Blind Image Super-Resolution: 
A Survey and Beyond

Anran Liu‡, Yihao Liu‡, Jinjin Gu, Yu Qiao‡, Senior Member, IEEE, and Chao Dong‡

Abstract—Blind image super-resolution (SR), aiming to super-resolve low-resolution images with unknown degradation, has attracted increasing attention due to its significance in promoting real-world applications. Many novel and effective solutions have been proposed recently, especially with powerful deep learning techniques. Despite years of efforts, it still remains as a challenging research problem. This paper serves as a systematic review on recent progress in blind image SR, and proposes a taxonomy to categorize existing methods into three different classes according to their ways of degradation modelling and the data used to solve the SR model. This taxonomy helps summarize and distinguish among existing methods. We hope to provide insights into current research states, as well as revealing novel research directions worth exploring. In addition, we make a summary on commonly used datasets and previous competitions related to blind image SR. Last but not least, a comparison among different methods is provided with detailed analysis on their merits and demerits using both synthetic and real testing images.

Index Terms—Deep learning, degradation modelling, image super-resolution

1 INTRODUCTION

Single-image super-resolution (SISR) has long been a fundamental problem in low-level vision, aiming to recover a high-resolution (HR) image from an observed low-resolution (LR) input. Years of efforts from the research community have brought about remarkable progress in this field, especially with the booming deep learning techniques [1], [2], [3], [4], [5]. However, most existing methods assume a pre-defined degradation process (e.g., bicubic downsampling) from an HR image to an LR one, which can hardly hold true for real-world images with complex degradation types. Towards filling this gap, growing attention has been paid in recent years to approaches for unknown degradations, namely blind SR. Despite many exciting improvements, these proposed methods tend to fail in many real-world scenarios, as their performance is usually limited to a certain kind of inputs and will drop dramatically in other cases. The main reason is that they still make some assumptions on the degradation types related to the input LR. Readers can see Fig. 1a for an illustration, which shows four different LR inputs with assumed degradation types of some state-of-the-art methods but targeting at the same HR. Therefore, when given an arbitrary input deviating from their assumed data distributions, these methods inevitably produce much less pleasing results. Fig. 1b demonstrates different SR results for a real-world image cropped from the famous film Forrest Gump, which are generated by four state-of-the-art methods. We may find none of these methods have lived up to our expectation for a good viewing experience, since this real-world image does not strictly follow their assumptions on inputs. In fact, it is not rare that we feel confused about which method to choose for a certain image at hand, or whether we could possibly reach a high-quality result using existing methods.

In this article, we try to relieve this confusion through a systematic survey on recent progress in blind SR with our own insight. What's more, it is highly necessary that we look back and reflect on existing methods to have a clear understanding of the current research state and remaining gaps. As stated above, we often have difficulty in selecting a proper method when facing so many ones: KernelGAN [6] can be trained with only a single image, and in what cases will it perform better than IKC [7] with iterative scheme or CinCGAN [8] with unpaired training data? Further, even if every single blind SR method is claimed to work well for real images, we may still struggle to obtain a satisfactory output for our own image, just like the case in Fig. 1. At this stage of development, it is time to ask: To what extent have we solved the problem? What is holding us back and where should we go for future endeavours?

Hence, this paper aims to serve much more than a list of recent progress. Specifically, we propose a taxonomy to effectively categorize existing approaches, which clearly distinguishes among different methods and naturally reveals some research gaps. Based on this taxonomy, our goal is to let each method have its own position within a broad picture composed of existing works. This picture can provide a guideline on reasonable and fair comparison between different kinds of methods in future work. In addition, we will make a summary on the application scope...
Our contributions are mainly three-fold: 1) We present a systematic survey on recent progress in blind image super-resolution, including improvements and limitations of different approaches. 2) We propose a taxonomy to effectively categorize existing methods and reveal some research gaps. 3) We provide insight into current research state and promising future directions.

In the following sections, we first introduce the mathematical formulations of some commonly used SR models in Section 2, and discuss about the challenges from real-world images that we could face when addressing blind SR in Section 3. Then we put forward our proposed taxonomy in Section 4. A quick review on non-blind SISR is provided in Section 5, considering that the research state in non-blind setting has set up the foundation for blind SR. Then we elaborate on methods of each category according to the proposed taxonomy in Sections 6 and 7, followed by a summary on commonly used datasets and previous competitions in the field of blind SR in Section 8. Quantitative and qualitative comparisons among some representative methods from each type are demonstrated in Section 9. Finally, we draw a conclusion on our insight through this survey as well as our perspective on future directions in Section 10.

2 PROBLEM FORMULATION

In this section, we introduce some mathematical formulations of the SISR problem. Specifically, SISR refers to the task of reconstructing an HR image from a given LR input, especially the high-frequency contents in HR. The underlying degradation process from HR to LR can be generally expressed with the following equation:

$$\mathbf{y} = f(\mathbf{x}; s),$$

where $\mathbf{x}$ and $\mathbf{y}$ denote HR image and LR image, $f$ is the degradation function with a scale factor $s$. Therefore, SR problem is equivalent to modelling and solving the inverse function $f^{-1}$. In the background of non-blind SR, $f$ is usually assumed to be bicubic downscaling:

$$\mathbf{y} = \mathbf{x} \downarrow_s \quad \text{bic},$$

or the combination of downsampling and a fixed Gaussian blur with kernel $k_s$:

$$\mathbf{y} = (\mathbf{x} \otimes k_s) \downarrow_s \quad \text{bic},$$

where $\otimes$ denotes convolutional operation. Under either assumption, the corresponding SR model is only able to handle LR inputs with this specific kind of degradation. For other LR images with different degradation types, the mismatch between SR model and intrinsic degradation of inputs may cause severe artifacts in SR results [7, 26]. Fig. 2 gives an illustration on this mismatch from the perspective of image domain adaptation: if an SR model corresponding to a pre-defined degradation is applied to an arbitrary LR input, there will be a large domain gap between the SR output and desired image samples from the target Natural HR domain, thus leading to a poor-quality result.
Some more practical degradation models have also been proposed recently, extending the classical model with high-order modelling [19] or random shuffle strategy [18]. Another group of methods leverage the internal statistics within a single image derived from the classical degradation model, thus requiring no external dataset for training, like ZSSR [26] and DGDML-SR [28]. In fact, internal statistics just reflects the patch recurrence property of an image, and readers can refer to Fig. 3b for an illustration.

Nevertheless, real-world degradations are usually too complex to be modelled with an explicit combination of multiple degradation types, as shown in Fig. 3c. Therefore, implicit modelling attempts to circumvent the explicit modelling function. Instead, it defines the degradation process implicitly through the data distribution, and all the existing approaches with implicit modelling require external datasets for training. Typically, these methods utilize data distribution learning with generative adversarial networks (GAN) [29] to grasp the implicit degradation model possessed within training dataset, like CinCGAN [8].

Although so many models have been put forward in blind SR, there is still a long way to go since we have only tackled a small set of real-world images. Existing methods often claim to focus on real-world settings, but they actually assume a certain scene, like images taken by some specific digital cameras [30], [31]. In fact, real-world images are greatly different in their underlying degradation types, and an SR model designed for a specific type could easily fail for another. In the next section, we will give a brief discussion on different kinds of real-world images, which have posed severe challenges to the field of blind SR.

3 CHALLENGES FROM REAL-WORLD IMAGES

With the development of modern imaging devices, we are now embracing the world with a surge of visual data. Such a variety of image sources also pose more challenges, especially...
in terms of degradation types. Generally speaking, there are three main factors causing different degradations:

1) Different imaging devices. This era of technology has given birth to a dazzling array of digital cameras, not to mention smartphones with advanced camera systems. However, these devices differ greatly in the characteristics of the photos taken. For example, a DSLR (digital single-lens reflex) camera is able to capture high-quality images with a sense of stereoscopy by adjusting its focal length, while a smartphone camera is nowhere near DSLR-quality, tending to produce a "flattened" and noisy scene due to its physical limitations in sensor size and lenses. Another type of low-quality imaging is surveillance video, which often suffers greatly from loss of focus. Readers can see Fig. 4 for some image examples. Images captured with different devices can thus have distinct degradations from one another.

2) Image processing algorithms. This problem is mostly related to digital and smartphone cameras, since it is an image signal processor on the chip that actually processes raw digital signals into visible images. The processing pipeline usually involves multiple steps, such as pixel correction, white balance correction, demosaicing, denoising and sharpening. During this process, complex unknown degradations can be introduced, which are unpredictable and varying among different devices.

3) Degradations rising from storage. To reduce the resource consumption for transmitting and storing data, images and videos are always compressed. Accompanying compressed images are compression artifacts, which will lead to degradations like blurring and blocky effects. In addition, time itself can gradually deteriorate images, especially for old photos and movies recorded on films. Such degradations are mainly caused by poor imaging equipments or erosion in the air, including film grain, sepia effect and color fading. Some example images are presented in Fig. 5. This kind of degradation can hardly be expressed with explicit functions or covered by a few external datasets, thus demanding more efforts in designing restoration algorithms.

The real-world images discussed above all bear their own degradations and challenges. Nonetheless, previous works usually focus on a single type of real image, like those taken by smartphones, which greatly limits their performance in diverse scenes. We expect to see more explorations on different types of real-world images in the future. Specifically, effective solutions for each distinct type, if not a general solution to all, should be the ultimate goal of our research community.

4 TAXONOMY

In this section, we elaborate on our proposed taxonomy to serve as the guideline for our review and analysis. According to Section 2, there have been two ways of modelling the degradation process involved in blind SR: explicit modelling based on the classical degradation model or its variants, and implicit modelling using data distribution among external datasets. The basic idea of explicit modelling is to learn an SR model with external training data covering a large set of degradations, which are usually parameterized with \( k \) and \( n \) in Equation (4). Representative approaches include SRMD [21], IKC [7], KMSR [36] and Real-ESRGAN [19]. Another group of methods propose to exploit internal statistics of patch recurrence, like KernelGAN [6] and ZSSR [26], which can directly work on a single input image. This kind of modelling is primarily based on the classical degradation model. On the other hand, methods with implicit modelling do not depend on any explicit parameterization, and they usually learn the underlying SR model implicitly through data distribution within external dataset. Among such methods are CinCGAN [8] and FSSR [37].

Therefore, we propose a taxonomy to effectively classify existing approaches according to their ways of degradation modelling and the used data for solving the SR model: explicit modelling or implicit modelling, external dataset or a single input image, as shown in Fig. 6. Our reasons for adopting this categorization are three-fold: first, distinguishing
between explicit and implicit modelling helps us to understand the assumption of a certain method, i.e., what kind of degradation this method aims to deal with; second, whether using external dataset or a single input image indicates different strategies of image-specific adaptation with explicit modelling; finally, after categorizing existing approaches into these classes, one remaining research gap naturally reveals itself - implicit modelling with a single image. We argue that this direction is promising in terms of addressing general real-world images with diverse contents, and we will also try to propose feasible suggestions for new solutions in this direction.

In addition to the main taxonomy shown in Fig. 6, we also provide more concrete classifications regarding some specific aspects. For methods with explicit modelling, we further categorize them according to the type of blur kernels used in their experiments. Since SR kernel plays a key role in explicit modelling, the choice of blur kernels may significantly impact the performance of SR model. We show the detailed classification in Table 1, including synthetic kernels (e.g., Gaussian, disk and motion blur) and more realistic ones estimated from real images. Besides, we distinguish different methods in terms of their learning scheme, i.e., whether supervised or unsupervised learning is adopted. Methods with unsupervised learning usually apply GAN framework to learn the data distribution, which differ greatly from supervised learning with paired data. We present detailed categorization in Table 2, and more descriptions of each method can be found in Sections 6 and 7. Readers may refer to a certain taxonomy according to their own needs and preference. Specifically, the main taxonomy presents a global picture to distinguish the overall frameworks among different methods and provides a guideline for fair comparison with existing works, while more specific categories could help readers choose proper experimental settings for their own works.

In the next sections, we first give a quick overview on non-blind SISR, which sets the basis for blind SR methods. Then methods with explicit modelling are introduced in Section 6, and those using implicit modelling are discussed in Section 7. For each type of method, we will unfold the review along its course of development, and make an analysis on their limitations to inspire future work.

5 Overview of Non-Blind Single-Image Super-Resolution

As explained in Section 2, non-blind SR assumes a fixed known degradation process to solve HR outputs. Before the development of deep learning techniques, many traditional techniques are example-based. [51], [52], [53], [54] learn the mapping function from LR to HR with external HR-LR exemplar pairs, where the mapping learning is usually based on a compact dictionary or manifold space. Some others [55], [56] utilize the property of internal self-similarity within a single image without employing external datasets. In 2014, the pioneering work of SRCNN [1] opened a new era of deploying convolutional neural network (CNN) to tackle this task, and it also set up the basic framework for later works based on deep learning.

The commonly adopted CNN framework for SISR task includes three main modules: shallow feature extraction to convert an input LR image into feature maps, deep feature extraction or mapping based on extracted shallow features, and finally SR output reconstruction. Residual learning has also been widely adopted to ease the training process, either in image-level [57] or feature-level [58]. Recent years have witnessed many improvements on deep feature extraction and SR reconstruction modules, such as introducing residual blocks [58], [59], [60], recursive or recurrent structure [61], [62], attention mechanism [5], [63], sub-pixel convolution [3],

### Table 1

| Blur kernel type          | Methods                                      |
|---------------------------|----------------------------------------------|
| Synthetic Gaussian        | SRMD [21], UDVD [38], DPSR [39], USRNet [22]; IKC [7], DAN [20], Cornillere et al. [40], CBISR [27], DRL-DASR [41], KOALANet [42], AMNet-RL [43], Ren et al. [44]; BSRGAN [18], Real-ESRGAN [19]; KernelGAN [6], FKP [45] |
| Motion/Disk               | DPSR [39], USRNet [22]                       |
| Realistic (estimated)     | KMSR [36], RealSR [46]                      |

The realistic kernels are those estimated from real images.
Table 1 for non-blind setting fail to handle degradation without Fig. 16. XPLICIT proposes to directly concatenate degradation D and ESRGAN to be known for a certain input. Yet, it could be easily adapted for blind SR by using an existing method for degradation estimation (e.g., KernelGAN [6] for kernel estimation). It could then receive the estimated degradation information as additional input and focus on how to utilize the given estimation for image-specific adaptation. In contrast, the second type pays special attention to degradation estimation along with the SR process, and combines them into a unified framework. An illustration of their overall frameworks is shown in Fig. 7. In addition, there has been an emerging trend in designing more complex degradation models built upon the classical one, which could better simulate real settings and tackle more challenging cases.

6 Explicit Degradation Modelling

This section covers recently proposed blind SR methods with explicit modelling of degradation process. These methods are mainly based on the classical degradation model shown by Equation (4), where the blur kernel k and additive noise n are two main degradation factors, especially the blur kernel. Table 1 shows different kinds of blur kernels that are used by specific methods. What’s more, these approaches can be further classified into two sub-classes according to whether they employ external dataset or rely on a single input image to solve the SR problem.

6.1 Classical Degradation Model With External Dataset

This kind of approach utilizes external dataset to train an SR model well adapted to a variety of SR blur kernels k and noises n, especially the former. Typically, the SR model is parameterized with a convolutional neural network (CNN), and an estimation on k or n for a specific LR image is used as conditional input to the SR model for feature adaptation purpose. After the training process, the model will be able to produce satisfactory results for LR inputs with degradation types covered in the training dataset. According to whether a certain approach includes degradation estimation in its proposed framework, we further divide these approaches into two types: image-specific adaptation without degradation estimation, and image-specific adaptation with degradation estimation. In fact, the first type is originally designed for non-blind setting, where the specific degradation is assumed to be known for a certain input. Yet, it could be easily adapted for blind SR by using an existing method for degradation estimation (e.g., KernelGAN [6] for kernel estimation). It could then receive the estimated degradation information etc. In addition, multiple loss functions are also proposed for better perceptual quality of SR results [4], [64], [65]. These techniques bring about remarkable progress in terms of both reconstruction accuracy and efficiency, and non-blind SISR with bicubic-downsampling assumption actually reaches maturity.

However, these non-blind models usually struggle to generalize to input images with more complex degradations deviating from their assumed one. This problem can be well observed in Fig. 1b and Figs. 16, 17, and 18, where SRResNet [4] and ESRGAN [64] for non-blind setting fail to handle blurry or noisy input images. Hence, it is demanding to propose effective methods for blind SR setting, which are the main focus of this survey and will be explored in detail in the following two sections.

| Learning scheme | Methods |
|-----------------|---------|
| Supervised      | SRMD [21], UDVD [38], DPSR [39], USRNet [22], IKC [7], DAN [20], KMSR [36], DRL-DASR [41], CBSR [27], Cornillere et al. [40], KOALAnet [42], AMNet-RL [43], Ren et al. [44], RealSR [46], BSRGAN [18], Real-ESRGAN [19] |
| Unsupervised (GAN-based) | KernelGAN [6], DGDM-SR [28], CinCGAN [8], unpaired image SR with pseudo-supervision [47], FSSR [37], Bulat et al. [48], Zhou et al. [49], DASR [50] |

Note that unsupervised methods are usually GAN-based methods.
To generate the degradation map for the LR image, it introduces a strategy called dimensionality stretching. Specifically, an SR blur kernel is encoded to a vector with principal component analysis (PCA) and concatenated with the noise level $\sigma$. The full encoding vector is stretched both vertically and horizontally to the same size with the LR image. This strategy can be easily extended to non-uniform maps for spatially variant degradations. The SR reconstruction network of SRMD is similar to those commonly adopted in non-blind SR. The whole pipeline is presented in Fig. 8a.

Following SRMD, UDVD [38] also uses the degradation map as an additional input for SR reconstruction, yet it makes one step forward by employing per-pixel dynamic convolution to more effectively deal with variational degradations across images. In addition, an improvement on the kernel coding operation is proposed by Cornillere et al. [40] to replace the PCA technique with a shallow neural network, which can potentially learn a kernel mapping more fitted to the specific SR model. The objective function can be split into two sub-problems using the half-quadratic splitting (HQS) algorithm: one addresses deblurring task and is related to parameter $k$, while the other aims to super-resolve a bicubic downsampled image with some virtual noise level $\mu$. Fortunately, the first sub-problem can be solved in a closed form with Fast Fourier Transform without any kernel coding, thus allowing the model to cope with more complex kernels. The second sub-problem can be modelled by a non-blind SR network capable of dealing with additive noise, and this network can be directly adapted from SRMD framework with a single noise model, especially for those with irregular patterns like motion blur.

Hence, another group of methods have been proposed based on an MAP framework, which requires no kernel coding for degradation map generation. Specifically, deep plug-and-play super-resolution (DPSR) [39] incorporates SR network into an MAP-based iterative optimization scheme. It solves the HR image by minimizing an objective function based on a degradation model modified from Equation (4):

$$y = (x_1 \odot k) + n.$$

This model decouples downsampling from blurring. The objective function can be split into two sub-problems using the half-quadratic splitting (HQS) algorithm: one addresses deblurring task and is related to parameter $k$, while the other aims to super-resolve a bicubic downsampled image with some virtual noise level $\mu$. Fortunately, the first sub-problem can be solved in a closed form with Fast Fourier Transform without any kernel coding, thus allowing the model to cope with more complex kernels. The second sub-problem can be modelled by a non-blind SR network capable of dealing with additive noise, and this network can be directly adapted from SRMD framework with a single noise model, especially for those with irregular patterns like motion blur.
map as additional input. Unfolding super-resolution network (USRNet) [22] also adopts the MAP framework but is based on the original degradation model in Equation (4), and the corresponding two sub-problems become super-resolving an LR image blurred by kernel $k$ and denoising an HR image with a virtual noise level $\mu$. It enhances the solution framework by unfolding the iterative optimization process of DPSR into an end-to-end trainable network with iterative scheme, enabling joint optimization between the two sub-problems. A comparison between solution frameworks of DPSR and USRNet is depicted by Figs. 8b and 8c. Besides, some earlier methods also exploit plug-and-play technique, including [66, 67, 68]. Specifically, [66] discovers that a denoiser could be adapted to a SISR solver using plug-and-play priors [69]. [67] presents a SISR method based on plug-and-play ADMM [70] with faster convergence speed. [68] integrates pre-trained CNN denoisers into model-based approaches to serve as effective image prior.

**Limitation:** In spite of the aforementioned progress, these methods have one obvious drawback: they all rely on an additional input of degradation estimation, especially the SR kernel $k$. However, estimating the correct kernel from an arbitrary LR image is not an easy task, and an inaccurate estimation input will cause kernel mismatch and greatly undermine the SR performance [7, 26]. Fig. 9 shows the comparison between SR results with correct and incorrect kernels based on the method SRMD. Therefore, only if one has some method at hand for reliable degradation estimation can he quickly obtain a satisfactory SR output, otherwise he may come to the tedious work of manually choosing a proper estimation input for better result. Hence, we introduce another kind of approach in the next section, which incorporates degradation estimation into SR framework for more robust performance.

### 6.1.2 Image-Specific Adaptation With Degradation Estimation

This type of method combines degradation estimation and SR process into a unified framework, where kernel estimation is usually the main focus of the former. Iterative kernel correction (IKC) [7] proposes to correct kernel estimation in an iterative way to gradually approach a satisfactory result. The highlight of this method is to take advantage of the intermediate SR results, since the artifacts within an SR image caused by kernel mismatch tend to have regular patterns. Specifically, a predictor network is applied for kernel initialization based on the LR input, and in subsequent iterations a corrector network is used to estimate the kernel correcting residual given an SR image conditioned on the current kernel. The updated kernel can generate a new SR result with fewer artifacts. Feature adaptation is accomplished with spatial feature transform [71] layers, which can be more effective than direct concatenation of inputs as proposed by SRMD. A more recent work, deep alternating network (DAN) [20], further enhances the IKC framework. It unifies the corrector and SR network into an end-to-end trainable one instead of training each sub-network separately as IKC does. This joint training strategy can make the two networks more compatible to each other. The overall frameworks of IKC and DAN are illustrated in Figs. 10a and 10b. The idea of making use of SR artifacts for kernel estimation is also employed in Cornillere et al. [40], yet they train a kernel discriminator to estimate the error map of an SR output instead of the kernel, and find the optimal kernel by minimizing the error of SR output during the inference stage, as shown by Fig. 10c. In addition to the SR kernel, estimation on more degradation types has also been studied. CBSR [27] combines two sub-networks for noise and kernel estimation with a non-blind SR network, thus forming a unified cascaded architecture for blind SR.

In fact, the iterative scheme adopted by IKC and DAN can be interpreted well from the perspective of domain adaptation: instead of producing the final SR output in a single stroke like SRMD, it chooses several intermediate SR results as interchange stations during the long trip from input LR to the target Natural HR domain, passing across the domain gap in Fig. 2 step-by-step. These two methods can have more robust performance than SRMD framework depending on the accuracy of kernel estimation input. Nevertheless, such an iterative scheme usually consumes more inference time and requires human intervention to choose the optimal number of iterations.

To tackle these issues, some recent works propose non-iterative frameworks by introducing more accurate degradation estimation or more efficient feature adaptation strategies. Unsupervised degradation representation learning for blind SR (DRL-DASR) [41] tries to estimate the degradation information with a trainable encoder in the latent space, and the degradation encoder is trained with contrastive information with a trainable encoder in the latent space, which incorporates degradation estimation into SR framework for more robust performance.
adaptive modulation network with reinforcement learning (AMNet-RL) [43], proposes a modified version of adaptive instance norm (AdaIN) [72] to incorporate kernel estimation into SR network, and it also pioneers in optimizing the blind SR model with in-differentiable perceptual metrics (e.g., NIQE [73]) using reinforcement learning.

There are also some other approaches proposing to learn a blind SR model by merely covering more degradations in the training dataset, especially more realistic kernels estimated from real images. For instance, kernel modelling super-resolution (KMSR) [36] builds a large kernel pool with data distribution learning based on some realistic SR kernels estimated from real LR images. Kernels from this pool are then used to synthesize HR-LR training pairs according to the classical degradation model, and the training process just follows non-blind setting with supervised learning. Usually, a more general training dataset enables the SR model to implicitly distinguish and adaptively deal with LR inputs with different degradations. In other words, the SR model will be implicitly endowed with more capacity for kernel estimation in the training process, thus avoiding explicit kernel estimation in the framework. However, such a direct way may not lead to top performance, as have been argued in [21]. A similar strategy is employed in RealSR [46] and Ren et al. [44] to build generic training datasets by covering more realistic degradations.

**Limitation:** Compared with approaches without degradation estimation, these methods practically save us from efforts in searching for degradation estimation algorithms, especially during inference stage, and have demonstrated impressive performance. Yet, they still cannot avoid the inherent disadvantage of explicit modelling: they cannot give out satisfactory results for images with degradations not covered in their model. As presented in Fig. 11, for an SR model focusing on the degradation caused by kernel $k$, like IKC, it can hardly deal with LR inputs with degradations out of its modelling scope. This limitation is really too tough for complex real-world images. Even if we are willing to retrain the model with more degradation types, it is impractical for us to explicitly model the degradation in an arbitrary LR and gather enough external training data, as stated in Section 3.

### 6.1.3 Complex Degradation Modelling

Due to the limited capacity of classical degradation model when facing real-world scenes, some recent works resort to more advanced degradation modelling, which are built upon the classical one but can greatly expand the degradation space with novel techniques. BSRGAN [18] follows the common scheme involving blur, downsampling and noise in the degradation process, and it proposes to apply different degradation operators with randomly shuffled sequences.
This strategy substantially enlarges the space of degraded LR with various combinations. Another work, Real-ESRGAN [19], formulates real-world degradations into a high-order model in contrast to the classical one with only first-order modelling. Specifically, an $n$-order model repeatedly applies the classical degradation model for $n$ times, and it could better simulate the multiple degradation processes that a real image could undergo during transmission. With a large set of HR-LR pairs synthesized from their new degradation models, these methods supervisedly train an SR network modified from popular SISR ones like ESRGAN [58]. This framework with complex degradation modelling has proven to be effective to tackle more challenging real images with a single large blind model.

**Limitation:** As shown in Section 9.2, complex degradation modelling does bring benefits, yet it still demonstrates limited application boundaries. For example, both of these methods struggle to yield clean and uniformly sharp details for the real image input in Fig. 12. Next, let us step into another type of method, which utilizes a single input image alone for image-specific SR modelling.

### 6.2 Single Image Modelling With Internal Statistics

SR modelling with a single image is based on the internal statistics of natural images: patches of a single image tend to recur within and across different scales of this image. The internal statistics has been quantified and proved to have more predictive power than external statistics for many natural images [75]. A theoretical formulation is given by [76], based on the classical degradation modelling in Equation (4) without noise $\mathbf{n}$. It proves that, for local patches in the same LR image with recurring pattern and different scales, their relationship is equivalent to patches from an HR and its LR version with the corresponding scaling ratio. This property can be used to estimate SR kernel $\mathbf{k}$ and solve the HR.

Glasner et al. [56] did the pioneering work in 2009 to introduce internal statistics into solving SISR problem from a single image. Michaeli & Irani [76] further extend this framework to blind SR setting. Specifically, [76] proposes an MAP framework to estimate the SR blur kernel, based on the observation that the optimal kernel $\mathbf{k}$ is the one that maximizes the similarity among recurring patches across different scales. In addition, it proves that the optimal $\mathbf{k}$ is not PSF of the camera but should be one with a smaller width, contrary to the common sense in its era.

The recent development of GAN gives birth to a new realization of using patch recurrence for blind kernel estimation. KernelGAN [6] interprets the maximization of patch recurrence within a single image as a data distribution learning problem. It assumes that the downsampled version of an LR image generated by the optimal $\mathbf{k}$ should share the same patch distribution with the original LR. Under GAN framework, a deep linear network is used as generator to parameterize the underlying SR kernel, and a discriminator distinguishes generated patches from those in original LR image. Once the training finishes, one can explicitly obtain the kernel estimation by convolving together all convolutional filters in generator. It is worth noting that the training process relies merely on the input LR without any external dataset, which can be seen as self-supervised learning. Flow-based kernel prior (FKP) [45] develops a more effective approach for kernel optimization, where a kernel prior in latent space is learned with normalizing flow (NF) [77], [78] technique. Thanks to the invertible mapping between latent and pixel spaces enabled by NF, the search for the optimal $\mathbf{k}$ can be conducted in the learned kernel manifold. This process can be more efficient than directly optimizing a randomly initialized deep network, thus leading to more robust kernel estimation results.

The idea of self-supervision based on patch recurrence property can also be directly applied to perform SR. Zero-
shot super-resolution (ZSSR) [26], developed by authors from the same group as Michaeli & Irani [76] and Kernel-GAN [6], made the very first attempt to train an image-specific CNN for super-resolving each input LR \( y \) without any pre-training step. During training, \( y \) is regarded as HR, and its paired LR patches are produced by downsampling with kernel \( k \). CNNs trained in this way are capable of inferring specific relationships across different scales of \( y \), which are then used to super-resolve \( y \).

In fact, ZSSR is still not well designed for blind setting: it requires estimated SR blur kernel \( k \) as input to guide the generation of LR images for training. A unified self-supervision framework is thus proposed in DGDM-SR [28] - depth guided degradation model for learning-based SR. It combines a degradation network and an SR network into a single architecture, where the former is trained to simulate the degradation process, similar to the function of KernelGAN, and the latter aims to perform SR task like ZSSR does. This joint framework allows directly using generated LR as input to SR network without explicit extraction of SR kernel. In addition, DGDM-SR proposes to sample HR and LR patches in an unpaired way according to the depth map of input image, assuming that patches with smaller depth are equivalent to HR and those with larger depth to LR. A two-cycle training scheme similar to CycleGAN [79] structure is employed to simultaneously train the two networks (see Fig. 13), where the unpaired HR and LR patches are used as real samples for data distribution learning.

**Limitation:** The idea of self-supervision with internal statistics seems attractive for solving SR images from LR with variant degradation types, since it requires no effort in gathering large external training dataset. Nevertheless, its basic assumption [75] may easily fail, especially for natural images with diverse contents (e.g., animals) or monotonous scenes (e.g., sky), since it is hard to exploit recurring information across scales to robustly perform SR with this kind of input image. Hence, these approaches can only produce favourable SR outputs for a very limited set of images with frequently recurring contents across scales, and new methods for single-image modelling are waiting to be explored for more general natural images.

So far, we have had an overview on approaches with explicit degradation modelling, as well as their merits and demerits. Explicit modelling of degradation process is clear and straightforward, yet it can be too simple to model unknown and complicated degradations in real scenarios. In fact, real-world images usually include multiple degradations, and we can hardly express these entangled factors with an explicit well-defined function. Hence, another group of methods propose to implicitly model the degradations through data distribution learning. To the best of our knowledge, so far there have only been approaches based on external dataset for implicit modelling, and we will talk about them in the following section.

### 7 Implicit Degradation Modelling

#### 7.1 Learning Data Distribution Within External Dataset

This kind of approach aims to implicitly grasp the underlying degradation model through learning with external dataset. For dataset with paired HR-LR images, supervised learning with cautious design of SR network may be already enough to achieve satisfactory results, just like the top solutions proposed in NTIRE 2018 [80] and AIM 2020 [31] challenges. A more difficult setting is learning with unpaired data, where ground truth of LR images with realistic degradations are unavailable. Existing approaches usually exploit data distribution learning with GAN framework [29], and one or more discriminators are used to distinguish generated image samples from real ones, pushing the generator towards the appropriate modelling direction. In most cases, two datasets are used to train the model, including HR and unpaired LR, and one may regard the two datasets as representing target and source domains to learn the domain adaptation.

Among the earliest attempts on implicit modelling with unpaired data is CinCGAN [8]. It proposes to first transform an LR input to the Bicubic LR domain before performing SR with a pre-trained non-blind model. The corresponding adaptation process is illustrated in Fig. 14a. The Bicubic LR domain is also regarded as Clean LR, because its samples are generated with bicubic downsampling from HR images and assumed to be without any noise. Two CycleGAN structures [79] are respectively applied to transformation from LR to Clean LR and to target HR, helping to maintain cycle consistency in the transformation process. In this way no paired data is required during training, thus forming an unsupervised training scheme.

However, unsupervised domain adaptation with GAN is a non-trivial task. Since paired data is not available, pixel-wise reconstruction loss cannot be directly used during
training. Note that it plays an important role in supervised learning to ensure reconstruction fidelity and optimization stability. For unsupervised learning, it only relies on a randomly-initialized discriminator to distinguish the target domain through optimization, hence it usually requires complex loss functions to ensure stable training and reduce artifacts.

To leverage both supervised learning and unsupervised learning with GAN, some approaches focus on learning the degradation process from HR to LR, and the generated LR samples with more realistic degradations are used to train the SR model in a paired manner. This process is depicted in Fig. 14b. Bulat et al. [48] adopt this scheme and combine a High-to-Low degradation generator with a Low-to-High SR one in a single framework. Specifically, the High-to-Low generator simulates the degradation from HR to LR with an LR discriminator, and the degraded LR image is used as input to Low-to-High generator for SR training.

supervision is employed to assign different loss weights to the LR inputs according to their domain distance to real LR, which is based on predictions of LR discriminator. This helps to reduce negative influence caused by generated far-away LR samples. Another work, unpaired image SR using pseudo-supervision [47] applies both generated and real LR images to train the SR networks, while the latter adopts a correction network similar to CinCGAN. “D” represents a discriminator.

Limitation: Though seemingly flexible and powerful, this kind of method is still far from a cure-all in blind SR. On the one hand, these methods must rely on large external datasets to learn the SR model through implicit data distribution, but this data-hungry manner is not suitable for certain tasks, including old photo restoration. On the other hand, most of them exploit GAN framework for unsupervised data distribution learning. The GAN-based framework may be difficult to train, and it will frequently produce severe artifacts in SR results. These artifacts are harmful to many real-world applications, such as high-definition display and old photo/film restoration. Readers can refer to Fig. 15 for illustration on the performance of a GAN-based method, FSSR.

7.2 Implicit Modelling With a Single Image: A Future Direction
Up till now, we have gone through a survey on different kinds of methods for blind SR, which steadily push forward the research frontier towards real-world applications. Nonetheless, as stated in previous sections, existing approaches
all have their limitations. If we try to figure out the part that lies outside the application scope of all these methods, we can find many cases that are commonly seen in our daily life, including surveillance video, old photos and films, as demonstrated in Section 3. These images could be the outliers of internal statistics with complex real-world degradations, posing great challenge to existing methods: they not only lack patch redundancy across scales to serve as hint for explicit modelling, but also cannot be covered with a few external datasets due to the complexity of unpredictable degradations. It is worth noting that accompanying this challenge is a new research direction: implicit modelling with a single LR image, as revealed in Fig. 6. So far, there have been no related works in this field, but we believe it is a worthwhile direction for future research.

The main difficulty lying in this direction is the lack of effective image prior to perform SR. As discussed in Section 6.2, general natural images with diverse scenes could deviate from the internal statistics that is currently used for single image modelling, yet no other effective prior has been proposed in literature to work on a single image. A more relevant related work is DIP [82], which unravels that a randomly-initialized network can serve as a hand-crafted prior for image restoration tasks. However, it could only work in non-blind setting and is not applicable to blind SR.

From our perspective, one possible solution to this new direction is to apply human intervention as additional information. In other words, we may resort to semi-automatic approaches to compensate for the lack of general image prior. This kind of approach could still be built upon the framework of blind SR methods, and users should only take minimal effort to guide the SR process towards a more satisfactory result. For example, the idea of adaptive feature modification [83] allows users to manipulate the SR reconstruction with one controlling coefficient. Besides, users can manually choose a high-quality image with similar contents to the LR input as an SR reference, which could help to recover more details without introducing artifacts [84], [85]. Such slight human intervention may be more effective than a fully automatic blind SR method, especially for highly degraded images. In summary, the key point of these proposals is to increase the amount of useful information to make blind SR possible.

8 DATASETS AND COMPETITIONS

8.1 Datasets

A large portion of methods covered in this paper, especially those with explicit degradation modelling and external dataset, require HR-LR image pairs to solve and evaluate SR models. However, due to the difficulty of obtaining real paired data, so far there have been only a few real-world datasets and most methods still adopt the way of synthesizing LR inputs from HR images. Section 8.1.1 discusses the common ways of building synthetic dataset, and Section 8.1.2 gives an introduction on a few available real-world datasets.

8.1.1 Synthetic Dataset

For methods with explicit degradation modelling, the process of synthesizing degraded LR images from HR ground truth usually follows Equation (4) or its more complex variants. For kernel $k$, Gaussian kernels have been the most widely adopted kernel type. A typical practice is introduced by SRMD [21]. An isotropic Gaussian kernel can be generated with a kernel width uniformly sampled from a pre-defined range, while an anisotropic one is characterized by a covariance matrix $\mathbf{Z}$, where the rotation angle of its eigenvectors and the corresponding eigenvalues determine the kernel shape. As for noise $n$, additive Gaussian noise is mostly used to simulate real-world noises, and the level $\sigma$ can also be sampled from a specific range. Examples of Gaussian kernels and different noise levels are included in Fig. 3a. In addition, one can also enlarge the dataset with more realistic kernel or noise types, just as done by KMSR [36] and RealSR [46].

The most popular HR datasets in the non-blind setting are also employed in blind SR. For example, DIV2K [86] and Flickr2K [87] are often used for training, while Set5 [88], Set14 [89], BSD100 [90] and Urban100 [24] are usually for testing. Specifically, a blind SR benchmark DIV2KRK is proposed in KernelGAN [6], where each image in DIV2K validation set is blurred and downsampled by a randomly generated anisotropic Gaussian kernel with some multiplicative noise to simulate more complex degradation. Other synthetic datasets with unknown degradation, e.g. DIV2K wild [80], are also used by some methods with implicit modelling like CinCGAN [8].

8.1.2 Real-World Dataset

Up till now, there have been several real-world datasets with paired HR-LR images, and these datasets are built with carefully designed techniques and advanced digital devices. Representatives are City100 [91], DRealSR [92] and RealSR [93]. Among them, DRealSR is the largest one with around 800 image pairs for each scale factor. Usually, an HR image and its corresponding LR observations are captured by adjusting the focal lengths of imaging devices, then HR-LR pairs are accurately aligned with image
registration and color rectification. Some other real-world datasets without HR ground truth have also been used as source domain images for SR network inputs, like DPED [25] in NTIRE 2020 Real-World Image SR challenge [30]. Compared with synthetic data, these real-world datasets serve as an important benchmark for investigating blind SR in real setting. However, building such a dataset is time-consuming and expensive, and also cannot cover all scenarios due to complicated variations among different imaging systems.

### 8.2 Competitions

In order to gauge and promote the development of superior solutions, some competitions have been held in the field of SISR, including some tracks with unknown degradation or blind SR setting. Here, we make a summary on previous competitions related to blind SR in Table 3, hoping to reveal some research trends from another perspective.

As the first competition related to blind setting, NTIRE 2018 challenge on SISR goes one step ahead from bicubic downscaling and employs unknown degradation operators to get LR images. More complex degradation types are then introduced by succeeding competitions, including real-world degradations arising from digital cameras. These competitions help to probe the state-of-the-art in this field. However, since paired training dataset is provided, their top solutions were largely focused on network structure enhancement as for non-blind SR. AIM 2019 challenge made the first attempt to address the real-world setting where paired training data is unavailable, hoping to stimulate research endeavours towards unsupervised learning. It can be seen that rising attention has been paid to this emerging task, especially from its follower NTIRE 2020, and the research community is expected to propose more novel solutions to fit into real-world problems.

### 9 Quantitative Comparison

This section includes detailed analysis on the performance and limitations of different methods with some testing examples. In fact, it is not an easy task to provide a comprehensive and fair comparison here, and we will state several problems hindering this practice in Section 9.1. Despite the difficulty, we still present some testing results based on officially released pre-trained models in Section 9.2, in order to provide readers with some useful information on the performance of each kind of method.

#### 9.1 Problem: Difficulty of Fair Comparison

- **Inaccessible Code**: Among the methods covered in our survey, some have released their full set of codes for training and testing, such as SRMD, IKC, RealSR,
different training data

Table 6 

The testing dataset is DIV2KRK [6]. “GT” denotes ground truth.

KernelGAN with ZSSR, and FSSR. But there are also some others which are not able to make their codes publicly available, and it is also non-trivial to reproduce these methods due to the lack of some important implementation details in original paper, especially for GAN-based approaches.

- Different Training Data: Even though the official implementations and pre-trained models are at hand, we still cannot sit back and relax for fair comparison since most of them adopt different datasets or different degradation types during training. These variables can greatly affect their performance and generalization capacity, especially for the blind setting with a large variety of distinct degradations.

Hence, it is necessary and demanding to set up a benchmark for existing approaches. This benchmark should provide a fair and comprehensive comparison based on a uniform platform, and help gauge and push forward the state-of-the-art for blind SR. In this survey, we present testing results using released pre-trained models of some representative methods, and also show some quantitative results from published papers for other methods.

9.2 Comparison and Analysis

This section includes an overview on both quantitative and qualitative performance of some representative approaches. Results in Tables 4 and 6 are from published papers. Results in Table 5 are tested by ourselves with released codes or pre-trained models. Also, each table mainly corresponds to a specific category in our proposed taxonomy. These results can provide readers with a straightforward comparison between different methods in each category.

For approaches with explicit modelling, we first show in Table 4 the comparison between two kernel estimation methods with single image modelling (Michaeli & Irani [76], KernelGAN [6]) combined with two SR methods requiring kernel estimation input (SRMD [21], ZSSR [26]). These results are reported in [6] on the testing dataset DIV2KRK synthesized with anisotropic Gaussian kernels. They indicate that methods using degradation information as additional input can well be fitted into blind setting if combined with an appropriate kernel estimation algorithm. However, there still remains a considerable performance gap between using ground truth kernel and the estimated one, since it is non-trivial to accurately estimate SR kernel from an arbitrary image. Also, these methods can hardly outperform those combining degradation estimation and SR into a single framework, like DAN, showing the advantage of joint optimization.

For methods with explicit modelling and incorporated degradation estimation, we present the quantitative results of some representative methods in Table 5. These results are produced using codes or pre-trained models released by authors of the corresponding papers. Specifically, we uniformly choose six isotropic Gaussian kernels from range [0.8, 3.8], and use them to synthesize LR inputs according to Equation (4) but without additive noise. Besides, we provide results of SRResNet trained with bicubic-downsampled LR to serve as baseline of non-blind setting. We can observe that the performance of SRResNet drops dramatically with the increase of blur level, gradually getting close to bicubic interpolation. On the other hand, blind SR methods still perform well for high blur levels, and DRL-DASR shows better performance than IKC due to its degradation estimation in latent space. Also, methods with non-iterative scheme, like KOALAnet and DRL-DASR, have great advantage in terms of inference time compared to IKC with iterative scheme.

For methods with implicit modelling, we show the results from DASR [50] in Table 6, including comparison among three representative data distribution learning models: CinCGAN [8], FSSR [37] and DASR [50]. Besides PSNR and SSIM, LPIPS [99] is used to better evaluate the perceptual quality of SR images generated from GAN. Two dataset

Table 5

Quantitative Comparison (PSNR(dB)) of Image-Specific Adaptation With Degradation Estimation (IKC [7], DRL-DASR [41], KOALAnet [42])

| Method      | Time  | BSD100          | Urban100         |
|-------------|-------|-----------------|-----------------|
|             |       | 0.8  | 1.4  | 2.0  | 2.6  | 3.2  | 3.8  | 0.8  | 1.4  | 2.0  | 2.6  | 3.2  | 3.8  |
| Bicubic     |       | 25.79 | 25.45 | 24.99 | 24.51 | 24.06 | 23.68 | 22.96 | 22.59 | 22.10 | 21.59 | 21.12 | 20.71 |
| SRResNet    | 0.016s| 27.39 | 26.50 | 25.59 | 24.85 | 24.26 | 23.79 | 25.57 | 24.06 | 22.84 | 21.98 | 21.33 | 20.83 |
| KernelGAN+ZSSR | 240s | 20.25 | 21.95 | 23.61 | 25.06 | 25.14 | 24.59 | 18.81 | 20.09 | 21.29 | 22.46 | 22.42 | 21.97 |
| KOALAnet    | 0.058s| 27.19 | 27.22 | 27.23 | 27.19 | 26.74 | 25.55 | 25.00 | 24.98 | 24.94 | 24.89 | 24.39 | 22.94 |
| IKC         | 0.232s| 27.61 | 27.33 | 27.28 | 27.20 | 26.79 | 26.28 | 26.01 | 25.45 | 25.25 | 25.02 | 24.58 | 23.99 |
| DRL-DASR    | 0.040s| 27.53 | 27.52 | 27.49 | 27.37 | 27.08 | 26.70 | 25.77 | 25.66 | 25.57 | 25.34 | 24.97 | 24.54 |

Inference time of a single image is averaged on BSD100 testset. Note that the comparison with KOALAnet is not perfectly fair, since its degradation space is different from IKC and DRL-DASR.
are used for testing: AIM [97] is released by AIM Challenge on Real World SR in ICCV 2019, while RealSR [93] is a real-world dataset composed of HR-LR pairs captured with different focal lengths of the camera. We can see that DASR demonstrates the best performance, especially in terms of visual quality, owing to its better training strategies for narrowing the domain gap.

Qualitative comparison of some testing examples is shown in Figs. 16, 17, and 18, which are produced with released pre-trained models. The first two degraded LR images are synthesized with isotropic Gaussian blur and additive Gaussian noise, and the third one is a real-world image from NTIRE real-world SR challenge [30] without ground truth. Note that for SRMD and USRNet, we directly use the ground truth kernel or noise level as additional inputs in order to validate the efficacy of the SR model alone. For RealSR and FSSR, we choose their pre-trained models trained with real images taken by mobile devices from DPED dataset [25]. In addition, we include two renowned non-blind SR models, SRResNet [4] and ESRGAN [58], into the comparison list for readers’ reference. Based on these testing examples, some important observations can be drawn as following:

1) For methods exploiting external dataset, their generalization largely depends on the coverage of degradation modelling or training data distribution. For example, approaches with explicit modelling can only handle noisy inputs if noise is directly covered as a degradation factor in the SR modelling, like SRMD, USRNet and BSRGAN. Otherwise, they will not be capable of noise removal and consequently bring unfavourable artifacts in SR results, like SRResNet and IKC. On the other hand, models trained on real image dataset hardly give out visually pleasing results on synthetic data, such as RealSR and FSSR.

2) Real-world images do include more complex degradations, which deviate a lot from some simple synthetic data distributions. Models with explicit modelling generally perform well on the synthetic images within their degradation coverage, like SRMD and IKC. However, except for BSRGAN and Real-ESRGAN, none of the methods generate satisfactory result for the third real image, including those with implicit modelling. Specifically, SRResNet and IKC tend to keep the noise texture and cause artifacts. SRMD and USRNet could alleviate the noisy artifacts to some extent but in the sacrifice of high-frequency details, and it also needs some effort

| Method       | PSNR(dB) | SSIM   | LPIPS | # |
|--------------|----------|--------|-------|---|
| AIM          | 40.54/0.7482/0.27 | 25.70/0.7487/0.20 |
| RealSR       | 40.54/0.7482/0.27 | 26.01/0.7482/0.39 |
| ESRGAN       | 40.54/0.7482/0.27 | 25.09/0.7459/0.41 |
| ZSSR         | 40.54/0.7482/0.27 | 25.99/0.7388/0.27 |
| CinCGAN      | 40.54/0.7482/0.27 | 26.78/0.7822/0.23 |
| FSSR         | 40.54/0.7482/0.27 | 25.60/0.6129/0.46 |
| DASR         | 40.54/0.7482/0.27 | 26.87/0.7640/0.34 |

Note that ESRGAN [58] is trained with paired data.

Fig. 16. Qualitative comparison of synthetic testing example with Gaussian blur.

Fig. 17. Qualitative comparison of synthetic testing example with Gaussian blur and noise.
to estimate kernel and noise level for a real image with unknown degradation. RealSR and FSSR do have better results in terms of removing noises and preserving sharp textures, yet they still generate fake textures or artifacts, leading to the unnatural-looking of SR images. Real-ESRGAN generally performs best in terms of restoration, but it also fails to generate uniformly sharp details in some local areas.

9.3 Suggestions on Fair Comparison

We would like to suggest that our readers make better use of our proposed taxonomy in their future work, especially for effective and fair comparison among different methods in their paper. One may first try to place their own method into the corresponding category, and then pay special attention to previous methods belonging to the same category for evaluation and comparison. As for methods out of the specific categorial scope, since they have employed different degradation modelling or data sources (external or internal), direct comparison may become unfair and unnecessary, sometimes causing difficulty and even confusion. Hence, we recommend here that future work may as well follow our taxonomy to make comparison among methods from the same category, both for proposing new solutions and benchmark setting.

10 Conclusion

In this article, we present a systematic survey on recent progress in blind image SR. In order to effectively classify and summarize existing methods, we propose a taxonomy according to their ways of degradation modelling and the data used to solve the SR model: explicit or implicit modelling with external dataset or a single LR image. Except implicit modelling with a single image, the other three categories all have representative existing approaches, and we make a conclusion on them as following:

- **Explicit modelling with external dataset:** representatives are SRMD and IKC, which utilize the classical degradation model or its variants for image-specific adaptation based on degradation information. Besides, complex degradation modelling, such as BSRGAN, is also attracting more attention. These methods perform well on degradation types covered in its modelling, but their performance will severely deteriorate on other degradations.

- **Explicit modelling with a single image:** including Michaeli & Irani and KernelGAN for blind kernel estimation, as well as ZSSR and DGDM-SR for SR. They leverage the internal statistics of natural images - patch recurrence across scales, which can also be theoretically derived from classical degradation model. However, these methods may fail for more general natural images with diverse or monotonous scenes, i.e., those without enough recurring clues.

- **Implicit modelling with external dataset:** such as CinCGAN, FSSR and DASR. These methods assume that real-world degradations are too complex to be explicitly modelled, but can be implicitly learned with data distribution learning under GAN framework. However, domain adaptation is a non-trivial task due to the large space of the natural image domain, making it hard to train a GAN-based model with good performance and generalization capacity. **Prospect:** Implicit modelling with a single image, which has not been proposed yet, is a direction worth exploring in future research, especially for general natural images with complex degradations and without strong internal statistics. One possible solution to this problem is utilizing human intervention to provide additional information as SR prior, and restoration network with modulation or manually choosing an HR reference image may be of help. Besides, since LR images could suffer from distinct degradation types in real-world scenarios, designing specific degradation modelling for certain scenes also deserves more attention from the research community. We hope this paper can inspire some new ideas and make contributions to the prosperity of blind image SR.

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis., 2014, pp. 184–199.
[2] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 391–407.
[3] W. Shi et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1874–1883.
[4] C. Ledig et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 4681–4690.
[5] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 286–301.
[6] S. Bell-Kligler, A. Shocher, and M. Irani, “Blind super-resolution kernel estimation using an internal-GAN,” in Proc. Int. Conf. Neural Inf. Process. Syst., 2019, pp. 284–293.

[7] J. Gu, H. Lu, W. Zuo, and C. Dong, “Blind super-resolution with iterative kernel correction,” in Proc. IEEE Conf. Comput. Vis. Pat- tern Recognit., 2019, pp. 1604–1613.

[8] Y. Yuan, S. Liu, Z. Yang, Z. Zhang, C. Dong, and L. Lin, “Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2018, pp. 701–710.

[9] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, “FSRNNet: End-to-end learning face super-resolution with facial priors,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 2492–2501.

[10] Y. Xu, R. Fernández-Pérdikli, and R. Hartsel, “Face super-resolution guided by facial component heatmaps,” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 217–233.

[11] H. Huang, R. He, Z. Sun, and T. Tan, “Wavelet-SRNNet: A wavelet-based CNN for multi-scale face super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 1689–1697.

[12] A. Bulat and G. Tzimiropoulos, “Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 109–117.

[13] C. Ma, Z. Jiang, Y. Rao, J. Lu, and J. Zhou, “Deep face super-resolution with iterative collaboration between attention recovery and landmark estimation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 3569–3578.

[14] T.-W. Hui, C. C. Loy, and X. Tang, “Depth map super-resolution by deep multi-scale guidance,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 353–369.

[15] C. Guo, C. Li, J. Guo, R. Cong, H. Fu, and P. Han, “Hierarchical features driven residual learning for depth map super-resolution,” IEEE Trans. Image Process., vol. 28, no. 5, pp. 2545–2557, May 2019.

[16] X. Song et al., “Channel attention based iterative residual learning for depth map super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., 2020, pp. 5631–5640.

[17] J. Lei, L. Li, H. Yue, F. Wu, N. Ling, and C. Hou, “Depth map super-resolution considering view synthesis quality,” IEEE Trans. Image Process., vol. 26, no. 4, pp. 1732–1745, Apr. 2017.

[18] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, “Designing a practical degradation model for deep blind image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2021, pp. 4791–4800.

[19] X. Wang, L. Xie, C. Dong, and Y. Shan, “Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data,” in Proc. IEEE Int. Conf. Comput. Vis., 2021, pp. 1905–1914.

[20] Y. Huang et al., “Unfolding the alternating optimization for blind super-resolution,” Proc. Conf. Neural Inf. Process. Syst., 2020, pp. 5056–5064.

[21] K. Zhang, W. Zuo, and L. Zhang, “Learning a single convolutional super-resolution network for multiple degradations,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3262–3271.

[22] K. Zhang, L. V. Gool, and R. Timofte, “Deep unfolding network for image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 3214–3223.

[23] S. Gu, A. Lugmayr, M. Danelljan, M. Fritsche, J. Lamour, and R. Timofte, “DIViSK: Diverse 8k resolution image dataset,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, 2019, pp. 3512–3516.

[24] J. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 5197–5206.

[25] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoevy, and L. V. Gool, “Dslr-quality photos on mobile devices with deep convolutional networks,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 3297–3305.

[26] A. Shocher, N. Cohen, and M. Irani, “zero-shot super-resolution using deep internal learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3118–3126.

[27] P. Liu, H. Zhang, Y. Cao, S. Liu, D. Ren, and W. Zuo, “Learning cascaded convolutional networks for blind single image super-resolution,” Neurocomputing, vol. 417, pp. 371–383, 2020.

[28] X. Chang, B. Z. Fu, D. Ghanem, F. Porikli, “HarMon: Image super-resolution with depth guided internal degradation learning,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 265–280.

[29] I. Goodfellow et al., “Generative adversarial nets,” in Proc. Int. Conf. Neural Inf. Process. Syst., 2014, pp. 2672–2680.
J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

G. Freedman and R. Fattal, “Image and video upscaling from local self-examples,” ACM Trans. Graph., vol. 30, no. 2, pp. 12:1–12:11, 2011.

D. Glasner, S. Bagon, and M. Irani, “Super-resolution from super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis., 2009, pp. 349–356.

J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1646–1654.

X. Wang et al., “ESRGAN: Enhanced super-resolution generative adversarial networks,” in Proc. Eur. Conf. Comput. Vis. Workshops, 2018, pp. 63–79.

B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced residual networks for single image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2017, pp. 1132–1140.

T. Dong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 4809–4817.

J. Kim, J. K. Lee, and K. M. Lee, “Deeply-recursive convolutional networks for super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1637–1645.

Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2790–2798.

T. Dai, J. Cai, Y. Zhang, S. Xia, and L. Zhang, “Second-order attention network for single image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3086–3095.

M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, “EnhanceNet: Single image super-resolution through automated texture synthesis,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 4501–4510.

W. Zhang, Y. Liu, C. Dong, and Y. Qiao, “RankSRGAN: Generative adversarial networks with ranker for image super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 3096–3105.

A. Brifman, Y. Romano, and M. Elad, “Turning a denoiser into a super-resolver using plug and play priors,” in Proc. IEEE Int. Conf. Image Process., 2016, pp. 1404–1408.

S. H. Chan, X. Wang, and O. A. Elgendi, “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications,” IEEE Trans. Comput. Imag., vol. 3, no. 1, pp. 84–98, Mar. 2017.

K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2808–2817.

S. V. Venkatashrini, C. A. Bouman, and B. Wohlberg, “Plug-and-play priors for model based reconstruction,” in Proc. IEEE Glob. Conf. Signal Informat. Process., 2013, pp. 945–948.

S. Boyd et al., “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, 2011.

X. Wang, K. Yu, C. Dong, and C. Change Loy, “Recovering realistic texture in image super-resolution by deep spatial feature transform,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 606–615.

X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 1501–1510.

A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a "completely blind" image quality analyzer,” IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, Mar. 2013.

S. A. Hussein, T. Tirer, and R. Giry, “Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 1428–1437.

S. Zhang and M. Irani, “Internal statistics of a single natural image,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 977–984.

T. Michaeli and M. Irani, “Nonparametric blind super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 945–952.

L. Dinh, D. Krueger, and Y. Bengio, “NICE: Non-linear independent components estimation,” in Proc. Int. Conf. Learn. Representations, Mar. 2015, pp. 1–13.

L. Dinh, J. Sohl-Dickstein, and S. Bengio, “Density estimation using real NVP,” in Proc. Int. Conf. Learn. Representations, 2017, pp. 1–32.

J. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2242–2251.

R. Timofte et al., “NTire 2018 challenge on single image super-resolution: Methods and results,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2018, pp. 852–863.

Y. Zhang, S. Liu, C. Dong, X. Zhang, and Y. Yuan, “Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution,” IEEE Trans. Image Process., vol. 29, pp. 1101–1112, 2020.

D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 9446–9454.

J. He, C. Dong, and Y. Qiao, “Modulating image restoration with continual levels via adaptive feature modification layers,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 11 056–11 064.

G. Shim, J. Park, and I. S. Kweon, “Robust reference-based super-resolution with single image super-resolution priors,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 8425–8434.

L. Lu, W. Li, X. Tao, J. Lu, and J. Jia, “MASA-SR: Matching acceleration and spatial adaptation for reference-based image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2021, pp. 6368–6377.

E. Agustsson and R. Timofte, “NTire 2017 challenge on single image super-resolution: Dataset and study,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2017, pp. 126–135.

R. Timofte et al., “NTIRE 2017 challenge on single image super-resolution: Methods and results,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2017, pp. 1110–1121.

M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi-Morel, “Low-complexity single-image super-resolution based on non-negative neighbor embedding,” in Proc. Brit. Mach. Vis. Conf., 2012, pp. 1–10.

R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” in Proc. Int. Conf. Curves Surf., 2010, pp. 711–730.

D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2001, pp. 416–423.

C. Chen, Z. Xiong, X. Tian, Z. Zha, and F. Wu, “Camera lens super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 1652–1660.

P. Wei et al., “Component divide-and-conquer for real-world image super-resolution,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 101–117.

J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, “Toward real-world single image super-resolution: A new benchmark and a new model,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, 2019, pp. 3575–3583.

X. Wang, K. Yu, T.-W. Hui, C. Dong, L. Lin, and L. C. Change, “Deep poly-dense network for image super-resolution,” NTIRE Challenge, 2019.

J. Cai et al., “NTIRE 2019 challenge on real image super-resolution: Methods and results,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop, 2019, pp. 2211–2223.

X. Liu, Y. Ding, D. He, C. Li, Y. Fu, and S. Wen, “ Cascaded U-shaped deep super-resolution,” NTIRE Challenge, 2019.

A. Lugmayr et al., “AIM 2019 challenge on real-world image super-resolution: Methods and results,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, 2019, pp. 101–117.

Z. Pan et al., “Real image super resolution via heterogeneous model ensemble using GP-NAS,” in Proc. Eur. Conf. Comput. Vis. Workshops, 2020, pp. 423–436.

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 586–595.
Yihao Liu received the BS degree from the University of Chinese Academy of Sciences, Beijing, in 2018. He is now working towards the PhD degree with Multimedia Laboratory, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences. He is supervised by Prof. Yu Qiao and Prof. Chao Dong. His research interests include computer vision and image/video enhancement.

Jinjin Gu received the BEng degree in computer science and engineering from the Chinese University of Hong Kong, Shenzhen, in 2020. He is currently working toward the PhD degree in Engineering and IT with the University of Sydney. His research interests include computer vision, image processing, interpretability of deep learning algorithms and the applications of machine learning in industrial.

Yu Qiao (Senior Member, IEEE) is currently a professor with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Science, and the director with the Institute of Advanced Computing and Digital Engineering. He has published more than 180 articles in international journals and conferences, including IEEE Transactions on Pattern Analysis and Machine Intelligence, International Journal of Computer Vision, IEEE Transactions on Image Processing, IEEE Transactions on Signal Processing, CVPR, and ICCV. His research interests include computer vision, deep learning, and bioinformation. He received the First Prize of the Guangdong Technological Invention Award, and the Jiaxi Lv Young Researcher Award from the Chinese Academy of Sciences. His group achieved the first runner-up at the ImageNet Large Scale Visual Recognition Challenge 2015 in scene recognition, and the winner at the ActivityNet Large Scale Activity Recognition Challenge 2016 in video classification. He has served as the program chair of the IEEE ICIST 2014.

Chao Dong received the PhD degree from The Chinese University of Hong Kong in 2016. He is currently an associate professor with the Shenzhen Institute of Advanced Technology, Chinese Academy of Science. In 2014, he first introduced deep learning method – SRCNN into the super-resolution field. This seminal work was chosen as one of the top ten “Most Popular Articles” of TPAMI in 2016. His team has won several championships in international challenges -- NTIRE2018, PIRM2018, NTIRE2019, NTIRE2020 AIM2020 and NTIRE2022. He worked in SenseTime from 2016 to 2018, as the team leader of Super-Resolution Group. In 2021, he was chosen as one of the World’s Top 2% Scientists. In 2022, he was recognized as the AI 2000 Most Influential Scholar Honorable Mention in computer vision. His current research interest focuses on low-level vision problems, such as image/video super-resolution, denoising, and enhancement.

For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csdl.