Blind Super Resolution with Reference Images and Implicit Degradation Representation

Huu-Phu Do[1⋆], Po-Chih Hu[1⋆], Hao-Chien Hsueh[1], Che-Kai Liu[1], Vu-Hoang Tran[2], and Ching-Chun Huang[1⋆⋆]

[1] National Yang Ming Chiao Tung University, Taiwan
[2] Ho Chi Minh City University of Technology and Education, Vietnam

Supplementary

Firstly, we introduce more details on the training process of our RDSR method, including hyperparammeter settings such as learning rate and the weights of the loss functions. Next, we explore the concept of designing a downsampling network. After that, we discuss further details about our reference add-on module based on ZSSR methods and our auto-selection reference image method. Finally, we present more visual results in Fig. 1 and Fig. 2.

1 Training Details of RDSR

As mentioned above, the initial learning rate of downsampler G_{dn} and upsampler G_{up} are 2×10^{-3} and 1×10^{-5}, respectively. Additionally, the initial learning rates of the degradation estimator $E_{k\ L\to H}$ and the discriminator D_{up} is 5×10^{-5} and 1×10^{-5}, respectively. Since our downsampling network is trained from scratch, we set a higher learning rate for it, and the upsampling network G_{up} which includes both the degradation estimator $E_{k\ L\to H}$ and upscaler, is set for a lower initial rate.

Refer to previous research [2, 5, 7], loss functions such as reconstruction loss, perceptual loss, and adversarial loss are commonly used in super-resolution task. Here we applied $\mathcal{L}_{Charb}$ and $\mathcal{L}_{vgg}$ and set the same value 1 for Charbonnier loss [4] and perceptual loss [3], taking into account the similarity of the pixel values and the perceptual similarity. Furthermore, we adopted the adversarial loss $\mathcal{L}_{gan}$ for the discriminator D_{up} as mentioned in section 3.3. Both $\mathcal{L}_{gan}$ and $\mathcal{L}_{vgg}$ encourage our networks to produce more visually pleasing results.

2 Criteria for Final Result

During the Fine-tune Phase of our training, we conduct the evaluation process every 50 iterations to determine the optimal image for our training purposes. Specifically, we compute the cycle consistency loss for the entire input image. Additionally, we calculate BRISQUE [6], a no-reference quality assessment, on both our SR output image and the super-resolution (SR) image generated by a

pre-trained model at Initial Phase. If the cycle consistency loss is minimized and the score obtained from our SR output image surpasses that of the SR image, we preserve the image as our output result. This mechanism acts as a precautionary measure against the inclusion of corrupted images from our SR network. Conversely, if we fail to identify a superior image during the evaluation stage, we retain the original image as the output result. In our experiments, approximately 6.5% of output images remain unchanged by our method, indicating that certain cases cannot be enhanced by our techniques.

3 Design of Downsampling Network

We adopt the methodology of KernelGAN [1] to formulate our downsampling network. This framework enables the extraction of an explicit kernel from the weights of deep linear networks, allowing for the direct computation of the kernel prior from downsampling networks G_{dn}. For instance, we utilize the architecture of G_{dn} resembling KernelGAN [1], comprising 5 linear convolutional layers with the first 3 filter are 7, 5, and 3, and the rest are 1. Consequently, this setup yields an explicit kernel with a receptive field size of 13, indicating the network's ability to accommodate linear degradation processes within a kernel size of 13. Hence, if the kernel size of the input low-resolution (LR) image exceeds 13×13, it is advisable to adjust the architecture of G_{dn}.

The downsampling network can be trained using either supervised or unsupervised methods. In the supervised method, akin to Section 3.2, G_{dn} is trained directly with high-resolution (HR) images from a pre-trained model. Conversely, the unsupervised method, exemplified by KernelGAN [1] as utilized in the reference add-on module based on ZSSR, provides an alternative approach considering various degradation conditions or image characteristics. In our experimental in section 4.4, we have determined that the supervised approach outperforms the unsupervised one. Hence, we opt to train the downsampling network G_{dn} directly, based on this rationale.

4 Training details of Reference Add-on Module

We will provide more details on how to integrate our reference add-on module into DualSR, as mentioned previously. Similar to RDSR method, our process includes Initial Phase and Fine-tune Phase. Initially, we refrained from directly activating the reference add-on module and instead ran pure DualSR for a set number of iterations to obtain the preliminary degradation process specific to the input target image. Enabling the reference branch from the start could potentially introduce a distinct degradation process compared to the input target image due to the intervention of the reference branch. In our implementation, the Fine-tune Phase begins at 1000 iterations and concludes at 2500 iterations, resulting in a total of 3000 training iterations. Notably, through monitoring the DualSR training process, we observed that the degradation kernel stabilizes after 1000 training iterations, leading us to select 1000 iterations for the Initial

Phase. Additionally, we provide supplementary visual results utilizing the reference add-on module based on DualSR, as depicted in Fig. 2.

Fig. 1: Qualitative comparison of the RDSR based on "Blind Super Resolution" method with the isotropic kernel at scale $\times 2$.

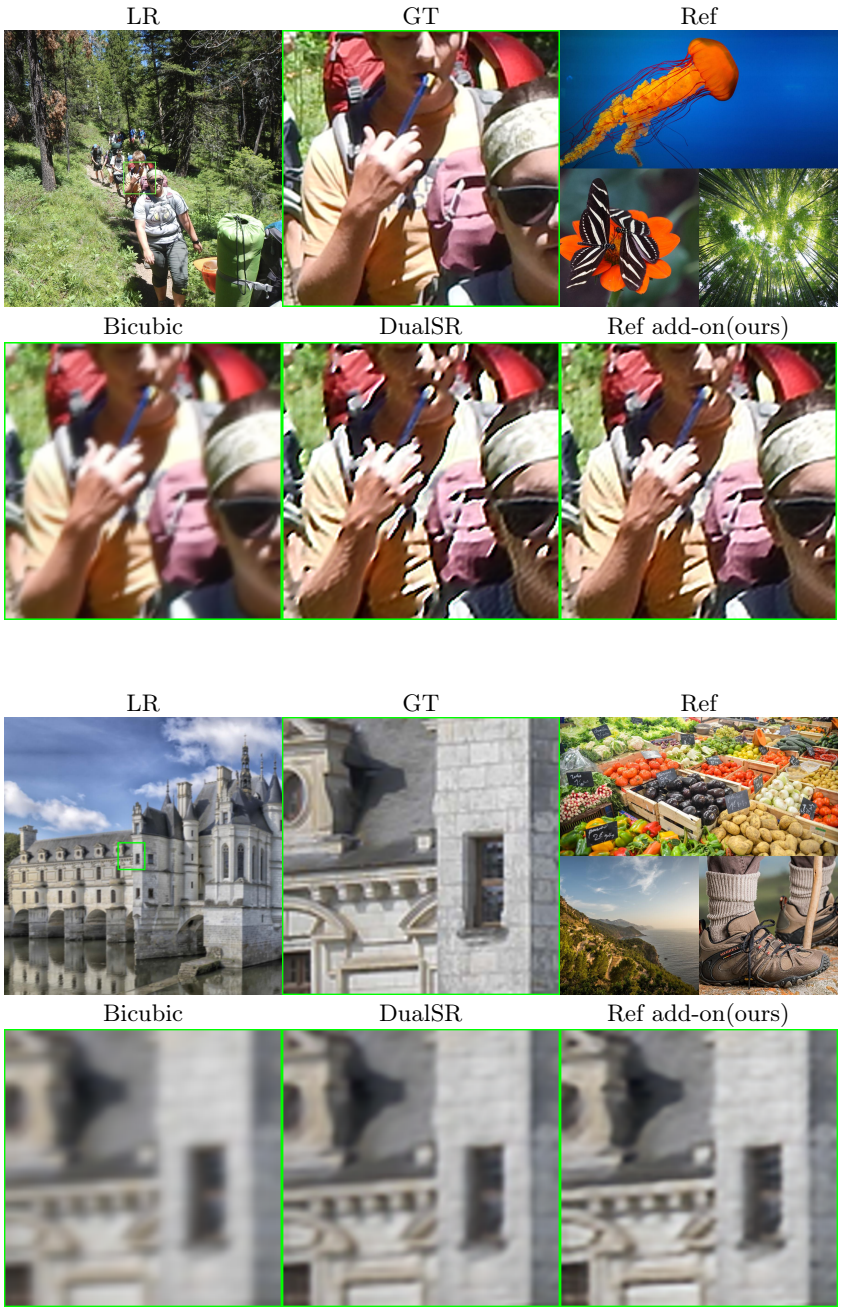

Fig. 2: Qualitative comparison of DualSR with the reference add-on module at scale $\times 2$.

5 Detail of Auto-select Reference Image Method

The auto-selection method process is summarized in Algorithm 1. We split the RGB channels from the images and calculate the mean values for the input target image. Then the mean for each reference image is to compute MSE with the input target images. Finally, we sort and take the top few reference images as our selection result. Using this approach, we have the ability to choose reference images from a pool of high-resolution images, surpassing the efficacy of random selection.

Algorithm 1 Reference images selection from a HR collection

Input: X_{LR}, $Ref^*_{HR} = \{R_1, R_2, ..., R_n\}$
Output: $Ref^*_{mse} = \{R_{1-mse}, R_{2-mse}, ..., R_{n-mse}\}$
$Ref^*_{mse} = \{\}$
X_r, X_g, X_b = `split_channel`(X_{LR})
$\overline{X_r}$, $\overline{X_g}$, $\overline{X_b}$ = `mean`(X_r, X_g, X_b)
for $i \leftarrow 1$ to n **do**
    R_{i-r}, R_{i-g}, R_{i-b} = `split_channel`(R_i)
    $\overline{R_{i-r}}$, $\overline{R_{i-g}}$, $\overline{R_{i-b}}$ = `mean`$(R_{i-r}, R_{i-g}, R_{i-b})$
    $R_{i-mse} = (\overline{X_r} - \overline{R_{i-r}})^2 + (\overline{X_g} - \overline{R_{i-g}})^2 + (\overline{X_b} - \overline{R_{i-b}})^2$
    $Ref^*_{mse} \leftarrow R_{i-mse}$
end for
`sort`(Ref^*_{mse})
return Ref^*_{mse}

References

1. Bell-Kligler, S., Shocher, A., Irani, M.: Blind super-resolution kernel estimation using an internal-gan. Advances in Neural Information Processing Systems **32** (2019)
2. Jiang, Y., Chan, K.C., Wang, X., Loy, C.C., Liu, Z.: Robust reference-based super-resolution via c2-matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2103–2112 (2021)
3. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. pp. 694–711. Springer (2016)
4. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Fast and accurate image super-resolution with deep laplacian pyramid networks. IEEE transactions on pattern analysis and machine intelligence **41**(11), 2599–2613 (2018)
5. Lu, L., Li, W., Tao, X., Lu, J., Jia, J.: Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6368–6377 (2021)
6. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing **21**(12), 4695–4708 (2012)

7. Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning texture transformer network for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5791–5800 (2020)