1. Details of Ablation Study

In the ablation study of block sizes and dilation rates, it should be noted that if the size of the LR block is $m \times n$, then the basic size of the Ref↓ block is set to $m_{H\text{Ref↓}} \times n_{W\text{Ref↓}}$, where $H_{Ref↓}$, $W_{Ref↓}$ and $H_{LR}$, $W_{LR}$ are the height and width of the Ref↓ feature and the LR feature, respectively. We multiply the basic size of different scale factors in the ablation study, and use the scale factors to denote the Ref↓ block size in Fig. 5(b) of body text and in the following descriptions.

**Influence of LR block sizes.** The FLOPS in Fig. 5(a) is computed by taking as input a $192 \times 192$ LR image and a $768 \times 768$ Ref image. The Ref↓ block size is 1.5 and the dilation is 1.

**Influence of Ref↓ block sizes.** The FLOPS in Fig. 5(b) is computed on a $128 \times 128$ LR image and a $512 \times 512$ Ref image. The LR block size is 8 and the dilation is 1.

**Influence of dilation rates.** The FLOPS in Fig. 5(c) is computed on a $120 \times 120$ LR image and a $480 \times 480$ Ref image. The LR block size is 12 and the Ref↓ block size is 1.5.

2. More Visual Results

We show more visual results of the proposed MASA and other state-of-the-art methods, including RCAN [6], HAN [2], ESRGAN [4], SRNTT [7] and TTSR [5]. RCAN and HAN are SISR methods that have achieved the best performance on PSNR, and ESRGAN is a GAN-based SISR method that is considered state-of-the-art in visual quality. SRNTT [7] and TTSR [5] are state-of-the-art RefSR methods. The visual comparison on CUFED5 [7] testing set are shown in Fig. 1 and Fig. 2, and the comparison on Sun80 [3] and Urban100 [1] are shown in Fig. 3 and Fig. 4, respectively.

It can be observed that our MASA can restore more regular structures and generate photo-realistic details.
| Input LR | RCAN | HAN | ESRGAN | Reference | SRNTT | TTSR | MASA (Ours) |
|---------|------|-----|--------|-----------|-------|------|-------------|
| ![Input LR](image1) | ![RCAN](image2) | ![HAN](image3) | ![ESRGAN](image4) | ![Reference](image5) | ![SRNTT](image6) | ![TTSR](image7) | ![MASA (Ours)](image8) |

Figure 1: Visual comparison among different SR methods on the CUFED5 [7] testing set.
| Input LR | RCAN | HAN | ESRGAN |
|---------|------|-----|--------|
| Reference | SRNTT | TTSR | MASA (Ours) |

Figure 2: Visual comparison among different SR methods on the CUFED5 [7] testing set.
Figure 3: Visual comparison among different SR methods on the Sun80 [3] dataset.
| Input LR | RCAN | HAN | ESRGAN |
|---------|------|-----|--------|
| Reference | SRNTT | TTSR | MASA (Ours) |

Figure 4: Visual comparison among different SR methods on the Urban100 [1] dataset.