Exploration for basic binary convolution unit

Wenxuan Zhang
Wzhang121@hawk.iit.edu

ABSTRACT

To be completed..

ACM Reference Format:
Wenxuan Zhang. 2024. Exploration for basic binary convolution unit. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 BACKGROUND

Image restoration is an important topic in image processing. Also known as image recovery, it is a technique in image processing aimed at improving the quality of a given image. When provided with a degraded or noise-contaminated image, the basic process of restoration involves reconstructing or restoring the original image using some prior knowledge of the degradation phenomenon. Image restoration, or IR, aims to recover high-quality (HQ) images from low-quality (LQ) images that have been degraded by various factors. Typical IR tasks include image denoising, super-resolution (SR), and removing compression artifacts. Due to its ill-posed nature and high practical value, image restoration is an active and challenging research topic in computer vision. For example, deep convolutional neural networks have achieved excellent performance by learning mappings from low-quality (LQ) to high-quality (HQ) image patches. However, most image restoration tasks require dense pixel predictions, and the strong performance of CNN-based models often relies on increasing model size and computational complexity, requiring significant computational and memory resources. Therefore, significantly reducing the computational and memory costs of CNN models while maintaining model performance is crucial for the generalization of image restoration models.

Binary Neural Network (BNN), also known as 1-bit CNN, has been proposed and considered one of the most promising neural network compression methods for deploying models on resource-constrained devices[1]. On specially designed processors, BNN can achieve a 32x memory compression and up to a 64x reduction in computational load. This paper comprehensively explores the impact of binary networks on restoration networks. After extensive experimentation and analysis, it proposes a basic binary convolution unit (BBCU) for image restoration along with a binarization scheme. Compared to previous binary restoration networks, significant performance improvements have been achieved in tasks such as super-resolution, denoising, and compression artifact reduction. For instance, maintaining the same performance in super-resolution tasks while achieving a tenfold speedup.

Currently, research on BNN is mainly focused on high-level vision tasks, especially image classification, but low-level vision tasks such as image super-resolution have not been fully explored. Considering the importance of BNN for deploying deep networks for image restoration and the differences between high-level and low-level vision tasks, there is an urgent need to explore the characteristics of BNN in low-level vision tasks and provide a simple, powerful, versatile, and scalable baseline for subsequent research and deployment.

2 A BRIEF SURVEY

Binary Neural Network (BNN) refers to a type of neural network where weights and activations are represented using only two values, +1 and -1. Compared to full-precision neural networks, BNNs can perform convolution operations using a simple combination of XNOR and popcount instead of float32 multiply-accumulate operations, thereby saving a significant amount of memory and computation. Compared to full-precision (32-bit) CNN models, Binary Neural Networks (BNNs) can achieve up to 32 times memory savings. Additionally, since activations are also binary, binary convolution operations can be performed using bitwise XNOR and popcount operations. Unlike full-precision convolutions that require floating-point operations, binary convolutions rely on bitwise operations, leading to up to 64 times computation savings.

2.1 Design of Basic Binary Convolution Units for Image Restoration

2.1.1 *Build for BBCU-V1.*

The principle and implementation process of Basic binary convolurion unit (BBCU) are relatively simple and intuitive. The core idea is to binarize the weights and activations of the neural network into +1 and -1 to reduce the model's storage space and computational complexity.

The implementation process is as follows:

1. Weight Binarization: Convert the weights of the neural network into binary values, usually +1 and -1. This can be achieved by comparing the weight parameters with a threshold and binarizing them based on the comparison result.

2. Activation Binarization: Binarize the activation values of the neural network, also using +1 and -1 representation.

3. Utilize Binary Convolution: Replace traditional floating-point multiplication and addition operations with bitwise operations such as bitwise XNOR and bit counting to perform convolution operations.

4. apply Appropriate Activation Function: use RPReLU to handle the binarized activation values.

2.1.2 *The variant of BBCU.*

Based on the basic BBCU model (BBCU-V1), four variants of BBCU

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

are used for binarization in this paper. Compared to BBCU-V1, BBCU-V2 reduced by 0.14dB and 0.24dB on Urban100 and Manga109, respectively. This is because BBCU-V2 only removes BatchNorm (BN), making the value range of binary convolution much larger than that of residual connections, covering more full-precision information (Figure 3).

BBCU-V3 adds a residual alignment scheme on the basis of BBCU-V2, solving the problem of value domain imbalance between binary convolution and residual connections, and removes BN. Since BN is detrimental to the image restoration network, BBCU-V3 exceeds BBCU-V1 by 0.17dB, 0.16dB, and 0.29dB on Set5, Urban100, and Manga109, respectively.

BBCU-V4 moves the activation function into the residual connection, retaining the full-precision negative value information in the residual connection (Figure 3). Therefore, the performance of BBCU-V4 is better than that of BBCU-V3.

3 HOW THE PROPOSED WORK IS DIFFERENTDIFFERENT

Based on the research on the basic binary convolutional unit for image restoration at ICLR2023, we can further explore the following research directions:

Improving Accuracy: Enhance the accuracy of image restoration models by refining the design and algorithms of the basic binary convolutional unit. This may involve optimizing network architectures, improving loss functions, and introducing more effective feature representations.

Reducing Time Lag and Enhancing Real-time Prediction: Investigate methods to decrease processing time lag and achieve better real-time prediction performance in image restoration tasks. This could include optimizing model inference processes, improving parallel computing strategies, and reducing processing steps.

Lowering Computation and Resource Consumption: Explore approaches to further reduce the computational and memory resource consumption of image restoration models. This could be achieved through model compression techniques, quantization methods, and lightweight network designs.

These research directions will contribute to advancing the development of image restoration techniques based on binary convolutional units, improving their performance in terms of accuracy, efficiency and resource consumption.

4 PRELIMINARY PLAN(MILESTONES)

- By Mar. 7, finish the investigation in literature.
- By Mar. 20, exploring methods to improve BBCU model.
- By Apr. 1, implement code for core experiment.
- By Apr. 15, finish Final Project Report.

REFERENCES

[1] Bin Xia, Yulun Zhang, Y. W. Y. T. W. Y. R. T. L. V. G. Basic binary convolution unit for binarized image restoration network. In *conference paper* (2023).