Super resolution

Critical analysis

Khalid Salman

210469877

Deep Learning and Computer Vision

ECS795P

Contents

Super Resolution Critical Analysis

Super resolution (SR) is a classic problem in digital image processing in which we aim to recover a high resolution Image from a given low resolution image. This problem is ill-posed as many solutions exist for a single low resolution image. Many strategies where adopted to learn a prior that can be enforced to get the best solution from the solution space. These strategies can be broadly categorized to prediction models, edge-based models, example-based methods (patch based), image statistical methods. A key assumption that underlies many SR techniques is that an image contains many redundant pixels which means the low resolution image has almost all the information it needs to be converted to the high resolution space. Figure (1) summarises the problem of super resolution and the strategies used to solve it. This report will focus on critically analysing deep learning methods for single image super resolution (SISR). It will start by discussing the first CNN model for super resolution, then It will talk about different approaches that were proposed to enhance the performance. performance metrics explanation and abbreviations can be found in Appendix B.

One of the most successful approaches to the super resolution problem was the sparse coding approach presented in [1] and [2]. This approach uses a pipeline that starts with cropping and prepossessing patches from the input image, then encoding these patches using low-resolution dictionary, then learning a high resolution dictionary to reconstruct the patches, and finally aggregating overlapping constructed patches to get the final output. Inspired by the this idea , a CNN model was developed by Chao Dong et al [3]. In this work, it was argued that the proposed CNN model -named SRCNN-, is equivalent to the aforementioned pipeline in the sparse coding technique. However, the key difference is that the SRCNN model learns an end-to-end mapping between the low and high resolution pairs. Although this method was the first deep learning method to outperform sparse coding techniques, it had some problems that are worth mentioning. First, the SRCNN architecture consist of only three convolutional layers which result in a low receptive field (13*13), this means for a certain patch, only a small portion of the image will be considered as it's context. Second, a very low learning rate is used to guarantee the convergence of the network (learning rate $= 10^{-5}$), this makes the network converge very slowly during training. Third, although the SRCNN architecture can be trained on any image size, it can only produce a single scale. Finally, the network needs prepossessing by up-sampling the image from the low resolution space to the high resolution space, this increases the computation cost making it difficult to use for real time application.

After SRCNN was introduced, attention was drawn towards CNNs and deep learning. Wenzhi Shi et al [4] suggested performing super resolution in the low resolution image space and then use sub-pixel convolution layer that learns filters to upscale the image. This way they avoid working on the high resolution space which makes the network more computationally efficient than SRCNN. The proposed network was fast enough to perform 1080p video super resolution using a single GPU. A similar idea to enhance SRCNN's efficiency was presented in [5] where an architecture named FSRCNN was introduced. This architecture also takes as an input the low resolution image in the low resolution space, but this time a deconvolution layer was used to upscale the image.

To address the problem of small receptive field in SRCNN, a very deep convolution network with 20 layers was introduced in [6] (VDSR). More convolutional layers, means higher receptive field, however, it also means more parameters (possibility of over-fitting) and the possibility of vanishing/exploding gradient. Gradient clipping was used to deal with the vanishing/exploding gradient. This work also solved the slow training problem in SRCNN by using a high learning rate and making the network learn to map the low resolution image to the residual image (the difference between the high and low resolution image) instead of mapping immediately to the high resolution image. VDSR needed only 4 hours of training to reach state-of-the-art performance. Another important feature of VDSR is that it can work for different scales. What I find very interesting is that training the network with multiple scales boosts the network performance - one might expect more tasks will lead to worse performance, but in fact different scales where helping each other to achieve an even better performance. Shortly after the introduction of VDSR, a deep recursive convolutional network (DRCN) was developed in [7]. The main idea behind this work is that VDSR has many identical convolutional layers that can be replaced using recursion to a single convolutional layer. This reduces the number of parameters significantly which makes the network less prone to over-fitting. It was able to achieve a high receptive field (41*41 comparing to 13*13 in SRCNN). Both VDSR and DRCN outperformed SRCNN in all performance metrics, which is expected because of the higher receptive field captured by deeper networks. An even deeper CNN network was proposed to perform the task of super resolution. [8] proposed a deep recursive residual network (DRRN), that uses both residual training and recursion (similar to DRCN) but with 52 layers network and skip connections betweem ResBlocks [9] to mitigate the difficulty of training such a deep network.

An idea that contributed notably to the performance of SR deep learning methods was the channel attention mechanism [10]. The main concept was to exploit the interdependence and interaction of the feature representation between different channels. Zhang et al. [11] introduced an architecture names RCAN that Incorporates the channel attention mechanism. This architecture gave superior results both in term of PSNR and SSIM, in fact RCAN currently achieves best results in those two metrics. Building on this work , Dai et al [12] proposed a novel architecture named SAN. This architecture uses a second order channel attention (SOCA). SOCA was used to adaptively re-scale channel wise features by using second order feature statistics to enable extracting more informative information.

A very interesting method was proposed in [13] where a generative adversarial network -named SRGAN- was used to get astonishing visual results. In this work, features were extracted from the low resolution image using a pre-trained VGG network [14], the VGG space loss was added to the adversarial loss to improve fine texture details. Although this approach gave poor results in terms of peak-signal-to-nose-ratio, it gave superior results in terms of how realistic the image is (mean opinion score). What is more astonishing is how the SRGAN was able to produce results nearly indistinguishable from the original image even at 16x scale. However, there was a downside to this approach, the network is very bad at reconstructing texts and numbers. This is not surprising because the network was trained on ImgaeNet [15] and hence it never saw texts or numbers.

In summary, almost every state of the art solution uses a deep CNN model with some measures to mitigate the difficulty of training such a deep model. Cleaver techniques where used to make deep CNNs learn better features, these technique include skip-connections and residual blocks, recursion to reduce parameters, channel attention mechanism to exploit interdependence between channels, and other methods. I find the SRGAN particularly interesting because of its outstanding visual results even for 16x scaling. Although the SRGAN gives very promising results, it fails in reconstructing numbers and texts, this means that this methods cannot be used in surveillance cameras which is one of the most important applications for super resolution. It would be interesting to explore ways to address this problem.

Appendix A

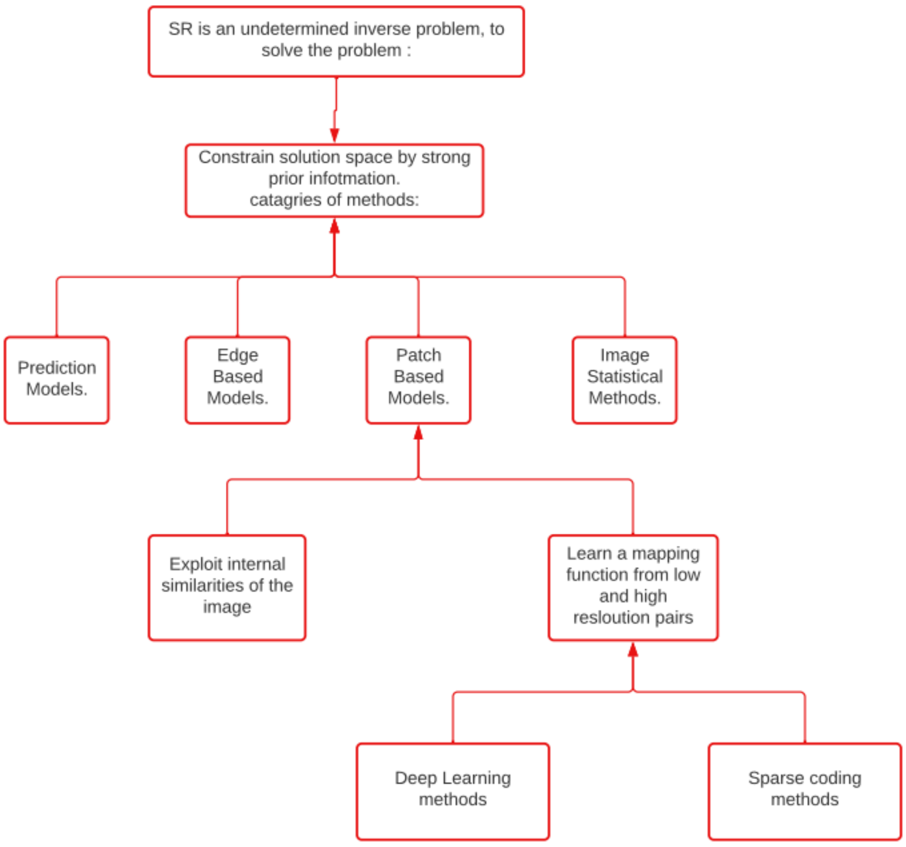
SR is an undetermined inverse problem, to solve the problem :
Constrain solution space by strong prior infotmation. catagries of methods:
Prediction Models.
Edge Based Models.
Patch Based Models.
Image Statistical Methods.
Exploit internal similarities of the image
Learn a mapping function from low and high resloution pairs
Deep Learning methods
Sparse coding methods

Figure 1: The problem with super resolution and the strategies used to solve it

Appendix B

Performance metrics:

- **peak signal to noise ratio**
  Inversely proportional to the logarithm of the Mean Squared Error (MSE) between the high resolution image and the generated image.
- **structural similarity index measure**
  Is a measure of similarity between the high resolution image and the generated image.
- **mean opinion score**
  Number of judges are asked to rate an image from 1 to 5 and the mean of all judgments is taken.

Abbreviations:

- PSNR: Peak signal-to-noise ratio.
- SSIM: Structural similarity index measure.
- SR: Super resolution.
- SISR: Single image super resolution.
- CNN: Convolutional neural network.
- SRCNN: Super resolution convolution neural network.
- FSRCNN: Fast Super resolution convolution neural network.
- VDSR : Very Deep Super Resolution.
- DRCN : Deep Recursive Convolutional Network.
- DRRN : Deep Recursive Residual Network.
- RCAN : Residual Channel Attention Networks.
- SOCA : Second Order Channel Attention.
- SAN : Second Order Attention Network.
- SRGAN : Super Resolution Generative Adverserial Network.

References

[1] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches", English (US), in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, ser. 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR ; Conference date: 23-06-2008 Through 28-06-2008, 2008, ISBN: 9781424422432. DOI: 10.1109/CVPR.2008.4587647.

[2] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation", *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010. DOI: 10.1109/TIP.2010.2050625.

[3] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016. DOI: 10.1109/TPAMI.2015.2439281.

[4] W. Shi, J. Caballero, F. Huszár, *et al.*, "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*, 2016, pp. 1874–1883.

[5] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network", in *European conference on computer vision*, Springer, 2016, pp. 391–407.

[6] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[7] ——, "Deeply-recursive convolutional network for image super-resolution", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645.

[8] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147–3155.

[9] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, 2015. arXiv: 1512.03385 [cs.CV].

[10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[11] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, *Image super-resolution using very deep residual channel attention networks*, 2018. arXiv: 1807.02758 [cs.CV].

[12] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution", in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 057–11 066. DOI: 10.1109/CVPR.2019.01132.

[13] C. Ledig, L. Theis, F. Huszár, *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*, 2017, pp. 4681–4690.

[14] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, 2015. arXiv: 1409.1556 [cs.CV].

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database”, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.