SRCNN Critical Analysis

Animesh Devendra Chourey – 210765551

Queen Mary University of London – February 2022

Super-Resolution Convolutional Neural Network

Image Super-resolution (SR) [15] refers to the task of restoring high-resolution (HR) images from one or more low resolution (LR) images of the similar scene. HR images from original images are a necessity in the field of computer vision for both pattern recognition and data analysis. It is an ill-posed problem as there exists several solutions and not a unique one. Deep learning has shown prominent superiority over other machine learning algorithms in many artificial domains, such as speech recognition [8], computer vision [11] and natural language processing [3]. Super-Resolution Convolutional Neural Network (SRCNN) [5] uses convolutional neural network (CNN) for single image SR which is a classical problem in computer vision. SRCNN learns an end-to-end mapping between LR and HR.

SRCNN proposes 3-layered CNN for SR. While constructing the architecture, each convolutional layer is interpretted in terms of traditional SR, and each layer is in charge of patch extraction, non-linear mapping and reconstruction respectively. The first layer creates feature maps from input images, the second one converts the feature maps into high dimensional feature vectors, and finally the last layer aggregates the feature maps to output final HR image. The LR image before being passed to the network is first upscaled to the desired size using bicubic interpolation [9]. SRCNN implements only the luminance components for training. The loss function for optimising SRCNN is mean square error (MSE). As for the sparse coding, the output being produced is of same resolution size as the input image which is our bicubic interpolated image. The overlapping patches in the reconstruction step are averaged instead of adding together with different weights by convolutions.

Main trends on the topic since the publication of the paper:-

The prospect of SR has led to explosive development of multi-frame SR processing. A novel fast spatio-temporal residual network (FSTRN) [13] adopts 3D convolutions for the video SR task in order to enhance the performance while maintaining a low computational load. SR has found applications in remote sensing where certain bandwidth limitations and pixel size restrictions are present; in security and surveillance and in medical imaging where desire to reduce irradiation dosage is paramount. Facial SR has also been implemented in various research domain with many papers incorporating GANs for finding solutions[1]. Super-resolution fluorescence microscopy (SRM) [4] methods have proven that it is possible to overcome the the hundred year old theoretical limit for the resolution potential of light microscopy which for decades has precluded a direct glimpse of the molecular machinery of life. Tasks such as removal of unwanted parts from image such as rain from single image [7], image fusion [14] also achieves it's desired results using methods similar to the SR.

Main problems solved/improvements over the original work:-

In SRCNN, the high computational cost still hinders it from practical usage that demands real-time performance. Fast Super-Resolution Convolutional Neural Networks (FSRCNN) [6], achieves a speed-up of more than 40× with even superior performance than the SRCNN. Here, the last convolution layer of SRCNN is replaced with a deconvolution layer. Also, the single mapping layer is replaced with the combination of a shrinking layer, 4 mapping layers and an expanding layer.

Even though SRCNN successfully introduced a deep learning technique into the SR problem, it still has limitations in three aspects: firstly it relies on the context of small image regions, secondly it's training converges too slowly, and lastly the network only works for a single scale Very Deep Super-Resolution (VDSR) [10] uses a relatively high initial learning rate to accelerate convergence and used gradient clipping to prevent the annoying gradient explosion problem. VDSR has made two more contributions. First one is that a single model is used for multiple scales and the second contribution is the residual learning.

MSE loss function is used in SRCNN.The restored output HR is learned with a high PSNR value, but sometimes produces blurry output. Therefore, a high PSNR value does not mean a clear image. Ledig applied Generative Adversarial Networks (GANs) to the SR field known as Super-Resolution Using a Generative Adversarial Networks (SRGANs) [12]. Here, generator network restores images and discriminator network that separates the ground truth and the output of the generator. It uses style transfer instead of existing MSE loss as well as GAN loss. Perceptual loss is proposed by using VGG loss as well. The experimental results suggest that PSNR value decreases, but it can produce a more plausible result for the human eye.

Classic single image SR, aims to enhance the resolution of bicubically degraded images, has recently obtained great success via deep learning. However, these existing methods do not perform well for real single image. An Encoder-Decoder Residual Network (EDRN) [2] adopts an encoder-decoder structure to encode highly effective features and embed the coarse-to-fine methods.The coarse-to-fine structure can gradually restore lost information and reduce noise effects. The encoder-decoder structure can extract features with more context information by the larger receptive field. Compared with state-of-the-art methods in classic single image super-resolution, EDRN can efficiently restore the corresponding high-resolution image from a degraded input image.

Remaining problems from the published works so far:-

There are various challenges that can be solved in the near future. The most significant issue can be that deep learning based algorithms requires many parameters and have a high computational load. Therefore, even more sophisticated model compression and optimization methods are needed to be apply these algorithms in the real world problems. Secondly, in addition to the bicubic downsampling method, the biggest task is to study a model that can perform well for images downsampled by the unknown method. Also, there are no unified and admitted metrics for assessing SR quality. Having a unified evaluation metric can also be done for the future works.

An unsolved problem on the topic most interesting to you to solve and why:-

It is difficult to collect images with different resolutions and even though unsupervised research work is underway, successful results from the research will allow us to train datasets without the need to pair LR-HR images. It will be a promising direction for me to solve and could prove highly beneficial in the future.

References

[1] Adrian Bulat and Georgios Tzimiropoulos. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–117, 2018.

[2] Guoan Cheng, Ai Matsune, Qiuyu Li, Leilei Zhu, Huaijuan Zang, and Shu Zhan. Encoder-decoder residual network for real super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[3] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167, 2008.

[4] Christoph Cremer and Udo Birk. Perspectives in super-resolved fluorescence microscopy: What comes next? *Frontiers in Physics*, 4:11, 2016.

[5] 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.

[6] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016.

[7] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3855–3863, 2017.

[8] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing magazine*, 29(6):82–97, 2012.

[9] Robert Keys. Cubic convolution interpolation for digital image processing. *IEEE transactions on acoustics, speech, and signal processing*, 29(6):1153–1160, 1981.

[10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[12] 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.

[13] Sheng Li, Fengxiang He, Bo Du, Lefei Zhang, Yonghao Xu, and Dacheng Tao. Fast spatio-temporal residual network for video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10522–10531, 2019.

[14] Yu Liu, Xun Chen, Zengfu Wang, Z Jane Wang, Rabab K Ward, and Xuesong Wang. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42:158–173, 2018.

[15] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.