This WACV 2021 paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

DualSR: Zero-Shot Dual Learning for Real-World Super-Resolution

Mohammad Emad
Eindhoven University of Technology
Eindhoven, The Netherlands
m.emad@tue.nl

Maurice Peemen
Thermo Fisher Scientific
Eindhoven, The Netherlands
maurice.peemen@thermofisher.com

Henk Corporaal
Eindhoven University of Technology
Eindhoven, The Netherlands
h.corporaal@tue.nl

Abstract

*Advanced methods for single image super-resolution (SISR) based upon Deep learning have demonstrated a remarkable reconstruction performance on downscaled images. However, for real-world low-resolution images (*e.g. *images captured straight from the camera) they often generate blurry images and highlight unpleasant artifacts. The main reason is the training data that does not reflect the real-world super-resolution problem. They train the network using images downsampled with an ideal (usually bicubic) kernel. However, for real-world images the degradation process is more complex and can vary from image to image. This paper proposes a new dual-path architecture (DualSR) that learns an image-specific low-to-high resolution mapping using only patches of the input test image. For every image, a downsampler learns the degradation process using a generative adversarial network, and an upsampler learns to super-resolve that specific image. In the DualSR architecture, the upsampler and downsampler are trained simultaneously and they improve each other using cycle consistency losses. For better visual quality and eliminating undesired artifacts, the upsampler is constrained by a masked interpolation loss. On standard benchmarks with unknown degradation kernels, DualSR outperforms recent blind and non-blind super-resolution methods in term of SSIM and generates images with higher perceptual quality. On real-world LR images it generates visually pleasing and artifact-free results.*

Figure 1: Example of 2x SR applied to a real-world image from RealSR [5] dataset.

1. Introduction

The aim of Single Image Super-Resolution (SISR) is to upsample a low-resolution (LR) image and reconstruct the high-resolution (HR) details. Recently, these super-resolution (SR) methods have entered our daily life by aiding low end smartphone cameras [31, 23]. Furthermore the restoration of historical LR photos to clean HR results is performed by novel SISR methods. Even old movies are converted to high-definition video quality. Next to the me-

dia industry these SR techniques have important other applications in medical imaging [33, 28], remote sensing [12], microscopy [10], surveillance [22] and so on.

The introduction of Convolutional Neural Networks (CNNs) have revolutionized computer vision and image processing techniques such as super-resolution. Many recently introduced SR methods based upon deep-learning, *e.g*. [8, 18, 24, 19, 30, 37, 7], learn the complicated LR-HR upsampling relations on huge datasets. Compared to traditional earlier methods these provide a significantly improved HR result. However, these pre-trained DL methods often perform much worse on captured images straight from a camera. They are trained on clean, noise-free, synthetically generated LR images, while the degradation process for real-world LR images is different from the ideal conditions. To a large extent, this is due to the supervised training scheme that is not representative for the real-world problem. Additionally, each cameras differs in acquisition parameters such as the Point Spread Function (PSF) of the sensor. Even images captured by the same camera will differ because of different light conditions, depth of field, blur due to shaking and so on. These conditions make it intractable to train a single CNN that performs well on all different image degradation conditions.

Blind SR methods solve the super-resolution problem with less assumptions on the degradation process. Often these assume that the LR image is the result of a downsampled HR image, convolved with a blur kernel k, and added noise n:

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

Many blind SR methods estimate the degradation process before they perform a parameterized super-resolution operation. The state-of-the-art techniques [2, 11, 3] use deep learning to learn an image-specific downsampler (degradation model parameters) that is used by the upsampler to super-resolve the input LR image. However, estimating a proper downsampler from a single input image is complicated. Especially in the presence of noise or other acquisition artifacts these methods often fail to estimate good degradation parameters. A wrong degradation severely reduces the effectiveness of the upsampler, and reduces the SR performance.

With the true downsampler one can determine the upsampler more accurately. On the other hand, with the true upsampler, one can correctly estimate the downsampler. In other words, the upsampler and downsampler are the inverse of each other and improving one can also improve the other. This relation motivated us to simultaneously train both the upsampler and downsampler in a single pipeline. Inspired by recent unsupervised methods such as CycleGAN [38] and DualGAN [32], we introduce DualSR, a dual-path architecture for super-resolution on real-world LR images. In DualSR, the downsampler learns the patch distribution of the input image using a KernelGAN based kernel estimation [2]. However, unlike KernelGAN, the upsampler and downsampler are trained simultaneously and improve each other using cycle consistency losses. The upsampler is constrained by a novel masked interpolation loss that gives better visual quality and eliminates undesired artifacts in the output result. On every new input image, the networks are trained from scratch using only patches of the new input image. We evaluate our method on existing synthesized and real-world benchmarks that shows how DualSR outperforms recent SR methods. Figure 1 compares the output of different methods on a real-world LR image.

In summery, the contributions of this work are threefold:

- Inspired by [38, 32], we propose a dual-path architecture optimized for blind super-resolution that can be trained in reasonable time using only the input LR image.
- We introduce a new masked interpolation loss that substantially reduces oversharpening and suppresses unwanted artifacts in the output image.
- We evaluate our method on existing synthetically generated and real-world benchmarks and we compare to the state-of-the-art blind and non-blind SR methods.

2. Related work

Super-resolution on LR images with an unknown degradation process is not completely new. Before the advent of deep learning, several learning-based methods [21, 29, 13, 14] have been introduced to address this problem. Very recently, some deep learning based methods for blind and non-blind SR have been proposed. Non-blind SR assumes the degradation process is known beforehand. For example, ZSSR [26] trains an image-specific CNN by using the recurrence of small patches across different scales in a single image. SRMD [36] uses dimensionality stretching that enables a convolutional SR network to take degradation parameters (*i.e*. blur kernel and noise level) as input. A grid search strategy is used to find the best configuration of these parameters. USRNet [35] is another non-blind SR method that tries to alternatively solve a data subproblem and a prior subproblem under the MAP (maximum a posteriori) framework.

On the other hand, Blind methods try to estimate the degradation process before upsampling. IKC [11] proposes an iterative correction scheme for blur kernel estimation and uses Spatial Feature Transform (SFT) layers to handle multiple blur kernels. KernelGAN [2] estimates the image-specific blur kernel by internal learning of a Generative Adversarial Network (GAN) using only LR test image. The

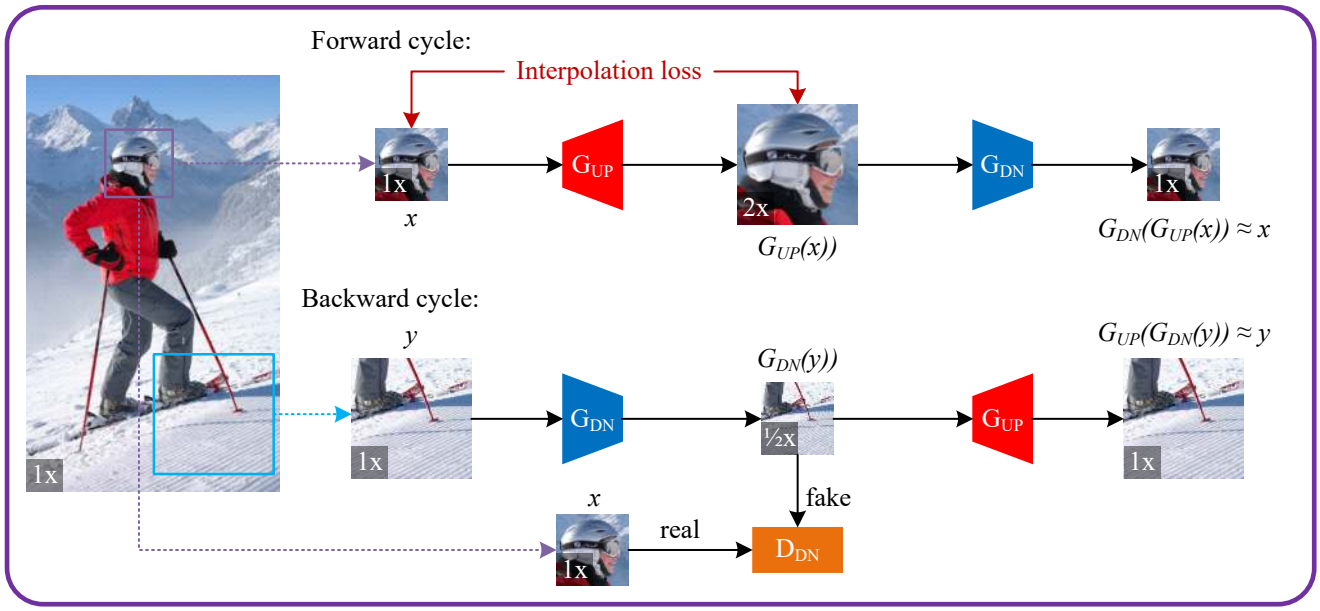

Figure 2: The network architecture of the proposed DualSR. G_{UP} is the upsampler, G_{DN} is the downsampler and D_{DN} is the discriminator. The top dataflow represents the forward cycle where we apply the upsampler before downsampler and the bottom part represents the backward cycle where upsampler is applied after downsampler.

estimated kernel can be used by non-blind methods such as ZSSR and USRNet to obtain super-resolved output image. Cornillère *et al*. [6] propose BlindSR that estimates the degradation setting by analysing the artifacts generated in the super-resolved HR output. They train a kernel discriminator that predicts errors present due to wrong kernel estimation and then they recover the correct kernel by minimizing the discriminator error. Very recently, based upon the generalized sampling theory, Hussein *et al*. [16] proposes a closed-form derivation of their correction filter to transform the degraded LR image such that it matches the bicubic downsampling result. Then, the modified LR image (similar to bicubic downsampling) is upsampled using existing state-of-the-art DNNs trained on bicubically downsampled images. In case of isotropic degradations, the transformation to bicubic gives acceptable results. However, for more complicated (non-isotropic) degradations the results are not reported.

Two similar works, CycleGAN [38] and DualGAN [32] introduce an effective architecture for the image-to-image translation where paired images are not available. They connect the main translator (generator) G:X $\rightarrow$ Y with its inverse translator F:Y $\rightarrow$ X. In addition, to reduce the solution space, they introduce cycle consistency loss that encourages F(G(x)) $\approx$ x and G(F(y)) $\approx$ y. Recently, new works have proposed similar structures for blind SR. Bulat *et al*. [3] propose a two-stage pipeline with High-to-Low and Low-to-High GANs. A cycle-in-cycle network structure is introduced in [34]. This model first maps the noisy and blurry input to cleaned-up LR space, next a pre-trained SR network is used to upsample the clean LR image. GAN-CIRCLE [33] uses a similar structure as CycleGAN, however the application is super-resolution for Computed Tomography (CT) images. All these CycleGAN-based methods need unpaired LR and HR datasets for training. As we explained, even for images from a single camera, the degradation process may be different and collecting these unpaired datasets is not always possible. This is in contrast to our method that does not require any other data than a test image.

3. Proposed method

We propose DualSR, a new blind SR method, that super-resolves real-world LR images with different acquisition processes (different downsampling kernels). It is a dual-path pipeline inspired by CycleGAN and DualGAN that can be trained in reasonable time using only patches of the LR input image. Figure 2 illustrates the proposed DualSR architecture. In this figure, G_{UP} is the upsampler that trains to upsample the specific LR image. G_{DN} is the downsampler that learns the degradation process. It aims to downsample the input LR image such that, at patch level, the distribution of downsampled image is as close as possible to the input image itself. It is shown in [2] that this internal distribution of patches can be learned using an internal GAN [25] trained on patches of the input image ($\mathcal{L}_{GAN}$ in equation 2). D_{DN} is the discriminator that learns to distinguish between real (patches of the input LR image) and G_{DN} generated output patches.

Ideally, the upsampler and downsampler should be the inverse of each other. This implies that $G_{DN}(G_{UP}(x)) = x$ and $G_{UP}(G_{DN}(y)) = y$ are valid. These conditions are demonstrated in figure 2 as forward and backward cycles and we refer to their loss functions as forward and backward cycle consistency losses ($\mathcal{L}_{cycle}$). In the forward cycle, we first apply the upsampler to generate a 2x upsampled image. Then the downsampler is applied and converts the upsampled image back to 1x. Similarly, in backward cycle, at first a ½x downsampled version of the input is generated by G_{DN}, and then G_{UP} upsamples the image back to the original scale. These two cycle consistency losses ensure that G_{UP} and G_{DN} can revert the operation done by the other.

In addition to the cycle consistency losses, the upsampler is constrained by a novel masked interpolation loss ($\mathcal{L}_{interp}$) that is applied between the input patch x and its 2x upsampled image $G_{UP}(x)$. This loss function encourages the upsampler to preserve the color composition of its input and eliminates unpleasant artifacts without making the output image blurry. The full loss function for training the upsampler and downsampler is:

$$\mathcal{L}_{total} = \mathcal{L}_{GAN} + \lambda_{cycle}\mathcal{L}_{cycle} + \lambda_{interp}\mathcal{L}_{interp} \quad (2)$$

Where λ values control the importance of the different loss function terms.

3.1. Adversarial loss

Accurate degradation model estimation is essential for blind SR. The importance of an accurate estimate of the blur kernel is significantly larger than that of a sophisticated prior (pre-training on an external dataset) [9]. To find the blur kernel, we use the idea of KernelGAN and embed it in our dual-path architecture. It estimates the image-specific degradation kernel using a Generative Adversarial Network (GAN) which tries to preserve the distribution of patches across scales of the LR image. The generator G_{DN} is a model of the degradation process and downsamples the input patch y such that, the output $G_{DN}(y)$ is indistinguishable by the discriminator from small input patches. We define the adversarial loss for the generator as:

$$\mathcal{L}_{GAN} = \mathbb{E}_y[D_{DN}(G_{DN}(y)) - 1]^2 + \mathcal{R} \quad (3)$$

Where the regularization term $\mathcal{R}$ applies realistic explicit priors on the estimated kernel (explained in [2]). On the other hand, the discriminator tries to distinguish fake images generated by G_{DN} from real patches of input LR image and its objective is:

$$\mathcal{L}_D = \mathbb{E}_x[(D_{DN}(x) - 1)^2] + \mathbb{E}_y[D_{DN}(G_{DN}(y))^2] \quad (4)$$

We also experimented with an adversarial loss for the upsampled image $G_{UP}(x)$ by adding a D_{UP} discriminator. However, without example HR images, it was not helpful and resulted in undesired artifacts in the output image.

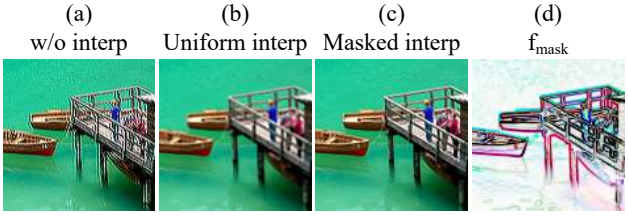

Figure 3: Effect of interpolation loss on the output. (a) SR result without interpolation loss. (b) SR result with uniform interpolation loss: $\|G_{UP}(x) - Bicubic(x)\|_1$. (c) SR result with masked interpolation loss represented in equation 7. (d) Frequency mask generate by equation 6.

3.2. Cycle consistency loss

Both forward and backward cycle consistency losses play an essential role in training of DualSR. The forward cycle facilitates the training convergence of the downsampler G_{DN} and also forces the upsampler G_{UP} to generate images invertible by G_{DN}. In the backward cycle, the upsampler learns to reconstruct the input LR image from its downsampled version generated by G_{DN}. Learning LR-HR relations from the downsampled version of the input image was firstly introduced by ZSSR [26], but in DualSR, it is implicitly part of the backward cycle loss. The final cycle consistency loss is the sum of the forward and backward losses:

$$\begin{aligned}\mathcal{L}_{cycle} = \mathbb{E}_x\|G_{DN}(G_{UP}(x)) - x\|_1 \\ + \mathbb{E}_y\|G_{UP}(G_{DN}(y)) - y\|_1\end{aligned} \quad (5)$$

3.3. Masked interpolation loss

Because there is no direct supervision for training G_{UP}, ringing effects often occur around sharp edges of the output image. In addition, unwanted artifacts show up especially in the low-frequency areas of the output image. This problem is even more severe when G_{DN} is not representing an accurate estimate of degradation model. Figure 3(a) shows the output of DualSR without interpolation loss when the estimated degradation kernel slightly differs from the ground-truth kernel. To eliminate these artifacts, we introduce a weighted cost function that minimizes the difference between a bicubically upsampled image and the output of G_{UP}.

It is well-known that bicubic interpolation correctly upsamples low-frequency areas, however it does not reconstruct high-frequency details. Hence, applying a uniform interpolation cost to all pixels generates an artifact-free but blurry result (see figure 3(b)). To avoid blurriness, we only apply an interpolation cost to low-frequency parts of the image. For this purpose, we generate a frequency mask (f_{mask}) by applying Sobel operator to the bicubic-

upsampled image:

$$f_{mask} = 1 - Sobel(Bicubic(x)) \quad (6)$$

This mask has higher values for pixels in low-frequency areas and lower values for pixels in high-frequency areas of the image (see figure 3(d)). We define masked interpolation loss as:

$$\mathcal{L}_{interp} = \mathbb{E}_x \| [G_{UP}(x) - Bicubic(x)] \times f_{mask} \|_1 \quad (7)$$

It encourages G_{UP} to follow bicubic interpolation in only low-frequency areas of the image. As it is shown in figure 3(c), it generates images with sharp edges without adding artifacts.

4. Implementation

Network configuration Unlike CycleGAN, our DualSR generators (G_{UP} and G_{DN}) have different network architectures. For a single image the LR-HR conversions can be performed by small networks unlike the huge networks that train on large datasets. For the upsampler, similar to ZSSR, a simple 8-layer fully convolutional network with ReLU activations is employed. There is a global residual connection between the input and output [18, 26]. We upscale the LR image to the output size before feeding it into the network. The network architecture from KernelGAN is used for downsampling and the corresponding discriminator. The generator used for downsampling is a deep linear network (without any activations) and the discriminator is a fully convolutional PatchGAN [17] with a receptive field of 7x7. The small receptive field enforces to use only local features (*e.g.* edges) of the LR image, instead of relying on high-level global features. Hence, the generator G_{DN} learns the kernel that generates images with a patch distribution similar to the input LR image.

Training details We train all networks G_{UP}, G_{DN} and D_{DN} from scratch for every input image. In each iteration, we train generators and discriminator successively, each time with a batch of two patches from the LR input. The patches are 64x64 and 128x128 (patches x and y in figure 2). Since SR kernels are not always symmetric, no geometric transformation is applied during training or test time. We tuned hyper parameters in equation 2 with a grid search strategy. We changed λ_{cycle} from 0 to 7.5 and λ_{interp} from 0 to 4 and calculated the average PSNR for the first 10 images in DIV2KRK [2] dataset. The results are demonstrated in figure 4. It shows that the best performance happens for $\lambda_{cycle} = 5$ and $\lambda_{interp} = 2$. We train the networks for 3000 iterations with ADAM optimizer. The initial learning rate is 0.001 for G_{UP} and 0.0002 for G_{DN} and D_{DN} and both are decreased every 750 iterations. The final super-resolved image is obtained by running the trained upsampler on the LR input image.

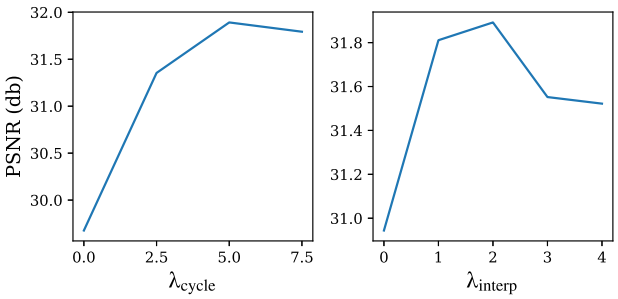

Figure 4: Performance analysis of DualSR with different values for λ_{cycle} and λ_{interp}. For the left plot, we set $\lambda_{interp} = 2$ and change λ_{cycle} from 0 to 7.5. For the right plot, we set $\lambda_{cycle} = 5$ and change λ_{interp} from 0 to 4. PSNR values are calculated on the first 10 images in DIV2KRK [2] dataset.

Run-time Since we train DualSR on fixed-size patches cropped from the input image, the training time is almost independent of image size. The average training+inference time for our method is 233 seconds on an RTX 2080 Ti GPU. For the combination of KernelGAN [2] + ZSSR [26] the run-time is 281 seconds and for BlindSR [6] it is 370 seconds. Supervised deep learning SR methods like SAN [7] have a very long training time and image size significantly influences their inference time. For SAN+, it takes 298 seconds, on average, to super-resolve each image of DIV2KRK benchmark [2].

5. Experiments and results

DualSR is designed to super-resolve real-world LR images. However, these images have no ground-truth, so these do not enable a quantitative evaluation. Hence, we use synthetically generated LR images with unknown degradation settings (from DIV2KRK [2], Urban100 [15] and NTIRE2017 [27] benchmarks) to compare DualSR against state-of-the-art methods. In addition, we experiment on real-world LR images (from RealSR [5] dataset) and compare the results qualitatively. In the end, we perform an ablation study to investigate the influence of each loss term in the overall loss function.

5.1. Evaluation on synthesized images

In this section, we evaluate DualSR on three benchmarks that simulate real-world LR images. The first one is DIV2KRK [2] benchmark which uses DIV2K [1] validation set (100 high-quality images) as HR ground-truth. It constructs the LR images by convolving an 11x11 anisotropic Gaussian blur kernel to the image before downsampling. Kernels vary for every image and have different shape and orientation. A uniform multiplicative noise is also applied to each kernel. We use the same degradation approach to generate LR images from Urban100 [15] dataset. As the

Category	Method	DIV2KRK [2]	Urban100 [15]	NTIRE2017 [27]
1st (Bicubically trained)	Bicubic Interpolation	28.73 / 0.8040	23.32 / 0.6859	27.72 / 0.7689
	Bicubic kernel + ZSSR [26]	29.10 / 0.8215	23.78 / 0.7143	27.80 / 0.7726
	EDSR+ [20]	29.17 / 0.8216	23.01 / 0.6857	27.78 / 0.7720
	SAN+ [7]	29.21 / 0.8232	23.81 / 0.7153	27.78 / 0.7721
2nd (Blind methods)	KernelGAN [2] + ZSSR	30.36 / 0.8669	24.66 / 0.7706	27.53 / 0.7572
	KernelGAN + ZSSR (masked interp)	30.40 / 0.8595	24.71 / 0.7692	28.04 / 0.7808
	KernelGAN + USRNet [35]	27.94 / 0.8084	21.81 / 0.6827	23.84 / 0.6662
	BlindSR [6]	31.36 / 0.8720	25.18 / 0.7742	28.35 / 0.7931
	DualSR (ours)	30.92 / 0.8728	25.04 / 0.7803	28.82 / 0.8045
3rd (Oracle kernel)	GT kernel + ZSSR	32.44 / 0.8955	26.38 / 0.8252	-
	GT kernel + USRNet	32.23 / 0.8981	24.89 / 0.7821	-
	GT kernel + DualSR	32.72 / 0.9030	26.04 / 0.8122	-

Table 1: Quantitative results (PSNR / SSIM) for 2x SR on different datasets. For comparison to other works, the PSNR and SSIM values of the Y channel are reported. The best performance numbers are highlighted in red and the second best numbers are highlighted in blue.

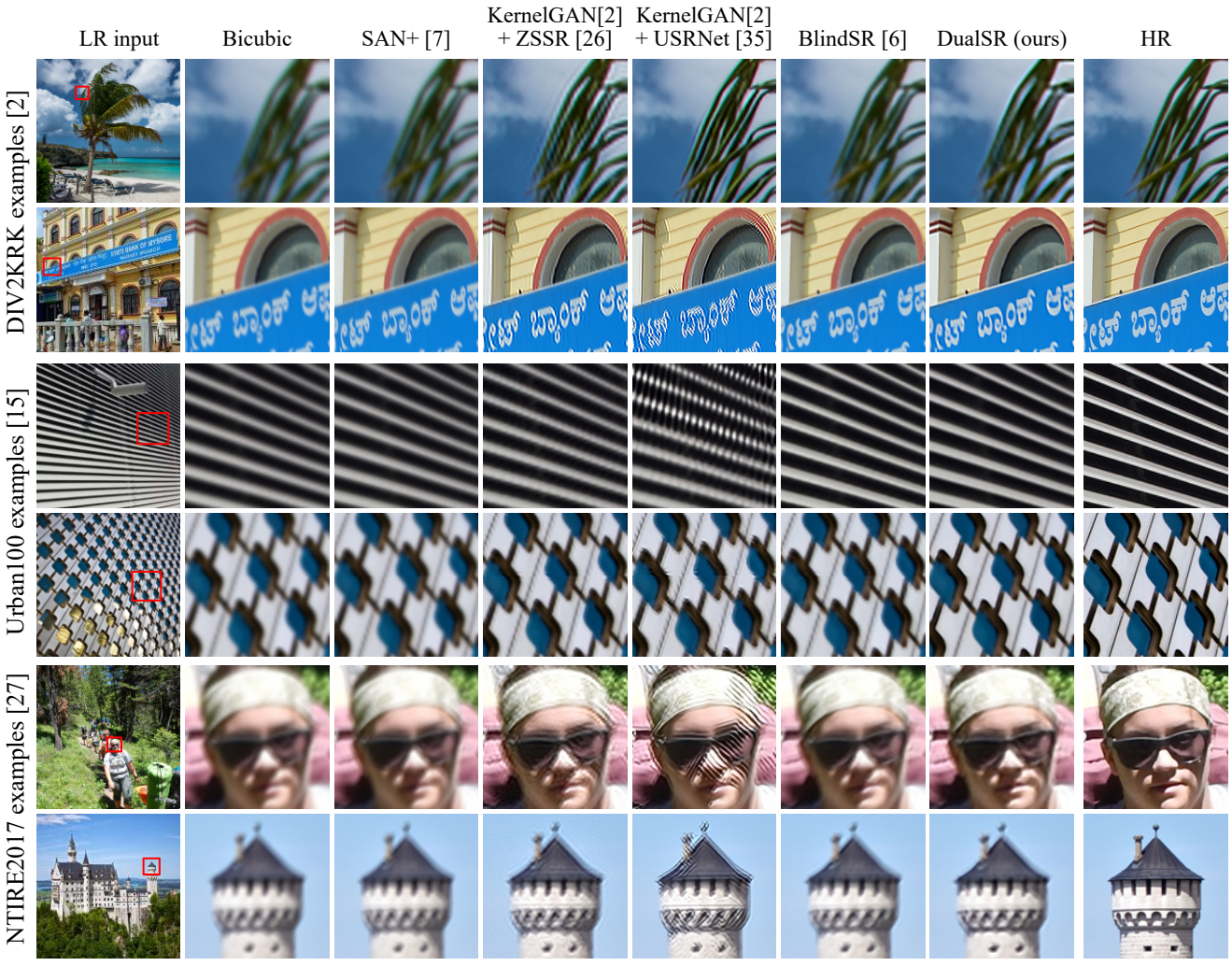

Figure 5: Qualitative comparison for 2x SR on synthetically generated datasets.

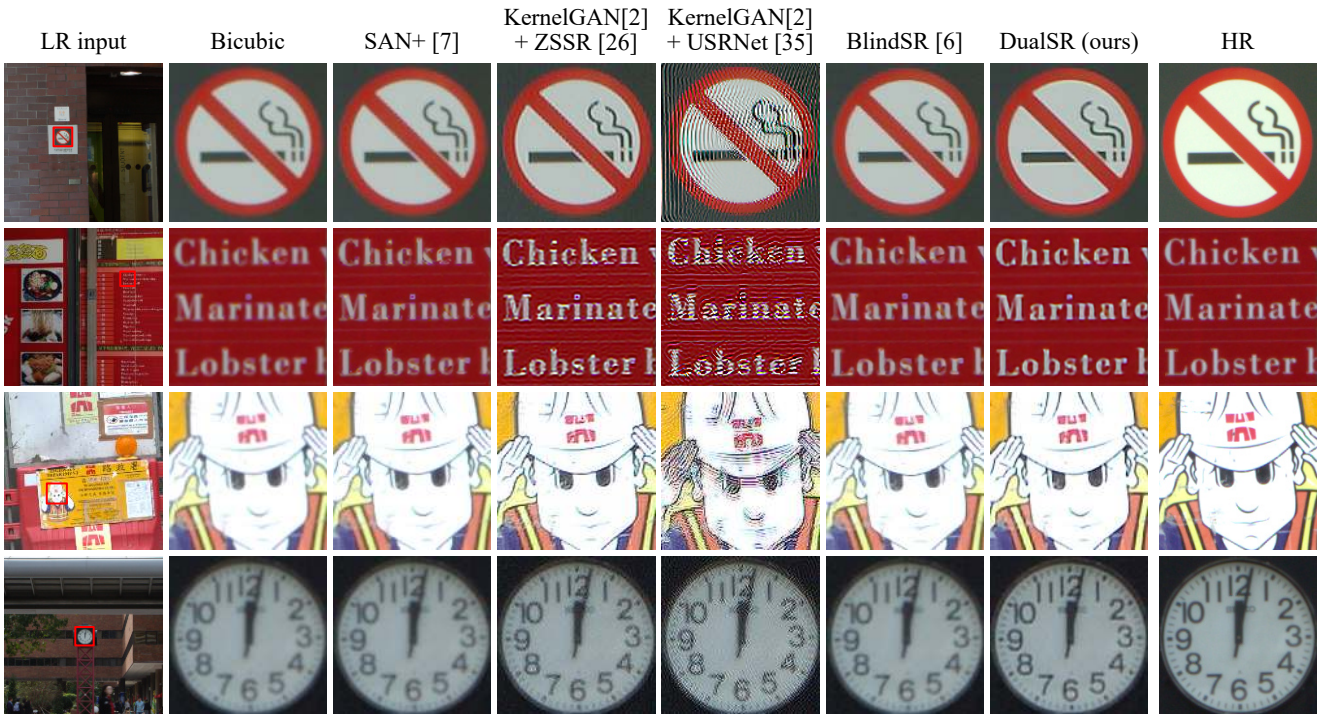

Figure 6: Qualitative comparison for 2x SR on real-world images (RealSR dataset).

third benchmark, we use the track 2 dataset of NTIRE2017 SISR challenge [27]. Similar to DIV2KRK, this benchmark uses DIV2K validation set as HR ground-truth and is different in synthesizing the LR counterparts. The downsampling operator is more complex and unknown for this benchmark.

Table 1 reports the PSNR and SSIM values. There are three categories of SR results in this table. The first category are SR methods which are trained on images downsampled with bicubic interpolation. To the best of our knowledge, SAN+ [7] is state-of-the-art in this category. For comparison, we also report the result of ZSSR [26] without kernel estimation (ZSSR using bicubic kernel as downsampler). Methods in this category are evaluated on a different degradation process than bicubic and due to their inflexibility these methods perform poorly. The second category contains blind SR results. For non-blind methods like ZSSR and USRNet [35], we use the blur kernel estimated by KernelGAN [2]. In contrast, BlindSR [6] and our DualSR methods have an integrated degradation kernel estimator. Finally, in the third category, the potential of previous methods is evaluated by providing the ground-truth blur kernels. There are no NTIRE2017 numbers, since the ground-truth kernel is not available.

According to table 1, the combination of KernelGAN and ZSSR performs better on DIV2KRK and Urban100 than methods from the first category. However, for the NTIRE2017 benchmark the combination of KenelGAN and ZSSR provides lower quality. This is mainly due to the degradation process that is more complex and is not well estimated by KernelGAN. Integrating our novel masked interpolation loss into the ZSSR architecture improves the results for NTIRE2017 substantially even if the estimated kernel is not accurate. We observe that USRNet performs well when a GT kernel is provided, but for realistic kernel estimation scenarios (kernel estimated by KernelGAN) the performance is inferior to other SR methods. Finally, BlindSR and DualSR methods provide best performance results. BlindSR assumes that the blur kernel is convolution of classic filters with an anisotropic Gaussian kernel. As a result, it provides the best PSNR numbers for DIV2KRK and Urban100 benchmarks where the blur kernel is Gaussian. However, for NTIRE2017 benchmark, the degradation process is not Gaussian so here DualSR outperforms BlindSR in both metrics. In addition, DualSR produces images with better perceptual quality and it outperforms other methods on all datasets in terms of SSIM.

The qualitative comparison of different SR methods is shown in Figure 5. Note that bicubic interpolation and SAN+ tends to produce blurry and oversmoothed images. On the other hand, the results of KernelGAN+ZSSR and KernelGAN+USRNet are oversharpended and contain severe ringing artifacts around edges and in smooth areas of the image. Thanks to the masked interpolation loss and improved kernel estimation, these artifacts are not present in the DualSR results. In comparison with BlindSR, DualSR generates sharper images with better visual quality.

5.2. Evaluation on real data

After experiments on synthesized images, we evaluate our method on real-world LR images. For this purpose, we use the new RealSR [5] dataset which is used in the NTIRE2019 competition [4]. This dataset contains raw images captured by DSLR cameras. Multiple images of the same scene have been captured with different focal lengths. Images taken by longer focal lengths contain finer details and can be considered as HR counterparts for images with shorter focal lengths. Although RealSR provides images on different scales, it is really hard to obtain image pairs that are totally aligned. That is because of complicated misalignment between images and changes in the imaging system introduced by adjusting the focal length. As a result, we only consider visual comparison of SR results. We use images captured with 28mm focal length as LR inputs and images taken at 50mm focal length as HR counterparts.

Figure 1 and figure 6 show the visual comparison of 2x upsampling on RealSR dataset. The degradation process is unknown and complicated for the LR images. Therefore, bicubic and SAN+ produce blurry results. Note that the ringing artifacts produced by KernelGAN+ZSSR and KernelGAN+USRNet are even more severe than in figure 5 and it shows that for real-world images KernelGAN does not find the correct kernel. Even BlindSR cannot estimate the degradation process correctly and produces blurry results. That is because, for real-world images, the degradation process is more complicated than a simple anisotropic Gaussian kernel. In contrast, DualSR produces artifact-free photo-realistic natural images.

5.3. Ablation study

To study the contribution of each term in the loss function for our dual-path architecture, we compare DualSR with ablations of the full version. Figure 7 illustrates visual comparison of different structures on two real-world images and table 2 shows PSNR and SSIM values on DIV2KRK benchmark. At first, we evaluate our method in the absence of masked interpolation loss. This model suffers from oversharpening and intense artifacts in the output image. Next we remove forward and backward cycle consistency losses in separate experiments (in the presence of interpolation loss). In both cases, the results are not as sharp as the proposed full DualSR output. Having cycle consistency losses and interpolation loss combined together in the objective function, makes DualSR capable of generating realistic natural images without unwanted artifacts.

Figure 7: 2x SR results on real-world images (RealSR) generated by different variations of DualSR.

Method	PSNR	SSIM
DualSR w/o masked interp loss	30.15	0.8658
DualSR w/o forward cycle loss	30.33	0.8505
DualSR w/o backward cycle loss	30.16	0.8495
DualSR (final)	30.92	0.8728

Table 2: Quantitative comparison between different variations of DualSR on DIV2KRK dataset.

6. Conclusion

Supervised deep learning methods cannot perform well when there is a mismatch between training and test data. This if very problematic for real-world SR problem where the acquisition process varies for every image. We proposed DualSR, a small dual-path architecture, which learns per image the specific LR-HR relations. It consists of a downsampler and upsampler that improve each other during training using cycle consistency losses. In addition, we introduced a new masked interpolation loss that removes artifacts from low-frequency areas of the image without smoothing the edges. Experimental results demonstrate a significant improvement over state-of-the-art SR methods. Not relying on external datasets makes DualSR very adaptive to various conditions, *e.g.* you do not have to retrain on a large dataset when your camera lens settings change. This flexibility is very important when the degradation process varies a lot. Our future work will aim to extend DualSR to larger scale factors and harder use-cases with more extreme degradation settings. Example applications could be in the domain of medical imaging or electron microscopy. In these domains a super-resolution model should not hallucinate but still generate a good high-resolution image.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017.

[2] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *Advances in Neural Information Processing Systems*, pages 284–293, 2019.

[3] Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. 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)*, pages 185–200, 2018.

[4] Jianrui Cai, Shuhang Gu, Radu Timofte, and Lei Zhang. Ntire 2019 challenge on real image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

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

[6] Victor Cornillère, Abdelaziz Djelouah, Wang Yifan, Olga Sorkine-Hornung, and Christopher Schroers. Blind image super resolution with spatially variant degradations. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH ASIA)*, 38(6), 2019.

[7] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11065–11074, 2019.

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

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

[10] Linjing Fang, Fred Monroe, Sammy Weiser Novak, Lyndsey Kirk, Cara R Schiavon, B Yu Seungyoon, Tong Zhang, Melissa Wu, Kyle Kastner, Yoshiyuki Kubota, et al. Deep learning-based point-scanning super-resolution imaging. *bioRxiv*, page 740548, 2019.

[11] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1604–1613, 2019.

[12] Juan Mario Haut, Ruben Fernandez-Beltran, Mercedes E Paoletti, Javier Plaza, Antonio Plaza, and Filiberto Pla. A new deep generative network for unsupervised remote sensing single-image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 56(11):6792–6810, 2018.

[13] He He and Wan-Chi Siu. Single image super-resolution using gaussian process regression. In *CVPR 2011*, pages 449–456. IEEE, 2011.

[14] Yu He, Kim-Hui Yap, Li Chen, and Lap-Pui Chau. A soft map framework for blind super-resolution image reconstruction. *Image and Vision Computing*, 27(4):364–373, 2009.

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

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

[17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[18] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

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

[20] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[21] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 945–952, 2013.

[22] Pejman Rasti, Tonis Uiboupin, Sergio Escalera, and Gholamreza Anbarjafari. Convolutional neural network super resolution for face recognition in surveillance monitoring. In *International conference on articulated motion and deformable objects*, pages 175–184. Springer, 2016.

[23] Yaniv Romano, John Isidoro, and Peyman Milanfar. Raisr: Rapid and accurate image super resolution. *IEEE Transactions on Computational Imaging*, 3(1):110–125, 2016.

[24] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[25] Assaf Shocher, Shai Bagon, Phillip Isola, and Michal Irani. Ingan: Capturing and retargeting the "dna" of a natural image. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019.

[26] Assaf Shocher, Nadav Cohen, and Michal Irani. Zero-shot super-resolution using deep internal learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3118–3126, 2018.

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

[28] Kensuke Umehara, Junko Ota, and Takayuki Ishida. Application of super-resolution convolutional neural network for

enhancing image resolution in chest ct. *Journal of digital imaging*, 31(4):441–450, 2018.

[29] Qiang Wang, Xiaoou Tang, and Harry Shum. Patch based blind image super resolution. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 709–716. IEEE, 2005.

[30] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[31] Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM Transactions on Graphics (TOG)*, 38(4):1–18, 2019.

[32] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE international conference on computer vision*, pages 2849–2857, 2017.

[33] Chenyu You, Guang Li, Yi Zhang, Xiaoliu Zhang, Hongming Shan, Mengzhou Li, Shenghong Ju, Zhen Zhao, Zhuiyang Zhang, Wenxiang Cong, et al. Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle). *IEEE Transactions on Medical Imaging*, 39(1):188–203, 2019.

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

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

[36] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3271, 2018.

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

[38] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.