Image-Adaptive Context Modeling for Compression Based on Implicit Neural Representations

Shoko Tanaka
Graduate School of FSE, Waseda University
Tokyo, Japan
shoko_official@akane.waseda.jp

Takahiro Shindo
Graduate School of FSE, Waseda University
Tokyo, Japan
taka_s0265@ruri.waseda.jp

Hiroshi Watanabe
Graduate School of FSE, Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*—Over the past decades, numerous approaches have been proposed to compress images while maintaining high image quality and reducing computational cost. Traditional approaches such as JPEG rely on linear transformations, whereas more recent neural network based methods often referred to as learned image compression achieve high performance by training on large datasets. Implicit Neural Representations (INRs) have also emerged as a promising alternative. INRs overfit a compact neural network to a single image, enabling competitive compression performance with a lightweight decoder. Additionally, context models have been used to predict image features in order to reduce the overall bitrate. However, existing INR-based methods typically utilize a fixed context across all images, which may limit its adaptability to image-specific structures. In this work, we propose an adaptive context modeling method that constructs image-specific contexts to improve the rate-distortion performance. The proposed adaptive context is integrated into the INR-based compression model, and its effectiveness is evaluated in comparison with conventional methods.**

***Index Terms*—Context model, implicit neural network.**

I. Introduction & Related works

Recent advances in image compression have introduced Implicit Neural Representations (INRs) as a promising alternative to conventional methods such as JPEG [1] and auto-encoder-based models. Unlike learned image compression models trained on large datasets, INR-based approaches overfit a compact neural network to a single image, offering competitive performance with significantly reduced decoder complexity.

A notable example is the Cool-ChiC (C3) [2], [3], which formulates image compression as an optimization problem balancing rate and distortion as follows:

$$\min D(x, \hat{x}) + \lambda R(\hat{x}). \tag{1}$$

The C3 model consists of four components: an autoregressive model f_ψ, multi-resolution latents $\hat{y}$, an upsampling network f_v, and a synthesis network f_θ. The autoregressive model serves as a context model to estimate latent values and guide entropy coding. The latent features are structured as a hierarchy of L spatial tensors at multiple resolutions, represented as follows:

$$\hat{y} = \{\hat{y}_k \in \mathbb{Z}^{H/2^k \times W/2^k} \mid k = 0, \ldots, L-1\}. \tag{2}$$

These latent features are first upsampled to match the resolution of $\hat{y}_0$ using the upsampling function f_v, and then decoded into the image using the synthesis network f_θ. In the original C3 implementation, the autoregressive model f_ψ utilizes a fixed context configuration that is applied to all input images. Although this approach demonstrates effectiveness, it lacks the flexibility to adapt to image-specific variations. To address this limitation, we propose an adaptive context mechanism that dynamically selects the most relevant features based on the content of each image. This approach aims to improve the accuracy of feature prediction.

II. Proposed method

To enhance the adaptability of INR-based image compression, we propose a context modeling approach that dynamically selects features for each image. This approach reduces the decoding bitrate by selecting the most relevant neighboring features for latent prediction. An overview of the proposed framework is shown in Fig. 1. The original C3 framework uses a fixed context of 24 neighboring features. In contrast, our method adaptively configures the context for each image

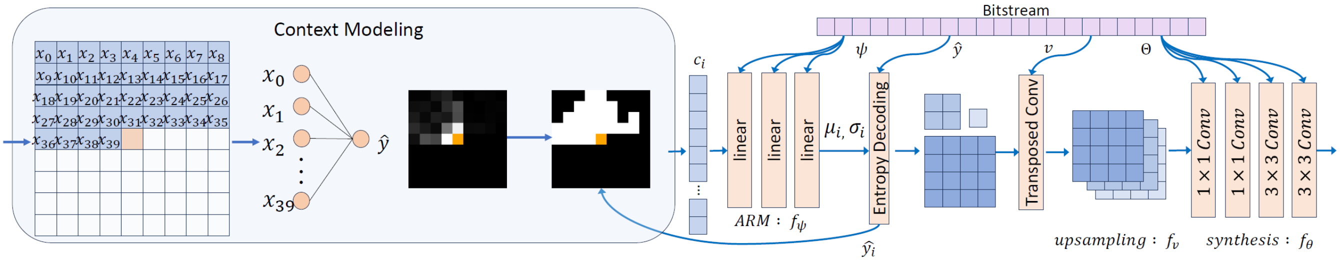

Fig. 1. Decoding process of the proposed method. Cool-chic decoder diagram is from Blard et al. [2].

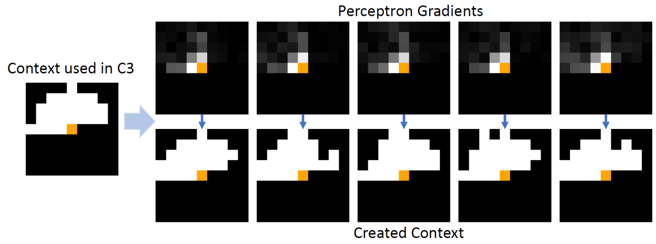

Fig. 2. Context created by the proposed method (target pixels in the middle, white area is the context used for prediction).

to improve the accuracy of latent predictions. To achieve this, we introduce a context modeling module that selects related features by evaluating the contribution of spatially adjacent features using a single-layer perceptron. Specifically, for each target y, a set of 40 neighboring features located in the upper-left region is fed into the perceptron, which is trained to predict y based on these inputs. After training, we compute the gradients with respect to each input to estimate their individual contributions. The 40 input features are ranked by gradient magnitude, and the top 10 are selected as the primary context. To complete the context configuration, an additional 14 features are selected from the remaining candidates in the original fixed context used in C3. These are chosen based on their contribution scores, excluding any features already selected in the top 10. Finally, the fixed 24-feature context in C3 is replaced with an image-specific configuration to improve prediction accuracy and rate-distortion performance.

III. Experiments

We integrate our context modeling approach into the C3 compression framework to evaluate its effectiveness. Its performance is then compared with that of the original C3 method. Experiments are conducted on 24 images from the Kodak dataset [4]. Each image is compressed using a customized context generated by our method. An example of the resulting contexts is shown in Fig. 2. Out of the 24 test images, 15 resulted in the same context configuration as the original C3, while the remaining 9 adopted distinct image-specific contexts.

To assess the impact of these adaptive contexts, we train the C3 model using the 9 images whose contexts differ from the default. The rate-distortion trade-off parameter λ in (1) is set to 0.0001, 0.00015, 0.0002, and 0.0003, while the number of resolution levels L in (2) is fixed at 7. We compare the proposed method with the original C3 in terms of compression efficiency. As shown in Fig. 3, our method achieves better performance at higher bitrates, demonstrating the effectiveness of adaptive context selection. However, the performance improvement becomes less significant at lower bitrates. In the high bitrate region, our method achieves a BD-rate reduction of 0.44% compared to the baseline.

IV. Discussion

Based on these results, we analyze the trends observed in compression performance based on these results. First, we examine why the proposed method outperforms the baseline at high bitrates but shows limited improvement at lower bitrates. The adaptive context is constructed based on the original image, which shares the same resolution as the latent representation $\hat{y}_0$ defined in (2). When the parameter λ in (1) is small, the optimization emphasizes fidelity reconstruction of the input image, which results in $\hat{y}_0$ retaining rich information. As λ increases, the optimization shifts toward bitrate reduction, leading to a decrease in the information content of $\hat{y}_0$ and thus diminishing the benefit of context adaptation.

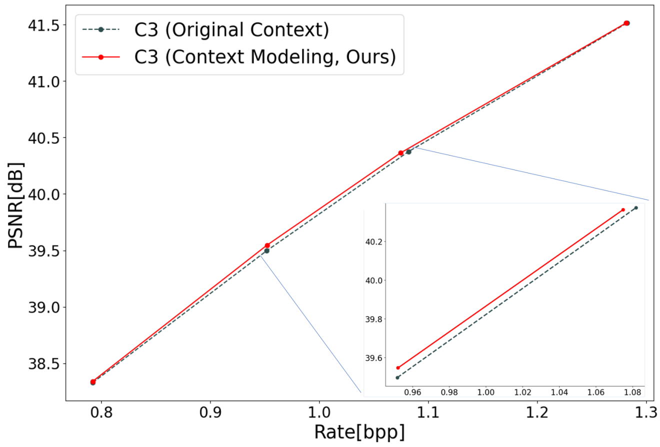

Fig. 3. Rate-distortion curves for the proposed method and original C3 at λ = 0.00015 and 0.0002 on the Kodak dataset.

Second, we examine why the compression gains remain relatively modest despite using adaptive contexts. As shown in Fig. 2, the selected contexts often overlap with those in the original C3. When the default context already includes most of the informative features, adaptive context offers limited additional benefit. Therefore, the effectiveness of the proposed approach depends on the extent to which contexts distinct from the original ones enhance prediction.

V. Conclusion

In this paper, we propose a context modeling approach that adaptively constructs image-specific contexts to enhance INR-based image compression. The proposed method leverages a perceptron to evaluate the contribution of neighboring features to prediction accuracy and selects the most informative features as context. Experimental results demonstrate that tailoring the context to each image leads to improved compression performance, particularly at high bitrates. Future research will focus on optimizing the number of contextual features to enhance rate-distortion efficiency.

References

[1] G. K. Wallace, “The JPEG still picture compression standard,” in IEEE Transactions on Consumer Electronics, vol. 38, no. 1, Feb. 1992, pp. 30-44.

[2] T. Ladune et al., “Cool-chic: Coordinate-based low complexity hierarchical image codec,” In: Proceedings of the IEEE/CVF International Conference on Computer Vision, Oct. 2023, pp. 13515-13522.

[3] H. Kim et al., “C3: High-performance and low-complexity neural compression from a single image or video,” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2024, pp. 9347-9358.

[4] Kodak image dataset. http://r0k.us/graphics/kodak/.