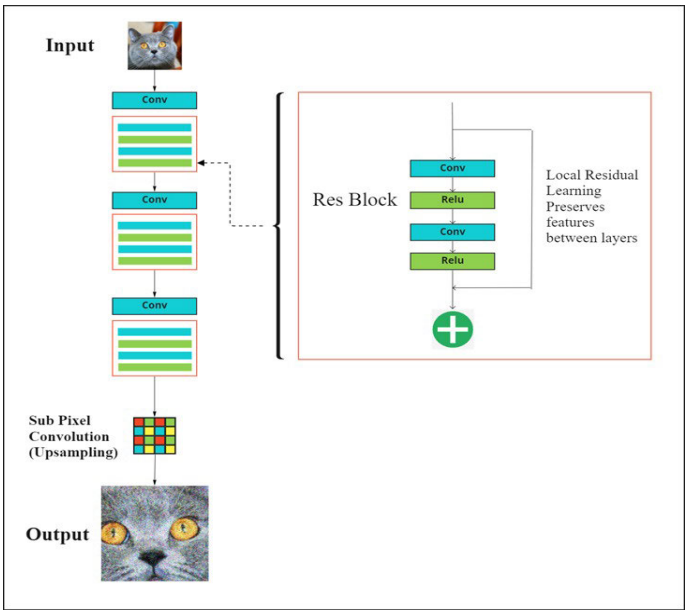

FIGURE 9. Local residual learning in EDSR [35].

passes into the next convolution block and not to the network's end.

Residual learning is also critical for training deep models since it mitigates the issues of vanishing/exploding gradients. Due to these reasons, it has been broadly used by many models such as SRGAN [30] and EDSR as shown in figure 9 [35]. EDSR improved over SRGAN and argued that the batch normalization layer reduces the range of the pixel values, which prohibits the network from learning textures accurately. It is also argued that batch normalization increases the model's memory usage since it consumes as much as its preceding convolutional layers.

3) RECURSIVE NETWORKS

Recursive networks emphasize an environment with reduced compute availability by introducing shared weights. Other networks emphasize increased perceptual quality at the expense of memory and processing. While this does come with an increase in image quality, such models are not feasible for mobile computing due to computation and memory constraints.

Kim et al. [37] approached this issue by making use of weight sharing amongst convolution layers in their model, DRCN. These layers, which share the weights amongst them, are arranged in blocks called recursive blocks that can be repeated any number of times; the more the block is, the better. This weight sharing decreases the size of the model and the flexibility in the number of blocks, which can be used as a trade-off between quality and performance. DRRN [36] improved on this by introducing more shared weights and using LRL within its network.

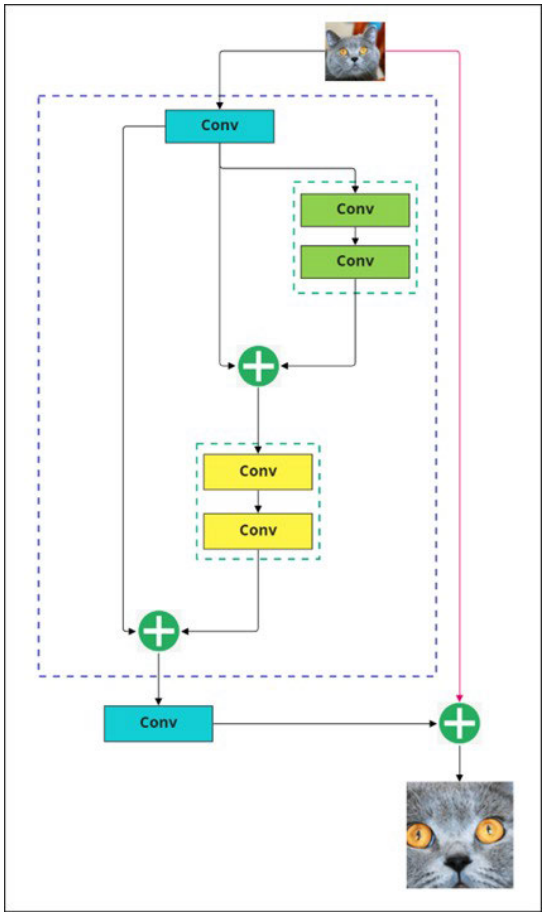

FIGURE 10. Two-tier recursive block in DRRN [36].

In figure 11, the dotted black line represents the global residual learning which carries the input to the end of the

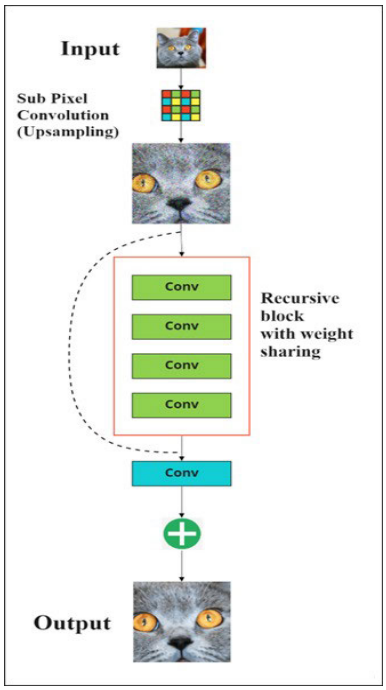

FIGURE 11. Linearly repetitive recursive block in DRCN [37].

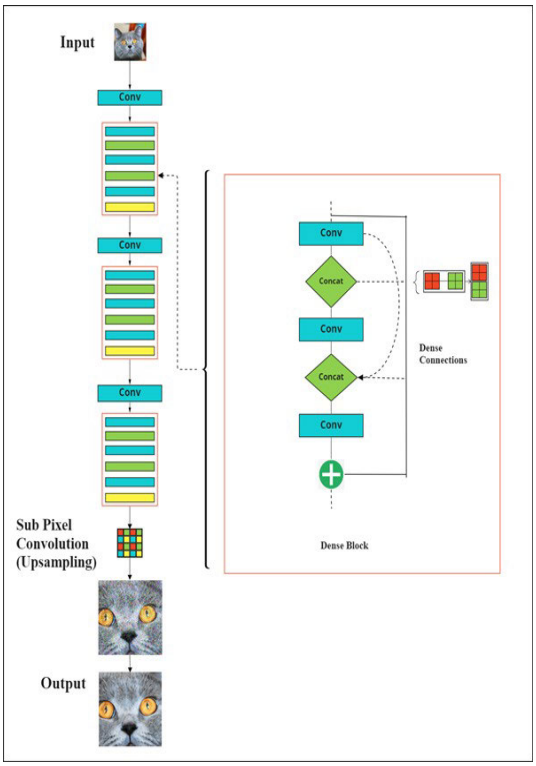

FIGURE 12. Dense connections in RDN [32].

network. The green box represents the recursive blocks repeated N times within the network. In figure 10, the black line represents the local residual learning which carries the feature maps between the convolution layers. The red line represents the global residual learning which carries the input to the end of the network. The layers with the same color share the same weights within a recursive block. For Example, the four convolution layers demonstrated in green color in a recursive block in DRCN have the same weights. Similarly, for DRRN, the two convolution layers represented by yellow color share the exact weights, and the same goes for the other two convolution layers in green. Reducing the parameter count also allows us to increase the depth of the network without worrying about the model size. A deeper network can learn high-level features, which ultimately improves network performance.

4) DENSE NETWORKS

Huang et al. [38] proposed the first dense structure in their DenseNet model. In Dense networks, the output of every convolution layer is concatenated with the output of all subsequent convolution layers. The final layer in a dense network outputs a feature map with reduced channels(3 in the case of RGB images), which curbs the channel growth. The above fundamental differences drastically change model retention and parameter count. The increased skip connections increase the network's bandwidth to retain more features across the convolutions, Which allows the network to perform inter-feature convolutions across the depth of the network. The reduced channel growth creates a substantially compact model from a performance perspective, allowing usage within memory-critical systems.

The fundamentals of DenseNet were implemented in single-image super-resolution by Tong et al. [24]. Zhang et al. [32] improved on dense connections by increasing the channel growth rate as seen in figure 12. Instead of adding the feature maps as is usually done in residual learning, the feature maps in dense connections are concatenated. Wang et al. [39] proposed a variant of dense connections that combined residual and dense connections inside a block in their work in ESRGAN. They coined this block as Residual-in-Residual Dense blocks (RRDB).

5) GENERATIVE ADVERSARIAL NETWORKS (GAN)

In general, a GAN model consists of 2 components. Generator and a Discriminator model. The Discriminator model aims to classify the output given by the Generator model as real or fake. The Generator gets feedback on loss from the Discriminator model and aims to minimize the loss. This repetitive cycle stops when a desired level of accuracy is achieved by the Generator model based on the feedback given by the Discriminator model. A review of several research work is already done that employed GAN-based techniques to bring out new methods to perform the crucial task of image resolution. Some of the methods are discussed below.

Dong et al. [40] discuss the image recovery problem in which the image is simultaneously corrupted by blur and impulse noise. The proposed model contains 3 terms:

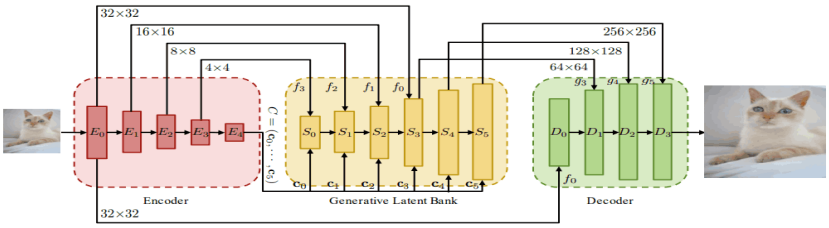

FIGURE 13. Glean architecture [42].

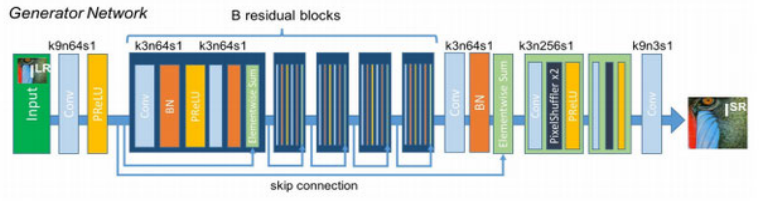

FIGURE 14. Resnet architecture for SRGAN generator [30].

sparse representation prior, total variation regularization, and data fidelity. To recover the original image, 2 steps process needs to be followed. The first step involves identifying the possible impulse noise positions, and the second is recovering the image via the patch-based model. He and Siu [41] presents an approach to image super-resolution without using any external training set. The proposed framework performs the magnification and de-blurring using only the original low-resolution image and its blurred version. In this method, each pixel is predicted by its neighbours through the Gaussian process regression. Discussed below are some GAN models commonly used to perform the task of Image Super-Resolution for varied scales.

- GLEAN - (Generative Latent Bank for Large-Factor Image Super-Resolution) [42] image super-resolution involves upscaling a low-resolution image to a desired high-resolution counterpart. While most GANs can effectively produce high-fidelity outputs for a smaller scale (2x, 3x, 4x), it fails to give a high-fidelity output when the scale factor increases beyond a particular scale. As a result, in higher scales, the GAN falls short of capturing the textures of the desired output, and the result is an upscaled image with just the core components of the image. This, however, is solved by using GLEAN (Generative Latent Bank For Large-Factor Image Super-Resolution). GLEAN applies the concept of using pre-trained GAN as a latent bank and can be employed as being used as an encoder-decoder architecture. The GLEAN architecture eases the process of training high-fidelity features by leveraging the results from a previously trained GAN. This pre-trained GAN already captures the rich texture before training the model. Unlike prevalent GAN inversion methods that require extensive image specific-optimization at runtime the GLEAN model only needs a single forward pass for restoration, resulting in high-fidelity outputs for a high-scale image super-resolution. Unlike conventional approaches, GAN is used here as an effective way of storing priors.
  Employing priors from different generative models allows GLEAN to be applied to diverse categories (human faces, cats, buildings, and cars etc.)
  The architecture for GLEAN model consists of essentially three main sections. 1) Encoder, 2) Generative Latent Bank, 3) Decoder following which the final output is given. The architecture for GLEAN is shown in figure 13. The encoder is used to extract features from the low resolution image. Further on the features extracted are analysed by passing them to a generative latent bank. The goal of the Generative-Latent Bank is to acquire useful knowledge regarded to the encoded features sent by the encoder. Once analysed and matched the features are then sent to the Decoder to convert them to a high resolution counterpart of the input low resolution image.
- SRGAN - SRGAN (Super Resolution Generative Adversarial Network) is used to upscale images to comparatively more minor scales than GLEAN. The generator network uses resnet architecture as shown in figure 9. Using a resnet architecture captures the more refined textures that are otherwise overlooked. Resnet architecture also proves effective against vanishing and diminishing gradient problems, allowing deeper models to be trained effectively. The resnet architecture used in SRGAN is shown in figure 14. The resolution of the image is increased by the two pixel shuffler layers (subpixel convolution). The generator also uses parametric RELU for more robustness as compared to the

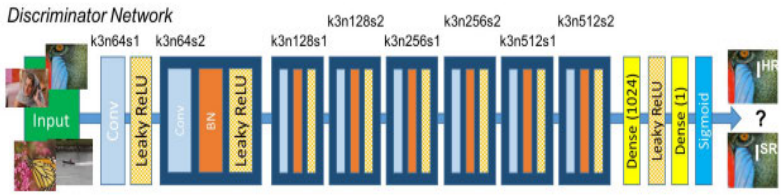

FIGURE 15. Resnet architecture for SRGAN discriminator [30].

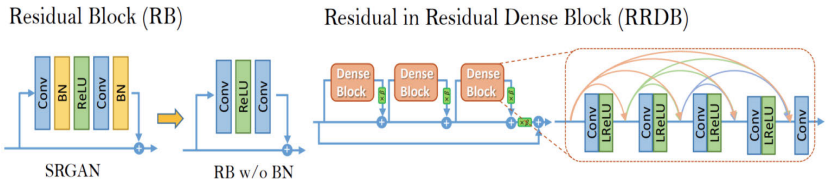

FIGURE 16. ESRGAN architecture [39].

LeakyRELU. The discriminator architecture consists of 8 convolutional layers (3 × 3 filters) which keeps on increasing from a factor of 2 to 512. Strided convolutions are put in place to reduce the image resolution each time the number of features taken into consideration are doubled.The discriminator utilized LeakyReLU as the activation function. The architecture is shown in figure 15. The generator used in SRGAN can also be used without a GAN architecture. The generator is known as the SRRESNET model. However, when the results were compared, it was found that the GAN architecture was more effective than the SRRESNET model. SRGAN uses perceptual loss function which is elaborated further in the II-C section.

- ESRGAN - (Enhanced Super Resolution Generative Adversarial Network) [39]. Building on the performance and the architecture of the SRGAN, ESRGAN's main aim is to reduce the complexity and the training time of the SRGAN, all the while increasing the model's performance. The changes made to the model were as follows.
  - The generator model now contains residual in residual networks, which leads to a better capture of the finer details that the normal residual networks may miss.
  - Batch normalization layers are removed while training the generator function.
  - The total loss to be considered now while training the generator function is the sum of GAN loss, perceptual loss, and the pixel-wise distance between the predicted images and the original low-resolution image [43].

  The architecture for the ESRGAN is as shown in figure 16. The loss function includes 3 components while training the generator function. Loss function is a linear function of Perceptual loss, Pixel wise absolute difference between real and fake image and Relativistic average loss between real and fake images during adversarial training. The results seemed to improve SRGAN's ability to improve the textures and reduce the training time to some extent.
- GMGAN - (GAN-Based Image Super-Resolution with a Novel Quality Loss) [44]. The SRGAN inspires the architecture of the GMGAN. To better the results produced by the SRGAN changes were made to the generator model.
  - The generator model now contains residual in residual networks, which leads to a better capture of the finer details that the normal residual networks may miss.
  - Batch normalization layers are removed while training the generator function.
  - The training instability of the generator model is now stabilized as the GAN is replaced by WGAN-GP to balance the training of the generator module.

  To gain better results while training the GAN architecture, a novel loss was introduced. This loss is known as Quality Loss. This loss is inspired by the gradient magnitude similarity deviation (GMSD) [43]. The results seemed to improve SRGAN's ability to improve the textures and reduce the training time to some extent.

6) BLIND SUPER RESOLUTION

There has been an increase in the performance of the DNNs recently. This change is attributed to change in the architecture and more resources being available to train the larger, complex architecture. But most of the DNNs method assume

that the blur kernel use is predefined as bi-cubic interpolation. However if we are to get even better results the blur kernels that are to be used for real life applications are much more complex in nature.Thus there should be more attention paid to the SR in the context of the blurring kernels put in place.

This type of approach is known as Blind SR [45], [46], [47]. This technique included one more undetermined variable i.e., the blur kernel (k) and the optimization also becomes more difficult.

- Practical Degradation Model for deep blind image super-resolution [48] - In this approach of blind SR several downsampling methods are chosen such as Nearest, Bilinear, Bicubic, Down-up. Based on the scale of the high-resolution image to be downscaled random downsampling methods are chosen from the mentioned choices. There can be various combinations possible. This leads to more robust and unpredictable blurring of a high resolution image to it low resolution counterpart [49], [50]. Further random noise can be injected to the low resolution image generated. This increases the complexity of the nature in which the low resolution images are obtained.
  Once low resolution images have been generated the training begins. For this paper, the model used has been borrowed from the classical ESRGAN model. The said model has been trained to identify the techniques used for blurring and give the best estimate i.e., the super resolved counterpart of the low resolution image. This process is more practical in real life applications where the conditions continuously keep changing and so does the blurring mechanism as well the noise associated with the image. This model BSRGAN has been able to outperform the classical model SRGAN. The PSNR/LPIPS ratio for the DIV2K4D dataset was monitored. The ESRGAN values were 23.68/0.599 while the BSRGAN model outperformed it by giving the metric of 24.58/0.361
- Real-ESRGAN - (Training real-world blind super-resolution with pure synthetic data) [51]. This approach is built upon classical ESRGAN model with some modifications to it. The degradation space considered here are several compression algorithms which can be due to imaging system of cameras, image editing and Internet transmission. As the degradation space is much larger than ESRGAN the training also becomes challenging. Improving the discriminator model a U-Net [52], [53] based architecture is put in place. This U-Net structure complicates the degradation process and increases the complexity in place. The model is named Real-ESRGAN since the training is done on real life images and is able to restore most real-world images achieving a better visual performance than the previous works. While BSRGAN improves the quality of the image by including robustness in training Real-SRGAN works better on real life images on which it has been trained on. The Real-ESRGAN model outperforms BSRGAN in terms restoring texture details and boosting visual sharpness.
- FeMaSR - (Real-World Blind Super-Resolution via Feature Matching) [54] The paper proposes a novel SR framework that outperforms Real-ESRGAN+. The framework here is based on the idea of feature matching. The model matches the LR features to a set of HR in the pre-trained implicit HR priors (HRP). The model has been inspired and taken from VQGAN. The HRP has been defined as the combination of a discrete codebook which consists of pre-defined number of feature vectors and the corresponding pre-trained decoder. The feature vectors then can be mapped to decode the targeted HR images. The task of SR is divided into 2 different parts. 1) Learning a high quality HRP 2) mapping the features to the codebook in the HRP for detailed recovery. To achieve this a novel framework named FeMaSR has been proposed for blind SR using the HRP encoded by a pre trained VQGAN network. This enables for better matching between the LR and the SR leading to the generation of more realistic images with less artifacts for the real world problem of Image Super Resolution. The proposed model outperformed the Real-ESRGAN model in terms of LPIPS (can better reflect texture quality). For the DIV2K Valid dataset the LPIPS score for the Real-ESGRAN model was 0.2993 and the score for the FeMaSR model was 0.2753.

C. LOSS FUNCTIONS

The loss function drives the back-propagation algorithm in any model, making it critical to select them according to the given task carefully. This section discusses the loss functions used by different approaches.

1) MEAN SQUARED LOSS

Mean square loss, also known as pixel loss, is the most common loss function used in deep learning. In the case of images, it finds the differences in the pixels of the predicted (I_{SR}) and ground-truth (I_{HR}) images. Mathematically, it is represented as

$$MSE = \frac{1}{W * H * C} \sum_{W*H*C} \{I_{HR}(W * H * C) - I_{SR}(W * H * C)\}^2. \quad (4)$$

Most models that use this loss function usually have a higher PSNR. However, this does not directly correlate to better image quality because, as can be seen from Eq. 5, PSNR and Mean squared error (MSE) have an inverse relationship, so as the loss decreases, the PSNR increases.

$$P.S.N.R. = 10 * log_{10}\left\{\frac{L^2}{MSE}\right\}. \quad (5)$$

Such a loss function restricts the optimization of the model since it is similar to the mean spatial filter with all weights

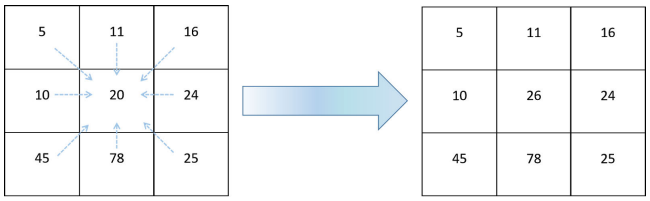

FIGURE 17. Mean filter.

as 1, shown in figure 17. The mean spatial filter applies a smoothing effect on the image by blurring the edges and textures, as it is an essential averaging operation. The same happens when we use pixel loss for learning in a network. While it reproduces the generic image details well, it cannot effectively generate high-frequency details.

2) PERCEPTUAL LOSS

The perceptual loss was first used for image super-resolution by Johnson et al. [25]. It works on the principle of transfer learning where we compare the values on a perceptual plane instead of comparing the pixel values as done in pixel loss in section II-C1. This is done by comparing the activation's of trained models such as VGG19 [55] by passing both the I_{HR} and I_{SR} to the model and then comparing the activation's on a particular layer. The difference in these activation's is then minimized, which is represented as follows:

$$Perceptual = \frac{1}{w*h*c} * \sum_{w*h*c} \left\{ \varphi^{i}_{HR}(w,h,c) - \varphi^{i}_{SR}(w,h,c) \right\}^2 \quad (6)$$

where, φ^{iSR} represents the i^{th} layer activation of pre-trained model on feeding I_{SR}.

φ^{iHR} represents the i^{th} layer activation of pre-trained model on feeding I_{HR}.

The perceptual loss uses pre-trained models already heavily trained for complex classification tasks such as the ImageNet [56]. Due to this, the layers of such models are adept at recognizing semantic differentiation, and they are more task-oriented for image super-resolution. For the sake of understanding, let us assume that a network reconstructs a perfect I_{SR}, which is highly similar to I_{HR}. However, the pixels of the I_{SR} are shifted one unit to the right compared to I_{HR}. This difference is insignificant to any observer but creates a tremendous loss value when we compute the pixel loss. This shows that the pixel loss is ill-suited for model training and deems PSNR as an ineffective performance metric. This also indicates that attention should be given to the image features instead of focusing on pixel comparison.

3) ADVERSARIAL LOSS

Ledig et al. [30] pioneered using GAN architecture for single image super-resolution. Their model utilized a discriminator that discriminates I_{HR} and I_{SR}. The loss comprises the loss of two networks, the generator, and the discriminator.

$$Adversarial_{Gen} = -log\left(Discrim\left(I_{SR}\right)\right) \quad (7)$$

$$Adversarial_{Discrim} = -log\left(Discrim\left(I_{HR}\right)\right) - log\left(1 - Discrim\left(I_{SR}\right)\right) \quad (8)$$

Introducing adversarial loss motivates the model to create better images to convince the discriminator. The discriminator, in turn, gets trained to better differentiate between I_{HR} and I_{SR}. This cycle improves the model as a whole. Recently, Jolicoeur-Martineau [57] proposed the relativistic discriminator, which improves on the original discriminator. The previous discriminator measured the absolute probability of the input belonging to the training set. This sets a limitation on the network where the adversarial loss depends on the training dataset. Using a relativistic discriminator solves this issue by redefining the loss as a difference in the probabilities of the image being authentic and the image being fake. Mathematically, the difference between the two kinds of generator can be depicted as follows

$$D\left(\overline{x}\right) = sigmoid\left(C\left(x_r\right) - C\left(x_f\right)\right) \quad (9)$$

$$D\left(x\right) = sigmoid\left(C\left(x\right)\right) \quad (10)$$

where,

$C(x)$ is a function which assigns a score to an image regarding its fakeness or realness where,

x_r represents the real image ($I_{HR} insuper - resolution$).

x_f represents the fake image ($I_{SR} insuper - resolution$).

4) QUALITY LOSS

Inspired by a image quality assessment (IQA) metric, a new loss was proposed for the GMGAN model, known as quality loss. MSE, which can be considered equivalent to PSNR, cannot capture the perceptual loss, leading to the need to develop other losses. The metric chosen for this loss is gradient magnitude similarity deviation (GMSD), which effectively can capture the perceptual loss from the output. The process of calculating GMSD takes place in a major three step process. Here h_x, and h_y stand for the Prewitt filters in the horizontal, and the vertical direction. The gradient magnitudes of I_{SR}, and I_{HR} at location i, denoted as $m_{SR}(i)$, and $m_{HR}(i)$ are calculated as follows.

$$m_{SR}(i) = \sqrt{(I_{SR} convolution h_x)^2(i) + (I_{SR} \otimes h_y)^2(i)} \quad (11)$$

$$m_{HR}(i) = \sqrt{(I_{HR} \otimes h_x)^2(i) + (I_{HR} \otimes h_y)^2(i)} \quad (12)$$

where, $\otimes$ symbol denotes the convolution operation. Then the gradient magnitude similarity (GMS), map is calculated as

$$GMS(i) = \frac{2m_{SR}(i).m_{HR}(i) + c}{m^2_{SR}(i) + m^2_{HR}(i) + c} \quad (13)$$

Here c depicts a positive constant. For here, the LQM of the I_{SR} has been acquired. Further, on average, pooling is applied to the GMS map to obtain GMSM.

$$GMSM = \frac{1}{N}\sum_{i=1}^{N} GMS(i) \quad (14)$$

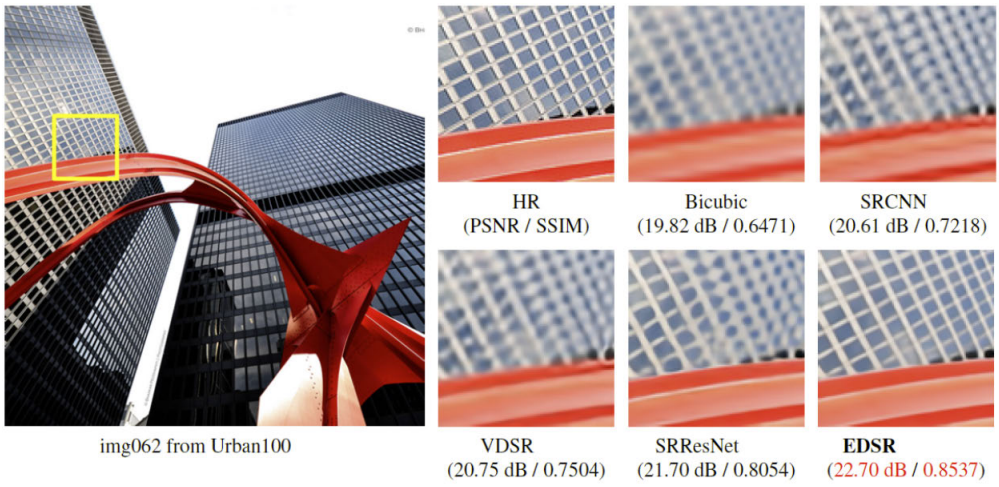

FIGURE 18. Model comparison on the test img062 image from Urban100 dataset.

Here N would refer to the number of pixels in I_{SR}. Since the average pooling offers little insight into the local quality degradation, the standard deviation of the GMS map is calculated to be the final IQA metric.

$$GMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(GMS(i) - GMSM)^2} \tag{15}$$

The final GMSD can now be used to reflect the distortion severity of the image in question, which can be used to optimize the training process. Therefore the quality loss in question can be defined as

$$I_Q = GMSD(G_\theta(I_{LR}), (I_{HR})) \tag{16}$$

A higher GMSD score depicts a higher distortion level. In practice, a single GMSD loss is calculated for each channel of I_{SR} and I_{HR} (R, G, B); After that, the three losses are combined to provide the final distortion score. There is a significant difference in how the loss function is used in backtracking and model training. While ESRGAN uses the sum of three different loss functions while training (GAN loss, perceptual loss, and pixel-wise loss); GMGAN, on the other hand, introduces a new loss function named Quality Loss that is used to capture the distortion score between the input image and the super resolved image.

D. PERFORMANCE METRICS

Evaluation of any model is paramount to determine its performance quantitatively against other models. However, since image perception is subjective, it isn't easy to analyze the performance of a super-resolution model. In this section, we will discuss the various approaches presented to quantify the perceptuality of an image and, thus, the performance of super-resolution models. Table 1 and Table 2 presents the performance based on two of the metrics mentioned below. The models have increasing complexity as we move from left to right in the tables, going from linear models on the left to residual and dense networks on the right, with the rightmost models (ESRGAN, SRGAN) also utilizing adversarial and perceptual loss. A sample image comparison has also been provided for the benchmark of the various models in Figure 18.

1) PSNR

Peak Signal-to-Noise Ratio is one of the most commonly used metrics for quality measurement in image processing and is defined as

$$P.S.N.R. = 10 * log_{10}\left\{\frac{L^2}{M.S.E.}\right\} \tag{17}$$

where, L represents the range of pixel values. However, as already discussed in the mean squared error in subsection II-C1, the PSNR is ineffective for measuring image super-resolution effectively since it does not consider perceptual image features. Table-1 shows the PSNR values for the different image super-resolution networks discussed in this paper. It can be generally noted from the given data that apart from networks using perceptual loss (SRGAN, ESRGAN), dense and residual models easily outperformed the linear models, with the performance gap increasing with higher upscaling. This can be attributed to the fact that as we grow the upscaling, it becomes increasingly difficult for models with lower feature retention to perform sufficient detail regeneration. As discussed, PSNR is very sensitive to per-pixel changes, which causes unstable metric values in the table. Therefore, it isn't easy to draw out long-running trends for PSNR since the metric has a high variance.

2) SSIM

The shortcomings of PSNR, such as its absoluteness, are improved upon by Structural Similarity Index (SSIM) [60] which considers image degradation as a change in the perceptuality by taking into consideration such luminance masking and contrast masking. It is usually calculated by a sliding

TABLE 1. PSNR benchmark matrix of datasets versus models.

DATASET	Scale	Bicubic	SRCNN	DRRN	SRDenseNet	RDN	EDSR	DRCN	SCN	REDNET	SRGAN	VDSR	ESRGAN
SET5 [1]	x2	33.66	36.66	37.74	-	38.24	38.11	37.63	37.14	37.66	-	37.53	-
	x3	30.39	32.75	34.03	-	34.71	34.66	33.82	33.26	33.82	-	33.66	-
	x4	28.42	30.48	31.68	32.02	32.47	32.5	31.53	31.04	31.51	29.4	31.35	32.73
SET14 [58]	x2	30.24	32.45	33.23	-	34.01	33.85	33.04	32.71	32.94	-	33.03	-
	x3	27.55	29.3	29.96	-	30.57	30.44	29.76	29.55	29.61	-	29.77	-
	x4	26	27.5	28.21	28.5	28.81	28.72	28.02	27.76	27.86	26.02	28.01	28.99
BSD100 [59]	x2	29.56	31.36	32.05	-	32.34	32.29	31.85	31.54	31.99	-	31.9	-
	x3	27.21	28.41	28.95	-	29.26	29.25	28.8	28.58	28.93	-	28.82	-
	x4	25.96	26.9	27.38	27.53	27.72	27.72	27.23	27.11	27.4	25.16	27.29	27.85
Urban100 [12]	x2	26.88	29.5	31.23	-	32.89	32.84	30.75	-	-	-	30.76	-
	x3	24.46	26.24	27.53	-	28.8	28.79	27.15	-	-	-	27.14	-
	x4	23.14	24.52	25.44	26.05	26.61	26.67	25.14	-	-	-	25.18	27.03

TABLE 2. SSIM benchmark matrix of datasets versus models.

DATASET	Scale	Bicubic	SRCNN	DRRN	SRDenseNet	RDN	EDSR	DRCN	SCN	REDNET	SRGAN	VDSR	ESRGAN
SET5 [1]	x2	0.9299	0.9542	0.9591	-	0.9614	0.9602	0.9588	0.9567	0.9599	-	0.9587	-
	x3	0.8682	0.909	0.9244	-	0.9296	0.928	0.9226	0.9167	0.923	-	0.9213	-
	x4	0.8104	0.8628	0.8888	0.8934	0.899	0.8973	0.8854	0.8775	0.8869	0.8472	0.8838	0.9011
SET14 [58]	x2	0.8688	0.9067	0.9136	-	0.9212	0.9198	0.9118	0.9095	0.9144	-	0.9124	-
	x3	0.7742	0.8215	0.8349	-	0.8468	0.8452	0.8311	0.8271	0.8341	-	0.8314	-
	x4	0.7027	0.7513	0.7721	0.7782	0.7871	0.7857	0.767	0.762	0.7718	0.7397	0.7674	0.7917
BSD100 [59]	x2	0.8431	0.8879	0.8973	-	0.9017	0.9007	0.8942	0.8908	0.8974	-	0.896	-
	x3	0.7385	0.7863	0.8004	-	0.8093	0.8091	0.7963	0.791	0.7994	-	0.7976	-
	x4	0.6675	0.7101	0.7284	0.7337	0.7419	0.7418	0.7233	0.7191	0.729	0.6688	0.7251	0.7455
Urban100 [12]	x2	0.8403	0.8946	0.9188	-	0.9353	0.9347	0.9133	-	-	-	0.914	-
	x3	0.7349	0.7989	0.8378	-	0.8653	0.8655	0.8276	-	-	-	0.8279	-
	x4	0.6577	0.7221	0.7638	0.7819	0.8028	0.8041	0.751	-	-	-	0.7524	0.8153

Gaussian window of dimensions $11 * 11$ according to the following equation:

$$SSIM = \frac{\left(2\mu_{I_{HR}}\mu_{I_{SR}} + O_1\right)\left(2\zeta_{I_{HR}}\zeta_{I_{SR}} + O_2\right)}{\left(\mu_{I_{HR}}^2 + \mu_{I_{SR}}^2 + O_1\right)\left(\sigma_{I_{HR}}^2 + \sigma_{I_{SR}}^2 + O_2\right)} \quad (18)$$

where,

μ represents Average across image patch.
ρ represents variance across image patch.
ζ represents covariance across image patch.
O_1 represents $(k_1 l_1)^2$.
O_2 represents $(k_2 l_2)^2$.
k_1, k_2 represents constants.
l_1, l_2 represents dynamic range of pixel values.

As seen from the Eq. 18, the averaging and the variance considers the nearby pixels to determine their similarity rather than the per-pixel similarity as calculated in PSNR. Table-2 shows the SSIM values for the different Image super-resolution networks discussed in this paper. Comparing the various networks shows that the dense networks (VDSR, ESRGAN) easily outperform the other networks, especially on the higher scales. It can also be noted that the difference in the scores of the models is more noticeable and consistent than PSNR since SSIM looks at the structure of the entire image rather than per-pixel similarity. Because of this, SSIM has a higher correlation with the perceived image quality than PSNR. Another point of interest is that all the networks trained on perceptual loss (ESRGAN, SRGAN) have consistently better SSIM values, validating that perceptual loss is a better network learning driver for image super-resolution.

The values for PSNR and the SSIM metrics are acquired from various papers written in the past addressing the issue of single image super-resolution. The dataset for which the values are written is uniform across all the papers considered while performing the scoring metrics. While performing the study, the scaling factors are 2x, 3x, and 4x. An interesting thing to note is the PSNR and the SSIM scores for the GAN-based models such as ESRGAN and SRGAN. It can be seen that the scores for the GAN-based models come into the picture only when the desired high-resolution image has a scale factor of 4x. The values are not available for the lower scale factors. This highlights that GAN models being more complex, are viable to use only when the desired high-resolution image is of a comparatively larger scale. Given the complexity of the GAN models, the time required for the GAN models outweigh the performance of the model for lower scales.

3) MA's SCORE

Ma et al. [61] proposed a score for evaluating the perceptual quality of the outputs of super-resolution models. Based on a large mean opinion score (MOS) test on the BSD dataset [59] of over nine models, they compiled a regression model which predicts a score based on various features of the image. Their seminal work was used as the benchmark test at the PIRM-SR 2018 challenge [62] to measure the performance of the multiple models submitted by the participants.

III. CHALLENGES AND RESEARCH DIRECTIONS

In this section, we discuss the latest developments in the field of image super-resolution, which has been proposed quite recently. However, they have not been put into the above taxonomy as these are singular, and not enough literature exists to warrant a separate category for the same.

- Coordinate convolution layer - While convolution layers are used in neural networks frequently, they work on a positional-invariant filter, i.e., the pixels can be

rearranged and still produce the same output for the same filters. This positional independence is essential for image processing tasks such as classification or segmentation. However, in the image super-resolution task, the spatial position information could lead to more efficient representations, especially for edges. To counter this problem, Liu et al. [63] introduced coordinate convolution layer in their work. This spatial information of the pixels has been explicitly accounted for by Zafeirouli et . [31] in their work using coordinate convolution layer. Their model shows that using a coordinate convolution layer can produce more refined details and better image reproduction.

- Importance of receptive field - While most of the research has focused on the depth of neural networks, recent work by Shamsolmoali et al. [33] has shown that the receptive field of the network layer is one of the most important parameters for image super-resolution. In this work, two models were compared with the same receptive field but varying network depth, and found that such models have similar performance. However, when the authors compared two models with the same network depth but varying receptive fields, they achieved higher performance, fewer parameters, and less execution time as the receptive field increased. Their work also proposes the use of a progressively dilating convolution layers block. Since dilated convolution layers have a higher receptive field for the same filter size as a convolution layer, such networks can speed up the execution of the network model.

In this section, we discuss the parts of the super-resolution domain that require attention and pinpoint specific areas of deep networks that can be investigated further. This section aims to provide a future research direction for the community to improve image super-resolution further.

- The performance metrics of any experiment are a fundamental aspect of analyzing the effectiveness of the proposed approaches. PSNR does not evaluate the perceptual quality of the image, and while SSIM does lean towards perceptual consideration, it is not good enough. Blind evaluation techniques such as Ma's score [61] need to be researched as they do not require a reference image for scoring. The need for a reference image is a limiting factor for evaluation since such images are not always available in real-time scenarios.
- Loss functions are the principle drivers of neural networks since they are responsible for parameter optimization. Perceptual-oriented loss functions need to be developed to drive the models to create realistic images. There is no best suitable loss function defined for image super-resolution. Perceptual loss as described in [25], [30], and [39] is a promising start in the direction of finding better loss functions.
- Various convolution layer varieties have been proposed, such as coordinate convolution layer [63] and dilated convolutions for image super-resolution. However, their various combinations in a neural network have not been explored. Imbalanced convolutions with unequal filter sizes (such as a filter of 1×3) and dilated convolutions reduce the computational cost, whereas coordinate convolutions add two extra positional channels. Research needs to be done to combine these layers in different combinations to determine what gives a superior performance.
- The work done by Shamsolmoali et al. [33] argues that the receptive field has a more significant impact on image quality than network depth. Research needs to be done in this direction to find the optimal network depth size and receptive field beyond which we get diminishing returns. A comparative study is required to understand the impact on super-resolution with minor parameter changes. Such analysis would enable the creation of compact models which can perform well at lower parameter counts.
- As models get more profound and their receptive fields wider, the parameters increase significantly. While this leads to better performance, it deems the models impractical for real-life scenarios such as smartphones or other mobile systems. A quick look at the PIRM-SR 2018 challenge [62] and NTIRE-SR2017 challenge [64] shows that most of the proposed methods with higher perceptual quality take a long time to process a single image that too on powerful hardware such as the NVIDIA TitanX. Such computationally expensive models do not make sense for real-life scenarios. Research needs to be done to optimize neural networks to reduce computational costs and their memory footprint to increase their applicability in less powerful hardware configurations.

  Apart from model optimizations, we also suggest using lightweight cloud computing architectures with load balancing algorithms [65] for super-resolution tasks for low-end systems. The input image can be divided into multiple parts, and each part can then be processed on cloud engines for faster computations.

FIGURE 19. Surveillance.

IV. USE CASES

Although image super-resolution has dramatically improved, it cannot be justified as necessary while there are no practical uses. This section explains how image super-resolution is helpful in several fields and how it is advancing many frontiers. Additionally, Dai et al. [66] demonstrated how image

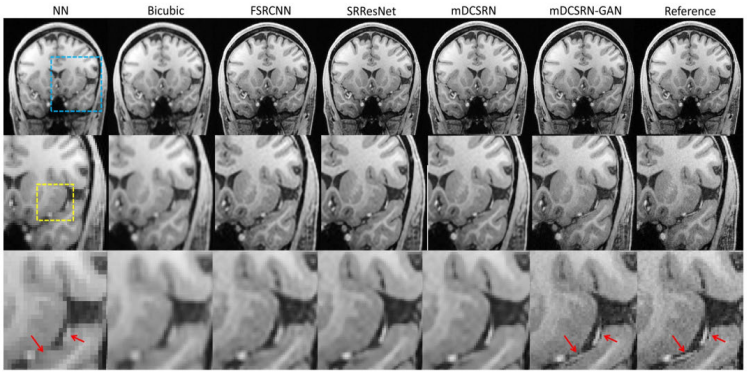

FIGURE 20. Medical image for MRI.

super-resolution enhances the efficiency of numerous image-processing tasks, including segmentation and classification.

A. HYPERSPECTRAL IMAGE SUPER-RESOLUTION

Hyperspectral images captured by Synthetic Aperture Radars (SAR) satellites such as the one described by Patel et al. [67] are critical for objectives such as area surveillance, crop health monitoring, and ecological analysis. The complexity of SAR satellites in terms of their weight and hardware depends on the precise resolution required for their output. These images are scaled to a resolution of several meters, represented by a single pixel. Due to this, their spatial resolution is of utmost importance. The concept of single image super-resolution is adopted for hyperspectral images to reduce the hardware requirements and stay within weight constraints in microwave remote-sensing. Yuan et al. [68] have proposed using a CNN model inspired by SRCNN to improve the spatial resolution of a given benchmark of a hyper-spectral image. Mei et al. [69] further proposed using a 3D convolutional layer to fully exploit the correlations between the spatial band and the pixels of the neighboring bands for better high-resolution image reproduction.

B. SURVEILLANCE MONITORING

CCTVs are a crucial part of today's security systems monitoring various areas. Many TV shows and movies show sci-fi scenes where a detective goes through surveillance footage and asks the operator to zoom in and ''enhance'' the blurred part of an image, revealing some clue in their investigation. While previously, such magical enhancement of low-quality images was a part of science fiction, mage super-resolution has made them a reality. Work by Rasti et al. [70] uses a model inspired by SRCNN [4] to improve facial recognition in low-quality images considerably as shown in figure 19. Such systems can be a handy tool for law enforcement agencies and security companies.

Apart from this, CCTVs are commonplace at crossroads to track vehicles and apply penalties and fines if the drivers are not driving correctly. In such applications, super-resolution proves helpful in processing the surveillance records of the systems. For example, license-Plate recognition for identifying vehicles and their owners can be improved by performing super-resolution as a pre-processing task in [71]. Furthermore, by increasing image resolution, Optical Character Recognition (OCR) tasks get more accurate, which is critical in license plate OCR, or the system can end up penalizing an innocent citizen. Apart from this, super-resolution can also be applied to surveillance records, as shown in [72] in which the authors use a multi-frame CNN model for image super-resolution to enhance the records.

Besides CCTVs, image super-resolution can be easily ported for wireless multimedia sensors such as RGB cameras. However, such systems have a limited resolution capability to preserve bandwidth and power usage. In such cases, image super-resolution can exponentially reduce hardware costs.

C. MEDICAL IMAGING

Magnetic Resonance Imaging (MRI) is a class of medical imaging techniques that is a critical resource used extensively for cancer diagnosis, brain abnormalities, and detecting and monitoring Alzheimer's and Parkinson's disease as shown in figure 20 MRI uses radio waves and strong magnetic fields to scan the targeted parts of the human body and generates detailed images of the body structures, such as organs and tissues. However, high-resolution MRI scans require longer scan times and reduce the spatial area that can be covered. Due to these reasons, it is critical to identify methods to generate higher resolution images without compromising on these factors. In addition, traditional methods such as bi-cubic interpolation cannot be used to upscale MRI scans effectively, leading to image blurring and even loss of details. The principles of image super-resolution address the shortcomings mentioned above. As a result, they can generate high-quality MRI scans without needing any special hardware equipment or increasing scan times.

Chen et al. [73] have used 3D convolutions with dense connections for image super-resolution that utilize the details from the neighboring scan slices to create super-resolution MRI images. Their work considerably outperforms the traditional methods. Since MRI scans are a particular niche, it is not always possible to get the high-resolution and low-resolution scan sequences required to scan neural networks.

The author in [74] uses a self super-resolution algorithm in which the MRI scans are degraded along the x-axis to generate the training dataset.

V. CONCLUSION

This paper survey focuses on major architectural elements of a network with single-image super-resolution. In particular, we have discussed various upscaling tools, network designs, loss functions created for model training, and current assessment metrics applied to multiple models. In addition, we measured the PSNR and the SSIM of these models and evaluated them using different datasets and scale factors. These results show a significant improvement in PSNR/Perceptual loss metric compared to traditional methods like normal CNN, interpolation, bi-cubic interpolation, etc. These findings have also been well-discussed in the corresponding assessment metric section. We also discussed several image super-resolution applications, including security monitoring and medical imaging, demonstrating that it is a critical domain that requires investigation and development. Finally, the paper advocates the future research directions that should be investigated to advance single-image super-resolution.

APPENDIX A

Acronym	Meaning
$\otimes$	Convolution Operation
I_{LR}	Represents the input image to the model to be upscaled and super-resolved
I_{HR}	Represents the higher resolution counterpart of the input image (used for training model)
I_{SR}	Represents the super-resolved image as outputted by the model
ϑ^{-1}	Inverse Operation to map I_{LR} to I_{HR} and vice versa
DRCN	Deep Recursive Convolutional Network
IQA	Image Quality Assessment
GRL	Global Residual Learning
LRL	Local Residual Learning
EDSR	Enhanced Deep Residual Network for Super Image Resolution
GAN	Generative Adversarial Network
SRGAN	Super Resolution Generative Adversarial Network
GLEAN	Generative Latent Bank for Large Factor Image Super Resolution
GMSM	Gradient Magnitude Similarity Mean
GMSD	Gradient Magnitude Similarity Deviation (Loss Function)
GMGAN	Generative Adversarial Network-based Image Super-Resolution with a Novel Quality Loss
ESRGAN	Enhanced Super Resolution Generative Adversarial Network
SRCNN	Super-Resolution Convolutional Neural Network
EDSR	Enhanced Deep Residual Network for Super Image Resolution
DRRN	Deep Recursive Residual Network
VDSR	Very Deep Super Resolution
BSRGAN	Blind SRGAN
LPIPS	Learned Perceptual Image Patch Similarity
FeMaSR	Feature Matching Super Resolution
HRP	High Resolution Priors
VQGAN	Vector Quantized Generative Adversarial Network

REFERENCES

[1] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.* London, U.K.: British Machine Vision Association, 2012, pp. 135.1–135.10, doi: 10.5244/C.26.135.

[2] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4544–4556, Nov. 2012.

[3] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3194–3205, Jul. 2012.

[4] 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. 2015.

[5] Y. Li, H. Huang, L. Jia, H. Fan, and S. Liu, "D2C-SR: A divergence to convergence approach for real-world image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2021, pp. 379–394.

[6] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Oct. 2016.

[8] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, *arXiv:1508.06576*.

[9] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 341–349.

[10] K. Chauhan, S. Jani, D. Thakkar, R. Dave, J. Bhatia, S. Tanwar, and M. S. Obaidat, "Automated machine learning: The new wave of machine learning," in *Proc. 2nd Int. Conf. Innov. Mech. Ind. Appl. (ICIMIA)*, Mar. 2020, pp. 205–212.

[11] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[12] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.

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

[14] K. Chauhan, H. Patel, R. Dave, J. Bhatia, and M. Kumhar, "Advances in single image super-resolution: A deep learning perspective," in *Proc. 1st Int. Conf. Comput., Commun., Cyber-Secur.*, P. K. Singh, W. Pawłowski, S. Tanwar, N. Kumar, J. J. P. C. Rodrigues, and M. S. Obaidat, Eds. Singapore: Springer, 2020, pp. 443–455.

[15] W. Yang, X. Zhang, Y. Tian, W. Wang, and J. Xue, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimedia*, vol. 21, no. 12, pp. 3106–3121, Dec. 2019.

[16] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sheriff, and C. Zhu, "Real-world single image super-resolution: A brief review," *Inf. Fusion*, vol. 79, pp. 124–145, Mar. 2022.

[17] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, May 2021.

[18] 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.

[19] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 370–378.

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

[21] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4539–4547.

[22] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. CVPR*, vol. 10, 2010, p. 7.

[23] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2802–2810.

[24] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4799–4807.

[25] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.

[26] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 391–407.

[27] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, p. 1–3, Oct. 2016.

[28] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," 2016, *arXiv:1606.00704*.

[29] W. Shi, J. Caballero, F. Huszar, 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. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.

[30] 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.

[31] K. Zafeirouli, A. Dimou, A. Axenopoulos, and P. Daras, "Efficient, lightweight, coordinate-based network for image super resolution," in *Proc. IEEE Int. Conf. Eng., Technol. Innov. (ICE/ITMC)*, Jun. 2019, pp. 1–9.

[32] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[33] P. Shamsolmoali, X. Li, and R. Wang, "Single image resolution enhancement by efficient dilated densely connected residual network," *Signal Process., Image Commun.*, vol. 79, pp. 13–23, Nov. 2019.

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

[35] 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. 136–144.

[36] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3147–3155.

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

[38] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[39] 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 *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 1–16.

[40] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[41] H. He and W.-C. Siu, "Single image super-resolution using Gaussian process regression," in *Proc. CVPR*, Jun. 2011, pp. 449–456.

[42] K. C. K. Chan, X. Wang, X. Xu, J. Gu, and C. C. Loy, "GLEAN: Generative latent bank for large-factor image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2020, pp. 14245–14254.

[43] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 684–695, Feb. 2014.

[44] X. Zhu, L. Zhang, L. Zhang, X. Liu, Y. Shen, and S. Zhao, "Generative adversarial network-based image super-resolution with a novel quality loss," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS)*, Dec. 2019, pp. 1–2.

[45] 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.

[46] Z. Luo, Y. Huang, S. Li, L. Wang, and T. Tan, "Unfolding the alternating optimization for blind super resolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 5632–5643.

[47] Y. Zhang, L. Dong, H. Yang, L. Qing, X. He, and H. Chen, "Weakly-supervised contrastive learning-based implicit degradation modeling for blind image super-resolution," *Knowl.-Based Syst.*, vol. 249, Aug. 2022, Art. no. 108984.

[48] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 4791–4800.

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

[50] C. Mou, Y. Wu, X. Wang, C. Dong, J. Zhang, and Y. Shan, "Metric learning based interactive modulation for real-world super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 723–740.

[51] 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. (ICCV) Workshop*, Oct. 2021, pp. 1905–1914.

[52] E. Schonfeld, B. Schiele, and A. Khoreva, "A U-Net based discriminator for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8204–8213.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[54] C. Chen, X. Shi, Y. Qin, X. Li, X. Han, T. Yang, and S. Guo, "Real-world blind super-resolution via feature matching with implicit high-resolution priors," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 1329–1338.

[55] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[56] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

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

[58] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surf.* Cham, Switzerland: Springer, 2010, pp. 711–730.

[59] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, Feb. 2001, pp. 416–423.

[60] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[61] 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.

[62] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 1–22.

[63] R. Liu, J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the CoordConv solution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9605–9616.

[64] 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.

[65] J. Bhatia and M. Kumhar, "Perspective study on load balancing paradigms in cloud computing," *Int. J. Comput. Sci. Commun.*, vol. 6, no. 1, pp. 112–120, 2015.

[66] D. Dai, Y. Wang, Y. Chen, and L. Van Gool, "Is image super-resolution helpful for other vision tasks?" in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9.

[67] H. Patel, G. Seth, and S. Trivedi, "Design and implementation of automated rotatory probe embedded system for measuring axial ratio," in *Proc. URSI Asia–Pacific Radio Sci. Conf.*, Mar. 2019, pp. 1–6.

[68] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.

[69] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, p. 1139, 2017.
[70] P. Rasti, T. Uiboupin, S. Escalera, and G. Anbarjafari, "Convolutional neural network super resolution for face recognition in surveillance monitoring," in *Proc. Int. Conf. Articulated Motion Deformable Objects*. Cham, Switzerland: Springer, 2016, pp. 175–184.
[71] Y. Yang, P. Bi, and Y. Liu, "License plate image super-resolution based on convolutional neural network," in *Proc. IEEE 3rd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2018, pp. 723–727.
[72] P. Shamsolmoali, M. Zareapoor, D. K. Jain, V. K. Jain, and J. Yang, "Deep convolution network for surveillance records super-resolution," *Multimedia Tools Appl.*, vol. 78, pp. 1–15, Apr. 2018.
[73] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, "Brain MRI super resolution using 3D deep densely connected neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 739–742.
[74] C. Zhao, A. Carass, B. E. Dewey, and J. L. Prince, "Self super-resolution for magnetic resonance images using deep networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, Apr. 2018, pp. 365–368.
[75] M. F. Mridha, A. A. Lima, K. Nur, S. C. Das, M. Hasan, and M. M. Kabir, "A survey of automatic text summarization: Progress, process and challenges," *IEEE Access*, vol. 9, pp. 156043–156070, 2021.
[76] A. Brifman, Y. Romano, and M. Elad, "Unified single-image and video super-resolution via denoising algorithms," 2018, *arXiv:1810.01938*.
[77] P. Shamsolmoali, M. Zareapoor, R. Wang, D. K. Jain, and J. Yang, "G-GANISR: Gradual generative adversarial network for image super resolution," *Neurocomputing*, vol. 366, pp. 140–153, Nov. 2019.
[78] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 606–615.
[79] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4491–4500.
[80] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6228–6237.
[81] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," 2016, *arXiv:1605.09782*.
[82] 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.*, Jul. 2021, pp. 457–466.
[83] Z. Zhang, P. Favaro, Y. Tian, and J. Li, "Learn to zoom in single image super-resolution," *IEEE Signal Process. Lett.*, vol. 29, pp. 1237–1241, 2022.
[84] D. Song, Y. Wang, H. Chen, C. Xu, C. Xu, and D. Tao, "AdderSR: Towards energy efficient image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15648–15657.
[85] Y. R. Musunuri and O.-S. Kwon, "Deep residual dense network for single image super-resolution," *Electronics*, vol. 10, no. 5, p. 555, Feb. 2021.

KARANSINGH CHAUHAN received the B.Tech. degree from Gujarat Technological University, in 2020. He is currently pursuing the M.S. degree in applied computer science with Dalhousie University, Halifax, Canada. He is currently a Technical Lead with OneForce. He has over two years of experience in senior positions in the industry and works across software engineering and machine learning domains. He was a recipient of the Best Interdisciplinary Project in 2019 by SSIP, Gujarat Government, and received funding for his undergraduate project from the state government. His main research interests include computer vision, image processing, and natural language processing. He has been a Guest Lecturer with DA-IICT, Gujarat, India, and the University of Waterloo, Canada. He has published/presented a number of articles and book chapters at reputed international and national level journals and conferences, such as IEEE and Springer.

SHAIL NIMISH PATEL is currently pursuing the B.Tech. degree in computer science and engineering with a minor degree in design with Ahmedabad University, Ahmedabad. He is currently a machine learning researcher to help and develop novel contextual-based sentimental analysis techniques. His research interests include deep learning, natural language processing, and autonomous vehicular systems. He is currently studying the impact of artificial intelligence in the field of physical training and gyms to help develop an autonomous system for tracking a person's physical progress. He has learnt and implemented various deep learning frameworks. This is his first coauthored paper in the field of deep learning.

MALARAM KUMHAR received the B.E. degree in computer engineering from Rajasthan University, Jaipur, in 2006, and the M.Tech. degree in computer science and engineering from Nirma University, Ahmedabad, India. He is currently pursuing the Ph.D. degree with Gujarat Technological University (GTU), Ahmedabad. He is also an Assistant Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University. He has more than 16 years of teaching experience. His research interests include the multimedia Internet of Things and wireless multimedia sensor networks.

JITENDRA BHATIA received the B.Tech. degree from Hemchandra Acharya North Gujarat University, in 2004, and the M.Tech. degree in computer science and engineering and the Ph.D. degree in vehicular ad hoc networks from the Institute of Technology, Nirma University, Ahmedabad, Gujarat, India, in 2012 and 2019, respectively. He is currently an Associate Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University. He has over 17 years of experience in teaching for undergraduate and postgraduate levels. He has published/presented a number of papers at reputed international and national level journals and conferences. He has authored or coauthored more than 40 technical research papers, published in leading journals and conferences from the IEEE, Elsevier, Springer, and Wiley. Some of his research findings are published in top cited journals, such as IEEE Internet of Things Journal, *Computer Communications* (Elsevier), *Peer-to-Peer Network Applications* (Springer), and *IJCS* (Wiley). His main research interests include protocol design for intelligent transportation systems, software-defined networking, vehicular networks, and cloud computing. He was a recipient of the "Best Idea2Innovate Award" in 2019 by Brihasti (Claris) Foundation. He was the session chair and a keynote speaker at various international conferences.

SUDEEP TANWAR (Senior Member, IEEE) is currently a Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University, India. He is also a Visiting Professor with Jan Wyzykowski University, Polkowice, Poland, and the University of Pitesti, Pitesti, Romania. He has authored two books, edited 13 books, and more than 270 technical papers, including top journals and conferences, such as IEEE Transactions on Network Science and Engineering, IEEE Transactions on Vehicular Technology, IEEE Transactions on Industrial Informatics, IEEE Wireless Communications, IEEE Network, ICC, GLOBECOM, and INFOCOM. He initiated the

research field of blockchain technology adoption in various verticals, in 2017. His H-index is 50. He actively serves his research communities in various roles. His research interests include blockchain technology, wireless sensor networks, fog computing, smart grids, and the IoT. He is a member of the Technical Committee on Tactile Internet of the IEEE Communication Society. He is also a Senior Member of CSI, IAENG, ISTE, and CSTA. He has received the Best Research Paper Awards from IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019. He has served many international conferences as a member of the organizing committee, such as the Publication Chair for FTNCT-2020, ICCIC 2020, and WiMob2019; a member of the Advisory Board for ICACCT-2021 and ICACI 2020; the Workshop Co-Chair for CIS 2021; and the General Chair for IC4S 2019 and 2020 and ICCSDF 2020. He is serving on the editorial boards for *Frontiers of Blockchain*, *Cyber Security and Applications*, *Computer Communications*, the *International Journal of Communication Systems*, and *Security and Privacy*.

INNOCENT EWEAN DAVIDSON (Senior Member, IEEE) received the B.Sc. (Eng.) (Hons) and M.Sc. (Eng.) degrees in Electrical Engineering from the University of Ilorin, in 1984 and 1987, respectively, the Ph.D. degree in electrical engineering from the University of Cape Town, in 1998; and the PG Diploma degree in Business Management, from the University of KwaZulu-Natal, in 2004. He also received the Associate Certificate in Sustainable Energy Management (SEMAC), from the British Columbia Institute of Technology, Burnaby, BC, Canada, in 2011, and the Course Certificate in Artificial Intelligence, from the University of California at Berkeley, USA in 2020. He was a Full Professor and Chair of the Department of Electrical Power Engineering; Research Leader of the Smart Grid Research Centre, and Program Manager of the DUT-DSI Space Science and CNS Research Program, Durban University of Technology (DUT), South Africa from 2016 to 2022. Currently, he is a Full Professor and Director, with the African Space Innovation Center (ASIC), and French South African Institute of Technology (F'SATI), Cape Peninsula University of Technology (CPUT), Bellville, South Africa. He has supervised five postdoctoral research fellows, and graduated 55 Ph.D./masters students and over 1200 engineers, technologists, and technicians. He is the author/co-author of over 350 technical papers in accredited journals, and peer-reviewed conference proceedings. He has managed over US$3 million in research funds. His current research interests include HVdc power transmission, grid integration of renewable energy, applied artificial Intelligence, and space technology. He is a fellow of the Institute of Engineering and Technology, U.K., and the South African Institute of Electrical Engineers; a Chartered Engineer in the U.K.; and a registered Professional Engineer (P Eng.), of the Engineering Council of South Africa. He is a member: Western Canada Group of Chartered Engineers (WCGCE); the Institute of Engineering and Technology (IET Canada) British Columbia Chapter; IEEE Collabratec Communities on Smart Cities and IEEE (South Africa Chapter). He was the General Chair of the 30th IEEE Southern Africa Universities Power Engineering Conference in 2022. He is the Host and Convener of the DSI-DUT-SANSA-ATNS Space Science and CNS Symposium, and Guest Speaker in several forums, including the Science Forum of South Africa and the International Conference on Sustainable Development. He was a recipient of numerous international Best Paper Awards, and from DUT's Annual Research and Innovation. He is a C2-rated researcher from the National Research Foundation (NRF), South Africa.

THOKOZILE F. MAZIBUKO received the National Diploma degree in engineering (electrical) from the Durban University of Technology (DUT), in 2008, the Bachelor of Technology degree in electrical from the Tshwane University of Technology (TUT), in 2011, and the master's degree (cum laude) in electrical engineering from TUT and RWTH Aachen University, Germany, in 2014. She is currently pursuing the Ph.D. degree with DUT. She served as a Tutor/a Students Assistant with DUT, from 2006 to 2007; and an Engineering Trainee with Anglo Platinum from 2007 to 2008. She joined Anglo American/Hatch in Rustenburg, South Africa, as a Planner Assistant and a Quality Control Coordinator from 2008 to 2009. She has been conducting research and implementation of real-time platforms, such as the application of PTP synchronized PMUs in power system small signal stability and transient stability analysis of a multi-machine system based on synchro-phasors. She was with the Council for Scientific and Industrial Research (CSIR), Pretoria, until 2016, and joined Rand Water, Johannesburg, from 2017 to 2018. She was a Lecturer with the University of Johannesburg, from 2018 to 2020, and joined DUT in January 2021, as a Lecturer. Her research interests include smart microgrids, network optimization, control, and applied artificial intelligence.

RAVI SHARMA is currently a Professor with the Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun, India. He is passionate in the field of business analytics and worked in various MNCs as a leader of various software development groups. He has contributed various articles in the area of business analytics, prototype building for a startup, and artificial intelligence. He is a consultant for leading academic institutions to uplift research activities in inter-disciplinary domains.