

Extensive Survey on Image Compression using Deep Learning

Mr. Anish U. Nagarsenker¹, Dr. Prasad D. Khandekar² and Mrs. Minal Deshmukh³

¹⁻²School of Electronics and Communication Engineering, Dr. Vishwanath Karad MIT World Peace University, Pune, India

Email: anishnagarsenker@gmail.com, Dean.engg@mitwpu.edu.in

³Vishwakarma Institute of Information and Technology, Pune, India
Email: minal.deshmukh@viit.ac.in

Abstract—Nowadays, Data Compression is adopted because it leads to pruning transmission bandwidth, storage hardware and data communication time. Deep Neural Networks (DNN) have confirmed exquisite achievements in several tasks such as Image Classification, Image Processing etc. Excellent performances with speech, audio or image input signals differentiates Convolutional Neural Networks (CNN) from other neural networks. Enhancing the video and lossy image compression visual quality is treated as a serious problem, but the modern advances in computing power along with the accessibility of sizeable training datasets has led to the usage of deep learning networks like CNNs to deal with image processing and image recognition tasks. Some of the image and video compression techniques (lossy as well as lossless techniques) include JPEG, JPEG2000, BPG, GIF, PNG, MPEG etc. Data Compression along with DNN ensures that there is not much loss in data, size is reduced and the accuracy also is preserved. In the following section, a comparison of various papers related to data compression and image compression using Deep Learning Techniques like CNN is provided along with suggestions for topics that can be taken up in the near future.

Index Terms— Data Compression, DNN, CNN, JPEG, MPEG, BPG.

I. INTRODUCTION

The mechanism of converting, customizing, or encrypting the bits format of data such that it exhausts limited storage on disk is called as Data Compression. It eliminates and changes the monotonous data characters and symbols to lower the amount of data, thus aids in clearing disk storage and leads to reduction of bit rate. Data Compression leads to pruning transmission bandwidth, storage hardware and data communication time. Data compression can be of the type lossy or lossless where lossy compression deletes unwanted data and decreases the complexity of actual data to lower its size whereas lossless compression eliminates bits by finding and discarding analytical redundancies to lower the size. Lossy compression causes notable trimming of file size with significant loss in quality and hence, it cannot be used in cases where image quality is a priority. Lossless compression causes a diminutive reduction in file size with no loss in quality but, this method is not convenient to compress documents of enormous sizes[2].

The major objective of compression of image relates to transferring or storing the image at reduced bit rates by minimizing its superfluity and irrelevance. Quality of image compression is validated using various parameters like compression ratio, PSNR (Peak Signal to Noise Ratio), MSE (Mean Squared Error), SSIM (Structural

Similarity Index Measure) etc. Compression ratio gives information about how much space is saved as compared to the original image size whereas PSNR, MSE and SSIM give information about how identical the compressed image is as compared to the original image. The compression ratio is given as the ratio of the size of compressed image to the size of the original image[1]. Some of the image and video compression techniques (lossy as well as lossless techniques) include JPEG, JPEG2000, BPG, GIF, PNG, MPEG etc. Figure 1 shows image compression process using JPEG image codec[5]. The correlation between the maximal achievable power of signal and the distorted noise power which influences the nature of its representation is called as PSNR. The aggregate discrepancy of all the picture elements in the image is referred to as MSE. An intuitive grade which evaluates deterioration of picture quality caused as a result of processing is called as SSIM[4]. High values of compression ratio along with high PSNR, SSIM and low MSE values for a particular image, indicates that the output image obtained after compression is optimal for use and is almost identical to the input image. The value of PSNR may vary between 0 and ∞ , while the value of SSIM varies from 0 to 1 where high values of PSNR and SSIM indicate that the compressed image(output) is almost identical to the original image. Ideal value in the case of MSE is 0 as it indicates the absence of error.

An extensive field of computer science dealing with the construction of some agile machines accomplished to execute functions that mainly need human intelligence is referred to as Artificial Intelligence(AI). AI processes enormous proportions of information rapidly and forms predictions much more accurately as compared to human

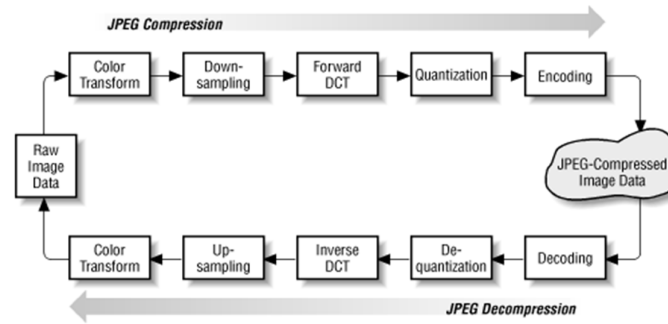


Figure 1: JPEG Block Diagram

processing. Some common advantages of AI include decreased time for large data processing, consistent accurate results and quality for meticulous jobs while the disadvantages include absence of capability to hypothesize between different tasks, needs strong technical proficiency and is a bit expensive. A class of AI which deals with software-based applications and makes them more accurate at forecasting consequences without even being programmed to do so, is Machine Learning(ML).It forecasts the fresh output values by utilizing historical statistics as input. Deep Learning is a subset of AI and ML, which includes predictive simulation and statistics[6]. It is fundamentally a neural network with multiple layers to increasingly extricate more advanced qualities from the input, where the learnings can be semi-supervised, unsupervised or supervised based on the applications at use and comprises predictive modelling and statistics. Deep Learning makes the task of accumulating, scrutinizing and interpreting substantial volumes of data significantly faster and easier[3]. Figure 2 explains the interconnection of AI, ML and Data Science where AI and Data Science are 2 separate sets and ML(Machine Learning) is a subset of AI(Artificial Intelligence) as well as Data Science containing neural networks,search,evolutionary algorithm and reinforcement learning.

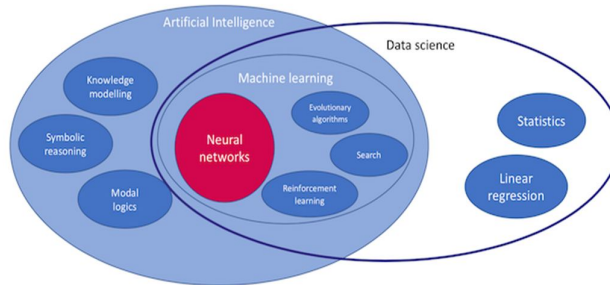


Figure 2:Relation between AI,ML and Data Science

A class of linearized artificial neural network employed in image processing and recognition especially made to measure pixel information is called as Convolutional Neural Network(CNN). CNN is basically an algorithm of Deep Learning that takes in the image, designates significance(learnable biases and weights) to numerous entities/features in the image and is capable to differentiate various images based on this as shown in figure 2. In Figure 3, a zebra is given as the input image and based on various feature extractions, output is predicted as being a horse, zebra or dog using SoftMax activation function with zebra having the highest output value of 0.7 indicating that there is a 70% probability that the image is that of a zebra. In contrast to other classification techniques, the preliminary processing in ConvNet is comparatively lower and ConvNets possess the capacity to master the attributes/filters. The CNNs provide high-resolution execution and increased compression artifact reduction in comparison with the conventional models of computer vision. The CNNs enhance the image grades like PSNR and SSIM which is a crucial part during the image compression stage[1].

In this paper, various research papers related to Compression using Deep Learning(CNN) are studied and reviewed. Some papers are related to Data compression using Deep Learning, while most of them are related to Image Compression using CNN. In the papers related to Data compression, parameters like compression ratios are used for comparison whereas in papers related to Image compression, parameters like PSNR and SSIM are used for comparison in addition to the compression ratios. The preceding sections include Literature review where analysis of 6 papers related to data compression using AI is done using different parameters like Compression Ratio, PSNR, SSIM etc followed by conclusions and references.

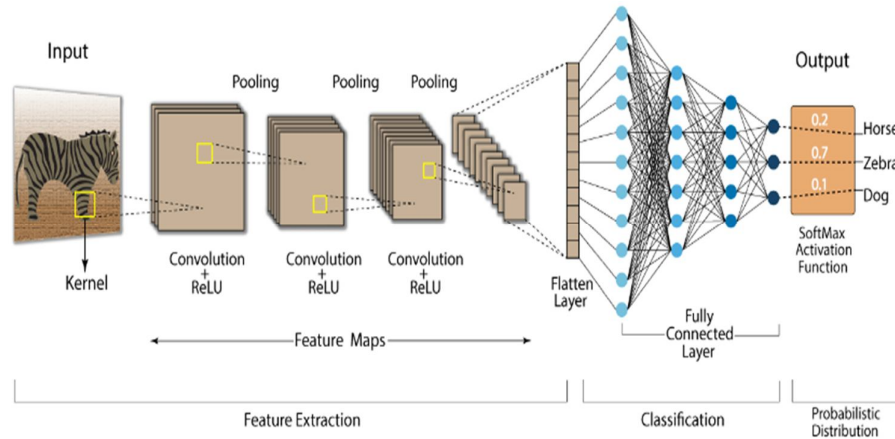


Figure 3: Working of CNN for identification of Zebra

II. LITERATURE REVIEW

In [1], an innovative compression framework is suggested based upon CNNs where 2 CNNs i.e ComCNN and RecCNN are ideally consolidated into a compression framework to achieve image compression of highest

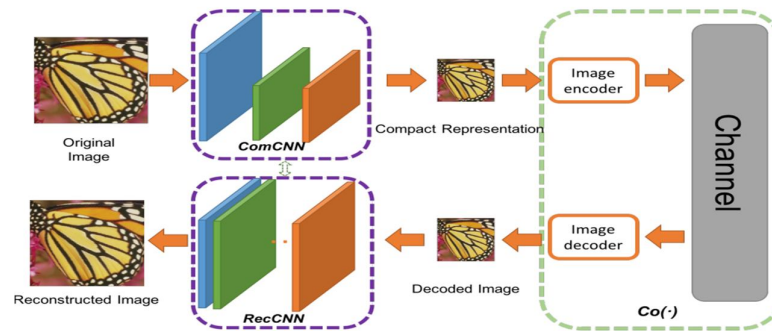


Figure 4: Paper 1 Architecture

standards at comparatively lower bit rates. Concurrently training ComCNN and RecCNN eases precise reconstruction of decrypted image using RecCNN. The ComCNN consists of 3 layers while the RecCNN consists of 20 layers and both CNNs are trained for 50 epochs each using 400 images of dimensions 180x180. It is also seen from the experimentation that minor changes in performances are observed for sizeable datasets of training. Post-Processing methods for removal of blocks and noise are also employed for superior image quality post compression. This framework achieves excellent performance for QF=5, while, for QF=10, the framework achieves a good performance but is outperformed by some post-processing methods like Zhang's and ARCNN for specific test images. This framework is compared with other post-processing and denoising methods using 6 test images and achieves an average PSNR and SSIM of 28.83 and 0.8535 respectively at QF=5 and an average PSNR and SSIM of 30.59 and 0.8895 respectively at QF=10. The run time needed in this framework is 0.017s in GPU and 1.56s in CPU, which is quicker than rest of the post-processing methods. The proposed framework maintains more of the higher frequency data plus retrieves honed edges and solid surfaces in the reformed picture. Datasets like General-100, Set14, Set5 and LIVE1 are used to exhibit the efficacy of the framework proposed. The framework is also compared with various image codecs like JPEG, JPEG2000 and BPG at different bpp(bits per pixel) and it is found that the framework significantly outperforms these image codecs when compared using PSNR and SSIM values. The architecture of this framework is shown in Figure 4.

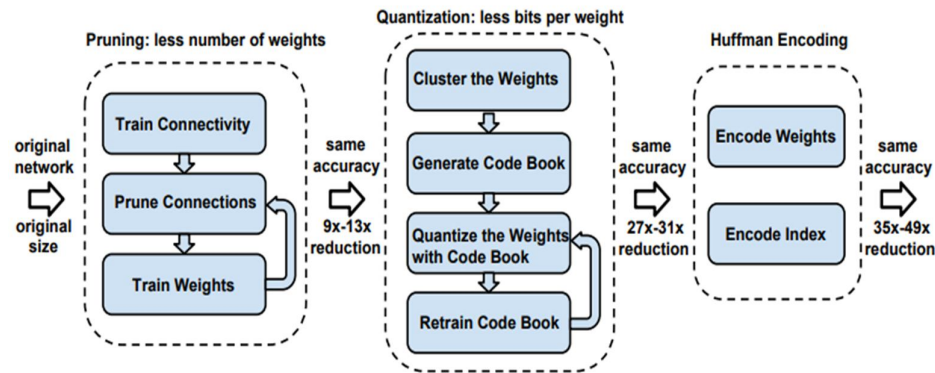


Figure 5: Paper 2 Architecture

In [2], Deep compression technique is used where a 3-phase channel consisting of trimming(eliminates the redundant networks), quantisation(various networks use equivalent weights, therefore only indices and codebook require to be reserved) and Huffman coding(to extract advantage of biased dissemination of efficient weights) is used for lowering the data storage needs of neural nets by 35x-49x with no harm to accuracy. Trimming and quantisation are capable of compacting the network without impeding one another, therefore result in higher compression of data. The trimming phase leads to 9x-13x data compression, the quantization phase leads to 27x-31x data compression while the Huffman coding phase(final phase) leads to 35x-49x data compression, thus reducing the storage size. This technique attempts to run deep neural networks on mobile owing to its advantages like real-time processing, better privacy and less network bandwidth. This framework reduces the storage consumption such that the data can be stored on the on-chip SRAM instead of the off-chip DRAM which significantly reduces the power consumption to as low as 5pJ/access. This compressed network also has 3x to 7x better energy efficiency as well as 3x to 4x layer wise speedup, being benchmarked on CPU, GPU and mobile GPU. This structure also eases the usage of compound neural networks in mobiles where download bandwidth and application storage are repressed. VGG-16, LeNet-5, AlexNet and LeNet-300-100 are the dataset models used in this paper for experimentation and training is carried out utilizing the Caffe framework.

In [3], in order to enfold the kind of deep stream data process mechanism of DNN architecture, an image compression framework named DeepN-JPEG is developed. It is found that DeepN-JPEG is able to achieve a compression rate enhancement of 3.5x and absorbs approximately 30% power of that needed for the typical JPEG techniques. As compared with HF-3 and SAME-Q4, DeepN-JPEG is able to achieve 2x and 3x power reduction respectively because of more efficient data compression. The utilization of Piece-wise Linear plot also decreases the quantization error based on the reformed quantization table and frequency component analysis, hence improves the compression amount without any precision depravity. DeepN-JPEG provides the finest compression ratios of 3.5x, conserving the analogous high precision similar to that of the indigenous datasets, thus lowering the data traffic price and edge mechanism storage for assignments of deep learning.

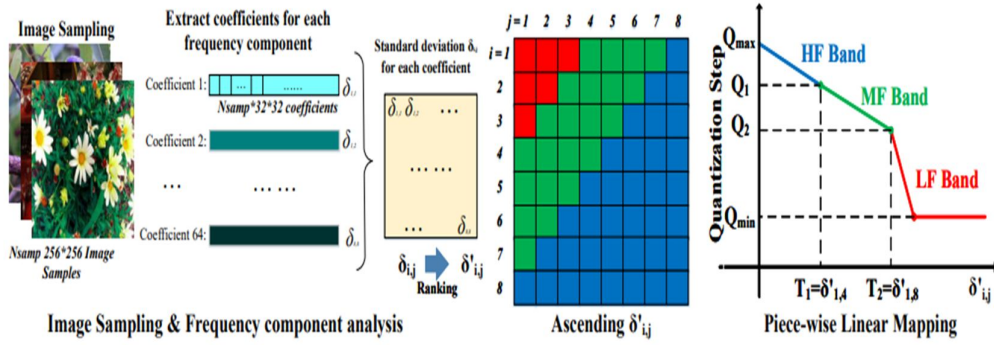


Figure 6: Working of DeepN-JPEG

In [4], higher visual quality is achieved through the use of CNN in lossy image compression, custom-fitted for the explicit task of semantic image perception. Multiple semantic areas can be identified in the image by designing and training a model which makes JPEG content-aware. In this technique, CNN-based image compression is obtained, which shows semantically-salient areas in a given image, through the use of a heat map(MS-ROI), so that those areas can be encrypted at a greater calibre than the background areas. PSNR and SSIM values were very high when this technique was applied on a given image in the testing dataset indicating good reconstruction image quality. Some of the MS-ROI heatmap examples are shown in figure 3 where objects in the image are identified and heat map is applied and these parts are then encrypted at a higher quality compared to other regions of the image. Caltech-256 dataset is used in the experimentation process containing 256 pictures of various objects, buildings, animals and plants etc thus providing a quality balance for training. The testing dataset consists of MIT dataset and Kodak PhotoCD set.



Figure 7: MSROI Heat Map Examples

In [5], by combining traditional image codecs and deep learning, an enhanced composite stacked image compression structure is presented. Convolutional neural network(CNN) is employed as the deep learning technique and image codecs used are FLIF(lossless), H.265 std or HEVC-based BPG. The values of MS-SSIM and PSNR in this technique are found to be better than some other image codecs over a sizeable domain of bit rates. The advanced deep structured learning-based stacked coding technique and conventional formats which include BPG in PSNR as well as MS-SSIM measure are outperformed by the suggested technique in this paper. The datasets utilized for experimentation are Tecnick and Kodak and the results obtained significantly outperforms deep learning-related stratified coding technique as well as the conventional codecs inclusive of BPG in MS-SSIM as well as PSNR standards over an extensive domain of bit rates, when the pictures are encrypted in RGB444 domain. Figure 8 shows the proposed block diagram for compression of image undergoing encoding and decoding to obtain image of high quality with significant compression.

In [6], 16-bit medical imaging volumes are compressed by using a novel lossless compression method. In the experiments performed, 2 confined sampling techniques were launched and assessed on actual volumetric datasets of medical imaging. The contrast with the indigenous data values are reduced and the excess are

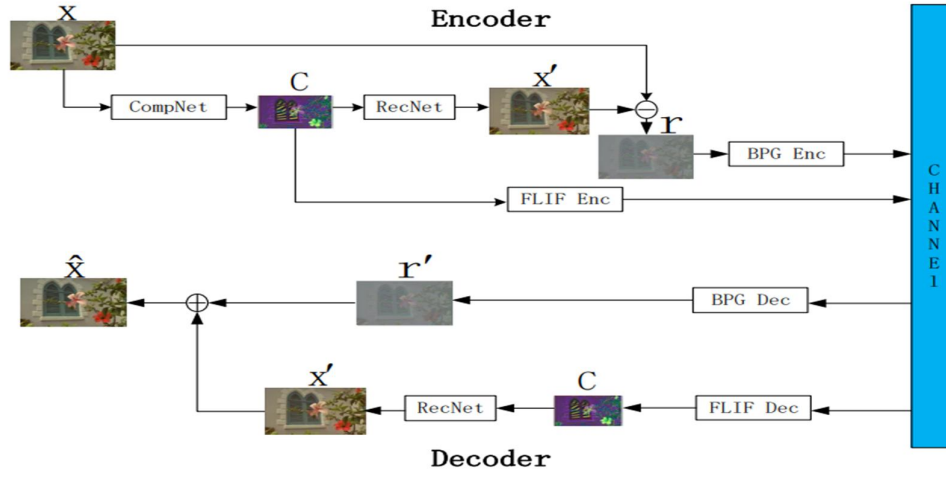


Figure 8: Paper 5 Proposed Block Diagram

compressed with the help of arithmetic coding when a neural network is trained using 3D data predictor. This technique is able to achieve a greater compression relationship as compared to JPEG2000, HEVC, JPEG-LS and JP3D. The performance remains satisfactory when the proposed method is generalized to unseen dataset. The generalization for this prototype to attain greater compression is having consistent data and structure identical to training set. Each and every dataset utilized in the experiments comprises of a group of DI-COM files that consists of patient's CT scans reserved as grayscale images of 16-bit. The training of prototypes was done using one training set while the testing was supervised on 2 distinct test sets.

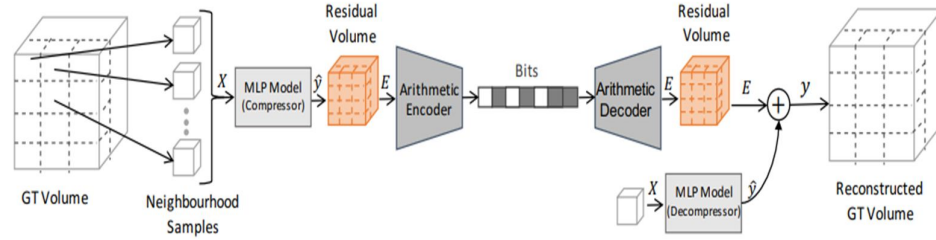


Figure 9: Paper 6 Proposed Block Diagram

III. CONCLUSION

In this paper, various research papers related to Compression using Deep Learning(CNN) are studied and reviewed. Some papers are related to Data compression using Deep Learning, while most of them are related to Image Compression using CNN. Data Compression leads to pruning transmission bandwidth, storage hardware and data communication time. The major objective of compression of image relates to transferring or storing the image at reduced bit rates by minimizing its superfluity and irrelevance. In the papers related to Data compression, parameters like compression ratios are used for comparison whereas in papers related to Image compression, parameters like PSNR, SSIM and MSE are used for comparison in addition to the compression ratios. It is observed that the Deep Learning Algorithms like CNNs play a crucial role and hence ensures reduced data loss during compression in comparison to methods that do not employ it. Different architectures and combinations of CNNs can be used for image compression to obtain compressed images which are almost similar to the original image as can be depicted by parameters like PSNR, SSIM, MSE etc. Some of the real-life applications where image compression is utilized includes Teleconferencing systems, Television broadcasting, Computer communications, Military communications via sonar and radar, Magnetic Resonance Imaging(MRI), CT scan etc.

REFERENCES

- [1] Feng Jiang, Wen Tao, Shaohui Liu, Jie Ren, Xun Guo, Debin Zhao, "An End-to-End Compression Framework based on Convolutional Neural Networks" 2017 IEEE Transactions on Circuits and Systems for Video Technology
- [2] Song Han, Huizi Mao, William J. Dally, "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding" 2016 Conference Paper at ICLR
- [3] Zihao Liu, Tao Liu, Wujie Wen, Lei Jiang, Jie Xu, Yanzhi Wang, Gang Quan, "DeepN-JPEG: A Deep Neural Network Favorable JPEG-based Image Compression Framework" 2018
- [4] Aaditya Prakash, Nick Moran, Solomon Garber, Antonella DiLillo, James Storer, "Semantic Perceptual Image Compression using Deep Convolution Networks" 2017
- [5] Haisheng Fu, Feng Liang, Bo Lei, Nai Bian, Qian Zhang, Mohammad Akbari, Jie Liang, Chengjie Tu, "Improved Hybrid Layered Image Compression using Deep Learning and Traditional Codecs" 2019
- [6] O. H. Nagoor, J. Whittle, J. Deng, B. Mora, M. W. Jones, "Lossless Compression for Volumetric Medical Images using Deep Neural Network with Local Sampling" 2020
- [7] Zhaoxiang Zhang, Guodong Xu and Jianing Song, "CubeSat cloud detection based on JPEG2000 compression and deep learning"
- [8] D. Sun and W.-K. Cham, "Postprocessing of low bit-rate block dct coded images based on a fields of experts prior," IEEE Transactions on Image Processing, vol. 16, no. 11, pp. 2743–2751, 2007
- [9] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," IEEE transactions on image processing, vol. 22, no. 12, pp. 4613–4626, 2013
- [10] J. Ren, J. Liu, M. Li, W. Bai, and Z. Guo, "Image blocking artifacts reduction via patch clustering and low-rank minimization," in Data Compression Conference (DCC), 2013. IEEE, 2013, pp. 516–516
- [11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Transactions on image processing, vol. 16, no. 8, pp. 2080–2095, 2007
- [12] H. Chang, M. K. Ng, and T. Zeng, "Reducing artifacts in jpeg decompression via a learned dictionary," IEEE transactions on signal processing, vol. 62, no. 3, pp. 718–728, 2014
- [13] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2862–2869.
- [14] J. Zhang, R. Xiong, C. Zhao, Y. Zhang, S. Ma, and W. Gao, "Concolor: constrained non-convex low-rank model for image deblocking," IEEE Transactions on Image Processing, vol. 25, no. 3, pp. 1246–1259, 2016.
- [15] C. Dong, Y. Deng, C. Change Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 576–584.