IDENet: Implicit Degradation Estimation Network for Efficient Blind Super Resolution

Asif Hussain Khan
University of Udine
Udine, Italy
khan.asifhussain@spes.uniud.it

Christian Micheloni
University of Udine
Udine, Italy
christian.micheloni@uniud.it

Niki Martinel
University of Udine
Udine, Italy
niki.martinel@uniud.it

Abstract

Blind image super-resolution (SR) aims to recover high-resolution (HR) images from low-resolution (LR) inputs hindered by unknown degradation. Existing blind SR methods exploit computationally demanding explicit degradation estimators hinging on the availability of ground-truth information about the degradation process, thus introducing a severe limitation in real-world scenarios where this is inherently unattainable. Implicit degradation estimators avoid the need for ground truth but perform poorly. Our model reduces this performance gap with (i) a novel loss component to implicitly learn the degradation kernel from the LR input only, and (ii) a novel learnable Wiener filter module that exploits the learned degradation kernel to efficiently solve the deconvolution task via a closed-form solution formulated in the Fourier domain. Systematic experiments show that our proposed approach outperforms existing implicit blind SR methods ($3dB$ PSNR gain and 8.5% SSIM improvement on average) and achieves comparable performance to explicit blind SR methods ($0.6dB$ and 0.5% difference in PSNR and SSIM, respectively). Remarkably, these results are obtained using 33% and 71% less parameters than implicit and explicit methods.

1. Introduction

Image super-resolution (SR) is the task of enhancing low-resolution (LR) to high-resolution (HR) images. It finds a plethora of applications in various domains [7, 17, 23, 26, 29, 49], especially where capturing an HR image is constrained by physical factors such as hardware limitations or bandwidth constraints.

The degradation process that relates a pair of LR and HR images can be formally defined as

$$\mathbf{I}_{LR} = (\mathbf{I}_{HR} \circledast \mathbf{K}) \downarrow_S + n \tag{1}$$

where $\mathbf{I}_{LR}$ is the degraded LR image resulting from an i.i.d. white Gaussian noise n added to the down-sampling $\downarrow_S$, with scaling factor S, of the convolution ($\circledast$) of the HR image $\mathbf{I}_{HR}$ with kernel $\mathbf{K}$.

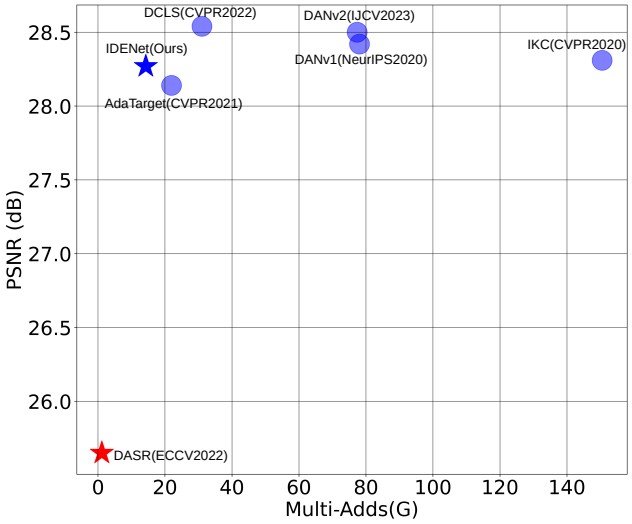

Figure 1. PSNR and Multi-Add(G) comparisons of implicit (stars) and explicit (dots) blind super-resolution methods. Results on the Set14 dataset for scale factor ($\times 4$).

In recent years, Convolution Neural Networks (CNNs) have become widely used for SR [7, 17, 17, 24, 32, 45, 45, 58–60]. These have achieved relevant performance improvements compared to traditional methods [6, 11] but face two fundamental challenges: 1) CNNs apply the same convolution kernels across the image, lacking context-specific adaptability; 2) CNNs are limited in capturing global context due to its focus on local processing; To address these challenges, Transformers [42], capturing long-range feature dependencies, have been successfully explored [24, 53].

Despite the success of such architectures in SR, a vast majority of methods [35, 36, 51, 55] assume a fixed and ideal blur kernel in (1)–*i.e.*, the bicubic kernel, referred to as *non-blind* SR (Figure 3(a)). In real-world scenarios with unknown blur kernels [3, 10, 52, 55], these methods significantly underperform. Thus, addressing the challenge of handling unknown blur kernels, commonly referred to as *blind SR*, has become a focal point of research in the field.

Existing blind SR methods (*e.g.*, [27, 44]) either assume the image degradation process outlined in (1) or consider a broader range of degradation factors, such as blur, noise, and JPEG compression (*e.g.*, [46, 58]). Most of the existing blind SR methods (*e.g.*, [14, 25, 27, 44, 46]) introduce *explicit* degradation estimators that are computationally intensive, rely on extensive parametrization, and, more importantly, hinge on ground-truth degradation information (Figure 3(b)). However, these ground-truth degradations are impossible to obtain or unavailable in real-world scenarios. This opened the recent exploration of *implicit* degradation estimators [25], modeling the degradation process without explicit ground-truth signals Figure 3(c)).

This paper introduces a novel computationally efficient blind SR approach that overcomes the challenges of explicit degradation estimation. The primary challenge in explicit blind SR lies in its dependency on ground truth degradation kernels, which are unattainable in practical, real-world environments. We introduced a novel loss and an implicit degradation estimator to elegantly address this issue. The novel loss directly exploits the degradation process in (1) to guide the estimator toward learning the degradation kernel directly from the low-resolution (LR) input image, thus eliminating the need for ground truth kernel information. The secondary challenge explicit blind SR approaches suffer from is introduced by the computationally demanding models that are required to perform the upscaling operation. Existing models hinge on multiple upscaling layers (*e.g.*, transposed convolution, pixel unshuffling, etc.) that are required to gradually increase the spatial resolution of a given input, hence having an impact on the network depth (*i.e.*, on the number of learnable parameters). With a novel learnable Wiener filter module performing deconvolution in the Fourier domain, we introduce an approach that bypasses such a requirement by working on any bilinear upscaled input to efficiently generate a high-resolution (HR) image. This module exploits the implicit degradation estimator and easily adapts to various degradation kernels by jointly learning multiple deconvolution filters. To enhance the HR image reconstruction further, we also incorporated an efficient transformer-based refinement module exploiting long-range pixel dependencies that are relevant to SR. As shown in Figure 1, combining such three novel components outperforms the existing implicit blind SR method and closes the gap with explicit SR techniques [14, 16, 25, 27], with notably reduced number of parameters.

Our contributions can be summarized as follows:

- We propose a novel loss for blind SR that is exploited by our proposed kernel estimation module to predict the degradation kernel without the need for ground truth information.
- By reformulating the problem of learning a Wiener filter in the Fourier domain, for which we derived a closed-form solution, we introduce a novel module that can adapt to multiple degradations while performing efficient deconvolution. Since the deconvolution is performed on a bilinear upscaled input, our approach removes the need for upscaling layers, thus enabling the generation of HR images with any scale factor at zero cost.
- Through extensive experiments, we show that our efficient approach outperforms the existing implicit blind SR methods, with $3dB$ PSNR and 8.5% SSIM gain (on average), and achieves comparable performance to explicit methods while having 33% and 71% fewer parameters, respectively.

2. Related Work

2.1. Non-Blind Super Resolution

Over the past few years, several non-blind SR techniques [7, 17, 29, 49, 56] have achieved excellent results on benchmark datasets. However, their performance drops when there is a gap between training and testing degradations. Some methods address this by using additional ground truth information like blur kernels and noise levels (*e.g.*, [35, 36, 51, 55–57]) as inputs. While these techniques can handle multiple degradation types with a single model, they rely on accurately estimating degradation parameters. Differently, our approach follows a blind super-resolution method, requiring no prior knowledge of image degradations.

2.2. Blind Super Resolution

Several blind SR methods have been introduced to address the challenge of handling unknown degradation(s).

Explicit degradation models (see Figure 3(b)) hinge on the availability of ground-truth kernels. Some methods [10, 26] combine a blur kernel estimator with SR networks, making the model adaptable to images degraded by various blur kernels [5, 20]. Others [27] introduced a reformulated degradation model to enhance kernel estimation and high-resolution restoration. These methods are specific to certain degradation types and require ground truth labels for multiple degradations, which are hard (if not impossible) to obtain in real-world settings. They can also be computationally demanding, involving two or more networks [10, 14], thus denying efficient inference. In contrast, we do not require knowledge of the ground truth degradation factors that are implicitly estimated through our novel loss by a single lightweight model.

Implicit degradation models (see Figure 3(c)) does not hinge on the availability of ground truth degradation kernels [18, 19, 61]. The initial exploration of this emerging approach was conducted in [25], which exploited metric-based learning to distinguish between various degradation types. Differently, we propose a kernel estimator with a novel loss formulation and a learnable Wiener filter that al-

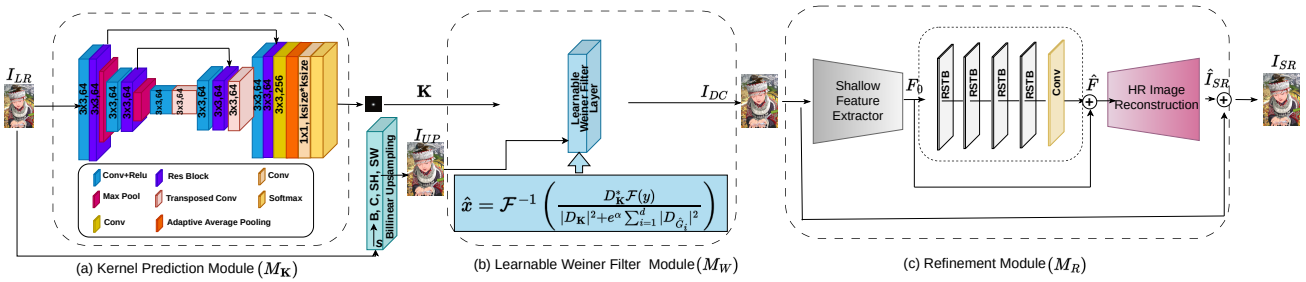

Figure 2. Illustration of the proposed IDENet. An LR image $\mathbf{I}_{LR}$ is fed to the (a) Kernel Prediction Module, which generates a degradation kernel $\mathbf{K}$. Then the (b) Learnable Wiener Filter Module exploits the upsampled image $\mathbf{I}_{UP}$ and the predicted kernel $\mathbf{K}$ to generate $\mathbf{I}_{DC}$. This is finally considered by the (c) Refinement Module to estimate the super-resolved image $\mathbf{I}_{SR}$.

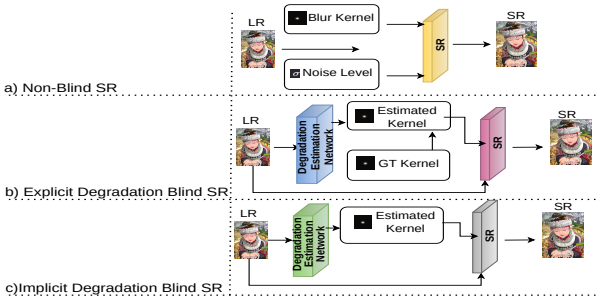

Figure 3. Categorization of existing SR approaches. a) Non-blind SR methods assume and use known degradation information during the SR process. b) Blind-SR methods exploit ground-truth degradation information to model it with full supervision. c) Blind SR methods implicitly estimate the degradation information to guide the SR process without the need for ground-truth supervision. Our proposed IDENet follows this approach.

low us to learn diverse degradation types exploited via efficient deconvolution in the Fourier domain.

2.3. Wiener Filter

In the context of SR, some recent works have explored the capabilities of the Wiener filter for deconvolution. In [37], Wiener deconvolution was introduced as an initial image preprocessing step, later employed in the feature space [8]. In [40], classical Wiener deconvolution was combined with a conventional CNN-based approach, relying on ground truth kernels during training and testing. All these methods have three notable limitations since they: (i) consider non-learnable Wiener filters, (ii) are limited by the inductive bias of CNNs, and (iii) assume a fixed and ideal kernel, hence fall in the category of non-blind SR methods. In contrast, we propose a novel learnable Wiener filter designed for implicit blind SR. We also perform the deconvolution with a closed-form approach in the Fourier domain while leveraging a self-attention mechanism to capture long-range pixel dependencies for HR image reconstruction.

3. Method

Our blind-SR approach, shown in Figure 2, consists of three main modules: the Kernel Prediction Module (KPM), the Learnable Wiener Filter Module (LWFM), and the Refinement Module (RM). The KPM implicitly estimates the degradation kernel $\mathbf{K} \in \mathbb{R}^{k\times k}$ from the LR input $\mathbf{I}_{LR} \in \mathbb{R}^{C\times H\times W}$, where C denotes the number of channels, H and W are spatial dimensions. With a parallel stream, $\mathbf{I}_{LR}$ is then upsampled (via bilinear interpolation with scale factor S) to generate $\mathbf{I}_{UP} \in \mathbb{R}^{C\times SH\times SW}$. $\mathbf{I}_{UP}$ and $\mathbf{K}$ are the inputs to the LWFM. We derived a novel formulation within such a module that leverages efficient operation in the frequency domain to learn the Wiener filter parameters with a closed-form solution. This computationally efficient approach yields to $\mathbf{I}_{DC} \in \mathbb{R}^{C\times SH\times SW}$ that the RM finally exploits to generate the SR image $\mathbf{I}_{SR} \in \mathbb{R}^{C\times SH\times SW}$.

3.1. Kernel Prediction Module (M_K)

This module estimates the degradation kernel $\mathbf{K} = M_{\mathbf{K}}(\mathbf{I}_{LR}; \theta_{M_{\mathbf{K}}})$ needed to effectively leverage the capabilities of the subsequent LWFM for deconvolution. $\theta_{M_{\mathbf{K}}}$ represent the module learnable parameters. We designed such a module as an encoder-decoder architecture (shown in Figure 2(a)). The encoder employs a combination of 3×3 Conv2D layers generating 64 feature maps via ReLU nonlinearity, followed by a set of 3×3 kernel residual blocks with skip connections and MaxPool operators. The decoder leverages the informative features computed by the encoder through two sets of 3×3 Conv2D layers, residual blocks with skip connections, and transposed convolutions for kernel prediction.

In our setup, as done by [10, 14, 27], we assume that the kernel downgrading HR to LR as in (1) is defined by a Gaussian probability distribution (that is also exploited when artificially generating LR images). Following the spirit of these existing methods, we added a softmax layer to ensure that the predicted degradation kernel $\mathbf{K}$ satisfies the probability distribution characteristics. It is worth noticing that this also prevents generation of negative (*i.e.*, invalid) pixel

values that may otherwise occur when the a kernel violating the non-negativity constraint is used (*e.g.*, the widely-used bicubic kernel).

3.2. Learnable Wiener Filter Module (M_W)

We introduce a learnable Wiener filter module to address the blurring effects and to efficiently exploit the predicted degradation kernel, through deconvolution. Deconvolution is an ill-posed inverse problem that can be addressed via a variational approach, *i.e.*, by finding the minima or maxima of a function via small changes in functions or functionals (variational derivatives). Following such an intuition, we derive a novel formulation that adapts to various degradation kernels learned through a closed-form solution.

We begin by assuming that the deconvolution task can be addressed by a module computing $\mathbf{I}_{DC} = M_W(\mathbf{I}_{UP}, \mathbf{K}; \theta_{M_W})$, where θ_{M_W} are learnable parameters. To learn the model parameters, for the sake of mathematical derivation and without loss of generality, we assume that $\mathbf{I}_{UP}$ that has been raster scanned in lexicographical order to obtain $\mathbf{y} \in \mathbb{R}^N$. This allows us to define our optimization objective as finding the variational approach

$$\hat{\mathbf{x}} = \underbrace{\arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \hat{\mathbf{K}}\mathbf{x}\|_2^2 + \frac{\alpha}{2}\sum_{i=1}^{d}\|\hat{\mathbf{G}}_i\mathbf{x}\|_2^2}_{F(\mathbf{x})} \tag{2}$$

where $\hat{\mathbf{K}} \in \mathbb{R}^{N\times N}$ is the point spread function (PSF) (obtained by symmetrically padding $\mathbf{K}$) and $\hat{\mathbf{G}}_i \in \mathbb{R}^{N\times N}$ is a learnable convolution kernel.

The minimization problem in (2) has a closed-form solution that corresponds to the Wiener-Kolmogorov deconvolution filter [48]. This is

$$\hat{\mathbf{x}} = (\hat{\mathbf{K}}^\top\hat{\mathbf{K}} + \alpha\sum_{i=1}^{d}\hat{\mathbf{G}}_i^\top\hat{\mathbf{G}}_i)^{-1}\hat{\mathbf{K}}^\top\mathbf{y} \tag{3}$$

where $\hat{\mathbf{K}}^\top$ and $\hat{\mathbf{G}}_i^\top$ denote the adjoint matrices for $\hat{\mathbf{K}}$ and $\hat{\mathbf{G}}_i$, respectively. Solving (3) requires the inversion of a large matrix, which can be computationally slow. We reformulate the closed-form solution in the Fourier domain to address this issue. This makes finding the Wiener filter a fast and efficient method, enabling the restoration of the underlying signal with low computational complexity.

Assuming periodic image boundary conditions, we can treat $\hat{\mathbf{K}}$ and $\hat{\mathbf{G}}_i$ as circulant matrices that can be diagonalized in the Fourier domain as:

$$\begin{aligned} \hat{\mathbf{K}} &= \boldsymbol{\mathcal{F}}^{-1}\mathbf{D}_{\hat{\mathbf{K}}}\boldsymbol{\mathcal{F}} \qquad \mathbf{D}_{\hat{\mathbf{K}}} = \boldsymbol{\mathcal{F}}\mathbf{S}_{\hat{\mathbf{K}}}\mathbf{P}_{\hat{\mathbf{K}}}\mathbf{k} \\ \hat{\mathbf{G}}_i &= \boldsymbol{\mathcal{F}}^{-1}\mathbf{D}_{\hat{\mathbf{G}}_i}\boldsymbol{\mathcal{F}} \qquad \mathbf{D}_{\hat{\mathbf{G}}_d} = \boldsymbol{\mathcal{F}}\mathbf{S}_{\hat{\mathbf{G}}_i}\mathbf{P}_{\hat{\mathbf{G}}_i}\mathbf{g}_i \end{aligned} \tag{4}$$

where $\boldsymbol{\mathcal{F}} \in \mathbb{C}^{N\times N}$ and $\boldsymbol{\mathcal{F}}^{-1} \in \mathbb{C}^{N\times N}$ are the Fourier matrix (DFT) and its inverse, respectively. $\mathbf{D}_{\hat{\mathbf{K}}}$ and $\mathbf{D}_{\hat{\mathbf{G}}_i} \in \mathbb{C}^{N\times N}$ are the diagonal matrices, $\mathbf{S}_{\hat{\mathbf{K}}}$ and $\mathbf{S}_{\hat{\mathbf{G}}_i} \in \mathbb{R}^{N\times N}$ are the corresponding circular shift operators. Corresponding zero-padding operators are $\mathbf{P}_{\hat{\mathbf{K}}} \in \mathbb{R}^{N\times M}$ and $\mathbf{P}_{\hat{\mathbf{G}}_i} \in \mathbb{R}^{N\times Li}$. $\mathbf{k} \in \mathbb{R}^M$ and $\mathbf{g}_i \in \mathbb{R}^{Li}$ are the blurring kernel and the regularization convolution kernel.

After reformulation of the problem in the Fourier domain, (3) can be rewritten as:

$$\hat{\mathbf{x}} = \boldsymbol{\mathcal{F}}^{-1}\left(\frac{\mathbf{D}_{\hat{\mathbf{K}}}^*\boldsymbol{\mathcal{F}}(\mathbf{y})}{|\mathbf{D}_{\hat{\mathbf{K}}}|^2 + e^{\alpha}\sum_{i=1}^{d}|\mathbf{D}_{\hat{\mathbf{G}}_\mathbf{i}}|^2}\right) \tag{5}$$

with $\mathbf{D}_{\hat{\mathbf{K}}}^*$ denoting the Hermitian transpose of the $\mathbf{D}_{\hat{\mathbf{K}}}$, and division is performed element-wise. Following [34], we consider the trade-off coefficient α a parameter to be learned together with the kernels during training.

3.3. Refinement Module (M_R)

To improve the LWFM deconvolution output $\mathbf{I}_{DC}$, which is based on a shallow network, we introduce a refinement module computing $\mathbf{I}_{SR} = M_R(\mathbf{I}_{DC}; \theta_{M_R})$. As shown in Figure 2, it has one component for shallow features, another for extracting deep features via a self-attention mechanism capturing long-range dependencies, and a last one for the generation of $\mathbf{I}_{SR}$. Parameters of all such components are denoted as θ_{M_R}.

Shallow and deep features extraction components start with a 3×3 Conv2D layer extracting shallow features $\mathbf{F}_0 \in \mathbb{R}^{C\times SH\times SW}$. We adopted a single convolution layer cause it has been proven beneficial in initial visual processing, enhancing optimization stability and superior outcomes [50]. Deep features are efficiently obtained through γ Residual Swin Transformer blocks (RSTB) [24], denoted as $L_{RSTB_i}(\cdot)$, followed by a 3×3 Conv2D layer, denoted as L_{conv}. We compute the deep features

$$\hat{\mathbf{F}} = L_{conv}(\mathbf{F}_\gamma) \in \mathbb{R}^{C\times SH\times SW} \tag{6}$$

with $\mathbf{F}_i = L_{RSTB_i}(\mathbf{F}_{i-1})$, for $i = 1, 2, 3, .., \gamma$.

HR image reconstruction is performed by aggregating the shallow and deep features as

$$\hat{\mathbf{I}}_{SR} = L_{REC}(\mathbf{F}_0 + \hat{\mathbf{F}}) \tag{7}$$

where $L_{REC}(\cdot)$ is 3×3 Conv2D reconstruction layer.

As discussed in [24], shallow features predominantly encompass the low-frequency components, whereas deep features concentrate on restoring the high-frequency details that may have been lost. By incorporating the long skip connection, the refinement module facilitates the direct transmission of low-frequency information to the reconstruction module. This arrangement assists the deep feature extraction module in prioritizing high-frequency information and enhancing training stability.

To finally obtain the super-resolved image $\mathbf{I}_{SR}$, we adopted a residual learning approach computing $\mathbf{I}_{SR} =$

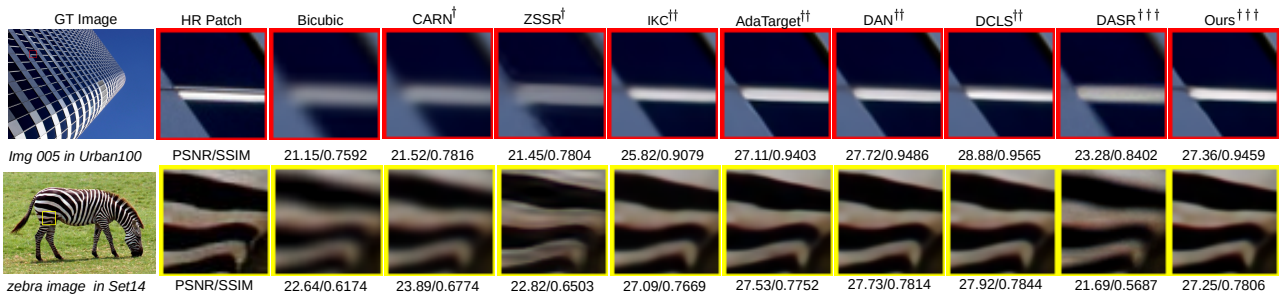

Figure 4. Visual comparison of *img 005* and *zebra* ($\times 4$ SR factor) from Urban100 and Set14 dataset. The isotropic kernel widths are set to 2.6 and 1.8, respectively.

$\hat{\mathbf{I}}_{SR} + \mathbf{I}_{DC}$. By doing this, we stabilize the gradients, achieving more efficient training and improved performance.

3.4. Optimization

We train our model using an end-to-end strategy optimizing

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{SR} + \lambda_2 \mathcal{L}_k + \lambda_3 \mathcal{L}_{TV} \tag{8}$$

where

$$\mathcal{L}_{SR} = \|\mathbf{I}_{HR} - \mathbf{I}_{SR}\|_1 \tag{9}$$

is the image reconstruction loss quantifying the accuracy of the super-resolved output $\mathbf{I}_{SR}$ against the HR ground-truth $\mathbf{I}_{HR} \in \mathbb{R}^{C \times SH \times SW}$.

The novel kernel estimation loss, $\mathcal{L}_k$, is a key component in our work. Considering the blurring process assumption in (1), we designed the function

$$\mathcal{L}_k = \|\mathbf{I}_{LR} - \left(\mathbf{I}_{HR} \circledast M_{\hat{\mathbf{K}}}(\mathbf{I}_{LR})\right) \downarrow_S \|_1 \tag{10}$$

which, through the L1 penalty, forces the model to learn a kernel that convolved with the ground truth HR would generate the same LR input sample, thus bypassing the need for explicit kernel definition.

To complement this, the total variation loss

$$\mathcal{L}_{TV} = \|\nabla \mathbf{I}_{HR} - \nabla \mathbf{I}_{SR}\|_1 \tag{11}$$

computes the difference between horizontal and vertical gradients (denoted with ∇) to encourage smoothness of the image by minimizing the variations in pixel intensities, hence reducing noise and unwanted artifacts. λ_1, λ_2 and λ_3 are balancing parameters.

4. Experiments

4.1. Datasets

Following [10, 14, 27], for all the experiments, we trained our model on 3450 2K HR images combined from DIV2K [1] and Flickr2K [39]. We adopted the protocol of [10, 14, 27] to synthesize LR images using the next two settings.

Isotropic Gaussian kernels are used to blurry downsampled ground-truth images to obtain the corresponding LR samples. We followed [10] and generated 21×21 filters with a kernel width uniformly sampled from $[0.2, 2.0]$, and $[0.2, 4.0]$ for SR scale factors $\times 2$ and $\times 4$, respectively. For evaluation, we used the Gaussian8 [10, 27] kernel setting [10] on five SR benchmarks: Set5 [4], Set14 [54], BSD100 [30], Urban100 [13] and Manga109 [31].

Anisotropic Gaussian kernels of size 11×11 and 31×31 for scale factors $\times 2$ and $\times 4$ are considered, as defined in [3]. During training, the anisotropic Gaussian kernels are generated by first selecting a random kernel width in $(0.6, 5)$, then by applying a random rotation in $[-\pi, \pi]$. We also apply uniform multiplicative noise and normalize it to sum to one. For evaluation, we used the DIV2KRK [3] dataset.

4.2. Implementation Details[1]

We trained our model with 32×32 LR image patches for 500k iterations. We used a batch size of 12 with the Adam optimizer [21] having $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. We set the learning rate to 10^{-4}, then reduced it by a factor of 2 after 250k, 400k, 450k, and 475k iterations. All Conv2D layers and residual blocks in the model have 3×3 kernels producing 64 output feature maps, except for L_{REC} emitting 3 feature maps. Within the RM module, we use $\gamma = 4$ RSTB blocks, each composed of 6 STL layers [24] with 96 feature maps. The LWFM considered $d = 24$ with a learnable wiener filter size of 5×5 [40], whose weights are initialized using the discrete cosine transform (DCT) basis. Random vertical and horizontal flipping and 90° rotations were used as data augmentation strategies. Following [41], we set $\lambda_1 = 10.0$ while assigned $\lambda_2 = \lambda_3 = 1.0$, respectively. We report on the PSNR and SSIM [47] evaluation metrics computed for the luminance channel of the YCbCr color space.

[1] Code will be made available upon acceptance.

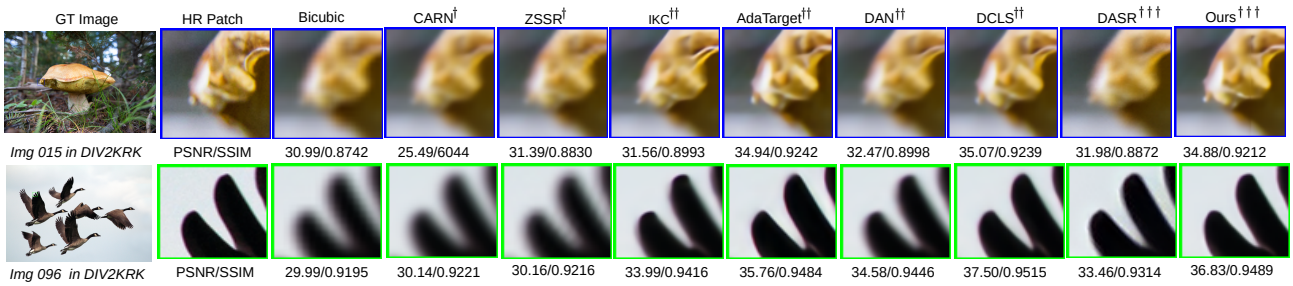

Figure 5. Visual comparison of *img 015* and *img 096* ($\times 4$ SR factor) from DIV2KRK dataset.

Degradation Estimation Approach	Method	Scale	Set5 PSNR↑/SSIM↑	Set14 PSNR↑/SSIM↑	BSD100 PSNR↑/SSIM↑	Urban100 PSNR↑/SSIM↑	Manga109 PSNR↑/SSIM↑	#Params (M) ↓	Multi-Adds (G)	Inference Time (s)
Non Blind + Blind SR †	Bicubic	x2	28.82/0.8577	26.02/0.7634	25.92/0.7310	23.14/0.7258	25.60/0.8498	-	-	-
	CARN[2]		30.99/0.8779	28.10/0.7879	26.78/0.7286	25.27/0.7630	26.86/0.8606	-	-	-
	Bicubic + ZSSR[35]		31.08/0.8786	28.35/0.7933	27.92/0.7632	25.25/0.7618	28.05/0.8769	-	-	-
	Deblurring[33] + CARN[2]		24.20/0.7496	21.12/0.6170	22.69/0.6471	18.89/0.5895	21.54/0.7946	-	-	-
	CARN[2]+ Deblurring[33]		31.27/0.8974	29.03/0.8267	28.72/0.8033	25.62/0.7981	29.58/0.9134	-	-	-
	MogaSRN[9]		38.19/0.9611	33.82/0.9196	32.30/0.9013	32.72/0.9340	39.16/0.9779	-	-	-
	Omni-SR[43]		38.22/0.9613	33.98/0.9210	32.36/0.9020	33.05/0.9363	39.28/0.9784	-	-	-
Explicit Blind SR ††	IKC[10]	x2	37.19/0.9526	32.94/0.9024	31.51/0.8790	29.85/0.8928	36.93/0.9667	5.32	-	-
	DANv1[14]		37.34/0.9526	33.08/0.9041	31.76/0.8858	30.60/0.9060	37.23/0.9710	4.33	-	-
	DANv2[28]		37.60/0.9544	33.44/0.9094	32.00/0.8904	31.43/0.9174	38.07/0.9734	4.71	-	-
	DCLS[27]		37.63/0.9554	33.46/0.9103	32.04/0.8907	31.69/0.9202	38.31/0.9740	13.63	-	-
Implicit Blind SR †††	DASR[25]	x2	NA	NA	NA	NA	NA	5.84	-	-
	IDENet (Ours)		37.16/0.9521	32.84/0.9025	31.65/0.8848	30.22/0.9004	36.86/0.9697	3.9	-	-
	Improvement		NA	NA	NA	NA	NA	**-33.22%**		
Non Blind + Blind SR †	Bicubic	x4	24.57/0.7108	22.79/0.6032	23.29/0.5786	20.35/0.5532	21.50/0.6933	-	-	-
	CARN[2]		26.57/0.7420	24.62/0.6226	24.79/0.5963	22.17/0.5865	21.85/0.6834	-	-	-
	Bicubic + ZSSR[35]		26.45/0.7279	24.78/0.6268	25.97/0.5989	22.11/0.5805	23.53/0.7240	-	-	-
	Deblurring[33] + CARN[2]		28.10/0.4843	16.59/0.3994	18.46/0.4481	15.47/0.3872	16.78/0.5371	-	-	-
	CARN[2] + Deblurring[33]		28.69/0.8092	26.40/0.6926	26.10/0.6528	23.46/0.6597	25.84/0.8035	-	-	-
	HPUN[38]		32.24/0.8950	28.66/0.7828	27.60/0.7371	26.12/0.7878	30.55/0.9089	-	-	-
	MogaSRN[9]		32.50/0.8987	28.81/0.7872	27.72/0.7417	26.53/0.8005	31.05/0.9154	-	-	-
	Omni-SR[43]		32.49/0.8988	28.78/0.7859	27.71/0.7415	26.64/0.8018	31.02/0.9151	-	-	-
Explicit Blind SR ††	IKC[10]	x4	31.67/0.8829	28.31/0.7643	27.37/0.7192	25.33/0.7504	28.91/0.8782	5.32	150.5	1.89
	DANv1[14]		31.89/0.8864	28.42/0.7687	27.51/0.7248	25.86/0.7721	30.50/0.9037	4.33	78.10	0.25
	DANv2[28]		32.00/0.8885	28.50/0.7715	27.56/0.7277	25.94/0.7748	30.45/0.9037	4.71	77.38	0.26
	AdaTarget[16]		31.58/0.8814	28.14/0.7626	27.43/0.7216	25.72/0.7683	29.97/0.8955	16.70	21.96	0.14
	DCLS[27]		32.12/0.8890	28.54/0.7728	27.60/0.7285	26.15/0.7809	30.86/0.9086	13.63	31.01	0.31
Implicit Blind SR †††	DASR[25]	x4	28.03/0.8061	25.65/0.6763	23.22/0.6628	25.51/0.6349	25.26/0.7955	5.84	1.13	0.27
	IDENet (Ours)		31.57/0.8846	28.27/0.7678	27.35/0.7235	25.39/0.7585	29.88/0.8988	3.9	14.27	0.08
	Improvement		**3.54/0.0785**	**2.62/0.0915**	**4.13/0.0607**	**-0.12/0.1236**	**4.62/0.1033**	**-33.22%**		

Table 1. Quantitative comparison on public SR benchmark datasets with Gaussian8 kernels. Improvements (in bold) are shown concerning the state-of-the-art methods that use the same implicit blind degradation estimation approach as our IDENet. For scale factor $\times 2$, there is no implicit blind SR method reporting on the performance of the benchmark datasets, hence the shown improvement values.

4.3. Comparison with State-of-The-Art Methods

We report on the results achieved by our direct competitors (*i.e.*, implicit blind SR methods, denoted with †††) as well as on the performance achieved by solutions that either assume the availability of the ground-truth kernel at training time (*i.e.*, explicit blind SR methods, denoted with ††) or at test time (*i.e.*, non-blind SR methods, denoted with †).

Isotropic Gaussian kernels evaluation is conducted using the Gaussian8 kernels defined in [10, 27]. Table 1 shows that some non-blind SR methods significantly underperform with respect to the blind SR alternatives. Among the latter, the best performances are obtained by explicit blind SR methods, particularly by DCLS [27]. These methods, however, hinge on the availability of ground-truth blur kernel information that is unattainable in real-world scenarios and often have many parameters (*e.g.*, AdaTarget [16] has 17M and DCLS [27] has 13M), thus limiting their adoption in computationally constrained settings. The best implicit blind SR method, *i.e.*, DASR [25], performs poorly with similar results to non-blind SR methods. In contrast, our approach achieves comparable performance (*e.g.*, less than $0.6dB$ PSNR and 0.5% SSIM difference on average for scale factor $\times 4$) with explicit blind SR methods. Most notably, we outperform our direct competitor, *i.e.*, DASR [25], with an average gain of more than $3dB$. All this with a model that has the smallest number of parameters with respect to every existing work, *i.e.* 33% and 71% fewer parameters than the best explicit (DCLS) and implicit (DASR) methods, respectively. These results highlight the ability of our approach to predict the isotropic kernels that are the foundation for the subsequent learnable wiener filter and refinement modules.

We also conducted a qualitative comparison considering different datasets. Results depicted in Figure 4 show that our IDENet method produces clearer and visually more pleasing results than many the blind SR methods, including our direct implicit blind SR competitor *i.e.*, DASR [25].

Degradation Estimation Approach	Method	DIV2KRK x2 PSNR↑/SSIM↑	DIV2KRK x4 PSNR↑/SSIM↑
Non Blind + Blind SR †	Bicubic	28.73/0.8040	25.33/0.6795
	Bicubic+ZSSR[35]	29.10/0.8215	25.61/0.6911
	EDSR[26]	29.17/0.8216	25.64/0.6928
	RCAN[59]	29.20/0.8223	25.66/0.6936
	DBPN[12]	29.13/0.8190	25.58/0.6910
	DBPN[12]+Correction[15]	30.38/0.8717	26.79/0.7426
	KernelGAN[3]+SRMD[55]	29.57/0.8564	27.51/0.7265
	KernelGAN[3]+ZSSR[35]	30.36/0.8669	26.81/0.7316
Explicit Blind SR ††	IKC[10]	NA	27.70/0.7668
	DANv1[14]	32.56/0.8997	27.55/0.7582
	DANv2[28]	32.58/0.9048	28.74/0.7893
	AdaTarget[16]	NA	28.42/0.7854
	KOALAnet[20]	31.89/0.8852	27.77/0.7637
	DCLS[27]	32.75/0.9094	28.99/0.7946
Implicit Blind SR †††	DASR[25]	NA	26.21/0.7082
	IDENet (Ours)	32.57/0.9010	28.59/0.7850
	Improvement	**32.57/0.9010**	**2.38/0.0768**

Table 2. Quantitative comparison on DIV2KRK dataset. Improvements (in bold) are shown concerning the state-of-the-art methods that use the same implicit blind degradation estimation approach as our IDENet. For scale factor $\times 2$, there is no implicit blind SR method reporting on the performance of the benchmark datasets, hence the shown improvement values.

Anisotropic Gaussian kernels present a more generalized and challenging scenario. Table 2 provides quantitative results on the DIV2KRK dataset. Similarly to what is obtained using the isotropic Gaussian kernel setting, our proposed IDENet outperforms the direct implicit degradation blind SR competitor, *i.e.*, DASR [25], with a gain of $2dB$ and a 7% improvement in PSNR and SSIM, respectively. Moreover, we perform remarkably similarly to the best explicit method (*i.e.*, less than 1% SSIM difference) with 71% fewer parameters.

Figure 5 visually demonstrate that our IDENet exhibits superior sharpness and cleanliness compared to DASR and other explicit blind SR methods.

Anisotropic kernels are complicated and diverse kernels. Results under this setting show that (i) our novel loss term is effective for driving the degradation kernel prediction, and that (ii) the novel learnable Wiener filter module greatly handles the diverse nature of anisotropic kernels. This substantiates the importance of these two components in producing an already upscaled input for the final refinement.

4.4. Ablation Study

We performed the ablation study on the three modules of our architecture. We also explored the impact of the various losses and the effects of learning the Wiener filter in the feature space rather than in the image space[2]. All the experiments have been carried out following the Isotropic Gaussian kernels settings with scale factor $\times 4$.

IDENet Modules. To analyze the importance of the three IDENet components, we altered our architecture by turning on and off the $M_{\mathbf{K}}$, M_W, and M_R modules. Results in Table 3 shows that using the kernel estimator and the Wiener filter alone we achieve similar performance to our closest DASR competitor. Using the refinement module alone significantly enhances the performance but the joint exploitation of all modules yields the best results with an average PSNR/SSIM gain of about $0.22dB$ and 1% over all benchmark datasets respectively. $M_{\mathbf{K}}$ and M_W bring improvements (*e.g.*, Manga109 $+0.48/+0.0107$ PSNR/SSIM) at a negligible computational cost: $M_{\mathbf{K}}$ and M_W account only for 0.50 GFLOP and $0.82M$ parameters, thus demonstrating the ability of such two modules to correctly estimate the blur kernels for deconvolution. Such an analysis might indicate that the learnable Wiener filter module is not sufficient to capture all the different degradation effects with a shallow network, which thus calls for a deeper module to refine the deconvolved image.

[2]For additional experiments please refer to the supplementary.

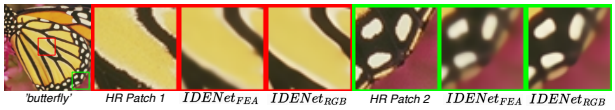

Figure 6. Visual comparison of Wiener Filtering in feature and RGB space on *Image 'butterfly'* (×4) from Set5 dataset. The kernel width is 3.2.

Impact of Losses. Table 4 presents the quantitative results for various combinations of the losses we used in (8). Performance demonstrates that each of the loss terms adds up to the final result, with the best performance achieved when all are jointly considered. To verify that the terms we introduced are complementary and relevant to SR, we analyzed the impact of adding the common perceptual loss ($\mathcal{L}_{per}$) function [17] to our optimization objective. This caused a drop of approximately $1dB$ in PSNR and 0.02% in average SSIM, respectively. Such a result strongly supports the choices we made, demonstrating the complementarity of the selected components for SR.

Wiener Filter. To assess the specific value of the Wiener deconvolution module, we analyze its performance in two different spaces: the standard image space and a deep feature space. For the deep feature space evaluation, we considered [8], where the Wiener deconvolution is applied to deep features rather than image pixels. In such a case, the refinement module utilizes the deconvolved deep features instead of the upsampled RGB image. Table 5 shows that performing Wiener deconvolution in the image space results in higher PSNR and SSIM values, *i.e.*, an average respective gain of $0.8dB$ and 2%, compared to deconvolution in the deep feature space. This might indicate that the filter better handles noise on colors, textures, and structures specific to the RGB representation, rather than in a feature space where artifacts are likely to be controlled by the learnable model parameters (see Figure 6 for some samples).

Kernel Estimation. To get more insights about the ability of the novel implicit kernel estimation loss, we computed the results in Figure 7. It shows the degradation kernels predicted by state-of-the-art methods while also report-

Method	$M_{\mathbf{K}}$	M_W	M_R	Set5 PSNR↑/SSIM↑	Set14 PSNR↑/SSIM↑	BSDS100 PSNR↑/SSIM↑	Urban100 PSNR↑/SSIM↑	Manga109 PSNR↑/SSIM↑
IDENet	✓	✓	✗	27.86/0.7918	25.72/0.6823	25.64/0.6463	22.74/0.6315	24.17/0.7575
IDENet	✗	✗	✓	31.41/0.8796	28.10/0.7601	27.27/0.7155	25.14/0.7442	29.40/0.8881
IDENet (Ours)	✓	✓	✓	31.57/0.8846	28.27/0.7678	27.35/0.7235	25.39/0.7585	29.88/0.8988

Table 3. Analysis of the impact of each proposed IDENet module. The best result for each dataset is in blue, and the second best is in red.

Method	$\mathcal{L}_{SR}$	$\mathcal{L}_k$	$\mathcal{L}_{TV}$	$\mathcal{L}_{per}$	Set5 PSNR↑/SSIM↑	Set14 PSNR↑/SSIM↑	BSDS100 PSNR↑/SSIM↑	Urban100 PSNR↑/SSIM↑	Manga109 PSNR↑/SSIM↑
IDENet	✓	✗	✗	✗	31.41/0.8796	28.10/0.7601	27.27/0.7155	25.14/0.7442	29.40/0.8881
IDENet	✓	✓	✗	✗	31.36/0.8829	28.14/0.7670	27.32/0.7234	25.22/0.7542	29.65/0.8954
IDENet (Ours)	✓	✓	✓	✗	31.57/0.8846	28.27/0.7678	27.35/0.7235	25.39/0.7585	29.88/0.8988
IDENet	✓	✓	✗	✓	30.17/0.8558	27.52/0.7402	26.61/0.6909	24.59/0.7311	28.47/0.8728
IDENet	✓	✓	✓	✓	30.43/0.8579	27.53/0.7395	26.70/0.6912	24.63/0.7278	28.40/0.8701

Table 4. Impact of the different optimization loss terms. The best result for each dataset is in blue, and the second best is in red.

Method	Set5 PSNR↑/SSIM↑	Set14 PSNR↑/SSIM↑	BSDS100 PSNR↑/SSIM↑	Urban100 PSNR↑/SSIM↑	Manga109 PSNR↑/SSIM↑
IDENet$_{FEA}$	31.18/0.8762	27.72/0.7541	27.09/0.7121	24.37/0.7136	27.65/0.8547
IDENet$_{RGB}$ (Ours)	31.57/0.8846	28.27/0.7678	27.35/0.7235	25.39/0.7585	29.88/0.8988

Table 5. Performance comparison obtained by performing Wiener deconvolution in the standard RGB space and in the deep feature space.

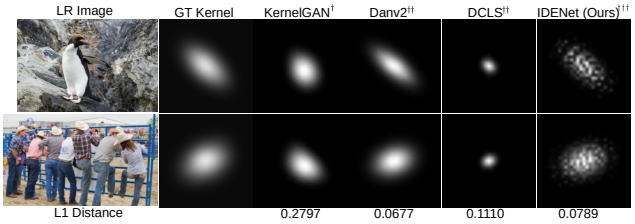

Figure 7. Visual results of estimated kernels of *img 001* and *img 004* from DIV2KRK datasets by different estimation methods.

ing on the L1 distance between the predicted kernels and the ground-truth kernels. The noisy values are due to how we implicitly estimate the degradation kernels through our novel loss –without the need for the ground truth degradation kernels. The implicit kernel loss term in (10) minimizes the difference between the GT ($\mathbf{I}_{LR}$) image and the downsampled GT ($\mathbf{I}_{HR}$) image (filtered with the learnable kernel $\mathbf{K}$). This does not guarantee the kernel is smooth: the bilinear downsampling operator (denoted as $\downarrow_S$) smooths neighboring values, thus canceling out the "imperfections" resulting from the noisy filter. Despite the noisy values shown in our estimated kernels (last column), the computed L1 distances demonstrate that our approach is notably better than KernelGAN and DCLS while performing very similarly to DANv2 (0.01 L1 difference). Our proposed implicit loss term is thus proven to accurately deduce the kernel, eliminating the dependency on ground-truth data (unavailable in real-world scenarios), which is a major leap forward in blind SR.

Performance on Real World Images. To showcase the effectiveness of our method in real-world scenarios where ground truth HR images and blur kernels are unavailable, we ran the experiment suggested in [27] using the historical images from [22]. The sample result in Figure 8 shows that

(a) LR image (b) Bicubic (c) DCLS (d) DASR (e) Ours)

Figure 8. Comparison of *image 006* (×4) from historic dataset.

our IDENet generates sharp edges and fine details, similarly to what the best explicit blind-SR method preserves, *i.e.*, DCLS [27]. Our implicit blind-SR competitor, *i.e.*, DASR [25], produces a similar-looking result but with less details. These visual results are aligned with the quantitative (*e.g.*, Table 1) and qualitative (*e.g.*, Figure 4) performance shown so far, thus, once again, demonstrating the ability of our model in producing excellent SR images with an efficient approach that overcomes the requirement for ground-truth blur kernel data.

5. Conclusion

In this work, we proposed an implicit degradation estimation network for blind image SR. We introduced a novel implicit kernel loss term that allowed us to design a network module estimating the blur kernel without the supervision of ground-truth data, which is unattainable in real-world settings. We also proposed a novel learnable Wiener module that leverages such a predicted kernel to perform deconvolution in the Fourier domain, via a closed-form solution. To further refine the deconvolved image, we added an efficient refinement module that exploits the attention mechanism to capture the long-range feature dependencies, crucial for SR image generation. Extensive experiments on different benchmark datasets show that our proposed approach outperforms the implicit blind SR state-of-the-art method and achieve comparable performance to explicit blind SR approaches, with a substantially lower number of learnable parameter.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, pages 126–135, 2017. 5

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 252–268, 2018. 6

[3] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *NeurIPS*, 32, 2019. 1, 5, 7

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 5

[5] Victor Cornillere, Abdelaziz Djelouah, Wang Yifan, Olga Sorkine-Hornung, and Christopher Schroers. Blind image super-resolution with spatially variant degradations. *ACM Transactions on Graphics (TOG)*, 38(6):1–13, 2019. 2

[6] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 1

[7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 1, 2

[8] Jiangxin Dong, Stefan Roth, and Bernt Schiele. Deep wiener deconvolution: Wiener meets deep learning for image deblurring. *Advances in Neural Information Processing Systems*, 33:1048–1059, 2020. 3, 7

[9] Garas Gendy, Nabil Sabor, and Guanghui He. Lightweight image super-resolution based multi-order gated aggregation network. *Neural Networks*, 166:286–295, 2023. 6

[10] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, pages 1604–1613, 2019. 1, 2, 3, 5, 6, 7

[11] Shuhang Gu, Nong Sang, and Fan Ma. Fast image super resolution via local regression. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 3128–3131. IEEE, 2012. 1

[12] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 7

[13] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 5

[14] Yan Huang, Shang Li, Liang Wang, Tieniu Tan, et al. Unfolding the alternating optimization for blind super resolution. *Advances in Neural Information Processing Systems*, 33:5632–5643, 2020. 2, 3, 5, 6, 7

[15] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1428–1437, 2020. 7

[16] Younghyun Jo, Seoung Wug Oh, Peter Vajda, and Seon Joo Kim. Tackling the ill-posedness of super-resolution through adaptive target generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16236–16245, 2021. 2, 6, 7

[17] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016. 1, 2, 7

[18] Asif Hussain Khan, Christian Micheloni, and Niki Martinel. Lightweight implicit blur kernel estimation network for blind image super-resolution. *Information*, 14(5):296, 2023. 2

[19] Asif Hussain Khan, Rao Muhammad Umer, Matteo Dunnhofer, Christian Micheloni, and Niki Martinel. Lbkenet: Lightweight blur kernel estimation network for blind image super-resolution. In *International Conference on Image Analysis and Processing*, pages 209–222. Springer, 2023. 2

[20] Soo Ye Kim, Hyeonjun Sim, and Munchurl Kim. Koalanet: Blind super-resolution using kernel-oriented adaptive local adjustment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10611–10620, 2021. 2, 7

[21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[22] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017. 8

[23] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1

[24] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1, 4, 5

[25] Jie Liang, Hui Zeng, and Lei Zhang. Efficient and degradation-adaptive network for real-world image super-resolution. In *European Conference on Computer Vision*, pages 574–591. Springer, 2022. 2, 6, 7, 8

[26] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE confer-*

ence on computer vision and pattern recognition workshops, pages 136–144, 2017. 1, 2, 7

[27] Ziwei Luo, Haibin Huang, Lei Yu, Youwei Li, Haoqiang Fan, and Shuaicheng Liu. Deep constrained least squares for blind image super-resolution. In *CVPR*, 2022. 2, 3, 5, 6, 7, 8

[28] Zhengxiong Luo, Yan Huang, Shang Li, Liang Wang, and Tieniu Tan. End-to-end alternating optimization for real-world blind super resolution. *International Journal of Computer Vision (IJCV)*, 2023. 6, 7

[29] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7769–7778, 2020. 1, 2

[30] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms. *Department of Electrical Engineering and Computer Sciences University of California, Berkeley*. 5

[31] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2017. 5

[32] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16*, pages 191–207. Springer, 2020. 1

[33] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Deblurring images via dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 40(10):2315–2328, 2017. 6

[34] Valeriya Pronina, Filippos Kokkinos, Dmitry V Dylov, and Stamatios Lefkimmiatis. Microscopy image restoration with deep wiener-kolmogorov filters. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, pages 185–201. Springer, 2020. 4

[35] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3118–3126, 2018. 1, 2, 6, 7

[36] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3516–3525, 2020. 1, 2

[37] Hyeongseok Son and Seungyong Lee. Fast non-blind deconvolution via regularized residual networks with long/short skip-connections. In *2017 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2017. 3

[38] Bin Sun, Yulun Zhang, Songyao Jiang, and Yun Fu. Hybrid pixel-unshuffled network for lightweight image super-resolution. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(2):2375–2383, 2023. 6

[39] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017. 5

[40] Rao Muhammad Umer, Gian Luca Foresti, and Christian Micheloni. Deep super-resolution network for single image super-resolution with realistic degradations. In *Proceedings of the 13th International Conference on Distributed Smart Cameras*, pages 1–7, 2019. 3, 5

[41] Rao Muhammad Umer, Gian Luca Foresti, and Christian Micheloni. Deep generative adversarial residual convolutional networks for real-world super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 438–439, 2020. 5

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 1

[43] Hang Wang, Xuanhong Chen, Bingbing Ni, Yutian Liu, and Jinfan Liu. Omni aggregation networks for lightweight image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22378–22387, 2023. 6

[44] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10581–10590, 2021. 2

[45] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 1

[46] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Realesrgan: Training real-world blind super-resolution with pure synthetic data supplementary material. *Computer Vision Foundation open access*, 1:2, 2022. 2

[47] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5

[48] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. The MIT press, 1949. 4

[49] Bin Xia, Yulun Zhang, Yitong Wang, Yapeng Tian, Wenming Yang, Radu Timofte, and Luc Van Gool. Basic binary convolution unit for binarized image restoration network. *arXiv preprint arXiv:2210.00405*, 2022. 1, 2

[50] Tete Xiao, Mannat Singh, Eric Mintun, Trevor Darrell, Piotr Dollár, and Ross Girshick. Early convolutions help transformers see better. *Advances in neural information processing systems*, 34:30392–30400, 2021. 4

[51] Yu-Syuan Xu, Shou-Yao Roy Tseng, Yu Tseng, Hsien-Kai Kuo, and Yi-Min Tsai. Unified dynamic convolutional network for super-resolution with variational degradations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12496–12505, 2020. 1, 2

[52] Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang. Single-image super-resolution: A benchmark. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 372–386. Springer, 2014. 1

[53] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 1

[54] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 5

[55] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3262–3271, 2018. 1, 2, 7

[56] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1671–1681, 2019. 2

[57] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3217–3226, 2020. 2

[58] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 1, 2

[59] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 7

[60] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 1

[61] Hongyi Zheng, Hongwei Yong, and Lei Zhang. Unfolded deep kernel estimation for blind image super-resolution. In *European Conference on Computer Vision*, pages 502–518. Springer, 2022. 2