CSC2529 Project Proposal: Reference-based Super-Resolution

Ke Zhao Haining Tan Tsz Fung Yau

Motivation

Single-image super-resolution (SISR) is a computer vision task that reconstructs a high-resolution (HR) image from a low-resolution (LR) image. Typically, estimating a high-resolution image from its low-resolution counterpart is an ill-posed inverse problem [2], meaning that there are infinitely many solutions that satisfy the measurements. This underdetermined nature of the problem is particularly pronounced for images with abundant high-frequency details. To reach an optimal solution to the inverse problem with respect to certain criteria, additional regularization terms need to be specified. However, no simple regularization term can be specified to cover the characteristics of all kinds of images, thus conventional SISR algorithms are usually poorly performed.

Reference-based super-resolution (RefSR) approaches explicitly exploit additional information from an external HR reference image to enhance the SISR process. Intuitively, sufficient information is encoded in a reference image that contains the same content as that on the LR image to facilitate texture restoration. However, as far as we know, such models can only take one reference image. We intend to improve upon that work and propose a new framework so that information from multiple reference images can be exploited to achieve better SISR performance. This new framework can be applied to a wide range of applications. For example, in video game scenes, HR patches for each are readily available as the texture to be mapped, and they can be used as reference images to perform super-resolution. This approach would be applied to any video game and would save huge computing resources compared to the state-of-the-art NVIDIA DLSS, which trains separate SR neural networks for each video game.

Related Work

CNN-based Super-resolution

With the popularity of Convolutional Neural Networks (CNN), learning-based approaches demonstrate significantly better performances given an appropriate training set. Early-stage CNN-based SISR models like SRCNN [1] choose pixel-level reconstruction errors such as MSE and MAE between the recovered HR image and ground truth as loss functions to optimize. Furthermore, significant improvements can be made by optimizing the standard CNN architecture. For instance, the approach EDSR proposed by Lim et al. [4] applied the residual network architecture to the SR task and achieved superior results. While these algorithms tend to maximize the peak signal-to-noise ratio (PSNR), they often result in smooth reconstruction lacking high-frequency details and are perceptually unsatisfying.

GAN-based Super-resolution

To state the problem of SISR in another way, downsampling an HR image is an irreversible compression process during which much high-frequency information is lost. Instead of trying to recover the lost information from nowhere, Ledig et al. [3] adopt Generative Adversarial Networks (GAN) and proposed SRGAN that generates "fake" texture details that are visually realistic. While these results are perceptually satisfying, texture details in these images are hallucinations and are often different from those in the ground-truth images, resulting in PSNR degradation. This deficiency makes the methods like SRGAN unsuited for fidelity-sensitive applications like medical imaging.

Additionally, pure GAN-based SISR approaches fail to produce satisfying results on test images with complicated components, often resulting in distorted color lumps.

Reference-based Super-resolution

Zheng et al. [7] proposed an end-to-end approach CrossNet based on fully convolutional neural networks that can perform spatial alignment between the reference features and the LR features. One issue of the early RefSR approaches is that they did not specify how suitable the reference image is and the relationship or similarity between the LR image and the reference. Motivated by this, Shim et al. [5] proposed a robust RefSR model that is aware of the relevancy of the reference image, leading to a more robust result when the reference images are irrelevant and can outperform SRGAN in terms of generating visually stratifying SR images while also achieving high PSNR. However, the performance of RefSR models relies on the choice of proper reference images. Zhang et al. [6] proposed to use dual zoomed observations (from a telephoto) as references and apply self-supervised techniques to that.

Project Overview

In this project, we will first reproduce the related work by doing an empirical study of single-reference-based SR models [6]. If time permits, we will try to build a framework for multi-reference-based super-resolution tasks to enhance the SR process, as shown below:

1. Create a benchmark with a collected dataset for performance evaluation of both single-reference-based and multi-reference-based super-resolution tasks. The dataset may also be used for model training or fine-tuning if required.
2. Reimplement several important methods in the research and development for the super-resolution task including SRCNN, SRGAN, and SelfDZSR with the dataset, then evaluate them using the benchmark created.
3. Modify the neural network structure of the existing models to allow the use of multiple inputs, which are also evaluated using the benchmark.

Goals & Milestones & Timeline

We divided the project into the following milestones:

- 20 Nov: Collect datasets for model testing and training + Create a benchmark for model evaluation
- 27 Nov: Reproduce important super-resolution models and compare their performance
- 4 Dec: Make attempts to design a more flexible model architecture to exploit information from multiple reference images to achieve performance improvement
- 8 Dec: Wrap up our work with poster presentation, source code submission, and final project report

References

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *CoRR*, abs/1501.00092, 2015.

[2] S Kathiravan and J Kanakaraj. An overview of sr techniques applied to images, videos and magnetic resonance images. *Smart CR*, 4(3):181–201, 2014.

[3] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, 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*, pages 4681–4690, 2017.

[4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 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.

[5] Gyumin Shim, Jinsun Park, and In So Kweon. 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] Zhilu Zhang, Ruohao Wang, Hongzhi Zhang, Yunjin Chen, and Wangmeng Zuo. Self-supervised learning for real-world super-resolution from dual zoomed observations. *arXiv preprint arXiv:2203.01325*, 2022.

[7] Haitian Zheng, Mengqi Ji, Haoqian Wang, Yebin Liu, and Lu Fang. 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.