Automatic Spectral Calibration of Hyperspectral Images: Method, Dataset and Benchmark

Supplementary Material

We present detailed benchmarks, including separate evaluations for visible light and near-infrared (NIR) spectra. Additionally, we provide illumination-specific benchmarks for the expansion dataset. To enhance understanding, we include more visual results from both the original and expansion datasets. Lastly, we offer a comprehensive comparison of existing hyperspectral imaging (HSI) datasets for natural scenes.

8. Detailed Evaluation of Calibration Methods: Visible and Near-Infrared Ranges

To further verify the quality of the spectral calibration, we evaluate the calibrated HSIs from the visible (400-700nm) and near-infrared (700-1000nm) perspectives. We conduct experiments on both the BJTU-UVA and BJTU-UVA-E datasets.

Comparison on the BJTU-UVA Table 4 presents the results. Compared to other learning-based methods, our approach demonstrates a significant advantage across all evaluation metrics in both the visible and near-infrared (NIR) ranges.

In the visible range, our proposed SIT model achieves superior scores in PSNR, RMSE, and ERGAS. Notably, the Gray-World method excels in visible range calibration, achieving the best SAM value. Leveraging the strengths of the Gray-World approach, our SIT model is designed to more effectively capture illumination compared to other learning-based methods.

In the NIR range, the performance of all methods declines considerably, underscoring the inherent challenges of this spectrum compared to the visible range. Despite this, our method consistently achieves the best scores across all metrics, reaffirming its robustness. The inclusion of the NIR spectrum in our dataset emphasizes the need to address these challenges and drive further advancements in this area.

Comparison on the BJTU-UVA-E On the BJTU-UVA-E dataset, our proposed SIT model consistently achieved the best scores across all evaluation metrics. In this context, the Gray-World method was not comparable to the learning-based approaches. Similarly, a significant performance drop was observed in the NIR range, highlighting the persistent challenges associated with this spectrum. Addressing the calibration limitations of the NIR range remains an open challenge for future research.

9. Evaluation on Different Illumination Conditions of the BJTU-UVA-E Dataset

To evaluate model performance under varying illumination conditions, we analyzed the PSNR calibration results across ten different illumination settings. The definitions and spectra of these illumination settings are illustrated in Fig. 3, and the results are summarized in Table 5.

The Gray-World method, which relies on linear functions of averaging and division, produces uniform predictions across different illumination settings, making it incomparable to other methods in this context. By contrast, our proposed SIT method consistently outperforms all competitors across all illumination settings.

Focusing on the five natural illuminations, we observe that dusk (du) and shadowy (sh) conditions pose greater calibration challenges due to their low-light environments. Even in these difficult settings, our method achieves the highest PSNR values, with 37.1 in "du" and 36.3 in "sh" for the NIR range.

For the five color-filtered illuminations, the red-filtered illumination presents the most significant challenge, as it limits visible light intensity between 400nm and 500nm. Despite this, our method maintains its superior performance.

10. More Qualitative Analysis

We present six examples from the BJTU-UVA dataset in Fig. 7 and Fig. 8, and six additional examples in Fig. 9 and Fig. 10. Each example includes the following evaluations:

- **First row:** The first two images represent the uncalibrated HSI and the ground truth for the calibrated HSI, both rendered as RGB images for visualization. Columns 3 to 7 display error heat maps of the calibration results from five methods across the full spectral range.
- **Second and third rows:** These rows visualize the error heat maps specifically for the visible and near-infrared spectral ranges, respectively.
- **Bottom-left corner:** The spectrum of a selected pixel is plotted, showing the output from various calibration methods for comparison.

For the BJTU-UVA heat maps, we use a threshold of 0.14, derived from the global 95th percentile of a randomly selected subset. For the BJTU-UVA-E heat maps, we set

		Val								Test							
		VIS				NIR				VIS				NIR			
		P↑	S ↓	R↓	E↓	P↑	S ↓	R↓	E↓	P↑	S ↓	R↓	E↓	P↑	S ↓	R↓	E↓
BJTU-UVA	Gray-World[5]	26.0	**2.5**	5.4	8.9	21.4	3.1	10.8	10.5	25.1	**2.8**	6.0	13.3	22.3	3.3	9.4	9.8
	DivIll[23]	25.6	5.7	6.3	9.6	21.7	5.3	9.6	9.8	25.0	6.2	6.0	10.6	22.5	5.9	8.7	9.2
	SERT[19]	24.1	10.5	6.3	12.4	20.2	11.2	9.9	13.2	23.6	10.6	6.1	14.2	20.8	11.5	9.1	12.7
	HCANet[13]	27.5	4.5	5.0	8.2	23.6	4.1	7.7	8.9	26.6	4.5	5.0	10.3	23.9	4.3	7.3	9.1
	SIT(proposed)	**28.5**	2.9	**4.5**	**8.1**	**24.1**	**3.0**	**7.3**	**8.7**	**27.4**	3.1	**4.6**	**10.0**	**24.5**	**3.3**	**6.6**	**8.9**
BJTU-UVA-E	Gray-World[5]	25.9	2.2	5.7	9.1	21.3	3.0	11.1	10.6	25.0	2.2	6.2	13.6	22.4	3.2	9.5	9.8
	DivIll[23]	36.6	3.3	1.7	2.9	33.4	3.3	2.5	2.6	37.0	3.7	1.5	3.1	33.1	3.7	2.6	2.9
	SERT[19]	37.3	3.1	1.8	2.9	33.3	3.1	2.6	3.0	37.8	3.3	1.5	3.0	33.1	3.3	2.5	3.0
	HCANet[13]	40.6	2.2	1.0	1.9	35.3	2.2	1.8	2.1	41.1	2.4	0.9	2.1	34.7	2.4	1.9	2.3
	SIT(proposed)	**42.5**	**1.7**	**0.8**	**1.5**	**37.2**	**1.7**	**1.5**	**1.7**	**43.2**	**1.8**	**0.7**	**1.6**	**36.5**	**1.8**	**1.6**	**1.9**

Table 4. Automatic spectral calibration evaluation on both visible-spectrum (VIS, 400nm-700nm) and near-infrared (NIR, 700nm-1000nm) HSI: PSNR(P), SAM(S), RMSE (R%), and ERGAS (E%).

		Validation Set										**Test Set**									
		sd	**cd**	**rd**	**du**	**sh**	**bl**	**re**	**ye**	**gr**	**pu**	**sd**	**cd**	**rd**	**du**	**sh**	**bl**	**re**	**ye**	**gr**	**pu**
Gray-World[5]	**vis**	25.9	25.9	25.9	25.9	25.9	25.9	25.9	25.9	25.9	25.9	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
	nir	21.3	21.3	21.3	21.3	21.3	21.3	21.3	21.3	21.3	21.3	22.4	22.4	22.4	22.4	22.4	22.4	22.4	22.4	22.4	22.4
	whole	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	23.2	23.2	23.2	23.2	23.2	23.2	23.2	23.2	23.2	23.2
DivIll[23]	**vis**	37.5	38.9	39.1	36.5	31.8	38.1	34.2	38.1	36.9	38.6	37.9	39.8	39.6	36.5	30.8	38.8	34.7	38.4	37.7	39.3
	nir	34.0	32.1	33.9	32.5	29.4	34.8	35.1	33.8	35.8	35.8	33.4	32.9	33.5	32.0	28.6	34.5	34.8	33.6	35.3	35.4
	whole	35.9	34.8	36.2	34.7	31.0	36.7	35.2	36.0	36.9	37.5	35.7	35.9	36.1	34.5	30.4	36.7	35.5	36.0	36.9	37.6
HCANet[13]	**vis**	43.1	43.0	42.6	40.8	41.4	40.8	34.5	39.6	38.4	41.8	43.9	44.4	43.5	41.8	42.9	41.7	35.2	40.1	38.9	42.2
	nir	35.8	34.4	34.6	34.3	33.8	36.0	36.2	35.7	36.2	36.2	35.3	34.6	34.0	34.2	33.9	35.6	35.8	35.3	35.8	35.8
	whole	38.6	37.4	37.4	37.0	36.7	38.3	35.7	37.7	37.6	38.6	38.4	37.9	37.1	37.2	37.1	38.3	36.1	37.6	37.6	38.5
SERT[19]	**vis**	40.5	42.6	38.2	36.1	28.1	38.0	35.2	38.8	38.1	40.7	41.1	43.4	38.7	36.4	29.6	38.4	35.2	39.4	38.4	41.2
	nir	32.9	34.3	32.2	31.4	25.4	35.4	35.9	36.2	35.7	36.2	32.4	34.6	32.2	31.1	26.0	35.0	35.5	35.5	35.4	35.9
	whole	35.8	37.3	34.8	33.8	26.9	37.1	36.1	37.7	37.2	38.5	35.7	37.8	35.1	33.9	28.1	37.1	36.1	37.5	37.3	38.4
SIT	**vis**	**44.5**	**47.7**	**46.2**	**42.7**	**43.9**	**42.5**	**36.4**	**40.3**	**39.3**	**44.8**	**45.3**	**48.4**	**47.2**	**43.1**	**44.8**	**42.8**	**36.8**	**41.0**	**40.0**	**45.5**
	nir	**37.4**	**37.6**	**37.4**	**37.1**	**36.3**	**37.8**	**38.0**	**37.7**	**37.9**	**38.3**	**36.5**	**37.1**	**36.8**	**36.4**	**35.7**	**37.0**	**37.2**	**36.9**	**37.0**	**37.6**
	whole	**40.2**	**40.8**	**40.3**	**39.6**	**39.1**	**40.0**	**37.6**	**39.2**	**38.9**	**41.0**	**39.7**	**40.5**	**40.1**	**39.2**	**38.9**	**39.6**	**37.5**	**39.1**	**38.8**	**40.6**

Table 5. Calibration results (PSNR) of visible, near-infrared, and full-range on the validation and test sets of BJTU-UVA-E.

the threshold to 0.03, based on the global 85th percentile of a similar subset, to enhance the clarity of the comparisons.

Qualitative Analysis on BJTU-UVA As illustrated in Fig. 7(c), the calibration errors of our proposed SIT method are the smallest across all regions, particularly for pure blue areas. In contrast, other methods struggle to achieve accurate calibration, especially in the right regions containing more red areas. The spectral analysis further confirms that the calibrated spectrum curve produced by SIT exhibits the highest similarity to the ground truth. From the visualization results in the visible (VIS) and near-infrared (NIR) spectral ranges, it is evident that the error maps of NIR from other methods display significantly more red areas compared to VIS. In contrast, the heatmap generated by SIT remains predominantly blue, demonstrating its ability to effectively handle calibration in the near-infrared range. In the remaining five examples, our method consistently achieves the best results.

Qualitative Analysis on BJTU-UVA-E Similarly, our SIT method outperforms other approaches on the BJTU-UVA-E dataset. For instance, in Fig. 9(c), the Gray-World method performs poorly, with a calibrated spectrum that deviates significantly from the ground truth. The shrubbery regions with shadow in this example present a particular challenge, especially in the NIR range, yet the SIT method achieves near-perfect calibration in these areas. Across other examples, it is apparent that the Gray-World method consistently underperforms on the BJTU-UVA-E dataset, while SIT maintains superior results.

11. Comparison with Existing Hyperspectral Datasets for Natural Scenes

As shown in Table 6, the rapid development of hyperspectral imaging devices has shifted research focus from

Dataset	Bands	Images	Spectral Range	Spatial Resolution	Denoising	Super-resolution	Spectral Recovery	Spectral Calibration	Year
CAVE [35]	31	32	400-700	512×512	✓	✓	✓		2010
Harvard [36]	31	50	400-700	1392×1040	✓	✓	✓		2011
NUS [37]	31	64	400-700	1392×1300	✓	✓	✓		2014
ICVL [38]	31	201	400-700	1392×1300	✓	✓	✓		2016
NTIRE'18 [39]	31	256	400-700	480×512			✓		2018
NTIRE'20 [40]	31	510	400-700	480×512			✓		2020
NTIRE'22 [41]	31	1000	400-700	480×512			✓		2022
FHRS [42]	31	607	400-700	512×512			✓		2023
BJTU-UVA	**204**	**765 (7650)**	**400-1000**	**512×512**	✓*	✓*	✓*	✓	**2024**

Table 6. Comparison of public hyperspectral image datasets. Only the proposed dataset BJTU-UVA is suitable for auto-calibration because other datasets don't provide raw data. Although this paper focuses on auto-calibration, other tasks are also possible by the proposed dataset, which will be released later. 7650 is the size of the expansion dataset.

traditional remote sensing to natural environments on the ground. In natural scenes, several hyperspectral datasets have been proposed for restoration tasks, including denoising [19], spatial super-resolution [43], and spectral super-resolution [42]. These datasets typically focus on the visible spectrum, ranging from 400 nm to 700 nm, and downsample the spectral resolution at intervals of 10 nm.

To enable comprehensive spectral recovery across both the visible and near-infrared ranges, our proposed dataset retains the full-spectrum data while also providing visible-range HSIs with 31 channels. Additionally, our dataset contains significantly more images compared to other HSI datasets, making it uniquely comprehensive. Importantly, only our proposed BJTU-UVA dataset is suitable for auto-calibration, as other datasets do not include raw data essential for this task.

Although this paper focuses on auto-calibration, the BJTU-UVA dataset is versatile and can support other tasks as well. The dataset will be publicly released in the future to facilitate further research.

Supplementary Reference

[35] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K. Nayar. Generalized assorted pixel camera: post-capture control of resolution, dynamic range, and spectrum. IEEE Transactions on Image Processing, 19(9):2241–2253, 2010.

[36] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 193–200, IEEE, 2011.

[37] Rang MH Nguyen, Dilip K. Prasad, and Michael S. Brown. Training-based spectral reconstruction from a single RGB image. In European Conference on Computer Vision, pages 186–201. Springer, 2014.

[38] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In European Conference on Computer Vision, pages 19–34. Springer, 2016.

[39] Boaz Arad, Ohad Ben-Shahar, and Radu Timofte. NTIRE 2018 challenge on spectral reconstruction from RGB images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 929–938, 2018.

[40] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D. Finlayson. NTIRE 2020 challenge on spectral reconstruction from an RGB image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 446–447, 2020.

[41] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. NTIRE 2022 spectral recovery challenge and dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 863–881, 2022.

[42] Du Z., Wei S., Liu T., et al. Exploring the applicability of spectral recovery in semantic segmentation of RGB images. IEEE Transactions on Multimedia, 2023.

[43] Zhang M., Zhang C., Zhang Q., et al. ESSAformer: Efficient transformer for hyperspectral image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 23073–23084, 2023.

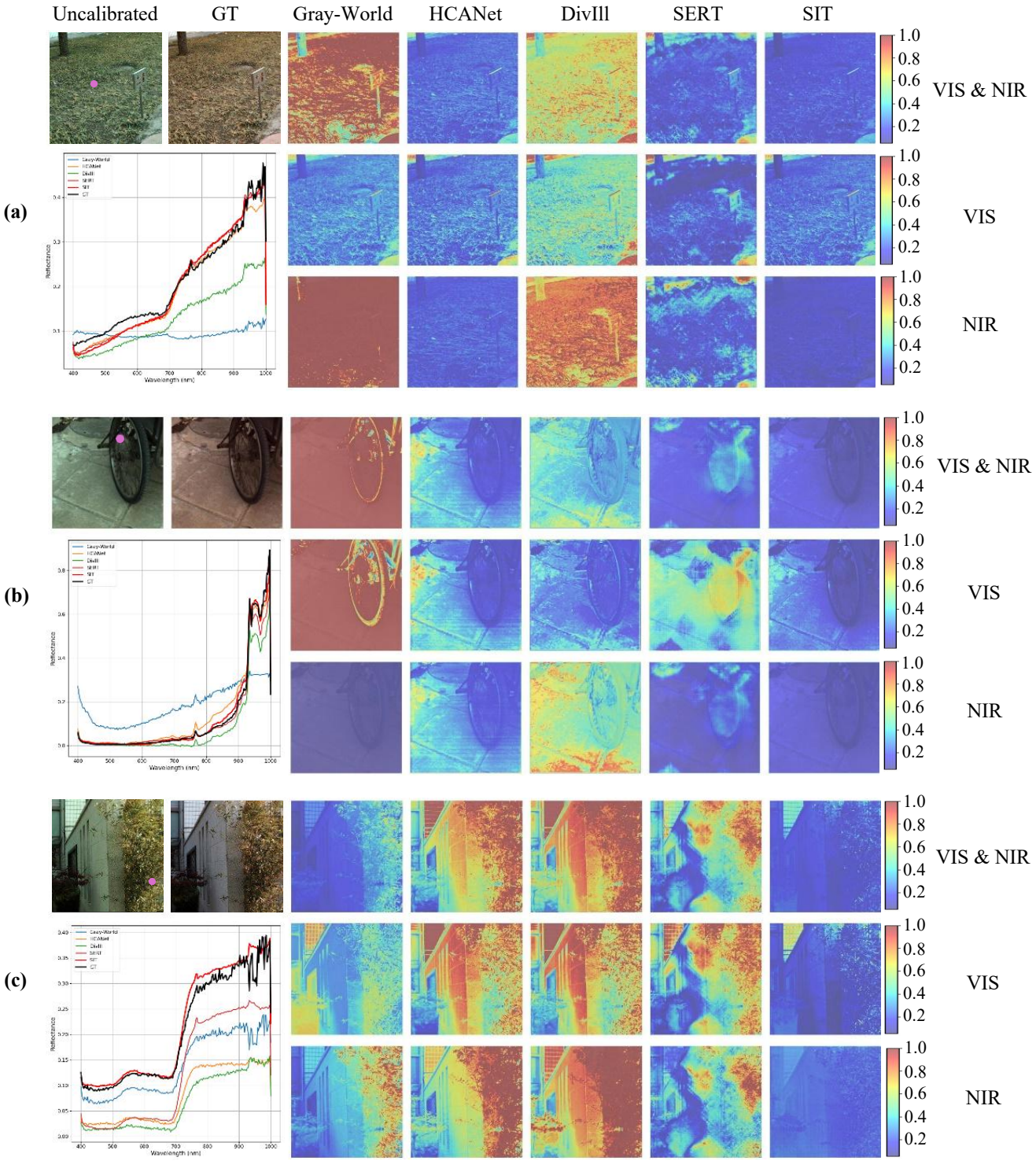

Figure 7. Visual Comparison of Absolute Error Using Heat Maps on BJTU-UVA. First row: Uncalibrated HSI, ground truth, and error maps (full spectrum) for five methods. Second and third rows: Error maps for VIS and NIR ranges. Bottom-left: Pixel spectra comparison across methods.

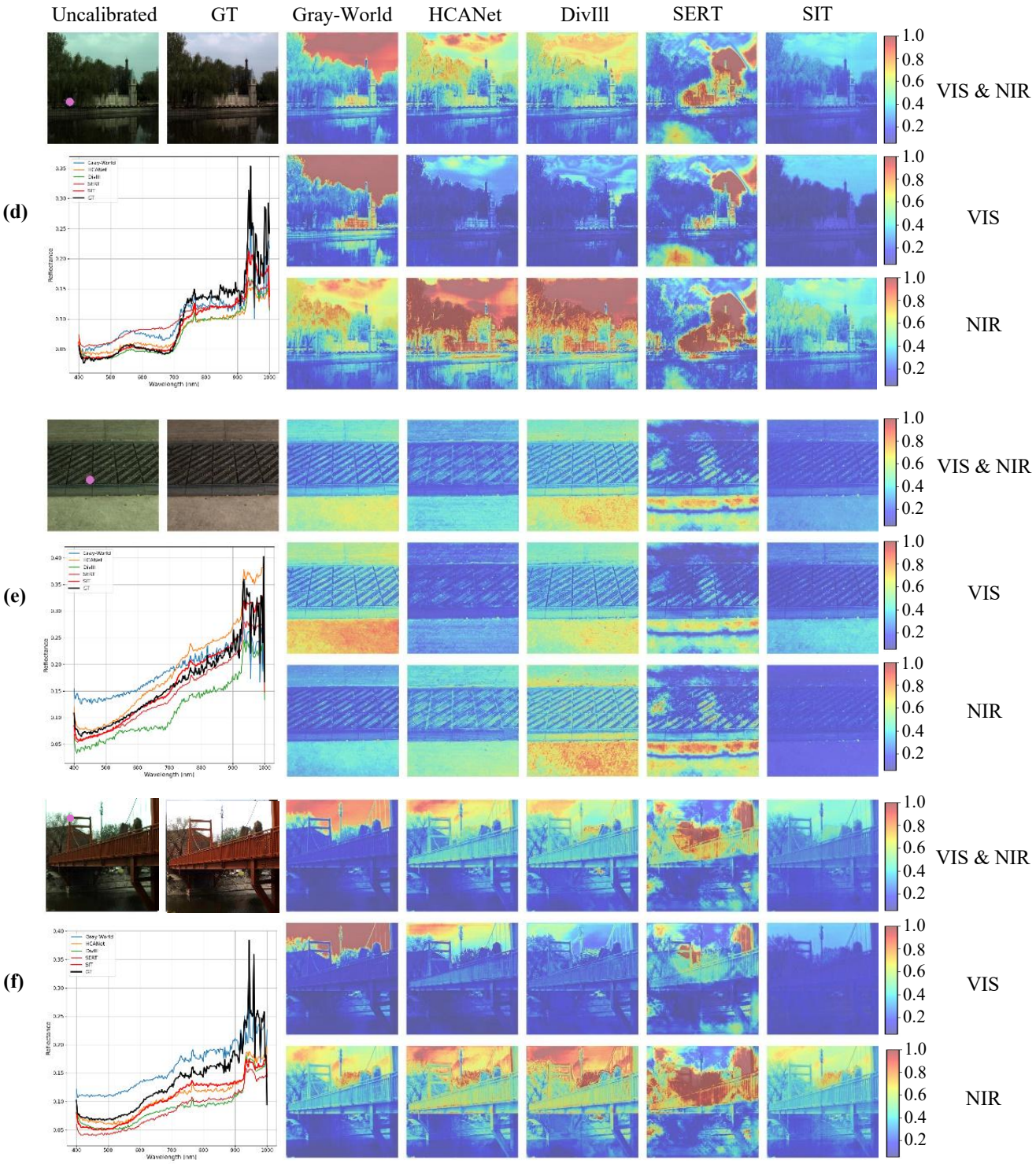

Figure 8. Visual Comparison of Absolute Error Using Heat Maps on BJTU-UVA.

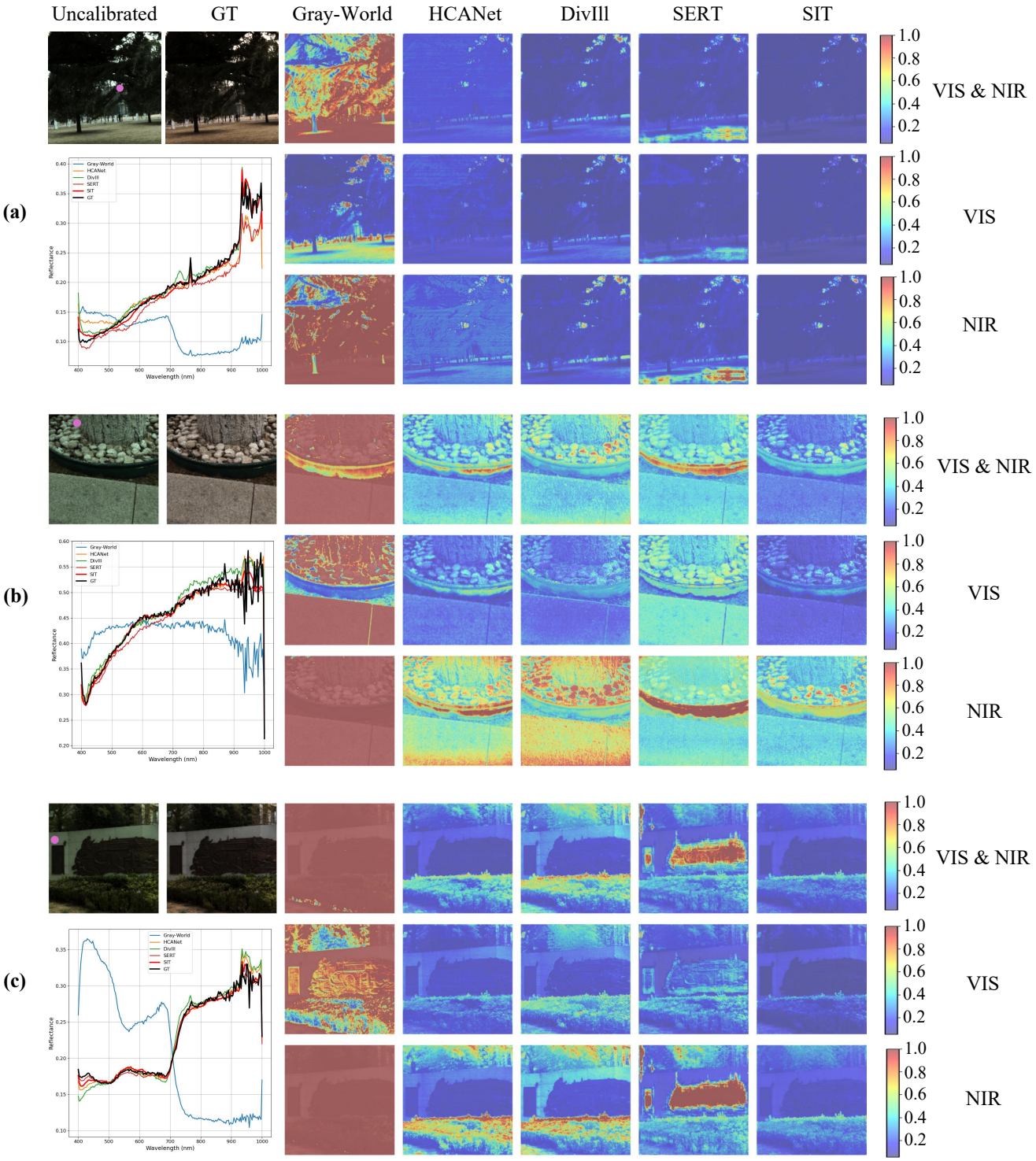

Figure 9. Visual Comparison of Absolute Error Using Heat Maps on BJTU-UVA-E.

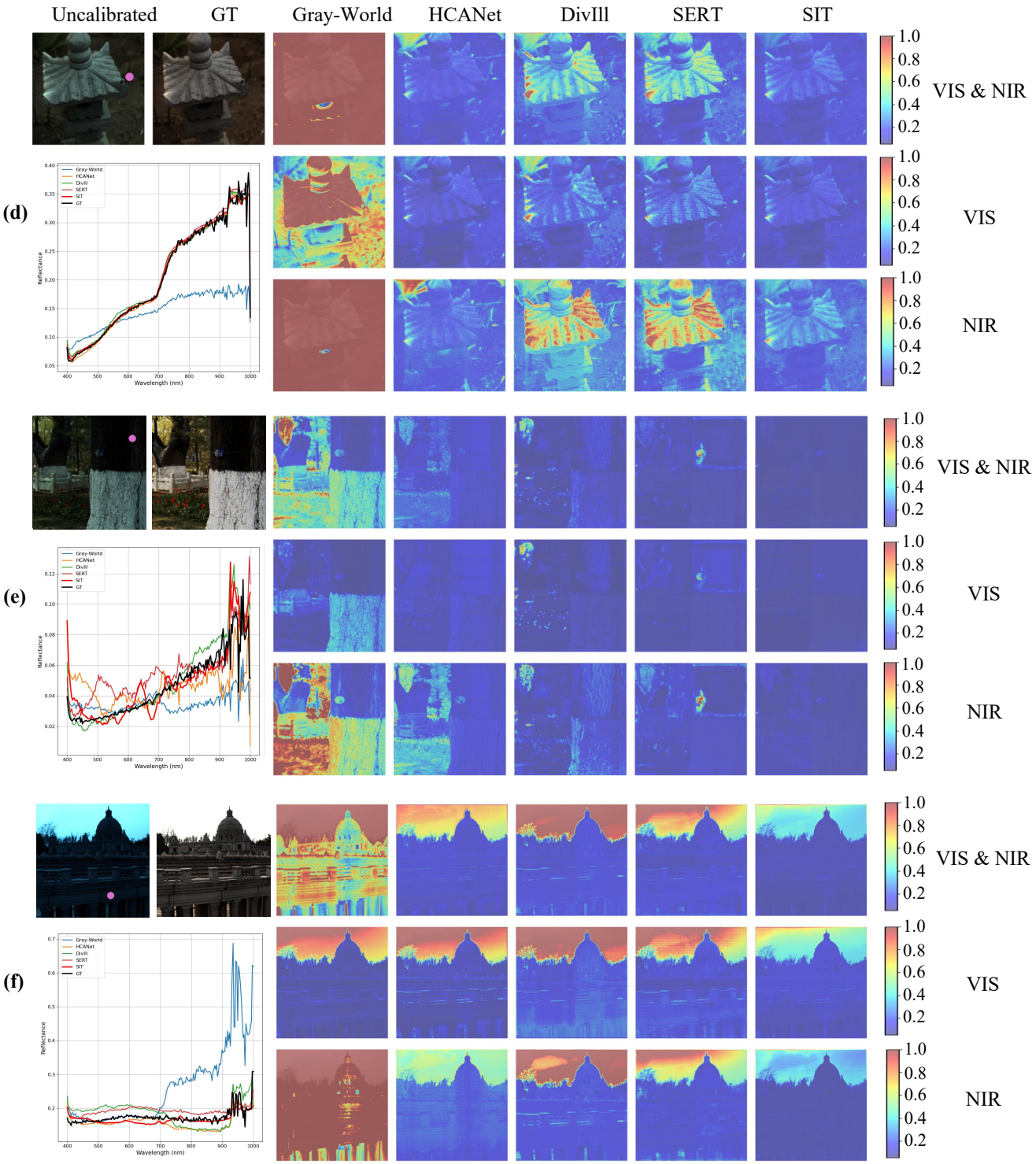

Figure 10. Visual Comparison of Absolute Error Using Heat Maps on BJTU-UVA-E.