ISSR-DIL: Image Specific Super-Resolution Using Deep Identity Learning

Sree Rama Vamsidhar S Jayadeep D Rama Krishna Gorthi
Indian Institute of Technology Tirupati, India
sreeramvds@gmail.com, ee18b015@iittp.ac.in, rkg@iittp.ac.in

Abstract

The advent of Deep Learning (DL) techniques has significantly improved the performance of Image Super-Resolution (ISR) algorithms. However, the primary limitation to extending the existing DL-based works for real-world instances is their computational and time complexities. Besides this, the presumed degradation process in their training datasets is another. In this paper, we present a lightweight and highly efficient zero-shot ISR model. The proposed algorithm first estimates the degradation kernel K from the given low-resolution (LR) image statistics. Later, we introduce "Deep Identity Learning (DIL)", a novel learning strategy, to compute the inverse of K by exploiting the identity relation *between the degradation and inverse degradation models. Contrary to the mainstream ISR works, the proposed model considers K alone as its input to learn the ISR task. We term the proposed approach as "Image Specific Super-Resolution Using Deep Identity Learning (ISSR-DIL)". In our experiments, ISSR-DIL demonstrated a competitive performance compared to state-of-the-art (SotA) works on benchmark ISR datasets while requiring, at least by order of 10, fewer computational resources.*

1. Introduction

Image Super Resolution (ISR) is a well-established low-level vision task whose objective is to generate a High-resolution (HR) image from the given corresponding LR observation(s) with many real-world applications in various prominent domains like medical imaging, satellite imaging, and surveillance, which demand the HR version of the scene of interest for its analysis and understanding.

In 2014, the first DL-based ISR work i.e. SRCNN [11] demonstrated remarkable improvement over the traditional methods like [32], [14], [16], [3], [33]. Further, various deep learning techniques, [28], [20], [34], [23], [8]. were proposed to generate super-resolved images with finer details and of better quality. These works consider the unknown degradation process as a learnable entity in either supervised or unsupervised approach from a large synthetic HR-LR image pair or LR image datasets respectively. The ISR works that presume the degradation model in their dataset to be MATLAB bicubic downscaling or some blurring kernel with additive or multiplicative noise followed by downscaling are known as non-blind approaches [11], [23], [8] [20]. Further, a line of works like [38], [35], [15], [36] etc, considered multiple types of degradations in their datasets and illustrated the improved performance over a wide range of degradations.

However, in practice, the degradation process from the HR image to the LR image is unknown and also may vary for every application based on factors like, but not limited to, the imaging system and the imaging environment. The appropriate prior knowledge of the degradation model is required to achieve the desired ISR task performance. Any naive assumption of the degradation process results in a domain gap between the synthetic and the realistic data as the real LR (RLR) images suffer unknown complex degradations. Hence the models trained and tested with the synthetic datasets perform poorly on RLR images. The progress of the existing DL-based ISR works can be largely attributed to the huge amount of computational resources, which made the deep models with millions of parameters to train on large ISR datasets available. Nonetheless, the existent DL-based ISR works as such cannot adapt to unseen LR images and yet demand substantial computational resources for the acquisition of robust datasets and to re-train their deep models.

Towards addressing these issues, LR image-specific ISR works like Zero-shot Super-resolution (ZSSR) [31], KernelGAN [4], Dual Back-Projection-Based Internal Learning for Blind Super-Resolution (DBPI) [19] etc, were proposed. These works are dataset-independent and rely on internal statistics of an image i.e., patch recurrence property [16]. The primary shortcoming of the existing zero-shot approaches is the inference time taken for a single image is very high. Because of this limitation, these methods are not recommendable for practical applications. Nevertheless, the assumption made by [16] may easily fail for images with diverse contents or monotonous scenes, since it is

hard to exploit recurring information across scales. Hence, these approaches are valid for a very limited set of images with frequently recurring contents across or within scales of the given LR image.

In this paper, we propose a lightweight and computationally efficient image-specific ISR model referred to as "Image Specific" Super-Resolution Using "Deep Identity Learning" (ISSR-DIL). Here, we introduce a novel learning objective i.e., Deep Identity Learning (DIL), which exploits the identity relation between the degradation kernel K and its inverse K^{-1} i.e., ISR kernel. In this work, initially, we estimate the K from the given input LR image using the variance reduction relation between the LR image and the K. Later, we estimate the K^{-1} characterized by a custom shallow Convolutional Neural Network (CNN), with no activation functions referred to as Linear-CNN (L-CNN). L-CNN was trained with DIL objective on an estimated K "alone" as input, to form the inverse degradation model i.e., the ISR model. At the inference stage, the test LR image is pre-upsampled to the required resolution, and then super-resolution enhancement is performed using the trained L-CNN for varying scale factor (sf)s like $\times 2, \times 3, \times 4$. The proposed ISSR-DIL super-resolves the given LR input image to the desired scale factor by at least five times less than in time and at least less by $O(10)$ in computational resources required by the current SotA zero-shot approaches. Our proposed image-specific ISR model can be categorized under zero-shot ISR approaches since the super-resolution task is carried out using a given single LR input alone similar to the existing zero-shot ISR methods [31], [19], [13]. The highlights of our proposition are as follows.

1. The proposed ISR framework transforms the task of learning the ISR problem from image datasets (LR images or HR-LR image pair datasets), to the task of "identity" learning from the convolution operation between K and K^{-1} with the proposed novel learning objective i.e., DIL with a dedicated regularization term.
2. The proposed lightweight image-specific K estimation algorithm is quite simple and robust, as compared to the existing highly complex and computationally demanding degradation kernel estimation techniques.
3. The proposed DIL-based ISR model performs on par with the existing SotA ZSSR methods, while being at least ten times more computationally efficient.

2. Related work

In this section, the relevant and existing DL-based zero-shot ISR methods were discussed and presented in brief.

A stream of works explores the internal statistics based on the recurrence property of a natural image to model the degradation from a given single test input, the LR image itself. The recurrence property of a natural image states that patches of a single image tend to recur within and across different scales of this image. Glasner et al. [16] proposed to capitalize the internal statistics within an image to tackle the Single ISR problem. Non-parametric blind super-resolution [25] utilized this framework for the BISR task. Based on this framework, Zero-Shot Super-Resolution (ZSSR) [31] proposed to train an image-specific CNN with HR-LR pairs generated from a single LR input, for super-resolving the same input LR image. KernelGAN [4] was proposed for blind kernel estimation using the patch recurrence property. For a given arbitrary LR image, the kernel recovery performance is limited and unstable. Flow-based kernel prior (FKP) [22], was proposed for robust kernel estimation, based on Normalizing flow (NF) [9], [10] to learn a kernel prior in latent space. Nonetheless, these kernel estimation works are to be associated with the existing ISR models to generate the super-resolved image. Later, Kim et al. [19] proposed a unified internal learning-based SR framework consisting of an SR network and a downscaling network. In the self-supervised training phase of DBPI, the SR network is optimized to reconstruct the LR input image from its downscaled version produced by the downscaling network. Meanwhile, the downscaling network is trained to recover the LR input image from its super-resolved version generated by the SR network. Similarly, Emad et al. [13] proposed DualSR: Zero-Shot Dual Learning for Real-World Super-Resolution, which jointly optimizes an image-specific down sampler and corresponding upsampler. More specifically, DualSR [13] is trained with the cycle-consistency loss, the masked interpolation loss, and the adversarial loss using the patches from the test image.

The organization of the rest of the paper is as follows. Section 3 presents a detailed explanation of our proposed method. In section 4, extensive experimental analysis and results are presented, followed by Conclusions in section 5.

3. Proposed method

In this section, we elaborate on the proposed ISSR-DIL model in greater detail.

3.1. Problem formulation

The entire ISR problem can be broadly observed as the degradation model followed by the inverse degradation model (as shown in Fig. 1). At first, the degradation model generates a degraded and downsampled version of the HR image i.e., the LR image. Later, the inverse degradation model restores the HR image from the obtained LR image i.e., the SR image. The degradation and inverse degradation models are explained below in detail.

Degradation model: The degradation model typically consists of two stages. The first is the convolution of the HR image with the degradation kernel, followed by the downsampling operation in the second stage. The mathemati-

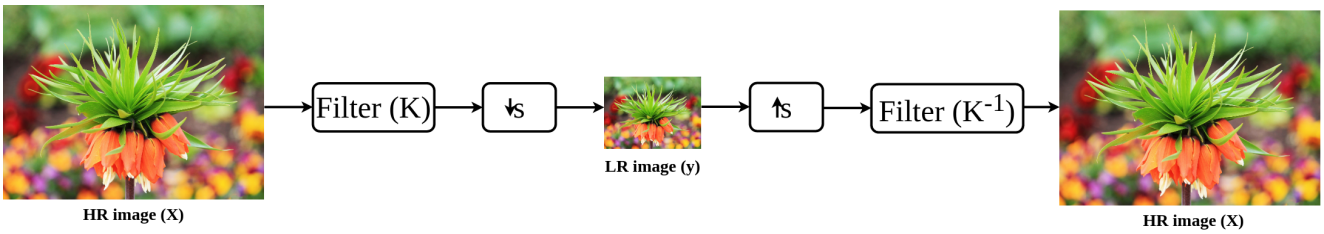


Figure 1. An end-to-end ISR model.

cal representation of the correspondence between HR image $X \in R^{M \times N}$ and LR image $y \in R^{m \times n}$ is given in Eq. (1) [4].

$$y = (X * K) \downarrow_s \tag{1}$$

Here K represents the degradation kernel, $*$ represents the convolution operation, s is the scale factor and the dimension $M \times N$ is equal to $sm \times sn$.

Inverse degradation model:. The objective of the ISR task is to achieve the inverse of the degradation model i.e., to reconstruct the HR image from the given LR image input. The mathematical representation of the equivalent inverse model of the degradation model (refer Eq. (1)) is given in Eq. (2).

$$(y \uparrow^s) * K^{-1} = X \tag{2}$$

Where K^{-1} represents the inverse degradation kernel, to be estimated.

End-to-end ISR model. The unified degradation and inverse degradation models form an end-to-end ISR problem. This unified framework represents an "identity model" with the same entity as the input and the output i.e., HR image (in the ideal case). The end-to-end ISR framework is presented in Fig. 1. The degradation and inverse degradation models are characterized and effectively represented by their respective kernels K and K^{-1}. In practice, the output of the inverse model is the super-resolved (SR) image from the LR image. Further, the simplified end-to-end ISR framework with only convolution operations involved can be obtained by the exclusion of mutually inverse operations i.e., downsampling and upsampling operations, under the assumption that the degraded HR images before the downsampling stage are bandlimited. In such a scenario, the downsampling and the immediate upsampling stages will not affect the ISR framework.

Based on the derived simplified end-to-end ISR model, we propose a simple, computationally efficient, and novel framework that learns the zero-shot ISR task from degradation kernel alone i.e., ISSR-DIL. The simplified end-to-end ISR framework is equivalent to an identity model with the HR image as its input and output. Therefore, the convolution operation between K and K^{-1} should result in an identity relation. Henceforth, we propose to learn the inverse degradation model forming an identity relation with the degradation kernel. The identity relation is given in Eq. (3).

$$K * K^{-1} = I \tag{3}$$

Where I is an Identity matrix of dimension $P \times Q$, and P, Q values are based on the dimensions of K & K^{-1}.

Here, the proposed formulation simplifies the task of learning ISR problem from image datasets using deep architectures with several millions of learnable parameters, to the task of "identity" learning from the convolution operation between degradation kernel K and the inverse degradation kernel K^{-1} via an L-CNN made up of purely convolution operations (discussed in Sec.3.3). Hence, in this work, instead of considering the image datasets with HR and LR images to represent their degradation relationship, the degradation kernel estimated from the given LR image is considered for modeling the ISR task.

3.2. Degradation kernel (K) estimation

In ISR literature, the degradation modeling is mainly carried out conventionally by using Gaussian prior [12],[25] and, in recent works, employing complex deep models like KMSR [40], KernelGAN [4], etc. To effectively capture the degradations in real LR images, K is modeled as unimodal [25] & Gaussian [29]. The existing well-known blind ISR works like [4], and [19] consider the K to be isotropic or anisotropic Gaussian models. Also, the work [21] demonstrated that with an appropriate estimation algorithm, blind deconvolution can be performed even with a weak Gaussian prior. Following these widely adopted premises, in our ISR work, we propose a simple and computationally inexpensive method to model and estimate the degradation as the anisotropic Gaussian kernel based on statistics of the given LR image.

Statement: When an image with standard deviation (SD) σ_i is convolved with a Gaussian filter having SD σ_f, then SD of the resulting image σ_o is approximately given in Eq. 6[1].

$$\sigma_0 \approx \frac{\sigma_i}{\sigma_f \, 2\sqrt{\pi}} \tag{4}$$

Here σ_f, σ_o, and σ_i refer to the SD of the Gaussian kernel, LR image, and HR image respectively. However, σ_i is un-

[1]Please refer to the supplementary material for the proof of Eq. (6)

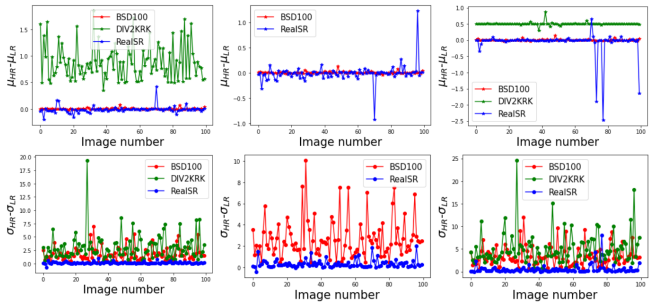


Figure 2. The visual plots depicting the mean preservation and the variations in standard deviations among HR-LR image pairs were given in row 1 and row 2 respectively for scale factors 2, 3, and 4 (left to right).

known for the zero-shot ISR framework and only σ_o can be known from the given LR image to compute the σ_f.

In our experiments, through empirical analysis, it was inferred that the mean of the HR images was approximately preserved while the Standard Deviation (SD) was always slightly higher than the LR images. In our study, we considered datasets with three different types of degradations i.e, MATLAB Bicubic downscaling (set5 [5], set14 [37], BSD100 [24] datasets), Random Gaussian Kernel degradation followed by downsampling (DIV2KRK dataset [4]), Real HR-LR image pair dataset captured by varying focal lengths of the device (RealSR dataset [17]). The visual plots with differences in mean and SD of HR-LR image pairs of the RealSR, DIV2KRK, and BSD100 datasets were shown in Fig. 2[2].

Therefore, we propose to model the SD of the HR images as the zero mean Gaussian random perturbations around the LR image SD values, i.e., we consider the $\sigma_i = \sigma_o + \delta_\sigma$. Thus, for the given LR image, the SD of its degradation kernel, σ_f is calculated as shown in Eq. (5).

$$\sigma_f \approx \frac{\sigma_o + \delta_\sigma}{\sigma_o \, 2\sqrt{\pi}} \tag{5}$$

In this work, we generate the anisotropic Gaussian degradation kernels K with dimensions 11×11, having SD σ_f given by Eq. (5) and rotation angle $\theta \in U[0, \pi]$.

3.3. Linear Convolutional Neural Network (L-CNN)

ISR is an ill-posed inverse problem for which a unique inverse will not exist. Computing a single-layer network to learn the inverse of the degradation kernel cannot serve the deblurring task's objective. This is because a matrix/single layer accepts only one set of parameters/global minima with convex loss. Also, the K can usually be a low-rank matrix. Further, it was empirically found that single-layer architecture does not converge to the correct solution [7]. Whereas the multi-layered linear networks have many good and equally valued local minima. This allows many valid optimal solutions to the optimization objective in the form of different factorizations of the same matrix [18], [30], [2]. Following these research results, we propose a multi-layer Linear CNN (L-CNN) with no activations to learn the inverse degradation kernel, with the degradation kernels (K) as its input. The proposed L-CNN is a computationally efficient architecture having a depth of five layers and a width of 32 with 3×3 filters across the depth. Here the L-CNN is chosen to maintain the dimensions of the input throughout the network i.e. the same output dimension at every layer. Therefore, at the inference stage, L-CNN operates on blur images to generate sharp images with fine details.

The general limitations of the networks that maintain the dimensions of the input at the output, like computational time, complexity, and memory are because of operations in high dimensional space. It is important to note that these limitations are not valid in this work as the input to the CNN is kernel K, a matrix of very small dimension (i.e., 11×11), compared to very high dimensional or 3-D tensor input images, in practice. The learning and inference methodology of the NSD-DIL model is depicted in Fig. 3.

3.3.1 Deep Restoration Kernel (DRK)

L-CNN with only convolution layers is proposed in the interest of extracting the inverse degradation kernel K^{-1}, from the trained network, by convolving all the filters of L-CNN sequentially with stride 1 with impulse as its input. The extracted K^{-1} from the L-CNN is referred to as "Deep Restoration Kernel (DRK)", as the proposed ISR model is a pre-upsampling based Network. The effective representation of the equivalent kernel K^{-1} from the L-CNN facilitates the application of the regularization constraints on it easily. We note that this is the first DL-based work to learn and extract explicit K^{-1} from the linear systems concepts of the ISR problem. We also note that this DRK alone could also be employed for the ISR task as per Eq. (2). The ISR performance of DRK was compared to traditional deconvolution methods like the Wiener filter and Moore-Penrose Pseudo inverse. The results were tabulated in Table 4. The visual depiction of the estimated degradation kernel K extracted DRK and the corresponding SR images for a few sample LR images was given in Fig. 4.

3.4. Loss function - Deep Identity learning (DIL)

The L-CNN is trained with the DIL objective given in Eq. (6).

$$Loss(L) = \; \mid K * K^{-1} - \; I \mid_2^2 + R \tag{6}$$

[2]Please refer to the supplementary material for the tabulated average mean, and standard deviation of the HR and LR images present in the benchmark image super-resolution datasets [5], [37], [24], [4], [17].

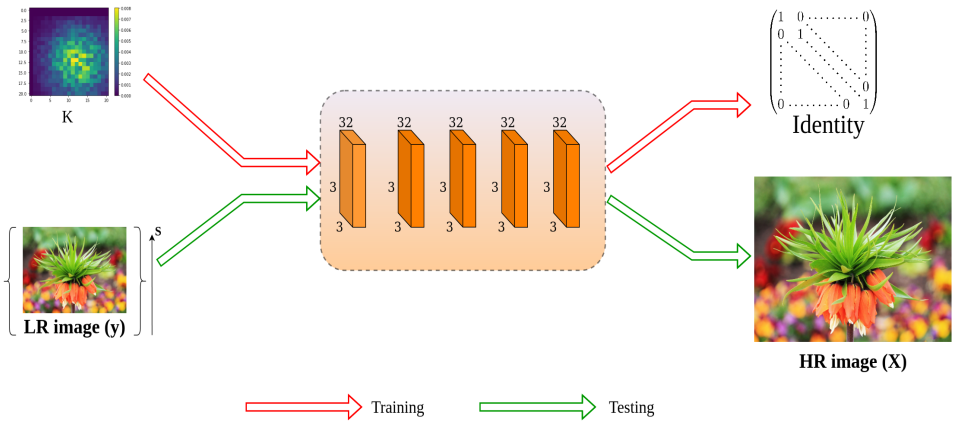


Figure 3. The training and inference methodology of the proposed ISSR-DIL model.

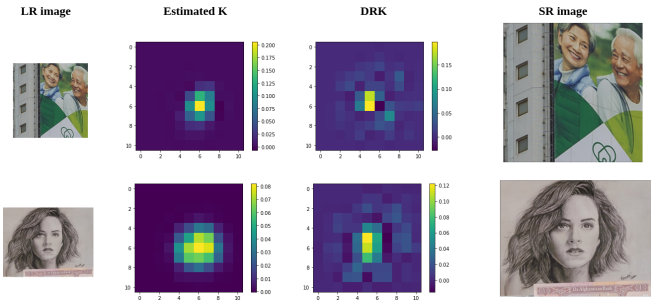


Figure 4. The sample visual depictions of the corresponding estimated K, DRK, and SR images using RealSR dataset.

Here, $\mid K * K^{-1} - I \mid_2^2$ term ensures the identity relation between the degradation and its inverse model and R is the proposed regularization, defined in Eq. (7).

$$R = \lambda_1 \times L_{pinv} + \lambda_2 \times L_{unity} + \lambda_3 \times L_{rs} \qquad (7)$$

where,
(i) $L_{pinv} = \mid K^{-1} - (MP\_PInv(K)) \mid$; encourages the inverse kernel K^{-1} to be close to Moore-Penrose Pseudo Inverse (MP_PInv) [27] of K. The MP_PInv is a substitute for matrix inverse in cases where it does not exist and guides the optimization by constraining the solution space in finding K^{-1}. Here, obtained K can be a singular matrix hence MP_PInv is used.
(ii) $L_{unity} = \mid 1 - \sum_{i,j} K^{-1}_{i,j} \mid$; encourages the sum of kernel K^{-1} elements to be equal to 1 and constraints the inverse degradation kernel space.
(iii) $L_{rs} = \mid K^{-1} - \left(\frac{K^{-1} - min(K^{-1})}{max(K^{-1}) - min(K^{-1})}\right) \mid$; encourages to have a good dynamic range in the model's parameters. Here $\lambda_{1,2,3}$ are hyper-parameters.

Table 1. Computational complexity comparison with SotA ISR methods.

Method	No.of parameters	Inference time (min)
EDSR	2.98 M	N.A.
ZSSR	0.29M	>= 10
KernelGAN + ZSSR	0.151M + 0.29M	>= 13
DBPI	0.5M	1
DualSR	0.45M	3.5
ISSR-DIL	$0.0028M$	0.16
DRK	$0.00001M$	< 0.016

4. Experiments

In this section, the implementation details, ISR results, comparisons, and the effect of the regularization term of the proposed ISSR-DIL method were discussed.

4.1. Training setup

We trained the proposed L-CNN (refer Sec.3.3 for architecture details of L-CNN) with the training objective given in Eq. (6) and the input K was estimated from the given LR image as discussed in Sec.3.2. The number of epochs was 50 and the learning rate was 0.1 with a step scheduler. The Adam optimization was used, with $\beta = 0.9$. The values of hyper-parameters used in the learning objective i.e, Eq. (7), set empirically, are as follows, $\lambda_1 = 0.1$, $\lambda_2 = 0.7$, $\lambda_3 = 0.1$. For K estimation, the range of δ_σ considered for RealSR and DIV2KRK datasets was $U[0.5, 1.5]$ and $U[1, 3]$ respectively[3]. The computational complexity of the ISSR-DIL model was compared concerning SotA in Table 1.

4.2. Results

The super-resolution ability of our approach was evaluated on the BISR benchmark datasets RealSR [6] and DIV2KRK [4] for sf 2, sf 4. In the RealSR dataset [6], the LR-HR image pairs are captured by adjusting the focal length of

[3]The code will be released soon publicly for reproducibility

the camera device. While the DIV2KRK [4] dataset is derived from the validation set of DIV2K [1] by convolving with random anisotropic Gaussian kernels perturbed with uniform multiplicative noise. In the above mentioned two benchmark datasets, [4] and [6], the noise distribution and the degradation kernel knowledge are completely unknown and do not truly follow the well-known distribution like Gaussian. For a fair comparison, we considered works like KernelGAN [4], ZSSR [31], DBPI [19], and DualSR [13] which consider single input LR image only for the ISR task as in our experiments. The results on the two datasets RealSR and DIV2KRK are outlined in Table 2. The qualitative and quantitative results demonstrate the efficacy of the proposed lightweight ISSR-DIL model across various sfs on both synthetic and real datasets, with > 10 fold less memory and at least five times fewer time requirements compared to SotA works.

In this work, the primary objective is not to identify an SR image that agrees well with HR image in terms of the quantitative image intensity based evaluation metric like PSNR. Rather, our objective is to generate SR images (i) that are close to natural images, (ii) that have finer details than the LR image, and (iii) qualitatively and quantitatively closer to the real HR images. However, due to the ill-posed nature of the ISR problem, one LR image can have multiple corresponding HR images. The proposed approach super-resolves the LR input image into one of the HR versions of it, based on the input estimated degradation kernel considered in our training. However, in practice, for zero-shot problem setting, the ground truth HR image is not available. The primary interest is to generate the HR version of the given LR image close to the natural image statistics. Hence we considered NIQE [26], a no-reference image quality assessment metric with a high correlation to human perception, to demonstrate the performance of our proposed method. Besides, to assess the ability of the proposed ISSR-DIL to restore the finer details during ISR process, we employ SSIM, and for qualitative and quantitative similarity with the HR image we provide the visual results and PSNR values respectively. From Table 2, in the case of our proposed method, it is worth noting that the superior NIQE score demonstrates the ability of our ISSR-DIL model to generate natural images whereas the good SSIM metric value, reflecting in better edge reconstruction in generated SR images. Since the proposed model did not train on either LR or LR-HR image pair data, a relative decline in reference-based metric (PSNR), at pixel level was observed as expected. The important note is that the proposed model did not produce artifacts in the generated super-resolved output images, unlike the compeer, SotA works like [4], [19]. The sample visual results were provided in Fig. 5.

In another set of experiments, two different cases were considered. In case-I, the estimated degradation kernel K was combined with the non-blind ISR works like ZSSR [31], USRNet [39] which consider K also as input along with input LR image. In case-II, the K was estimated by the work KernelGAN [4] and KMSR [40] and fed as input to the proposed ISSR-DIL model. The results were tabulated in Table 3. These results indicate that the proposed approach is quite generic as i) it has an edge over existing approaches in terms of the kernel estimation (case-II results) and ii) the kernels estimated by the existing SotA methods can also be employed through proposed DIL for effective SR task, despite being a very lightweight model (case-I results).

Additionally, the comparison of ISR performance on **real captured LR images** provided in Fig. 6, showcases that our method super-resolves with essential fine details, and without introducing artifacts, unlike its compeers. Here, the noise distribution and the degradation kernel information are unknown. *Thus our ISR method demonstrated its generalizability without any knowledge of the corresponding blur kernel, has no strict limitations on the blur kernel estimation, and is also not restrictive to the assumptions considered in our degradation kernel estimation*.

Furthermore, we observed the *skewness of the uniform regions*, to measure the non-Gaussianity/asymmetry, in the LR images from i) the synthetic dataset [4], ii) a real dataset [6], and iii) from real captured LR image using an old smartphone camera (refer to Figure 6). Our experimental analysis do confirm that i) the (background of the) LR images are degraded not purely by a Gaussian distribution, and ii) the proposed approach is robust to perform the ISR task for more complex and real degradations than Gaussian. The skewness measure along with the histogram depicting the pixel distribution in the randomly selected uniform regions of LR images are provided in the supplementary material.

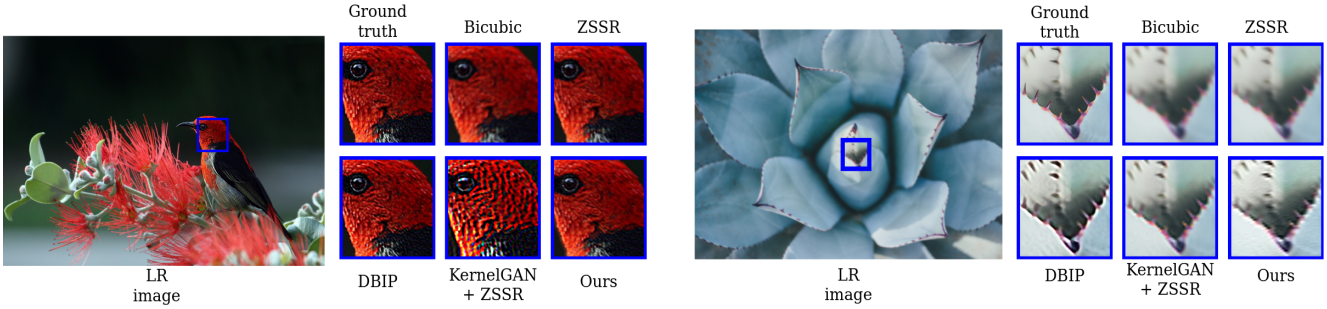
4.3. DRK for ISSR

The extracted DRK from the L-CNN was evaluated for various sfs like $\times 2, \times 3$ using the DIV2KRK dataset. In this experiment, the given LR image is bicubically upsampled to the desired sf and then convolved with the DRK to obtain the super-resolved image output. The results were tabulated in Table 4. From the experimental results, it was observed that the ISR performance is not altered when the deep model L-CNN is replaced by a matrix DRK and the L-CNN model produced images that were more natural than DRK.

We also compared with the standard traditional deconvolution method i.e., the Wiener filter, which also depends only on the image statistics for degradation kernel estimation. It was observed that the performance of the Wiener filter is not robust across the dataset and demonstrated 50% poor values in all three metrics. Further, the Wiener filter requires an appropriately estimated K and estimated noise additionally to perform better deconvolution operation. In the case of the proposed model, a weak prior of the degra-

Table 2. Comparison of ISR results on RealSR and DIV2KRK datasets in terms of NIQE↓/SSIM↑/PSNR↑ were given below. Here, the red indicates the best score and the blue indicates the second-best score.

Test Dataset	Scale	Bicubic	EDSR	ZSSR	KernelGAN + ZSSR	DBPI	DualSR	Ours
RealSR	2	5.71/0.8736/30.27	5.43/0.8706/30.29	5.34/0.8786/ 30.56	6.22/ 0.8907/30.24	5.87/0.8226/27.83	4.97/0.8570/28.01	5.17/0.8890/27.33
	4	6.11/0.7413/25.74	6.25/0.7449/ 25.98	5.35/0.7434/25.83	6.10/0.7243/24.09	5.73/0.6508/22.21	-/-/-	5.32/ 0.8058/24.64
DIV2KRK	2	5.26/0.7846/27.24	5.11/0.8034/29.17	4.70/0.7925/27.51	5.95/0.8379/28.24	5.19/ 0.8684/ 30.77	4.75/0.8538/29.38	4.45/0.8188/23.65
	4	6.21/0.6478/23.89	6.43/0.6615/25.63	5.69/0.6550/24.05	5.92/0.6799/24.76	5.20/0.7368/ 26.86	-/-/-	4.97/ 0.7611/22.40



(a) **Scale factor 2** (b) **Scale factor 4**

Figure 5. Qualitative results of different ISR methods.

Table 3. Robustness analysis of the proposed ISSR-DIL for sf 2 and sf 4 on DIV2KRK dataset for case I (ISSR-DIL on SotA generated Ks) and case II (SotA ISRs on estimated image specific K) as detailed in Sec. 4.2.

	Method	Scale factor	NIQE	SSIM	PSNR
Case I	KernelGAN + ZSSR	X2	5.95	0.8379	28.24
		X4	5.92	0.6799	24.76
	KernelGAN + USRNet	X2	4.54	0.8281	24.18
		X4	5.39	0.7060	20.16
	KMSR + USRNet	X2	4.83	0.8777	27.17
		X4	6.33	0.7731	23.26
	KernelGAN + ISSR-DIL	X2	4.78	0.7610	22.06
		X4	5.36	0.7416	21.52
	KMSR + ISSR-DIL	X2	4.55	0.7906	22.34
		X4	5.11	0.7481	21.54
Case II	Our K + ZSSR	X2	4.62	0.8843	27.43
		X4	5.74	0.7872	24.02
	Our K + USRNet	X2	4.95	0.8020	23.78
		X4	6.34	0.7510	22.26

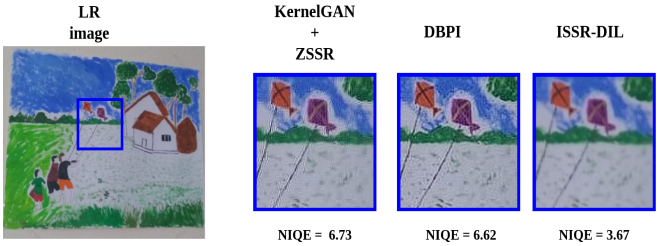


Figure 6. Visual results for the cropped regions of a real captured image using Samsung J7 Prime (13MP, f/1.9) for the SR sf 2.

Table 4. Comparison of the ISR performance of the proposed NSSR-DIL, DRK, and the traditional deconvolution i.e. Wiener filter, Moore-Penrose Pseudo inverse (MP-Pinv) methods on DIV2KRK dataset using SSIM↑/PSNR↑ metrics.

Test dataset	Scale factor	MP-Pinv	Wiener filter	NSSR-DIL (Ours)	DRK (Ours)
DIV2KRK	**x2**	0.000036/-195.58	0.3442/14.50	0.8644/26.02	0.8736/26.54
	x4	0.000076/-195.27	0.3609/15.35	0.8058/24.64	0.797/24.49

dation model is sufficient to have robust performance across a wide range of degradations in the ISR task. In this experiment, we considered the degradation kernels estimated by the KernelGAN [4] as input to the Wiener filter.

Besides the Wiener filter, we also included MP-Pinv in our comparisons. Unlike DRK i.e. a convolution inverse, MP-Pinv is a replacement for the matrix inverse for the cases where the inverse does not exist. Here, we computed the MP-Pinv of the degradation kernels estimated by the KernelGAN [4] and performed the ISR task employing the MP-Pinv. From the results, it is realized that the MP-Pinv was not able to do the ISR task, and the proposed DRK-based ISR outperforms the established matrix inverse technique by many folds.

4.4. Ablation study

In this section, we discussed the effect of the proposed regularization term (R).

In the proposed learning objective (refer Eq. (6)), we introduced three constraints through the regularization term (R) (refer Eq. (7)) to obtain the appropriate inverse degra-

dation model. The influence of each entity in the proposed regularization term (R) is quantified and is presented in Table 5 and qualitative results are presented in Fig.7. It was observed that the presence of L_{unity} and L_{rs} in Eq. (7) facilitated the proposed ISR model to generate the SR images with fine edge details. Hence, from the given LR image, to generate the SR image containing the sharp edges and good dynamic range, the regularization term $R = L_{unity} + L_{rs}$ is sufficient to include in the loss function given in Eq. (6). Further, the terms L_{pinv} aided the ISSR-DIL model to produce smoother and more realistic SR images with greater quality metrics.

Table 5. The effect of the proposed regularization term R on the ISSR-DIL model's performance with the RealSR dataset for sf 2.

Loss (L)	NIQE↓	SSIM↑	PSNR↑
L_I	6.50	0.2845	9.72
$L_I + \lambda_1 \times L_{pinv}$	6.08	0.4684	12.37
$L_I + \lambda_2 \times L_{unity}$	5.54	0.8652	26.14
$L_I + \lambda_2 \times L_{unity} + \lambda_3 \times L_{rs}$	5.94	0.7911	20.82
$L_I + \lambda_1 \times L_{pinv} + \lambda_2 \times L_{unity}$	5.54	0.8657	26.22
$L_I + \lambda_1 \times L_{pinv} + \lambda_2 \times L_{unity} + \lambda_3 \times L_{rs}$	5.17	0.8890	27.33

Figure 7. Two SR image results and their corresponding DRKs given by ISSR-DIL for $L = L_I + R$ & $L = L_I + L_{unity} + L_{rs}$ using RealSR dataset images of sf 2.

5. Conclusion

We have established a new problem formulation in terms of DIL for the ISR task and also proposed a computationally efficient approach to carry out ISR for higher sfs. The proposed ISSR-DIL is both memory and computationally efficient with reliable performance, hence eligible for real-time ISR tasks. The proposed method's performance is quite comparable to the SotA works, despite using only the image-specific degradation kernel maps. Thus, this work demonstrates tremendous hope for improving the ISR capability without the need for HR-LR image pair data. Our ISSR-DIL method paves the path to have a deeper look into the learning and understanding of the inverse degradation kernel since the explicit representation of it in terms of DRK is made possible.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 6

[2] Sanjeev Arora, Nadav Cohen, and Elad Hazan. On the optimization of deep networks: Implicit acceleration by overparameterization. In *International Conference on Machine Learning*, pages 244–253. PMLR, 2018. 4

[3] Selen Ayas and Murat Ekinci. Single image super resolution using dictionary learning and sparse coding with multi-scale and multi-directional gabor feature representation. *Information Sciences*, 512:1264–1278, 2020. 1

[4] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems*, 32, 2019. 1, 2, 3, 4, 5, 6, 7

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

[6] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 5, 6

[7] Anna Choromanska, Mikael Henaff, Michael Mathieu, Gérard Ben Arous, and Yann LeCun. The loss surfaces of multilayer networks. In *Artificial intelligence and statistics*, pages 192–204. PMLR, 2015. 4

[8] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11057–11066, 2019. 1

[9] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014. 2

[10] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016. 2

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1

[12] Netalee Efrat, Daniel Glasner, Alexander Apartsin, Boaz Nadler, and Anat Levin. Accurate blur models vs. image priors in single image super-resolution. In *2013 IEEE International Conference on Computer Vision*, pages 2832–2839, 2013. 3

[13] Mohammad Emad, Maurice Peemen, and Henk Corporaal. Dualsr: Zero-shot dual learning for real-world super-resolution. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1629–1638, 2021. 2, 6

[14] W.T. Freeman, T.R. Jones, and E.C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002. 1

[15] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3599–3608. IEEE, 2019. 1

[16] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. pages 349 – 356, 2009. 1, 2

[17] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1914–1923, 2020. 4

[18] Kenji Kawaguchi. Deep learning without poor local minima. *Advances in neural information processing systems*, 29, 2016. 4

[19] Jonghee Kim, Chanho Jung, and Changick Kim. Dual back-projection-based internal learning for blind super-resolution. *IEEE Signal Processing Letters*, 27:1190–1194, 2020. 1, 2, 3, 6

[20] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP, 2017. 1

[21] Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971, 2009. 3

[22] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte. Flow-based kernel prior with application to blind super-resolution. *arXiv preprint arXiv:2103.15977*, 2021. 2

[23] 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 conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 1

[24] 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. 8th Int'l Conf. Computer Vision*, pages 416–423, 2001. 4

[25] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *2013 IEEE International Conference on Computer Vision*, pages 945–952, 2013. 2, 3

[26] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 6

[27] Roger Penrose. A generalized inverse for matrices. In *Mathematical proceedings of the Cambridge philosophical society*, pages 406–413. Cambridge University Press, 1955. 5

[28] Kalpesh Prajapati, Vishal Chudasama, and Kishor Upla. A light weight convolutional neural network for single image super-resolution. *Procedia Computer Science*, 171:139–147, 2020. Third International Conference on Computing and Network Communications (CoCoNet'19). 1

[29] Gernot Riegler, Samuel Schulter, Matthias Rüther, and Horst Bischof. Conditioned regression models for non-blind single image super-resolution. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 522–530, 2015. 3

[30] A Saxe, J McClelland, and S Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. International Conference on Learning Represenatations 2014, 2014. 4

[31] 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

[32] Jian Sun, Zongben Xu, and Heung-Yeung Shum. Image super-resolution using gradient profile prior. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 1

[33] Radu Timofte, Vincent De, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *2013 IEEE International Conference on Computer Vision*, pages 1920–1927, 2013. 1

[34] 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

[35] 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

[36] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 701–710, 2018. 1

[37] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, 2010. 4

[38] 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

[39] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3217–3226, 2020. 6

[40] Ruofan Zhou and Sabine Süsstrunk. Kernel modeling super-resolution on real low-resolution images. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2433–2443, 2019. 3, 6