Supplementary Material: Texture Hallucination for Large-Factor Painting Super-Resolution

Anonymous ECCV submission

Paper ID 181

A Network Details

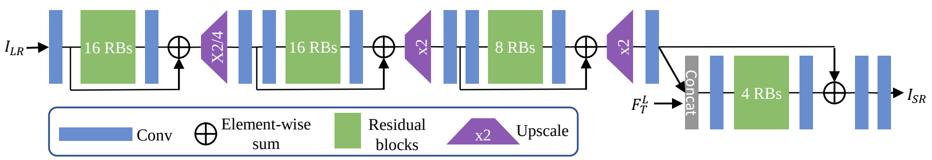

Fig. A.1: Illustration of the proposed network for large-scale painting super-resolution. "Conv" denotes convolutional layer, and "RBs" indicates residual blocks removed the batch normalization

An overview of the proposed network structure is illustrated in Fig. A.1. All batch normalization (BN) layers are removed from the residual blocks (RBs) since BN layers shrink range flexibility from networks by normalizing the features [1,2]. Therefore, the structure of a residual block is Conv-ReLU-Conv with a short cut. The convolution kernel is set to be 3×3 for all convolutional (Conv) layers, and zero-padding is conducted to preserve the feature map size after convolution. Instead, the concatenation layer adopts a kernel size of 1×1. The activation function is ReLU (except for the output layer that uses tanh), and the number of channels at each intermedia convolutional layers is set to be 64. The upscaling layers employ the sub-pixel convolution [3]. The network difference between 8× and 16× upscaling is that the first upscaling layer will perform 2× and 4× upscaling, respectively.

B Visual Comparisons

This section will demonstrate more visual comparison between the state-of-the-art methods and the proposed method on the newly collected PaintHD dataset. More specifically, our method is compared to a state-of-the-art SISR method RCAN [4] and a representative Ref-SR method SRNTT [5]. The comparison will be conducted at 8× and 16×, respectively.

B.1 Results of 8×

The visual comparison of 8× upscaling is shown in Figs. B.1 and B.2. Each example spans two rows, where the upper and lower figures in the first column are the LR input and reference, respectively. The rest columns are results from corresponding methods. For better visual comparison, only two zoom-in areas are cropped from the original results as indicated by the color-coded boxes in the LR input. The input and reference may be patches from the same original painting, but they avoid large overlap and would show different scales, angles, and styles of the stroke.

B.2 Results of 16×

By the same token, the visual comparison is conducted at 16× in the similar way as shown in Figs. B.3 and B.4.

B.3 Effect of Different References

For Ref-SR methods, investigation on the effect from references is an interesting and opening problem, e.g., how the references affect SR results, how to control (i.e., utilize or suppress) such effect, etc. This section intends to explore the effect of references in the proposed Ref-SR method. As shown in Figs. B.5 and B.6, the same LR input is super-resolved using different reference images, respectively.

In general, the local texture in the results would vary with the reference texture. In Fig. B.5, the stroke/texture scale in Reference 1 is relatively smaller than that in Reference 2, thus the texture presented in the results using Reference 1 would be of smaller scale, i.e., more details and visually sharper. In Fig. B.6, Reference 2 shows stronger canvas texture, which is transferred to the results. The proposed method transfers the texture from different references to the results, while preserving the content of the LR input.

Fig. B.1: Visual results with scaling factor 8×

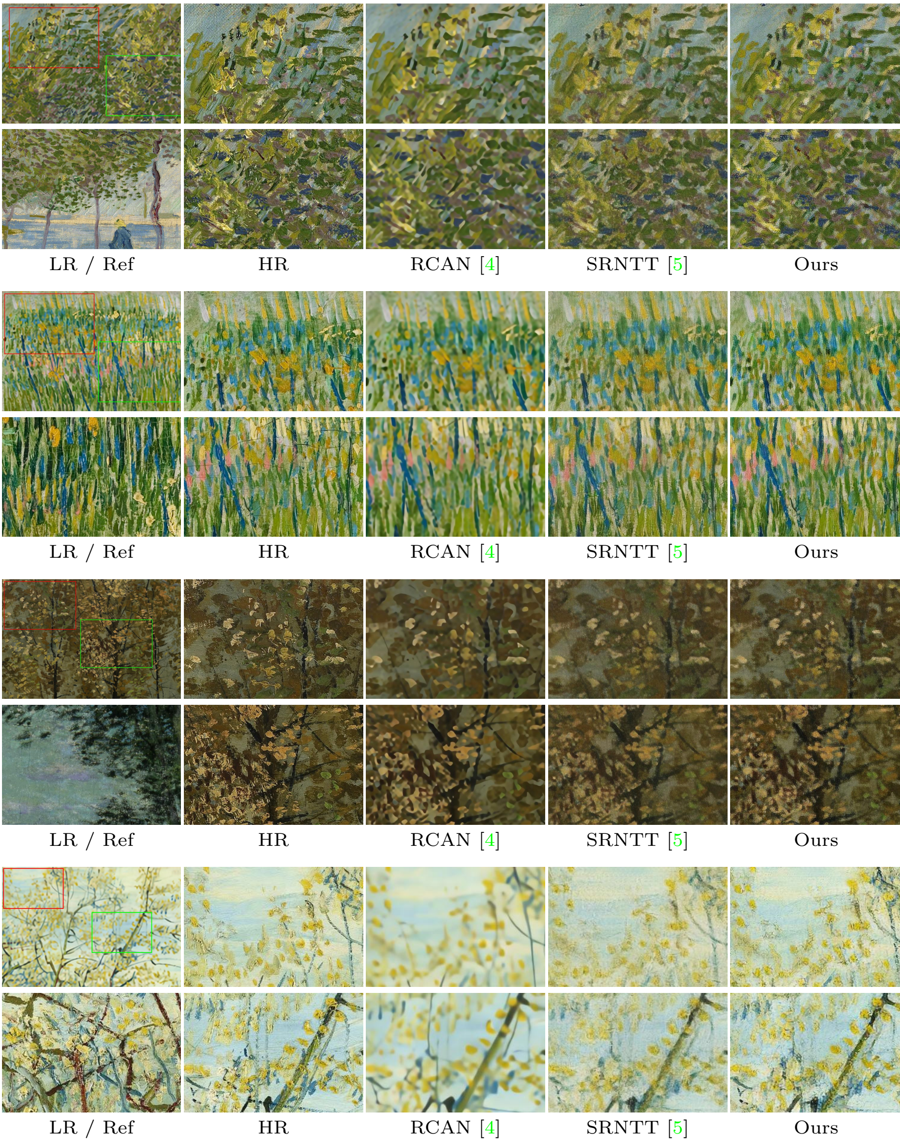

Fig. B.2: Visual results with scaling factor 8×

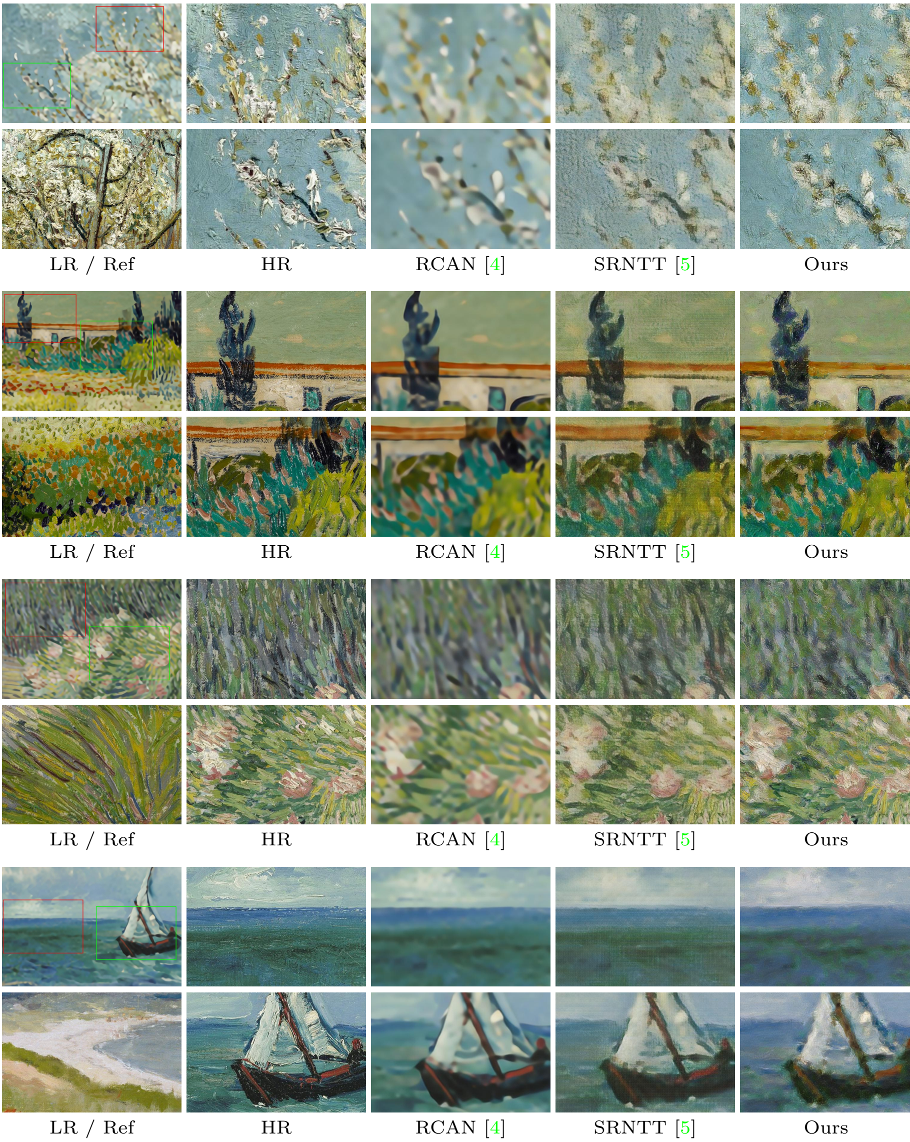

Fig. B.3: Visual results with scaling factor 16×

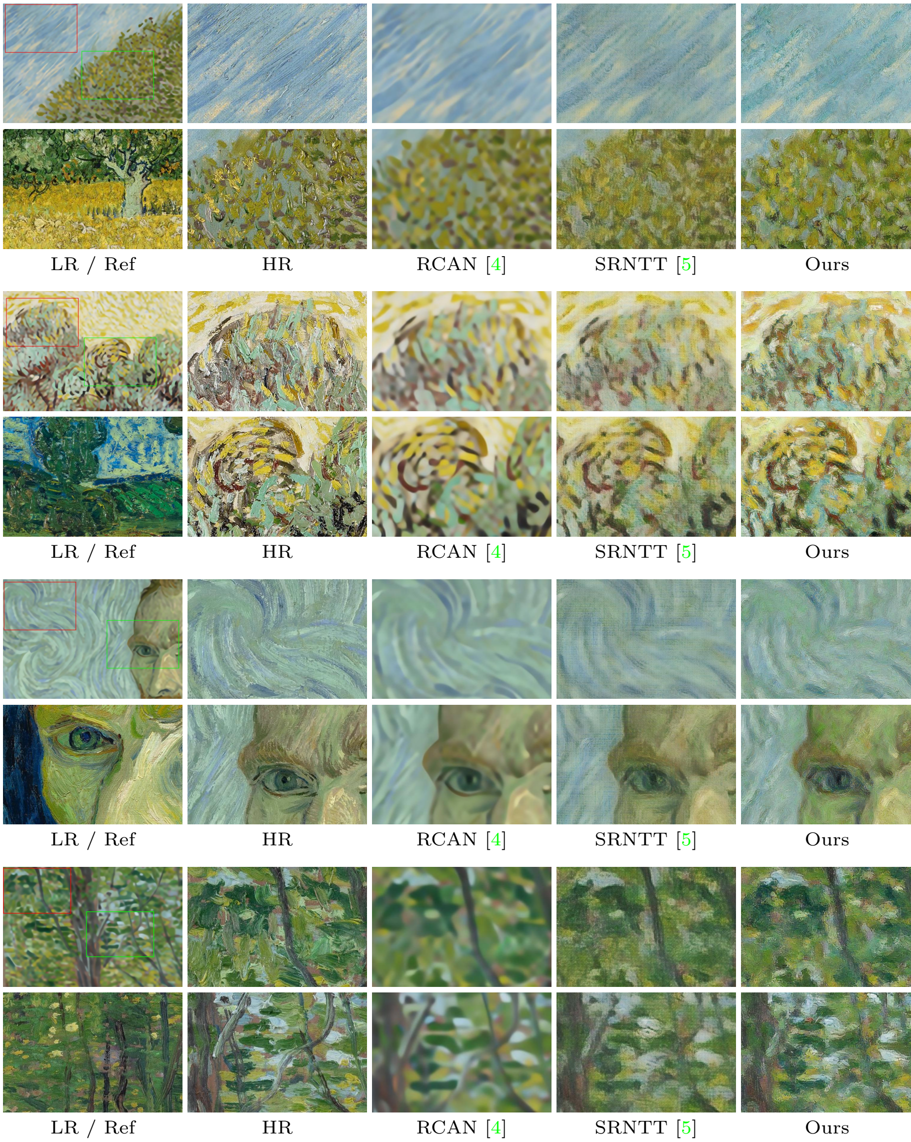

Fig. B.4: Visual results with scaling factor 16×

Fig. B.5: Visual results with scaling factor 8× using different reference images. For each reference image, we show three patches extracted from the corresponding results

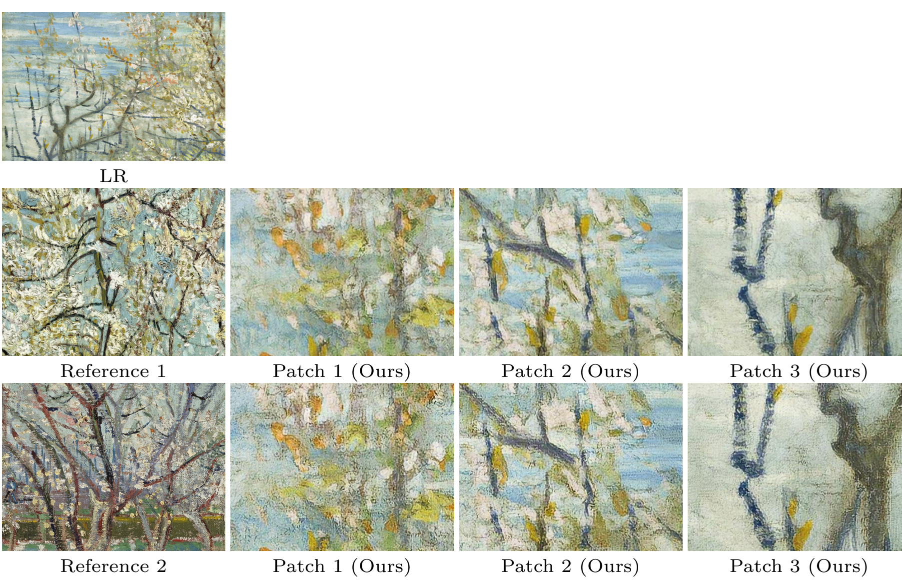

Fig. B.6: Visual results with scaling factor 8× using different reference images. For each reference image, we show three patches extracted from the corresponding results

References

1. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: CVPRW (2017)
2. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: CVPR (2017)
3. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR (2016)
4. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV (2018)
5. Zhang, Z., Wang, Z., Lin, Z., Qi, H.: Image super-resolution by neural texture transfer. In: CVPR (2019)