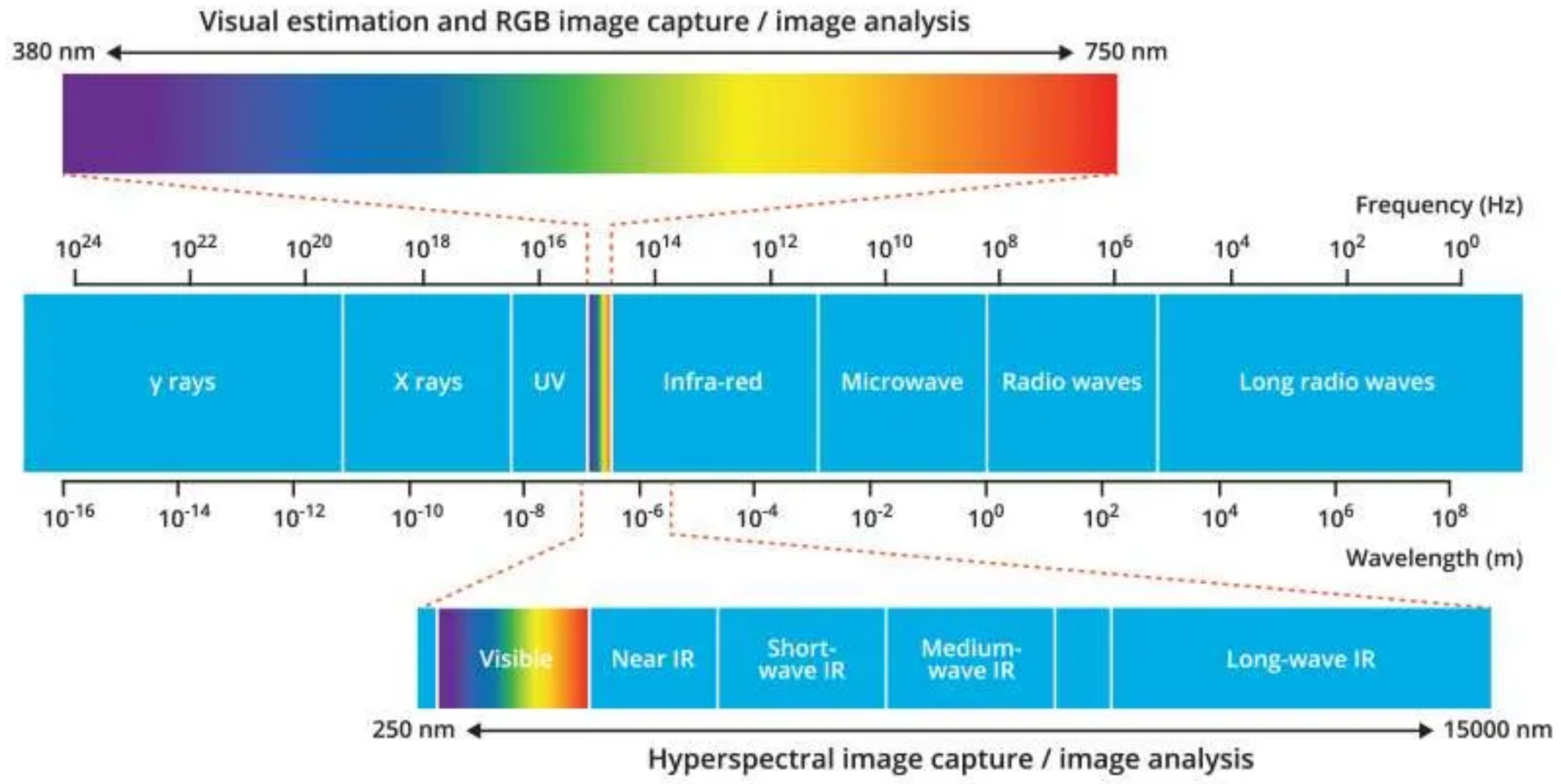
A Shared Encoder for Muti-source Hyperspectral Image

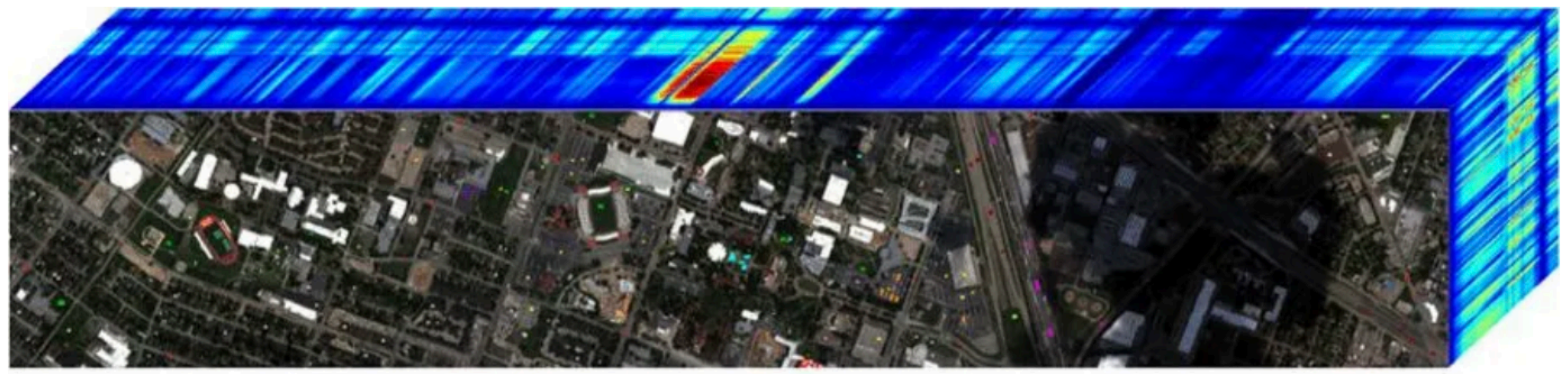
Harbin Engineering University
School of Information and Communication Engineering
WEILI KONG, BAISEN LIU

Characteristics of Hyperspectral Image(HSI)

Hyperspectral Images provide rich information of the observed objects. It enables researchers to conduct meticulous analysis of them beyond what is possible with RGB image or multispectral Image. This has made hyperspectral image be the preferred tool of choice and a key component in a wide range of Earth Observation applications, including land use/land cover mapping, weather forecasting, energy resource development, biodiversity conservation, and geological exploration.

- ✦RGB images
  - 3 bands in 430*nm*-750*nm*
- ✦Multispectral images
  - tens of bands within 400*nm*-2400*nm*
- ✦Hyperspectral images
  - hundreds of bands within 400*nm*-2400*nm*
    - wider spectral coverage
    - high spectral resolution
    - Detailed information





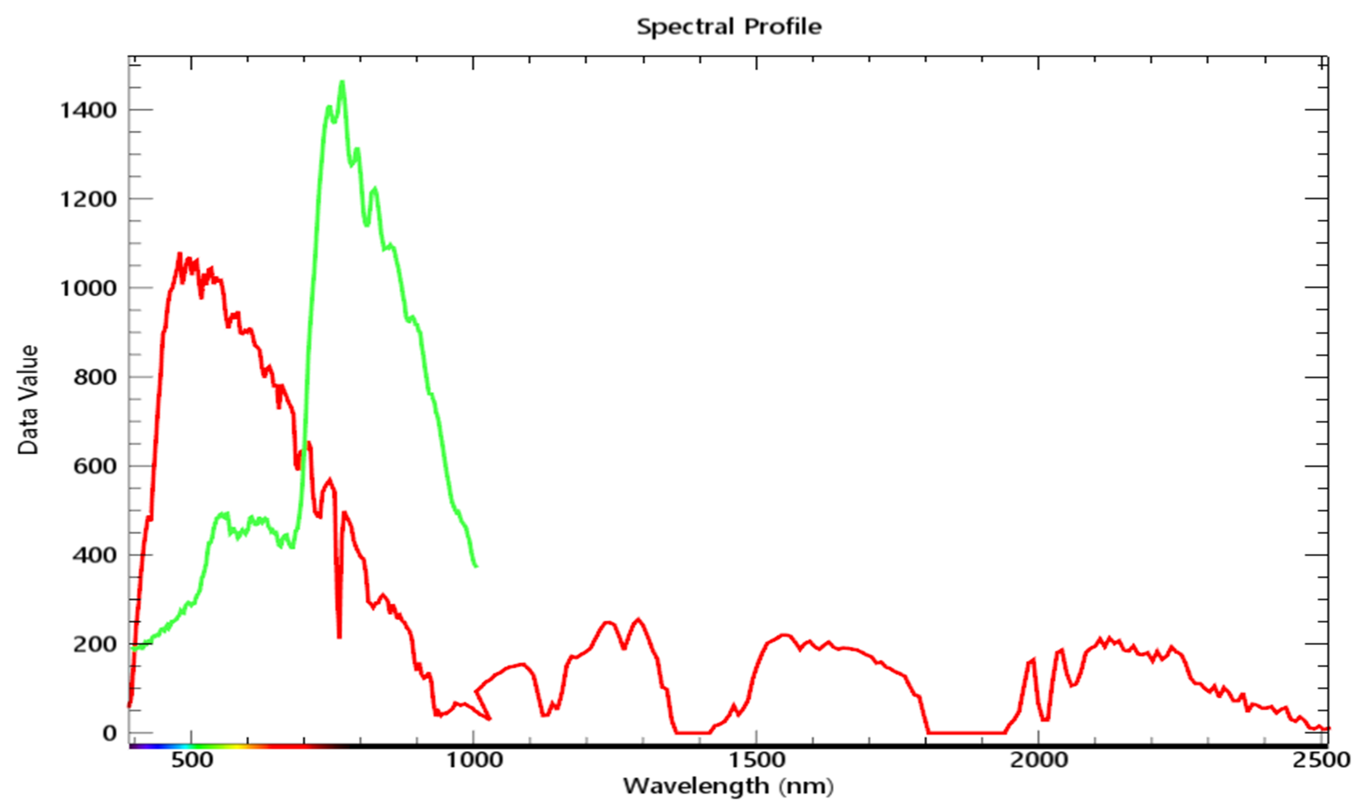
IEEE GRSS Huston2013. Wavelength from 364nm-1046nm with 144 bands.

When Deep Learning Comes to Hyperspectral Images

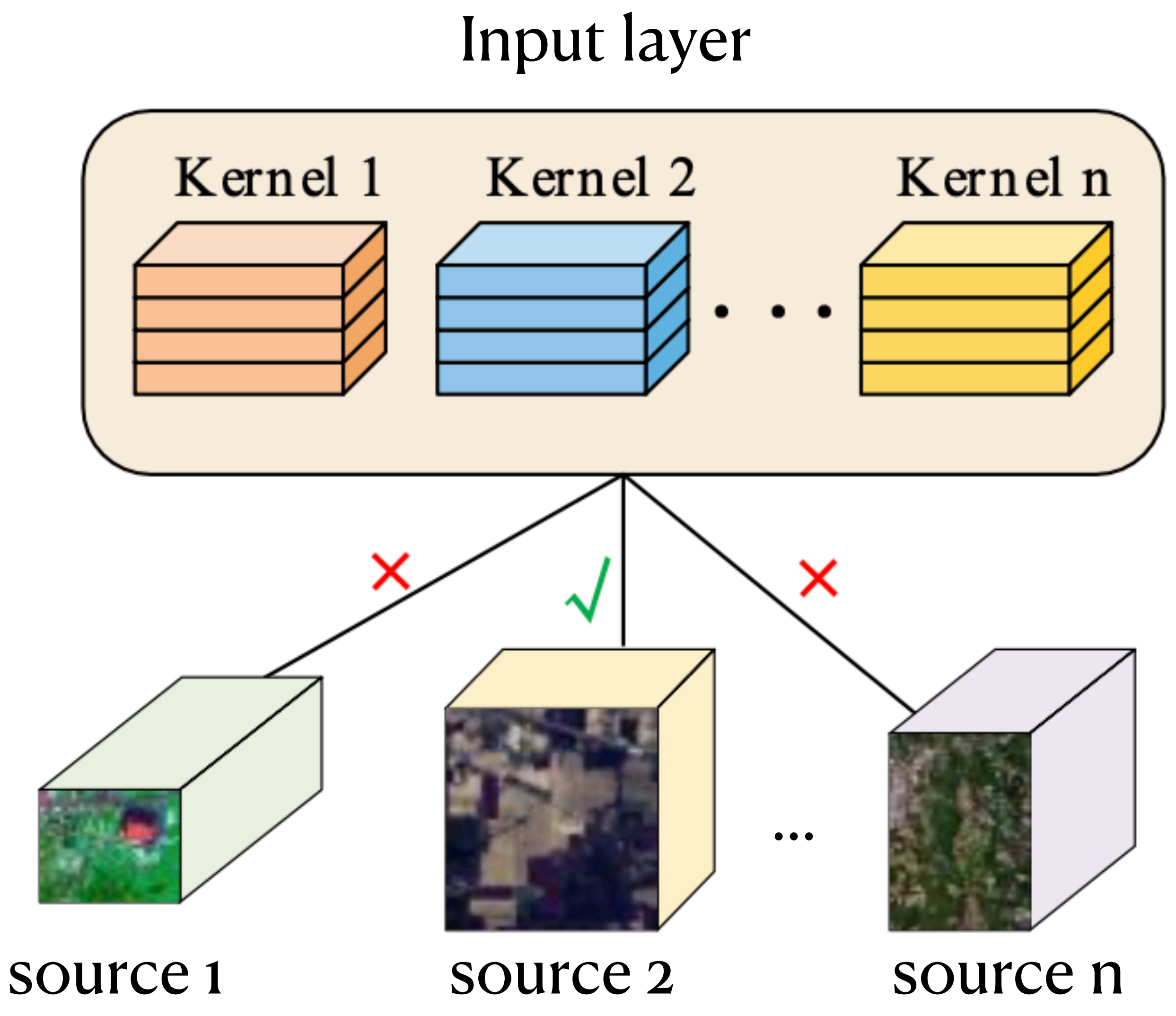
—Limitations also arise due to more spectral information

Limitations:

- High correlations between the bands.
- High input dimension, more model parameters.
- Extremely high labelling cost, limited training samples, overfitting.
- Independent model for each data source, weak transfer ability, low data utilization.



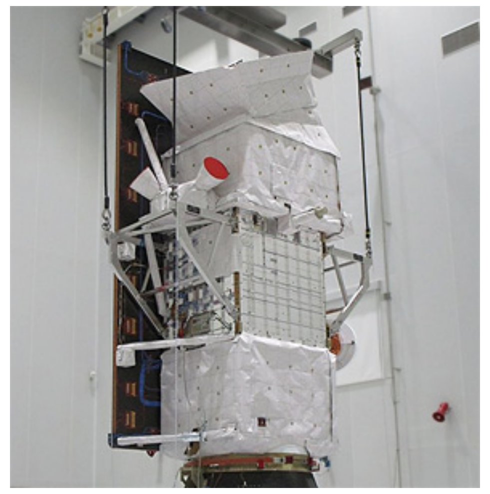
Different distribution

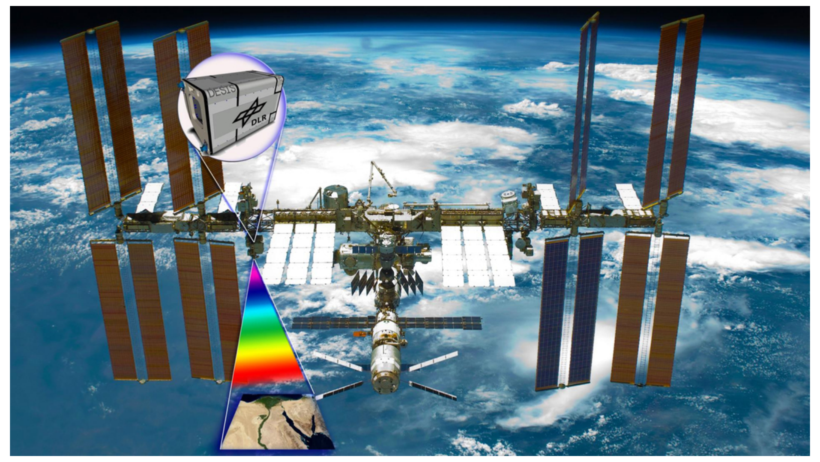


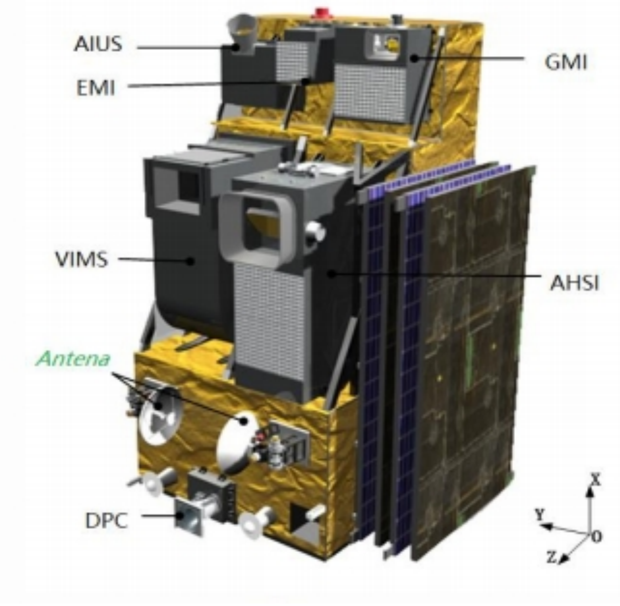
Different structure

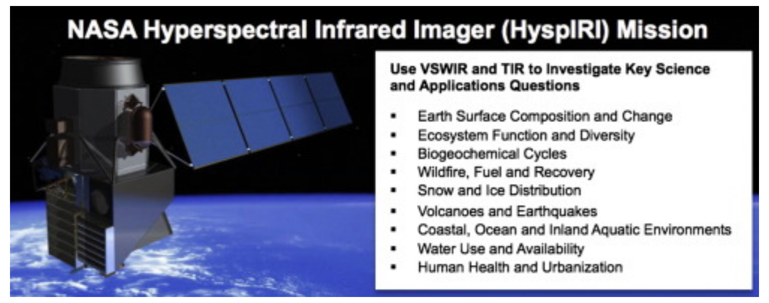
Motivation

Nowadays, an increasing number of HSI capture missions are being deployed, such as MODIS, HypSEO, DESIS, Gaofen-5, EnMap, HyspIRI, and so on. Due to the different features 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









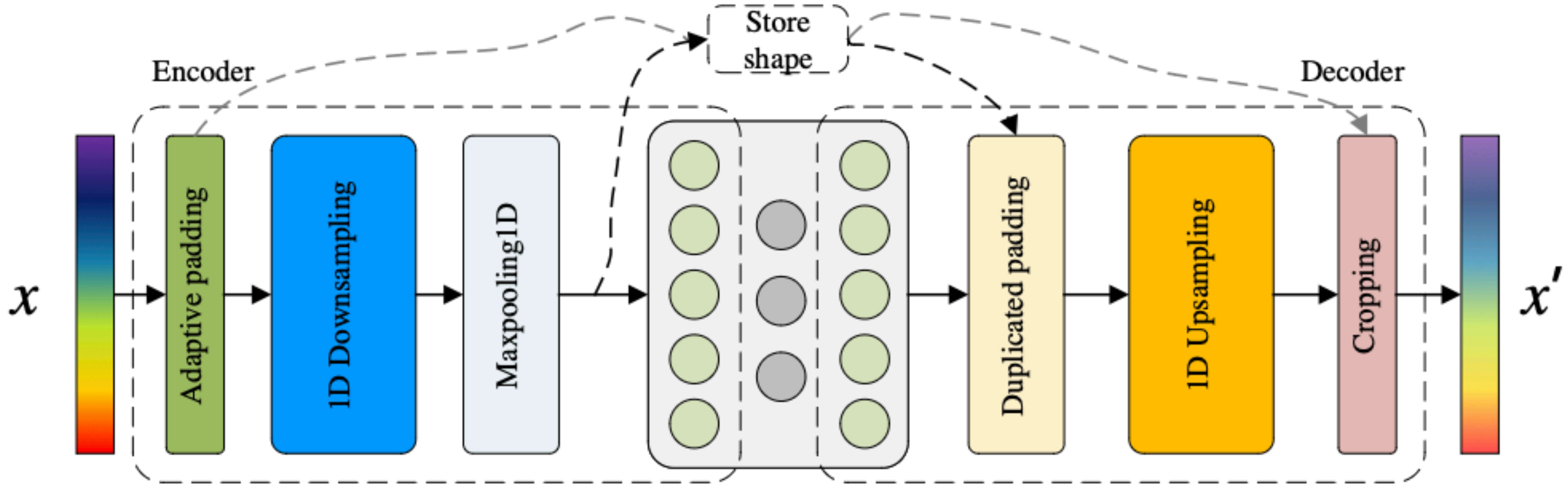
Platforms	spectral range	spectral resolution	spectral bands
MODIS	400nm-14400n	-	32
HypSEO	400-2500nm	12 nm	239
DESIS	400 – 1000 nm	2.55 nm	235
Gaofen-5	400-2500nm	5nm	330
EnMap	420-2450nm	5nm	244
HyspIRI	380–2500nm	10nm	212

Shared encoder provides a unified feature space for multi-source hyperspectral data, allowing a large amount of multi-source hyperspectral data to be used to train pre-trained foundational models. Thereby reducing the cost of labeling, improving the adaptability of the model as well as the utilization of data.

How?

Methodology

- Adaptive padding & adaptive cropping
- 1D-CNN based Autoencoder
- Maxpolling and repetitive padding



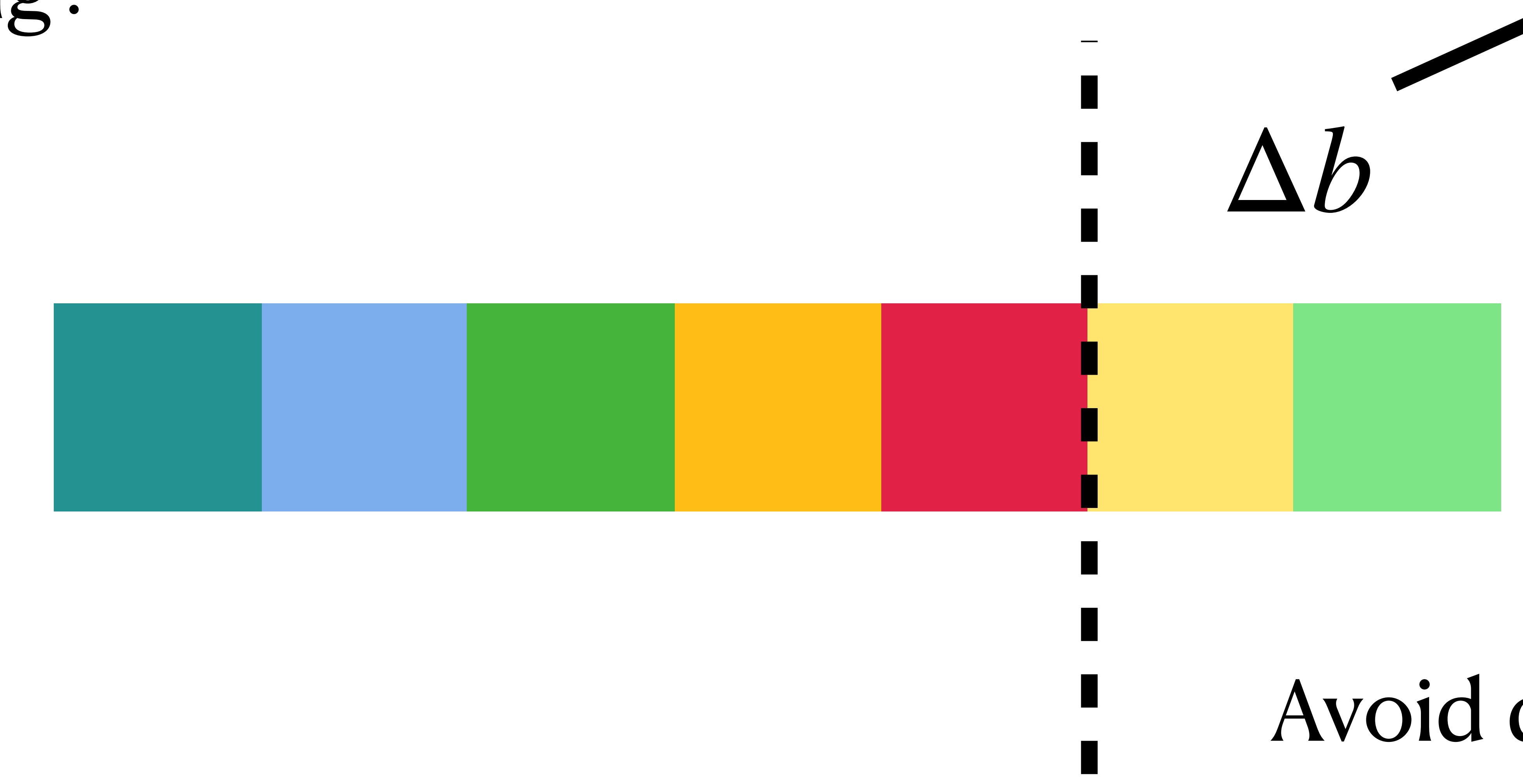
- Adaptive padding & adaptive cropping

Enable model with the ability to dynamically process and transfer across various spectral signature dimensions

$$\Delta b = b - \lfloor b/r \rfloor \times r$$

Reflection padding:

Δb



Δb will be saved in model to decide how many bands will be cropped in the decoder section.

Avoid disrupting data continuity.

- 1D-CNN based Autoencoder

We employ a 1D CNN based autoencoder as the backbone of our method. Its encoder is a downsampling 1D CNN, and the decoder is an upsampling 1D CNN.

$$f_D : \mathbb{R}^{1\times(b+\Delta b)} \to \mathbb{R}^{c\times\hat{b}}$$

$$f_U : \mathbb{R}^{c\times\hat{b}} \to \mathbb{R}^{1\times(b+\Delta b)}$$

The shared encoder will be optimized as follow:

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

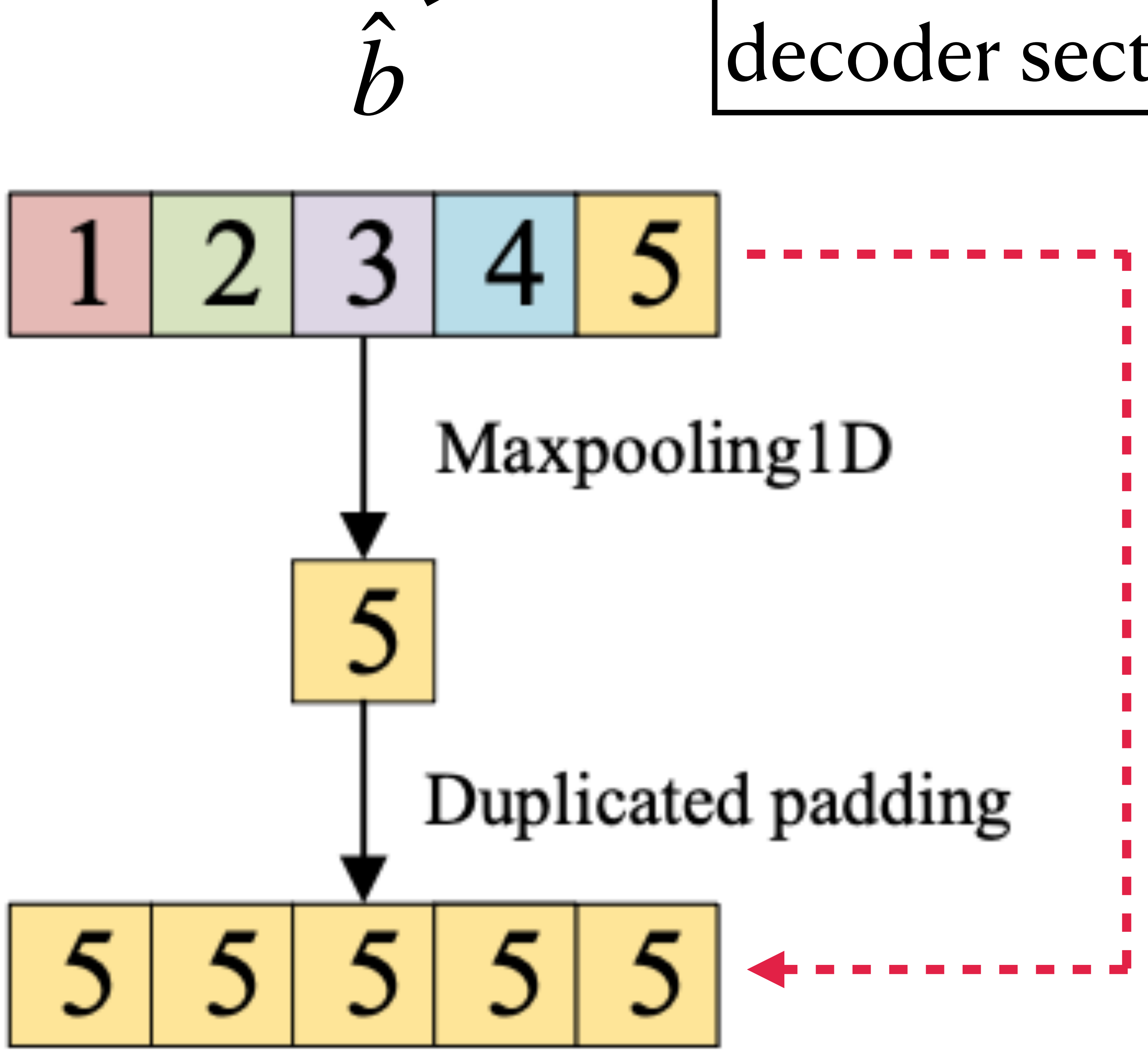
- Maxpooling and Duplicated padding
  - Maxpooling operation is used to extract the most significant factors and eliminate the dimension differences across multi-source HSIs.

$$f_{max} : \mathbb{R}^{c \times \hat{b}} \rightarrow \mathbb{R}^{c}$$

  - Duplicated padding operation is used to keep the dimension before maxpooling.

$$f_{Dup} : \mathbb{R}^{c} \rightarrow \mathbb{R}^{c \times \hat{b}}$$

$\hat{b}$ will also be saved in model to decide how many times should be repeated in the decoder section.



Experiments

In our experiments, we primarily discuss two aspects:

(1) The accuracy and generalization performance of shared encoder.

(2) The promotion of classification performance by pre-trained model.

- Datasets

We utilize a total of 12 hyperspectral images from distinct sensors as experimental materials. Details of these datasets are provided in table 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

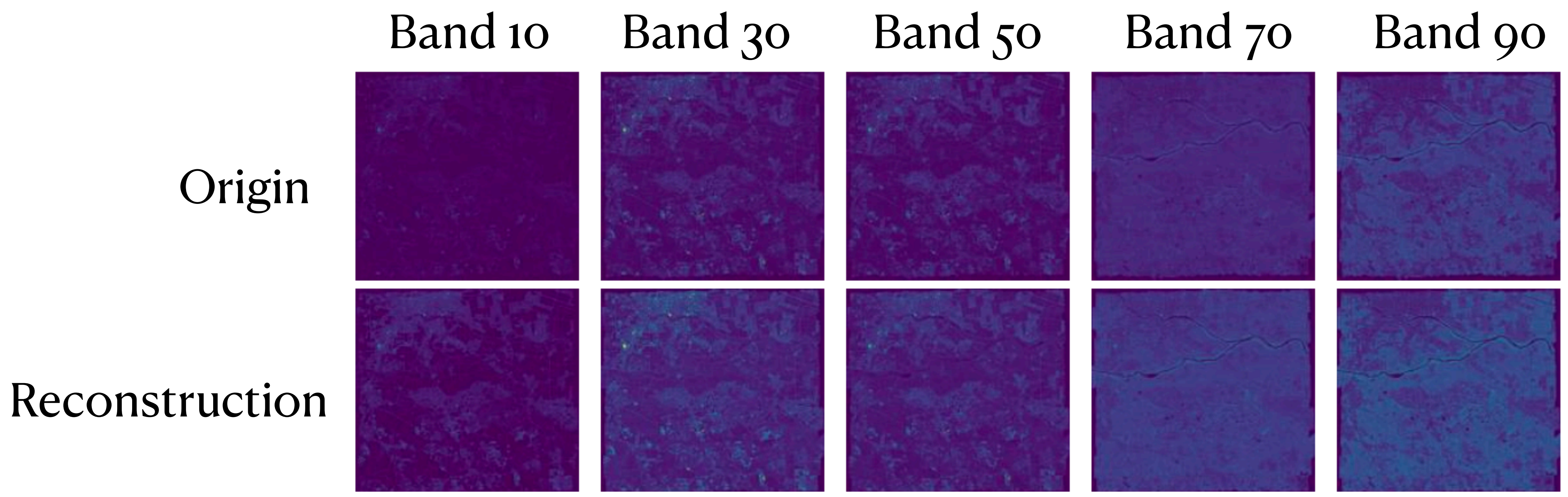
Table 1: The detailed information of the datasets used.

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

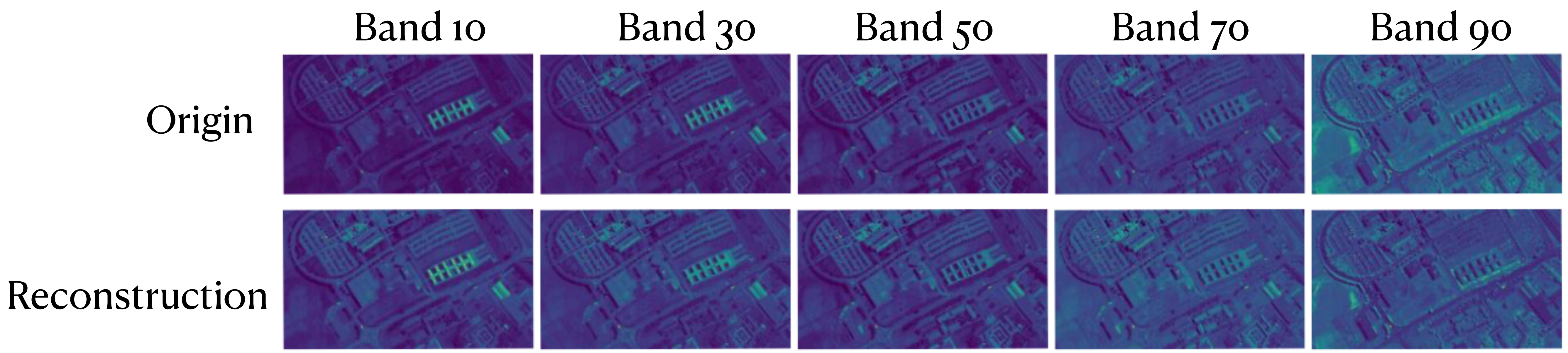
No.	Class	1% Training	5% Training	Total
1	Asphalt	66	331	6631
2	Meadows	186	932	18649
3	Gravel	20	104	2099
4	Trees	30	153	3064
5	Mental sheets	13	67	1345
6	Bare soil	50	251	5029
7	Bitumen	13	66	1330
8	Bricks	36	184	3682
9	Shadow	9	47	947
	Total	423	2135	42776

- Reconstruction results

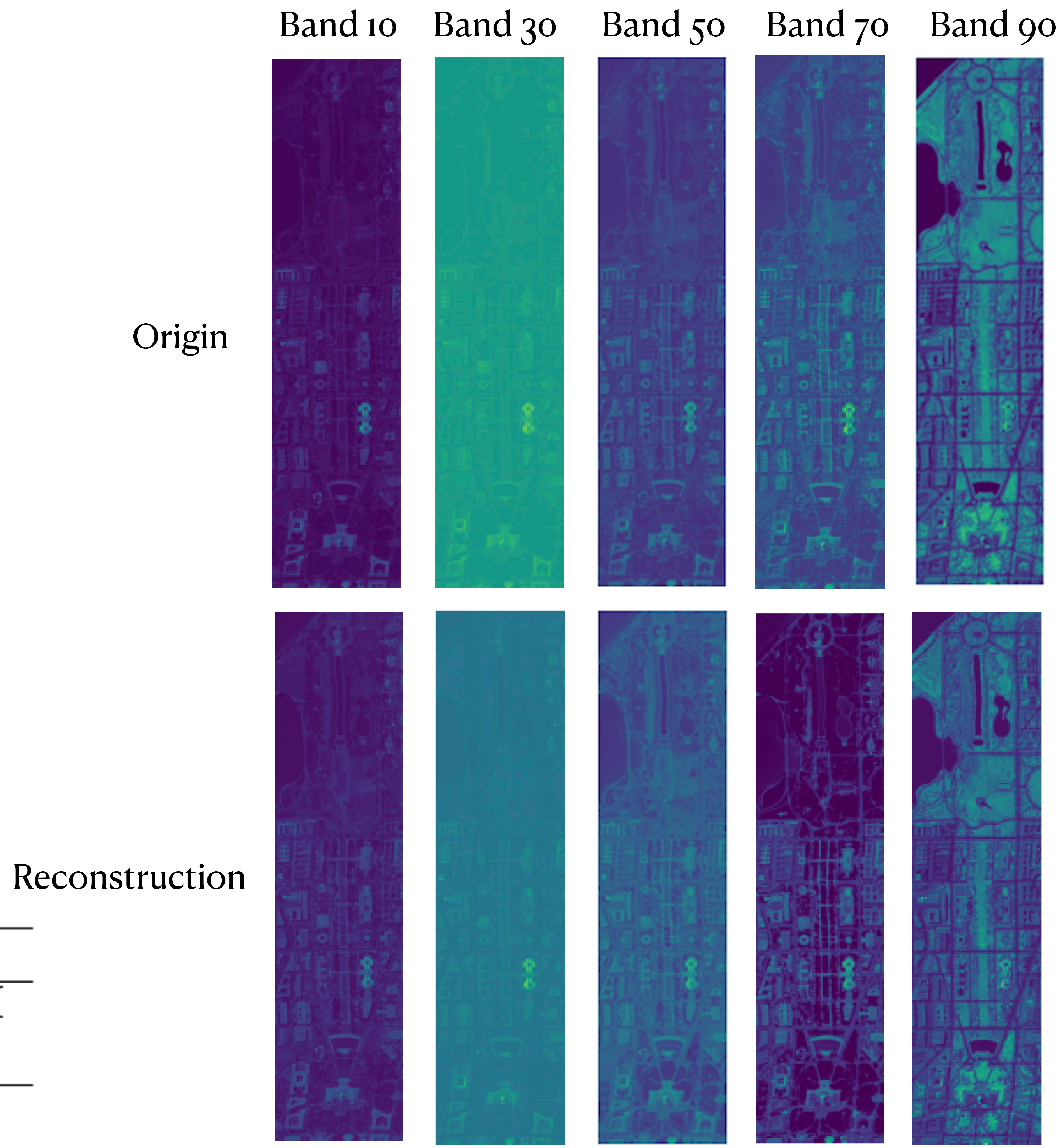
Chikuse



PaviaU



WashingtonDC



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**

- Classification results

	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

Conclusion

Shared encoder which constructs a unified feature space for multi-source HSIs. The model is highly compatible with HSIs from various sources, requiring no model structure adjustments after pretraining to accommodate different data structures. The uniform representation provides the possibility to design a large foundation 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 focus on the development of large foundational pre-training models to enhance the practical applicability of hyperspectral data in production environments.

Thank you !

Harbin Engineering University
School of Information and Communication Engineering
WEILI KONG, BAISEN LIU
E-Mail: kkweil@hrbeu.edu.cn