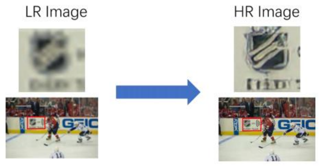
Multi-Reference Image Super-Resolution: A Posterior Fusion Approach

Ke Zhao, Haining Tan, Tsz Fung Yau
David Lindell, University of Toronto

Motivation

Single-Image Super-Resolution (SISR)

- Recover the high-resolution version of a image
- Ill-posed inverse problem with infinite solutions
- No universal regularization to guide optimization



Single-Reference-based Super-Resolution (SRefSR)

- High-resolution reference provides additional information for texture restoration
- Only one reference supported

Our Goal

- Utilize information from multiple references
- Should be easily integrated into any SRefSR pipeline to improve performance
- Low computational cost

Potential Applications: Video game scenes

- NVIDIA DLSS: Train separate SR neural networks for each video game
- Multiple HR texture patches readily available
- Could use single model for any video game, and save huge amount of computing resources

Related Work

SISR models

Convolutional Neural Networks

- SRCNN [1], EDSR [2]
- Smooth reconstruction

Generative Adversarial Networks

- SRGAN [3]
- Generate "fake" texture details
- Texture details different from ground-truth

SRefSR models

Convolutional Neural Networks

- CrossNet [4]
- Align reference and LR features
- Not specify suitable references (by similarity)

Similarity-Aware Deformable Convolution

- SSEN [5]
- Detect relevancy of reference
- More robust with irrelevant references
- Performance still relies on reference choice

Dual zoomed observations with self-supervision

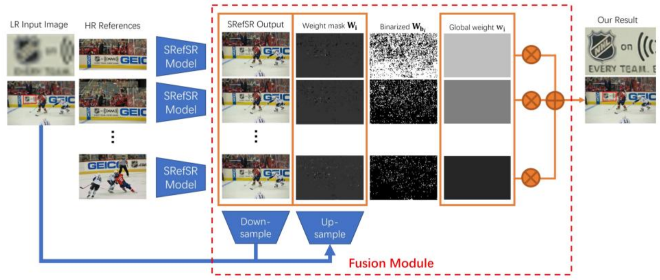
- SelfDZSR [6]
- Telephoto as references

References

[1] C. Dong et al., "Image super-resolution using deep convolutional networks," CoRR, vol. abs/1501.00092, 2015.
[2] B. Lim et al., "Enhanced deep residual networks for single image super-resolution," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, July 2017.
[3] C. Ledig 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, 2017, pp. 4681–4690.
[4] H. Zheng et al., "Crossnet: An end-to-end reference-based super resolution network using cross-scale warping," in Proceedings of the European Conference on Computer Vision (ECCV), September 2018.
[5] G. Shim et al., "Robust reference-based super-resolution with similarity-aware deformable convolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.
[6] Z. Zhang et al., "Self-supervised learning for real-world super-resolution from dual zoomed observations," arXiv preprint arXiv:2203.01325, 2022
....

New Technique

Multi-Reference-based Super-Resolution pipeline



Part 1: Single-Reference-based Super-Resolution (SRefSR) Module

- Align input and reference image spatially and semantically
- One LR image + Multiple HR reference images → Multiple SRefSR outputs

Part 2: Image Fusion Module

- Goal: Combine the best regions of each SRefSR output
- Step 1: Adaptive Weight Masking

$$W_i = \mathcal{U}\left(\exp\left(-\beta\left(\mathcal{D}(I_i) - I_{input}\right)^2\right)\right) \qquad \hat{I}(p) = \frac{1}{\sum_{i=1}^{N} W_i(p)} \sum_{i=1}^{N} I_i(p) W_i(p)$$

(The idea: higher weights for matching regions)

- Step 2: Globally Reference-Quality-based Weighted Averaging

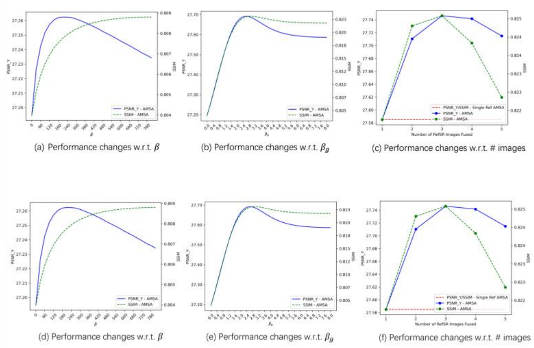
$$W_{b_i}(p) = \begin{cases} 1 \; if \; i = argmax_i \, W_i(p) \\ 0 \; otherwise \end{cases} \qquad w_i = \exp\left(\beta_g \sum_p W_{b_i}(p)\right) \qquad I_{fused} = \frac{1}{\sum_i w_i} \sum_i w_i \hat{I}_i$$

(The idea: higher overall weights for high-quality SRefSR output)

Experimental Results

Qualitative Comparison

Quantitative Evaluations



- When $\beta, \beta_g = 0$: Naive Fusion
- **1^{st} & 2^{nd} col:** PSNR_Y/SSIM rise as β & β_g increase, showing distorted regions are penalized
- **3^{rd} col:** RefSR images with less and less quality are fused, showing the fusion method is resistant to low-quality inputs.
- **1^{st} & 2^{nd} row**: The proposed method is tested on two RefSR models, showing consistent performance improvement