A comprehensive review of deep learning-based single image super-resolution

Syed Muhammad Arsalan Bashir[1,2], Yi Wang[2], Mahrukh Khan[3], Yilong Niu[4]

[1] Quality Assurance, Pakistan Space and Upper Atmosphere Research Commission, Karachi, Sindh, Pakistan.

[2] School of Electronics and Information, Northwestern Polytechnical University, Xi'an, Shaanxi, China

[3] Department of Computer Science, National University of Computer and Emerging Sciences, Karachi, Sindh, Pakistan

[4] School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi, China

Corresponding Authors:
Yilong Niu and Syed Muhammad Arsalan Bashir
Email address: yilong_niu@nwpu.edu.cn; smarsalan@mail.nwpu.edu.cn

Abstract

Image super-resolution (SR) is one of the vital image processing methods that improve the resolution of an image in the field of computer vision. In the last two decades, significant progress has been made in the field of super-resolution, especially by utilizing deep learning methods. This survey is an effort to provide a detailed survey of recent progress in single-image super-resolution in the perspective of deep learning while also informing about the initial classical methods used for image super-resolution. The survey classifies the image SR methods into four categories, i.e., classical methods, supervised learning-based methods, unsupervised learning-based methods, and domain-specific SR methods. We also introduce the problem of SR to provide intuition about image quality metrics, available reference datasets, and SR challenges. Deep learning-based approaches of SR are evaluated using a reference dataset. Some of the reviewed state-of-the-art image SR methods include enhanced deep SR network (EDSR), cycle-in-cycle GAN (CinCGAN), multiscale residual network (MSRN), meta residual dense network (Meta-RDN), recurrent back-projection network (RBPN), second-order attention network (SAN), SR feedback network (SRFBN) and wavelet-based residual attention network (WRAN). Finally, this survey is concluded with future directions and trends in SR and open problems in SR to be addressed by the researchers.

1. Introduction

The image-based computer graphics models lack resolution independence (Freeman, Jones & Pasztor, 2002) as the images cannot be zoomed beyond the image sample resolution without compromising the quality of images. This is the case, especially in realistic images, for instance, natural photographs. Thus, simple image interpolation will lead to the blurring of features and edges within a sample image.

The concept of super-resolution was first used by (Gerchberg, 1974) to improve the resolution of an optical system beyond the diffraction limit. In the past two decades, the concept of super-resolution (SR) is defined as the method of producing high-resolution (HR) images from a corresponding low-resolution (LR) image. Initially, this technique was classified as spatial resolution enhancement (Tsai & Huang, 1984). The applications of super-resolution include computer graphics (Kim, Lee & Lee, 2016a,b; Tao et al., 2017), medical imaging (Bates et al., 2007; Fernández-Suárez & Ting, 2008; Huang et al., 2008; Hamaide et al., 2017; Jurek et al., 2020; Teh et al., 2020; Bashir & Wang, 2021a), security, and surveillance (Zhang et al., 2010; Shamsolmoali et al., 2018; Lee, Kim & Heo, 2020), which shows the importance of this topic in recent years.

Although being explored for decades, image super-resolution remains a challenging task in computer vision. This problem is fundamentally ill-posed because there can be several HR images with slight variations in camera angle, color, brightness, and other variables for any given LR image. Furthermore, there are fundamental uncertainties among the LR and HR data since the downsampling of different HR images may lead to a similar LR image, making this conversion a many-to-one process (Yang & Yang, 2013).

59 The existing methods of image super-resolution can be categorized into single-image super-
60 resolution (SISR) and multiple-image approaches. In single image SR, the learning is performed
61 for single LR-HR pair for a single image, while in multiple-image SR, the learning is performed
62 for a large number of LR-HR pairs for a particular scene, thereby enabling the generation of an
63 HR image from a scene (multiple images) (Kawulok et al., 2020). Video super-resolution deals
64 with multiple successive images (frames) and utilizes the relationship within the frames to super-
65 resolve a target frame; it is a special type of multiple image SR where the images are part of a
66 scene containing different frames (Liu et al., 2020b).
67 In the past, classical SR methods such as statistical methods, prediction-based methods, patch-
68 based methods, edge-based, and sparse representation methods were used to achieve super-
69 resolution. However, recently the advances in computational power and big data have made
70 researchers use deep learning (DL) to address the problem of SR. In the past decade, deep
71 learning-based SR studies have reported superior performance than the classical methods, and
72 DL methods have been used frequently to achieve SR. Researchers have used a range of methods
73 to explore SR, ranging from the first method of Convolutional Neural Network (CNN) (Dong et
74 al., 2014) to the recently used Generative Adversarial Nets (GAN) (Ledig et al., 2017). In
75 principle, the methods used in deep learning-based SR methods vary in hyper-parameters such as
76 network architecture, learning strategies, activation functions, and loss functions.
77 In this study, a brief overview of the classical methods of SR is outlined initially, whereas the
78 main focus is given to give an overview of the most recent research in SR using deep learning.
79 Previous studies have explored the literature on SR, but most of these studies emphasize the
80 classical methods (Borman & Stevenson, 1998; Park, Park & Kang, 2003; van Ouwerkerk, 2006;
81 Yang, Ma & Yang, 2014; Thapa et al., 2016), additionally (Yang, Ma & Yang, 2014; Thapa et
82 al., 2016) used human visual perception to gauge the performance of SR methods.
83 In recent years, there are some reviews (Ha et al., 2019; Yang et al., 2019; Zhang et al., 2019c;
84 Zhou & Feng, 2019; Li et al., 2020) focused on deep learning-based image super-resolution. The
85 study by Yang et al. (Yang et al., 2019) was focused on the deep learning methods for single
86 image super-resolution Zhang et al. (Zhang et al., 2019c) limited the scope of image SR to
87 CNN-based methods for space applications, thereby only reviewing four methods namely,
88 SRCNN, FSRCNN, VDSR, and DRCN. Ha et al. (Ha et al., 2019) reviewed the state-of-the-art
89 SISR methods and classified them based on the type of framework, i.e., CNN, RNN-CNN-based
90 methods, and GAN-based methods. Zhou et al. (Zhou & Feng, 2019) briefly reviewed some of
91 the state-of-the-art SISR methods and provided an introduction of some of the methods without
92 any evaluation of comparison of methods, while Li et al. (Li et al., 2020) reviewed the state-of-
93 the-art methods in image SR while emphasizing on the methods based on CNNs and GANs for
94 real-time applications. These review papers did not encompass the domain of super-resolution as
95 a whole, and this paper fills that research gap by providing an overview of both classical and
96 deep learning-based methods. At the same time, we have reviewed the deep learning-based
97 methods into subdomain based on the functional blocks, i.e., upsampling methods, SR networks,
98 learning strategies, SR framework and other improvements. This review paper fills the gap of a

99   comprehensive review where a reader could access the overall progress of image super-
100   resolution with appropriate section for the overall image quality metrics, SR methods, datasets,
101   applications, and challenges in the field of image SR.
102   This survey is a comprehensive overview of the recent advances in SR, emphasizing deep
103   learning-based approaches and their achievements in systematically achieving SR. Table S1 and
104   Table S2 respectively show the complete list of symbols and acronyms used in this study.
105   The key features of this study are:

1.   We highlight the brief overview of the classical methods in SR and their contributions in
106   light of past studies to give perspective.
108   2.   We provide a detailed survey of deep learning-based SR, including the definition of the
109   problem, dataset details, performance evaluation, deep learning methods used for SR, and
110   specific applications where these SR methods were used and their performance.
111   3.   We compare and contrast the recent advances in deep learning-based SR methods by
112   summarizing the bounds of the methods by providing details of components of the SR
113   methods used structurally.
114   4.   Finally, the open problems in SR and critical challenges that require further probing are
115   highlighted in this survey to provide future directions in SR.

116

117   This study is organized as follows:
118   In Section 1, we have introduced the concept of SR and the overall overview of this study. In
119   Fig. 1, we have summarized the hierarchical structure of this review. There are four main
120   sections: classical methods, deep learning-based methods, applications of SR, Discussion, and
121   future directions. In Section 2, we put forward the problem definition and details of the
122   evaluation dataset. Section 3 discusses the methodology for the selection of studies included
123   within this review. In Section 4, we compare and contrast the classical methods of SR, whereas,
124   in Section 5, the SR methods based on supervised deep learning are explored. Section 6 covers
125   the studies that used unsupervised deep learning-based methods for SR, and in Section 7, various
126   field-specific applications of SR in recent years are discussed. Section 8 summarizes open
127   challenges and limitations in current SR methods and puts forward future research directions,
128   while Section 9 highlights the conclusions.

129

130 2. Super-resolution: Definitions and Terminologies

131   In this section, the problem definition and the associated concepts of image super-resolution are
132   discussed in light of the literature review.

133

134 2.1. Single Image Super-resolution – Problem Definition

135   The image SR focuses on the recovery of an HR image from LR image input as and in principle,
136   the LR image I_{xLR} can be represented as the output of the degradation function, as shown in (1).

137
$$I_{xLR} = d\left(I_{yHR}, \partial\right) \tag{1}$$

138 Where d is the SR degradation function that is responsible for the conversion of HR image to LR
139 image, I_{yHR} is the input HR image (reference image), whereas ∂ depicts the input parameters of
140 the image degradation function. Degradation parameters are usually scaling factor, blur type, and
141 noise. In practice, the degradation process and dependent parameters are unknown, and only LR
142 images are used to get HR images by the SR method. The SR process is responsible for
143 predicting the inverse of the degradation function d, such that $g = d^{-1}$

$$g\left(I_{xLR}, \delta\right) = d^{-1}\left(I_{xHR}\right) = I_{yE} \approx I_{yHR} \tag{2}$$

144

145 Where g is the SR function, δ depicts the input parameters to the function g, and I_{yE} is the

146 estimated HR corresponding to the input I_{xLR} image. It is also worth noticing that the super-

147 resolution function, as in (2), is ill-posed, as the function g is a non-injective function; thus, there

148 are infinite possibilities of I_{yE} for which the condition $d\left(I_{yE}, \partial\right) = I_{xLR}$ will hold.

149 The degradation process for the input LR images is unknown, and this process is affected by
150 numerous factors such as sensor-induced noise, artifacts created because of lossy compression,
151 speckle noise, motion blur, and misfocused images. In the literature, most of the studies have
152 used a single downsampling function as the image degradation function:

$$d\left(I_{yHR}, \partial\right) = \left(I_{yHR}\right)\downarrow_{s_f}, \{s\} \subseteq \partial \tag{3}$$

153

154 Where $\downarrow_{s_f}$ is the downsampling operator with s_f being the scaling factor. One of the frequently

155 used downsampling functions in SR is the bicubic interpolation (Shi et al., 2016; Zhang & An,
156 2017; Shocher, Cohen & Irani, 2018) with antialiasing. In some studies, like (Zhang, Zuo &
157 Zhang, 2018), researchers have used more operations in the downsampling function, and the
158 overall downsampling operation is:

$$d\left(I_{yHR}, \partial\right) = \left(I_{yHR} \otimes \kappa\right)\downarrow_{s_f} + n_\sigma, \{\kappa, s, \sigma\} \subseteq \partial \tag{4}$$

159

160 Where $I_{yHR} \otimes \kappa$ depicts the convolution of the HR image I_{yHR} with the blurring kernel κ, n_σ

161 represents the additive white Gaussian noise with a standard deviation of σ. The degradation
162 function defined in (4) and Fig. 2 is closer to the actual function as it considers more parameters
163 than the simple downsampling degradation function (Zhang, Zuo & Zhang, 2018).
164
165 Finally, the purpose of SR is to minimize the loss function as follows:

$$\hat{\phi} = \left[\min \mathsf{L}\left(I_{yE}, I_{yHR}\right)\right]_\phi + h\Psi\left(\phi\right) \tag{5}$$

166

167 Where $\mathsf{L}\left(I_{yE}, I_{yHR}\right)$ is the loss function between the output HR image of SR and the actual HR

168 image, h is the tradeoff parameter, whereas $\Psi\left(\phi\right)$ is the regularization term. The most common

169 loss function used in SR is the pixel-based mean square error (MSE), which can also be referred
170 to as pixel loss. In recent years, researchers have used a combination of various loss functions,

171 and these combinations are further explored in later sections. Further mathematical modeling of
172 the SR problem is discussed in (Candès & Fernandez-Granda, 2014).
173

174 **2.2. Methods for quality of SR images**
175 Image quality can have several definitions as per the measurement methods, and it is generally a
176 measure of the quality of visual attributes and perception of the viewers. The image quality
177 assessment (IQA) methods are characterized into subjective methods (human perception of an
178 image is natural and of good quality) and objective methods (quantitative methods by which
179 image quality can be numerically computed) (Thung & Raveendran, 2009).
180 Quality-related visual aspects of an image are mostly a good measure, but this method requires
181 more resources, especially if the dataset is large (Wei, Yuan & Cai, 1999); thus, in SR and
182 computer vision tasks, the more suitable methods are objective. As per (Saad, Bovik & Charrier,
183 2012), the IQA methods are primarily categorized into three categories, i.e., reference image-
184 based features from the actual image and blind IQA with no information about the ground truth.
185 In this section, IQA methods primarily used in the domain of SR are further explored.
186

187 ***2.2.1. Peak Signal-to-Noise Ratio***
188 In information systems, the peak signal-to-noise ratio (PSNR) is a measurement technique for
189 analyzing the signal power compared to the noise power, especially in images; the PSNR is used
190 as a quantitative measure of the compression quality of an image. In super-resolution, the PSNR
191 of an image is defined by the maximum pixel value and the mean square error between the
192 reference image and the SR image, also known as the power of image distortion noise. For a
193 given maximum pixel value (M) and the reference image (I_r) having t pixels and the SR image
194 (I_y), the peak signal-to-noise ratio is defined as:

$$PSNR = 10\log_{10}\left(\frac{M^2}{MSE}\right)$$

195 (6)

196 Where M is mostly for 8-bit color space depth, i.e., the max value of 255 and *MSE* is given by:

$$MSE = \frac{1}{t}\sum_{i=1}^{t}\left(I_r(i) - I_y(i)\right)^2$$

197 (7)

198 As seen from (6), the PSNR is related to the individual pixel intensity values of the SR image
199 and reference image and is a pixel-based metric of image quality. In some cases (Almohammad
200 & Ghinea, 2010; Horé & Ziou, 2010; Goyal, Lather & Lather, 2015), this quality metric can be
201 misleading as the overall image might not be visually similar to that of the reference image. This
202 metric is still used for image comparisons, especially comparing the results of SR algorithms
203 with previously published results to compare the working of any new method in the field of SR.
204 MSE for color images averaged for color channels, and an alternate approach is to measure
205 PSNR for luminance and or greyscale channels separately as the human eye is more sensitive to
206 changes in luminance in contrast to changes in chrominance (Dabov et al., 2006).
207

208 **2.2.2. Structural Similarity Index**
209 The visual perception of humans is efficient in extracting the structural information within an
210 image, and PSNR does not consider the structural composition of the image (Rouse & Hemami,
211 2008). The structural similarity index metric (SSIM) was proposed by (Wang et al., 2004) to
212 measure the structural similarity between images by comparing the contrast, luminance, and
213 structural details within the reference image.
214 An image I_r with total pixels P; the contrast C_I, and luminance L_I can be denoted as the standard
215 deviation and the mean of the image intensity given by:

$$L_I = \frac{1}{M}\sum_{i=1}^{M} I_r(i)$$

216 (8)

$$C_I = \sqrt{\left(\frac{1}{M-1}\sum_{i=1}^{M}\left(I_r(i)-L_I\right)^2\right)}$$

217 (9)

218 The ith pixel of the reference image is denoted by $I_r(i)$. The comparisons based on the contrast
219 and luminance between the reference image I_r and the estimated image $\hat{I}$ are:

$$Com_l\left(I_r,\hat{I}\right) = \frac{2L_{I_r}L_{\hat{I}} + \mu_1}{L_{I_r}^2 + L_{\hat{I}}^2 + \mu_1}$$

220 (10)

$$Com_c\left(I_r,\hat{I}\right) = \frac{2C_{I_r}C_{\hat{I}} + \mu_2}{C_{I_r}^2 + C_{\hat{I}}^2 + \mu_2}$$

221 (11)

222
223 Where $\mu_1 = (k_1 S)^2$ and $\mu_2 = (k_2 S)^2$, these constant terms ensure stability by ensuring $k_1 << 1$ and
224 $k_2 << 1$.

225 Normalized pixel values $I_r - \frac{L_{I_r}}{C_{I_r}}$ represent the image structure, while the inner product of

226 these is the equivalent of structural similarity between the reference image I_r and the estimated
227 image $\hat{I}$. The covariance $\sigma_{I_r,\hat{I}}$ is given by:

$$\sigma_{I_r,\hat{I}} = \frac{1}{M-1}\sum_{i=1}^{M}\left(I_r(i)-L_{I_r}\right)\left(\hat{I}(i)-L_{\hat{I}}\right)$$

228 (12)

229 Function for structural comparison $Com_s(I_r,\hat{I})$ is given by:

$$Com_s\left(I_r,\hat{I}\right) = \frac{\sigma_{I_r,\hat{I}} + \mu_3}{C_{I_r}C_{\hat{I}} + \mu_3}$$

230 (13)

231 Where μ_3 is stability constant, the final structural similarity index (SSIM) is given by:

232
$$SSIM\left(I_r,\hat{I}\right)=\left\{Com_l\left(I_r,\hat{I}\right)\right\}^{\alpha}\left\{Com_c\left(I_r,\hat{I}\right)\right\}^{\beta}\left\{Com_s\left(I_r,\hat{I}\right)\right\}^{\gamma} \tag{14}$$

233 The control parameters α, β, and γ can be adjusted to increase the importance of luminance,
234 contrast, and structural comparison in calculating the $SSIM$.
235 Conventionally, PSNR is used in computer vision tasks for evaluation, but SSIM is based on
236 human perception of structural information within an image. Thus this method is widely used for
237 comparing the structural similarity between images (Blau & Michaeli, 2018; Sara, Akter &
238 Uddin, 2019). In medical images where the variance or luminance of the reference images are
239 low, SSIM could be very unstable, thus reporting false results; however, this is not the case for
240 natural images (Pambrun & Noumeir, 2015).
241

242 ### 2.2.3. Opinion Scoring
243 Opinion scoring is a qualitative method, which lies in the subjective category of IQA. In this
244 method, the quality testers are asked to grade the quality of images based on specific criteria,
245 e.g., sharpness, natural look, and color, where the final graded score is the mean of the rated
246 scores.
247 This method has limitations such as non-linearity between the scores, variation in results due to
248 changes in test criteria, and human error. In SR, certain methods have reported good objective
249 quality scores but scored poorly in subjective results, especially in human face reconstruction
250 (Ledig et al., 2017), (Nasrollahi & Moeslund, 2014; Chen et al., 2018b). Thus, the opinion
251 scoring method is also used in studies (Wei, Yuan & Cai, 1999), (Deng, 2018; Ravì et al., 2018,
252 2019; Vasu, Thekke Madam & Rajagopalan, 2019) to measure the quality of human perception.
253

254 ### 2.2.4. Perceptual Quality
255 Opinion scoring used human raters for manual evaluation of the images; while this method can
256 provide accurate results as far as human perception is concerned, this method requires many
257 resources, especially large datasets (Viswanathan & Viswanathan, 2005). Initially, (Kim & Lee,
258 2017) proposed a CNN-based full reference image quality assessment (FR-IQA) model where
259 human behavior was learned using an IQA database that contained distorted images, subjective
260 scores, and error maps, and this method was called DeepQA.
261 In (Ma et al., 2017a), the authors used quality-discriminable image pairs (DIP) for training, and
262 the system was called dipQA (DIP inferred quality index); they used RankNet with L2R
263 algorithm to learn blind opinion IQA, whereas in (Ma et al., 2018) a multi-task end-to-end
264 optimized deep neural network (MEON) was proposed. MEON used two stages, in the first
265 stage, distortion type learning using large datasets already available, and in the second stage, the
266 output of the first stage was used to train the quality assessment network using stochastic
267 gradient descent. In (Talebi & Milanfar, 2018), the authors used CNN to develop a no-reference
268 IQA method known as NIMA; NIMA was trained on pixel-level and aesthetic quality datasets.
269 RankIQA (Liu, Weijer & Bagdanov, 2017) trained a Siamese network to grade the quality of
270 images using datasets with known image distortions, CNNs were used to learn the IQA, and this

271   method even outperformed full-reference methods without using the reference image. IQA
272   proposed in (Bosse et al., 2018) included ten convolution layers and five pooling layers for
273   feature extraction while there were two fully connected layers for regression; this method
274   performed significantly well for both no-reference and full-reference IQA.
275   Even though opinion scoring and perceptual quality-based methods do exhibit human perception
276   in IQA, but the quality we require is still an open question (i.e., if we want images to be more
277   natural or similar to the reference image); thus, PSNR and SSIM are the primarily used methods
278   in computer vision and SR.
279
2.2.5. Task-based Evaluation
281   Although the primary purpose of image SR is to achieve better resolution, as mentioned earlier,
282   SR is also helpful in other computer vision tasks (Kim, Lee & Lee, 2016b), (Tao et al., 2017),
283   (Teh et al., 2020), (Liu et al., 2018a). The performance achieved in these can indirectly measure
284   the performance of the SR methods used in those tasks. In the case of medical images, the
285   researchers used the original and SR constructed images to see the performance in the training
286   and prediction phases. In general, computer vision tasks such as classification (Krizhevsky,
287   Sutskever & Hinton, 2012; Cai et al., 2019), face recognition (Nasrollahi & Moeslund, 2014; Liu
288   et al., 2015; Chen et al., 2018b), and object segmentation (Martin et al., 2001; Lin et al., 2014;
289   Wang et al., 2018b) can be done using SR images. The performance of these computer vision
290   tasks can be used as a metric to assess the performance of the SR method.
2.2.6. Miscellaneous IQA Methods
292   The development of IQA methods is an open field, and in recent years various researchers have
293   proposed SR metrics, but these methods were not used widely by the SR community. Feature
294   similarity (FSIM) index metric (Zhang et al., 2011) evaluates image quality by extracting feature
295   points considered by the human visual system based on gradient magnitude and phase
296   congruency. The multi-scale structural similarity (MS-SSIM) (Wang, Simoncelli & Bovik, 2003)
297   used multi-scale to incorporate variations in the viewing conditions to measure the image quality
298   and proposed that MS-SSIM provides form flexibility in the measurement of image quality than
299   single-scale SSIM. In (Li & Bovik, 2010), the authors claimed that SSIM and MS-SSIM do not
300   perform well on distorted and noisy images; thus, they used a four-component-based weighted
301   method that adjusted the weight of scores based on the local feature, whereas in the case of
302   contrast-distorted images (Yao & Liu, 2018) like TID2013 and CSIQ datasets SSIM does not
303   perform well.
304   According to (Blau & Michaeli, 2018), the perceptual quality and image distortion are at odds
305   with each other; as the distortion decreases, the perceptual quality should also be worse; thus, the
306   accurate measurement of SR image quality is still an open area of research.
307   The comparison of image quality assessment metrics for super-resolution is shown in Table 1. It
308   depends on the requirements of the methods; most of the methods use PSNR and SSIM to
309   evaluate the performance as these are quantitative methods.
310

2.3. Operating Color Channels

In most datasets, RGB color space is used; thus, SR methods mostly employ RGB images, YCbCr space is also used in SR (Dong et al., 2016). The Y component in YCbCr is the luminance component, which represents the light intensity, while Cb and Cr are the chrominance components (i.e., blue-differenced and red-differenced Chroma channels) (Shaik et al., 2015). In recent years, most of the SR challenges and datasets use the RGB color space, limiting the use of RGB space for comparison with state of the art. Furthermore, the results of IQA based on PSNR vary if the color space in the testing stage is different from the training/evaluation stage.

2.4. Details of the reference dataset

The datasets used in evaluating the SR algorithms are summarized in this section; the various datasets discussed in this section vary in the total number of example images, image resolution, quality, and imaging hardware setup. A few of the datasets comprise paired LR-HR images for training and testing SR algorithms. In contrast, the rest of the datasets include HR images, and the corresponding LR images are usually generated by using bicubic interpolation with antialiasing as performed in (Shi et al., 2016; Zhang & An, 2017; Shocher, Cohen & Irani, 2018). Matlab function imresize (I, scale), where the default method is bicubic interpolation with antialiasing, and scale is the downsampling factor input to the function.

Table 2 comprises a list of datasets frequently used in SR and information on total image count, image format, pixel count, HR resolution, type of dataset, and classes of images.

Most of the datasets for SR are unpaired data, and the LR images are generated using various scale factors using bicubic interpolation with antialiasing. Other than the mentioned datasets in Table 2, datasets like General-100 (Dong, Loy & Tang, 2016), L20 (Timofte, Rothe & Gool, 2016), and ImageNet (Deng et al., 2009) are also used in computer vision tasks. In recent times, researchers have preferred the use of multiple datasets for training/evaluation and testing the SR models; for instance, in (Bashir & Ghouri, 2014; Lai et al., 2017; Sajjadi, Scholkopf & Hirsch, 2017; Tong et al., 2017), the researchers used SET5, SET14, BSDS100, and URBAN100 for training and testing.

2.5. Super-resolution Challenges

The most prominent SR challenges NTIRE (Agustsson & Timofte, 2017), (Timofte et al., 2017), and PIRM (Blau et al., 2018), are discussed in this section.

The New Trends in Image Restoration and Enhancement (NTIRE) challenge (Agustsson & Timofte, 2017; Timofte et al., 2017) was in collaboration with the Conference on Computer Vision and Pattern Recognition (CVPR). NTIRE includes various challenges like colorization, image denoising, and SR. In the case of SR, the DIV2K dataset (Agustsson & Timofte, 2017) was used, which included bicubic downscaled image pairs and blind images with realistic but unknown degradation. This dataset has been widely used to evaluate SR methods under known and unknown conditions to compare against the state-of-the-art methods.

350   The perceptual image restoration and manipulation (PRIM) challenges were in collaboration
351   with the European Conference on Computer vision (ECCV), and like NTIRE, it contained
352   multiple challenges. Apart from the three challenges mentioned in NTIRE, PRIM also focused
353   on SR for smartphones and compared perceptual quality with generation accuracy (Blau et al.,
354   2018). As mentioned by (Blau & Michaeli, 2018), the models that focus on distortion often give
355   visually unpleasant SR images, while the models focusing on the perceptual image quality do not
356   perform well on information fidelity. Using the image quality metrics NIQE (Mittal,
357   Soundararajan & Bovik, 2013) and (Ma et al., 2017b), the methods that performed best in
358   achieving perceptual quality (Blau & Michaeli, 2018) was the winner. In contrast, in a sub-
359   challenge (Ignatov et al., 2018b), SR methods were evaluated using limited resources to evaluate
360   SR performance for smartphones using the PSNR, MS-SSIM, and opinion scoring metrics. Thus,
361   PIRM encouraged the researchers to explore the perception-distortion tradeoff domain and SR
362   for smartphones.
363

3. Survey Methodology

365   The majority of the studies included in this review paper are peer-reviewed publications to
366   ensure the validity of the methods; these studies include conference proceedings and journal
367   papers. The included papers include early access and a published version of recent papers for
368   super-resolution from 2008 to 2021. However, some in classical methods, some papers, and
369   initial papers on image SR were included before this range to develop the review and give the
370   background of the classical methods developed before the deep learning-based methods overtook
371   the field. Google Scholar, IEEE Xplore, and Science Direct were queried to collect the initial list
372   of papers in this research. Specific keywords were used to search the databases and based on the
373   abstract. A further selection of papers was made using the reference sections of the selected
374   papers as they contain additional relevant studies in image super-resolution. The last query was
375   made on May 08, 2021. The collected papers were segregated based on their relevance with the
376   Section; for example, papers with supervised learning were stored separately for review in
377   Section 5, and studies highlighting the applications of SR methods were grouped for discussion
378   in Section 7.
379   The relevant search terms include image super-resolution, super-resolution, deep learning super-
380   resolution, convolutional neural networks, image upsampling methods, super-resolution
381   frameworks, supervised super-resolution, unsupervised super-resolution, super-resolution
382   review, image interpolation, pixel-based methods, super-resolution application, assisted
383   diagnosis using deep learning. The search keywords were not limited to single-image SR because
384   our target was to report other aspects of super-resolution, including classical methods,
385   applications, and datasets for SR.
386   Logical operators and wildcards were used to combine the keywords further and perform the
387   additional search. Initial screening of the collected papers was performed following the
388   inclusion/exclusion criteria shown in Table 3. The whole process is graphically shown in Fig. 3,

389   where 653 studies were collected over one year. A total of 242 studies were included from the
390   initial 653 collected research studies.
391

4. Conventional Methods of Super-Resolution

393   Classical methods of SR are briefly discussed in this section to encompass the overall
394   development cycle of the SR. The classical methods include prediction-based, edge-based,
395   statistical, patch-based, and sparse representation methods.
396   The primary methods were based on prediction, and the first method (Duchon, 1979) was based
397   on Lanczos filtering, which filtered the digital data using sigma factors (with modifiable weight
398   function), and a similar frequency-domain filtering approach was used in (Tsai & Huang, 1984)
399   for image resampling. In contrast, cubic convolution (Keys, 1981) was used for resampling the
400   image data, and the results showed that this prediction method was more accurate than the
401   nearest-neighbor prediction algorithm and linear interpolation of image data (Parker, Kenyon &
402   Troxel, 1983). In (Tsai & Huang, 1984), the authors did not consider the blur in the imaging
403   process, while (Irani & Peleg, 1991) used the knowledge of the imaging process and the relative
404   displacements for image interpolation when the sampling rate was kept constant and this method
405   reduced to deblurring.
406   The patch-based approach was used in (Freeman, Jones & Pasztor, 2002); the authors used a
407   training set where various patches within the training set were extracted as training patterns,
408   which helped generate detailed high-frequency images using the patch texture information. In
409   (Chang, Yeung & Xiong, 2004), the authors used locally linear embedding to use local patches
410   for generating high-resolution images based on the local patch features. In contrast, (Glasner,
411   Bagon & Irani, 2009) used the concept of reoccurrence of geometrically similar patches in
412   natural images to select the best possible pixel value based on the patch redundancy on the same
413   scales. In (Baker & Kanade, 2002), the authors introduced the concept of hallucination, where
414   they extracted local features within the LR image first and used these to map the HR image.
415   Edge-based methods use edge smoothness priors to upsample images, and in (Sun, Xu & Shum,
416   2008), a generic image prior, gradient prior profile was used to smoothen the edges within an
417   image to achieve super-resolution in natural images. In (Freedman & Fattal, 2011), the authors
418   used specially designed filters to search for similar patches using the local self-similarity
419   observation, which performed lower nearest patch computations; this method was able to
420   reconstruct realistic-looking edges, whereas it performed poorly in clustered regions with fine
421   details.
422   Statistical methods were used to perform image super-resolution (Kim & Kwon, 2010), where
423   the authors used Kernel ridge regression (KRR) with gradient descent to learn the mapping
424   function from the image example pairs. Adaptive regularization was used to supervise the energy
425   change during the image resampling iterative process. This provided more accurate results as the
426   energy map was used to limit the energy change per iteration, which reduced the noise while
427   maintaining the perceptual quality (Xiong, Sun & Wu, 2010) while (Yang et al., 2010), (Yang et

428   al., 2008) used sparse representation methods to perform image super-resolution which used the
429   concept of compressed sensing.
430   The robust SR method proposed in (Zomet, Rav-Acha & Peleg, 2001) used the information of
431   outliers to improve the performance of SR in patches where other methods introduce noise due to
432   these outliers. Additionally, (Yang et al., 2007) proposed a post-processing model that enhanced
433   the resolution of a set of images using a single reference image up to 100x scaling factor.
434   Another way to achieve SR is to use LR images to achieve a single HR image (Tipping &
435   Bishop, 2003). The conventional upsampling methods, such as interpolation-based, use the
436   information within the LR image to generate HR images, and these methods do not add any new
437   information to the image (Farsiu et al., 2004a). Furthermore, they also introduce some inherent
438   problems, such as noise amplification and blur enhancement. Thus, in recent years, the
439   researchers have shifted to learning-based upsampling methods explored in Section 5.
440

5. Supervised Super-resolution

442   Various deep learning methods were developed over the years to solve the SR problem; in this
443   section, the models discussed are trained using both low and high-resolution images (LR-HR
444   pairs). Although there are significant differences in the supervised SR models, and the models
445   can be classified based on the components like the upsampling method employed, deep learning
446   network, learning algorithm, and model frameworks. Any supervised image SR model is based
447   on the combinations of these components, and in this section, we summarize the employed
448   methods for these four components in light of recent supervised image SR research studies.
449   The component-based review of various methods is performed in this section, and the basic
450   overview of the models is shown in Fig. 1.
451

5.1. Upsampling Methods

453   The upsampling is essential in deep learning-based SR methods such as its positioning, and the
454   method performed for upsampling has a significant impact on the training and test performance
455   of the model. There are some commonly used methods (Yang et al., 2010), (Yang et al., 2008),
456   (Lee, Yang & Oh, 2015; Timofte, De Smet & Van Gool, 2015), which use the conventional
457   CNNs for end-to-end learning. In this subsection, various deep learning-based upsampling layers
458   are discussed.
459   As mentioned in Section 4, the interpolation-based methods of upsampling do not add any new
460   information; hence, learning-based methods are used in image SR in the last decade.
461

5.1.1. Sub-pixel Layer

463   The end-to-end learning layer (Shi et al., 2016), called the sub-pixel layer, performs upsampling
464   by generating several additional channels using convolution, and by reshaping these channels,
465   this layer performs upsampling, as shown in Fig. 4. In this layer, convolution is applied to s_f^2
466   where s_f is the scaling factor, as shown in Fig. 4b. Since the input image size is $h \times w \times c$ where

467   h is height, w is width, and c depicts color channels, the resulting convolution is $h \times w \times cs_f^2$. In

468   order to achieve the final image, a reshuffling (Shi et al., 2016) operation is performed to get the

469   final output image $s_f h \times s_f w \times c$, as shown in Fig. 4c. Since it is an end-to-end layer, this layer is

470   frequently used in SR models (Ledig et al., 2017), (Zhang, Zuo & Zhang, 2018), (Ahn, Kang &

471   Sohn, 2018a), (Zhang et al., 2018b).

472   This layer has a wide receptive field, which helps learn more contextual information that

473   generates realistic details, whereas this layer may generate some false artifacts at the boundaries

474   of complex patterns due to its uneven distribution of the respective field. Furthermore, predicting

475   the neighborhood pixels in a block-type region sometimes results in unsmooth outputs that do

476   not look realistic when compared with the true HR image; to address this issue, PixelTCL (Gao

477   et al., 2020) was proposed that used the interdependent prediction layer, which used the

478   information of the interlinked pixels during upsampling. The results were smooth and more

479   realistic when compared with the ground truth image.

480

481   **5.1.2. Deconvolution Layer**

482   The deconvolution layer also referred to as transposed convolution layer (Zeiler et al., 2010), is

483   the converse of the convolution, i.e., predicting the probable input HR-image based on the

484   feature maps from the LR image. In this process, additional zeros are inserted to increase the

485   resolution, and afterwards, convolution is performed. For instance, taking scaling factor 2 for the

486   SR image, a convolution kernel of 3×3 (as shown in Fig. 5a, 5b, and 5c), the input LR image is

487   expanded twice by inserting zeros, convolution with the kernel is performed by using a stride and

488   padding of 1.

489   The deconvolution layer is widely used in SR methods (Sroubek, Cristobal & Flusser, 2008;

490   Hugelier et al., 2016; Lam et al., 2017; Tong et al., 2017; Haris, Shakhnarovich & Ukita, 2018),

491   as it generates HR images in an end-to-end way, and it has compatibility with the vanilla

492   convolution. As per (Odena, Dumoulin & Olah, 2017), in some cases, this layer may cause the

493   problem of uneven overlapping within the generated HR image as the patterns are replicated in a

494   check-like format and may result in a non-realistic HR image, thereby decreasing the

495   performance of the SR method.

496

497   **5.1.3. Meta Upscaling**

498   The scaling factor was predefined in the previously mentioned methods, thereby training

499   multiple upsampling modules with different factors, which is often inefficient and is not the

500   actual requirement of an SR method. A meta upscaling module (Hu et al., 2019) was proposed;

501   this module uses arbitrary scaling factors to generate SR image-based in meta-learning. Meta

502   scaling module projects every position in the required HR image to a small patch in the given LR

503   feature maps $j \times j \times c_i$, where j is arbitrary, and ci is the total number of channels within the

504   extracted feature map (in (Hu et al., 2019) this was 64). Additionally, it also generates the

505   convolution weights $\left(j \times j \times \left(c_i \times c_o \right) \right)$, where c_o represents the output image channels, and it is

506 usually 3. Thus, the meta upscaling module continuously uses arbitrary scaling factors within a
507 single model and using a substantial training set, a large number of factors are simultaneously
508 trained. The performance of this layer even surpasses the results produced with fixed factor
509 models, and even though this module predicts the weights during the inference time, the overall
510 execution time for weight prediction is 100 times less than the total time required for feature
511 extraction (Hu et al., 2019). In cases where there is a need for larger magnifications, this module
512 may become unstable as it predicts the convolution weights for every pixel independent of the
513 image information within those pixels.
514 This upscaling method is frequently used in recent years, particularly in post-upsampling
515 frameworks (Section 5.4.2). The high-level representations extracted from the low-level
516 information are used to construct an HR image using meta upscaling in the last layer of the
517 model, making this method an end-to-end SR approach.
518 The comparison of upsampling methods is shown in Table 4; most SR methods use
519 deconvolution or sub-pixel layers for upscaling. However, for multiple scale factors, meta
520 upscaling is used.
521

5.2. Deep learning SR Networks

523 The network design and advancements in design architecture are recent trends in deep learning,
524 and in SR, researchers have tried several design implications along with the SR framework (as
525 seen in Section 5.4) for designing the overall SR network. Some of the fundamental and recent
526 network designs are discussed in this section.
527

5.2.1. Recursive Learning

529 One of the basic network-based learning strategies is to use the same module for recursively
530 learning high-level features. This method also minimizes the parameters as the strategy is based
531 on the same module being updated recursively, as shown in Fig. 6a.
532 One of the most used recursive networks is the Deeply-recursive Convolutional Network
533 (DRCN) (Kim, Lee & Lee, 2016b). Utilizing a single convolution layer DRCN reaches up to a
534 41×41 repetitive field without requiring additional parameters, which is very deep compared to
535 the Super-resolution Convolution Neural Network SRCNN (Thapa et al., 2016) (13×13). The

536 Deep Recursive Residual Network (DRRN) (Tai, Yang & Liu, 2017) utilized a ResBlock (He et
537 al., 2016) as part of the recursive module for a total of 25 recursions and was reported to achieve
538 better performance than the baseline ResBlock. Using the concept of DRCN, (Tai et al., 2017)
539 proposed a memory block-based method MemNet which contained six recursive ResBlocks.
540 Whereas the Cascading Residual Network (CARN) (Ahn, Kang & Sohn, 2018a) also used
541 ResBlocks as recursive units. In this approach, the network shares the weights globally in
542 recursion using an iterative up-and-down sampling-based approach. Apart from end-to-end
543 recursions, the researchers also used Dual-state Recurrent Network (DSRN) (Han et al., 2018),
544 which shared the signals between the LR and generated HR states within the network.

545 Overall, while reducing the parameters, recursive learning networks can learn the complex
546 representation of the data at the cost of computational performance. Additionally, the increase in
547 computational requirements may result in an exploding or vanishing gradient. Thus, recursive
548 learning is often used in combination with multi-supervision or residual learning for minimizing
549 the risk of exploding or vanishing gradient (Kim, Lee & Lee, 2016b; Tai et al., 2017; Tai, Yang
550 & Liu, 2017; Han et al., 2018).
551

5.2.2. Residual Learning

553 Residual learning was widely used in the field of SR (Bevilacqua et al., 2012), (Timofte, De
554 Smet & Van Gool, 2015), (Timofte, De & Gool, 2013), until ResNet (He et al., 2016) was
555 proposed for learning residuals, as shown in Fig. 6b. Overall, there are two approaches, local and
556 global residual learning.
557 The local residual learning approach mitigates the degradation problem (He et al., 2016) caused
558 by increased network depth. Furthermore, the local residual learning also improved the learning
559 rate and reduced the training difficulty; this is frequently used in the SR field (Protter et al.,
560 2009; Mao, Shen & Yang, 2016; Han et al., 2018; Li et al., 2018).
561 The global residual learning is an approach used in which the input and the final output are
562 correlated, and in image SR, the output HR is highly correlated with the input LR image; thus,
563 learning the global residuals between LR and HR image is significant in SR. In global residual
564 learning, the model only learns the residual map that transforms the LR image into an HR image
565 by generating the missing high-frequency details in the LR image. Furthermore, the residuals are
566 minimal; thereby, the learning difficulty and model complexity are significantly reduced in
567 global residual-based learning. This method is also frequently used in SR methods (Kim, Lee &
568 Lee, 2016a), (Tai, Yang & Liu, 2017), (Tai et al., 2017), (Hui, Wang & Gao, 2018).
569 Overall, both methods use residuals to connect the input image with the output HR image; in the
570 case of global residual learning, the connection is directly made, which in local residual learning
571 various layers of different depth to connect the input (using local residuals) with the output.
572

5.2.3. Dense Connection-based Learning

574 This learning method uses dense blocks to address SR, like DenseNet (Huang et al., 2017). The
575 dense block utilizes all the features maps generated by the previous layers as inputs and its
576 feature inputs, leading to $l(l-1)/2$ connections in an l-layer $(l \geq 2)$ dense block. Using dense
577 blocks will increase the reusability of the features while resolving the gradient vanishing
578 problem. Furthermore, the dense connections also minimize the model size by utilizing a small
579 growth rate and enfolding the channels using concatenated input features.
580 Dense connections are used in SR to connect the low-level and high-level features maps for
581 reconstructing a high-quality fine-detailed HR image, as shown in Fig. 6c. SRDenseNet (Tong et
582 al., 2017) proposed a 69-layer network containing dense connections within the dense blocks and
583 dense connections among the dense blocks. In SRDenseNet, the feature maps from the prior
584 blocks and the feature maps were used as inputs of all preceding blocks. RDN (Zhang et al.,

585 2018b), CARN (Ahn, Kang & Sohn, 2018a), MemNet (Tai et al., 2017), and ESRGAN (Wang et
586 al., 2019c) also used layer or block-level dense connection, while DBPN (Wang et al., 2018b)
587 only used the dense connection between the upsampling and downsampling units.
588
589 ### 5.2.4. Multi-path Learning
590 In multi-path learning, the features are transferred to multiple paths for different representations,
591 and these representations are later combined to gain improved performance. Scale-specific, local,
592 and global multi-path learnings are the main types.
593 For different scales, the super-resolution models use different feature extraction; in (Lim et al.,
594 2017), the authors proposed a single network-based multi-path learning for multiple scales. The
595 intermediate layers of the model were shared for feature extraction, while scale-specific paths,
596 including pre-processing and upsampling, were at the end of the models, i.e., the start and end of
597 the network. During training, the scale relative paths are enabled and updated accordingly, and
598 the proposed deep super-resolution MDSR method (Lim et al., 2017) also decreases the overall
599 model size because of the sharing of parameters across the scales. Like MDSR, a similar multi-
600 path-based approach is also implemented in ProSR and CARN.
601 Local multi-path learning is inspired using a new block, the inception module (Szegedy et al.,
602 2015), for multi-scale feature extraction, as performed in MSRN (Li et al., 2018) (shown in Fig.
603 6d). The additional block consists of 3×3 and 5×5 kernel size convolution layers, which
604 simultaneously extracts the features. After combining the outputs of the two convolution layers,
605 the final output goes through a 5×5 kernel convolution. Furthermore, a path links the input and
606 output by element-wise addition and uses this local multi-path learning; this method extracts
607 features efficiently than multi-scale learning.
608 Another variation of multi-path learning is global multi-path learning; in this method, various
609 features are extracted from multi-paths that can interact. In DSRN (Han et al., 2018), there are
610 two paths for extracting low and high-level information, and there is a continuous sharing of
611 features for improved learning. In contrast, in pixel recursive SR (Dahl, Norouzi & Shlens,
612 2017), a conditioning path is responsible for extracting global structures, and the prior path
613 further finds the serial codependence among the generated pixels. A different method was
614 employed by (Ren, El-Khamy & Lee, 2017), where multi-path learning was performed for
615 unbalanced structures, which were later combined in the final layer to get the SR output.

616 ### 5.2.5. Advanced Convolution-based Learning
617 In SR, the methods explored depend on the convolution operation, and various research studies
618 have attempted to modify the convolution operation for better performance. In recent years,
619 research studies have shown that group convolution, as shown in Fig. 6e, decreased the total
620 number of parameters at the cost of small loops in performance (Hui, Wang & Gao, 2018),
621 (Johnson, Alahi & Fei-Fei, 2016). In CARN-M (Ahn, Kang & Sohn, 2018a) and IDN (Hui,
622 Wang & Gao, 2018), group convolution was used instead of vanilla convolution. In dilated
623 convolution, the contextual information is used to generate realistic-looking SR images (Zhang

624 et al., 2017); dilated convolution was used to double the receptive field, resulting in better
625 results.
626 Another type of convolution is depthwise separable convolution (Howard et al., 2009); although
627 this convolution significantly reduces the total number of parameters, it reduces the overall
628 performance.
629

5.2.6. Attention-based Learning

631 In deep learning, attention learning is the idea where certain factors are given more preference,
632 which processes the data than others; here, two types of attention-based learning mechanisms are
633 discussed in SR. In channel attention, a particular block is added in the model where global
634 average pooling (GAP) squeezes the input channels; two fully connected layers process these
635 constants to generate channel-wise residuals (Hu, Shen & Sun, 2018), as shown in Fig. 6f. This
636 technique has been incorporated in SR, known as RCAN (Zhang et al., 2018a), which has
637 improved performance. Instead of GAP, (Dai et al., 2019) used the second-order channel
638 attention (SPCA) module, which used second-order feature metric for extracting more data
639 representation using channel-based attention
640 In SR, most of the models use local fields for the generation of SR pixels, while in a few cases,
641 some textures or patches which are far apart are necessary for generating accurate local patches.
642 In (Zhang et al., 2019b), local and non-local attention blocks were used to extract local and non-
643 local representations between pixel data. Similarly, the non-local attention technique was
644 incorporated by (Dai et al., 2019) to capture contextual information using a non-local attention
645 method. Chen et al. proposed an SR reconstruction method with feature maps to facilitate the
646 reconstruction of the image using an attention mechanism (Chen et al., 2021), while Yang et al.
647 proposed a channel attention and spatial graph convolutional network (CASGCN) for a more
648 robust feature obtaining and feature correlations modeling (Yang & Qi, 2021).
649

5.2.7. Wavelet transform-based Learning

651 Wavelet transform (WT) (Daubechies & Bates, 1993; Griffel & Daubechies, 1995) represents
652 textures using high-frequency sub-bands and global structural information in low-frequency sub-
653 bands in a highly efficient way. WT was used in SR to generate the residuals of the HR sub-
654 bands using the sub-bands of the interpolated LR wavelet. Using the WT, the LR image is
655 decomposed, while the inverse WT provides the reconstruction of the HR image in SR. Other
656 examples of WT based SR are Wavelet-based residual attention network (WRAN) (Xue et al.,
657 2020), multi-level wavelet CNN (MWCNN) (Liu et al., 2018b), and (Ma et al., 2019); these
658 approaches used a hybrid approach by combining WT with other learning methods to improve
659 the overall performance.
660

5.2.8. Region-recursive-based Learning

662 In SR, most methods follow the underlying assumption that it is a pixel-independent process;
663 thus, there is no priority to the interdependence among the generated pixels. Using the concept of

664 PixelCNN (Van Den Oord et al., 2016), (Dahl, Norouzi & Shlens, 2017) proposed a method for
665 pixel recursive learning, which performed SR by pixel-by-pixel generation using two networks.
666 The two networks (Dahl, Norouzi & Shlens, 2017) captured information about pixel dependence
667 and global contextual information within the pixel recursive SR method. Using the mean opinion
668 scoring-based evaluation method (Dahl, Norouzi & Shlens, 2017) performed well compared to
669 other methods for generating SR face images using the pixel recursive method. The attention-
670 based face hallucination method (Cao et al., 2017a) also utilized the concept of a path-based
671 attention shifting mechanism to enhance the details in the local patches.
672 While the region-recursive methods perform marginally better than other methods, the recursive
673 process exponentially increases the training difficulty and computation costs due to long
674 propagation paths.
675
676 **5.2.9. Other Methods**
677 Other SR networks are also used by researchers, such as Desubpixel-based learning (Vu et al.,
678 2019), xUnit-based learning (Kligvasser, Shaham & Michaeli, 2018) and Pyramid Pooling-based
679 learning (Zhao et al., 2017).
680 To improve the computational speed, the desubpixel-based approach was used to extract features
681 in a low-dimensional space, which does the inverse task of the sub-pixel layer. By segmenting
682 the images spatially and using them as separate channels, the desubpixel-based learning avoids
683 any information loss; after learning the data representations in low-dimensional space, the
684 images are upsampled to get a high-resolution image. This technique is particularly efficient in
685 applications with limited resources such as smartphones.
686 In xUnit learning, a spatial activation function was proposed for learning complicated features
687 and textures. In xUnit, the ReLU operation was replaced by xUnit to generate the weight maps
688 through Gaussian gating and convolution. The model size was decreased by 50% using xUnit at
689 the cost of increased computational demand without compromising the SR performance
690 (Kligvasser, Shaham & Michaeli, 2018).
691
692 **5.3. Learning Strategies**
693 Learning strategies also dictate the overall performance of any SR algorithm as the evaluations
694 are dependent upon the choice of the learning strategy selected. In this section, recent research
695 studies are discussed using the learning strategy utilized in SR, and some of the critical strategies
696 are discussed in detail.
697
698 **5.3.1. Loss Functions**
699 For any application in deep learning, the selection of the loss functions is critical, and in SR,
700 these functions are used to measure the error in the reconstruction of HR, which further helps
701 optimize the model iteratively. Since the necessary element of the images is a pixel, initial
702 research studies employed the pixel loss, L2, but it was evaluated that the pixel loss cannot
703 wholly represent the quality of reconstruction (Ghodrati et al., 2019). Thus, in SR, different loss

704 functions such as content loss (Johnson, Alahi & Fei-Fei, 2016) or adversarial loss (Ledig et al.,
705 2017) are used to measure the error in the generation these loss functions have been widely used
706 in the field of SR. Various loss functions are explored in this section, and the notation follows
707 the previously defined variables except where defined otherwise.
708 Content Loss. The perceptual quality, as mentioned previously, is essential in the evaluation of
709 an SR model, and this loss was used in SR (Johnson, Alahi & Fei-Fei, 2016), (Dosovitskiy &
710 Brox, 2016) to measure the differences between the generated and ground-truth images using an
711 image classification network (N). Let the high-level data representation on the lth lth layer is
712 $r^l(I)$, the content loss is defined as the Euclidean among the high-level representations of the
713 two images I and $\hat{I}$, where I is the original image and $\hat{I}$ is the generated SR image as below:

714
$$L\left(I,\hat{I};N_c,l\right)=\frac{1}{h_l w_l c_l}\sqrt{\sum_{i,j,k}\left(r^l_{i,j,k}\left(\hat{I}\right)-r^l_{i,j,k}\left(I\right)\right)^2} \tag{15}$$

715 Where h_l, w_l, and c_l respectively are height, width, and several channels of the image
716 representations in the l layer.
717 Content loss aims to share information about image features from the image classification
718 network N_c to the SR network. This loss function ensures the visual similarity between the
719 original image (I) and the generated image ($\hat{I}$) by comparing the content and not the individual
720 pixels. Thus, this loss function helps in producing visually perceptible and more realistic looking
721 images in the field of SR as in (Ledig et al., 2017), (Wang et al., 2018b), (Sajjadi, Scholkopf &
722 Hirsch, 2017), (Wang et al., 2019c), (Johnson, Alahi & Fei-Fei, 2016), (Bulat & Tzimiropoulos,
723 2018) where the networks used as pre-trained CNNs were ResNet (He et al., 2016) and VGG
724 (Simonyan & Zisserman, 2015).
725 Adversarial Loss. In recent years, after the development of GANs (Goodfellow et al., 2014),
726 GANs have received more consideration due to their ability to learn and self-supervise. A GAN
727 combines dual networks performing generation and discrimination tasks, i.e., generating the
728 actual output and using a discriminator network to evaluate the results of the generative network.
729 While training the GANs, two continuous updates were performed, i.e. (i) Adjust the generator
730 for better results while training the discriminator to discriminate more efficiently and (ii) Adjust
731 the discriminator while training the generator. This is a recursive training network, and through
732 many iterations of training and evaluation, the generator can generate the output that conforms to
733 the distribution of the actual data. The discriminator is unable to differentiate between real and
734 generated information.
735 In terms of image SR, the purpose of a generative network is to generate an HR image, while
736 another discriminator network will be used to evaluate if the image is of the same distribution as
737 the input data. This method was first introduced in SR as SRGAN (Ledig et al., 2017), the
738 adversarial loss in (Ledig et al., 2017) was represented by:

739
$$L_{GAN\_CE\_g}\left(\hat{I};D\right)=-\log D\left(\hat{I}\right) \tag{16}$$

740
$$L_{GAN\_CE\_d}\left(\hat{I},I_s;D\right)=-\left\{\log D\left(I_s\right)+\log\left(1-D\left(\hat{I}\right)\right)\right\} \tag{17}$$

741 Where $L_{GAN\_CE\_g}$. is the adversarial loss function of the generator in the SR model, while

742 $L_{GAN\_CE\_d}$ is the adversarial loss function of the discriminator D, which is a binary classifier.

743 In (17), the randomly sampled ground truth image is denoted by I_s. The same loss functions

744 were reported by (Sajjadi, Scholkopf & Hirsch, 2017).

745 Other than binary classification error, the studies (Yuan et al., 2018) and (Wang et al., 2018a)

746 used mean square error for improved training and better results compared to (Ledig et al., 2017),

747 the loss functions are given in (18) and (19):

748
$$L_{GAN\_LS\_g}\left(\hat{I};D\right)=\left(D\left(\hat{I}\right)-1\right)^2 \tag{18}$$

749
$$L_{GAN\_LS\_d}\left(\hat{I},I_s;D\right)=\left\{\left(D\left(\hat{I}\right)\right)^2+\left(D\left(I_s\right)-1\right)^2\right\} \tag{19}$$

750 Contrary to the loss functions mentioned in (18) and (19), (Park et al., 2018) showed that in some

751 cases, pixel-level discriminator network generates high-frequency noise; thus, we used another

752 discriminator network to evaluate the first discriminator network for high-frequency

753 representations. Using the two discriminator networks, (Park et al., 2018) were able to capture all

754 attributes accurately.

755 Various opinion scoring systems have been used regressively to test the performance of the SR

756 model that uses adversarial loss. Although the SR models attained lower PSNR than the pixel-

757 loss-based SR on perceptual quality metrics like opinion scoring, these adversarial loss-based SR

758 methods scored very high (Ledig et al., 2017), (Sajjadi, Scholkopf & Hirsch, 2017). The use of a

759 discriminator as the control network for the generator GANs was able to regenerate some

760 intricate patterns that were very difficult to learn using ordinary deep learning methods. The only

761 drawback of the GANs is their training stability (Arjovsky, Chintala & Bottou, 2017; Gulrajani

762 et al., 2017; Lee et al., 2018a; Miyato et al., 2018).

763 Pixel Loss. As evident from the name, this loss function performs a pixel-wise comparison

764 between the reference image and the generated image, and there are two types of comparisons,

765 i.e., an $L1$ loss, which is also termed as mean absolute error and $L2$ loss, which is the mean

766 square error (MSE)

767
$$L_{PIX\_L1}\left(I,\hat{I}\right)=\frac{1}{hwc}\sum_{i,j,k}\left|I_{i,j,k}-\hat{I}_{i,j,k}\right| \tag{20}$$

768
$$L_{PIX\_L2}\left(I,\hat{I}\right)=\frac{1}{hwc}\sum_{i,j,k}\left|I_{i,j,k}-\hat{I}_{i,j,k}\right|^2 \tag{21}$$

769 The $L1$ loss in some cases becomes numerically unstable to compute; thus, another variant of the

770 L1 loss called the Charbonnier loss (Farsiu et al., 2004b; Barron, 2017, 2019; Lai et al., 2017) is

771 given by:

772
$$\mathsf{L}_{PIX\_CH}\left(I,\hat{I}\right)=\frac{1}{hwc}\sum_{i,j,k}\sqrt{\left|I_{i,j,k}-\hat{I}_{i,j,k}\right|^2-e^2} \tag{22}$$

773   Here e is a constant which ensures numerical stability.

774   The pixel loss function ensures that the generated HR image $\hat{I}$ has the same pixel values as the
775   HR image I. Furthermore, the L2 loss used the square of pixel-value errors, giving more
776   weightage to high-value differences than lower ones; thus, this loss function may give either the
777   too variable result (in case of outliers) or give too smooth results (in case of minimal error
778   values). Therefore, the L1 loss function is widely used over L2 loss (Zhao et al., 2016; Lim et al.,
779   2017; Ahn, Kang & Sohn, 2018a). Furthermore, the PSNR equation is closely related to the
780   definition of L1 loss, and minimizing L1 loss always leads to increased PSNR. Thus, researchers
781   have often used the L1 loss to maximize the PSNR; as mentioned earlier, the pixel loss function
782   does not cater to perceptual quality or textures. Thus, SR networks based on this loss function
783   may have less high-frequency details, resulting in smooth but unrealistic HR images (Wang,
784   Simoncelli & Bovik, 2003; Wang et al., 2004).
785   Style Reconstruction Loss. Ideally, the reconstructed HR image should have comparable styles to
786   the actual HR image (colors, textures, gradient, contrast), thus using the research studies (Sajjadi,
787   Scholkopf & Hirsch, 2017), (Gatys, Ecker & Bethge, 2015), style reconstruction loss was used in
788   SR to match the texture details of the reference image with the generated image. The correlation
789   between the feature maps of different channels as given by the Gram matrix (Levy & Goldberg,
790   2014) $G^{(l)}$. $G_{i,j}^{(l)}$ is the dot product of the features i, and j in the layer l, it is which is given by:

791
$$G_{i,j}^{(l)}=vec\left(ch_i^{(l)}(I)\right).vec\left(ch_j^{(l)}(I)\right) \tag{23}$$

792   Where $vec(\ )$ is the vectorization operation and $ch_i^{(l)}$ denoted the ith channel of feature maps
793   in the layer l. Now the texture loss is given by (24)

794
$$\mathsf{L}_{TEX}\left(I,\hat{I};ch,l\right)=\frac{1}{c_l^2}\sqrt{\sum_{i,j}\left(G_{i,j}^{(l)}(I)-G_{i,j}^{(l)}(\hat{I})\right)^2} \tag{24}$$

795   Using the texture loss function in (24), EnhanceNet (Sajjadi, Scholkopf & Hirsch, 2017) reported
796   more realistic results that look visually similar to the reference HR image. Although an
797   optimized texture loss function-based SR generates more realistic-looking images, the selection
798   of patch size is still an open field of research. The selection of small patch size leads to the
799   generation of artifacts in the textured region, while selecting a big patch size generates artifacts
800   across the whole image as the patches are averages over the whole image.
801   Total Variation Loss. Using the pixel values of the neighboring pixels, the total variation loss
802   (Rudin, Osher & Fatemi, 1992) was defined as the sum of the absolute difference among the
803   values of the neighboring pixels as:

804
$$\mathsf{L}_{TV}\left(\hat{I}\right)=\frac{1}{hwc}\sum_{i,j,k}\sqrt{\left(\hat{I}_{i+1,j,k}-\hat{I}_{i,j,k}\right)^2-\left(\hat{I}_{i,j+1,k}-\hat{I}_{i,j,k}\right)^2} \tag{25}$$

805   Total variation loss was used in (Ledig et al., 2017; Yuan et al., 2018) to ensure smoothness

806   across sharp edges/transitions within the generated image.

807   Cycle Consistency Loss. Using the CycleGAN (Zhu et al., 2017a) image SR method was

808   presented in (Yuan et al., 2018) using the cyclic consistency loss function. Using the generated

809   HR image $\hat{I}$, the network generated another LR image I'_{LR}, which is further compared with the

810   input LR image I_{LR} for cyclic consistency.

811   In practice, various loss functions are used as a combination in SR to ensure various aspects of

812   the generation process in the form of a weighted average as in (Kim, Lee & Lee, 2016a), (Wang

813   et al., 2018b), (Sajjadi, Scholkopf & Hirsch, 2017), (Lai et al., 2017). The selection of

814   appropriate weights of the loss functions in itself is another learning problem as the results vary

815   significantly by varying the weights of the loss function in image SR.

816

5.3.2. Curriculum Learning

818   In the Curriculum learning technique (Bengio et al., 2009), the method adapts itself to the

819   variable difficulty of tasks, i.e., starting from simple images with minimum noise to complex

820   images. Since SR always suffers from adverse conditions, the curriculum approach is mainly

821   applied to its learning difficulty and network size. For reducing the training difficulty of the

822   network in SR, small scaling factor, SR is performed in the beginning; in the curriculum

823   learning-based SR, the training starts with 2x upsampling, and gradually the following scaling

824   factors $4\times$, $8\times$, and so on are generated using the output of previously trained networks. ProSR

825   (Wang et al., 2018a) uses the upsampled output of the previous level and linearly trains the next

826   level using the previous one, while ADRSR (Bei et al., 2018) concatenates the HR output of the

827   previous levels and further adds another convolution layer. In CARN (Ahn, Kang & Sohn,

828   2018b), the previously generated image is entirely replaced by the next level generated image,

829   updating the HR image in sequential order.

830   Another alternative is to transform the image SR problem into N subsets and gradually solving

831   these problems; as in (Park, Kim & Chun, 2018), the $8\times$ upsampling problem was divided into

832   three problems (i.e., $1\times to\, 2\times$; $2\times to\, 4\times$ and $4\times to\, 8\times$) and three separate networks were used to

833   solve these problems. Using a combination of the previous reconstruction, the next level was

834   finetuned in this method. The same concept was used in (Li et al., 2019b) to train the network

835   from low image degradations to high image degradations, thus gradually increasing the noise in

836   the LR input image. Curriculum learning reduces the training difficulty; hence, the total

837   computational time is also reduced.

838

5.3.3. Batch Normalization

840   Batch normalization (BN) was proposed by (Ioffe & Szegedy, 2015) to stabilize and accelerate

841   the deep CNNs by reducing the internal covariate shift of the network. Every mini-batch was

842   normalized, and two additional parameters were used per channel to preserve the representation

843   ability. Batch normalization is responsible for working on the intermediate feature maps; thus, it

844   resolves the vanishing gradient issue while allowing high learning rates. This technique is widely

845   used in SR models such as (Ledig et al., 2017), (Zhang, Zuo & Zhang, 2018), (Tai, Yang & Liu,
846   2017), (Tai et al., 2017), (Ledig et al., 2017; Sønderby et al., 2017; Tai et al., 2017; Tai, Yang &
847   Liu, 2017; Liu et al., 2018b; Zhang, Zuo & Zhang, 2018). In contrast, (Lim et al., 2017) claimed
848   that batch normalization-based networks lose the scale information of the generated images.
849   Thus, there is a lack of flexibility in the network; hence, (Lim et al., 2017) removed batch
850   normalization and used the additional memory to design a large model with superior
851   performance compared to the BN-based network. Other studies (Wang et al., 2019c), (Wang et
852   al., 2018a), (Chen et al., 2018a) also implemented this technique to achieve marginally better
853   performance.
854

5.3.4. Multi-supervision

856   Using numerous supervision signals within the same model for improving the gradient
857   propagation and evading the exploding/vanishing gradient problem is called multi-supervision.
858   In (Kim, Lee & Lee, 2016b), multi-supervision is incorporated within the recursive units to
859   address the gradient problems. In SR, the multi-supervision learning technique is implemented
860   by catering to a few other factors in the loss function, which improves the back-propagation path
861   and reduces the training difficulty of the model.
862

5.4. SR Frameworks

864   SR being an ill-posed problem; thus, upsampling is critical in defining the performance of the SR
865   method. Based on learning strategies, upsampling methods, and network types, there are several
866   frameworks for SR; here, four of them are discussed in detail, especially in light of the
867   upsampling method used within the framework, as shown in Fig. 7 – 10.
868

5.4.1. Pre-upsampling SR

870   Learning the mapping functions for upsampling from an LR image directly to an HR image is
871   done using this framework, where the LR image is upsampled in the beginning, and various
872   convolution layers are used to extract representations in an iterative way using deep neural
873   networks. Using this concept (Dong et al., 2014), (Dong et al., 2016) introduced the pre-
874   upsampling-based SR framework (SRCNN), as shown in Fig. 7. SRCNN was used to learn the
875   end-to-end mapping of LR-HR image conversion using CNNs. Using the classical methods of
876   upsampling as discussed in Section 4, the LR image is firstly converted to an HR image, and
877   then deep CNNs were used to learn the representations for mapping the HR image.
878   Since the pre-upsampling layer already performs the actual pixel conversion task, the network
879   needs to refine the results using CNNs; this results in reduced learning difficulty. Compared to
880   single-scale SR (Kim, Lee & Lee, 2016a), which uses specific scales of input, these models can
881   handle any random size image for refinement and have similar performance. In recent years,
882   many application-oriented research studies have used this framework (Kim, Lee & Lee, 2016b),
883   (Shocher, Cohen & Irani, 2018), (Tai, Yang & Liu, 2017), (Tai et al., 2017), the differences in
884   these models are in the deep learning layers employed after the upsampling. The only drawback

885   in this model is the use of a predefined classical method of pre-upsampling, which often results
886   in the introduction of image blur, noise amplification in the upsampled image, which later affects
887   the quality of the concluding HR image. Moreover, the dimensions of the image are increased at
888   the start of the method. Thus, the computational cost and memory requirements of this
889   framework are higher (Shi et al., 2016).
890

5.4.2. Post-upsampling SR

892   To minimize the memory requirements and increase computational efficiency, the post-
893   upsampling method was used in SR to utilize deep learning to learn the mapping functions in
894   low-dimensional space. This concept was first used in SR by (Shi et al., 2016) and (Dong, Loy &
895   Tang, 2016), and the network diagram is shown in Fig. 8.
896   Due to low computational costs and the use of low-dimensional space for deep learning, this
897   model has been widely used in SR because this reduces the complexity of the model (Ledig et
898   al., 2017; Lim et al., 2017; Tong et al., 2017; Han et al., 2018).
899

5.4.3. Iterative Up-and-down Sampling SR

901   Since the LR-HR mapping is an ill-posed problem, efficient learning using the LR-HR image
902   pair using back-propagation (Irani & Peleg, 1991) was used in SR (Timofte, Rothe & Gool,
903   2016). The SR network is called the iterative up-down sampling SR, as shown in Fig. 9. This
904   model refines the image using recursive back-propagation, i.e., continuously measuring the error
905   and refining the model based on the reconstruction error. The DBPN method proposed in (Haris,
906   Shakhnarovich & Ukita, 2018) used this concept to perform continuous upsampling and
907   downsampling, and the final image was constructed using the intermediate generations of the HR
908   image.
909   Similarly, SRFBN (Li et al., 2019b) used this technique with densely connected layers for image
910   SR, while RBPN (Haris, Shakhnarovich & Ukita, 2019) used recurrent back-propagation with
911   iterative up-down upsampling for video SR. This framework has shown significant improvement
912   over the other frameworks; still, the back-propagation modules and their appropriate use require
913   further exploration as this concept is recently introduced.
914

5.4.4. Progressive-upsampling SR

916   Since the post-upsampling framework uses a single layer at the end of upsampling and the
917   learning is fixed for scaling factors; thus, multi-scale SR will increase the computational cost of
918   the post-upsampling framework. Thus, using progressive upsampling within the framework to
919   gradually achieve the required scaling was proposed, as seen in Fig. 10. An example of this
920   framework is the LapSRN (Lai et al., 2017), which uses cascaded CNN-based modules
921   responsible for mapping a single scaling factor, and the output of one module acts as the input
922   LR image to the other module. This framework was also used in ProSR (Wang et al., 2018a) and
923   MS-LapSRN (Lai et al., 2017).

924 This model achieves higher learning rates as the learning difficulty is less since the SR operation
925 is segregated into several small upscaling tasks, which is more straightforward for CNNs to
926 learn. Furthermore, this model has built-in support for multi-scale SR as the images are scaled
927 with various intermediate scaling factors. Training stability and convergence are the main issues
928 with this framework, and this requires further research.
929

930 **5.5. Other Improvements**
931 Apart from the four primary considerations in image SR, other factors have a significant effect
932 on the performance of a super-resolution method, and in this section, a few are discussed in light
933 of recent research.
934

935 *5.5.1. Data Augmentation*
936 Data augmentation is a common technique in deep learning, and this concept is used to further
937 enhance the performance of a deep learning model by generating more training data using the
938 same dataset. In the case of image super-resolution, some of the augmentation techniques are
939 flipping, cropping, angular rotation, skew, and color degradation (Timofte, Rothe & Gool, 2016;
940 Lai et al., 2017; Lim et al., 2017; Tai, Yang & Liu, 2017; Han et al., 2018). Recoloring the image
941 using channel shuffling in the LR-HR image pair is also used as data augmentation in image SR
942 (Bei et al., 2018).
943

944 *5.5.2. Enhanced Prediction*
945 This data augmentation method affects the output HR image as multiple LR images are
946 augmented using rotation and flipping functions (Timofte, Rothe & Gool, 2016). These
947 augmented are fed to the model for reconstruction, the reconstructed outputs are inversely
948 transformed, and the final HR image is based on the mean (Timofte, Rothe & Gool, 2016),
949 (Wang et al., 2018a) or median (Shocher, Cohen & Irani, 2018) pixel values of the
950 corresponding augmented outputs.
951

952 *5.5.3. Network Fusion and Interpolation*
953 This technique used multiple models to predict the HR image, and each prediction acts as the
954 input to the following network, like in context-wise network fusion (CNF) (Ren, El-Khamy &
955 Lee, 2017). The CNF was based on three individual SRCNNs, and this model achieved the
956 performance, which was compared with the state-of-the-art SR models (Ren, El-Khamy & Lee,
957 2017).
958 In the SR network, network interpolation is a model that uses PSNR-based and GAN-based
959 models for image SR to boost SR performance. Network interpolation strategy (Wang et al.,
960 2019c), (Wang et al., 2019b) used a PSNR-based model for training. In contrast, a GAN-based
961 model was used for fine-tuning while the parameters were interpolated to get the weights of
962 interpolation, and their results had few artifacts and look realistic.
963

5.5.4. Multi-task Learning

Multi-task learning is used for learning various problems and getting a generalized model for representations found in learning, for example, image segmentation, object detection, and facial recognition (Caruana, 1997; Collobert & Weston, 2008). In the field of super-resolution, (Wang et al., 2018b) used semantic maps as input to the model and predicted the parameters of the affine transformation on the transitional feature maps. The SFT-GAN in (Wang et al., 2018b) generated more realistic and crisp-looking images with good visual details regarding the textured regions. While in DNSR (Bei et al., 2018), a denoising network was proposed to denoise the output generated by the SR network; thus, using this closed-loop system (Bei et al., 2018) was able to achieve good results. Like DNSR, (Yuan et al., 2018) proposed an unsupervised SR using the cycle-in-cycle GAN (CinCGAN) for denoising during the SR task. Using a multi-tasking framework may increase the computational difficulty, but the system's performance is also enhanced in terms of PSNR and perceptual quality indexes.

5.6. State-of-the-art SR Methods

The recent year has excelled in developing SR models, especially using supervised deep learning; thus, the models have excelled in achieving state-of-the-art performance. Previously various aspects of the SR models and their underlying components were discussed in light of their strengths and weaknesses. In recent times the use of multiple learning strategies is common, and most of the state-of-the-art methods have used a combination of these strategies.

The first innovation was using the dual-branched network (DBCN) (Gao, Zhang & Mou, 2019) to increase the computational efficiency of the single-branched network by using a smaller number of convolutional layers for representation. Furthermore, in RCAN (Zhang et al., 2018a), attention-based learning was used combined with residual learning, L1 pixel loss function, and subpixel upsampling method to achieve the state-of-the-art results in image SR. Furthermore, various models and their reported results and some key factors are summarized in Table 5.

In previous sections, we discussed various strategies and compared and contrasted them; while these are important, the performance of any SR algorithm in comparison to the computational cost and parameters is also vital. In Fig. 11, we have graphically shown the performance of SR methods using PSNR metrics compared to their size (represented as several parameters) and computational cost (measured by the number of Multi-Adds). The datasets used in measurements are Set14, B100, and Urban 100; the overall PSNR is the average score over the three datasets, while the scaling factor for these models was fixed to $2\times$.

As evident in Fig. 11, the five best-performing methods on the selected datasets based on PSNR are WRAN (34.790dB), RCAN (34.540dB), SAN (34.480dB), Meta-RDN (34.400dB). and EDSR (34.330dB). The variation in the average PSNR reported by these methods only varies in the range of 0.46dB. There is a significant difference in the number of parameters reported by these five methods; WRAN reported only 2.710 million parameters while EDSR reported the highest parameters among the five methods, i.e., 40.74 million. WRAN and RCAN performed

1003   well in terms of PSNR, the number of parameters, and computational cost, thereby making them
1004   one of the best methods for image super-resolution.
1005

6. Unsupervised Super-resolution

1007   In this section, the methods of unsupervised SR are discussed, which does not require LR-HR
1008   pairs. The limitation of the supervised learning methods is that the LR images are usually
1009   generated using known degradations. In supervised learning, the model learns the reverse
1010   transformation function of the degradation function to convert the LR image into the HR image.
1011   Thus, using the unsupervised model to upsample the LR images is a field of growing interest,
1012   where the model learns the real-world image degradation to achieve SR using the information of
1013   unpaired LR and HR images. A few of the unsupervised SR models are discussed in the sub-
1014   sections.
1015

6.1. Weakly-supervised Super-resolution

1017   The first method to address the use of known degradation in the model for the generation of LR
1018   images using weakly supervised deep learning, this method utilized the unpaired LR and HR
1019   images for training the model. Although this model still requires both LR and HR images, the
1020   associations are not defined. Thus, there are two possible approaches; the first one is to learn the
1021   degradation function first, which can generate the degraded LR images and train the model to
1022   generate the HR images. The other method is to employ degradation function learning and LR-
1023   HR mapping cyclically, thus validating the results with each other (Ignatov et al., 2018a).
1024

6.1.1. Cyclic weakly-supervised SR

1026   Using the unpaired LR and HR images and referring them to two separated uncorrelated datasets,
1027   this method uses a cycle-in-cycle approach to predict the mapping function of these two datasets,
1028   i.e., from LR to HR and HR to LR images. This is a recursive process where the mapping
1029   functions generate images with equal distribution, and these images are fed to the second
1030   prediction cyclically.
1031   Using the deep learning-based CycleGAN (Zhu et al., 2017a), a cycle-in-cycle SR framework
1032   was proposed in (Yuan et al., 2018); this framework used a total of four generators, while there
1033   were two discriminators; the two GANS learned the representation of degraded LR to LR and
1034   LR to HR mappings. In (Yuan et al., 2018), the first generator is a simple denoising element that
1035   generates similar scale denoised LR images; these denoised images act as input to the second
1036   generator to regenerate the HR image, which is further validated by the adversarial network, i.e.,
1037   a discriminator. Thus, using different loss functions, the CycleGAN achieves image SR using
1038   weakly supervised learning.
1039   Although this method has achieved comparable results, especially in very noisy images where
1040   classical degradation functions in supervised learning cannot be used, there is room for research
1041   to decrease the learning difficulty of the computational cost of this method.
1042

1043

1044

1045   ***6.1.2. Learning the degradation function***

1046   A similar concept to the cyclic SR, but the two networks, i.e., a degradation learning network

1047   and LR-HR mapping network, are independently trained. In (Bulat, Yang & Tzimiropoulos,

1048   2018), a two-staged method of image SR was proposed, where a GAN learns the representations

1049   of the HR to LR transformation while the second GAN is trained using the paired output of the

1050   first GAN to learn the mapping representations of the LR to HR transformation. This two-stage

1051   model outperformed the state-of-the-art in Fréchet Inception Distance (FID) (Heusel et al., 2017)

1052   with 10% failure cases. This method reported superior reconstruction of HR human facial

1053   features.

1054

1055   **6.2. Zero-shot Super-resolution**

1056   Using the training concept at the time of the test, the zero-shot SR (ZSSR) (Shocher, Cohen &

1057   Irani, 2018) uses a single image to train a deep learning network using image augmentation

1058   techniques to learn the degradation function. ZSSR was used (Michaeli & Irani, 2013) to predict

1059   the degradation kernel, which was further used to generate scaled and augmented images. The

1060   final step was to train an SRCNN network to learn the representations of this dataset, and in this

1061   way, the ZSSR uses augmentation and input image data to achieve SR. This model outdid the

1062   state-of-the-art for non-bicubic, noisy, and blurred LR images by 1dB in the case of estimated

1063   kernels and 2dB for known kernels.

1064   Since this model requires training for every input image at the test time, the overall inference

1065   time is substantial.

1066

1067   **6.3. Image Prior in SR**

1068   The low-level details in any learning problem can be mapped using CNNs, thus using a

1069   randomly initialized CNN as an image prior (Ulyanov, Vedaldi & Lempitsky, 2020) to perform

1070   SR. The network is not trained; instead, it uses a random vector v as input to the model, and it

1071   generates the HR image I_{yHR}. This method aims to determine an image $\hat{I}_{yHR}$ that, when

1072   downsampled, returns an LR image that is similar to the input LR image $\hat{I}_{xHR}$. The model

1073   performed 2dB the state-of-the-art methods but reported superior results than the conventional

1074   bicubic upsampling method by 1dB.

1075

1076   **7. Domain-Specific Applications of Super-Resolution**

1077   In this section, various applications of SR grouped by the application domains are discussed.

1078

1079   **7.1. Face Image Super-resolution**

1080   Face hallucination (FH) is perhaps the ultimate target utility of the image SR for face-

1081   recognition-based tasks such as (Gunturk et al., 2003), (Taigman et al., 2014; Korshunova et al.,

1082   2017; Zhang et al., 2018c; Grm, Scheirer & Štruc, 2020). The facial images contain facial-
1083   structured information; thus, using image priors in FH has been a common approach to achieve
1084   FH.
1085   Using techniques such as in CBN (Zhu et al., 2016), the generated HR images can be constrained
1086   to face-related features, forcing the model to output HR images containing facial features. In
1087   CBN, this was achieved by using a facial prior and a dense correspondence field estimation.
1088   While in FSRNet (Chen et al., 2018b) facial parsing maps and facial landmark heatmaps were
1089   used as priors to the learning network to achieve face image SR, SICNN (Zhang et al., 2018c)
1090   used a joint training approach to recover the real identity using a super-identity loss function.
1091   Super-FAN (Bulat & Tzimiropoulos, 2018) approached FH using end-to-end learning with FAN
1092   to ensure the generated images are consistent with human facial features.
1093   Using implicit methods for solving the face misalignment problem is another way to approach
1094   FH; for instance, in (Yu & Porikli, 2017), the spatial transformation is achieved using
1095   transformation networks (Jaderberg et al., 2015). Another method based on (Jaderberg et al.,
1096   2015) is TDAE (Wang et al., 2019b), which uses a three-module approach for FH, using a D-E-
1097   D (decoder-encoder-decoder) model to achieve FH; the first decoder performs denoising and
1098   upsampling while encoder downsamples the denoised image which is fed to the final decoder for
1099   FH. Another approach is to use HR exemplars from datasets to decompose the facial features of
1100   an LR image and project the HR features into the exemplar dataset to achieve FH (Yang, Liu &
1101   Yang, 2018). In (Song et al., 2018), an adversarial discriminative network was proposed for
1102   feature learning on both feature space and raw pixel space; this method performed well for
1103   heterogenous face recognition (HFR).
1104   In other research studies, human perception of attention shifting (Najemnik & Geisler, 2005) was
1105   used in Attention-FH (Cao et al., 2017a) to learn face patches for local enhancement of FH. In
1106   (Xu et al., 2017), a multi-class GAN network was proposed for FH, which composed of multiple
1107   generators and discriminators, while in (Yu & Porikli, 2016), the authors adopted a network
1108   model analogous to SRGAN (Ledig et al., 2017). Using conditional GAN (Gauthier, 2014), the
1109   studies (Lee et al., 2018b; Yu et al., 2018) used additional facial features to achieve FH with
1110   predefined attributes. Gao et al. proposed an efficient multilayer locality-constrained matrix
1111   regression (MLCMR) framework for face super-resolution of highly degraded LR images (Gao
1112   et al., 2021).
1113
1114   **7.2. Real-world Image Super-resolution**
1115   In real-world images, the sensors used to capture them already introduce degradations as the
1116   final RGB (8-bit) image is converted from the raw image (usually more than 14-bit or higher).
1117   Thus, using these images as a reference for SR is not optimal as the images have already been
1118   degraded (Wang, Chen & Hoi, 2020). To approach this problem, research studies such as (Zhang
1119   et al., 2019a) and (Chen et al., 2019) have proposed methods for developing real-world image
1120   datasets. In (Zhang et al., 2019a), the SR-RAW dataset was developed by the authors, which
1121   contained raw-HR-LR(RGB) pairs generated using the optical zoom in cameras, while in (Chen

1122   et al., 2019), image resolution and its relationship with the field of view (FoV) were explored by
1123   the authors to generate a real-world dataset called City100.
1124
1125   **7.3. Depth Map Super-resolution**
1126   In the field of computer vision, problems like image segmentation (Zaitoun & Aqel, 2015; Yu &
1127   Koltun, 2016; Kirillov et al., 2019)and pose estimation (Wei et al., 2016; Cao et al., 2017b; Chen
1128   & Ramanan, 2017) have been approached by using depth maps. Depth maps retain the distance
1129   information of the scene and the observer, although these depth maps are of low-resolution
1130   because of the hardware constraints of the modern camera systems. Thus, image SR is used in
1131   this regard to increase the resolution of the depth maps.
1132   Using multiple cameras to record the same scene and generate multiple HR images is the most
1133   suitable way of doing depth map SR. In (Hui, Loy & Tang, 2016), the authors used two separate
1134   CNNs to downsample HR image concurrently and upsample the LR depth map; after the
1135   generation of RGB features from the downsampling CNN, these features were used to fine-tune
1136   the upsampling process of depth maps. While (Riegler, Rüther & Bischof, 2016) used the energy
1137   minimization model (such as (Bashir & Ghouri, 2014)) to guide the model for generating HR
1138   depth maps without the need for reference images.
1139
1140   **7.4. Remote Sensing and Satellite Imaging**
1141   The use of SR in improving the resolution of remote sensing and satellite imaging has increased
1142   in the past years (Shermeyer & Van Etten, 2019). In (Li et al., 2017b), the authors used the
1143   concept of multi-line cameras to utilize multiple LR images to generate a high-quality HR image
1144   from the ZY-3 (TLC) satellite image dataset. In (Zhu et al., 2017b; Benecki et al., 2018), the
1145   authors argued that the conventional methods of evaluation of the SR techniques are not valid for
1146   satellite imaging as the degradation functions and operation conditions of the satellite hardware
1147   are entirely in a different environment and thus (Benecki et al., 2018) proposed a new way for
1148   validation of SR methods for satellite image SR methods. An adaptive multi-scale detail
1149   enhancement (AMDE-SR) was proposed in (Zhu et al., 2018) to use the multi-scale SR method
1150   to generate high-detailed HR images with accurate textual and high-frequency information.
1151   GAN-based methods provide superior performance for remote sensing image SR; Liu et al.
1152   developed a novel cascaded conditional Wasserstein generative adversarial network (CCWGAN)
1153   to generate HR images for remote sensing (Liu et al., 2020a). Bashir et al. proposed a YOLOv3-
1154   based small-object detection framework SRCGAN-RFA-YOLO (Bashir & Wang, 2021b), where
1155   the authors used residual feature aggregation and cyclic GAN to improve the resolution of
1156   remote sensing images before performing object detection.
1157
1158   **7.5. Video Super-resolution**
1159   In video SR, multiple frames represent the same scene; thus, there is inter and intra-frame spatial
1160   dependency in the video, which includes the information of brightness, colors, and relative
1161   motion of objects. Using the optical flow-based method (Sun, Roth & Black, 2010; Liao et al.,

2015), Sun et al. and Liao et al. proposed a method to generate probable HR candidate images and ensemble these images using CNNs. Using the Druleas (Drulea & Nedevschi, 2011) algorithm, CVSRnet (Kappeler et al., 2016) addressed the effect of motion by using CNNs for the images in successive frames to generate HR images.

Apart from direct learning motion compensation, a trainable spatial transformer (Jaderberg et al., 2015) was used in VESPCN (Caballero et al., 2017) to motion compensation mapping using data from successive frames for end-to-end mapping. Using a sub-pixel layer-based module, (Tao et al., 2017) achieved super-resolution and motion compensation simultaneously.

Another approach is to use recurrent networks to indirectly grasp the spatial and temporal interdependency to address the motion compensation. In STCN (Guo & Chao, 2017), the authors used a bidirectional LSTM (Graves, Fernández & Schmidhuber, 2005) and deep CNNs to extract the temporal and spatial information from the video frames, while BRCN (Huang, Wang & Wang, 2015) utilized RNNs, CNNs, and conditional CNNs respectively for temporal, spatial and temporal-spatial interdependency mapping. Using 3D convolution filters of small size to replace the large-sized filter, FSTRN (Li et al., 2019a) achieves state-of-the-art performance using deep CNNs, sustaining a low computational cost. A novel spatio-temporal matching network (STMN) for video SR was proposed, which worked on the wavelet transform to minimize the dependence on motion estimations (Zhu et al., 2021).

7.6. SR for Medical Imaging

Other fields also used the concept of image super-resolution to achieve high-resolution images, such as in (Mahapatra, Bozorgtabar & Garnavi, 2019); the authors proposed the use of progressive GANs to enhance the image quality of magnetic resonance (MR) images. DeepResolve (Chaudhari et al., 2018) used image SR methods to generate thin-sliced knee MR images from the thick-sliced input images. Since the diffusion MRI has high image acquisition time and low resolution, Super-resolution Reconstruction Diffusion Tensor Imaging (SRR-DTI) reconstructed HR diffusion parameters from LR diffusion-weighted (DW) images (Van Steenkiste et al., 2016). (Hamaide et al., 2017) also used SRR-DTI to find the structural sex variances in the adult zebra finch brain.

Assisted diagnosis using super-resolution has been a recent trend; for instance, researchers used deep learning-based SR methods to assist the diagnosis of movement disorders like isolated dystonia (Bashir & Wang, 2021a).

7.7. Other Applications

Other applications of SR include object detection (Li et al., 2017a; Tan, Yan & Bare, 2018), stereo image SR (Duan & Xiao, 2019; Guo, Chen & Huang, 2019; Wang et al., 2019a), and super-resolution in optical microscopy (Qiao et al., 2021). Overall, SR plays a vital role in multi-disciplines, from medical science, computer vision to satellite imaging and remote sensing.

8. Discussion and Future Directions

1202   This paper gives an overall review of literature for image super-resolution, and the contribution
1203   of this paper is discussed in this section.
1204
1205   **8.1. Learning Strategies**
1206   Learning strategies in image SR are introduced in Section 5.3; while the learning strategies are
1207   well matured in image SR, there are research directions in the development of alternate loss
1208   functions and alternative of batch normalization
1209   There are various loss functions in SR, and the choice of SR depends upon the task, while it is
1210   still an open research area to find an optimal loss function that fits all SR frameworks. A
1211   combination of loss functions is currently used to optimize the learning process, and there are no
1212   standard criteria for the selection of loss function; thus, exploring various probable loss functions
1213   for super-resolution is a promising future direction.
1214   Batch normalization is a technique that performs well in computer vision tasks and reduces the
1215   overall runtime of the training, and enhances the performance; however, in SR batch
1216   normalization proved to be sub-optimal (Lim et al., 2017), (Wang et al., 2018a), (Chen et al.,
1217   2018a). In this regard, normalization techniques for super-resolution should be explored further.
1218
1219   **8.2. Network Design**
1220   Network design strategies require further exploration in SR as the network design inherently
1221   dictates the overall performance of any SR method. Some of the key research areas are
1222   highlighted in this section
1223   As discussed in Section 5.1, current upsampling methods have significant drawbacks for the
1224   deconvolution layer and may produce checkerboard artifacts. In contrast, the sub-pixel layer is
1225   susceptible to the non-uniform distribution of receptive fields; the meta-scale method has
1226   stability issues, while the interpolation-based methods lack end-to-end learning. Thus, further
1227   research is required to explore upsampling methods that can be generic to SR models and can be
1228   applied to LR images with any scaling factors.
1229   For human perception in SR, further research is required in attention-based SR, where the
1230   models may be trained to give more attention to some image features than others like the human
1231   visual system does.
1232   Using a combination of low and high-level representations simultaneously to accelerate the SR
1233   process is another field in network design for fast and accurate reconstruction of the HR image.
1234   Exploring network architectures that can be implemented in practical applications since current
1235   methods use deep neural networks, which increases the performance of the SR at the expense of
1236   higher computational cost; thus, research in the development of network architecture that is
1237   minimal and provides optimal performance is another promising research direction.
1238
1239   **8.3. Evaluation Metrics**
1240   The image quality metrics used in SR act as the benchmark score, while the two most commonly
1241   used metrics, PSNR and SSIM, help gauge the performance of SR, but these metrics introduce

1242 inherent issues in the generated image. Using PSNR as an evaluation metric usually introduces
1243 non-realistic smooth surfaces, while SSIM works with textures, structures, brightness, and
1244 contrast to imitate human perception. These metrics cannot completely grasp the perceptual
1245 quality of images (Ledig et al., 2017), (Sajjadi, Scholkopf & Hirsch, 2017). Opinion scoring is a
1246 metric that ensures perceptual quality, but this metric is impractical for implementing SR
1247 methods for large datasets; thus, a probable research direction is developing a universal quality
1248 metric for SR.
1249

8.4. Unsupervised Super-resolution

1250
1251 In the past two years, unsupervised SR methods have gained popularity, but still, the task of
1252 collecting various resolution scenes for a similar pose is difficult; thus, bicubic interpolation is
1253 used instead to generate an unpaired SR dataset. In actuality, the unsupervised SR methods learn
1254 the inverse mapping of this interpolation for the reconstruction of HR images, and the actual
1255 learning of SR is still an open research field using unsupervised learning methods.
1256

9. Conclusion

1257
1258 A detailed survey of classical SR and recent advances in SR with deep learning are explored in
1259 this survey paper. The central theme of this survey was to discuss deep learning-based SR
1260 techniques and the application of SR in various fields. Although image SR has achieved a lot in
1261 the last decade, some open problems are highlighted in Section 8. This survey is intended for the
1262 researchers in the field of SR and researchers from other fields to use image SR in their
1263 respective fields of interest.
1264

Acknowledgment

1265
1266 We show our gratitude to the authors of all referred research studies for sharing results,
1267 especially to the authors of (Kim, Lee & Lee, 2016b), (Ledig et al., 2017), (Dong, Loy & Tang,
1268 2016), (Lai et al., 2017), (Haris, Shakhnarovich & Ukita, 2018), (Hu et al., 2019), (Tai, Yang &
1269 Liu, 2017), (Tai et al., 2017), (Li et al., 2018), (Lim et al., 2017), (Zhang et al., 2018a), (Dai et
1270 al., 2019), (Xue et al., 2020), (Caballero et al., 2017).
1271

References

1272
1273 Agustsson E, Timofte R. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution:
1274   Dataset and Study. In: *IEEE Computer Society Conference on Computer Vision and Pattern
1275   Recognition Workshops*. 126–135. DOI: 10.1109/CVPRW.2017.150.
1276 Ahn N, Kang B, Sohn KA. 2018a. Fast, accurate, and lightweight super-resolution with
1277   cascading residual network. In: *Lecture Notes in Computer Science (including subseries
1278   Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 252–268.
1279   DOI: 10.1007/978-3-030-01249-6_16.
1280 Ahn N, Kang B, Sohn KA. 2018b. Image super-resolution via progressive cascading residual
1281   network. In: *IEEE Computer Society Conference on Computer Vision and Pattern
1282   Recognition Workshops*. 791–799. DOI: 10.1109/CVPRW.2018.00123.

Almohammad A, Ghinea G. 2010. Stego image quality and the reliability of PSNR. In: *2010 2nd International Conference on Image Processing Theory, Tools and Applications, IPTA 2010*. 215–220. DOI: 10.1109/IPTA.2010.5586786.

Arjovsky M, Chintala S, Bottou L. 2017. Wasserstein generative adversarial networks. In: *34th International Conference on Machine Learning, ICML 2017*.

Baker S, Kanade T. 2002. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24:1167–1183. DOI: 10.1109/TPAMI.2002.1033210.

Barron JT. 2017. *A More General Robust Loss Function*.

Barron JT. 2019. A general and adaptive robust loss function. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 4331–4339. DOI: 10.1109/CVPR.2019.00446.

Bashir SMA, Ghouri FAK. 2014. Perspective texture synthesis based on improved energy optimization. *PLoS ONE* 9. DOI: 10.1371/journal.pone.0110622.

Bashir SMA, Wang Y. 2021a. Deep learning for the assisted diagnosis of movement disorders, including isolated dystonia. *Frontiers in Neurology* 12:638266. DOI: 10.3389/fneur.2021.638266.

Bashir SMA, Wang Y. 2021b. Small Object Detection in Remote Sensing Images with Residual Feature Aggregation-Based Super-Resolution and Object Detector Network. *Remote Sensing* 13:1854. DOI: 10.3390/rs13091854.

Bates M, Huang B, Dempsey GT, Zhuang X. 2007. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* 317:1749–1753. DOI: 10.1126/science.1146598.

Bei Y, Damian A, Hu S, Menon S, Ravi N, Rudin C. 2018. New techniques for preserving global structure and denoising with low information loss in single-image super-resolution. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 874–881. DOI: 10.1109/CVPRW.2018.00132.

Benecki P, Kawulok M, Kostrzewa D, Skonieczny L. 2018. Evaluating super-resolution reconstruction of satellite images. *Acta Astronautica* 153:15–25. DOI: 10.1016/j.actaastro.2018.07.035.

Bengio Y, Louradour J, Collobert R, Weston J. 2009. Curriculum learning. In: *Proceedings of the 26th International Conference On Machine Learning, ICML 2009*. 41–48.

Bevilacqua M, Roumy A, Guillemot C, Morel MLA. 2012. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: *BMVC 2012 - Electronic Proceedings of the British Machine Vision Conference 2012*. DOI: 10.5244/C.26.135.

Blau Y, Mechrez R, Timofte R, Michaeli T, Zelnik-Manor L. 2018. The 2018 PIRM challenge on perceptual image super-resolution. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. DOI: 10.1007/978-3-030-11021-5_21.

Blau Y, Michaeli T. 2018. The Perception-Distortion Tradeoff. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 6228–6237. DOI: 10.1109/CVPR.2018.00652.

Borman S, Stevenson RL. 1998. Super-resolution from image sequences-A review. In: *Midwest Symposium on Circuits and Systems*. 374–378. DOI: 10.1109/MWSCAS.1998.759509.

Bosse S, Maniry D, Müller KR, Wiegand T, Samek W. 2018. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing* 27:206–219. DOI: 10.1109/TIP.2017.2760518.

1329   Bulat A, Tzimiropoulos G. 2018. Super-FAN: Integrated Facial Landmark Localization and
1330   Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses with GANs. In:
1331   *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*
1332   *Recognition*. 109–117. DOI: 10.1109/CVPR.2018.00019.
1333   Bulat A, Yang J, Tzimiropoulos G. 2018. To learn image super-resolution, use a GAN to learn
1334   how to do image degradation first. In: *Proceedings of the European conference on computer*
1335   *vision (ECCV)*. 185–200. DOI: 10.1007/978-3-030-01231-1_12.
1336   Caballero J, Ledig C, Aitken A, Acosta A, Totz J, Wang Z, Shi W. 2017. Real-time video super-
1337   resolution with spatio-temporal networks and motion compensation. In: *Proceedings - 30th*
1338   *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 4778–4787.
1339   DOI: 10.1109/CVPR.2017.304.
1340   Cai D, Chen K, Qian Y, Kämäräinen JK. 2019. Convolutional low-resolution fine-grained
1341   classification. *Pattern Recognition Letters* 119:166–171. DOI:
1342   10.1016/j.patrec.2017.10.020.
1343   Candès EJ, Fernandez-Granda C. 2014. Towards a mathematical theory of super-resolution.
1344   *Communications on Pure and Applied Mathematics* 67:906–956. DOI: 10.1002/cpa.21455.
1345   Cao Q, Lin L, Shi Y, Liang X, Li G. 2017a. Attention-aware face hallucination via deep
1346   reinforcement learning. In: *Proceedings - 30th IEEE Conference on Computer Vision and*
1347   *Pattern Recognition, CVPR 2017*. 690–698. DOI: 10.1109/CVPR.2017.180.
1348   Cao Z, Simon T, Wei SE, Sheikh Y. 2017b. Realtime multi-person 2D pose estimation using part
1349   affinity fields. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern*
1350   *Recognition, CVPR 2017*. 7291–7299. DOI: 10.1109/CVPR.2017.143.
1351   Caruana R. 1997. Multitask Learning. *Machine Learning* 28:41–75. DOI:
1352   10.1023/A:1007379606734.
1353   Chang H, Yeung DY, Xiong Y. 2004. Super-resolution through neighbor embedding. In:
1354   *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*
1355   *Recognition*. I–I. DOI: 10.1109/cvpr.2004.1315043.
1356   Chaudhari AS, Fang Z, Kogan F, Wood J, Stevens KJ, Gibbons EK, Lee JH, Gold GE,
1357   Hargreaves BA. 2018. Super-Resolution Musculoskeletal MRI Using Deep Learning.
1358   *Magnetic Resonance in Medicine* 80:2139–2154. DOI: 10.1002/mrm.27178.
1359   Chen Y, Liu L, Phonevilay V, Gu K, Xia R, Xie J, Zhang Q, Yang K. 2021. Image super-
1360   resolution reconstruction based on feature map attention mechanism. *Applied*
1361   *Intelligence*:1–14. DOI: 10.1007/s10489-020-02116-1.
1362   Chen R, Qu Y, Zeng K, Guo J, Li C, Xie Y. 2018a. Persistent memory residual network for
1363   single image super resolution. In: *IEEE Computer Society Conference on Computer Vision*
1364   *and Pattern Recognition Workshops*. 809–816. DOI: 10.1109/CVPRW.2018.00125.
1365   Chen CH, Ramanan D. 2017. 3D human pose estimation = 2D pose estimation + matching. In:
1366   *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*
1367   *2017*. 7035–7043. DOI: 10.1109/CVPR.2017.610.
1368   Chen Y, Tai Y, Liu X, Shen C, Yang J. 2018b. FSRNet: End-to-End Learning Face Super-
1369   Resolution with Facial Priors. In: *Proceedings of the IEEE Computer Society Conference on*
1370   *Computer Vision and Pattern Recognition*. DOI: 10.1109/CVPR.2018.00264.
1371   Chen C, Xiong Z, Tian X, Zha ZJ, Wu F. 2019. Camera lens super-resolution. In: *Proceedings of*
1372   *the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. DOI:
1373   10.1109/CVPR.2019.00175.
1374   Collobert R, Weston J. 2008. A unified architecture for natural language processing: Deep neural

networks with multitask learning. In: *Proceedings of the 25th International Conference on Machine Learning*. 160–167.

Dabov K, Foi A, Katkovnik V, Egiazarian K. 2006. Color image denoising via sparse 3D collaborative filtering with grouping constraint in luminance-chrominance space. In: *Proceedings - International Conference on Image Processing, ICIP*. I–313. DOI: 10.1109/ICIP.2007.4378954.

Dahl R, Norouzi M, Shlens J. 2017. Pixel Recursive Super Resolution. In: *Proceedings of the IEEE International Conference on Computer Vision*. 5439–5448. DOI: 10.1109/ICCV.2017.581.

Dai T, Cai J, Zhang Y, Xia ST, Zhang L. 2019. Second-order attention network for single image super-resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 11065–11074. DOI: 10.1109/CVPR.2019.01132.

Daubechies I, Bates BJ. 1993. Ten Lectures on Wavelets. *The Journal of the Acoustical Society of America*. DOI: 10.1121/1.406784.

Deng X. 2018. Enhancing Image Quality via Style Transfer for Single Image Super-Resolution. *IEEE Signal Processing Letters* 25:571–575. DOI: 10.1109/LSP.2018.2805809.

Deng J, Dong W, Socher R, Li L-J, Kai Li, Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition*. 248–255. DOI: 10.1109/cvpr.2009.5206848.

Dong C, Loy CC, He K, Tang X. 2014. Learning a deep convolutional network for image super-resolution. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. DOI: 10.1007/978-3-319-10593-2_13.

Dong C, Loy CC, He K, Tang X. 2016. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38:295–307. DOI: 10.1109/TPAMI.2015.2439281.

Dong C, Loy CC, Tang X. 2016. Accelerating the super-resolution convolutional neural network. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 391–407. DOI: 10.1007/978-3-319-46475-6_25.

Dosovitskiy A, Brox T. 2016. Generating images with perceptual similarity metrics based on deep networks. In: *Advances in Neural Information Processing Systems*. 658–666.

Drulea M, Nedevschi S. 2011. Total variation regularization of local-global optical flow. In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. 318–323. DOI: 10.1109/ITSC.2011.6082986.

Duan C, Xiao N. 2019. Parallax-Based Spatial and Channel Attention for Stereo Image Super-Resolution. *IEEE Access* 7:183672–183679. DOI: 10.1109/ACCESS.2019.2960561.

Duchon CE. 1979. Lanczos filtering in one and two dimensions. *Journal of applied meteorology* 18:1016–1022. DOI: 10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2.

Farsiu S, Robinson D, Elad M, Milanfar P. 2004a. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology* 14:47–57. DOI: 10.1002/ima.20007.

Farsiu S, Robinson MD, Elad M, Milanfar P. 2004b. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing* 13:1327–1344. DOI: 10.1109/TIP.2004.834669.

Fernández-Suárez M, Ting AY. 2008. Fluorescent probes for super-resolution imaging in living cells. *Nature Reviews Molecular Cell Biology* 9:929–943. DOI: 10.1038/nrm2531.

1421 Freedman G, Fattal R. 2011. Image and video upscaling from local self-examples. *ACM*
1422   *Transactions on Graphics* 30:1–11. DOI: 10.1145/1944846.1944852.
1423 Freeman WT, Jones TR, Pasztor EC. 2002. Example-based super-resolution. *IEEE Computer*
1424   *Graphics and Applications* 22:56–65. DOI: 10.1109/38.988747.
1425 Gao G, Yu Y, Xie J, Yang J, Yang M, Zhang J. 2021. Constructing multilayer locality-
1426   constrained matrix regression framework for noise robust face super-resolution. *Pattern*
1427   *Recognition* 110:107539. DOI: 10.1016/j.patcog.2020.107539.
1428 Gao H, Yuan H, Wang Z, Ji S. 2020. Pixel Transposed Convolutional Networks. *IEEE*
1429   *Transactions on Pattern Analysis and Machine Intelligence* 1:1–1. DOI:
1430   10.1109/TPAMI.2019.2893965.
1431 Gao X, Zhang L, Mou X. 2019. Single Image Super-Resolution Using Dual-Branch
1432   Convolutional Neural Network. *IEEE Access* 7:15767–15778. DOI:
1433   10.1109/ACCESS.2018.2889760.
1434 Gatys LA, Ecker AS, Bethge M. 2015. Texture synthesis using convolutional neural networks.
1435   In: *Advances in Neural Information Processing Systems*. 262–270.
1436 Gauthier J. 2014. Conditional generative adversarial nets for convolutional face generation.
1437   *Technical report*.
1438 Gerchberg RW. 1974. Super-resolution through error energy reduction. *Optica Acta* 21:709–720.
1439   DOI: 10.1080/713818946.
1440 Ghodrati V, Shao J, Bydder M, Zhou Z, Yin W, Nguyen KL, Yang Y, Hu P. 2019. MR image
1441   reconstruction using deep learning: Evaluation of network structure and loss functions.
1442   *Quantitative Imaging in Medicine and Surgery* 9:1516. DOI: 10.21037/qims.2019.08.10.
1443 Glasner D, Bagon S, Irani M. 2009. Super-resolution from a single image. In: *Proceedings of the*
1444   *IEEE International Conference on Computer Vision*. 349–356. DOI:
1445   10.1109/ICCV.2009.5459271.
1446 Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio
1447   Y. 2014. Generative adversarial nets. In: *Advances in Neural Information Processing*
1448   *Systems*. 2672–2680.
1449 Goyal M, Lather Y, Lather V. 2015. Analytical Relation & Comparison of PSNR and SSIM on
1450   Baboon Image and Human Eye Perception Using MATLAB. *International Journal of*
1451   *Advanced Research in Engineering and Applied Sciences* 4:108–119.
1452 Graves A, Fernández S, Schmidhuber J. 2005. Bidirectional LSTM networks for improved
1453   phoneme classification and recognition. In: *Lecture Notes in Computer Science (including*
1454   *subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
1455   799–804.
1456 Griffel DH, Daubechies I. 1995. Ten Lectures on Wavelets. *The Mathematical Gazette* 79:224–
1457   227. DOI: 10.2307/3620105.
1458 Grm K, Scheirer WJ, Štruc V. 2020. Face Hallucination Using Cascaded Super-Resolution and
1459   Identity Priors. *IEEE Transactions on Image Processing* 29:2150–2165. DOI:
1460   10.1109/TIP.2019.2945835.
1461 Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. 2017. Improved training of
1462   wasserstein GANs. In: *Advances in Neural Information Processing Systems*. 5767–5777.
1463 Gunturk BK, Batur AU, Altunbasak Y, Hayes MH, Mersereau RM. 2003. Eigenface-domain
1464   super-resolution for face recognition. *IEEE Transactions on Image Processing* 12:597–606.
1465   DOI: 10.1109/TIP.2003.811513.
1466 Guo J, Chao H. 2017. Building an end-to-end spatial-temporal convolutional network for video

1467   super-resolution. In: *31st AAAI Conference on Artificial Intelligence, AAAI 2017*.

1468 Guo C, Chen D, Huang Z. 2019. Learning Efficient Stereo Matching Network with Depth
1469   Discontinuity Aware Super-Resolution. *IEEE Access* 7:159712–159723. DOI:
1470   10.1109/ACCESS.2019.2950924.

1471 Ha VK, Ren JC, Xu XY, Zhao S, Xie G, Masero V, Hussain A. 2019. Deep Learning Based
1472   Single Image Super-resolution: A Survey. *International Journal of Automation and*
1473   *Computing* 16:413–426. DOI: 10.1007/s11633-019-1183-x.

1474 Hamaide J, De Groof G, Van Steenkiste G, Jeurissen B, Van Ruijssevelt L, Verhoye M, Van der
1475   Linden A, Cornil C, Sijbers J. 2017. Exploring sex differences in the adult zebra finch
1476   brain: In vivo diffusion tensor imaging and ex vivo super-resolution track density imaging.
1477   *NeuroImage* 146:789–803. DOI: 10.1016/j.neuroimage.2016.09.067.

1478 Han W, Chang S, Liu D, Yu M, Witbrock M, Huang TS. 2018. Image Super-Resolution via
1479   Dual-State Recurrent Networks. In: *Proceedings of the IEEE Computer Society Conference*
1480   *on Computer Vision and Pattern Recognition*. 1654–1663. DOI:
1481   10.1109/CVPR.2018.00178.

1482 Haris M, Shakhnarovich G, Ukita N. 2018. Deep Back-Projection Networks for Super-
1483   Resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision*
1484   *and Pattern Recognition*. 1664–1673. DOI: 10.1109/CVPR.2018.00179.

1485 Haris M, Shakhnarovich G, Ukita N. 2019. Recurrent back-projection network for video super-
1486   resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision*
1487   *and Pattern Recognition*. 3897–3906. DOI: 10.1109/CVPR.2019.00402.

1488 He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In:
1489   *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*
1490   *Recognition*. 770–778. DOI: 10.1109/CVPR.2016.90.

1491 Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. 2017. GANs trained by a two
1492   time-scale update rule converge to a local Nash equilibrium. In: *Advances in Neural*
1493   *Information Processing Systems*. 6626–6637.

1494 Horé A, Ziou D. 2010. Image quality metrics: PSNR vs. SSIM. In: *Proceedings - International*
1495   *Conference on Pattern Recognition*. 2366–2369. DOI: 10.1109/ICPR.2010.579.

1496 Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Marco A, Adam H. 2009.
1497   MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. In:
1498   *Computer Vision and Pattern Recognition*. DOI: 10.1016/S1507-1367(10)60022-3.

1499 Hu X, Mu H, Zhang X, Wang Z, Tan T, Sun J. 2019. Meta-SR: A magnification-arbitrary
1500   network for super-resolution. In: *Proceedings of the IEEE Computer Society Conference on*
1501   *Computer Vision and Pattern Recognition*. 1575–1584. DOI: 10.1109/CVPR.2019.00167.

1502 Hu J, Shen L, Sun G. 2018. Squeeze-and-Excitation Networks. In: *Proceedings of the IEEE*
1503   *Computer Society Conference on Computer Vision and Pattern Recognition*. 7132–7141.
1504   DOI: 10.1109/CVPR.2018.00745.

1505 Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. 2017. Densely connected convolutional
1506   networks. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern*
1507   *Recognition, CVPR 2017*. 4700–4708. DOI: 10.1109/CVPR.2017.243.

1508 Huang B, Wang W, Bates M, Zhuang X. 2008. Three-dimensional super-resolution imaging by
1509   stochastic optical reconstruction microscopy. *Science* 319:810–813. DOI:
1510   10.1126/science.1153529.

1511 Huang Y, Wang W, Wang L. 2015. Bidirectional recurrent convolutional networks for multi-
1512   frame super-resolution. In: *Advances in Neural Information Processing Systems*. 235–243.

1513   Hugelier S, De Rooi JJ, Bernex R, Duwé S, Devos O, Sliwa M, Dedecker P, Eilers PHC,
1514   Ruckebusch C. 2016. Sparse deconvolution of high-density super-resolution images.
1515   *Scientific Reports* 6:21413. DOI: 10.1038/srep21413.
1516   Hui TW, Loy CC, Tang X. 2016. Depth map super-resolution by deep multi-scale guidance. In:
1517   *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*
1518   *Intelligence and Lecture Notes in Bioinformatics)*. 353–369. DOI: 10.1007/978-3-319-
1519   46487-9_22.
1520   Hui Z, Wang X, Gao X. 2018. Fast and Accurate Single Image Super-Resolution via Information
1521   Distillation Network. In: *Proceedings of the IEEE Computer Society Conference on*
1522   *Computer Vision and Pattern Recognition*. 723–731. DOI: 10.1109/CVPR.2018.00082.
1523   Ignatov A, Kobyshev N, Timofte R, Vanhoey K, Van Gool L. 2018a. WESPE: Weakly
1524   supervised photo enhancer for digital cameras. In: *IEEE Computer Society Conference on*
1525   *Computer Vision and Pattern Recognition Workshops*. 691–700. DOI:
1526   10.1109/CVPRW.2018.00112.
1527   Ignatov A, Timofte R, Van Vu T, Luu TM, Pham TX, Van Nguyen C, Kim Y, Choi JS, Kim M,
1528   Huang J, Ran J, Xing C, Zhou X, Zhu P, Geng M, Li Y, Agustsson E, Gu S, Van Gool L, de
1529   Stoutz E, Kobyshev N, Nie K, Zhao Y, Li G, Tong T, Gao Q, Hanwen L, Michelini PN,
1530   Dan Z, Fengshuo H, Hui Z, Wang X, Deng L, Meng R, Qin J, Shi Y, Wen W, Lin L, Feng
1531   R, Wu S, Dong C, Qiao Y, Vasu S, Thekke Madam N, Kandula P, Rajagopalan AN, Liu J,
1532   Jung C. 2018b. PIRM challenge on perceptual image enhancement on smartphones: Report.
1533   In: *Proceedings of the European Conference on Computer Vision (ECCV)*. DOI:
1534   10.1007/978-3-030-11021-5_20.
1535   Ioffe S, Szegedy C. 2015. Batch normalization: Accelerating deep network training by reducing
1536   internal covariate shift. In: *32nd International Conference on Machine Learning, ICML*
1537   *2015*.
1538   Irani M, Peleg S. 1991. Improving resolution by image registration. *CVGIP: Graphical Models*
1539   *and Image Processing* 53:231–239. DOI: 10.1016/1049-9652(91)90045-L.
1540   Jaderberg M, Simonyan K, Zisserman A, Kavukcuoglu K. 2015. Spatial transformer networks.
1541   In: *Advances in Neural Information Processing Systems*. 2017–2025.
1542   Johnson J, Alahi A, Fei-Fei L. 2016. Perceptual losses for real-time style transfer and super-
1543   resolution. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in*
1544   *Artificial Intelligence and Lecture Notes in Bioinformatics)*. 694–711. DOI: 10.1007/978-3-
1545   319-46475-6_43.
1546   Jurek J, Kociński M, Materka A, Elgalal M, Majos A. 2020. CNN-based superresolution
1547   reconstruction of 3D MR images using thick-slice scans. *Biocybernetics and Biomedical*
1548   *Engineering* 40:111–125. DOI: 10.1016/j.bbe.2019.10.003.
1549   Kappeler A, Yoo S, Dai Q, Katsaggelos AK. 2016. Super-resolution of compressed videos using
1550   convolutional neural networks. In: *Proceedings - International Conference on Image*
1551   *Processing, ICIP*. 1150–1154. DOI: 10.1109/ICIP.2016.7532538.
1552   Kawulok M, Benecki P, Piechaczek S, Hrynczenko K, Kostrzewa D, Nalepa J. 2020. Deep
1553   Learning for Multiple-Image Super-Resolution. *IEEE Geoscience and Remote Sensing*
1554   *Letters* 17:1062–1066. DOI: 10.1109/LGRS.2019.2940483.
1555   Keys RG. 1981. Cubic Convolution Interpolation for Digital Image Processing. *IEEE*
1556   *Transactions on Acoustics, Speech, and Signal Processing* 29:1153–1160. DOI:
1557   10.1109/TASSP.1981.1163711.
1558   Kim KI, Kwon Y. 2010. Single-image super-resolution using sparse regression and natural

1559   image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32:1127–
1560   1133. DOI: 10.1109/TPAMI.2010.25.

1561 Kim J, Lee S. 2017. Deep learning of human visual sensitivity in image quality assessment
1562   framework. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern
1563   Recognition, CVPR 2017*. 1676–1684. DOI: 10.1109/CVPR.2017.213.

1564 Kim J, Lee JK, Lee KM. 2016a. Accurate image super-resolution using very deep convolutional
1565   networks. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision
1566   and Pattern Recognition*. 1646–1654. DOI: 10.1109/CVPR.2016.182.

1567 Kim J, Lee JK, Lee KM. 2016b. Deeply-recursive convolutional network for image super-
1568   resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision
1569   and Pattern Recognition*. 1637–1645. DOI: 10.1109/CVPR.2016.181.

1570 Kirillov A, He K, Girshick R, Rother C, Dollar P. 2019. Panoptic segmentation. In: *Proceedings
1571   of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
1572   9404–9413. DOI: 10.1109/CVPR.2019.00963.

1573 Kligvasser I, Shaham TR, Michaeli T. 2018. XUnit: Learning a Spatial Activation Function for
1574   Efficient Image Restoration. In: *Proceedings of the IEEE Computer Society Conference on
1575   Computer Vision and Pattern Recognition*. 2433–2442. DOI: 10.1109/CVPR.2018.00258.

1576 Korshunova I, Shi W, Dambre J, Theis L. 2017. Fast Face-Swap Using Convolutional Neural
1577   Networks. In: *Proceedings of the IEEE International Conference on Computer Vision*.
1578   3677–3685. DOI: 10.1109/ICCV.2017.397.

1579 Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional
1580   neural networks. In: *Advances in neural information processing systems*. 1097–1105. DOI:
1581   10.1145/3065386.

1582 Lai WS, Huang J Bin, Ahuja N, Yang MH. 2017. Deep laplacian pyramid networks for fast and
1583   accurate super-resolution. In: *Proceedings - 30th IEEE Conference on Computer Vision and
1584   Pattern Recognition, CVPR 2017*. 624–632. DOI: 10.1109/CVPR.2017.618.

1585 Lam F, Cladière D, Guillaume C, Wassmann K, Bolte S. 2017. Super-resolution for everybody:
1586   An image processing workflow to obtain high-resolution images with a standard confocal
1587   microscope. *Methods* 115:17–27. DOI: 10.1016/j.ymeth.2016.11.003.

1588 Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J,
1589   Wang Z, Shi W. 2017. Photo-realistic single image super-resolution using a generative
1590   adversarial network. In: *Proceedings - 30th IEEE Conference on Computer Vision and
1591   Pattern Recognition, CVPR 2017*. 4681–4690. DOI: 10.1109/CVPR.2017.19.

1592 Lee S, Kim JH, Heo JP. 2020. Super-resolution of license plate images via character-based
1593   perceptual loss. In: *Proceedings - 2020 IEEE International Conference on Big Data and
1594   Smart Computing, BigComp 2020*. 560–563. DOI: 10.1109/BigComp48618.2020.000-1.

1595 Lee K, Xu W, Fan F, Tu Z. 2018a. Wasserstein Introspective Neural Networks. In: *Proceedings
1596   of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
1597   3702–3711. DOI: 10.1109/CVPR.2018.00390.

1598 Lee D, Yang MH, Oh S. 2015. Fast and accurate head pose estimation via random projection
1599   forests. In: *Proceedings of the IEEE International Conference on Computer Vision*. 1958–
1600   1966. DOI: 10.1109/ICCV.2015.227.

1601 Lee CH, Zhang K, Lee HC, Cheng CW, Hsu W. 2018b. Attribute augmented convolutional
1602   neural network for face hallucination. In: *IEEE Computer Society Conference on Computer
1603   Vision and Pattern Recognition Workshops*. 721–729. DOI: 10.1109/CVPRW.2018.00115.

1604 Levy O, Goldberg Y. 2014. Neural word embedding as implicit matrix factorization. In:

1605   *Advances in Neural Information Processing Systems*. 2177–2185.

1606   Li C, Bovik AC. 2010. Content-partitioned structural similarity index for image quality
1607   assessment. *Signal Processing: Image Communication* 25:517–526. DOI:
1608   10.1016/j.image.2010.03.004.

1609   Li J, Fang F, Mei K, Zhang G. 2018. Multi-scale residual network for image super-resolution. In:
1610   *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*
1611   *Intelligence and Lecture Notes in Bioinformatics)*. 517–532. DOI: 10.1007/978-3-030-
1612   01237-3_32.

1613   Li S, He F, Du B, Zhang L, Xu Y, Tao D. 2019a. Fast spatio-temporal residual network for video
1614   super-resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer*
1615   *Vision and Pattern Recognition*. 10522–10531. DOI: 10.1109/CVPR.2019.01077.

1616   Li J, Liang X, Wei Y, Xu T, Feng J, Yan S. 2017a. Perceptual generative adversarial networks
1617   for small object detection. In: *Proceedings - 30th IEEE Conference on Computer Vision and*
1618   *Pattern Recognition, CVPR 2017*. 1222–1230. DOI: 10.1109/CVPR.2017.211.

1619   Li L, Wang W, Luo H, Ying S. 2017b. Super-resolution reconstruction of high-resolution
1620   satellite ZY-3 TLC images. *Sensors (Switzerland)* 17:1062. DOI: 10.3390/s17051062.

1621   Li X, Wu Y, Zhang W, Wang R, Hou F. 2020. Deep learning methods in real-time image super-
1622   resolution: a survey. *Journal of Real-Time Image Processing* 17:1885–1909. DOI:
1623   10.1007/s11554-019-00925-3.

1624   Li Z, Yang J, Liu Z, Yang X, Jeon G, Wu W. 2019b. Feedback network for image super-
1625   resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision*
1626   *and Pattern Recognition*. 3867–3876. DOI: 10.1109/CVPR.2019.00399.

1627   Liao R, Tao X, Li R, Ma Z, Jia J. 2015. Video super-resolution via deep draft-ensemble learning.
1628   In: *Proceedings of the IEEE International Conference on Computer Vision*. 531–539. DOI:
1629   10.1109/ICCV.2015.68.

1630   Lim B, Son S, Kim H, Nah S, Lee KM. 2017. Enhanced Deep Residual Networks for Single
1631   Image Super-Resolution. In: *IEEE Computer Society Conference on Computer Vision and*
1632   *Pattern Recognition Workshops*. 136–144. DOI: 10.1109/CVPRW.2017.151.

1633   Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. 2014.
1634   Microsoft COCO: Common objects in context. In: *European Conference on Computer*
1635   *Vision*. 740–755. DOI: 10.1007/978-3-319-10602-1_48.

1636   Liu B, Li H, Zhou Y, Peng Y, Elazab A, Wang C. 2020a. A super resolution method for remote
1637   sensing images based on cascaded conditional wasserstein GANs. In: *2020 3rd IEEE*
1638   *International Conference on Information Communication and Signal Processing, ICICSP*
1639   *2020*. 284–289. DOI: 10.1109/ICICSP50920.2020.9232066.

1640   Liu Z, Luo P, Wang X, Tang X. 2015. Deep learning face attributes in the wild. In: *Proceedings*
1641   *of the IEEE International Conference on Computer Vision*. 3730–3738. DOI:
1642   10.1109/ICCV.2015.425.

1643   Liu H, Ruan Z, Zhao P, Shang F, Yang L, Liu Y. 2020b. Video super resolution based on deep
1644   learning: A comprehensive survey. *arXiv*.

1645   Liu Di, Wang Z, Fan Y, Liu X, Wang Z, Chang S, Wang X, Huang TS. 2018a. Learning
1646   Temporal Dynamics for Video Super-Resolution: A Deep Learning Approach. *IEEE*
1647   *Transactions on Image Processing* 27:3432–3445. DOI: 10.1109/TIP.2018.2820807.

1648   Liu X, Weijer J Van De, Bagdanov AD. 2017. RankIQA: Learning from Rankings for No-
1649   Reference Image Quality Assessment. In: *Proceedings of the IEEE International*
1650   *Conference on Computer Vision*. 1040–104. DOI: 10.1109/ICCV.2017.118.

1651 Liu P, Zhang H, Zhang K, Lin L, Zuo W. 2018b. Multi-level wavelet-CNN for image restoration.
1652   In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*
1653   *Workshops*. 773–782. DOI: 10.1109/CVPRW.2018.00121.
1654 Ma K, Liu W, Liu T, Wang Z, Tao D. 2017a. DipIQ: Blind Image Quality Assessment by
1655   Learning-to-Rank Discriminable Image Pairs. *IEEE Transactions on Image Processing*
1656   26:3951–3964. DOI: 10.1109/TIP.2017.2708503.
1657 Ma K, Liu W, Zhang K, Duanmu Z, Wang Z, Zuo W. 2018. End-To-end blind image quality
1658   assessment using deep neural networks. *IEEE Transactions on Image Processing* 27:1202–
1659   1213. DOI: 10.1109/TIP.2017.2774045.
1660 Ma W, Pan Z, Guo J, Lei B. 2019. Achieving Super-Resolution Remote Sensing Images via the
1661   Wavelet Transform Combined with the Recursive Res-Net. *IEEE Transactions on*
1662   *Geoscience and Remote Sensing* 57:3512–3527. DOI: 10.1109/TGRS.2018.2885506.
1663 Ma C, Yang CY, Yang X, Yang MH. 2017b. Learning a no-reference quality metric for single-
1664   image super-resolution. *Computer Vision and Image Understanding* 158:1–16. DOI:
1665   10.1016/j.cviu.2016.12.009.
1666 Mahapatra D, Bozorgtabar B, Garnavi R. 2019. Image super-resolution using progressive
1667   generative adversarial networks for medical image analysis. *Computerized Medical Imaging*
1668   *and Graphics* 71:30–39. DOI: 10.1016/j.compmedimag.2018.10.005.
1669 Mao XJ, Shen C, Yang Y Bin. 2016. Image restoration using very deep convolutional encoder-
1670   decoder networks with symmetric skip connections. In: *Advances in Neural Information*
1671   *Processing Systems*. 2802–2810.
1672 Martin D, Fowlkes C, Tal D, Malik J. 2001. A database of human segmented natural images and
1673   its application to evaluating segmentation algorithms and measuring ecological statistics. In:
1674   *Proceedings of the IEEE International Conference on Computer Vision*. 416–423. DOI:
1675   10.1109/ICCV.2001.937655.
1676 Michaeli T, Irani M. 2013. Nonparametric blind super-resolution. In: *Proceedings of the IEEE*
1677   *International Conference on Computer Vision*. 945–952. DOI: 10.1109/ICCV.2013.121.
1678 Mittal A, Soundararajan R, Bovik AC. 2013. Making a "completely blind" image quality
1679   analyzer. *IEEE Signal Processing Letters* 20:209–212. DOI: 10.1109/LSP.2012.2227726.
1680 Miyato T, Kataoka T, Koyama M, Yoshida Y. 2018. Spectral normalization for generative
1681   adversarial networks. In: *6th International Conference on Learning Representations, ICLR*
1682   *2018 - Conference Track Proceedings*.
1683 Najemnik J, Geisler WS. 2005. Optimal eye movement strategies in visual search. *Nature*
1684   434:387–391. DOI: 10.1038/nature03390.
1685 Nasrollahi K, Moeslund TB. 2014. Super-resolution: A comprehensive survey. *Machine Vision*
1686   *and Applications* 25:1423–1468. DOI: 10.1007/s00138-014-0623-4.
1687 Odena A, Dumoulin V, Olah C. 2017. Deconvolution and Checkerboard Artifacts. *Distill* 1:e3.
1688   DOI: 10.23915/distill.00003.
1689 Van Den Oord A, Kalchbrenner N, Vinyals O, Espeholt L, Graves A, Kavukcuoglu K. 2016.
1690   Conditional image generation with PixelCNN decoders. In: *Advances in Neural Information*
1691   *Processing Systems*. 4790–4798.
1692 van Ouwerkerk JD. 2006. Image super-resolution survey. *Image and Vision Computing* 24:1039–
1693   1052. DOI: 10.1016/j.imavis.2006.02.026.
1694 Pambrun JF, Noumeir R. 2015. Limitations of the SSIM quality metric in the context of
1695   diagnostic imaging. In: *Proceedings - International Conference on Image Processing, ICIP*.
1696   2960–2963. DOI: 10.1109/ICIP.2015.7351345.

Park D, Kim K, Chun SY. 2018. Efficient module based single image super resolution for multiple problems. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 882–890. DOI: 10.1109/CVPRW.2018.00133.

Park SC, Park MK, Kang MG. 2003. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine* 20:21–36. DOI: 10.1109/MSP.2003.1203207.

Park SJ, Son H, Cho S, Hong KS, Lee S. 2018. SRFeat: Single Image Super-Resolution with Feature Discrimination. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 439–455. DOI: 10.1007/978-3-030-01270-0_27.

Parker JA, Kenyon R V., Troxel DE. 1983. Comparison of Interpolating Methods for Image Resampling. *IEEE Transactions on Medical Imaging*. DOI: 10.1109/TMI.1983.4307610.

Protter M, Elad M, Takeda H, Milanfar P. 2009. Generalizing the nonlocal-means to super-resolution reconstruction. *IEEE Transactions on Image Processing* 18:36–51. DOI: 10.1109/TIP.2008.2008067.

Qiao C, Li D, Guo Y, Liu C, Jiang T, Dai Q, Li D. 2021. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nature Methods* 18:194–202. DOI: 10.1038/s41592-020-01048-5.

Ravì D, Szczotka AB, Pereira SP, Vercauteren T. 2019. Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy. *Medical Image Analysis* 53:123–131. DOI: 10.1016/j.media.2019.01.011.

Ravì D, Szczotka AB, Shakir DI, Pereira SP, Vercauteren T. 2018. Effective deep learning training for single-image super-resolution in endomicroscopy exploiting video-registration-based reconstruction. *International Journal of Computer Assisted Radiology and Surgery* 13:917–924. DOI: 10.1007/s11548-018-1764-0.

Ren H, El-Khamy M, Lee J. 2017. Image Super Resolution Based on Fusing Multiple Convolution Neural Networks. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 54–61. DOI: 10.1109/CVPRW.2017.142.

Riegler G, Rüther M, Bischof H. 2016. ATGV-net: Accurate depth super-resolution. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 268–284. DOI: 10.1007/978-3-319-46487-9_17.

Rouse DM, Hemami SS. 2008. Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM. *Human Vision and Electronic Imaging XIII* 6806:680615. DOI: 10.1117/12.768060.

Rudin LI, Osher S, Fatemi E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60:259–268. DOI: 10.1016/0167-2789(92)90242-F.

Saad MA, Bovik AC, Charrier C. 2012. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Transactions on Image Processing* 21:3339–3352. DOI: 10.1109/TIP.2012.2191563.

Sajjadi MSM, Scholkopf B, Hirsch M. 2017. EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. In: *Proceedings of the IEEE International Conference on Computer Vision*. 4491–4500. DOI: 10.1109/ICCV.2017.481.

Sara U, Akter M, Uddin MS. 2019. Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study. *Journal of Computer and Communications* 7:8–18. DOI: 10.4236/jcc.2019.73002.

Shaik KB, Ganesan P, Kalist V, Sathish BS, Jenitha JMM. 2015. Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space. *Procedia Computer*

1743   *Science* 57:41–48. DOI: 10.1016/j.procs.2015.07.362.

1744 Shamsolmoali P, Zareapoor M, Jain DK, Jain VK, Yang J. 2018. Deep convolution network for
1745   surveillance records super-resolution. *Multimedia Tools and Applications* 78:23815–23829.
1746   DOI: 10.1007/s11042-018-5915-7.

1747 Shermeyer J, Van Etten A. 2019. The effects of super-resolution on object detection performance
1748   in satellite imagery. In: *IEEE Computer Society Conference on Computer Vision and
1749   Pattern Recognition Workshops*. 0–0. DOI: 10.1109/CVPRW.2019.00184.

1750 Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z. 2016. Real-
1751   Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel
1752   Convolutional Neural Network. In: *Proceedings of the IEEE Computer Society Conference
1753   on Computer Vision and Pattern Recognition*. 1874–1883. DOI: 10.1109/CVPR.2016.207.

1754 Shocher A, Cohen N, Irani M. 2018. Zero-Shot Super-Resolution Using Deep Internal Learning.
1755   In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern
1756   Recognition*. 3118–3126. DOI: 10.1109/CVPR.2018.00329.

1757 Simonyan K, Zisserman A. 2015. Very deep convolutional networks for large-scale image
1758   recognition. In: *3rd International Conference on Learning Representations, ICLR 2015 -
1759   Conference Track Proceedings*.

1760 Sønderby CK, Caballero J, Theis L, Shi W, Huszár F. 2017. Amortised map inference for image
1761   super-resolution. In: *5th International Conference on Learning Representations, ICLR 2017
1762   - Conference Track Proceedings*. DOI: 10.17863/CAM.51994.

1763 Song L, Zhang M, Wu X, He R. 2018. Adversarial discriminative heterogeneous face
1764   recognition. In: *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*.

1765 Sroubek F, Cristobal G, Flusser J. 2008. Simultaneous super-resolution and blind deconvolution.
1766   In: *Journal of Physics: Conference Series*. 012048. DOI: 10.1088/1742-6596/124/1/012048.

1767 Van Steenkiste G, Jeurissen B, Veraart J, Den Dekker AJ, Parizel PM, Poot DHJ, Sijbers J. 2016.
1768   Super-resolution reconstruction of diffusion parameters from diffusion-weighted images
1769   with different slice orientations. *Magnetic Resonance in Medicine* 75:181–195. DOI:
1770   10.1002/mrm.25597.

1771 Sun D, Roth S, Black MJ. 2010. Secrets of optical flow estimation and their principles. In:
1772   *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern
1773   Recognition*. 2432–2439. DOI: 10.1109/CVPR.2010.5539939.

1774 Sun J, Xu Z, Shum HY. 2008. Image super-resolution using gradient profile prior. In: *26th IEEE
1775   Conference on Computer Vision and Pattern Recognition, CVPR*. 1–8. DOI:
1776   10.1109/CVPR.2008.4587659.

1777 Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich
1778   A. 2015. Going deeper with convolutions. In: *Proceedings of the IEEE Computer Society
1779   Conference on Computer Vision and Pattern Recognition*. 1–9. DOI:
1780   10.1109/CVPR.2015.7298594.

1781 Tai Y, Yang J, Liu X. 2017. Image super-resolution via deep recursive residual network. In:
1782   *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR
1783   2017*. 3147–3155. DOI: 10.1109/CVPR.2017.298.

1784 Tai Y, Yang J, Liu X, Xu C. 2017. MemNet: A Persistent Memory Network for Image
1785   Restoration. In: *Proceedings of the IEEE International Conference on Computer Vision*.
1786   4539–4547. DOI: 10.1109/ICCV.2017.486.

1787 Taigman Y, Yang M, Ranzato M, Wolf L. 2014. DeepFace: Closing the gap to human-level
1788   performance in face verification. In: *Proceedings of the IEEE Computer Society Conference*

on *Computer Vision and Pattern Recognition*. 1701–1708. DOI: 10.1109/CVPR.2014.220.

Talebi H, Milanfar P. 2018. NIMA: Neural Image Assessment. *IEEE Transactions on Image Processing* 27:3998–4011. DOI: 10.1109/TIP.2018.2831899.

Tan W, Yan B, Bare B. 2018. Feature Super-Resolution: Make Machine See More Clearly. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3994–4002. DOI: 10.1109/CVPR.2018.00420.

Tao X, Gao H, Liao R, Wang J, Jia J. 2017. Detail-Revealing Deep Video Super-Resolution. In: *Proceedings of the IEEE International Conference on Computer Vision*. 4472–4480. DOI: 10.1109/ICCV.2017.479.

Teh I, McClymont D, Carruth E, Omens J, McCulloch A, Schneider JE. 2020. Improved compressed sensing and super-resolution of cardiac diffusion MRI with structure-guided total variation. *Magnetic Resonance in Medicine*:1–13. DOI: 10.1002/mrm.28245.

Thapa D, Raahemifar K, Bobier WR, Lakshminarayanan V. 2016. A performance comparison among different super-resolution techniques. *Computers and Electrical Engineering* 54:313–329. DOI: 10.1016/j.compeleceng.2015.09.011.

Thung KH, Raveendran P. 2009. A survey of image quality measures. In: *International Conference for Technical Postgraduates 2009, TECHPOS 2009*. DOI: 10.1109/TECHPOS.2009.5412098.

Timofte R, Agustsson E, Gool L Van, Yang M-H, Zhang L, Lim B, Son S, Kim H, Nah S, Lee KM, Wang X, Tian Y, Yu K, Zhang Y, Wu S, Dong C, Lin L, Qiao Y, Loy CC, Bae W, Yoo J, Han Y, Ye JC, Choi J-S, Kim M, Fan Y, Yu J, Han W, Liu D, Yu H. 2017. NTIRE 2018 challenge on single image super-resolution: Methods and results. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1110–1121. DOI: 10.1109/CVPRW.2017.149.

Timofte R, De V, Gool L Van. 2013. Anchored neighborhood regression for fast example-based super-resolution. In: *Proceedings of the IEEE International Conference on Computer Vision*. 1920–1927. DOI: 10.1109/ICCV.2013.241.

Timofte R, Rothe R, Gool L Van. 2016. Seven Ways to Improve Example-Based Single Image Super Resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1865–1873. DOI: 10.1109/CVPR.2016.206.

Timofte R, De Smet V, Van Gool L. 2015. A+: Adjusted anchored neighborhood regression for fast super-resolution. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 111–126. DOI: 10.1007/978-3-319-16817-3_8.

Tipping ME, Bishop CM. 2003. Bayesian image super-resolution. In: *Advances in Neural Information Processing Systems*. 1303–1310.

Tong T, Li G, Liu X, Gao Q. 2017. Image Super-Resolution Using Dense Skip Connections. In: *Proceedings of the IEEE International Conference on Computer Vision*. 4799–4807. DOI: 10.1109/ICCV.2017.514.

Tsai R, Huang TS. 1984. Multiframe Image Restoration and Registration. *Advances in Computer Vision and Image Processing* 1:317–339.

Ulyanov D, Vedaldi A, Lempitsky V. 2020. Deep Image Prior. *International Journal of Computer Vision*. DOI: 10.1007/s11263-020-01303-4.

Vasu S, Thekke Madam N, Rajagopalan AN. 2019. Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*

1835   *Bioinformatics)*. DOI: 10.1007/978-3-030-11021-5_8.

1836 Viswanathan M, Viswanathan M. 2005. Measuring speech quality for text-to-speech systems:
1837   Development and assessment of a modified mean opinion score (MOS) scale. *Computer*
1838   *Speech and Language* 19:55–83. DOI: 10.1016/j.csl.2003.12.001.

1839 Vu T, Nguyen C V., Pham TX, Luu TM, Yoo CD. 2019. Fast and efficient image quality
1840   enhancement via desubpixel convolutional neural networks. In: *Lecture Notes in Computer*
1841   *Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*
1842   *Bioinformatics)*. DOI: 10.1007/978-3-030-11021-5_16.

1843 Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. 2004. Image quality assessment: From error
1844   visibility to structural similarity. *IEEE Transactions on Image Processing* 13:600–612.
1845   DOI: 10.1109/TIP.2003.819861.

1846 Wang Z, Chen J, Hoi SCHC. 2020. Deep Learning for Image Super-resolution: A Survey. *IEEE*
1847   *Transactions on Pattern Analysis and Machine Intelligence*:1–1. DOI:
1848   10.1109/TPAMI.2020.2982166.

1849 Wang Y, Perazzi F, McWilliams B, Sorkine-Hornung A, Sorkine-Hornung O, Schroers C.
1850   2018a. A fully progressive approach to single-image super-resolution. In: *IEEE Computer*
1851   *Society Conference on Computer Vision and Pattern Recognition Workshops*. 864–873.
1852   DOI: 10.1109/CVPRW.2018.00131.

1853 Wang Z, Simoncelli EP, Bovik AC. 2003. Multi-scale structural similarity for image quality
1854   assessment. In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*.
1855   1398–1402. DOI: 10.1109/acssc.2003.1292216.

1856 Wang L, Wang Y, Liang Z, Lin Z, Yang J, An W, Guo Y. 2019a. Learning parallax attention for
1857   stereo image super-resolution. In: *Proceedings of the IEEE Computer Society Conference*
1858   *on Computer Vision and Pattern Recognition*. 12250–12259. DOI:
1859   10.1109/CVPR.2019.01253.

1860 Wang X, Yu K, Dong C, Change Loy C. 2018b. Recovering Realistic Texture in Image Super-
1861   Resolution by Deep Spatial Feature Transform. In: *Proceedings of the IEEE Computer*
1862   *Society Conference on Computer Vision and Pattern Recognition*. 606–615. DOI:
1863   10.1109/CVPR.2018.00070.

1864 Wang X, Yu K, Dong C, Tang X, Loy CC. 2019b. Deep network interpolation for continuous
1865   imagery effect transition. In: *Proceedings of the IEEE Computer Society Conference on*
1866   *Computer Vision and Pattern Recognition*. 1692–1701. DOI: 10.1109/CVPR.2019.00179.

1867 Wang X, Yu K, Wu S, Gu J, Liu Y. 2019c. ESRGAN : Enhanced Super-Resolution Generative
1868   Adversarial Networks. In: Leal-Taixé L, Roth S eds. *Leal-Taixé L., Roth S. (eds) Computer*
1869   *Vision – ECCV 2018 Workshops. ECCV 2018*. Springer, Cham, 63–79. DOI:
1870   https://doi.org/10.1007/978-3-030-11021-5_5.

1871 Wei SE, Ramakrishna V, Kanade T, Sheikh Y. 2016. Convolutional pose machines. In:
1872   *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*
1873   *Recognition*. 4724–4732. DOI: 10.1109/CVPR.2016.511.

1874 Wei Z, Yuan J, Cai Y. 1999. Picture quality evaluation method based on human perception. *Tien*
1875   *Tzu Hsueh Pao/Acta Electronica Sinica* 4.

1876 Xiong Z, Sun X, Wu F. 2010. Robust web image/video super-resolution. *IEEE Transactions on*
1877   *Image Processing* 19:2017–2028. DOI: 10.1109/TIP.2010.2045707.

1878 Xu X, Sun D, Pan J, Zhang Y, Pfister H, Yang MH. 2017. Learning to Super-Resolve Blurry
1879   Face and Text Images. In: *Proceedings of the IEEE International Conference on Computer*
1880   *Vision*. 251–260. DOI: 10.1109/ICCV.2017.36.

1881 Xue S, Qiu W, Liu F, Jin X. 2020. Wavelet-based residual attention network for image super-
1882   resolution. *Neurocomputing* 382:116–126. DOI: 10.1016/j.neucom.2019.11.044.
1883 Yang CY, Liu S, Yang MH. 2018. Hallucinating Compressed Face Images. *International*
1884   *Journal of Computer Vision* 126:597–614. DOI: 10.1007/s11263-017-1044-4.
1885 Yang CY, Ma C, Yang MH. 2014. Single-image super-resolution: A benchmark. In: *Lecture*
1886   *Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and*
1887   *Lecture Notes in Bioinformatics)*. 372–386. DOI: 10.1007/978-3-319-10593-2_25.
1888 Yang Y, Qi Y. 2021. Image super-resolution via channel attention and spatial graph
1889   convolutional network. *Pattern Recognition* 112:107798. DOI:
1890   10.1016/j.patcog.2020.107798.
1891 Yang J, Wright J, Huang T, Ma Y. 2008. Image super-resolution as sparse representation of raw
1892   image patches. In: *26th IEEE Conference on Computer Vision and Pattern Recognition,*
1893   *CVPR*. 1–8. DOI: 10.1109/CVPR.2008.4587647.
1894 Yang J, Wright J, Huang TS, Ma Y. 2010. Image super-resolution via sparse representation.
1895   *IEEE Transactions on Image Processing* 19:2861–2873. DOI: 10.1109/TIP.2010.2050625.
1896 Yang CY, Yang MH. 2013. Fast direct super-resolution by simple functions. In: *Proceedings of*
1897   *the IEEE International Conference on Computer Vision*. 561–568. DOI:
1898   10.1109/ICCV.2013.75.
1899 Yang Q, Yang R, Davis J, Nistér D. 2007. Spatial-depth super resolution for range images. In:
1900   *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*
1901   *Recognition*. 1–8. DOI: 10.1109/CVPR.2007.383211.
1902 Yang W, Zhang X, Tian Y, Wang W, Xue JH, Liao Q. 2019. Deep Learning for Single Image
1903   Super-Resolution: A Brief Review. *IEEE Transactions on Multimedia* 21:3106–3121. DOI:
1904   10.1109/TMM.2019.2919431.
1905 Yao J, Liu G. 2018. Improved SSIM IQA of contrast distortion based on the contrast sensitivity
1906   characteristics of HVS. *IET Image Processing* 12:872–879. DOI: 10.1049/iet-
1907   ipr.2017.0209.
1908 Yu X, Fernando B, Hartley R, Porikli F. 2018. Super-Resolving Very Low-Resolution Face
1909   Images with Supplementary Attributes. In: *Proceedings of the IEEE Computer Society*
1910   *Conference on Computer Vision and Pattern Recognition*. 908–917. DOI:
1911   10.1109/CVPR.2018.00101.
1912 Yu F, Koltun V. 2016. Multi-scale context aggregation by dilated convolutions. In: *4th*
1913   *International Conference on Learning Representations, ICLR 2016 - Conference Track*
1914   *Proceedings*.
1915 Yu X, Porikli F. 2016. Ultra-resolving face images by discriminative generative networks. In:
1916   *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*
1917   *Intelligence and Lecture Notes in Bioinformatics)*. 318–333. DOI: 10.1007/978-3-319-
1918   46454-1_20.
1919 Yu X, Porikli F. 2017. Face hallucination with tiny unaligned images by transformative
1920   discriminative neural networks. In: *31st AAAI Conference on Artificial Intelligence, AAAI*
1921   *2017*.
1922 Yuan Y, Liu S, Zhang J, Zhang Y, Dong C, Lin L. 2018. Unsupervised image super-resolution
1923   using cycle-in-cycle generative adversarial networks. In: *IEEE Computer Society*
1924   *Conference on Computer Vision and Pattern Recognition Workshops*. 701–710. DOI:
1925   10.1109/CVPRW.2018.00113.
1926 Zaitoun NM, Aqel MJ. 2015. Survey on Image Segmentation Techniques. *Procedia Computer*

1927   *Science* 65:797–806. DOI: 10.1016/j.procs.2015.09.027.

1928   Zeiler MD, Krishnan D, Taylor GW, Fergus R. 2010. Deconvolutional networks. In:
1929   *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*
1930   *Recognition*. 2528–2535. DOI: 10.1109/CVPR.2010.5539957.

1931   Zhang YN, An MQ. 2017. Deep Learning- and Transfer Learning-Based Super Resolution
1932   Reconstruction from Single Medical Image. *Journal of Healthcare Engineering* 2017. DOI:
1933   10.1155/2017/5859727.

1934   Zhang X, Chen Q, Ng R, Koltun V. 2019a. Zoom to learn, learn to zoom. In: *Proceedings of the*
1935   *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 3762–
1936   3770. DOI: 10.1109/CVPR.2019.00388.

1937   Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. 2018a. Image super-resolution using very deep
1938   residual channel attention networks. In: *Lecture Notes in Computer Science (including*
1939   *subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
1940   DOI: 10.1007/978-3-030-01234-2_18.

1941   Zhang Y, Li K, Li K, Zhong B, Fu Y. 2019b. Residual non-local attention networks for image
1942   restoration. In: *7th International Conference on Learning Representations, ICLR 2019*.

1943   Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. 2018b. Residual Dense Network for Image Super-
1944   Resolution. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision*
1945   *and Pattern Recognition*. 2472–2481. DOI: 10.1109/CVPR.2018.00262.

1946   Zhang H, Wang P, Zhang C, Jiang Z. 2019c. A comparable study of CNN-based single image
1947   super-resolution for space-based imaging sensors. *Sensors (Switzerland)* 19:3234. DOI:
1948   10.3390/s19143234.

1949   Zhang K, Zhang Z, Cheng CW, Hsu WH, Qiao Y, Liu W, Zhang T. 2018c. Super-Identity
1950   Convolutional Neural Network for Face Hallucination. In: *Proceedings of the European*
1951   *Conference on Computer Vision (ECCV)*. 183–198. DOI: 10.1007/978-3-030-01252-6_12.

1952   Zhang L, Zhang L, Mou X, Zhang D. 2011. FSIM: A feature similarity index for image quality
1953   assessment. *IEEE Transactions on Image Processing* 20:2378–2386. DOI:
1954   10.1109/TIP.2011.2109730.

1955   Zhang L, Zhang H, Shen H, Li P. 2010. A super-resolution reconstruction algorithm for
1956   surveillance images. *Signal Processing* 90:848–859. DOI: 10.1016/j.sigpro.2009.09.002.

1957   Zhang K, Zuo W, Zhang L. 2018. Learning a Single Convolutional Super-Resolution Network
1958   for Multiple Degradations. In: *2018 IEEE/CVF Conference on Computer Vision and*
1959   *Pattern Recognition*. 3262–3271. DOI: 10.1109/CVPR.2018.00344.

1960   Zhao H, Gallo O, Frosio I, Kautz J. 2016. Loss Functions for Image Restoration With Neural
1961   Networks. *IEEE Transactions on Computational Imaging* 3:47–57. DOI:
1962   10.1109/tci.2016.2644865.

1963   Zhao H, Shi J, Qi X, Wang X, Jia J. 2017. Pyramid scene parsing network. In: *Proceedings -*
1964   *30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. 2881–
1965   2890. DOI: 10.1109/CVPR.2017.660.

1966   Zhou L, Feng S. 2019. A Review of Deep Learning for Single Image Super-Resolution. In:
1967   *ICIIBMS 2019 - 4th International Conference on Intelligent Informatics and Biomedical*
1968   *Sciences*. 139–142. DOI: 10.1109/ICIIBMS46890.2019.8991477.

1969   Zhu X, Li Z, Lou J, Shen Q. 2021. Video super-resolution based on a spatio-temporal matching
1970   network. *Pattern Recognition* 110:107619. DOI: 10.1016/j.patcog.2020.107619.

1971   Zhu S, Liu S, Loy CC, Tang X. 2016. Deep cascaded Bi-network for face hallucination. In:
1972   *Proceedings of the European Conference on Computer Vision (ECCV)*. 614–630. DOI:

1973   10.1007/978-3-319-46454-1_37.

1974   Zhu JY, Park T, Isola P, Efros AA. 2017a. Unpaired Image-to-Image Translation Using Cycle-
1975   Consistent Adversarial Networks. In: *Proceedings of the IEEE International Conference on*
1976   *Computer Vision*. 2223–2232. DOI: 10.1109/ICCV.2017.244.

1977   Zhu H, Tang X, Xie J, Song W, Mo F, Gao X. 2018. Spatio-temporal super-resolution
1978   reconstruction of remote-sensing images based on adaptive multi-scale detail enhancement.
1979   *Sensors (Switzerland)* 18:498. DOI: 10.3390/s18020498.

1980   Zhu XX, Tuia D, Mou L, Xia G-S, Zhang L, Xu F, Fraundorfer F. 2017b. Deep Learning in
1981   Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and*
1982   *Remote Sensing Magazine* 5:8–36. DOI: 10.1109/MGRS.2017.2762307.

1983   Zomet A, Rav-Acha A, Peleg S. 2001. Robust super-resolution. In: *Proceedings of the IEEE*
1984   *Computer Society Conference on Computer Vision and Pattern Recognition*. I–I. DOI:
1985   10.1109/cvpr.2001.990535.

1986

Figure 1

Hierarchical classification of this survey.

Four main categories are (a) classical methods of image super-resolution, (b) deep learning-based methods for SR, (c) applications of super-resolution, (d) future research and directions in SR. *Green* color represent first-level sections, the *blue* color is for second-level subsections, and *orange* color represent third level subsections.

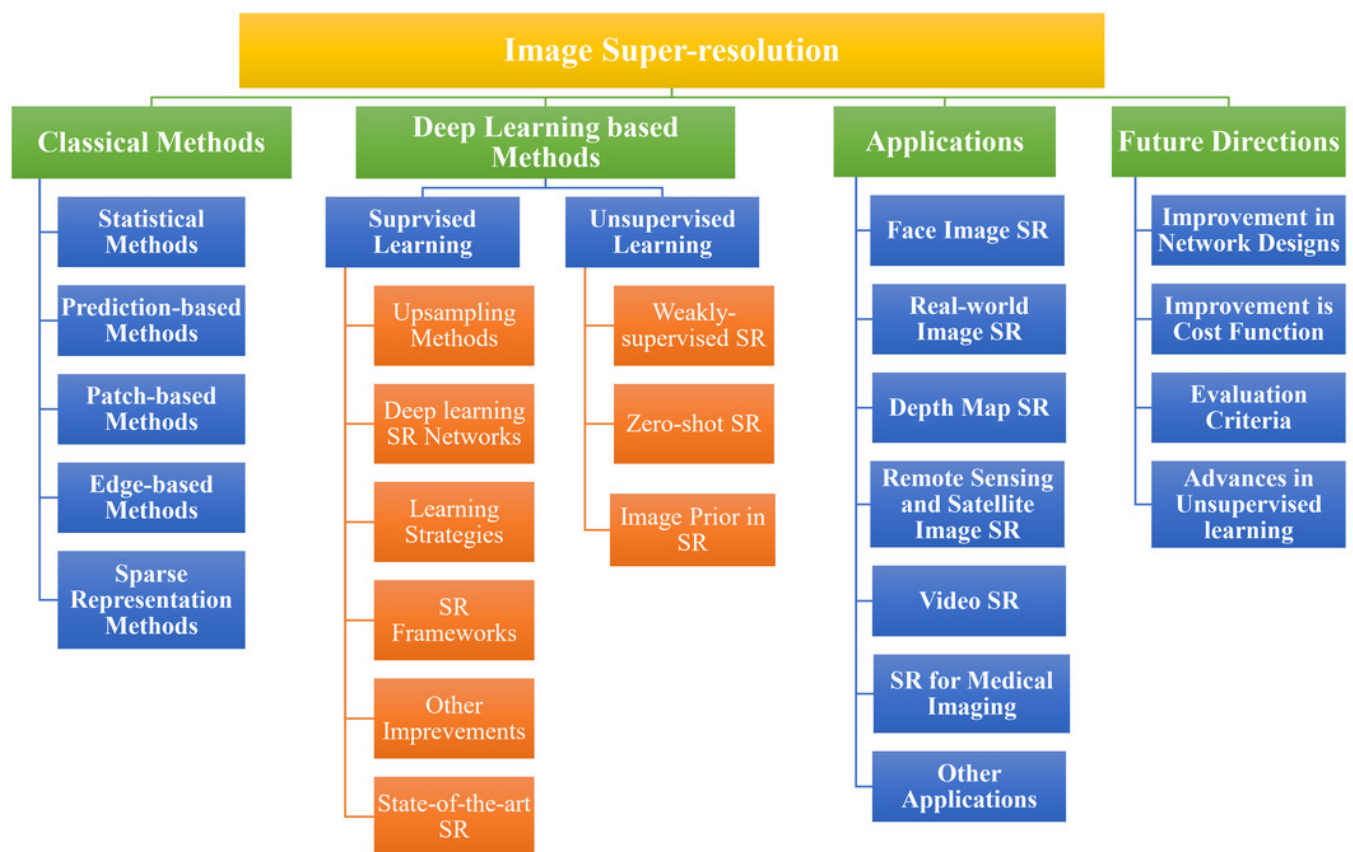



Figure 2

Downsampling and upsampling in super-resolution.

Noise is added to simulate realistic degradation within an image.

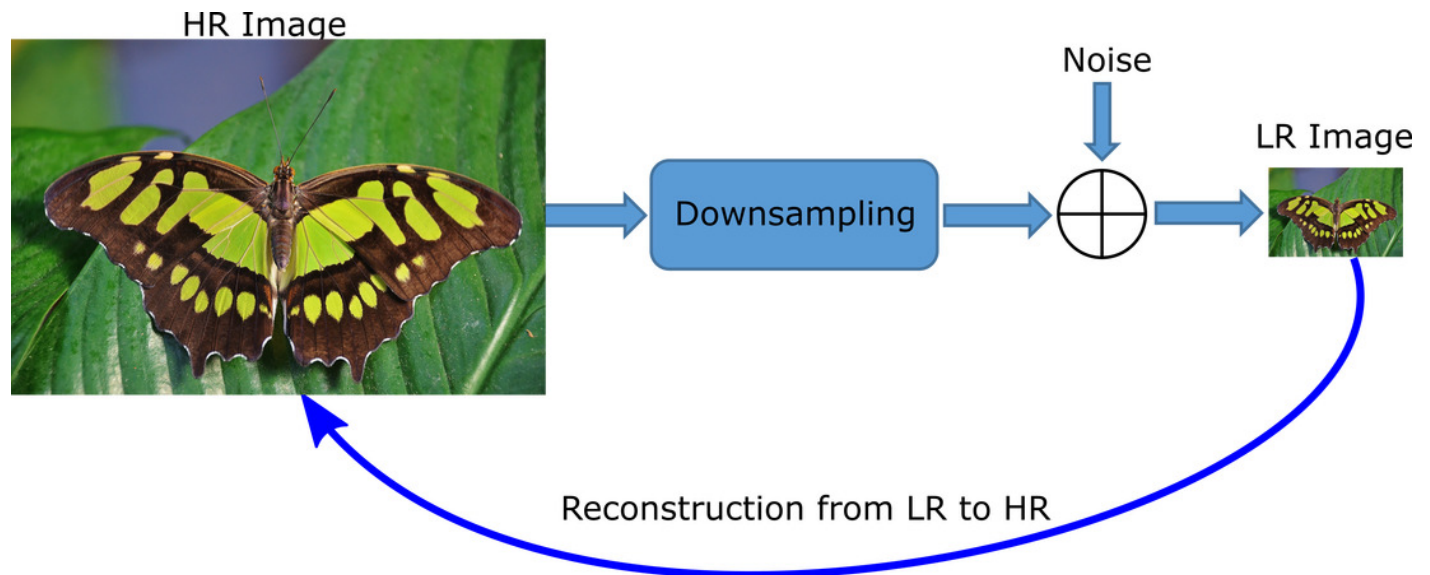



Figure 3

Methodology for the collection of studies.

Sample details based on inclusion/exclusion criteria defined in Table 5.

Google scholar (n = 538)
IEEE Xplore (n = 76)
Science Direct (n = 39)

Total articles identified
using databases,
n = 653

Articles after initial
screening and duplicate
removal
n = 195

Articles identified from
bibliography of already
searched papers
n = 47

Total articles identified
using databases,
n = 242

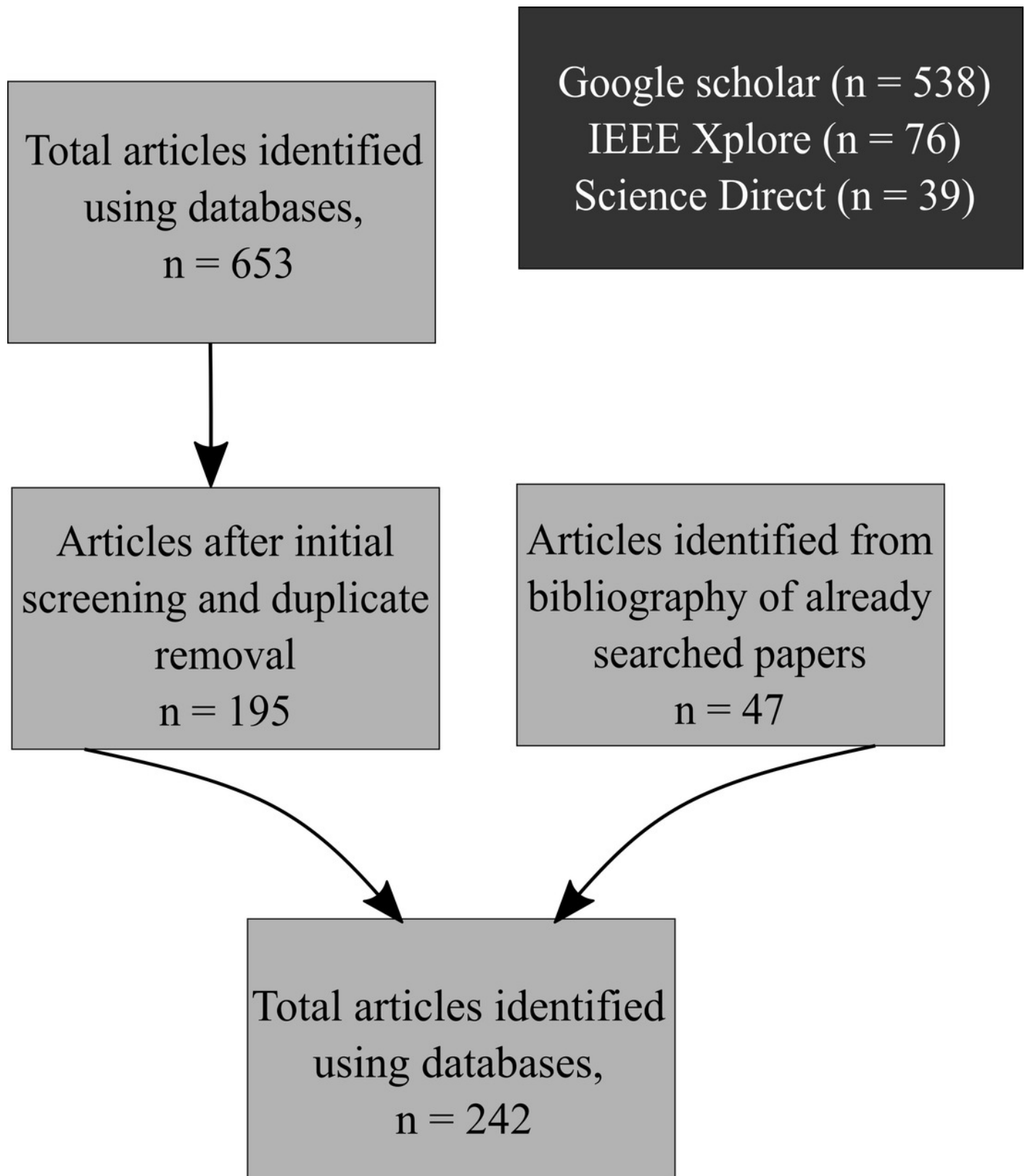



Figure 4

Sub-pixel layer. *Blue* color represents the input convolution, and output feature maps are represented in other colors.

(a) Input (b) Convolution (c) Reshaping

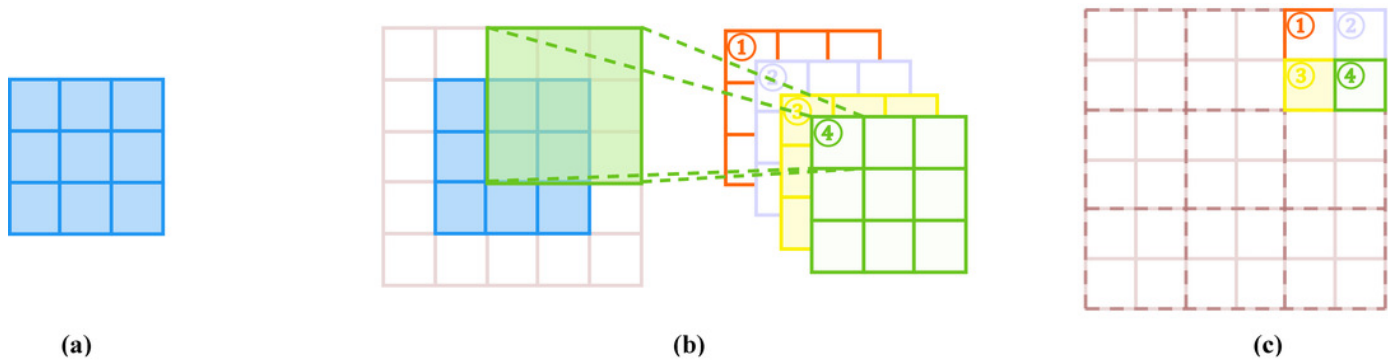



Figure 5

Deconvolution layer. The *blue* color represents the input, and the *green* color represents the convolution operation.

(a) Input (b) Expansion (c) Convolution

(a)   (b)   (c)

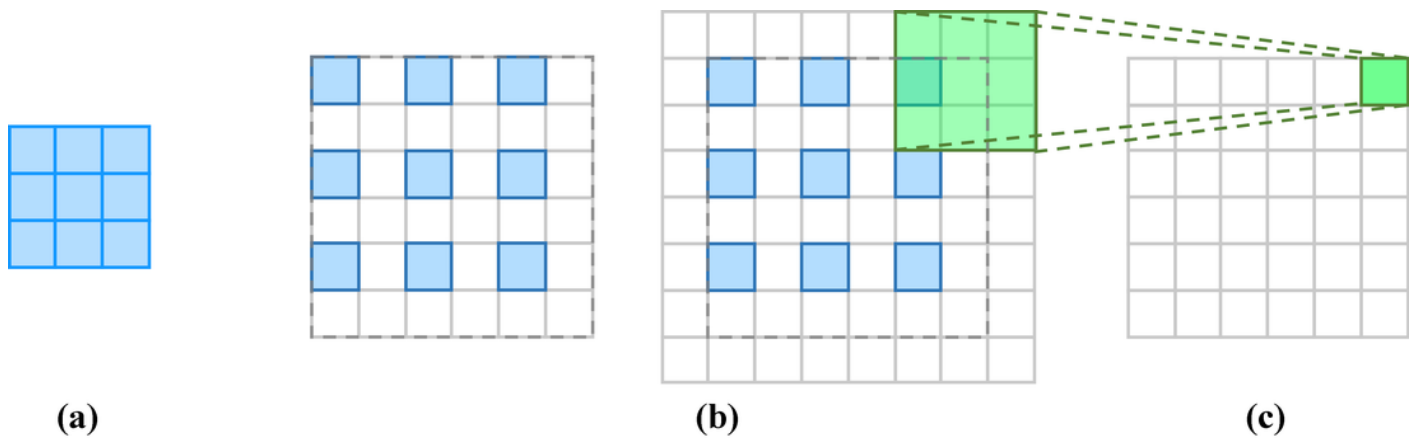



Figure 6

Deep-learning network structures for super-resolution.

(a) recursive learning, (b) residual learning, (c) dense connection-based learning, (d) multiscale learning, (e) advanced convolution-based learning, (f) attention-based learning

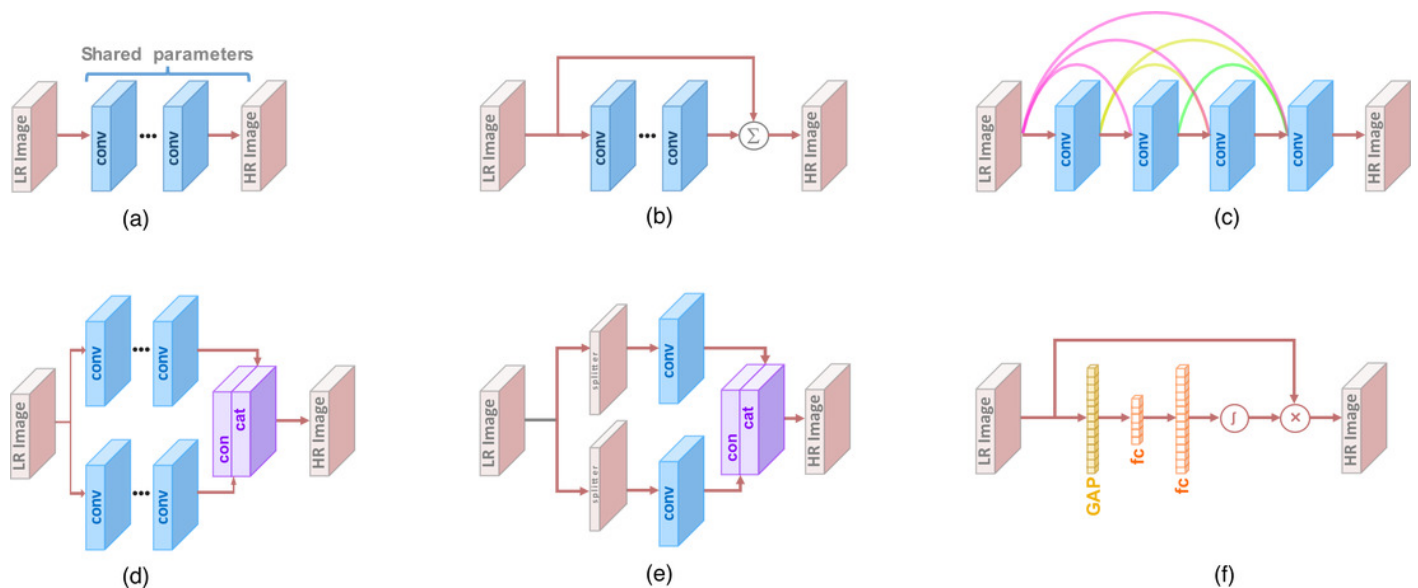



Figure 7

Pre-upsampling-based super-resolution network pipeline.

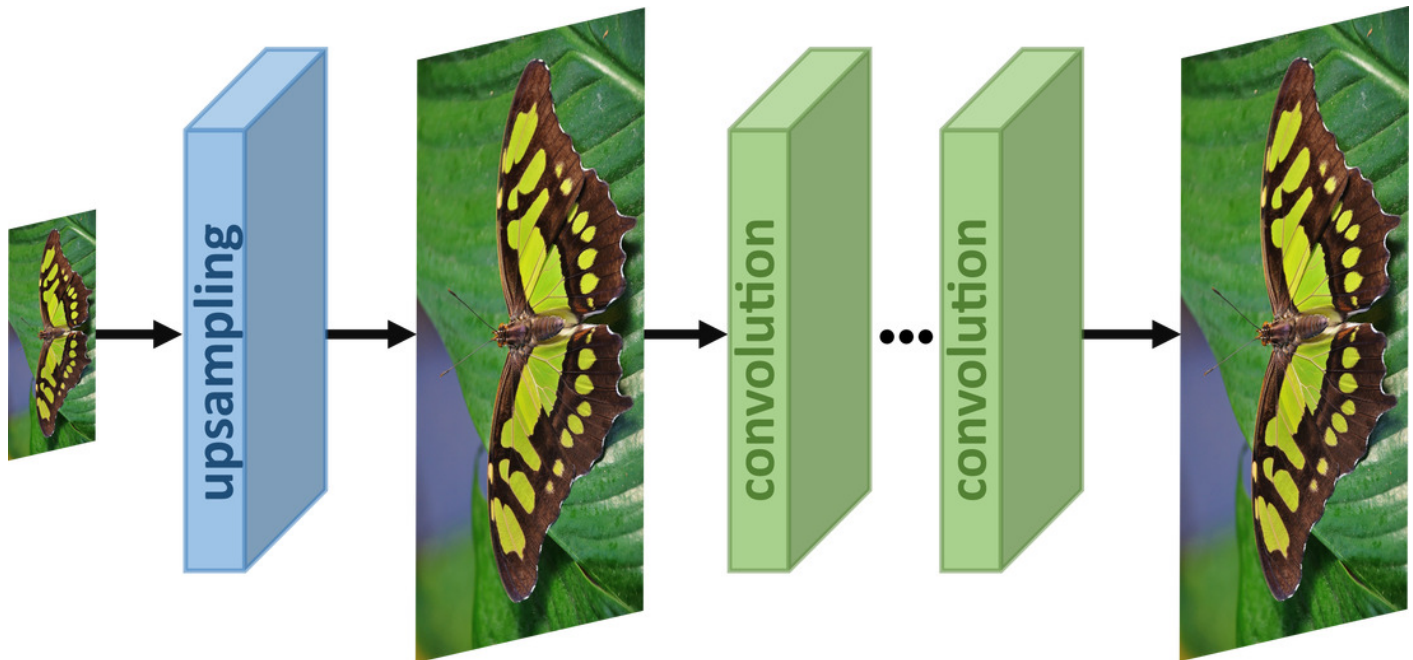



Figure 8

Post-upsampling-based super-resolution network pipeline.

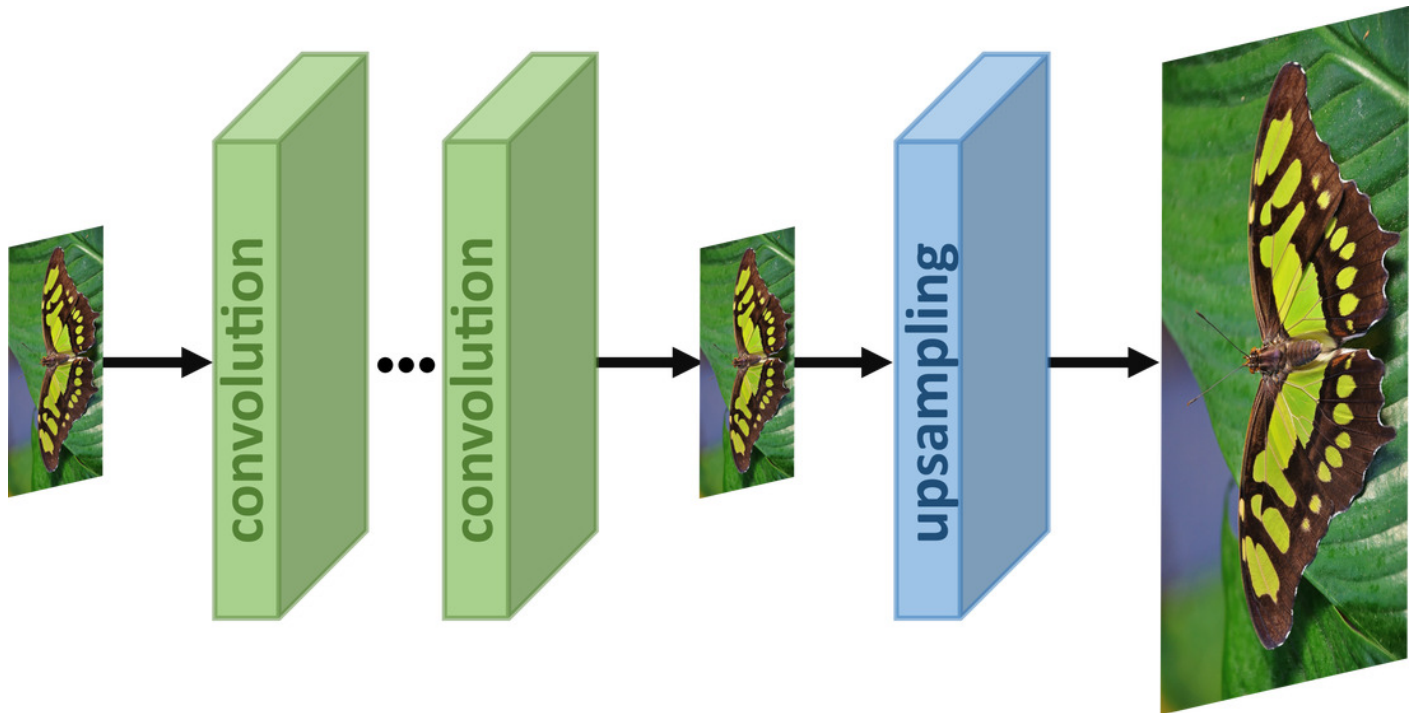



Figure 9

Iterative up-and-down sampling-based super-resolution network pipeline.

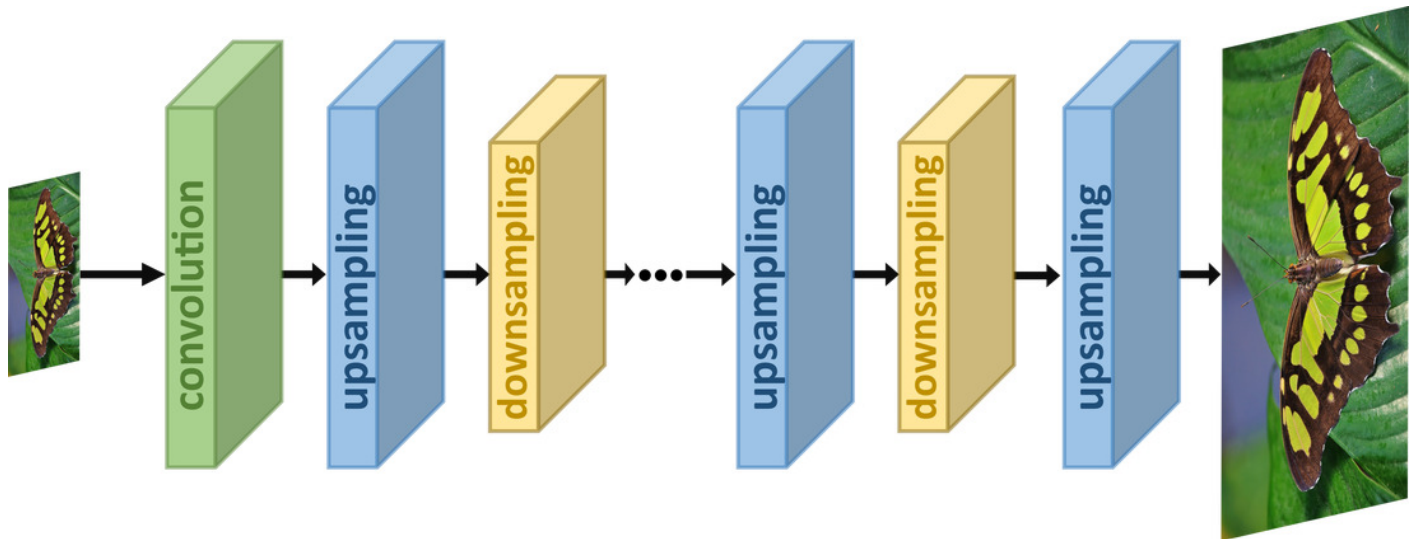



Figure 10

Progressive sampling-based super-resolution network pipeline.

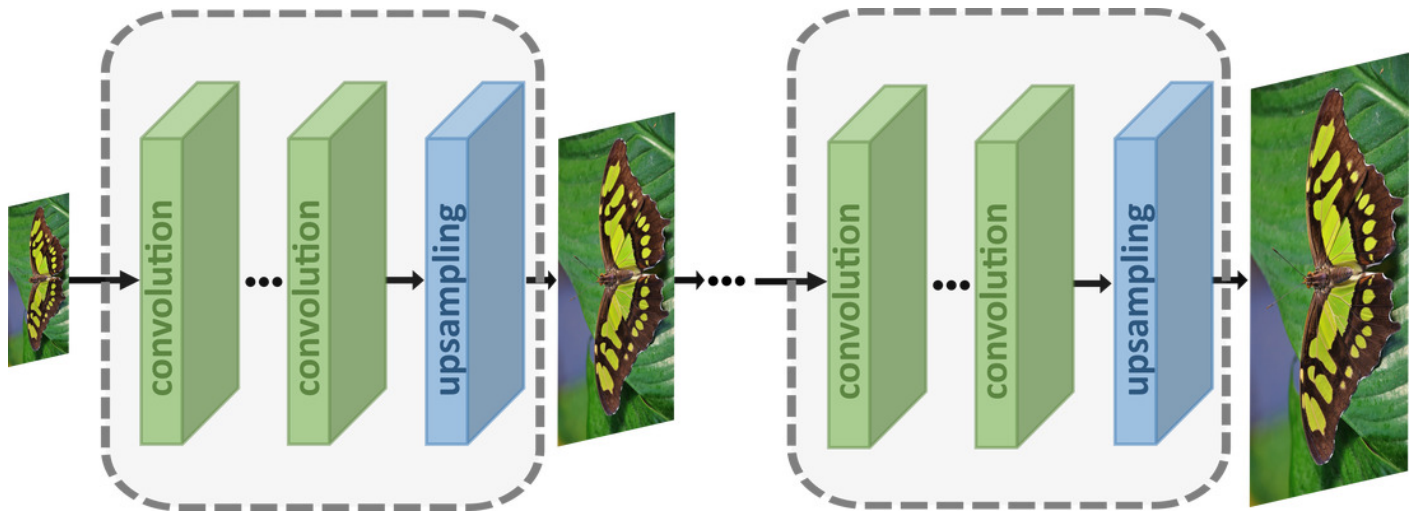
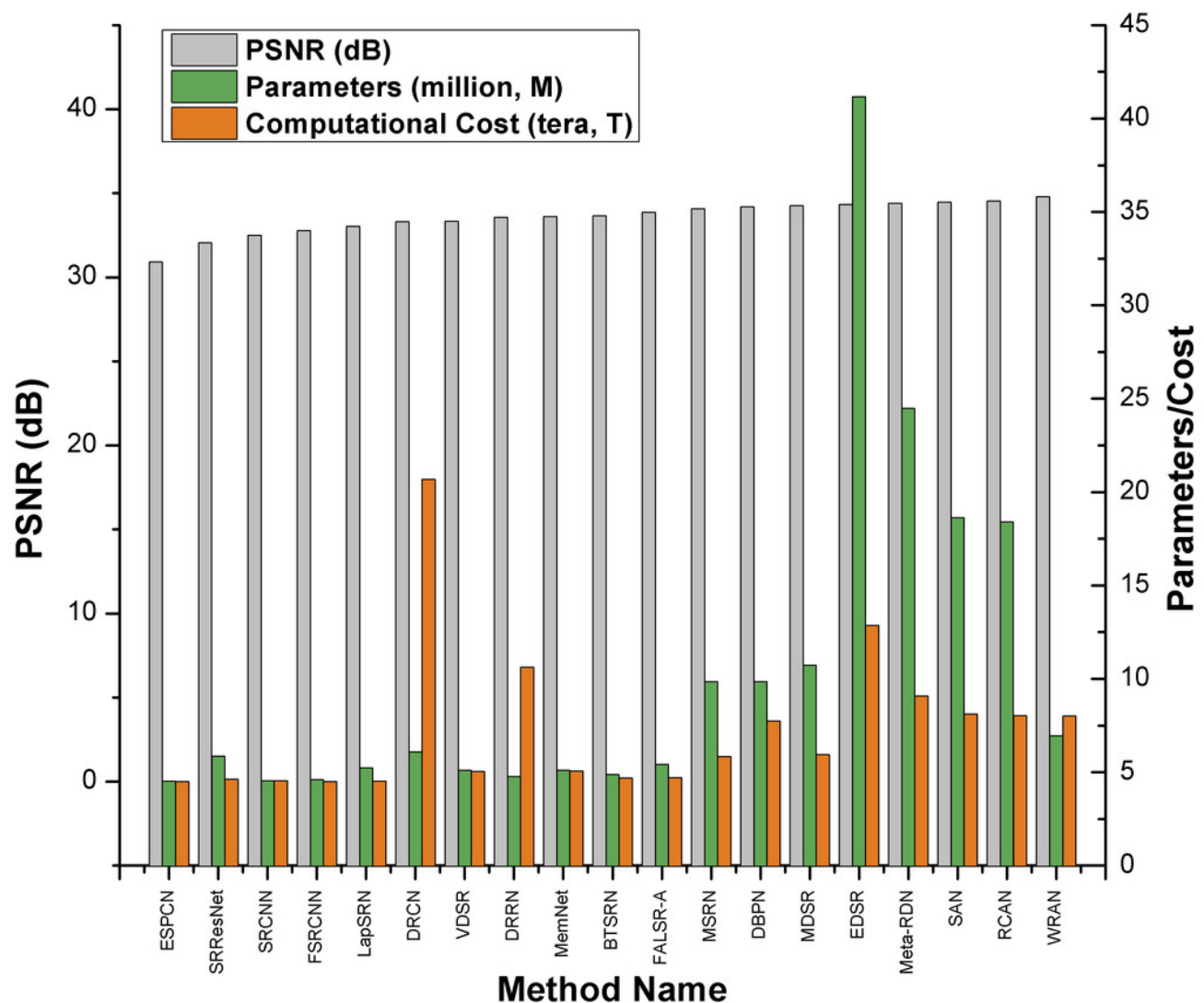



Figure 11

Benchmarking of super-resolution models.

Image quality index is represented by PSNR (in *blue* color), which is a significant evaluation indicator of any super-resolution method; the total number of parameters learned by every method is shown in *green*. The computational efficiency is measured in tera multiply-adds, and it is shown in *orange* color.

Comparison of image quality metrics for super-resolution.

Method	Strengths	Weaknesses
PSNR	• Most commonly used quality assessment metric; thus, it is easy to compare results with other methods. • Quantitative and is based on MSE	• Since this metric is pixel-based, the overall score could be misleading in some cases where two images could be visually different, but the PSNR would still be high. • This method does not give consider the structural information within the image
SSIM	• After PSNR, this metric is the most commonly used IQA metric; thus, comparing results with other methods is easier. • Quantitatively scores an image based on its structural similarity with the original image with the possibility to change the weights of luminance, contrast, and structural comparison.	• SSIM is unstable in cases where the variance or luminance of the reference image is low; thus, in medical imaging, this metric could give inconsistent results.
Opinion Scoring	• Opinion scoring is a subjective quality metric, and human testers grade image quality based on predefined parameters such as sharpness, color, and natural look. • This method is particularly suitable for human face reconstruction methods.	• Limitations include non-linear scoring among the testers, human error, and changes in test parameters. • Scoring takes much time, especially for large datasets
Perceptual Quality	• This method is similar to opinion scoring, but human testers are replaced by models that learn the behavior of testers using deep learning. • Very fast compared to opinion scoring.	• Requires additional resources for training the network to learn the features for the quality assessment network. • It depends on annotated datasets to learn human behavior.
Task-based Evaluation	• This metric is appropriate if the SR images are used to perform another task, for example, object detection/classification and diagnosis. • It helps in measuring the performance of the whole task, which uses SR images.	• Highly dependent on the performance of the associated task. • Same SR images will give different scores if there is a change in task parameters.

List of benchmark datasets used in super-resolution.

Name	Number of images/ pairs	Image format	Type	Resolution	Details of images
BSD100 (Martin et al., 2001)	100	PNG	Unpaired	(480, 320)	100 images of animals, people, buildings, scenic views etc.
BSDS300 (Martin et al., 2001)	300	JPG	Unpaired	(430, 370)	300 images of animals, people, buildings, scenic views, plants, etc.
BSDS500 (Arbeláez et al., 2010)	500	JPG	Unpaired	(430, 370)	Extended version of BSD 300 with additional 200 images
CelebA (Liu et al., 2015)	202,599	PNG	Unpaired	(2048, 1024)	Over 40 attribute defined categories of celebrities
DIV2K (Agustsson & Timofte, 2017)	1000	PNG	Paired	(2048, 1024)	Objects, People, Animals, scenery, nature
Manga109 (Fujimoto et al., 2016)	109	PNG	Unpaired	(800, 1150)	109 manga volumes drawn by professional manga artists in Japan
MS-COCO (Lin et al., 2014)	164,000	JPG	Unpaired	(640, 480)	Labeled objects with over 80 object categories
OutdoorScene (Wang et al., 2018)	10624	PNG	Unpaired	(550, 450)	Outdoor scenes including plants, animals, sceneries, water reservoirs, etc.
PIRM (Blau et al., 2018)	200	PNG	Unpaired	(600, 500)	Sceneries, people, flowers, etc.
Set14 (Zeyde, Elad & Protter, 2012)	14	PNG	Unpaired	(500, 450)	Faces, animals, flowers, animated characters, insects, etc.
Set5 (Bevilacqua et al., 2012)	5	PNG	Unpaired	(300, 340)	Only 5 images including, butterfly, baby, bird, head, and women.
T91 (Yang et al., 2010)	91	PNG	Unpaired	(250, 200)	91 images of fruits, cars, faces, etc.
Urban100 (Huang, Singh & Ahuja, 2015)	100	PNG	Unpaired	(1000, 800)	Urban buildings, architecture
VOC2012 (Everingham et al., 2014)	11,530	JPG	Unpaired	(500, 400)	Labelled objects with over 20 classes

Inclusion and exclusion criteria.

Section	Inclusion	Exclusion
Introduction	Methods that defined image interpolation and performed some practical form of image interpolation, i.e., super-resolution	• Studies that solely defined model • Review articles
Classical Methods	Methods that performed pixel, neighborhood, or any classical image interpolation	• Application research where applications of classical methods were discussed • Review papers
Deep learning-based methods	Development of image super-resolution using deep learning methods, including review papers	• Papers that emphasize video super-resolution as these papers give priority to frame per second (FPS) and inference time were not included
Applications	Direct applications of super-resolution methods in the six fields defined in Section 7 were included	• Applications that combined other methods with image super-resolution and SR was a limited part were not included • Review papers

Comparison of upsampling methods.

Method	Strengths	Weaknesses
Sub-pixel layer	• It uses convolution in an end-to-end manner, so it is frequently used in SR models. • This layer has a wide receptive field, which helps learn more contextual information	• This layer may generate some false artifacts at the boundaries of complex patterns due to its uneven distribution of the respective field. • Upscaling factor is fixed
Deconvolution layer	• This layer is most commonly used in SR methods, and it generates HR images in an end-to-end manner. • Compatible with vanilla convolution	• In some cases, due to uneven overlapping within the generated HR image, the patterns are replicated in a check-like format and may result in a non-realistic HR image. • Upscaling factor is fixed
Meta upscaling	• This method uses arbitrary scaling factors to generate the SR image. • Extracts more information from the LR feature maps, which helps to construct an HR image using meta upscaling in the last layer of the SR models, which makes this method an end-to-end SR approach	• The whole process may become unstable for high-scale factors as it predicts the convolution weights for every single pixel independent of the image information within those pixels.

SR method details of various SR algorithms

"US," "Rec.," "Res.," "Attent.," "Dense," "Pre.," "Post.," "Iter.," and "Prog." represent
upsampling methods, recursive learning, residual learning, attention-based learning, dense
connections, pre-upsampling framework, post-upsampling framework, iterative up-down
upsampling framework, and progressive upsampling framework respectively.

Year	Method Name	US	Network	Framework	Loss Function	Details
2014, ECCV	SRCNN (Dong et al., 2014)	Bicubic	CNN	Pre	L_{L2}	First deep learning-based SR
2016, CVPR	DRCN (Kim et al., 2016)	Bicubic	Res., Rec.	Pre	L_{L2}	Recursive layers
2016, ECCV	FSRCNN (Dong et al., 2016)	Deconv		Post	L_{L2}	Lightweight
2017, CVPR	ESPCN (Caballero et al., 2017)	Sub-pixel		Pre	L_{L2}	Sub-pixel
2017, CVPR	LapSRN (Lai et al., 2017)	Bicubic	Res.	Prog	L_{L1}, L_{PIX_CH}	Cascaded CNN
2017, CVPR	DRRN (Tai et al., 2017a)	Bicubic	Res., Rec.	Pre	L_{L2}	Recursive layer blocks
2017, CVPR	SRResNet (Ledig et al., 2017)	Sub-pixel	Res.	Post	L_{L2}	Content loss
2017, CVPR	SRGAN (Ledig et al., 2017)	Sub-pixel	Res.	Post	L_{GAN}	GAN-based loss
2017, CVPR	EDSR (Lim et al., 2017)	Sub-pixel	Res.	Post	L_{L1}	Compact design
2017, ICCV	EnhanceNet (Sajjadi et al., 2017)	Bicubic	Res.	Pre	L_{GAN}	GAN-based loss
2017, ICCV	MemNet (Tai et al., 2017b)	Bicubic	Res., Rec., Dense	Pre	L_{L2}	Memory layers blocks
2017, ICCV	SRDenseNet (Tong et al., 2017)	Deconv	Res., Dense	Post	L_{L2}	Fully connected layers
2018, CVPR	DBPN (Haris et al., 2018)	Deconv	Res., Dense	Iter	L_{L2}	Back-prop. Based
2018, CVPR	DSRN (Han et al., 2018)	Deconv	Res., Rec.	Pre	L_{L2}	Dual-state network
2018, CVPRW	ProSR, ProGanSR (Wang et al., 2018)	Progressive Upscale	Res., Dense	Prog	L_{LS}	Least square loss
2018, ECCV	MSRN (Li et al., 2018)	Sub-pixel	Res.	Post	L_{L1}	Multi-path
2018, ECCV	RCAN (Zhang et al., 2018)	Sub-pixel	Res., Attent.	Post	L_{L1}	Attention-based loss
2018, ECCV	ESRGAN (Wang et al., 2019)	Sub-pixel	Res., Dense	Post	L_{L1}	GAN-based loss
2019, CVPR	Meta-RDN (Hu et al., 2019)	Meta Upscale	Res., Dense	Post	L_{L1}	Multi-scale model
2019, CVPR	Meta-SR (Hu et al., 2019)	Meta Upscale	Res., Dense	Post	L_{L1}	Arbitrary scale factor as input
2019, CVPR	RBPN (Haris et al., 2019)	Sub-Pixel	Rec.	Post	L_{L1}	Used SISR and MISR together for VSR
2019, CVPR	SAN (Dai et al., 2019)	Sub-Pixel	Res., Attent.	Post	L_{L1}	2nd order attention
2019, CVPR	SRFBN (Li et al., 2019)	Deconv	Res., Rec., Dense	Post	L_{L1}	Feedback path
2020, Neuro-computing	WRAN (Xue et al., 2020)	Bicubic	Res., Attent.	Pre	L_{L1}	Wavelet-based