

Image Compression with Recurrent Neural Network and Generalized Divisive Normalization

Khawar Islam¹, L. Minh Dang¹, Sujin Lee², Hyeonjoon Moon^{1,2}

Computer Vision and Pattern Recognition Lab¹, Sejong University, South Korea

Department of Artificial Intelligence^{1,2}, Sejong University, South Korea

Fkhawar, minhdang@sejong.ac.kr, genegraphy@gmail.com, hmoon@sejong.ac.kr

Abstract

Image compression is a method to remove spatial redundancy between adjacent pixels and reconstruct a high-quality image. In the past few years, deep learning has gained huge attention from the research community and produced promising image reconstruction results. Therefore, recent methods focused on developing deeper and more complex networks, which significantly increased network complexity. In this paper, two effective novel blocks are developed: analysis and synthesis block that employs the convolution layer and Generalized Divisive Normalization (GDN) in the variable-rate encoder and decoder side. Our network utilizes a pixel RNN approach for quantization. Furthermore, to improve the whole network, we encode a residual image using LSTM cells to reduce unnecessary information. Experimental results demonstrated that the proposed variable-rate framework with novel blocks outperforms existing methods and standard image codecs, such as George's [11] and JPEG in terms of image similarity. The project page along with code and models are available at https://github.com/khawar512/cvpr_image_compress

1. Introduction

Recent deep learning approaches for lossy image compression have been gained significant interest in machine learning and achieves more promising results. It plays a significantly important role in streaming a large amount of image and video data under adequate storage, and low bandwidth [4, 5, 7]. Deep learning has made tremendous contributions in neural network-based image coding.

Various image compression methods [2, 9, 11] have achieved remarkable performance. Ballé et al. [3] proposed a generalized divisive normalization (GDN) based framework which comprises three main steps: convolutional layer, sub-sampling scheme, and nonlinear GDN lay-

ers. Recently, a novel model for learned image compression is context-adaptive [6] receives significant performance and outperforms all image codec. Johannes et al. [3] designed a simple hyperprior model to add a greater number of bits in the entropy module. Minnen et al. [8] combined an autoregressive with a hierarchical module to obtain better results in the context of image compression. This work [6] constructed a similar technique [3] by combining two different context models; one is a bit consuming, and the other one is bit-free contexts to understand a context-adaptive model.

Although CNN-based image compression methods [6, 7, 8], already improved compression accuracy, commonly mostly methods are increasing more layers and make deeper network to achieve compression results. These methods substantially increased model complexity and computational cost. As far as we know, the development of image compression networks based on RNN is relatively small compared to CNN and auto-encoders. These techniques [11, 12] proposed a full image resolution network using residual scaling, recurrent neural network (RNN), and entropy coding based on deep learning. This network simultaneously generated three models during training.

In this paper, two spatial adaptive blocks called analysis and synthesis are proposed in the variable-rate encoder and decoder side based on a convolution layer and generalized divisive normalization process. Then, it embeds with a variable-rate framework. The analysis block generates a powerful spatial and channel representation with a down-sampling feature for encoding. Simultaneously, synthesis block up-sampling the decoded image. To the best of our knowledge, constructing recurrent neural network-based image compression networks is very limited. This is the first work that employs a generalized divisive normalization scheme in a variable rate network to handle multiple bit rates in the RNN network. Furthermore, visual quality performance and scalability of the proposed blocks and end-to-end network are jointly validated by the popular evaluation with PSNR and MS-SSIM. Comprehensive experiments have been performed on Kodak to describe that

