

Supplementary Materials: Spatial-Temporal Context Model for Remote Sensing Imagery Compression

Anonymous Authors

Due to the space limitation in the main paper, we provide more details on implementation and experimental results in the supplementary material.

1 DOWNSTREAM TASKS VALIDATION ON SEMANTIC SEGMENTATION

In the main paper, we present the validation results for downstream remote sensing tasks in scene classification based on the UC Merced Land dataset. Additionally, we evaluate the effects of reconstructed images from various compression methods on the performance of a fine-grained downstream task, specifically semantic segmentation, using the ISPRS-Potsdam dataset. Details of these datasets are shown in Table 1.

The effects of reconstructed images for semantic segmentation tasks are shown in Table 2. We marked the results with similar bpp in blue for easy comparison. Our context model exhibits the least degradation in downstream tasks at similar bpp. It is also noteworthy that when our context model is combined with the backbone network of HL_RSCompNet, our approach achieves relatively satisfactory results even at extremely low bpp, as indicated by the underlined data.

2 IMPLEMENTATION DETAILS

2.1 Detailed Network Structures of our STCM

In this section, we provide a detailed network illustration of our Spatial-Temporal Context Model (STCM). Figure 1 shows the network structures of the channel-wise, spatial, and temporal context-predicting modules in our context model and the structure of the parameter aggregation block used in the entropy model to combine context priors and hyper priors. In our approach, the stacking layer number of spatial context predicting can be adjusted arbitrarily according to data characteristics and model performance. We follow the uneven group division in ELIC[2] for the channel-wise context prediction, whose channel number ranges from 8 to 96 for each group.

2.2 Training and Evaluation Details

In this section, we provide more implementation details of the training process of our context model and other experimental details including downstream tasks.

Training data organization. The fMoW-full dataset includes RS images with different numbers of bands, such as 3, 4, and 8. To organize the training data, we randomly select three bands from the images to guide the model in learning to deal with spectral data with different wavelengths. The original fMoW dataset can be downloaded according to their official GitHub repository¹.

Downstream task implementation details. For scene classification tasks validation, we train the MSMatch[1] on the UCM. The

Table 1: Datasets Details for compression models and downstream application validation. UCM is for scene classification tasks. ISPRS-Potsdam dataset is for semantic segmentation task.

Datasets	GSD(m)	Resolution	Train/Test Num	Source
UCM	0.3	256×256	1680/420	Aerial
ISPRS-Potsdam	0.05	6000×6000	36/2	Aerial

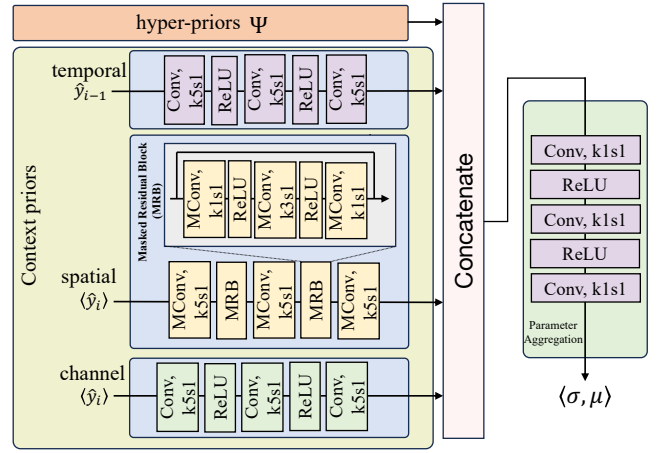


Figure 1: The network architectures of the channel, spatial, temporal context prediction modules in context model and parameter aggregation block in entropy model.

experiment is trained from scratch and without using additional data. The UCM can be obtained from the official website². We use 1680 labels in total for the training on UCM. For the division train and test sets, the test set contains 20% of the data in the UCM dataset (420 images). The code of MSMatch can be acquired from the authors' GitHub open-source repository³. For the semantic segmentation task, we train an LSKNet-T on the ISPRS-Potsdam dataset, which is based on the DCSwin[3]. We use a pretrained backbone provided by the authors and fine-tune it on ISPRS-Potsdam. All the data preparation procedures follow the instructions given in the GitHub repository.

REFERENCES

- [1] Pablo Gómez and Gabriele Meoni. 2021. MSMatch: Semisupervised multispectral scene classification with few labels. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), 11643–11654.
- [2] Dailan He, Ziming Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. 2022. Elic: Efficient learned image compression with unevenly grouped space-channel

²Official UCM dataset download website: <http://weegee.vision.ucmerced.edu/datasets/landuse.html>

³MSMatch open-source repository: <https://github.com/gomezms/MSMatch>

¹Official fMoW dataset repository: <https://github.com/fMoW/dataset>

Table 2: Comparison of downstream semantic segmentation task results based on reconstructed images from various methods and quality settings (i.e., bpp), using the ISPRS-Potsdam dataset. Blue highlights indicate results with similar bpp, while red highlights indicate the best result among these methods.

	No Comp	VVC		ELIC(Ori)		MLIC++		Ours+HL_RSComp		Ours+ELIC	
bpp	-	1.15	2.45	1.13	7.58	1.15	1.93	0.05	1.24	1.14	2.23
mIoU	0.8230	0.2807	0.3520	0.6771	0.7922	0.7968	0.8003	0.6554	0.8009	0.7937	0.7989
OA	0.8975	0.5646	0.6352	0.7973	0.8834	0.8813	0.8841	0.8008	0.8870	0.8829	0.8860

contextual adaptive coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5718–5727.

[3] Libo Wang, Rui Li, Chenxi Duan, Ce Zhang, Xiaoliang Meng, and Shenghui Fang. 2022. A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 19 (2022), 1–5.