ECS795P
Individual Reports on Critical Analysis
Super-Resolution
161189284
Guanzhen Wu

Super-resolution imaging (SR) is a class of techniques that enhance or increase the resolution of an imaging system. The authors of discussed paper present a convolutional neural network for image super-resolution called SRCNN which shows that the traditional sparse coding-based SR method can be reformulated as a deep convolutional neural network. The proposed method, SRCNN, achieves better performance than the state-of-the-art methods with a lightweight structure [1]. Since then, the super-resolution algorithm based on SRCNN has been widely studied. Due to the application of deep convolutional neural networks in SR and the benefits from deep learning, subsequent researchers started to delve deeper in the field of deep learning.

The core problem in SR deep learning model is the conflict between complexity of model and resolution image quality. To improve the quality of resolution image, we need to deepen the network or increase the number of parameters [1,2], which also increases the computational complexity and makes the model huge and slow. Taking SRCNN as an example, the author experimented with three different filter sizes, (9-1-5, 9-3-5, 9-5-5), and their parameter numbers were 8,032, 24,416, and 57,184 respectively. It can be seen that 9-5-5 has more than twice the number of parameters than 9-3-5, but the performance improvement is marginal. Therefore, recent papers focus on how to maximize the quality of the resolution image while increasing the computational complexity as little as possible. The main improvements are mainly in the following aspects: training method [3], network architecture [4], number of network layers [2], parameters of filters [2], loss function [5], etc. Meanwhile, with the rise of GAN networks in recent years, the generative adversarial model is gradually showing its powerful ability in the field of SR [6].

In particular, the core ideas of these aspects and the specific ways to enhance them are shown as below:

1. Proposed a new structure-enhancing loss, coined gradient variance loss, for the SISR task to alleviate the issue of over-smoothness commonly existing in the generated SR images when trained with the L1 or L2 loss. [5]
2. By cascading small filters many times in a deep network structure to exploit contextual information over large image regions efficiently. [7]
3. To make full use of the relative information between the local features and the overall features, residual features learning is introduced to the branch structure of large scale-feature mapping to further boost the reconstruction performance. [3]
4. Create internal dataset using the test image itself and exploits a double-branch structure to capture and train the image features at different scales. [3]
5. Introducing Spatial Attention Block (SAB) which can substantially improve the quality of super-resolution, without increasing the complexity of the CNN network. [8]

However, even though there are various improvements in model speed and image reconstruction quality, the SR algorithm still faces many problems. Most of the training data of previous deep learning models are obtained by downsampling high-resolution images, but this fuzzy way is too homogeneous, and although good results will be obtained in validation set, the generalization ability of the model is still not robust. Meanwhile, the commonly used Bicubic Interpolation method of reconstruction does not perform well in the text images because the spatial structure of text image is quite different.

For the first problem, among the existing studies, the use of internal dataset and blind set up can solve the problems to some extent and enhance the robustness of the model, but the improvement is limited [9]. The lack of a large amount of training data and the vague interpolation approach leads to the concept of adding perceptual knowledge to the SR algorithm [10]. GAN algorithm is an example followed this concept. However, I think adding perceptual knowledge to neural networks can be investigated in more details. The most interesting question to me is how we can analogize to computers the process of human perception of things or sensations. Combined with human cognition, the neural network should be able to have a processing method that is more inclined to the human brain. The attention mechanism in deep learning is an example, and the application of the attention mechanism to the SR algorithm has indeed achieved obvious success. Especially on the reconstruction of feature details [8]. I believe that if we can understand how the human brain reconstructs and replenishes the blurred images or lost frames in video, the SR algorithm can be more powerful.

Reference:

[1] Dong, C., Loy, C.C., He, K. and Tang, X., 2015. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), pp.295-307.
[2] Du, X., Qu, X., He, Y. and Guo, D., 2018. Single image super-resolution based on multi-scale competitive convolutional neural network. *Sensors*, *18*(3), p.789.
[3] Liu, J., Xue, Y., Zhao, S., Li, S. and Zhang, X., 2020. A convolutional neural network for image super-resolution using internal dataset. *IEEE Access*, *8*, pp.201055-201070.
[4] Kim, J., Lee, J.K. and Lee, K.M., 2016. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637-1645).
[5] Abrahamyan, L., Truong, A.M., Philips, W. and Deligiannis, N., 2022. Gradient Variance Loss for Structure-Enhanced Image Super-Resolution. *arXiv preprint arXiv:2202.00997*.
[6] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).
[7] Kim, J., Lee, J.K. and Lee, K.M., 2016. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).
[8] Liu, Z.S., Wang, L.W., Li, C.T., Siu, W.C. and Chan, Y.L., 2019, October. Image super-resolution via attention based back projection networks. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (pp. 3517-3525). IEEE.
[9] Riegler, G., Schulter, S., Ruther, M. and Bischof, H., 2015. Conditioned regression models for non-blind single image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 522-530).
[10] Johnson, J., Alahi, A. and Fei-Fei, L., 2016, October. Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision (pp. 694-711). Springer, Cham.