Divya Hitesh Chhipani
200369536

# [ECS795P] Deep Learning and Computer Vision
# Critical Analysis Report on Super Resolution

---

**Super-resolution**

Super-resolution (SR) is one of the prominent tasks in the field of computer vision which aims to get a high-resolution (HR) output from one or more low-resolution (LR) image of the same scene. According to the number of input LR images, the SR can be classified into single image super-resolution (SISR) [1] and multi-image super-resolution (MISR) [2]. Super-resolution is an ever-growing research area and has seen many applications such as in the field of face hallucination, medical imaging, security and surveillance to name a few.

**Main trends**

Super Resolution has seen many applications not only in Image Resolution. [21] proposes CT image synthesis from MRI images using GAN for Medical Imaging purpose outperforming other state-of-the art methods such as SRCNN. SR has also been used for local motion estimation and pixel synthesis for Video Super Resolution[22]. Facial Super Resolution has found its own set of research domain with many papers employing GAN [26] and FSRNet[25] as solutions. Other imaging tasks such as removal of unwanted parts of image such as rain streak [23], image fusion [24], image restoration, and inpainting; have also been achieved using similar methods as that for Super Resolution.

**Key ideas**

Discussing about SISR, mainstream algorithms of SISR are mainly divided into three categories namely - interpolation-based methods, reconstruction-based methods and learning-based methods. Interpolation-based SISR methods, such as linear interpolation, bicubic interpolation [3] and Lanczos resampling [4], are very speedy and straightforward but they usually yield solutions with overly smooth textures thereby suffer from accuracy shortcomings. Reconstruction-based SR methods [5, 6], restrict the possible solution space by using sophisticated prior knowledge resulting in flexible and sharp details. However, the performance of many these methods degrades rapidly by upscaling factors, and these methods are usually time-consuming. Learning-based SISR methods employ machine learning algorithms such as neighbour embedding [7] utilizing similar local geometry between LR and HR, sparse signal recovery theory based sparce-coding [9, 10], random forest [8] and artificial neural networks; to analyse statistical relationships between the LR and its corresponding HR image from substantial training examples. Deep-learning based architectures have recently been widely adopted due to their demonstration of greater superiority over above-mentioned methods.

Deep learning algorithms aim to learn informative hierarchical representations automatically and then leverage them to achieve the ultimate purpose. Two multi-perceptron algorithms which have played a major role for SISR are convolutional neural network (CNN) [1,11] and recurrent neural network (RNN) [12]. With the progress of network architectures, deeper structures have been explored and achieved significant improvement. The structure of ResNet [13] is popularly applied to make deep networks. In VDSR [15], Kim et al. introduced skip connection into super-resolution and demonstrated that residual learning is more efficient than direct learning. Ledig et al. [13] proposed SRResNet which stacked residual blocks to build a deep network. Increasing the depth is beneficial for representation power but meanwhile under-use the feature information from shallow layers which represent low-level features. To deal with this issue, methods proposed are SRDenseNet [17], RDN [18], MemNet [19] that introduce various skip connections and concatenation operations between shallow layers and deep layers. Alongside the depth of the networks, increasing the width of the networks has also been experimented with the width of the networks such as seen in WDSR [20]. They also discuss about normalization of parameters suggesting that batch normalization is not suitable for training deep SR networks and introduce weight normalization for faster convergence and better accuracy. Most of these methods use Mean Square Error as loss function apart from GANs which combines Adversarial Loss. All of these methods utilize PSNR as an evaluation metric.

**Issues of SRCNN**

Dong et al. [1] proposed the first CNN based SISR method but the strategy incurred a huge training time because of the interpolation of LR image. FSRCNN [11] is a compact hourglass-shape CNN structure with deconvolution operations for reducing training time and improving the performance. Costly nonlinear mapping in SRCNN, is avoided by shrinking and expanding layer at the beginning and the end of the mapping layer separately to limit mapping in a low-dimensional feature space in FSRCNN [11]. SRCNN [1] also faced the problem of upscaling factors for which Ledig et al [13] proposed SRGAN capable of inferring photo-realistic natural images for 4× upscaling factors using a discriminative network which is trained to distinguish between super-resolution images and original photo-realistic images. In SRCNN [1], the experiments indicate "deeper is not better" but in DRCN [12], Kim et al. designed a convolutional network that models long-range pixel dependencies with limited capacity and the network recursively widens the receptive field without increasing model capacity. Increasing the number of layers is still a problem in SRCNN [1] but Kim et al.[12] proposed DRCN with upto 16 recursions. Having more than 16 recursions is still a challenge to be overcome to use image level context. It is also interesting to note from [1] that while larger filter size significantly improves the performance but however deployment speed also decreases. The methods mentioned above assume that low resolution images are down-sampled from high-resolution images given. In the real-world, however, low-resolution images are degraded more complicatedly. Therefore, the performance becomes poor when existing methods are directly used for solving real-world low-resolution image in practical. EDRN [14] introduced an encoder-decoder structure with coarse-to-fine methods. The encoder-decoder structure can extract features with more context information by the larger receptive field. The coarse-to-fine structure can gradually restore lost information and attenuate the effects of noise. To conclude, there are many interesting problems which have been solved under the domain of super-resolution and the issue of working with real-life low images from different sources is one of the interesting use cases which I would like to work with.

**References –**

1. Dong, Chao, et al. “Image super-resolution using deep convolutional networks” IEEE transactions on pattern analysis and machine intelligence 38.2 (2015)
2. M. Kawulok, P. Benecki, S. Piechaczek, K. Hrynczenko, D. Kostrzewa and J. Nalepa, "Deep learning for multiple-image super-resolution", arXiv:1903.00440, 2019
3. R. G. Keys, Cubic Convolution Interpolation for Digital Image Processing, IEEE Trans. Acoust., Speech, Signal Proces., Vol.29, No.6, Dec. 1981
4. B. Karch and R. C. Hardie, “Robust super-resolution by fusion of interpolated frames for color and grayscale images,” Frontiers Phys., vol. 3, no. 28, pp. 1–14, 2015.
5. S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, “SoftCuts: A soft edge smoothness prior for color image super-resolution,” IEEE Trans. Image Process., vol. 18, no. 5, pp. 969–981, May 2009.
6. J. Sun, J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1–8, 2008.
7. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: CVPR (2004).
8. S. Schulter, C. Leistner, and H. Bischof. Fast and accurate image upscaling with super-resolution forests. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
9. R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” in International conference on curves and surfaces, 2010.
10. J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
11. C. Dong, C. C. Loy, and X. Tang, “Accelerating the superresolution convolutional neural network,” in ECCV, 2016.
12. J. Kim, J. Kwon Lee, and K. Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in CVPR, 2016.
13. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, ´ A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang et al., “Photorealistic single image super-resolution using a generative adversarial network.”
14. G. Cheng, A. Matsune, Q. Li, L. Zhu, H. Zang, and S. Zhan. Encoder-decoder residual network for real super-resolution. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2019.
15. J. Kim, J. Kwon Lee, and K. Mu Lee, “Accurate image superresolution using very deep convolutional networks,” in CVPR, 2016.
16. B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in CVPRW, 2017.
17. T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in ICCV, 2017.
18. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in CVPR, 2018.
19. Y. Tai, J. Yang, X. Liu, and C. Xu, “Memnet: A persistent memory network for image restoration,” in CVPR, 2017.
20. Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xinchao Wang, and Thomas Huang. Wide activation for efficient and accurate image super-resolution. arXiv preprint arXiv:1808.08718, 2018.
21. D. Nie et al., ‘‘Medical image synthesis with context-aware generative adversarial networks,’’ in Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Intervent., 2017, pp. 417–425.
22. S. Li, F. He, B. Du, L. Zhang, Y. Xu, and D. Tao, “Fast spatiotemporal residual network for video super-resolution,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
23. X. Fu, J. Huang, D. Z. Y. Huang, X. Ding, and J. Paisley, “Removing rain from single images via a deep detail network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017.
24. Y. Liu, X. Chen, Z. Wang, Z.J. Wang, R.K. Ward, X. Wang , “Deep learning for pixel-level image fusion: Recent advances and future prospects” Inf. Fusion, 42 (2018), pp. 158-173.
25. Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, “FSRNet: End-to-end learning face super-resolution with facial priors,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018.
26. Bulat, A., Tzimiropoulos, G.: Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. arXiv preprint arXiv:1712.02765 (2017).