Compressing and Decompressing 360-degree Images - A Neural Network based Approach

1st Sayed Tanjim Alam
Dept. of Computer Science and Engineering
BRAC University)
Dhaka, Bangladesh

2nd Syed Ahsan Raizan
Dept. of Computer Science and Engineering
BRAC University)
Dhaka, Bangladesh

***Abstract*—In this paper, we present a novel neural network-based approach for compressing and decompressing 360-degree images, which are widely used in virtual reality (VR) and panoramic applications. The exponential growth of such images demands efficient compression techniques to reduce storage and transmission requirements. Leveraging the power of deep learning, we propose a comprehensive framework that employs a convolutional neural network (CNN) architecture tailored specifically for 360-degree image data. Our approach capitalizes on the inherent spatial and structural characteristics of these images to achieve superior compression performance. By training the network on a large-scale dataset of 360-degree images, we enable it to learn complex patterns and efficiently represent the content within the images. Through extensive experiments and comparative analyses, we demonstrate that our proposed method outperforms state-of-the-art techniques in terms of compression ratio and visual quality, while maintaining low decoding latency. Our approach not only provides substantial gains in storage efficiency but also contributes to enhancing the user experience by enabling faster and more immersive delivery of 360-degree content. The results obtained highlight the potential of neural network-based approaches in revolutionizing the compression and decompression of 360-degree images, thereby opening up new avenues for efficient data representation in VR and panoramic applications.**

***Index Terms*—360 images, compression, decompression, panoramic images, omnidirectional images**

I. Introduction

The rapid advancement of virtual reality (VR) and panoramic applications has led to an exponential growth in the usage and demand for 360-degree images. These immersive visuals provide users with a seamless and encompassing experience, enabling them to explore virtual environments with unprecedented realism. However, this increase in image complexity poses significant challenges in terms of storage and transmission, necessitating the development of efficient compression techniques specifically tailored to the unique characteristics of 360-degree images.

In this paper, we propose a neural network-based approach to address the critical problem of compressing and decompressing 360-degree images. Our objective is to leverage the capabilities of deep learning algorithms to learn the intricate features and transformations inherent in these images, enabling highly efficient compression and faithful reconstruction. We present an innovative approach - Convolutional Feature Learning, which is designed to capture the spatial and structural intricacies of 360-degree images and exploit them for superior compression performance.

The approach, Convolutional Feature Learning, aims to exploit the inherent spatial dependencies present in 360-degree images. By inputting the images in an equirectangular format, we engineer the dataset to generate multiple images, each representing a vertical cross-section of the original image. These images are then processed through a convolutional neural network (CNN) to learn their distinctive features. The learned features are subsequently utilized to train a neural network that outputs the pixel values for each pixel in the original 360-degree image. During decompression, the stored learned weights and feature maps are utilized to faithfully recreate the original image, ensuring high fidelity while achieving efficient compression.

Through extensive experiments and comparative analyses, we evaluate the performance of our proposed approaches against state-of-the-art compression techniques for 360-degree images. We measure compression ratios, visual quality, and decoding latencies to demonstrate the superiority of our neural network-based methods. The results obtained highlight the significant potential of our approaches in revolutionizing the compression and decompression of 360-degree images, ultimately enhancing storage efficiency and enabling faster and more immersive delivery of VR and panoramic content.

In summary, this paper presents an innovative and comprehensive neural network-based approach for compressing and decompressing 360-degree images. Our novel method, Convolutional Feature Learning, exploits the unique characteristics of 360-degree images to achieve superior compression performance. The findings from this study have implications for the efficient representation and delivery of VR and panoramic content, contributing to the ongoing advancements in immersive visual experiences. However, this research would also like to point out that the proposed method may not work and/or may need some significant alterations to improve the results to an acceptable standard.

II. Literature Review

Efficient compression of 360-degree images poses unique challenges compared to traditional 2D image compression methods. As highlighted in the first reference paper OSLO [1], extending convolutional neural network (CNN) models, which have shown promising results for 2D image compression, to the domain of omnidirectional images is not straightforward. Omnidirectional images, also known as spherical images, possess specific spatial and statistical properties that cannot be adequately captured by existing CNN models. Furthermore, the mathematical operations fundamental to CNN architectures, such as translation and sampling, are ill-defined on the spherical domain.

To address these challenges, the OSLO paper proposes a novel framework for learning representation models for omnidirectional images. The authors leverage the properties of HEALPix uniform sampling on the sphere to redefine the mathematical tools used in deep learning models for omnidirectional images. They introduce a new convolution operation on the sphere, preserving the expressiveness and low complexity of traditional 2D convolutions. Additionally, standard CNN techniques like stride, iterative aggregation, and pixel shuffling are adapted to the spherical domain. The proposed framework is applied to the task of omnidirectional image compression, demonstrating improved compression gains compared to learned models applied to equirectangular images. The solution supports more expressive filters, preserving high frequencies and enhancing the perceptual quality of compressed images.

In the second reference paper [2], the authors address the unbalanced sampling problem in 360-degree image compression, which arises due to the transformation of spherical images into planar images using projections like equirectangular projection (ERP). The unbalance in pixel density across different circles of latitude presents challenges for planar compression methods, especially those based on deep neural networks (DNNs). To overcome this problem, the authors propose a latitude adaptive coding scheme for DNNs. The scheme allocates variant numbers of codes for different regions based on the latitude on the sphere. A flexible regional adaptive rate loss is introduced to control the rate allocation, considering both the number of allocated codes for each region and their entropy. Latitude adaptive constraints are further incorporated to prevent excessive code usage in oversampled regions. The authors also introduce a viewport-based distortion loss to optimize compression quality. Experimental results on a large dataset of 360-degree images demonstrate the superiority of the proposed latitude adaptive coding scheme, outperforming existing image compression standards.

Both reference papers provide valuable insights into the challenges and advancements in the compression of 360-degree images. The first paper emphasizes the need for tailored approaches that exploit the spherical nature of omnidirectional images, proposing a framework that redefines mathematical tools and CNN techniques for the spherical domain. The second paper addresses the unbalance sampling problem introduced during the transformation of spherical images to planar images, proposing a latitude adaptive coding scheme that improves compression performance compared to traditional planar compression methods. These findings highlight the importance of developing specialized techniques to effectively compress and decompress 360-degree images, leading to enhanced storage efficiency, transmission, and overall user experience in virtual reality and panoramic applications.

Building upon the advancements highlighted in the reference papers, our research aims to contribute to the field of 360-degree image compression by proposing a neural network-based approach specifically designed to efficiently compress and decompress these immersive images.

The first reference paper acknowledges the limitations of existing CNN models in capturing the spatial and statistical properties of omnidirectional images. Our approach, inspired by the concepts presented in the paper, seeks to address these limitations by developing a neural network architecture capable of effectively learning the features and transformations unique to 360-degree images. By leveraging the insights from the proposed new convolution operation on the sphere, we aim to design a network that captures the complex spatial dependencies and preserves the high expressiveness necessary for faithful compression and reconstruction of omnidirectional content.

Furthermore, the second reference paper EtEO [2] highlights the challenges arising from the unbalanced sampling problem when transforming spherical images to planar representations for compression. Drawing inspiration from their latitude adaptive coding scheme, our research aims to explore adaptive rate allocation strategies tailored specifically for neural network-based codecs. By considering the varying pixel densities across different regions of the spherical images, we aim to develop an innovative approach that dynamically allocates resources and optimizes the compression quality for each region, ensuring efficient representation and preservation of high-frequency details.

In addition to the compression performance metrics such as compression ratio and visual quality, we recognize the importance of considering practical factors like computational complexity and decoding latency. Our research will evaluate and analyze these aspects to ensure that the proposed approach is not only efficient in terms of compression gains but also feasible for real-time applications and practical implementations.

To validate the effectiveness of our approach, we will conduct extensive experiments on a large-scale dataset of 360-degree images, capturing diverse scenes and content. Comparative analyses will be performed against state-of-the-art compression techniques, including those presented in the reference papers, as well as established image compression standards. Through these experiments, we aim to demonstrate the superior compression performance and improved perceptual quality achieved by our neural network-based approach, showcasing its potential in revolutionizing the compression and decompression of 360-degree images.

The outcomes of our research have significant implications for various applications relying on 360-degree images, such as virtual reality, panoramic photography, and immersive multimedia experiences. By efficiently compressing these high-dimensional images, we can reduce storage requirements, enable faster transmission, and enhance the overall user experience by facilitating seamless delivery of immersive content.

In summary, our research builds upon the insights provided by the reference papers, aiming to develop a neural network-based approach for efficient compression and decompression of 360-degree images. By addressing the challenges unique to omnidirectional content and leveraging advanced techniques, our research seeks to contribute to the advancement of 360-degree image compression, paving the way for improved storage efficiency, transmission, and immersive visual experiences in a wide range of applications.

III. Methodology

A. Convolutional Feature Learning Approach

This research explores the application of Convolutional Neural Networks (CNNs) for the task of feature extraction and image regeneration, using 360-degree images in an equirectangular format as the primary data. The goal is to leverage the inherent learning capabilities of a neural network to understand the intricate spatial relations within the 360-degree images, subsequently using these learnt features for image regeneration.

The process commences with the conversion of equirectangular images into a suitably processed dataset. In a typical equirectangular image, every pixel encapsulates a view of the scene from a different angle. To ensure the utilization of this comprehensive information, we generate a set of new images equivalent to the count of pixels in the image height (N). Each generated image then represents a unique perspective of the original 360-degree image.

Following the generation of this new dataset, these images are subjected to a convolutional layer to facilitate feature learning. A convolutional layer applies a series of filters to the image, which are optimized through the training process to extract relevant features, thereby creating a higher-dimensional representation of the original image. These features serve as the foundation for understanding the inherent structures and patterns within the original images, capturing the spatial and temporal dependencies in the data.

Once these features have been learnt, they are utilized to train a subsequent neural network. This network is designed to associate the extracted features with the corresponding pixel values from the original 360-degree image. The objective here is to effectively learn a mapping from the feature space to the original pixel space, effectively allowing the network to recreate any given input image.

The trained neural network model is then preserved, along with the learned weights and feature maps. These stored assets form the basis of the decompression mechanism. The concept of decompression here pertains to the regeneration of the original image using only the saved neural network model, weights, and feature maps. This approach promises potential for efficient image storage and transfer, minimizing data requirements while preserving image integrity.

During the decompression process, the saved neural network model is reloaded, and the corresponding weights are applied to the feature maps. The process results in the recreation of the original image from the learned feature representations.

IV. Dataset

The research uses **360 Viewer** dataset by Will Buzan from Kaggle [3] to explore the aforementioned approach. The dataset contains equirectangular format RGB images of a store, with 18 images in JPG format.

V. Results

In this section, the outcomes of the research experiment is presented, evaluating the effectiveness of it in the task of image compression and decompression using 360-degree images. The quantitative assessments of the proposed method's model performance is detailed, followed by a discussion of the findings.

The model uses 3 convolutional layers along with 2 max pool layers in between, before being passed on to a flattening layer and finally to two dense layers. It tries to map a tensor containing only ones as input to each channel of the three RGB values of each pixel of the image, essentially labelling channels between 0-255. This would be the compression stage. The experiment was repeated twice using two loss functions - Sparse Categorical Crossentropy [4] and Kullback-Leibler Divergence [5], with the former showing better accuracy results than the latter, but only by a small margin.

Unfortunately, the model seems to be failing very badly to perform this mapping. The result accuracy are as low as 1%, over a span of 10 epochs, without any visible trend that would suggest the possibility of an improvement in the model performance over a larger number of epochs. We theorize that this can be attributed to the model architecture design, thereby can be termed as a technical issue. Since the compression stage is failing, we cannot go forward with the decompression stage of the experiment. We will triage more on the model architecture before the final submission.

References

[1] N. Mahmoudian Bidgoli, R. G. de A. Azevedo, T. Maugey, A. Roumy and P. Frossard, "OSLO: On-the-Sphere Learning for Omnidirectional Images and Its Application to 360-Degree Image Compression," in IEEE Transactions on Image Processing, vol. 31, pp. 5813-5827, 2022, doi: 10.1109/TIP.2022.3202357.

[2] M. Li, J. Li, S. Gu, F. Wu and D. Zhang, "End-to-End Optimized 360° Image Compression," in IEEE Transactions on Image Processing, vol. 31, pp. 6267-6281, 2022, doi: 10.1109/TIP.2022.3208429.

[3] https://www.kaggle.com/datasets/willbuzan/360-panos

[4] Mannor, Shie, Peleg, Dori and Rubinstein, Reuven. (2005). The cross entropy method for classification. 561-568. 10.1145/1102351.1102422.

[5] R. Kwitt and A. Uhl, "Image similarity measurement by Kullback-Leibler divergences between complex wavelet subband statistics for texture retrieval," 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 2008, pp. 933-936, doi: 10.1109/ICIP.2008.4711909.