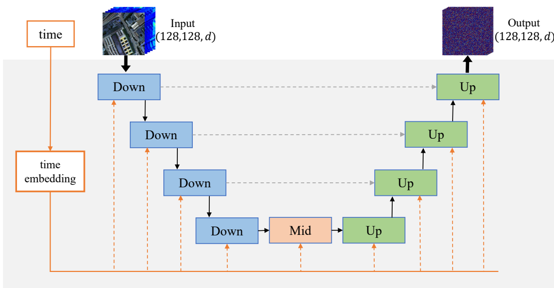
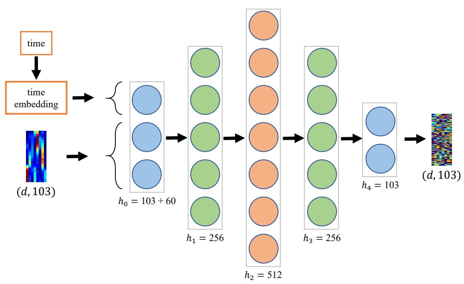
Self-Learning Hyperspectral and Multispectral Image Fusion via Adaptive Residual Guided Subspace Diffusion Model

Supplementary Material

1. Details of Network Architecture



(a) Spatial Network



(b) Spectral Network

Figure 1. The detailed network structures using the Pavia dataset as an example.

The detailed network structure is illustrated in Figure 1. To explore the most cost-effective spatial-spectral network, we conducted ablation experiments on these two networks. Specifically, for the spatial network, we performed an ablation study on the number of channels within the network, controlled by the `channel_mult` parameter. For the spectral network, we examined the effect of varying the dimensions of hidden layers, determined by the `hidden_dims` parameter. The experimental results are summarized in Table 1. Considering both accuracy and model size, we selected the spatial network with `channel_mult` = {1,2,3,4} and the spectral network with `hidden_dims` = {256,512,256}. (**Note:** Based on our existing experimental results, the ablation study of the spatial network is conducted with the optimal spectral network configuration by default, and vice versa for the spectral network ablation study.)

Table 1. Ablation analysis of spectral and spatial networks.

Networks	channel_mult/hidden_dims	PSNR	Params (M)
Spa. Net	1,1,2,4	41.55	15.60
	1,2,3,4	42.33	21.46
	1,2,4,8	42.40	57.03
Spe. Net	256,256	41.64	0.19
	256,512,256	42.33	0.39
	256,512,512,256	42.36	0.65

2. Impact of Two Components

Since our method simultaneously updates both the spectral and spatial components, we conduct an ablation study to investigate their individual contributions by evaluating the cases where only the spectral component or only the spatial component is updated. Specifically, when updating only the spectral component, we fix the spatial component by deriving $\mathcal{A}$ from $\mathcal{Y}$ (HR-MSI) using the Archetypal Analysis unmixing method [1]. Conversely, when updating only the spatial component, we fix the spectral component by estimating $\mathbf{E}$ from $\mathcal{X}$ (LR-HSI) following the approach of PLRDiff [2]. The final results are presented in Table 2, showing that when only the spectral component is updated, the fusion accuracy significantly decreases. Additionally, when only the spatial component is updated, the PSNR of the fusion results remains approximately 2 dB lower than when both components are updated simultaneously.

Table 2. Ablation study on two components.

Update	PSNR↑	SAM↓	EGARS↓	SSIM↑
Only $\mathbf{E}$	21.77	11.82	13.59	0.296
Only $\mathcal{A}$	40.53	3.28	1.77	0.969
$\mathbf{E}$ & $\mathcal{A}$	42.33	2.64	1.49	0.977

3. Sensitivity Analysis of Hyperparameters

We present the parameter analysis results of the balance weights λ_1 and λ_2 in Table 3. Since the performance is optimal and relatively stable around $(1, 1)$, we set both values to 1.

References

[1] Adele Cutler and Leo Breiman. Archetypal analysis. *Technometrics*, 36(4):338–347, 1994. 1

Table 3. Sensitivity analysis of the parameters λ_1 and λ_2.

λ_2 \ λ_1	0.1	0.5	1	2	5
0.1	42.08	42.22	42.33	41.91	41.44
0.5	42.15	42.31	42.25	41.90	41.63
1	42.26	42.30	42.33	41.99	41.59
2	42.04	42.06	42.13	42.33	41.40
5	41.61	41.80	41.41	41.19	40.00

[2] Xiangyu Rui, Xiangyong Cao, Li Pang, Zeyu Zhu, Zongsheng Yue, and Deyu Meng. Unsupervised hyperspectral pansharpening via low-rank diffusion model. *Information Fusion*, 107: 102325, 2024. 1