Supplementary Material of "Tensor Compressive Sensing Fused Low-rankness and Local-smoothness"

1 Proofs of Main Theorems

1.1 Some Key Lemmas

First of all, it can be easily derived from the definitions that for any tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we have

$$\frac{1}{\sqrt{\ell}}\|\boldsymbol{\mathcal{X}}\|_F \leq \|\boldsymbol{\mathcal{X}}\|_{\star_L} \leq \sqrt{\frac{rn_3}{\ell}}\|\boldsymbol{\mathcal{X}}\|_F, \tag{1}$$

where $r = \text{rank}_t(\boldsymbol{\mathcal{X}})$. Suppose $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star_L \boldsymbol{\mathcal{S}} \star_L \boldsymbol{\mathcal{V}}^T$, it can be equivalently rewritten as

$$\boldsymbol{\mathcal{X}} = \sum_{i=1}^{r} \boldsymbol{\mathcal{U}}(:,i,:) \star_L \boldsymbol{\mathcal{S}}(i,i,:) \star_L \boldsymbol{\mathcal{V}}(:,i,:)^T. \tag{2}$$

Moreover, for a matrix $\boldsymbol{X} \in \mathbb{C}^{n_1 \times n_2}$, we use $\|\boldsymbol{X}\|_{1,2} = \sum_{i=1}^{n_1} \sqrt{\sum_{j=1}^{n_2} x_{ij}^2}$ to denote the matrix $\ell_{1,2}$ norm.

Lemma 1 (ε-net for S_r)**.** *Let $\varepsilon > 0$, there exists a subset $\bar{S}_r$ of S_r such that*

$$\sup_{\boldsymbol{\mathcal{X}} \in S_r} \inf_{\boldsymbol{\mathcal{X}}' \in \bar{S}_r} \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}'\|_F \leq \varepsilon, \tag{3}$$

and

$$|\bar{S}_r| \leq \left(\frac{9}{\varepsilon}\right)^{(n_1+n_2+1)rn_3}. \tag{4}$$

Proof. The proof is similar to Lemma 3 of [2], we omit it here. □

Lemma 2. *Suppose $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are two tensors of the same size, if $\boldsymbol{\mathcal{X}}^T \star_L \boldsymbol{\mathcal{Y}} = \boldsymbol{0}$ and $\boldsymbol{\mathcal{X}} \star_L \boldsymbol{\mathcal{Y}}^T = \boldsymbol{0}$, then*

$$\|\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Y}}\|_{\star_L} = \|\boldsymbol{\mathcal{X}}\|_{\star_L} + \|\boldsymbol{\mathcal{Y}}\|_{\star_L}.$$

Proof. Since $\boldsymbol{\mathcal{X}}^T \star_L \boldsymbol{\mathcal{Y}} = \boldsymbol{0}$ and $\boldsymbol{\mathcal{X}} \star_L \boldsymbol{\mathcal{Y}}^T = \boldsymbol{0}$, we have $\bar{\boldsymbol{X}}^T\bar{\boldsymbol{Y}} = \boldsymbol{0}$ and $\bar{\boldsymbol{X}}\bar{\boldsymbol{Y}}^T = \boldsymbol{0}$. By the definition of tensor nuclear norm, we further obtain

$$\begin{aligned}
\|\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Y}}\|_{\star_L} =& \frac{1}{\ell}\|\bar{\boldsymbol{X}} + \bar{\boldsymbol{Y}}\|_* \\
=& \frac{1}{\ell}(\|\bar{\boldsymbol{X}}\|_* + \|\bar{\boldsymbol{Y}}\|_*) \\
=& \|\boldsymbol{\mathcal{X}}\|_{\star_L} + \|\boldsymbol{\mathcal{Y}}\|_{\star_L},
\end{aligned}$$

where the second equal sign holds by Lemma 2.3 of [1]. □

Lemma 3. *Let $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{Y}}$ be two tensors of the same size, and* $\mathrm{rank}_t(\boldsymbol{\mathcal{X}}) \le r$. *Then there exists two tensors $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ such that*
(1) $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{Y}}_1 + \boldsymbol{\mathcal{Y}}_2$;
(2) $\langle \boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{Y}}_2 \rangle = 0$;
(3) $\mathrm{rank}_t(\boldsymbol{\mathcal{Y}}_1) \le 2r$;
(4) $\boldsymbol{\mathcal{X}}^T \star_L \boldsymbol{\mathcal{Y}}_2 = \mathbf{0}$ *and* $\boldsymbol{\mathcal{X}} \star_L \boldsymbol{\mathcal{Y}}_2^T = \mathbf{0}$.

Proof. Consider a full tensor singular value decomposition of $\boldsymbol{\mathcal{X}}$, i.e. $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star_L \boldsymbol{\mathcal{S}} \star_L \boldsymbol{\mathcal{V}}^T$, and let

$$\hat{\boldsymbol{\mathcal{Y}}} = \begin{bmatrix} \hat{\boldsymbol{\mathcal{Y}}}_{11} & \hat{\boldsymbol{\mathcal{Y}}}_{12} \\ \hat{\boldsymbol{\mathcal{Y}}}_{21} & \hat{\boldsymbol{\mathcal{Y}}}_{22} \end{bmatrix} = \boldsymbol{\mathcal{U}}^T \star_L \boldsymbol{\mathcal{Y}} \star_L \boldsymbol{\mathcal{V}}.$$

Next, we define

$$\boldsymbol{\mathcal{Y}}_1 = \boldsymbol{\mathcal{U}} \star_L \begin{bmatrix} \hat{\boldsymbol{\mathcal{Y}}}_{11} & \hat{\boldsymbol{\mathcal{Y}}}_{12} \\ \hat{\boldsymbol{\mathcal{Y}}}_{21} & \mathbf{0} \end{bmatrix} \star_L \boldsymbol{\mathcal{V}}^{\boldsymbol{T}}, \quad \boldsymbol{\mathcal{Y}}_2 = \boldsymbol{\mathcal{U}} \star_L \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\mathcal{Y}}}_{22} \end{bmatrix} \star_L \boldsymbol{\mathcal{V}}^{\boldsymbol{T}},$$

It is easy to verify that $\boldsymbol{\mathcal{Y}}_1$ and $\boldsymbol{\mathcal{Y}}_2$ satisfy properties (1) $\sim$ (4). □

Lemma 4. *Let r, r' be two non-negative integers, suppose that two tensors $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}'$ have the same size and* $\mathrm{rank}_t(\boldsymbol{\mathcal{X}}) \le r$, $\mathrm{rank}_t(\boldsymbol{\mathcal{X}}') \le r'$, $\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}' \rangle = 0$ *and a linear map $\mathcal{A}$ satisfies the TRIP condition, we have*

$$\left| \langle \mathcal{A}(\boldsymbol{\mathcal{X}}), \mathcal{A}(\boldsymbol{\mathcal{X}}') \rangle \right| \le \delta_{r+r'} \|\boldsymbol{\mathcal{X}}\|_F \|\boldsymbol{\mathcal{X}}'\|_F. \tag{5}$$

Proof. Without loss of generality, we assume $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{X}}'$ have unit Frobenius norms. Since $\mathrm{rank}_t(\boldsymbol{\mathcal{X}} \pm \boldsymbol{\mathcal{X}}') \le r + r'$, we get from the TRIP condition

$$(1 - \delta_{r+r'}) \|\boldsymbol{\mathcal{X}} \pm \boldsymbol{\mathcal{X}}'\|_F^2 \le \|\mathcal{A}(\boldsymbol{\mathcal{X}} \pm \boldsymbol{\mathcal{X}}')\|_2^2 \le (1 + \delta_{r+r'}) \|\boldsymbol{\mathcal{X}} \pm \boldsymbol{\mathcal{X}}'\|_F^2. \tag{6}$$

Besides, based on the assumption $\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}' \rangle = 0$, we have

$$\|\boldsymbol{\mathcal{X}} \pm \boldsymbol{\mathcal{X}}'\|_F^2 = \|\boldsymbol{\mathcal{X}}\|_F^2 + \|\boldsymbol{\mathcal{X}}'\|_F^2 = 2. \tag{7}$$

Now, by using the parallelogram identity we can deduce

$$\begin{aligned} \left| \langle \mathcal{A}(\boldsymbol{\mathcal{X}}), \mathcal{A}(\boldsymbol{\mathcal{X}}') \rangle \right| &= \frac{1}{4} \left| \|\mathcal{A}(\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{X}}')\|_2^2 - \|\mathcal{A}(\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}')\|_2^2 \right| \\ &\le \delta_{r+r'}, \end{aligned}$$

where we use the results form (6) and (7) in the inequality. This completes the proof. □

Lemma 5. *Let $\boldsymbol{B}, \boldsymbol{C} \in \mathbb{C}^{n_1 \times n_2}$ be two matrices, and $\boldsymbol{B}\boldsymbol{B}^H = \boldsymbol{C}\boldsymbol{C}^H$, then there exist $\boldsymbol{U} \in \mathbb{C}^{n_2 \times n_2}$ satisfying $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{U}$ and $\boldsymbol{U}\boldsymbol{U}^H = \boldsymbol{I}_{n_1}$.*

Proof. Since $\boldsymbol{B}\boldsymbol{B}^H = \boldsymbol{C}\boldsymbol{C}^H$, we obtain that $\boldsymbol{B}$ and $\boldsymbol{C}$ have the same singular values $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \cdots, \sigma_r, 0, \cdots, 0]$. Let $\boldsymbol{B} = \boldsymbol{U}_1 \,\mathrm{diag}(\boldsymbol{\sigma}) \boldsymbol{V}_1^H$ and $\boldsymbol{C} = \boldsymbol{U}_2 \,\mathrm{diag}(\boldsymbol{\sigma}) \boldsymbol{V}_2^H$. The condition $\boldsymbol{B}\boldsymbol{B}^H = \boldsymbol{C}\boldsymbol{C}^H$ again implies $\boldsymbol{U}_1 = \boldsymbol{U}_2$. Let $\boldsymbol{U} = \boldsymbol{V}_1 \boldsymbol{V}_2^H$, it is easy to check that the conclusion holds. □

Lemma 6. *Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, and $\boldsymbol{L} \in \mathbb{C}^{n_3 \times n_3}$ be an invertible linear transform such that $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{L}^T\boldsymbol{L} = \ell \boldsymbol{I}_{n_3}$, then $\|\boldsymbol{\mathcal{X}}\|_{\star_L} \le \sqrt{n_3/\ell} \|\boldsymbol{\mathcal{X}}\|_1$.*

Proof. By Lemma 5, there exists $\boldsymbol{U} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$ such that

$$\begin{aligned} \|\boldsymbol{\mathcal{X}}\|_{\star_L} &= \frac{1}{\ell} \|\bar{\boldsymbol{X}}\|_* = \frac{1}{\ell} \,\mathrm{tr}\left(\sqrt{\bar{\boldsymbol{X}} \bar{\boldsymbol{X}}^H} \right) = \frac{1}{\ell} \,\mathrm{tr}(\bar{\boldsymbol{X}} \boldsymbol{U}^H) \\ &= \frac{1}{\ell} \,\mathrm{tr}(\mathrm{bdiag}(\boldsymbol{\mathcal{X}} \times_3 \boldsymbol{L}) \boldsymbol{U}^H). \end{aligned} \tag{8}$$

Let

$$\boldsymbol{L}\otimes\boldsymbol{I}_{n_1}=\begin{bmatrix}\boldsymbol{L}_1\\ \boldsymbol{L}_2\\ \vdots\\ \boldsymbol{L}_{n_3}\end{bmatrix},\ \text{and}\ \boldsymbol{U}^H=[\boldsymbol{U}_1^H,\boldsymbol{U}_2^H,\cdots,\boldsymbol{U}_{n_3}^H],$$

it holds from equation (8) that

$$\|\boldsymbol{\mathcal{X}}\|_{\star_L}=\frac{1}{\ell}\operatorname{tr}(\operatorname{diag}(\boldsymbol{L}\otimes\boldsymbol{I}_{n_1}\cdot\operatorname{unfold}(\boldsymbol{\mathcal{X}}))\boldsymbol{U}^H), \tag{9}$$

where

$$\operatorname{diag}(\boldsymbol{L}\otimes\boldsymbol{I}_{n_1}\cdot\operatorname{unfold}(\boldsymbol{\mathcal{X}}))=\begin{bmatrix}\boldsymbol{L}_1\operatorname{unfold}(\boldsymbol{\mathcal{X}}) & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{L}_2\operatorname{unfold}(\boldsymbol{\mathcal{X}}) & \cdots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{L}_{n_3}\operatorname{unfold}(\boldsymbol{\mathcal{X}})\end{bmatrix},\ \text{and}\ \operatorname{unfold}(\boldsymbol{\mathcal{X}})=\begin{bmatrix}\boldsymbol{\mathcal{X}}^{(1)}\\ \boldsymbol{\mathcal{X}}^{(2)}\\ \vdots\\ \boldsymbol{\mathcal{X}}^{(n_3)}\end{bmatrix}.$$

Now, equation (9) implies

$$\begin{aligned}\|\boldsymbol{\mathcal{X}}\|_{\star_L}=&\frac{1}{\ell}\operatorname{tr}\left([\boldsymbol{U}_1^H,\boldsymbol{U}_2^H,\cdots,\boldsymbol{U}_{n_3}^H]\begin{bmatrix}\boldsymbol{L}_1\operatorname{unfold}(\boldsymbol{\mathcal{X}}) & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{L}_2\operatorname{unfold}(\boldsymbol{\mathcal{X}}) & \cdots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{L}_{n_3}\operatorname{unfold}(\boldsymbol{\mathcal{X}})\end{bmatrix}\right)\\ =&\frac{1}{\ell}\operatorname{tr}\left(\sum_{i=1}^{n_3}\boldsymbol{U}_i^H\left[\boldsymbol{0},\cdots,\boldsymbol{L}_i\operatorname{unfold}(\boldsymbol{\mathcal{X}}),\cdots,\boldsymbol{0}\right]\right)\\ =&\frac{1}{\ell}\operatorname{tr}\left(\sum_{i=1}^{n_3}\boldsymbol{U}_i^H\boldsymbol{L}_i\left[\boldsymbol{0},\cdots,\operatorname{unfold}(\boldsymbol{\mathcal{X}}),\cdots,\boldsymbol{0}\right]\right)\\ =&\frac{1}{\ell}\operatorname{tr}\left(\operatorname{unfold}(\boldsymbol{\mathcal{X}})\sum_{i=1}^{n_3}(\boldsymbol{U}_i^H\boldsymbol{L}_i)_i\right)\end{aligned}$$

where we use the notation $(\boldsymbol{U}_i^H\boldsymbol{L}_i)_i$ to denote the submatrix of size $n_2\times n_1n_3$ consisting of the $(i-1)n_2+1$ to in_2 rows of $\boldsymbol{U}_i^H\boldsymbol{L}_i$. Besides, it is easy to check that the ℓ_2 norm of each column vector of $\sum_{i=1}^{n_3}(\boldsymbol{U}_i^H\boldsymbol{L}_i)$ satisfies $\leq\sqrt{\ell n_3}$, we obtain

$$\begin{aligned}\|\boldsymbol{\mathcal{X}}\|_{\star_L}\leq&\sqrt{\frac{n_3}{\ell}}\|\operatorname{unfold}(\boldsymbol{\mathcal{X}})\|_{1,2}\\ \leq&\sqrt{\frac{n_3}{\ell}}\|\boldsymbol{\mathcal{X}}\|_1,\end{aligned} \tag{10}$$

This completes the proof. □

1.2 Proof of Theorem 1

Proof. Consider all tensors from S_r, it follows from concentration inequality condition together with the union bound that

$$\begin{aligned}&P\left(\max_{\boldsymbol{\mathcal{X}}'\in\bar{S}_r}\left|\|\mathcal{A}(\boldsymbol{\mathcal{X}}')\|_2^2-\|\boldsymbol{\mathcal{X}}'\|_F^2\right|>t\|\boldsymbol{\mathcal{X}}'\|_F^2\right)\\ \leq&|\bar{S}_r|\times c_1\exp(-c_2m)\\ \leq&\left(\frac{9}{\varepsilon}\right)^{r(n_1+n_2+1)n_3}\times c_1\exp(-c_2m)\\ =&c_1\exp\left(r(n_1+n_2+1)n_3\ln\frac{9}{\varepsilon}-c_2m\right),\end{aligned} \tag{11}$$

where the second inequality uses Lemma 1. Choose $t = \delta$, $\varepsilon = \delta/4\sqrt{2}$ where $\delta = \delta_r$, (11) implies that with probability greater than $1 - c_1 \exp\left(r(n_1+n_2+1)n_3 \ln \frac{36\sqrt{2}}{\delta} - c_2 m\right)$, we have

$$\max_{\boldsymbol{\mathcal{X}}' \in \bar{S}_r} \left| \|\mathcal{A}(\boldsymbol{\mathcal{X}}')\|_2^2 - \|\boldsymbol{\mathcal{X}}'\|_F^2 \right| \leq \frac{\delta}{2}. \tag{12}$$

We continue by showing that

$$(1-\delta)\|\tilde{\boldsymbol{\mathcal{X}}}\|_F^2 \leq \|\tilde{\mathcal{A}}(\tilde{\boldsymbol{\mathcal{X}}})\|_2^2 \leq (1+\delta)\|\tilde{\boldsymbol{\mathcal{X}}}\|_F^2 \tag{13}$$

holds for any $\tilde{\boldsymbol{\mathcal{X}}} \in S_r$. To this end, set

$$\nu_r = \sup_{\tilde{\boldsymbol{\mathcal{X}}} \in S_r} \|\mathcal{A}(\tilde{\boldsymbol{\mathcal{X}}})\|_2, \tag{14}$$

By Lemma 1, for any $\tilde{\boldsymbol{\mathcal{X}}} \in S_r$, there exists $\boldsymbol{\mathcal{X}}' \in \bar{S}_r$ such that $\|\tilde{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}'\|_F < \delta/4\sqrt{2}$. Therefore, it follows by the triangle inequality and (12) that

$$\begin{aligned} \|\mathcal{A}(\tilde{\boldsymbol{\mathcal{X}}})\|_2 \leq & \|\mathcal{A}(\tilde{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}')\|_2 + \|\mathcal{A}(\boldsymbol{\mathcal{X}}')\|_2 \\ \leq & \|\mathcal{A}(\tilde{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}')\|_2 + 1 + \frac{\delta}{2}. \end{aligned} \tag{15}$$

Put $\boldsymbol{\mathcal{D}} = \tilde{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}'$, we have $\mathrm{rank}_t(\boldsymbol{\mathcal{D}}) \leq 2r$. We further decompose $\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}_1 + \boldsymbol{\mathcal{D}}_2$ by the T-SVD where $\langle \boldsymbol{\mathcal{D}}_1, \boldsymbol{\mathcal{D}}_2 \rangle = 0$ and $\mathrm{rank}_t(\boldsymbol{\mathcal{D}}_i) \leq r$ $(i = 1, 2)$. Note that $\boldsymbol{\mathcal{D}}_1/\|\boldsymbol{\mathcal{D}}_1\|_F$, $\boldsymbol{\mathcal{D}}_2/\|\boldsymbol{\mathcal{D}}_2\|_F \in S_r$, we obtain

$$\begin{aligned} \|\mathcal{A}(\boldsymbol{\mathcal{D}})\|_2 \leq & \|\mathcal{A}(\boldsymbol{\mathcal{D}}_1)\|_2 + \|\mathcal{A}(\boldsymbol{\mathcal{D}}_2)\|_2 \\ \leq & \nu_r(\|\boldsymbol{\mathcal{D}}_1\|_F + \|\boldsymbol{\mathcal{D}}_2\|_F) \\ \leq & \sqrt{2(\|\boldsymbol{\mathcal{D}}_1\|_F^2 + \|\boldsymbol{\mathcal{D}}_2\|_F^2)}\nu_r \\ = & \sqrt{2}\|\boldsymbol{\mathcal{D}}\|_F \nu_r \\ \leq & \frac{\delta}{4}\nu_r. \end{aligned} \tag{16}$$

Combining (15) and (16) implies

$$\|\mathcal{A}(\tilde{\boldsymbol{\mathcal{X}}})\|_2 \leq \frac{\delta}{4}\nu_r + 1 + \frac{\delta}{2}. \tag{17}$$

Theorefore, $\nu_r \leq \delta/4\nu_r + 1 + \delta/2$, and thus $\nu_r \leq \frac{1+\delta/2}{1-\delta/4} \leq 1 + \delta$ which essentially indicates the right inequality of (13). Now that this is established, the lower bound follows from

$$\begin{aligned} \|\mathcal{A}(\tilde{\boldsymbol{\mathcal{X}}})\|_2 \geq & \|\mathcal{A}(\boldsymbol{\mathcal{X}}')\|_2 - \|\mathcal{A}(\boldsymbol{\mathcal{D}})\|_2 \\ \geq & 1 - \frac{\delta}{2} - \frac{\delta(1+\delta)}{4} \\ \geq & 1 - \delta. \end{aligned}$$

Up to now, we have shown (13). This completes the proof. □

1.3 Proof of Theorem 2

Proof. Let $\boldsymbol{\mathcal{D}} = \hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}$, for $i \in [3]$, we decompose $\nabla_i \boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_c^{(i)}$ by Lemma 3 such that

$$\begin{aligned} &(1)\ \mathrm{rank}_t(\boldsymbol{\mathcal{D}}_0^{(i)}) \leq 2r; \qquad (2)\ \langle \boldsymbol{\mathcal{D}}_0^{(i)}, \boldsymbol{\mathcal{D}}_c^{(i)} \rangle = 0; \\ &(3)\ (\nabla_i \boldsymbol{\mathcal{X}}_{r,s}) \star_L (\boldsymbol{\mathcal{D}}_c^{(i)})^T = \mathbf{0};\ (4)\ (\nabla_i \boldsymbol{\mathcal{X}}_{r,s})^T \star_L (\boldsymbol{\mathcal{D}}_c^{(i)}) = \mathbf{0}. \end{aligned}$$

Let $\boldsymbol{\mathcal{X}}_{r,s}^c = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}_{r,s}$, the decomposition of $\nabla_i \boldsymbol{\mathcal{D}}$ implies

$$\begin{aligned}\|\nabla_i(\boldsymbol{\mathcal{X}}+\boldsymbol{\mathcal{D}})\|_{\star_L} =& \|\nabla_i(\boldsymbol{\mathcal{X}}_{r,s}^c + \boldsymbol{\mathcal{X}}_{r,s}) + \boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_c^{(i)}\|_{\star_L} \\ \geq& \|\nabla_i \boldsymbol{\mathcal{X}}_{r,s} + \boldsymbol{\mathcal{D}}_c^{(i)}\|_{\star_L} - \|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L} - \|\boldsymbol{\mathcal{D}}_0^{(i)}\|_{\star_L} \\ =& \|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}\|_{\star_L} + \|\boldsymbol{\mathcal{D}}_c^{(i)}\|_{\star_L} - \|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L} - \|\boldsymbol{\mathcal{D}}_0^{(i)}\|_{\star_L}, \end{aligned} \tag{18}$$

where we use Lemma 2 in the last equality. Besides, since $\hat{\boldsymbol{\mathcal{X}}}$ is the optimal solution of the original model, we have

$$\begin{aligned}\sum_{i=1}^3 \|\nabla_i(\boldsymbol{\mathcal{X}}+\boldsymbol{\mathcal{D}})\|_{\star_L} \leq& \sum_{i=1}^3 \|\nabla_i \boldsymbol{\mathcal{X}}\|_{\star_L} \\ \leq& \sum_{i=1}^3 (\|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}\|_{\star_L} + \|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}). \end{aligned} \tag{19}$$

Combining (18) and (19) shows

$$\sum_{i=1}^3 \|\boldsymbol{\mathcal{D}}_c^{(i)}\|_{\star_L} \leq \sum_{i=1}^3 \left(\|\boldsymbol{\mathcal{D}}_0^{(i)}\|_{\star_L} + 2\|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}\right) \tag{20}$$

Now, we further carry out a decomposition $\boldsymbol{\mathcal{D}}_c^{(i)} = \sum_{j\geq 1} \boldsymbol{\mathcal{D}}_j^{(i)}$ based on the T-SVD of $\boldsymbol{\mathcal{D}}_c^{(i)} = \boldsymbol{\mathcal{U}} \star_L \boldsymbol{\mathcal{S}} \star_L \boldsymbol{\mathcal{V}}^T$ such that

$$\boldsymbol{\mathcal{D}}_j^{(i)} = \sum_{k=2r(j-1)+1}^{2rj} \boldsymbol{\mathcal{U}}(:,k,:) \star_L \boldsymbol{\mathcal{S}}(k,k,:) \star_L \boldsymbol{\mathcal{V}}(:,k,:)^T.$$

Such a decomposition indicates two facts: (1) $\mathrm{rank}_t(\boldsymbol{\mathcal{D}}_j^{(i)}) \leq 2r$; (2) $\boldsymbol{\mathcal{D}}_j^{(i)}$s are orthogonal to each other as well as $\boldsymbol{\mathcal{D}}_0^{(i)}$. With these preparations, direct computation shows

$$\begin{aligned}\sum_{j\geq 2} \|\boldsymbol{\mathcal{D}}_j^{(i)}\|_F \leq& \sum_{j\geq 2} \frac{1}{\sqrt{\ell}} \|\bar{\boldsymbol{D}}_j^{(i)}\|_* \leq \sum_{j\geq 2} \frac{1}{\sqrt{2r\ell}} \left\|\bar{\boldsymbol{D}}_{j-1}^{(i)}\right\|_* \\ \leq& \frac{1}{\sqrt{2r\ell}} \|\bar{\boldsymbol{D}}^{(i)}\|_* = \sqrt{\frac{\ell}{2r}} \|\boldsymbol{\mathcal{D}}_c^{(i)}\|_{\star_L}, \end{aligned} \tag{21}$$

where we use (1) in the first inequality. Next, (20) together with (21) gives

$$\begin{aligned}\sum_{i=1}^3 \sum_{j\geq 2} \|\boldsymbol{\mathcal{D}}_j^{(i)}\|_F \leq& \sqrt{\frac{\ell}{2r}} \sum_{i=1}^3 \left(\|\boldsymbol{\mathcal{D}}_0^{(i)}\|_{\star_L} + 2\|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}\right) \\ \leq& \sum_{i=1}^3 \left(\sqrt{n}\|\boldsymbol{\mathcal{D}}_0^{(i)}\|_F + \sqrt{\frac{2\ell}{r}} \|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}\right), \end{aligned} \tag{22}$$

where we use (1) again in the last inequality.

Additionally, since $\mathrm{rank}_t(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}) \leq 4r$, the TRIP conditions suggest

$$(1-\delta_{4r})\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F \leq \|\mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)})\|_2^2 \leq (1+\delta_{4r})\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F^2, \tag{23}$$

where

$$\begin{aligned}\|\mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)})\|_2^2 =& \left\langle \mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}), \mathcal{A}_i\big(\nabla_i \boldsymbol{\mathcal{D}} - \sum_{j\geq 2} \boldsymbol{\mathcal{D}}_j^{(i)}\big) \right\rangle \\ =& \langle \mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}), \mathcal{A}_i(\nabla_i \boldsymbol{\mathcal{D}}) \rangle - \left\langle \mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}), \mathcal{A}_i\big(\sum_{j\geq 2} \boldsymbol{\mathcal{D}}_j^{(i)}\big) \right\rangle. \end{aligned} \tag{24}$$

For the first term in the right hand side of (24), the Cauchy-Schwarz inequality implies

$$\begin{aligned}\langle \mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}), \mathcal{A}_i(\nabla_i \boldsymbol{\mathcal{D}})\rangle \leq & \|\mathcal{A}_i(\nabla_i \boldsymbol{\mathcal{D}})\|_2 \times \|\mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)})\|_2 \\ \leq & 2\varepsilon\sqrt{1+\delta_{4r}}\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F, \end{aligned} \tag{25}$$

where we use the tube constraint of gradients derived by the proof procedure from Theorem 1 of [3] and the upper bound from (23). For the second term in the right hand side of (24), since $\boldsymbol{\mathcal{D}}_j^{(i)}$s are orthogonal to each other, we can derive from Lemma 4 that

$$\begin{aligned}\big\langle \mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}), \mathcal{A}_i\big(\sum_{j\geq 2} \boldsymbol{\mathcal{D}}_j^{(i)}\big)\big\rangle =& \big\langle \mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)}), \mathcal{A}_i\big(\sum_{j\geq 2} \boldsymbol{\mathcal{D}}_j^{(i)}\big)\big\rangle + \big\langle \mathcal{A}_i(\boldsymbol{\mathcal{D}}_1^{(i)}), \mathcal{A}_i\big(\sum_{j\geq 2} \boldsymbol{\mathcal{D}}_j^{(i)}\big)\big\rangle \\ \leq & \delta_{4r}(\|\boldsymbol{\mathcal{D}}_0^{(i)}\|_F + \|\boldsymbol{\mathcal{D}}_1^{(i)}\|_F)\sum_{j\geq 2}\|\boldsymbol{\mathcal{D}}_j^{(i)}\|_F \\ \leq & \sqrt{2}\delta_{4r}(\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F)\sum_{j\geq 2}\|\boldsymbol{\mathcal{D}}_j^{(i)}\|_F \end{aligned} \tag{26}$$

where the second inequality holds because $\boldsymbol{\mathcal{D}}_0^{(i)}$ and $\boldsymbol{\mathcal{D}}_1^{(i)}$ are orthogonal. Now, substituting the bounds (25) and (26) into (24) implies

$$\|\mathcal{A}_i(\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)})\|_2^2 \leq \|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F\Big(2\varepsilon\sqrt{1+\delta_{4r}} + \sqrt{2}\delta_{4r}\sum_{j\geq 2}\|\boldsymbol{\mathcal{D}}_j^{(i)}\|_F\Big). \tag{27}$$

Combining the lower bound from (23) and the upper bound from (27), simple simplification shows

$$\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F \leq \frac{2\sqrt{1+\delta_{4r}}}{1-\delta_{4r}}\varepsilon + \frac{\sqrt{2}\delta_{4r}}{1-\delta_{4r}}\sum_{j\geq 2}\|\boldsymbol{\mathcal{D}}_j^{(i)}\|_F. \tag{28}$$

Putting the upper bound from (22) into (28), simple simplification again implies

$$\sum_{i=1}^3\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F \leq \frac{6\sqrt{1+\delta_{4r}}}{1-(1+\sqrt{2n})\delta_{4r}}\varepsilon + \frac{2\delta_{4r}\sqrt{\ell}}{1-(1+\sqrt{2n})\delta_{4r}}\sum_{i=1}^3\frac{\|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}}{\sqrt{r}}. \tag{29}$$

Hence, by (22) again

$$\begin{aligned}\sum_{i=1}^3\|\nabla_i\boldsymbol{\mathcal{D}}\|_F \leq & \sum_{i=1}^3\Big(\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F + \sum_{j\geq 2}\|\boldsymbol{\mathcal{D}}_j^{(i)}\|_F\Big) \\ \leq & \sum_{i=1}^3\Big((1+\sqrt{n})\|\boldsymbol{\mathcal{D}}_0^{(i)} + \boldsymbol{\mathcal{D}}_1^{(i)}\|_F + \sqrt{2\ell}\frac{\|\nabla_i\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}}{\sqrt{r}}\Big) \\ \leq & C_1\varepsilon + C_2\sum_{i=1}^3\frac{\|\nabla_i\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}}{\sqrt{r}}, \end{aligned} \tag{30}$$

where we use (29) in the last inequality and $C_1 = \frac{6(1+\sqrt{n})\sqrt{1+\delta_{4r}}}{1-(1+\sqrt{2n})\delta_{4r}}$, $C_2 = \frac{(1+(\sqrt{2}-1)\delta_{4r})\sqrt{2\ell}}{1-(1+\sqrt{2n})\delta_{4r}}$. Moreover, since $\|\mathcal{B}(\boldsymbol{\mathcal{D}})\|_2 \leq 2\varepsilon$, Theorem 4 of [3] gives that there exist an universal constant C_3 such that

$$\|\boldsymbol{\mathcal{D}}\|_F \leq C_3\frac{\|\boldsymbol{\mathcal{D}}\|_{\mathrm{TV}}}{\sqrt{s}}\log n + \varepsilon. \tag{31}$$

By the Cauchy-Schwarz inequality again, we have

$$\begin{aligned}\frac{\|\boldsymbol{\mathcal{D}}\|_F}{n^{3/2}} \leq & C_3\frac{\log n}{\sqrt{s}}\sum_{i=1}^3\|\nabla_i\boldsymbol{\mathcal{D}}\|_F + \frac{\varepsilon}{n^{3/2}} \\ \leq & C_3\frac{\log n}{\sqrt{s}}\Big(C_1\varepsilon + C_2\sum_{i=1}^3\frac{\|\nabla_i\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}}{\sqrt{r}}\Big) + \frac{\varepsilon}{n^{3/2}} \\ = & C_3C_2\log n\frac{\sum_{i=1}^3\|\nabla_i\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}}{\sqrt{rs}} + \Big(C_3C_1\frac{\log n}{\sqrt{s}} + \frac{1}{n^{3/2}}\Big)\varepsilon. \end{aligned} \tag{32}$$

This completes the proof. □

1.4 Proof of Corollary 1

Proof. It is enough to check that

$$\|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\mathrm{TCTV}} \lesssim \min\{\|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\mathrm{TV}}, \|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}\}. \tag{33}$$

Firstly, Lemma 6 implies that

$$\|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\mathrm{TCTV}} = \sum_{i=1}^{3} \|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L} \leq \sqrt{\frac{n}{\ell}} \sum_{i=1}^{3} \|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_1 = \sqrt{\frac{n}{\ell}} \|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\mathrm{TV}}. \tag{34}$$

Moreover

$$\|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\mathrm{TCTV}} = \sum_{i=1}^{3} \|\nabla_i \boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L} \leq 6 \|\boldsymbol{\mathcal{X}}_{r,s}^c\|_{\star_L}. \tag{35}$$

(34) and (35) directly implies (33). This completes the proof. □

1.5 Proof of Theorem 3

Proof. Our main technique is transforming the problem (19) of the manuscript into the two-block ADMM case. Concretely, Let $\boldsymbol{\mathcal{Z}} = [\boldsymbol{\mathcal{Z}}_1; \boldsymbol{\mathcal{Z}}_2; \boldsymbol{\mathcal{Z}}_3]$, $\boldsymbol{\mathcal{D}} = [\boldsymbol{\mathcal{D}}_1; \boldsymbol{\mathcal{D}}_2; \boldsymbol{\mathcal{D}}_3]$, $f(\boldsymbol{\mathcal{X}}) = \frac{1}{2}\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2$, $g(\boldsymbol{\mathcal{Z}}) = \lambda \sum_{i=1}^{3} \|\boldsymbol{\mathcal{Z}}_i\|_{\star_L}$, problem (19) of the manuscript can be equivalently written as

$$\min_{\boldsymbol{\mathcal{X}}} f(\boldsymbol{\mathcal{X}}) + g(\boldsymbol{\mathcal{Z}}),\ s.\ t.\ \boldsymbol{\mathcal{Z}} - \begin{bmatrix} \nabla_1 \\ \nabla_2 \\ \nabla_3 \end{bmatrix} \boldsymbol{\mathcal{X}} = \boldsymbol{0}. \tag{36}$$

Problem (36) involves two variable blocks, and the corresponding ADMM is guaranteed to converge to its global optimum [6]. What's more, the Lagrangian function of problem (36) is

$$\mathcal{L}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{D}}, \mu) = \frac{1}{2}\|\boldsymbol{y} - \mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2 + g(\boldsymbol{\mathcal{Z}}) + \frac{\mu}{2}\left\|\boldsymbol{\mathcal{Z}} - \begin{bmatrix} \nabla_1 \\ \nabla_2 \\ \nabla_3 \end{bmatrix} \boldsymbol{\mathcal{X}} + \frac{\boldsymbol{\mathcal{D}}}{\mu}\right\|_F^2 + C, \tag{37}$$

where C is only multipliers dependent squared items. According to the update rule of ADMM, variables are updated alternately.

- For term $\boldsymbol{\mathcal{X}}$, we update it by solving the following linear system obtained by taking derivative on both sides of the Lagrangian function with respect to $\boldsymbol{\mathcal{X}}$

$$(\mathcal{A}^T\mathcal{A} + \sum_{i=1}^{3} \mu^k \nabla_i^T \nabla_i)(\boldsymbol{\mathcal{X}}) = \mathcal{A}^T(\boldsymbol{y}) + \sum_{i=1}^{3} \nabla_i^T(\mu^k \boldsymbol{\mathcal{Z}}_i^k + \boldsymbol{\mathcal{D}}_i^k). \tag{38}$$

  Equations (38) can be solved by off-the-shelf techniques such as the preconditioned conjugate gradient method [7].

- For terms $\boldsymbol{\mathcal{Z}}$, we update it by solving

$$\min_{\boldsymbol{\mathcal{Z}}}\ g(\boldsymbol{\mathcal{Z}}) + \frac{\mu}{2}\left\|\boldsymbol{\mathcal{Z}} - \begin{bmatrix} \nabla_1 \\ \nabla_2 \\ \nabla_3 \end{bmatrix} \boldsymbol{\mathcal{X}} + \frac{\boldsymbol{\mathcal{D}}}{\mu}\right\|_F^2, \tag{39}$$

  which is equivalent to solve the following three problems

$$\begin{aligned} \boldsymbol{\mathcal{Z}}_i^{k+1} =& \arg\min_{\boldsymbol{\mathcal{Z}}_i} \lambda\|\boldsymbol{\mathcal{Z}}_i\|_{\star_L} + \frac{\mu^k}{2}\left\|\boldsymbol{\mathcal{Z}}_i - \nabla_i \boldsymbol{\mathcal{X}}^{k+1} + \frac{\boldsymbol{\mathcal{D}}_i^k}{\mu}\right\|_F^2 \\ =& \mathcal{D}_{\lambda/\mu^k}\left(\nabla_i \boldsymbol{\mathcal{X}}^{k+1} - \frac{\boldsymbol{\mathcal{D}}_i^k}{\mu^k}\right), \quad i \in [3] \end{aligned} \tag{40}$$

  where $\mathcal{D}_{\lambda/\mu}(\cdot)$ denotes the Tensor Singular Value Thresholding (T-SVT) operator which can be found in [8].

- For term $\boldsymbol{\mathcal{D}}$, we update it by

$$\boldsymbol{\mathcal{D}}^{k+1} = \boldsymbol{\mathcal{D}}^{k} + \mu^{k}\big(\boldsymbol{\mathcal{Z}}^{k+1} - \begin{bmatrix} \nabla_1 \\ \nabla_2 \\ \nabla_3 \end{bmatrix} \boldsymbol{\mathcal{X}}^{k+1}\big), \tag{41}$$

  which is equivalent to update $\boldsymbol{\mathcal{D}}_i$ $i \in [3]$ one by one, i.e.

$$\boldsymbol{\mathcal{D}}_i^{k+1} = \boldsymbol{\mathcal{D}}_i^{k} + \mu^{k}\big(\boldsymbol{\mathcal{Z}}_i^{k+1} - \nabla_i \boldsymbol{\mathcal{X}}^{k+1}\big). \tag{42}$$

Now, it is easy to find that the update procedures (38), (40), (42) are the same as line 2, line 3, line 4 of Algorithm 1 of the manuscript, respectively. Therefore, we have provided a theoretical guarantee for Algorithm 1 of the manuscript. □

2 More Experimental Results

In this section, we conduct more experiments to demonstrate the superiority of our models.

2.1 More Exhibitions for HSI Compressive Sensing

2.1.1 Spectral Signatures of Both HSI Datasets

We show the spectral signatures of a pixel (120, 120) of HYDICE Washington DC Mall and HYDICE Urbanpart in Fig. 1. The curves of our methods are closer to the original, which indicates that our approaches are better than the competing ones.

2.1.2 HSI Compressive Sensing with Noise

Because of imaging conditions in practice such as weather conditions and data transmission procedures, the observations of HSIs are often contaminated with noise. To illustrate the effectiveness of our models in this situation, we assume that the observations are contaminated by Gaussian noise $\mathcal{N}(0, \sigma^2)$. With a fixed SR = 5%, we conduct the experiment by varying $\sigma \in \{0.01, 0.02, 0.03, 0.04\}$. The results are shown in Table 1. It can be easily find that our models still behave better in almost all standard deviations under the noisy setting.

2.2 Application to Video Compressive Sensing

2.2.1 More Exhibitions for Both Video Cubes

We further show more visual comparisons of both video cubes under SR = 10% in Fig. 2 and Fig. 3. It can be seen that our models always present better visual quality across all the comparative frames.

2.2.2 Video Compressive Sensing on More Dataset

Further, we apply our models to another two video cubes coming from the I2R dataset [4]: Fountain and Shopping Mall. The sizes of Fountain and Shopping Mall used in this experiment are 128×160×64 and 144×176×64, respectively. As before, we compare our method with four other methods TV-DCT, MC-BCS-SPL, 3DTV and TNN.

Fig. 4 shows that our models obtain higher PSNR and SSIM values among all SRs. What is worth mentioning, on the extremely low SRs such as 1%, 5% and 10%, our models behave observably better than other four methods on both video cubes. For a fixed SR=10%, Fig. 5 further illustrates that our models are better than the others among almost all frames. The MC-BCS-SPL seems better in the key frames since it utilizes a higher SR=0.7. Visual comparisons about more frames of both video cubes are exhibited in Fig. 6 and Fig. 7. It can be seen that our models present closer visual effect to the original frames.

Table 1: Comparisons about quality indices of different models under different standard deviations and a fixed SR=5% on the two HSI datasets.

σ	Quality indices	Models					
		SparCS	TVAL3	JTTV	TNN	TCTV-DCT	TCTV-DFT
HYDICE Washington DC Mall							
0.01	PSNR	25.149	23.794	25.857	28.934	**33.728**	33.315
	SSIM	0.7500	0.6248	0.8293	0.8947	**0.9579**	0.9531
	ERGAS	242.77	237.41	206.36	131.01	**76.09**	79.41
0.02	PSNR	23.944	23.494	25.159	27.892	**31.153**	30.578
	SSIM	0.7052	0.6111	0.8114	0.8655	**0.9246**	0.9131
	ERGAS	272.93	243.93	216.42	147.18	**100.64**	107.28
0.03	PSNR	23.797	23.080	24.404	26.724	**29.005**	28.436
	SSIM	0.6903	0.5857	0.7862	0.8231	**0.8782**	0.8609
	ERGAS	281.05	254.42	229.68	167.47	**128.32**	137.12
0.04	PSNR	23.730	22.614	23.550	25.624	**27.288**	26.753
	SSIM	0.6829	0.5593	0.7536	0.7753	**0.8255**	0.8040
	ERGAS	280.57	268.09	248.00	189.32	**157.00**	167.35
HYDICE Urbanpart							
0.01	PSNR	26.305	24.897	34.071	31.205	**34.985**	34.659
	SSIM	0.7064	0.6162	0.9419	0.9001	**0.9440**	0.9381
	ERGAS	187.57	204.61	76.21	99.68	**68.49**	70.73
0.02	PSNR	25.279	24.680	32.415	29.796	**32.518**	32.145
	SSIM	0.6569	0.6038	0.9026	0.8640	**0.9075**	0.8971
	ERGAS	211.52	209.1	90.526	116.14	**87.299**	90.931
0.03	PSNR	24.917	24.338	30.516	28.503	**30.517**	30.142
	SSIM	0.6464	0.5821	0.8534	0.8188	**0.8605**	0.8465
	ERGAS	216.25	216.84	110.71	134.12	**108.27**	113.02
0.04	PSNR	24.846	23.960	**29.121**	27.339	28.888	28.542
	SSIM	0.6396	0.5558	**0.8620**	0.7687	0.8080	0.7919
	ERGAS	221.75	225.76	**128.71**	152.81	130.10	135.56

2.3 Application to Multispectral Image Compressive Sensing

Multispectral Image (MSI) is similar to HSI. But it is composed of wider wave bands. CS had also been applied to MSI recovery [5]. We choose toy and cloth from the CAVE dataset [1] in this experiment and both are of the size 512×512×31. As in the HSI experiment, we use four other methods SparCS, TVAL3, JTTV and TNN for comparisons.

Table 2 collects the quality indices of recovered MSIs. It can be seen that compared with the other four competing methods, our models achieve the highest PSNR and SSIM values and the lowest ERGAS values among all SRs. In the extremely low SRs such as 0.3% and 1%, the difference between our method and the second best one is particularly obvious. TCTV-DCT and TCTV-DFT behave similarly in all SRs. For a fixed SR=1%, Fig. 8 further displays the quality indices of both MSI datasets on all bands. Our models are always dramatically better than others in terms of neither PSNR, SSIM, or ERGAS values. Moreover, Fig. 9 shows the visual comparisons of the recovered images which are composited as RGB images (R: band 23, G: band 13, B: band 4), and the left and right bottoms of each RGB image denote the locally magnified image with a magnification 2 and its residual image, respectively. Our models are distinctly closer to the original images. At last, we examine the effect of noise coming from the observation process. Similar to the HSI setting, we assume that the observations are contaminated by Gaussian noise from $\mathcal{N}(0, \sigma^2)$. For a fixed SR = 5%, we conduct the experiment by varying $\sigma \in \{0.01, 0.02, 0.03, 0.04\}$ as in the HSI case. The results are shown in Table 3. Our models still achieve the best quality indices in all standard deviations under the noisy setting.

[1] https://www.cs.columbia.edu/CAVE/databases/multispectral/

Table 2: Comparisons about quality indices of different models under different SRs on two MSI datasets.

SR	Quality indices	Models					
		SparCS	TVAL3	JTTV	TNN	TCTV-DCT	TCTV-DFT
				toy			
0.3%	PSNR	11.880	19.878	13.094	13.633	**26.070**	25.851
	SSIM	0.1811	0.3650	0.1536	0.2297	**0.5854**	0.5668
	ERGAS	1117.80	408.99	874.96	822.00	**202.56**	208.22
1%	PSNR	12.031	23.401	14.871	17.868	**29.374**	29.220
	SSIM	0.1575	0.3632	0.2246	0.3582	**0.7646**	0.7478
	ERGAS	1169.90	282.55	715.92	507.05	**138.37**	141.37
5%	PSNR	14.936	30.538	22.108	31.882	36.706	**37.001**
	SSIM	0.2361	0.8898	0.5386	0.9100	**0.9565**	0.9541
	ERGAS	772.47	125.50	313.86	102.31	59.21	**57.54**
10%	PSNR	19.221	34.74	31.849	37.120	41.651	**42.302**
	SSIM	0.3979	0.9513	0.8923	0.9694	0.9871	**0.9872**
	ERGAS	548.45	84.70	112.78	56.31	33.52	**31.30**
				cloth			
0.3%	PSNR	17.104	18.534	17.945	17.441	**21.150**	20.800
	SSIM	0.2100	0.2878	0.3330	0.2397	**0.4932**	0.4534
	ERGAS	573.95	434.19	467.58	496.39	**320.77**	337.54
1%	PSNR	17.157	19.655	18.736	18.814	**23.801**	23.633
	SSIM	0.2195	0.3851	0.4480	0.3480	**0.7063**	0.6933
	ERGAS	590.33	381.35	429.85	423.21	**234.48**	239.57
5%	PSNR	19.031	22.712	21.883	26.669	30.411	**30.473**
	SSIM	0.5013	0.6594	0.7245	0.8645	**0.9312**	0.9298
	ERGAS	412.08	261.96	297.37	164.57	107.73	**106.87**
10%	PSNR	21.125	25.076	26.427	31.380	34.689	**35.00**
	SSIM	0.6752	0.8012	0.8886	0.9508	0.9730	**0.9736**
	ERGAS	319.66	198.79	171.05	96.40	67.33	**64.98**

Table 3: Comparisons about quality indices of different models under different standard diviations and fixed SR=5% on two MSI datasets.

σ	Quality indices	Models					
		SparCS	TVAL3	JTTV	TNN	TCTV-DCT	TCTV-DFT
				toy			
0.01	PSNR	15.055	30.408	21.857	31.309	**35.550**	35.545
	SSIM	0.2411	0.8859	0.5285	0.8977	**0.9313**	0.9245
	ERGAS	758.21	127.17	323.00	109.40	**68.22**	68.59
0.02	PSNR	14.784	29.848	21.842	30.162	**33.506**	33.367
	SSIM	0.2289	0.8644	0.5268	0.8680	**0.8667**	0.8551
	ERGAS	790.55	135.34	323.48	125.16	**87.74**	89.62
0.03	PSNR	14.670	29.122	21.505	29.092	**31.729**	31.564
	SSIM	0.2280	0.8443	0.5132	0.8317	**0.7863**	0.7728
	ERGAS	791.50	144.29	336.30	142.16	**109.57**	112.36
0.04	PSNR	14.721	28.198	21.099	28.154	**30.275**	30.106
	SSIM	0.2249	0.7967	0.4966	0.7946	**0.7059**	0.6920
	ERGAS	788.81	162.62	352.34	159.25	**131.75**	135.20
				cloth			
0.01	PSNR	19.029	22.694	21.853	26.453	29.999	**30.013**
	SSIM	0.4992	0.6564	0.7225	0.8582	**0.9231**	0.9210
	ERGAS	411.54	262.55	298.56	168.89	113.05	**112.79**
0.02	PSNR	18.979	22.625	21.713	25.907	**29.068**	29.030
	SSIM	0.4926	0.6533	0.7137	0.8417	**0.9020**	0.8989
	ERGAS	413.14	264.85	303.44	180.34	**126.26**	126.84
0.03	PSNR	18.871	22.513	21.495	25.243	**28.011**	27.964
	SSIM	0.4848	0.6479	0.7001	0.8165	**0.8713**	0.8683
	ERGAS	420.79	268.30	311.52	195.33	**143.35**	144.33
0.04	PSNR	18.771	22.308	21.276	24.614	**27.001**	26.957
	SSIM	0.4757	0.6384	0.6866	0.7916	**0.8373**	0.8348
	ERGAS	426.85	274.99	319.54	210.62	**162.18**	163.41

References

[1] B. Recht, M. Fazel, P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471-501, Aug. 2010.

[2] F. Zhang, W. D. Wang, J. Y. Hou, J. J. Wang, and J. W. Huang, "Tensor restricted isometry property analysis for a large class of random measurement ensembles," *Sci. China-Inf. Sci.*, vol. 64, no. 1, pp. 1-3, Jan. 2021.

[3] D. Needell and R. Ward, "Near-Optimal Compressed Sensing Guarantees for Total Variation Minimization," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3941-3949, Oct. 2013.

[4] L. Li, W. Huang, I. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection, *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 14591472, 2004.

[5] P. Liu and K. B. Eom, "Compressive Sensing of Noisy Multispectral Images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1931-1935, Nov. 2014.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1-122, Jul. 2011.

[7] O. Axelsson and G. Lindskog, "On the rate of convergence of the preconditioned conjugate gradient method," *Numer. Math.*, vol. 48, no. 5, pp. 499-523, Sep. 1986.

[8] C. Y. Lu, X. Peng, and Y. C. Wei, "Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5996-6004 2019.

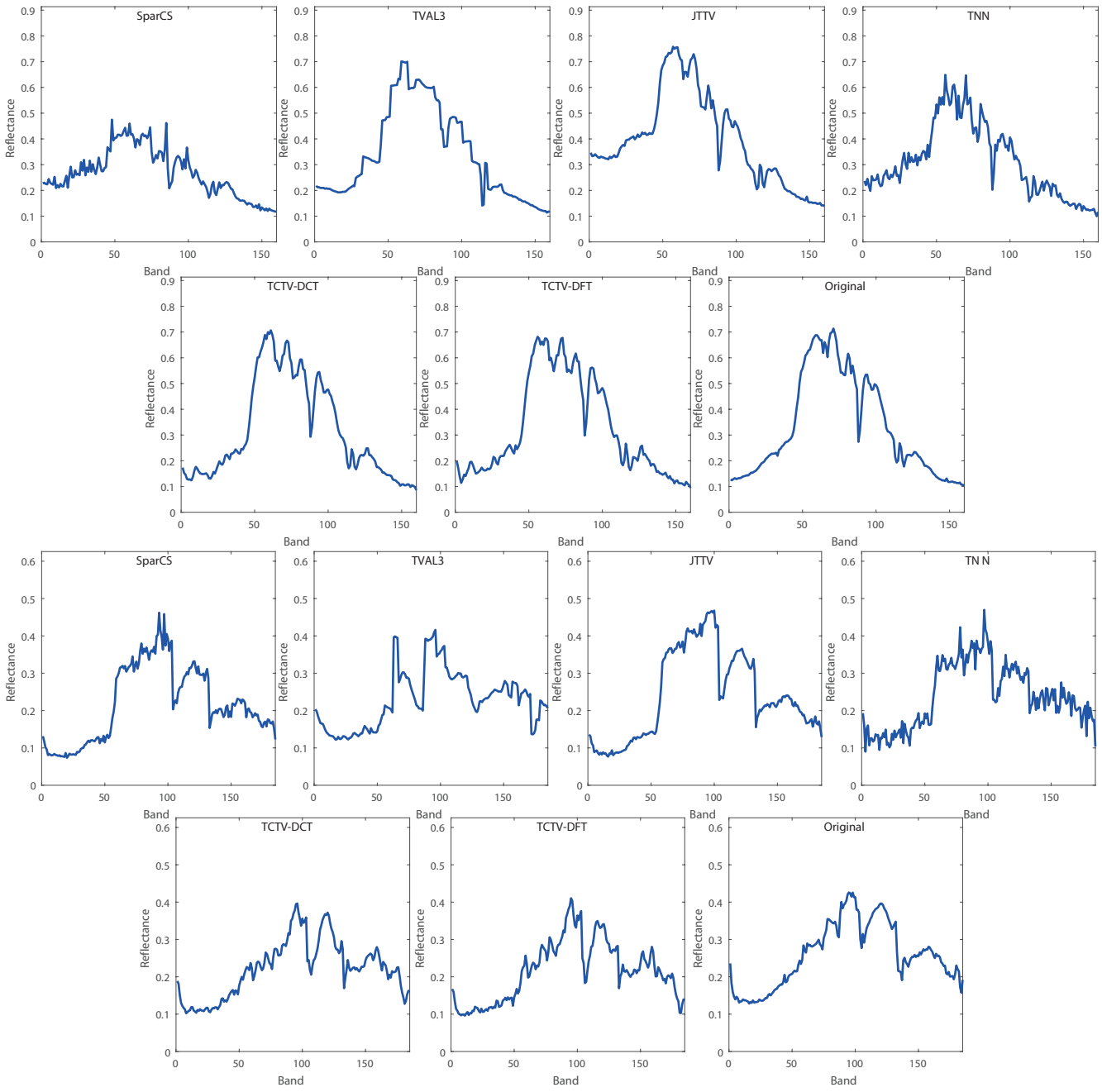

Fig. 1: Spectral signatures of a pixel (120, 120) of both HSI datasets under SR = 1%. Up: HYDICE Washington DC Mall, Down: HYDICE Urbanpart.

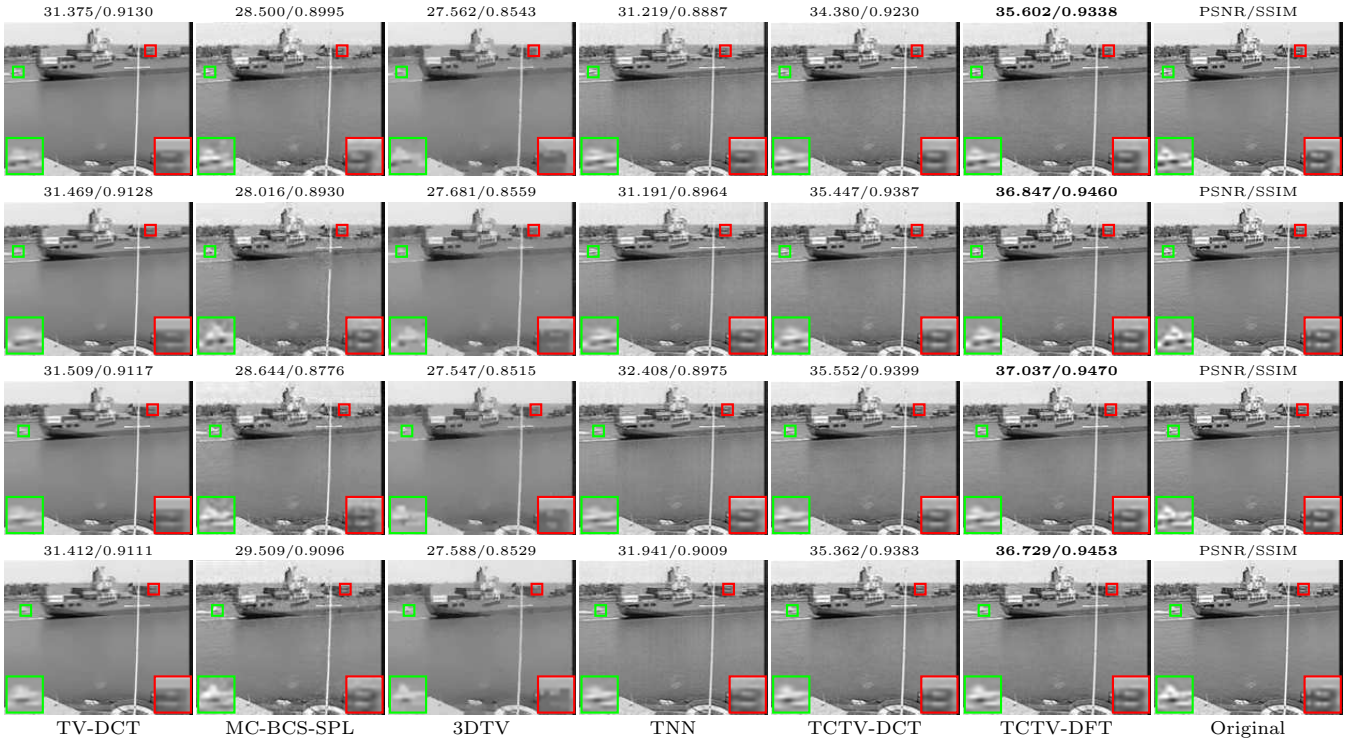

Fig. 2: Visual comparisons of the proposed approaches over four other competing methods on frames 5, 20, 35, 50 of Container under SR=10%.

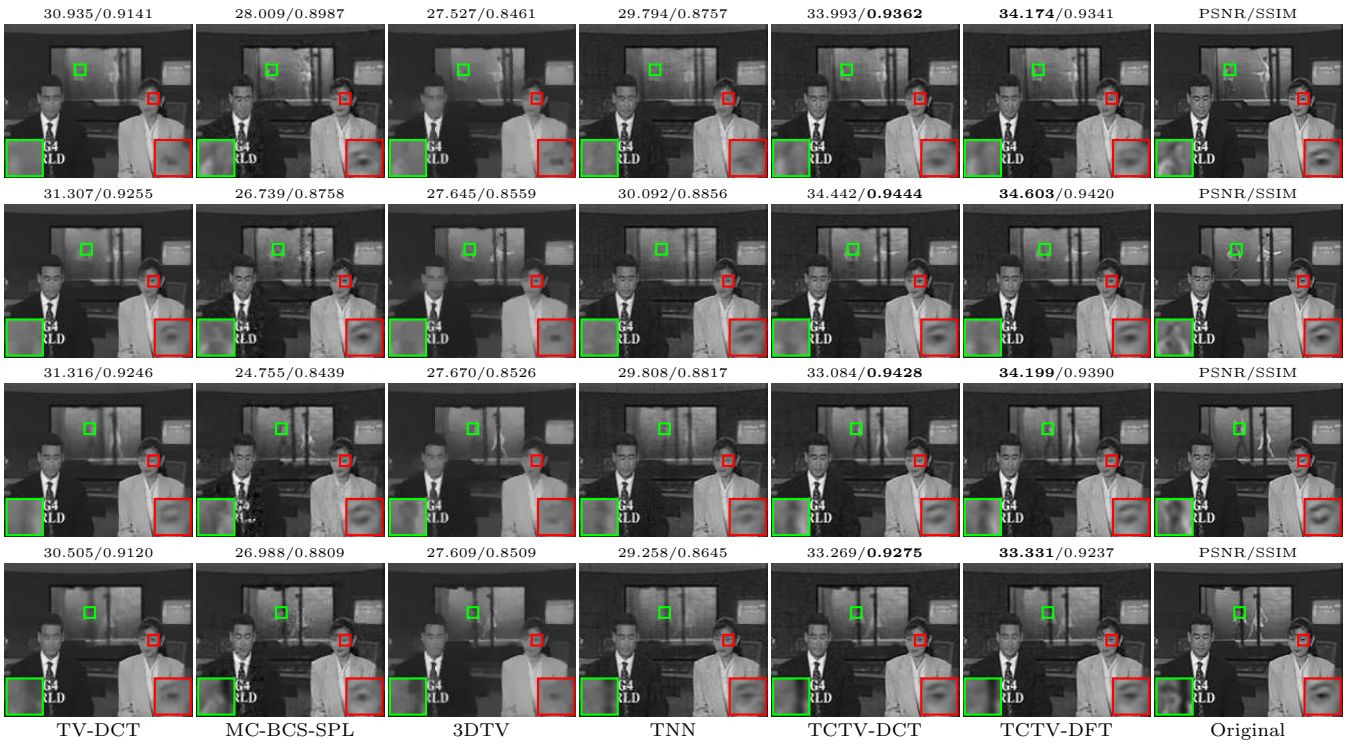

Fig. 3: Visual comparison of the proposed approaches over four other competing methods on frames 15, 30, 45, 60 of News under SR=10%.

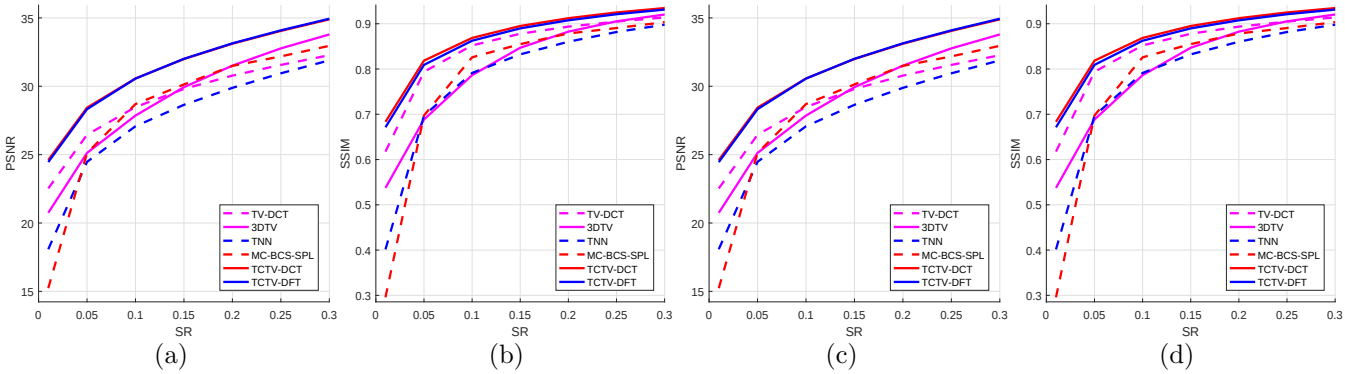

Fig. 4: PSNR and SSIM values comparisons of the proposed approaches over four other competing methods with varying SRs under SR=10%. (a) and (b): Fountain, (c) and (d): Shopping Mall.

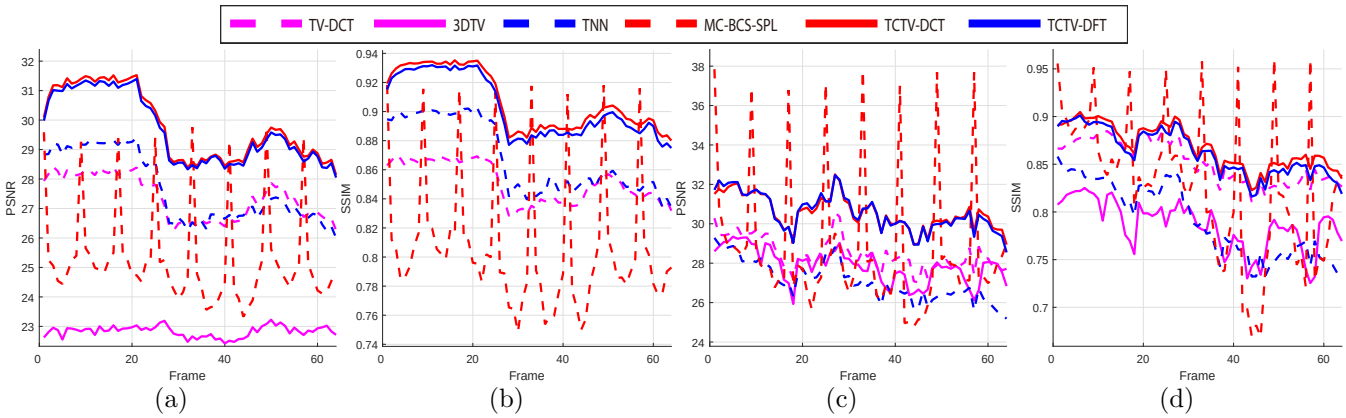

Fig. 5: PSNR and SSIM values comparisons of the proposed approaches over four other competing methods on all frames under SR=10%. (a) and (b): Fountain, (c) and (d): Shopping Mall.

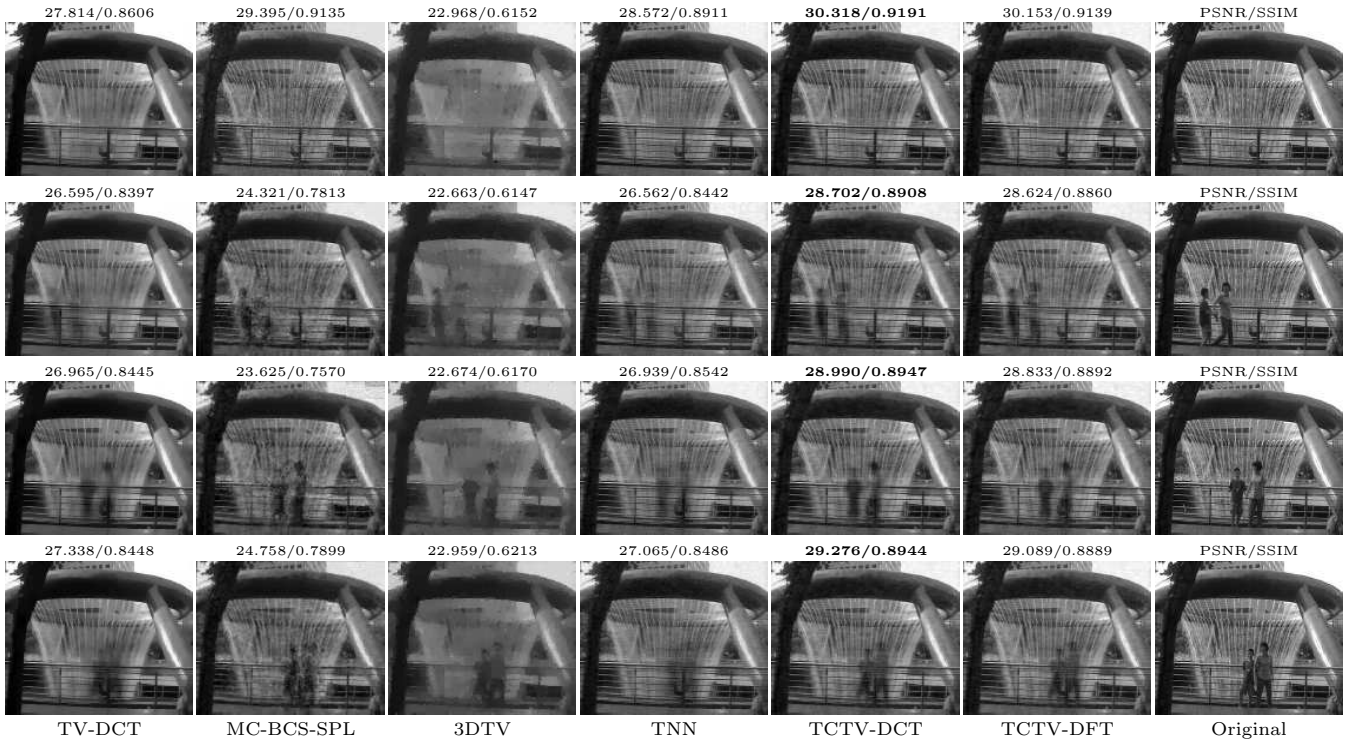

Fig. 6: Visual comparison of the proposed approaches over four other competing methods on frames 25, 35, 45, 55 of Fountain under SR=10%.

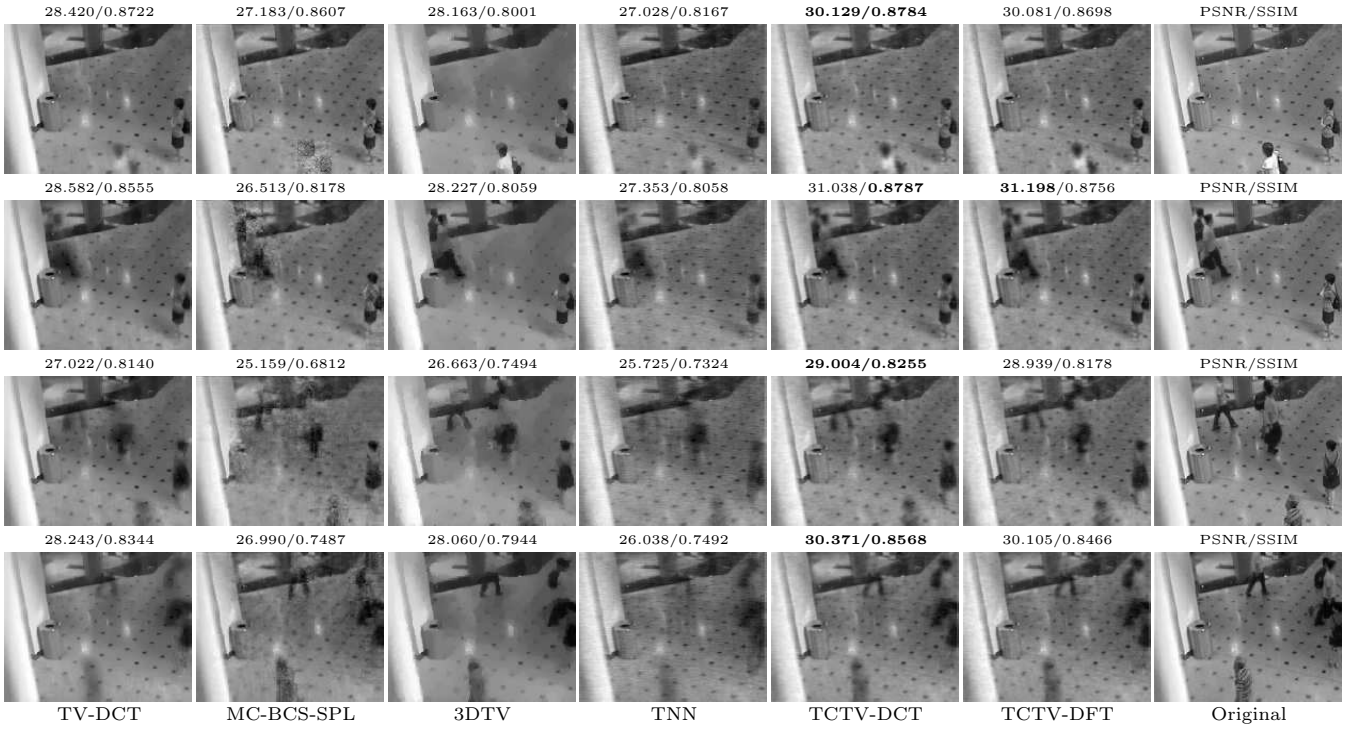

Fig. 7: Visual comparison of the proposed approaches over four other competing methods on frames 15, 30, 45, 60 of Shopping Mall under SR=10%.

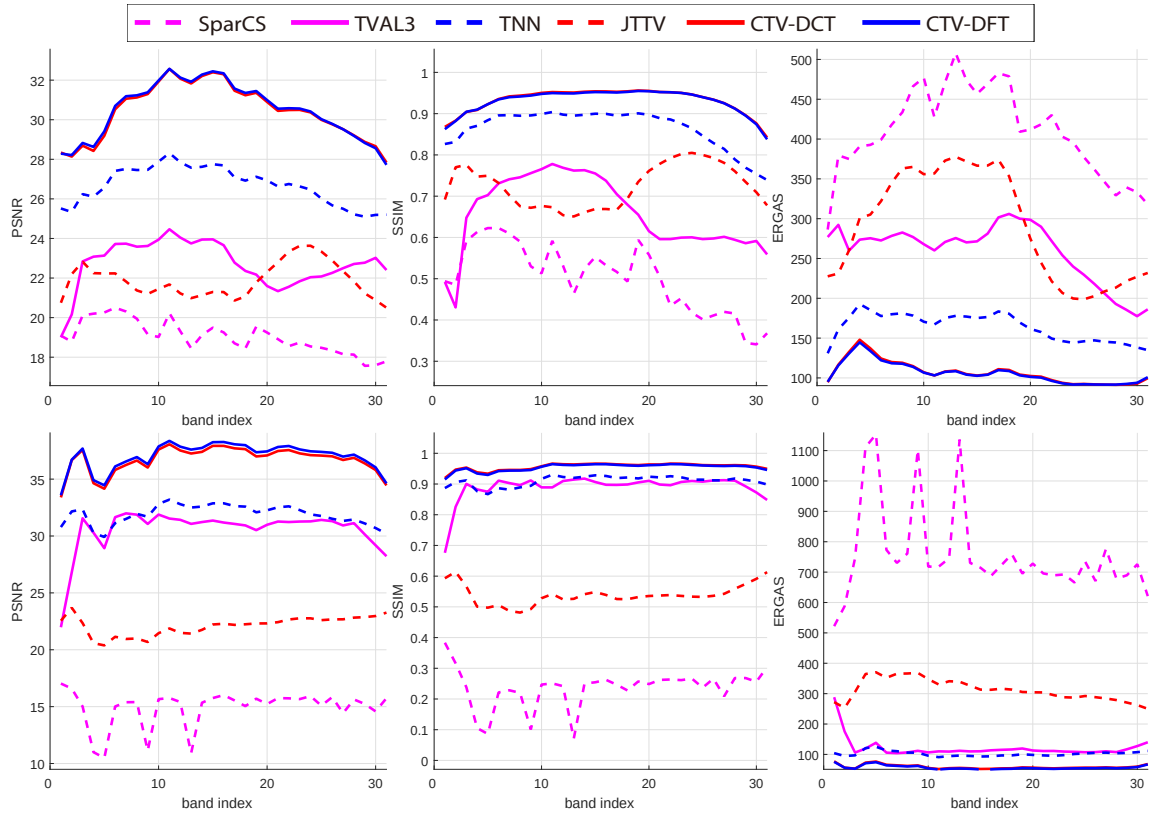

Fig. 8: PSNR, SSIM and ERGAS values comparisons of the proposed approaches over four other competing methods on all bands of both MSI datasets under SR=1%. Up: toy, Down: cloth.

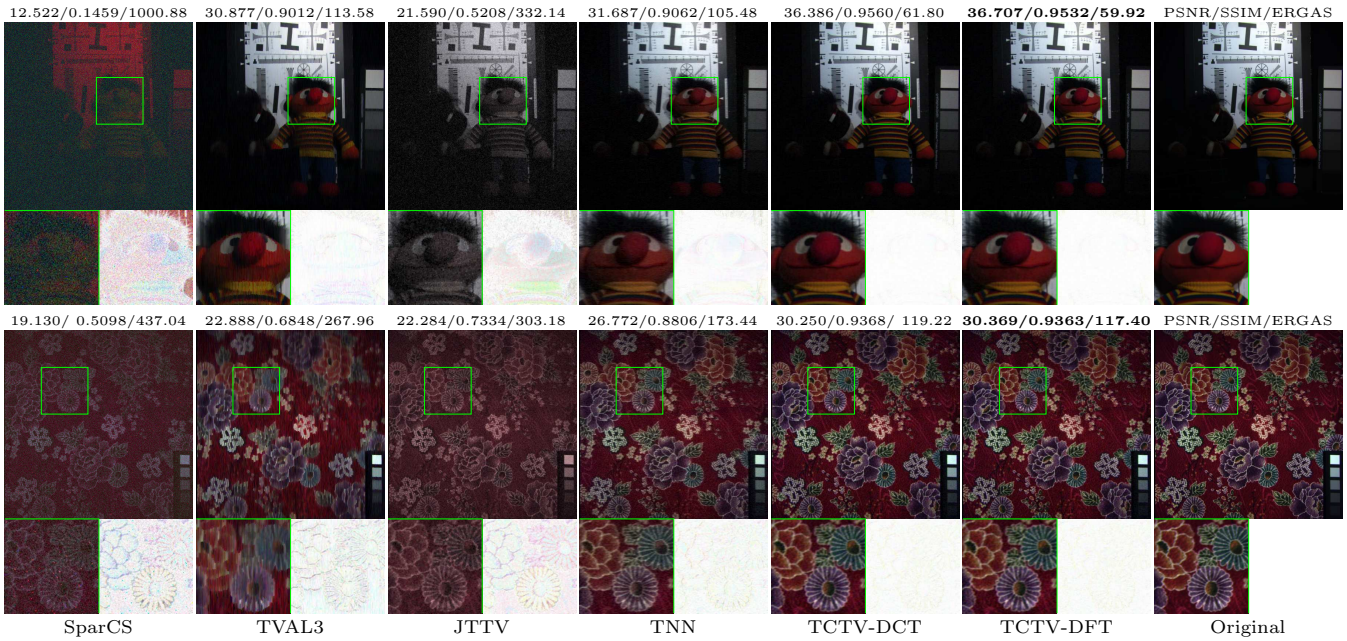

Fig. 9: Visual comparisons of the proposed approaches over four other competing methods on three bands (composite to an RGB image, R:23, G:13, B:4) of both MSI datasets under SR=1%. Up: toy, Down: cloth.