A classical problem in computer vision is single image super-resolution that aims to recuperate a high-resolution image from a low-resolution image [1]. Super-resolution has gained attention because of its usefulness in various fields like medical imaging, elevating gaming experience, capturing images with ease without the need for sophisticated expensive cameras to meteorological satellite image and security surveillance to name a few. SRCNN, was a first successful attempt for using convolutional layers in deep learning network architecture. In this critical analysis report, trends post this attempt in the field of super-resolution task is covered.

The main trends since the SRCNN was inclusion of deconvolutional layer for up sampling [3], expansion of channels of the output features to increase resolution by adding a sub-pixel convolution layer [4], and encoder-decoder composed of multiple symmetrically linked convolutional and deconvolutions with skip layer connections [5]. Recursive networks (which utilizes same convolution kernel in non-linear mapping) with supervision to eliminate vanishing/exploding gradient along with skip connection to save network capacity in order to store recursive input signal [6] was innovative. Skip connection back-propagate to bottom layers and pass image details to top layers, making training of end-to-end mapping easier and effective [5]. Another advancement is residual learning to learn the mapping between bicubic HR image, accelerating convergence; along with usage of gradient clipping to speed up training [4]. In order to make full use of the hierarchical features due to multiple objects in the images, residual networks were appreciated; these algorithms learn the residue of high frequencies between the input and ground-truth [2]. This was also followed by combination with densely connected networks [7]. An interesting progress of densely connected networks is the deep back projection, which includes mutually connected up and down sampling stages to learn non-linear relation between HR and LR image with iterative error feedback mechanism for rectification followed by reconstruction of image [8]. A few other trends were progressive reconstruction design to deal with large scale factors by predicting outputs in multiple phases [2] and fusion of individually trained CNN that constructs mapping from HR to LR spaces and finally sums each output [9]. Furthermore, attention based networks [2] that selectively attends only few features at a given layer [10] showed significant improvement for SR. Multiple degradation handling networks like zero-shot super resolution uses internal image statistics with deep neural networks and aims to predict the test image from LR image initially created from the test image and after training predicts SR image using the test image [11]. A more recent area is GAN models that focus on the realistic perception of the input image using a discriminator trained to differentiate true HR image or a generated one and assists the generator to generate high- frequency structural features rather than noisy artifacts [14]. GAN based models are perceptually driven and aim to enhance the visual quality of generated outputs [2]. A variation of the SRGAN is using orthogonal projection in the output layer of generator to facilitate learning high frequency components while keeping the original low frequency components [13].

The key ideas of the related published works include transposed convolution layer, also called deconvolutional layer [7] that involves post-up sampling without interpolation instead of pre-up sampling that can be computationally expensive and learning from original LR image accelerates the process [3]. Interpolation by padding the subpixels with zeros simplifies the deconvolution layer [4] and enlarges LR images at various magnification ratios and removes the bicubic interpolation used in input layer of SRCNN. Residual learning focuses on addition of skip connections, namely, global connections, local connections, recursive connections, and dense connections [2], increasing the number of convolutional layers and avoiding vanishing gradient. In some models, batch normalization is removed as they decrease flexibility of CNNs and consume memory, tackled by parameter sharing. Depth of the networks' architectures to build even deeper neural networks improved the image quality. Selection units was introduced, a cascade connection of ReLU, followed by 1 x 1 convolutional filter and sigmoid activation [10]. Additionally, various parameter settings of kernel size, usage of activation functions like leaky ReLU and parameterized ReLU, variations in loss function from mean squared error l2 losses (that focusses on erroneous predictions) to absolute mean error loss l1 (considers balanced error distribution), along with perceptual loss considered by GANs [2], and parallelization of the networks were some more ideas that encompassed the published works. ResNet architecture forms the basis of SRGANs that includes many residual blocks composed of two convolution layers, two batch normalization layers, and one rectified linear unit (ReLU). Some more improvements are: better feature representations [15], like hierarchical feature representations to learn the localized patterns, deep Laplacian pyramid networks [16] focusing on learning multiple SR scales to obtain distinct features [12] and zero-resolution, an unsupervised learning that learns SR model on down sampled version of the given image with aim to learn internal image statistics.

The methodologies discussed above overcome some of the inherent issues of SRCNN as follows: (i) SRCNN up samples the LR image to match the ground truth HR image with Bicubic interpolation, which is computationally expensive [2], and is slow in processing, (ii) has a shallow network architecture that can be slow in convergence, (iii) lacks the capability to process images in real-time, (iv) considers all spatial locations (v) ignores hierarchical features.

As mentioned in the paper, super-resolution is inherently ill-posed due to the lack of unified solution [1]. In spite of various methodologies and innovations there exists multiple trailing problems mentioned in various published work: (i) sub-optimal results for newer domains like medical imaging for which large training data is not widely available, (ii) choosing the loss or objective function: l1 and l2 loss is used commonly which does not consider visual perception and hence lack of universally accepted perceptual metric to evaluate image quality for human perception, (iii) uunsupervised methods will fail if the input image itself is of very poor resolution, (iv) networks trained on artificially created degradations do not generalize well to actual LR images in practical scenarios [17], (v) implementing extreme super resolution like higher than 8x is challenging and (vi) computational complexity still persists for high performing models.
The most interesting and significant aforementioned problems might be the implementation of extreme super resolution. With more than ever video conferencing usage and lack of social contact, there is a cry for enhanced real-time super resolution for better work, medical and educational experience. Furthermore, it can be advantageous for historians to reconstruct ancient manuscripts and documents, that might be key to many unanswered questions in the history of human civilization, and useful to transfer those knowledges for the current era to have better future of living beings on earth.

REFERENCES

[1] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE, 2015.

[2] Anwar, S., Khan, S. and Barnes, N., A Deep Journey into Super-resolution: A survey. arXiv 2019. arXiv preprint arXiv:1904.07523.

[3] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in ECCV, 2016.

[4] Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.H. and Liao, Q., 2019. Deep learning for single image super-resolution: A brief review. IEEE.

[5] X. Mao, C. Shen, and Y.-B. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” in NIPS, 2016.

[6] J. Kim, J. Kwon Lee, and K. Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[7] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in CVPR, 2018.

[8] M. Haris, G. Shakhnarovich, and N. Ukita, “Deep backprojection networks for super-resolution,” in CVPR, 2018.

[9] H. Ren, M. El-Khamy, and J. Lee, “Image super resolution based on fusing multiple convolution neural networks,” in CVPRW, 2017.

[10] J. Choi and M. Kim, “A deep convolutional neural network with selection units for super-resolution,” in CVPRW, 2017.

[11] A. Shocher, N. Cohen, and M. Irani, “Zero-shot super-resolution using deep internal learning,” CVPR, 2018.

[12] S. Anwar and N. Barnes, “Densely residual laplacian super- resolution,” arXiv, 2019.

[13] H. Yamamoto, D. Kitahara and A. Hirabayashi, "Image Super-Resolution via Generative Adversarial Network Using an Orthogonal Projection," 2020 28th European Signal Processing Conference (EUSIPCO.

[14] S.-J. Park, H. Son, S. Cho, K.-S. Hong, and S. Lee, “Srfeat: Single image super-resolution with feature discrimination,” in ECCV, 2018.

[15] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in ICCV, 2017.

[16] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, “Fast and accurate image super-resolution with deep laplacian pyramid networks,” TPAMI, 2018.

[17] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, “Learning enriched features for real image restoration and enhancement,” arXiv preprint 2003.06792, 2019.