Enhanced Super-Resolution Training via Mimicked Alignment for Real-World Scenes

Omar Elezabi, Zongwei Wu*, and Radu Timofte

Computer Vision Lab, CAIDAS & IFI, University of Würzburg
{omar.elezabi,zongwei.wu,radu.timofte}@uni-wuerzburg.de

Abstract. Image super-resolution methods have made significant strides with deep learning techniques and ample training data. However, they face challenges due to inherent misalignment between low-resolution (LR) and high-resolution (HR) pairs in real-world datasets. In this study, we propose a novel plug-and-play module designed to mitigate these misalignment issues by aligning LR inputs with HR images during training. Specifically, our approach involves mimicking a novel LR sample that aligns with HR while preserving the degradation characteristics of the original LR samples. This module seamlessly integrates with any SR model, enhancing robustness against misalignment. Importantly, it can be easily removed during inference, therefore without introducing any parameters on the conventional SR models. We comprehensively evaluate our method on synthetic and real-world datasets, demonstrating its effectiveness across a spectrum of SR models, including traditional CNNs and state-of-the-art Transformers. The source codes will be publicly made available at github.com/omarAlezaby/Mimicked_Ali

Keywords: Super Resolution, Alignment, Computational Photography

1 Introduction

Single image super-resolution (SISR) aims to recover fine details from a low-resolution (LR) image while upscaling it. This process is crucial for enhancing image quality, particularly with the surge of streaming services and augmented reality (AR) applications [27, 65]. With the development of deep learning methods [33, 51, 60, 74] and the availability of large-scale data [1, 43], recent works have shown great success in improving the image quality [31, 55].

The effectiveness of learning models is highly dependent on the quantity and quality of LR-HR (high-resolution) image pairs. Synthetic data [2, 32] has been a popular choice due to its ability to provide ample high-quality training pairs. However, synthetic data often oversimplify the SR problem by using uniform degradations like bicubic or Gaussian blurring [2,53]. Consequently, models trained on such data struggle to generalize to real-world scenarios with complex degradations [30]. As shown in Fig. 2, models trained on synthetic datasets produce less sharp images with noise when applied to real data. To address this,

* Corresponding Author

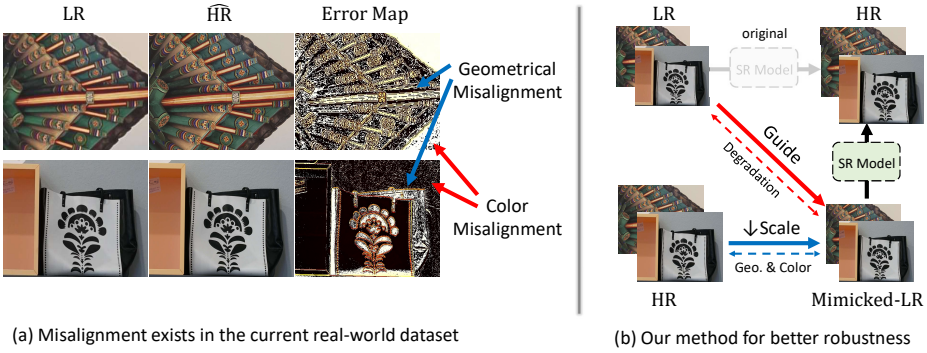

Fig. 1: Motivation: Despite significant effort, misalignment issues persist in real-world datasets [7, 67], limiting the potential of SR models. To address this challenge, we propose a novel alignment method using mimicked-LR, which maintains the same geometry and color properties as the HR input while sharing the same degradation type as the LR. By training our network with this newly generated mimicked-LR, we fully leverage the potential of SR models. $\widehat{HR}$ is the downscaled HR.

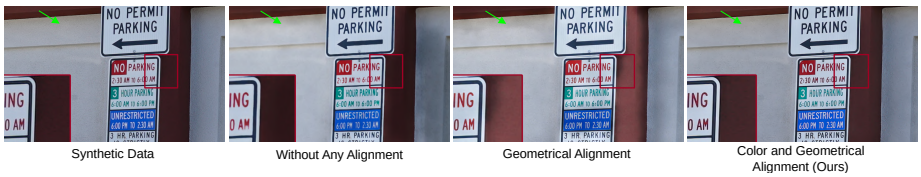

Fig. 2: The importance of Color and Geometrical Alignment between the LR and HR images for the training on Realistic SR Datasets

some previous works have proposed more complex degradation pipelines to create better-simulated datasets [3, 64]. However, these efforts often fail to capture specific system degradations, such as those introduced by different lenses.

A more realistic approach involves creating real-world super-resolution datasets using DSLR cameras with various zoom lenses [7, 67]. Despite producing images with real degradations, these datasets often suffer from misalignment issues between LR and HR images. Practical systems, when adjusting lens settings, introduce significant geometric misalignments, such as translations, scaling, and rotation. Additionally, changes in optical settings result in varying light readings for the same scene, affecting pixel colors, brightness, and illumination. These misalignments significantly limit the effectiveness of existing SR models.

In this work, we thoroughly analyze the existing alignment efforts with the goal of developing a more generic, holistic, and easy-to-implement method to improve the robustness of SR models. We study the problem under more rigorous settings. We propose training existing models on misaligned LR-HR data and evaluating their performance on perfectly aligned synthetic datasets to focus solely on the detail-recovery capabilities of SR models. Current approaches involve creating alternative inputs that better match the HR images [52, 71].

For example, Zhang et al. [71] uses the center crop of short-focus and telephoto images to construct LR-HR pairs. However, this approach is limited to small and simple shifts and fails with more complex misalignments. Sun et al. [52] leverages unpaired images to learn real-world degradations and generate alternative inputs, but this method does not fully utilize paired LR-HR data and complete supervision. Regarding color mismatch, few works have addressed this issue [9, 70]. A related work for under-display cameras [17] introduces a domain adaptation module to transfer color information between a high-quality reference and a degraded image. In addition to another module for geometrical alignment. However, these modules are trained separately which results in error accumulation that produces unsatisfactory visual results for our SR problem.

To address these issues, we propose a modification module to create a more realistic and aligned LR image, denoted as LR Mimicking Module. Our method takes the LR (LR) and downscaled HR ($\widehat{HR}$) as input and aims to transform the downscaled HR into the Mimicked-LR (Mim_{LR}) by incorporating characteristics from the LR input. This approach ensures that the produced Mim_{LR} is geometrically and color-aligned with the HR image, facilitating pixel alignment for loss computations. Additionally, the Mim_{LR} retains the degradation and quality of the original LR image, allowing it to be replaced by the original LR during inference without additional manipulation. Our generation module is plug-and-play, integrating seamlessly with SR models in an end-to-end learning manner. We show that our model significantly improves the performance of existing SR models, from CNN to Transformer architectures, and from large models to efficient ones, outperforming state-of-the-art alignment attempts.
To conclude, the contribution of this paper is twofold:

- We address the misalignment issue in existing SR works and thoroughly revise the existing alignment processes by establishing comprehensive benchmarks on synthetic and realistic datasets.
- We propose a novel plug-and-play module that mimics the aligned LR to improve the SR learning stage, enhancing the robustness of SR models against misalignment without adding any learnable parameters during inference.
- Extensive comparisons across diverse SR models, ranging from CNNs to Transformers and from lightweight to heavy models, validate the effectiveness and generalization ability of our proposed method.

2 Related Work

Realistic Super Resolution Most research on Single Image Super Resolution (SISR) [13, 22, 32, 35, 54, 66, 69] relies on synthetic datasets with simple degradation models. To develop more robust SR models, researchers have been working on generating synthetic SR datasets with more complex degradation. These methods often involve degradation pipelines [25, 30, 64] or parameterized degradation processes to create blur kernels. Degradation parameters can be derived through numerical methods [46, 47] or optimized jointly using deep learn-

ing [12,19,21]. Additionally, some approaches leverage deep learning to generate training samples from pre-estimated degradation parameters [3, 6, 59, 73].

Another line of research tackles real-world super-resolution through domain adaptation [10, 18, 28, 36, 38, 45, 61, 68]. In these works, real-world LR images and synthetic clean LR images are treated as different domains, with the goal of narrowing the gap between them. To make SR models more generalizable, self-learning-based methods have been employed to minimize the discrepancy between training and testing data [15, 44, 49, 50, 71, 72].

To address real-world SR in a fully supervised manner, several realistic SR datasets have been introduced to create paired image data. These datasets often try to capture LR and HR images of the same scene by varying the focal length of a zoom lens [7,8,58,67]. Other approaches utilize pixel binning [30] and beam-splitter techniques [26]. However, all of these datasets suffer from misalignment issues due to variations in the capture systems, optical setups, or imaging conditions. In our work, we focus on real-world SR using realistic SR datasets captured with zoom lenses.

Alignment of Paired Images Misalignment is a common challenge in low-level vision. When capturing realistic datasets, multiple devices, varying optical systems, or changing capture settings are often required, leading to misalignment between the input and GT pairs. This problem is evident in Learned-Based ISP [14,23,24,48], Realistic SR [7,67,71], and image restoration [17]. A standard approach to mitigate misalignment is to apply global alignment techniques like affine transformations [7,24] or dense alignment using deep learning methods [23].

However, global alignment alone is insufficient to achieve pixel-level correspondence between image pairs. To address this, the Contextual Loss (CXLoss) [41] and the Contextual Bilateral Loss (CoBi Loss) [41] were introduced for training on misaligned image pairs by computing the loss based on the most similar features between the compared pair. Differently, zhang et al. [70] proposed an alignment that densely aligns the GT patch with a color-reconstructed version of the RAW input using optical flow

Other works generate a new pair that is better aligned. Zhang et al. [71] generates an auxiliary LR that is better aligned with the GT and used to train the restoration model. Feng et al. [17] instead produces a pseudo-GT better aligned with the input. They utilize a domain alignment module, for colors and degradations, and a geometric alignment module for structural alignment.

3 Methodology

Motivation: We propose a module designed for seamless integration with any SR model to address alignment issues. This module generates a Mimicked-LR image that replaces the original LR image during training. The Mimicked-LR must meet three key criteria: First, it should be geometrically well-aligned with the GT HR image. Second, it must have the same color characteristics as the GT HR image to prevent model confusion and color artifacts. Third, the Mimicked-LR should accurately replicate the degradation characteristics of the original LR

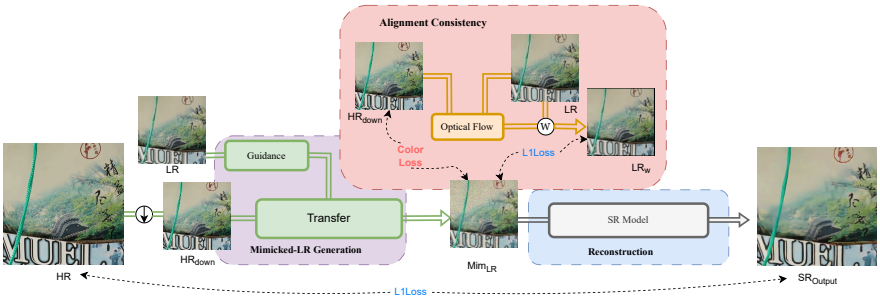

Fig. 3: Overview. In the training stage, we employ a modification module for alignment purposes to create a Mimicked-LR input with improved consistency with the GT. Such module generates or mimics a new LR (Mim_{LR}) which (i) has the same degradations as the LR input image, (ii) is geometrically aligned with the HR, and (iii) is color consistent with the GT (same colors, brightness, etc). During inference, we remove the generation/alignment module and replace the Mimicked-LR with the normal LR input, allowing for a sharper output without color changes or distortion.

image, enabling a direct replacement during testing without performance loss. Hence, the model learns the SR task without being hindered by non-SR-related differences in the dataset. The overall pipeline is shown in Fig. 3.

3.1 Mimicked-LR Generation

To generate the Mimicked-LR, we design a straightforward yet effective architecture without any specialized layers. As depicted in Fig. 3, we process the downscaled HR and LR images, transforming the downscaled HR to match the degradation of the LR image. The model relies on simple convolutions, ensuring efficient computation and minimal training cost, which allows for easy integration with any existing SR models without extra computational burden.

Guidance Network: Realistic degradation is non-uniform and can be scene-specific, varying for each image patch during training. To address this, we incorporate information from the LR image through a guidance network. As shown in Fig. 4, the guidance network comprises a sequence of CNN layers that extract a guidance vector from the LR and downscaled HR images. We follow the shallow layers from ResNet, enhancing them with additional convolutions for deep feature extraction. The extracted features are then transformed into a guidance vector using global average pooling, capturing the overall style and domain information from the LR input. This vector further serves as an attention mechanism to guide the Mimicked-LR generation.

Transfer Network: The transfer network aims to transform the downscaled HR to match the LR degradation without introducing color or geometric shifts. Initially, we apply the guidance vector as a direct scale and shift to process the downscaled HR, projecting features from the HR domain into the target LR domain. The combined features are then processed through an encoder-decoder architecture for image generation.

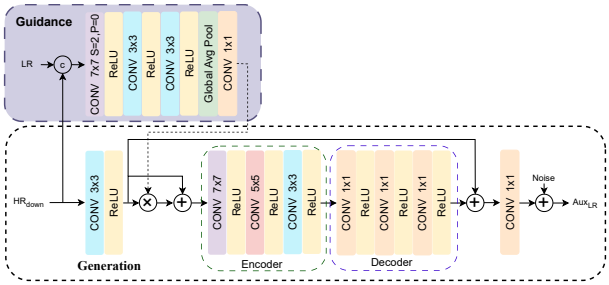

Fig. 4: Illustration of the proposed LR Mimicking Architecture.

Specifically, the encoder consists of convolutions with varying receptive fields to extract features in a coarse-to-fine manner, understanding global dependencies and better modeling the gap to align the two domains. The decoder reconstructs the degraded HR patch (Mimicked-LR) from the encoded features, using only 1×1 convolution layers to avoid pixel shifting for better alignment. We incorporate conventional practices in low-level vision by adding skip connections and a final convolution layer. Additionally, we introduce noise at the end to simulate complex degradation, utilizing a mix of Gaussian and JPEG noise. The modification process maintains the same spatial size, preventing artifacts and shifts that typically result from downsampling and upsampling.

3.2 Objective Functions

Geometrical Alignment: To ensure the geometrical alignment between the Mimicked-LR and GT, we use a downsampled version of the HR image (to the same size as the LR image) modified to mimic the characteristics of the LR. Inspired by [70], we adapt a misalignment loss using an optical flow network to densely align the compared image patches before applying the loss.

Specifically, we calculate the optical flow between the downsampled HR patch $\widehat{HR}$ and the LR patch LR, then use the optical flow to align the LR patch LR_w. We compute the L1 loss between LR_w and the output Mimicked-LR Mim_{LR}, ensuring geometrical alignment between Mim_{LR} and GT HR while retaining the characteristics of the LR. By limiting the optical flow alignment to the loss calculation, we avoid warping artifacts in the generated images.

Unlike previous works [52,71], our method modifies $\widehat{HR}$ to match the degradation of the LR rather than generating a new image, simplifying the task and avoiding generative artifacts. The misalignment loss is defined as follows:

$$LR_w = \mathcal{W}(LR, \psi), \qquad \psi = \mathcal{F}(\widehat{HR}, LR) \tag{1}$$

$$\mathcal{L}_{deg} = ||m \cdot (LR_w - Mim_{LR})||_1, \qquad Mim_{LR} = \mathcal{M}(\widehat{HR}, LR) \tag{2}$$

$$m_i = \begin{cases} 1, [\omega(1, \psi)]_i \geq 1 - \epsilon \\ 0, \quad otherwise \end{cases} \tag{3}$$

where $\mathcal{F}(.)$ and $\mathcal{W}(.)$ denote the optical flow and the warping operation, respectively. m_i is the mask representing valid positions in the optical flow, and $\mathcal{M}(.)$ is the Mimicked-LR creation module. $\mathcal{L}_{deg}$ represents the degradation loss. **Color Alignment**: Ensuring color consistency between the input and GT during training is crucial to avoid model confusion and color artifacts. Inspired by [14], we employ the Color Difference Network (CDNet) [57] as an off-the-shelf color loss function. CDNet, trained on large-scale color difference datasets, achieves state-of-the-art performance and is fully differentiable, making it suitable for backpropagation. We formulate our color loss as follows:

$$\mathcal{L}_{CD} = CD(Mim_{LR}, \widehat{HR}) \tag{4}$$

3.3 Learning Diagram

Training: In the training stage, we integrate our Mimicking module with the SR model, replacing the conventional misaligned LR-HR pair with the Mimicked-LR-HR pair. We jointly train the Mimicking module and the SR model for optimization. To prevent the SR model from interfering with the Mimicking module's output, we detach the Mim_{LR} image from the computation graph. This ensures that the SR model's backpropagation does not affect the weights of the Mimicking module, maintaining the desired characteristics of the Mim_{LR} image and preventing it from being optimized to simplify the SR task. The complete loss function is described in the following equation:

$$\mathcal{L}_{res} = ||SR_{Out} - HR||_1, \qquad SR_{Out} = \mathcal{R}(Mim_{LR}) \tag{5}$$

$$\mathcal{L}_{total} = \mathcal{L}_{res} + \mathcal{L}_{deg} + \lambda \mathcal{L}_{CD} \tag{6}$$

where $\mathcal{R}(.)$ is the SR Model. λ is the balancing factor between the degradation loss $\mathcal{L}_{deg}$ and color difference loss $\mathcal{L}_{CD}$ for the mimicking module optimization. **Inference**: During inference, we simply remove our Mimicking module and replace Mim_{LR} with the conventional LR. Our process ensures that Mim_{LR} has the same degradation characteristics as the LR image, allowing this substitution to be made seamlessly without introducing artifacts or performance drop.

4 Experiments

4.1 Overall Evaluation Protocol

This work aims to assess the robustness of the previous alignment process for SR models. To achieve this, we combine these alignment methods with widely used SR models and ***retrain them on real-world datasets*** that inherently contain misalignment, while ***evaluating*** their performance using ***synthetic datasets*** with perfectly aligned LR-HR pairs. Notably, conventional SR metrics (PSNR/SSIM) cannot be reliably evaluated on the real-world dataset due to misalignment in the testing set; therefore, we compare the non-reference image quality metrics (NIQE [42], NRQM [37], PI [5]).

4.2 Datasets

Training: Specifically, our experiments involve training on two distinct real-world datasets: RealSR [7] and SR-RAW [67].
RealSR [7] is a realistic dataset captured using DSLR cameras equipped with zoom lenses. LR and HR images were obtained by varying the focal length (28, 35, 50, and 105 mm) to explore different scaling factors. The dataset employs an image registration framework to achieve pixel-wise alignment. This process begins by cropping a central region from the HR image to mitigate severe distortions, serving as a reference for aligning images captured with other focal lengths. Optimization involves iteratively adjusting an affine transformation matrix and luminance parameters. Despite these efforts, residual misalignment remains evident, as depicted in Fig. 1. We focus our experiments on the $\times 4$ scaling factor.
SR-RAW [67], on the other hand, is captured using a similar process but includes a broader range of focal lengths (24–240 mm) and offers both RAW and RGB images. For our study, we utilize RGB images and focus on specific pairs (24/100, 35/150, and 50/240) to form a $4\times$ scale dataset. The initial alignment process utilizes the Euclidean motion model [16], resulting in more complex scenes and higher-resolution images compared to RealSR. However, this registration method also leads to image pairs with noticeable pixel-wise misalignment.
Testing: For comprehensive evaluation under full-reference conditions, we utilize synthetic SR benchmark datasets renowned for their absence of misalignment issues: Set5 [4], Set14 [63], BSD100 [39], Urban100 [20], Manga109 [40], and DIV2KRK [3], derived from the DIV2K dataset [2] with complex degradation pipelines. Additionally, we incorporate non-reference image quality metrics to assess performance on the previously mentioned real-world datasets RealSR [7] and SR-RAW [67], using their official training/testing splits.

4.3 Implementation Details.

For a consistent and fair comparison, we employ identical hyperparameters across all alignment processes tested on various SR models. We perform random cropping of patches sized $128\times$ 128, augmented with random flipping and rotation. The reconstruction loss used for all models is L1 loss. Optimization is carried out using Adam optimizer [29] for 200,000 iterations, with a batch size of 32 and an initial learning rate of 1e-3. To schedule the learning rate, we employ a Cosine Annealing LR scheduler [34]. All experiments are implemented in the PyTorch framework on a single NVIDIA RTX 4090 GPU.

4.4 SOTA Alignment Counterparts

For a comprehensive benchmark, we test 5 alignment processes including processes created for different tasks. In Tab. 1, 2, 3 Alignment Loss is the training process presented in [70]. It densely aligns the gt patch with the input patch using optical flow. CXLoss is the contextual loss proposed in [41] and the CoBi Loss is Contextual Bilateral Loss presented in [67]. They compute loss on similar

features between the compared images. Auxiliary Input uses the auxiliary-LR generation module from [71] to generate an auxiliary LR that replaces the LR during training. lastly, Pseudo GT is the alignment process proposed in [17]. They apply domain and geometrical alignment using two models to create a Pseudo GT that is used during training. We adapt these processes to work for the realistic SR task.

Moreover, we evaluate the effectiveness of the tested alignment processes across four state-of-the-art (SOTA) SR models. Our selection includes two CNN-based models: RRDBNet [56] and SeeMore [62], as well as two transformer-based models: SwinIR [31] and DAT [11]. These models differ not only in their architectural foundations but also in terms of model complexity and size. This diverse selection allows us to assess alignment techniques across a spectrum of SR methodologies.

4.5 Full-Reference Quantitative Benchmark on Synthetic Data

We employ synthetic benchmarks (Evaluation in Sec. 4.2) for quantitative full-reference evaluation to ensure no misalignment between the input and the ground truth (GT). Despite the different distributions of our training data (realistic SR datasets), which typically yield lower performance on synthetic datasets, they facilitate a valuable relative comparison between the alignment approaches. By evaluating the same SR model trained with different alignment processes, we aim to identify the optimal alignment strategy that produces sharper outputs without introducing color shifts. Such outputs typically achieve superior performance in synthetic benchmarks. Conversely, inadequate alignment processes during training lead to SR models that introduce geometric and color shifts, resulting in blurred outputs with noticeable color artifacts. These outputs typically perform poorly on synthetic benchmarks, underscoring the importance of effective misalignment correction strategies.

Training with SR-RAW Dataset: The results of our models trained with different alignment processes on the SR-RAW dataset [67] and evaluated on synthetic data are presented in Tab. 1. Due to significant misalignment in the dataset, we omitted evaluation based on L1 loss, which resulted in severely blurred outputs. Across all tested methods at a $4\times$ scale SR, our alignment process consistently achieves superior performance compared to other alignment methods. This performance advantage holds across various SR models, demonstrating the robustness and consistency of our approach regardless of the specific model architecture. In contrast, alternative alignment methods exhibit varying performance across different SR models, indicating disparities in effectiveness. Moreover, our method demonstrates balanced performance in both PSNR and SSIM metrics, unlike other methods that may prioritize one metric over the other due to the sensitivity of PSNR to blurred outputs. This underscores again the efficacy of our approach in achieving high-quality outputs.

Our approach leverages a simple module that is jointly trained with the SR model, eliminating the need for additional separate models (e.g., Pseudo GT) or complex calculations (e.g., CX Loss, CoBi Loss). Unlike alignment loss methods

Table 1: Comparison of the different alignment processes on the synthetic benchmark. All models are trained on **SR-RAW** dataset [67]. Best and second best performances are highlighted. Our training processes archive the best performance in all tested methods. Please refer to Training Processes Evaluated Sec. 4.4 for information about training processes.

Methods	Training Processes	Set5		Set14		BSD100		Urban100		Manga109		DIV2KRK	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RDDBNet [56]	Alignment Loss	20.07	0.7486	17.69	0.6381	18.00	0.6019	14.41	0.5755	17.33	0.7104	15.46	0.5166
	CXLoss	22.42	0.7017	20.83	0.5902	21.15	0.5446	18.85	0.5641	19.44	0.6939	19.69	0.5888
	CoBi Loss	22.02	0.5625	20.60	0.4650	20.49	0.4049	19.35	0.4798	20.10	0.5895	21.02	0.4909
	Auxiliary Input	18.20	0.6996	16.88	0.5809	19.59	0.5601	14.24	0.4937	15.73	0.6460	15.10	0.5258
	Pseudo GT	22.11	0.7479	20.18	0.6338	20.19	0.5964	17.61	0.5844	20.05	0.7196	17.66	0.5654
	Ours	23.03	0.7599	22.73	0.6618	23.26	0.6356	19.99	0.6399	21.42	0.7811	22.95	0.6342
SwinIR [31]	Alignment Loss	22.74	0.7779	22.26	0.6676	22.04	0.6337	19.91	0.6350	21.35	0.7526	20.40	0.6125
	CXLoss	22.39	0.6144	21.05	0.5247	20.52	0.4743	19.27	0.5362	20.46	0.6435	18.93	0.2947
	CoBi Loss	21.18	0.5835	19.08	0.4693	17.86	0.3890	18.44	0.4874	19.44	0.5901	19.19	0.3143
	Auxiliary Input	21.01	0.7433	19.33	0.6272	21.09	0.5987	17.79	0.5865	19.94	0.7222	19.76	0.6076
	Pseudo GT	22.96	0.7453	22.48	0.6406	23.10	0.6091	20.45	0.6011	20.88	0.7247	19.58	0.5945
	Ours	23.73	0.7748	23.26	0.6725	24.04	0.6434	20.98	0.6555	22.90	0.7947	23.41	0.6464
DAT [11]	Alignment Loss	21.99	0.7632	22.76	0.6644	22.40	0.6257	20.73	0.6361	21.41	0.7478	21.46	0.6132
	CXLoss	22.02	0.5625	20.60	0.4650	20.49	0.4049	19.35	0.4798	20.10	0.5895	19.07	0.2772
	CoBi Loss	22.34	0.5849	20.53	0.4796	19.09	0.3925	19.21	0.4764	20.50	0.5955	19.92	0.2873
	Auxiliary Input	22.67	0.7496	19.79	0.6226	21.65	0.5937	18.89	0.5830	20.68	0.7141	21.49	0.6164
	Pseudo GT	23.70	0.7399	23.34	0.6442	23.37	0.6029	20.92	0.6033	21.53	0.7294	20.63	0.6046
	Ours	22.87	0.7652	22.40	0.6690	23.39	0.6425	20.13	0.6497	22.48	0.7925	23.61	0.6533
SeeMore [62]	Alignment Loss	21.85	0.7718	20.90	0.6668	20.69	0.6246	18.45	0.6207	20.23	0.7439	17.84	0.5843
	CXLoss	21.02	0.5355	19.84	0.4397	18.68	0.3766	17.46	0.4297	19.28	0.5394	16.51	0.2488
	CoBi Loss	20.85	0.5751	18.55	0.4614	17.64	0.3706	17.63	0.4600	18.67	0.5738	18.02	0.2721
	Auxiliary Input	18.49	0.7087	17.25	0.6035	18.80	0.5851	14.901	0.5526	17.94	0.6866	17.50	0.5813
	Pseudo GT	22.45	0.7429	21.54	0.6389	21.76	0.6036	19.20	0.5923	19.40	0.7159	17.60	0.5498
	Ours	23.90	0.7626	23.28	0.6669	23.56	0.6348	21.00	0.6410	22.35	0.7797	23.33	0.6398

applied during the reconstruction stage, we integrate alignment considerations into the LR Mimicking stage, enhancing efficiency and performance. Furthermore, our method exhibits robustness compared to limited alignment modules used in the Auxiliary Input approach [71].

Training with RealSR Dataset: We can notice the consistent pattern on the models trained with RealSR dataset [7] as shown in Tab. 2. Despite RealSR's relatively good alignment compared to the SR-RAW dataset [67], we observe notable improvements when employing alignment processes. Even minor misalignment present in the dataset significantly impacts model performance, particularly evident in the poorer results obtained with L1 loss. This underscores the sensitivity of SR models to alignment issues, highlighting the necessity and effectiveness of robust alignment strategies for enhancing performance across diverse datasets.

4.6 Realistic Data Benchmark

Since our focus is on training with realistic SR datasets, evaluating alignment processes on these datasets' images is crucial. Due to the presence of misalignment issues in both training and validation datasets, traditional full-reference evaluation methods like PSNR and SSIM are not suitable. Therefore, we employed No Reference Evaluation methods for a more accurate comparison.

Table 2: Comparison of the different alignment processes on the synthetic benchmark. All models are trained on **RealSR** dataset [7]. Best and second best performances are highlighted. Our training processes archive the best performance in all tested methods. Please refer to Training Processes Evaluated Sec. 4.4 for information about training processes.

Methods	Training Processes	Set5		Set14		BSD100		Urban100		Manga109		DIV2KRK	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RDDBNet [56]	L1Loss	22.40	0.6788	21.40	0.6025	21.067	0.5641	19.24	0.5643	21.02	0.7079	23.89	0.6657
	Alignment Loss	22.48	0.6900	21.65	0.6078	21.28	0.5700	19.61	0.5751	20.95	0.7266	23.89	0.6626
	CXLoss	20.76	0.5948	19.45	0.5141	18.88	0.4584	17.86	0.4994	19.66	0.6507	23.48	0.6268
	CoBi Loss	20.51	0.6102	19.11	0.5234	18.09	0.4570	17.30	0.5054	19.91	0.6764	23.42	0.6418
	Auxiliary Input	22.94	0.6796	22.05	0.5853	22.59	0.5574	19.51	0.5206	20.49	0.6786	22.72	0.6302
	Pseudo GT	22.86	0.7014	21.38	0.6080	21.04	0.5707	18.94	0.5625	20.74	0.7139	23.98	0.6703
	Ours	24.62	0.7587	23.45	0.6515	23.75	0.6289	21.16	0.6239	22.76	0.7815	23.56	0.6679
SwinIR [31]	L1Loss	23.51	0.7153	21.85	0.6200	21.66	0.5883	19.77	0.5901	21.95	0.7476	24.10	0.6732
	Alignment Loss	23.92	0.7188	22.53	0.6342	22.08	0.5930	20.61	0.6126	22.84	0.7673	24.10	0.6715
	CXLoss	21.56	0.6043	20.40	0.5354	19.67	0.4808	19.14	0.5451	21.16	0.6917	22.49	0.4946
	CoBi Loss	20.65	0.5791	19.57	0.5155	18.43	0.4528	18.39	0.5223	20.71	0.6839	22.37	0.5285
	Auxiliary Input	23.96	0.7035	22.61	0.6000	23.02	0.5706	19.97	0.5413	21.19	0.7053	23.79	0.6501
	Pseudo GT	24.11	0.7397	22.52	0.6491	22.09	0.6093	20.26	0.6125	22.29	0.7643	24.01	0.6720
	Ours	24.69	0.7647	23.86	0.6680	23.83	0.6357	21.50	0.6462	23.81	0.7992	24.27	0.6797
DAT [11]	L1Loss	23.05	0.6950	21.70	0.6058	21.12	0.5633	19.42	0.5621	21.46	0.7246	24.14	0.6738
	Alignment Loss	24.04	0.7280	22.50	0.6324	22.32	0.6015	20.46	0.5956	22.21	0.7490	24.13	0.6726
	CXLoss	21.31	0.5816	20.60	0.5268	20.20	0.4937	18.98	0.5251	20.82	0.6657	22.59	0.5179
	CoBi Loss	20.25	0.5667	19.43	0.4982	18.61	0.4442	18.19	0.4980	20.37	0.6542	22.19	0.5035
	Auxiliary Input	24.10	0.7193	22.83	0.6138	23.12	0.5858	20.29	0.5684	21.55	0.7222	23.95	0.6612
	Pseudo GT	23.67	0.7112	21.99	0.6246	21.44	0.5812	19.68	0.5859	21.77	0.7439	24.13	0.6766
	Ours	24.82	0.7710	24.05	0.6739	23.86	0.6433	21.80	0.6592	23.77	0.8045	24.35	0.6885
SeeMore [62]	L1Loss	23.45	0.7047	22.24	0.6275	21.73	0.5838	20.01	0.5849	21.83	0.7305	23.80	0.6631
	Alignment Loss	23.75	0.7067	22.76	0.6424	22.64	0.6216	20.45	0.5998	21.56	0.7290	23.06	0.6412
	CXLoss	21.96	0.5923	21.33	0.5410	20.92	0.4965	19.46	0.5098	21.68	0.6697	22.00	0.4789
	CoBi Loss	21.38	0.6052	20.35	0.5325	19.83	0.4893	18.70	0.5231	20.48	0.6535	22.34	0.5259
	Auxiliary Input	24.12	0.7201	22.87	0.6108	23.35	0.5860	20.16	0.5584	21.43	0.7084	23.22	0.6464
	Pseudo GT	24.39	0.7354	22.65	0.6478	22.32	0.6139	20.45	0.6074	21.87	0.7346	23.64	0.6590
	Ours	24.15	0.7483	23.08	0.6478	23.17	0.6194	20.86	0.6237	22.88	0.7813	24.14	0.6773

As observed from the results in Tab. 3, similar to the synthetic benchmark, our alignment process achieves superior performance by attaining the highest scores across the NIQE and PI, underscoring the high quality of our process outputs. Furthermore, our method demonstrates remarkable consistency in performance across different SR models. This improvement is primarily attributed to the consistent output quality and minimal color distortions, aspects that are effectively evaluated by no-reference metrics.

Note that we also include the no-reference metric, NRQM, in our evaluation. This metric is less robust to artifacts. For instance, methods using CXLoss and CoBi Loss may achieve higher NRQM scores but typically perform worse on NIQE and PI metrics. Please refer to the qualitative results in Sec. 4.7 for the visual artifacts when training with CXLoss and CoBi. Despite this, our method consistently outperforms all other methods in terms of NRQM score, indicating our effectiveness compared to alternative approaches.

4.7 Qualitative Results

Alignment Output: First, we show the effectiveness of our alignment process in producing a better-aligned input. In Fig. 5 we show the improved alignment

Table 3: Evaluation with Realistic SR Datasets on No Reference Image Quality Metrics. Models are trained and evaluated on the same dataset. Best and second best performances are highlighted.

Methods	Training Processes	SR-RAW			RealSR		
		NIQE↓	NRQM↑	PI↓	NIQE↓	NRQM↑	PI↓
RDDBNet [56]	Synthetic Data	7.42	3.24	7.19	8.74	2.92	7.99
	Alignment Loss	7.88	3.19	7.47	7.50	3.23	7.17
	CXLoss	6.16	4.95	6.84	7.61	4.42	6.53
	CoBi Loss	8.96	5.22	5.67	7.41	4.46	6.53
	Auxiliary Input	7.02	3.09	7.52	7.63	3.09	7.35
	Pseudo GT	8.18	2.81	7.78	7.53	3.26	7.19
	Ours	6.67	4.11	6.40	7.07	3.47	6.79
SwinIR [31]	Synthetic Data	7.41	3.25	7.18	8.72	2.93	7.97
	Alignment Loss	7.74	3.13	7.42	7.49	3.17	7.19
	CXLoss	26.61	6.35	14.90	16.189	4.97	10.39
	CoBi Loss	27.06	5.68	15.28	17.14	4.92	10.80
	Auxiliary Input	7.39	3.37	7.13	7.65	3.14	7.33
	Pseudo GT	8.45	2.86	7.88	7.40	3.21	7.15
	Ours	6.94	3.73	6.71	6.94	3.39	6.78
DAT [11]	Synthetic Data	7.43	3.23	7.20	8.71	2.94	7.96
	Alignment Loss	7.96	3.15	7.51	7.59	3.16	7.26
	CXLoss	32.84	6.15	17.91	15.98	4.95	10.23
	CoBi Loss	24.18	5.93	13.73	17.62	5.08	11.07
	Auxiliary Input	7.48	3.34	7.19	7.56	3.20	7.27
	Pseudo GT	8.69	2.90	7.97	7.49	3.24	7.18
	Ours	6.93	3.93	6.58	6.98	3.42	6.78
SeeMore [62]	Synthetic Data	7.25	3.29	7.09	8.64	2.92	7.93
	Alignment Loss	7.73	3.12	7.40	7.55	3.16	7.23
	CXLoss	27.23	6.43	15.10	19.56	4.97	11.96
	CoBi Loss	25.18	5.98	14.25	15.29	4.92	9.91
	Auxiliary Input	7.37	3.24	7.20	7.51	3.14	7.23
	Pseudo GT	8.50	2.78	7.93	7.51	3.20	7.17
	Ours	6.91	3.74	6.68	6.84	3.51	6.65

compared to the SOTA counterpart based on optical flow (OF). Quantitatively, compared to the OF alignment, our method reduces the per-pixel error by 69.36% and 9.77% on SR-Raw (less aligned) and RealSR (better aligned), respectively. Additionally, we can notice the improved color alignment in the error map of the uniform areas which is not addressed by the OF method. Likewise, we can appreciate our method's ability to capture the LR image degradation producing a Mimicked LR image with similar characteristics.

SR Output: For qualitative evaluation, we present the outputs of various SR models trained using different alignment processes. In Fig. 6, we showcase the results of the SwinIR [31] and Fig. 7 of the SeeMoRe [62] model trained on the SR-RAW dataset under different alignment methods. Our proposed alignment process consistently yields the best outcomes, producing sharp images devoid of color artifacts while preserving structural fidelity. In contrast, other alignment methods exhibit noticeable color misalignment issues, resulting in visible artifacts in their outputs. In Fig. 6 following the green arrow we can see severe color artifacts in the other alignment processes. Additionally, in the zoomed area we can see our method producing sharper details with better texture.

Similarly, in Fig. 7, we can see severe color artifacts in the building with bad color reproduction. In the zoomed area we can see our method producing better shadows, better details, and texture. However, the Pseudo GT method mitigates color artifacts to some extent by considering dataset domain disparities,

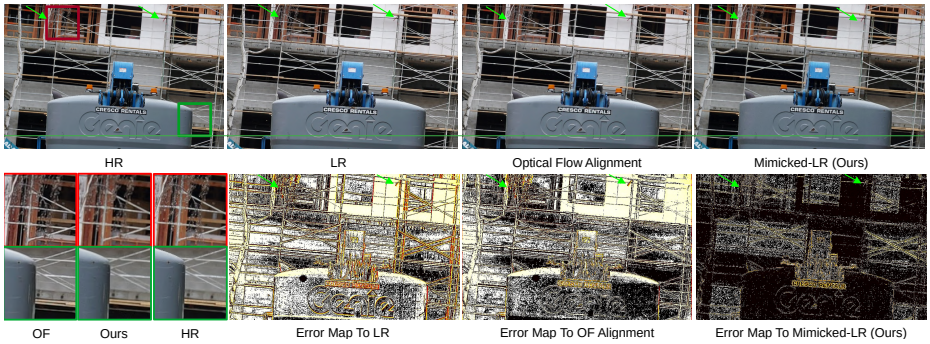

Fig. 5: Visualization of our Mimicking alignment quality. We compare to dense alignment using Optical Flow. The error maps illustrate that our method produces outputs better aligned with the HR image.

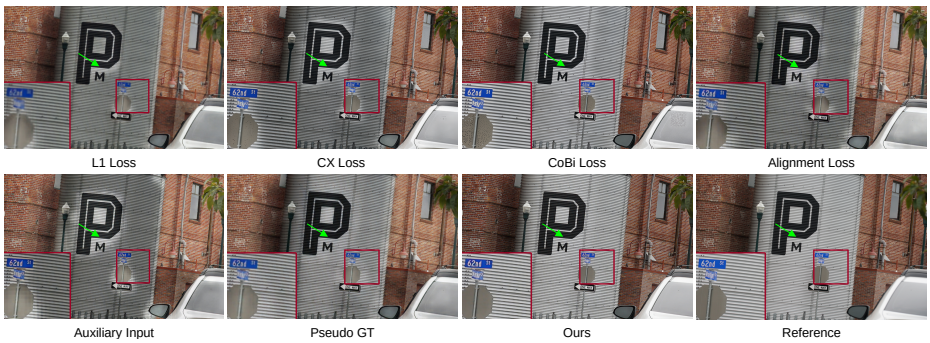

Fig. 6: Qualitative Results of SwinIR [31] Model Trained on SR-RAW Dataset

albeit with reduced robustness due to error accumulation from separately trained models. Models trained with CXLoss and CoBi Loss exhibit cartoon artifacts, as these methods compare features at the expense of fine image details.

Additionally, we show the output of the different alignment processes on the RealSR dataset in Fig. 8. The performance improvement achieved through our method is evident. Our process yields sharper outputs with more details compared to the blurred and washed-out results from other methods. Following the green arrows, we observe enhanced textures and details in the clouds and water. These visual results, combined with the previously presented quantitative data, validate the effectiveness of our model in handling color and geometric misalignment in realistic SR datasets. Furthermore, our approach proves robust across different SR models and datasets, regardless of varying levels of misalignment.

Fig. 7: Qualitative Results of SeeMore [62] Model Trained on SR-RAW Dataset

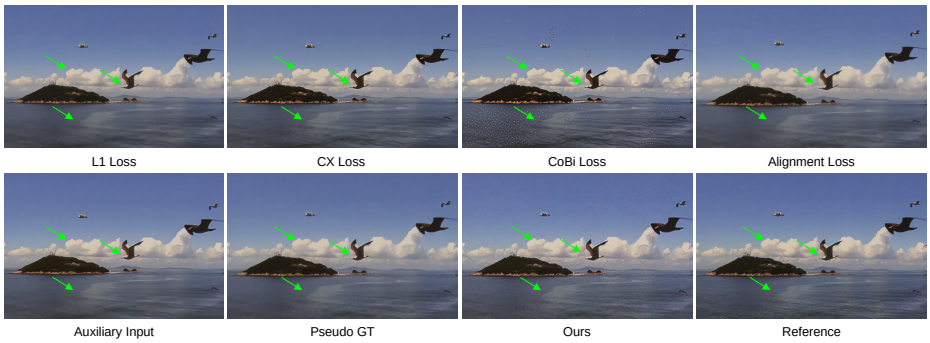

Fig. 8: Qualitative Results of DAT [11] Model Trained on RealSR Dataset.

5 Conclusion

We propose a novel training process designed to effectively train SR models on realistic SR datasets. Our approach involves a simple module that creates an image mimicking the LR image while maintaining alignment with the HR image. This module is plug-and-play, allowing it to be integrated with any SR model without additional modifications. Moreover, we introduce an extensive benchmark for various alignment processes across different datasets and SR models. Our training process consistently outperforms all other tested alignment methods. The superior visual quality of our method is evident when compared to other alignment processes, showcasing sharper, more detailed outputs without color changes or distortions. Our extensive evaluation demonstrates the effectiveness and generality of our method across different datasets and models. We believe our work lays the foundation for a new approach to addressing misalignment issues in fully supervised real-world datasets, potentially leading to significant advancements in the field of Super-Resolution.

Acknowledgments: This work was supported by The Alexander von Humboldt Foundation.

References

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1692–1700 (2018)
2. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 126–135 (2017)
3. Bell-Kligler, S., Shocher, A., Irani, M.: Blind super-resolution kernel estimation using an internal-gan. Advances in Neural Information Processing Systems **32** (2019)
4. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012)
5. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The 2018 pirm challenge on perceptual image super-resolution. In: Proceedings of the European conference on computer vision (ECCV) workshops. pp. 0–0 (2018)
6. Bulat, A., Yang, J., Tzimiropoulos, G.: To learn image super-resolution, use a gan to learn how to do image degradation first. In: Proceedings of the European conference on computer vision (ECCV). pp. 185–200 (2018)
7. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: 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. pp. 3086–3095 (2019)
8. Chen, C., Xiong, Z., Tian, X., Zha, Z.J., Wu, F.: Camera lens super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1652–1660 (2019)
9. Chen, H., He, X., Qing, L., Wu, Y., Ren, C., Sheriff, R.E., Zhu, C.: Real-world single image super-resolution: A brief review. Information Fusion **79**, 124–145 (2022)
10. Chen, S., Han, Z., Dai, E., Jia, X., Liu, Z., Xing, L., Zou, X., Xu, C., Liu, J., Tian, Q.: Unsupervised image super-resolution with an indirect supervised path. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 468–469 (2020)
11. Chen, Z., Zhang, Y., Gu, J., Kong, L., Yang, X., Yu, F.: Dual aggregation transformer for image super-resolution. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 12312–12321 (2023)
12. Cornillere, V., Djelouah, A., Yifan, W., Sorkine-Hornung, O., Schroers, C.: Blind image super-resolution with spatially variant degradations. ACM Transactions on Graphics (TOG) **38**(6), 1–13 (2019)
13. Dong, C., Loy, C.C., He, K., Tang, X.: 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. pp. 184–199. Springer (2014)
14. Elezabi, O., Conde, M.V., Timofte, R.: Simple image signal processing using global context guidance. arXiv preprint arXiv:2404.11569 (2024)
15. Emad, M., Peemen, M., Corporaal, H.: Dualsr: Zero-shot dual learning for real-world super-resolution. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1630–1639 (2021)
16. Evangelidis, G.D., Psarakis, E.Z.: Parametric image alignment using enhanced correlation coefficient maximization. IEEE transactions on pattern analysis and machine intelligence **30**(10), 1858–1865 (2008)
17. Feng, R., Li, C., Chen, H., Li, S., Gu, J., Loy, C.C.: Generating aligned pseudo-supervision from non-aligned data for image restoration in under-display camera.

In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5013–5022 (2023)
18. Fritsche, M., Gu, S., Timofte, R.: Frequency separation for real-world super-resolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3599–3608. IEEE (2019)
19. Gu, J., Lu, H., Zuo, W., Dong, C.: Blind super-resolution with iterative kernel correction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1604–1613 (2019)
20. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5197–5206 (2015)
21. Huang, Y., Li, S., Wang, L., Tan, T., et al.: Unfolding the alternating optimization for blind super resolution. Advances in Neural Information Processing Systems **33**, 5632–5643 (2020)
22. Hui, Z., Gao, X., Yang, Y., Wang, X.: Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th acm international conference on multimedia. pp. 2024–2032 (2019)
23. Ignatov, A., Chiang, C.M., Kuo, H.K., Sycheva, A., Timofte, R.: Learned smartphone isp on mobile npus with deep learning, mobile ai 2021 challenge: Report. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2503–2514 (2021)
24. Ignatov, A., Van Gool, L., Timofte, R.: Replacing mobile camera isp with a single deep learning model. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 536–537 (2020)
25. Ji, X., Cao, Y., Tai, Y., Wang, C., Li, J., Huang, F.: Real-world super-resolution via kernel estimation and noise injection. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 466–467 (2020)
26. Joze, H.R.V., Zharkov, I., Powell, K., Ringler, C., Liang, L., Roulston, A., Lutz, M., Pradeep, V.: Imagepairs: Realistic super resolution dataset via beam splitter camera rig. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 518–519 (2020)
27. Khani, M., Sivaraman, V., Alizadeh, M.: Efficient video compression via content-adaptive super-resolution. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4521–4530 (2021)
28. Kim, G., Park, J., Lee, K., Lee, J., Min, J., Lee, B., Han, D.K., Ko, H.: Unsupervised real-world super resolution with cycle generative adversarial network and domain discriminator. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 456–457 (2020)
29. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
30. Köhler, T., Bätz, M., Naderi, F., Kaup, A., Maier, A., Riess, C.: 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)
31. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1833–1844 (2021)
32. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)

33. Liu, J., Zhang, W., Tang, Y., Tang, J., Wu, G.: Residual feature aggregation network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2359–2368 (2020)
34. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
35. Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L., Zeng, T.: Transformer for single image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 457–466 (2022)
36. Lugmayr, A., Danelljan, M., Timofte, R.: Unsupervised learning for real-world super-resolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3408–3416. IEEE (2019)
37. Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. Computer Vision and Image Understanding **158**, 1–16 (2017)
38. Maeda, S.: Unpaired image super-resolution using pseudo-supervision. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 291–300 (2020)
39. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings eighth IEEE international conference on computer vision. ICCV 2001. vol. 2, pp. 416–423. IEEE (2001)
40. Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. Multimedia tools and applications **76**, 21811–21838 (2017)
41. Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. In: Proceedings of the European conference on computer vision (ECCV). pp. 768–783 (2018)
42. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters **20**(3), 209–212 (2012)
43. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3883–3891 (2017)
44. Park, S., Yoo, J., Cho, D., Kim, J., Kim, T.H.: Fast adaptation to super-resolution networks via meta-learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16. pp. 754–769. Springer (2020)
45. Prajapati, K., Chudasama, V., Patel, H., Upla, K., Ramachandra, R., Raja, K., Busch, C.: Unsupervised single image super-resolution network (usisresnet) for real-world data using generative adversarial network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 464–465 (2020)
46. Shao, W.Z., Elad, M.: Simple, accurate, and robust nonparametric blind super-resolution. In: Image and Graphics: 8th International Conference, ICIG 2015, Tianjin, China, August 13–16, 2015, Proceedings, Part III. pp. 333–348. Springer (2015)
47. Shao, W.Z., Ge, Q., Wang, L.Q., Lin, Y.Z., Deng, H.S., Li, H.B.: Nonparametric blind super-resolution using adaptive heavy-tailed priors. Journal of Mathematical Imaging and Vision **61**, 885–917 (2019)
48. Shekhar Tripathi, A., Danelljan, M., Shukla, S., Timofte, R., Van Gool, L.: Transform your smartphone into a dslr camera: Learning the isp in the wild. In: European Conference on Computer Vision. pp. 625–641. Springer (2022)

49. Shocher, A., Cohen, N., Irani, M.: "zero-shot" super-resolution using deep internal learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3118–3126 (2018)
50. Soh, J.W., Cho, S., Cho, N.I.: Meta-transfer learning for zero-shot super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3516–3525 (2020)
51. Sun, L., Pan, J., Tang, J.: Shufflemixer: An efficient convnet for image super-resolution. Advances in Neural Information Processing Systems **35**, 17314–17326 (2022)
52. Sun, W., Gong, D., Shi, Q., van den Hengel, A., Zhang, Y.: Learning to zoom-in via learning to zoom-out: Real-world super-resolution by generating and adapting degradation. IEEE Transactions on Image Processing **30**, 2947–2962 (2021)
53. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 114–125 (2017)
54. Wang, H., Chen, X., Ni, B., Liu, Y., Liu, J.: Omni aggregation networks for lightweight image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22378–22387 (2023)
55. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2654–2662 (2023)
56. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: The European Conference on Computer Vision Workshops (ECCVW) (September 2018)
57. Wang, Z., Xu, K., Yang, Y., Dong, J., Gu, S., Xu, L., Fang, Y., Ma, K.: Measuring perceptual color differences of smartphone photographs. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(8), 10114–10128 (2023)
58. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component divide-and-conquer for real-world image super-resolution. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16. pp. 101–117. Springer (2020)
59. Xiao, J., Yong, H., Zhang, L.: Degradation model learning for real-world single image super-resolution. In: Proceedings of the Asian Conference on Computer Vision (2020)
60. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.H., Liao, Q.: Deep learning for single image super-resolution: A brief review. IEEE Transactions on Multimedia **21**(12), 3106–3121 (2019)
61. Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., Lin, L.: Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 701–710 (2018)
62. Zamfir, E., Wu, Z., Mehta, N., Zhang, Y., Timofte, R.: See more details: Efficient image super-resolution by experts mining. In: Forty-first International Conference on Machine Learning (2024)
63. Zeyde, R., Elad, M., Protter, M.: 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. pp. 711–730. Springer (2012)

64. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4791–4800 (2021)
65. Zhang, K., Li, D., Luo, W., Ren, W., Stenger, B., Liu, W., Li, H., Yang, M.H.: Benchmarking ultra-high-definition image super-resolution. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 14769–14778 (2021)
66. Zhang, X., Zeng, H., Guo, S., Zhang, L.: Efficient long-range attention network for image super-resolution. In: European conference on computer vision. pp. 649–667. Springer (2022)
67. Zhang, X., Chen, Q., Ng, R., Koltun, V.: Zoom to learn, learn to zoom. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3762–3770 (2019)
68. Zhang, Y., Liu, S., Dong, C., Zhang, X., Yuan, Y.: Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution. IEEE transactions on Image Processing **29**, 1101–1112 (2019)
69. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2472–2481 (2018)
70. Zhang, Z., Wang, H., Liu, M., Wang, R., Zhang, J., Zuo, W.: Learning raw-to-srgb mappings with inaccurately aligned supervision. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4348–4358 (2021)
71. Zhang, Z., Wang, R., Zhang, H., Chen, Y., Zuo, W.: Self-supervised learning for real-world super-resolution from dual zoomed observations. In: European Conference on Computer Vision. pp. 610–627. Springer (2022)
72. Zhao, C., Dewey, B.E., Pham, D.L., Calabresi, P.A., Reich, D.S., Prince, J.L.: Smore: a self-supervised anti-aliasing and super-resolution algorithm for mri using deep learning. IEEE transactions on medical imaging **40**(3), 805–817 (2020)
73. Zhou, R., Susstrunk, S.: Kernel modeling super-resolution on real low-resolution images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2433–2443 (2019)
74. Zhou, Y., Li, Z., Guo, C.L., Bai, S., Cheng, M.M., Hou, Q.: Srformer: Permuted self-attention for single image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12780–12791 (2023)