Article

Learning the Frequency Domain Aliasing for Real-World Super-Resolution

Yukun Hao and Feihong Yu *

College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China; yukunhao@zju.edu.cn
* Correspondence: feihong@zju.edu.cn

Abstract: Most real-world super-resolution methods require synthetic image pairs for training. However, the frequency domain gap between synthetic images and real-world images leads to artifacts and blurred reconstructions. This work points out that the main reason for the frequency domain gap is that aliasing exists in real-world images, but the degradation model used to generate synthetic images ignores the impact of aliasing on images. Therefore, a method is proposed in this work to assess aliasing in images undergoing unknown degradation by measuring the distance to their alias-free counterparts. Leveraging this assessment, a domain-translation framework is introduced to learn degradation from high-resolution to low-resolution images. The proposed framework employs a frequency-domain branch and loss function to generate synthetic images with aliasing features. Experiments validate that the proposed domain-translation framework enhances the visual quality and quantitative results compared to existing super-resolution models across diverse real-world image benchmarks. In summary, this work offers a practical solution to the real-world super-resolution problem by minimizing the frequency domain gap between synthetic and real-world images.

Keywords: super-resolution; real-world; domain-translation

1. Introduction

Single-image super-resolution (SR) aims to enhance the quality of low-resolution images by reconstructing them into high-resolution counterparts. Learning-based methods [1–10], such as SRCNN [1], VDSR [2], LapSRN [3], RCAN [4], SRGAN [5], and ESRGAN [6], have made significant advancements in achieving impressive results. Typically, these methods require pairs of high-resolution (HR) and low-resolution (LR) images for training. However, acquiring such HR–LR image pairs for real-world scenarios proves challenging. Even if obtained through methods like optical zoom [11–16], they fail to address issues stemming from differences in depth of field, illumination, and perspective [17]. These disparities impact the performance of end-to-end learning models and result in blurry SR outcomes [17]. Moreover, these methods are constrained by limited degradation diversity, making it challenging to apply them effectively to various scenarios with differing degradation distributions. As a result, their adoption remains limited, primarily due to issues related to misalignment and a lack of adaptability.

Considering that there are currently many widely used HR datasets [18–22], a practical approach for real-world SR involves creating LR images from HR images by using a degradation model. This process leads to the formation of HR–LR image pairs, which can serve as the training dataset. Some methods [1–6] use HR images and LR images generated by blurring and downsampling to train SR models. When real-world images are taken as input, SR results are disturbed by severe artifacts and noise. This is due to the mismatch between the degradation model and the real-world degradation process. Some recent methods extend the classical degradation model through random strategies, or recombine steps such as blur, downsampling, and noise [23–26]. These methods greatly expand

the range of degradation distributions covered and are able to deal with more complex degradation scenarios. However, at the cost of improved generalization capabilities, these methods lack detail-recovery capabilities and tend to produce over-smoothed results. Furthermore, these predefined, constrained degradation steps often prove inadequate for covering the complexity of degradation in real-world images.

Since the degradation processes in the real world are complex and diverse, domain-translation-based methods are proposed to adaptively learn degradation models for images with different degradation distributions [27]. The reference HR image is considered to be from the source domain, while the real-world LR image is considered to be from the target domain. The degradation model is considered to be the domain-translation process from the source domain to the target domain. Since the image contents in the reference HR and real-world LR are not directly consistent, the translation process from the source domain to the target domain is learned under an adversarial training framework [28] to generate synthetic LR images that conform to the degradation distribution of the target domain. Then, image pairs consisting of HR and synthetic LR images can be used to train the SR model in a supervised learning manner. KernelGAN [29] first generates LR images from HR using a deep linear generator. Bulat et al. [30] used a generator with an encoder–decoder structure for the domain-translation task. The CycleGAN [31] architecture for image-to-image translation is also used to learn domain translation from HR to LR images, as well as LR to HR images [32–34]. In this framework, the degradation process and the SR model are jointly trained, and cycle consistency loss ensures that the content of the image does not change during the degradation and SR processes. To facilitate integration with more advanced SR methods, more methods train the degradation and SR processes separately [17,30,35–40]. Zhou et al. [37] proposed a color-guided domain-mapping network to alleviate color shift in domain-translation processes. Luo et al. [38] proposed a probabilistic degradation model, which studies the degradation process by modeling it as a random variable and covers more diverse degradation distributions. Son et al. [40] proposed an adaptive data loss to allow the downsampler to adaptively learn the degradation process of real-world images. These learning-based methods can learn the degradation process of real-world images to generate LR images with smaller gaps from the real-world domain, thereby training SR models that can restore better image reconstruction quality.

However, these methods are all based on classical degradation models, focusing on establishing accurate spatial blur kernels and noise, without paying attention to the frequency-domain gap between synthetic LR images and real-world LR images. The frequency distributions between the synthetic LR image and the real-world LR image show obvious differences [27]. Ji et al. [39] focused on the frequency-domain features of synthetic images and used a frequency-consistent adaptation to construct images that were consistent with real-world images. However, when the image contents are inconsistent, it is inappropriate for the frequenc- consistent adaptation to directly compare their frequency spectral densities. Furthermore, this frequency domain consistency adaptation method does not point out the reason for the frequency-domain gap between synthetic images and real-world images, and adopts an indiscriminate comparison of densities at all frequencies.

This work points out that the main reason for the frequency-domain gap between synthetic LR images and real-world LR images is the aliasing phenomenon, which is widely present in real-world images [41–43]. However, existing degradation models [17,25,26,29,30,35–37] ignore the impact of aliasing on images [39,44]. Aliasing is an inherent property of sampling equipment [45–47]. Aliasing occurs when the spatial frequency of the scene surpasses the Nyquist limit defined by the imaging device's resolution [48,49]. Therefore, aliasing widely exists in images captured by acquisition devices such as cameras and mobile phones in real scenes [41–43]. Aliasing reduction [50–52] has always been a classical problem in the fields of signal processing and image processing. However, in the field of real-world image super resolution, few works [39] focus on the impact of aliasing on images.

Therefore, this paper proposes a method to evaluate aliasing in LR images that have undergone unknown degradation. The method involves generating an alias-free copy of

the LR image to be evaluated and then computing the *L*1 distance between this copy and the LR image. The greater the distance calculated, the more serious the aliasing. Compared with the previously discussed frequency-consistent adaptation method, this approach provides a more intuitive and accurate means of quantifying the degree of aliasing. The consistency of image contents between the copy and the LR image enhances the precision of the results.

On the basis of this measurement method, this work proposes a domain-translation-based method that generates frequency features similar to real-world images. On the one hand, a frequency-domain branch extracts the aliasing features in the degradation process. On the other hand, a loss function is proposed to guide the generation of synthetic images with aliasing features. Both the domain-translation generator and the frequency-domain loss function reduce the frequency-domain gap between synthetic LR images and real-world LR images, so the proposed method can help the SR model reconstruct more realistic image details.

The main contributions of this work are as follows:

- This work points out the frequency-domain gap between the synthetic and real-world image. A method is proposed to measure the degree of frequency-domain aliasing in images that undergo unknown degradation;
- A domain-translation framework is proposed to generate frequency-domain features that are similar to real-world images, including a branch to extract aliasing features and a loss function related to the degree of aliasing;
- The proposed domain-translation framework is proven to help the SR model achieve better reconstruction quality on real-world images.

2. Related Work

2.1. Image Pair-Based Methods

The challenge in real-world SR primarily stems from the shortage of real-world HR–LR image pairs in most scenarios [27]. To address this issue, various approaches have been proposed [11–16]. These methods involve capturing images of the same scene at different resolutions, effectively creating datasets containing HR and LR images derived directly from real-world sources. Some methods [11–14] involve using zoom lenses to acquire HR and LR images with long and short focal lengths, respectively. SR models trained with these datasets have the capacity to learn the degradation distribution present in real-world scenarios, resulting in improved performance on LR images characterized by the same degradation distribution. Nonetheless, challenges persist due to disparities in depth of field, lighting conditions, and perspective, leading to inevitable misalignment between HR and LR images [17]. These alignment discrepancies significantly impact the efficacy of SR models. Moreover, existing datasets often lack diversity in terms of scenes and degradation distributions, making it challenging to generalize to real-world images that exhibit varying degradation distributions.

To address complex and diverse real-world SR challenges, degradation models that are consistent with real-world degradation processes and have generalization capabilities are necessary. Alternatively, domain-translation methods for unpaired HR and LR images offer a viable solution.

2.2. Degradation Modeling-Based Methods

In many cases, only reference HR images and real-world LR images are available due to the absence of HR–LR image pairs. There is no direct content correspondence between reference HR and real-world LR images. To reconstruct SR images using reconstruction-based SR methods, it is necessary to determine the parameter set for the degradation model between the HR and LR images. Efrat et al. [53] proposed that an accurate degradation model is more important than using advanced image priors. Classical degradation models typically involve steps like blurring, downsampling, and noise. Most current research efforts [54–61] concentrate on the precise recovery of blur kernels. Several approaches [57–61]

optimize degradation parameters and intermediate super-resolved images through iterative prediction and correction techniques. However, when compared to the vast and diverse degradation space in the real world, these methods still exhibit limitations in terms of diversity and face the risk of failure in realistic scenarios.

For learning-based SR methods, it is crucial to determine the degradation model that transforms reference HR images into their corresponding LR counterparts, in order to train the SR model. One intuitive approach [62] involves synthesizing various LR images with multiple degradation distributions to train the SR model. This strategy aims to enhance the model's generalization performance. Several methods seek to expand upon the classical degradation model by introducing new degradation frameworks [23–26]. For instance, Zhou et al. [23] employed Generative Adversarial Network [28] (GAN) to construct a blur kernel pool and used these kernels in conjunction with HR images to generate LR images. Ji et al. [24] proposed RealSR, a degradation framework that employed blind kernel estimation to construct blur kernels and noise pools. The degradation process involves random sampling of blur kernels and noise from these pools to degrade HR images. Zhang et al. [25] used a random shuffled strategy to extend the classic degradation model and proposed a practical degradation model BSRGAN. Wang et al. [26] proposed Real-ESRGAN, a high-order degradation model that simulates ringing artifacts through a *sinc* filter.

Nevertheless, in their pursuit of improved generalization, these methods often produce over-smoothed results and require extensive datasets and training resources.

2.3. Domain-Translation-Based Methods

Consider the reference HR image and the real-world LR image as residing in different domains: HR images in the source domain and LR images in the target domain. In this context, both the degradation process and the SR process can be treated as domain-translation challenges. One approach is to leverage the well-established CycleGAN architecture [31], typically used in style transfer. This strategy [32–34] involves the joint training of the degradation and SR stages. Cycle-consistency loss enforces content correspondence within the cycle, while style consistency between different domains is often assessed through adversarial loss. These methods possess the ability to adaptively learn the degradation and SR processes for images with varying distributions. These methods can achieve more accurate recovery of image details compared to degradation models based on generalization ability. However, these methods exhibit limited flexibility, making it challenging to integrate with advanced SR methods. Some methods [17,35] leverage the cycle structure solely within the degradation process, opting to train the degradation and SR processes separately. But, unlike the style transfer problem, which usually assumes a one-to-one correspondence between elements in domains, SR poses an ill-posed challenge. In the field of SR, a single LR image corresponds to countless HR images. The deterministic degradation relationship provided by the cycle structure struggles to account for the random variables in real-world degradation.

Another approach [29,30,36–38] is to directly employ GAN [28] for unsupervised learning of domain translation within the degradation process, resulting in the generation of HR–LR image pairs. The subsequent SR process is executed by deep convolution neural network (CNN) based SR methods through supervised learning. In contrast to the explicit modeling of degradation found in degradation modeling-based methods, this strategy constitutes implicit modeling of degradation. Fritsche et al. [36] proposed DSGAN to learn the domain-translation process from HR images to unpaired LR images in an unsupervised manner. DSGAN trains two models for different datasets, including SDSR [36] and TDSR [36]. Zhou et al. [37] proposed a color-guided domain-mapping network to alleviate color shift during domain translation, named CARB [37]. Luo et al. [38] proposed a probabilistic degradation model (PDM), which studies the degradation process as a random variable and models the mapping from a prior random variable to the degradation process. Son et al. [40] simulated the distribution of LR images in the target domain by generalizing

low-frequency loss and designed an adaptive data loss (ADL). These methods can cover a wider range of unknown degradation distributions and is more robust when dealing with degradations that lack definition or are affected by random factors. Compared with methods based on the CycleGAN [31] framework, this strategy of separately modeling degradation can be easily integrated with advanced SR methods. With no cycle-consistency constraint, these methods rely on content consistency measured through the distance between low-frequency components of HR images and their corresponding LR images. Style consistency is still represented by adversarial loss.

However, the domain gap exists in the frequency domain between synthetic LR images and real-world LR images, which is overlooked by the existing methods [39]. Real-world images exhibit frequency-domain aliasing phenomena, while synthetic images tend to reduce aliasing features due to the low-pass filtering [40]. When applied to real-world images, SR models trained on such synthetic images often mistake aliasing phenomena as image content and amplify these features [44].

3. Method

3.1. Classical Degradation Model

Generally, LR image I_{LR} can be obtained from HR image I_{HR} by the classical degradation model [63,64]:

$$I_{LR} = (I_{HR} \otimes k) \downarrow_s + n \tag{1}$$

where k denotes a degradation kernel, $\otimes$ is a spatial convolution, $\downarrow_s$ indicates downsampling and decimation by a scale factor s, and n is a noise term. The order of these steps is not fixed; the classical degradation model can also be expressed as

$$I_{LR} = (I_{HR} \downarrow_s) \otimes k + n \tag{2}$$

The classical degradation model is widely used to synthesize HR–LR image pairs by degradation modeling-based methods. Furthermore, the classical degradation model is also combined with the probabilistic model in domain-translation methods [38].

3.2. Frequency-Domain Aliasing

There is a frequency-domain gap between real-world LR images and LR images synthesized by Equation (1) or Equation (2). Image downsampling process $\downarrow_s$ can be understood as sampling and decimation, which causes aliasing. Figure 1 shows the impact of several degradation processes on signals in the frequency domain. Figure 1a represents direct downsampling. Figure 1b and Figure 1c represent the degradation processes of Equation (1) and Equation (2), respectively. The influence of noise is not considered. Among them, ω_M is the highest inherent frequency of the image I_{HR}, ω_S is the sampling frequency, and ω_K is the cutoff frequency of the low-pass filter k. Compared with direct downsampling, the degradation processes represented by Equations (1) and (2) limit the frequency of I_{LR} within low-pass windows, reducing aliasing. Assume that I_{HR} generates I_{LR} through an unknown degradation process $\mathcal{D}^{\times s}(\cdot)$, and the scale factor between I_{HR} and I_{LR} is s, which can be expressed as

$$I_{LR} = \mathcal{D}^{\times s}(I_{HR}) \tag{3}$$

Figure 2 selects three sets of image samples and shows the frequency-domain images of I_{HR} and I_{LR} after the above-mentioned degradations. The direct downsampling is denoted as $\mathcal{D}_B^{\times s}(\cdot)$. And the degradation process in Equations (1) and (2) are expressed as $\mathcal{D}_1^{\times s}(\cdot)$ and $\mathcal{D}_2^{\times s}(\cdot)$, respectively ($k$ using Gaussian low-pass filter). The samples come from the real-world dataset City100 [11], so an LR image that has experienced the real-world degradation process $\mathcal{D}_R^{\times s}(\cdot)$ could be used as a reference. The first row of each group is the spatial-domain image. The second row of each group is the corresponding frequency-domain image. Consistent results with the previous analysis can be obtained from the samples. Compared with $\mathcal{D}_B^{\times s}(\cdot)$, $\mathcal{D}_1^{\times s}(\cdot)$ and $\mathcal{D}_2^{\times s}(\cdot)$ limit the frequency-domain

information to low-pass windows. Real-world LR images $\mathcal{D}_R^{\times s}(\boldsymbol{I}_{HR})$ exhibit frequency-domain characteristics that are more like $\mathcal{D}_B^{\times s}(\boldsymbol{I}_{HR})$, rather than $\mathcal{D}_1^{\times s}(\boldsymbol{I}_{HR})$ or $\mathcal{D}_2^{\times s}(\boldsymbol{I}_{HR})$. In the spatial domain, $\mathcal{D}_R^{\times s}(\boldsymbol{I}_{HR})$ is not as smooth as $\mathcal{D}_1^{\times s}(\boldsymbol{I}_{HR})$ and $\mathcal{D}_1^{\times s}(\boldsymbol{I}_{HR})$, but have jagged edges and background stripes, similar to $\mathcal{D}_B^{\times s}(\boldsymbol{I}_{HR})$. These phenomena indicate that the classical degradation model cannot simulate the aliasing phenomenon existing in real-world images. The step-by-step degradation models do not accurately represent complex degradation processes in the real world.
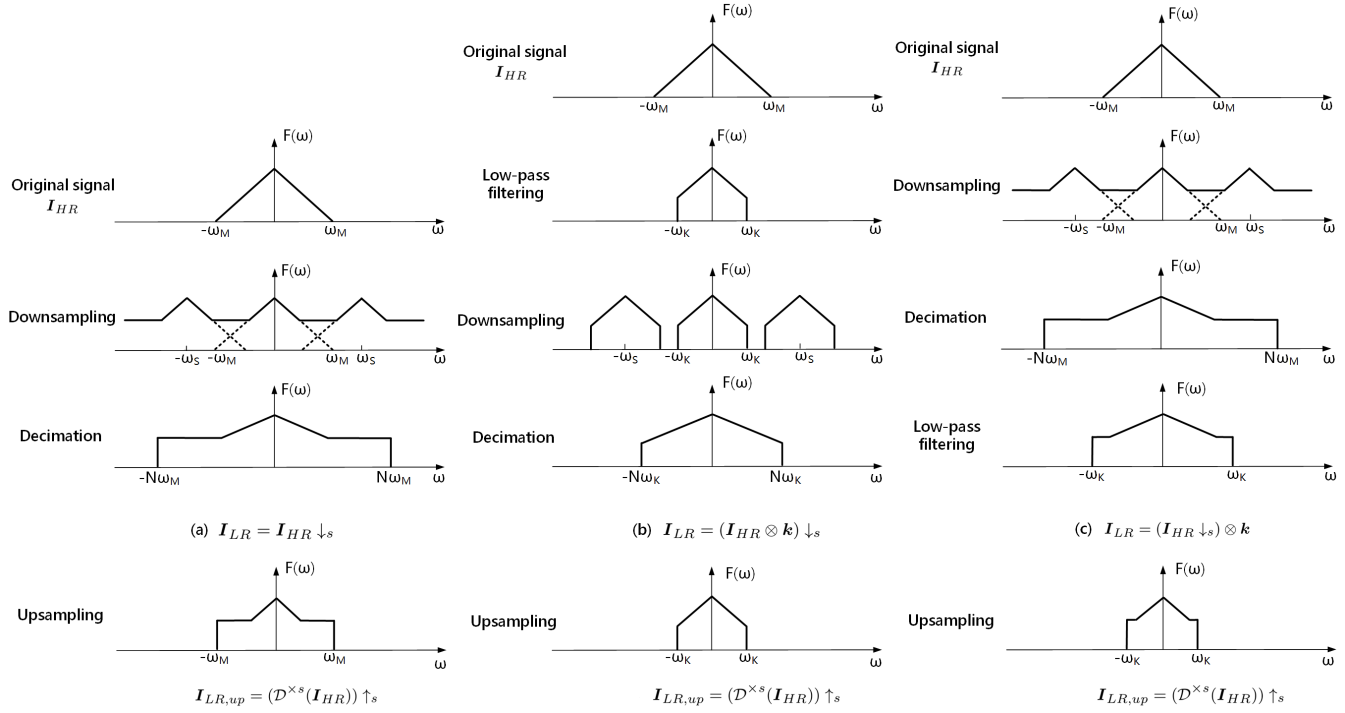
Figure 1. Overview of the impact of several degradation processes on signals, including (**a**) direct downsampling, (**b**) the degradation process represented by Equation (1), (**c**) the degradation process represented by Equation (2). The last row is the results of LR image $\boldsymbol{I}_{LR}$ being upsampled again, recorded as $\boldsymbol{I}_{LR,up}$.

In order to more accurately describe the aliasing features of these LR images generated through different degradation processes, this work proposes a method to evaluate aliasing in LR images. Interpolate $\boldsymbol{I}_{LR}$ in Equation (3) to the size of $\boldsymbol{I}_{HR}$, expressed as

$$\boldsymbol{I}_{LR,up} = (\mathcal{D}^{\times s}(\boldsymbol{I}_{HR})) \uparrow_s \tag{4}$$

where $\uparrow_s$ indicates upsampling by a scale factor s. In order to measure the degree of aliasing caused by the degradation process $\mathcal{D}^{\times s}(\cdot)$ on $\boldsymbol{I}_{HR}$, assume that there is an alias-free copy of $\boldsymbol{I}_{LR}$, which is the solution to the following optimization problem:

$$\hat{\boldsymbol{I}}_F = \underset{\boldsymbol{I}_F}{\arg\min} \|\boldsymbol{I}_F - \boldsymbol{I}_{LR,up}\| + \lambda \mathcal{H}(\boldsymbol{I}_F) \tag{5}$$

where the first item represents that the contents of $\boldsymbol{I}_F$ and $\boldsymbol{I}_{LR,up}$ are consistent. $\mathcal{H}(\cdot)$ is a constraint that ensures that there is no aliasing in $\boldsymbol{I}_F$. λ is the regularization constant. As shown in the last row of Figure 1, $\boldsymbol{I}_F$ can be obtained from $\boldsymbol{I}_{HR}$ through a low-pass filter $\boldsymbol{k}$ in the absence of aliasing. Therefore, $\mathcal{H}(\cdot)$ can be defined as

$$\mathcal{H}(\boldsymbol{I}_F) = \mathcal{C}(\boldsymbol{I}_F, \boldsymbol{k}) = \|\boldsymbol{I}_{HR} \otimes \boldsymbol{k} - \boldsymbol{I}_F\| + \mu \mathcal{G}(\boldsymbol{k}) \tag{6}$$

where $\mathcal{C}(\cdot)$ is the non-aliasing constraint on I_F that introduces variable k, and $\mathcal{G}(\cdot)$ is the prior constraint of the filter k, usually using regularization term related to gradient. μ is the regularization constant. Equation (5) is rewritten as

$$(\hat{I}_F, \hat{k}) = \operatorname*{argmin}_{I_F, k} \|I_F - I_{LR,up}\| + \lambda \|I_{HR} \otimes k - I_F\| + \mu \mathcal{G}(k) \tag{7}$$

When a signal is aliased, its high-frequency components additively spill into the low-frequency components [65]. Treat aliasing as an additive noise [66–68] and denote it as η; then,

$$I_{LR,up} = I_F + \eta \tag{8}$$

Then, k can be obtained by solving

$$\hat{k} = \operatorname*{argmin}_{k} \|I_{HR} \otimes k - I_{LR,up} + \eta\| + \mu \mathcal{G}(k) \tag{9}$$

During the downsampling process, aliasing mostly occurs near areas containing high-frequency information, that is, areas where gradient changes are dramatic. η has less impact on areas where gradient changes are relatively gentle. Divide the image $I_{LR,up}$ and I_{HR} into N sample blocks. The i-th sample is denoted as $I_{LR,up}^i$ and I_{HR}^i, and x^i is the center pixel of this window. The metric to measure the usefulness of gradients in each block [69] is defined as

$$r(x^i) = \frac{\|\sum_{y \in I_{HR}^i} \nabla I_{HR}(y)\|}{\sum_{y \in I_{HR}^i} \|\nabla I_{HR}(y)\| + \beta} \tag{10}$$

where ∇ is the operator of first-order spatial derivatives and β is a constant coefficient. A small r implies that a flat region is involved, which causes neutralization of many gradient components. A large r implies existing strong image structures in the local region [69]. Next, define a weight w that is negatively related to r:

$$w(x^i) = \exp(-\|r(x^i)\|^\alpha) \tag{11}$$

where α is a constant coefficient. Use w as the weight to find the average blur kernel of N blocks. Since η of the flat area is very small and has different directions, it will be ignored by the averaging process. The weight w of such a block is set large. The η of blocks containing high-frequency areas will interfere with the solution of the blur kernel as noise. Thus, the weight w of such blocks is set small. The average blur kernel can be approximated as

$$\overline{k} = \operatorname*{argmin}_{k} \sum_{i=1}^{N} w(x^i) \|I_{HR}^i \otimes k - I_{LR,up}^i\| + \mu \mathcal{G}(k) \tag{12}$$

In the case of $\lambda > 1$, $\hat{I}_F$ can be obtained from Equation (7) as

$$\hat{I}_F = I_{HR} \otimes \overline{k} \tag{13}$$

η can be obtained from Equation (8) as

$$\eta = I_{LR,up} - \hat{I}_F \tag{14}$$

This work proposes to use the L1 distance $\|\eta\|_1$ to quantify the degree of aliasing of I_{LR} obtained by I_{HR} through unknown degradation $\mathcal{D}^{\times s}(\cdot)$. Figure 2 shows the visualization of the result I_{LR} and η produced by different degradation processes. There are three sets of samples. The first row of each set is I_{HR} and I_{LR} generated by the corresponding degradation processes. The second row is the corresponding images in the frequency domain, and the third row is the visualized images of η calculated by the proposed method. The corresponding L1 distance $\|\eta\|_1$ is marked below the visualization of η. The visualization

and the L1 distance of η both show that the degree of aliasing caused by the bicubic downsampling degradation $\mathcal{D}_B^{\times s}(\cdot)$ is similar to the real-world degradation $\mathcal{D}_R^{\times s}(\cdot)$. In contrast, the currently commonly used degradation models $\mathcal{D}_1^{\times s}(\cdot)$ and $\mathcal{D}_2^{\times s}(\cdot)$ have obvious gaps with the real-world degradation model $\mathcal{D}_R^{\times s}(\cdot)$.
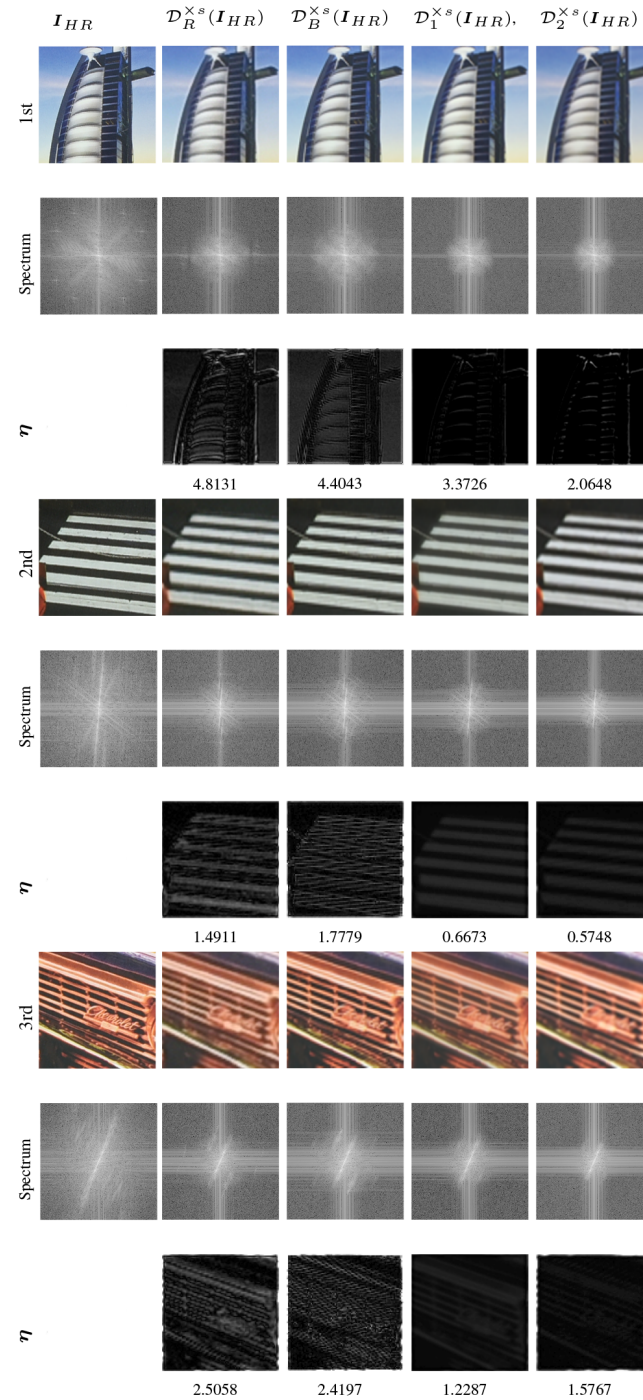

Figure 2. Visualization of the result I_{LR} and η produced by different degradation processes. The degradation processes include real-world degradation process $\mathcal{D}_R^{\times s}(\cdot)$, bicubic downsampling $\mathcal{D}_B^{\times s}(\cdot)$, and the degradation processes of Equations (1) and (2), denoted as $\mathcal{D}_1^{\times s}(\cdot)$ and $\mathcal{D}_2^{\times s}(\cdot)$, respectively ($k$ uses a Gaussian blur kernel). Three sets of samples are presented. The first row of each set shows I_{HR} and I_{LR}. The second row shows the corresponding frequency domain image. The third row has the visualized images of η calculated by the proposed method. The corresponding L1 distance $\|\eta\|_1$ is marked below η.

3.3. Downsampling with Domain Translation

To reduce the frequency-domain gap between synthetic and real-world images, this work proposes a domain-translation-based downsampling framework to generate aliasing features through the frequency-domain branch. The framework is based on the probabilistic degradation model [38]. Since the step-by-step degradation process cannot simulate real-world aliasing, this paper proposes to separate the aliasing information from the reference image through the frequency-domain branch, and then compensate the image before downsampling to generate LR images with aliasing. As shown in Figure 3, the degradation process of the framework involves three steps:

$$\begin{cases} \boldsymbol{I}_F = \boldsymbol{I}_{HR} \otimes \boldsymbol{k} \\ \boldsymbol{I}_C = \text{netD}(\text{netE}_1(\boldsymbol{I}_F) + \text{netE}_2(\boldsymbol{\eta})) \downarrow_s \\ \boldsymbol{I}_{LR} = \boldsymbol{I}_C + \boldsymbol{n} \end{cases} \tag{15}$$

where $\boldsymbol{k}$ is the blur kernel generated by the Kernel Generator, $\boldsymbol{n}$ is the spatial-domain noise generated by the Noise Generator, netE_1 and netE_2 are two encoders, and netD is a decoder. $\boldsymbol{\eta}$ is the aliasing feature produced by the frequency-domain branch. Compensate $\boldsymbol{I}_F$ with $\boldsymbol{\eta}$ before downsampling.
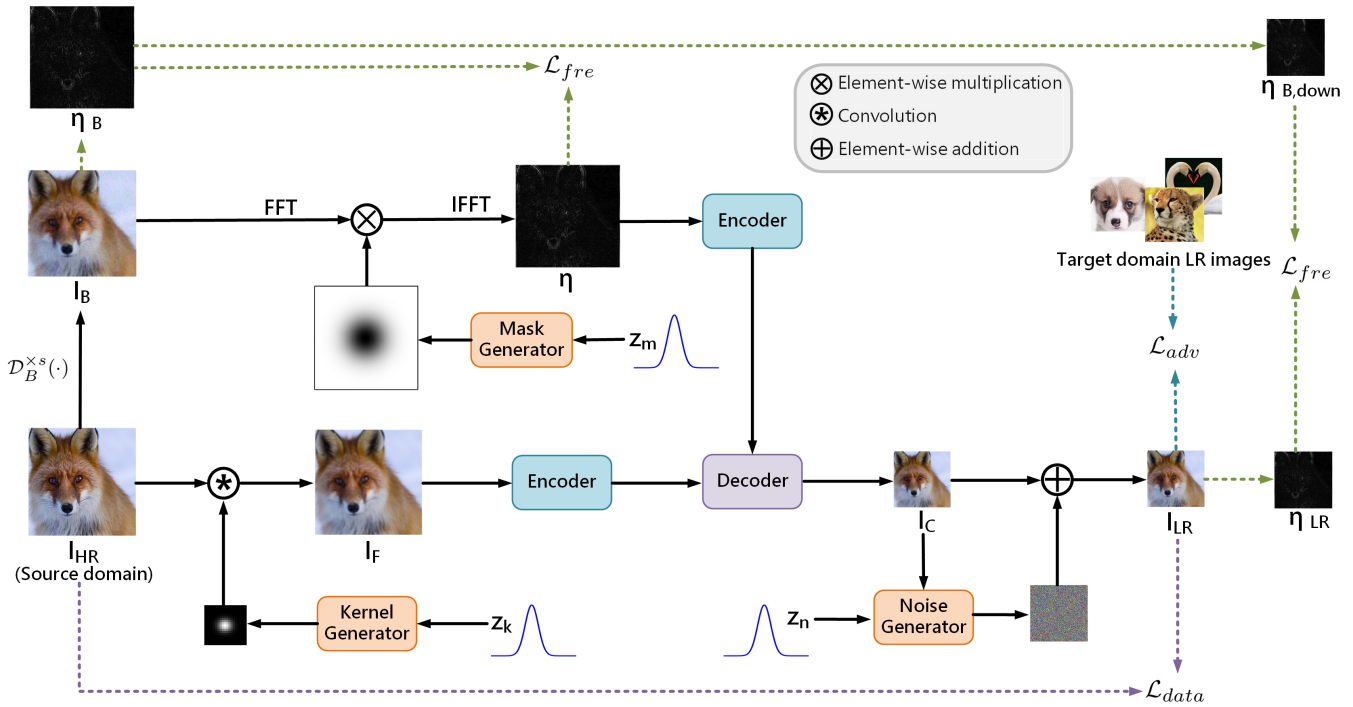
Figure 3. Overview of the proposed domain-translation-based downsampling framework.

The main function of the frequency-domain branch is to generate the aliasing feature $\boldsymbol{\eta}$. Compared with the method of obtaining $\boldsymbol{\eta}$ mentioned in Section 3.2, a more intuitive method is to directly use a frequency-domain filter $\boldsymbol{K}_H$ to separate this part of aliasing information from the aliased image:

$$\boldsymbol{\eta} = \text{IFFT}\{\boldsymbol{K}_H * \text{FFT}\{(\mathcal{D}_B^{\times s}(\boldsymbol{I}_{HR})) \uparrow_s\}\} \tag{16}$$

where the image undergoes the degradation process $\mathcal{D}_B^{\times s}(\cdot)$ and is used as a reference image to provide aliasing information.

To model the distribution of the blur kernel $\boldsymbol{K}_H$, a prior random variable $\boldsymbol{z}_m$ is defined, and a generator netM is used to learn the mapping from $\boldsymbol{z}_m$ to the frequency domain filter:

$$\boldsymbol{K}_H = \text{netM}(\boldsymbol{z}_m) \qquad \boldsymbol{z}_m \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{1}) \tag{17}$$

If the height, width, and number of channels of $\boldsymbol{I}_{HR}$ are h, w, and c, then

$$\boldsymbol{z}_m \in \mathbb{R}^{f_m \times h \times w} \qquad \boldsymbol{K}_H \in \mathbb{R}^{h \times w \times c} \tag{18}$$

where f_m is the dimension of the normal distribution $\boldsymbol{z}_m$.

Finally, the aliasing information $\boldsymbol{\eta}$ is integrated into $\boldsymbol{I}_F$ for compensation before downsampling. In this way, compared to classical degradation processes, $\boldsymbol{I}_{LR}$ generated by this proposed framework contains aliasing information and has a smaller domain gap compared to real-world images.

3.4. Frequency-Domain Loss

In this work, a frequency-domain loss is proposed to guide the domain-translation network to generate frequency-domain features that are consistent with real-world images. In order to guide the frequency-domain branch to generate $\boldsymbol{\eta}$ that conforms to the aliasing distribution, $\boldsymbol{\eta}_B$ that corresponds to $\mathcal{D}_B^{\times s}(\cdot)$ degradation is used as prior knowledge. The optimization of netM is formulated as

$$\underset{\text{netM}}{\text{argmin}} \, \| \, \text{IFFT}\{\text{netM}(\boldsymbol{z}_m) * \text{FFT}\{(\mathcal{D}_B^{\times s}(\boldsymbol{I}_{HR})) \uparrow_s\}\} - \gamma \boldsymbol{\eta}_B \|_1 \tag{19}$$

where γ is a constant coefficient. Equation (19) can be simply written as

$$\underset{\boldsymbol{\eta}}{\text{argmin}} \, \|\boldsymbol{\eta} - \gamma \boldsymbol{\eta}_B\|_1 \tag{20}$$

In order to ensure that the generated image $\boldsymbol{I}_{LR}$ contains aliasing features, the aliasing degree $\|\boldsymbol{\eta}_{LR}\|_1$ related to $\boldsymbol{I}_{LR}$ is measured through the method proposed in Section 3.2. $\boldsymbol{\eta}_{B,down}$ is a downsampled version of $\boldsymbol{\eta}_B$, which serves as a prior. In this way, the overall frequency-domain loss function is defined as

$$\mathcal{L}_{fre} = \|\boldsymbol{\eta} - \gamma \boldsymbol{\eta}_B\|_1 + \upsilon \, \| \, \|\boldsymbol{\eta}_{B,down}\|_1 - \tau \|\boldsymbol{\eta}_{LR}\|_1 \, \| \tag{21}$$

where υ and τ are weights for the regularizer term. The entire frequency-domain loss function consists of two parts. The first part uses $\boldsymbol{\eta}_B$ to guide the generator to generate $\boldsymbol{\eta}$ that conforms to the aliasing distribution. The second part uses the downsampled version $\boldsymbol{\eta}_{B,down}$ to constrain the final generated $\boldsymbol{I}_{LR}$ to satisfy the aliasing degree.

Since the degradation process represented by $\mathcal{D}_B^{\times s}(\cdot)$ is deterministic, $\overline{\boldsymbol{k}}_B$ of the processes $\mathcal{D}_B^{\times s}(\cdot)$ is independent of image and is only related to the downsampling factor s. As such, replace $\overline{\boldsymbol{k}}_B$ with fixed blur kernel calculated through a large number of samples in advance. This approach avoids recalculating $\overline{\boldsymbol{k}}_B$ every time the loss is calculated.

3.5. Overall Loss

Synthetic image $\boldsymbol{I}_{LR}$ needs to meet the following two requirements: the image content is consistent with $\boldsymbol{I}_{HR}$ and the image distribution is consistent with the target domain images. In order to preserve the image content across different scales, the data loss $\mathcal{L}_{data}$ is defined to measure the distance between the low-frequency components of $\boldsymbol{I}_{HR}$ and the synthetic image $\boldsymbol{I}_{LR}$:

$$\mathcal{L}_{data} = \|\mathcal{P}^{\times s}(\boldsymbol{I}_{HR}) - \mathcal{P}(\boldsymbol{I}_{LR})\|_1 \tag{22}$$

where $\mathcal{P}(\cdot)$ is a low-pass filtering and $\mathcal{P}^{\times s}(\cdot)$ is a combination of low-pass filtering and downsampling by s. The adversarial loss $\mathcal{L}_{adv}$ is used to enforce $\boldsymbol{I}_{LR}$ to follow the distribution of the target domain images:

$$\begin{cases} \mathcal{L}_{adv}^g = -\mathbb{E}[\log(\mathcal{F}(\boldsymbol{I}_{LR}))] \\ \mathcal{L}_{adv}^d = -\mathbb{E}[\log(\mathcal{F}(\boldsymbol{I}_{LR}^{real}))] + \mathbb{E}[\log(1 - \mathcal{F}(\boldsymbol{I}_{LR}))] \end{cases} \tag{23}$$

where $\mathcal{L}_{adv}^{g}$ and $\mathcal{L}_{adv}^{d}$ are the adversarial losses of the generator and discriminator, respectively. $\mathcal{F}(\cdot)$ is the discriminator and $\boldsymbol{I}_{LR}^{real}$ is the real-world image of the target domain.

The total loss function of the degradation model is

$$\begin{cases} \mathcal{L}^{g} = \mathcal{L}_{data} + \mathcal{L}_{adv}^{g} + \mathcal{L}_{fre} \\ \mathcal{L}^{d} = \mathcal{L}_{adv}^{d} \end{cases} \tag{24}$$

4. Experiments

4.1. Datasets and Training Details

Datasets. In order to test the effect of the proposed method on real-world images, 400 images were selected from the HR dataset DIV2K [18] as the source domain HR images. Real-world LR datasets or datasets with unknown and complex degradation were selected as target domain images, including DRealSR [12] and NTIRE2020 [70] Track1 and Track2. The source domain HR image and the target domain LR image were unpaired, which means that their contents were inconsistent. A portion of the target domain LR images were selected for training, and another portion of the target domain images were chosen for testing and validation. The proposed domain-translation-based downsampling method was used to learn the translation from the source domain to the target domain, and the loss function proposed in Equation (24) was used.

Evaluation metrics. For validation sets with ground-truth references, such as NTIRE2020 Track1, we used PSNR, SSIM [71], and LPIPS [72] as evaluation metrics. For the validation set without ground-truth references, we used NIQE [73], NRQM [74], and PI [75] as evaluation metrics.

Implementation and training details. During training, z_m was set as $f_m = 128$. The HR images were cropped into 128×128, and the LR images were cropped into 64×64 for scale factor 2 and 32×32 for scale factor 4. In comparative experiments and ablation experiments, the network was trained for 40 epochs with a learning rate of 1×10^{-4}. The batch size was set as 24.

4.2. Comparison with Other Domain-Translation Based Methods

The proposed method is compared with several state-of-the-art real-world SR methods based on domain translation, including SDSR [36], TDSR [36], CARB [37], PDM [38], and ADL [40]. For fair comparison, the SR model used by the proposed method is consistent with the SR model used by these methods, including EDSR [76] and ESRGAN [6]. EDSR is PSNR-oriented and supervised by L1 and L2 losses. ESRGAN is perceptual-oriented and supervised by perceptual loss. Therefore, when EDSR is the SR model, PSNR and SSIM are used as the main evaluation metrics. When ESRGAN is the SR model, LPIPS (based on perceptual similarity) is used as the main evaluation metric.

The results on NTIRE2020 Track 1 [70] are presented in Table 1. When combined with the PSNR-oriented network EDSR, the proposed method in this work attains the best performance in terms of PSNR and SSIM. Similarly, when combined with the perceptual-oriented network ESRGAN, it achieves the best LPIPS results, indicating superior perceptual quality. This underscores that, regardless of being PSNR-oriented or focused on perceptual quality, incorporating the method in this work leads to the best results. The LR images in NTIRE2020 Track 1 are obtained by an undisclosed degradation operator on the HR dataset Flickr2k [19]. This degradation operator [70] generates structured artifacts produced by the kind of image-processing pipelines found on very low-end devices. Figure 4 provides a visual comparison of three sets of samples. The SR model in the first row of each group of samples is EDSR; both LR images are amplified by bicubic interpolation, and ground-truth HR images are used as references. The SR model in the second row is ESRGAN. In the results with EDSR as the SR model, the method proposed in this work achieved the sharpest edge while maintaining the denoising ability. As a perception-oriented method, ESRGAN obviously has stronger detail-recovery capabilities when used as the SR model, but it also bears the risk of amplifying complex noise. In each set of samples, "SDSR + ESRGAN" and

"TDSR + ESRGAN" are plagued by complex noise. "ADL + ESRGAN" over-enlarges some details that do not exist in the ground-truth HR image. "CARB + ESRGAN" is too blurry to take advantage of the perception-oriented methods. "Ours + ESRGAN" is cleaner than other methods, indicating that the proposed method can better model complex noise in the spatial domain and frequency domain.
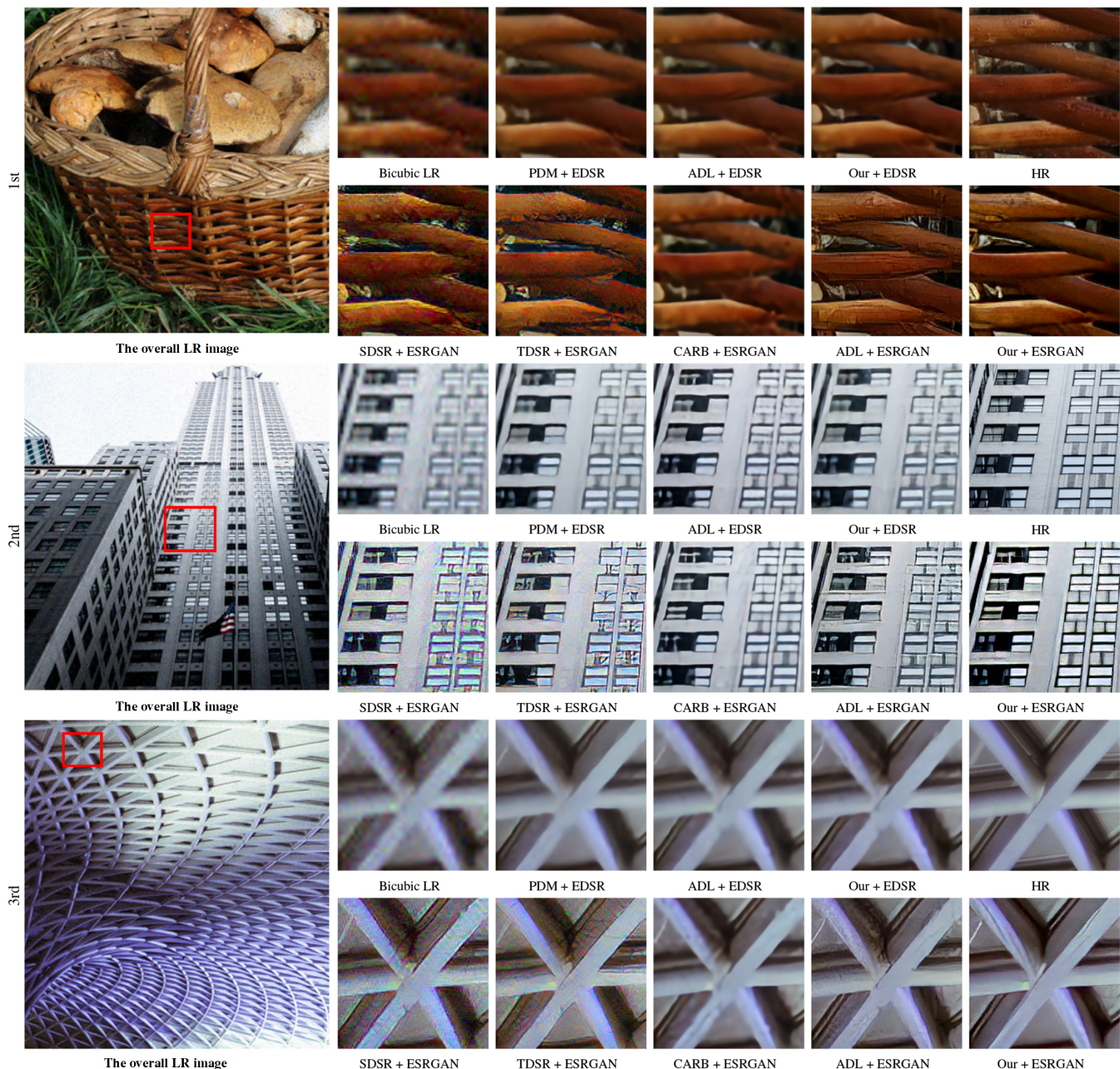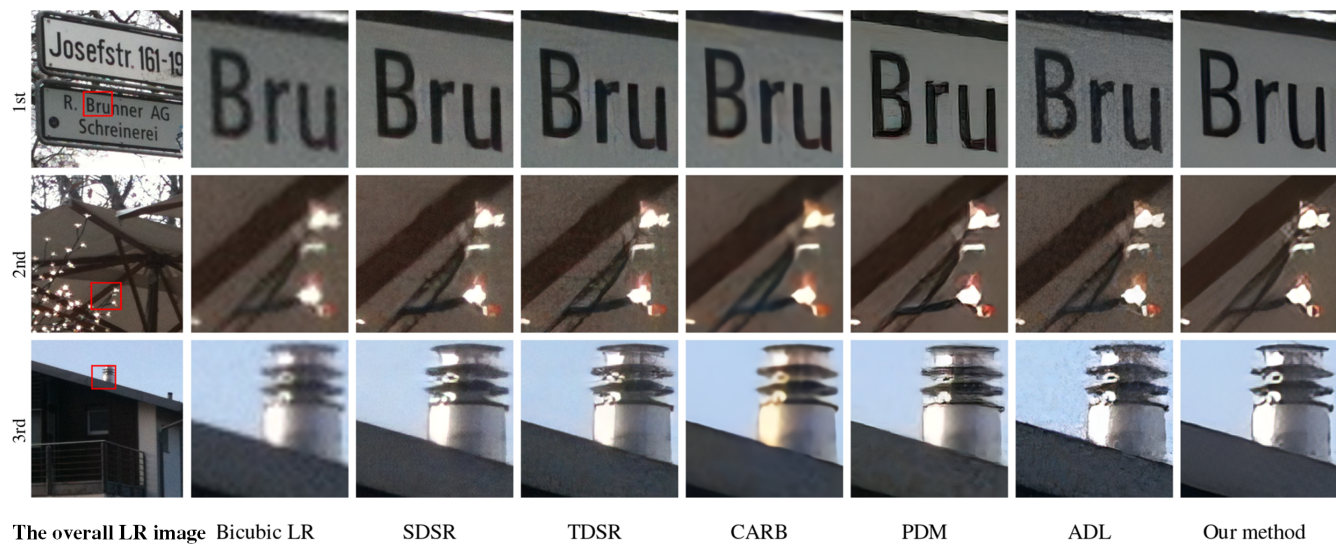

Figure 4. Comparison of our method with state-of-the-art real-world SR methods based on domain translation. Test images from dataset NTIRE2020 Track 1. Inside the red box is the enlarged area.

Table 2 shows the results of these methods on NTIRE2020 Track 2 [70]. The method proposed in this work still achieves the best results in NRQM and PI . The LR images in NTIRE2020 Track 2 are real-world images taken by an iPhone3 and come from DPED [22]. NTIRE2020 Track 2 contains more unknown and complex real-world degradation processes. Since the methods used for comparison all provide the perceptual-oriented SR model ESRGAN, the method proposed in this work also uses ESRGAN as the SR model. Figure 5 shows the visual comparison of three sets of samples. LR images amplified by bicubic

interpolation are used as references. The results of "SDSR", "TDSR", and "ADL" are all affected by severe noise, while the results of "CARB" are still blurry. "PDM" has a better ability to remove noise in the spatial domain, but there is still some complex noise related to the image structure, as shown in the first sample. The method proposed in this work obtains the cleanest results, and there is no structural noise like in the results of "PDM".

Figure 5. Comparison of our method with state-of-the-art real-world SR methods based on domain translation. Test images from dataset NTIRE Track 2. Inside the red box is the enlarged area.

Table 1. Quantitative comparison with domain-translation-based methods on NTIRE2020 Track 1 with ground-truth references. ↑ denotes the larger the better. ↓ denotes the smaller the better. The best results are denoted in red, and the second best are denoted in blue.

Methods	PSNR ↑	SSIM↑	LPIPS↓
PDM + EDSR	21.099	0.6044	0.3794
ADL + EDSR	28.942	0.8004	0.3248
SDSR + ESRGAN	23.096	0.4479	0.5619
TDSR + ESRGAN	21.949	0.3901	0.6024
CARB + ESRGAN	28.483	0.7968	0.3285
ADL + ESRGAN	24.688	0.6437	0.3063
Ours + EDSR	29.366	0.8033	0.3107
Ours + ESRGAN	25.661	0.7636	0.2930

Table 2. Quantitative comparison with domain-translation-based methods on NTIRE2020 Track 2 without ground-truth references. ↑ denotes the larger the better. ↓ denotes the smaller the better. The best results are denoted in red, and the second best are denoted in blue.

| Methods	NIQE↓	NRQM↑	PI↓
SDSR + ESRGAN	6.744	4.630	6.057
TDSR + ESRGAN	4.365	4.985	4.690
CARB + ESRGAN	8.459	2.256	8.101
PDM + ESRGAN	6.714	4.231	6.241
ADL + ESRGAN	5.229	3.352	5.938
Ours + ESRGAN	4.423	5.158	4.632

4.3. Comparison with Other Degradation Modeling-Based Methods

In order to have an intuitive comparison of the effects of different methods, comparisons with several methods based on degradation modeling are also provided, including

RealSR [24], BSRGAN [25], and Real-ESRGAN [26]. These methods are all based on pre-defined degradation steps and have strong generalization capabilities for real-world LR images. Since these methods are perceptual-oriented and use ESRGAN or improved ESRGAN [26] as SR models, the method proposed in this work is also combined with improved ESRGAN [26] for comparison. The real-world dataset DRealSR [12] is selected as the validation set. The LR images of DRealSR are derived from the real world, and have ground-truth images as references. PSNR, SSIM, and LPIPS are evaluation metrics. Due to the perceptual quality orientation, the results of LPIPS are more important.

Table 3 shows the results of theses methods on DRealSR. The method proposed in this work achieved the best LPIPS and SSIM results. Figure 6 shows the visual comparison of three samples. LR images amplified by bicubic interpolation and ground-truth HR images are used as references. The results show that, at the expense of generalization ability, "BSRGAN" and "Real-ESRGAN" tend to produce over-smoothed results that ignore texture details, such as the shape of the petals in the first sample. Since the predefined degradation model may not match the real degradation model, there is a gap between the results of "RealSR" and the ground truth. The method proposed in this work can learn the specific degradation process, so it achieves the visual results that are the closest to ground truth.

| The overall LR image | Bicubic LR | RealSR | BSRGAN | Real-ESRGAN | Our method | HR |
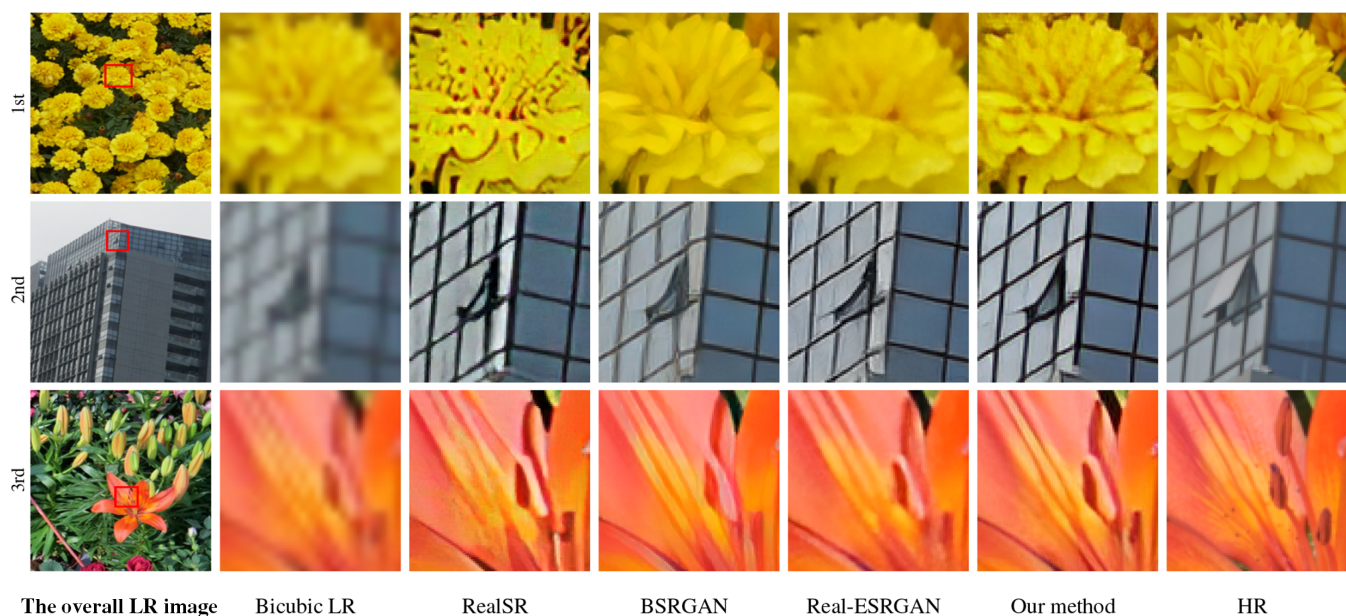
Figure 6. Comparison of our method with state-of-the-art real-world SR methods based on degra-dataion modeling. Test images from dataset DRealSR. Inside the red box is the enlarged area.

Table 3. Quantitative comparison with degradation modeling-based methods on DRealSR with ground-truth references. ↑ denotes the larger the better. ↓ denotes the smaller the better. The best results are denoted in red, and the second best are denoted in blue.

Methods	PSNR↑	SSIM ↑	LPIPS↓
RealSR	23.088	0.7122	0.2438
BSRGAN	28.147	0.8128	0.1824
Real-ESRGAN	26.656	0.8013	0.1875
Ours + ESRGAN	26.799	0.8188	0.1779

4.4. Ablation Studies

In order to verify the effect of the proposed frequency-domain branch and frequency-domain loss function, comparative experiments were designed. The probabilistic degradation model without frequency-domain branch and frequency-domain loss function is used

as a baseline. On this basis, the effects of the frequency-domain branch and the frequency-domain loss function are independently verified. The results are shown in Table 4. The effect of having frequency-domain branch or frequency-domain loss function is better than the baseline. When both are available, SR has the best effect.

Table 4. Ablation studies on frequency-domain branch and frequency-domain loss. The test set is NTIRE2020 Track 1 with ground-truth references. ↑ denotes the larger the better. ↓ denotes the smaller the better. The best results are denoted in red.

Frequency Branch	Frequency Loss	PSNR↑	SSIM ↑	LPIPS↓
		26.397	0.7335	0.3919
✓		26.719	0.7416	0.4005
	✓	28.842	0.7909	0.3242
✓	✓	29.366	0.8033	0.3107

In order to verify the role of the random variable, z_m is fixed to zero in the comparative experiment. Table 5 shows that, when z_m is randomly sampled, a better SR effect is achieved than when z_m is fixed to zero. The reason is speculated that z_m can simulate the influence of random factors and avoid overfitting.

Table 5. Ablation studies on the random variable z_m. The test set is NTIRE2020 Track 1 with ground-truth references. ↑ denotes the larger the better. ↓ denotes the smaller the better. The best results are denoted in red.

z_m	PSNR ↑	SSIM ↑	LPIPS↓
	28.909	0.7996	0.3456
✓	29.366	0.8033	0.3107

5. Conclusions

This work proposes a method to evaluate frequency-domain aliasing in images suffering from unknown degradation. This is the first time that the effect of aliasing on images has been considered in a degradation model. Furthermore, a domain-translation framework is proposed that aims to leverage unpaired HR and LR images to generate synthetic datasets for training. This framework combines frequency-domain branch and frequency-domain loss function to generate synthetic images with real-world aliasing characteristics, thereby reducing the domain gap between synthetic images and real-world images. Experimental results demonstrate that the proposed method could improve the visual effects and quantitative results of super-resolution models on real-world images. This illustrates the practicality of the method proposed in this work and marks the possibility of improving super-resolution technology for real-world scenarios. In the future, the proposed method will be extended to estimate degradation models with real-world noise. Another direction is to combine domain-translation methods with reconstruction-based methods, with the goal of accurately estimating the degradation model and reconstructing high-resolution images, even when there is only a single low-resolution image or there are real-world images from different sources.

Author Contributions: Methodology, Y.H.; software, Y.H.; validation, Y.H.; formal analysis, Y.H.; investigation, Y.H.; resources, F.Y.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, F.Y.; visualization, Y.H.; supervision, F.Y.; project administration, F.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All data generated or analyzed during this study are included in this manuscript, and the associated code required to generate the data will be published as soon as possible.

Conflicts of Interest: All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
2. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
3. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
4. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
5. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
6. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
7. Yu, C.; Hong, L.; Pan, T.; Li, Y.; Li, T. ESTUGAN: Enhanced Swin Transformer with U-Net Discriminator for Remote Sensing Image Super-Resolution. *Electronics* **2023**, *12*, 4235. [CrossRef]
8. Shao, G.; Sun, Q.; Gao, Y.; Zhu, Q.; Gao, F.; Zhang, J. Sub-Pixel Convolutional Neural Network for Image Super-Resolution Reconstruction. *Electronics* **2023**, *12*, 3572. [CrossRef]
9. Shi, Y.; Jiang, C.; Liu, C.; Li, W.; Wu, Z. A Super-Resolution Reconstruction Network of Space Target Images Based on Dual Regression and Deformable Convolutional Attention Mechanism. *Electronics* **2023**, *12*, 2995. [CrossRef]
10. Ye, S.; Zhao, S.; Hu, Y.; Xie, C. Single-Image Super-Resolution Challenges: A Brief Review. *Electronics* **2023**, *12*, 2975. [CrossRef]
11. 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, Long Beach, CA, USA, 16–20 June 2019; pp. 1652–1660.
12. 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 Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 101–117.
13. 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, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3086–3095.
14. 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, Long Beach, CA, USA, 16–20 June 2019; pp. 3762–3770.
15. 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 Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2944–2959. [CrossRef]
16. 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, Seattle, WA, USA, 14–19 June 2020; pp. 518–519.
17. 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 Trans. Image Process.* **2021**, *30*, 2947–2962. [CrossRef]
18. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 126–135.
19. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 114–125.
20. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 606–615.
21. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4401–4410.
22. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L. Dslr-quality photos on mobile devices with deep convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 3277–3285.
23. Zhou, R.; Susstrunk, S. Kernel modeling super-resolution on real low-resolution images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2433–2443.

24. 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, Seattle, WA, USA, 14–19 June 2020; pp. 466–467.

25. 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, Montreal, QC, Canada, 11–17 October 2021; pp. 4791–4800.

26. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 1905–1914.

27. Chen, H.; He, X.; Qing, L.; Wu, Y.; Ren, C.; Sheriff, R.E.; Zhu, C. Real-world single image super-resolution: A brief review. *Inf. Fusion* **2022**, *79*, 124–145. [CrossRef]

28. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets.   In Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2014), Montreal, QC, Canada, 8–13 December 2014.

29. Bell-Kligler, S.; Shocher, A.; Irani, M. Blind super-resolution kernel estimation using an internal-gan. In Proceedings of the 33rd Conference on Neural Information Processing Systems 32 (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.

30. 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), Munich, Germany, 8–14 September 2018; pp. 185–200.

31. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 2223–2232.

32. 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, Seattle, WA, USA, 14–19 June 2020; pp. 468–469.

33. 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, Seattle, WA, USA, 14–19 June 2020; pp. 456–457.

34. Wang, W.; Zhang, H.; Yuan, Z.; Wang, C. Unsupervised real-world super-resolution: A domain adaptation perspective. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 11–17 October 2021; pp. 4318–4327.

35. Lugmayr, A.; Danelljan, M.; Timofte, R. Unsupervised learning for real-world super-resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3408–3416.

36. Fritsche, M.; Gu, S.; Timofte, R. Frequency separation for real-world super-resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3599–3608.

37. Zhou, Y.; Deng, W.; Tong, T.; Gao, Q. Guided frequency separation network for real-world super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 428–429.

38. Luo, Z.; Huang, Y.; Li, S.; Wang, L.; Tan, T. Learning the degradation distribution for blind image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 6063–6072.

39. Ji, X.; Tao, G.; Cao, Y.; Tai, Y.; Lu, T.; Wang, C.; Li, J.; Huang, F. Frequency consistent adaptation for real world super resolution. *Aaai Conf. Artif. Intell.* **2021**, *35*, 1664–1672. [CrossRef]

40. Son, S.; Kim, J.; Lai, W.S.; Yang, M.H.; Lee, K.M. Toward real-world super-resolution via adaptive downsampling models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 8657–8670. [CrossRef] [PubMed]

41. Lee, W.; Son, S.; Lee, K.M. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 17725–17734.

42. Glotzbach, J.W.; Schafer, R.W.; Illgner, K. A method of color filter array interpolation with alias cancellation properties. In Proceedings of the 2001 International Conference on Image Processing (Cat. No. 01CH37205), Thessaloniki, Greece, 7–10 October 2001; IEEE: Piscataway, NJ, USA, 2001; Volume 1, pp. 141–144.

43. Forsey, A.; Gungor, S. Demosaicing images from colour cameras for digital image correlation. *Opt. Lasers Eng.* **2016**, *86*, 20–28. [CrossRef]

44. Hao, Y.; Yu, F. Super-Resolution Degradation Model: Converting High-Resolution Datasets to Optical Zoom Datasets. In *IEEE Transactions on Circuits and Systems for Video Technology*; IEEE: Piscataway, NJ, USA, 2023; pp. 6374–6389.

45. Lettington, A.H.; Hong, Q.H.; Tzimopoulou, S. Superresolution by spatial-frequency aliasing. In Proceedings of the Infrared Technology and Applications XXII, Orlando, FL, USA, 8–12 April 1996; SPIE: St Bellingham, WA, USA, 1996; Volume 2744, pp. 583–590.

46. Hoshino, H.; Okano, F.; Yuyama, I. A study on resolution and aliasing for multi-viewpoint image acquisition. *IEEE Trans. Circuits Syst. Video Technol.* **2000**, *10*, 366–375. [CrossRef]

47. Pusey, E.; Yoon, C.; Anselmo, M.L.; Lufkin, R.B. Aliasing artifacts in MR imaging. *Comput. Med Imaging Graph.* **1988**, *12*, 219–224. [CrossRef]

48. Schöberl, M.; Schnurrer, W.; Oberdörster, A.; Fössel, S.; Kaup, A. Dimensioning of optical birefringent anti-alias filters for digital cameras. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 4305–4308.

49. Muammar, H.; Dragotti, P.L. An investigation into aliasing in images recaptured from an LCD monitor using a digital camera. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 2242–2246.

50. Esqueda, F.; Bilbao, S.; Välimäki, V. Aliasing reduction in clipped signals. *IEEE Trans. Signal Process.* **2016**, *64*, 5255–5267. [CrossRef]

51. Bilinskis, I. *Digital Alias-Free Signal Processing*; John Wiley & Sons: Hoboken, NJ, USA, 2007.

52. Vandewalle, P.; Süsstrunk, S.; Vetterli, M. A frequency domain approach to registration of aliased images with application to super-resolution. *Eurasip J. Adv. Signal Process.* **2006**, *2006*, 1–14. [CrossRef]

53. Efrat, N.; Glasner, D.; Apartsin, A.; Nadler, B.; Levin, A. Accurate blur models vs. image priors in single image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2832–2839.

54. Luo, Z.; Huang, H.; Yu, L.; Li, Y.; Fan, H.; Liu, S. Deep constrained least squares for blind image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 17642–17652.

55. Tao, G.; Ji, X.; Wang, W.; Chen, S.; Lin, C.; Cao, Y.; Lu, T.; Luo, D.; Tai, Y. Spectrum-to-kernel translation for accurate blind image super-resolution. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22643–22654.

56. Wang, L.; Wang, Y.; Dong, X.; Xu, Q.; Yang, J.; An, W.; Guo, Y. Unsupervised degradation representation learning for blind super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 10581–10590.

57. 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, Long Beach, CA, USA, 16–20 June 2019; pp. 1604–1613.

58. Zhang, K.; Gool, L.V.; Timofte, R. Deep unfolding network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3217–3226.

59. Huang, Y.; Li, S.; Wang, L.; Tan, T. Unfolding the alternating optimization for blind super resolution. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 5632–5643.

60. Yue, Z.; Zhao, Q.; Xie, J.; Zhang, L.; Meng, D.; Wong, K.Y.K. Blind image super-resolution with elaborate degradation modeling on noise and kernel. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 2128–2138.

61. Hussein, S.A.; Tirer, T.; Giryes, R. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1428–1437.

62. Zhang, K.; Zuo, W.; Zhang, L. Learning a single convolutional super-resolution network for multiple degradations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3262–3271.

63. Elad, M.; Feuer, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.* **1997**, *6*, 1646–1658. [CrossRef]

64. Liu, C.; Sun, D. On Bayesian adaptive video super resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 346–360. [CrossRef]

65. Vasconcelos, C.; Larochelle, H.; Dumoulin, V.; Roux, N.L.; Goroshin, R. An effective anti-aliasing approach for residual networks. *arXiv* **2020**, arXiv:2011.10675.

66. Xiaoxi, W.; Yingjie, Y.; Jianbin, H. Aliasing fringe pattern denoising based on deep learning. In Proceedings of the AOPC 2021: Novel Technologies and Instruments for Astronomical Multi-Band Observations, Beijing, China, 23–25 July 2021; SPIE: St Bellingham, WA, USA, 2021; Volume 12069, pp. 178–183.

67. Vollmerhausen, R.H.; Driggers, R.G.; Wilson, D.L. Predicting range performance of sampled imagers by treating aliased signal as target-dependent noise. *JOSA A* **2008**, *25*, 2055–2065. [CrossRef]

68. Hennenfent, G.; Herrmann, F.J. Irregular sampling–from aliasing to noise. In Proceedings of the 69th EAGE Conference and Exhibition Incorporating SPE EUROPEC 2007, London, UK, 11–14 June 2007; p. cp–27.

69. Xu, L.; Jia, J. Two-phase kernel estimation for robust motion deblurring. In Proceedings of the Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; Proceedings, Part I 11; Springer: Berlin/Heidelberg, Germany, 2010; pp. 157–170.

70. Lugmayr, A.; Danelljan, M.; Timofte, R. Ntire 2020 challenge on real-world image super-resolution: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 494–495.

71. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

72. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 586–595

73. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [CrossRef]

74. Ma, C.; Yang, C.Y.; Yang, X.; Yang, M.H. Learning a no-reference quality metric for single-image super-resolution. *Comput. Vis. Image Underst.* **2017**, *158*, 1–16. [CrossRef]

75. 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, Munich, Germany, 8–14 September 2018.

76. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.