Image Super-Resolution: A Critical Analysis Post SRCNN

220432052
t.d.joloko@se22.qmul.ac.uk

***Abstract*—Since the emergence of the SRCNN deep learning model for single image super-resolution, several other models that build on this work have been developed. These models seek to improve and address the limitations of SRCNN, and in general, other issues that may arise from using deep learning methods to solve computer vision problems. This report provides a critical analysis of the trends, challenges, and improvements in image super-resolution after the release of the SRCNN model.**

I. INTRODUCTION

Super resolution is the task of producing a high-resolution (HR) image from a specified low-resolution (LR) image. This is an ill-posed problem because numerous high-resolution images can exist for a single low-resolution image and likewise, numerous low-resolution images can exist for a high-resolution image [1]. Despite this inherent fact, researchers continue to make advancement in this field by providing approximate solutions, with accuracy measured using standardized assessment criterions.

One of such advancements is the Super Resolution Convolutional Neural Network (SRCNN) model [1] which pioneers the use of a deep convolutional neural network (CNN) for single image super-resolution. This method consists of 3 layers – a patch extraction and representation layer, a non-linear mapping layer and a reconstruction layer that are jointly optimized during the training phase where the network learns the mapping between the low-resolution input and its high-resolution counterpart, also used as the ground truth [1]. After training, this network can be used as a fully feed-forward network [1].

This model is widely used as a comparison benchmark [2], [3], [4], [5], [6], [8] for new novel techniques developed as its performance was superior to the previously existing methods at the time of publication. (who has used it as a benchmark)

The subsequent section of this report analyses the key works and trends following the SRCNN model and discusses the improvements and remaining problems on the topic of super-resolution using deep learning.

II. KEY WORKS AND TRENDS

The SRCNN model takes as input a low-resolution image gotten by down sampling the high-resolution ground truth image and then up sampling again via bicubic interpolation [1]. This pre-processing has been identified to increase the computational complexity of the model as input images with larger spatial dimensions can slow down the speed of reconstructing the new image [2], [3]. FSRCNN (Fast Super-Resolution Convolutional Neural Networks) and ESPCN (Efficient Sub-Pixel Convolutional Neural Network) were proposed by Dong et al. [2] and Shi et al. [3] to make SRCNN perform faster to allow for real time usage. Both methods substitute the bicubic interpolation layer with an upscaling layer at the end of the network. The former uses a deconvolution layer while the latter uses a sub-pixel aggregation convolution layer. Both methods allow for the low-resolution to high-resolution mapping to be done in the low-resolution space and directly learn an upscaling filter [1]. This allows the networks to be able to perform faster.

Though Dong et al in the original SRCNN paper proposed that the performance might not be improved by a deeper network, subsequent works in [4], [6], [8], [9] have been able to prove otherwise. Therefore, deep networks can provide better accuracy for the super-resolution. Despite this, there may be tradeoffs between accuracy and speed as very deep networks have been shown to be computationally slow.

Most deep learning methods mentioned in this report have also utilized supervised learning methods as the model training approach. An exception to this is the CinCGAN method [7] which uses an unsupervised learning approach to model scenarios where the actual down sampling factors that result in a LR image are unknown resulting in the unavailability of an LR-HR image combinations, and so no direct mapping to learn by the network.

III. DISCUSSION

Super-Resolution (SR) models are commonly evaluated using peak signal-to-noise ratio (PSNR), a pixel-wise comparison of two images and more recently the structural similarity index measure (SSIM). While maximizing the PSNR value shows improvement in a model's performance, it is unable to represent the perceptual quality of an SR image [6]. Ledig et al. [6] proposed a generative adversarial network (GAN) based model optimized for perceptual loss, rather than the commonly used Mean Squared Error (MSE) which only considers spatial positioning of pixels for optimization. This method by Ledig et al. is also the first model to use GANs for super-resolution and provides the best pho-realistic output since the introduction of SRCNN [8]. The generative model is trained to generate an HR image from a LR input, while the adversarial part of the network is trained to discriminate between real HR images and super-resolved HR images. Other works have been developed to improve SRGAN [8].

As there continues to be advancements in super-resolution techniques, it is still hard to find models that can jointly provide a high perceptual quality while meeting the requirements for real-time performance [8]. Thus, models that provide a high perceptual quality can be more suited to high-risk critical applications such as medical imaging and facial recognition, while negligible risk applications with real-time demands can handle little reductions in perceptual quality for quicker results.

For the deeper models [4], [6], [8], [9], there is no information of what these models are learning at the different layers. Models with few layers like [1], [2], [3], [5] provide insight into the models by describing what the model layers learns and represents. Therefore, it is yet to be seen as to whether the interpretability of what happens at each layer may help further help in improving accuracy and performance to provide information for fine-tuning.

VI. CONCLUSION

This report provided an overview of the trends and improvements following the introduction of deep learning based SRCNN model for image super-resolution. Different evaluation criterions were also discussed. Perceptual quality as an evaluation criterion for SR model performance should be a major consideration for future works because in the real-world applications, human ability to make deductions from the SR images produced is dependent on what they can perceive in the images. Apart from image super-resolution, some authors also propose that their methods can be useful for solving other image restoration tasks.

REFERENCES

[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] Dong, C., Loy, C.C. and Tang, X., 2016. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 391-407). Springer International Publishing.

[3] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1874-1883).

[4] 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).

[5] 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).

[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] Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C. and Lin, L., 2018. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 701-710).

[8] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. and Change Loy, C., 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0).

[9] Lim, B., Son, S., Kim, H., Nah, S. and Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE*

conference on computer vision and pattern recognition workshops (pp. 136-144).