Supplementary Materials to " Unmixing before Fusion: A Generalized Paradigm for Multi-Source-based Hyperspectral Image Synthesis"

Yang Yu*, Erting Pan*, Xinya Wang, Yuheng Wu, Xiaoguang Mei†, Jiayi Ma
Electronic Information School, Wuhan University, Wuhan, China
{yuyang1995,panerting,wangxinya,yuhengwu}@whu.edu.cn, {meixiaoguang,jyma2010}@gmail.com

Abstract

This supplementary manuscript provides the network configuration of the proposed paradigm, details of the experimental settings, and more qualitative evaluation results for generated synthetic hyperspectral images (HSIs). Finally, we discuss the potential benefits, limitations, as well as future work of this study.

1. Network Configuration

Table s1 lists the detailed constitutions of the residual spectral attention block (RSA) in the proposed unmixing network. Table s2 lists the detailed architecture of the proposed unmixing network.

Our experiments involved three types of typical and advanced generative models, including the VDVAE, StyleGAN3, and DDPM. Therefore, We offer the specific implementations and parameters for different models separately. For the VDVAE, we choose the 62 layers with a hidden size of 512 and a bottleneck size of 128. The dimension of the latent feature(the latent dim) per layer is set to 16. Duiring the training process, we set the VDVAE learning rate to 0.00015 and utilize the Adam optimizer. Additionally, we apply an exponential moving average (EMA) rate of 0.00015 and set the skip threshold to 180. For the StyleGAN3, the mapping network consists of 8 fully-connected layers, and the dimensionality of the input and output activations (W, Z) is set to 512. During the traning process, the learning rate of the generator and discriminator is set to 0.002 and choose the Adam optimizer. For the DDPM, we set the variance to increase linearly from $1e{-}6$ to $1e{-}2$ with a step size of 2000 during the forward process. We construct a denoising U-Net with depth multipliers of $[1, 2, 2, 4, 4]$ for the reversal process. The setting details of the VDVAE, StyleGAN3, and DDPM are listed in Table s3, s4, s5, respectively.

*Equal contribution
†Corresponding author

Module Name		Layer Name	Kernel size	Channel	Inputs
RSA	Spatial block	Conv1	3×3	64	Input
		ReLU1	-	64	Conv1
		Conv2	3×3	64	ReLU1
		Sum1	-	64	Input,Conv2
	Spectral attention block	Conv3	3×3	64	Sum_1
		ReLU2	-	64	Conv3
		Conv4	3×3	64	ReLU2
		Sum2	-	64	Sum1,Conv4
		Conv5	1×1	64	Sum2
		ReLU3	-	64	Conv5
		Conv6	1×1	64	ReLU3
		Sigmoid1	-	64	Conv6
		Sum3	-	64	Sigmoid1, Sum2
	Residual	Sum4	-	64	Sum1,Sum3

Table s1. Detailed constitutions of RSA in the proposed unmixing network.

2. Experimental Settings

We provide detailed descriptions of datasets adopted for HSI synthesis in the remote sensing scenario in Table s6 and the natural mixed-ground scenario in Table s7, respectively.

For the remote sensing scenario, an HSI dataset Chikusei [5] collected over agricultural and urban areas, is employed to train the unmixing network. We conduct validation experiments on the HSI dataset HSRS-SC [4]. Noteworthy, these two datasets have different spectral ranges and resolutions due to sensor differences. As presented in Table s6, we resample and align their spectra within the visible range. External RGB dataset AID [3] is utilized for training the abundance-based generation. Similarly, HSI datasets ARAD [1], Harvard [2], and external RGB dataset Place [6] are used for the natural scenario.

Network	Layer	Kernel size	Channel	Output size
	Input	-	-	(W,H,3)
Encoder	Conv2D	3×3	32	(W,H,32)
	RSA	-	64	(W,H,64)
	RSA	-	128	(W,H,128)
	RSA	-	96	(W,H,96)
	RSA	-	48	(W,H,48)
	Conv2D	3×3	5	(W,H,5)
	Softmax	-	-	(W,H,5)
Decoder	Conv2D	1×1	C	(W,H,C)
	Output	-	-	(W,H,C)

Note: W and H indicate the spatial width and height of the input, while C symbolizes the spectral resolution of the original HSI.

Table s2. The detailed architecture of the unmixing network.

3. Additional Visual Results

This research contributes to the HSI synthesis by introducing an innovative paradigm for high-dimensional data generation. Instead of generating HSIs in a high-dimensional cube space, we propose generating them in abundance space. This significantly simplifies the generation process. Taking DDPM as an example, Figure 1 showcases this generation process over time steps for abundance synthesis.

Using the proposed 'unmixing before fusion' paradigm, we employ StyleGAN3 and DDPM to generate some results, which are displayed in Fig. s2. For HSI synthesis in remote sensing scenarios, we select the Chikusei HSI datasets for training the unmixing network, and HSRS-SC for its validation, and inference on the RGB dataset AID. The RGB rendering examples are shown in Fig. s1 (a). As illustrated in Fig. s1 (b), these generated results predominantly feature three typical scenes: farmland, city, and building. It is evident that, under the guidance of the proposed paradigm, various generation models can yield a plethora of results with rich diversity, natural textures, and high quality. Figs. s3 and s4 display the band-by-band images of representative generated HSIs. These figures demonstrate that our generated results exhibit reasonable spectral reflection within the visible light band range.

To further exemplify the potential of our generation model, we provide the training and testing results of hyperspectral scene classification in Table 3. This is achieved by integrating the generated results with the original HSRS dataset and applying a standard classification model. We reserved 100 HSIs from each category to form the test set, leaving the remaining original data as the pre-augmentation training set. This left only 385, 134, and 54 samples for each category, leading to a severe class imbalance problem. To solve this issue, we supplemented these three categories with the HSIs generated in this study, bringing each category up to 1000 samples. The enhanced test accuracy across different models suggests that increasing the sample size with generated data can effectively boost model performance.

Moreover, we also generate results in naturally mixed ground scenes. Fig. s5 (a) exhibits the multimodal training and validation dataset used, while Fig. s5 (b) shows some generated results using StyleGAN3 and DDPM under the proposed framework. Fig. s6 shows some less satisfactory generated results. Figs. s7 and s7 display the band-by-band images of representative generated HSIs. Two main factors contribute to these less than ideal results: data issues and model limitations. From a data perspective, despite incorporating an RGB dataset to enhance sample diversity, the training samples still suffer from data imbalance, which can lead to unreasonable results in certain scenes. From a model perspective, limitations in model architecture and difficulty achieving global optimization during training may also result in less than satisfactory outcomes.

4. Discussion

4.1. Potential Benefits

Common RGB datasets such as ImageNet have already amassed over 10 million images, greatly stimulating the development of various computer vision tasks. However, the largest existing HSI dataset, ARAD, consists of only 1,000 HSIs with a size of $482 \times 512 \times 31$, hindering the promotion and application of hyperspectral imaging. The proposed unmixing before fusion paradigm can generate extensive, diverse, and reliable HSI samples, alleviating the data scarcity issue of HSI.

More than that, it also has a plethora of potential benefits from both algorithm and application perspectives. For instance, new synthesis HSI samples can be used to correct the class skew problem in the existing dataset, provide more validated training samples, reducing the overfitting problem that is commonly found in the existing HSI classification and target detection models. Furthermore, this technique can lead to novel algorithms for HSI analysis and interpretation, aiding in the detection and identification of objects or materials of interest. It also has the potential to drive and revolutionize applications in a range of fields, such as inferring missing information for cultural heritage restoration, offering realistic fake data for camouflage target recognition, and others.

4.2. Limitations and Future Work

From the model architecture perspective, while our unmixing network has shown competitive performance in various evaluations, it may benefit from the inclusion of specific modules to further improve its efficacy. Nevertheless, the

Network parameter					Training parameter			
Num of layers	Hidden size	Bottleneck size	Latent dim	Skip threshold	Learning rate	Optimizer	Weight decay	EMA rate[1]
62	512	128	16	180	0.00015	Adam	0.0	0.00015

[1] EMA rate: Exponential Moving Average rate.

Table s3. The key hyperparameters of the UBF+VDVAE.

Network parameter					Training parameter	
Latent space Z,W	Mapping network	Leaky ReLU	R1 regularization	Lazy regularization	G,D Learning rate	Optimizer
512	8 fully-connected layers	0.2	8.2	4	0.002	Adam

Table s4. The key hyperparameters of the UBF+StyleGAN3.

Network parameter						Training parameter	
Channel multiplier	Attention channel	Residual blocks	Dropout	Variance	Steps	Learning rate	Optimizer
[1,2,2,4,4]	16	2	0.2	[1e-6,1e-2]	2000	1e-4	Adam

Table s5. The key hyperparameters of the UBF+DDPM.

primary aim of this study is to demonstrate that the endmember and abundance representation can be adapted to images other than HSIs. While endmember and abundance representations are valid low-dimensional features for HSIs, they also capture the typical composition and spatial distribution of similar scenes for RGB images. It is worth noting that our objective is to provide a general and versatile backbone for HSI synthesis, rather than a heavily tuned architecture.

On the other hand, to effectively apply the synthetic HSIs we generate to downstream tasks, precise annotation of attributes or semantic labels is essential. Fortunately, external RGB data is already equipped with highly detailed labels that can be leveraged for various visual tasks. Thus, our future research efforts will not only involve incorporating image information from external datasets, but also their corresponding label in semantic or properties to the generation process, ultimately leading to the generation of labeled synthetic HSIs.

Moreover, while our proposed unmixing before fusion paradigm for HSI synthesis scheme has been demonstrated using two modalities, HSI and RGB, we anticipate that it could be extended to other modalities, such as PAN, MSI, CT, MRI, and beyond.

References

[1] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Luc Van Gool, Shuai Liu, Yongqiang Li, Chaoyu Feng, Lei Lei, Jiaojiao Li, Songcheng Du, Chaoxiong Wu, Yihong Leng, Rui Song, Mingwei Zhang, Chongxing Song, Shuyi Zhao, Zhiqiang Lang, Wei Wei, Lei Zhang, Renwei Dian, Tianci Shan, Anjing Guo, Chengguo Feng, Jinyang Liu, Mirko Agarla, Simone Bianco, Marco Buzzelli, Luigi Celona, Raimondo Schettini, Jiang He, Yi Xiao, Jiajun Xiao, Qiangqiang Yuan, Jie Li, Liangpei Zhang, Taesung Kwon, Dohoon Ryu, Hyokyoung Bae, Hao-Hsiang Yang, Hua-En Chang, Zhi-Kai Huang, Wei-Ting Chen, Sy-Yen Kuo, Junyu Chen, Haiwei Li, Song Liu, Sabarinathan Sabarinathan, K Uma, B Sathya Bama, and S. Mohamed Mansoor Roomi. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 862–880, 2022. 1

[2] A. Chakrabarti and T. Zickler. Statistics of Real-World Hyperspectral Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 193–200, 2011. 1

[3] Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu. Aid: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7):3965–3981, 2017. 1

[4] Kejie Xu, Peifang Deng, and Hong Huang. Hsrs-sc: a hyperspectral image dataset for remote sensing scene classification. journal of image and graphics. *Journal of Image and Graphics*, 26(8):1809–1822, 2021. 1

[5] N. Yokoya and A. Iwasaki. Airborne hyperspectral data over chikusei. Technical Report SAL-2016-05-27, Space Application Laboratory, University of Tokyo, Japan, 2016. 1

[6] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 1

Tasks	Training for Unmixing	Validation for Unmixing	Training for Generation
Datasets	Chikusei (HSI)	HSRS-SC (HSI)	AID (RGB)
Imaging Sensor	The Headwall Hyperspec-VNIR-C imaging sensor	The Compact airborne spectro-graphic imager, CASI	The Google Earth imagery
Spectral Range	363-1018nm→398-698nm	380-1050nm→398-698nm	-
Spectral Resolution	128→59	48→59	3
Spatial Resolution	2.5m	1m	0.5-8m
Patch Size	2517×2335×128→128×128×59	256×256×59	256×256×3
Samples	1 HSI→840 patches	700	1902

Table s6. Detailed description of datasets adopted for HSI synthesis in the remote sensing scenario.

Tasks	Training for Unmixing	Validation for Unmixing	Training for Generation
Datasets	ARAD (HSI)	Harvard (HSI)	Place (RGB)
Imaging Sensor	Specim IQ mobile hyperspectral camera	The Nuance CRi Multi-spectral Imaging System-FX	From online image search engines
Spectral Range	400-700nm	420-720nm	-
Spectral Resolution	31	31	3
Spatial Resolution	-	-	-
Patch_Size	482×512×31→256×256×31	1300×1300×31	256×256×3
Samples	900 HSIs→3600 patches	45	10000

Table s7. Detailed description of datasets adopted for HSI synthesis in the natural mixed-ground scenario.

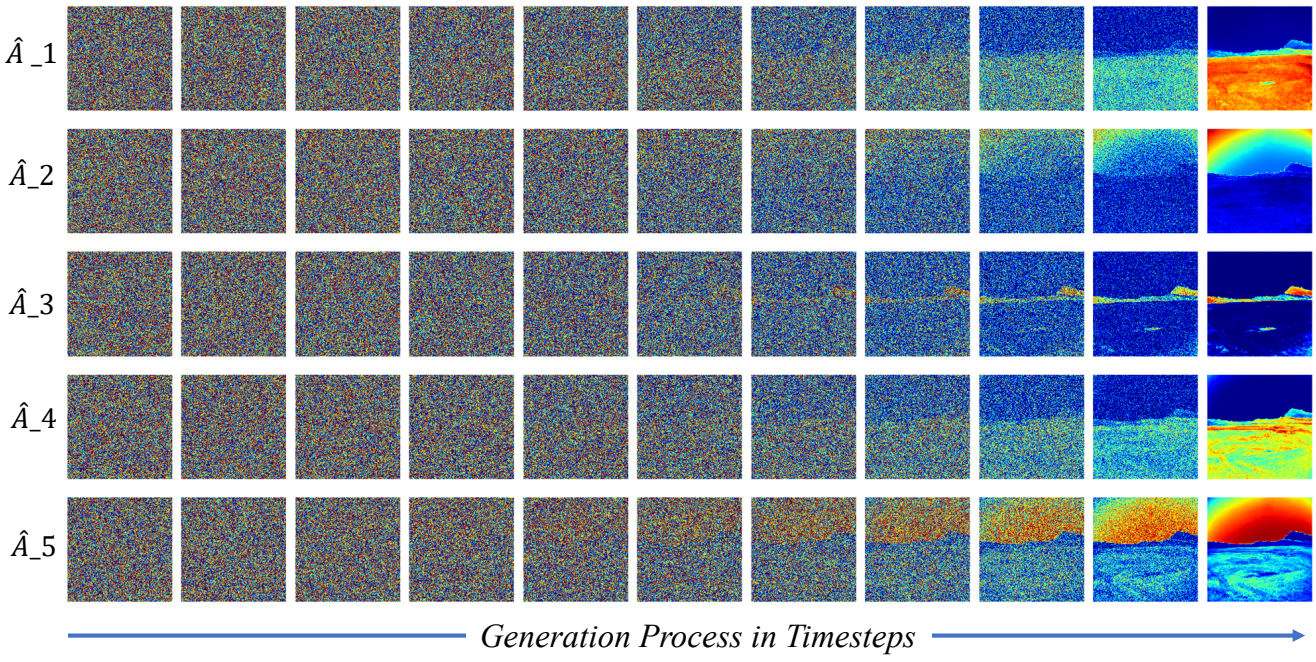
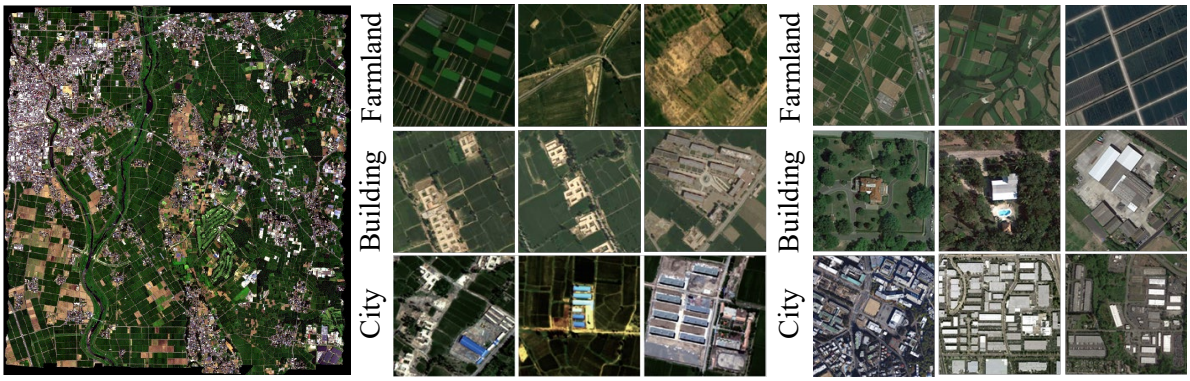


Figure s1. Illustration of the synthetic abundances generation process within timesteps in the UBF+DDPM.



(a) Examples in adopted datasets

(b) Examples of synthetic HSIs generated by the UBF+StyleGAN3

(c) Examples of synthetic HSIs generated by the UBF+DDPM

Figure s2. RGB rendering examples of (a) adopted datasets, (b) generated synthetic HSIs by the UBF+Stylegan3, and (c) generated synthetic HSIs by the UBF+DDPM (in bands 15,25,49).

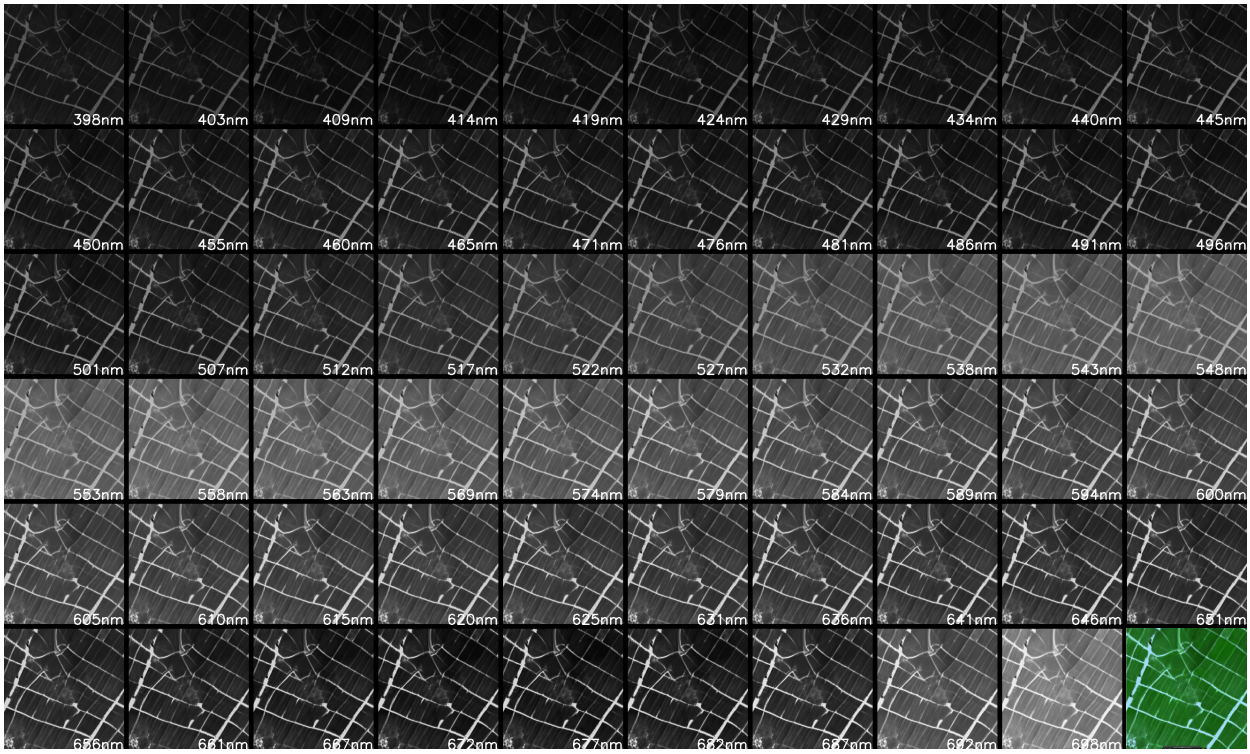


Figure s3. Illustration on each band of the generated synthetic HSIs by the StyleGAN3 in the remote sensing scenario, and the last one is the corresponding false-color image (based on bands 15, 32, 49).

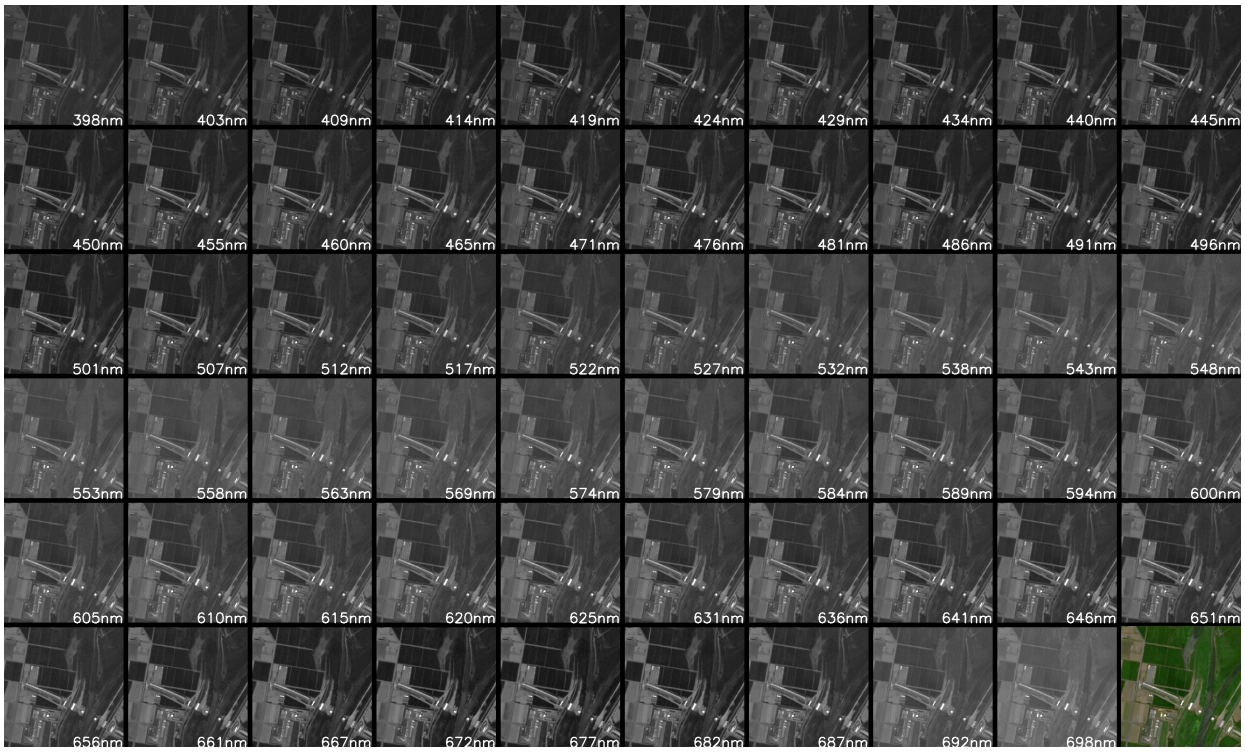
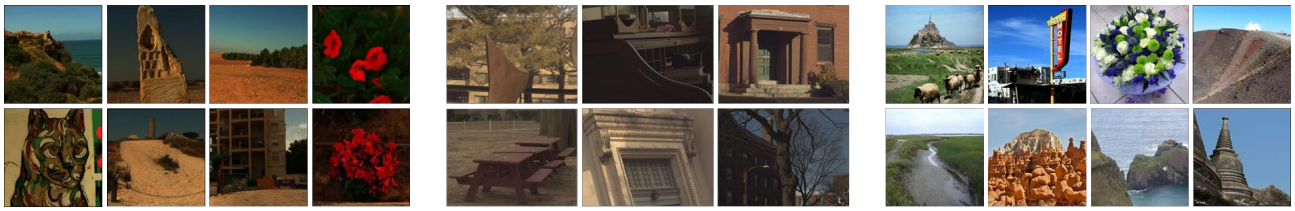


Figure s4. Illustration on each band of the generated synthetic HSIs by the DDPM in the remote sensing scenario, and the last one is the corresponding false-color image (based on bands 15, 32, 49).



ARAD (HSI) Harvard (HSI) Place (RGB)

(a) Examples in adopted datasets

(b) Examples of synthetic HSIs generated by the UBF+StyleGAN3

(c) Examples of synthetic HSIs generated by the UBF+DDPM

Figure s5. RGB rendering examples of (a) adopted datasets, (b) generated synthetic HSIs by the UBF+Stylegan3, and (c) generated synthetic HSIs by the UBF+DDPM (in bands 7,17,27).

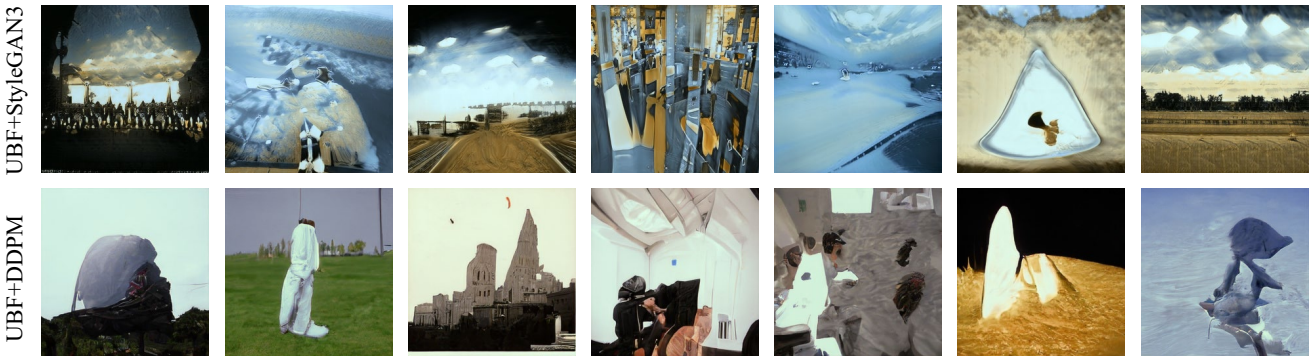


Figure s6. Illustration of the unsatisfactory synthesis results.

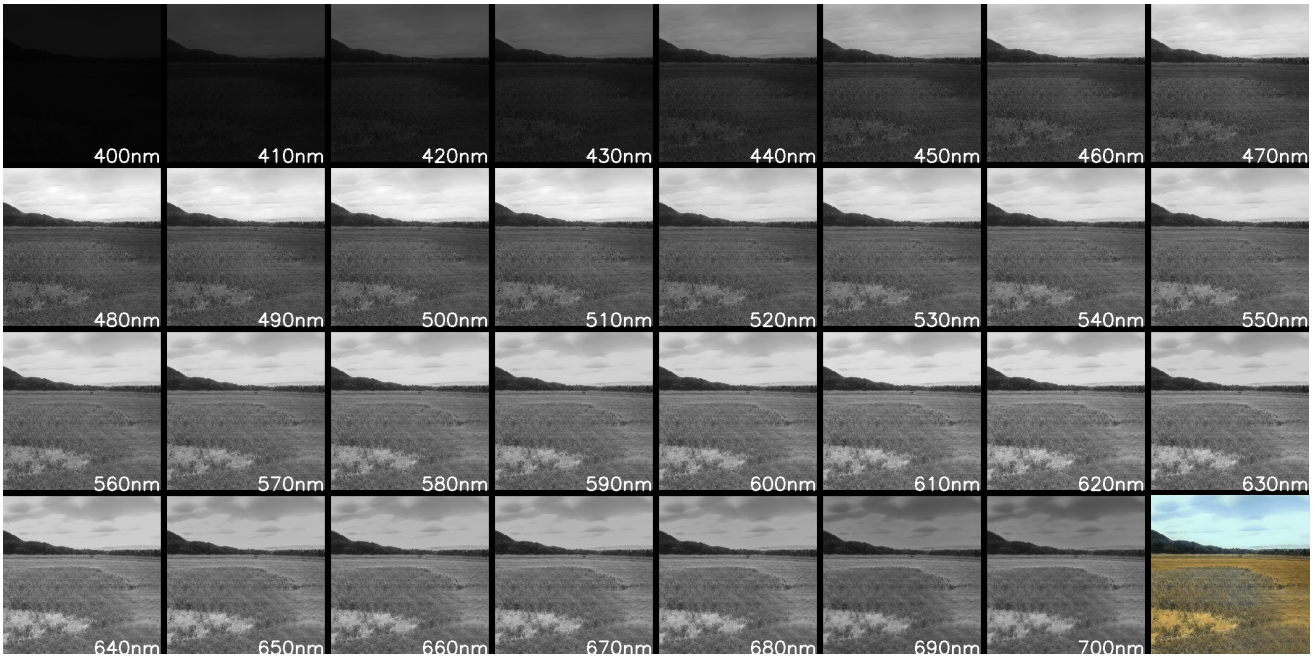


Figure s7. Illustration on each band of the generated synthetic HSIs by the StyleGAN3 in the natural mixed-ground scenario, and the last one is the corresponding false-color image (based on bands 7, 17, 27).

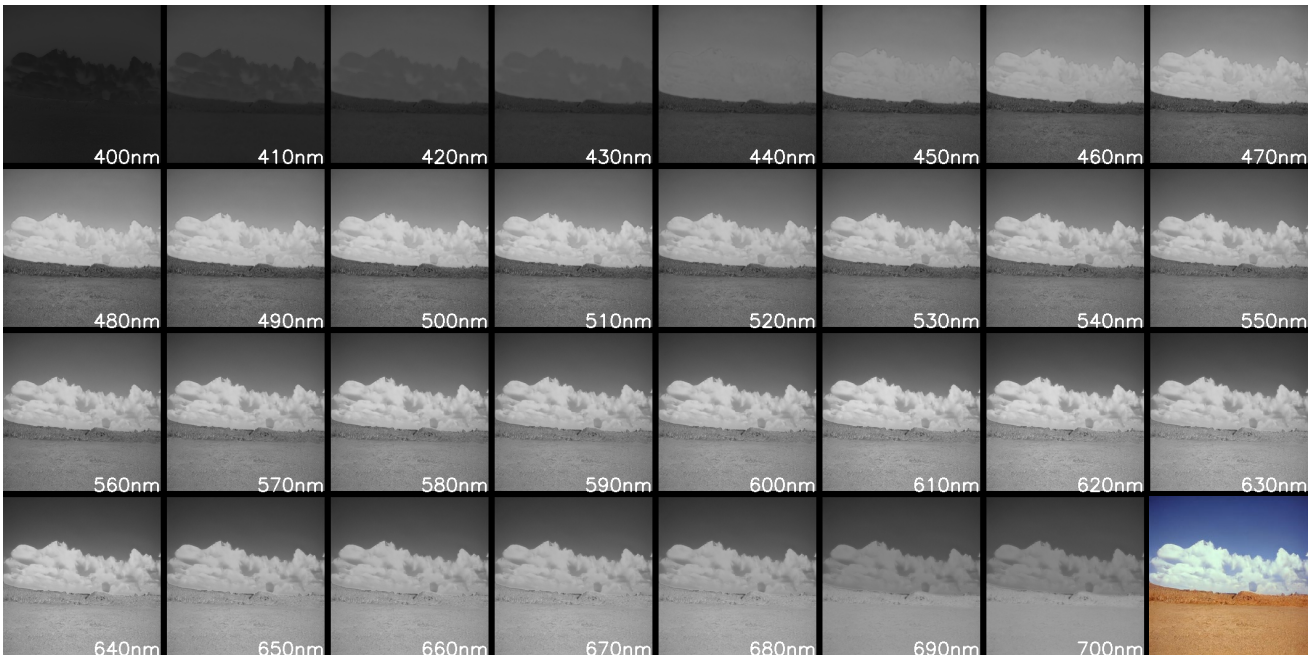


Figure s8. Illustration on each band of the generated synthetic HSIs by the DDPM in the natural mixed-ground scenario, and the last one is the corresponding false-color image (based on bands 7, 17, 27).