A SHARED ENCODER FOR MULTI-SOURCE HYPERSPECTRAL IMAGES

Weili Kong, Baisen Liu*& Xiaojun Bi
School of Information and Communication Engineering
Harbin Engineering University
{kkweil,liubaisen,bixiaojun}@hrbeu.edu.cn

Jiaming Pei
School of computer science
The University of Sydney
jpei0906@uni.sydney.edu.au

ABSTRACT

Multi-source hyperspectral images(HSIs) which captured from diverse sensors commonly possess varying bands. When employing deep learning techniques for their processing, individual models are necessitated for each source due to the disparate dimensions. To tackle this problem, we propose a shared encoder to project all HSIs into a unified feature space. It establishes a general framework for the representation of multi-source HSIs, providing foundational conditions for the development of a universal HSI analysis model.

1 INTRODUCTION

Hyperspectral imaging technology is a method that utilizes spectral sensors to capture spectral information across hundreds or thousands of contiguous bands. It offers rich spectral features, facilitating object identification, environmental monitoring, and geoscientific research (Kong et al. (2023)). Nowadays, an increasing number of HSI capture missions are being deployed, such as MODIS, HypSEO, DESIS, Gaofen-5, EnMap, HyspIRI, and so on (Hong et al. (2021)). Owing to the distinct characteristics of data collected by various sensors, current methodologies for HSI feature extraction often necessitate individual models and training from scratch when dealing with multi-source HSIs[e.g., PCA (Jiang et al. (2018)), 3D-CAE (Nalepa et al. (2020)), CENet (Guo et al. (2022)) and so on]. In the very recent period, researchers proposed a 1D-CNN based autoencoder (Kuester et al. (2023)), which can adaptively process multi-source HSIs and has strong transferability. However, this method cannot construct a unified feature space because the dimensions of output features are not uniform. Hence, this method still falls short of addressing the stringent requirements of deep learning networks regarding data dimensions. In this work, we design a shared encoder to learn the general features of multi-source HSIs, it can project multi-source HSIs into the same feature space and ensure that they have the same feature dimension. Additionally, we further validated the generality of our method on multi-source HSIs through a series of experiments.

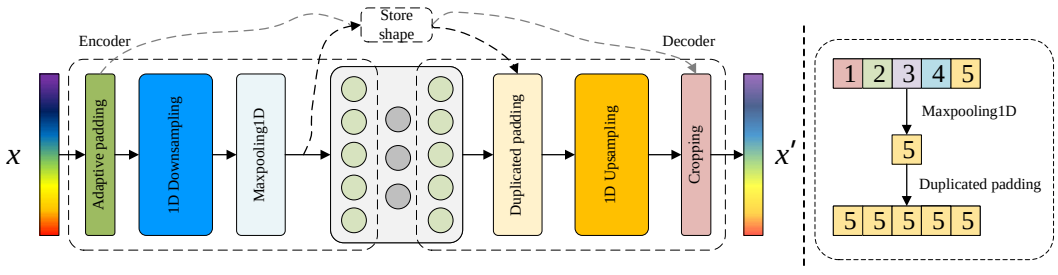

Figure 1: The architecture of shared encoder.

*Corresponding author and team leader. He is also with professor in Heilongjiang Institute of Technology.

2 METHODOLOGY

Notation: The shared encoder aims to provide a unified representation of multi-source HSIs, which only accepts one-dimensional spectral input. Consider 3D hyperspectral data as $\mathbf{H} \in \mathbb{R}^{h \times w \times b}$, where h, w and b represent the height, weigth and bands of HSI. Since spatial information is not considered in our method, we flatten the spatial dimensions $\mathbf{H} \rightarrow \mathbf{X} \in \mathbb{R}^{n \times b}$, where $n = h \times w$. The input to the model is a spectral signature $\boldsymbol{x} \in \mathbb{R}^{1 \times b}$.

Methodology: The overall architecture of shared encoder is illustrated in figure 1, where *'Store shape'* is used to save the two parameters Δb and $\hat{b}$ mentioned below. To enable the model to adaptively process the spectral signatures with different dimensions, we incorporated an adaptive padding layer and a cropping layer at the input and output end. The padding band numbers $\Delta b = \lceil b/r \rceil \times r - b$, where r represents the downsampling rate. In encoder section, we employ a 1D CNN based downsampling moudel $f_D : \mathbb{R}^{1 \times (b + \Delta b)} \rightarrow \mathbb{R}^{c \times \hat{b}}$ and a maxpooling layer $f_{max} : \mathbb{R}^{c \times \hat{b}} \rightarrow \mathbb{R}^{c}$ to extract the most significant factors and eliminate the dimension differences across multi-source HSIs, where $\hat{b} = (b + \Delta b)/r$ and c is the output channel of f_D which can be preset. It is worth noting that both Δb and $\hat{b}$ are computed by parameter-free operations, thus they will be preserved during the model inference process for reconstruction by the decoder. To extract higher-level features, we construct a bottleneck-structured MLP, the bottleneck feature is selected as the feature vector, which is utilized for subsequent tasks. On the decoder side, a repetitive padding layer, a 1D transposed CNN based upsampling moudel and a cropping layer are used to perform the reverse operations corresponding to the encoder. The shared encoder is trained as equation 1, where θ^* represents the parameters of shared encoder, $||\theta||_2^2$ represents the L2 regularization of θ and α is its coefficient.

$$\theta^* = \arg\min_{\theta} \frac{1}{n} \sum_{i=0}^{n} |f(\boldsymbol{x}_i, \theta) - \boldsymbol{x}_i|^2 + \alpha ||\theta||_2^2 \tag{1}$$

3 EXPERIMENTS

In our experiments, PSNR and SSIM are employed to quantify the similarity between reconstructed and original images, thereby assessing the correctness of the features. Ablation experiments are conducted using a classification task as the downstream task to assess the performance improvement of our method on it. Given the absence of a universal model applicable to multi-source HSIs at present, there are no analogous methods available for comparative analysis. The details of the datasets and experimental results are provided in *Appendix A.1* and *A.2*, respectively.

Table 4 presents our reconstruction results on muti-source HSIs from different and unknown sensors. These results indicate that our model has achieved excellent reconstruction performance. On the three testsets, when $r = 32$, the PSNR are $26.84dB$, $28.81dB$, $29.40dB$, and the SSIM are 0.78, 0.83, 0.85, respectively. This demonstrates that our model can extract highly reliable universal features from hyperspectral images of different sources. Besides, we fine-tuned the pre-trained model on a dense classification task, and the results are presented in table 5. Clearly, our approach demonstrates a significant performance improvement in classifying hyperspectral images from different sensors. With a 5% training sample, the OA increased by 1.95% and 6.2% on PU and IN, respectively. When utilizing only 1% of the training samples, the OA on PU and IP saw a more substantial improvement, with increases of 2.71% and 8.75%, respectively.

4 CONCLUSION

In this paper, we design a shared encoder which constructs a unified feature space for muti-source HSIs. The model is highly compatible with HSIs from various sources, requiring no structural adjustments after pretraining to accommodate different data structures. The uniform representation of muti-source HSIs provides the possibility to design a general model for HSI analysis. Experimental results indicate a significant improvement in the performance of our method on classification tasks across different datasets. In future work, we will emphasize the development of a general hyperspectral pre-training model to enhance the practical applicability of hyperspectral data in production environments.

ACKNOWLEDGEMENTS

The work was supported by the Natural Science Foundation of Heilongjiang Province for Key projects, China (Grant no. ZD2021F004) and the 2023 Heilongjiang Province Art and Science Planning Key Projects (Grant no. 2023A010).

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

REFERENCES

Marion F. Baumgardner Larry L. Biehl and David A. Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3, Sep 2015. URL https://purr.purdue.edu/publications/1947/1.

Yuanyuan Guo, Yulong Tao, Yanwen Chong, Shaoming Pan, and Miao Liu. Edge-guided hyperspectral image compression with interactive dual attention. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–17, 2022.

Danfeng Hong, Wei He, Naoto Yokoya, Jing Yao, Lianru Gao, Liangpei Zhang, Jocelyn Chanussot, and Xiaoxiang Zhu. Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, 9(2):52–87, 2021.

Junjun Jiang, Jiayi Ma, Chen Chen, Zhongyuan Wang, Zhihua Cai, and Lizhe Wang. Superpca: A superpixelwise pca approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4581–4593, 2018.

Weili Kong, Baisen Liu, Xiaojun Bi, Jiaming Pei, and Zheng Chen. Instructional mask autoencoder: A scalable learner for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

Jannick Kuester, Wolfgang Gross, Simon Schreiner, Wolfgang Middelmann, and Michael Heizmann. Adaptive two-stage multisensor convolutional autoencoder model for lossy compression of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

Jakub Nalepa, Michal Myller, Yasuteru Imai, Ken-ichi Honda, Tomomi Takeda, and Marek Antoniak. Unsupervised segmentation of hyperspectral images using 3-d convolutional autoencoders. *IEEE Geoscience and Remote Sensing Letters*, 17(11):1948–1952, 2020.

Kun Tan, Fuyu Wu, and Xue Wang. Xuzhou hyspex dataset, 2018. URL https://dx.doi.org/10.21227/t3c9-h862.

Yanfei Zhong, Xin Hu, Chang Luo, Xinyu Wang, Ji Zhao, and Liangpei Zhang. Whu-hi: Uav-borne hyperspectral with high spatial resolution (h2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with crf. *Remote Sensing of Environment*, 250:112012, 2020. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2020.112012. URL https://www.sciencedirect.com/science/article/pii/S0034425720303825.

A APPENDIX

A.1 DATASETS

We utilized a total of 12 hyperspectral images from distinct sensors as experimental materials. Among them, Botswana, Houston, Indian Pines(IN)(Biehl & Landgrebe (2015)), KSC, NewXiongAn, Salinas, Xuzhou(Tan et al. (2018)), WHU-Hi-HanChuan, and WHU-Hi-LongKou(Zhong et al. (2020)) were employed as the training set for training the shared encoder, while Chikuse, WashingtonDC and the Pavia University(PU) served as the test set to evaluate the generality of model and the

accuracy of the extracted features. The detailed information for these datasets is provided in table 1. For the classification task, we selected via University and Indian Pines as benchmark datasets. The detailed category information is provided in table 2 and table 3. Note that when the training samples for each class are less than 1, they are automatically supplemented to reach 1.

Dataset	Sensor	Band	Spectral Range/**nm**	Spatial Resolution/**m**
Botswana	Hyperion	145	400-2500	30
Houston	ITRES CASI-1500	144	380-1050	2.5
Indian Pines	AVIRIS	220	400-2500	20
KSC	AVIRIS	176	400-2500	18
NewXiongAn	PHI	250	400—1000	0.5
Salinas	AVIRIS	224	400-2500	3.7
Xuzhou	HYSPEX	436	415-2508	0.73
WHU-Hi-HanChuan	Nano-Hyperspec	274	400-1000	0.109
WHU-Hi-LongKou	Nano-Hyperspec	270	400-1000	0.463
Chikusei	Hyperspec-VNIR-C	128	343 – 1018	2.5
WashingtonDC	Hydice	191	400-2400	-
Pavia University	ROSIS	103	430-860	1.3

Table 1: The detailed information of the datasets used.

No.	Class	1% Training	5% Training	Total
1	Alfalfa	1	2	46
2	Corn-notill	14	71	1428
3	Corn-mintill	8	41	830
4	Corn	2	11	237
5	Grass-pasture	4	24	483
6	Grass-trees	7	36	730
7	Grass-pasture-mowed	1	1	28
8	Hay-windrowed	4	23	478
9	Oats	1	1	20
10	Soybean-notill	9	48	972
11	Soybean-mintill	24	122	2455
12	Soybean-clean	5	29	593
13	Wheat	2	10	205
14	Woods	12	63	1265
15	Buildings-Grass-Trees	3	19	386
16	Stone-Steel-Towers	1	4	93
	Total	98	505	10249

Table 2: Land cover classes illustration and numbers of training and testing samples for India Pines.

No.	Class	1% Training	5% Training	Total
1	Asphalt	66	331	6631
2	Meadows	186	932	18649
3	Gravel	20	104	2099
4	Trees	30	153	3064
5	Mental sheets	13	67	1345
6	Bare soil	50	251	5029
7	Bitumen	13	66	1330
8	Bricks	36	184	3682
9	Shadow	9	47	947
	Total	423	2135	42776

Table 3: Land cover classes illustration and numbers of training and testing samples for Pavia University.

A.2 Experimental Results

Table 5 documents the performance of our method on the classification task. Given the limited samples, we conducted ablation experiments on our approach, testing the accuracy with randomly

Shared Encoder	Chikuse		WashingtonDC		PaviaU	
	PSNR $[dB]$	SSIM []	PSNR $[dB]$	SSIM []	PSNR $[dB]$	SSIM []
$r = 8$	24.57	0.66	26.98	0.69	27.89	0.73
$r = 16$	25.82	0.71	27.64	0.77	28.63	0.79
$r = 32$	**26.84**	**0.78**	**28.81**	**0.83**	**29.40**	**0.85**

Table 4: Evaluation of the reconstruction accuracy of shared encoder for the downsampling rates $r = 8$, $r = 16$, and $r = 32$ on different datasets from different hyperspectral sensors based on the metric PSNR, and SSIM. The best results are highlighted in bold.

initialized model parameters and the accuracy with parameters initialized using pretraining. We employed overall accuracy (OA), average accuracy (AA), and the Kappa coefficient as performance metrics.

	1% IN			1% PU			5% IN			5% PU		
	OA [%]	AA [%]	KAPPA [%]	OA [%]	AA [%]	KAPPA [%]	OA [%]	AA [%]	KAPPA [%]	OA [%]	AA [%]	KAPPA [%]
W/O Pertrain	43.44	39.42	35.40	87.48	81.71	83.51	71.32	59.98	67.14	91.77	87.02	89.06
Pertrain	52.19	41.82	45.75	90.19	84.32	86.52	77.52	73.21	74.34	93.72	90.86	91.67
Differ	8.75	2.4	10.35	2.71	2.61	3.01	6.2	13.23	7.2	1.95	3.84	2.61

Table 5: The results of classification. W/O: Without.