NSSR-DIL: Null-Shot Image Super-Resolution Using Deep Identity Learning

Sree Rama Vamsidhar S
sreeramvds@gmail.com

Rama Krishna Gorthi
rkg@iittp.ac.in

Indian Institute of Technology (IIT)
Tirupati
India

Abstract

The present State-of-the-Art (SotA) Image Super-Resolution (ISR) methods employ Deep Learning (DL) techniques using a large amount of image data. The primary limitation to extending the existing SotA ISR works for real-world instances is their computational and time complexities. In this paper, contrary to the existing methods, we present a novel and computationally efficient ISR algorithm that is independent of the image dataset to learn the ISR task. The proposed algorithm reformulates the ISR task from generating the Super-Resolved (SR) images to computing the inverse of the kernels that span the degradation space. We introduce "Deep Identity Learning", exploiting the identity relation between the degradation and inverse degradation models. The proposed approach neither relies on the ISR dataset nor on a single input low-resolution (LR) image (like the self-supervised method i.e. ZSSR) to model the ISR task. Hence we term our model as "Null-Shot Super-Resolution Using Deep Identity Learning (NSSR-DIL)". The proposed NSSR-DIL model requires fewer computational resources, at least by an order of 10, and demonstrates a competitive performance on benchmark ISR datasets. Another salient aspect of our proposition is that the NSSR-DIL framework detours retraining the model and remains the same for varying scale factors like $\times 2, \times 3, \times 4$. This makes our highly efficient ISR model more suitable for real-world applications.

1 Introduction

Super Resolution 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). Real-world applications in prominent domains like medical imaging, satellite imaging, and surveillance 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 [10] had demonstrated remarkable improvement over the existing dominant example-based ISR methods [13, 15, 16]. Further, various deep learning techniques [7, 23, 24, 27, 29, 34, 38, 45] were proposed to generate the 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

existing works [7, 10, 23, 27, 45] are non-blind as they presume the degradation model in their dataset to be MATLAB bicubic downscaling or some blurring kernel with additive or multiplicative noise followed by downscaling operation. Further, a line of works like [14, 37, 40, 42, 43] considered multiple types of degradations in their datasets and illustrated the improved performance over a wide range of degradations in LR images.

Another line of works address the fundamental problem in ISR algorithms i.e., the domain gap between the synthetic and the Real LR (RLR) images by acquiring the realistic datasets. The real HR-LR image pairs datasets were captured with various techniques like the focal length adjusting-based approach [3, 4, 35, 44], the hardware binning-based approach [22], and the beam splitter-based approach [18]. However, misalignment between LR and HR image pairs despite image registration introduces blurring artifacts in the reconstructed HR images. Moreover, capturing and preparing a realistic dataset is a hindering task requiring lots of human effort, time, and, computational resources. Further, it is not scalable for the generality of all types of real-world degradations.

The other stream of works are zero-shot approaches [2, 12, 20, 32], that attempt LR image-specific super-resolution. These works are training dataset-independent and rely on internal statistics of an image i.e., patch recurrence property [15]. The primary shortcoming of the existing zero-shot approaches is that the inference time taken for a single image is very high and is not suitable for practical applications.

In this paper, we redefine the task of learning the ISR model from image data to computing an inverse of the degradation model. Here, the degradation model is constructed using a wide set of anisotropic Gaussian kernels (Ks). We propose a novel learning strategy called "Deep Identity Learning (DIL)" that exploits the identity relation between the degradation model and its inverse (K^{-1}) i.e., the ISR model. In this work, a custom lightweight and computationally efficient CNN with no activation functions, referred to as Linear-CNN (L-CNN), is trained on the generated degradation model with DIL as its objective. Unlike existing self-supervised zero-shot ISR frameworks [32], [20], and [12] the proposed NSSR-DIL model is image independent and doesn't require LR-HR pair or LR image data at any stage of the ISR task learning. The highlights of our proposition in this paper are as follows.

1. To the best of our knowledge, NSSR-DIL is the first image data-independent ISR model and is contrary to the mainstream SoTA works that do "distribution mapping from LR to HR image data" using DL frameworks.

2. A novel learning objective exploiting the identity relation between the degradation model and its inverse i.e., DIL, is proposed.

3. The proposed ISR model is at least ten times more computationally efficient and delivers competitive ISR performance.

2   Related work

In this section, the relevant DL-based zero-shot ISR methods were discussed 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. [15] proposed to capitalize the internal statistics within an image to tackle the Single ISR problem. Non-parametric blind
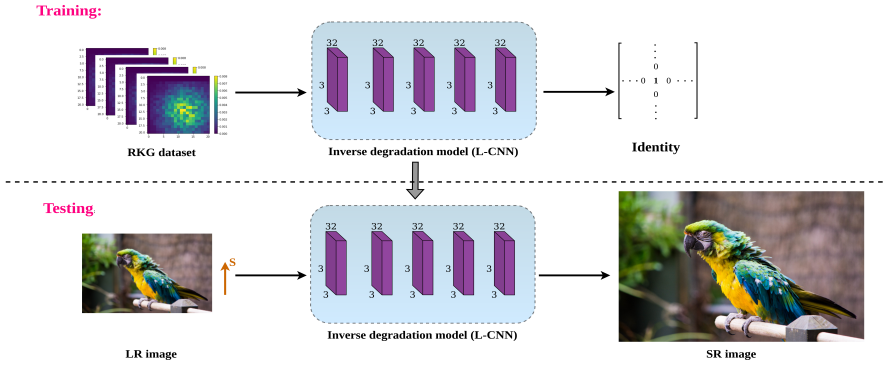
Figure 1: The training and inference methodology of the proposed NSSR-DIL model.

super-resolution [28] utilized this framework for the blind ISR task. Based on this framework, Zero-Shot Super-Resolution (ZSSR) [32] 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 [2] 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. In recent times, Reference-based Zero-Shot Super-Resolution with Depth Guided Self-Exemplars (RZSR) [39] was proposed based on ZSSR [32] and reference-based ISR techniques and also utilizes the predicted depth information of the given LR image in its ISR framework. Flow-based kernel prior (FKP) [26], was proposed for robust kernel estimation, based on Normalizing flow (NF) [8], [9] 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. [20] 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. [12] proposed DualSR: Zero-Shot Dual Learning for Real-World Super-Resolution, which jointly optimizes an image-specific down sampler and corresponding upsampler with the cycle-consistency loss, the masked interpolation loss, and the adversarial loss using the patches from the test image. The most recent zero-shot ISR method, Zero-Shot Dual-Lens Super-Resolution [36] learns an image-specific SR model from the dual-lens pair given at test time.

In this work, towards addressing the limitations of data dependency, time, and computational complexities for the practical use case of real-time implementation, we proposed a lightweight yet robust purely image data-independent model to restore the distorted LR image.

3 Proposed method

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

Problem introduction: The end-to-end ISR framework can be broadly observed as the degradation model followed by the inverse degradation model. 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 discussed 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 down-sampling operation in the second stage. The mathematical 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) [2].

$$y = (X * K) \downarrow_s \qquad (1)$$

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

Inverse degradation model: The objective of the inverse of the degradation model is to reconstruct the HR image from the given LR image input. The mathematical representation of the equivalent inverse model of the degradation model is given in Eq. (2).

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

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

Unified ISR model. The unified degradation and inverse degradation models framework represents an "identity model" with the same entity as the input and also as the output i.e. HR image (in the ideal case). Here 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. In the unified ISR framework, it is assumed that the degraded HR images before the downsampling stage are adequately band-limited such that the downsampling and the immediate upsampling stages can be ignored. Then, a simplified ISR framework with a degradation model directly followed by the inverse degradation model is obtained with an HR image as both 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).

$$K * K^{-1} = \delta \qquad (3)$$

Where δ is a two-dimensional discrete impulse function.

Based on the derived simplified form of the unified ISR model, we propose a simple, computationally efficient, and novel framework that is independent of images to learn the ISR task. 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 between the degradation kernel K and the inverse degradation kernel K^{-1} through a custom linear CNN (discussed in Sec.3.2).

3.1 Random Kernel Gallery (RKG) Dataset

To realize the identity learning task, a set of degradation maps sampled from the degradation kernel space is generated. In ISR literature, the degradation modeling is mainly carried out conventionally by using Gaussian prior [11, 28] and, in recent works, using complex deep models like KMSR [46], KernelGAN [2] with their own underlying assumptions in the degradation model (K) estimation. Especially for real LR images, theoretically, K is

modeled as unimodal [28] & Gaussian [30]. The existing well-known blind ISR works like [2, 20] consider the K to be isotropic or anisotropic Gaussian models. Also, the work [25] 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 constructed the degradation model with the anisotropic Gaussian kernel.

In this work, a set of anisotropic Gaussian kernels with dimensions 21×21, having $\sigma_1, \sigma_2 \in U[0.175, 6]$ and rotation angle $\theta \in U[0, \pi]$ were randomly generated as representative samples of the degradation kernel space. The number of K samples required to effectively learn the inverse degradation kernel K^{-1} is determined empirically as 3,200. The reference output to learn the inverse degradation model is the two-dimensional impulse function mentioned in Eq.3 and its dimensions are based on the size of K & K^{-1}. We term this dataset as the "Random Kernel Gallery (RKG)" dataset.

The proposed NSSR-DIL method is not limited to the "RKG-dataset" with anisotropic Gaussian kernels. NSSR-DIL can be easily re-trained on any other kernel distribution if proven to be a more accurate representation of the degradations kernel space in RLR images.

3.2 Linear Convolutional Neural Network (L-CNN)

ISR is an ill-posed inverse problem for which a unique inverse will not exist. Computing the direct (or pseudo) inverse of the provided input degradation kernel K or a single-layer network to learn the inverse of the degradation kernel cannot serve the ISR task's objective. This is because a matrix/single layer accepts only one set of parameters/global minima with convex loss [2]. 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 [6]. 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 [19], [51], [0]. 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) from the RKG-dataset 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 be a pre-upsample network [10], which maintains the same output dimension at every layer. Therefore, at the inference stage, L-CNN operates on traditional upsampled LR images of any desired scale factor to generate SR images with fine details. This implies that the inverse degradation kernel K^{-1} remains the same for varying scale factors like $\times 2, \times 3, \times 4$.

The general limitations of the pre-upsample networks like complexity, computational time, and memory, because of operations in high dimensional space, are not valid in this work as the input to the CNN is kernel K, a matrix of very small dimension (i.e., 21×21), compared to very high dimensional or 3-D tensor input images, in practice. The learning and inference methodology of the NSSR-DIL model is depicted in Fig. 1.

3.3 Deep Identity Learning (DIL):

The proposed DIL objective is to train the L-CNN on the RKG dataset as given in Eq. (4).

$$Loss(L) = \|K * K^{-1} - \delta\|_2^2 + R \tag{4}$$

Here, R is the proposed regularization term, defined in Eq. (5), and is provided to constrain the solution space to have ISR capability, in addition to learn a valid inverse.

$$R = \lambda_1 \times L_{ConvArea} + \lambda_2 \times L_{center} \tag{5}$$

where,

1. $L_{ConvArea} = |1 - \sum_{i,j} K_{i,j}^{-1}|$; This term holds one of the important properties of the convolution operation i.e. the area property since the proposed ISR approach learns the inverse degradation model only from the convolution operation. The area property states that the product of the area under the two input signals is equal to the area under the convolution output signal. Here, the two input signals are the degradation kernel and its inverse i.e. K and K^{-1}, while the output signal is a 2-D impulse function. The area property of the convolution operation is discussed in detail below.

   The Fourier transform of a 2-D discrete Time-domain signal can be written as

   $$\mathcal{F}(K(m,n) = \hat{K}(p,q) = \sum_m \sum_n K(m,n) \exp^{-j\frac{2\pi pm}{N_1} + \frac{2\pi qn}{N_2}} \tag{6}$$

   Here, $\mathcal{F}(.)$ represent the Fourier transform operation, (m,n) and (p,q) represents the index locations of the degradation and inverse degradation kernel matrices in the Time-domain and Frequency-domain signals respectively, and N_1, N_2 represent the dimensions of degradation, inverse degradation kernel.

   Consider the identity relation given in Eq. 3, and apply the Fourier transform on both sides i.e., $\mathcal{F}(K * K^{-1}) = \mathcal{F}(\delta)$;

   The above expression can be written as $\hat{K}(p,q).\hat{K^{-1}}(p,q) = 1$, since the Convolution operation in the Time domain is equivalent to multiplication in the Frequency domain (and vice-versa). We know that the Fourier transform of the 2-D impulse function is 1 i.e. $\mathcal{F}(\delta) = 1$. Here '.' represents the element-wise multiplication operation.

   Keeping the values of $(p,q) = (0,0)$ in the Eq. 6, we arrive at another important property of the Convolution operation that states the area under the Time-domain signal is equal to the value at the origin in the Frequency-domain form of that signal. Therefore, $\hat{K}(0,0) = \sum_m \sum_n K(m,n)$. Here, the $\sum_m \sum_n K(m,n)$ is set to be 1 in our RKG dataset, since a degradation kernel. Therefore, to satisfy the relation $\hat{K}(p,q).\hat{K^{-1}}(p,q) = 1$ when $(p,q) = (0,0)$, the area under the K^{-1} is imposed to be 1 i.e. $\sum_m \sum_n K^{-1}(m,n) = 1$ and introduced as a regularizer in DIL objective.

2. $L_{center} = |1 - K.K^{-1^T}|$; This term encourages the output of the LCNN to have a unit value at the center. Considering the Eq.3, $[K * K^{-1}](m,n) = \delta(m,n) = 1, if (m,n) = (0,0)$. Here, '.' represents the dot product operation, and $\lambda_{1,2}$ are hyper-parameters.

4 Experiments

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

4.1 Training setup

We trained the proposed L-CNN (refer Sec.3.2 for architecture details of L-CNN) on the RKG dataset (refer Sec. 3.1) with the learning objective given in Eq. (4). The number of

epochs was 50 and the learning rate was 0.1 with a step scheduler. The Adam [21] optimization was used, with $\beta = 0.9$. The values of hyper-parameters used in the learning objective i.e., Eq. (5), set empirically, are as follows, $\lambda_1 = 0.8$, $\lambda_2 = 0.2$. We have used a computer with an NVIDIA-GTX 2080 Ti 11GB GPU in all our experiments.

4.2   Results

Super-Resolution. The ISR ability of our approach was evaluated on the benchmark datasets RealSR [3] for sf 2, sf 3, sf 4 and DIV2KRK [2] for sf 2, sf 4. The results are outlined in Table 1. For a fair comparison, we considered works like KernelGAN [2], ZSSR [32], DBPI [20], and DualSR [12] which consider single input LR image only for modeling the ISR task as in our experiments. Besides, we included the recent SotA supervised ISR methods Dual Aggregation Transformer (DAT) [5], and MIRNetV2 [41], in our comparison experiments for adequate validation. We note that the proposed NSSR-DIL is only the DL approach in the literature that learns the inverse degradation kernel from the degradation kernel itself, without the need for LR image input. While, the compared SotA methods need supervised learning from images for K estimation and/or ISR model learning.

Computational complexity. The proposed L-CNN is an efficient, lightweight ISR model. The comparison of computational complexity in terms of parameters and super-resolution time (inference time) in minutes with standard self-supervised and zero-shot methods is given in Table 1. The proposed NSSR-DIL requires 15 times fewer memory resources even when compared with the least among the SotA methods i.e. ZSSR. Besides, the inference time required for the NSSR-DIL is reduced to less than a minute and less by an order of 100s than compeer works. Here, the recent SotA methods are excluded from the complexity comparison experiments that employ sophisticated deep models trained on large ISR datasets.

In practice, for real test instances, the reference image is not available and the primary interest is to generate the HR version of the given LR image close to the natural image statistics. Hence, we considered Neural Image Assessment (NIMA) [33], 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 NSSR-DIL to restore the finer details during SR processes, we employ Structural Similarity Index Measure (SSIM), and for qualitative and quantitative similarity with the HR image we provide the visual results and Peak Signal to Noise Ratio (PSNR) values. Despite being very lightweight the proposed NSSR-DIL demonstrated its effectiveness significantly in two out of three standard metrics on real and synthetic ISR datasets. Since the proposed model did not train on either LR or LR-HR image pair data, a relative decline in reference-based metric at pixel level i.e., PSNR 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 [2], [20], [41]. The sample visual results were provided in Fig. 3. We would like to emphasize that the RealSR dataset contains real LR images with unknown noise levels and degradation kernel information [17]. Furthermore, the DIV2KRK dataset includes synthetically generated LR images degraded by the Gaussian kernels perturbed by uniform multiplicative noise [2]. Even though we didn't consider the additive noise explicitly in our degradation model represented by Eq. 1, the proposed model showcased its ability to generate super-resolved images with superior SSIM and NIMA scores for both datasets. This highlights the ability of the proposed NSSR-DIL model to be robust to unseen degradations and to generate natural images with fine details in ISR.

ISR of real captured LR image. Additionally, the comparison of ISR performance on RLR

Table 1: Comparison of the computational complexity of SotA ISR methods and their performance on RealSR and DIV2KRK datasets in terms of NIMA↑/SSIM↑/PSNR↑ were given below. Here, the red indicates the best score and the blue indicates the second-best score.

|		ZSSR	KernelGAN + ZSSR	DBPI	DualSR	NSSR-DIL (Ours)	MIRNetV2	DAT
No.of Parameters (Million)		0.29	0.151+0.29	0.5	0.45	0.028	N.A.	N.A.
Inference time (min)		>=10	>=13	>=1	3.5	0.005	N.A.	N.A.
RealSR	x2	3.996/0.8786/30.56	3.81/0.8907/30.24	3.909/0.8226/27.83	5.0020/0.8570/28.01	4.075/0.8836/25.29	4.9329/0.8625/29.94	4.9413/0.8709/30.30
	x3	4.008/0.6210/20.53	-/-/-	-/-/-	-/-/-	4.095/0.8395/24.85	4.9372/0.7765/26.97	4.9322/0.7884/27.29
	x4	4.058/0.7434/25.83	4.068/0.7243/24.09	4.020/0.6508/22.21	-/-/-	4.118/0.8079/24.15	4.9273/0.7283/25.61	4.8973/0.1211/11.69
DIV2KRK	x2	4.111/0.7925/27.51	4.071/0.8379/28.24	4.049/0.8684/30.77	5.005/0.8538/29.38	4.161/0.8644/26.02	5.020/0.7845/26.85	4.9676/0.8046/27.81
	x4	4.156/0.6550/24.05	4.089/0.6799/24.76	4.146/0.7368/26.86	-/-/-	4.170/0.7926/23.58	5.001/0.6991/24.92	4.9081/0.6631/24.23
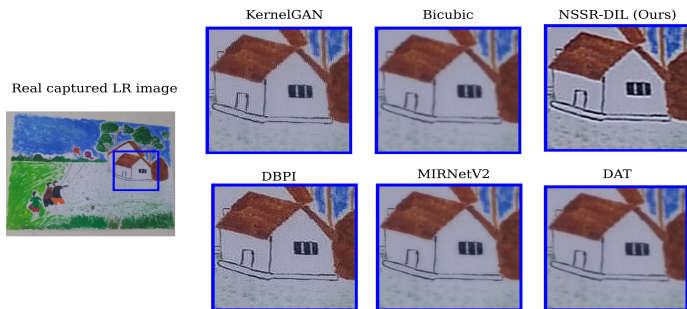
Figure 2: Visual results of the ISR of the cropped regions of a real captured image using Samsung J7 Prime (13MP, f/1.9) for the SR sf 2.

image provided in Fig. 2, 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.

4.3 Ablation study

In this section, the significance of the proposed regularization term (R) (given in Eq. 5), the performance of the NSSR-DIL method for various sizes of the RKG dataset, and the performance of the NSSR-DIL model on unseen synthetic test dataset were discussed.

4.3.1 Regularization term (R)

In the proposed learning objective (refer Eq. 4), we introduced two constraints through the regularization term (R) (refer Eq. 5) to obtain the inverse degradation model and with the reliable ISR performance. The influence of each entity in the proposed R is quantified and presented in Table 2. It was observed that the presence of $L_{ConvArea}$ in R has shown a greater
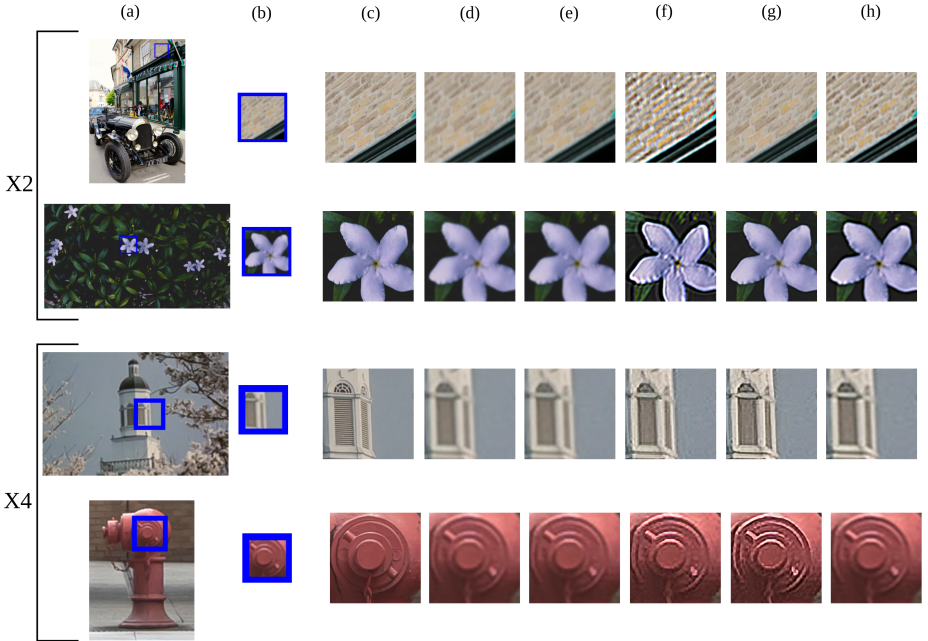
Figure 3: Qualitative results of different ISR methods for the scale factor 2 & 4 using DIV2KRK and RealSR datasets respectively. Here (a) - LR image, (b) - LR patch, (c) - HR image, (d) - Bicubic, (e) - ZSSR, (f) - KernelGAN + ZSSR, (g) - DBPI, (h) - NSSR-DIL.

impact on the performance of the proposed method relative to the L_{center}. The term L_{center} influenced the perceptual quality of the images. The term $L_{ConvArea}$ together with the L_{center} assisted the NSSR-DIL model to attain a reliable performance.

Note: The ISR performance results of the NSSR-DIL method for various sizes of the RKG dataset, and also on unseen synthetic test dataset are provided in the supplementary material.

Table 2: The effect of each of the proposed regularization terms in R on the NSSR-DIL model's performance with the DIV2KRK dataset for sf 2.

Loss (L)	$\|K * K^{-1} - \delta\|_2^2$	$\|K * K^{-1} - \delta\|_2^2 +$ $\lambda_1 \times L_{ConvArea}$	$\|K * K^{-1} - \delta\|_2^2 +$ $\lambda_2 \times L_{Center}$	$\|K * K^{-1} - \delta\|_2^2 +$ $\lambda_1 \times L_{ConvArea} + \lambda_2 \times L_{Center}$
NIMA↑/SSIM↑/PSNR↑	3.9138/0.0153/6.58	4.1550/0.8096/24.22	3.099/0.6360/12.69	4.161/0.8644/26.02

5   Conclusion

We have established a new problem formulation in terms of DIL for the ISR task. The proposed computationally efficient NSSR-DIL is the first image data-independent DL-based ISR model, for any scale factor. The proposed method's performance is quite comparable to the SotA works, despite being independent of image data in the ISR model design. Thus, this work demonstrates tremendous hope for improving the ISR capability without the need for image datasets (i.e., both supervised and unsupervised ISR datasets). Our NSSR-DIL

method paves the path to have a deeper look into the learning and understanding of the inverse degradation model from a linear systems perspective. The experimental results and computational performance comparisons with the SotA indicate that the proposed NSSR-DIL are remarkably suitable for real-time ISR tasks and embedded ISR applications.

References

[1] 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.

[2] 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.

[3] 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.

[4] Chang Chen, Zhiwei Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera lens super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1652–1660, 2019.

[5] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *ICCV*, 2023.

[6] 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.

[7] 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. doi: 10.1109/CVPR.2019.01132.

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

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

[10] 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.

[11] 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. doi: 10.1109/ICCV.2013.352.

[12] 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. doi: 10.1109/WACV48630.2021.00167.

[13] W.T. Freeman, T.R. Jones, and E.C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002. doi: 10.1109/38.988747.

[14] 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.

[15] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. pages 349 – 356, 11 2009. doi: 10.1109/ICCV.2009.5459271.

[16] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5197–5206, 2015. doi: 10.1109/CVPR.2015.7299156.

[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. doi: 10.1109/CVPRW50498.2020.00241.

[18] Hamid Reza Vaezi Joze, Ilya Zharkov, Karlton Powell, Carl Ringler, Luming Liang, Andy Roulston, Moshe Lutz, and Vivek Pradeep. Imagepairs: Realistic super resolution dataset via beam splitter camera rig. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 518–519, 2020.

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

[20] 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. doi: 10.1109/LSP.2020.3005043.

[21] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

[22] Thomas Köhler, Michel Bätz, Farzad Naderi, André Kaup, Andreas Maier, and Christian Riess. Toward bridging the simulated-to-real gap: Benchmarking super-resolution on real data. *IEEE transactions on pattern analysis and machine intelligence*, 42(11):2944–2959, 2019.

[23] 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, 10 2017. doi: 10.1109/TPAMI.2018.2865304.

[24] 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.

[25] 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. doi: 10.1109/CVPR.2009. 5206815.

[26] 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.

[27] 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.

[28] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *2013 IEEE International Conference on Computer Vision*, pages 945–952, 2013. doi: 10. 1109/ICCV.2013.121.

[29] 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. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2020.04. 015. URL https://www.sciencedirect.com/science/article/pii/ S1877050920309807. Third International Conference on Computing and Network Communications (CoCoNet'19).

[30] 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. doi: 10.1109/ICCV.2015.67.

[31] 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.

[32] 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.

[33] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE transactions on image processing*, 27(8):3998–4011, 2018.

[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.

[35] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *European Conference on Computer Vision*, pages 101–117. Springer, 2020.

[36] Ruikang Xu, Mingde Yao, and Zhiwei Xiong. Zero-shot dual-lens super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9130–9139, June 2023.

[37] 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.

[38] Wenming Yang, Wei Wang, Xuechen Zhang, Shuifa Sun, and Qingmin Liao. Lightweight feature fusion network for single image super-resolution. *IEEE Signal Processing Letters*, 26(4):538–542, 2019. doi: 10.1109/LSP.2018.2890770.

[39] Jun-Sang Yoo, Dong-Wook Kim, Yucheng Lu, and Seung-Won Jung. Rzsr: Reference-based zero-shot super-resolution with depth guided self-exemplars. *IEEE Transactions on Multimedia*, pages 1–13, 2022. doi: 10.1109/TMM.2022.3202018.

[40] 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.

[41] 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 *CVPR*, 2022.

[42] 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.

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

[44] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3762–3770, 2019.

[45] 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.

[46] 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. doi: 10.1109/ICCV.2019.00252.