"Learning An Explicit Weighting Scheme for Adapting Complex HSI Noise": Supplementary Material

Xiangyu Rui[1], Xiangyong Cao[1], Qi Xie[1], Zongsheng Yue[1], Qian Zhao[1], Deyu Meng[1,2*]
[1]Xi'an Jiaotong University; [2]Pazhou Lab, Guangzhou
xyrui@outlook.com, zsyzam@gmail.com
{caoxiangyong, xie.qi, timmy.zhaoqian, dymeng}@mail.xjtu.edu.cn

Abstract

In this supplementary material, we first present the implementation details on HW-NAILRMA, HW-NGmeet and HW-LLRT, and more comparison results. Then, we discuss loss objective functions of the proposed method for training network and demonstrate experiments for comparison. Finally, we provide details on the combination of the proposed HWnet and the existing denoising networks and present comparison results.

1. Implementation Details on Transfer Experiments

In this section, we introduce more detail on HW-NAILRMA, HW-NGmeet and HW-LLRT. Without additional annotations, '/' with respect to matrix or tensor means elementwise division.

1.1. On HW-NAILRMA

The original NAILRMA [8] solves the denoising problem:

$$\min_{X} ||Y - X||_F^2, \quad \text{s.t.} rank(X) \leq r, \tag{1}$$

where $Y \in \mathbb{R}^{hw \times b}$ denotes the observed noisy HSI. h, w represent height and width respectively and b is spectral number. The original NAILRMA performs two-stage updating in each iteration. Specifically, in the k-th iteration, the patchwise LRMA (PLRMA) denoising process is conducted on the input image u^{k-1}, and noise-adjusted iterative regularization is performed on u^{k-1} and PLRMA output f^k

$$f^k = \text{PLRMA}_r(u^{k-1}), \tag{2}$$

$$u_i^k = (1 - \delta_i) f_i^k + \delta_i u_i^{k-1}, \quad i = 1, 2, ..., b. \tag{3}$$

In the paper, δ_i is defined as

$$\delta_i = e^{-c\mathbf{W}(\mathbf{i},\mathbf{i})}, \quad i = 1, 2, ..., b, \tag{4}$$

where $\mathbf{W}(\mathbf{i}, \mathbf{i})$ is the noise variance estimated by multiple regression theory-based approach [4][1] on noisy HSI Y and fixed during iterations.

We design two schemes to replace noise-adjust parameter δ_i in (4). Remember that we estimate weights W^2 by pretrained HWnet C_θ as the mode of $\Gamma(\alpha, \beta)$ where α, β are outputs of C_θ. The first scheme is to directly replace $\mathbf{W}(\mathbf{i}, \mathbf{i})$ in (4) by $1/W$:

$$\tilde{\delta}_1 = e^{-c/W}. \tag{5}$$

Note that in (4), the noise variance $\mathbf{W}(\mathbf{i}, \mathbf{i})$ is estimated band by band while W is estimated on the pixel level and has the same size as the noisy HSI Y.

The second scheme is

$$\tilde{\delta}_2 = \frac{W^2}{W^2 + \mu}. \tag{6}$$

With this modified $\tilde{\delta}_2$, (3) is equal to the closed-form solution of the problem

$$\min_{u} ||W \odot (u - u^{k-1})||_F^2 + \mu ||u - f_i^k||_F^2. \tag{7}$$

Here we set $\mu = 0.5 * mean(W^2)$ as in the WLRMF problem.

Tab. 1 shows comparison results of the proposed two schemes (5) and (6) on ICVL dataset. Both proposed noise-adjust schemes achieve better performance than the original NAILRMA. It could be seen that the second scheme (6) performs relatively better than the first scheme (5). Through out all our experiments, we thus implemented the second scheme (6) in our HW-NAILRMA.

1.2. On HW-NGmeet

The original NGmeet denoising problem [6][7] aims to solve

$$\min_{A,\mathcal{M}} \frac{1}{2} ||\mathcal{Y} \times_3 A^T - \mathcal{M}||_F^2 + \lambda ||\mathcal{M}||_{\text{NL}}, \ \text{s.t.} A^T A = I, \tag{8}$$

*Corresponding author

Table 1: PSNR values of denoising results of HW-NAILRMA under two noise-adjust shemes on ICVL dataset. The best results are in **bold**.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
NAILRMA	32.47	31.11	30.64	29.79	26.25	27.89	24.71
scheme 1 (5)	35.29	33.31	34.12	32.25	28.05	30.17	27.27
scheme 2 (6)	**36.52**	**34.96**	**34.47**	**33.83**	**29.07**	**31.62**	**28.13**

where $\mathcal{Y} \in \mathbb{R}^{h\times w\times b}$ represents the observed noisy image. The original NGmeet applies half-quadratic splitting algorithm by introducing an auxiliary variable $\mathcal{Z}$ to relax (8) as

$$\begin{aligned}\min_{A,\mathcal{M},\mathcal{Z}} F(A,\mathcal{M},\mathcal{Z}) = \frac{1}{2}||\mathcal{Y}-\mathcal{Z}||_F^2 + \lambda||\mathcal{M}||_{\text{NL}} \\ + \frac{\mu}{2}||\mathcal{Z}-\mathcal{M}\times_3 A||_F^2, \quad \text{s.t.} A^TA = I.\end{aligned} \tag{9}$$

It is straightforward to modify (8) as its weighted vision HW-NGmeet as:

$$\begin{aligned}\min_{A,\mathcal{M}} \frac{1}{2}||\mathcal{W}\odot(\mathcal{Y}\times_3 A^T - \mathcal{M})||_F^2 + \lambda||\mathcal{M}||_{\text{NL}}, \\ \text{s.t.} A^TA = I,\end{aligned} \tag{10}$$

and the corresponding relaxed problem is

$$\begin{aligned}\min_{A,\mathcal{M},\mathcal{Z}} F^{HW}(A,\mathcal{M},\mathcal{Z}) = \frac{1}{2}||\mathcal{W}\odot(\mathcal{Y}-\mathcal{Z})||_F^2 \\ + \tilde{\lambda}||\mathcal{M}||_{\text{NL}} + \frac{\tilde{\mu}}{2}||\mathcal{Z}-\mathcal{M}\times_3 A||_F^2, \ \text{s.t.} A^TA = I.\end{aligned} \tag{11}$$

The alternative minimization strategy is readily used to solve (9) and (11). Note that in the k-th iteration, the $\mathcal{M}$-subproblem and A-subproblem are the same in (9) and (11). In the original NGmeet, by fixing $\mathcal{M}_{k-1}$ and A_{k-1}, the $\mathcal{Z}$-subproblem in (9) and its closed-form solution is

$$\begin{aligned}\mathcal{Z}_k &= \arg\min_{\mathcal{Z}} \frac{1}{2}||\mathcal{Y}-\mathcal{Z}||_F^2 + \frac{\mu}{2}||\mathcal{Z}-\mathcal{M}_{k-1}\times_3 A_{k-1}||_F^2 \\ &= \frac{\mathcal{Y}+\mu\mathcal{M}_{k-1}\times_3 A_{k-1}}{1+\mu}.\end{aligned} \tag{12}$$

In HW-NGmeet, the $\mathcal{Z}$-subproblem in (11) and its closed-form solution is

$$\begin{aligned}\tilde{\mathcal{Z}}_k &= \arg\min_{\mathcal{Z}} \frac{1}{2}||\mathcal{W}\odot(\mathcal{Y}-\mathcal{Z})||_F^2 + \frac{\tilde{\mu}}{2}||\mathcal{Z}-\tilde{\mathcal{M}}_{k-1}\times_3 \tilde{A}_{k-1}||_F^2 \\ &= \frac{\mathcal{W}^2\odot\mathcal{Y}+\tilde{\mu}\tilde{\mathcal{M}}_{k-1}\times_3 \tilde{A}_{k-1}}{\mathcal{W}^2+\tilde{\mu}}.\end{aligned} \tag{13}$$

As the original NGmeet set $\mu = 9$ in (12), here we set $\tilde{\mu} = 9 * mean(\mathcal{W}^2)$ in (13). The other parameters are similarly specified as in the original NGmeet. If all elements in $\mathcal{W}^2$ are 1, then HW-NGmeet problem (10) degrades to original NGmeet (8) and (13) degrades to (12).

1.3. On HW-LLRT

The original LLRT [3] proposes the following denoising problem:

$$\begin{aligned}\min_{\mathcal{X},\mathcal{L}_i} \frac{1}{2}||\mathcal{Y}-\mathcal{X}||_F^2 + \mu||\nabla_z\mathcal{X}||_p \\ + \omega\sum_i[\frac{1}{\lambda_i^2}||\mathcal{R}_i\mathcal{X}-\mathcal{L}_i||_F^2 + \text{rank}_2(\mathcal{L}_i)].\end{aligned} \tag{14}$$

The alternative minimization scheme is applied to solve (14). Fisrt fix the variable $\mathcal{X}$ and optimize $\mathcal{L}_i$ with its corresponding subproblem. Then LLRT fixes $\mathcal{L}_i$ and applies alternative direction multiplier method (ADMM) [2] to solve the $\mathcal{X}$-subproblem as

$$\min_{\mathcal{X}} \frac{1}{2}||\mathcal{Y}-\mathcal{X}||_F^2 + \mu||\nabla_z\mathcal{X}||_p + \omega\sum_i(\frac{1}{\lambda_i^2}||\mathcal{R}_i\mathcal{X}-\mathcal{L}_i||_F^2). \tag{15}$$

LLRT introduces two auxiliary variables $\mathcal{D}$ and $\mathcal{Z}$ and formulates the augmented Lagrangian function as:

$$\begin{aligned}L(\mathcal{X},\mathcal{D},\mathcal{Z}) = \frac{1}{2}||\mathcal{Y}-\mathcal{X}||_F^2 + \mu||\mathcal{D}||_p \\ + \omega\sum_i\frac{1}{\lambda_i}||\mathcal{R}_i\mathcal{Z}-\mathcal{L}_i||_F^2 + \frac{\beta}{2}||\mathcal{D}-\nabla_z\mathcal{X}-\frac{\mathcal{J}_1}{\beta}||_F^2 \\ + \frac{\alpha}{2}||\mathcal{Z}-\mathcal{X}-\frac{\mathcal{J}_2}{\alpha}||_F^2,\end{aligned} \tag{16}$$

where $\mathcal{J}_1$ and $\mathcal{J}_2$ are Lagrangian multipliers.

The proposed HW-LLRT replaces the term $||\mathcal{Y}-\mathcal{X}||_F^2$ in (14) with $||\mathcal{W}\odot(\mathcal{Y}-\mathcal{X})||_F^2$ and the denoising problem then becomes

$$\begin{aligned}\min_{\mathcal{X},\mathcal{L}_i} \frac{1}{2}||\mathcal{W}\odot(\mathcal{Y}-\mathcal{X})||_F^2 + \tilde{\mu}||\nabla_z\mathcal{X}||_p \\ + \tilde{\omega}\sum_i[\frac{1}{\lambda_i^2}||\mathcal{R}_i\mathcal{X}-\mathcal{L}_i||_F^2 + \text{rank}_2(\mathcal{L}_i)].\end{aligned} \tag{17}$$

Applying the same optimization strategy, the $\mathcal{L}_i$-subproblem is unaltered, while the $\mathcal{X}$-subproblem of HW-LLRT becomes

$$\begin{aligned}\min_{\mathcal{X}} \frac{1}{2}||\mathcal{W}\odot(\mathcal{Y}-\mathcal{X})||_F^2 + \tilde{\mu}||\nabla_z\mathcal{X}||_p \\ + \tilde{\omega}\sum_i(\frac{1}{\lambda_i^2}||\mathcal{R}_i\mathcal{X}-\mathcal{L}_i||_F^2).\end{aligned} \tag{18}$$

We also apply ADMM to solve (18), and an additional auxiliary variable $\mathcal{U}$ is introduced. The corresponding augmented Lagrangian function is

$$\begin{aligned}L^{HW}(\mathcal{X},\mathcal{D},\mathcal{Z},\mathcal{U}) = \frac{1}{2}||\mathcal{W}\odot(\mathcal{Y}-\mathcal{U})||_F^2 + \tilde{\mu}||\mathcal{D}||_p \\ + \tilde{\omega}\sum_i\frac{1}{\tilde{\lambda}_i}||\mathcal{R}_i\mathcal{Z}-\mathcal{L}_i||_F^2 + \frac{\tilde{\beta}}{2}||\mathcal{D}-\nabla_z\mathcal{X}-\frac{\tilde{\mathcal{J}}_1}{\tilde{\beta}}||_F^2 \\ + \frac{\tilde{\alpha}}{2}||\mathcal{Z}-\mathcal{X}-\frac{\tilde{\mathcal{J}}_2}{\tilde{\alpha}}||_F^2 + \frac{\tilde{\tau}}{2}||\mathcal{U}-\mathcal{X}-\frac{\tilde{\mathcal{J}}_3}{\tilde{\tau}}||_F^2.\end{aligned} \tag{19}$$

In the original LLRT, the optimization of $\min_{\mathcal{X},\mathcal{D},\mathcal{Z}} L(\mathcal{X},\mathcal{D},\mathcal{Z})$ in the k-th iteration consists of

$$\begin{aligned}\mathcal{X}_k &= \arg\min_{\mathcal{X}} \frac{1}{2}||\mathcal{Y}-\mathcal{X}||_F^2 \\ &+ \frac{\beta_{k-1}}{2}||\mathcal{D}_{k-1}-\nabla_z\mathcal{X}-\frac{(\mathcal{J}_1)_{k-1}}{\beta_{k-1}}||_F^2 \\ &+ \frac{\alpha_{k-1}}{2}||\mathcal{Z}_{k-1}-\mathcal{X}-\frac{(\mathcal{J}_2)_{k-1}}{\alpha_{k-1}}||_F^2, \end{aligned} \tag{20}$$

$$\begin{aligned}\mathcal{Z}_k &= \arg\min_{\mathcal{Z}} \omega\sum_i \frac{1}{\lambda_i}||\mathcal{R}_i\mathcal{Z}-\mathcal{L}_i||_F^2 \\ &+ \frac{\alpha_{k-1}}{2}||\mathcal{Z}-\mathcal{X}_k-\frac{(\mathcal{J}_2)_{k-1}}{\alpha_{k-1}}||_F^2, \end{aligned} \tag{21}$$

$$\mathcal{D}_k = \arg\min_{\mathcal{D}} \mu||\mathcal{D}||_p + \frac{\beta_{k-1}}{2}||\mathcal{D}-\nabla_z\mathcal{X}_k-\frac{(\mathcal{J}_1)_{k-1}}{\beta_{k-1}}||_F^2, \tag{22}$$

$$(\mathcal{J}_1)_k = (\mathcal{J}_1)_{k-1}+\beta_{k-1}(\nabla_z\mathcal{X}_k-\mathcal{D}_k), \tag{23}$$

$$(\mathcal{J}_2)_k = (\mathcal{J}_2)_{k-1}+\alpha_{k-1}(\mathcal{X}_k-\mathcal{Z}_k), \tag{24}$$

$$\alpha_k = \rho*\alpha_{k-1}, \tag{25}$$

$$\beta_k = \rho*\beta_{k-1}. \tag{26}$$

And in HW-LLRT, the optimization of $\min_{\mathcal{X},\mathcal{D},\mathcal{Z},\mathcal{U}} L^{HW}(\mathcal{X},\mathcal{D},\mathcal{Z},\mathcal{U})$ in the k-th iteration consists of

$$\begin{aligned}\tilde{\mathcal{X}}_k &= \arg\min_{\mathcal{X}} \frac{\tilde{\tau}_{k-1}}{2}||\tilde{\mathcal{U}}_{k-1}-\frac{(\tilde{\mathcal{J}}_3)_{k-1}}{\tilde{\tau}_{k-1}}-\mathcal{X}||_F^2 \\ &+ \frac{\tilde{\beta}_{k-1}}{2}||\tilde{\mathcal{D}}_{k-1}-\nabla_z\mathcal{X}-\frac{(\tilde{\mathcal{J}}_1)_{k-1}}{\tilde{\beta}_{k-1}}||_F^2 \\ &+ \frac{\tilde{\alpha}_{k-1}}{2}||\tilde{\mathcal{Z}}_{k-1}-\mathcal{X}-\frac{(\tilde{\mathcal{J}}_2)_{k-1}}{\tilde{\alpha}_{k-1}}||_F^2, \end{aligned} \tag{27}$$

$$\begin{aligned}\tilde{\mathcal{Z}}_k &= \arg\min_{\mathcal{Z}} \tilde{\omega}\sum_i \frac{1}{\tilde{\lambda}_i}||\mathcal{R}_i\mathcal{Z}-\mathcal{L}_i||_F^2 \\ &+ \frac{\tilde{\alpha}_{k-1}}{2}||\mathcal{Z}-\tilde{\mathcal{X}}_k-\frac{(\tilde{\mathcal{J}}_2)_{k-1}}{\tilde{\alpha}_{k-1}}||_F^2, \end{aligned} \tag{28}$$

$$\begin{aligned}\tilde{\mathcal{U}}_k &= \arg\min_{\mathcal{U}} \frac{1}{2}||\mathcal{W}\odot(\mathcal{Y}-\mathcal{U})||_F^2 \\ &+ \frac{\tilde{\tau}_{k-1}}{2}||\mathcal{U}-(\tilde{\mathcal{X}}_k+\frac{(\tilde{\mathcal{J}}_3)_{k-1}}{\tilde{\tau}_{k-1}})||_F^2, \end{aligned} \tag{29}$$

$$\tilde{\mathcal{D}}_k = \arg\min_{\mathcal{D}} \tilde{\mu}||\mathcal{D}||_p + \frac{\tilde{\beta}_{k-1}}{2}||\mathcal{D}-\nabla_z\tilde{\mathcal{X}}_k-\frac{(\tilde{\mathcal{J}}_1)_{k-1}}{\tilde{\beta}_{k-1}}||_F^2, \tag{30}$$

$$(\tilde{\mathcal{J}}_1)_k = (\tilde{\mathcal{J}}_1)_{k-1}+\tilde{\beta}_{k-1}(\nabla_z\tilde{\mathcal{X}}_k-\tilde{\mathcal{D}}_k), \tag{31}$$

$$(\tilde{\mathcal{J}}_2)_k = (\tilde{\mathcal{J}}_2)_{k-1}+\tilde{\alpha}_{k-1}(\tilde{\mathcal{X}}_k-\tilde{\mathcal{Z}}_k), \tag{32}$$

$$(\tilde{\mathcal{J}}_3)_k = (\tilde{\mathcal{J}}_3)_{k-1}+\tilde{\tau}_{k-1}(\tilde{\mathcal{X}}_k-\tilde{\mathcal{U}}_k), \tag{33}$$

$$\tilde{\alpha}_k = \rho*\tilde{\alpha}_{k-1}, \tag{34}$$

$$\tilde{\beta}_k = \rho*\tilde{\beta}_{k-1}, \tag{35}$$

$$\tilde{\tau}_k = \rho*\tilde{\tau}_{k-1}. \tag{36}$$

Updating of $\tilde{\mathcal{Z}}_k$ and $\tilde{\mathcal{D}}_k$ in (28) and (30) are unaltered. Updating of $\tilde{\mathcal{X}}_k$ in (27) has the same form as that in (20). Updating of $\tilde{\mathcal{U}}_k$ in (29) has the closed-form solution:

$$\tilde{\mathcal{U}}_k = \frac{\mathcal{W}^2\odot\mathcal{Y}+\tilde{\tau}_{k-1}\tilde{\mathcal{X}}_k+(\tilde{\mathcal{J}}_3)_{k-1}}{\mathcal{W}^2+\tilde{\tau}_{k-1}}. \tag{37}$$

We set parameter $\tilde{\tau}_0 = 0.5*mean(\mathcal{W}^2)$ through out all the experiments. It is the only additional parameter introduced by HW-LLRT. The other parameters are set as the default settings in the original LLRT. If we set all elements in $\mathcal{W}$ in HW-LLRT as 1, the HW-LLRT problem (17) reduces to LLRT problem (14), but the optimizations are slightly different. Thus, we conduct ablation experiments for HW-LLRT by setting all elements in $\mathcal{W}$ as 1.

2. Compared with LRTV

As the reviewer suggested, we also transfer the proposed weighting scheme to LRTV [9], a widely used hyperspectral mixed noise removal method. The optimization problem of LRTV is formulated as follows:

$$\begin{aligned}&\min_{L,X,S}||L||_* + \tau||X||_{\mathrm{HTV}} + \lambda||S||_1, \\ &\text{s.t.}||Y-L-S||_F^2 \le \varepsilon, \mathrm{rank}(L)\le r, L=X,\end{aligned} \tag{38}$$

where $L \in \mathbb{R}^{hw\times b}$ represents the restored HSI. (38) is solved by ALM method, and the corresponding augmented Lagrangian funtion is

$$\begin{aligned}l(L,X,S,\Gamma_1,\Gamma_2) &= ||L||_* + \tau||X||_{\mathrm{HTV}} + \lambda||S||_1 \\ &+\langle\Gamma_2, Y-L-S\rangle + \langle\Gamma_1, X-L\rangle \\ &+\frac{\mu_2}{2}||Y-L-S||_F^2 + \frac{\mu_1}{2}||X-L||_F^2 \\ &\text{s.t.}\mathrm{rank}(L)\le r.\end{aligned} \tag{39}$$

Note that the original LRTV model (38) is a kind of low-rank and sparse decomposition and uses robust L_1 loss to constrain the sparse noise. We fomulate the HW-LRTV model as

$$\begin{aligned}&\min_{L,X}||L||_* + \tau||X||_{\mathrm{HTV}} + \kappa||W\odot(Y-Z)||_F^2, \\ &\text{s.t.}L=Z, X=Z, \mathrm{rank}(L)\le r.\end{aligned} \tag{40}$$

Applying ALM method to solve (40), the augmented Lagrangian function is

$$\begin{aligned}l(Z,L,X) &= ||L||_* + \tau||X||_{\mathrm{HTV}} + \kappa||W\odot(Y-Z)||_F^2 \\ &+\langle\Gamma_1, L-Z\rangle + \langle\Gamma_2, X-Z\rangle \\ &+\frac{\mu_1}{2}||L-Z||_F^2 + \frac{\mu_2}{2}||X-Z||_F^2.\end{aligned} \tag{41}$$

In original LLRT, the update scheme in the k-th iteration

is

$$L_k = \arg\min ||L||_* + \frac{(\mu_1)_{k-1} + (\mu_2)_{k-1}}{2} ||L - \frac{\alpha_{k-1}}{(\mu_1)_{k-1} + (\mu_2)_{k-1}}||_F^2, \quad (42)$$
$$\alpha_{k-1} = (\mu_1)_{k-1} X_k + (\mu_2)_{k-1}(Y - S_k) + (\Gamma_1)_{k-1} + (\Gamma_2)_{k-1},$$

$$X_k = \arg\min \tau ||X||_{\text{HTV}} + \frac{(\mu_1)_{k-1}}{2} ||X - (L_k - \frac{(\Gamma_1)_{k-1}}{(\mu_1)_{k-1}})||_F^2 \quad (43)$$

$$S_k = \arg\min \lambda ||S||_1 + \frac{(\mu_2)_{k-1}}{2} ||S - (Y - L_k + \frac{(\Gamma_2)_{k-1}}{(\mu_2)_{k-1}})||_F^2, \quad (44)$$

$$(\Gamma_1)_k = (\Gamma_1)_{k-1} + \mu_1(X_k - L_k), \quad (45)$$
$$(\Gamma_2)_k = (\Gamma_2)_{k-1} + \mu_2(Y - L_k - S_k), \quad (46)$$
$$(\mu_1)_k = \rho * (\mu_1)_{k-1}, \quad (47)$$
$$(\mu_1)_k = \rho * (\mu_1)_{k-1}. \quad (48)$$

The update scheme in the k-th iteration to solve HW-LRTV problem is

$$L_k = \arg\min ||L||_* + \frac{(\mu_1)_{k-1}}{2} ||L - (Z_{k-1} - \frac{(\Gamma_1)_{k-1}}{(\mu_1)_{k-1}})||, \quad (49)$$

$$X_k = \arg\min \tau ||X||_{\text{HTV}} + \frac{(\mu_2)_{k-1}}{2} ||X - (Z_{k-1} - \frac{(\Gamma_2)_{k-1}}{(\mu_2)_{k-1}})||_F^2 \quad (50)$$

$$Z_k = \arg\min \kappa ||W \odot (Y - Z)||_F^2 + \frac{(\mu_1)_{k-1}}{2} ||L_k - Z + \frac{(\Gamma_1)_{k-1}}{(\mu_1)_{k-1}}||_F^2 + \frac{(\mu_2)_{k-1}}{2} ||X_k - Z + \frac{(\Gamma_2)_{k-1}}{(\mu_2)_{k-1}}||_F^2, \quad (51)$$

$$(\Gamma_1)_k = (\Gamma_1)_{k-1} + \mu_1(L_k - Z_k), \quad (52)$$
$$(\Gamma_2)_k = (\Gamma_2)_{k-1} + \mu_2(X_k - Z_k), \quad (53)$$
$$(\mu_1)_k = \rho * (\mu_1)_{k-1}, \quad (54)$$
$$(\mu_1)_k = \rho * (\mu_1)_{k-1}. \quad (55)$$

The parameter λ in original LRTV is set as $35/\sqrt{hw}$ and the other parameters are set as the default setting. In HW-LRTV optimization problem, we set parameter $\kappa = 11 * [(\mu_1)_0 + (\mu_2)_0]/mean(W^2)$ and τ =1e-3 through out all the experiments.

Table 2: Quantitative comparison of LRTV and HW-LRTV on ICVL dataset. The best results are in **bold**

Case	LRTV		HW-LRTV	
	PSNR	SSIM	PSNR	SSIM
1	34.10	0.9357	**35.09**	**0.9509**
2	32.32	0.8780	**34.48**	**0.9397**
3	38.13	0.9578	**38.92**	**0.9769**
4	32.71	0.9284	**32.81**	**0.9497**
5	**35.26**	0.9240	32.32	**0.9251**
6	30.36	0.9082	**31.37**	**0.9490**
7	**30.85**	0.8979	30.73	**0.9333**

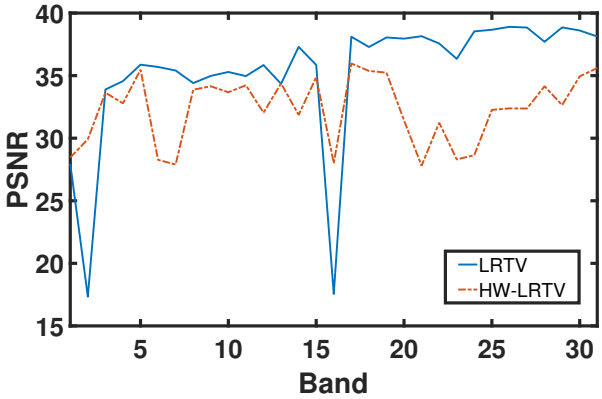


Figure 1: PSNR of each band of restored HSIs by LRTV and HW-LRTV for 'Gaussian + impulse' noise case.

Tab. 2 shows the denoising results of LRTV and HW-LRTV on ICVL dataset. For most cases, HW-LRTV achieves best results while LRTV is very suitable for removing 'Gaussian + impulse' noise. From Fig. 1 we could see that the average PSNR value of LRTV is better than HW-LRTV, however, extreme noise on some bands fails to be removed. For all cases, HW-LRTV better preserves structural similarity.

3. More Experimental Results

In this section, we present more experimental results on CAVE, PaviaU and HYDICE Urban datasets. From Tab. 3 and Tab. 4 one can see that the proposed HW-LRMF achieves the best results under most cases. For cases 2,4 and 6, the SSIM values of HW-LRMF are very competitive to the best results. And for experiments on PaviaU, the proposed HW-LRMF is superior to all competitive methods under all cases with significant improvements. Fig. 3 shows visual denoising results of HW-LRMF and compared methods. It can be seen that the proposed method successfully retrieves more details and achieves the best visual results

among all competing methods.

Tab. 5 and Tab. 6 show quantitative comparison results on NAILRMA, NGmeet, LLRT and the corresponding weighted visions. The proposed HW-NAILRMA, HW-NGmeet and HW-LLRT are evidently superior to original methods under complex noise cases. Fig. 4 and Fig. 5 show the visual denoising results. From Fig. 4 we can see, for severe Gaussian and impulse noises, the proposed weigthed methods produce evidently higher image quality while the original methods can not efficiently remove impulse noises. Fig. 6 shows the denoising results of real noisy HYDICE Urban HSI. Since the groundtruth image is missing, noise level is estimated in band-by-band manner by multiple regression theory-based approach [1] as suggested in [6]. We plot the mean value of square-root of mode $1/W^2$, *i.e.* $\sqrt{\beta/(\alpha+1)}$, of each band and the estimated noise level by [1] in Fig. 2. It is easy to see that most bands of Urban are relatively clean while a few bands are severely damaged. Fig. 2 shows the ability of our HWnet for revealing noise insights.

4. More details on Loss Objective of The Proposed Method

We train the parameters θ in HWnet C_θ by optimizing over objective function:

$$\min_{\theta} KL[q(X, W^2|Y)||p(X, W^2||Y)]. \tag{56}$$

As we have aforementioned, this objective function contains three parts. The third part L_3 is equivalent to the mean square error(MSE) loss between the algorithm output $X_{(N)}$ and groundtruth HSI X_{gt}. MSE loss can also be straightforwardly used to train C_θ

$$\min_{\theta} ||X_{(N)} - X_{gt}||_F^2. \tag{57}$$

To compare the above two loss objectives, we conduct an additional experiment of training C_θ by minimizing MSE loss (57). The training settings are similar except that instead of outputing α, β, the HWnet C_θ directly outputs weight W^2.

Tab. 7 shows the quantitative comparison results of HW-LRMF, HW-NAILRMA, HW-NGmeet and HW-LLRT with pre-trained C_θ that is trained under the above two loss objectives, respectively. It is observed that for i.i.d Gaussian noise, results of MSE loss are slightly better than KL divergence loss for all comparing methods. For HW-NGmeet, denoising result of MSE loss is significantly better than that of KL divergence loss under noise case 7, whereas on the contrary is under noise case 2 and 3. In summary, the overall performance of KL divergence loss is better and more stable than MSE loss.

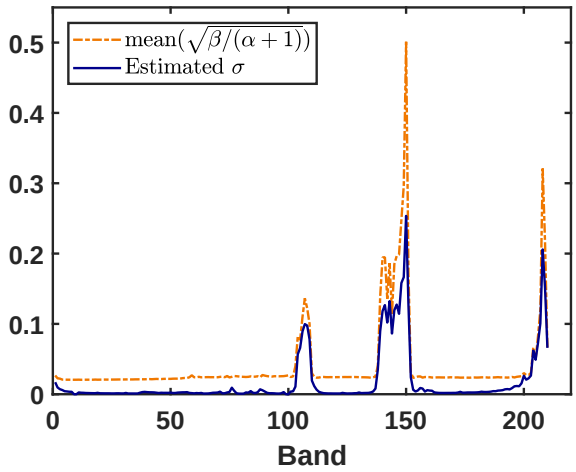


Figure 2: Comparison between the calculated mean value of $\sqrt{\beta/(\alpha+1)}$ of each band by the proposed method and the estimated noise level by multiple regression theory-based approach [1] on HYDICE Urban HSI.

5. Combination of HWnet and Blind Denoising Networks

Like FFDnet [10] and CBDnet [5], we concatenate the estimated $1/W$ with noisy image as input to blind denoising networks. HSI-DeNet and DSSnet are set as the baseline networks. We denote the corresponding modified networks as HSI-DeNet+ and DSSnet+. DSSnet is a 3D convolutional neural network and we simply double the input channel to construct DSSnet+. Besides, we apply parameter sharing strategy for both networks. HSI-DeNet is a 2D convolutional neural network. Since each band of $1/W$ is mainly related to the corresponding band of noisy image, we add an additional 3D conv layer before the first 2D conv layer of the HSI-DeNet. Specifically, the 3D conv layer takes $1/W$ and noisy image as input, the output feature map is then concatenated again with noisy image as input to HSI-DeNet with double input channel number. We follow the training settings similarly in their original paper. Cases 2,3 and 4 types of noise are added to clean patches to create paired training patches.

Tab. 8 shows the comparison results on ICVL dataset. It can be easily seen that with additional input $1/W$, the modified networks outperforms the original networks both for noise cases that appear in training dataset and that do not appear, which reveals the ability of extracting noise information of our HWnet.

References

[1] José M Bioucas-Dias and José MP Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008. 1, 5

Table 3: Quantitative comparison of HW-LRMF and other denoising methods on CAVE dataset. The best results are in **bold**.

Case	Index	Methods							
		Noisy	SVD	LRMR	LRTA	PARAFAC	NMoG	TDL	HW-LRMF
1	PSNR	18.59	27.30	29.65	30.68	24.17	27.53	**34.58**	30.24
	SSIM	0.5909	0.8410	0.9103	0.8890	0.6615	0.8774	**0.9644**	0.8967
2	PSNR	17.07	25.19	26.59	27.41	28.88	29.45	27.23	**30.77**
	SSIM	0.3739	0.7207	0.7454	0.7238	0.8113	**0.8720**	0.7847	0.8688
3	PSNR	17.27	25.74	24.92	17.62	30.58	28.85	19.49	**31.90**
	SSIM	0.5329	0.7622	0.7977	0.5404	0.9202	0.8915	0.6132	**0.9295**
4	PSNR	16.14	26.49	28.55	27.01	29.68	29.87	27.06	**30.96**
	SSIM	0.2407	0.7002	0.6755	0.6047	0.7179	**0.8146**	0.6750	0.7669
5	PSNR	13.18	19.71	25.59	23.64	23.45	24.29	23.24	**26.23**
	SSIM	0.2545	0.5403	0.7640	0.5986	0.6674	0.7729	0.6407	**0.8124**
6	PSNR	17.87	24.97	29.67	26.17	31.91	29.93	25.47	**33.09**
	SSIM	0.2774	0.5891	0.7356	0.6163	**0.8749**	0.8470	0.6077	0.8745
7	PSNR	14.76	20.40	24.37	20.97	23.43	15.56	20.15	**31.02**
	SSIM	0.2455	0.4495	0.6310	0.4402	0.5889	0.3490	0.4476	**0.8236**

Table 4: Quantitative comparison of HW-LRMF and other denoising methods on PaviaU. The best results are in **bold**.

Case	Index	Methods							
		Noisy	SVD	LRMR	LRTA	PARAFAC	NMoG	TDL	HW-LRMF
1	PSNR	18.59	32.86	32.64	31.35	30.39	33.41	34.96	**35.38**
	SSIM	0.2784	0.8593	0.8816	0.8481	0.8382	0.8902	0.9368	**0.9390**
2	PSNR	17.06	30.07	29.31	23.73	29.25	33.89	24.71	**34.75**
	SSIM	0.2453	0.7923	0.7768	0.4931	0.7858	0.8978	0.5563	**0.9204**
3	PSNR	17.33	30.85	29.22	17.47	29.99	34.54	18.31	**36.87**
	SSIM	0.3446	0.8017	0.8023	0.3476	0.8200	0.9170	0.3988	**0.9539**
4	PSNR	16.10	29.51	30.05	22.64	27.39	33.37	23.97	**33.46**
	SSIM	0.2178	0.7726	0.8004	0.4415	0.7108	0.8884	0.5155	**0.9101**
5	PSNR	13.94	24.69	29.64	21.33	24.39	30.23	21.63	**31.75**
	SSIM	0.1653	0.6079	0.7985	0.3899	0.6139	0.8226	0.4120	**0.8843**
6	PSNR	16.52	29.11	29.53	22.02	26.60	32.26	22.73	**32.59**
	SSIM	0.2280	0.7827	0.8021	0.4164	0.6823	0.8897	0.4653	**0.9035**
7	PSNR	14.00	24.41	26.23	18.91	22.17	29.01	19.24	**32.61**
	SSIM	0.1915	0.6146	0.7078	0.2787	0.4975	0.8161	0.3264	**0.8984**

[2] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011. 2

[3] Yi Chang, Luxin Yan, and Sheng Zhong. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4260–4268, 2017. 2

[4] Lianru Gao, Qian Du, Bing Zhang, Wei Yang, and Yuanfeng Wu. A comparative study on linear regression-based noise estimation for hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):488–498, 2013. 1

[5] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1712–1722, 2019. 5

[6] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, and Qibin Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6861–6870. IEEE, 2019. 1, 5

[7] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, Qibin Zhao, Hongyan Zhang, and Liangpei Zhang. Non-local meets global: An integrated paradigm for hyperspectral image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 1

[8] Wei He, Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen. Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):3050–3061, 2015. 1

Table 5: Quantitative comparison of transfer experiments on CAVE dataset. The best results are in **bold**.

Case	Index	NAILRMA	HW-NAILRMA	NGmeet	HW-NGmeet	LLRT	HW-LLRT	ablation
1	PSNR	30.99	**33.98**	**37.66**	37.32	36.55	**36.70**	36.55
	SSIM	0.9352	**0.9533**	**0.9822**	0.9793	0.9789	**0.9793**	0.9789
2	PSNR	30.02	**32.06**	30.69	**32.32**	33.53	**33.66**	33.53
	SSIM	0.8795	**0.9200**	0.8958	**0.9077**	0.9310	**0.9326**	0.9310
3	PSNR	27.67	**31.68**	29.29	**36.96**	36.93	**37.81**	36.93
	SSIM	0.8679	**0.9397**	0.9049	**0.9755**	0.9844	**0.9851**	0.9844
4	PSNR	32.67	**35.62**	33.00	**37.82**	38.89	**39.14**	38.89
	SSIM	0.7813	**0.8999**	0.8984	**0.9109**	**0.9621**	0.9619	0.9621
5	PSNR	23.99	**27.05**	28.03	**28.63**	28.48	**28.55**	28.48
	SSIM	0.7235	**0.8060**	0.8237	**0.8558**	0.8406	**0.8416**	0.8406
6	PSNR	32.28	**36.18**	30.02	**37.33**	36.08	**36.21**	36.08
	SSIM	0.8299	**0.9353**	0.8550	**0.9282**	**0.9663**	0.9650	0.9663
7	PSNR	24.71	**27.64**	27.31	**29.38**	26.00	26.45	26.00
	SSIM	0.6798	**0.7299**	0.7650	**0.7965**	0.7027	0.7136	0.7026

Table 6: Quantitative comparison of transfer experiments on PaviaU. The best results are in **bold**.

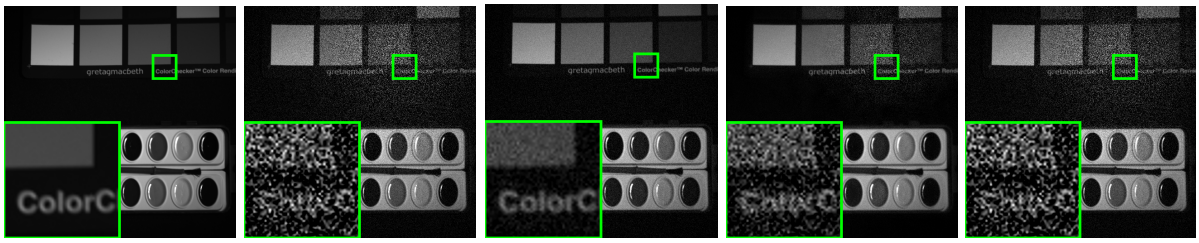
Case	Index	NAILRMA	HW-NAILRMA	NGmeet	HW-NGmeet	LLRT	HW-LLRT	ablation
1	PSNR	33.95	**35.87**	**37.84**	37.65	36.37	**36.55**	36.37
	SSIM	0.8990	**0.9466**	**0.9659**	0.9641	0.9530	**0.9537**	0.9530
2	PSNR	32.57	**34.74**	30.53	**33.59**	33.95	**34.85**	33.95
	SSIM	0.8803	**0.9342**	0.8093	**0.9203**	0.9262	**0.9326**	0.9262
3	PSNR	31.62	**35.65**	32.56	**35.32**	32.43	**34.44**	32.43
	SSIM	0.8654	**0.9465**	0.8715	**0.9424**	0.9027	**0.9252**	0.9027
4	PSNR	31.43	**33.15**	30.21	**33.49**	33.29	**34.08**	33.29
	SSIM	0.8662	**0.9213**	0.7919	**0.9234**	0.9157	**0.9222**	0.9157
5	PSNR	26.03	**27.95**	27.31	**27.77**	26.43	**26.89**	26.43
	SSIM	0.7158	**0.8057**	0.7666	**0.7951**	0.7800	**0.7903**	0.7800
6	PSNR	30.60	**32.19**	29.34	**32.23**	32.03	**32.82**	32.03
	SSIM	0.8687	**0.9188**	0.7884	**0.9105**	0.9001	**0.9077**	0.9001
7	PSNR	24.99	**27.52**	26.20	**27.72**	26.21	**27.00**	26.21
	SSIM	0.6953	**0.8056**	0.7570	**0.7780**	0.7717	**0.7919**	0.7717

[9] Wei He, Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE transactions on geoscience and remote sensing*, 54(1):178–188, 2015. 3

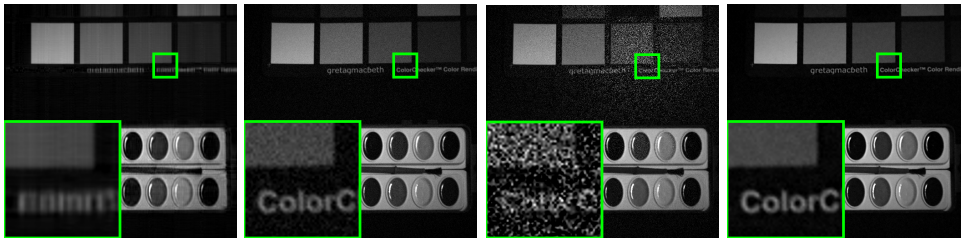
[10] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 5

Table 7: Quantitative comparison of MSE loss (57) and KL divergence loss (56). The best results are in **bold**.

Case	Index	HW-LRMF		HW-NAILRMA		HW-NGmeet		HW-LLRT	
		MSE	KL	MSE	KL	MSE	KL	MSE	KL
1	PSNR	**35.04**	34.93	**36.63**	36.52	**40.02**	39.98	**39.55**	39.48
	SSIM	0.9407	**0.9431**	0.9678	**0.9680**	0.9797	**0.9798**	**0.9771**	0.9769
2	PSNR	29.48	**32.52**	33.68	**34.96**	34.29	**37.18**	38.66	**38.82**
	SSIM	0.8094	**0.8946**	0.8618	**0.9287**	0.9198	**0.9424**	0.9620	**0.9626**
3	PSNR	30.60	**33.67**	34.24	**34.47**	34.77	**38.35**	38.42	**39.24**
	SSIM	0.8850	**0.9295**	0.9277	**0.9508**	0.9435	**0.9708**	0.9700	**0.9734**
4	PSNR	30.34	**31.80**	32.93	**33.83**	34.56	**34.95**	35.79	**35.92**
	SSIM	0.9060	**0.9258**	0.9463	**0.9555**	0.9556	**0.9571**	0.9601	**0.9612**
5	PSNR	30.71	**31.59**	**29.42**	29.07	30.49	**30.81**	30.82	**30.84**
	SSIM	0.8830	**0.8960**	**0.8724**	0.8578	0.8688	**0.9000**	0.8812	**0.8814**
6	PSNR	29.37	**31.26**	30.21	**31.62**	31.43	**31.73**	29.34	**29.56**
	SSIM	0.8984	**0.9324**	0.9304	**0.9513**	0.9272	**0.9284**	0.8879	**0.8899**
7	PSNR	27.77	**28.34**	**28.33**	28.13	**28.86**	27.57	26.63	**26.89**
	SSIM	0.8687	**0.8838**	**0.8957**	0.8913	**0.8785**	0.8630	0.8110	**0.8206**



(a) Original (PSNR,SSIM) (b) Noisy (18.97,0.6166) (c) SVD (26.84,0.8210) (d) LRMR (24.98, 0.8053) (e) LRTA (19.15,0.6201)

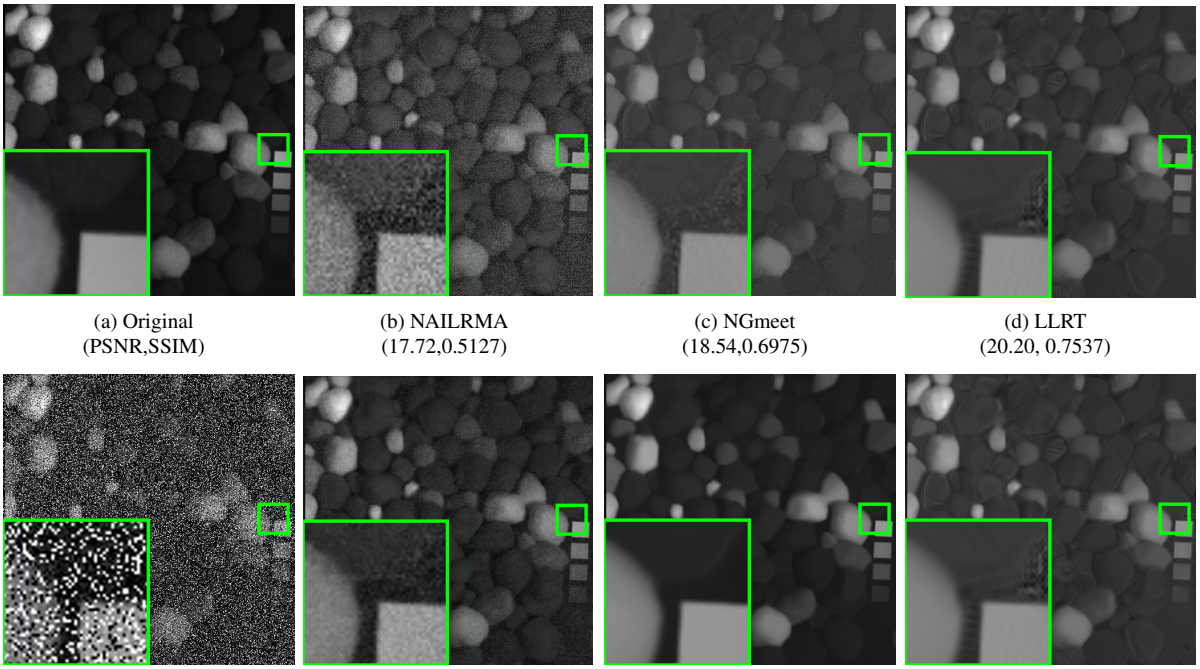


(f) PARAFAC (30.31,**0.9261**) (g) NMoG (29.77,0.8966) (h) TDL (20.49, 0.6664) (i) HW-LRMF (**33.20**,0.9152)

Figure 3: Visual comparison results at the 15^{th} band of image *paints* in CAVE dataset. The noisy HSI is corrupted by spatial-spectral variant Gaussian noise.

Table 8: PSNR values of denoising results of supervised denoising networks on ICVL dataset. '+' means the modified network which takes both noisy image and $1/W$ as input.

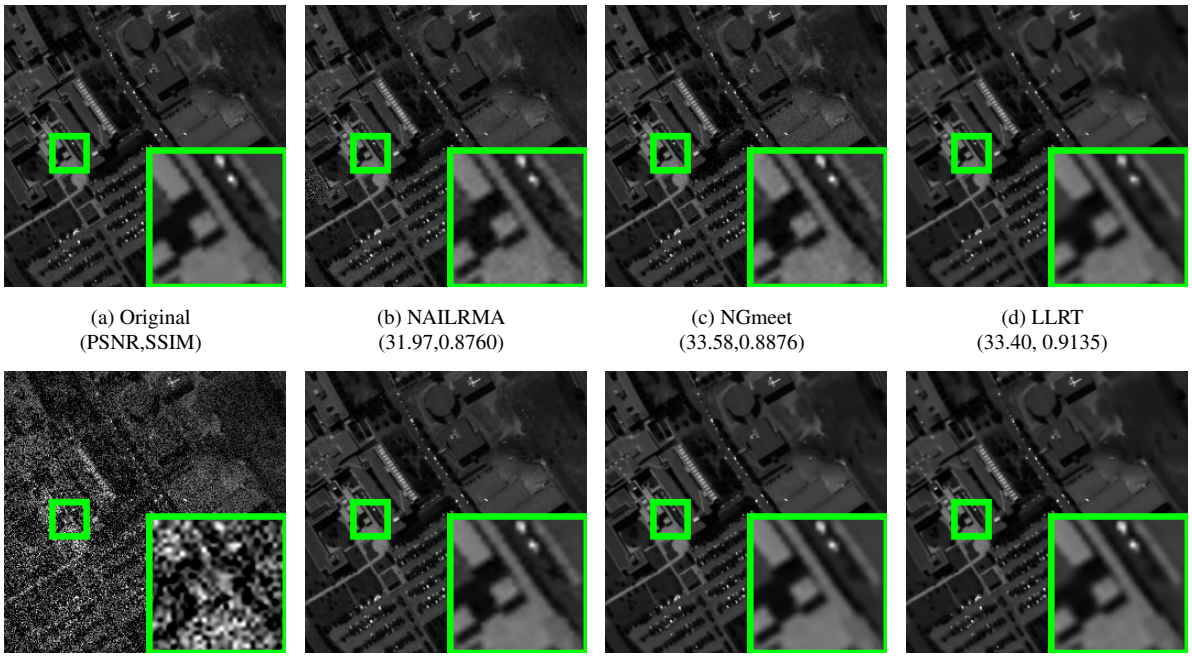
Case	HSI-DeNet	HSI-DeNet+	DSSnet	DSSnet+
1	36.06	**36.42**	40.51	**40.59**
2	34.40	**34.93**	40.55	**40.60**
3	35.52	**35.84**	42.28	**42.32**
4	33.89	**34.38**	40.43	**40.46**
5	26.60	**27.14**	32.56	**32.88**
6	31.44	**32.24**	39.47	**39.71**
7	27.00	**27.17**	32.88	**33.22**



(a) Original (PSNR,SSIM) | (b) NAILRMA (17.72,0.5127) | (c) NGmeet (18.54,0.6975) | (d) LLRT (20.20, 0.7537)

(e) Noisy (9.0284,0.0710) | (f) HW-NAILRMA (**24.99**,**0.8057**) | (g) HW-NGmeet (**29.66**,**0.9286**) | (h) HW-LLRT (**20.36**,**0.7574**)

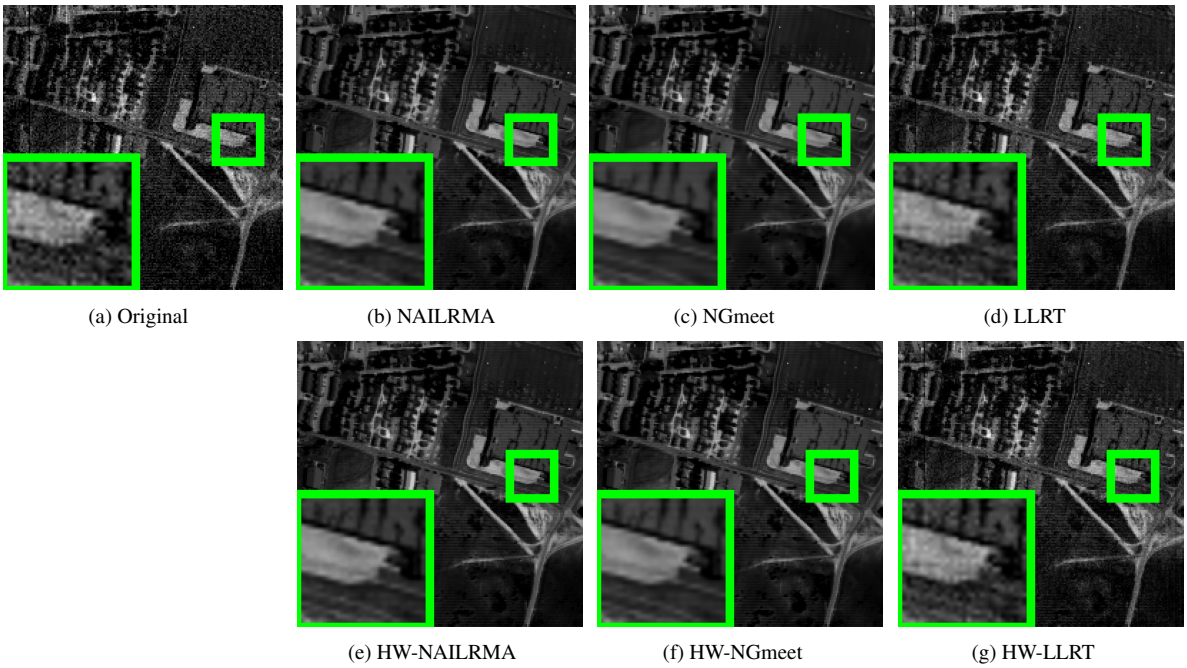
Figure 4: Transfer experiments: visual denoising results at the 3^{rd} band of image *pompoms* in CAVE dataset. The noisy HSI is corrupted by Gaussian and impulse noise.



(a) Original (PSNR,SSIM) | (b) NAILRMA (31.97,0.8760) | (c) NGmeet (33.58,0.8876) | (d) LLRT (33.40, 0.9135)

(e) Noisy (15.14,0.1823) | (f) HW-NAILRMA (**36.53**,**0.9564**) | (g) HW-NGmeet (**36.47**,**0.9512**) | (h) HW-LLRT (**35.48**,**0.9325**)

Figure 5: Transfer experiments: visual denoising results at the 39^{th} band of PaviaU. The noisy HSI is corrupted by spatial-spectral variant Gaussian noise.



(a) Original (b) NAILRMA (c) NGmeet (d) LLRT

(e) HW-NAILRMA (f) HW-NGmeet (g) HW-LLRT

Figure 6: Transfer experiments: real noisy HSI denoising results at the 204^{th} band of HYDICE Urban.