Supplementary of Semantically Structured Image Compression via Irregular Group-Based Decoupling

Ruoyu Feng[1*] Yixin Gao[1*] Xin Jin[2] Runsen Feng[1] Zhibo Chen [1,†]
[1]University of Science and Technology of China [2]Eastern Institute of Technology, Ningbo

This supplementary material provides additional experimental comparison, implementation details of semantically-aware encryption, and ablation study of taking all foreground objects as regions of interest.

A. Comparison with Other Methods

In this section, we compare our proposed method with VTM 18.2 and three other methods designed for image coding for machines: the traditional codec based RoI bit-allocation scheme [4], the learning based joint training codec [5], and the general representation learning based approach [1]. The evaluation task is instance segmentation on COCO 2017 [6]. It is noteworthy that the bitstreams of RoI bit-allocation [4] and the task-driven joint training methods are tailored to the corresponding task, and the general representation learning based method [1] necessitates retraining the task model with new source data (i.e., the learned representation). Nevertheless, our method still yields significant improvement compared with other methods.

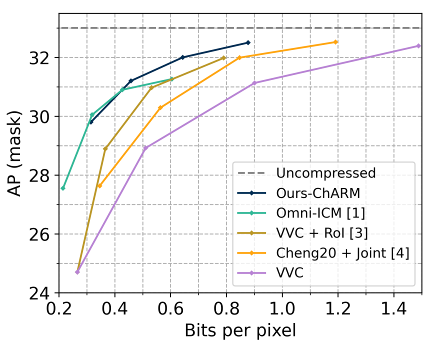
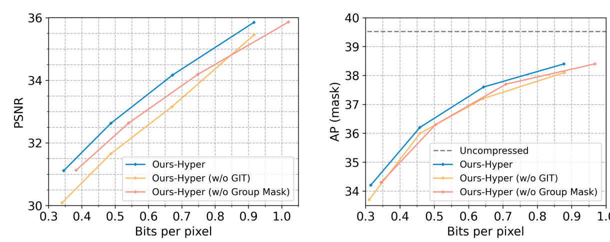


Figure 1. Instance segmentation on MS COCO.

B. Semantically-Aware Encryption

We conduct encryption on latent variables. Specifically, for each group to be encrypted, the chosen variables are first reshaped into a one dimensional vector, and then perturbed by Fisher-Yates shuffle algorithm [2]. The perturbance is achieved through a cryptographic random seed. And, the perturbed vector is rearranged into its original shape and encoded into the corresponding bitstream. Correspondingly, the likelihood is shuffled in the same way for entropy coding. Then, entropy coding are conducted on the latent variables of current group, resulting in the encrypted bitstream. At last, the semantically structured bitstream consists of both encrypted and non-encrypted ones according to requirements. Furthermore, the user can perform different levels of encryption on different groups, resulting in the layered encrypted bitstream.

C. Ablation Study



(a) RoI reconstruction. (b) Instance Segmentation.

Figure 2. Ablation study.

In Section 5 of the main text, we perform ablation studies by regarding people as the regions of interest (RoI). In this section, we provide additional experimental results, where we consider all foreground objects belonging to MS COCO's categories as the RoI. The results shown in Figure 2 indicate the same conclusion as Section 5 of the main text, that our proposed group mask can significantly save bitrate while the group-independent transform can ensure the reconstruction quality of selective transmission and reconstruction.

D. Complexity Analysis

D.1. Parameter-Performance Trade-Off

We evaluated the parameter-performance trade-off of our approach under two scenarios: with and without a structured bitstream (Struc.), and compared it with the transformer-based method Zhu2022 [7] and the CNN-based

---

* First two authors contributed equally.
† Corresponding author.

method ELIC [3]. The comparison includes decoding time (Dec.), parameters (Par.), complexity (FLOPs) and coding efficiency (BD rate). The results presented in Table 1 indicate that our method can achieve comparable RD performance and decoding complexity while saving parameters compared with the other two SOTA codecs. For the entire image reconstruction, semantically structured bitstream (SSB) causes a increase in decoding time and a slight BD-rate drop. This is because the bitstreams of all groups are individually stored during encoding, resulting in a small amount of additional overhead and the traversal of all groups decoding. However, given the flexibility and functionality of SSB, we consider the additional cost of decoding complexity and bitrate to be acceptable.

Table 1. Comparison of model complexity and BD-rate (%) over VTM-18.2 for entire image reconstruction on Kodak.

Model	Struc.	Dec.	Par. (M)	FLOPs (G)	BD-rate
Zhu2022	×	84	60.55	230.37	-2.32
ELIC (repro.)	×	90	28.29	179.69	-4.10
Ours-ChARM	×	96	23.13	193.42	-2.78
Ours-ChARM	✓	258	23.13	193.42	-1.89

D.2. Analysis regarding run time for different aspects.

Table 2 shows the run-time analysis for different aspects of our approach. Semantically structured bitstream enables us only decode the bitstreams of corresponding groups for partial reconstruction or specific downstream tasks, resulting in a proportional reduction of YDec.

Table 2. Decoding complexity (ms) of our model under scenarios of with and without structured bitstream (Struc.), in terms of the whole decoding time (Dec.), time of decoding latent $\hat{\boldsymbol{y}}$ (YDec.), time of decoding hyper $\hat{\boldsymbol{z}}$ (ZDec.), time of the synthesis transform in the decoding process (Inf.).

Testset	Struc.	Dec.	YDec.	ZDec.	Inf.
Kodak (entire)	×	96	54	2	40
Kodak (entire)	✓	258	214	2	42
Kodak (partial)	✓	149	106	2	41

References

[1] Ruoyu Feng, Xin Jin, Zongyu Guo, Runsen Feng, Yixin Gao, Tianyu He, Zhizheng Zhang, Simeng Sun, and Zhibo Chen. Image coding for machines with omnipotent feature learning. In *ECCV*, pages 510–528. Springer, 2022. 1

[2] Ronald Aylmer Fisher, Frank Yates, et al. *Statistical tables for biological, agricultural and medical research, edited by ra fisher and f. yates*. Edinburgh: Oliver and Boyd, 1963. 1

[3] Dailan He, Ziming Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In *CVPR*, pages 5718–5727, 2022. 2

[4] Zhimeng Huang, Chuanmin Jia, Shanshe Wang, and Siwei Ma. Visual analysis motivated rate-distortion model for image coding. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021. 1

[5] Nam Le, Honglei Zhang, Francesco Cricri, Ramin Ghaznavi-Youvalari, and Esa Rahtu. Image coding for machines: An end-to-end learned approach. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1590–1594. IEEE, 2021. 1

[6] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755. Springer, 2014. 1

[7] Yinhao Zhu, Yang Yang, and Taco Cohen. Transformer-based transform coding. In *ICLR*, 2022. 1