MA797 project proposal: Image Super-resolution in real-world

Jie Hu, Peiran Wang, Chanae Ottley

October 2021

1 Topic

Single image super-resolution (SISR) is a classical problem in computer vision. It aims to obtain a high-resolution output from one of its low-resolution versions. This topic has many applications in the real world, e.g., old-video restoration, photo enhancement, etc. With the help of emerging, powerful deep learning algorithms, we want to mention current challenges in this field, reproduce and improve the current algorithms in the SISR literature [1, 4] to achieve better performance, and conclude with future trends in SISR.

2 Methodology

Our baseline method is a super-resolution convolutional neural network (SRCNN) [1], which uses an end-to-end mapping between the low and high-resolution images in the YCbCr color space. The mapping is achieved as a deep CNN that takes the low-resolution image as the input and outputs the high-resolution image. It only contains three layers, and each layer has a convolution layer with an activation function. Three layers of SRCNN limit its ability to express and reconstruct the targeted image features, e.g., edge, texture, and could bring some unrelated features into account. When the trained dataset includes different global luminance and colors and we only focus on improving the details in the image, we may want to decompose the image and capture the details first, instead of taking every information into account. Besides, the recent usage of Generative Adversarial Network (GAN) also helps generate missing details in the image, e.g., [2]. We want to explore more types of neural networks in the literature and apply them to this SISR topic.

We plan to use three performance metrics in our project: peak signal to noise ratio (PSNR), mean squared error (MSE), and structural similarity (SSIM). The quality of the predict image will be better if the PSNR value is higher, so is MSE. SSIM is a method for comparing the similarity between two images. If SSIM approaches 1, we say the two images are identical.

3 Data set

We will use the dataset in [4] that includes both high-resolution (HR) low-resolution (LR) images with size 1024×512 captured from a variety of scenes (500 places), including indoor and outdoor scenes. Many existing datasets [1, 3] come from simple bicubic degradation by down-sampling the original image and adding noise. The dataset in [4] is collected from iPhone 11 Pro Max so that it naturally contains many complex degradations for the LR images. With the help of a multi-camera system with different focal lengths, they can capture the images with different levels of details from different lenses at the same time, minimizing the mismatching effect in image content. Moreover, the dataset has been post-processed, i.e., lens angle correction and image alignment, to eliminate the alignment artifacts around the boundary.

The link of the dataset to be used is: `https://drive.google.com/drive/folders/1-8MvMEYMOeOE713DjI7TJKyRE-LnrM3Y`.

References

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

[2] 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.

[3] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *International Journal of Computer Vision*, 127(8):1106–1125, 2019.

[4] Xi Yang, Wangmeng Xiang, Hui Zeng, and Lei Zhang. Real-world video super-resolution: A benchmark dataset and a decomposition based learning scheme. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4781–4790, 2021.