Hipandas: Hyperspectral Image Joint Denoising and Super-Resolution by Image Fusion with the Panchromatic Image

SUPPLEMENTARY MATERIALS

Abstract

This supplementary material serves to provide additional findings and discussions presented in the main manuscript titled "Hipandas: Hyperspectral Image Joint Denoising and Super-Resolution by Image Fusion with the Panchromatic Image." It is designed to offer a more detailed exploration and a clearer understanding of the various components and outcomes of the research. The supplementary content includes the following detailed sections:

1. *A description of the baseline configurations that were utilized for the comparative evaluation of different methods, which is detailed in Sec. S-1;*
2. *A sensitivity analysis that delves into the weights assigned to the loss function, stage transition timing, and stage learning rates, is discussed in Sec. S-2;*
3. *A visual comparison of the fusion results obtained from the different methods presented in Sec. S-4;*
4. *Additional results from the application of the proposed method to real-world datasets provided in Sec. S-5.*
5. *An analysis of rank selection for the proposed UHipandas method outlined in Sec. S-6.*
6. *A comparison of inference times across different methods detailed in Sec. S-3.*
7. *An exploration into the interaction between pandenoising and pansharpening processes discussed in Sec. S-7.*
8. *An investigation into the intermediate results and the relationship between pandenoising and pansharpening within the UHipandas method presented in Sec. S-8.*

This supplementary material is intended to enhance the reader's comprehension of the Hipandas method and its components.

S-1. Baseline configuration

This section provides a description of the configuration parameters and setup for the baseline methods that were employed in our experimental evaluations.

- PWTV is a model-based optimization algorithm specifically designed for the pandenoising task. We conducted the grid search to obtain the optimal performance of PWTV within our experimental framework.
- SwinIR, a state-of-the-art Transformer-based network originally developed for RGB image restoration, presents a challenge when applied to our dataset due to its limited number of images. Consequently, instead of attempting to finetune SwinIR on our dataset, pretrained weights were utilized in the experiments.
- PLRD, HIRD, and RPNN are three methods that have been previously proposed for HS pansharpening or HSI denoising. For the purpose of a fair and direct comparison with our proposed method, we have carefully reimplemented these algorithms on our dataset.

S-2. More ablation experiments

S-2.1. On loss function weights

In the main manuscript, the components of our loss function were combined by simple summation, without assigning specific weights to individual terms. In this supplementary material, we delve deeper into the significance of each term by introducing weights to investigate their relative contributions to the overall optimization process. The modified loss functions are defined as follows:

$$\begin{aligned} \mathcal{L}^{(1)} &= \lambda_1 \mathcal{L}_D + \lambda_2 \mathcal{L}_S^{(1)} + \lambda_3 \mathcal{L}_Q, \\ \mathcal{L}^{(2)} &= \lambda_1 \mathcal{L}_D + \lambda_2 \mathcal{L}_S^{(2)} + \lambda_3 (\mathcal{L}_Q + \mathcal{L}_P). \end{aligned} \tag{1}$$

Here, λ_1, λ_2, and λ_3 represent the weights assigned to the denoising loss ($\mathcal{L}_D$), the first and second stage super-resolution loss ($\mathcal{L}_S^{(1)}$ and $\mathcal{L}_S^{(2)}$), and the PAN image's texture loss ($\mathcal{L}_Q$ and $\mathcal{L}_P$), respectively. The results presented in Table S-1 suggest that the adjustment of these weights does not lead to substantial changes in the overall performance.

S-2.2. On stage transition timing

The stage transition timing t represents a critical hyperparameter in our two-stage training process, determining when the model switches from the first stage (pre-training) to the second stage (fine-tuning). Our experiments investi-

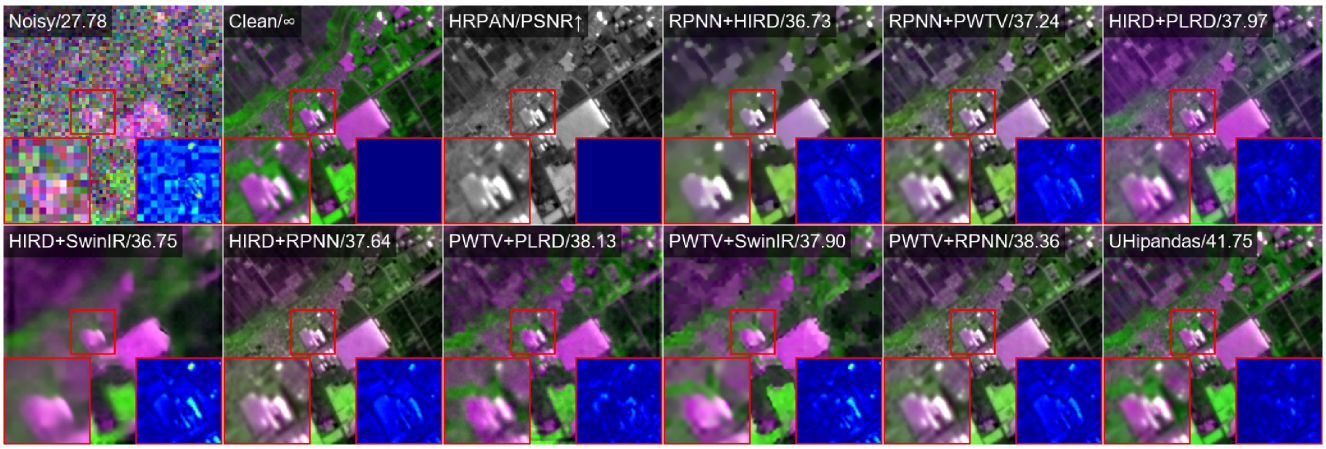

Figure S-1. Restoration results on the Dongying_8_0 image for Gaussian noise with $\sigma = 10$.

Table S-1. Sensitivity analysis of loss weights in UHipandas.

$\lambda_1 : \lambda_2 : \lambda_3$	1:1:1 (ours)	2:1:1	4:1:1	1:2:1	1:4:1	1:1:2	1:1:4
PSNR↑	36.61	**36.68**	36.07	36.48	36.37	36.00	35.03
SSIM↑	**0.9227**	0.9213	0.9108	0.9218	0.9211	0.9202	0.9162
ERGAS↓	45.26	44.20	42.41	45.42	45.41	43.88	**40.44**
SAM↓	**8.68**	8.80	8.75	8.71	8.75	8.81	8.95

Table S-2. Ablation results on the stage transition timing.

t	100	200	300	400	500	600	700
PSNR↑	35.18	35.88	36.31	36.61	36.77	37.03	**37.19**
SAM↓	**8.18**	8.27	8.49	8.68	8.90	9.00	9.14

Table S-3. Ablation results on the relative learning rates.

q	0.4	0.5	0.6	0.7	0.8	0.9	1
PSNR↑	36.56	**36.61**	36.60	36.53	36.50	36.60	**36.61**
SAM↓	**8.62**	**8.62**	8.74	8.79	8.85	8.91	8.68

gate the impact of varying t from 100 to 700 epochs while maintaining a fixed fine-tuning epochs.

As shown in Tab. S-2, we observe an interesting trade-off between spatial quality (measured by PSNR) and spectral fidelity (measured by SAM). Earlier transitions (e.g., $t = 100$) yield better spectral preservation (SAM = 8.18) but at the cost of spatial quality (PSNR = 35.18 dB). Conversely, later transitions improve PSNR up to 37.19 dB at $t = 700$, but with increased spectral distortion (SAM = 9.14).

The balance in our experiments occurs at $t = 400$ epochs, achieving a PSNR of 36.61 dB while maintaining reasonable spectral quality (SAM = 8.68). This configuration allows sufficient time for the network to establish robust feature representations in Stage 1 before specializing for super-resolution in Stage 2.

S-2.3. On stage learning rates

It is well-acknowledged that the fine-tuning stage often requires a relatively smaller learning rate than that of the pre-training stage. Let learning rates (η_1 and η_2) in two stages be proportional, $\eta_2 = q\eta_1$, where q is the learning rate proportion between two stages. We evaluate q values ranging from 0.4 to 1.0 while keeping η_1 fixed at 10^{-3}. S-3 reveals several key insights:

- The method demonstrates remarkable stability across different q values, with PSNR variations less than 0.11 dB;
- Contrary to conventional wisdom about fine-tuning, we find equal learning rates ($q = 1$) work well.

The robustness to learning rate variations can be attributed to the PAN guidance that provides stable supervisory signals throughout training. This stability is particularly valuable in practical applications where hyperparameter tuning may be limited. The choice of $q = 1$ simplifies implementation while maintaining competitive performance across all metrics.

S-3. Inference time comparison

Tab. S-4 provides a comprehensive comparison of the inference times for various methods when tested against a challenging mixture noise scenario with a noise probability of $p = 0.55$. Our proposed method, UHipandas, exhibits competitive inference times, showcasing its efficiency in pro-

Table S-4. Inference time comparison for the compared methods.

Method	PWTV+SwinIR	SwinIR+PWTV	HIRD+SwinIR	PWTV+RPNN	RPNN+PWTV	HIRD+RPNN
Time	8	10	14	31	33	37
UHipandas (ours)	SwinIR+HIRD	RPNN+HIRD	PWTV+PLRD	PLRD+PWTV	HIRD+PLRD	PLRD+HIRD
46	69	92	518	520	524	579

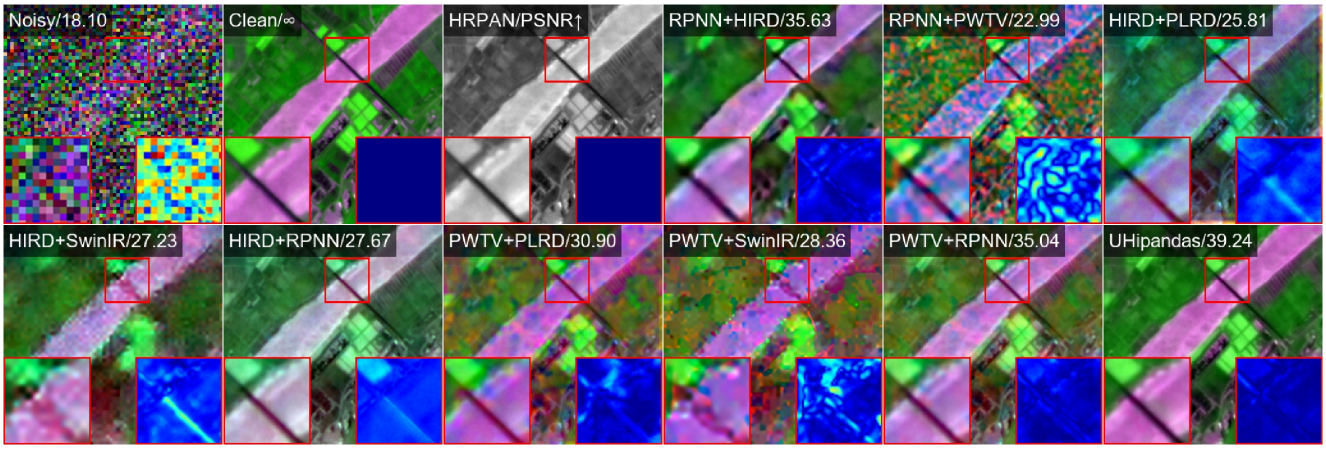

Figure S-2. Restoration results on the Dongying_5_6 image for non-i.i.d. Gaussian noise.

cessing hyperspectral images. Notably, UHipandas outperforms methods that incorporate PLRD, HIRD, and RPNN, which are characterized by their reliance on diffusion-based networks or necessitate individual fine-tuning for each hyperspectral image test case.

S-4. Visual comparison

Figs. S-1 to S-4 present additional visual comparison results that underscore the effectiveness of our method in noise removal and the restoration of intricate textures within hyperspectral images. The PSNR values displayed in the figures further validate the efficacy of UHipandas, thereby confirming its robustness and suitability for hyperspectral image enhancement tasks.

S-5. Details of experiments on the real-world dataset

In this section, we delve into the details of the experiments conducted on a real-world hyperspectral dataset. The dataset comprises the NLRHS image with dimensions $144 \times 144 \times 49$, and the corresponding HRPAN image of size 864×864. The spatial resolution ratio between the HRPAN and NLRHS images is $s = 6$. The visual results of the restoration process are presented in Fig. S-5. The figure showcases the effectiveness of our method in enhancing the spatial and spectral quality of the NLRHS image by fusing it with the HRPAN image.

S-6. Rank selection

The selection of appropriate ranks for the GDN and GSRN components within UHipandas is crucial for achieving satisfactory performance. Fig. S-6 provides a detailed illustration of the rank sensitivity for both components. The analysis reveals that while there is a slight degradation in performance as the rank increases, the overall impact is not large. The highest PSNR value of 37.43 dB is achieved with a GDN rank of 9 and a GSRN rank of 7, indicating an effective balance between denoising and super-resolution capabilities. Even at the worst-case scenario, with a PSNR value of 37.17 dB, the performance of UHipandas remains superior to that of the second-best method, PWTV+RPNN, which has a PSNR of 36.70 dB. This demonstrates the robustness of UHipandas to variations in rank selection and its consistent high performance across different configurations.

S-7. Interaction of pandenoising and pansharpening

The intricate relationship between pandenoising and pansharpening in the context of hyperspectral image processing is a critical aspect of our study. Fig. S-7 provides a visual representation of the denoised images, illustrating the distinct outcomes of each process. Here, Fig. S-7(b) provides a visual representation of the denoised images, illustrating the distinct outcomes of each process. That is, this image is produced by omitting the GSRN component from

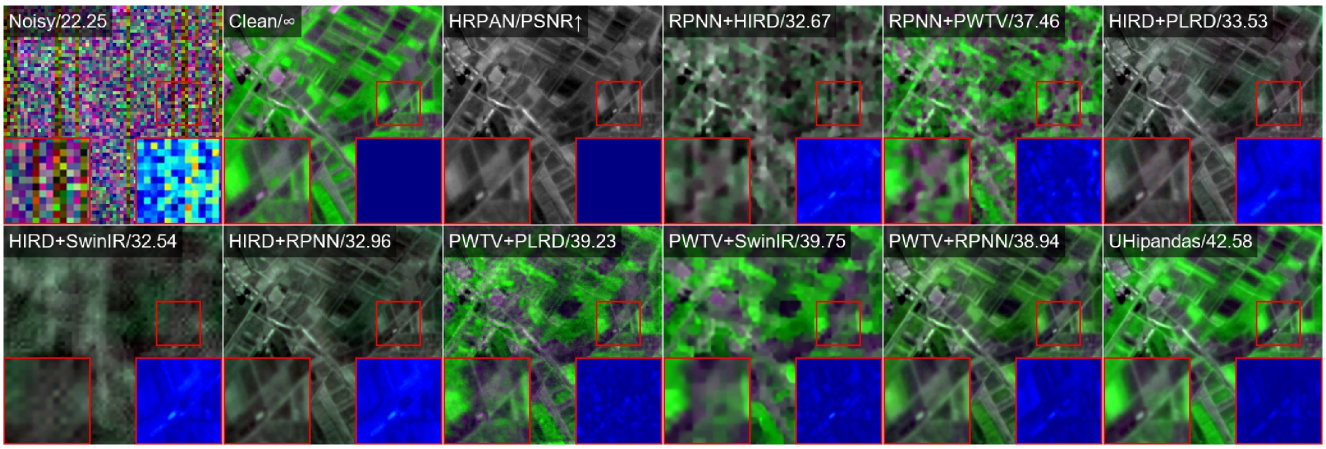

Figure S-3. Restoration results on the Dongying_0_8 image for mixture noise with $p = 0.15$.

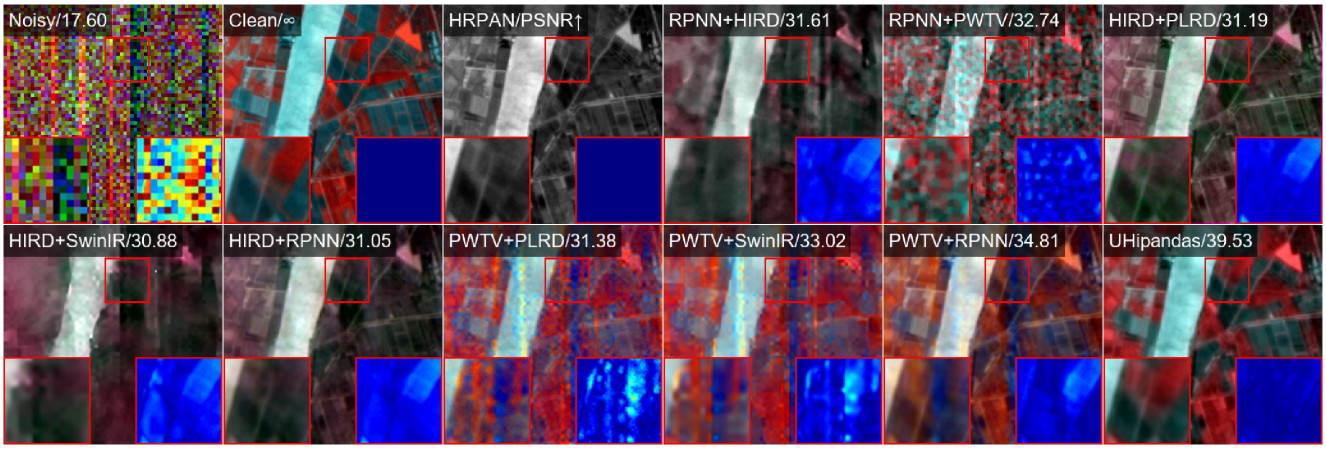

Figure S-4. Restoration results on the Dongying_2_8 image for mixture noise with $p = 0.35$.

the UHipandas method. In contrast, Fig. S-7(c) represents the output of the full Hipandas process, which integrates both pandenoising and pansharpening. The visual comparison clearly demonstrates that the inclusion of pansharpening in the Hipandas method not only enhances the pandenoising performance but also effectively mitigates spectral distortion, thereby preserving the integrity of the spectral information.

S-8. The intermediate results for UHipandas

To gain a deeper understanding of the UHipandas method and its underlying mechanisms, we present the intermediate results of the two-stage process in Fig. S-8. The initial stage, which involves pretraining, results in a denoised LRHS image that has successfully removed noise. However, this stage introduces a notable spectral/color distortion, which is a common challenge in hyperspectral image processing. The subsequent stage, as depicted in Fig. S-8, aims to address this issue. In stage 2, the GSRN component is applied, which not only refines the spatial resolution but also makes a significant contribution to the reduction of spectral distortion. This reduction can be interpreted as the recovery of lost spectral information, highlighting the GSRN's dual role in restoring both spatial and spectral details of the hyperspectral image. The intermediate results underscore the importance of the sequential and complementary nature of the stages within UHipandas, each playing a vital part in achieving the overall high-quality restoration of the hyperspectral image.

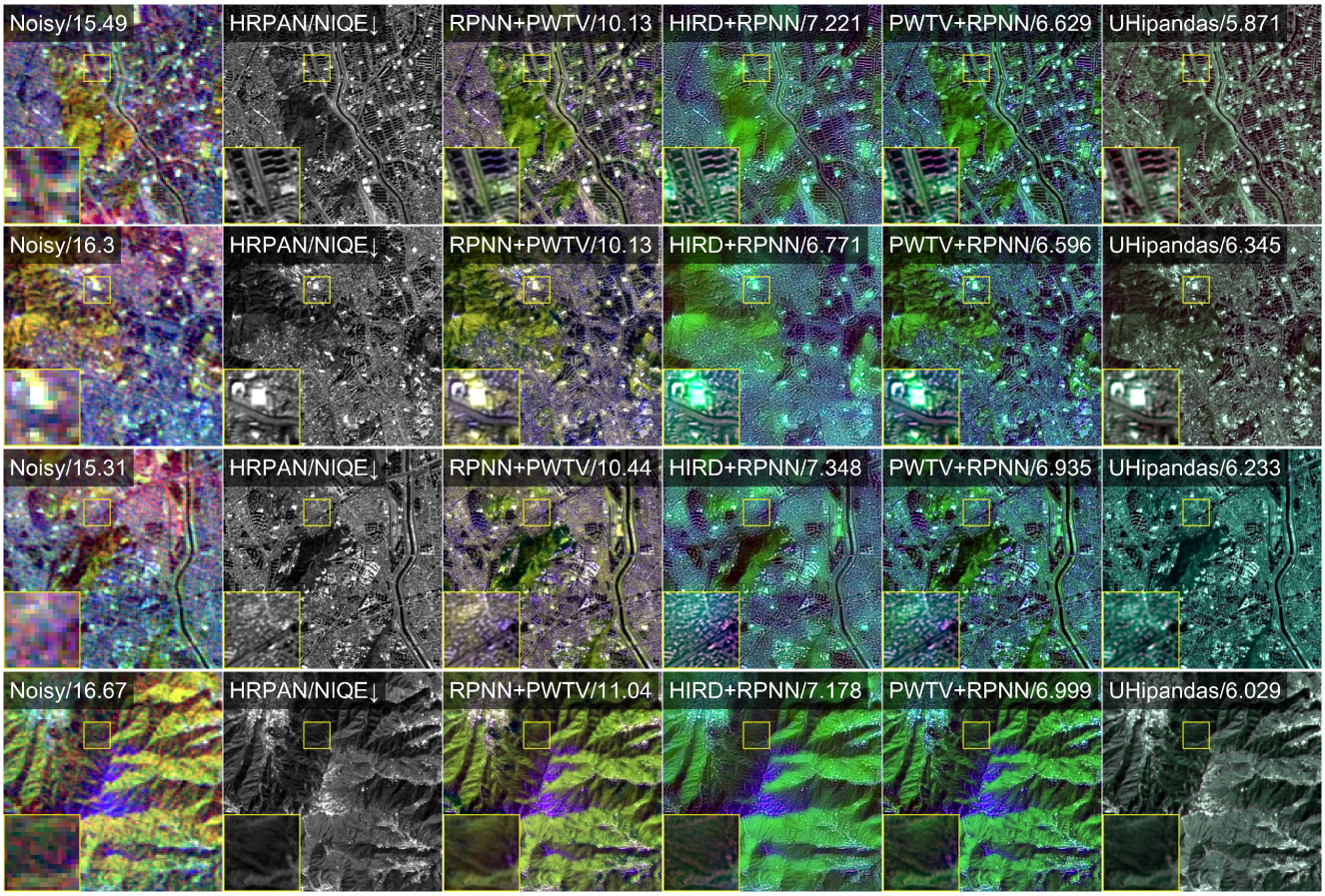

Figure S-5. Restoration results on the real-world image.

GDN \ GSRN	3	4	5	6	7	8	9	10	11	12	13	14	15	colormap
3	37.31	37.34	37.33	37.37	37.41	37.35	37.38	37.39	37.38	37.24	37.37	37.26	37.35	37.41
4	37.33	37.35	37.38	37.38	37.38	37.39	37.34	37.37	37.36	37.34	37.33	37.35	37.37	37.39
5	37.33	37.33	37.33	37.33	37.35	37.27	37.33	37.29	37.28	37.34	37.32	37.32	37.34	37.37
6	37.38	37.35	37.40	37.38	37.37	37.34	37.35	37.37	37.37	37.28	37.39	37.31	37.28	37.35
7	37.37	37.33	37.36	37.36	37.33	37.36	37.37	37.36	37.34	37.24	37.34	37.34	37.35	37.33
8	37.35	37.39	37.38	37.28	37.33	37.41	37.31	37.27	37.34	37.40	37.34	37.40	37.33	37.31
9	37.36	37.35	37.29	37.37	**37.43**	37.34	37.24	37.39	37.37	37.33	37.37	37.38	37.39	37.29
10	37.38	37.35	37.28	37.37	37.33	37.29	37.33	37.33	37.34	37.35	37.35	37.32	37.29	37.27
11	37.34	37.33	37.38	37.31	37.38	37.34	37.38	37.39	37.33	37.36	37.25	37.34	37.33	37.25
12	37.21	37.26	37.22	37.32	37.37	37.30	37.31	37.37	37.39	37.34	37.35	37.25	37.28	37.23
13	37.18	37.24	37.24	37.21	37.24	37.27	37.17	37.19	37.26	37.24	37.32	37.19	37.26	37.21
14	37.24	37.21	37.27	37.22	37.19	37.21	37.28	37.26	37.25	37.18	37.35	37.28	37.25	37.19
15	37.27	37.31	37.18	37.21	37.28	37.21	37.26	37.19	37.20	37.23	37.27	37.25	37.19	37.17

Figure S-6. PSNR values of UHipandas with different ranks for the non-i.i.d. Gaussian noise case.

(a) Clean LRHS (b) Only Pandenoising (PSNR: 40.46) (c) UHipandas (PSNR: 41.35)

Figure S-7. Visual comparison for the interaction of pandenoising and pansharpening

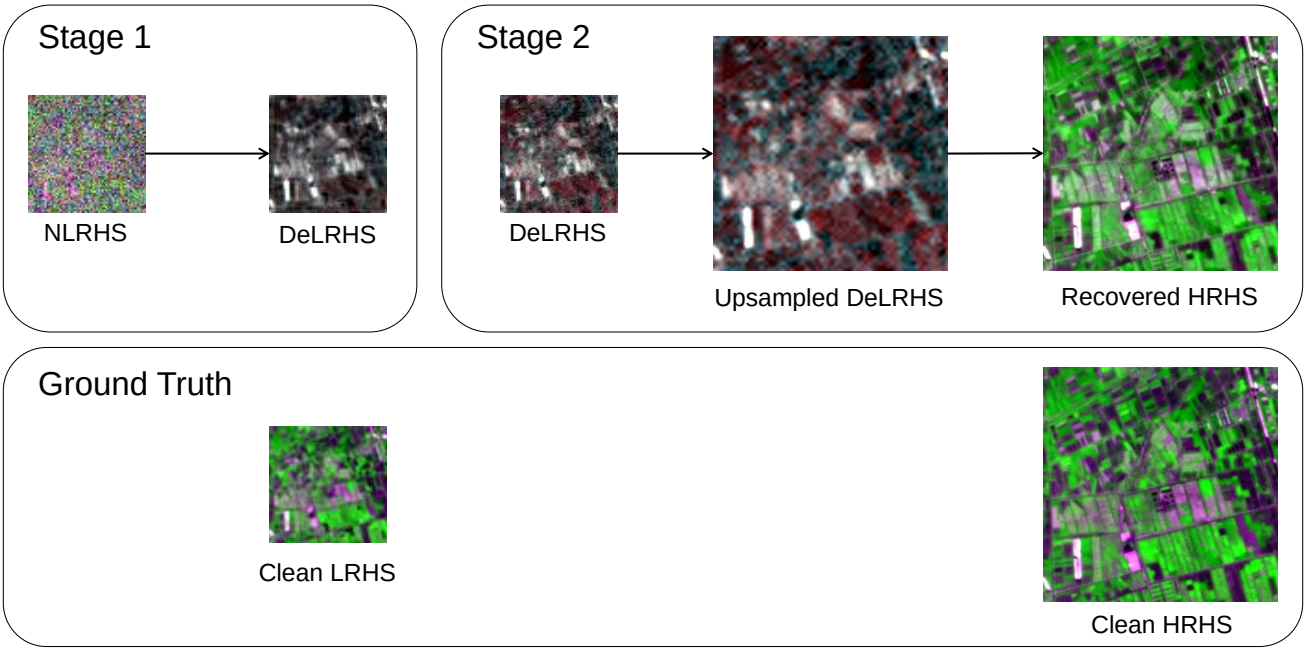

Figure S-8. The intermediate results for UHipandas.