Tensor Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Tensors via Convex Optimization

Canyi Lu[1], Jiashi Feng[1], Yudong Chen, Wei Liu, Zhouchen Lin[2], Shuicheng Yan[1]   [1]National University of Singapore, [2] Peking University

Robust PCA and Tensor RPCA

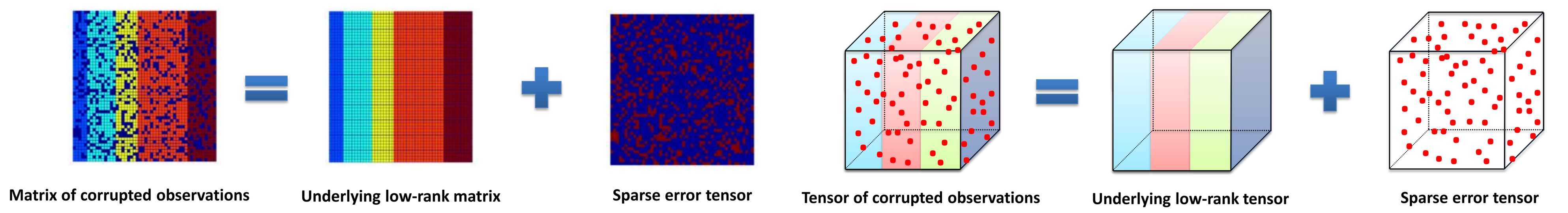
Robust PCA by convex optimization

$$\min_{L,S} \|L\|_* + \lambda\|S\|_1, \text{ s.t. } X = L + S.$$

(Candès, et al. Robust principal component analysis? JACM, 2011).

Tensor RPCA: given $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, how to recover the underlying low-rank tensor $\mathcal{L}_0$ and sparse tensor $\mathcal{S}_0$ perfectly?

$$\min_{\mathcal{L},\mathcal{S}} \|\mathcal{L}\|_* + \lambda\|\mathcal{S}\|_1, \text{ s.t. } \mathcal{X} = \mathcal{L} + \mathcal{S}. \quad (1)$$



Notations and Preliminaries

- For $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, denote $A^{(i)}$ as the i-th frontal slice of $\mathcal{A}$.
- $\bar{\mathcal{A}} = \texttt{fft}(\mathcal{A}, [], 3)$ is obtained by discrete Fourier transformation of $\mathcal{A}$ along the 3-rd dimension.
- Block circulant matrix $\texttt{bcirc}(\mathcal{A}) = \begin{bmatrix} A^{(1)} & A^{(n_3)} & \cdots & A^{(2)} \\ A^{(2)} & A^{(1)} & \cdots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_3)} & A^{(n_3-1)} & \cdots & A^{(1)} \end{bmatrix} \in \mathbb{R}^{n_1n_3\times n_2n_3}.$
- $\bar{A} = \texttt{bdiag}(\bar{\mathcal{A}}) = \begin{bmatrix} \bar{A}^{(1)} & & & \\ & \bar{A}^{(2)} & & \\ & & \ddots & \\ & & & \bar{A}^{(n_3)} \end{bmatrix}$, $\texttt{unfold}(\mathcal{A}) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n_3)} \end{bmatrix}$, $\texttt{fold}(\texttt{unfold}(\mathcal{A})) = \mathcal{A}$.
- Key property: block circulant matrix can be block diagonalized in the Fourier domain, i.e.,

$$(F_{n_3} \otimes I_{n_1}) \cdot \texttt{bcirc}(\mathcal{A}) \cdot (F_{n_3}^{-1} \otimes I_{n_2}) = \bar{A},$$

where $F_{n_3} \in \mathbb{R}^{n_3\times n_3}$ is Discrete Fourier Transform (DFT) matrix with property $F_{n_3}^* F_{n_3} = F_{n_3} F_{n_3}^* = n_3 I$.

Tensor-Tensor Product and Tensor Singular Value Decomposition [1]

- **T-product:** For $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, $\mathcal{B} \in \mathbb{R}^{n_2\times l\times n_3}$, $\mathcal{A} * \mathcal{B} = \texttt{fold}(\texttt{bcirc}(\mathcal{A}) \cdot \texttt{unfold}(\mathcal{B})) \in \mathbb{R}^{n_1\times l\times n_3}$. $\mathcal{C} = \mathcal{A} * \mathcal{B}$ is equivalent to $\bar{C} = \bar{A}\bar{B}$.
- **Conjugate transpose $\mathcal{A}^*$:** , obtained by conjugate transposing each frontal slice of $\mathcal{A}$.
- **Identity tensor $\mathcal{I} \in \mathbb{R}^{n\times n\times n_3}$:** the first frontal slice is an identity matrix, and others are zeros.
- **Orthogonal tensor:** $\mathcal{Q} \in \mathbb{R}^{n\times n\times n_3}$ is orthogonal if $\mathcal{Q}^* * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^* = \mathcal{I}$.
- **F-diagonal Tensor:** all frontal slices are diagonal matrices.
- **Tensor SVD (T-SVD):** Let $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$. Then it can be factored as

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*,$$

where $\mathcal{U} \in \mathbb{R}^{n_1\times n_1\times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2\times n_2\times n_3}$ are orthogonal, and $\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is a f-diagonal tensor.

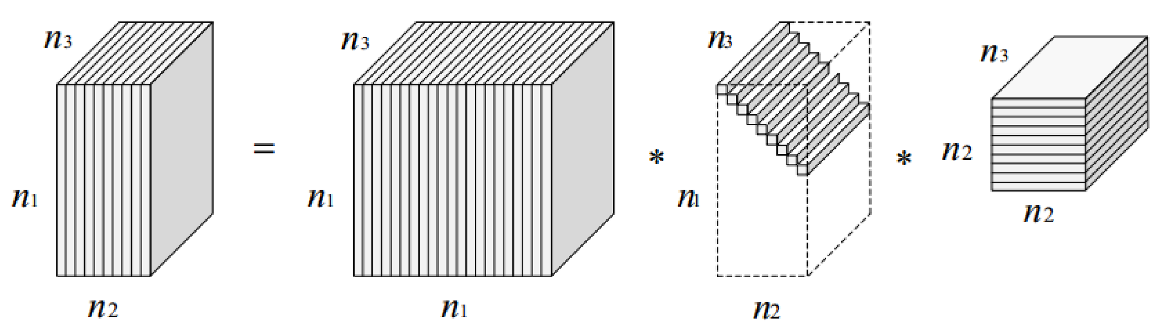


Figure: Illustration of the t-SVD of an $n_1 \times n_2 \times n_3$ tensor.

[1] Kilmer M E, Martin C D. Factorization strategies for third-order tensors. Linear Algebra and its Applications, 2011, 435(3): 641-658.

Tensor Nuclear Norm

Tensor tubal rank: the number of nonzero singular tubes of $\mathcal{S}$ in t-SVD.

denoted as $\text{rank}_t(\mathcal{A}) = \#\{i : \mathcal{S}(i,i,:) \neq \mathbf{0}\} = \max_i \text{rank}(\bar{A}^{(i)})$.

Tensor average rank: $\text{rank}_a(\mathcal{A}) = \frac{1}{n_3}\text{rank}(\texttt{bcirc}(\mathcal{A})) = \frac{1}{n_3}\text{rank}(\bar{A})$.
Note that $\text{rank}_a(\mathcal{A}) \leq \text{rank}_t(\mathcal{A})$.
Tensor nuclear norm: $\|\mathcal{A}\|_* = \frac{1}{n_3}\|\texttt{bcirc}(\mathcal{A})\|_* = \frac{1}{n_3}\|\bar{A}\|_*$.
Tensor spectral norm: $\|\mathcal{A}\| = \|\texttt{bcirc}(\mathcal{A})\| = \|\bar{A}\|$.
Theorem: On the set $\{\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3} | \|\mathcal{A}\| \leq 1\}$, the convex envelop of the tensor average rank $\text{rank}_a(\mathcal{A})$ is the tensor nuclear norm $\|\mathcal{A}\|_*$.

Main Result: Exact Recovery Guarantee

- Assume that $\mathcal{X} = \mathcal{L}_0 + \mathcal{S}_0$, where $\mathcal{L}_0$ has low tubal rank and $\mathcal{S}$ is sparse. How to perfectly recover both $\mathcal{L}_0$ and $\mathcal{S}_0$ from $\mathcal{X}$?
- **Tensor Incoherence Conditions.** For $\mathcal{L}_0 \in \mathbb{R}^{n_1\times n_2\times n_3}$, assume that $\text{rank}_t(\mathcal{L}_0) = r$ and it has the skinny t-SVD $\mathcal{L}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$, where $\mathcal{U} \in \mathbb{R}^{n_1\times r\times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2\times r\times n_3}$ satisfy $\mathcal{U}^* * \mathcal{U} = \mathcal{I}$ and $\mathcal{V}^* * \mathcal{V} = \mathcal{I}$, and $\mathcal{S} \in \mathbb{R}^{r\times r\times n_3}$ is f-diagonal. Then $\mathcal{L}_0$ is said to satisfy the tensor incoherence conditions with parameter μ if

$$\max_{i=1,\cdots,n_1} \|\mathcal{U}(i,:,:)\|_F \leq \sqrt{\frac{\mu r}{n_1 n_3}}, \quad (2)$$

$$\max_{j=1,\cdots,n_2} \|\mathcal{V}(j,:,:)\|_F \leq \sqrt{\frac{\mu r}{n_2 n_3}}, \quad (3)$$

$$\|\mathcal{U} * \mathcal{V}^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2 n_3^2}}. \quad (4)$$

Define $n_{(1)} = \max(n_1, n_2)$ and $n_{(2)} = \min(n_1, n_2)$. Then $n_3 \leq \mu \leq \frac{n_{(1)}n_3}{r}$.
- **Theorem:** Suppose $\mathcal{L}_0 \in \mathbb{R}^{n\times n\times n_3}$ obeys (2)-(4). Fix any $n \times n \times n_3$ tensor $\mathcal{M}$ of signs. Suppose that the support set $\boldsymbol{\Omega}$ of $\mathcal{S}_0$ is uniformly distributed among all sets of cardinality m, and that $\text{sgn}([\mathcal{S}_0]_{ijk}) = [\mathcal{M}]_{ijk}$ for all $(i,j,k) \in \boldsymbol{\Omega}$. Then, there exist universal constants $c_1, c_2 > 0$ such that with probability at least $1 - c_1 n^{-c_2}$ (over the choice of support of $\mathcal{S}_0$), $\{\mathcal{L}_0, \mathcal{S}_0\}$ is the unique minimizer to (1) with $\lambda = 1/\sqrt{nn_3}$, provided that

$$\text{rank}_t(\mathcal{L}_0) \leq \frac{\rho_r n n_3}{\mu(\log(nn_3))^2} \text{ and } m \leq \rho_s n^2 n_3,$$

where ρ_r and ρ_s are positive constants. If $\mathcal{L}_0 \in \mathbb{R}^{n_1\times n_2\times n_3}$ has rectangular frontal slices, TRPCA with $\lambda = 1/\sqrt{n_{(1)}n_3}$ succeeds with probability at least $1 - c_1 n_{(1)}^{-c_2}$, provided that $\text{rank}_t(\mathcal{L}_0) \leq \frac{\rho_r n_{(2)} n_3}{\mu(\log(n_{(1)}n_3))^2}$ and $m \leq \rho_s n_1 n_2 n_3$.
- **Remarks:** (1) The perfect recovery is guaranteed with high probability for $\text{rank}_t(\mathcal{L}_0)$ on the order of $nn_3/(\mu(\log nn_3)^2)$ and $\|\mathcal{S}_0\|_0$ on the order of n^2n_3. (2) When $n_3 = 1$, TRPCA and its recovery guarantee reduce to RPCA.

Optimization by ADMM

- The main per-iteration cost lies in the update of $\mathcal{L}_{k+1}$, which requires computing FFT and n_3 SVDs of $n_1 \times n_2$ matrices.
- The per-iteration complexity of ADMM for TRPCA is $O\left(n_1n_2n_3\log(n_3) + n_{(1)}n_{(2)}^2n_3\right)$.

$$\operatorname*{argmin}_{\mathcal{L}} \lambda\|\mathcal{L}\|_* + \frac{1}{2}\|\mathcal{L} - \mathcal{A}\|_F^2 = \operatorname*{argmin}_{\mathcal{L}} \frac{\lambda}{n_3}\|\bar{L}\|_* + \frac{1}{2n_3}\|\bar{L} - \bar{A}\|_F^2$$

$$\Leftrightarrow \operatorname*{argmin}_{\{\bar{L}^{(i)}\}_{i=1}^{n_3}} \lambda\|\bar{L}^{(i)}\|_* + \frac{1}{2}\|\bar{L}^{(i)} - \bar{A}^{(i)}\|_F^2$$

Experiment: Low-rank Tensor Recovery on Random Data

- Set $\lambda = 1/\sqrt{n_{(1)}n_3}$ in all experiments.
- Generate $\mathcal{L}_0 = \mathcal{P} * \mathcal{Q}$, $\mathcal{P} \in \mathbb{R}^{n\times r\times n}$, $\mathcal{Q} \in \mathbb{R}^{r\times n\times n} \sim \mathcal{N}(0, 1/n)$.
- The support set $\boldsymbol{\Omega}$ (with size m) of $\mathcal{S}_0$ is chosen uniformly at random.

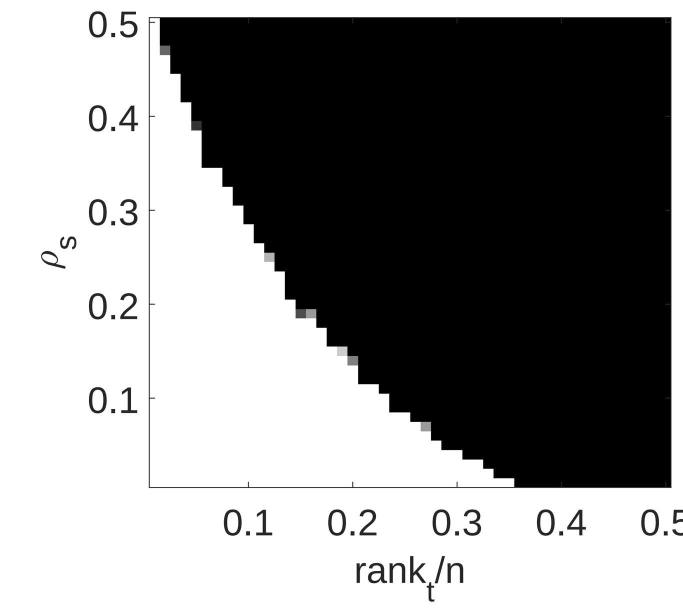
Table: Correct recovery for random problems of varying size.

$r = \text{rank}_t(\mathcal{L}_0) = 0.1n$, $m = \|\mathcal{S}_0\|_0 = 0.1n^3$

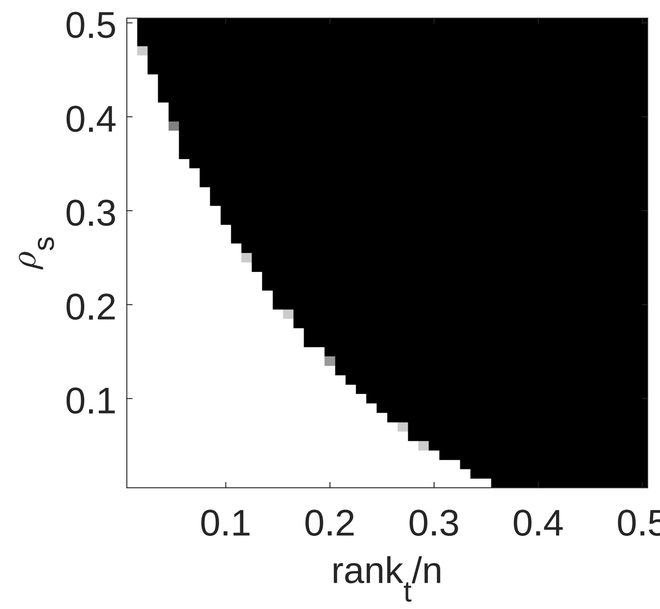
n	r	m	$\text{rank}_t(\hat{\mathcal{L}})$	$\|\hat{\mathcal{S}}\|_0$	$\frac{\|\hat{\mathcal{L}}-\mathcal{L}_0\|_F}{\|\mathcal{L}_0\|_F}$	$\frac{\|\hat{\mathcal{S}}-\mathcal{S}_0\|_F}{\|\mathcal{S}_0\|_F}$
100	10	1e5	10	101,952	4.8e−7	1.8e−9
200	20	8e5	20	815,804	4.9e−7	9.3e−10
300	30	27e5	30	2,761,566	1.3e−6	1.5e−9

$r = \text{rank}_t(\mathcal{L}_0) = 0.1n$, $m = \|\mathcal{S}_0\|_0 = 0.2n^3$

n	r	m	$\text{rank}_t(\hat{\mathcal{L}})$	$\|\hat{\mathcal{S}}\|_0$	$\frac{\|\hat{\mathcal{L}}-\mathcal{L}_0\|_F}{\|\mathcal{L}_0\|_F}$	$\frac{\|\hat{\mathcal{S}}-\mathcal{S}_0\|_F}{\|\mathcal{S}_0\|_F}$
100	10	2e5	10	200,056	7.7e−7	4.1e−9
200	20	16e5	20	1,601,008	1.2e−6	3.1e−9
300	30	54e5	30	5,406,449	2.0e−6	3.5e−9



(a) $n_1 = n_2 = 100$, $n_3 = 50$



(b) $n_1 = n_2 = 200$, $n_3 = 50$

Figure: Correct recovery for varying rank and sparsity. Fraction of correct recoveries across 10 trials, as a function of $\text{rank}_t(\mathcal{L}_0)$ (x-axis) and sparsity of $\mathcal{S}_0$ (y-axis).

Experiment: TRPCA for Image Recovery

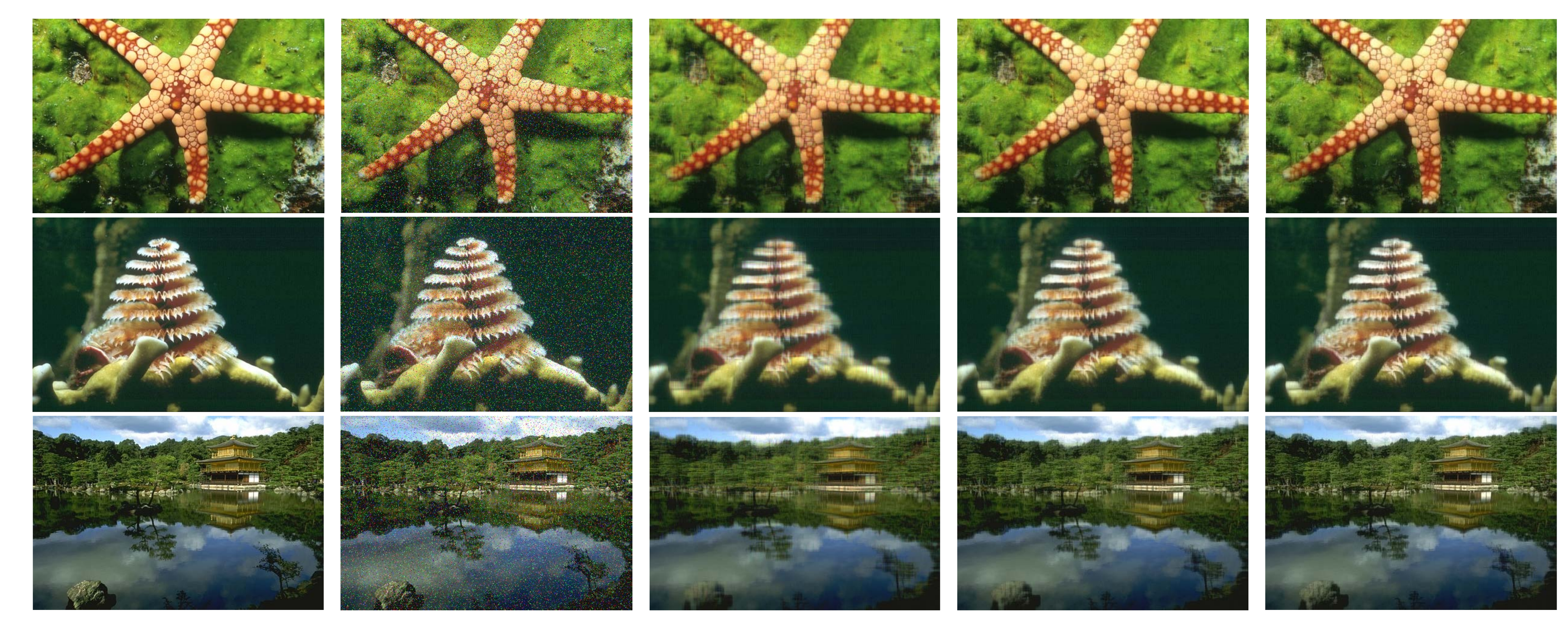


Figure: From left to right: original image; noisy image; recovered images by RPCA, SNN (Sum of Nuclear Norms) and TRPCA, respectively.

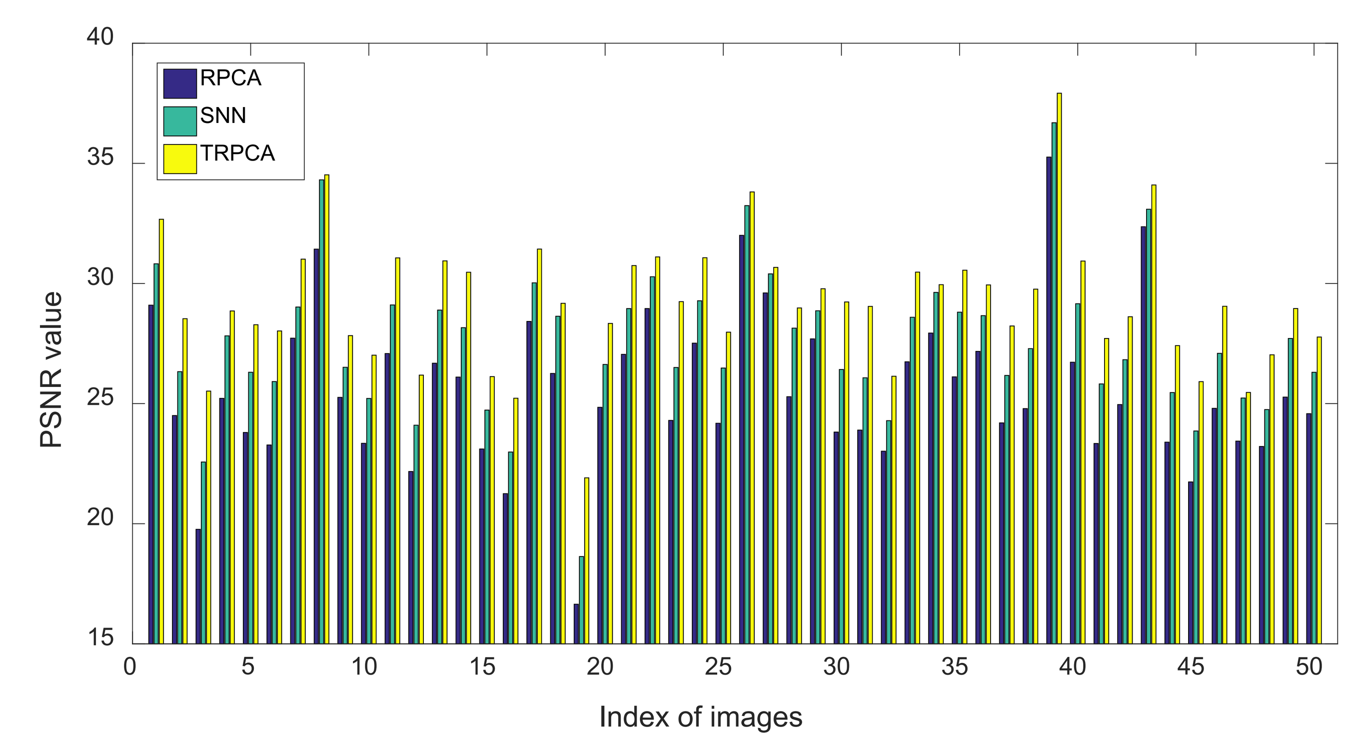


Figure: Comparison on the PSNR values for image denoising on 50 images.