

BCSCN: Reducing Domain Gap through Bézier Curve basis-based Sparse Coding Network for Single-Image Super-Resolution

Anonymous Authors

The supplementary material is divided into three main sections. In the first part, we provide a detailed description of the BCSCN network architecture. The second part visualizes the intermediate steps in the search for the Bézier curve basis within BCSCN, helping readers understand its operational mechanism. Finally, we present additional qualitative results for the experiments of generalization on BCSCN (Sections 4.5) and comparisons with blind super-resolution methods (Sections 4.6).

1 NETWORK STRUCTURE

1.1 Actor Network

The input to the Actor network is the state $s_t = \{I, \hat{I}_t, t\}$, where t is expanded into a tensor of the same size as I . As illustrated in Figure 1, the inputs $\{I, \hat{I}_t, t\}$ are concatenated and fed into the Actor network. The processing sequence within the network begins with a convolutional layer followed by eight resblocks, marked in blue. Notably, the last two residual blocks employ convolutions with a stride of 2, which serve to downsample the feature maps. Finally, an average pooling layer precedes a fully connected network that generates the Bézier curve parameters a_t .

Output size	Layer
$16 \times 16 \times 7$	(I, \hat{I}_t, t)
$16 \times 16 \times 64$	Conv, BN, ReLu
$16 \times 16 \times 64$ $16 \times 16 \times 64$	Conv, BN, ReLu Conv, BN, ReLu
$16 \times 16 \times 128$ $16 \times 16 \times 128$	Conv, BN, ReLu Conv, BN, ReLu
$16 \times 16 \times 256$ $8 \times 8 \times 256$	Conv, BN, ReLu Conv stride 2, BN, ReLu
$8 \times 8 \times 512$ $4 \times 4 \times 512$	Conv, BN, ReLu Conv stride 2, BN, ReLu
$1 \times 1 \times 512$	AvgPool stride 4
11	FC, Sigmoid
11	Basis Param

Figure 1: Actor network structure

1.2 Critic Network

The Critic network receives the state $\{I, \hat{I}_t, \hat{I}_{t+1}, t\}$ as input, where t is expanded to match the size of I . As shown in Figure 2, the concatenated inputs $\{I, \hat{I}_t, \hat{I}_{t+1}, t\}$ are fed into the Critic network. Similar to the Actor network, the initial layers of the Critic consist of a convolutional layer followed by 8 resblocks. After these, the data passes through an average pooling layer, which is connected to a fully connected network that outputs the action-state value function Q .

Output size	Layer
$16 \times 16 \times 3$	$(I, \hat{I}_t, \hat{I}_{t+1}, t)$
$16 \times 16 \times 64$	Conv, ReLu
$16 \times 16 \times 64$ $16 \times 16 \times 64$	Conv, ReLu Conv, ReLu
$16 \times 16 \times 128$ $16 \times 16 \times 128$	Conv, ReLu Conv, ReLu
$16 \times 16 \times 256$ $8 \times 8 \times 256$	Conv, ReLu Conv stride 2, ReLu
$8 \times 8 \times 512$ $4 \times 4 \times 512$	Conv, ReLu Conv stride 2, ReLu
$1 \times 1 \times 512$	AvgPool stride 4
1	FC, Sigmoid

Figure 2: Critic network structure

1.3 Differentiable Neural Renderer

As shown in Figure 3, the left image illustrates the training process of the differentiable neural renderer (DNR). The basis parameters consist of 11 components, where (x_i, y_i) represent the coordinates of the control points, and (w_0, w_1) are the widths at the start and end of the curve, respectively. Initially, these basis parameters are rendered into pixel space Bézier curves using computer graphics techniques, serving as the ground truth. Subsequently, these parameters are then input into the DNR network, which is trained by minimizing the L2 distance between the network output and the ground truth. During the training of BCSCN, the parameters of the DNR are fixed to rapidly produce rendering results from the basis parameters. The right image depicts the network structure of the DNR, which takes the Bézier curve basis parameters as input and outputs the corresponding Bézier curve in the pixel space. The basis parameters are initially uplifted through four fully connected layers to a dimension of 4096, and then reshaped into a $(16, 16, 16)$ structure. The reshaped data then passes through two convolutional layers, resulting in the Bézier curve in the pixel space.

Output size	Layer
11	Basis Param
512	FC, ReLu
1024	FC, ReLu
2048	FC, ReLu
4096	FC, ReLu
$16 \times 16 \times 16$	Reshape
$16 \times 16 \times 32$	Conv, ReLu
$16 \times 16 \times 3$	Conv, Sigmoid
$16 \times 16 \times 3$	Bézier Curve Basis

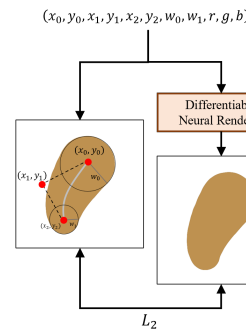


Figure 3: Differentiable neural renderer network

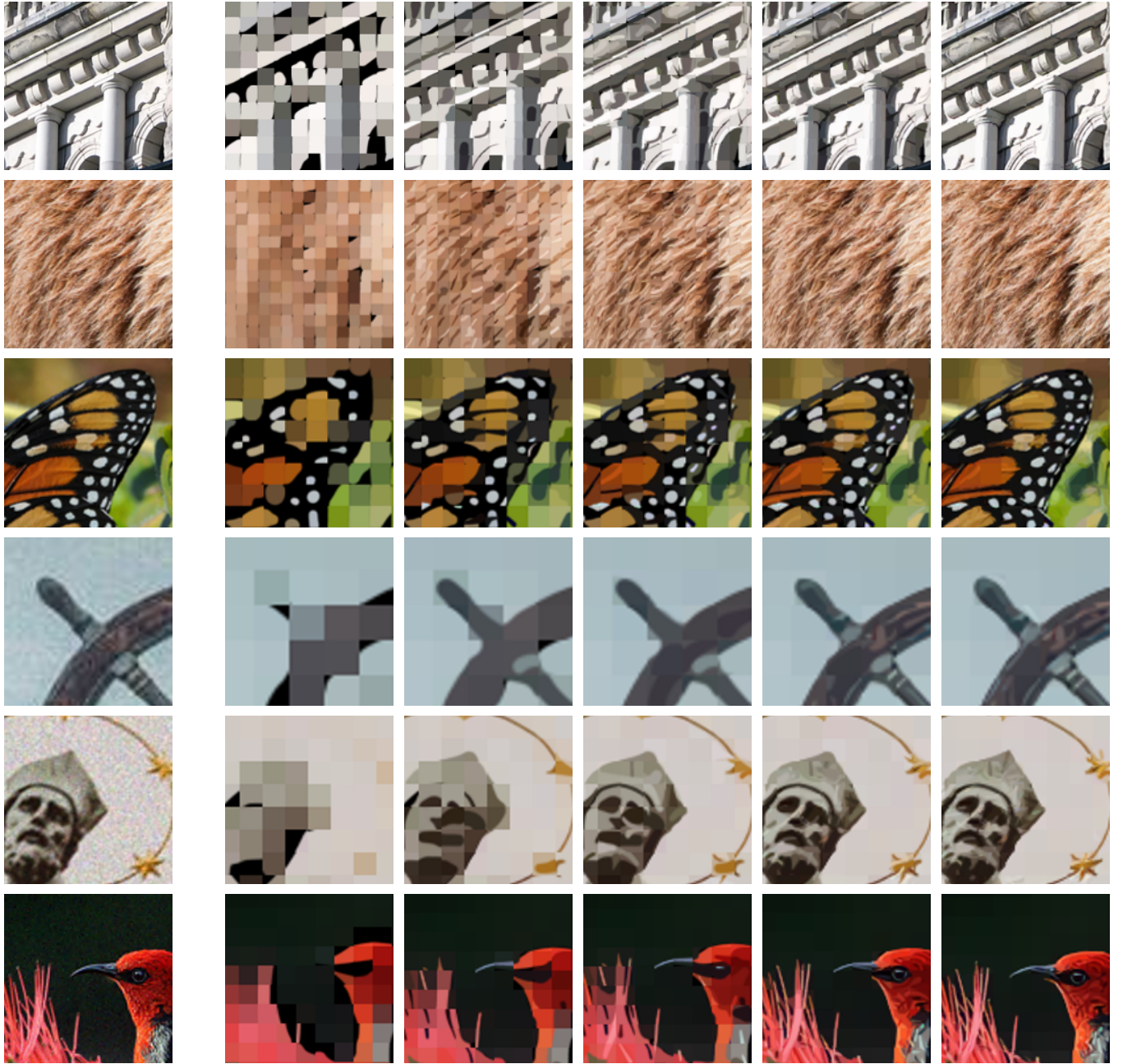


Figure 4: Illustrates the process of searching for Bézier curve bases based on image patches. The first column displays the target image, followed by the subsequent columns showing the progression of the BCSCN basis search.

2 DETAILED PROCESS OF SEARCHING

Figure 4 demonstrates the process of searching for Bézier curve bases within an image. The image is segmented into patches of size 16×16 , and each patch undergoes a search for 20 bases. This search is conducted in parallel across all patches. From the figure, it is evident that BCSCN not only reconstructs the target image's structure and fine details but also effectively eliminates artifacts, residual

structures, and noise related to degradation kernels, resulting in a “clean” reconstructed outcome.

3 ADDITIONAL VISUAL RESULTS

3.1 Experiments of Generalization

We provide additional qualitative comparisons to analyze the impact of BCSCN on the generalization performance of foundational

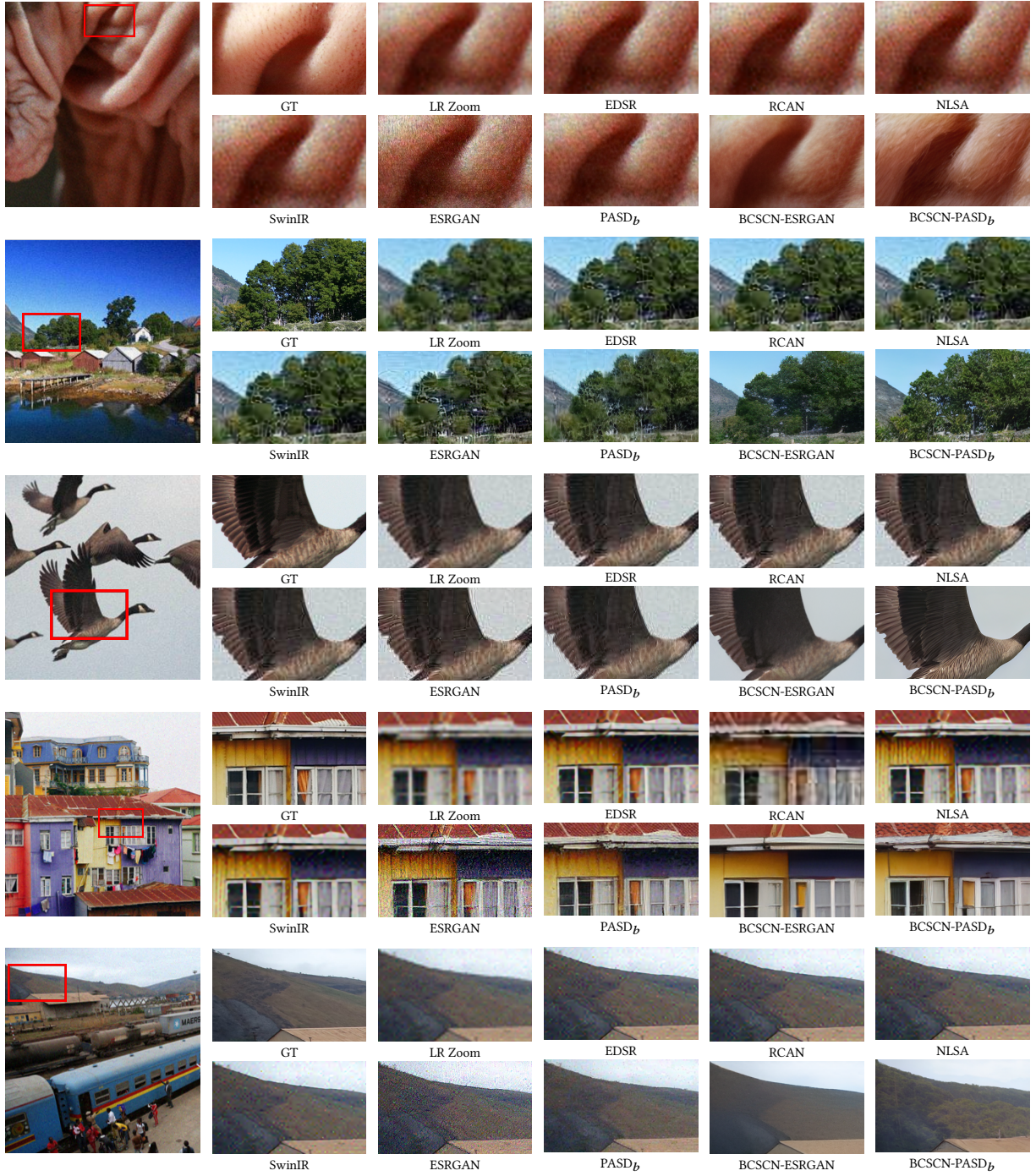


Figure 5: More qualitative results from comparisons with Single Image Super-Resolution methods. (Zoom in for details)

SISR models. As shown in Figure 5, both BCSCN-ESRGAN and BCSCN-PASD_b effectively eliminate degradation factors, restoring clean results that significantly surpass other baseline methods. This demonstrates the BCSCN preprocessing network’s ability to enhance the generalization performance of SISR models.

3.2 Compared with Blind SR Method

We provide additional qualitative results from comparisons with state-of-the-art blind super-resolution methods, as shown in Figure 6. The outcomes demonstrate that BCSCN-PASD_b generates

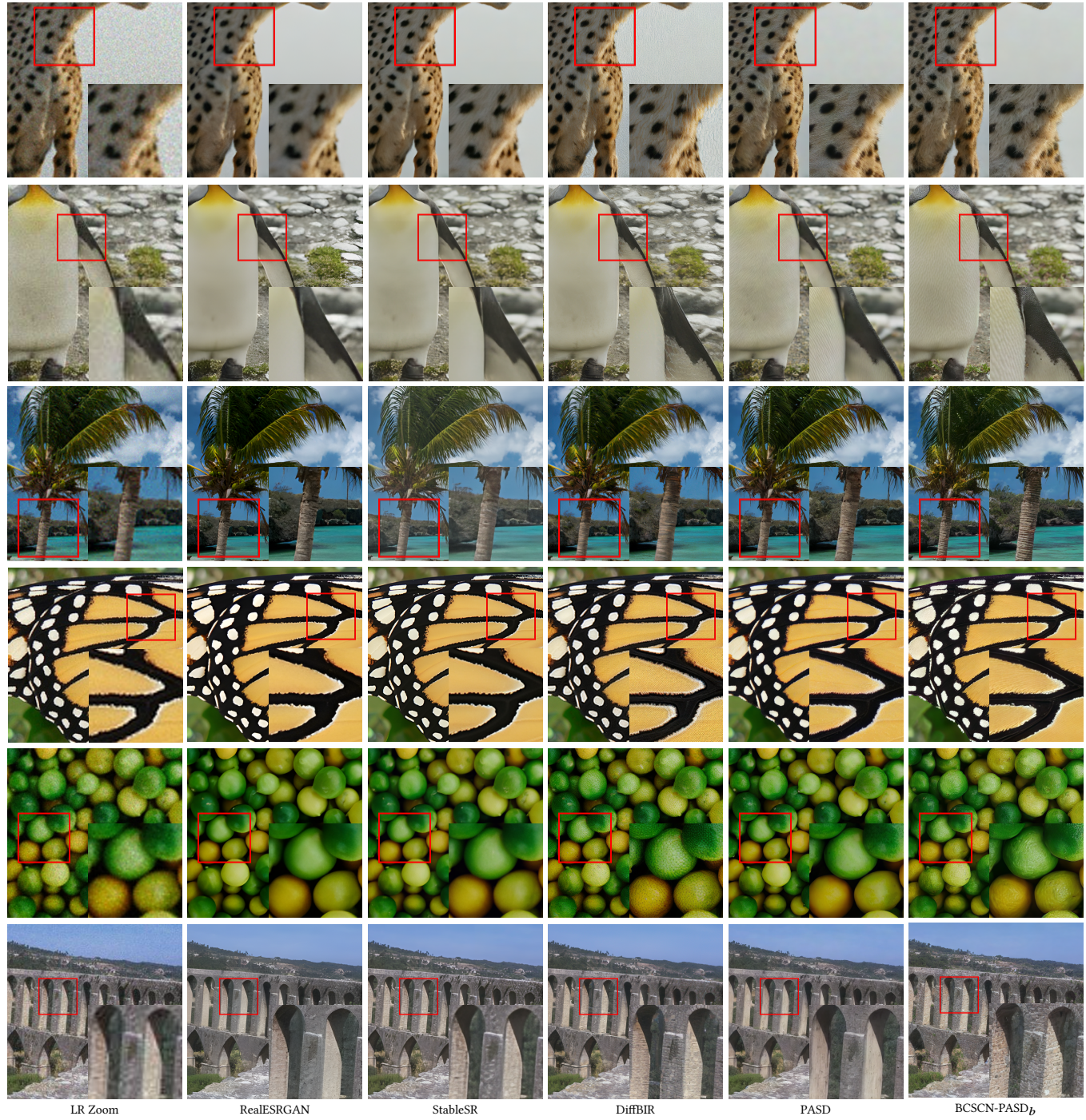


Figure 6: More qualitative results from comparisons with blind super-resolution methods. BCSCN-PASD_b achieved competitive visual outcomes. (Zoom in for details)

clean and sharp results while producing rich and realistic textures, achieving competitive visual effects.