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

In the realm of AI, data serves as a pivotal resource. Real-world hyperspectral images (HSIs), bearing wide spectral characteristics, are particularly valuable. However, the acquisition of HSIs is always costly and time-intensive, resulting in a severe data-thirsty issue in HSI research and applications. Current solutions have not been able to generate a sufficient volume of diverse and reliable synthetic HSIs. To this end, our study formulates a novel, generalized paradigm for HSI synthesis, i.e., unmixing before fusion, that initiates with unmixing across multi-source data and follows by fusion-based synthesis. By integrating unmixing, this work maps unpaired HSI and RGB data to a low-dimensional abundance space, greatly alleviating the difficulty of generating high-dimensional samples. Moreover, incorporating abundances inferred from unpaired RGB images into generative models allows for cost-effective supplementation of various realistic spatial distributions in abundance synthesis. Our proposed paradigm can be instrumental with a series of deep generative models, filling a significant gap in the field and enabling the generation of vast high-quality HSI samples for large-scale downstream tasks. Extension experiments on downstream tasks demonstrate the effectiveness of synthesized HSIs. The code is available at HSI-Synthesis.github.io.

1. Introduction

Hyperspectral image (HSI), with its high-resolution spectral information, holds great potential to revolutionize myriad research fields [8, 34, 38]. However, a confluence of factors including prohibitive collection costs, limited sensor availability, intricate data management, and environmental constraints have conspired to render HSI data scarce [11, 27, 49]. This scarcity of HSI data, coupled with the complexity and time-intensive nature of HSI acquisition and pre-processing, is a significant impediment to the construction of comprehensive HSI datasets and the overall advancement of AI in HSI [15, 42].

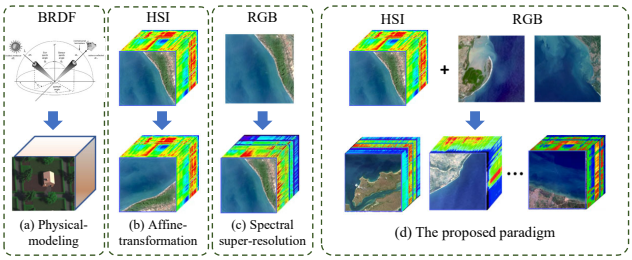

Figure 1. Comparisons of existing HSI synthesis techniques. (a) Physical modeling-based methods. (b) Affine transformation-based methods. (c) Spectral reconstruction-based methods. (d) The proposed paradigm (multi-source-based).

In light of the burgeoning demand for high-quality data, the advent of synthetic or artificially generated data is a welcome innovation [35, 47]. It offers a viable solution to the chronic data shortage, and the domain of HSI synthesis has seen an influx of research interest [14]. Existing studies can be broadly categorized into three groups:

1. Physical modeling-based HSI synthesis [12, 17, 21] (refer to Fig. 1 (a)) mainly focus on mimicking physical phenomena, such as the Bidirectional Reflectance Distribution Function (BRDF) or the Gaussian Mixture Model (GMM). Despite its meticulous manual design, it often produces unreliable HSIs due to too ideal assumptions.
2. Affine transformation-based HSI synthesis [16, 36] (refer to Fig. 1 (b)) involves enriching data through a series of affine transformations, including rotation, scaling, and shearing. It serves to augment the quantity of training samples, albeit without enriching their diversity.
3. Spectral super-resolution-based HSI synthesis [1, 4] (refer to Fig. 1 (c)) attempts to expand the spectral dimension from multi-spectral or RGB images. Regrettably, they are hard to accurately reconstruct the spectral signatures and are unable to generate new samples.

While these techniques have their merits, they also grapple

*Equal contribution
†Corresponding author

with significant limitations such as lack of authenticity, restricted diversity, and inability to produce new HSI samples.

Generative AI, encompassing techniques such as Variational Autoencoder (VAE) [31, 32, 37], Generative Adversarial Network (GAN) [3, 9, 13, 22], Normalizing Flow (Flow) [25, 30], and Denoising Diffusion Probabilistic Model (Diffusion) [7, 19, 24, 33], offers an appealing alternative. These techniques have been the bedrock of numerous synthetic data architectures, covering images, audio, and text data [20, 28, 43, 48]. Nonetheless, the high dimensionality of spectral signatures makes HSI synthesis via generative AI a challenging task. Up to now, to the best of our knowledge, few publicly available research has appeared on this task.

To address these issues, this work raises a novel paradigm for HSI synthesis, unmixing before fusion. In our observation, similar scenes have salient and common low-rank characteristics, empowering a small scale of endmembers to describe the entire scene with high efficiency, and the diversity of various scenes can also be manifested in their abundance maps. Noteworthy, existing HSIs are limited in quantity and quality, the incorporation of unpaired RGB images that cover similar scenes is feasible and economical. Motivated by this, we customize the concept of unmixing for multi-source data, decomposing unpaired HSI and RGB images that record similar scenes into fixed endmembers and varied abundances. Further, we propose a fusion within these abundances to jointly learn various and realistic spatial distributions of real scenes. Notably, our work differs from existing unmixing-based fusion ideas for paired data [44, 46] in several aspects, whether it comes to data, fusion rules, or objectives.

Hence, in this paper, we assume that multi-source data recording similar scenes share endmembers. Building upon this, we have tailored the unmixing and integrated external RGB data, developing the unmixing before fusion paradigm. Specifically, we first decompose unpaired HSI and RGB data into fixed endmembers and unique abundances. Then, we synthesize diverse abundance maps via fusion-based deep generative models. Finally, we fuse the estimated HSI endmembers and synthetic abundance maps to generate new HSIs that closely resemble real-world data.

Our main contributions are summarized below:

- Formulate a generalized paradigm for multi-source-based HSI synthesis, incorporating a series of deep generative models to produce diverse and abundant HSI samples.
- Bridge the dimensional gap between RGB and HSIs in the abundance space and incorporate multi-source data to alleviate the issue of limited sample availability.
- Pioneer to synthesis abundance (low-dimensional) instead of HSI sample (high-dimensional), mitigating the intricacies tied to high-dimensional sample synthesis.

Extensive experiments, especially extension on downstream tasks, have revealed that the proposed paradigm promises to surmount the limitations of HSI data scarcity, paving the way for substantial advancements in the field.

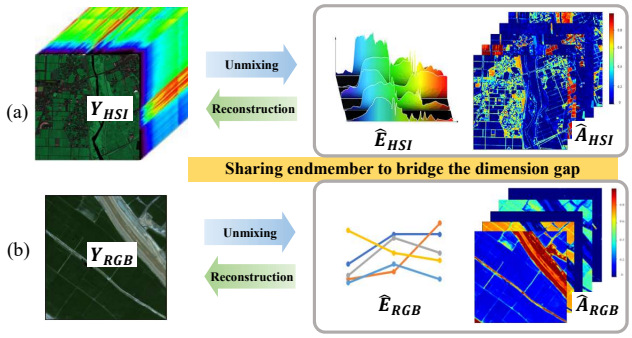

Figure 2. Illustration of the unmixing and reconstruction process on (a) an HSI, and (b) an unpaired RGB image.

2. Motivation

2.1. Why unmixing?

A recognized fact is that a single pixel in HSIs, even within modern imaging systems, can span tens of meters and represent a blend of different substances. The Linear Mixing Model (LMM) [5] is a powerful tool for unmixing that decomposes each pixel in HSIs into a blend of pure spectral signatures, or "endmembers". Each endmember signifies a distinct material or component present within the scene, offering a comprehensive elucidation of the image's composition [29]. The real charm of LMM lies in its "abundances" or coefficients, which denote the proportion of each endmember within a pixel, constructing an in-depth portrayal of the pixel's composition.

Denote $Y_{\text{HSI}} \in \mathbb{R}^{B\times W\times H}$ as an HSI, which contains B channels in the spectral domain, and $W \times H$ pixels in the spatial domain. Its unmixing process based on LMM can be articulated as follows :

$$Y_{\text{HSI}} = \hat{E}_{\text{HSI}} \cdot \hat{A}_{\text{HSI}} + \epsilon, \tag{1}$$

where $\hat{E}_{\text{HSI}} \in \mathbb{R}^{B\times k}$ represents the endmembers, $\hat{A}_{\text{HSI}} \in \mathbb{R}^{k\times W\times H}$ denotes the abundances, B signifies the total number of spectral bands, k corresponds to the number of endmembers, and W and H record the spatial size.

According to Eq. (1), an HSI cube, marked by high spectral dimensions reaching into the hundreds, is unmixing into two unique, low-dimensional features: endmembers and abundances. Noteworthy, these two unravel intricate structures hidden with HSIs with a straightforward physical interpretation [10]. Intriguingly, as shown in Fig. 2(a), the unmixing progress guided by LMM is reversible, which enables us to reconstruct the original HSI $\hat{Y}_{\text{HSI}}$. In this case, we provide a novel perspective for HSI synthesis. If we focus on generating abundance and sharing the specific endmember, the whole HSI could be easily reconstructed. It

significantly streamlines the HSI synthesis and sidesteps the challenges in dealing directly with high-dimensional data.

2.2. Why fusion?

Another challenge encountered in HSI synthesis is the limited training samples in both quantity and quality. To overcome this issue, we suggest the integration of unpaired RGB data, which are more easily accessible, as auxiliary information. RGB images, essentially composed of the primary colors of red, green, and blue, offer a unique perspective and a wide range of spatial details. As illustrated in Fig. 2(b), under the assumption of multi-source data recorded similar scenes share endmembers, the LMM-based unmixing procedure can also be applied to unpaired RGB images, represented by $Y_{\text{RGB}} \in \mathbb{R}^{3\times W\times H}$. The equation for the unmixing procedure is as follows:

$$Y_{\text{RGB}} = \hat{E}_{\text{RGB}} \cdot \hat{A}_{\text{RGB}} + \epsilon, \tag{2}$$

where $\hat{E}_{\text{RGB}} \in \mathbb{R}^{3\times k}$ signifies the endmembers, $\hat{A}_{\text{RGB}} \in \mathbb{R}^{k\times W\times H}$ is the abundances with three color channels.

Interestingly, if we assume a consistent quantity (k) of shared endmembers, then the abundance in HSIs ($\hat{A}_{\text{HSI}}$) and RGB ($\hat{A}_{\text{RGB}}$) will have an identical shape, i.e., $k \times W \times H$. This suggests the possibility of finding an alignment in the abundance space between unpaired HSI and RGB.

Based on this, we propose blending the abundances from unpaired HSI and RGB images to significantly boost HSI synthesis. Incorporating multi-source data empowers the model to learn various and realistic spatial distributions of real scenes at a low cost, which helps to extract more stable and generalizable features from it. It also maximizes the use of readily available RGB data, setting the stage for more comprehensive and reliable HSI synthesis.

3. Unmixing before Fusion as a paradigm

To ensure the most accurate representation of objects' spatial characteristics - while preserving the integrity of their spectral signatures - we propose a pioneering solution, aptly named "Unmixing before Fusion". This innovative approach, visually encapsulated in Fig. 3, promises to revolutionize our understanding and application of HSIs. A detailed exploration of this transformative approach follows.

3.1. Unmixing across multi-source data

As previously delineated, the unmixing process facilitates the alignment between two distinct data modalities, HSI and RGB. This alignment is contingent upon the consistent assumption of endmembers' quantity. While traditional HSI unmixing networks have proven to be effective, their usability is intrinsically limited due to their dependence on unsupervised training in a singular HSI scenario. It poses a significant impediment to the unmixing across modalities with a different count of spectral channels.

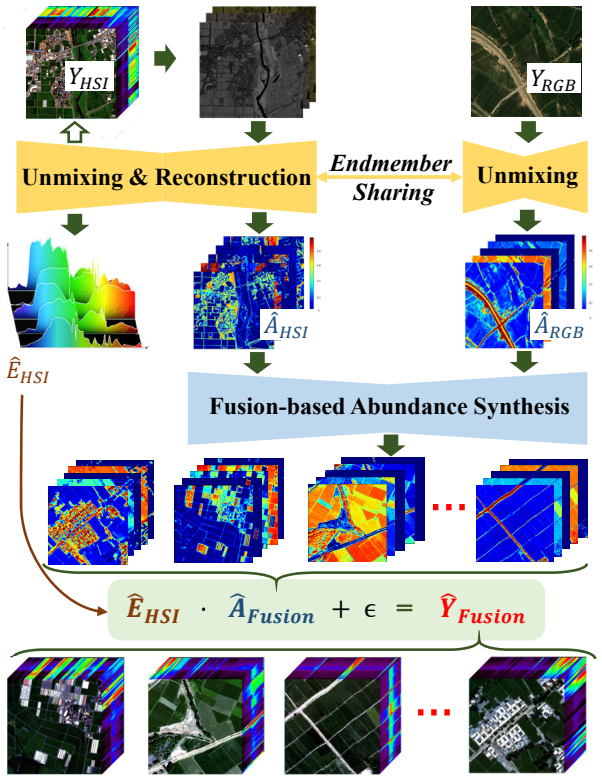

Figure 3. Illustration of the proposed "Unmixing before Fusion" pipeline, a novel paradigm for generating synthetic HSIs. It involves unmixing the abundances from the HSIs and external RGB images, fusing them to generate a greater variety of synthesized abundance maps, and ultimately, generating new HSI samples guided by the LMM.

To this end, we propose a unique unmixing approach, as illustrated in Fig. 3, that capacitates an HSI-trained unmixing model, $\mathcal{U}(\cdot)$, to infer the abundance within RGB data with robust precision. Given the HSI Y_{HSI}, we first train the unmixing net $\mathcal{U}(\cdot)$ to acquire the endmembers $\hat{E}_{\text{HSI}}$ and abundance maps $\hat{A}_{\text{HSI}}$, following:

$$\hat{E}_{\text{HSI}}, \hat{A}_{\text{HSI}} = \mathcal{U}(\Psi(Y_{\text{HSI}})), \tag{3}$$

where $\Psi(\cdot)$ here symbolizes the band selection operation to extract representative three-band data Y_{bs} (corresponding to RGB) from HSIs. Notably, the reconstruction target is $\hat{Y}$, not Y_{bs}. The fractional abundance maps $\hat{A}$ should be governed by the abundance non-negative constraint (ANC) and the abundance sum-to-one constraint (ASC) [2].

Our proposed unmixing model, $\mathcal{U}(\cdot)$, deviates from the conventional symmetrical encoder-decoder structure, adopting an unsymmetrical one instead. The encoder comprises several residual spectral attention modules (RSA) [39] and culminates in a softmax layer. Conversely, the decoder merely includes a 1×1 convolutional layer, which simulates the LMM. The weights of the decoder represent the estimated endmembers $\hat{E}_{\text{HSI}}$. Such a design en-

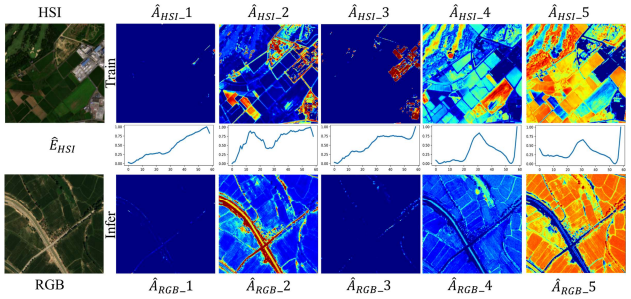

Figure 4. Example results of unmixing across multi-source data. It is trained on the Chikusei dataset (HSI) and inferred on the AID dataset (RGB).

sures that the output of the encoder precisely reflects the spatial distribution of abundances while enabling the weight of the decoder to indicate endmembers with corresponding physical significance.

The loss function of the proposed unmixing net is tripartite: the mean absolute error (MAE) loss $\mathcal{L}_{\text{MAE}}$ to ensure the pixel-wise reconstruction accuracy, the spectral angle distance (SAD) loss $\mathcal{L}_{\text{SAD}}$ to govern the fidelity of spectral signatures, and the endmembers total variation (ETV) loss $\mathcal{L}_{\text{ETV}}$ to preserve the spectral smoothness of the extracted endmembers. The total loss function $\mathcal{L}$ can be expressed as:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{\text{MAE}} + \alpha \cdot \mathcal{L}_{\text{SAD}} + \beta \cdot \mathcal{L}_{\text{ETV}} \\ &= ||Y, \hat{Y}||_1 + \alpha \cdot \arccos\left(\frac{(Y, \hat{Y})}{||Y||_2 ||\hat{Y}||_2}\right) + \beta \cdot \sum_{i}^{C} (e_{i+1} - e_i),\end{aligned} \tag{4}$$

where e_i represents the value of i_{th} band in each spectral vector of endmembers, and α and β are used to balance convergency for each item. α and β are setting as 0.1 and $1e-3$, empirically. By optimizing our network using this loss function, we can ensure stable and desirable unmixing results to a significant degree.

Our proposed paradigm is predicated on the claim that data, whether HSI or RGB, recording similar scenarios can be represented by a finite set of fixed endmembers and customized abundance maps. Here, the term "endmembers" refers to the typical compositional constituents in a given scene, a departure from the traditional unmixing concept of pure pixels composed of a singular material. This fundamental idea provides a low-dimensional space to align data in different modalities and enhances the applicability of our unmixing network across a range of modalities.

Next, leveraging the robustly trained unmixing network, we apply it to external RGB datasets that fall within the same scenario category. This allows us to infer their abundance, denoted as $\hat{A}_{\text{RGB}}$. following:

$$\hat{A}_{\text{RGB}} = \mathcal{U}(Y_{\text{RGB}}; \hat{E}_{\text{HSI}}). \tag{5}$$

The efficacy of our method is corroborated by the example results, which are depicted in Fig. 4.

Algorithm 1: Unmixing across multi-source data

Input: $Y_{\text{HSI}} \in \mathbb{R}^{B \times W \times H}$; $Y_{\text{RGB}} \in \mathbb{R}^{3 \times W \times H}$
Output: $\hat{E}_{\text{HSI}} \in \mathbb{R}^{B \times k}$; $\hat{A}_{\text{HSI}} \in \mathbb{R}^{k \times W \times H}$; $\hat{A}_{\text{RGB}} \in \mathbb{R}^{k \times W \times H}$

/* Training on HSI data. */
$Y_{\text{bs}} = \Psi(Y_{\text{HSI}})$ ▷ Band Selection ($Y_{\text{bs}} \in \mathbb{R}^{3 \times W \times H}$);
while *not converged* **do**
  $\hat{E}_{\text{HSI}}, \hat{A}_{\text{HSI}} = \mathcal{U}(Y_{\text{bs}})$;
  $\theta \leftarrow \mathcal{L}(\theta)$; ▷ Refer to Eq. (4);
end
return $\mathcal{U}(y; \theta)$; $\hat{E}_{HSI}$; $\hat{A}_{HSI}$;
/* Inferring on RGB data. */
return $\hat{A}_{\text{RGB}} = \mathcal{U}(Y_{\text{RGB}}, \hat{E}_{\text{HSI}}; \theta)$

Algorithm 2: Fusion-based synthesis

Input: $\hat{A}_{\text{HSI}}$; $\hat{A}_{\text{RGB}}$; $\hat{E}_{\text{HSI}}$
Output: $\hat{A}_{\text{Fusion}} \in \mathbb{R}^{k \times W \times H}$; $\hat{Y}_{\text{Fusion}} \in \mathbb{R}^{B \times W \times H}$

/* Synthesizing abundances. */
$\hat{a}_{\text{Fusion}} = \mathcal{G}(\hat{A}_{\text{RGB}}, \hat{A}_{\text{HSI}})$; ▷ Training ;
$\hat{A}_{\text{Fusion}} \sim \hat{a}_{\text{Fusion}}$; ▷ Sampling ;
/* Synthesizing HSI samples. */
return $\hat{Y}_{\text{Fusion}} = \hat{E}_{\text{HSI}} \cdot \hat{A}_{\text{Fusion}} + \epsilon$

3.2. Fusion-based synthesis

The act of projecting HSIs into an abundance space notably reduces the complexity associated with the generation tasks. The rapid technological advancements have given rise to an array of advanced and effective generative models. However, it should be underscored that the primary objective of this paper is not to design an exceptional generative model but rather to propose a practical and insightful pipeline.

With the estimated abundances $\hat{A}_{\text{RGB}}$ and $\hat{A}_{\text{HSI}}$, we have the capability to synthesize abundance utilizing a generative model, denoted as $\mathcal{G}(\cdot)$. We can synthesize abundances by:

$$\hat{A}_{\text{Fusion}} = \mathcal{G}(\hat{A}_{\text{RGB}}, \hat{A}_{\text{HSI}}), \tag{6}$$

where $\mathcal{G}(\cdot)$ here is not limited to a specific type but can encompass any variety. In the subsequent experiments, we have employed various generative models, involving VDVAE [6], StyleGAN3 [23], and DDPM [19], and customized them from the image space to the abundance space as $\mathcal{G}(\cdot)$ to provide a comprehensive understanding. Implementation details can be found in *suppl. material*.

After obtaining the generated synthetic abundance maps $\hat{A}_{\text{Fusion}}$, our attention returns to the original LMM model. In fact, the generation of synthetic HSI is also a fusion process of multi-modal information, *i.e.*, synthetic abundance maps $\hat{A}_{\text{Fusion}}$ from given HSIs and RGB images, and endmembers estimated from HSIs $\hat{E}_{\text{HSI}}$. Mathematically, the final step

Figure 5. The false-color image of HSIs synthesized by existing techniques and typical generative models under the proposed paradigm.

can be formulated as:

$$\hat{Y}_{\text{Fusion}} = \hat{E}_{\text{HSI}} \cdot \hat{A}_{\text{Fusion}} + \epsilon, \tag{7}$$

where $\hat{Y}_{\text{Fusion}} \in \mathbb{R}^{C\times W\times H}$ indicates the synthetic HSI, $\hat{E}_{\text{HSI}} \in \mathbb{R}^{C\times k}$ represents the estimated endmembers, and $\hat{A}_{\text{Fusion}} \in \mathbb{R}^{k\times W\times H}$ symbolizes the generated abundance maps. The detailed designs and the procedure of the proposed paradigm are in Algorithms 1 and 2. Examples of synthesized HSIs are presented in Fig. 5.

4. Experiments

We present a comprehensive evaluation of the proposed paradigm through a range of experimental procedures, including comparative, ablation, and expansion experiments.

4.1. Experimental settings

Datasets. In this research, we utilized multi-source datasets to validate the robustness and generalizability of our proposed method. We trained the unmixing model using the Chikusei [45] HSI dataset [1]. This HSI dataset measures $2517\times2335\times128$ in size and covers a spectral range of 363 to 1018 nm. Following this, we used the AID [40] dataset [2] for abundance inference. The AID dataset is a conventional RGB dataset used for scene classification, exhibiting scenes similar to those found in the Chikusei dataset.

Metrics. The efficacy of synthetic HSI generation is predominantly evaluated on two primary criteria: diversity and reliability. The assessment of diversity is inherently subjective, hinging on the visual quality of the data generated. Conversely, reliability is measured in terms of the quality of the synthesized abundance and the generated HSI. Given the novelty of the generated data and the absence of a reference, evaluation metrics that do not require a ground truth, such as the Fréchet Inception Distance (FID) [18], Precision-and-Recall [26], are employed. To further corroborate the reliability of the inferred abundance from RGB, we introduce supplementary quantitative evaluation metrics in the ablation study, including the Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Spectral Angle Distance (SAD).

Implementation details. Our experiment involved data from two different modalities, HSI and RGB. We aimed to ensure consistency in the physical meaning of the abundance in both types of data. Therefore, we performed spectral alignment, which resulted in the retention of 59 bands from 400 to 700 nm in Chikusei. In the unmixing experiment, the encoder was designed with 3 stacked RSA modules. The encoder feature maps' dimension was set to $[3, 32, 64, 128, 96, 48, 5]$ and concluded with a softmax layer, which facilitated the satisfaction of the abundance constraints of ASC and ANC. The decoder consisted of a bias-free 1×1 convolution layer, which simulated the LMM process. The initial learning rate for the unmixing training was set to $1e-4$ and we employed the Adam optimizer for 40 epochs. In the fusion generation experiment, we extended the proposed paradigm to different types of widely-used generative models, including VDVAE [6], StyleGAN3 [23], and DDPM [19]. We used the official implementations for all experiment codes, and the training hyperparameters were referred from the configuration of synthesis experiments for datasets with a spatial size of 256×256. All generative models were trained to converge under the same computational resources. The experiments were carried out using four NVIDIA 3090 GPUs.

4.2. With different generative models

Fig. 5 provides a false-color visualization of HSIs produced by various algorithms. The GMM [50] produces results that deviate significantly from the real-life scenario. This discrepancy occurs because GMM uses ideal mathematical distributions and does not mimic the actual distribution of objects. On the other hand, the Multi-Stage Spec-

[1] https://naotoyokoya.com/Download.html
[2] https://captain-whu.github.io/AID/

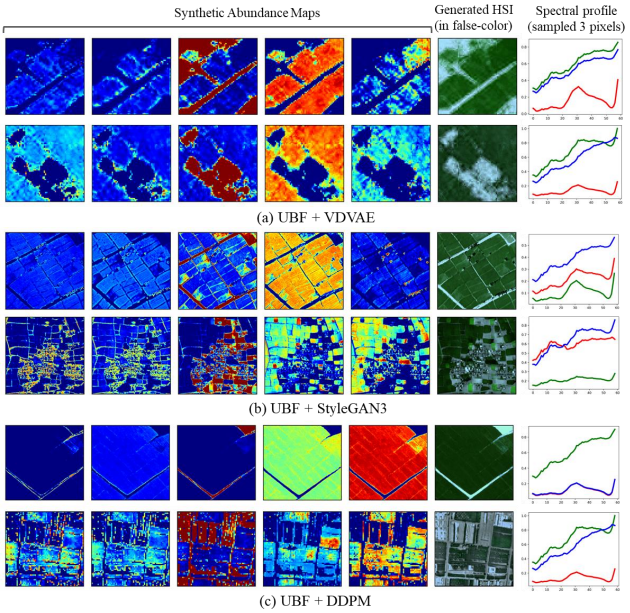

Figure 6. Typical examples of synthesis abundances and generated HSIs in different generative models under the proposed "Unmixing before Fusion" paradigm.

Metrics	FID ↓	Recall ↑	Precision ↑	Params	Sampling
UBF+VDVAE	39.13	0.209	0.184	178.78M	0.15s
UBF+StyleGAN3	7.69	0.556	0.502	58.4M	0.03s
UBF+DDPM	8.23	0.509	0.484	99.7M	90s

Table 1. Quantitative evaluation for synthetic HSIs generated in different generative models.

tral Transformer (MST++) [4] just produces HSIs that align closely with the input RGB. However, their spatial quality is marginally subpar. Notably, spectral super-resolution methods like MST++ do not generate new data. The latter three methods involve VDVAE [6], StyleGAN3 [23], and DDPM [19] used in conjunction with the proposed paradigm. Fig. 6 presents some representative examples of the abundance synthesized and HSIs generated under these three. Quantitative evaluation indicators for the generated data are listed in Table 1. When comparing these models, it is clear that both the GAN and Diffusion models have a marked advantage over the VAE. The VDVAE tends to produce results with blurry textures, low informational content, and overall poor quality. In contrast, the GAN and Diffusion models can generate HSI samples that closely resemble the spatial distribution found in actual remote sensing scenarios. It reveals that, when combined with our proposed paradigm, advanced generative models can provide a diverse, wide-ranging, and reliable set of HSI samples.

4.3. With/without unmixing

HSI reconstruction with/without unmixing. The efficacy of our proposed HSI synthesis paradigm is contingent upon

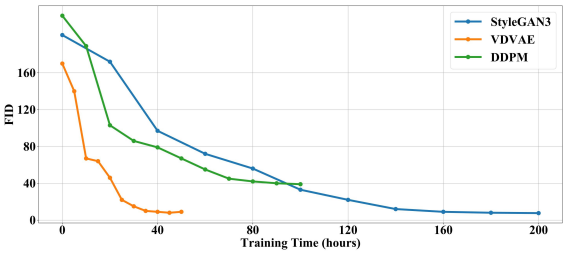

Figure 7. The FID curves within the training process by different generative models.

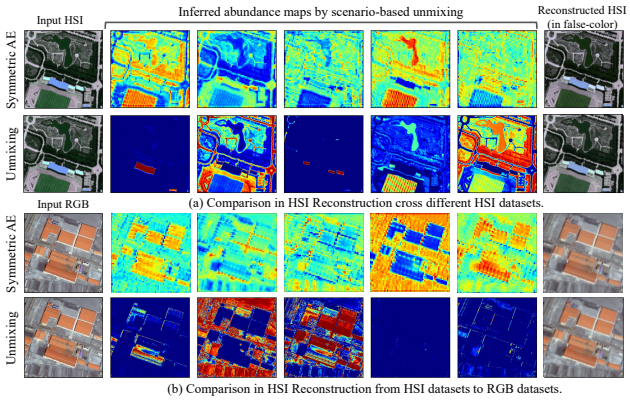

Figure 8. Illustration of HSI reconstruction comparison based on traditional AE or our scenario-based unmixing. The validation data comes from (a) the HSRS dataset (HSI) and (b) the AID dataset (RGB), respectively.

robust unmixing. In order to substantiate the merits of unmixing, we undertook an ablation study. This study juxtaposed the performance of HSI reconstruction, which involved latent feature estimation based on a traditional autoencoder (AE), and abundance estimation utilizing unmixing. The traditional AE used here has a symmetrical U-net structure, while the unmixing network, due to the introduction of LMM, has an asymmetrical structure and its decoder consists only of a linear layer. To ensure a fair comparison, the encoders in both models are identical. Additionally, the latent features in the AE and the abundance in the unmixing model share an equivalent number of channels. The findings presented in Fig. 8 offer compelling evidence of the substantial benefits of the unmixing model. When viewing this from a feature perspective, it is apparent that abundance presents a more effective representation of the spatial distribution of objects, providing a more explicit physical interpretation. From a reconstruction performance perspective, our HSI reconstructed based on unmixing has higher fidelity in both spatial and spectral dimensions.

HSI synthesis with/without unmixing. Take DDPM as an example, we compare the results generated in abundance space, original HSI cube space, and latent feature space in Fig. 9. The original HSI-based DDPM is trained on the Chikusei dataset with 128 bands, but the generation of

Generation space (dimension)	Reconstruction quality				Generation quality			Cost	
	RMSE ↓	PSNR ↑	SSIM ↑	SAD ↓	FID ↓	Recall ↑	Precision ↑	Params	Training steps
HSI Cube(128)	-	-	-	-	-	-	-	393.24M	-
Abundance(15)	0.022	36.48	0.944	4.56	49.15	0.11	0.08	100.85M	6.9 million
Abundance(8)	0.027	35.89	0.938	4.81	15.19	0.36	0.31	99.78M	4.5 million
Abundance(5)	0.034	34.26	0.931	5.47	8.23	0.50	0.48	99.76M	2.8 million
Abundance(3)	0.041	30.95	0.927	9.04	8.77	0.50	0.41	99.75M	2 million

Table 2. The quantitative performance on unmixing and synthesis under different numbers of endmembers, which equals the dimension of the abundance maps.

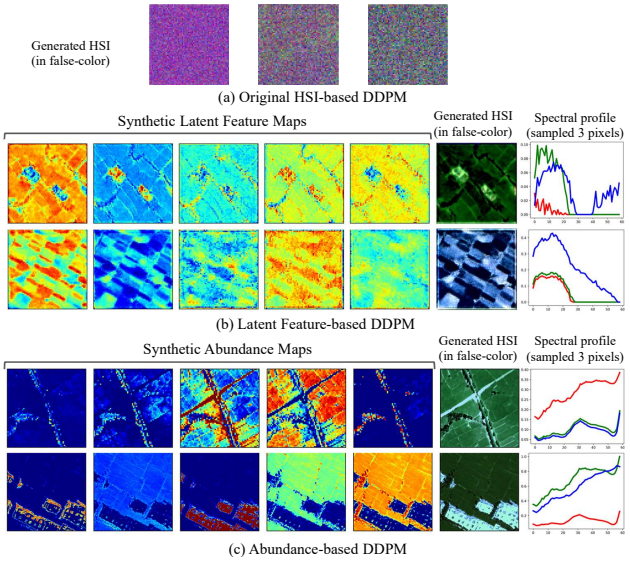

Figure 9. Illustration of HSI generation comparison on diffusion in different feature spaces.

meaningful images becomes exceptionally challenging due to the high dimensionality of the data. After 5 million iterations of the generator, the resulting image remains a random noise without any meaningful interpretation, as depicted in Fig. 9 (a). In contrast, the latent features-based DDPM provides some improvement in mitigating the curse of dimensionality. Nevertheless, despite approximately 2 million diffusion steps, the quality of the synthesized latent features and the reconstructed HSI remains unsatisfactory. The lack of physical meaning of latent features and the inherent instability of these shallow features restrict the effective reconstruction and generation of high-quality HSI. Moreover, the spectral curve of the generated HSI exhibits noticeable distortion, as demonstrated in Fig. 9 (b). Finally, the proposed abundance-based DDPM generates high-quality HSI in a more realistic style and with physical meaning, as shown in Fig. 9 (c). It implies that the utilization of abundance allows us to overcome the challenge of generating high-quality HSI in high-dimensional space.

HSI reconstruction and synthesis in different feature space. As depicted in Fig. 9 (a), generation directly in the HSI cube space still yielded meaningless results even after training for 5 million steps, making it ineffective for

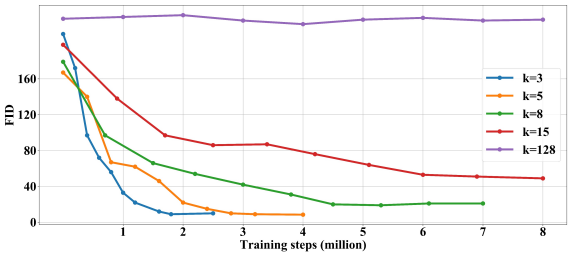

Figure 10. Comparison of the convergence rates and synthesis quality with different endmember numbers.

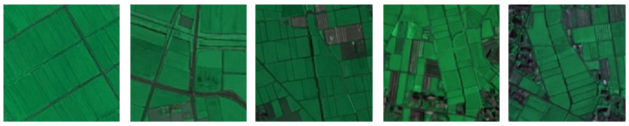

Figure 11. Examples of the false-color image of typical synthesized HSIs with only HSI training samples.

evaluating the generation quality. Fortunately, generation in the abundance space noticeably alleviated the issue. Notably, the dimension of the abundance equals the quantity of endmembers. Here, we list the evaluation on reconstruction quality, generation quality, and computational cost with different quantities of endmembers $\{15, 8, 5, 3\}$ in Table 2. We also display Fig. 10 to show the FID curves of the corresponding training process. It suggests that setting 5 endmembers is the most efficient option. It offers a trade-off between generation quality and computational consumption.

4.4. With/without fusion

Due to limited quantity, training with HSI samples alone results in generation outputs in monotonous, as shown in Fig. 11. Introducing more varied and accessible data (*i.e.*, RGB) is to provide rich and realistic spatial distributions. Boosting by such external guidance, the abundance-based diffusion is empowered to learn more robust and accurate feature representation, generating various and vastness synthesized abundance. Leveraging the strengths of both of them, the proposed method produces the generated HSIs with a diverse and reasonable spatial distribution.

4.5. Extension experiments in natural scenario

To broaden the application of our proposed HSI synthesis paradigm, we extend it to complex natural scenes. The

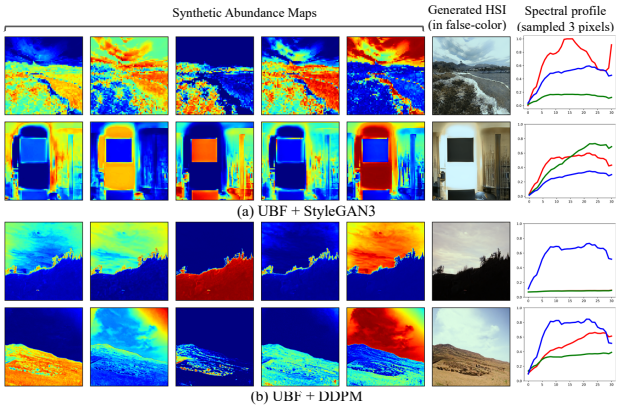

Figure 12. Typical examples of synthesis abundances and generated HSIs in different generative models under the proposed "Unmixing before Fusion" paradigm on the natural HSI dataset.

results we achieved, displayed in Fig. 12, attest to the robustness of our proposed paradigm, whether in conjunction with StyleGAN or DDPM. The synthesized HSIs demonstrate convincing visual effects, portraying a variety of natural scenes with remarkable fidelity. They yield abundance maps that accurately reflect actual spatial distributions and exhibit distinct spectral features. It underlines the effectiveness and generalization ability of our proposed paradigm in discerning the spatial distribution of different substances, even amidst the huge uncertainty of natural scenes. Particularly in Fig. 12(b), where the spectral curves align with our anticipated understanding that the spectral reflection in real backlight areas generally contains minimal information.

4.6. Extension experiments in downstream task

Our study confirms the beneficial impact of synthetic data on downstream tasks, particularly scene classification. Existing HSI datasets are limited in scale, exemplified by HSRS-SC [41], which encompasses only five categories (farmland, city building, building, water, and idle region) with sample sizes ranging from 154 to 485, totaling 1385. We have selected the former four categories and three typical classification models to test the benefits of using different augmentation strategies in constructing a larger and more balanced dataset. Related results are presented in Table 3. Regarding the results under the same augmentation scale, it is verified that HSIs synthesized by the proposed UBF paradigm hold superior diversity and reliability as opposed to those synthesized via traditional affine transformation. These synthesized HSIs generated by UBF offer considerable advantages in the training of classification models. It also demonstrates that synthetic HSI generated by our proposed UBF paradigm can enhance the diversity and scale of existing limited datasets, mitigate issues like sample scarcity and class imbalance, and potentially benefit other downstream tasks.

Augmentation	Training set scale	AlexNet	VGG-16	ResNet-18
✗	761	89.51%	87.30%	37.14%
Affine Trans.	4k	91.11%	88.84%	41.75%
Our UBF	4k	92.70%	93.97%	44.33%
Our UBF	8k	94.29%	94.60%	45.76%

Table 3. The overall scene classification accuracy on the HSRS-SC HSI dataset with/without augmentation with synthetic HSIs.

5. Discussion

Our experiments demonstrate that StyleGAN3 slightly outperforms DDPM in synthesis quality and excels in speed and efficiency. It produces realistic and high-quality HSIs resembling real-world remote sensing scenes but suffers from interclass similarity. DDPM-based generation, however, offers greater diversity. Notably, our results represent specific model instances, and variations or different architectures could potentially outperform them. We encourage exploring different GAN architectures, or alternative diffusion models with diverse priors or strategies, for potentially superior results. For limited computational resources, we suggest using GAN-based models for HSI Synthesis. If resources are abundant, we believe that an advanced diffusion model has the potential to show promising performance.

6. Conclusion

Standing at the forefront of high-dimensional data synthesis, we have boldly introduced a novel and generalized paradigm for the synthesis of HSI. This paradigm shifts the focus of HSI synthesis from the traditionally high-dimensional HSI space to a more manageable, lower-dimensional, and multi-source abundance space. This pivotal move not only gives us a dimensionality advantage but also revolutionizes how we interpret multi-source data with the spatial distribution of scenes. Furthermore, we have bridged the dimension gap between HSI and RGB by abundance, unlocking a promising future where RGB and HSI can be seamlessly fused, and even data from other sources like PAN, MSI, *etc.*, can be incorporated. With the application of advanced generative models, we have already produced a vast quantity of diverse, high-fidelity synthetic HSIs, which have shown initial positive impacts on downstream tasks, especially on scene classification. It demonstrates a closed-loop from method proposal to application verification. The proposed generalized paradigm is a significant step in the field of high-dimensional data synthesis, and we believe it has great potential to inspire and revolutionize trustworthy AI-based HSI applications.

Acknowledgments

This work was supported by NSFC (U23B200344) and NSFC of Guangdong Province (2023A1515012834).

References

[1] Naveed Akhtar and Ajmal Mian. Hyperspectral recovery from rgb images using gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1): 100–113, 2018. 1

[2] Ricardo Augusto Borsoi, Tales Imbiriba, José Carlos Moreira Bermudez, Cédric Richard, Jocelyn Chanussot, Lucas Drumetz, Jean-Yves Tourneret, Alina Zare, and Christian Jutten. Spectral variability in hyperspectral data unmixing: A comprehensive review. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):223–270, 2021. 3

[3] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *Proceedings of the International Conference on Learning Representations*, 2019. 2

[4] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 745–755, 2022. 1, 6

[5] Chein-I Chang, Shao-Shan Chiang, James A Smith, and Irving W Ginsberg. Linear spectral random mixture analysis for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 40(2):375–392, 2002. 2

[6] Rewon Child. Very deep {vae}s generalize autoregressive models and can outperform them on images. In *International Conference on Learning Representations*, 2021. 4, 5, 6

[7] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2

[8] Viktor Dremin, Zbignevs Marcinkevics, Evgeny Zherebtsov, Alexey Popov, Andris Grabovskis, Hedviga Kronberga, Kristine Geldnere, Alexander Doronin, Igor Meglinski, and Alexander Bykov. Skin complications of diabetes mellitus revealed by polarized hyperspectral imaging and machine learning. *IEEE Transactions on Medical Imaging*, 40(4): 1207–1216, 2021. 1

[9] Bin Fan, Yuzhu Yang, Wensen Feng, Fuchao Wu, Jiwen Lu, and Hongmin Liu. Seeing through darkness: Visual localization at night via weakly supervised learning of domain invariant features. *IEEE Transactions on Multimedia*, 2022. 2

[10] Xin-Ru Feng, Heng-Chao Li, Rui Wang, Qian Du, Xiuping Jia, and Antonio Plaza. Hyperspectral unmixing based on nonnegative matrix factorization: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:4414–4436, 2022. 2

[11] Hang Fu, Genyun Sun, Li Zhang, Aizhu Zhang, Jinchang Ren, Xiuping Jia, and Feng Li. Three-dimensional singular spectrum analysis for precise land cover classification from uav-borne hyperspectral benchmark datasets. *ISPRS Journal of Photogrammetry and Remote Sensing*, 203:115–134, 2023. 1

[12] Eloi Grau and Jean-Philippe Gastellu-Etchegorry. Radiative transfer modeling in the earth–atmosphere system with dart model. *Remote Sensing of Environment*, 139:149–170, 2013. 1

[13] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, and Jieping Ye. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):3313–3332, 2021. 2

[14] Sanghui Han and John P Kerekes. Overview of passive optical multispectral and hyperspectral image simulation techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11):4794–4804, 2017. 1

[15] Wei Han, Xiaohan Zhang, Yi Wang, Lizhe Wang, Xiaohui Huang, Jun Li, Sheng Wang, Weitao Chen, Xianju Li, Ruyi Feng, et al. A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202:87–113, 2023. 1

[16] Juan Mario Haut, Mercedes E Paoletti, Javier Plaza, Antonio Plaza, and Jun Li. Hyperspectral image classification using random occlusion data augmentation. *IEEE Geoscience and Remote Sensing Letters*, 16(11):1751–1755, 2019. 1

[17] Xiaoyu He and Xiaojian Xu. Physically based model for multispectral image simulation of earth observation sensors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5):1897–1908, 2017. 1

[18] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30, 2017. 5

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 2, 4, 5, 6

[20] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *The Journal of Machine Learning Research*, 23(1):2249–2281, 2022. 2

[21] Qiwen Jin, Yong Ma, Fan Fan, Jun Huang, Xiaoguang Mei, and Jiayi Ma. Adversarial autoencoder network for hyperspectral unmixing. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8):4555–4569, 2023. 1

[22] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019. 2

[23] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34:852–863, 2021. 4, 5, 6

[24] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in Neural Information Processing Systems*, 34:21696–21707, 2021. 2

[25] Ivan Kobyzev, Simon JD Prince, and Marcus A Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, 2020. 2

[26] Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems*, 32, 2019. 5

[27] Hongmin Liu, Fan Jin, Hui Zeng, Huayan Pu, and Bin Fan. Image enhancement guided object detection in visually degraded scenes. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 1

[28] Kangfu Mei and Vishal Patel. Vidm: Video implicit diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9117–9125, 2023. 2

[29] Burkni Palsson, Johannes R Sveinsson, and Magnus O Ulfarsson. Blind hyperspectral unmixing using autoencoders: A critical comparison. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:1340–1372, 2022. 2

[30] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research*, 22(1):2617–2680, 2021. 2

[31] Jialun Peng, Dong Liu, Songcen Xu, and Houqiang Li. Generating diverse structure for image inpainting with hierarchical vq-vae. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10775–10784, 2021. 2

[32] Ali Razavi, Aaron Van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae-2. *Advances in Neural Information Processing Systems*, 32, 2019. 2

[33] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 2

[34] Linus Scheibenreif, Michael Mommert, and Damian Borth. Masked vision transformers for hyperspectral image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2165–2175, 2023. 1

[35] Pourya Shamsolmoali, Masoumeh Zareapoor, Eric Granger, Huiyu Zhou, Ruili Wang, M Emre Celebi, and Jie Yang. Image synthesis with adversarial networks: A comprehensive survey and case studies. *Information Fusion*, 72:126–146, 2021. 1

[36] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019. 1

[37] Arash Vahdat and Jan Kautz. Nvae: A deep hierarchical variational autoencoder. *Advances in Neural Information Processing Systems*, 33:19667–19679, 2020. 2

[38] Sander Veraverbeke, Philip Dennison, Ioannis Gitas, Glynn Hulley, Olga Kalashnikova, Thomas Katagis, Le Kuai, Ran Meng, Dar Roberts, and Natasha Stavros. Hyperspectral remote sensing of fire: State-of-the-art and future perspectives. *Remote Sensing of Environment*, 216:105–121, 2018. 1

[39] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision*, pages 3–19, 2018. 3

[40] 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. 5

[41] 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. 8

[42] Yonghao Xu, Tao Bai, Weikang Yu, Shizhen Chang, Peter M Atkinson, and Pedram Ghamisi. Ai security for geoscience and remote sensing: Challenges and future trends. *IEEE Geoscience and Remote Sensing Magazine*, 11(2):60–85, 2023. 1

[43] Dongchao Yang, Jianwei Yu, Helin Wang, Wen Wang, Chao Weng, Yuexian Zou, and Dong Yu. Diffsound: Discrete diffusion model for text-to-sound generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023. 2

[44] Jing Yao, Danfeng Hong, Jocelyn Chanussot, Deyu Meng, Xiaoxiang Zhu, and Zongben Xu. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. In *Proceedings of the European Conference on Computer Vision*, pages 208–224. Springer, 2020. 2

[45] Naoto Yokoya and Akira Iwasaki. Airborne hyperspectral data over chikusei. *Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27*, 5:5, 2016. 5

[46] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2011. 2

[47] Fangneng Zhan, Yingchen Yu, Rongliang Wu, Jiahui Zhang, Shijian Lu, Lingjie Liu, Adam Kortylewski, Christian Theobalt, and Eric Xing. Multimodal image synthesis and editing: A survey. *arXiv preprint arXiv:2112.13592*, 2021. 1

[48] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 2

[49] 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. 1

[50] Yuan Zhou, Anand Rangarajan, and Paul D Gader. A gaussian mixture model representation of endmember variability in hyperspectral unmixing. *IEEE Transactions on Image Processing*, 27(5):2242–2256, 2018. 5