Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A review of deep-learning-based super-resolution: From methods to applications

Hu Su [a,b,1], Ying Li [a,1], Yifan Xu [b], Xiang Fu [b], Song Liu [b,c,*]

[a] The State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[b] The School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[c] Shanghai Engineering Research Center of Intelligent Vision and Imaging, Shanghai, China

ARTICLE INFO

Keywords:
Deep learning
Super-resolution
Single image super-resolution
Multiple image super-resolution
Degradation model

ABSTRACT

Super-resolution (SR), aiming to super-resolve degraded low-resolution image to recover the corresponding high-resolution counterpart, is an important and challenging task in computer vision, and with various applications. The emergence of deep learning (DL) has significantly advanced SR methods, surpassing the performance of traditional techniques. This paper presents a comprehensive survey of DL-based SR methods encompassing single image super resolution (SISR) and multiple image super resolution (MISR) methods, along with their applications and limitations. In SISR methods, addressing individual images independently, we review blind and non-blind SR methods. Additionally, within MISR, we delve into multi-frame, multi-view, and reference-based SR methods. DL-based SR methods are categorized from the application perspective and a taxonomy is proposed. Finally, we present research prospects and future directions.

Contents

1. Introduction 2
2. Outline 2
3. Single image super-resolution methods 3
3.1. Non-blind single image super-resolution methods 3
3.1.1. Non-blind SISR with single degradation model 3
3.1.2. Non-blind SISR with multiple degradation models 3
3.2. Blind single image super-resolution methods 3
3.2.1. Blind SISR with degradation modeling under the ground-truth supervision 3
3.2.2. Blind SISR with degradation modeling without the ground-truth supervision 3
3.2.3. Blind SISR with implicit degradation modeling using unpaired images 5
4. Discussion on SISR methods 5
4.1. How to increase resolution during SISR? 5
4.2. What role does degradation model play in SISR? 5
4.3. How to incorporate degradation model into the input of CNN? 6
5. Multiple images super-resolution methods 6
5.1. Multi-frame super-resolution methods 6
5.1.1. Non-blind MFSR with explicit alignment 6
5.1.2. Blind MFSR with explicit alignment 6
5.1.3. Non-blind MFSR with implicit alignment 7
5.2. Multi-view super-resolution methods 7
5.3. Reference-based image super-resolution methods 7
6. Discussion on MISR methods 8
6.1. How multiple inputs improve SR performance compared with single input? 8
6.2. How multiple inputs are combined and used in different scenarios? 8

[*] Corresponding author at: The School of Information Science and Technology, ShanghaiTech University, Shanghai, China.
E-mail addresses: hu.su@ia.ac.cn (H. Su), liusong@shanghaitech.edu.cn (S. Liu).

[1] The first two authors contributed equally.

https://doi.org/10.1016/j.patcog.2024.110935
Received 2 April 2024; Received in revised form 22 August 2024; Accepted 22 August 2024
Available online 24 August 2024

0031-3203/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

6.3. How to obtain the reference image in RefSR? 8
7. Datasets, metrics, and performance 9
7.1. Datasets 9
7.2. Metric and performance 9
8. Applications of SR 9
8.1. Applications of SISR methods 9
8.2. Applications of MVSR methods 9
8.3. Applications of MFSR methods 9
9. Future works and conclusions 9
9.1. Future works 9
9.1.1. Feasible degradation modeling 9
9.1.2. Assessment of SR results 9
9.1.3. Multiple images acquisition with robot motion 10
9.1.4. Motion process SR under physical guidance 10
9.2. Conclusions 11
CRediT authorship contribution statement 11
Declaration of competing interest 11
Data availability 11
Acknowledgments 11
References 11

1. Introduction

Super-resolution (SR) aims to recover a high-resolution (HR) image from a low-resolution (LR) input. Although imaging devices have simplified visual data acquisition, the resulting images often suffer from blurring and blocky effects due to degradations introduced by the devices, image processing algorithms, and image compression [1]. SR has long been a fundamental problem in computer vision with practical applications across various fields. Early methods [2,3] involving explicit HR image reconstruction models were inadequate for real-world scenarios. With the rise of deep learning (DL), DL-based SR methods have shown great potential in addressing SR tasks. Dong et al. [4] pioneered the use of a super-resolution convolutional neural network (SRCNN) for single image super-resolution, marking a breakthrough in the field. SRCNN autonomously learns an end-to-end mapping between LR and HR images, outperforming traditional methods by margins of 0.15, 0.17, and 0.13 dB in terms of peak signal-to-noise ratio (PSNR) across three diverse datasets, i.e., Set5 [5], Set14 [6] and BSD200 [7]. However, SRCNN requires paired LR-HR images for training and involves extra pre/post-processing, which limits its practicality. Since then, DL-based SR methods have been extensively studied.

SR plays a crucial role as a pre-processing step, enhancing the performance of downstream tasks especially for low-resolution images [8]. Researchers made the attempts to improve the performances of low-quality [9] or small object detection [10] and classification [11] with the utilization of SR. Thus, focusing on SR's diverse applications, rather than just HR reconstruction accuracy, is essential. However, previous works and surveys have concentrated on the methods and the reconstruction performance on synthetic data, with little discussion on real-world applications. This paper reviews DL-based SR methods from recent years with an emphasis on their practicability, to help readers understand their application scopes and limitations.

Existing SR solutions include both single-image (SISR) [1,12] and multiple-image SR (MISR) [13]. SISR processes each image independently, while MISR utilizes relationships between multiple images for super-resolution. Early research focused on SISR, and MISR is gaining attention due to the growing popularity of multimedia data. This paper introduces a framework dominated by SISR and MISR to offer a comprehensive survey of recent advances in DL-based SR methods.

Several surveys on DL-based SR have been published [1,12–14]. Surveys [1,12,14] provide overviews of SISR methods. The former [1] focuses primarily on the blind SR methods while the latter two [12,14] introduces both non-blind and blind methods. The recent survey [13] reviews the progress on video super-resolution methods. The related methods as described in [13] are referred to as multi-frame super-resolution (MFSR), which are categorized as a type of MISR method. However, these surveys often focus on technical advances in neural network architectures and optimization objectives, with less emphasis on experimental settings and applications. Additionally, these surveys typically cover only a portion of SR methods, leading to potential confusion for readers. This paper reviewed representative DL-based SR methods in both SISR and MISR topics, categorizing them by method characteristics and discussing the experimental settings and practicability, while technical advances are only briefly introduced. A taxonomy is established and further research directions are discussed. The contributions of this paper are: (1) a comprehensive review of DL-based SR methods, including the latest developments, (2) categorization of methods from an application perspective, providing a taxonomy to guide method selection, and (3) a focus on experimental settings and practicality, with visual analysis and discussions on key points, offering prospects for future research.

2. Outline

Super-resolution (SR) has been a longstanding challenge with significant research efforts. As depicted in Fig. 1, the development of SR can be divided into two periods: non-deep and deep learning. The key turning point was the introduction of SRCNN by Dong et al. in 2014 [4], which combined deep learning with SR. In 2019, Gu et al. [15] introduced the blind SISR method, iterative kernel correction (IKC). Guo et al. [16] proposed the MFSR method which leverage inter-frame information for better HR restoration, marking the emergence of MISR methods. Other notable MISR methods, such as reference-based SR (RefSR) method [17] and multi-view SR (MVSR) [18] method, are proposed. While achieving superior performance, MISR methods have become the significant ideology for SR.

As illustrated in Fig. 1, a taxonomy classifies DL-based SR methods based on input and experimental settings, highlighting their application scenarios. DL-based SR methods are categorized into SISR (Section 3) and MISR (Section 4). SISR includes non-blind methods with known degradation models and blind methods with unknown models. MISR methods are divided into MFSR, MVSR, and RefSR, depending on input image relationships. Fig. 1 provides typical implementations of the methods with the annotations. Applications of the SR methods are discussed (Section 8) and future directions of SR are presented in the survey (Section 9).

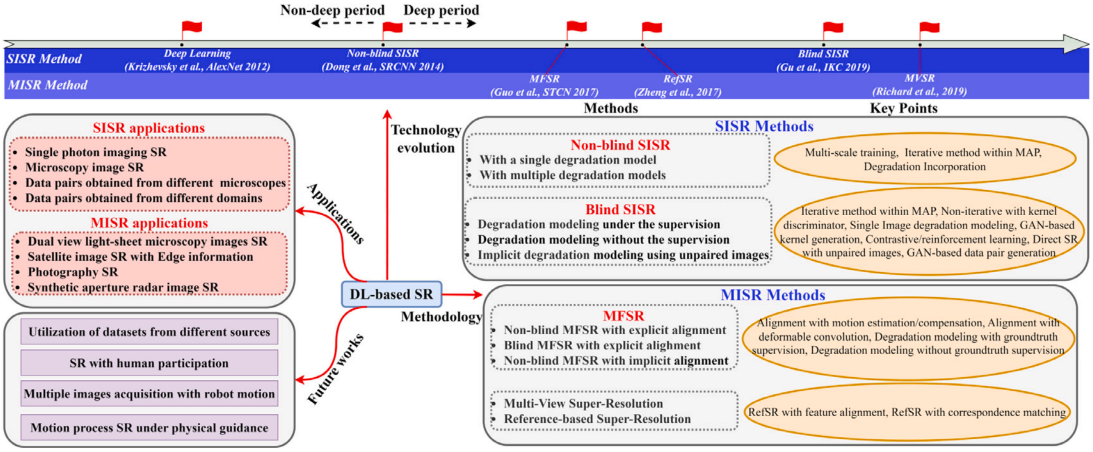
Non-deep period
Deep period
SISR Method
MISR Method
Deep Learning
(Krizhevsky et al., AlexNet 2012)
Non-blind SISR
(Dong et al., SRCNN 2014)
MFSR
(Guo et al., STCN 2017)
RefSR
(Zheng et al., 2017)
Blind SISR
(Gu et al., IKC 2019)
MVSR
(Richard et al., 2019)
Methods
Key Points
SISR applications
Single photon imaging SR
Microscopy image SR
Data pairs obtained from different microscopes
Data pairs obtained from different domains
MISR applications
Dual view light-sheet microscopy images SR
Satellite image SR with Edge information
Photography SR
Synthetic aperture radar image SR
Applications
Technology evolution
DL-based SR
Methodology
Future works
Utilization of datasets from different sources
SR with human participation
Multiple images acquisition with robot motion
Motion process SR under physical guidance
SISR Methods
Non-blind SISR
With a single degradation model
With multiple degradation models
Multi-scale training, Iterative method within MAP, Degradation Incorporation
Blind SISR
Degradation modeling under the supervision
Degradation modeling without the supervision
Implicit degradation modeling using unpaired images
Iterative method within MAP, Non-iterative with kernel discriminator, Single Image degradation modeling, GAN-based kernel generation, Contrastive/reinforcement learning, Direct SR with unpaired images, GAN-based data pair generation
MISR Methods
MFSR
Non-blind MFSR with explicit alignment
Blind MFSR with explicit alighment
Non-blind MFSR with implicit alignment
Alignment with motion estimation/compensation, Alignment with deformable convolution, Degradation modeling with groundtruth supervision, Degradation modeling without groundtruth supervision
Multi-View Super-Resolution
Reference-based Super-Resolution
RefSR with feature alignment, RefSR with correspondence matching

Fig. 1. Outline of our survey, including evolution, methodology, future works, and applications of SR methods.

3. Single image super-resolution methods

Regarding the LR image as the degradation of its HR counterpart, SISR methods devote to accomplishing the inverse of the degradation process. Degradation modeling plays a central role in the construction of the HR-LR mapping. In most cases, the degradation is the superposition of blurring, down-sampling, and noise, which is represented as:

$$y = (x \otimes k) \downarrow_s + n \tag{1}$$

where, x is the HR image, y is the LR image, s is the down-sampling factor, k is the blur kernel, n is the noise, $\otimes$ is the convolution operation. Degradation modeling is then to determine the parameters k and n. The uncertainty on degradation makes SISR to be an ill-posed problem (e.g., multiple-to-multiple mapping). By identifying the degradation process, SISR is further transformed into a one-to-one mapping. SISR methods solve the inverse to reconstruct the HR.

3.1. Non-blind single image super-resolution methods

3.1.1. Non-blind SISR with single degradation model

The DL-based SR technique originated with SRCNN [4], a non-blind SISR method using three CNN layers for patch extraction, feature representation, non-linear mapping, and HR reconstruction. Key developments include very deep super-resolution (VDSR) [19], which employs a residual learning strategy to enhance training convergence and accuracy, and Laplacian pyramid super-resolution network (LapSRN) [20], known for large-scale iterative up-sampling with supervised residuals. For greater detail at high magnifications, diffusion models (DM) like the denoising diffusion probabilistic model used in super-resolution via repeated refinement (SR3) [21] improve image-to-image translation. Subsequent advancements have enhanced diffusion model efficiency by operating in residual [22,23] or latent spaces [24], optimizing convergence and computational efficiency. These methods rely on a singular, fixed degradation model, restricting their effectiveness in varied real-world applications.

3.1.2. Non-blind SISR with multiple degradation models

This study extends the investigation to multiple degradations. VDSR [19] implements multi-scale training to handle various downscale factors. The unfolding super-resolution network (USRNet) [25] and the deep plug-and-play super-resolution (DPSR) [26] address SISR within the MAP framework, promoting iterative methods to manage different blur kernels and downscale factors in Eq. (1), as depicted in Fig. 2(a). These methods, however, face challenges with LR images that have multiple or combined degradations. To tackle this, degradation information is encoded and integrated into the input for one-to-one mapping [27,28]. Wang et al. [29] introduced the spatial feature transform (SFT) to adjust feature maps based on degradation. While effective for multiple degradations, these methods depend on synthesized HR-LR pairs and assume a known degradation process. Non-blind methods are not the focus of the survey. Only a brief introduction is presented to ensure comprehensiveness.

3.2. Blind single image super-resolution methods

Blind methods are more practical than non-blind ones as they do not rely on a predefined degradation process. A key aspect is degradation estimation, learned during training, and used during testing for SISR. Blind SISR methods are classified into those with and without ground-truth supervision, and those using implicit modeling. The first two combine degradation estimation with a non-blind SISR method, differing in learning strategies, while the latter learns degradation implicitly.

3.2.1. Blind SISR with degradation modeling under the ground-truth supervision

With a requirement of the ground-truth degradation models, degradation estimation is learned in a supervised manner. The MAP framework motivates iterative pipelines, differentiating between non-blind and blind settings, as shown in Fig. 2. In Eq. (1), there are three variables, i.e., x, k and n, to be determined in blind SISR. By using a denoise algorithm, blind SISR only needs to focus on solving x and k. The deep alternating network (DAN) [30] employs two CNNs – restorer and estimator – updated iteratively for blind SISR. [38] improves DAN by iterating in feature space. The IKC method [15] uses a predictor to estimate blur kernels from LR images and a corrector to refine them before generating HR images. The kernel-oriented adaptive local adjustment (KOALAnet) [31], a non-iterative method, enhances SR by estimating degradation kernels with a discriminator network, as illustrated in Fig. 3(a). Other non-iterative methods [39] estimate reformulated degradation models, but KOALAnet excels at handling spatially-variant degradations. While MAP-based methods are limited by the need for ground-truth degradation models, KOALAnet addresses this with a predefined kernel space. However, practical concerns remain regarding the complexity, additional time costs, and modeling of complex degradation processes with finite kernels.

3.2.2. Blind SISR with degradation modeling without the ground-truth supervision

These methods eliminate the need for ground-truth degradation models by extracting internal statistics from a single image for degradation estimation. Michaeli and Irani [40] suggest that the optimal SR-kernel maximizes patch similarity across scales. Based on the priori,

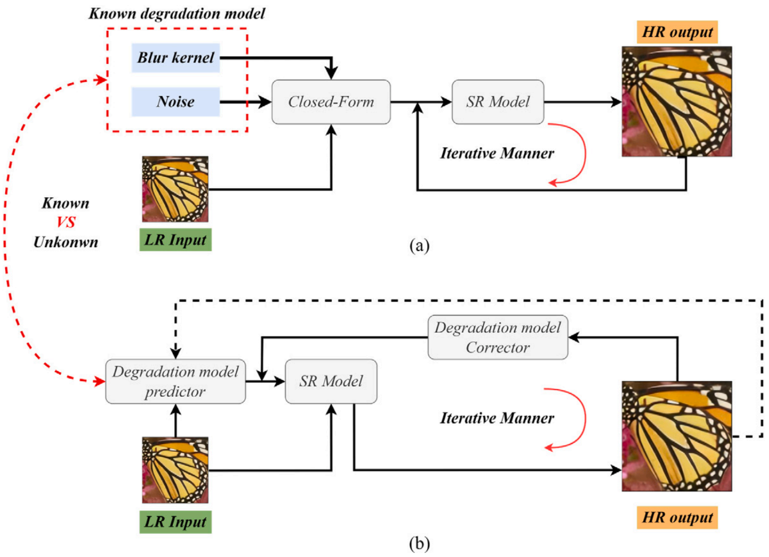

Fig. 2. Iterative pipelines of MAP-based SISR methods in (a) non-blind and (b) blind settings. For example, USRNet [25] and DPSR [26] are non-blind SISR methods that adopt the pipeline in (a). DAN [30] is a blind SISR method that adopt the pipeline in (b).

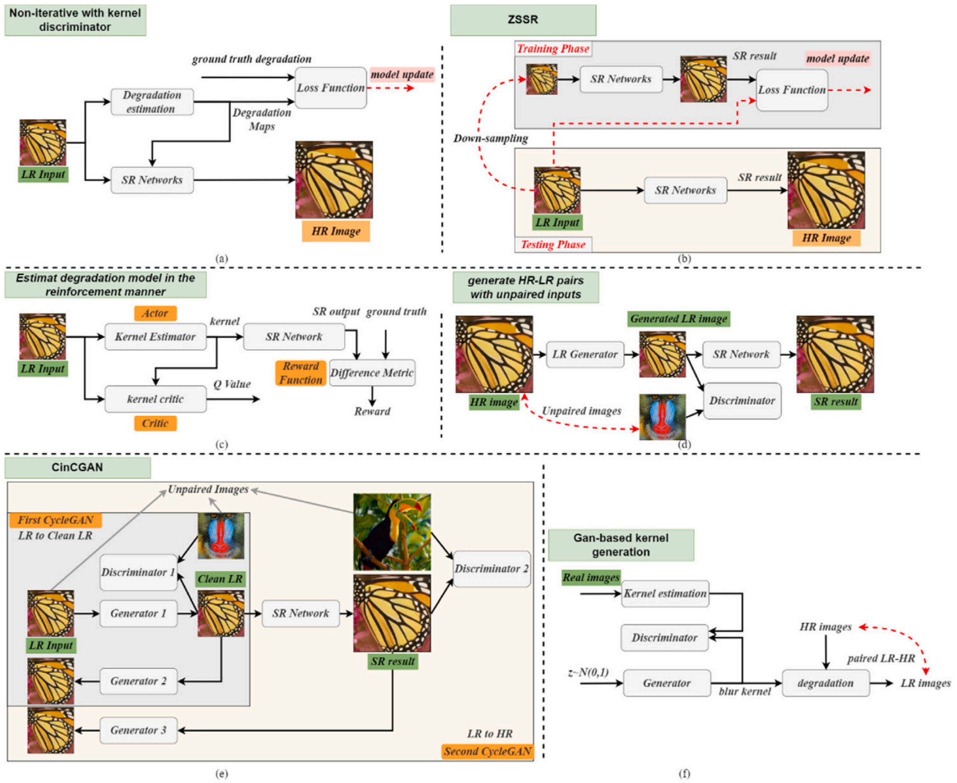

Fig. 3. Frameworks of blind SISR methods. (a) non-iterative methods with kernel discriminator, KOALAnet [31]. The reconstruction loss in KOALAnet is not indicated in the figure for simplicity. (b) ZSSR [32], input image and the down-scaling image are used to learn image-specific relation. (c) AMNet-RL [33], the kernel estimation is optimized by reinforcement learning. (d) and (e) indicate implicit degradation modeling using unpaired images. (d) generate HR-LR pairs with unpaired inputs. (e) CinCGAN [34] decouples the denoising and up-sampling processes. (f) Zhou et al. [35], kernel generation using the GAN framework.

the zero-shot super-resolution (ZSSR) method [32], and similar self-supervised approaches [41], use the input image alone to create HR-LR pairs for training and result prediction, as shown in Fig. 3(b). KernelGAN [36] estimates degradation models based on patch similarity, serving as a plug-and-play kernel estimator combined with ZSSR, while Liang et al. [42] improved this approach with a normalizing flow-based kernel prior. Without direct supervision, degradation models can be estimated through methods like reinforcement or self-supervised learning, with approaches varying in assumptions about degradation consistency within and across images, employing strategies like contrastive [37,43] and metric learning [44] to refine SR models. Another method [45] involves estimating the blur kernel with self-supervised learning, relying solely on the LR input. In the adaptive modulation network with reinforcement learning (AMNet-RL) [33], the kernel estimation is optimized by reinforcement learning (Fig. 3(c)). AMNet-RL incorporates perceptual performance into the optimization process,

Table 1
Summary of Single Image Super-Resolution (SISR), methods, settings and practicability.

Method	*Setting*	*Validation dataset*	*Application scenario*	*Practicability description*
SRCNN [4]	Paired HR-LR images with a known degradation model	Synthetic image	Predefined degradation model	Impractical
VDSR [19], USRNet [25], DPSR [26], [27–29]	Paired HR-LR images with known degradation models	Synthetic image	Predefined multiple degradation models	Impractical
DAN [30], IKC [15], KOALAnet [31]	Paired HR-LR images with known degradation models only for training	Synthetic image	Predefined multiple degradation models	Impractical
ZSSR [32], KernelGAN [36]	LR images	Real image	Corresponding patches with similar distributions	Practical
DASR [37]	LR images	Synthetic image	Same degradation model within each image while diverse models among different images	Practical for special cases
AMNet-RL [33]	LR images	Synthetic image	Quantified perceptual performance	Practical for special cases
[35]	Unpaired HR-LR images	Real image	Degradation process maintaining high-frequency details	Practical
DM-based methods [21–24]	Paired HR-LR images with a known degradation model	Synthetic image	Predefined degradation model	Impractical

facing challenges posed by complex degradations and the difficulty in evaluating SR performance. The methods mentioned in this section offer increased practicality. However, each has specific limitations, as summarized in Table 1.

3.2.3. Blind SISR with implicit degradation modeling using unpaired images

SISR with unpaired images is challenging but is close to real-world setting. Generative adversarial networks (GANs) offer a feasible solution by capturing latent information. As shown in Fig. 3(e), the cycle-in-cycle GAN (CinCGAN) [34] decouples denoising and up-sampling using two CycleGANs. However, training such a two-stage GAN-based method can be challenging due to its complex network structure. A more lightweight method was proposed by Liu et al. [46], using an invertible neural network (INN) to handle degradation and SR as reverse processes.

Other methods shown in Fig. 3(d) generate HR-LR pairs with unpaired inputs using powerful generators. Bulat et al. [47] designed a unified framework with two GANs—one for learning the degradation model and another for training SR with paired image. The domain gap exists between generated LR and real LR images. To alleviate this problem, unlabeled real LR images are incorporated into SR model training for domain adaption [48]. Fritsche et al. [49] introduced a down-sample GAN (DSGAN) to generate LR images that match source characteristics, while Zhou et al. [50] further improved domain transformation with a color-guided network. Despite their potential, these methods face limitations in real-world scenarios, such as the need for large-scale data and challenges in GAN training and convergence.

4. Discussion on SISR methods

The methods reviewed in Section 3 are summarized in Table 1, which categorizes their feasibility as "impractical", "practical for specific cases", or "practical". Methods like SRCNN, which rely on paired HR-LR images with known degradation models, are deemed "impractical" because these conditions are rarely met in real-world scenarios, and validation is done only on synthetic datasets. In contrast, methods like ZSSR and DASR, which make more realistic assumptions and are validated on real images, are considered "practical for specific cases" but still have limited applicability. Methods designed for broader real-world use are labeled "practical".

Note that this is not a traditional classification but rather a description of method practicability. It does not reflect the method's application extent or performance but instead indicates how well it aligns with real-world scenarios and its potential for practical use. This section aims to elucidate the fundamental principles of SISR methods by highlighting key aspects.

4.1. How to increase resolution during SISR?

Early SISR methods [20,51] use interpolation to increase image resolution, followed by deep CNNs to recover details. Later methods replaced interpolation with learning-based techniques, like transposed convolution and sub-pixel layers [14], enabling end-to-end training. Transposed convolution increases resolution by expanding the image with zeros and applying convolution, while the sub-pixel layer upsamples by generating multiple channels through convolution and reshaping them. Within the sub-pixel layer, a convolution is firstly applied for producing outputs with s^2 times channels, where s is the scaling factor. Assuming the input size is $h \times w \times c$, the output size will be $h\times w\times s^2c$. After that, the reshaping operation is performed to produce outputs with size $sh\times sw\times c$. The sub-pixel layer has a larger receptive field than the transposed convolution layer. Meanwhile, due to the reshaping operation in the sub-pixel layer, blocky regions in the output feature/image share the same receptive field, resulting in artifacts near the boundaries of different blocks.

4.2. What role does degradation model play in SISR?

Accurate degradation model estimation is crucial for SR model performance; errors can drastically reduce results. Gu et al. [15] noted that SR models are highly sensitive to estimation errors, as shown in Fig. 4, where kernel mismatches lead to over-smoothing or ringing artifacts. Only the correct blur kernel produces natural results. Non-blind SISR methods use identified models to generate paired HR-LR images or guide the SR process, while blind methods estimate the

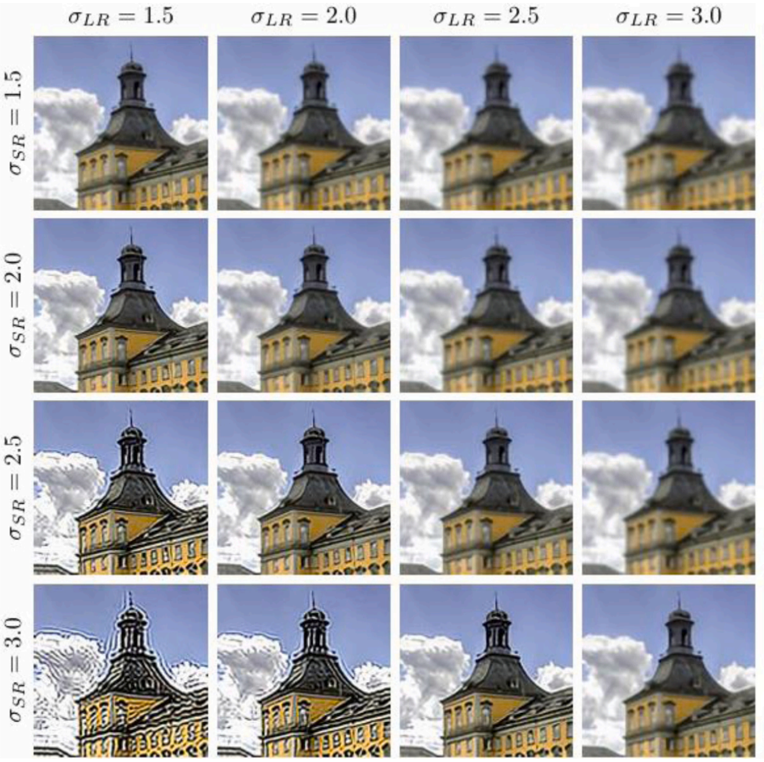

Fig. 4. SR sensitivity to the kernel mismatch [15]. δ_{LR} denotes the kernel used for downsampling and δ_{SR} denotes the kernel used for SR. In the upper-right region, $\delta_{SR} < \delta_{LR}$, which means that the kernel used for SR is smoother than the real one. In the lower-left region, $\delta_{SR} > \delta_{LR}$, which means that the kernel used for SR is sharper than the real one. In the diagonal, correct blur kernels are used.

degradation model explicitly or implicitly. Estimation approaches include supervised [15,30,31], self-supervised [37], reinforcement [33], and adversarial [35] methods, but degradation estimation is ill-posed. As mentioned in [1], different LR inputs may correspond to the same HR and vice versa. Degradation estimation is still an open problem especially in real application scenarios.

4.3. How to incorporate degradation model into the input of CNN?

There are primarily two ways to incorporate degradation model. The first [15] is to directly concatenate degradation maps with the feature maps. The second [28,31] is to transform the feature maps with the parameters learned as the representation of the degradation. The first deals with more complex degradation models, especially for those which vary in different parts of an image. However, the stretched degradation maps are not the real images, and they do not include image information. Thus, the first method would bring unsuspected noise. In the second method, transformation layers work similar to the Batch Normalization (BN) layer. The second method is learning-based and is particularly suited for deeper SR models.

5. Multiple images super-resolution methods

5.1. Multi-frame super-resolution methods

In MFSR methods, multiple continuous video frames y_{t-N}, …, y_{t-1}, y_t, y_{t+1}, y_{t+N} are captured where N is the time radius. MFSR aims to reconstruct the HR frame x_t at time t with the LR frames. Similar to SISR, the degradation process is modeled, which is represented as [2,3,61]

$$y_{t+i} = SKF_{t\to t+i}x_t + n_{t+i}, i \in [-N, N] \tag{2}$$

where, S represents the down-sampling scale, K represents blur operation, F is the warping operation with the motion from x_t to x_{t+i}, n_{t+i} is the noise.

MFSR typically involves an alignment procedure to extract spatial–temporal information. This process aligns features across frames by choosing a reference frame, extracting features from each frame, calculating the transformations, and applying them to achieve alignment. The aligned frames are used for feature extraction and HR reconstruction. MFSR methods are categorized into the following groups.

5.1.1. Non-blind MFSR with explicit alignment

In non-blind MFSR, known degradation models generate synthetic HR-LR pairs, necessitating precise frame alignment. The motion estimation and motion compensation (MEMC) pipeline estimates and aligns inter-frame motion as shown in Fig. 5(a). While traditional MEMC uses optical flow [52,62] for motion estimation and bilinear interpolation for compensation, newer approaches employ CNN architectures like recurrent convolutional network [63], ConvLSTM [64], and bidirectional recurrent networks [65,66] to handle MEMC and SR simultaneously. However, MEMC methods can struggle with motion inaccuracies, especially under significant motion or lighting changes, impacting SR quality. Learning-based deformable convolution methods [67,68] address these issues by adapting receptive fields, though they require more computational power. Generally, the constraints of non-blind settings limit their practicality in real-world applications.

5.1.2. Blind MFSR with explicit alignment

Blind MFSR reconstructs HR images without known degradation models. Traditional approaches use priors [2,3], whereas deep-learning methods [69] employ CNN architectures to estimate degradation. While traditional methods [2] assume a consistent blur kernel across frames, advanced techniques [53,54] estimate variable kernels per frame using CNNs in a supervised manner, as shown in Fig. 5(b). Based on the estimated blur kernel, [53] allows for the construction of intermediate latent HR images, enhancing HR restoration. Methods like [70] combine blur estimation [30] with non-blind techniques [67] to implement blind MFSR without needing kernel ground-truth. Additionally, GANs

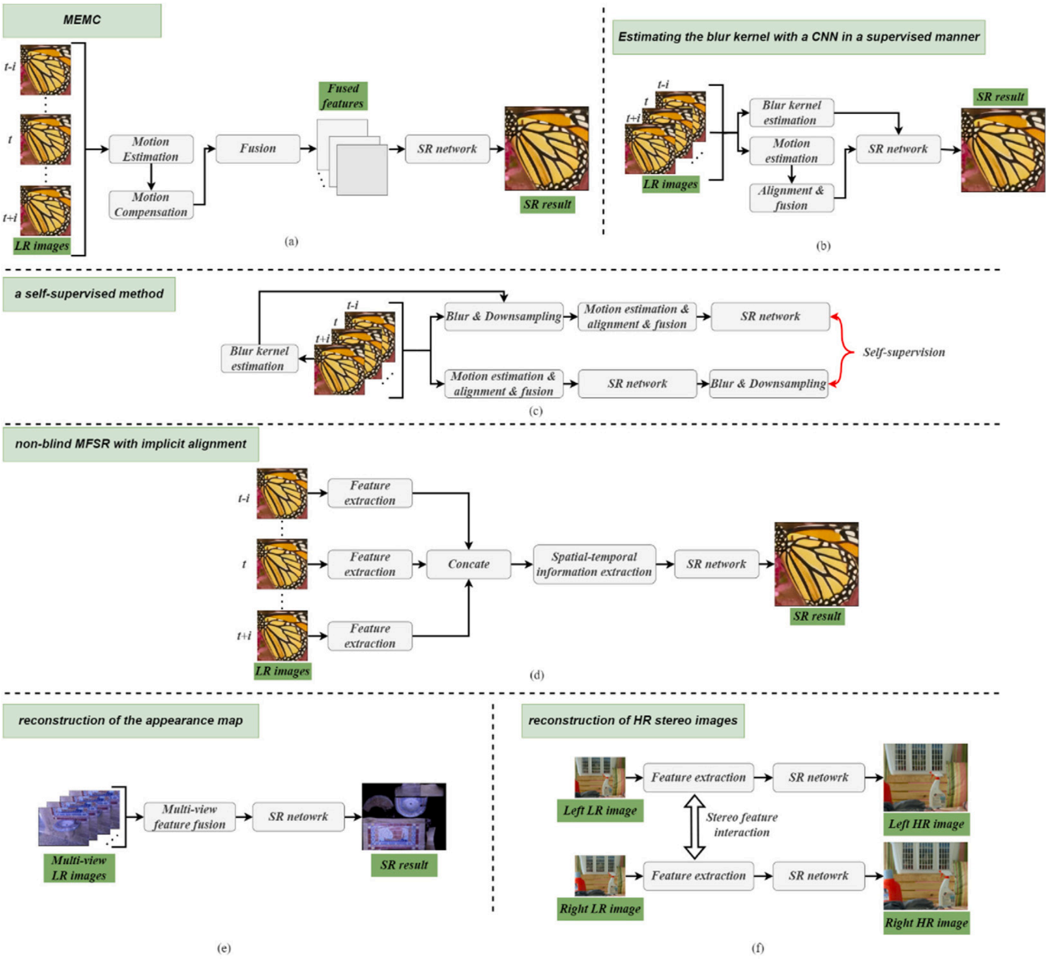

Fig. 5. The frameworks of the MISR methods. (a) MSFR methods [52] with MEMC pipeline, MEMC estimates inter-frame motion and accomplishes alignment based on motion information. (b) MFSR methods [53,54] estimate the blur kernel with a CNN in a supervised manner. (c) Self-supervised MFSR method [55] only uses LR frames to train the blur kernel estimation network and the SR network. (d) Nonblind MFSR with implicit alignment [56,57], which involves feature extraction and spatial–temporal information extraction. (e) MVFR in appearance map reconstruction [18,58], aggregates multi-view redundancy from all LR input images and computes an super-resolved texture atlas. (f) MVFR in stereoscopic reconstruction [59,60], adopts a symmetric architecture with stereo feature interaction.

are used for video super-resolution [71] to extend blind SISR methods by exploiting high-order feature relations. Unlike other approaches that require paired images, a self-supervised method introduced in [55] only needs LR frames to train both the blur kernel estimation and SR network, as depicted in Fig. 5(c). Blind MFSR methods, integrating motion and degradation model estimation, offer enhanced practicality for real-world SR tasks.

5.1.3. Non-blind MFSR with implicit alignment

In non-blind settings with predefined degradation models, methods extract spatial–temporal information from multiple frames, as shown in Fig. 5(d). These processes involve either extracting features from each frame individually before pooling them across feature maps or simply concatenating frames for simultaneous feature and spatial–temporal extraction [72]. Advanced CNN architectures like 3D convolution [73] and recurrent networks [16] enhance this process, with adversarial learning [74] further improving SR realism. However, these methods are constrained by the need for known degradation models, which complicates their application in dynamic MFSR scenarios. Blind settings, in contrast, use implicit alignment to improve efficiency and applicability, with techniques such as down-sampling and up-sampling networks [75] that allow MFSR using only LR frames, similar to blind SISR methods [41]. This integration of blind SISR principles into MFSR highlights the critical role of inter-frame information and positions implicit alignment in blind settings as a viable approach for practical real-world applications.

5.2. Multi-view super-resolution methods

Multi-view super-resolution (MVSR) uses multiple independent measurements of the same sample to enhance estimation accuracy, handling significant variations among input images, unlike MFSR which deals with minor motion-related differences. Typical applications include reconstructing appearance maps from calibrated LR images [18, 58] and HR stereo images from LR counterparts [59,60]. As shown in Fig. 5(e), the method [18] first aggregates redundancy across all LR images to create a super-resolved texture atlas, then enhances it using statistical priors. In stereoscopic reconstruction, the methods [59,60] use symmetric dual-branch architectures for left and right views, as Fig. 5(f) shows. Stereo feature interactions are embedded in each step to utilize the correlation information in stereoscopic images. To further stress the correspondence between cross views, an edge-guided stereo attention mechanism is designed in [59] and a disparity loss in [60]. In MVFR, addressing the efficient use of input image correlations remains a challenge, with most methods [18,58,59] requiring additional inputs that increase computational demands, suggesting potential areas for enhancement.

5.3. Reference-based image super-resolution methods

Reference-based super-resolution (RefSR) uses an HR reference image to enhance the resolution of a similar-viewpoint LR image by

Table 2
Summary of MISR methods and the applications.

Method	*Setting*	*Validation dataset*	*Application scenario*	*Practicability description*
Non-blind MFSR methods with explicit alignment [52], and those with implicit alignment [56]-[74]	Paired HR-LR frames with a known degradation model	Synthetic video frames	Predefined degradation model	Impractical
[2]	LR frames	Real-world video sequences	The same blur kernel	Practical for special cases
[53]	Paired HR-LR frames with known degradation models only for training	Synthetic video frames	Predefined multiple degradation models	Impractical
GVSR [54]		Synthetic and real-world video frames		Practical for special cases
[70]	Paired HR-LR frames with known degradation model	synthetic video frames	Predefined degradation models	Impractical
[55]	Unpaired HR-LR frames	Synthetic and real-world video frames	Corresponding patches with similar distributions	Practical for special cases
[75]	Input LR frames	synthetic video frames		practical for special cases
[18]	Data pair consisting of calibrated LR images and the HR texture map	Synthetic images	Predefined multiple degradation models	Impractical
[58]	Paired HR-LR images and the corresponding normal maps	Synthetic images	Predefined multiple degradation models	Impractical
MESFINet [59], IMSSRnet [60]	Data pair consisting of LR stereo image pair and HR stereo image pair, and additional edge probability maps in [59]	Synthetic images	Predefined multiple degradation models	Practical for special cases
RefSR methods [76]	Data pair consisting of LR image, reference image, the corresponding HR image, and occasionally up-sampling and down-sampling versions of the input images	Synthetic images	Predefined multiple degradation models	Practical for special cases

utilizing the HR reference's rich texture to compensate for details lost in the LR image, thus easing the ill-posed nature of SISR and enhancing performance. A major focus of RefSR is integrating reference information through feature alignment and correspondence matching. Feature alignment typically involves aligning the reference and LR feature maps using optical flow estimators [76] or deformable convolution [77]. In [78], Ref and LR images are segmented into patches, matched by cosine distance [79], and then coarsely warped and finely aligned.

Correspondence matching methods search for useful information in the reference image based on similarities [17]. Subsequent techniques [80] perform this matching in feature space to handle color and illumination variances, using methods like cosine distance for multi-level matching and creating swapped feature maps to assist in resolving LR images. The transformer architecture in [81] treats LR and Ref images as queries and keys, enhancing deep feature correspondence through attention, transferring relevant textures from the reference to the LR image. [82] explicitly deals with transformation gap and resolution gap in the matching by using the proposed C^2-matching. [83] performs C^2-matching in a coarse-to-fine manner, achieving improved real-time performance. The success of RefSR depends on the precise construction of correspondences between the LR and reference images, with errors significantly impacting SR quality. Aforementioned MISR methods are summarized in Table 2.

6. Discussion on MISR methods

6.1. How multiple inputs improve SR performance compared with single input?

Multiple inputs in MFSR, derived from sequential frames, capture subtle differences to preserve dynamics lost with single static inputs, also mitigating overfitting risks associated with deep CNNs [72]. In MVSR, images from different views provide more detailed information compared to MFSR, enhancing SR performance beyond what a single view offers. In RefSR, an HR reference image supplements the LR input with additional high-frequency information, easing the challenges of SISR.

6.2. How multiple inputs are combined and used in different scenarios?

Effectively utilizing multiple inputs is key. For instance, aligned images can be concatenated as the SR model's input, and this can also be done in the feature space [73]. In MFSR, inputs are aligned using motion information or through implicit feature extraction before concatenation for spatial–temporal analysis. In MVSR, fusion is guided step-by-step by edge information [59] or disparity loss [60]. In RefSR, aligned or matched features are concatenated [78] to enhance SR results.

6.3. How to obtain the reference image in RefSR?

Reference images can be obtained from various sources like photo albums, video frames, web image search, etc. In [78], the RefSR is applied to images captured by the smartphone which has dual cameras with wide-angle and telephoto lenses, each with different field of views (FoV). The quality of RefSR may significantly decline when the reference image is less similar to the LR input. An elegant matching scheme enables the model to gracefully degrade its performance to that of SISR when confronted with less relevant Ref inputs.

7. Datasets, metrics, and performance

7.1. Datasets

The datasets involved in SISR and MISR tasks are summarized in Table 3. Popular SISR datasets include DIV2K [84] and Flickr2K [85], featuring 2K resolution images. Common benchmarks for SISR evaluation are Set5 [5], Set14 [6], BSD100 [7], Urban100 [86], and Manga109 [87]. The datasets [5–7,84–87] synthesize LR images from provided HR images for SISR experiments. Real-world datasets [88,89] involve both LR and HR images captured with specific techniques, serving as vital benchmarks for real-world SISR method evaluation. Details of these SISR datasets are in the upper portion of Table 3. For MISR, the REDS dataset [90], part of the NTIRE 2019 Challenge, includes 100-frame video sequences. Vimeo-90K [91] is divided into two subsets: Triplet dataset and Septuplet dataset. Septuplet dataset is for MISR task such as denoising and deblockingg, which consists of 91701 7-frame sequences. Other MISR benchmarks include Vid4 [2] and SPMCS [64], are two common benchmark datasets for MISR evaluation. The MISR datasets mentioned above are detailed in the lower portion of Table 3.

7.2. Metric and performance

PSNR and SSIM are two common metrics to evaluate the performance of SR methods, which are calculates as folows. $PSNR = 10log_{10}\left(\frac{MAX_I{}^2}{MSE}\right)$, where, MAX_I represents the max value in the image, *MSE* is the mean square error between two images. SSIM evaluates the structural similarity between two images. $SSIM = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}$, where, C_1 and C_2 are constants, μ and σ_x^2 are the mean value and the variances, respectively, σ_{xy} is the covariance matrix between x and y.

Directly comparing non-blind and blind methods is challenging due to their different setups. Performance comparisons in Table 4 show non-blind methods generally outperform blind ones, with the latter facing challenges due to unknown degradation models. The ZSSR method [32] scores the lowest due to its simplicity and real-world applicability. Table 5 illustrate the performances of MISR methods, revealing that while SISR methods may show over-smoothing or artifacts, MFSR methods reduce these issues using multiple frames. RefSR methods deliver the highest visual quality by leveraging HR reference.

8. Applications of SR

Previous sections highlight significant advancements in SR, but some methods rely on known degradation models, limiting their real-world applicability despite excellent performance on synthetic data. This section explores practical SR applications derived from real-world scenarios.

8.1. Applications of SISR methods

SISR without Sufficient Data Pairs. Synthetic methods are vital in data augmentation and pair generation, especially when training data is scarce. Various synthetic approaches for image restoration tasks include modifying well-resolved lateral slices to mimic anisotropic axial slices when real biological images are not available [95], and generating time-series LR images using calibrated parameters of fluorophores and stochastic models [96]. Physical noise models are used to create realistic single-photon images [97]. The models trained on synthetic data could perform well on real-world tasks.

SISR with Modality Gap. SR applications extend beyond spatial resolution improvement to other domains, such as cross-modality SR. A scene of cross-modality SR is discussed in [98], where the microscopic images of one modality can be transformed to match the resolution obtained from the other modality. Another task is cross-modality imaging, where the confocal microscopy images can be transformed to match which obtained by the stimulated emission depletion (STED) microscope. In cases where paired data from different microscopic modalities is hard to obtain, unpaired data is used for domain adaptation. A task-assisted cycleGAN model [99] transforms fixed cell images to live cell images, facilitating domain adaptation with unpaired training data.

8.2. Applications of MVSR methods

In the scenarios, the sample can be imaged by two or more cameras, which provides supplementary information for image SR. Guo et al. [100] proposed optical microscopy image SR methods by fusing multi-view images. It is based on Richard-son–Lucy deconvolution (RLD), and can be further accelerated by a deep learning method. Two input images of cells are captured by a dual-view light-sheet microscopy. Leveraging the information from the two views, the proposed framework can produce high-quality results which is much similar to the ground truth.

8.3. Applications of MFSR methods

Multiple frames provide additional spatial and temporal information for superior SR results compared to a single frame. For instance, high-frequency information is enhanced by subtracting the blurred part from the LR image, as demonstrated in [101] that upscales 720p and 1080p to 4K. In another application [72], a multi-frame SR model reduces flickering artifacts in the fluorescence time-lapse imaging of fast-moving subcellular organelles, resulting in higher PSNR. Alon et al. [102] improved the spatial–temporal resolution of single-molecule localization microscopy (SMLM) data to better analyze live cell dynamics. MFSR often utilizes prior knowledge to enhance outcomes, as seen in Shen et al.'s edge-guided video SR framework [103], which excels in processing real satellite video imagery from Jilin-1 and OVS-1.

9. Future works and conclusions

9.1. Future works

9.1.1. Feasible degradation modeling

While current methods excel with synthetic images, a significant domain gap between these models and real-world degradation limits their generalization to actual scenarios. Bridging this gap is crucial for practical application, a topic explored in Section 3.3.2 which suggests using single-image degradation modeling. This method requires patch redundancy across scales, a challenge for SR in surveillance, old photos, and films. Another strategy involves utilizing diverse datasets, though acquiring large-scale high-quality HR and LR pairs is difficult in real-world settings. Typically, available datasets are unpaired, with LR and HR images from different sources or captured by different devices, leading to notable variations that can affect performance. Addressing these issues, one approach from [34] uses two CycleGANs to ensure consistency across different source images. Furthermore, Xu et al. [104] suggest leveraging pre-trained models, which might contain useful degradation-related information for image restoration. Despite the potential, integrating these pre-trained models with SR tasks is challenging due to their heterogeneity. While research in this area is still limited, it presents a promising direction for future exploration.

9.1.2. Assessment of SR results

As noted in Section 7.2, common metrics like PSNR and SSIM may not align with human visual perception in evaluating SR results, given their focus on pixel accuracy rather than perceptual quality. Integrating human preferences into the SR process offers a solution, allowing for nuanced evaluations beyond simple metric scores. Neural networks can utilize these preferences to enhance SR performance, viewing this

Table 3
Dataset details. The datasets above the bold line are SISR datasets, while those below the bold line are MISR datasets.

Dataset	*Image number and the resolution*	*Description*	Train/test Setting
DIV2K [84]	1000 images, 2*K*	Real-world HR images with synthetic LR counterparts.	Train: 800 images, Validation: 100 images, Test: 100 images.
Flickr2K [85]	2650 images, 2*K*	Real-world HR images with synthetic LR counterparts.	No explicit setting.
City100 [88]	100 images, 1218 × 870	Real-world LR-HR pairs.	No explicit setting.
DRealSR [89]	884 380 × 380 images, 783 272 × 272 images, 840 192 × 192 images.	Real-world LR-HR pairs.	Test: 83, 84, and 93 image pairs at the three scales, Train: the remaining data.
Set5 [5]	5 images, 313 × 336	Real-world HR images with synthetic LR counterparts.	Only for test.
Set14 [6]	14 images, 492 × 446	Real-world HR images with synthetic LR counterparts.	Only for test.
BSD100 [7]	100 images, 432 × 370	Real-world HR images with synthetic LR counterparts.	Only for test.
Urban100 [86]	100 images, 984 × 797	Real-world HR images with synthetic LR counterparts.	Only for test.
Manga109 [87]	109 images, 826 × 1169	Real-world HR images with synthetic LR counterparts.	Only for test.
REDS [90]	300 videos, 1280 × 720	HR images, synthetic LR images.	Train: 240 videos, Validation: 30 videos, Test: 30 videos.
Vimeo-90K [91]	91701 videos, 448 × 256	HR images, synthetic LR images.	No explicit setting.
Vid4 [2]	4 videos, 576 × 720, 576 × 704, 480 × 720	Real-world HR images with synthetic LR counterparts.	No explicit setting.
SPMCS [64]	30 videos, 540 × 960	Real-world HR images with synthetic LR counterparts.	No explicit setting.

Table 4
Performance of some SISR methods (PSNR/SSIM with 4×Upscaling).

Method category	*Methods*	*Set5 [5]*	*Set14 [6]*	*BSD100 [7]*	*Urban100 [86]*	*Manga109 [87]*
Non-Blind SISR	VDSR [19] (2016)	31.35/0.883	28.02/0.768	27.29/0.726	25.18/0.754	28.83/0.887
	RCAN [92] (2018)	32.63/0.900	28.87/0.789	27.77/0.744	26.82/0.809	31.22/0.917
	LapSRN [20] (2019)	31.54/0.885	28.19/0.772	27.32/0.727	25.21/0.756	29.09/0.890
	ACT [93] (2023)	32.97/0.903	29.18/0.795	27.95/**0.751**	27.74/0.831	32.20/0.927
	HAT [94] (2023)	**33.04**/**0.905**	**29.23**/**0.797**	**28.00**/**0.751**	**27.97**/**0.837**	**32.48**/**0.929**
Blind SISR	ZSSR [32] (2018)	26.49/0.753	24.93/0.681	25.36/0.652	22.39/0.632	24.43/0.781
	IKC [15] (2019)	28.04/0.808	25.85/0.726	26.01/0.695	23.21/0.694	25.82/0.836
	DAN [30] (2020)	31.89/0.930	28.43/0.769	27.51/0.808	25.86/**0.782**	30.50/0.904
	KOALAnet [31] (2021)	30.28/0.866	27.20/0.754	26.97/0.717	24.71/0.743	28.48/0.881
	DCLS [39] (2022)	**32.12**/**0.889**	**28.54**/**0.773**	**27.60**/**0.729**	**26.15**/0.781	**30.86**/0.909

integration as a cross-task challenge potentially addressed by multi-task learning. However, human involvement is labor-intensive. The use of pre-trained multi-modal foundation models may alleviate this by automating preference-based evaluations, where the model scores images based on human-like preferences during SR model training.

While multi-modal models help quantify human preferences, their internal workings remain opaque. Understanding these mechanisms requires further exploration. High-quality SR aligns with human aesthetics, a subjective quality that goes beyond mere ground truth accuracy. Insights from cognitive science on contexts or textures that attract human interest could inform the development of evaluation systems inspired by brain-like intelligence.

9.1.3. Multiple images acquisition with robot motion

Section 5 reviews MISR methods including MFSR, MVSR, and RefSR, where multiple LR images from varied sources such as video sequences or multi-view setups consistently serve as input. In MFSR and MVSR, synergizing the image capture process with robotic movements can enhance SR performance. For example, mounting a camera on a robot allows adjustments in proximity, scale, or perspective, transforming SISR tasks into MISR tasks with multi-view or reference images. Investigating robotic motion strategies to optimize image capture for improved SR results, possibly through an iterative process where adjustments are made based on SR quality, is a promising research direction. Additionally, evaluating SR results based on human preferences remains relevant.

9.1.4. Motion process SR under physical guidance

Image super-resolution (SR) typically involves increasing image resolution. Another variant, motion process SR, aims to generate additional frames between two consecutive frames captured by a camera with limited frame rates. This is especially useful when observing microorganisms under a microscope or fast-moving objects with a low-frame rate camera, as it can produce clearer images and reveal details that are usually obscured by device limitations. This enhancement is crucial for analyzing dynamics and interpreting motion mechanisms. Motion process SR can be guided by known physical properties such as motion speed, direction, camera distance, frame rate, and imaging mode. However, effectively incorporating these physical priors into the motion process SR poses a significant challenge.

Table 5
Performance of some MISR methods (PSNR/SSIM with 4×Upscaling).

Method category	*Methods*	*Vid4*	*SPMCS*
Non-Blind MISR	[57] (2020)	27.21/0.822	29.74/0.871
	PSRT-recurrent [62] (2022)	**28.07**/**0.8485**	–
	RRCN [63] (2019)	25.54/0.754	–
Blind MISR	[53] (2021)	24.47/0.745	27.53/0.802
	[55] (2022)	**24.59**/**0.763**	**27.77**/**0.818**

9.2. Conclusions

The paper offers a comprehensive review of DL-based SR methods, focusing on applications rather than technical enhancements. We discuss SISR and MISR methods and introduce a taxonomy from an application-oriented perspective that assesses method practicality across settings, validation datasets, application scenarios, and practicability. Key issues such as the degradation model in SISR and the benefits of using multiple inputs in MISR are explored. Finally, we outline several areas for future research to provide insights for the field.

CRediT authorship contribution statement

Hu Su: Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Ying Li:** Writing – original draft, Visualization, Validation, Methodology, Data curation. **Yifan Xu:** Writing – original draft, Visualization, Investigation, Data curation. **Xiang Fu:** Writing – review & editing, Validation, Data curation. **Song Liu:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

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

Acknowledgments

This work is supported in part by the National Key Research and Development Program of China under Grant 2021ZD0114505, and in part by the National Natural Science Foundation of China under Grants 62303321.

References

[1] A. Liu, Y. Liu, J. Gu, Y. Qiao, C. Dong, Blind image super-resolution: a survey and beyond, IEEE Trans. Pattern Anal. Mach. Intell. 45 (5) (2023) 5461–5480.
[2] C. Liu, D. Sun, On bayesian adaptive video super resolution, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 346–360.
[3] Z. Ma, R. Liao, X. Tao, L. Xu, J. Jia, E. Wu, Handling motion blur in multi-frame super-resolution, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5224–5232.
[4] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 295–307.
[5] M. Bevilacqua, A. Roumy, C. Guillemot, M.A. Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding, in: Proceedings of the British Machine Vision Conference, 2012.
[6] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse-representations, in: Proceedings of the International Conference on Curves and Surfaces, 2010, pp. 711–730.
[7] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proceedings of the IEEE International Conference on Computer Vision, 2001, pp. 416–423.
[8] D. Dai, Y. Wang, Y. Chen, L. Van Gool, Is image super-resolution helpful for other vision tasks? in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, WACV, 2016, pp. 1–9.
[9] B. Wang, T. Lu, Y. Zhang, Feature-driven super-resolution for object detection, in: Proceedings of the International Conference on Control, Robotics and Cybernetics, CRC, 2020, pp. 211–215.
[10] X. Fang, H. Fan, M. Yang, T. Zhu, B. Ran, Z. Zhang, Z. Gao, Small object detection in remote sensing images based on super-resolution, Pattern Recognit. Lett. 153 (2021) 107–112.
[11] R. Sharmaa, B. Dekab, V. Fuscoa, O. Yurdusevena, Integrated convolutional neural networks for joint super-resolution and classification of radar images, Pattern Recognit. 150 (2024) 110351.
[12] W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue, Q. Liao, Deep learning for single image super-resolution: a brief review, IEEE Trans. Multimed. 21 (12) (2019) 3106–3121.
[13] H. Liu, Z. Ruan, P. Zhao, C. Dong, F. Shang, Y. Liu, L. Yang, R. Timofte, Video super-resolution based on deep learning: A comprehensive survey, Artif. Intell. Rev. 55 (8) (2022) 5981–6035.
[14] Z. Wang, J. Chen, S.C.H. Hoi, Deep learning for image super-resolution: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (10) (2021) 3365–3387.
[15] J. Gu, H. Lu, W. Zuo, C. Dong, Blind super-resolution with iterative kernel correction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1604–1613.
[16] J. Guo, H. Chao, Building an end-to-end spatial–temporal convolutional network for video super-resolution, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2017, pp. 4053–4060.
[17] H. Zheng, M. Ji, H. Wang, Y. Liu, L. Fang, Learning cross-scale correspondence and patch-based synthesis for reference-based super-resolution, in: Proceedings of the British Machine Vision Conference, 2017.
[18] A. Richard, I. Cherabier, M.R. Oswald, V. Tsiminaki, M. Pollefeys, K. Schindler, Learned multi-view texture super-resolution, in: Proceedings of the International Conference on 3D Vision, 2019, pp. 533–543.
[19] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.
[20] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Fast and accurate image super-resolution with deep Laplacian pyramid networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (11) (2019) 2599–2613.
[21] C. Saharia, J. Ho, W. Chan, T. Salimans, D.J. Fleet, et al., Image super-resolution via iterative refinement, IEEE Trans. Pattern Anal. Mach. Intell. 45 (4) (2023) 4713–4726.
[22] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, Y. Chen, Srdiff: single image super-resolution with diffusion probabilistic models, Neurocomputing 479 (2022) 47–59.
[23] S. Shang, Z. Shan, G. Liu, J. Zhang, Combining cnn and diffusion model for image super-resolution, 2023, arXiv:2303.08714.
[24] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2022, pp. 10674–10685.
[25] K. Zhang, L. Van Gool, R. Timofte, Deep unfolding network for image super-resolution, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3214–3223.
[26] K. Zhang, W. Zuo, L. Zhang, Deep plug-and-play super-resolution for arbitrary blur kernels, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1671–1681.
[27] K. Zhang, W. Zuo, L. Zhang, Learning a single convolutional super-resolution network for multiple degradations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3262–3271.
[28] Y. Xu, S.R. Tseng, Y. Tseng, H. Kuo, Y. Tsai, Unified dynamic convolutional network for super-resolution with variational degradations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12493–12502.
[29] X. Wang, K. Yu, C. Dong, C. Change Loy, Recovering realistic texture in image super-resolution by deep spatial feature transform, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 606–615.

[30] Z. Luo, Y. Huang, S. Li, L. Wang, T. Tan, Unfolding the alternating optimization for blind super resolution, in: Proceedings of Conference on Neural Information Processing Systems, 2020, pp. 5632–5643.

[31] S.Y. Kim, H. Sim, M. Kim, Koalanet: Blind super-resolution using kernel-oriented adaptive local adjustment, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10606–10615.

[32] A. Shocher, N. Cohen, M. Irani, Zero-shot super-resolution using deep internal learning, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3118–3126.

[33] Z. Hui, J. Li, X. Wang, X. Gao, Learning the non-differentiable optimization for blind super-resolution, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2093–2102.

[34] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, L. Lin, Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 701–710.

[35] R. Zhou, S. Susstrunk, Kernel modeling super-resolution on real low-resolution images, in: Proceedings of IEEE/CVF International Conference on Computer Vision, 2019, pp. 2433–2443.

[36] S. Bell-Kligler, A. Shocher, M. Irani, Blind super-resolution kernel estimation using an internal-GAN, in: Proceedings of Advances in Neural Information Processing Systems, 2019, pp. 284–293.

[37] L. Wang, Y. Wang, X. Dong, Q. Xu, J. Yang, W. An, Y. Guo, Unsupervised degradation representation learning for blind super-resolution, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10576–10585.

[38] Z. Luo, Y. Huang, S. Li, L. Wang, T. Tan, End-to-end alternating optimization for real-world blind super resolution, Int. J. Comput. Vis. 131 (2023) 3152–3169.

[39] Z. Luo, H. Huang, L. Yu, Y. Li, H. Fan, S. Liu, Deep constrained least squares for blind image super-resolution, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17621–17631.

[40] T. Michaeli, M. Irani, Nonparametric blind super-resolution, in: Proceedings of IEEE International Conference on Computer Vision, 2013, pp. 945–952.

[41] H. Chen, X. He, H. Yang, Y. Wu, L. Qing, R.E. Sheriff, Self-supervised cycle-consistent learning for scale-arbitrary real-world single image super-resolution, Expert Syst. Appl. 212 (2023) 118657.

[42] J. Liang, K. Zhang, S. Gu, L. Van Gool, R. Timofte, Flow-based kernel prior with application to blind super-resolution, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10596–10605.

[43] Y. Xiao, Q. Yuan, K. Jiang, J. He, Y. Wang, L. Zhang, From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution, Inf. Fusion 96 (2023) 297–311.

[44] C. Mou, Y. Wu, X. Wang, C. Dong, J. Zhang, Y. Shan, Metric learning based interactive modulation for real-world super-resolution, in: Proceedings of European Conference on Computer Vision, 2022, pp. 723–740.

[45] J. Dong, H. Bai, J. Tang, J. Pan, Deep unpaired blind image super-resolution using self-supervised learning and exemplar distillation, Int. J. Comput. Vis. (2023).

[46] H. Liu, M. Shao, Y. Qiao, Y. Wan, D. Meng, Unpaired image super-resolution using a lightweight invertible neural network, Pattern Recognit. 144 (2023) 109822.

[47] A. Bulat, J. Yang, G. Tzimiropoulos, To learn image super-resolution use a gan to learn how to do image degradation first, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 185–200.

[48] Y. Wei, S. Gu, Y. Li, R. Timofte, L. Jin, H. Song, Unsupervised real-world image super resolution via domain-distance aware training, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 13385–13394.

[49] M. Fritsche, S. Gu, R. Timofte, Frequency separation for real-world super-resolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop, 2019, pp. 3599–3608.

[50] Y. Zhou, W. Deng, T. Tong, Q. Gao, Guided frequency separation network for real-world super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 428–429.

[51] Y. Zhang, K. Li, K. Li, B. Zhong, Y. Fu, Residual non-local attention networks for image restoration, in: Proceedings of the International Conference on Learning Representations, 2019.

[52] T.H. Kim, M.S. Sajjadi, M. Hirsch, B. Schölkopf, Spatio-temporal transformer network for video restoration, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 106–122.

[53] J. Pan, H. Bai, J. Dong, J. Zhang, J. Tang, Deep blind video super-resolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4791–4800.

[54] Z. He, D. He, X. Li, R. Qu, Blind superresolution of satellite videos by ghost module-based convolutional networks, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–19.

[55] H. Bai, J. Pan, Self-supervised deep blind video super-resolution, 2022, arXiv:2201.07422.

[56] P. Yi, Z. Wang, K. Jiang, J. Jiang, J. Ma, Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3106–3115.

[57] W. Sun, J. Sun, Y. Zhu, Y. Zhang, Video super-resolution via dense non-local spatial–temporal convolutional network, Neurocomputing 403 (2020) 1–12.

[58] Y. Li, V. Tsiminaki, R. Timofte, M. Pollefeys, L.V. Gool, 3D appearance super-resolution with deep learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9663–9672.

[59] J. Wan, H. Yin, Z. Liu, Y. Liu, S. Wang, Multi-stage edge-guided stereo feature interaction network for stereoscopic image super-resolution, IEEE Trans. Broadcast. 69 (2) (2023) 357–368.

[60] J. Lei, Z. Zhang, X. Fan, B. Yang, X. Li, Y. Chen, Q. Huang, Deep stereoscopic image super-resolution via interaction module, IEEE Trans. Circuits Syst. Video Technol. 31 (8) (2021) 3051–3061.

[61] S. Farsiu, M.D. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super resolution, IEEE Trans. Image Process. 13 (10) (2004) 1327–1344.

[62] S. Shi, J. Gu, L. Xie, X. Wang, Y. Yang, C. Dong, Rethinking alignment in video super-resolution transformers, in: Proceedings of the Advances in Neural Information Processing Systems, 2022, pp. 36081–36093.

[63] D. Li, Y. Liu, Z. Wang, Video super-resolution using non-simultaneous fully recurrent convolutional network, IEEE Trans. Image Process. 28 (3) (2019) 1342–1355.

[64] X. Tao, H. Gao, R. Liao, J. Wang, J. Jia, Detail-revealing deep video super-resolution, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4482–4490.

[65] K.C. Chan, X. Wang, K. Yu, C. Dong, C.C. Loy, BasicVSR: the search for essential components in video super-resolution and beyond, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 4945–4954.

[66] K.C. Chan, S. Zhou, X. Xu, C.C. Loy, BasicVSR++: Improving video super-resolution with enhanced propagation and alignment, 2021, arXiv:2104.13371v1.

[67] X. Wang, K.C.K. Chan, K. Yu, C. Dong, C.C. Loy, EDVR: video restoration with enhanced deformable convolutional networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 1954–1963.

[68] S. Zhang, W. Mao, Z. Wang, An efficient accelerator based on lightweight deformable 3D-CNN for video super-resolution, IEEE Trans. Circuits Syst. I. Regul. Pap. 70 (6) (2023) 2384–2397.

[69] Y. Xiao, Q. Yuan, Q. Zhang, L. Zhang, Deep blind super-resolution for satellite video, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–16.

[70] L. Xiang, R. Lee, M. Abdelfattah, N. Lane, H. Wen, Temporal kernel consistency for blind video super-resolution, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2021, pp. 3470–3479.

[71] R. Chen, Y. Mu, Y. Zhang, High-order relational generative adversarial network for video super-resolution, Pattern Recognit. 146 (2024) 110059.

[72] L. Fang, et al., Deep learning-based point-scanning super-resolution imaging, Nat. Methods 18 (2021) 406–416.

[73] Y. Jo, S.W. Oh, J. Kang, S.J. Kim, Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 3224–3232.

[74] A. Lucas, S. López-Tapia, R. Molina, A.K. Katsaggelos, Generative adversarial networks and perceptual losses for video super-resolution, IEEE Trans. Image Process. 28 (7) (2019) 3312–3327.

[75] Z. He, D. He, X. Li, J. Xu, Unsupervised video satellite super-resolution by using only a single video, IEEE Geosci. Remote Sens. Lett. 19 (2022) 1–5.

[76] H. Zheng, M. Ji, H. Wang, Y. Liu, L. Fang, CrossNet: an end-to-end reference-based super resolution network using cross-scale warping, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 87–104.

[77] G. Shim, J. Park, I.S. Kweon, Robust reference-based super-resolution with similarity-aware deformable convolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8422–8431.

[78] T. Wang, J. Xie, W. Sun, Q. Yan, Q. Chen, Dual-camera super-resolution with aligned attention modules, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1981–1990.

[79] V. Boominathan, K. Mitra, A. Veeraraghavan, Improving resolution and depth-of-field of light field cameras using a hybrid imaging system, in: Proceedings of the IEEE International Conference on Computational Photography, 2014, pp. 1–10.

[80] Z. Zhang, Z. Wang, Z. Lin, H. Qi, Image super-resolution by neural texture transfer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7974–7983.

[81] F. Yang, H. Yang, J. Fu, H. Lu, B. Guo, Learning texture transformer network for image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5790–5799.

[82] Y. Jiang, K.C.K. Chan, X. Wang, C.C. Loy, Z. Liu, Reference-based image and video super-resolution via C2-matching, IEEE Trans. Pattern Anal. Mach. Intell. 45 (7) (2023) 8874–8887.

[83] B. Xia, Y. Tian, Y. Hang, W. Yang, Q. Liao, J. Zhou, Coarse-to-fine embedded patch match and multi-scale dynamic aggregation for reference-based super-resolution, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 2768–2776.

[84] E. Agustsson, R. Timofte, Ntire 2017 challenge on single image super-resolution: dataset and study, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2017, pp. 126–135.

[85] R. Timofte, et al., NTIRE 2017 challenge on single image super-resolution: Methods and results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 1122–1131.

[86] J.-B. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5197–5206.

[87] A. Fujimoto, T. Ogawa, K. Yamamoto, Y. Matsui, T. Yamasaki, K. Aizawa, Manga109 dataset and creation of metadata, in: Proceedings of the International Workshop on Comics Analysis, Processing and Understanding, 2016, pp. 1–5.

[88] C. Chen, Z. Xiong, X. Tian, Z. Zha, F. Wu, Camera lens super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 1652–1660.

[89] P. Wei, Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, L. Lin, Component divide-and-conquer for real-world image super-resolution, in: Proceedings of the European Conference on Computer Vision, 2020, pp. 101–117.

[90] S. Nah, S. Baik, S. Hong, G. Moon, S. Son, R. Timofte, K. Mu Lee, NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 1996–2005.

[91] T. Xue, B. Chen, J. Wu, D. Wei, W.T. Freeman, Video enhancement with task-oriented flow, Int. J. Comput. Vis. 127 (2019) 1106–1125.

[92] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, Y. Fu, Image super-resolution using very deep residual channel attention networks, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 286–301.

[93] J. Yoo, T. Kim, S. Lee, S.H. Kim, H. Lee, T. Hyun Kim, Enriched CNN-transformer feature aggregation networks for super-resolution, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023, pp. 4945–4954.

[94] X. Chen, X. Wang, J. Zhou, C. Dong, Activating more pixels in image super-resolution transformer, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 22367–22377.

[95] M. Weigert, et al., Content-aware image restoration: Pushing the limits of fluorescence microscopy, Nat. Methods 15 (2018) 1090–1097.

[96] Y. Li, et al., DLBI: deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy, Bioinformatics 34 (13) (2018) i284–i294.

[97] L. Bian, et al., High-resolution single-photon imaging with physics-informed deep learning, Nature Commun. 14 (2023) 5902.

[98] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L.A. Bentolila, C. Kural, A. Ozcan, Deep learning enables cross-modality super-resolution in fluorescence microscopy, Nat. Methods 16 (1) (2019) 103–110.

[99] C. Bouchard, et al., Resolution enhancement with a task-assisted GAN to guide optical nanoscopy image analysis and acquisition, Nat. Mach. Intell. 5 (2023) 830–834.

[100] M. Guo, et al., Rapid image deconvolution and multiview fusion for optical microscopy, Nat. Biotechnol. 38 (2020) 1337–1346.

[101] E. Zamfir, M. Conde, R. Timofte, Towards real-time 4K image super-resolution, in: Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2023, pp. 1522–1532.

[102] A. Saguy, et al., DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning, Nat. Methods 20 (2023) 1939–1948.

[103] H. Shen, Z. Qiu, L. Yue, L. Zhang, Deep-learning-based super-resolution of video satellite imagery by the coupling of multiframe and single-frame models, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14.

[104] X. Xu, S. Kong, T. Hu, Z. Liu, H. Bao, Boosting image restoration via priors from pre-trained models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 2900–2909.

Hu Su received the B.Sc. and M.Sc. degrees in information and computation science from Shandong University (SDU), Jinan, China, in 2007 and 2010, respectively, and the Ph.D. degree in control science and engineering from the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China, in 2013. Currently, he is Associate Researcher with the State Key Laboratory of Multimodal Artificial Intelligence Systems, CASIA. His current research interests include intelligent control and optimization and computer vision.

Ying Li received his B.Sc. degree in control science and engineering from North China Electric Power University (Baoding), China in 2016, and received the Ph.D. degree in control science and engineering from the University of Chinese Academy of Sciences, Beijing, China in 2021. His research interests include deep-learning, neural network quantization, large language model.

Yifan Xu is currently pursuing B.Sc. degree in Computer Science and Technology from ShanghaiTech University, Shanghai, China. His current research interest is image super resolution and image reconstruction.

Xiang Fu received the B.S. degree in information and computer science from ShanghaiTech University, Shanghai, China, in 2021, where he is currently pursuing the master's degree with the Advanced Micro-Nano Robots Laboratory. His current research interests include image processing and nano-robotic manipulation.

Song Liu received the B.S. degree in sensing technology and instrumentation from Shandong University, Jinan, China, in 2012, and the Ph.D. degree in control science and engineering from the University of Chinese Academy of Sciences, Beijing, China, and the City University of Hong Kong (CityU), Hong Kong, in 2017. From July 2017 to 2018, he was a Post-Doctoral Fellow with the Robot Vision Research Laboratory, Department of Mechanical Engineering, CityU. From January 2019 to September 2020, he worked as a Post-Doctoral Scholar with the MEMS Group, University of Southern California, Los Angeles, CA, USA. He is currently serving as a Tenure-Track Assistant Professor with ShanghaiTech University, Shanghai, China, where he is also the Director of the Advanced Micro-Nano Robots Laboratory. His current research interests include ultrasonic MEMS, acoustic holography, nano-robotic manipulation, and robot learning. Visit Google Scholar for more publications at 891 https://scholar.google.com/citations?user=n2rWxVQAAAAJ&hl=en.