Queen Mary, University of London
School of Electronic Engineering and Computer Science

CNN model for super-resolution

ECS795P - Deep Learning and Computer Vision

Mughees Asif | 180288337

Due: 17 February 2022

The deployment of deep convolutional neural networks (CNNs) to resolve the classical problem of single image super-resolution (SISR) has had a tangible and positive contribution towards the computer vision domain. The research methodology denoted as the Super-Resolution Convolutional Neural Network (SRCNN) architecture, involves the implementation of an end-to-end mapping between low- and high-resolution images that enables the optimisation of the pre- and post-processing steps. Salient features of SRCNN that augment with the sparse-coding-based method include a fully connected feed-forward network with a limited number of layers and filters, thereby, simplifying the overall architecture and providing adequate speed for online usage whilst preserving accuracy. Moreover, SRCNN can use three channels of colour concurrently to achieve enhanced performance and superior reconstruction of low-resolution images. However, the SRCNN model represents considerable challenges including sensitivity to hyperparameter tuning, initialisation of filter weights, and the constant vigilance required for the modulation of the learning rate being considered extraneous.

The most important aspects of research in SISR involve improvements in performance and output image quality via a myriad of techniques including deeper networks, modifying hyperparameters, using advanced loss functions, and developing ensemble methods. As mentioned by Yang et al., leveraging CNNs to enhance single-image resolution has given rise to various production-grade applications characterized by the usage of adaptive increments in the resolution and increasing the width and/or depth of the network [1]. These new techniques with the availability of large datasets, coupled with increased computing power are driving SISR into mainstream adoption.

Kim et al. showed the development of a deep neural network architecture by extending the design of Visual Geometry Group (VGG) nets and integrating multiple convolution layers with a high learning rate, thereby, improving accuracy and highlighting the efficient evaluation of larger image regions [2]. Additionally, the optimizer was changed from the stochastic implementation (Stochastic Gradient Descent) to Adam (Adaptive Moment Estimation) with gradient clipping to offset large gradient shifts (exploding gradient problem). Consequently, Kim et al. also showed a recursive approach applied to deep networks, although the training of the networks suffered from the vanishing gradient problem that obstructed convergence [3]. To solve this issue, residual blocks as shortcut connections designed to bypass several convolution layers were implemented.

To counteract the degradation of training accuracy with deep models, the usage of residual learning frameworks has enabled "... explicitly [reformulating] the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions" [4]. Moreover, Johnson et al. introduced the usage of perceptual loss functions, instead of per-pixel loss functions, to improve high-level feature extraction and enhance the quality of the output image [5]. In a similar vein, Ledig et al. defined a Super-Resolution Generative Adversarial Network (SRGAN) that enabled the recuperation of finer texture details, usually lost during the upscaling process, by using adversarial (perceptual similarity) and content loss (mean-opinion-score) functions as opposed to the vanilla Mean Squared Error (MSE) used in traditional SRCNNs [6]. The architecture of SRGAN also uses batch normalization for each block that exhibits increased stability in the learning process during training.

Lai et al. proposed high-quality reconstruction for SISR by using a pyramid model based on using sub-band high frequency residuals at each level, therefore, positively influencing the pre-processing by negating the requirement of bicubic interpolation [7]. As an extension to this research, Wang et al. designed a scalable GAN model that progressively improved the upscaling of the input images whilst labelling the learning process from easy to hard [8]. Thus, improving the upscaling done by Lai et al. by a factor of two for high upsampling factors.

Shi et al. developed a methodology to enable reconstruction of the input image before upscaling by employing feature extraction in the low resolution (LR) space, instead of performing the same operation in high resolution (HR) [9]. This negates the usage of bicubic interpolation where "... the [LR] input image is upscaled to the [HR] space using a single filter" [9], thereby, reducing the overall computational complexity and improving performance.

As generally agreed in the research community, the super-resolution problem is an inference problem which is ill-posed at the best of times. Indeed, research still needs to be conducted in this nascent domain where problems such as the development of networks capable of executing simultaneous upscaling factors and addressing issues related to image blurring. Typically, researchers constrain the solution space by using prior information such as exploiting the texture of the input image where similarities can be correlated, or by using a pre-trained network that has been trained on generic sample pairs. An array of solutions exists for these issues, yet, due to the undetermined nature of the base problem, no unique solution can be guaranteed.

An interesting topic that needs further research is the combined incorporation of operating channels such as RBG or YCbCr colour spaces [10]. Currently, "... operating (training or evaluation) on different color spaces or channels can make the evaluation results differ greatly (up to 4 dB)" [11]. Addressing this would expand the scope of multi-dimensional SRCNNs to produce greater colour depths whilst performing upscaling on the input image.

References

[1] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. *Deep Learning for Single Image Super-Resolution: A Brief Review*. 2019. DOI: 10.1109/TMM.2019.2919431 (page 1).

[2] Jiwon Kim, Jung Lee, and Kyoung Mu Lee. *Accurate Image Super-Resolution Using Very Deep Convolutional Networks*. Nov. 2015 (page 1).

[3] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. *Deeply-Recursive Convolutional Network for Image Super-Resolution*. 2016. arXiv: 1511.04491 [cs.CV] (page 1).

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition*. 2016. DOI: 10.1109/CVPR.2016.90 (page 1).

[5] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*. 2016. arXiv: 1603.08155 [cs.CV] (page 1).

[6] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*. 2017. arXiv: 1609.04802 [cs.CV] (page 1).

[7] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. *Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution*. 2017. arXiv: 1704.03915 [cs.CV] (page 2).

[8] Yifan Wang, Federico Perazzi, Brian McWilliams, Alexander Sorkine-Hornung, Olga Sorkine-Hornung, and Christopher Schroers. *A Fully Progressive Approach to Single-Image Super-Resolution*. 2018. arXiv: 1804.02900 [cs.CV] (page 2).

[9] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. *Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network*. 2016. arXiv: 1609.05158 [cs.CV] (page 2).

[10] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. *Deep Learning for Image Super-resolution: A Survey*. 2020. arXiv: 1902.06068 [cs.CV] (page 2).

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. *Image Super-Resolution Using Deep Convolutional Networks*. 2016. DOI: 10.1109/TPAMI.2015.2439281 (page 2).