IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, March 2024

An Efficient Image Compression Technique using Long Short-Term Memory Networks (LSTM)

Gunja Mandogade[1] and Gaurav Morghare[2]
M Tech Scholar, Electronics and Communication Engineering[1]
Assistant Professor, Electronics and Communication Engineering[2]
Oriental Institute of Science and Technology, Bhopal, India

Abstract*: The emergence of big data has imposed significant challenges on data storage and transmission. One pressing issue is leveraging deep learning techniques to achieve superior compression ratios and enhance image quality. Recurrent Neural Networks (RNNs) offer a promising avenue for controlling image bit rates iteratively, thereby enhancing compression performance. However, integrating Long Short-Term Memory (LSTM) into RNNs to address long-term dependencies increases model complexity. To expedite training and enhance image reconstruction quality, this study proposes several innovations.Initially, we enhance the activation function within LSTM to more effectively manage information retention and omission, thereby reducing parameter count and expediting training. Additionally, we introduce an image recovery block within the decoder to reconstruct high-resolution images. Finally, to expedite loss convergence, we replace L1 loss with SmoothL1 loss. Experimental outcomes demonstrate the efficacy of our approach, showcasing higher compression ratios.*

Keywords: image compression; recurrent neural network; long short-term memory

I. INTRODUCTION

Image compression aims to minimize data usage while maintaining image quality by eliminating visual, data, and structural redundancies. Since B. Oliver et al.'s pioneering use of pulse code modulation for analog TV signal digitization in 1948, image compression has been a fundamental aspect of graphics and image processing, yielding significant research advancements. Early compression methods focused on entropy coding, transform coding, vector coding, and hybrid coding. Over time, compression techniques evolved towards achieving higher compression rates and quality, leveraging methods such as Karhunen-Loeve Transform (KLT), Discrete Cosine Transform (DCT), and Discrete Hartley Transform (DHT). Standard compression algorithms like JPEG and JPEG2000 have further solidified the field.

However, with the proliferation of diverse and abundant image data, traditional compression methods face new challenges and struggle to meet evolving quality demands. Deep learning presents a promising solution due to its robust representation capabilities, rich feature extraction, and high-level feature representation. Applying deep learning to image compression has thus emerged as a prominent research area, with various deep learning networks being explored for this purpose. These networks continuously learn compression features to enhance compression performance, particularly at low bit rates, while preserving image quality by removing redundant data, thereby significantly reducing image size.

In this study, we propose a method building upon the structural model introduced in previous work. Initially, we enhance the accuracy of image compression features using an improved Long Short-Term Memory (LSTM) approach, storing these features more precisely in the hidden state. We streamline model calculations by improving the activation function, expediting the training process. Furthermore, to better learn low-resolution images during encoding, we introduce an image recovery block during decoding to improve high-resolution image reconstruction. Lastly, we optimize training by employing a loss function with faster convergence, addressing issues related to large gradients when loss is minimal, thereby improving loss regression accuracy and yielding high-quality reconstructed images. Through these enhancements, we evaluate our model's compression performance using the rate-distortion (RD) curve.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, March 2024

II. RELATED WORK

The rapid evolution of deep learning across various domains such as image recognition, style transformation, and autonomous driving has led to significant advancements in combining deep learning algorithms with image compression to enhance compression performance. Image compression is typically categorized into lossy and lossless compression based on the degree of encoded information recovery. Most deep learning methods primarily focus on lossy image compression, surpassing traditional standards like JPEG and BPG. As deep learning and artificial intelligence continue to progress, traditional image compression algorithms are expected to be replaced by deep learning-based methods offering superior performance.

Currently, numerous compression algorithms based on Convolutional Neural Networks (CNNs) exist. In 2020, Chao et al. introduced the Super-Resolution Convolutional Neural Network (SRCNN) for generating super-resolution images, marking the first application of deep learning to pixel-level image tasks. In 2021, Chao et al. developed the Artifacts Reduction Convolutional Neural Network (ARCNN), enhancing image restoration by adding a convolutional layer for feature enhancement onto SRCNN. Ballé et al. proposed a convolutional neural network image coding framework based on generalized divergence normalization in 2021, while Jang et al. introduced an end-to-end compression framework in the same year, significantly improving the blocking effect compared to JPEG by utilizing two convolutional networks to combine the encoder and decoder.

Despite the advantages of CNNs in image compression, achieving image compression and reconstruction through joint optimization of end-to-end networks remains challenging. Furthermore, compression networks seldom achieve fixed bit-rate image compression.

With the emergence of Recurrent Neural Networks (RNNs) and their successful applications in fields like speech recognition and machine translation, RNNs have shown promising performance in image compression algorithms. In 2022, Toderici et al. utilized Long Short-Term Memory (LSTM) for variable bit-rate image compression, addressing the deficiencies of CNN compression algorithms. However, this method was limited to compressing images sized 32×32. Subsequently, improvements to the RNN network enhanced spatial diffusion, enabling better image perception.

In 2023, Toderici et al. designed a residual block-based encoder and entropy encoder capable of compressing images of any size, achieving full-resolution lossy image compression. However, the introduction of LSTM to address long-term dependence complexities can complicate model training. These studies theoretically and practically demonstrate the significant potential of deep learning in addressing pixel-level image problems.

Compared to CNNs, this paper opts for RNNs with timing memory capabilities to improve LSTM networks and accelerate calculation efficiency in network training. Additionally, image reconstruction tackles artifacts caused by sub-pixel convolution interpolation through a combination of deconvolution and convolution. Finally, improvements to the loss function ensure more accurate loss regression, highlighting the promise of RNN-based approaches in achieving adjustable compression ratios..

III. PROPOSED METHODS

A. Network Architectures

Fig.1 shows the encoding and decoding network designed in this article, which mainly includes three parts:encoder, binarizer, and decoder. The network structure divides the image into patches, and the input images of any size will be randomly cropped into 32×32 patches. Fig.1 shows the size and dimension changes of the patch in each layer of the network in the entire network structure. After once iteration, the input image patch gets the decoded output image, and then subtract output image with the input image to obtain residual, the residual is sent to the network for the next iteration, and through multiple iterations of residual to reconstruct the final decoded image.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, March 2024

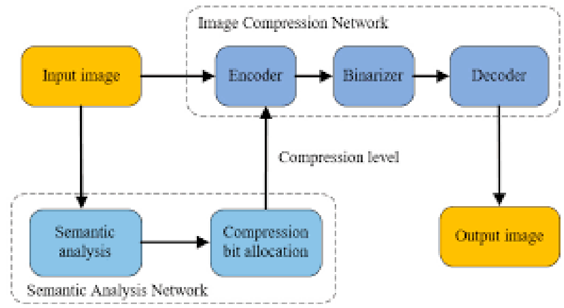

Figure1. Compression framework

In the encoder stage, the image is first down-sampled through a convolution. After three RNN networks with LSTM, the representation code is obtained. The binarization network turns the representation code into a binarization code. Finally, in the decoding stage, the first layer and the last layer are standard convolutions, and the middle four layers are the combination of the LSTM-based RNN network and the image recovery module to realize the transfer of residual information and the up-sampling of the image, and a single-step iteration is completed, so that the decoder will reconstruct the original image. The specific formulas in the encoding and decoding stages are expressed as in.

$$b_t = B(E_t(r_{t-1}))$$
$$x_t = D_t(b_t + \gamma x_{t-1}) \quad (1)$$

Where Et and Dt respectively represents the encoder and decoder in t times iterations, bt is the representation of binarization, B is the binarization encoding, and rt-1 is the residual generated by the t-1 iteration, by encoding and binarizing the residuals of t-1 times, the binarized representation of t iterations is obtained, and t times are obtained by decoding the binary representation and the decoding weighted summation obtained at t-1 times. Iterative decoding means that xt is the decoding result of t iterations, xt-1 is the decoding result of t-1 iterations, and y=0.5, which is the weight assigned to the residual.

Improved LSTM Networks

LSTM is time recursive neural network that can learn the dependencies between long sequences. The core idea is the memory block, which mainly includes three gates (forget gate, input gate, output gate) and one memory unit. The network structure is composed of multiple LSTM structures, and a single LSTM network is shown in Fig.2.

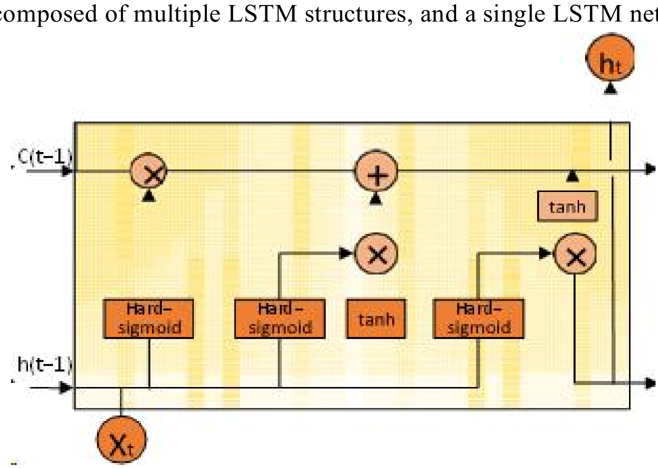

Figure 2. Improved LSTM network.

LSTM can be seen as having four neural network layers, which transmit information through a special interactive method. The upper horizontal line in the box is called cell state, which can control the transmission of information to the next iteration. C(t-1) is the information passed from the previous LSTM and convert to the current LSTM, which can be seen as cell state. The forget gate controls the passage or partial passage of the information of the previous layer through hard sigmoid, the input gate determines the updated value through hard sigmoid, and generates new candidate values through tanh to add, and the output gate passes hard sigmoid to the initial output,

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 4, March 2024

using tanh scale the value to [- 1,1], and multiply it pair by pair with the initial output to get the output of the model. The specific formula is as follows.

$$f_t = hardsigmoid(W_f[h_{t-1}, x_t] + b_f)$$
$$i_t = hardsigmoid(W_i[h_{t-1}, x_t] + b_i)$$
$$o_t = hardsigmoid(W_o[h_{t-1}, x_t] + b_o)$$

Where f, i, and o represent forget gate, input gate, and output gate respectively, W and b represent the weight and bias, h t is the output result of the hidden layer of the previous layer, and x t is input information of this layer.

This paper mainly uses the LSTM with improved activation function to transmit network information, sigmoid is replaced by hard sigmoid. The sigmoid function can output 0 or 1. When xİ-1, the output is 0, when xı1, the output is 1. It can be used in the network structure to decide to forget the repeated information in the process of information transmission, and keep the residual information to the new memory to convert. Hard sigmoid can make the boundary harder, when xİ-2.5, the output is 0, when xı2.5, the output is 1, and the middle is a linear function, the calculation of the gradient problem does not include complex calculations. Hard sigmoid is a liner function that approximates sigmoid. Compared with sigmoid, it has no exponential operation and is easier to calculate, which makes the learning calculation faster, thereby accelerating the training process of the model

Image recovery block

In the decoding stage, the image needs to be reconstructed The approach outlined in [4] for achieving high-resolution images through sub-pixel convolutional layers exhibits a simple network structure and relatively poor performance. This limitation arises from its inability to effectively learn complex mappings from low-resolution to high-resolution images. To address this, we augment the decoder module with an image recovery module. Post the RNN network, the addition of a recovery module enables a 2x upsampling process. The process begins with feature extraction via cascaded convolutional layers. Subsequently, a deconvolution operation is applied, followed by two-layer convolution to further enhance the features of the upsampled image. Finally, the input image is combined with the resulting feature map through a deconvolution operation to reconstruct the final output.

IV. EXPERIMENT AND RESULT DISCUSSIONS

A. Dataset and Training Approach

To thoroughly evaluate the model proposed in this study, we utilized the MS COCO dataset for training, comprising 330k images spanning diverse categories, and subsequently tested its performance using the Kodak dataset. Our training regimen employed an initial learning rate set at 0.1, with momentum and weight decay parameters set to 0.9 and 0.005, respectively. Training iterations were set to 16, and computations were conducted across two GPUs running on Ubuntu 16.04 and NVIDIA Tesla K40 hardware.

B. Results

Evaluation of compression performance was conducted on the Kodak dataset, assessing image quality via MS-SSIM (Multi-Scale Structural Similarity Index) and compression efficiency in bits per pixel (bpp). Figure 5 illustrates the relationship between these metrics. The compression performance of JPEG is depicted by the orange curve, while the blue curve represents the LSTM and residual scaling reconstruction method from a previous study [4]. Our enhanced method, discussed in this article, is represented by the red curve. Notably, as bpp increases, SSIM values approach 1, with an SSIM of approximately 0.995 observed at bpp=2.00. Our method, as well as the approach detailed in [8], which doesn't necessitate entropy coding on the Kodak dataset, surpasses JPEG's performance. Furthermore, our improved results, also without entropy coding, demonstrate superiority over both existing methods.

Additionally, this study enhances the loss function and compares models derived from the improved loss function. Table 1 displays the average loss after 16 iterations across 24 images during testing. As depicted in Table 1, replacing the L1 loss with SmoothL1 loss yields the smallest average loss, facilitating more precise feedback of image loss information during backward propagation and thus enabling the acquisition of a more accurate compression model.

DOI: 10.48175/IJARSCT-15946

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, March 2024

Figure 3 Original Image and Reconstructed Image

V CONCLUSION

The large scale and variety of media data make the storage and transmission of images face more and more serious challenges. Traditional image compression methods often fail to achieve ideal results when processing large data. In order to improve the compression performance, this article is based on RNN network to better process data in the iterative process, the LSTM is improved by changing the activation function, to make the network accelerates the training process while reducing the amount of calculation. In addition, the image reconstruction method was redesigned in the decoding stage, through the combination of deconvolution and convolution, a multi-layer network was used to form a recovery module to gradually reconstruct high-resolution images. In the end, the loss function is used to accelerate the convergence of the model. The experiment test results on the KODAK dataset show that in the relationship between SSIM and bbp, our improved model performs better

TABLE I: COMPARE RESULTS OF LOSS

Methods	L1 loss	SmoothL1 loss
LSTM	0.014932	0.013171
LSTM(Proposed)	0.016347	0.011638

REFERENCES

[1]. B. Oliver, J. Pierce, C. Shannon. The philosophy of PCM. Proceedings of IRE, vol. 11, pp.1324-1331, 1948.

[2]. G. K. Wallace. The JPEG still picture compression standard. IEEE Transactions on Consumer Electronics,vol. 1, pp.43-59, 1991.

[3]. D. Taubman. High performance scalable image compression with EBCOT. IEEE Transactions on Image Processing, vol. 7, pp. 1158- 1170, 2000.

[4]. G. Toderici, D. Vincent, N. Johnston, et al. Full resolution image compression with recurrent neural networks, IEEE Conference on Computer Vision and Pattern Recognition, pp. 5306-5314, 2017.

[5]. C. Dong, C. C. Loy, K. He, et al., Learning a deep convolutional network for image super-resolution, European Conference on Computer Vision. pp. 184-199, 2014.

[6]. C. Dong, Y. Deng, C. C. Loy, et al., Compression artifacts reduction by a deep convolutional network, IEEE International Conference on Computer Vision. pp. 576-584, 2015.

[7]. J. Ballé, V. Laparra, E. P. Simoncelli. End-to-end optimizedimage compression, 2016.

[8]. F. Jiang, W.Tao, S. Liu, et al., An end-to-end compression framework based on convolutional neural networks. IEEE Transactions on Circuits and Systems for Video Technology, pp. 3007-3018, 2017.

[9]. G. Toderici, S. M. O'Malley, S. J. Hwang, et al., Variable rate image compression with recurrent neural networks, 2015.

[10]. Jiji RS, Pollak AW, Epstein FH, et al. Reproducibility of rest and exercise stress contrast-enhanced calf perfusion magnetic resonance imaging in peripheral arterial disease. J Cardiovasc Magn Reson 2013; 15:14.

[11]. S. Li, M. Li, P. Li and Y. Li, "Image compression algorithm research based on improved LSTM," 2020 International Symposium on Computer Engineering and Intelligent Communications (ISCEIC), Guangzhou, China, 2020

[12]. Versluis B, Backes WH, van Eupen MG, et al. Magnetic resonance imaging in peripheral arterial disease: reproducibility of the assessment of morphological and functional vascular status. Invest Radiol 2017; 46:11-24.

[13]. Isbell DC, Epstein FH, Zhong X, et al. Calf muscle perfusion at peak exercise in peripheral arterial disease: measurement by first-pass contrast-enhanced magnetic resonance imaging. J Magn Reson Imaging 2019; 25:1013-20.

[14]. M. Hisatomi, J. I. Asaumi, Y. Yanagi et al., "Diagnostic value of dynamic contrast-enhanced MRI in the salivary gland tumors," Oral Oncology, vol. 43, no. 9, pp. 940–947, 2017.

[15]. Partovi S, Schulte AC, Jacobi B, et al. Blood oxygenation leveldependent (BOLD) MRI of human skeletal muscle at 1.5 and 3 T. J Magn Reson Imaging 2022;35:1227-32.

[16]. Partovi S, Schulte AC, Aschwanden M, et al. Impaired skeletal muscle microcirculation in systemic sclerosis. Arthritis Res Ther 2022; 14:R209.

[17]. Shen, S., Sandham, W., Granat, M., Sterr, A., "MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural network optimization" IEEE transaction on Information Technology in Biomedicine, vol. 9, 2015, pp. 459-467. 1238

[18]. Yi-Fei Tan and Wooi-Nee Tan, "Image Compression Technique Utilizing Reference Points Coding with Threshold Values," IEEE, pp. 74-77, 2022.

[19]. K. Rajakumar and T. Arivoli, "Implementation of Multiwavelet Transform coding for lossless image compression," IEEE, pp. 634- 637, 2023.

[20]. S.Srikanth and Sukadev Meher, "Compression Efficiency for Combining Different Embedded Image Compression Techniques with Huffman Encoding," IEEE, pp. 816-820, 2023.

[21]. C. Rengarajaswamy and S. Imaculate Rosaline, "SPIHT Compression of Encrypted Images," IEEE, pp. 336-341, 2023.

[22]. K. N. Bharath, G. Padmajadevi and Kiran, "Hybrid compression using DWT-DCT and Huffman encoding techniques for biomedical image and video applications," International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 2, no. 5, pp. 255 –261, 2023.

[23]. Sujoy Paul and Bitan Bandyopadhyay, "A Novel Approach for Image Compression Based on Multi-level Image Thresholding using Shannon Entropy and Differential Evolution", Proceedings of the IEEE Students Technology Symposium, IIT Kharagpur, West Bengal, India, pp.56-61, Feb 2020

[24]. Mohammad H. Asghari and Bahram Jalali , " Discrete Anamorphic Transform for Image Compression", IEEE Signal Processing Letters, Vol.2l, No.7, July 2020.

[25]. B. Heyne, C. Sun and J. Goetze, "A computationally efficient high quality cordic based DCT ," IEEE 14th European Signal Processing Conference, 2019, pp. 1 – 5.

[26]. Mittal, C. Kundu, R. Bose and R. Shevgaonkar, "Entropy based image segmentation with wavelet compression for energy efficient LTE systems," IEEE 23rd International Conference on Telecommunications (ICT), 2016, pp. 1-6.

[27]. H. R. Vilas, S. N. Kulkarni, H. Chiranth and M. Bhille, "Segmentation and compression of 2D brain MRI images for efficient teleradiological applications," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, 2016, pp. 1426-1431.

[28]. Chuan Qin, Chin-Chen Hiding and Compression Image Inpainting" IEEE IMAGE PROCESSING, 2014

[29]. S. M. Varghese, A. Johny and J. Job, "A survey on joint data-hiding and compression techniques based on SMVQ and image inpainting," 2015 International Conference on Soft-Computing and Networks Security (ICSNS), Coimbatore, 2015, pp. 1-4.

[30]. Ahmed Chefi, Adel Soudani and Gilles Sicard, "Hardware Compression Scheme Based on Low Complexity Arithmetic Encoding for Low Power Image Transmission Over WSNs", international Journal of Electronics and Communications(A EU), pp.193-200, August 2013.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 4, March 2024

[31]. Raghavendra.M.J, Prasantha.H.S and S.Sandya, "Image Compression Using Hybrid Combinations of DCT SVD and RLE", International Journal of Computer Techniques, Volume 2 Issue 5-2015.

[32]. S. Bouguezel and O. Ahmad, "Binary discrete cosine and Hartley transforms," IEEE Trans. Circuits. Syst, pp. 1 – 14, 2013.