



## Image super-resolution: A comprehensive review, recent trends, challenges and applications



Dawa Chyophel Lepcha <sup>a</sup>, Bhawna Goyal <sup>a,\*</sup>, Ayush Dogra <sup>b</sup>, Vishal Goyal <sup>c</sup>

<sup>a</sup> Department of ECE, Chandigarh University, Mohali, Punjab 140413, India

<sup>b</sup> Ronin Institute, Montclair, NJ 07043, USA

<sup>c</sup> Department of ECE, GLA University, Mathura, India

### ARTICLE INFO

#### Keywords:

Low resolution (LR)  
High resolution (HR)  
Deep learning  
Convolutional neural network (CNN)  
Generative adversarial network (GAN)  
Super-resolution (SR)  
Image quality assessment (IQA)  
Learning strategies  
Survey

### ABSTRACT

Super resolution (SR) is an eminent system in the field of computer vision and image processing to improve the visual perception of the poor-quality images. The key objective of image super resolution is to address the limitations of imaging systems mainly due to hardware problems and requirements for clinical processing of medical imaging using post-processing operations. Numerous super resolution strategies have been put-forward in the computer vision community to improve and achieve high-resolution images over the years. In the past few years, there has been a significant advancement in image super-resolution algorithms. This paper aims to provide the detailed survey on recent advancements in image super-resolution in terms of traditional, deep learning and the latest transformer-based algorithms. The in-depth taxonomy of broadly classified super-resolution techniques within these categories has been broadly discussed. An extensive survey has been carried out on deep learning techniques in terms of parameters, architecture, network complexity, depth, learning rate, framework, optimization, and loss function. Furthermore, we also address some of the significant parameters such as problem definition, evaluation metrics, publicly benchmarks datasets, loss functions and applications. In addition, we have performed an experimental analysis and comparison of various benchmark algorithms on publicly available datasets both qualitatively and quantitatively. Lastly, we conclude our survey by emphasizing some of the prospective future directions and open issues that the community need to address in the future.

### 1. Introduction

In recent years, an importance of image super resolution has earned significant attention from the scientific fraternity to reconstruct the good resolution of the images. It is intended to transform a blurred, fuzzy, indistinct low-resolution images having unrefined details into an apparent, clear high-resolution images along with improved visual perception and enhanced details. Alternative terms i.e., interpolation, image scaling, enlargement and upscaling are generally related to image super resolution. A single image or multiple images may be used to create a higher resolution raster image. Because of empirical assessments, this illustration concentrates primarily on the single image super resolution since it has been researched extensively because of its difficult environment. There are number of seminal works reported in literature [1–7] for super resolution (SR) in large dimension inputs i.e., videos and 3D scans. High-resolution images produce enhanced information of locations and inherent objects which is important in numerous

applications i.e., high-definition TV sets, bigger computer screens, portable devices such as cameras, laptops, mobile phones etc. Super resolution plays an important role in other fields such as the identification of objects in the locations [8], detecting small objects [9], face identification in surveillance videos [10], medical diagnostics [11], improved analysis in remote sensing images [12], astronomy [13] and forensics [14].

Reconstruction techniques for SR have become acknowledged in the medical image resolution improvement community. For accurate disease diagnosis, descriptions of small anatomical landmarks and pathologies are vital. Small changes in the microvasculature around the tumor are the important biomarker for the diagnosis of cancer [15] and non-apparent soft exudates are important pathologies for the diagnosis of retinal disease [16]. However, due to imaging equipment and imaging parameters (e.g., scanning time for CT images), many real medical images suffer from poor spatial resolution. Such poor medical image resolution prevents minor anatomical locations and pathologies from being

\* Corresponding author.

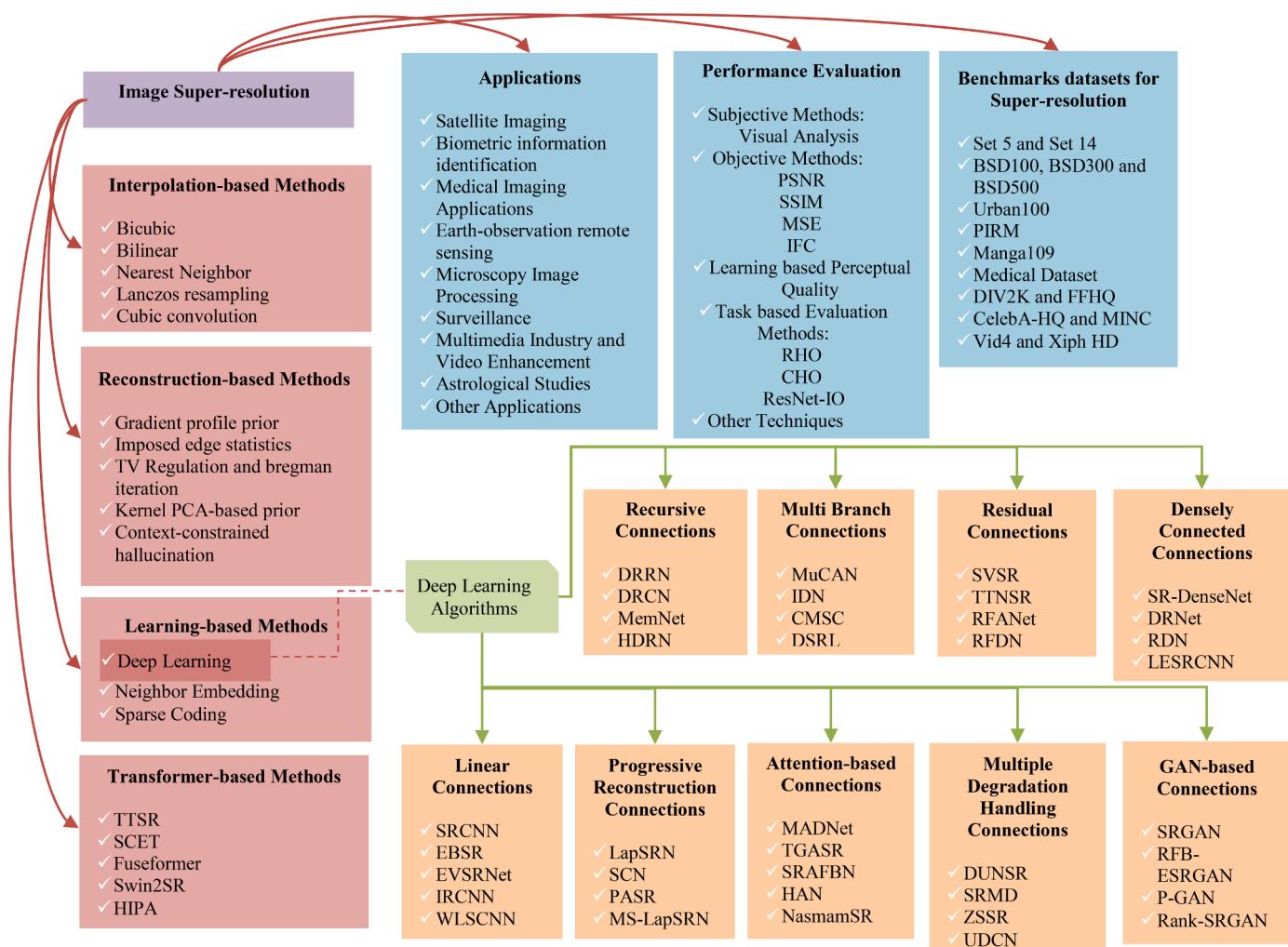
E-mail address: [bhawnagoyal28@gmail.com](mailto:bhawnagoyal28@gmail.com) (B. Goyal).

correctly identified or segmented and impedes the accurate detection of some of the serious diseases at an early stage. The most significant techniques to disease diagnosis are X-ray, computed tomography (CT) scans, etc. However, once the radiation exposure increase, X-ray/CT scans will increase the risk of cancer to the patients [17]. In addition, reduction of the radiation doses increases the effect of noise on reconstruction images which can compromise the quality of the diagnosis. Significant number of works has been carried out over the years to improve the resolution of medical images. These techniques enhance the resolution of images such as simple bicubic interpolation and their variant, but they typically suffer from a significant loss of sharp-edged information and high local contrast enhancement [18]. The reconstruction of the SR allows doctors to clearly observe the possibility of diagnosing and curing the diseases. Therefore, super resolution is currently playing a pivotal role for reconstruction of poor-quality images in numerous computer vision community. Super-resolution is considered as the significant benefits for several reasons yet considered as a difficult and available research challenges in computer vision. First, SR is a difficult ill posed problem which tries to get the HR output from the corresponding LR input. However, once the scaling factor starts to increase, complexity for the problem rises. A recovery of lost scene become much highly complicated with high scaling factor and thus often times leads to replication of improper information. In addition, quality of performance measurement still unclear since the quantitative matrices such as peak-signal-to-noise ratio (PSNR), structural similarity (SSIM) and root mean squared error (RMSE) are still more or less corresponded

with human perceptive.

The main contributions of this paper are as follows:

- 1) Taxonomy of broadly classified super resolution approaches in four categories have been presented that include reconstruction-based, interpolation-based, learning-based algorithms and transformer-based algorithms.
- 2) A comparative analysis has been performed between the algorithms in terms of architecture, network complexity, number of epochs, dataset, learning rate, filter size, optimizer, depth, loss function, parameters.
- 3) An extensive comparative review has been performed on SR algorithms based on deep learning algorithms
- 4) The study illustrated some of the most widely used datasets, objective metrics, SR competitions, applications and learning strategies in detail.
- 5) This paper demonstrates the performance evaluation of some of the SR algorithms i.e., experimental results is presented (both subjective and objective) of the state-of-the-art SR algorithms on six widely used datasets.
- 6) Experimental evaluation demonstrates that there is a steady and rapid growth in accuracy of algorithms over the last couple of years in addition to corresponding increase in computational complexity and deep algorithms.
- 7) Lastly some of the potential research directions has been discussed in detail.



**Fig. 1.** Taxonomy of the existing state-of-the-art super resolution techniques.

It is indicated that the recent algorithms substantially outperformed the pioneering methods outlined as the benchmark for image super-resolution. Instead, the success in recent years, many limitations of recent algorithms have been found and potential research directions have been given to solve the prevailing open challenges. In the following sections, this paper cover numerous parameters of existing super-resolution techniques on both traditional, deep learning and transformer-based approaches. The in-depth taxonomy of super resolution discussed in the survey is shown in Fig. 1 in the hierarchically ordered manner. Section 2 defines the statement and examines the most widely used datasets, evaluation metrics and SR competitions. The third section discusses the comprehensive details on state-of-the-art techniques and learning strategies. Experimental details along with qualitative and quantitative analysis has been presented in Section 4. In addition, some well-known applications of SR have been discussed in Section 4, while Section 5 discusses the conclusion and the future directions.

## 2. Problem definition and terminology

### 2.1. Problem statement

Image super-resolution is a classic ill posed inverse problem [19], that can be denoted by the observation model:

$$y = Hx \quad (1)$$

where  $x$  indicates an unknown HR image,  $y$  indicates an observed LR image. The  $H$  matrix represents an imaging system that consists of many operations such as blurring, warping, and down-sampling. In order to obtain an approximation  $\hat{x}$  to the unknown HR image  $x$  using SR reconstruction to solve the above-mentioned inverse problem. Since the problem is underdetermined, infinitely many HR images satisfy the above reconstruction constraint. For this reason, different methods have been suggested to use various priors to regularize the SR problem:

$$\hat{x} = \|y - Hx\| + \lambda E(x) \quad (2)$$

where  $E(x)$  is the prior of the HR image. An initial expression in Eq. (2) represents a fidelity of the information and the second expression represents a desired HR image property. A regularization parameter  $\lambda$  represents a tradeoff among data closeness and prior to the HR image.

It is possible to implicitly or explicitly to define the prior of HR images. Tikhonov regularization [20] and TV regularization [21] are the two renowned appropriate regularization approaches such as letting  $E(x) = \|Cx\|_2^2$  or  $\|Cx\|_1$  in which  $C$  represents Laplacian function. Though, these techniques based on predefined priors merely lead to achieve adequate performance. This is mainly because it is difficult to recognize what and how much priors should be utilized for super-resolution problem beforehand. Most SR methods limit the techniques to mitigate and resolve a problem having robust prior knowledge that can be learned through learning from the collection of datasets using different methods of regularization by considering a certain distribution of image properties [22].

### 2.2. Image quality assessment

Image quality is a term that represents to the visual properties of images and targets on perception of the observers. Image quality assessment (IQA) technique in general incorporate qualitative techniques based on human perception such as how realistic the image appears and quantitative experimental techniques. The former is more in accordance with the requirements however it is often time consuming and costly therefore the latter is now the standard. These techniques still are not always consistent with one another since objective techniques are often not able to adequately capture the human visual perception resulting in considerable differences in IQA results [23,24]. Furthermore, there are three types of objective IQA methods [24]: full reference

techniques that carry out assessment utilizing reference images, reduced reference technique related to comparing extracted features and non-reference technique such as blind IQA that do not use any reference images. Furthermore, we will be introducing different mostly utilized IQA techniques including both qualitative and quantitative techniques.

#### 2.2.1. Performance evaluation metrics

To quantify the reconstruction performance and compare it to other reconstructions approaches, RMSE (root mean square error), PSNR (peak signal to noise ration) and SSIM (structure similarity index measure) and Information fidelity criterion (IFC) have been utilized as metrics in literature [25,26], which are specified as

$$\left\{ \begin{array}{l} RMSE = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N [Y(i,j) - X(i,j)]^2}{MN}} \\ PSNR = 10 \lg \frac{255^2 MN}{\sum_{i=1}^M \sum_{j=1}^N [Y(i,j) - X(i,j)]^2} \\ SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_x\sigma_y + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \begin{cases} c_1 = (k_1 L)^2 \\ c_2 = (k_2 L)^2 \end{cases} \end{array} \right. \quad (3)$$

where  $\mu_x, \mu_y$  are mean values and  $\sigma_x, \sigma_y$  are standard variance image  $X$  and  $Y$ , respectively.  $M$  and  $N$  is the image size of row and column, respectively.  $c_1, c_2$  are two stabilizing constants that describe the dynamics of a pixel value, with  $k_1$  and  $k_2$  are constants set to 0.01 and 0.03, respectively.

**Information fidelity criterion (IFC):** IFC is the mutual details between the input and distorted image of the given statistical model for the input and the distortion (channel). Where it first obtains mutual details for one sub-band initially then generalize to multiple sub-bands.

Let  $N$  elements from  $C$  are denoted by  $C^N = (C_1, C_2, \dots, C_N)$ . In this portion, it assumes that the underlying  $RF_\mu$  is uncorrelated (and thus  $C$  is a RF with conditionality independent elements of given  $S$ ) and distortion model parameters  $G$  and  $\sigma_V^2$  are terms as *a priori*. Let  $D^N = (D_1, D_2, \dots, D_N)$  indicate the corresponding  $N$  elements from  $D$ . As  $I(C^N; D^N)$  denotes the mutual information between them. It is significantly simple to analyse the mutual details assuming  $S$  is known because of nonlinear dependence among the  $C^N$  by way of  $S$ . This conditioning "tunes" the GSM model for the particular reference image and thus, models the input more specifically. Thus, the IFC introduces in this paper is the conditional mutual information  $I(C^N; D^N | S^N = s^N)$ , where  $S^N = (S_1, S_2, \dots, S_N)$  are the corresponding  $N$  elements of  $S$  and  $s^N$  represents a realisation of  $S^N$ . Furthermore  $I(C^N; D^N | S^N = s^N)$  denoted by  $I(C^N; D^N | s^N)$ . With the stated assumptions on  $C$  and distortion model [26], one can denote:

$$I(C^N; D^N | s^N) = \sum_{j=1}^N \sum_{i=1}^N I(C_i; D_j | C^{i-1}, D^{j-1}, s^N) \quad (4)$$

$$= \sum_{i=1}^N I(C_i; D_i | C^{i-1}, D^{i-1}, s^N) \quad (5)$$

$$= \sum_{i=1}^N I(C_i; D_i | s_i) \quad (6)$$

where Eq. (4) is obtained using the chain rule [27] and Eqs. (5) and (6) are obtained using conditional independent of  $C$  given  $S$ , independent of noise  $V$ , the fact where the distortion model  $D_i$  keeps independent of  $C_j$ ,  $\forall i \neq j$  and that given  $S_i, C_i$  and  $D_i$  are independent of  $S_j$ ,  $\forall i \neq j$ . From given  $S_i$ ,  $C_i$  is Gaussian and  $V_i$  are also gaussian with variance  $\sigma_V^2$ , we get:

$$I(C^N; D^N s^N = \sum_{i=1}^N I(C_i; D_i s_i) \quad (7)$$

$$= \sum_{i=1}^N (h(D_i s_i) - h(D_i C_i, s_i)) \quad (8)$$

$$= \sum_{i=1}^N (h(g_i C_i + V_i s_i) - h(V_i)) \quad (9)$$

$$= \frac{1}{2} \sum_{i=1}^N \log_2 \left( 1 + \frac{g_i^2 s_i^2 \sigma_U^2}{\sigma_V^2} \right) \quad (10)$$

where  $h(X)$  represents differential entropy of the continuous random variable  $X$  and  $h(X) = 1/2 \log_2 2\pi e \sigma^2$  [27] for  $X$  distributed as  $N(\mu, \sigma^2)$ . Eq. (10) was obtained for one sub-band. Using separate GSM RFs for modelling each sub-band of interest in the image is simple. The RF modelling wavelet coefficients of reference image in the  $k$ th sub-band can be denoted as  $C^k$  and the test (distorted) image will be denoted as  $D^k$  and assume that  $C^k$  being independent of each other. Then assume that each sub-band is distorted on its own. Thus, the RFs  $V^k$  are also independent of each other. *IFC* is then calculated by adding all of the sub-bands together:

$$IFC = \sum_{K \in \text{subbands}} I(C^{N_k, k}; D^{N_k, k} | s^{N_k, k}) \quad (11)$$

where  $N_k$  coefficients from the RF  $C^k$  of the  $k$ th sub-band are denoted by  $C^{N_k, k}$  and similarly for  $D^{N_k, k}$  and  $s^{N_k, k}$ . The *IFC* in Eq. (11) which measures the statistical information share between the input and distorted images. The important features of the criterion like mean square error and some other computational fidelity indexes, it does not necessitate parameters based on display device physics, data from visual psychology experiments, viewing configuration details or stabilizing constants that determine the adequacy of HSV-based FR QA techniques i.e., some structural ones too. However, *IFC* does not need any training dataset. But, once Eq. (11) is implemented, some implementation parameters will inevitably arise. The *IFC* is a fidelity criterion not a distortion metric. It can theoretically be anywhere between zero (non-fidelity) to infinity (perfectly fidelity within a nonzero multiplicative constant in the absent of noise).

### 2.2.2. Mean opinion score

The mean opinion score (MOS) is often utilized qualitative IQA technique in which human raters are quizzed to evaluate perceptual quality score to test images. The scores usually range from 1 (poor) to 5 (excellent). The arithmetic mean overall ratings are utilized to determine the final MOS. Although MOS testing appears to be a reliable IQA approach, it possesses some flaws like nonlinearly perceptual scales, biases, and rating criterion fluctuation. In practical, several super-resolution techniques perform badly in standard IQA measures i.e., PSNR, yet greatly outperform other technique in case of perceptual quality therefore MOS testing is extremely necessary IQA technique for reliably evaluating perceptual quality [23,28–33].

### 2.2.3. Learning-based perceptual quality

Researchers strive to analyze image perceptual quality by learning dataset for better assessment of perceptual quality while eliminating manual intervention. Chao et al. [34] and Hossein et al. [35] introduces no reference that are learned from visual perceptual ratings and straightway predict quality scores without the use of original images. Whereas Kim et al. [36] propose Deep QA through training the triplets' distorted images, subjective scores, and objective error maps to predict visual similarity of images. Zhang et al. [37] gathers a large-scale perceptual identical dataset and evaluate learned perceptual image patch similarity (LPIPS) based on dissimilarity in deep feature through

trained deeper networks and display that CNN-learned deep feature model perceptual similarity is higher than measures without CNN. These techniques perform superior to capture human visual perception however the type of perceptual quality requirements such as more natural image or constant identity to the original image is still unknown and remains a question to explores hence quantitative IQA techniques such as PSNR, SSIM and RMSE still remain the current mainstreams.

### 2.2.4. Task based evaluation

In fact, super resolution models can often assist other visual tasks [7, 28,38,39], evaluating reconstruction performance using additional tasks is another feasible way. In particular, researcher feed both source and restructured HR images into trained models by comparing the impacts on the prediction performance to assess reconstruction quality. The vision tasks were utilized for assessment incorporate recognizing objects [28,40], face recognition [41,42], face alignment and parsing [43,44] etc. Deep super-resolution networks have been shown in the literature to increase image quality are broadly evaluated by traditional physical metrics including RMSE, PSNR and SSIM. However, it is unclear if these metrics accurately reflect the diagnostic value of the reconstructed SR images. The efficiency and performance of DL-SR algorithms are evaluated quantitatively using a task-based super-resolution (SR) image quality assessment [45]. The purpose of a Rayleigh task is to see how signal length and super-resolution network complexity impact on s binary detection performance. To evaluate Rayleigh task performance, numerical observers (NOs) such as the regularized Hoteling Observer (RHO), the anthropomorphic Gabor channelized observers (Gabor CHO), and the ResNet-approximated ideal observer (ResNet-IO) have recently been introduced.

### 2.2.5. Other IQA methods

Other less well-known SR measures exist in addition to the above IQA techniques. The MS-SSIM (multiscale structural similarity) [46] provides high flexibility in considering differences in viewing conditions than the single-scale SSIM. To assess image quality, the FSIM (feature similarity) [47] extract feature points of human interest related to phase congruency and image gradient magnitude. The natural image quality evaluator (NIQE) [48] uses measurable deviations from statistical regularities seen in natural images without exposure to distorted images to evaluate image quality. Blau et al. [49] recently demonstrated analytically that distortion such as PSNR and SSIM, and perceptual quality such as MOS, are at odds with each other and demonstrate that as distortions diminishes, perceptual quality will deteriorate. As a result, determining ways to effectively quantify SR quality remains an urgent challenge to be solved.

## 2.3. Datasets for image supe resolution

There is variety of datasets of image SR are available with largely vary in their image quantities, resolution, quality, diversity and among other things. Few of them generate low resolution-high resolution image pairs whereas other just probed high resolution images where LR images can be usually obtain using different function in MATLAB or any other programming language such by bicubic interpolation with anti-aliasing. In Fig. 1, we have demonstrated some of the image datasets which has been extensively utilized by the SR community such as Set 5 [50], Set 14 [51], BSD100, BSD300, and BSD500 [52,53], Urban100 [54], PIRM [55], Mangal109 [56], Medical Dataset [57], DIV2K [58], Flickr-Faces-HQ(FFHQ) [59], CelebA-HQ [60], MINC [61], Vid4 and XiphHD [62]. Other datasets popularly utilized for vision tasks are also used for super resolution like ImageNet [63], MS-COCO [64] and VOC2012 [65]. Further, combining different image datasets for training is also common i.e., combining T91 and BSD 300 [66–69], combining DIV2K and Flickr2K [70,71].

## 2.4. Super-resolution competitions

A key purpose for fast paced research in the domain of single or multiple SR recently emerges from the competitions organized by the companies and conferences. In this section, we will in brief discuss on two most popular competitions for image super resolution i.e., new trends in image restoration and enhancement (NTIRE) [72] and perceptual image restoration and manipulation (PIRM) [55,73].

### 2.4.1. NTIRE

The NTIRE [72] challenge was organized in 2017 to benchmark single image super resolution (SISR). The datasets used to train, and test is named as Diverse2K (DIV2K). There are two tracks for the challenge to determine the participants. Track-1 is used for classical bicubic degradation and Track-2 is used for unknown down sampling. The down sampling operator is only identified in Track-2 by training LR and HR pairs. In addition, only decimation and blur are used without adding any noise. Using the factor of  $2 \times$ ,  $3 \times$  and  $4 \times$ , the images in the challenge are down sampled. The goals of this challenge are:

- ii An implementation of novel DIV2K dataset
- ii To progress the current existing super resolution algorithms
- ii In order to compare different state-of-the-art methods
- ii To implement in different settings

In addition, the NTIRE challenge is also applied to low-level activities performed annually in combination with computer vision and pattern recognition (CVPR).

### 2.4.2. PIRM

The PIRM is the next challenge for super-resolution [55,73]. This challenge focuses on the image perceptual consistency and jointly quantifies PSNR accuracy. Therefore, it provides an ability to advance perceptual guided algorithms alongside PSNR targeted algorithms. The challenge of PIRM uses  $4\times$  factor to assess the competing algorithms. Using bicubic kernel degradation, the images are down sampled. The challenge assessment is based on standard overall reference metrics i.e., SSIM, RMSE, PSNR [25], fidelity criterion (FC) [26], learned perceptual image patch similarity (LPIPS) [37] along with non-reference algorithm by Ma et al. [34], Natural image quality evaluator (NIQE) [48], Blind/referenceless image spatial quality (BRISQUE) [74]. Ma et al. and NIQE [48] measure the perceptual index. A hundred images of two groups test the techniques. The groups consist of extremely diversified data, such as pedestrians, plants, objects etc. The HR images of the ground truth are not accessible to the participants during the period of competition. The authors send their SR images to the website online. In addition, the participants select model training datasets. A PIRM challenge workshop is conducted at the European conference on computer vision (ECCV).

## 3. Taxonomy of image super-resolution techniques

Numerous SR techniques has been introduced by the researchers over the decades, these techniques have been widely categorized within four groups namely interpolation-based algorithms, reconstructions-based algorithms, learning or learning based algorithms and transformer-based algorithms as demonstrated in Fig. 1. The details overview of each category has been mentioned below accordingly:

### 3.1. Interpolation-based methods

An elementary kind of super-resolution techniques and commonly utilized for the simplicity of zooming images. By estimating the pixel intensities on an up-sampled grid, interpolation techniques produce a HR image from its LR images. In this technique, permanent function kernels, local variance coefficients [75] and adaptive structure kernels

[76,77] is commonly utilized. While these methods are simple and reliable for the applications of SR in real-time, they are likely to produce visual artifacts in certain instances when scale factors increase. In addition, a comprehensive study of this kind of super resolution techniques along with its limitations is presented in [78]. Zhang et al. [79] proposed a technique for converting a LR input image into a HR image while preserving textural and structural information. First, they build a new type of bivariate rational fractal interpolation model and study the analytical features. This model can be used to better characterize image features than current interpolation approaches since it have different forms of expression with different values of the scaling factors and shape parameters. Furthermore, this model combines the benefits of rational and fractal interpolation, and its efficiency has been proven by theoretical study.

### 3.2. Reconstruction-based methods

Reconstruction as known as regularization-based methods assumes that the LR images detected are derived from multiple degradations components i.e., blurring, warping, down sampling with focuses the restriction of reconstruction. As a result, the down sampled edition of the selected high-resolution images must be same as the corresponding low-resolution images. The detailed systematic overview of popular current regularization-based super-resolution algorithms in the last decade has been discussed in [80] for reconstructing high-resolution images. Different priors i.e., as edge directed priors [81–83] and similarity redundancy [84] are commonly used in this form of super resolution approach for achieving a definitive solution. In order to construct the prior design for face image SR in [85], Kernel PCA was used. Lepcha et al. [86] introduced an effective SR method based on wiener filtering and adaptive line search. This method initially processed through wiener filter which uses inverse filtering of poor-quality input images to recover visual features. Further, an adaptive line approach is used for SR rapid convergence with an approximate analytical term for step size provided to avoid setting it empirically. Further, a proposed line search strategy is modified to be more adaptable in diverse SR scenarios. Finally, the method employs recursive filtering in the transform domain which effectively preserves the edges of source images. The reconstruction-based techniques, however, often produce HR images incorporating more sharpness and unwanted edges and produce ringing artefacts particularly towards the salient edges.

### 3.3. Learning-based methods

Traditional algorithms have been using over the decades, but nowadays learning-based algorithms are outperforming these classical methods. Taking a benefit of the machine learning approaches, SR methods based on learning is achieving superior and challenging results and have received growing attention in recent years. By using the knowledge given by an image database, the learning-based approach achieves its objective. The three major types in example of learning based super resolution techniques are neighbor embedding (NE), sparse coding and deep learning approaches.

Typically, NE-based algorithms assume small patches of LR, and HR make low dimensional non-linear manifolds having identical local geometry. A NE approach presented in [87], which uses locally linear embedding (LLE) [88] to create patches of high resolution, is one of the representative works. LLE assumes that as long as appropriate samples are available, every sample along with its neighbor's rest on or nearby a locally linear patch and then discover a low dimensional embedding which safeguards the local image geometry. Marco et al. [50] suggested a further neighbor embedding super resolution technique based on non-negative NE and a least square (LS) approximation of non-negative LR patches and demonstrated the efficiency of the option of non-negative weights. Further, Multiview properties and local spatial neighbors of patches are explored in current dual geometric neighbor

embedding (DGNE) [89] technique for super resolution to find an SR spatial multiple embedding function. A neighbor embedding technique also need to look for a high reference dataset and often hinder effectiveness for practice for adequately render patches along with complex composition. Through learning a more complete low resolution and high-resolution directory pair, coding-based approaches have shown huge prospective for better quality image SR.

Yang et al. [90,91] proposes a sparse coding related SR technique in which the high-resolution property can be retrieved through considering the LR and HR characteristics share the similar sparse components of reconstruction as a sparse linear combination of the learned HR dictionary atoms. The optimization function used in [90,91] requires the regularity of the  $L_1$  norm which makes the super resolution technique deteriorate from highly intensive calculation for each low-resolution feature input. The issue was discussed further discussed in [51], since orthogonal matching pursuit (OMP) and principal component analysis (PCA) [92] are utilized for decreasing dimensionality of low-resolution characteristics and overcome sparse representation, respectively. Jiang et al. [93] proposes a SR method that improves resolution and visual quality of images simultaneously by presenting a practical solution combining manifold regularization with sparse support regression. Similarly, they introduce Graph Embedding Super-Resolution (GESR) [94], an effective algorithmic solution to SR to super resolve a HR image from a single low-resolution observation. The primary purpose behind GESR is to learn the projection matrix which maps the low-resolution patch to the high-resolution image patch by preserving the original HR image patch manifold's intrinsic geometrical structure. Using domain transform filter (DTF) and a weighted least squares (WLS) optimization strategy, Lepcha et al. [95] introduce a piecewise linear regression strategy for learning image specific low- to high resolution mapping. First, this method used a WLS optimization strategy to gradually coarsen the source images to extract multiscale information via construction of edge preserving multiscale decompositions. Then uses RF (*recursive filtering*) in transform domain to adequately preserve the edges of medical images. Furthermore, Hadamard patterns generated from LT training patches are then utilized to classify the LR-HR training patch pairings into different classes. Finally, a piecewise linear regression strategy is employed to learn the mapping relation between low resolution and high-resolution space for each class which is then utilized to produce the desired high-resolution image.

A zoom-based super-resolution method [96] is proposed for SR of low-resolution images captured from different camera zoom-lens. This approach uses three LR images of the similar static scene that are obtained with three different zoom factors. The experimental results demonstrated better results comparing to some of the existing methods. An advanced SVR (support vector regression) with sparse image representation in [97] that offers high degree of generalization when it comes to modelling the relationship between images and their associated SR versions. A statistical approach by using one to many correspondences in between low resolutions and high-resolution patches of image was introduced by Xiong et al. [98]. With respect to learned dictionaries for patches of image with and without preventing artefacts, Kang et al. [99] proposes an image sparse representation for modelling the correlation between LR and HR patches. With the view to learn the overcomplete dictionaries in [100], a Bayesian approach using a beta method was used based on sparse representation of LR and HR image patches. A statistical prediction technique was utilized in [101] which enhances the super resolution outcomes at the expenses of high computational time in case of both dictionary learning and super resolution phases. Simultaneously, by incorporating dictionary learning or clustering and regression, rapid SR algorithms were proposed in Regressors, and sparse representations anchored to the dictionary atoms were trained through ANR (anchored neighborhood regression) [102] along with its refined version A+ [103] achieved good performance through studying the regressors in complete feature training spaces. Collaborative representation was used by both ANR [102] and A+ [103] to measure the offline mapping matrices for

quick and precise super resolution. simple function (SF) [104] is used to assemble feature space in multiple subspaces along with learn efficient mapping functions in case of each subspace through linear regression. [105] have presented a new technique related to self-example learning and sparse representation for fast super resolution process. Using a similar framework [104], in each clustered feature subspace, multiple linear mapping (MLM) [105] learned low resolution/high resolution sub dictionaries, succeeded by a quick nonlocal means (NLM) [98] super-resolution technique. While these quick SR techniques are computationally effective but due to the lack of diversity in each clustered feature subspace, ANR [102], SF [104], and multiple linear mapping (absence of non-local means enhancement) [106] was ineffective for significantly enhancing the super resolution quality. Zhang et al. [107] suggested an approach which can be used to reconstruct main high frequency (MHF) and residual high frequency (RHF) respectively, consisting of the main dictionary and residual learning process.

In addition, Yongbing et al. [108] introduced first by clustering the LR feature space in manifold low-resolution feature subspaces along with grouping the corresponding HR feature subspaces. A local geometry learned through the clustering technique are utilized to gather multiple subsets of neighboring low resolution and high-resolution features through entire feature spaces in case of each cluster Centre. The multiple projection matrices are evaluated to map LR feature subspaces to high resolution subspaces by collaborative representation. The effective HR output can be calculated based on the projection matrix for an arbitrary input LR function who's corresponding low resolution cluster are closest to input matrix. Moreover, this clustering-based SR technique will moreover reduce calculation time in the reconstruction process through learning statistical priors through the clustering system. A new technique based on the Hadamard pattern and tree search has been proposed by Luo et al. [109] to significantly reduce running time. Related to Hadamard patterns created through low resolution training patches, the LR-HR training patch pairs are categorized into different groups. The mapping relationship for each class between low-resolution and high-resolution space is learned and utilized for SR. In realistic and real-time applications, quick and high-performance super resolution techniques are certainly highly preferable. Thus, most recent super-resolution methods depend on data driven deep learning algorithms to recover the essential features for desired SR. Deep learning is a part of the large family of machine learning systems based on neural networks with representation learning which seeks for learning relation between input and output directly from data. In the literature, the super-resolution problem has been comprehensively discussed based on deep learning techniques [52,110]. Deep learning algorithms have been classified into nine categories according to the most distinctive features. Fig. 1 illustrates the complete taxonomy used in the literature.

**Linear Connections.** Linear connections consist of basic configuration including of only one signal flow path without skipping ties or manifold branches. Multiple convolutional layers are stockpile on top of each other in such network designs and the input flows from first to subsequent layer sequentially. The manner in which up scaling is conducted that is, early upscaling (upsampling) or late upscaling (upsampling) varies in linear networks. It should be noted that certain linear networks are learned to replicate residual images which is the difference between LR and HR images [66,111,112]. In such instances, due to that fact that the network structure is linear, we categorize them as linear networks. Particularly in contrast to the residual networks with skip connections in their architecture. In these two subcategories below, we illustrated notable linear designs. An early upsampling model is the linear networks which first up-sample the LR input to the required high-resolution output dimension and then learn hierarchical representations of features to produce results. Bicubic interpolation which is an expensive estimation process is a popular upsampling method used for this purpose. The SR convolutional neural network (SRCNN) is the seminal task focuses on this pipeline. The initial effective attempt to use convolutional layers only for SR was SRCNN [113,114]. This attempt

can rightly be regarded as the first work SR based on deep learning algorithm which influenced many subsequent efforts in this approach. The architecture of SRCNN is straightforward which consist of only convolutional layers in which each layer or other than the last layer is succeeded by ReLU layer, stacked linearly together. The authors called the layers according to their functionality, even though the layers are similar that is convolutional layers. The initial convolutional layers are called patch extraction or extraction of features since it produces feature maps through the input images. Nonlinear mapping is called the second convolutional layer, which transforms feature map into high dimension feature vector. A final convolutional layer combines the feature maps to create the resultant HR image. A training data set are synthesized from HR images through extraction of non overlapping dense patches of size  $32 \times 32$ . The low-resolution patch at initial is down-sampled and then up-sampled applying bicubic interpolation of the similar scale in accordance with HR output image. SRCNN is an end-to-end trainable network that reduces the difference between HR image obtained from the output and HR images from the ground truth using the MSE loss function. A very deep SR (VDSR) [66] is related to deep CNN structure initially presented in [115] as compares to the shallow network structure proposes in fast SR convolutional neural network (FSRCNN) [116] and SRCNN. The structure is widely identified as VGG-net and in all network layers it uses fixed size convolutions ( $3 \times 3$ ). They propose two effective strategies in order to prevent slow convergence deep networks i.e., with 20 weighted layers. First, instead of creating a high-resolution image directly, however they learn residual mapping which provides a distinction between a LR and HR image. As a consequence, it offers a simpler purpose, and the design focuses only on HF (high frequency) data. Next, the gradient is clipped within the range of  $-θ$  to  $+θ$  that allows immensely high learning rate for accelerating training operation. The findings confirm the argument that where deep network is capable of providing better contextualization and learn generalized representation which could be utilized in multiscale super resolution. Instead of latent super-resolved images, denoising convolutional neural network (DnCNN) [111] learns to directly forecast high frequency residual. The variation between LR and HR images is essentially a residual image. DnCNN architecture is easy and same to SRCNN because it contains includes convolutional, batch normalization and ReLU layers. While both models have been capable to produce favorable outputs, their achievements rely tediously on the perfection for estimation of noise without requirements of underlying textures and structures information in the image. In addition, due to the batch normalization operations after each convolutional layer, they are computationally costly. Image restoration convolutional neural network (IRCNN) [112] suggests the series of denoisers related to CNN which could be utilizes mutually for the variety of lower-level computer vision works i.e., image deblurring, denoising and SR. The approach seeks to combine high-performance discriminatory convolutional neural networks along with model related optimization methods for achieving greater generalizability through image reconstructing works. In particular, HQS method is utilized in the observation model [117] to distinguish regularization and fidelity terms. Due to its exceptional modelling capability and test time quality, denoising prior is discriminatively learned by using CNN. The CNN denoiser consists of the stockpile of seven dilated convolutional layers intertwined with batch normalization along with non-linearity layer ReLU. A dilution process assists to model a bigger context by including a larger area of reception. Same as previous structures e.g., VDSR [66], deeply recursive convolutional network (DRCN) [118] and deeply recursive residual network (DRRN) [69] where residual image learning is carried out to speed up the learning method. The authors have proposed using small training samples and zero padding for preventing boundary artefacts because of convolution process. The scale of noise levels (0,50) which is jointly utilized for image reconstruction tasks is trained with a set of 25 denoisers. The proposed unified approach offers good output on image deblurring, denoising and SR simultaneously. As discussed in earlier examples, linear network

normally carries out early upsampling of source images. Subsequently the network architecture increases in proposition to handle with larger inputs, but the process can be costly computationally. In order to resolve this issue, post upsampling network train on LR inputs and sample the features nearby the result. The approach contributes to successful low memory footprint strategies. FSRCNN enhances velocity along with performance over SRCNN. A target is for carrying the computing speed in real-time (24 fps) as compares to SRCNN (1.3 fps). The FSRCNN also carry basic design having four convolutional layers and one deconvolutional later. Although convolutional activities are carried out by the first four layers, FSRCNN terms each layer in accordance with its purpose that is feature extractions, nonlinear mapping, shrinking and layer expansions. The feature extraction stage of the function is identical to SRCNN, with the only difference being the filter and input size. A SRCNN input is an unsampled patches of bicubic interpolation however input of FSRCNN is the initial patches without being unsampled.

The next layer is called the shrinking, because of its potentiality for minimizing the sizes of the function i.e., total number of factors through adopting small filter size (i.e.  $f = 1$ ) to effectively enhance the computation. The next step of convolutional layer is the nonlinear mapping since this is the crucial step in both SRCNN and FSRCNN according to the researchers as it assists to learn nonlinear functions and as a result it has a strong influence on efficiency. The scale of filters in nonlinear mapping is assign to three for experiment, while the quantity of channels is assigned equal to previous layer. Finally, a last convolutional layer is known as expending, it is an inverse shrinking phase to enlarge the sizes. The final layer contributes to a 0.3dB output improvement. The last part of the structure is a deconvolutional layer for upsampling and aggregation details, i.e., is an inverse of convolution method. An image is transformed with a stride in the convolution filter in the convolution operation, and the result of this convolution layer are 1/stride of the input. A function of the filters is however exactly inverse in the deconvolution layer in which stride performs as an upscaling factor. Another approach terms as EBSR (enhanced burst super resolution) [119] that separates the multi-frame super resolution problem into three parts: alignment, fusion, and reconstruction. To align several LR bust images in the feature level, this method proposes a FEPDC (feature enhanced pyramid cascading and deformable convolution) module. A cross nonlocal fusion (CNLF) module is then fused the aligned features together. The long-range concatenation network (LRCN) finally reconstructs the SR image. Furthermore, this approach creates a cascading residual pathway structure to boost performance. Similarly, the use of the parametric rectified linear unit (PReLU) [120] rather than rectified linear unit (ReLU) after every convolutional layer is other subtle variation from SRCNN. A same cost function as SRCNN is used by FSRCNN that is MSE. FSRCNN utilized 91 images [90] along with 100 more images acquired through internet for training. In order for increasing the quantity of images by 19 times and data augmentation i.e., rotation, flipping and scaling is often utilized. Wang et al. proposes a wavelet multiscale convolutional neural network (WMCNN) [121] based on wavelet analysis by training convolutional neural networks. For multi-scale representations, wavelet decomposition was carried out in areal image. Train multiscale convolutional neural networks (CNNs) separately to approximate the wavelet multiscale representations. Thus, trained multiple CNNs characterize aerial images in multiple directions and multiscale frequency bands, thus making image reconstruction subject to sophisticated culture variability. From a low-resolution aerial image, the trained CNNs regress wavelet multiscale images followed by wavelet synthesis which obtained a desired HR aerial image. In attempts to restore the perceptually sharp HR medical images, recently Goyal et al. [122] proposes an effective SR technique based on weighted least squares optimization via multiscale convolutional neural networks (CNNs). The weighted least squares optimization progressively coarsens and extract multiscale details from the input images. Then subsequently utilizes CNNs based on wavelet analysis to carry out SR in order to produce desired HR image. Shi et al. proposed a technique known as

efficient subpixel convolutional neural network (ESPCN) [62] for extracting feature information directly from LR images and successfully generate HR images. In general, it is simple SR technique which can perform for both videos and images in real-time scenarios. As described above, conventional super resolution methods initially map a LR image with bicubic interpolation to a high resolution and then learn the super resolution design in the large dimensional space. However, ESPCN represents that this pipeline leads to high computational necessities, therefore perform to extract features in low resolution space. ESPCN utilizes the sub pixel convolution layer at last step to collect low resolution feature maps after the features are extracted and simultaneously projected in high dimensional space to restore high resolution image. Transforming image feature at low resolution space greatly reduces the memory requirements and computation. An operation of sub-pixel convolution used in this task is approximately same as the operation of deconvolution or transpose of convolution [123], in which a fractional kernel stride is utilized for improving a spatial resolution of input feature maps. To map each function map, a separate upsampling kernel is utilized to provide more flexibility in modelling low-resolution to high resolution mapping. To train the overall network, a  $l_1$  loss is used. ESPCN offers competitive SR results with the efficiency of processing 1080p videos in real time in a one GPU. The super-resolution outcomes of CNN-based methods have considerably surpassed traditional methods since the development of CNN. In particular, CNN based single image super-resolution algorithms generates excellent results. Although video sequences carry more information than images, there are few videos super-resolution solutions that can be used on mobile devices due to the high computational requirements which limits the use of video SR. For real-time video SR, the efficient video super-resolution network (EVSNet) [124] uses neural architecture search. Extensive experiments demonstrate that this method provides a good balance of quality and efficiency.

**Residual Connections.** Residual learning networks when comparing to linear utilizes network architecture having skip connections for preventing gradient vanishing and create the possibility of designing very deeper networks. Importance of this model was initially illustrated in the problem of image classifications [125]. Several networks have recently provided a boost to super resolution output through residual learning network (RLN) [70,126]. These techniques learn the residue such as high frequencies between source image and the resultant image. The recent residual learning networks (RLNs) [127–129] are categorized on the basis of the number of phases used in the networks. An individual network consists of a single stage design; an enhanced deep residual network (EDRN) [70] for SR modifies a ResNet structure [125] primarily proposed for image classification in cooperation with the super-resolution tasks. In particular, the removal of BN (batch normalization) layer from every residual block and ReLU activation outside residual block shows huge improvements. This model also expanded their single scale method to work on different scaling factors similar to VDSR. However, via majority of shared parameters, their method multiscale deep super-resolution (MDSR) architecture decreases the quantity of parameters. Scales of specific layers are implemented in only parallel near the input and output blocks for learning a scale dependent representation. Using  $l_1$  loss, the proposes deeper architecture is trained. In order to create a ‘self-ensemble’ data augmentation (rotations and flips) is utilized which means transformed inputs is transferred along the network and added jointly to generate a single output.

The researchers observed that self-ensemble method does not need several different modalities to be trained, however outputs are beneficial as compares to traditional ensemble related models. For quantitative measurements such as PSNR, MDSR and EDSR obtains better result as compares to previous design such as VDSR, FSRCNN and other similar based design based on ResNet such as super-resolution generative adversarial network (SRGAN) [23]. ResNet blocks [125] are used by cascading residual network (CARN) [126] for learning about the

correlation between LR inputs and HR outputs. An existence of global and local cascading modules is actually difference between the models. The intermediate layer features are cascaded and clustered into a convolutional  $1 \times 1$  layer. Except that the blocks are simple residual blocks, the local links are similar to global cascading networks. Due to multi-level representations and several shortcut connections, this method makes data propagation efficient. The model is trained with data augmentation using  $64 \times 64$  patches from Berkeley segmentation dataset (BSD) [52,53], Yang et al. [90] and DIV2K dataset [58] along with  $\text{AptCommandmathcall}_1$  loss. Adam [130] are utilized to optimize the first learning rate of  $10^{-4}$  which get reduce to halves after each  $4 \times 10^5$  steps. The residual feature distillation network (RFDN) [131] proposes a feature distillation connection (FDC) which functionally comparable to channel splitting process while being more lightweight and adaptable. In order to learn discriminative feature representations. RFDN employs multiple feature distillation connections. It also proposes a SRB (shallow residual block) as RFDN’s main construction block so that the network could reap maximum benefits from residual learning while being lightweight. Detailed experimental performance indicates that the RFDN provides a superior trade-off in terms of performance and model complexity than other algorithms. Xiang et al. [132] presented a one stage space time video super resolution (SVSR) architecture that produces a HR slow motion video directly from LR video. Instead of synthesizing the missing low-resolution video frames like VFI networks do, the proposed feature temporal interpolation network first temporally interpolates low resolution frame properties in missing low-resolution video frames to captures local temporal contexts. The researchers then proposed a deformable ConvLSTM to align and collect temporal information at the same time in order to use global temporal contexts. Finally, a deep reconstruction network is used to predict the HR slow motion video frames.

There are many sub-networks in a multi-stage architecture that are typically successively trained [127,128]. The initial subnet generally predicts the coarse characteristics, whereas other subnet enhances the initial prediction. It also includes encoder-decoder model such as [129] which initially use an encoder to down-sample the input and then up-sampled using decoder i.e., two separate stages. The architecture super resolved the images in numerous stages. The FormResNet is illustrated in [127] that constructs on DnCNN. This method is made up of dual networks, both are identical to DnCNN but differs in loss layer. An initial network known as formatting layer involves Euclidean and perceptual losses. This formatting layer can also be substituted by classical algorithms such as BM3D. The “DiffResNet”, a next deep network is same as DnCNN where the input of the network is fed from the first layer. In uniform fields correspond to formatting layers eliminates HF corruption, whereas DiffResNet learns the structured areas. FormResNet increases DnCNN’s outputs by a slight margin. The LR stage and a HR stage are composed of balanced two-stage residual networks [128]. The function maps have a smaller size in the low-resolution level which is similar to input patches. A feature maps is upscale utilizing deconvolutional layer followed by upscaling of the nearest neighbor. Then upscaled feature map is then fed in the HR stages using a residual block variant [125] known as projected convolution. As feature map projection, the residual block include  $1 \times 1$  convolutional layer to minimize the input size of  $3 \times 3$  convolutional features. There are six residual blocks in LR phase while four residual blocks are present in HR phase. The model is trained on 900 DIV2K dataset [58] in which 800 training images along with 100 validation images are presented in competitor in NTIRE 2017 [72]. The images are cropped into  $108 \times 108$  dimensional patches during training along with extended utilizing rotation and flipping procedures. The first learning rate is set to 0.001 that reduces by the factor 0.6 exponentially for each iteration. Adam [130] was used for optimization. There are around 128 feature maps as inputs and 64 as outputs in the residual block. The distance  $l_2$  is utilized for measuring the variation between prediction performance and the

ground truth.

The SR technique utilizing an encoder such as relating to convolutional layer and a decoder has been proposed lately due to the success of UNet [129,133]. REDNet [129] refers as Residual Encoder Decoder Network and contains predominantly convolutional and symmetric deconvolutional. After each convolutional and deconvolutional layer, a ReLU layer is add on. Although by maintaining the object structures and eliminating degradations, the convolutional layer extract feature maps. The deconvolutional layers, on the other hand, reconstruct the missing image information. In addition, skip connections is included between the convolutional and the symmetric deconvolution layer. Until applying nonlinear rectification, the feature maps of convolutional is added up with the result of mirrored deconvolutional layer. Bicubic interpolated images is input to the network along with a HR image is the result of final deconvolutional layer. The given network is trainable and end to end convergence is obtained through minimizing  $\text{AptCommandmathcall}_2$ -norm between the device performance and a ground truth. The authors suggested three REDNet structure variants in which overall architecture stays the similar however the quantity of convolution and deconvolution layers are reversed. There is 30 weight layers in the best performing architecture, each with 64 feature maps. In addition, BSD luminance channel is utilized for producing the training image dataset. The  $50 \times 50$  patch sizes are extracted as the ground truth with a regular stride, and the input patches is obtained by sampling the patches through the ground truth and later up sampled it to original size through bicubic interpolation. They extract patches from the 91 images [90] and network is trained using MSE loss as a loss function. The patch size for input is  $9 \times 9$  and output are of  $5 \times 5$ , respectively. The patches were normalize using its variance and means that is eventually added to the corresponding reconstructed resultant HR image. In addition, kernel possess dimension of  $5 \times 5$  along with 128 feature channels. On the other hand, existing approaches do not fully use the hierarchical properties of residual branches. The residual feature aggregation (RFA) architecture for more effective feature extraction was developed to address this issue. By adding skip connections, the RFA framework connects several residual modules together and immediately sends the features on each local residual branch. Thus, the RFA framework can combine these useful residual features to build more representative features. This method also introduces ESA (enhanced spatial attention) block to form residual characteristics highly focused on essential spatial contents in order to maximize the power of RFA framework. The ESA block was created to be both lightweight and efficient. The final RFANet [134] is designed using RFA framework with ESA blocks. Recently, a new recurrent structure detail network (RSDN) [135] for video super-resolution has been developed that is both efficient and effective in exploiting prior frames to super-resolve the current frames. It divides the input into structure and detail coefficients which is then fed into the recurrent unit made up of many proposed two-stream structure detail blocks. Furthermore, a hidden state adaption module is also developed which allows the present frame to selectively employ details from hidden state to improve its effectiveness to appearance change and error accumulation. A BasicVSR [136] is introduces to untangle the knots and readdress several highly important components for video super resolution (VSR) using four basic functions such as alignment, propagation, aggregation and upsampling. BasicVSR shows a concise pipeline that offers compelling advantages concerning to speed and reconstruction quality as compares to several exiting methods by reusing some existing components with less redesigns. This strategy uses a methodical approach to describe how such gains can be made and to explore the potential difficulties. To demonstrate the effectiveness of BasicVSR, a coupled propagation technique and information refill mechanism has been used to facilitate extracting information adequately. The model based on Dirac convolution, an enhanced Dirac residual block was introduced [137], that utilize the trainable parameters to adaptively control the balance of the convolution and skip connection. In addition, global skip connection is realized to enhance

the nonlinear mapping ability of the model. The model shows excellent performance both on terms of visually and quantitatively.

**Progressive Reconstruction Connections.** Convolutional neural networks methods usually determine the result in single step; yet for large scaling factors it may not be feasible. Some algorithms [30,67], estimate the performance in several steps to deal with large factors i.e.,  $2 \times$ , followed by  $4 \times$  and so on. Wang et al. [30] proposes a method for consolidating the standard of sparse coding [90] with deep neural networks. It aims to create a compact model and boost efficiency with this combination. In order to construct the multi-layer neural network, a recommended sparse coding-based network (SCN) [30] imitates the LISTA (learned iterative shrinkage and thresholding algorithm) network. The initial convolution layer extracts feature through LR patches and thus fed into the LISTA network similar to SRCNN. The LISTA network comprises of limited number of recurring phases to obtain sparse code for each feature. A LISTA operation comprises of two linear layers along with a non-linear layer having a threshold activation function that are learn during training. Researchers performed non-linear neuron decomposition into two linear scaling layers and a unit threshold neuron to simplify training. The two scaling layers is the diagonal metrics that is reciprocal to each other, it means if there is a multiplication scaling layer then it follows division after the threshold unit.

The primary HR patches is reconstructed after a LISTA network by multiplication of sparse code and HR dictionary in next linear layer. In the last phase, the HR patches are again placed in the primary position in the images using a linear layer to achieve the high-resolution output. A pyramidal structure is used by the Lap-SRN [67]. Lap-SRN is made up of three subnet that progressively determine residual images up to  $8 \times$  factor. In order to generate SR images, the remaining image of subnet is applied to low resolution input image. A  $4 \times$  residue is given by the output of the initial sub-network and the last one provides the  $8 \times$  residual image. To generate the final super resolution image, those residual images are applied to the corresponding unsampled images. Researchers relates residual prediction branch as feature extractions though adding bicubic interpolation images with residue is known as image reconstruction branch. LapSRN network consists of element forms such as leaky ReLU, convolutional layers and deconvolutional layers. The convolution layers come before leaky ReLU that allows the negative slope of 0.2 and convolutional layer at the final stage of subnetwork according to the CNN convention to increase size if residual images to the corresponding range. Lap-SRN utilizes a distinguishable  $\text{AptCommandmathcall}_1$  loss function variant called Charbonnier which can accommodate outliers. In every sub network, the loss is engaged, resembles to a multi-loss structure. Further, the convolutional and deconvolutional layer filter sizes are  $3 \times 3$  and  $4 \times 4$  respectively that contain 64 channels each. The training data is close to that of SRCNN that is 91 images with 200 images of BSD dataset [52,53], respectively. Three different models are used for  $2 \times$ ,  $4 \times$  and  $8 \times$  SR performance in the LapSRN model. They also propose single design called Multi-scale (MS)-LapSRN which learns for managing various super resolution scales together [33]. However, a single MS-LapSRN design improve an output as compares to other three different designs. The main reason is that design uses common inter-scale traits which supports to achieve further detailed outputs. Zhao et al. proposed a technique that proposes two building blocks for main branch and the reconstruction for super-resolution on the basis of pixel attenuation (PASR) [138]. First, PASR block is similar structure as self-calibrated convolution but with pixel attenuation (PA) layer. Due to its two-branch structure and attention system, this block is substantially highly effective than traditional residual and dense blocks. Next, U-PA block combines nearest-neighbor upsampling, convolution and PA layers. It increases the quality of final reconstruction at low-cost parameters.

**Recursive Connections.** Recursive networks [69,118,139], either uses recursive linked convolution layers or recursive linked units, as the name suggests. The key motive behind these models is to gradually

break down the more complicated issue of SR into a series of easier ones which are simple to solve. The same convolution layers are extended many times by the DRCN (deep recursive convolutional network) [118]. The benefits of this approach on which the quantity of parameters for further recursions remains constant. DRCN consists of three small networks known as an embedding network, an inference network and reconstruction network. A first sub-network, known as embedding network, transforms an input to the feature maps i.e., either grayscale or color images. Super-resolution is performed through the successive sub-network called an inference network that analyze image areas by recursively using single layer that consist of convolutional and ReLU layer. After each recursion, a receptive field size is increased. HR feature maps are converted to greyscale/color using reconstruction network after inference network. The deep CNN design with conservative parametric complexity is proposed by DRRN. This design presents an even deep structure having many as 52 convolution layers as compares to earlier designs i.e., VDSR [66], REDNet [129] and DRCN [118]. Similarly, the size of the network is decreased by factors of  $14 \times$ ,  $6 \times$ ,  $2 \times$  for DRCN, VDSR and REDNet accordingly. This is accomplished through integrating residual image learning [66] and local identity ties within the network between small block of layers. The researchers emphasize that such parallel data flow offers robust training for deep structures. DRRN uses recursive learning same as DRCN [118] that reproduces a simple skip connection block many times in order to obtain the multi-path network block. The computational complexity and memory cost are greatly reduced as parameters are shared between the replications and by stacking several recursive blocks to obtain final architecture. For parameter learning, DRCN utilized standard stochastic gradient descent (SGD) optimizer with gradient clipping [66]. Same as other traditional architectures, the loss layer is based on MSE loss. A coherent improvement over the previous models is registered in the proposed architecture, supporting the scenario of deep recursive structures and residual learning.

The novel super-resolution image persistent memory network called MemNet is available accessible through Tai et al. [139]. MemNet are split in three phases which is same as SRCNN. A first phase is known as feature extracting block that extract features from input images. This phase is similar to previous models such as [62,114,116]. A sequence of memory blocks stacked together consists of the second phase where this phase plays a most important role the network. The gate unit and recursive unit form a memory block. A recursive phase is same as ResNet architecture [125] that contain two convolutional layers along with the pre-activation operation and dense networks to the gate unit. The convolutional layer having kernel size of  $1 \times 1$  in each gate unit. MemNet adopts the MSE loss function. It uses 200 image of BSD dataset along with 91 images of [90] and the experimental setting is same as VDSR. The network includes of six memory blocks and six recursions. The overall layers count in MemNet is 80 layers. These methods have delivered exceptional results particularly for dense connection-based networks, but they also bring higher complexity and more parameters. Jiang et al. [140] propose a hierarchical dense recursive network (HDRN) for image SR that considers both reconstruction performance and efficiency. First, using a hierarchical matrix structure design that construct a hierarchical dense residual block (HDB) to improve feature representation while reducing memory footprint. In this way, it can provide more interleaved pathways for information fusion and gradient optimization however with shallower depth comparing to earlier networks.

A group of convolutional layers with a small size ( $1 \times 1$ ) are embedded in HDB allowing the computational burden and parameters to be released by rescaling the feature dimensions. Furthermore, HDBs are linked together in a sharing fashion, allowing the network to combine features at different stages. The multi-scale features from these HDBs are then merged into the global fusion module (GFM) for global fusion and representation and the final profile enriched residual map is obtained by realigning and sub-pixel upsampling the fusion maps. MemNet is often

used where it produces promising results for other image reconstruction activities i.e., image denoising and JPEG deblocking. The SR feedback network (SRFBN) based on recurring architectural model was proposed by Li et al. [141]. To generate a corresponding HR image, a LR input is refined recursively. A model is based on feedback block (FB) consisting of many projection groups. Each projection group initially extract features of HR via deconvolution and then produces features of LR via convolution. With each FB, there are dense relations between LR and HR representations. Inputs are recursively transferred to FB at various time steps that learn residual signal due to presence of global residual relation. In the case of multiple types of degradation in low-resolution images, SRFBN is trained utilizing the curriculum learning strategy. HR images of increasing complexity are introduced as ground truth to the model in this method. The design is trained with an  $l_1$  objective along with total of four recursive iterations is utilised during training. In addition to usual bicubic down sampling, evaluations have been observed for other degradations i.e., Gaussian blur. The recursive architecture enables this technique to operate with a comparatively reduced number of trainable parameters.

**Attention based Connections.** Both spatial positions and channels are of uniform significance for super resolution in the previously mentioned network designs. It aids to selectively deal with just the limited features on a specific layer in several instances. The attention related techniques [40,142], enable this versatility and taken into account in which all features are necessary for SR however but encounter varied consequence. Recent attention-based models, combined with deep networks, have shown major improvements in SR. A new selection entity for the image SR network known as SelNet was proposed by Choi et al. [142]. A selection entity acts as a gateway between convolution layers granting only the values selected through the feature maps to be permitted. A selection entity consists of identity mapping,  $1 \times 1$  convolution, ReLU, a sigmoid and cascade layer. The overall 22 layers of convolution consist of SelNet, along with selection entity is added up after each convolution layers. Same as VDSR [66], SelNet [142] also employs residual learning and gradient switching (a variant of gradient clipping) for faster learning. The LR  $120 \times 120$  patches are the network inputs that is crop from DIV2K dataset [58]. Along with learning rate of  $10^{-1}$ , a number of epochs is set at 50. The loss utilized for the SelNet training is  $l_2$ . Zhang et al. [40] recently proposes a deep CNN structure for SR known as residual channel attenuation network (RCAN) [40]. The architecture key highlights consist of the recursive residual model in which there is residual connection within every block of global residual network (GRN) and every local residual carry the mechanism of channel attention in such a manner that filter activations collapse from  $h \times c \times w$  to vector with size of  $l \times l \times c$  after passing through bottleneck which serves as the selective attention over channel maps. The initial breakthrough offers numerous paths for the transfer of information from the first to the last layers. The next contribution enables network to concentrate in selective feature maps which is largely important for final task and design the correlations between the feature maps efficiently. RCAN utilizes  $l_1$  loss function for network training. The recursive residual type of architecture has been observed to lead for superior convergence characteristics on very deep networks. Furthermore, when compares with modern methods such as VDSR [66], IRCNN [112], residual dense network (RDN) [143] generates better results. In order to illustrates the effectiveness for real time super resolution for low level vision assignments, Lui et al. presents a self-attention negative feedback network (SRAFBN) [144]. Through the self-attention negative feedback model, the network model constrains the image mapping space and picks the important information of the image resulting in high quality images that meet human visual perception. The shortcoming of this method possesses high computational complexity ( $\times 15$  million parameters for  $4 \times$  super resolution) as compares to VDSR [66], LapSRN [67] and MemNet [139] respectively. At recent, the densely residual Laplacian network (DRLN) [145] has been implemented to super-solve

images. The configuration of the network is modular and hierarchical i.e. (a) densely connected residual units, (b) modular architecture, (c) cascading operation and (d) Laplacian attenuation are the key highlights of the network. In addition to the cascaded ones, DRLN exploits distinct connections such as long skips, median skips, local skips. Likewise, the residual units are closely linked in each block to learn a dense representation. In the same block, the obtained features are then weighted using Laplacian attention. The configuration is replicated in every block in the network. DRLN currently offers the best performance for all datasets. DRLN [145] has a  $l_1$  loss function similar to RCAN [40] for training network. The settings for training are same as RCAN where the quantity of epochs, optimizer, and training patches. A creative module with Laplacian attention and cascading structure can be attributed to the improvement of DRLN [145]. Compared to RCAN, the number of traditional DRLN layers is slightly lesser. Although in case of DRLN, the number of parameters is high, but it is computationally less expensive because of channel concatenation as opposite to RCAN where costly operation is used, i.e., channel addition. Most existing CNN-based SR models necessitate a lot of computational power which limits their real-world applications. Furthermore, most CNN-based algorithms do not investigate intermediate features that are useful for final image recovery.

A dense lightweight network known as MADNet [146] was introduced to overcome these challenges by introducing for stronger multi-scale feature expression and feature correlation learning. To improve the informative multiscale feature representation capabilities, a residual multiscale module with an attention mechanism (RMAM) is designed. This method also includes a dual residual-path block (DRPB) that makes use of hierarchical features from LR images. Dense networks between blocks are used to take benefit of the multilevel features. On the other hand, channel attention interprets each convolutional layer as an independent operation that missed the relationship between different layers. To address these problems, a holistic attention network (HAN) [147] was developed to model the holistic interdependences among layers, channels, and positions. It consists of LAM (layer attention module) and the CSAM (channel spatial attention module). The proposed LAM considers layer correlations to adaptively emphasize hierarchical features. Meanwhile, CSAM learns the level of trust at all positions of each channel so that it may selectively capture more informative features. The latest research introduces a super-resolution residual attention module (SRRAM) [148] that concentrates on the attention blocks utilized for single image SR. In order for comparing their performance and respective advantages/ disadvantages, they analyze the variety of attention mechanisms with popular super resolution architectures. SRRAM is similar in structure to RCAN, as both approaches are influenced by EDSR [70]. The SRRAM can be divided into three sections: extraction of features, upsampling of features and reconstruction of features. The first and last sections are close to the methods previously discussed [113,116]. It features upscaling section that consists of residual attention module (RAM). The RAM is an integral part of SRRAM system includes the residual blocks along with channel attention and special attention to learn inter and intra channel dependencies. A model is trained with data augmentation employing randomly cropped patch sizes of  $48 \times 48$  from the DIV2K dataset [58]. Adam [130] employs  $l_1$  loss as optimizer compose first leaning learning rate at  $10^{-4}$  and the final model utilises a total of 64 RAM blocks. Isobe et al. [149] proposed a video SR with temporal group attention (TGASR) technique that may successfully include temporal details in the hierarchical way with the goal of generating a HR video from its corresponding LR version. The input sequence is split into numerous groups where each of them corresponds to a different frame rate which is again combines with an attention module and a deep intra group fusion module to provides complementary information to restore missing information in the reference frame. In addition, a quick spatial alignment is proposed to handle videos with a lot of motion. Recently, Yang et al. [150] proposes

an effective SR approach based on neural architecture search and attention mechanism. This method uses global residual learning and further added a down-sampling for reducing the features map's size and computation. Further, they establish lightweight search space and joints rewards to search an ideal network structure. Additionally, they introduce Bayesian algorithm for hyper parameter tuning and improves the efficiency of model based on an ideal sub-network search out.

**Multi Branch Connections.** The objective of multi branch network is to gain the diverse set of features at different context scale in the linear (single stream) contract and skip connections related models. Such additional details are then combined for achieving better high-resolution image reconstructions. This model also allows for multi path signal flow, lead to better details sharing during training in forward backward stages. In many other computer vision activities, multi branch designs are becoming popular as well. Ren et al. [151] proposes by combining various CNNs for imaging. The researchers proposed a CNN based context-wise network fusion (CNF) where each SRCNN are designed with certain number of layers. Each output of SRCNN fed into single convolutional layer and then combined utilizing a sum pooling. The 20 million patches obtained from open image datasets [152] are trained on the model. Only luminance channel of each patch size  $33 \times 33$  pixels. First for 50 epochs, each SRCNN is separately trained with learning rate of  $10^{-4}$  and then the combined network is trained with similar learning rate with 10 epochs. Corresponding progressive learning approach is same as curriculum learning where it begins with the basic operation and further goes on to higher complicated operation of combined optimizing multiple sub-networks to achieve enhanced super resolution. A MSE is utilized as a loss for training the networks. A dual super resolution learning (DSRL) [153] has been proposed as a simple and flexible two-stream system for improving segmentation accuracy without adding extra computing expenses. The proposed method is made up of three parts: semantic segmentation SR (SSSR), single image SR (SISR) and feature affinity (FA) module that can preserve high-resolution representations with low-resolution input while decreasing model computation complexity. It can also be simply generalized to other tasks such as human posture estimation. This simple strategy produces powerful representations and as proven by promising results in semantic segmentation and human posture estimation.

The cascaded multi scale cross-network (CMSC) [154] consists of features extraction layer, a reconstruction network and cascaded subnet. The feature layer executes the similar role as described for SRCNN and FRCNN cases. Each subnet consists of blocks of merge and run (MR). Each MR block consists of two parallel networks, each having two convolution layers. A residual connection from each leg is gathered jointly and then added individually to the output of both legs. Each CMSC subnet consists of four MR blocks with different  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$  receptive fields to capture multiscale contextual details. In addition, batch normalization and Leaky-ReLU [155] are further performed in individual convolution layers in the merge and run (MR) block. The final output is created through the last reconstruction layer. Loss function is  $l_1$  that uses a balance term to combine the intermediate outputs with the final one. Bicubic interpolation is used to up sample the input to the network with  $41 \times 41$  patch sizes. The design is trained utilizing a first learning rate of  $10^{-1}$  with 291 images comparable to VDSR [66], decreasing by a factor of 10 for a total of 50 epochs after every ten epochs. As compares to EDSR and its MDSR variant, CMSC lags in performance. There are three blocks in IDN (information distillation network) [156], i.e., the feature extraction block, several information block and recontraction block. The feature extraction block consists of two convolutional layers for extracting features. A distillation block consists of two another blocks, a compression unit along with an enhancement unit. There are six convolutional layers to the enhancement unit followed by leaky ReLU. The third convolutional layer output is segmented where half batch is concatenated along with input and another half is utilized as the fourth convolutional input layer. A

concatenated component result is added to an enhancement block output. Four enhancement blocks in total are used. After each enhancement block, the compression unit are realized utilizing a  $1 \times 1$  convolutional layer. The deconvolutional layer having a kernel dimeson of  $17 \times 17$  is the reconstruction block. Using absolute mean square error loss, the network is first trained and then finely tuned using the MSE error loss. A training images is similar to [139]. The size of input patches is  $26 \times 26$ . For total of  $10^5$  iterations, the first leaning rate is configured at  $1e^{-4}$  using Adam [130] as an optimizer. An embedded block residual network (EBRN) [157] is related to the principle where various image frequencies appearing in an image need diverse processing levels, i.e., example, a shallow network may restore low-frequency information, whereas higher complex, high frequency content requires a deep network for proper modelling. For that reason, they suggested a multi branch structure where during super-resolution, the higher complex signal is transferred in deep modules. There are ten block residual modules (BRM) in the final model, leading total ten parallel networks in the structure. A novel fusion method is utilized to integrate multiple branch outputs. The recursive fusion method is used rather than simple summation of all outputs that only the outputs of two adjacent branches is gradually merged before they arrive the main LR branch. The design is initially trained with  $AptCommandmathcall_1$ -loss and then fine-tuned by the objective of  $AptCommandmathcall_2$  to penalise outliers in expected outputs. Altogether compared to the existing models of  $2 \times$ ,  $4 \times$  and  $8 \times$  SR models, the proposed approach offers impressive results. Video super resolution (VSR) intends to provide HR prediction for each frame by utilizing multiple low-resolution frames. Inter-and intra-frames are the primary source to exploit temporal and spatial information in this process. But existing VSR approaches have a few drawbacks. Initially, optical flow is frequently utilized to create one-on-one temporal correspondences. However, flow estimation is prone to errors which has a significant impact on the final recovery result. Second, for the VSR task, similar patterns present in the natural images are rarely used. Based on these findings, Li et al. [158] proposed a temporal multi-correspondence aggregation network (MuCAN) to exploit most similar patches across frames and the cross scale non-local correspondence aggregation scheme to analyze self-similarity of the images across scales. They construct an efficient multi-correspondence aggregation network for video SR based on these two modules. In order to addresses some of the problems of previous models, Yang et al. [159] recently proposed a method based on multi branch attention SR model. The network consists of multiple residual bodies composed of several residual units. In addition, they constructed multi branch mechanism that divides all channel into equal parts. Further network learns the correlation between channels and concentrates more on the high frequency channels.

**Densely Connected Connections.** With the success of dense convolutional network (DenseNet) [160] in image classification models, SR methods based on densely connected convolutional layers has been introduced to enhance performance. In such architecture, the primary motive id to integrate hierarchical cues available along the network to accomplish large flexibility along with rich features representations.

The SRDenseNet [161] is based on DenseNet that utilizes dense networks between the layers where layer operates directly on the output of all prior layers. Such details transfer from low to high level feature layer prevents vanishing gradient problem and allows learning compact designs that accelerates the training procedures. SRDenseNet utilizes a few deconvolutional layers for upscaling the inputs to the rear part of the network. The authors suggest three SRDenseNet versions, (1) a sequential dense block structure succeeded by deconvolutional layers. Only high-level features are utilized in this way to recreate the final super resolution image. (2) Low level features are combined from first layers prior to the final reconstruction. The skip network is utilized to integrate low and high features. (3) Uses various skip connection between low level features and dense blocks and furthermore all features

are combined to allow the direct transfers of information for better high-resolution reconstruction. An integration of all feature maps provides the better output as compares to other variants of SRDenseNet as corresponding features are encoded at various steps in the network. A MSE is utilized s loss function to train the complete model. In general, SRDenseNet designs display the clear performance enhancement over designs that does not employs dense connections between layers. The RDN [143] integrates residual skip connections (i.e., likewise SRDenseNet) having dense connections. A primary inspiration is to make good use of the hierarchical feature representations for learning local patterns. Residual links are formed at steps for this purpose, local and global. A new residual dense block (RDB) was proposed at the local level in which input for each block i.e., an image or resultant of earlier block is transferred to entire layers in the residual dense block and further applied to the output of the block such that each block concentrates high on the residual pattern. However dense connection easily guides to higher dimensional results; therefore, a local feature image integration approach was used in each RDB to minimize the dimensions with  $1 \times 1$  convolution. Multiple RDB outputs are combined at the global level (through concatenation and  $1 \times 1$ convolutional procedures) and global residual learning (GRL) carried out to integrate features form various network blocks. The residual links support to stabilize the training of the network and lead to enhancement over SRDenseNet. RDN uses the  $y$  loss function in comparison to  $x$  loss utilized in SRDenseNet and supports its enhanced converge features and network training is carried out on randomly chosen  $32 \times 32$  patches in each batch. The data augmentation by flips and rotation is used as a regularization step. Researchers also analyze settings where various source of deterioration i.e., noise and artefacts are available in low resolution images. This method demonstrates strong resistance to such degradations and recovers much improved images of SR. A Dual Regression Networks (DRNet) [162] for SISR was created to address the existing concerns with deep neural networks. This dual regression approach reduces the space of potential functions by introducing an additional constraint on low resolution data. In particular this method learns an additional dual regression mapping that estimate the down sampling kernel and reconstructs low resolution images besides mapping from LR to HR images which makes the closed loop to generate additional supervision.

It is important since the dual regression operation is not reliant on HR images as they could learn directly from LR images. Hence this method makes it simple to adopt SR algorithms to real world data such as raw video frames from YouTube. The deep back projection networks (DBPN) for SR [71] is inspired from convolutional super-resolution method i.e., Irani et al. [163], where iteratively execute back propagations for learning the signal of feedback error between LR and HR images. An explanation in fact the model of mapping from LR and HR images is not suitable for a feed forward technique however a feedback mechanism will greatly help to achieve superior performance. The proposed architecture involves a sequence of upward and downward sampling layers which is densely connected with one another for this purpose. In such a way, high resolution images are combined from different depths of the network to retrieve the resultant output. A simple case of single connection from prior layer is introduced for the compactness. The integration of upsampling results for the input feature maps and residual signal is the significant factor of this model. In the up-sampled feature maps, the explicit inclusion of the residual signal generates error feedback and urge network to concentrate on fine information and employs a standard  $l_1$ - loss function for training the network. The DBPN has a comparably high computation complexity for  $4 \times$  super-resolution of  $10 \times$  million parameters that lead to a slight decrease in performance. A lightweight improved SRCNN (LESRCNN) [164] was recently proposed that consists of three successive sub-blocks, an information extraction and enhancement block (IEEB), a reconstruction block (RB) and information refinement block (IRB). IEEB extracts hierarchical low-resolution features and collects an extracted features step by step to boost the memory ability of the shallow layers on

deep layer of super resolution. In order to eliminate unnecessary data, IEEB has a heterogeneous design. The reconstruction block converts low frequency (LF) features into high frequency (HF) features by fusing local and global features which is complementary to IEEB to address the long-term reliance issue. Lastly, IRB utilizes coarse HF to learn more precise super resolution features from the RB and constructs the super resolution image. The proposed LESRCNN uses a model to generate a high-quality image at various scales.

**Multiple Degradation Handling Connections.** Bicubic degradations are considered in the SR networks discussed so far in [66,113]. However, maybe this is not a feasible assumption since several degradations may prevail simultaneously in reality. The following approaches are proposed to deal with such real-world situations. Zero-shot super-resolution (ZSSR) [165] that exceed the steps of traditional techniques by using internal image statistics to SR images using the power of deep neural networks. A basic network architecture is used by the ZSSR which is trained using a down-sampled test version. The objective is to foresee the test images produced through test images from the LR images. The similar network is utilized for predicting the super resolution image utilizing the test images as input when the network discovers the correlation between the low-resolution test images and original test images. Therefore, for a particular degradation, it does not involve training images and it learns an image specific network during inference. A Zero-shot SR consists of eight convolutional layers, succeeded by a 64-channel ReLU Similar to [66,70], ZSSR [165] uses  $l_1$  norm to learn the residue image. The concatenated LR images and its degradation maps are taken by SR network for multiple degradation (SRMD) [166]. The SRMD architecture is same as [111–113]. First, a cascade of  $3 \times 3$  kernel size convolutional layers is used to extract features followed by the series of convolutional, ReLU and the batch normalization layers. In addition, a convolutional procedure is used to extract high resolution sub-images similar to [62], alongside the various high-resolution sub images are transferred to final HR result as the end step. SRMD learns high resolution images directly instead of the image residues. A variant known as super-resolution network for multiple degradation from noise free (SRNDNF) [166] that learns through noise free degradation was also introduced by the authors. The networks from first noise level maps in the convolutional layer is eliminated in SRNDNF network but the rest of the structure is same as SRMD. In comparison to multiscale training, the researchers trained each model for every upsampling scales.  $l_1$  loss is used and the training patch sizes is fixed to  $40 \times 40$ . The number of convolutional layers is fixed at 12 with 128 feature maps in each layer. The training is carried out on 5944 images of BSD dataset [52,53], DIV2K [58] and Waterloo [167]. Early learning rate is set at  $10^{-3}$  that are subsequently reduced to  $10^{-5}$ . Learning rate minimizing parameters are related to change in error between subsequent epochs. The peak signal-to-noise ratio record of previous super resolution networks e.g., EDS, MDSR and CMSC cannot be broken by both the SRMD and its version. However, it offers a unique ability to jointly tackle many degradations.

Deep learning techniques continue to outperform traditional based algorithms in terms of efficiency and effectiveness due to end to end training. Deep learning algorithms carry flexibility which could handle super resolution issue with different scale factors, blur kernels and noise levels inside a unified maximum a posteriori (MAP) framework. Zhang et al. [168] proposes a deep unfolding network for super resolution (DUNSNR) which is an end-to-end trainable unfolding network that uses learning-based algorithms to address these challenges. The fixed number of iterations that alternately addresses the data sub-problem and a previous sub-problem can be obtained by unfolding the MAP inference using a half-quadratic splitting technique. The two sub-problems can be solved using neural modules which result in the trainable iterative network from start to finish. Thus, this network inherits the flexibility of model-based algorithms to super resolve blurry and noisy images for several scale factors using single model while maintaining the benefits of

learning-based algorithms. Recent studies have included many degrading effects in order to reflect real-world scenarios by considering only a single degradation. However, the majority of works either assume a fixed set of degrading effects or train a single network to handle a variety of combinations. Instead of training a single network for wide ranging, degradations are a more practical technique. To meet this need, Xu et al. [169] proposed a unified network that can adapt the variations from both inter image (cross image variations) and intra image (spatial variations). Unlike previous works, this approach utilizes dynamic convolution that is significantly more flexible option for dealing with numerous variations. The proposed unified dynamic convolutional network for variable degradations (UDVD) [169] is assessed on both synthetic and real images with an extensive range of variations in super resolution with non-blind framework.

**Generative Adversarial Network (GAN) Models.** GAN models [170,171] uses the game theoretic technique where two model components viz. a generator and discriminator which attempt to fool the later. A generator obtains SR images which could not be differentiated by a discriminator as a true HR image or an artificial SR image. In general, the corresponding PSNR values gets reduced by highlighting the issue where common objective metrics in SR literature do not encloses the perceptual reception of high-resolution results. GAN-based models for SR proposes in [23,28]. A single image SR is very challenging for large-upscaling factors. SRGAN proposes using the adversarial objective feature where it encourages SR outputs similar to the multiple of natural images. The multitask loss expression consist of three key components: (1) a mean square error loss which encode pixelwise similarity, (2) a perceptual similarity metric with regard to distance metric defined on top of higher-level image representations such as deep network features, and (3) an adversarial loss balancing a min-max game between generator and discriminator such as standard GAN [170]. The architecture of this method essentially favors output which is perceptually same as high dimensional images. They proposed a novel mean square score (MOS) to quantify this capability that are manually assigned by human raters indicates poor/ exceptional performance of each SR image. However other methods typically learn to optimize direct data dependent metrics such as errors. SRGAN [23] overcomes its contenders on perceptual quality metric by a huge margin. In HR super-resolved image, Enhancement [28] focuses on creating faithful texture information. The inability to comply with perceptual quality of the images is an important issue along with standard image performance measures i.e., PSNR. This results in excessively smooth images with no sharp textures. To solve this problem besides the standard pixel level MSE loss, Enhancenet uses two other loss functions: (1) a *perceptual loss function* is specified in the form of  $AptCommandmathcall_1$  distance on the intermediate feature representations the pre-trained network [172]. (2) a *texture matching* loss is used and specified as  $AptCommandmathcall_1$  loss between gram metrics calculated from deep features to match the texture of low resolution and high-resolution images. While a SR network aims it to fool the discriminator network where the entire network design is adversely trained. The framework utilized by Enhancenet is related to full convolutional network [173] and the principle of residual learning [66]. Their findings exhibit that while the finest PSNR is obtained only when the pixel level loss is utilized and further loss function with an adversarial training method contribute to high practical and perceptual performance. However, when super resolving highly textured regions, the proposed adversarial training would generate visible artefacts. The latest work on high perceptual efficiency SR [174] has further explores these shortcomings. Another super-resolution algorithm based on GAN is SRFNet [175] with feature discrimination. The research concentrates on natural perception of the source image using additional discriminator which supports the generator for producing structural features of high frequency instead of noisy artefacts. This requirement is accomplished by differentiating between the synthetic features and the actual images. To extract features, this network utilizes  $9 \times 9$  convolutional layer. Residual block same as [125]

along with long skip connections with  $1 \times 1$  convolution is then used. The feature maps are up sampled by pixel shuffler layer to meet the desired output size. The 16 residual blocks were used by the researchers for two separate feature map settings i.e., 64 and 128. An integration of adversarial /perceptual loss and pixel level (*AptCommandmathcall<sub>2</sub>*) functions optimized with an Adam optimizer in this model [130]. The system's input resolution is  $74 \times 74$  which only produces  $296 \times 296$  images. For pre-training the generator, the network utilizes 120k images from the ImageNet [63], follows by fine tuning an augmented DIV2K dataset [58] and utilizes  $10^{-4}$  to  $10^{-6}$  leaning rates. An enhanced SR generative adversarial network (ESRGAN) [174] construct on SRGAN by eliminating the batch normalization and add dense blocks where the input of each block is also connected to the output individual block by allowing the residual connection over each dense block. To enforce residual learning, ESRGAN also carry a global residual connection (GRC). In addition, an improved discriminator called relativistic GAN (RGAN) [176] is also used by the researchers. The training is carried out on overall 3450 images from DIV2K and Flicker2K [58] using augmentation first through *AptCommandmathcall<sub>1</sub>*-loss feature and further utilizing the trained model with perceptual loss. A patch size for training is set to  $128 \times 128$  and has of 23 blocks. There are five convolutional layers in each block and each having 64 feature maps. The perceptual result is relatively better as compares to RCAN [40] but it fails to quantitative tests where RCAN achieves better result. Kligler *et al.* proposed a “KernelGAN” [177] which is an image specific internal-GAN that learns its internal distribution of patches by training just on the low-resolution test image in test time. Its generator is trained to generate a down-scaled replica of low-resolution test image so that its discriminator could not tell the difference between the patch distribution of the down-scaled image and patch distribution of the original low-resolution image. The generator performs the downscaling GRL means global residual learning, LRL stands for local residual learning and MST stands for multiscale training operation using the correct image specific super-resolution kernel one it gets trained. KernelGAN is completely unsupervised that requires no training data other than the input image itself and when combined with existing SR algorithms, it produces excellent results in Blind-SR. While current approaches perform better on standard datasets however, they often fail to deliver strong results in real word scenes.

Since real world images can be corrupted by things like sensor noise which is severally changed by bicubic down sampling. Thus, the models never observe real world images during training that limit their capabilities to generalize. Furthermore, collecting paired LR and HR images in the same source domain is time-consuming. The down sample generative adversarial network (DSGAN) [178] was developed to solve these issues by introducing natural image properties into bicubically downsampled images. It can be trained on HR images in an unsupervised manner resulting in low resolution images with same features as an origin an image. The data is then used to train the super resolution model which significantly enhances its performance on real world images. This method proposes that low-and high image frequencies be separated and treated differently during training due to down sampling processes preserve the low frequencies whereas it only needs adversarial training to change the high frequencies. This concept is also applicable to SR operations. The perceptual extreme SR (PESR) for SISR is incredibly difficult to perform due to vast differences in texture details between images. A novel SR network with receptive field block based on enhanced SRGAN known as (RFB-ESRGAN) [179] was presented to address above problems. First, this method used receptive field block to extract multiscale information and improve feature discriminability. In terms of object detection and classification, RFB has achieved competitive results. Second, rather than employing large convolution kernels in multiscale receptive field block, RFB employs numerous small kernels which allow to extract detailed features while reducing computing complexity. Third, in the upsampling stage, this method alternately

**Table 1**  
Comparative survey on some of the popular deep learning-based SR algorithms.

Parameters	Techniques	SRCNN End-to-end mapping	VDSR VGG-net	FSCNN End-to-end mapping	SCN LISTA network	DRCN End-to-end mapping	EDSR SRResNet	WMCNN End-to-end mapping	MDSR Multi-scale dense network	LapsRN End-to-end mapping	DCNN End-to-end mapping	ESPCN End-to-end mapping	DRRN ResNet
Architecture													
Epochs	20	10, 80	10	9 $\times$ 9, 5 $\times$ 5	3 $\times$ 3	5	5 $\times$ 5	5	12	1000	5	100	10
Filter size													
Scale Factors	2, 3, 4	2, 3, 4	3, 4	9 $\times$ 9, 5 $\times$ 5, 1 $\times$ 1, 3 $\times$ 3	2, 3, 4	2, 3, 4	2, 4	2, 3	2, 4, 8	2, 4, 8	2 $\times$ 10 <sup>-2</sup> ,	3, 4,	2, 3, 4
Learning Rate	$10^{-4}, 10^{-5}$	$10^{-1}, 10^{-3}$	$10^{-3}, 10^{-4}$	$10^{-2}, 10^{-6}$	$10^{-4}$	$10^{-2}$	$10^{-2}$	$10^{-3}$	$10^{-4}$	$10^{-5}$	$2 \times 10^{-5}$	$10^{-2}, 10^{-4}$	$10^{-1}$
Weight decay	-	1	1	-	BSD100	$10^{-4}$	DIV2K, BSD100, Urban100	RSSCN7	DIV2K	URBAN100, BSDS100, MANGA109	$10^{-4}$	-	BSD300, BSD500, ImageNet, Xiph, Ultra-Video Group
Datasets	91 images, ImageNet	MatConvNet	91 images, General-100, BSD200, BSD500	-	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Optimizer	Adam	Adam	-	MSE	MSE	MSE	MSE	MSE	-	-	MSE	MSE	-
Cost function	MSE	MSE	-	-	-	-	-	-	-	-	0.8	0.8	-
Momentum	-	-	-	-	-	-	-	-	-	-	-	-	0.9
parameter	-	-	-	-	-	-	-	-	-	-	-	-	-
Activation	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	tanh
function	-	-	-	-	-	-	-	-	-	-	-	-	-
Performance	PSNR, SSIM	PSNR, SSIM	PSNR, SSIM, IFC	PSNR, SSIM	PSNR, SSIM	PSNR, SSIM	PSNR, SSIM	PSNR, SSIM	PSNR, SSIM	PSNR	PSNR	PSNR	PSNR, SSIM
matrices	-	-	-	-	-	-	-	-	-	-	-	-	-
Stride	-	-	-	-	-	-	-	-	-	-	-	-	21
Gradient	-	-	-	-	-	-	-	-	-	-	-	-	Stochastic
descent	-	-	-	-	-	-	-	-	-	-	-	-	-
No of Layers	3	20	1	5	20	-	3	20, 30	3	-	-	-	3
													52

utilizes different upsampling algorithms to lower the high computing complexity while maintaining satisfactory performance. Fourth, it employs an ensemble of ten models of different iteration to enhance model resilience and reduce noise caused by each model. However, existing methods are unable to directly optimize indistinguishable perceptual metrics which have been demonstrated to significantly correlate with human perceptions. Zhang et al. [180] proposed a SRGAN with ranker (RankSRGAN) to generative optimize the generator in the direction of distinct perceptual properties to address the problem. This strategy involves first training a ranker which learn the behavior of perceptual metrics and then introduce a new rank content loss to improve perceptual quality. The most attractive aspect is that this technique combines the advantages of various SR methods to generate superior outcomes. This method also proposed two simple and efficient ways for providing multi-dimension restrictions for the generator using either a single ranker or multiple rankers. Experimental results indicated that the RankSRGAN delivers visually appealing results and achieves excellent performance in perceptual metrics and quality. The comparison of different parameters for various deep learning SR methods is shown in Tables 1 and 2. Direct reconstruction algorithms carry-out single step up-sampling to establish low-resolution to high-resolution space however progressive reconstruction predicts high resolution image in numerous upsampling steps. In a long path from input to the output for super-resolution, depth indicates number of conventional and transposed convolutional layers. The global residual learning (GRL) shows that the network is learning the difference between high resolution image of a ground truth and the sampled low-resolution images in order to make sure by applying bicubic interpolation or learned filters. A local

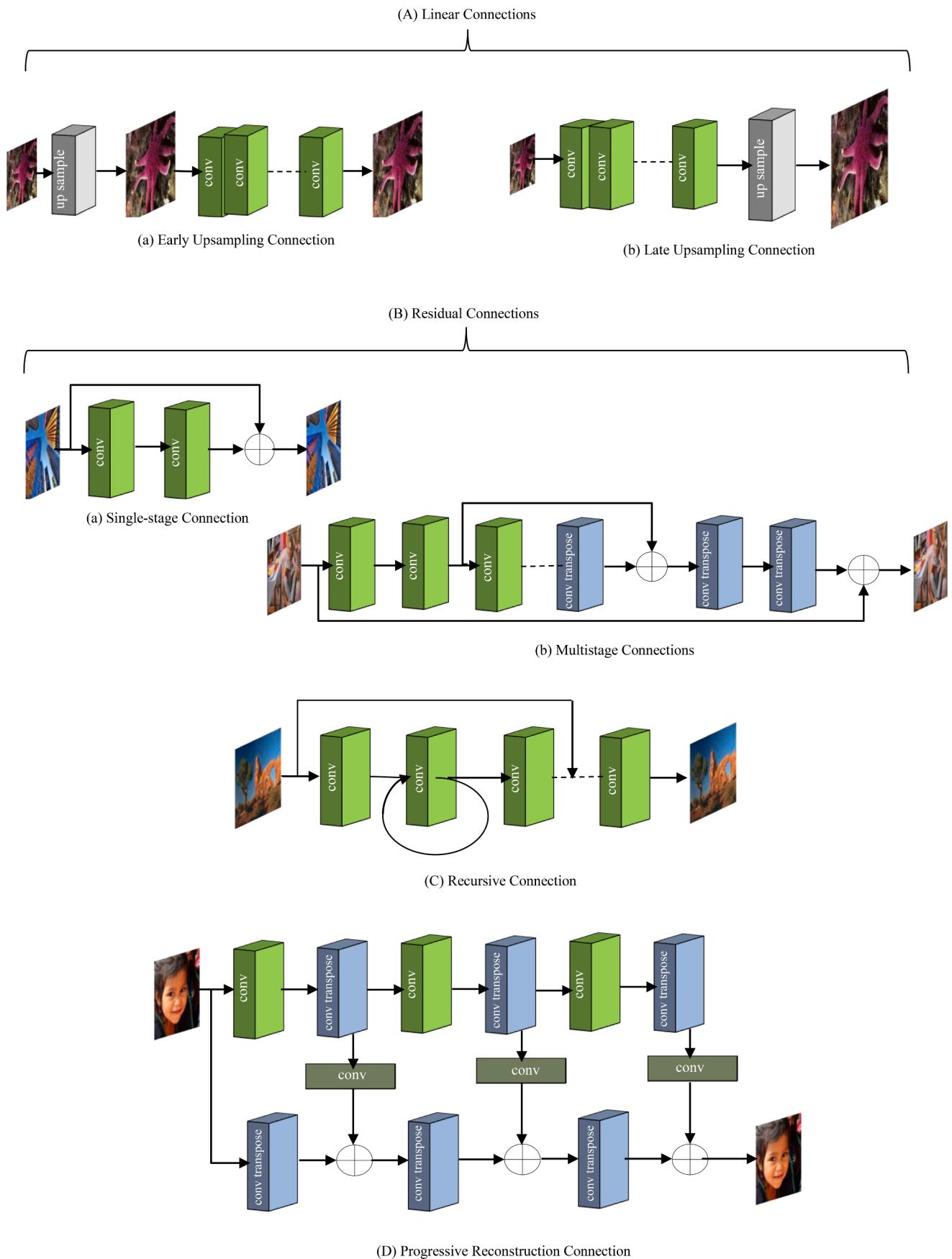
residual learning (LRL) represents a local skip connection between intermediate convolutional layers. The method for performing late upsampling has substantially low-computational costs than method performing early upsampling in the network pipeline. A model based on multiscale generative adversarial network (GAN) was proposed [181]. The multiscale pyramid inside the generator may extracts the features consists of high frequency information and further high-resolution image with the results of bicubic interpolation is reconstructed. The model discriminator is utilized to identify the authenticity of the input image after refactoring. A glimpse of the diverse range of network architectures used for image super-resolution using CNN and GAN-based models are shown in Fig. 2.

### 3.4. Transformer-based methods

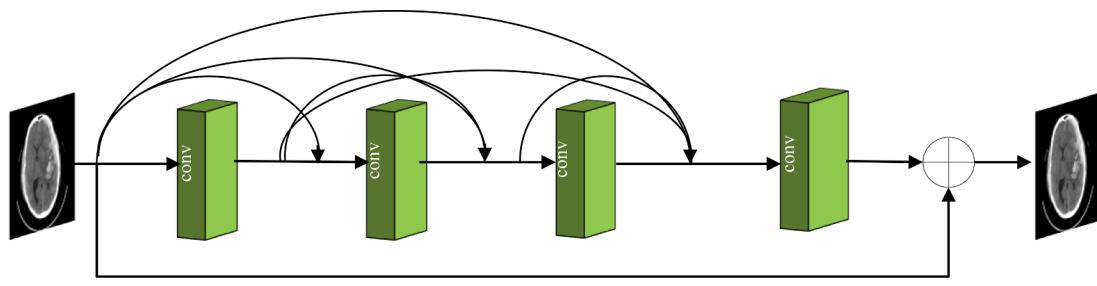
In the domain of computer vision, the problems of state-of-the-art (SOTA) models in natural language processing (NLP) faced to modeled long term dependencies efficiently before transformers. For example, a long-short term memory (LSTM) networks and gated recurrent units (GRUs) are popular recurrent neural networks (RNNs). The architecture of these models makes them robust to exploding and vanishing gradients, a common problem in RNNs which enables them to keep track of quite long dependencies between elements in the sequence, although processing tokens sequentially and relying on keeping their information in memory is not suitable at all when dependencies are distant. This sequential nature also makes them difficult to scale or parallelize efficiently. In particular, convolutions have been popular for NLP tasks due to their efficiency and scalability when trained

**Table 2**  
Detailed comparative analysis on popular deep learning-based SR techniques.

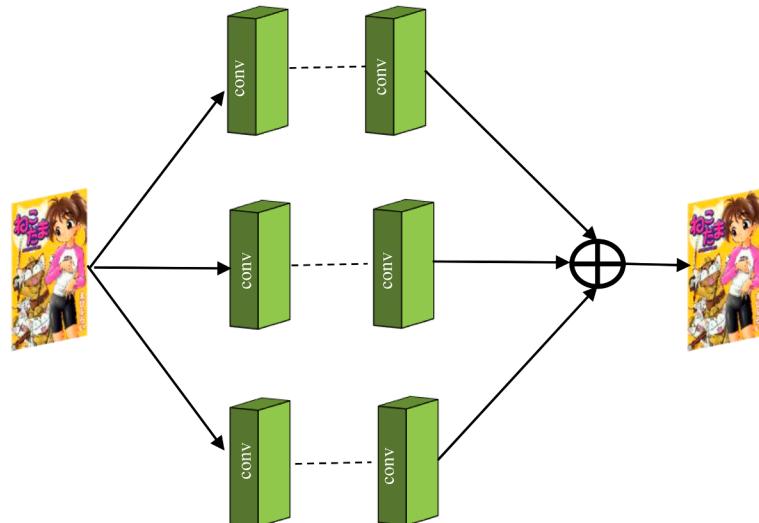
Algorithm	Input	Blocks	Depth	MST	LRL	GRL	Framework	Parameters	Filters	Loss	Output
SCN	bicubic	Yes	10	-	-	-	Cuda- CovNet	42k	128	<i>AptCommandmathcall_2</i>	Prog.
REDNet	bicubic	-	30	-	Yes	Yes	Caffe	4,131k	128	<i>AptCommandmathcall_2</i>	direct
MemNet	bicubic	Yes	80	Yes	Yes	Yes	Caffe	677k	64	<i>AptCommandmathcall_2</i>	direct
RDN	LR	Yes	149	-	Yes	Yes	Torch	21,900k	64	<i>AptCommandmathcall_1</i>	direct
SRFeat	LR	Yes	54	-	Yes	Yes	TensorFlow	6,189k	128	<i>AptCommandmathcall_2, AptCommandmathcall_p, GAN</i>	direct
RCAN	LR	Yes	500	Yes	Yes	Yes	Pytorch	16,000k	64	<i>AptCommandmathcall_1</i>	direct
DRLN	LR	Yes	160	Yes	Yes	Yes	Pytorch	34,000k	64	<i>AptCommandmathcall_1</i>	direct
SRCNN	bicubic	-	3	-	-	-	Caffe	57k	64	<i>AptCommandmathcall_2</i>	direct
ESPCN	LR	-	3	-	-	-	Theano	20k	64	<i>AptCommandmathcall_2</i>	direct
IDN	LR	Yes	31	-	Yes	Yes	Caffe	796k	64	<i>AptCommandmathcall_1, AptCommandmathcall_2</i>	direct
CARN	LR	Yes	32	Yes	Yes	Yes	PyTorch	1,592k	64	<i>AptCommandmathcall_1</i>	direct
SelNet	LR	-	22	-	Yes	Yes	MatConvNet	974k	64	<i>AptCommandmathcall_2</i>	direct
SRRAM	LR	Yes	64	Yes	Yes	Yes	TensorFlow	1,090k	64	<i>AptCommandmathcall_1</i>	direct
EnhanceNet	LR	Yes	24	-	Yes	-	TensorFlow	889k	64	<i>AptCommandmathcall_1, AptCommandmathcall_t, GAN</i>	direct
FSRCNN	LR	-	8	-	-	-	Caffe	12k	56	<i>AptCommandmathcall_2</i>	direct
DRCN	bicubic	-	20	-	-	Yes	Caffe	1,775k	256	<i>AptCommandmathcall_2</i>	direct
IRCNN	bicubic	-	7	Yes	-	-	MatConvNet	188k	64	<i>AptCommandmathcall_2</i>	direct
LapSRN	LR	Yes	24	-	-	Yes	MatConvNet	812k	64	<i>AptCommandmathcall_1</i>	prog.
ESRGAN	LR	Yes	115	-	Yes	Yes	Pytorch	38,549k	64	<i>AptCommandmathcall_1</i>	direct
SRDenseNet	LR	Yes	64	-	Yes	Yes	TensorFlow	5,452k	16-128	<i>AptCommandmathcall_2</i>	direct
VDSR	bicubic	-	20	-	Yes	Yes	Caffe	665k	64	<i>AptCommandmathcall_2</i>	direct
EDSR	LR	Yes	65	-	Yes	Yes	Torch	43,000k	256	<i>AptCommandmathcall_1</i>	direct
BTSRN	LR	Yes	22	-	Yes	Yes	TensorFlow	410k	64	<i>AptCommandmathcall_2</i>	direct
D-DBPN	LR	Yes	46	-	Yes	Yes	Caffe	10,000k	64	<i>AptCommandmathcall_2</i>	direct
ZSSR	LR	-	8	-	-	Yes	TensorFlow	225k	64	<i>AptCommandmathcall_1</i>	direct
MDSR	LR	Yes	162	Yes	Yes	Yes	Torch	8,000k	64	<i>AptCommandmathcall_1</i>	direct
CNF	bicubic	-	15	-	-	-	Caffe	337k	64	<i>AptCommandmathcall_2</i>	direct
CMSC	bicubic	Yes	35	Yes	Yes	Yes	PyTorch	1220k	64	<i>AptCommandmathcall_2</i>	direct
DnCNN	bicubic	-	17	Yes	-	-	MatConvNet	566k	64	<i>AptCommandmathcall_2</i>	direct
EBRN	LR	Yes	173	-	Yes	-	Pytorch	7,900k	64	<i>AptCommandmathcall_1, AptCommandmathcall_2</i>	direct
SRFBN	LR	Yes	28	Yes	Yes	Yes	Pytorch	3,500k	64	<i>AptCommandmathcall_1</i>	direct
FormResNet	bicubic	Yes	20	Yes	-	Yes	MatConvNet	671k	64	<i>AptCommandmathcall_2, AptCommandmathcall_TV</i>	direct
MS-LapSRN	LR	Yes	84	Yes	Yes	Yes	MatConvNet	222k	64	<i>AptCommandmathcall_1</i>	prog.
SRMD	LR	-	112	-	-	-	MatConvNet	1,482k	128	<i>AptCommandmathcall_2</i>	direct
DRRN	bicubic	Yes	52	Yes	Yes	Yes	Caffe	297k	128	<i>AptCommandmathcall_2</i>	direct
SRGAN	LR	Yes	52	Yes	Yes	Yes	Caffe	297k	128	<i>AptCommandmathcall_2</i>	direct



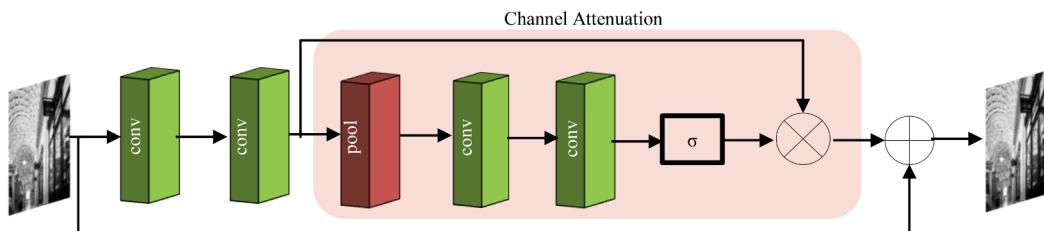
**Fig. 2.** Network architectures of CNN and GAN-based models.



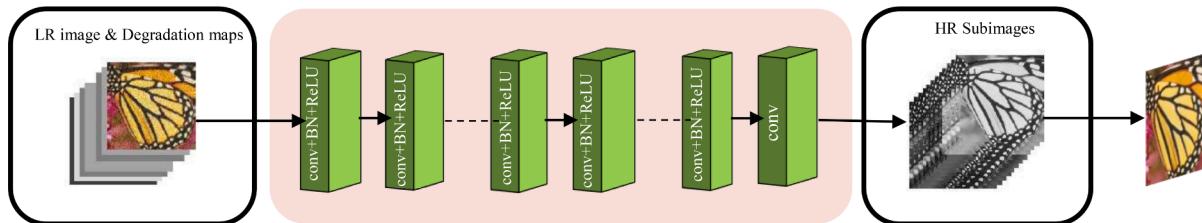
(E) Densely Connected Network



(F) Multi-branch Connection



(G) Attenuation-based Connection

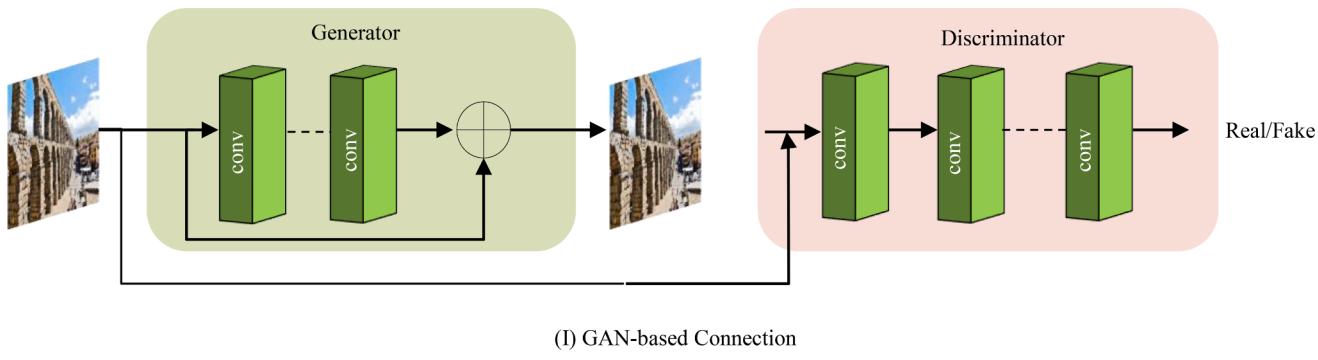


(H) Multiple Degradation Handling Connection

**Fig. 2. (continued).**

using graphics processing units (GPUs). In the same way that 2D convolutions can extract features from an image, these models use 1D filters to extract information from texts, which are represented as a 1D sequence. The receptive field on these kinds of CNNs depends on the size of their filters and the number of convolutional layers used. Increasing the value of these hyperparameters increases the complexity

of the model, which can produce vanishing gradients or even models impossible to train. Residual connections and diluted convolutions have also been used to increase the receptive fields of these models, but the way convolutions operate over texts always present limitations and tradeoffs on the receptive field that it can capture. Deep learning based performs with remarkable performance in computer vision such as



(I) GAN-based Connection

Fig. 2. (continued).

RNNs, CNNs and GANs, but most of the models focus on building more complex network with a large numbers of layers which can entail heavy computational costs and memory storage.

Transformers appeared in 2017 as a simple and scale way to obtain

SOTA results in language translation. They were soon used in other NLP tasks becoming a new SOTA of several benchmarks [182]. It is common to train large versions of these models and fine tune them for different tasks, so they are useful even when the data is scarce. Performance in

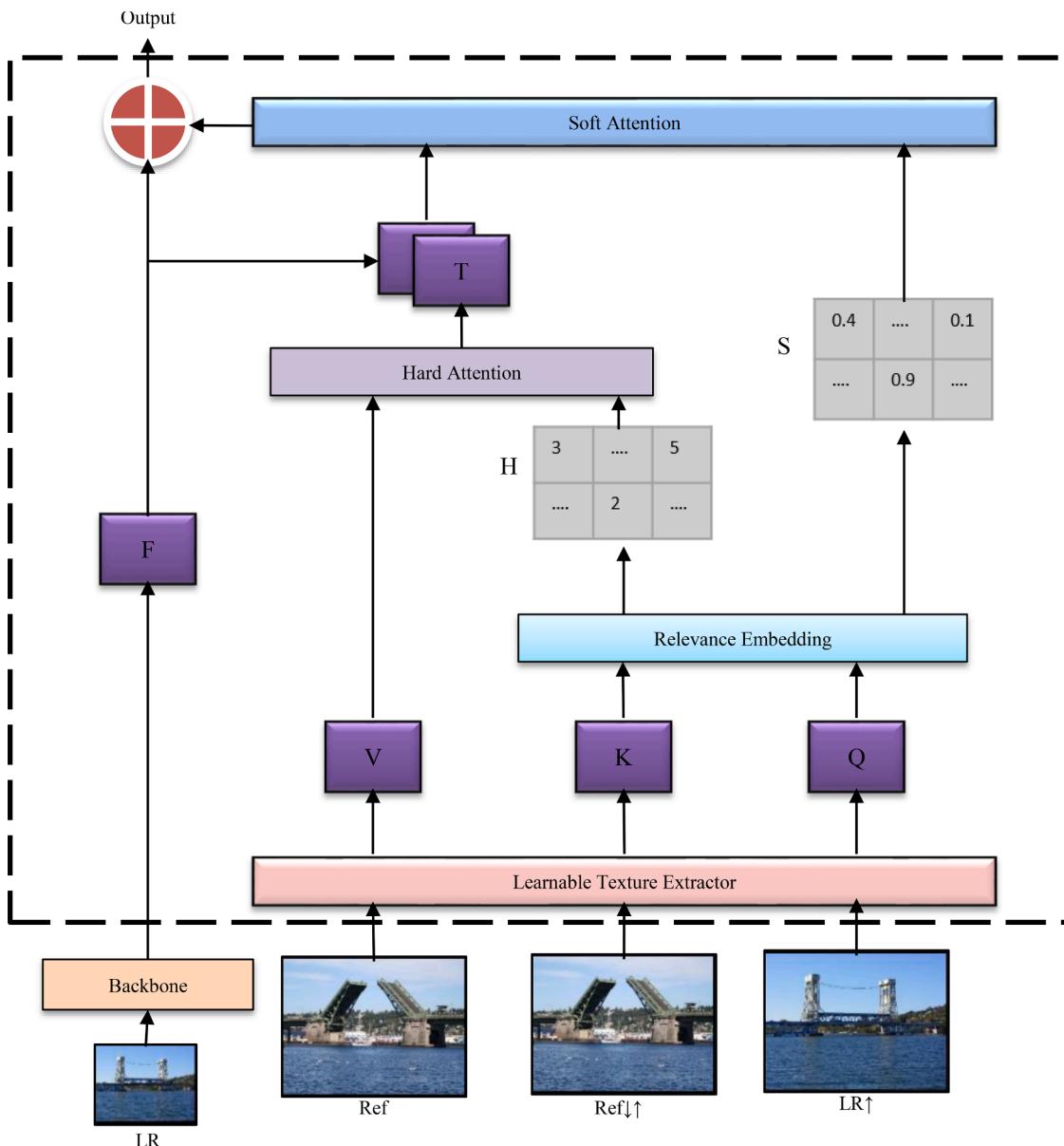


Fig. 4. Transformer-based architecture for image super-resolution.

these models, even with billions of parameters, do not seem to saturate. The larger the model, the more accurate the results are, and the more interesting the emerging knowledge that the model presents. In parallel to how transformers leveraged self-attention to modelized long range dependencies in a text, novel works have presented techniques that use self-attention to overcome the limitations presented by inductive convolutions biases in an efficient way. These works have already shown promising results in multiple computer vision benchmarks in fields such as object detection, video-classification, image classification, image generation, etc. Some of these architectures can match or outperforms SOTA results even when getting rid of convolutional layers and relying solely on self-attention. In recent years, numerous transformers-based works has been emerged in the domain of super resolution. Yang *et al.* proposes a texture transformer network for image super resolution (TTSR) [183] as shown in Fig. 4, where the LR and HR images are formulated as queries and the keys in a transformer, respectively. It contains of four closely based modules optimized for image generations tasks. The learnable texture extractor by DNN, a hard attention module for transfer, a relevant embedding module and a soft attention module for texture synthesis constituted TTSR which is optimized for image production tasks. This model encourages collaborative feature learning across low resolution and Ref images allowing attention to uncover deep feature correspondences and appropriate texture characteristics to be transferred. The proposed texture converter could be stacked in the cross-scale manner to recover textures from different factors such as  $1 \times$  to  $4 \times$  magnification. The lightweight self-calibrated efficient transformer (SCET) [184] was proposed for image super resolution, where architecture of SCET are mainly contains of the self-calibrated module and efficient transfer block, where the self-calibrated module adopts the pixel attention mechanism to extract image features effectively. Further, a transformer-based architecture known as Fusformer [185] are proposed for the hyperspectral image super resolution. Due to the excellent potentiality of feature representations, especially by the self-attention (SA) in the transformer, this method can globally explore the intrinsic relationship within features. This method estimates the spatial residual between the up sampled LR multispectral image and the desired HR multispectral image, thus reduces the burden of training the whole data in a smaller mapping space. Recently, a novel Swin transformer V2 [186] was proposed for improving SwinIR for image super-resolution, and in particular, the compressed input scenario. Through this method, it can tackle the major problems in training transformer vision models, such as training insatiability, resolution gaps between pre-training and fine-tuning, and hunger on data. Cai *et al.* presents a novel transformer architecture based on hierarchical patch (HIPA) [187] that progressively recovers the HR image using a hierarchical patch partition. In addition, a hybridization of deep learning and transformer models are emerged successful in the domain of image super-resolution tasks in recent years [188–190].

### 3.5. Learning strategies

#### 3.5.1. Loss functions

Loss function is utilized in the SR discipline to estimate reconstruction error and guide model optimization. At initial researchers generally utilizes pixelwise  $L_2$  loss but later discovered that, it could not reliably quantify reconstruction quality. Thus, a variety of loss function is utilized to improve the measurement reconstruction error and produce more realistic along with high-quality results such as content loss [172], adversarial loss [23], reconstruction loss [191]. These loss functions are increasingly playing key role in recent years. In this section, we will explore some most widely utilized loss functions:

**Reconstruction loss.** The first loss is the reconstruction loss. It is defined as

$$L_{rec} = \frac{1}{CHW} \| I^{HR} - I^{SR} \|_1, \quad (12)$$

where  $(C, H, W)$  is the size of the HR. In comparison to  $L_2$  loss, we use  $L_1$  loss, which has been shown to be sharper in terms of performance and easier to convergence.

**Texture loss.** The texture loss also known as style reconstruction loss is introduces in SR since the reconstructed image should carry same style such as colors, textures, contrast as the target image and it is inspired by the style representation from Gatys *et al.* [191,192]. An image texture is defined as the relationship between different feature channels according to [191,192] and specified as the Gram matrix  $G^{(l)} \in R^{c_l \times c_l}$ , where  $G_{ij}^{(l)}$  is the inner product between the vectorized feature maps  $i$  and  $j$  on layer  $l$ :

$$G_{ij}^{(l)}(I) = \text{vec} \left( \phi_i^{(l)}(I) \right) \cdot \text{vec} \left( \phi_j^{(l)}(I) \right), \quad (13)$$

where  $\phi_i^{(l)}(I)$  indicates the  $i - th$  channel of the feature maps on layer  $l$  of  $I$  and  $\text{vec}(\cdot)$  denotes the vectorization process. The texture loss is then denoted by:

$$\mathcal{L}_{\text{texture}}(\hat{I}, I; \phi, l) = \frac{1}{c_l^2} \sqrt{\sum_{i,j} \left( G_{ij}^{(l)}(\hat{I}) - G_{ij}^{(l)}(I) \right)^2} \quad (14)$$

The Enhancement [28] proposes by Sajjadi *et al.* provides far highly natural textures and generates visually pleasing results by utilizing texture loss. However, to compute the patch size to match textures is still based on trial and error. Since texture statistics are averaged over areas of different textures since excessively small patches produce an artifact in textured regions and excessively large patches cause artifacts throughout the image.

**Conventional loss.** Most SR approaches use a common pixelwise loss to optimize the intricately designed networks that is effective for super resolution tasks as evaluated by PSNR. The average pixel difference between reconstructed images and original images can be reduced using this metric however the results may be excessive smooth to preserve sharp edges for visual effects. Further this loss is still broadly utilized to improve SR performance and accelerate convergence:

$$L_{SR}^{Pix} = E_{SR} \| G(I^{LR}) - I^{HR} \|_1 \quad (15)$$

Johnson *et al.* [172] proposes the perceptual loss to increase the perceptual quality of reconstructed images. A pre-trained VGG network [115] extracts features containing semantic information. In perceptual loss, the Euclidean distance between the features of HR and super-resolution images are minimized.

$$L_{SR}^{Per} = E_{SR} \| \phi_i(G(I^{LR})) - \phi_i(I^{HR}) \|_1 \quad (16)$$

where  $\phi_i(\cdot)$  represents  $i$ -th layer output of the VGG model. In the SR problem, methods based on generative adversarial networks (GANs) [170,171,176,193–195] also play an important role. A two-player game is used to optimize the discriminator  $D_I$  and the generator  $G$  as follows:

$$L_{SR}^{DisI} = -E_{SR} [\log(1 - D_I(I^{SR}))] - E_{I^{HR}} [\log D_I(I^{HR})], \quad (17)$$

$$L_{SR}^{AdvI} = -E_{SR} [\log D_I(G(I^{LR}))] \quad (18)$$

We use relativistic average GAN (RaGAN) to obtain superior optimization in practice as described in [174,176]. The image-space constraint for images is only considered by models supervised by the above objective functions but the semantically structured information provided by the gradient space is ignored. While the obtained results appear to be photorealistic, this method still have numbers of geometric distortions which is undesirable. Thus, the gradient loss has been introduced to address this problem.

**Gradient loss.** The gradient loss [196] is formulated by reducing the distance between the gradient maps extracted from SR image and one form the associated HR image. The generator may learn fine appearance and attach priority to avoid detailed geometric distortions under supervision in both images and gradient domains. Thus, the method

constructs two loss terms to penalize the difference between HR images and gradient maps of super resolution. The one is based on pixelwise loss as shown below:

$$L_{SR}^{PixGM} = \mathbb{E}_{I^{SR}} \| M(G(I^{LR})) - M(I^{HR}) \|_1 \quad (19)$$

The other is to differentiate whether a gradient patch is from the high-resolution gradient map or not. To accomplish this, a new gradient discriminator network is introduced as:

$$L_{SR}^{DisGM} = -\mathbb{E}_{I^{SR}} [\log(1 - D_{GM}(M(I^{SR})))] - \mathbb{E}_{I^{HR}} [\log D_{GM}(M(I^{HR}))] \quad (20)$$

By using adversarial learning, the gradient discriminator can also supervise the generation of super resolution outcomes.

$$L_{SR}^{AdvGM} = -\mathbb{E}_{I^{SR}} [\log D_{GM}(M(G(I^{LR})))] \quad (21)$$

Note that each step of  $M(\cdot)$  process is differentiable. Thus, the model with gradient loss could be trained in end-to-end procedure. Gradient loss is also convenient to use as supplementary guidance in any generative because of its concise formulation and high transferability.

**Cyclic Consistency Loss.** Inspired from CycleGAN introduced by Zhu et al. [197], Yuan et al. [198] proposes a cycle-to-cycle technique for SR. This method not only super-resolve LR image  $I$  to HR image  $\hat{I}$  however also down-sample  $\hat{I}$  through another CNN back to another LR image  $I'$ . Since the regenerated  $I'$  must be same to the  $I$ , a cycle consistency loss is used to limit their pixel level consistency.

$$\mathcal{L}_{cycle}(I', I) = \frac{1}{hwc} \sqrt{\sum_{i,j,k} (I'_{i,j,k} - I_{i,j,k})^2} \quad (22)$$

**Pixel Loss.** Pixel loss is a metric that calculates pixelwise difference between two images and mostly include  $L_1$  loss such as mean absolute error and  $L_2$  loss such as mean square error.

$$\mathcal{L}_{pixel-l1}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} |\hat{I}_{i,j,k} - I_{i,j,k}| \quad (23)$$

$$\mathcal{L}_{pixel-l2}(\hat{I}, I) = \frac{1}{hwc} \sum_{i,j,k} (\hat{I}_{i,j,k} - I_{i,j,k})^2 \quad (24)$$

The height, weight and number of channels of reconstructed images are represented by  $h$ ,  $w$  and  $c$ , respectively. In addition, there is a Charbonnier loss [67,199] which is the version of pixel  $L_1$  loss defined as

$$L_{pixel} \underset{Cha(I, I)}{\sim} = \frac{1}{hwc} \sum_{i,j,k} \sqrt{(\hat{I}_{i,j,k} - I_{i,j,k})^2 + \epsilon^2} \quad (25)$$

For numerical stability where  $\epsilon$  is the constant (e.g.,  $10^{-3}$ ). On the pixel values, the pixel loss constraints the obtained high-resolution image  $\hat{I}$  to be close enough to original image  $I$ . When compared to  $L_1$  loss,  $L_2$  loss prohibits greater errors while being more tolerant on small errors which result in overly smooth output. In fact, the  $L_1$  loss outperforms the  $L_2$  loss in terms of performance and convergence [70,126, 200]. Since the pixel-wise difference is significantly correlated with the description of PSNR (Section 2.2.1) and thus decreasing pixel loss directly maximizes PSNR while pixel loss gently become most often utilized loss function. But since the pixel loss ignores image quality such as perceptual quality [172] and textures [28], the results often lack high frequency features and perceptually unsatisfactory with overly smooth textures [23,24,46,172].

**Perceptual Loss.** This loss has been introduced to improve visual quality and has employed previously in [23,24,28,201]. The main goal of the perceptual loss is to improve the similarity in feature space between predicted image and the target image. There are two parts in perceptual loss specified as:

$$L_{per} = \frac{1}{C_i H_i W_i} \| \phi_i^{vgg}(I^{SR}) - \phi_i^{vgg}(I^{HR}) \|_2^2 + \frac{1}{C_j H_j W_j} \| \phi_j^{lre}(I^{SR}) - T_2^2 \| \quad (26)$$

where the first part is a typical perceptual loss and  $(C_i, H_i, W_i)$  indicates the shape of the feature map at that layer and  $\phi_i^{vgg}(\cdot)$  represents the  $i$ -th feature map layers of VGG19. The predicted super resolution image is  $I^{SR}$ . The second part of perceptual loss is a transferred perceptual loss where  $\phi_j^{lre}(\cdot)$  represents the texture map extracted from the  $j$ -th layer of proposed LTE and  $(C_i, H_i, W_i)$  represents the shape of that layer.  $T$  stands for transferred high resolution texture features. This transferred perceptual loss restricts the predicted super resolution image to have same texture features to the transferred feature  $T$  making SR approaches for transferring the Ref textures more effectively.

**Adversarial loss.** The effectiveness of generative adversarial networks [170] in generating clear and visually appealing images has been proven. Furthermore, WGAN-GP [195] proposes replacing weight clipping with a penalization of gradient norm resulting in more stable training and improving results. The loss could be defined as:

$$L_G = \mathbb{E}_{\tilde{x} \sim P_g} [D(\tilde{x})] - \mathbb{E}_{\tilde{x} \sim P_r} [D(x)] + \lambda \mathbb{E}_{\tilde{x} \sim P_g} [(\| \nabla_{\tilde{x}} D(\tilde{x}) \|_2 - 1)^2] \quad (27)$$

$$L_G = \mathbb{E}_{\tilde{x} \sim P_g} [D(\tilde{x})] \quad (28)$$

**Content loss.** It is introduced in SR [172,195] to evaluate the perceptual quality of an image. It uses a pre-trained image classification network to estimate the semantic differences between images. This loss is expressed as the Euclidean distance between high level representations of two images of a network as  $\phi$  and the extracted high-level representation on

$$l-th \text{ layer as } \phi^{(l)}(I) : \mathcal{L}_{content}(\hat{I}, I; \phi, l) = \frac{1}{h_l w_l c_l} \sqrt{\sum_{i,j,k} (\phi_{i,j,k}^{(l)}(\hat{I}) - \phi_{i,j,k}^{(l)}(I))^2} \quad (29)$$

The height, width and number of channels representing layer  $l$  are denoted by  $h_l$ ,  $w_l$  and  $c_l$ , respectively. This loss essentially transfers learned hierarchical image feature knowledge from classification network  $\phi$  to super resolution network. Unlike pixel loss, content loss facilitates output image  $\hat{I}$  to be perceptually same as target image  $I$  rather than forcing them to match pixels exactly. Thus, it generates visually highly perceptible performance and broadly utilized in the domain such as [23,28,43,172] where the most typically utilized pre-trained CNNs are VGG [115] and ResNet [125].

**Prior based Loss.** Besides above-mentioned loss functions, external prior knowledge is used to restrict the generation. Li et al. [3] concentrates on face super-resolution and presented a FAN (face alignment network) to limit a consistency of facial landmarks. A FAN is pretrained and combined to provide face alignment priors, both super-resolution and FAN are trained together. Thus, the mentioned super-FAN increases both low resolution face alignment and face image SR performance. In reality both texture loss and content loss where both uses the classification network which generally generate prior knowledge of hierarchical image characteristics for super-resolution. The SR performance can be increased even further by introducing more prior knowledge. In practices for restricting different aspects of the generation operation, researcher often integrate multiple loss functions via weighted average [23,28,29,67,202] notably for the distortion-perception trade-off [23, 174,203]. Further weights of the various loss functions need a great deal of empirical investigation and to determine how to integrate them in a reasonable and effective manner remains a challenge.

**Total Variation Loss.** Aly et al. [204] introduced a total variation loss [21] in super-resolution to reduce noise in reconstructed images. It

estimates how much noise in the images and is interpreted as the sum of the difference between neighboring pixels.

$$\mathcal{L}_{TV}(\hat{I}) = \frac{1}{hwc} \sum_{i,j,k} \sqrt{(\hat{I}_{i,j+1,k} - \hat{I}_{i,j,k})^2 + (\hat{I}_{i+1,j,k} - \hat{I}_{i,j,k})^2} \quad (30)$$

The TV loss is also adopted by Lie et al. and Yuan et al. [128] to introduce spatial smoothness.

### 3.5.2. Batch normalization

Sergey et al. [205] introduced batch normalization (BN) to decrease internal covariate shift to expedite and stabilize deep CNN training. To preserve the representation ability, they execute normalization for each mini batch and train two additional transformation measures for each channel. BN allows for using high learning rates and less cautious initialization because it calibrates the intermediate feature distortion and reduces vanishing gradients. Thus, SR models [23,69,139,166,206] widely uses this technique. Furthermore, Lin et al. [15] claims that BN losses scale details of each image and eliminates network range flexibility. Thus, they eliminate BN and uses the reduced memory cost up to 40% to construct a much larger model thus resulting in a significant gain in performance sustainability. Other models such as [174,207] incorporate this experience and increase performance.

### 3.5.3. Curriculum learning

Starting with a simple task and gradually increases the difficulties in terms as curriculum learning [141]. Since SR is an ill posed problem that is prone to problems like large scale factors, noise and blurring, the curriculum learning is combined to make learning easier. Wang et al. [208], Bei et al. [209] and Ahn et al. [126] introduces ProSR, ADRSR and progressive CARN to minimize the difficulties of SR with large scaling factors. ProSR, ADRSR and progressive CARN are progressive not only on structures but also non training process. Training starts from  $2 \times$  upscaling factor and after completing the training, the parts with  $4 \times$  or large scaling factors are gradually mounted and blended with prior parts. The ProSR blends by integrating the output of this level with up-sampled output of previous levels [210], the ADRSR concatenate them and adds another convolutional layer. Similarly, the progressive CARN replaces the previous reconstruction block with one that generates an image in double resolution. Further, Park et al. [211] divides  $8 \times$  SR problem into three sub problems such as  $1 \times$  to  $2 \times$ ,  $2 \times$  to  $4 \times$  and  $4 \times$  to  $8 \times$  and separate networks for each problem. Then the first two are concatenated and fine-tuned followed by third. They also decompose the  $4 \times$  SR through tough conditions into  $1 \times$  to  $2 \times$ ,  $2 \times$  to  $4 \times$  and denoising or deblurring sub problems. However, SRFBN [141] employs this technique for super resolution under adverse conditions which start from simple degradation and gradually increases the complexity of degradation. Curriculum learning when compared to traditional training methods considerably lessen the training difficulties and reduces the total training time particularly for large scaling factors.

### 3.5.4. Multi-supervision

Multi-supervision is the process of incorporating multiple supervision signals into a model to improve gradient propagation and avoid vanishing and bursting gradients. The DRCN [118] incorporates multi-supervision with recursive units to avoid the gradient difficulties introduced by recursive learning. They feed each recursive output unit in the reconstruction module to produce a high-resolution image and then combine all of the intermediate reconstructions to get the final prediction. MemNet [139] and DSRN [212] which is also based on recursive training uses similar techniques. Further, since the LapSRN [67] produces intermediate outputs of multiple scales during the propagation under progressive upsampling framework, it is simple to implement a multi-supervision method. In particular, the intermediate findings are forced to match the intermediate images down sampled from the original high-resolution images. In practical, this multi

supervision approach is enacted by including some terms in the loss function which allows the supervision signals to be back propagated more efficiently thus reducing the training complexity and improving model training.

## 4. Experimental settings

Experimental environment consists of hardware devices and software configurations. The PC configuration is Intel(R) Core (TM) i7 6700 CPU@ 3.40 GHz and an NVIDIA GeForce GTX 1050-Ti GPU. The experimental platform is equipped with 64-bit windows 10, Caffe, MatlabR2019b, CUDA Toolkit v8.0 and Anaconda2. For deep learning approaches, the entire network in our setup includes the training network and the testing network. In the iteration process of a training network, the basic learning rate is  $10^{-4}$ , the weight decay is zero and the momentum is 0.9. Mini-batch gradient descent (MBGD) is the gradient descent algorithm we've chosen during the experiment.

### 4.1. Datasets and metrics

We evaluated existing methods on the publicly available benchmark datasets such as FFHQ [59], Set14 [51], BSD100 [53], Urban100 [54], DIV2K [58], and Medical Dataset [57]. Fig. 3 demonstrates the representative images from all of the datasets. The publicly available benchmark datasets are used in the training phase that contain Set14 [51] which comprises of 14 test photos. BSD100 [53] dataset proposed by Martin et al. which is a classical dataset contain 100 test images. The dataset contains a wide range of images including nature images as well as object specific images such as people, plants, and food, Urban100 [54] dataset proposed by Huang et al. which is relatively more recent dataset. The total number of image s is same as that of BSD100 [53] but the composition is completely different. The images focus on man-made structures such as urban scenes. The DIV2K dataset [58] is used in the NITRE challenge. The images have a resolution of 2K and are made up of 800 images for training and 100 images for testing along with validation. Flickr-Faces-HQ (FFHQ) [59] is a high-resolution image dataset of human face. The dataset contains 70,000 high quality PNG images resolution of  $1024 \times 1024$  pixels with a wide range of ethnicity, age, and image background. It covers good coverage of accessories such as eyeglasses, hats, sunglasses etc. IDI (I Do Imaging) [57] is a publicly available dataset for medical dataset. IDI is a resource location to find the free and open-source medical imaging software with approximately 300 projects that are nicely categorized, ranked, and searchable. Testing is carried out using the same above datasets. Existing algorithms are implemented using the available codes in the computer with the GTX 1050-Ti. Table 3 presented the experimental results. The PSNR and SSIM [25] has been used as an evaluation metrics to examine and compare the reconstruction results quantitatively.

### 4.2. Evaluation results and analysis

In order to investigate and compare the current state-of-the-art procedures, we have considered fourteen existing super resolution methods including Bicubic, GESR [94], MMSR [93], CCR [105], DualDic [107], RFISR [79], HTSR [109], ALS [25], SRCNN [113], WMCNN [121], SRGAN [23], FSRCNN [116], SRMD [166], LapSRN [67], and WLSCNN [122]. All the experiments are conducted with scaling factors  $2 \times$ ,  $3 \times$ , and  $4 \times$  between low resolution and high-resolution images.

**Qualitative Evaluation.** As shown in Figs. 5 and 6, some of the algorithms performs better in terms of visual quality. In the Figs. 5 and 6, deep learning-based algorithms performed better in terms of visual performance such as SRGAN, SRMD, WLSCNN, FSRCNN and LapSRN. These methods are able to transfers more appropriate high-resolution textures from reference image to produce best results. Also, the deep learning-based algorithms are able to extract fine textures from local regions and able to transfers important textures into final SR output. As



**Fig. 3.** Image representatives from SR datasets used for training and testing algorithms.

observed from the visual results some of the traditional algorithms are also able to preserve spatial details of the images. We demonstrate the experimental results of all fourteen comparative existing algorithms to verify the superiority of their visual quality.

**Quantitative Evaluation.** The quantitative evaluation findings are shown in Table 3. The highest scores are indicated by red colors while the second highest scores are indicated by blue colors. On all six testing datasets, WLSCNN, LapSRN and SRMD produces higher objective metrics values than other recent SR algorithms as shown in the comparative results. In most of the cases, WLSCNN produces better performance among other approaches in case of most of the datasets, however, the values are marginally higher than other methods. The quantitative comparison results indicate that the recent SR methods have shown huge improvements in case of performance evaluation metrics results as compared to previous benchmark SR algorithms.

#### 4.3. Discussion

This study illustrates the vital need for research in the domain of image super-resolution [52,213]. Numerous techniques for improving the resolution of the images were studied in this survey and categorized based on four groups. Most of the algorithms are compared based on visual quality and quantitative metrics performance such as peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) measures as discussed in Section 2.2.1. The methods are presented the results for  $2 \times$ ,  $3 \times$  and  $4 \times$  and  $8 \times$  super resolution. The PSNR and SSIM results is better for recent deep learning-based methods compared to other methods. But it is challenging to declare one or two methods to be clear winner as compares to rest as there are numerous factors involved such as depth of the network, training data, network complexity, number of feature maps, patch size for training, etc. The fair comparison is only possible by keeping all the parameters consistent. We also discussed the visual quality comparison between the state-of-the-art

methods that aims to enhance the resolution of the images. We have considered traditional, convolutional neural network (CNN) [143,160] and generative adversarial network (GAN) [170] based algorithms for comparison. Thanks to the power of deep CNNs, the reconstruction efficiency is heavily improved. But the high-frequency information and low-resolution images still can hardly be reconstructed. The GAN based methods [174,180] which are perceptually driven and aims to improve the visual quality far better. As one can observed, the results obtained in case of most methods are generally crisper for some methods, but their PSNR and SSIM scores are relatively poor when compares to algorithms that optimize pixel level measures. All the competing algorithms are generally evaluated on the standard datasets up to  $2 \times$ ,  $3 \times$ ,  $4 \times$  and  $8 \times$  super resolution. When these methods are tested for higher magnification levels, the artifacts in the images became more visible. It is clear from the images that most of the state-of-the-art algorithms struggle to reproduce the textures in high magnified versions of the images. Also, the most popular choices for network loss are either mean square error  $\text{AptCommandmathcall}_2$  or mean absolute error  $\text{AptCommandmathcall}_1$  in the CNN.

Likewise, the GANs also uses perceptual loss or adversarial loss in addition to the pixel level losses, namely the MSE. It is evident that the initial CNN models [1,70,116] were trained utilizing  $\text{AptCommandmathcall}_2$  loss. And it gives poor results. However, there is shift in the trend towards  $\text{AptCommandmathcall}_1$  more presently and absolute mean difference measure  $\text{AptCommandmathcall}_1$  had demonstrated more robust as compares to  $\text{AptCommandmathcall}_2$ . Since  $\text{AptCommandmathcall}_2$  puts more emphasis on more erroneous predictions while  $\text{AptCommandmathcall}_1$  considers a more balanced error distribution. From the experimental results on all six testing datasets, the recent methods are able to produces higher objective metrics values than other state-of-the-art algorithms. In most of the cases, recent deep learning methods produces better performance among other algorithms in case of most of the datasets, however, the values are marginally

**Table 3**

The PSNR/SSIM of different SR methods are compared on six different datasets.

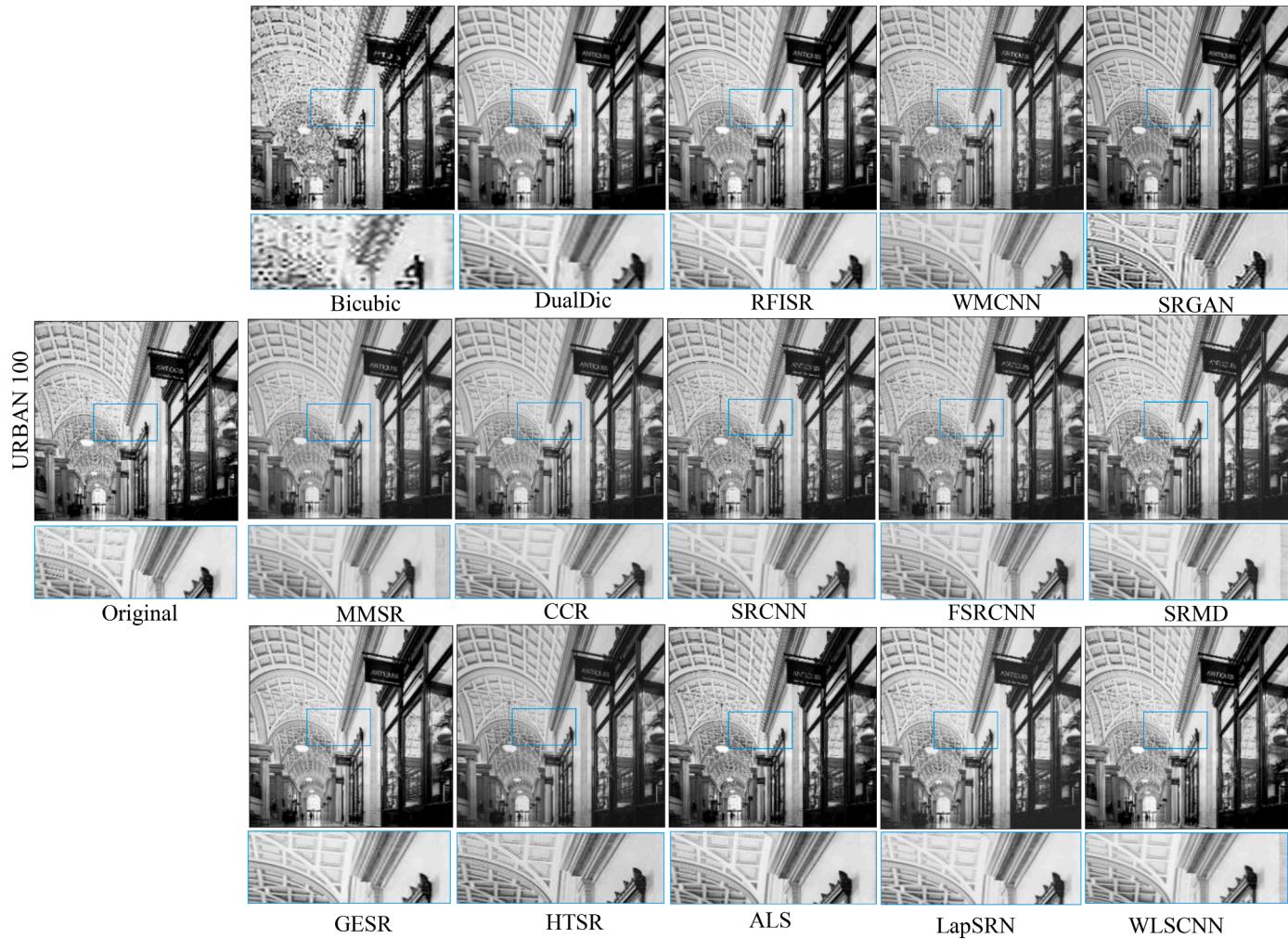
Method	Scale	Urban100		BSD100		Medical Dataset		FFHQ	Set14		DIV2K		
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		PSNR	SSIM	PSNR	SSIM	
Bicubic	×2	20.4474	0.6919	26.4012	0.7281	27.8934	0.8613	33.4927	0.8501	25.9982	0.7628	25.1337	0.6316
GESR		22.8826	0.7965	28.7262	0.7982	31.1924	0.9641	35.8822	0.9082	26.8822	0.8072	26.9827	0.7462
MMSR		21.7726	0.8262	28.8826	0.8266	28.7622	0.9366	36.9827	0.8972	27.7722	0.8399	27.9927	0.7172
CCR		22.8826	0.8109	28.9922	0.7825	31.8826	0.9525	37.8726	0.9276	27.8811	0.8466	26.8827	0.7255
DualDic		21.7722	0.7862	30.8722	0.8027	29.8762	0.9025	35.8827	0.9266	27.8826	0.8162	27.8726	0.6962
RFISR		21.8049	0.6737	28.8385	0.775	30.6171	0.8545	35.9967	0.8126	28.0041	0.7315	26.6133	0.6127
HTSR		24.1632	0.8234	30.9822	0.8452	32.8826	0.9266	36.8826	0.8826	26.8826	0.8263	27.7726	0.7452
ALS		24.7363	0.7835	29.8722	0.8325	34.8827	0.9562	36.8826	0.9352	27.8822	0.8162	26.9922	0.7562
SRCNN		23.2328	0.856	30.5772	0.8733	33.575	0.9558	38.0779	0.9482	28.6153	0.8274	28.2801	0.7651
WMCNN		23.3673	0.8632	30.6568	0.8727	33.7953	0.961	38.1029	0.9489	26.623	0.8378	28.0795	0.7761
SRGAN		24.6524	0.8062	29.9827	0.8562	31.9827	0.9672	35.8826	0.9572	27.8922	0.8562	27.2822	0.7672
FSRCNN		23.4437	0.8645	30.6917	0.8719	32.5049	0.963	36.2078	0.9463	27.3076	0.8463	28.188	0.7676
SRMD		24.8826	0.8427	30.8726	0.8672	32.8826	0.9762	36.7726	0.9672	28.8826	0.8462	27.9272	0.7725
LapSRN		23.5913	0.8309	29.4745	0.8534	32.3435	0.9631	37.6439	0.9355	27.5669	0.8699	28.5235	0.7875
WLSCNN		24.9964	0.8663	31.1338	0.875	33.9662	0.9724	38.7382	0.9605	28.101	0.8774	28.3883	0.7861
Urban100		BSD100		Medical Dataset		FFHQ		Set14		DIV2K			
Method	Scale	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	×3	18.5135	0.6465	23.5895	0.6812	24.6654	0.8489	31.2688	0.8006	23.2487	0.7053	22.582	0.5664
GESR		20.7762	0.6762	26.8827	0.7462	27.8826	0.8926	32.9827	0.9162	25.8827	0.7477	25.9927	0.6288
MMSR		21.9927	0.7255	25.0928	0.7562	28.8826	0.9066	31.8171	0.9027	25.8826	0.7492	24.8827	0.6472
CCR		22.8822	0.7152	24.8822	0.7455	29.8826	0.9026	32.9927	0.9176	26.8822	0.7328	25.8826	0.6399
DualDic		22.9927	0.7072	27.8272	0.7672	31.8272	0.8826	33.2722	0.9026	27.8822	0.7027	26.6826	0.6452
RFISR		20.8272	0.6962	26.8385	0.755	30.8826	0.8926	32.8826	0.8826	26.1941	0.7315	25.8822	0.6572
HTSR		21.8826	0.7166	25.8822	0.7672	29.8826	0.9073	31.8822	0.8926	25.8822	0.7288	26.8822	0.6562
ALS		20.8827	0.7072	26.8827	0.7562	30.8826	0.9026	32.8827	0.9027	24.9922	0.7499	25.9927	0.6452
SRCNN		22.5682	0.7224	28.2189	0.7225	32.1012	0.9122	32.8083	0.9249	26.5449	0.755	26.1595	0.6518
WMCNN		21.8826	0.6826	27.8826	0.7452	29.2872	0.7825	32.9927	0.9176	26.8827	0.7572	26.8826	0.6462
SRGAN		22.9927	0.7293	26.8822	0.7672	30.8822	0.8926	33.8827	0.9026	25.8826	0.7672	26.8272	0.6328
FSRCNN		22.6762	0.7326	26.2939	0.7556	31.584	0.9021	33.0618	0.9276	26.4179	0.7629	25.173	0.6286
SRMD		21.9827	0.6976	27.8763	0.7725	31.9822	0.8826	32.8826	0.9306	26.8272	0.7725	25.9027	0.6426
LapSRN		21.2394	0.7332	27.332	0.7522	33.0383	0.9182	33.9636	0.9215	25.6112	0.7768	26.8324	0.6473
WLSCNN		23.3552	0.7314	28.8733	0.7862	32.8862	0.9252	34.9272	0.9266	27.7623	0.7724	27.8826	0.6672
Urban100		BSD100		Medical Dataset		FFHQ		Set14		DIV2K			
Method	Scale	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	×4	16.4919	0.5358	23.6702	0.6027	22.865	0.8064	27.5986	0.8317	22.1459	0.668	20.731	0.5182
GESR		19.9911	0.5725	24.9272	0.6982	26.9827	0.8627	30.8827	0.8562	24.8826	0.6965	23.8161	0.5526
MMSR		18.8816	0.5826	24.8827	0.7082	27.8287	0.8462	31.8822	0.8466	23.8826	0.7055	25.8122	0.5626
CCR		21.9917	0.5622	24.2622	0.6826	28.9827	0.8624	30.2882	0.8262	24.8827	0.7244	24.8161	0.5725
DualDic		20.8816	0.5826	25.9262	0.7082	27.8827	0.8562	29.8927	0.8526	23.8822	0.7155	25.7722	0.5825
RFISR		19.9917	0.6072	24.9729	0.7152	26.8826	0.8672	30.8262	0.8627	24.8826	0.7244	24.7562	0.5626
HTSR		20.9817	0.6342	25.9927	0.6826	25.9827	0.8562	31.8824	0.8627	23.9826	0.7121	25.7992	0.5725
ALS		20.8827	0.6244	25.8826	0.7072	27.8826	0.8624	32.9927	0.8725	24.8827	0.7093	25.7672	0.5825
SRCNN		21.347	0.6282	25.1317	0.706	28.0519	0.8635	32.8828	0.8868	24.7634	0.7221	25.2526	0.5873
WMCNN		21.4408	0.6365	26.175	0.711	29.1733	0.8751	32.9741	0.8783	24.7326	0.7268	25.0781	0.5821
SRGAN		20.8822	0.6142	25.0928	0.6982	28.8826	0.8626	31.8826	0.8627	24.8826	0.7155	25.7735	0.5627
FSRCNN		21.4235	0.6339	24.1583	0.7079	29.2195	0.8734	32.0953	0.8891	25.1425	0.724	26.0499	0.5889
SRMD		21.9927	0.6026	25.9822	0.6826	28.2727	0.8725	33.9927	0.8725	24.7722	0.7241	24.7722	0.5725
LapSRN		22.626	0.5917	26.3292	0.7364	27.0973	0.8627	32.0845	0.9035	25.566	0.7095	25.9599	0.5993
WLSCNN		22.1951	0.6442	26.9158	0.7104	29.3119	0.881	32.8938	0.8996	25.1621	0.736	26.2938	0.5841

The best results are indicated by red color while the second-best results are indicated by blue color

higher than other methods. Although the deeper CNN [40,114,165,214] can bring large receptive fields, it will enhance the computation that makes the training difficult and reduce performance. In addition, the feature details obtained by each channel plays an important role in the information recovery during the SR process. Sometimes GAN driven and flow-based methods [4,16,174,181] suffers from over smoothing, mode collapse and large model footprint issues, respectively. The rapid utilization of deep learnings shows its flexibility and versatility. The improved accuracy rates associated with the deep learning algorithms demonstrates the importance of this technology, clearly emphasizing the tendency for research. Encouraged by the latest success of deep learning and the transformer-based algorithms in terms of numerous computer vision tasks, researchers are working on novel image synthesis algorithms and designing of robust, data driven solution for image super-resolution [184,185,214,215].

#### 4.4. Super-resolution applications

**Satellite Imaging.** Higher resolution images are often desired in the domain of satellite imaging. In order to execute this task SR plays a prominent role. Satellite image processing domain include image rectification, restorations, enhancement and also information extraction. Super resolution techniques are often required in all of these fields. The number of pixels in a super-resolved image increase which improves the visual perception of the digital image. Furthermore, it may aid in the removal of distortions and the enhancement of geographical information. SR can also be combined in additional classification of areas or geographical locations during processing. It could also incorporate learning-based strategies that are effective in land map constructions [216]. Genitha et al. [217] proposed utilizing Hopfield neural networks (HNN) to map satellite images with super-resolution. They have used neural-based approaches for the classification of spatial fields. This is useful for super resolution mapping task. It provides an appropriate and reliable way for identifying target fields using remote sensing imagery. It



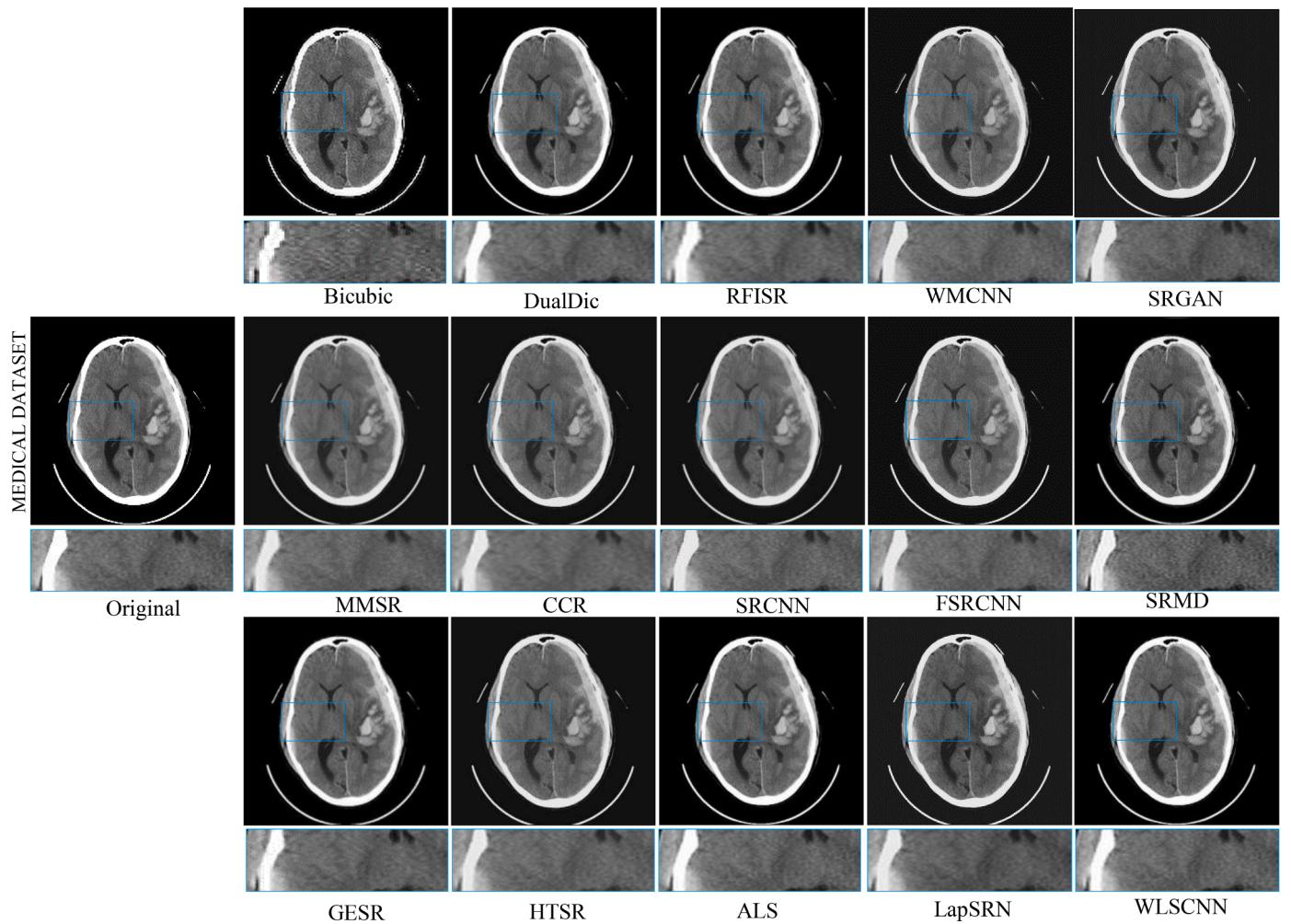
**Fig. 5.** Qualitative comparison of different existing algorithms on Urban100 dataset.

is one of the uses of SR in mapping with satellite imagery. The multi-angle remote sensing imaging [218] is one of the other SR applications. The information in the multi angle direction of the regions may be identical but it is not exactly same. However, with the use of SR, it is possible to extract additional information from images. The adaptive based approaches to merge multi angle information have been proposed by Zhang et al. [218]. To lower the limitations of the different resolutions, an adaptive weighted super-resolution reconstruction technique has been employed. Another aspect of a satellite imaging that may be improved using SR is the classification of different types of regions. Spectrum information provided by sensors can be improved by super resolution mapping. Tolpakin et al. [219] have done research in this field and proposes a contextual classification strategy to improve super-resolution mapping. Then applied it in identification and extraction of tree crown objects [219].

**Biometric information identification.** Super resolution is also crucial in biometric recognition such as to enhance the resolution of faces [220–222], fingerprints [223] and iris images [224]. The resolution of the biometric images is crucial in recognition and detection system. A popular way to dealing with LR observations is to generate high quality images from multiple LR images. Example based single frame super resolution with an external dataset is an effective approach for enhancing resolution based on redundancy and similarity in the structural features of biometric images [90]. Examples of biometric image reconstruction are shown in [222,223,225]. The details of shapes and structural textures are visibly increased using SR while global

structure is effectively preserved by potentially improving recognition ability in relevant applications.

**Medical Imaging.** Super resolution plays an important role in the domain of medical industry. In the literature there is some works that concentrates for increasing the quality of medical images [226]. High contrast and good enhancement images are required in CT scans, MRIs, and other medical imaging modalities which can only be achieved with SR technologies. The majority of images in the medical field have low resolutions, geometric deformations, and low contrast, e.g., X-rays have low contrast, ultrasound images are noisy and so on. Furthermore, if extra time is allotted for imaging due to patient movement blurring may occur. As a result, super-resolution of images may be used to address these concerns [227]. Todays, many medical imaging applications are now faster and more accurate with the use of SR. Detecting the disease at an early stage is always desirable. However, imaging of the matter at that early time is often result with poor contrast. With the development of innovative imaging methods, clinicians and medical researchers can now have higher resolution and early visualization of diagnosis. Super resolution does not necessitate a lot of hardware changes and it is usually done using software [11,228,229]. In the medical field, super resolution is used in functional magnetic resonance imaging (fMRI) [230], positron emission tomography (PET) imaging system [231,232], X-ray digital mammography and Optical Coherence Tomography (OCT). Since medical imaging is so vital for diagnosis therefore having a greater resolution helps corrective treatment significantly. Furthermore, a better resolution may significantly improve the results of automatic



**Fig. 6.** Qualitative comparison on medical test datasets using different algorithms.

detection and image segmentation [11]. Medical imaging algorithms for diagnosis are usually based on basic SR algorithms. The results of the approaches described in the literature show that SR techniques have the potential to be used in practical medical applications [11].

**Earth observation remote sensing.** The first super resolution concept [233] was inspired by the need to improve the resolution of Landsat remote sensing images. The concept of using super resolution algorithms to remote sensing imaging has been around for decades. Although data to meet the demands for super resolution is difficult to come by however there has been few successful real-world examples [218,234–238]. The super resolution of two 5-m images generated by shifting the double CCD array by half a sapling interval which is most effective case [234,239] allows the resolution of the panchromatic image acquired through SPOT-5 to reach 2.5 m. Shen *et al.* [235] also developed a MAP technique and tested it with remote sensing images from the moderate resolution imaging spectroradiometer (MODIS). Further, satellites such as Landsat, CBERS, and WorldView-2 can obtain multispectral or Multiview images for the same area thus produce an opportunity for super resolution [218,238]. An example is given in [218] that includes 5 angular an image from WorldView-2 satellites for super resolution. Sub-pixel mapping [240–242] is a technique for obtaining a fine resolution maps of class labels using super resolution for spectral unmixing of fraction images. The example-based methods have also been used to remotely sensed image SR [243,244] by researchers. Skybox Imaging recently announced plans to launch a group of 24 small satellites that will use SR techniques to produce real time videos with a sub-meter resolution [245,246]. SkySat-1 and SkySat-2 has been

lunched and currently operational. The ground-based distance (GSD) of the final image can be reduced to 4/5 of the original data by incorporating roughly 20 frames [245]. This is a huge opportunity to incorporate super resolution techniques in our everyday lives.

**Microscopy image processing.** In the processing of microscopic images, super resolution also plays a crucial role. According to the literature, there has been a lot of progress in this field recently. SR is highly beneficial for visualizing biological structures like as cells and tissues. One of the most important fields in microscopic imaging is super-resolution fluorescence microscopy. Super-resolved fluorescence microscopy was also awarded a Nobel Prize in Chemistry in 2014. Fluorescence microscopes were once considered one of the most important tools for examining pathways, biological molecules, living cells, tissues, and even full subjects [247]. When compared to electron microscopy, it is more useful. Other techniques such as magnetic resonance imaging (MRI) or optical coherence tomography (OCT) can resolutions in the tens of centimeters and micrometers. However, by using super resolution, the fluorescence range can be extended much farther. Techniques based on super resolution may be able to deliver information up to nanometer scale. It contains techniques such as switchable fluorophores [248] and advanced localization algorithms. Multiple images taken in switching mode can be combined to get better resolution. Furthermore, [248] has explored the area-based difficulties. SR has been shown to be quite useful in the processing of microscopic images.

**Surveillance.** Digital video recorder (DVR) tools are ubiquitous nowadays and they serve an important role in applications such as traffic surveillance and security monitoring. Though equipping large-scale HR

devices is currently unfeasible. Thus, image SR technique should be explored in this field as well. Two examples of super resolution such as a Walk Sequence [249] and a UAV surveillance sequence [250]. However, these techniques have improved over the time. The practical use of video super resolution remains a difficulty. First, outdoor video systems are susceptible to the effects of weather conditions. Further, video data typically contains a large amount of information along with complicated motion. Although certain algorithms can handle motion outliers, but their computational efficiency restricts their use. An attempt on compressed video super-resolution has been made in [251,252].

**Multimedia industry and video enhancement.** Multimedia-based applications are becoming increasingly popular in today's world. In the multimedia industry, super resolution is also used [213]. Movies, animations, and visual effects all require HD data in today's world. As a result, SR has been proven to be a useful technique in video enhancements. Numerous methods in multimedia-based applications use the SR method for images and videos enhancement. SR-based techniques were also used to improve the quality of cell phone-based applications such as images and movies.

**Astrological Studies.** Super resolution is also involved as the important method in the realm of astrological studies. High-resolution astronomical images are always preferred for better computation. To achieve a better view, many blurred and noisy images might be combined. The SR was employed by [253] to increase the quality of astrological images. Many closely clustered stars and distant objects can be visualized more clearly. Many unidentified objects may also be visualized more clearly in this area many times.

**Other Applications.** Besides the previous applications, SR can be used in the area such as object detection, the automotive sector, real-time processing, scanning, surveillance, military, and forensics. In the field of surveillance, [254] has been proposed a conjugate gradient (CG) optimization method. Super resolution is utilized in military surveillance in a similar way [254]. SR is finding recent applications in the automobile industry. It is used as a supporting approach in auto classification and robotics. SR-based approaches are also used in forensic applications. So overall SR research-based work is active in a variety of fields.

## 5. Conclusion and future directions

Super-resolution is a challenging research problem along with significant real-time applications. Exceptional success of recent super resolution techniques has garnered in rapid growth for improving the quality and visual perception of the images in the field of computer visual and image processing. Beside impressive advancement in network architectures and learning mechanism, a diverse set of algorithms has been proposed recently. This survey offers a comprehensive discussion on traditional, deep learning and recent advances on transformer-based SR algorithms. This paper addresses some important parameters such as benchmark datasets, evaluation matrices, loss functions and applications. We had conducted an experiment to some of the benchmarks and recent SR techniques to compare their performance both qualitatively and quantitatively. The top performing algorithms possess greater computational complexity and have deep networks than their counterparts. We notice that the SR performance has been improved along with corresponding increases in network complexity in recent years. Despite considerable success there are still various unsolved problems. Existing algorithms still suffers from constraints that prohibits their implementation in key real-world scenarios such as insufficient metrics, higher complexity of the algorithms, incompetence to cope with real life degradations etc. We hope that this survey not only gives a better understanding of image SR for researchers but also facilitates for future research activities and application developments in this domain. We discuss some of the potential research problems and its future directions below:

**In real work scenarios.** In the real-world scenarios i.e., unknown degradation or missing paired low-to-high resolution images, images

super-resolution is greatly constrained. Bicubic interpolation is commonly used in recent SR works to generate LR images. In comparison to those produced synthetically by bicubic interpolation, real-world images tend to suffer from degradation such as blurring, additive noise and artifacts. As a result, artificial datasets generated manually often badly in real world scenes. A recent attempt to address this problem through learning a GAN to design the real-world degradations [255]. Other current initiative aims to enhance the features by retaining the original spatial resolution and interchanging multiscale details for real-image SR [256] along with feature processing way. A comprehensive challenge for real image SR was lately organized in NTIRE to encourage advancement on this important research problems [257, 258].

**Evaluation metrics.** Nowadays the PSNR and SSIM are two most often utilized SR evaluation metrics. However, PSNR tends to produce excessive smoothness with results varying dramatically between nearly undistinguishable images. Similarly, SSIM [25] generally evaluates brightness, contrast a structure but it is still unable to reliably quantify perceptual quality [23,28]. Further, MOS is the closest to the human visual perception, but it requires lot of efforts and also non-reproducible. Despite the fact that researchers have presented several metrics as presented in Fig. 1 but there is presently no unified and accepted SR proper quality evaluation metrics. As a result, more appropriate metrics for assessing reconstruction quality are urgently required. More recent ones are designed for human perception such as LPIPS [37] and PieAPP [259]. However, each of these metrics has individual cases of failure. Thus, there is no universal perceptual metrics which operates perfectly in all situations and evaluates image performance perfectly. To design a novel evaluation metric is therefore an open challenge of research.

**Loss Functions.** Existing loss functions can be considered as imposing limitations among LR/HR/SR images and guide optimization depending on whether or not these limitations are met. These loss functions are often weighted together in reality and ideal loss function for super resolution is yet unknown. Exploring the potential relationships between these images and look for more appropriate loss functions is perhaps one of the most promising directions.

**Normalization.** Although batch-normalization (BN) is popularly utilized in vision tasks which considerably accelerates training and enhance performance however it is proven to be inferior for super-resolution [70,207,208]. Thus, novel efficient SR normalizing strategies must be investigated.

**High super-resolution rates.** In general, existing algorithms do not discuss the high SR that can be beneficial in instances e.g., super solving faces in the crowd scene. Extremely limited works aim super resolution rates above  $8 \times$  such as  $16 \times$  and  $32 \times$  in [67]. It is difficult to maintain precise local information in the images in such extreme upscaling conditions. In addition, an obvious question is how to retain high perceptual performance in SR images.

**Incorporation of priors.** Recent deep neural networks for super-resolution are data driven algorithms which are learned in an end-to-end manner. Although these methods usually demonstrate superior performance, however when a specific category of degradation arises in which the significant quantities of training data does not exist, then it shows sub-optimal performance, such as in case of medical imaging. In such instances, if the sensors, image objects/ scene and acquisition environments information is established, then helpful priors can be configured for acquiring images of higher resolutions. Both deep network [260] along with sparse coding [261] related priors have been proposed for best SR by current works concentrating on this direction.

**Network improvements.** A well-designed network not only defines a hypothesis space with high performance upper bound, but it also assists in the efficient learning of representations with less spatial and computational redundancy. A large receptive field delivers more contextual details and helps to develop more realistic performance by combining local and global information. Therefore, by combining local and global information to provide contextual information at different

scales for image SR appears to be more promising. Also combining low- and high-level information, network layers prefer to extract low level features such as colors and edges while deep layers learn high level representations such as object identities. Therefore, combining low level features with high level semantics can be quite beneficial for HR reconstruction.

**Context specific attention.** People prefers to care about different aspects of images in different contexts. For example, individuals may be more concerned with local colors and textures in the grass region whereas in the animal body region, people may be more concerned with species and related heir information. Thus, to incorporates an attention mechanism to enhance the concentration on key features helps to produce a realistic information.

**Need for unified solutions.** In real life situations, two or more degradations sometimes occur concurrently. In such instances, how to mutually retrieve images with HR, low artefacts and improved information is an important consideration. In general, existing models built for super resolution are limited to only single case and fails from further degradation. In addition, particular model problems vary in their architectures, training information and loss functions. Designing combined models that work better for many low-level vision operations at the same time is a huge challenge.

## Funding

No funding was received to assist with the preparation of this manuscript.

## CRediT authorship contribution statement

**Dawa Chyophel Lepcha:** Data curation, Writing – original draft, Formal analysis. **Bhawna Goyal:** Methodology, Supervision, Writing – original draft, Formal analysis, Validation. **Ayush Dogra:** Software, Investigation, Writing – review & editing, Resources. **Vishal Goyal:** Investigation, Project administration.

## Declaration of Competing Interest

The authors declare that they have no conflict of interest.

## Data Availability

No data was used for the research described in the article.

## Acknowledgments

We are thankful to the reviewers and editors for their valuable suggestions. We appreciate the insightful comments which helped us in improving the manuscript significantly.

## References

- [1] S.M.A. Bashir, Y. Wang, M. Khan, Y. Niu, A comprehensive review of deep learning-based single image super-resolution, PeerJ Comput. Sci. 7 (2021) 1–56, <https://doi.org/10.7717/PEERJ-CS.621/SUPP-3>. Jul.
- [2] R. Tan, Y. Yuan, R. Huang, J. Luo, Video super-resolution with spatial-temporal transformer encoder, in: 2022 IEEE International Conference on Multimedia and Expo (ICME), 2022, pp. 1–6, <https://doi.org/10.1109/ICME52920.2022.9859774>. Jul.
- [3] H. Li, P. Zhang, Spatio-temporal fusion network for video super-resolution, in: Proceedings of the International Joint Conference on Neural Networks 2021-July, 2021, <https://doi.org/10.1109/IJCNN52387.2021.9533610>. Jul.
- [4] O. Thawakar, P.W. Patil, A. Dudhane, S. Murala, U. Kulkarni, Image and video super resolution using recurrent generative adversarial network, in: 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2019, 2019, <https://doi.org/10.1109/AVSS.2019.8909900>. Sep.
- [5] E. Smith, S. Fujimoto, D. Meger, Multi-view silhouette and depth decomposition for high resolution 3D object representation, Adv. Neural Inf. Process Syst. 31 (2018). Accessed: Sep. 06, 2022. [Online]. Available, <https://github.com/EdwardSmith1884/Multi-View-Silhouette-and-Depth-Decomposition-for-High>.
- [6] B. Li, X. Li, Y. Lu, S. Liu, R. Feng, Z. Chen, HST: Hierarchical Swin Transformer for Compressed Image Super-resolution, 2022, <https://doi.org/10.48550/arxiv.2208.09885>. Aug.
- [7] H. Li, et al., SRDiff: Single image super-resolution with diffusion probabilistic models, Neurocomputing 479 (2022) 47–59, <https://doi.org/10.1016/J.NEUROCOM.2022.01.029>. Mar.
- [8] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based convolutional networks for accurate object detection and segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) (2016) 142–158, <https://doi.org/10.1109/TPAMI.2015.2473784>. Jan.
- [9] Y. Bai, Y. Zhang, M. Ding, B. Ghanem, SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network, 2018, pp. 206–221.
- [10] S.P. Mudunuri, S. Biswas, Low resolution face recognition across variations in pose and illumination, IEEE Trans. Pattern Anal. Mach. Intell. 38 (5) (2016) 1034–1040, <https://doi.org/10.1109/TPAMI.2015.2469282>. May.
- [11] H. Greenspan, Super-resolution in medical imaging, Comput. J. 52 (1) (2009) 43–63, <https://doi.org/10.1093/COMJNL/BX075>. Jan.
- [12] “Remote Sensing and Image Interpretation - Thomas Lillesand, Ralph W. Kiefer, Jonathan Chipman.”
- [13] A.P. Lobanov, Resolution limits in astronomical images (2005), <https://doi.org/10.48550/arxiv.astro-ph/0503225>. Mar.
- [14] A. Swaminathan, M. Wu, K.J.R. Liu, Digital image forensics via intrinsic fingerprints, IEEE Trans. Inform. Forens. Security 3 (1) (2008) 101–117, <https://doi.org/10.1109/TIFS.2007.916010>. Mar.
- [15] F. Lin, J.D. Rojas, P.A. Dayton, Super resolution contrast ultrasound imaging: analysis of imaging resolution and application to imaging tumor angiogenesis, IEEE Int. Ultrason. Sympos. IUS 2016-November (2016), <https://doi.org/10.1109/ULTSYM.2016.7728735>. Nov.
- [16] D. Mahapatra, B. Bozorgtabar, R. Garnavi, Image super-resolution using progressive generative adversarial networks for medical image analysis, Comput. Medical Imag. Graphics 71 (2019) 30–39, <https://doi.org/10.1016/J.COMPMEDIMAG.2018.10.005>. Jan.
- [17] D.J. Brenner, E.J. Hall, Computed tomography — an increasing source of radiation exposure, The New England Journal of Medicine 357 (22) (Nov. 2007) 2277–2284, <https://doi.org/10.1056/NEJMra072149>.
- [18] T.M. Lehmann, C. Gönner, K. Spitzer, Survey: Interpolation methods in medical image processing, IEEE Trans. Med. Imaging 18 (11) (1999) 1049–1075, <https://doi.org/10.1109/42.816070>.
- [19] N.A. Dodgson, Quadratic interpolation for image resampling, IEEE Trans. Image Processing 6 (9) (1997) 1322–1326, <https://doi.org/10.1109/83.623195>.
- [20] R.A. Willoughby, Solutions of Ill-posed problems, in: (A. N. Tikhonov (Ed.), “Arsenin”), Society for Industrial and Applied Mathematics Review or SIAM Review 21 (2) (Feb. 2012) 266–267, <https://doi.org/10.1137/1021044>.
- [21] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1–4) (1992) 259–268, [https://doi.org/10.1016/0167-2789\(92\)90242-F](https://doi.org/10.1016/0167-2789(92)90242-F). Nov.
- [22] X. Lu, Y. Yuan, P. Yan, Image super-resolution via double sparsity regularized manifold learning, IEEE Trans. Circuits Syst. Video Technol. 23 (12) (2013) 2022–2033, <https://doi.org/10.1109/TCSVT.2013.2244798>. Dec.
- [23] C. Ledig, et al., Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, 2017, pp. 4681–4690.
- [24] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Tran. Image Processing 13 (4) (2004) 600–612, <https://doi.org/10.1109/TIP.2003.819861>. Apr.
- [25] Y. Wang, J. Yang, C. Xiao, W. An, Fast convergence strategy for multi-image superresolution via adaptive line search, IEEE Access 6 (2018) 9129–9139, <https://doi.org/10.1109/ACCESS.2018.2799161>. Jan.
- [26] H.R. Sheikh, A.C. Bovik, G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Trans. Image Process. 14 (12) (2005) 2117–2128, <https://doi.org/10.1109/TIP.2005.859389>. Dec.
- [27] “Elements of Information Theory - T. M. Cover.”
- [28] M.S.M. Sajjadi, B. Scholkopf, M. Hirsch, EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis, 2017, pp. 4491–4500.
- [29] X. Wang, K. Yu, C. Dong, C.C. Loy, Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform, 2018, pp. 606–615.
- [30] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep Networks for Image Super-Resolution With Sparse Prior, 2015, pp. 370–378.
- [31] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, M.-H. Yang, Learning to Super-Resolve Blurry Face and Text Images, 2017, pp. 251–260. Accessed: Sep. 06, 2022. [Online]. Available, [https://sites.google.com/view/xiangyuxu/deblusr\\_iccv17](https://sites.google.com/view/xiangyuxu/deblusr_iccv17).
- [32] R. Dahl, M. Norouzi, J. Shlens, G. Brain, Pixel Recursive Super Resolution, 2017, pp. 5439–5448.
- [33] W.S. Lai, J. bin Huang, N. Ahuja, M.H. Yang, Fast and accurate image super-resolution with deep laplacian pyramid networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (11) (2019) 2599–2613, <https://doi.org/10.1109/TPAMI.2018.2865304>. Nov.
- [34] C. Ma, C.Y. Yang, X. Yang, M.H. Yang, Learning a no-reference quality metric for single-image super-resolution, Comput. Vision Image Understand. 158 (2017) 1–16, <https://doi.org/10.1016/J.CVIU.2016.12.009>. May.
- [35] H. Talebi, P. Milanfar, NIMA: neural image assessment, IEEE Trans. Image Processing 27 (8) (2018) 3998–4011, <https://doi.org/10.1109/TIP.2018.2831899>. Aug.

- [36] J. Kim, S. Lee, Deep Learning of Human Visual Sensitivity in Image Quality Assessment Framework, 2017, pp. 1676–1684.
- [37] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, 2018, pp. 586–595. Accessed: Sep. 07, 2022. [Online]. Available, <https://www.github.com/richhang/PerceptualSimilarity>.
- [38] D. Dai, Y. Wang, Y. Chen, L. van Gool, Is image super-resolution helpful for other vision tasks?, in: 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, 2016, <https://doi.org/10.1109/WACV.2016.7477613>. May.
- [39] M. Haris, G. Shakhnarovich, N. Ukita, Task-driven super resolution: object detection in low-resolution images, Commun. Comput. Inform. Sci. 1516 CCIS (2021) 387–395, [https://doi.org/10.1007/978-3-030-92307-5\\_45/COVER](https://doi.org/10.1007/978-3-030-92307-5_45/COVER).
- [40] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image Super-Resolution Using Very Deep Residual Channel Attention Networks, 2018, pp. 286–301.
- [41] C. Fookes, F. Lin, V. Chandran, S. Sridharan, Evaluation of image resolution and super-resolution on face recognition performance, J. Vis. Commun. Image Represent 23 (1) (2012) 75–93, <https://doi.org/10.1016/J.JVCIR.2011.06.004>. Jan.
- [42] K. Zhang, et al., Super-Identity Convolutional Neural Network for Face Hallucination, 2018, pp. 183–198.
- [43] A. Bulat, G. Tzimiropoulos, Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses With GANs, 2018, pp. 109–117.
- [44] Y. Chen, Y. Tai, X. Liu, C. Shen, J. Yang, FSRNet: End-to-End Learning Face Super-Resolution With Facial Priors, 2018, pp. 2492–2501. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/tysiwo/FSRNet>.
- [45] V.A. Kelkar, X. Zhang, J. Granstedt, H. Li, M.A. Anastasio, Task-based evaluation of deep image super-resolution in medical imaging, 11599, 2021, pp. 207–213, <https://doi.org/10.1111/12.2582011>. Feb.
- [46] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, Conf. Record Asilomar Conf. Signals Syst. Comput. 2 (2003) 1398–1402, <https://doi.org/10.1109/ACSSC.2003.1292216>.
- [47] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: A feature similarity index for image quality assessment, IEEE Trans. Image Processing 20 (8) (2011) 2378–2386, <https://doi.org/10.1109/TIP.2011.2109730>. Aug.
- [48] A. Mittal, R. Soundararajan, A.C. Bovik, Making a ‘completely blind’ image quality analyzer, IEEE Signal Process Lett. 20 (3) (2013) 209–212, <https://doi.org/10.1109/LSP.2012.2227726>.
- [49] Y. Blau, T. Michaeli, The Perception-Distortion Tradeoff, 2018, pp. 6228–6237.
- [50] M. Bevilacqua, A. Roumy, C. Guilleminot, M.L.A. Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: BMVC 2012 - Electronic Proceedings of the British Machine Vision Conference, 2012, p. 2012, <https://doi.org/10.5244/C.26.135>.
- [51] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6920 LNCS (2012) 711–730, [https://doi.org/10.1007/978-3-642-27413-8\\_47/COVER](https://doi.org/10.1007/978-3-642-27413-8_47/COVER).
- [52] S. Anwar, S. Khan, N. Barnes, A Deep Journey into Super-resolution, ACM Comput. Surveys (CSUR) 53 (3) (May 2020), <https://doi.org/10.1145/3390462>.
- [53] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the IEEE International Conference on Computer Vision 2, 2001, pp. 416–423, <https://doi.org/10.1109/ICCV.2001.937655>.
- [54] J.-B. Huang, A. Singh, N. Ahuja, Single Image Super-Resolution From Transformed Self-Exemplars, 2015, pp. 5197–5206.
- [55] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, L. Zelnik-Manor, The 2018 PIRM Challenge on Perceptual Image Super-Resolution, 2018, 0–0.
- [56] A. Fujimoto, T. Ogawa, K. Yamamoto, Y. Matsui, T. Yamasaki, K. Aizawa, Manga109 dataset and creation of metadata, in: ACM International Conference Proceeding Series, 2016, <https://doi.org/10.1145/3011549.3011551>. Dec.
- [57] Free Medical Imaging Software : I Do Imaging, 2022, <https://idoomaging.com/home>. accessed Sep. 07..
- [58] DIV2K Dataset | Kaggle, 2022. <https://www.kaggle.com/datasets/joe1995/div2k-dataset>. accessed Sep. 07..
- [59] Flickr-Faces-HQ Dataset (FFHQ) | Kaggle, 2022. <https://www.kaggle.com/datasets/arnaud58/flickrfaceshq-dataset-fbhq>. accessed Sep. 07..
- [60] “celeba-hq Kaggle.” <https://www.kaggle.com/datasets/lamsimon/celebahq>.
- [61] S. Bell, P. Upchurch, N. Snavely, K. Bala, Material Recognition in the Wild With the Materials in Context Database, 2015, pp. 3479–3487.
- [62] W. Shi, et al., Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network, 2016, pp. 1874–1883.
- [63] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, ImageNet: A large-scale hierarchical image database, 2010, pp. 248–255, <https://doi.org/10.1109/CVPR.2009.5206848>. Mar.
- [64] T.Y. Lin, et al., Microsoft COCO: Common objects in context, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8693 LNCS (5) (2014) 740–755, [https://doi.org/10.1007/978-3-319-10602-1\\_48/COVER](https://doi.org/10.1007/978-3-319-10602-1_48/COVER). PART.
- [65] M. Everingham, S.M.A. Eslami, L. van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The Pascal visual object classes challenge: a retrospective, Int. J. Comput. Vision 111 (1) (2014) 98–136, <https://doi.org/10.1007/S11263-014-0733-5>, 2014 111:1Jun.
- [66] J. Kim, J.K. Lee, K.M. Lee, Accurate Image Super-Resolution Using Very Deep Convolutional Networks, 2016, pp. 1646–1654.
- [67] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution, 2017, pp. 624–632. Accessed: Sep. 07, 2022. [Online]. Available, <http://vlab1.ucmerced.edu/wlai24/LapSRN>.
- [68] S. Ramchandra Kinge, S. Rani, R. Raju Taksande, S. Kumar Kinge, B. Sheela Rani, M.S. Sutaone, A Survey of Segmentation of Colored Images based on Impressive Algorithms, 2017, <https://doi.org/10.15680/IJRSET.2017.0609200>.
- [69] Y. Tai, J. Yang, X. Liu, Image Super-Resolution via Deep Recursive Residual Network, 2017, pp. 3147–3155.
- [70] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced Deep Residual Networks for Single Image Super-Resolution, 2017, pp. 136–144.
- [71] M. Haris, G. Shakhnarovich, N. Ukita, Deep Back-Projection Networks for Super-Resolution, 2018, pp. 1664–1673.
- [72] R.Timofte Eirikur, et al., NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results, 2017, pp. 114–125. Accessed: Sep. 07, 2022. [Online]. Available, <http://www.vision.ee.ethz.ch/npire17/>.
- [73] A. Ignatov, et al., PIRM Challenge on Perceptual Image Enhancement on Smartphones: Report, 2018, 0–0Accessed: Sep. 07, 2022. [Online]. Available, <http://ai-benchmark.org>.
- [74] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Processing 21 (12) (2012) 4695–4708, <https://doi.org/10.1109/TIP.2012.2214050>.
- [75] R.G. Keys, Cubic convolution interpolation for digital image processing, IEEE Trans. Acoust. 29 (6) (1981) 1153–1160, <https://doi.org/10.1109/TASSP.1981.1163711>.
- [76] X. Li, M.T. Orchard, New edge-directed interpolation, IEEE Trans. Image Processing 10 (10) (2001) 1521–1527, <https://doi.org/10.1109/83.951537>. Oct.
- [77] L. Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, IEEE Trans. Image Processing 15 (8) (2006) 2226–2238, <https://doi.org/10.1109/TIP.2006.877407>, Aug.
- [78] Philippe Thévenaz, Thierry Blu, Unser Michael, Image interpolation and resampling, Handbook of medical imaging, processing and analysis 1 1 (2000) 393–420.
- [79] Y. Zhang, Q. Fan, F. Bao, Y. Liu, C. Zhang, Single-image super-resolution based on rational fractal interpolation, IEEE Trans. Image Processing 27 (8) (2018) 3782–3797, <https://doi.org/10.1109/TIP.2018.2826139>. Aug.
- [80] M.M. Khattab, A.M. Zeki, A.A. Alwan, A.S. Badawy, Regularization-based multi-frame super-resolution: a systematic review, J. King Saud Univ. Comput. Inform. Sci. 32 (7) (2020) 755–762, <https://doi.org/10.1016/J.JKSUCL.2018.11.010>. Sep.
- [81] J. Sun, J. Zhu, M.F. Tappen, Context-constrained hallucination for image super-resolution, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 231–238, <https://doi.org/10.1109/CVPR.2010.5540206>.
- [82] H. Xu, G. Zhai, X. Yang, Single image super-resolution with detail enhancement based on local fractal analysis of gradient, IEEE Trans. Circuits Syst. Video Technol. 23 (10) (2013) 1740–1754, <https://doi.org/10.1109/TCSVT.2013.2248305>.
- [83] L. Wang, S. Xiang, G. Meng, H. Wu, C. Pan, Edge-directed single-image super-resolution via adaptive gradient magnitude self-interpolation, IEEE Trans. Circuits Syst. Video Technol. 23 (8) (2013) 1289–1299, <https://doi.org/10.1109/TCSVT.2013.2240915>.
- [84] K. Zhang, X. Gao, D. Tao, X. Li, Single image super-resolution with non-local means and steering kernel regression, IEEE Trans. Image Processing 21 (11) (2012) 4544–4556, <https://doi.org/10.1109/TIP.2012.2208977>.
- [85] A. Chakrabarti, A.N. Rajagopalan, R. Chellappa, Super-resolution of face images using kernel PCA-based prior, IEEE Trans. Multimedia 9 (4) (2007) 888–892, <https://doi.org/10.1109/TMM.2007.893346>. Jun.
- [86] D.C. Lepcha, B. Goyal, Medical-Modality Super-resolution for increased visualisation of Intracranial tissue details and structural details, in: 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICrito 2021, 2021, <https://doi.org/10.1109/ICRITO51393.2021.9596440>.
- [87] H. Chang, D.Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1, 2004, <https://doi.org/10.1109/CVPR.2004.1315043>.
- [88] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326, <https://doi.org/10.1126/science.290.5500.2323>, 1979Dec.
- [89] S. Yang, Z. Wang, L. Zhang, M. Wang, Dual-geometric neighbor embedding for image super-resolution with sparse tensor, IEEE Trans. Image Process. 23 (7) (2014) 2793–2803, <https://doi.org/10.1109/TIP.2014.2319742>.
- [90] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873, <https://doi.org/10.1109/TIP.2010.2050625>. Nov.
- [91] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution as sparse representation of raw image patches, in: 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008, <https://doi.org/10.1109/CVPR.2008.4587647>.
- [92] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666, <https://doi.org/10.1109/TIT.2007.909108>.
- [93] J. Jiang, R. Hu, Z. Wang, Z. Han, S. Dong, Manifold regularized sparse support regression for single image super-resolution, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2013, pp. 1429–1433, <https://doi.org/10.1109/ICASSP.2013.6637887>. Oct.

- [94] J. Jiang, R. Hu, Z. Han, K. Huang, T. Lu, Efficient single image super-resolution via graph embedding, in: Proc (IEEE Int Conf Multimed Expo), 2012, pp. 610–615, <https://doi.org/10.1109/ICME.2012.102>.
- [95] D.C. Lepcha, B. Goyal, A. Dogra, S.H. Wang, An efficient medical image super-resolution based on piecewise linear regression strategy using domain transform filtering, *Concurr. Comput.* 34 (20) (2022) e6644, <https://doi.org/10.1002/CPE.6644>. Sep.
- [96] M. Doshi, P. Gajjar, A. Kothari, Zoom based image super-resolution using DCT with LBP as characteristic model, *J. King Saud Univ. Comput. Inform. Sci.* 34 (2) (2022) 72–85, <https://doi.org/10.1016/J.JKSUCI.2018.10.005>. Feb.
- [97] M.C. Yang, Y.C.F. Wang, A self-learning approach to single image super-resolution, *IEEE Trans. Multimedia* 15 (3) (2013) 498–508, <https://doi.org/10.1109/TMM.2012.2232646>. Apr.
- [98] Z. Xiong, D. Xu, X. Sun, F. Wu, Example-based super-resolution with soft information and decision, *IEEE Trans. Multimedia* 15 (6) (2013) 1458–1465, <https://doi.org/10.1109/TMM.2013.2264654>.
- [99] L.W. Kang, C.C. Hsu, B. Zhuang, C.W. Lin, C.H. Yeh, Learning-based joint super-resolution and deblocking for a highly compressed image, *IEEE Trans. Multimedia* 17 (7) (2015) 921–934, <https://doi.org/10.1109/TMM.2015.2434216>. Jul.
- [100] L. He, H. Qi, R. Zaretzki, Beta Process Joint Dictionary Learning for Coupled Feature Spaces with Application to Single Image Super-Resolution, 2013, pp. 345–352, <https://doi.org/10.1109/CVPR.2013.51>.
- [101] T. Peleg, M. Elad, A statistical prediction model based on sparse representations for single image super-resolution, *IEEE Trans. Image Process.* 23 (6) (2014) 2569–2582, <https://doi.org/10.1109/TIP.2014.2305844>.
- [102] R. Timofte, V. de Smet, L. van Gool, Anchored Neighborhood Regression for Fast Example-Based Super-Resolution, 2013, pp. 1920–1927, <https://doi.org/10.1109/ICCV.2013.241>.
- [103] R. Timofte, V. de Smet, L. van Gool, A+: Adjusted anchored neighborhood regression for fast super-resolution, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9006 (2015) 111–126, [https://doi.org/10.1007/978-3-319-16817-3\\_8/COVER](https://doi.org/10.1007/978-3-319-16817-3_8/COVER).
- [104] C.-Y. Yang, M.-H. Yang, Fast Direct Super-Resolution by Simple Functions, 2013, pp. 561–568, <https://doi.org/10.1109/ICCV.2013.75>.
- [105] Z. Zhu, F. Guo, H. Yu, C. Chen, Fast single image super-resolution via self-example learning and sparse representation, *IEEE Trans. Multimedia* 16 (8) (2014) 2178–2190, <https://doi.org/10.1109/TMM.2014.2364976>. Dec.
- [106] K. Zhang, D. Tao, X. Gao, X. Li, Z. Xiong, Learning multiple linear mappings for efficient single image super-resolution, in: 24, 2015, pp. 846–861, <https://doi.org/10.1109/TIP.2015.2389629>. Mar.
- [107] J. Zhang, C. Zhao, R. Xiong, S. Ma, D. Zhao, Image super-resolution via dual-dictionary learning and sparse representation, in: ISCAS 2012 - 2012 IEEE International Symposium on Circuits and Systems, 2012, pp. 1688–1691, <https://doi.org/10.1109/ISCAS.2012.6271583>.
- [108] Y. Zhang, Y. Zhang, J. Zhang, Q. Dai, CCR: Clustering and Collaborative Representation for Fast Single Image Super-Resolution, *IEEE Trans. Multimedia* 18 (3) (2016) 405–417, <https://doi.org/10.1109/TMM.2015.2512046>. Mar.
- [109] J. Luo, X. Sun, M. Lung Yiu, L. Jin, X. Peng, Piecewise linear regression-based single image super-resolution via Hadamard transform, *Inf. Sci. (N Y)* 462 (2018) 315–330, <https://doi.org/10.1016/J.INS.2018.06.030>. Sep.
- [110] Z. Wang, J. Chen, S.C.H. Hoi, Deep learning for image super-resolution: a survey, *IEEE Trans. Pattern Anal. Mach. Intell.* 43 (10) (2021) 3365–3387, <https://doi.org/10.1109/TPAMI.2020.2982166>. Oct.
- [111] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising, *IEEE Trans. Image Process.* 26 (7) (2017) 3142–3155, <https://doi.org/10.1109/TIP.2017.2662206>. Jul.
- [112] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep CNN Denoiser prior for image restoration, 2017, pp. 3929–3938.
- [113] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8692 LNCS (4) (2014) 184–199, [https://doi.org/10.1007/978-3-319-10593-2\\_13/COVER](https://doi.org/10.1007/978-3-319-10593-2_13). PART.
- [114] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (2) (2016) 295–307, <https://doi.org/10.1109/TPAMI.2015.2439281>. Feb.
- [115] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2014, <https://doi.org/10.48550/arxiv.1409.1556>. Sep.
- [116] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9906 LNCS (2016) 391–407, [https://doi.org/10.1007/978-3-319-46475-6\\_25/FIGURES/7](https://doi.org/10.1007/978-3-319-46475-6_25/FIGURES/7).
- [117] D. Geman, C. Yang, Nonlinear image recovery with half-quadratic regularization, *IEEE Trans. Image Process.* 4 (7) (1995) 932–946, <https://doi.org/10.1109/83.392335>.
- [118] J. Kim, J.K. Lee, K.M. Lee, Deeply-Recursive Convolutional Network for Image Super-Resolution, 2016, pp. 1637–1645.
- [119] Z. Luo, et al., EBSR: Feature Enhanced Burst Super-Resolution With Deformable Alignment, 2021, pp. 471–478. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/Algolzw/EBSR>.
- [120] K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, 2015, pp. 1026–1034.
- [121] T. Wang, W. Sun, H. Qi, P. Ren, Aerial image super resolution via wavelet multiscale convolutional neural networks, *IEEE Geosci. Remote Sensing Lett.* 15 (5) (2018) 769–773, <https://doi.org/10.1109/LGRS.2018.2810893>. May.
- [122] B. Goyal, A. Dawa, C. Lepcha, A. Dogra, S.-H. Wang, D.C. Lepcha, A weighted least squares optimisation strategy for medical image super resolution via multiscale convolutional neural networks for healthcare applications, *Complex Intell. Syst.* 8 (4) (2021) 3089–3104, <https://doi.org/10.1007/S40747-021-00465-Z>, 2021 8: 4Jul.
- [123] W. Shi, et al., Is the deconvolution layer the same as a convolutional layer? (2016), <https://doi.org/10.48550/arxiv.1609.07009>. Sep.
- [124] S. Liu, et al., EVSRNet: Efficient Video Super-Resolution With Neural Architecture Search, 2021, pp. 2480–2485.
- [125] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2016, pp. 770–778. Accessed: Sep. 07, 2022. [Online]. Available, <http://image-net.org/challenges/LSVRC/2015/>.
- [126] N. Ahn, B. Kang, K.-A. Sohn, Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network, 2018, pp. 252–268.
- [127] J. Jiao, W.-C. Tu, S. He, R.W.H. Lau, FormResNet: Formatted Residual Learning for Image Restoration, 2017, pp. 38–46.
- [128] Y. Fan, et al., Balanced Two-Stage Residual Networks for Image Super-Resolution, 2017, pp. 161–168. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/ychfan/>.
- [129] X.-J. Mao, C. Shen, Y.-B. Yang, Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, *Adv. Neural Inf. Process. Syst.* 29 (2016).
- [130] D.P. Kingma, J.L. Ba, Adam: a method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2014, <https://doi.org/10.48550/arxiv.1412.6980>. Dec.
- [131] J. Liu, J. Tang, G. Wu, Residual feature distillation network for lightweight image super-resolution, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12537 LNCS (2020) 41–55, [https://doi.org/10.1007/978-3-03-67070-2\\_2/COVER](https://doi.org/10.1007/978-3-03-67070-2_2/COVER).
- [132] X. Xiang, Y. Tian, Y. Zhang, Y. Fu, J.P. Allebach, C. Xu, Zooming Slow-Mo: Fast and Accurate One-Stage Space-Time Video Super-Resolution, 2020, pp. 3370–3379.
- [133] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9351 (2015) 234–241, [https://doi.org/10.1007/978-3-319-24574-4\\_28/COVER](https://doi.org/10.1007/978-3-319-24574-4_28/COVER).
- [134] J. Liu, W. Zhang, Y. Tang, J. Tang, G. Wu, Residual Feature Aggregation Network for Image Super-Resolution, 2020, pp. 2359–2368.
- [135] T. Isobe, X. Jia, S. Gu, S. Li, S. Wang, Q. Tian, Video super-resolution with recurrent structure-detail network, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12357 LNCS (2020) 645–660, [https://doi.org/10.1007/978-3-03-58610-2\\_38/COVER](https://doi.org/10.1007/978-3-03-58610-2_38/COVER).
- [136] Chan, Kelvin CK, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. "BasicVSR: The search for essential components in video super-resolution and beyond." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4947-4956. 2021.
- [137] X. Yang, T. Xie, L. Liu, D. Zhou, Image super-resolution reconstruction based on improved Dirac residual network, *Multidimensional Systems and Signal Processing* 32 (4) (2021) 1065–1082, <https://doi.org/10.1007/S11045-021-00773-0>, 2021 32:4Mar.
- [138] H. Zhao, X. Kong, J. He, Y. Qiao, C. Dong, Efficient image super-resolution using pixel attention, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12537 LNCS (2020) 56–72, [https://doi.org/10.1007/978-3-03-67070-2\\_3/COVER](https://doi.org/10.1007/978-3-03-67070-2_3/COVER).
- [139] Y. Tai, J. Yang, X. Liu, C. Xu, MemNet: A Persistent Memory Network for Image Restoration, 2017, pp. 4539–4547.
- [140] K. Jiang, Z. Wang, P. Yi, J. Jiang, Hierarchical dense recursive network for image super-resolution, *Pattern Recognit.* 107 (2020), 107475, <https://doi.org/10.1016/J.PATCOG.2020.107475>. Nov.
- [141] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, W. Wu, Feedback Network for Image Super-Resolution, 2019, pp. 3867–3876. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com>.
- [142] J.-S. Choi, M. Kim, A Deep Convolutional Neural Network With Selection Units for Super-Resolution, 2017, pp. 154–160.
- [143] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual Dense Network for Image Super-Resolution, 2018, pp. 2472–2481.
- [144] X. Liu, S. Chen, L. Song, M. Woźniak, S. Liu, Self-attention negative feedback network for real-time image super-resolution, *J. King Saud Univ. Comput. Inform. Sci.* 34 (8) (2022) 6179–6186, <https://doi.org/10.1016/J.JKSUCI.2021.07.014>. Sep.
- [145] S. Anwar, N. Barnes, Densely residual laplacian super-resolution, *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (3) (2022) 1192–1204, <https://doi.org/10.1109/TPAMI.2020.3021088>. Mar.
- [146] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, X. Luo, MADNet: a fast and lightweight network for single-image super-resolution, *IEEE Trans. Cybern.* 51 (3) (2021) 1443–1453, <https://doi.org/10.1109/TCYB.2020.2970104>. Mar.
- [147] B. Niu, et al., Single image super-resolution via a holistic attention network, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12357 LNCS (2020) 191–207, [https://doi.org/10.1007/978-3-03-58610-2\\_12/COVER](https://doi.org/10.1007/978-3-03-58610-2_12/COVER).

- [148] J.-H. Kim, J.-H. Choi, M. Cheon, J.-S. Lee, RAM: residual attention module for single image super-resolution, arXiv Comput. Sci. (2019). Oct.
- [149] T. Isobe, et al., Video Super-Resolution With Temporal Group Attention, 2020, pp. 8008–8017.
- [150] X. Yang, J. Fan, C. Wu, D. Zhou, T. Li, NasmamSR: a fast image super-resolution network based on neural architecture search and multiple attention mechanism, Multimedia Systems 28 (1) (2021) 321–334, <https://doi.org/10.1007/S00530-021-00841-2>, 202128:1Sep.
- [151] H. Ren, M. El-Khamy, J. Lee, Image Super Resolution Based on Fusing Multiple Convolution Neural Networks, 2017, pp. 54–61.
- [152] A. Kuznetsova, et al., The open images dataset V4, Int. J. Comput. Vision 128 (7) (2020) 1956–1981, <https://doi.org/10.1007/S11263-020-01316-Z>, 2020 128: 7Mar.
- [153] L. Wang, D. Li, Y. Zhu, L. Tian, Y. Shan, Dual Super-Resolution Learning for Semantic Segmentation, 2020, pp. 3774–3783.
- [154] Y. Hu, X. Gao, J. Li, Y. Huang, H. Wang, Single Image Super-Resolution via Cascaded Multi-Scale Cross Network, 2018, <https://doi.org/10.48550/arxiv.1802.08808>, Feb.
- [155] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier Nonlinearities Improve Neural Network Acoustic Models, 2013.
- [156] Z. Hui, X. Wang, X. Gao, Fast and Accurate Single Image Super-Resolution via Information Distillation Network, 2018, pp. 723–731. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/Zheng222/IDN-Caffe>.
- [157] Y. Qiu, R. Wang, D. Tao, J. Cheng, Embedded Block Residual Network: A Recursive Restoration Model for Single-Image Super-Resolution, 2019, pp. 4180–4189.
- [158] W. Li, X. Tao, T. Guo, L. Qi, J. Lu, J. Jia, MuCAN: multi-correspondence aggregation network for video super-resolution, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12355 LNCS (2020) 335–351, [https://doi.org/10.1007/978-3-030-58607-2\\_20/COVER](https://doi.org/10.1007/978-3-030-58607-2_20).
- [159] X. Yang, Y. Guo, Z. Li, D. Zhou, Image super-resolution network based on a multi-branch attention mechanism, Signal Image Video Process. 15 (7) (2021) 1397–1405, <https://doi.org/10.1007/S11760-021-01870-0>, 2021 15:7Feb.
- [160] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, 2017, pp. 4700–4708. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/liuzhuang13/DenseNet>.
- [161] T. Tong, G. Li, X. Liu, Q. Gao, Image Super-Resolution Using Dense Skip Connections, 2017, pp. 4799–4807.
- [162] Y. Guo, et al., Closed-Loop Matters: Dual Regression Networks for Single Image Super-Resolution, 2020, pp. 5407–5416.
- [163] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: Graph. Models Image Process. 53 (3) (1991) 231–239, [https://doi.org/10.1016/1049-9652\(91\)90045-L](https://doi.org/10.1016/1049-9652(91)90045-L). May.
- [164] C. Tian, et al., Lightweight image super-resolution with enhanced CNN, Knowl. Based Syst. 205 (2020), 106235, <https://doi.org/10.1016/J.KNOSYS.2020.106235>. Oct.
- [165] A. Shocher, N. Cohen, M. Irani, ‘Zero-Shot’ Super-Resolution Using Deep Internal Learning, 2018, pp. 3118–3126.
- [166] K. Zhang, W. Zuo, L. Zhang, Learning a Single Convolutional Super-Resolution Network for Multiple Degradations, 2018, pp. 3262–3271.
- [167] K. Ma, et al., Waterloo exploration database: New challenges for image quality assessment models, IEEE Trans. Image Process. 26 (2) (2017) 1004–1016, <https://doi.org/10.1109/TIP.2016.2631888>, Feb.
- [168] K. Zhang, L. van Gool, R. Timofte, Deep Unfolding Network for Image Super-Resolution, 2020, pp. 3217–3226. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/cszn/USRNet>.
- [169] Y.-S. Xu, S.-Y.R. Tseng, Y. Tseng, H.-K. Kuo, Y.-M. Tsai, Unified Dynamic Convolutional Network for Super-Resolution With Variational Degradations, 2020, pp. 12496–12505.
- [170] I. Goodfellow, et al., Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144, <https://doi.org/10.1145/3422622>. Oct.
- [171] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, in: 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, 2015, <https://doi.org/10.48550/arxiv.1511.06434>. Nov.
- [172] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses supplementary, ArXiv 9906 LNCS (2016) 1–5. Accessed: Sep. 07, 2022. [Online]. Available, <http://arxiv.org/abs/1603.08155>.
- [173] J. Long, E. Shelhamer, T. Darrell, Fully Convolutional Networks for Semantic Segmentation, 2015, pp. 3431–3440.
- [174] X. Wang, et al., ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks, 2018, 0–0Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/xintao/ESRGAN>.
- [175] S.-J. Park, H. Son, S. Cho, K.-S. Hong, S. Lee, SRFeat: Single Image Super-Resolution with Feature Discrimination, 2018, pp. 439–455.
- [176] A. Jolicoeur-Martineau, The relativistic discriminator: a key element missing from standard GAN, in: 7th International Conference on Learning Representations, ICLR 2019, 2018, <https://doi.org/10.48550/arxiv.1807.00734>. Jul.
- [177] S. Bell-Klijger, A. Shocher, M. Irani, Blind Super-Resolution Kernel Estimation using an Internal-GAN, Adv. Neural Inf. Process. Syst. 32 (2019).
- [178] M. Fritzsche, S. Gu, R. Timofte, Frequency separation for real-world super-resolution, in: Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019, 2019, pp. 3599–3608, <https://doi.org/10.1109/ICCVW.2019.00445>. Oct.
- [179] T. Shang, Q. Dai, S. Zhu, T. Yang, Y. Guo, Perceptual Extreme Super-Resolution Network With Receptive Field Block, 2020, pp. 440–441.
- [180] W. Zhang, Y. Liu, C. Dong, Y. Qiao, RankSRGAN: Generative Adversarial Networks With Ranker for Image Super-Resolution, 2019, pp. 3096–3105. Accessed: Sep. 07, 2022. [Online]. Available, <https://wenlongzhang0724.github.io/>.
- [181] J. Daihong, Z. Sai, D. Lei, D. Yueming, Multi-scale generative adversarial network for image super-resolution, 26, Soft Computing, 2022, pp. 3631–3641, <https://doi.org/10.1007/S00500-022-06822-5>, 2022 26:8Feb.
- [182] N. Parmar, et al., Image Transformer, PMLR, 2018, pp. 4055–4064. Jul. 03, Accessed: Oct. 02, 2022. [Online]. Available, <https://proceedings.mlr.press/v80/parmar18a.html>.
- [183] F. Yang, H. Yang, J. Fu, H. Lu, B. Guo, Learning Texture Transformer Network for Image Super-Resolution, 2020, pp. 5791–5800.
- [184] W. Zou, T. Ye, W. Zheng, Y. Zhang, L. Chen, Y. Wu, Self-Calibrated Efficient Transformer for Lightweight Super-Resolution, 2022, pp. 930–939. Accessed: Oct. 02, 2022. [Online]. Available, <https://github.com>.
- [185] J.F. Hu, T.Z. Huang, L.J. Deng, H.X. Dou, D. Hong, G. Vivone, Fusformer: a transformer-based fusion network for hyperspectral image super-resolution, in: IEEE Geosci. Remote Sensing Lett., 19, 2022, <https://doi.org/10.1109/LGRS.2022.3194257>.
- [186] M. v. Conde, U.-J. Choi, M. Burchi, R. Timofte, Swin2SR: SwinV2 Transformer for Compressed Image Super-Resolution and Restoration, 2022, <https://doi.org/10.48550/arxiv.2209.11345>. Sep.
- [187] Q. Cai, et al., HIPA: Hierarchical Patch Transformer for Single Image Super Resolution, 2022, <https://doi.org/10.48550/arxiv.2203.10247>. Mar.
- [188] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, T. Zeng, Transformer for Single Image Super-Resolution, 2022, pp. 457–466.
- [189] J. Fang, H. Lin, X. Chen, K. Zeng, A Hybrid Network of CNN and Transformer for Lightweight Image Super-Resolution, 2022, pp. 1103–1112. Accessed: Oct. 02, 2022. [Online]. Available, <https://github.com/ljhthp/HNCT>.
- [190] K. Jin, et al., SwinPASSR: Swin Transformer Based Parallax Attention Network for Stereo Image Super-Resolution, 2022, pp. 920–929.
- [191] L.A. Gatys, A.S. Ecker, M. Bethge, Texture synthesis using convolutional neural networks, Adv. Neural Inf. Process. Syst. 28 (2015). Accessed: Sep. 07, 2022. [Online]. Available, <http://github.com/leongatys/DeepTextures>.
- [192] L.A. Gatys, A.S. Ecker, M. Bethge, Image Style Transfer Using Convolutional Neural Networks, 2016, pp. 2414–2423.
- [193] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein Generative Adversarial Networks, PMLR, 2017, pp. 214–223. Jul. 17, Accessed: Sep. 07, 2022. [Online]. Available, <https://proceedings.mlr.press/v70/arjovsky17a.html>.
- [194] D. Berthelot, T. Schumm, L. Metz Google, BEGAN: boundary equilibrium generative adversarial networks, ArXiv (Mar. 2017) 1–9, <https://doi.org/10.48550/arxiv.1703.10717>.
- [195] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of Wasserstein GANs, Adv. Neural Inf. Process. Syst. 30 (2017). Accessed: Sep. 07, 2022. [Online]. Available, [https://github.com/igul222/improved\\_wgan\\_training](https://github.com/igul222/improved_wgan_training).
- [196] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, J. Zhou, Structure-Preserving Super Resolution With Gradient Guidance, 2020, pp. 7769–7778. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/Maclory/SPSR>.
- [197] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, B.A. Research, Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks, 2017, pp. 2223–2232. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/junyanz/CycleGAN>.
- [198] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, L. Lin, Unsupervised Image Super-Resolution Using Cycle-in-Cycle Generative Adversarial Networks, 2018, pp. 701–710. Accessed: Sep. 07, 2022. [Online]. Available, <https://competitions.codalab.org/competitions/18024>.
- [199] A. Bruhn, J. Weickert, C. Schnörr, Lucas/Kanade Meets Horn/Schunck: combining local and global optic flow methods, Int. J. Comput. Vision 61 (3) (2005) 211–231, <https://doi.org/10.1023/B:VISL.0000045324.43199.43>, 2005 61:3Feb.
- [200] H. Zhao, O. Gallo, I. Froisio, J. Kautz, Loss functions for image restoration with neural networks, IEEE Trans. Comput. Imaging 3 (1) (2016) 47–57, <https://doi.org/10.1109/TCI.2016.2644865>. Dec.
- [201] Z. Zhang, A. Research, Z. Wang, Z. Lin, H. Qi, Image Super-Resolution by Neural Texture Transfer, 2019, pp. 7982–7991. Accessed: Sep. 07, 2022. [Online]. Available: <https://github.com/ZZUTK/SRNTT>.
- [202] Y. Guo, et al., Dual Reconstruction Net for Image Super-Resolution with Gradient Sensitive Loss, 2018, <https://doi.org/10.48550/arxiv.1809.07099>. Sep.
- [203] S. Vasu, N. Thekké Madam, A.N. Rajagopalan, Analyzing Perception-Distortion Tradeoff using Enhanced Perceptual Super-resolution Network, 2018, 0–0.
- [204] H.A. Aly, E. Dubois, Image up-sampling using total-variation regularization with a new observation model, IEEE Trans. Image Process. 14 (10) (2005) 1647–1659, <https://doi.org/10.1109/TIP.2005.851684>. Oct.
- [205] S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, PMLR, 2015, pp. 448–456. Jun. 01..
- [206] C.K. Sønderby, J. Caballero, L. Theis, W. Shi, F. Huszár, Amortised MAP inference for image super-resolution, in: 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2016, <https://doi.org/10.48550/arxiv.1610.04490>. Oct.
- [207] R. Chen, Y. Qu, K. Zeng, J. Guo, C. Li, Y. Xie, Persistent Memory Residual Network for Single Image Super Resolution, 2018, pp. 809–816.
- [208] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, C. Schroers, A Fully Progressive Approach to Single-Image Super-Resolution, 2018, pp. 864–873.

- [209] Y. Bei, A. Damian, S. Hu, S. Menon, N. Ravi, C. Rudin, New Techniques for Preserving Global Structure and Denoising With Low Information Loss in Single-Image Super-Resolution, 2018, pp. 874–881.
- [210] T. Karras, T. Aila, S. Laine, J. Lehtinen, Progressive growing of GANs for improved quality, stability, and variation, in: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2017, <https://doi.org/10.48550/arxiv.1710.10196>. Oct.
- [211] D. Park, K. Kim, S. Young Chun, Efficient Module Based Single Image Super Resolution for Multiple Problems, 2018, pp. 882–890.
- [212] W. Han, S. Chang, D. Liu, M. Yu, M. Witbrock, T.S. Huang, Image Super-Resolution via Dual-State Recurrent Networks, 2018, pp. 1654–1663. Accessed: Sep. 07, 2022. [Online]. Available, <https://github.com/weihan3/dsrn>.
- [213] K. Malczewski, R. Stasiński, Super resolution for multimedia, image, and video processing applications, Stud. Comput. Intell. 231 (2009) 171–208, [https://doi.org/10.1007/978-3-642-02900-4\\_8/COVER](https://doi.org/10.1007/978-3-642-02900-4_8/COVER).
- [214] P. Wang, B. Bayram, E. Sertel, A comprehensive review on deep learning based remote sensing image super-resolution methods, Earth Sci. Rev. 232 (2022), 104110, <https://doi.org/10.1016/J.EARSCIREV.2022.104110>. Sep.
- [215] H. Chen, et al., Real-world single image super-resolution: a brief review, Inform. Fusion 79 (2022) 124–145, <https://doi.org/10.1016/J.INFFUS.2021.09.005>. Mar.
- [216] E. Gm and M. Shah, “PROCESSING OF SATELLITE IMAGE USING DIGITAL IMAGE PROCESSING”.
- [217] C.H. Genitha, K. Vani, Super resolution mapping of satellite images using hopfield neural networks, in: Proceedings of the International Conference on “Recent Advances in Space Technology Services and Climate Change - 2010”, RSTS and CC-2010, 2010, pp. 114–118, <https://doi.org/10.1109/RSTCC.2010.5712813>.
- [218] H. Zhang, Z. Yang, L. Zhang, H. Shen, Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences, Remote Sensing 6 (1) (2014) 637–657, <https://doi.org/10.3390/RS6010637>, 2014, Vol. 6, Pages 637–657.Jan.
- [219] V. A. Tolpekin, J. P. Ardila, and W. Bikker, “SUPER-RESOLUTION MAPPING FOR EXTRACTION OF URBAN TREE CROWN OBJECTS FROM VHR SATELLITE IMAGES”.
- [220] Y. Zhuang, J. Zhang, F. Wu, Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation, Pattern. Recognit. 40 (11) (2007) 3178–3194, <https://doi.org/10.1016/J.PATCOG.2007.03.011>. Nov.
- [221] X. Wang, X. Tang, Hallucinating face by eigentransformation, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 35 (3) (2005) 425–434, <https://doi.org/10.1109/TSMCC.2005.848171>. Aug.
- [222] W. Liu, D. Lin, X. Tang, Hallucinating faces: TensorPatch super-resolution and coupled residue compensation, in: Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005 II, 2005, pp. 478–484, <https://doi.org/10.1109/CVPR.2005.172>.
- [223] Z. Yuan, J. Wu, S.I. Kamata, A. Ahrary, P. Yan, Fingerprint image enhancement by super resolution with early stopping, in: Proceedings - 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS 2009 4, 2009, pp. 527–531, <https://doi.org/10.1109/ICICISYS.2009.5357637>.
- [224] Y.-H. Li, Doctoral Thesis in the field of Language Technologies Long Range and Unconstrained Iris Recognition, 2010.
- [225] R. Barnard, et al., High-resolution iris image reconstruction from low-resolution imagery, 6313, 2006, pp. 116–128, <https://doi.org/10.1117/12.681930>. Aug.
- [226] M. Dirk Robinson, S.J. Chiu, C.A. Toth, J.A. Izatt, J.Y. Lo, S. Farsiu, New applications of super-resolution in medical imaging, Super-Resolution Imaging (2017) 383–412, <https://doi.org/10.1201/9781439819319-13>. Dec.
- [227] G.S. Sable, A.N. Gaikwad, A Novel Approach for Super Resolution in Medical Imaging, 2022. Accessed: Sep. 07,[Online]. Available, [www.ijetae.com](http://www.ijetae.com).
- [228] H. Greenspan, G. Oz, N. Kiryati, S. Peled, MRI inter-slice reconstruction using super-resolution, Magn. Reson. Imaging 20 (5) (2002) 437–446, [https://doi.org/10.1016/S0730-725X\(02\)00511-8](https://doi.org/10.1016/S0730-725X(02)00511-8). Jun.
- [229] S. Peled, Y. Yesurun, Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging, Magn. Reson. Med. 45 (2001) 29–35, <https://doi.org/10.1002/1522-2594>.
- [230] R.R. Peeters, et al., The use of super-resolution techniques to reduce slice thickness in functional MRI, Int. J. Imaging Syst. Technol. 14 (3) (2004) 131–138, <https://doi.org/10.1002/IMA.20016>. Jan.
- [231] J.A. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, H. Azhari, Super-resolution in PET imaging, IEEE Trans. Med. Imaging 25 (2) (2006) 137–147, <https://doi.org/10.1109/TMI.2005.861705>. Feb.
- [232] J.A. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, H. Azhari, Improved image fusion in PET/CT using hybrid image reconstruction and super-resolution, Int. J. Biomed. Imaging (2007) 2007, <https://doi.org/10.1155/2007/46846>.
- [233] J. Liu, H. Yan, J. de Sun, Multiframe image restoration and registration, Adv. Comput. Visual image Process. 1 (1) (1984) 317–339, <https://doi.org/10.1587/TRANSINF.E92.D.59>.
- [234] H. Wang, D. Wen, The progress of sub-pixel imaging methods, 9142, 2014, pp. 166–170, <https://doi.org/10.1117/12.2054205>. Feb.
- [235] H. Shen, M.K. Ng, P. Li, L. Zhang, Super-Resolution reconstruction algorithm to MODIS remote sensing images, Comput. J. 52 (1) (2009) 90–100, <https://doi.org/10.1093/COMJNL/BXM028>. Jan.
- [236] M.T. Merino, J. Núñez, Super-resolution of remotely sensed images with variable-pixel linear reconstruction, IEEE Trans. Geosci. Remote Sensing 45 (5) (2007) 1446–1457, <https://doi.org/10.1109/TGRS.2007.893271>. May.
- [237] M. Elbakary, M. Alam, Superresolution construction of multispectral imagery based on local enhancement, IEEE Geosci. Remote Sens. Lett. 5 (2) (2008) 276–279, <https://doi.org/10.1109/LGRS.2008.915935>. Apr.
- [238] J.C.-W. Chan, J. Ma, F. Canters, A comparison of superresolution reconstruction methods for multi-angle CHRIS/Proba images, 7109, 2008, pp. 23–33, <https://doi.org/10.1117/12.800256>. Oct.
- [239] C. Mongkolsawat, N. Wattanakajit, T. Kamchai, K. Mongkolsawat, D. Chuyakhai, Super-resolution for SPOT5 - Beyond supermode, Scopus 1 (2009) 652–657.
- [240] Y. Zhong, L. Zhang, Remote sensing image subpixel mapping based on adaptive differential evolution, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 42 (5) (2012) 1306–1329, <https://doi.org/10.1109/TSMCB.2012.2189561>.
- [241] D. Vikhamar, R. Solberg, Subpixel mapping of snow cover in forests by optical remote sensing, Remote Sens. Environ. 84 (1) (2003) 69–82, [https://doi.org/10.1016/S0034-4257\(02\)00098-6](https://doi.org/10.1016/S0034-4257(02)00098-6). Jan.
- [242] Y. Ge, S. Li, V.C. Lakhan, Development and testing of a subpixel mapping algorithm, IEEE Trans. Geosci. Remote Sens. 47 (7) (2009) 2155–2164, <https://doi.org/10.1109/TGRS.2008.2010863>. Jul.
- [243] R. Patel and M. v Joshi, “SUPER-RESOLUTION OF HYPERSPECTRAL IMAGES USING COMPRESSIVE SENSING BASED APPROACH”.
- [244] F.A. Mianji, Y. Zhang, Y. Gu, Resolution enhancement of hyperspectral images using a learning-based super-resolution mapping technique, in: International Geoscience and Remote Sensing Symposium (IGARSS) 3, 2009, pp. 813–816, <https://doi.org/10.1109/IGARSS.2009.5417891>.
- [245] K. Murthy, M. Shearn, B.D. Smiley, A.H. Chau, J. Levine, M.D. Robinson, SkySat-1: very high-resolution imagery from a small satellite, 9241, 2014, pp. 367–378, <https://doi.org/10.1117/12.2074163>. Oct.
- [246] P. D'Angelo, G. Kuschl, P. Reinartz, Evaluation of Skybox Video and Still Image products, “International Archives of the Photogrammetry, Remote Sens. Spatial Inform. Sci. - ISPRS Arch. 40 (1) (2014) 95–99, <https://doi.org/10.5194/ISPRSARCHIVES-XL-1-95-2014>.
- [247] ZEISS Microscopy Online Campus | Introduction to Superresolution Microscopy, 2022. [https://zeiss-campus.magnet.fsu.edu/articles/superresolution/introduction\\_n.html](https://zeiss-campus.magnet.fsu.edu/articles/superresolution/introduction_n.html), accessed Sep. 07..
- [248] A. Small, S. Stahlheber, Fluorophore localization algorithms for super-resolution microscopy, Nature Methods 11 (3) (2014) 267–279, <https://doi.org/10.1038/nmeth.2844>, 2014 11:3Feb.
- [249] C. Liu, D. Sun, A Bayesian approach to adaptive video super resolution, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011, pp. 209–216, <https://doi.org/10.1109/CVPR.2011.5995614>.
- [250] Y. Wang, R. Fevig, R.R. Schultz, Super-resolution mosaicking of UAV surveillance video, in: Proceedings - International Conference on Image Processing, ICIP, 2008, pp. 345–348, <https://doi.org/10.1109/ICIP.2008.4711762>.
- [251] X. Zhang, M. Tang, R. Tong, Robust super resolution of compressed video, Visual Comput. 28 (12) (Jan. 2012) 1167–1180, <https://doi.org/10.1007/S00371-011-0666-8>, 2012 28:12.
- [252] C.A. Segall, R. Molina, A.K. Katsaggelos, High-resolution images from low-resolution compressed video, IEEE Signal Process. Mag. 20 (3) (2003) 37–48, <https://doi.org/10.1109/MSP.2003.1203208>.
- [253] H.E. Payne, R.I. Jedrzejewski, R.N. Hook, Astronomical data analysis software and systems XII, ASPC 295 (2003).
- [254] L. Zhang, H. Zhang, H. Shen, P. Li, A super-resolution reconstruction algorithm for surveillance images, Signal Process. 90 (3) (2010) 848–859, <https://doi.org/10.1016/J.SIGPRO.2009.09.002>. Mar.
- [255] A. Bulat, J. Yang, G. Tzimiropoulos, To learn image super-resolution, use a GAN to learn how to do image degradation first, 2018, pp. 185–200.
- [256] S.W. Zamir, et al., Learning enriched features for real image restoration and enhancement, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 12370 LNCS (2020) 492–511, [https://doi.org/10.1007/978-3-030-58595-2\\_30](https://doi.org/10.1007/978-3-030-58595-2_30)/COVER.
- [257] J. Cai, et al., NTIRE 2019 Challenge on Real Image Super-Resolution: Methods and Results, 2019, 0–0Accessed: Sep. 07, 2022.
- [258] S. Nah, et al., NTIRE 2019 Challenge on Video Deblurring and Super-Resolution: Dataset and Study, 2019, 0–0.
- [259] E. Prashnani, H. Cai, Y. Mostofi, P. Sen, PieAPP: Perceptual Image-Error Assessment Through Pairwise Preference, 2018, pp. 1808–1817, <https://doi.org/10.7919/F4GX48M7>.
- [260] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep Image Prior, 2018, pp. 9446–9454. Accessed: Sep. 07, 2022. [Online]. Available, <http://www.cs.tut.fi/>.
- [261] F. Wu, W. Dong, G. Shi, X. Li, Learning Hybrid Sparsity Prior for Image Restoration: Where Deep Learning Meets Sparse Coding, 2018, <https://doi.org/10.48550/arxiv.1807.06920>. Jul.