

Tensor Optimization Models and Algorithms based on Tucker Decomposition and their Applications in Data Completion

Gong Wenwu

Supervisor: Prof. Yang Lili
Major: Mathematics

Department of Statistics and Data Science

Southern University of Science and Technology

2024.05.12



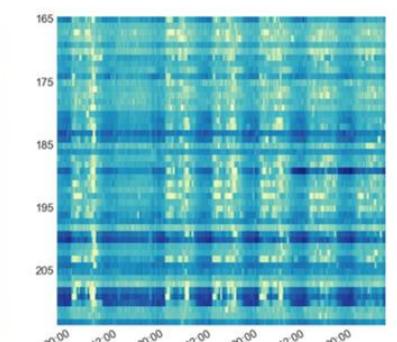
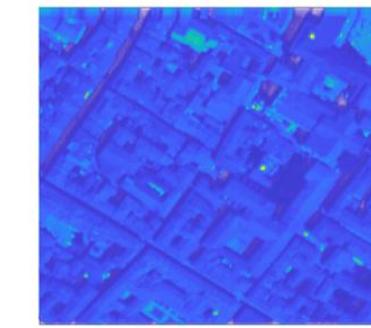
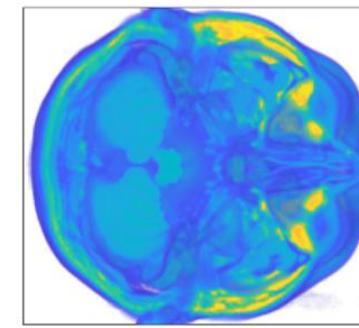
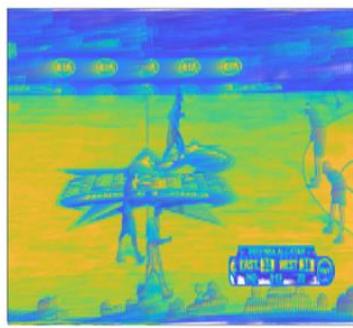
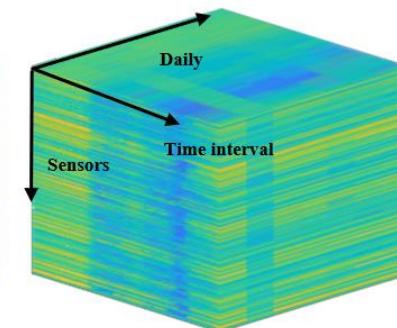
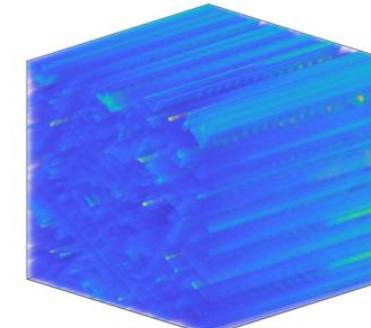
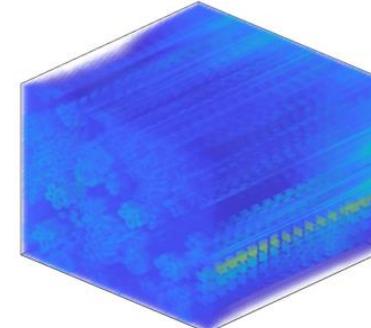
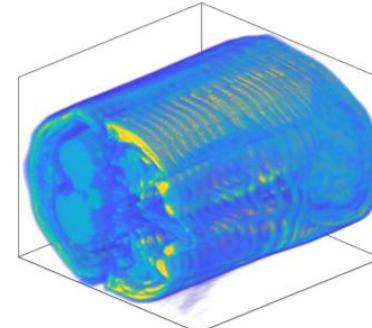
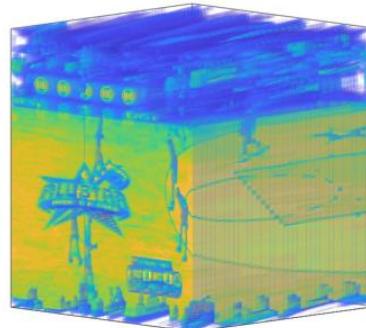
Content

- **Introduction**
- **Literature Review**
- **Proposed Models**
 - **Sparsity-based Tucker decomposition model (SparsityTD)**
 - **Enhanced low rankness and smoothness priors Tucker decomposition (ELST)**
 - **Gradient-based Tucker decomposition model (GradientTD)**
- **Conclusion**

■ Background

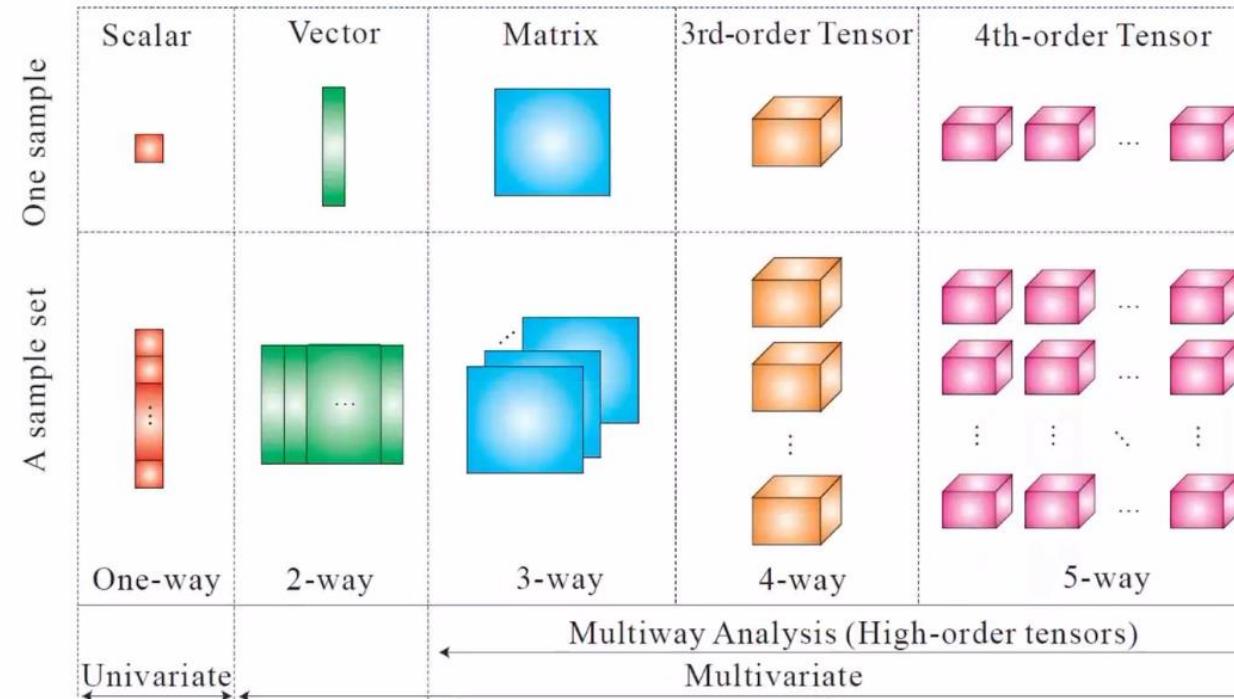


In **image and signal processing**, the data one acquires tends to be high dimensional and deals with high-dimensional visual data and multi-dimensional traffic data is a significant challenge



There is **global correlation and local similarity** between the high dimensional data

□ **Tensor structure:** the multidimensional array^[1] (N -dimensional, $N \geq 3$) collection of numerical values, is an extension of matrix structure



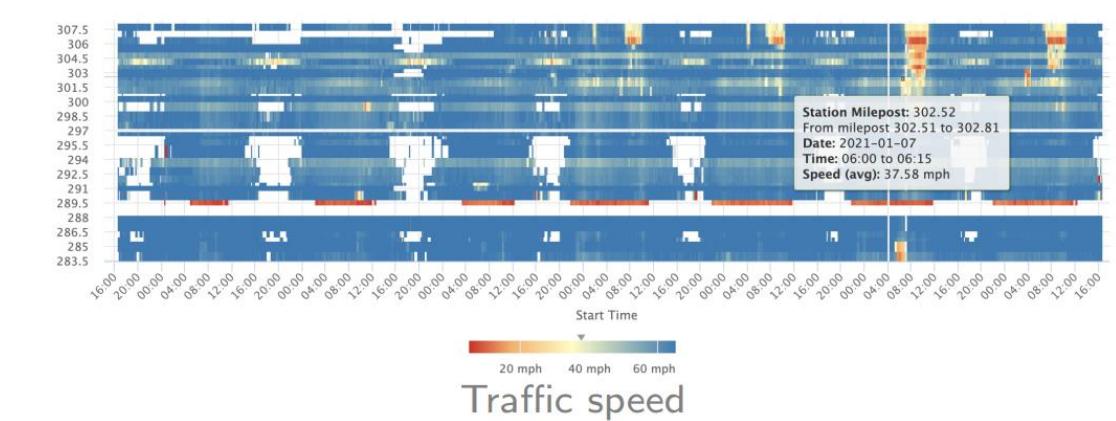
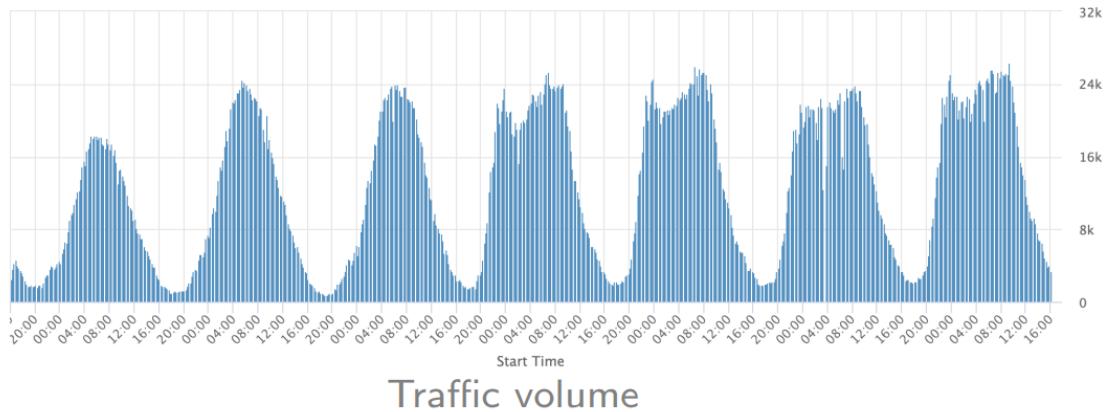
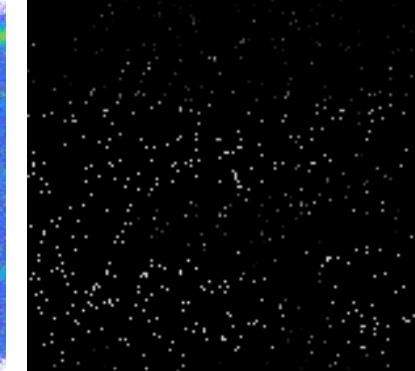
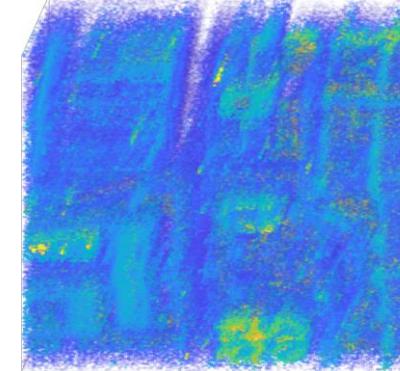
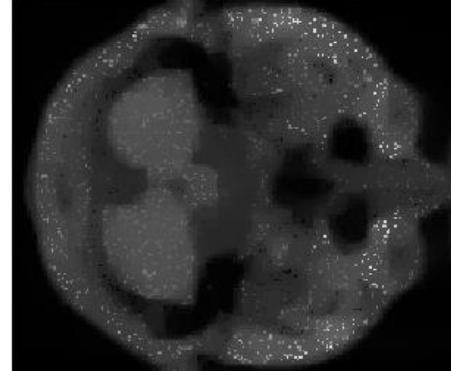
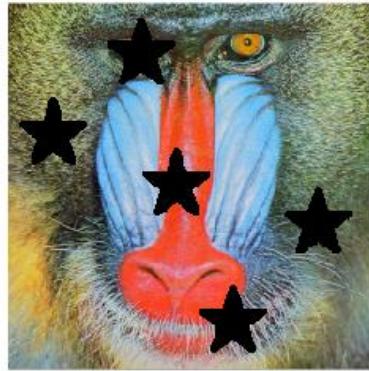
Unlike traditional matrix representation, tensor

Tensor Learning !

- provides **more complicated structural information** to represent high-dimensional data
- allows for a **more comprehensive characterization of the data's redundancies and correlations**
- avoids the problem of the curse of dimensionality

[1] Song, Qing Quan, Han Cheng Ge, James Caverlee, and Xia Hu. 2019. "Tensor Completion Algorithms in Big Data Analytics." ACM Transactions on Knowledge Discovery from Data.

- **Tensorial data missing:** the collected tensor exhibits degradation and incompleteness due to **information loss or sensor failure**



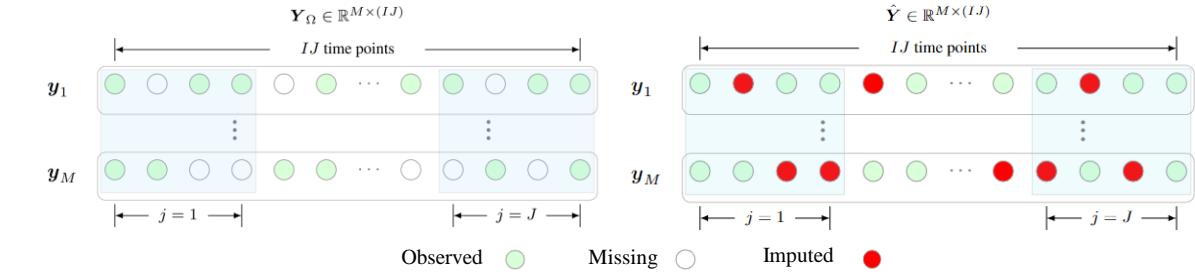
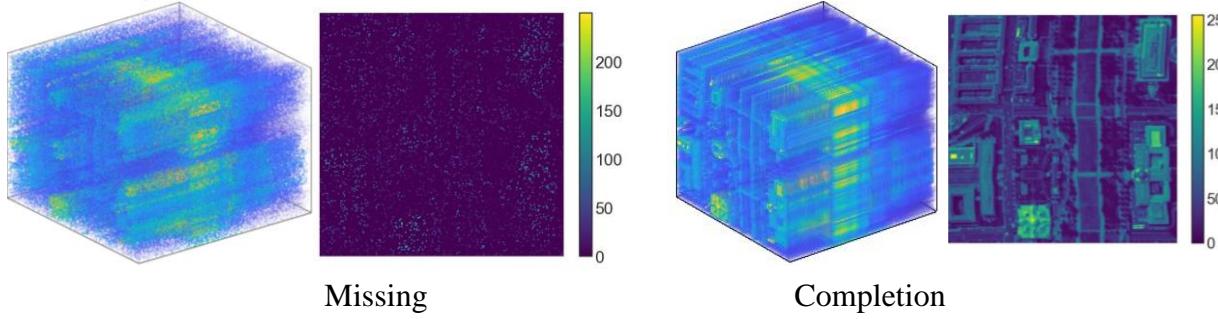
Tensor Completion (TC) !

TC aims to estimate high-quality data through degraded high-dimensional data

■ Problem Formulation



□ **Task:** high-dimensional visual data completion and multi-dimensional traffic data imputation



Task A: high-dimensional visual data completion

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{T}\|_F^2, \quad \text{s. t., } \mathcal{X}_\Omega = \mathcal{T}_\Omega$$
$$\mathcal{X}, \mathcal{T}, \Omega \in R^{I_1 \times \dots \times I_N}$$

Task B: multi-dimensional traffic data imputation

Task C: extreme missing tensorial data completion

Motivations



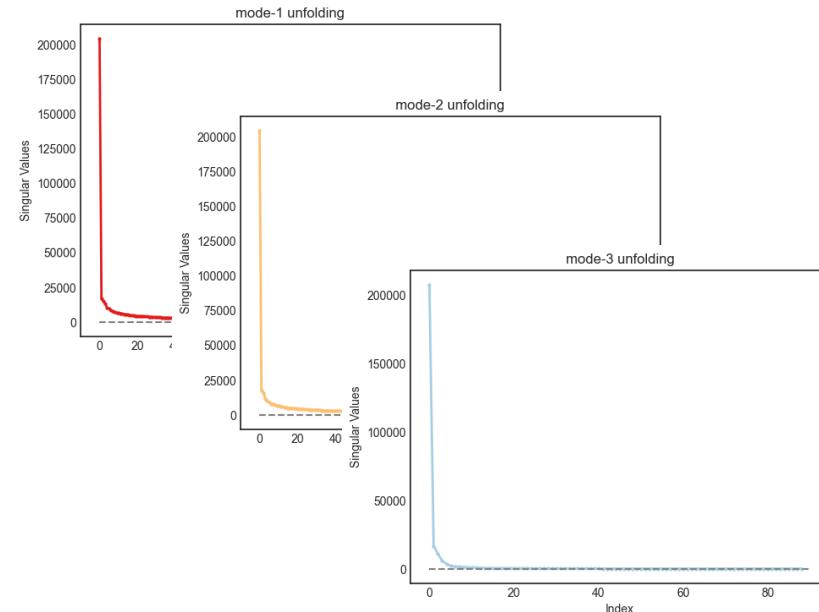
■ **Prior modeling:** the TC problem is expressed as the following maximum a posteriori (MAP) model

$$\begin{aligned}\hat{\mathcal{X}} &= \operatorname{argmax}_{\mathcal{X}} P(\mathcal{X} | \mathcal{T}) = \operatorname{argmax}_{\mathcal{X}} \frac{P(\mathcal{T} | \mathcal{X})P(\mathcal{X})}{P(\mathcal{T})} \\ &= \operatorname{argmax}_{\mathcal{X}} \{\log P(\mathcal{T} | \mathcal{X}) + \log P(\mathcal{X})\}.\end{aligned}$$

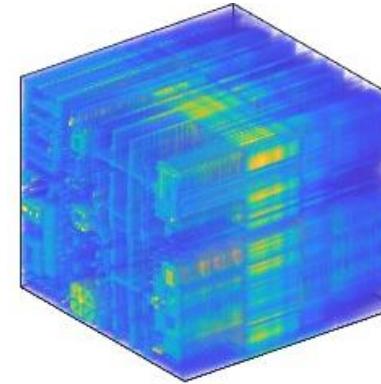


$$\min_{\mathcal{X}} \mathcal{R}(\mathcal{X}), \quad \text{s.t., } \mathcal{X}_{\Omega} = \mathcal{T}_{\Omega}$$

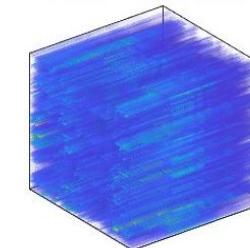
$\mathcal{R}(\mathcal{X}) = -\log P(\mathcal{X})$ the intrinsic structure of tensors



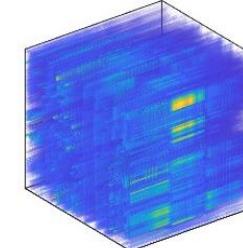
low rankness (L) captures the global correlations



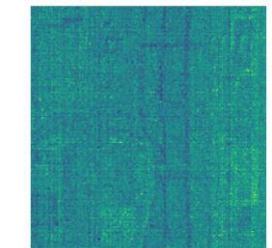
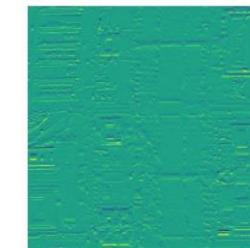
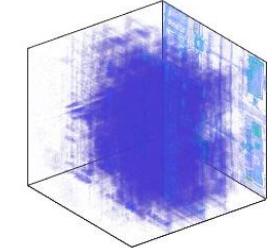
mode-1 difference



mode-2 difference



mode-3 difference

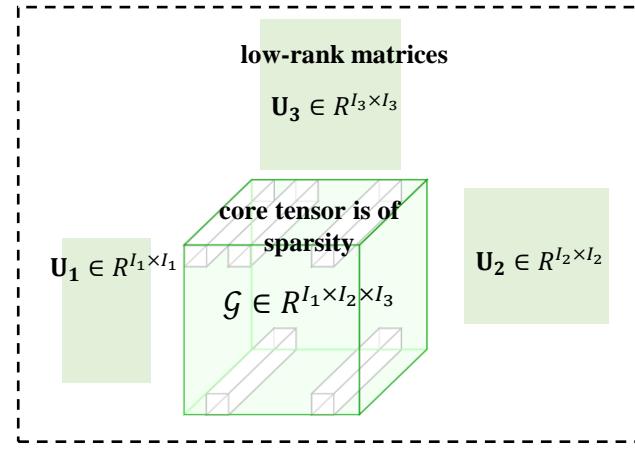


smoothness (S) characterizes local similarity

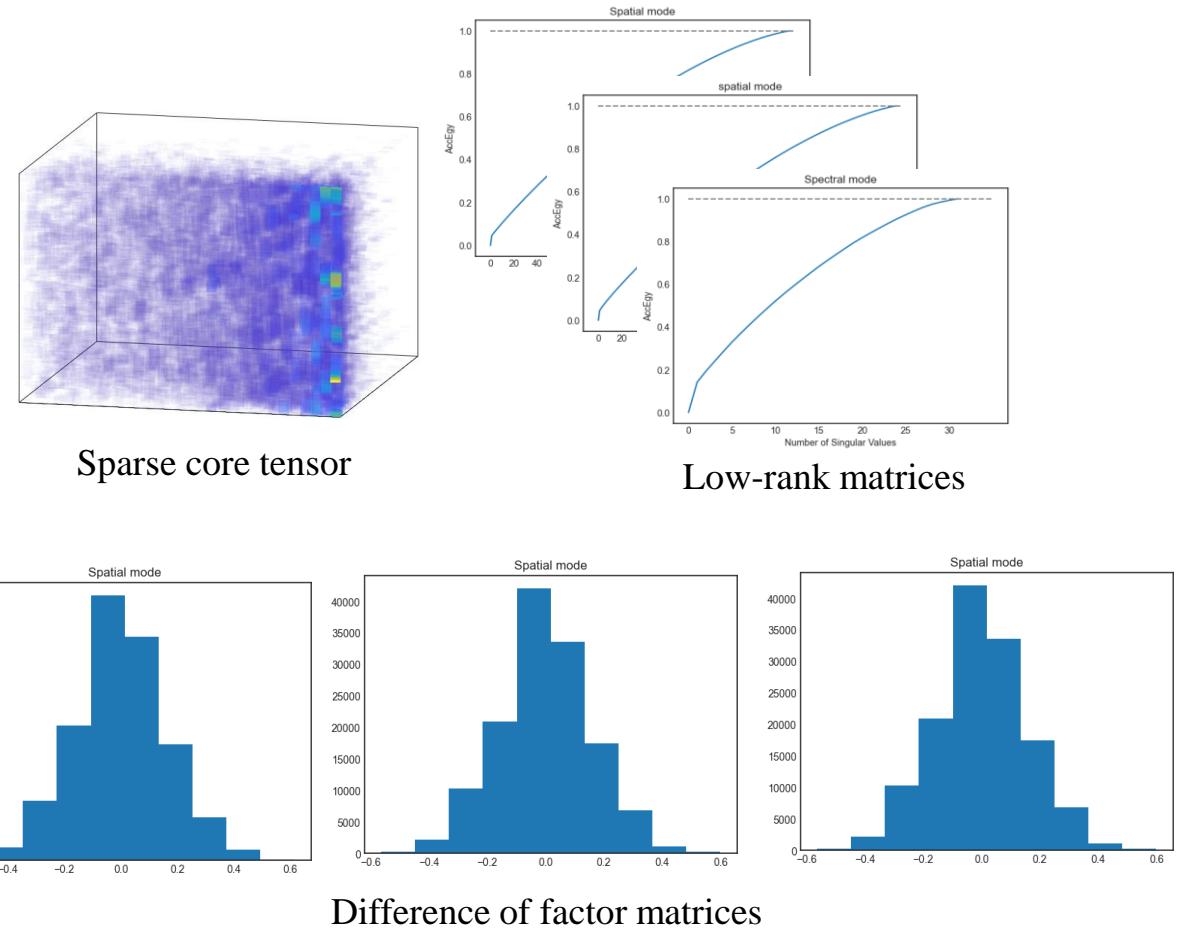
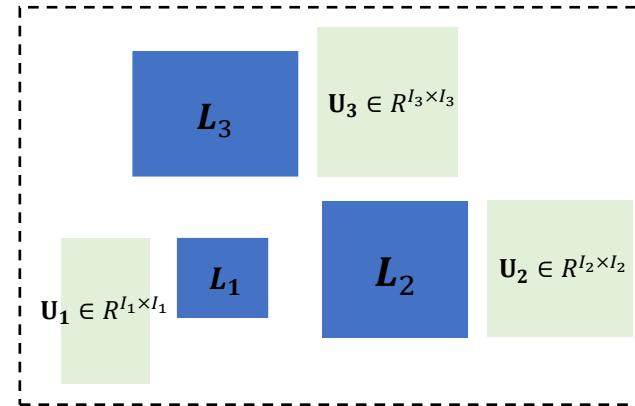
Joint Low-rank and Smooth Tensor Completion !

□ Tucker-based TC model captures low-rank structures while preserving smooth structure in a latent space

Low rankness



Smoothness

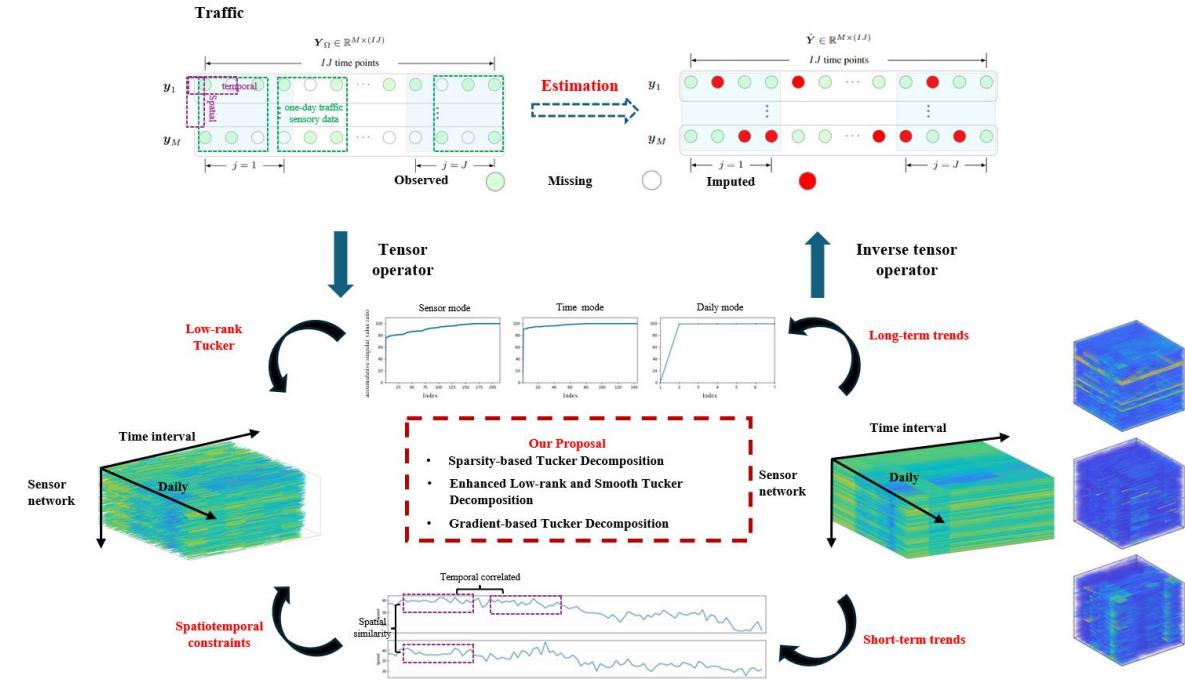
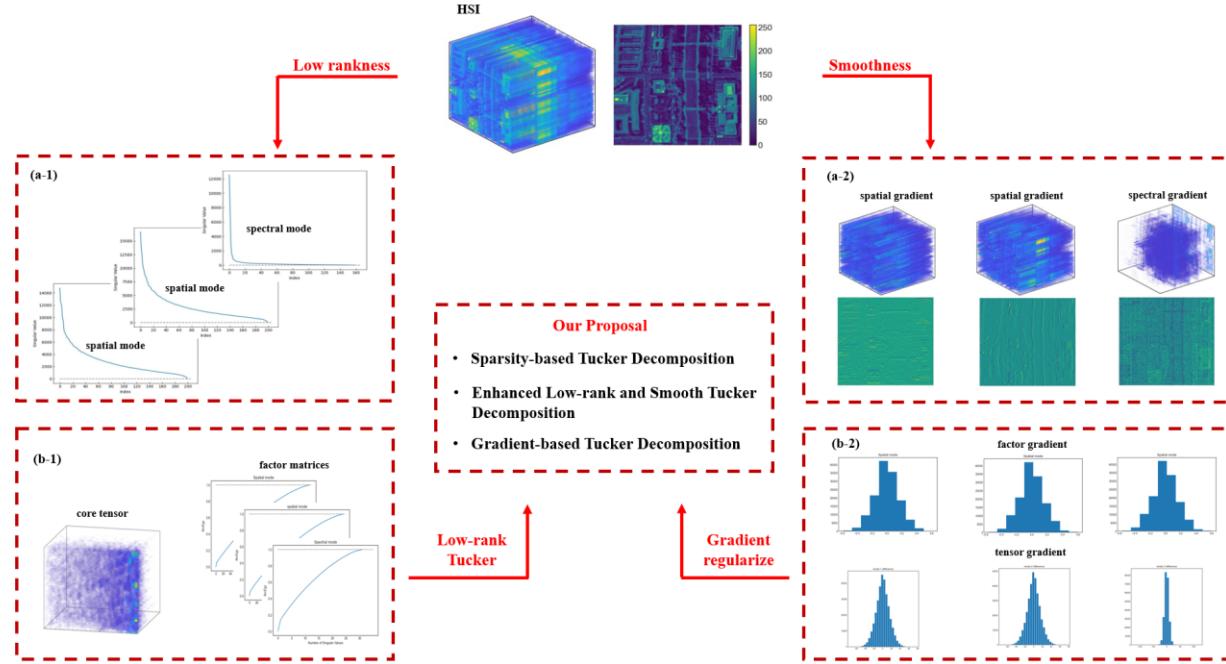


Enhanced Tucker Decomposition Models !

■ Research Content



□ **Research Problems:** develop novel Tucker-based TC models and algorithms for high-dimensional visual data completion and multi-dimensional traffic data imputation



□ **Research Goals:**

1. Development of enhanced Tucker decomposition
2. Establishment of TC optimization models
3. Design of high-performance and convergent algorithms



Content

➤ Introduction

➤ Literature Review

➤ Proposed Models

- Sparsity-based Tucker decomposition model (SparsityTD)
- Enhanced low rankness and smoothness priors Tucker decomposition (ELST)
- Gradient-based Tucker decomposition model (GradientTD)

➤ Conclusion

□ Low-rank tensor completion (LRTC): 1) minimizing the tensor rank (TRM)

➤ TRM lies in approximating the tensor rank

$$\min_{\mathcal{X}} \text{rank}(\mathcal{X}), \quad \text{s.t., } \mathcal{X}_{\Omega} = \mathcal{T}_{\Omega}$$

which includes the summation of nuclear norm minimization [2] and its nonconvex relaxation [3], parallel matrix factorization [4], tubal rank [5], and sparsity measure [6]

$$\begin{aligned} \text{rank}_{\text{nn}}(\mathcal{X}) &= \sum_{n=1}^N \alpha_n \|\mathbf{X}_{(n)}\|_* & \text{rank}_{\text{nonconvex}}(\mathcal{X}) &= \sum_{i=1}^{r_n} P_{\lambda} \left(\sigma_i(\mathbf{x}_{(n)}) \right) & \text{rank}_{\text{tmac}}(\mathcal{X}) &= \sum_{n=1}^N \frac{\alpha_n}{2} \|\mathbf{X}_{(n)} - \mathbf{A}_n \mathbf{Y}_n\|_F^2 \\ \text{rank}_{\text{tubal}}(\mathcal{X}) &= \#\{i: \mathcal{S}(i, i, :, \dots, :) \neq \mathbf{0}\} & \text{rank}_{\text{kbr}}(\mathcal{X}) &= (1 - \alpha) \prod_{n=1}^N \|\mathbf{X}_{(n)}\|_* + \alpha \|\mathcal{G}\|_1 \end{aligned}$$

However

1. TRM and its relaxations **disrupt tensor structure and incur high computational costs** due to unfolding matrix SVD
2. TRM-based models **cannot incorporate smooth structure**, which is crucial for the performance of TC models

[2] Liu, Ji, Przemyslaw Musalski, Peter Wonka, and Jie Ping Ye. 2009. “Tensor Completion for Estimating Missing Values in Visual Data.” IEEE Transactions on Pattern Analysis and Machine Intelligence.

[3] Cao, Wen Fei, Yao Wang, Can Yang, Xiang Yu Chang, Zhi Han, and Zong Ben Xu. 2015. “Folded-Concave Penalization Approaches to Tensor Completion.” Neurocomputing.

[4] Xu, Yang Yang, Ru Ru Hao, Wo Tao Yin, and Zhi Xun Su. 2015. “Parallel Matrix Factorization for Low-Rank Tensor Completion.” Inverse Problems and Imaging.

[5] Zhang, Ze Min, and Shuchin Aeron. 2017. “Exact Tensor Completion Using T-SVD.” IEEE Transactions on Signal Processing.

[6] Xie, Qi, Qian Zhao, De Yu Meng, and Zong Ben Xu. 2018. “Kronecker-Basis-Representation Based Tensor Sparsity and Its Applications to Tensor Recovery.” IEEE Transactions on Pattern Analysis and Machine Intelligence.

□ Low-rank tensor completion (LRTC): 2) leveraging the low-rank tensor decomposition structure (LRTA)

➤ LRTA uses the low-rank tensor structure \mathcal{H} to address the TC problem

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{H}\|_F^2, \quad \text{s.t., } \mathcal{X}_\Omega = \mathcal{T}_\Omega$$

which takes various forms such as the CP model [7], Tucker model [8], Tensor train model [9], and tensor network form [10]

$$\max_{\mathcal{X}} p(\mathcal{X}_\Omega \mid \{\mathbf{A}^n\}_{n=1}^N, \tau) \prod_{n=1}^N p(\mathbf{A}^n \mid \lambda) p(\lambda) p(\tau)$$

$$\min_{\mathcal{X}, \{\mathbf{U}_n\}, \{\mathbf{V}_n\}} \sum_{n=1}^{N-1} \frac{\alpha_n}{2} \|T(\mathcal{X})_{(n)} - \mathbf{U}_n \mathbf{V}_n\|_F^2$$

$$\min_{\mathcal{X}} \sum_{n=1}^N \alpha_n \|\mathbf{X}_{(n)}\|_* + \frac{\lambda}{2} \|\mathcal{X} - \mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_N \mathbf{U}_N\|_F^2$$

$$\min_{\mathcal{X}, \{\mathcal{G}_n\}} \frac{1}{2} \|\mathcal{X} - \text{FNTC}(\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_N)\|_F^2$$

However

1. How to **determine tensor rank**
2. How to **construct appropriate decomposition structures** to represent underlying low-rank structures

[7] Zhao, Qi Bin, Li Qing Zhang, and Andrzej Cichocki. 2015. “Bayesian CP Factorization of Incomplete Tensors with Automatic Rank Determination.” IEEE Transactions on Pattern Analysis and Machine Intelligence.

[8] Gandy, Silvia, Benjamin Recht, and Isao Yamada. 2011. “Tensor Completion and Low-N-Rank Tensor Recovery via Convex Optimization.” Inverse Problems.

[9] Bengua, Johann A., Ho N. Phien, Hoang Duong Tuan, and Minh N. Do. 2017. “Efficient Tensor Completion for Color Image and Video Recovery: Low-Rank Tensor Train.” IEEE Transactions on Image Processing.

[10] Luo, Yi Si, Xi Le Zhao, De Yu Meng, and Tai Xiang Jiang. 2022. “HLRTF: Hierarchical Low-Rank Tensor Factorization for Inverse Problems in Multi-Dimensional Imaging.” CVPR.



□ Joint low-rank and smooth tensor completion: low rankness (L) plus smoothness (S)

➤ Smooth structure is introduced to portray information about the local similarity of the tensor data [11]

SNN-based

$$\mathcal{R}(\mathcal{X}) = \frac{1}{N} \sum_{n=1}^N \|\mathbf{X}_{(n)}\|_* + \sum_{n=1}^N \beta_n |\mathbf{A}_n \mathbf{X}_{(n)}|, \text{ s.t., } \mathcal{X}_\Omega = \mathcal{T}_\Omega$$

Total Variation, TV

Quadratic Variation, QV

CP-based

$$\mathcal{R}(\mathcal{X}) = \frac{1}{2} \|\mathcal{X} - \mathcal{Z}\|_F^2 + \sum_{r=1}^R \frac{g_r^2}{2} \sum_{n=1}^N \beta_n \|\mathbf{A}_n \mathbf{u}_r^n\|_p^p, \text{ s.t. } \mathcal{Z} = \sum_{r=1}^R g_r \mathbf{u}_r^1 \circ \mathbf{u}_r^2 \circ \dots \circ \mathbf{u}_r^N, \mathcal{X}_\Omega = \mathcal{T}_\Omega$$

Tucker-based

$$\mathcal{R}(\mathcal{X}) = \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \sum_{n=1}^N \frac{\beta_n}{2} \|\mathbf{A}_n \mathbf{X}_{(n)}\|_F^2, \text{ s.t., } \mathcal{G} \in \mathbb{R}^{r_1 \times \dots \times r_N}, \mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \dots \times_N \mathbf{U}_N, \mathcal{X}_\Omega = \mathcal{T}_\Omega$$

TT-based

$$\mathcal{R}(\mathcal{X}) = \sum_{n=1}^{K-1} \frac{\alpha_n}{2} \|\mathcal{K}(\mathcal{X})_{(n)} - \mathbf{U}_n \mathbf{V}_n\|_F^2 + \beta |\mathbf{A}_n \mathbf{X}_{(n)}|, \text{ s.t. } \mathcal{K}(\mathcal{X}) = \mathcal{Z} \in R^{I_1 \times I_2 \times \dots \times I_K}, \mathcal{X}_\Omega = \mathcal{T}_\Omega$$

Tubal-based

$$\mathcal{R}(\mathcal{X}) = \frac{1}{N} \sum_{n=1}^N \|\mathcal{X} \times_n \mathbf{A}_n\|_{\otimes, \Omega}, \text{ s.t. } \mathcal{X}_\Omega = \mathcal{T}_\Omega.$$

However

How to construct smooth structure and the hyper-parameter tuning of L and S is challenging



□ The related works considering joint L and S priors

TC models	Low rankness (L)		Smoothness (S)		
	TRM	LRTA	TV term	QV term	Others
QTTSRTD ^[37]		√		√	
LRSETD ^[46]	√	√	√		
TRGFR ^[43]		√		√	
tCTV ^[62]	√		√		
ATVTC ^[36]	√		√		
BayesianTD ^[39]		√		√	√
NonnagativeTD ^[96]		√		√	
SBCD ^[45]	√	√		√	√

How to more effectively combine low rankness and smoothness in Tucker decomposition models?

TC models	Low rankness (L)		Smoothness (S)		
	TRM	LRTA	TV term	QV term	Others
LATC ^[15]		√		√	
stTT ^[42]				√	√
ESP ^[44]		√			√
FATC ^[41]				√	√
TTTV ^[26]				√	√
FCP ^[79]				√	√
TTMacTV ^[88]		√		√	
SNNTV ^[24]		√	√	√	
STMac ^[92]		√			√
MacTV ^[25]		√			√
gHOI ^[50]				√	√
SPC ^[38]				√	√
BGCP ^[23]				√	√
SNTD ^[51]				√	√
stLRTC ^[1]				√	
STDC ^[47]	√	√			√



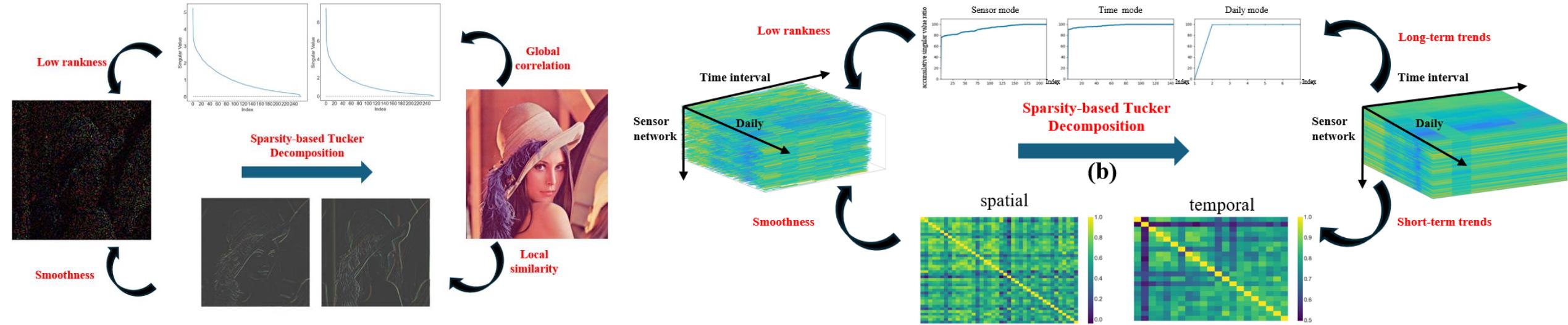
□ Research gaps

- Joint low-rank and smooth Tucker decomposition
- Traditional Tucker decomposition methods **requiring pre-given rank**
- Designing effective regularization to **portray smoothness**
- Development of **high-performance and convergent algorithms**
- Tucker-based TC optimization for **multi-dimensional traffic data imputation**

Content

- Introduction
- Literature Review
- Proposed Models
 - Sparsity-based Tucker decomposition model (SparsityTD)
 - Enhanced low rankness and smoothness priors Tucker decomposition (ELST)
 - Gradient-based Tucker decomposition model (GradientTD)
- Conclusion

■ Sparsity-based Tucker Decomposition



$$\min_{\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}} \alpha \|\mathcal{G}\|_1 + \sum_{n=1}^{\Gamma} \frac{\beta_n}{2} \text{tr}(\mathbf{U}_n^T \mathbf{L}_n \mathbf{U}_n) + \sum_{n=\Gamma+1}^N \frac{\beta_n}{2} \|\mathbf{U}_n \mathbf{T}_n\|_F^2$$

$$\text{s.t., } \mathcal{X} = \mathcal{G} \times_{n=1}^N \mathbf{U}_n, \quad \mathbf{U}_n \in \mathbb{R}_+^{I_n \times I_n}, n = 1, \dots, N, \quad \mathcal{X}_{\Omega} = \mathcal{T}_{\Omega}$$

➤ Proposed algorithm

We use the proximal gradient to transform the objective function into multiple solvable subproblems

$$\min_{\mathcal{G}, \{\mathbf{U}_n \in \mathbb{R}_+^{I_n \times I_n}\}, \mathcal{X}} \mathbb{F}(\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}) \equiv \frac{1}{2} \left\| \mathcal{X} - \mathcal{G} \underset{n=1}{\overset{N}{\times}} \mathbf{U}_n \right\|_F^2 + \alpha \|\mathcal{G}\|_1 + \sum_{n=1}^{\Gamma} \frac{\beta_n}{2} \text{tr}(\mathbf{U}_n^T \mathbf{L}_n \mathbf{U}_n) + \sum_{n=\Gamma+1}^N \frac{\beta_n}{2} \|\mathbf{U}_n \mathbf{T}_n\|_F^2$$

Algorithm 3-1 APG-based solver for the SparsityTD model

1: **Input:** Missing tensor $\mathcal{T} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_N}$, Ω containing indices of observed entries, and the parameters $\alpha = 1$, $\beta_n \geq 0$, $\text{tol} = 1e^{-4}$, and $K = 300$.
 2: **Output:** Reconstructed tensor $\hat{\mathcal{X}}$.
 3: **construct** positive semi-definite similarity matrix \mathbf{W}_n and Toeplitz matrix \mathbf{T}_n ;
 4: **initialize** $\mathcal{G}^0, \mathbf{U}_n^0 \in \mathbb{R}_+^{I_n \times I_n}$ ($1 \leq n \leq N$);
 5: $\mathcal{X}_\Omega = \mathcal{T}_\Omega$, $\mathcal{X}_{\bar{\Omega}} = \text{mean}(\mathcal{T}_{\bar{\Omega}})$; Acceleration [12]
 6: **for** $k = 1$ to K **do**
 7: Optimize \mathcal{G} according to Eq. (3-14);
 8: **for** $n = 1$ to N **do**
 9: Optimize \mathbf{U}_n using Eq. (3-15);
 10: **end for**
 11: Update \mathcal{Z}^k using Eq. (3-18);
 12: Whenever $\mathbb{F}(\mathcal{G}^k, \mathbf{U}_k) < \mathbb{F}(\mathcal{G}^{k-1}, \mathbf{U}_{k-1})$, we re-update $\tilde{\mathcal{G}}, \{\tilde{\mathbf{U}}_n\}$ using Eq. (3-16) and Eq. (3-17) **until** stopping conditions (3-19) are satisfied.
 13: **end for**
 14: **return** $\hat{\mathcal{X}}_\Omega = \mathcal{T}_\Omega$, $\hat{\mathcal{X}}_{\bar{\Omega}} = \mathcal{Z}^k_{\bar{\Omega}}$.

$$\omega_k = \min\left\{\frac{t^{k-1} - 1}{t^k}, 0.999 \sqrt{\frac{L_{\mathbf{U}}^{k-1}}{L_{\mathbf{U}}^k}}\right\} \text{ for } k \geq 1$$

$$t^k = \frac{1 + \sqrt{4(t^{k-1})^2 + 1}}{2}, t^0 = 1$$

$$\mathcal{G}^{k+1} = \mathcal{S}_{\frac{\alpha}{L_{\mathcal{G}}^k}} \left(\tilde{\mathcal{G}}^k - \frac{1}{L_{\mathcal{G}}^k} \nabla_{\mathcal{G}} f(\tilde{\mathcal{G}}^k) \right), \tilde{\mathcal{G}}^k = \mathcal{G}^k + \omega_k (\mathcal{G}^k - \mathcal{G}^{k-1})$$

$$\mathbf{U}_n^{k+1} = \mathcal{P}_+ \left(\tilde{\mathbf{U}}_n^k - \frac{1}{L_{\mathbf{U}_n}^k} \nabla_{\mathbf{U}_n} \ell(\tilde{\mathbf{U}}_n^k) \right), \tilde{\mathbf{U}}_n^k = \mathbf{U}_n^k + \omega_k (\mathbf{U}_n^k - \mathbf{U}_n^{k-1})$$

$$\mathcal{X}^{k+1}_\Omega = \mathcal{T}_\Omega + \phi(\mathcal{X}^k_\Omega - \mathcal{Z}^k_\Omega), \quad \mathcal{X}^{k+1}_{\bar{\Omega}} = \mathcal{Z}^k_{\bar{\Omega}}$$

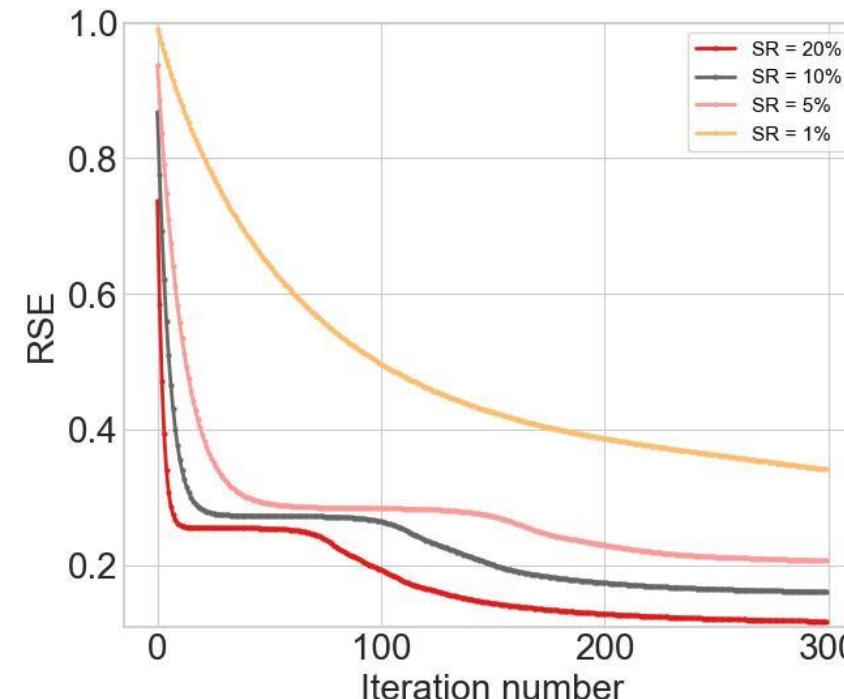
➤ Algorithm analysis

Theorem 1. If the hyper-parameters α and $\{\beta_n\}$ are non-negative, then any limit point of sequence $\Theta^k = \{\{\mathbf{U}_n^k\}, \mathcal{G}^k\}$ generated by Algorithm 3-1 is a stationary point.

Step 1, the positivity of hyper-parameters implies the boundedness of Θ^k , with bounded $L_{\mathcal{G}}, L_{\mathbf{U}}$ also existing;

$$\sum_{k=1}^{\infty} (\|\mathcal{G}^{k-1} - \mathcal{G}^k\|_F^2 + \|\mathbf{U}_n^{k-1} - \mathbf{U}_n^k\|_F^2) \leq \infty$$

Step 2, verifies the first-order optimality conditions confirms that the limit point $\hat{\Theta}$ is stationary.

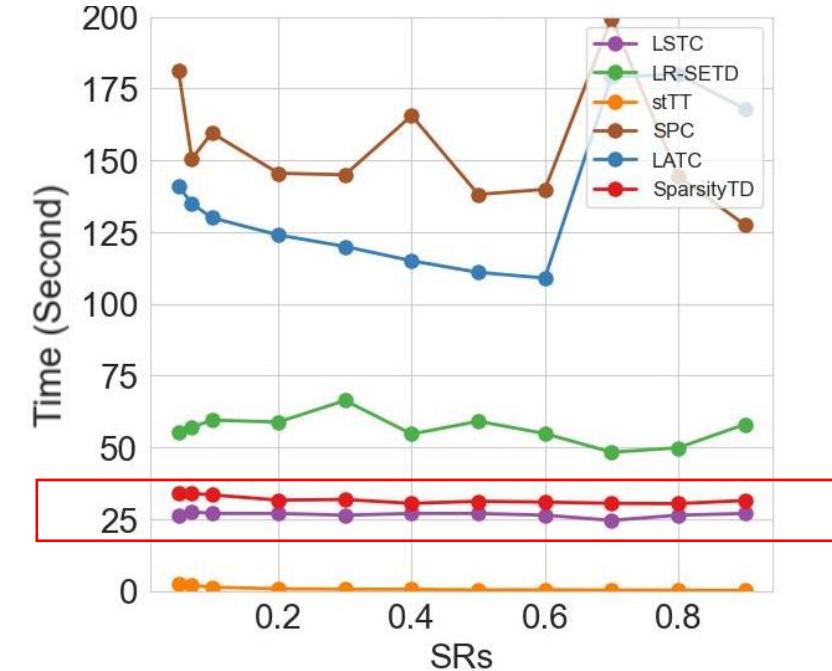
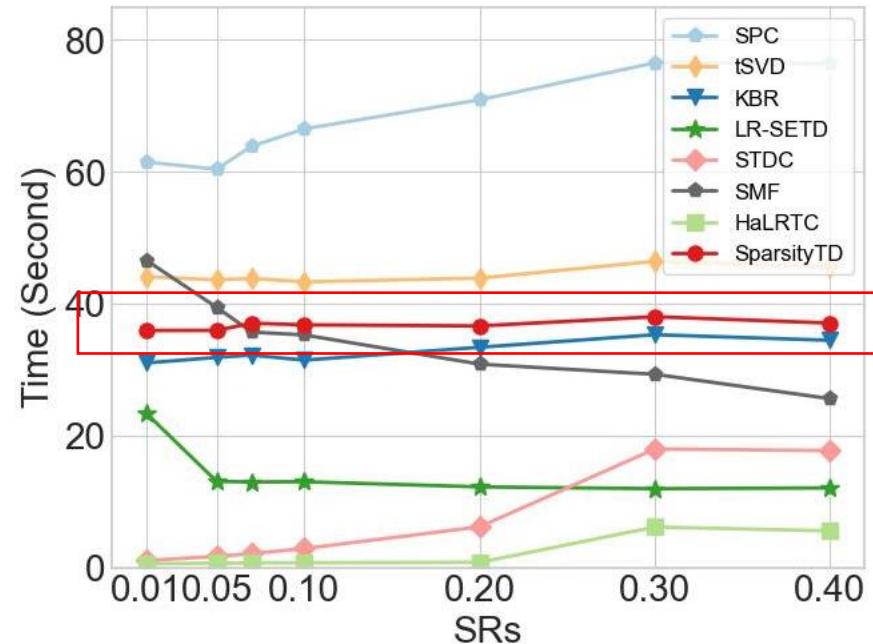


➤ Computational complexity

Theorem 2. Assuming that the APG-based algorithm converges in the K iterations, we summarize the per-iteration time complexity of the Algorithm 3-1

$$\mathcal{O} \left((N+1) \sum_{n=1}^N \left(\prod_{i=1}^n I_i \right) \left(\prod_{j=n}^N I_j \right) \right).$$

The per-iteration cost is relevant to the tensor sizes $\prod_{i=1}^n I_i$, and the proposed algorithm is theoretically efficient





➤ Numerical results

- Experimental settings: we perform both **random missing (RM)** and **structural missing (SM)** experiments on popular RGB-color images; we consider three missing scenarios, i.e., **RM**, **no-random missing (NM)**, and **black-out missing (BM)** for traffic data imputation
- Parameter settings: we set $\alpha = 1$ and choose $\phi = 0.2$. For all experiments, we calculate the parameters β_n using mode- n tensor unfolding matrices
- Performance evaluation:

$$\text{PSNR} = 10 \cdot \log_{10} \frac{(\mathcal{X}_{\max})^2}{\|\hat{\mathcal{X}} - \mathcal{X}_{\text{true}}\|_F^2 / |\bar{\Omega}|} \quad \text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad \text{RSE} = \frac{\|\hat{\mathcal{X}} - \mathcal{X}_{\text{true}}\|_F}{\|\mathcal{X}_{\text{true}}\|_F}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100, \quad \text{NMAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i|}$$

- Baselines:

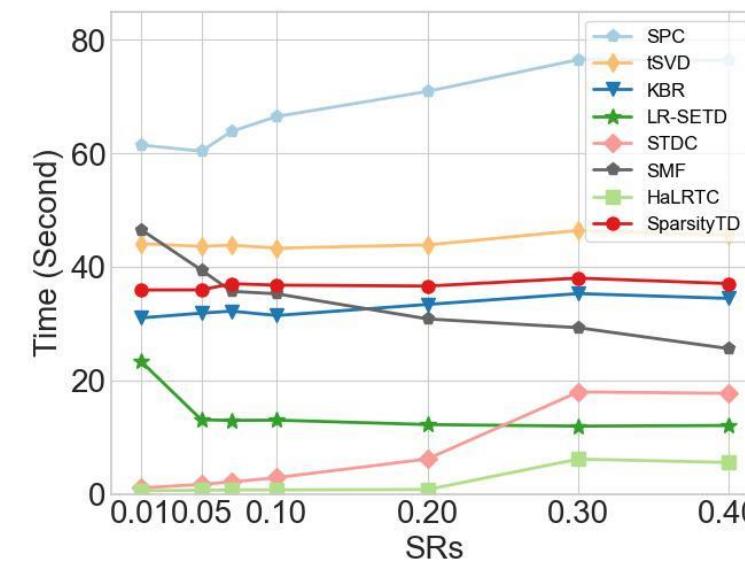
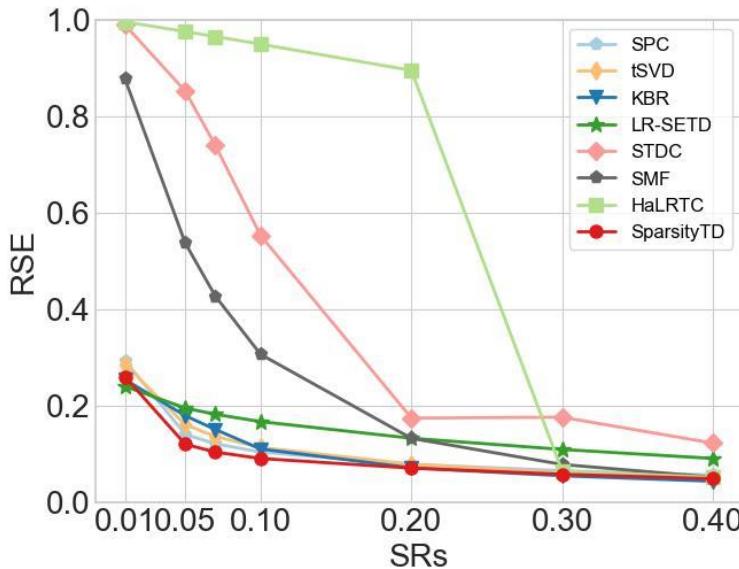
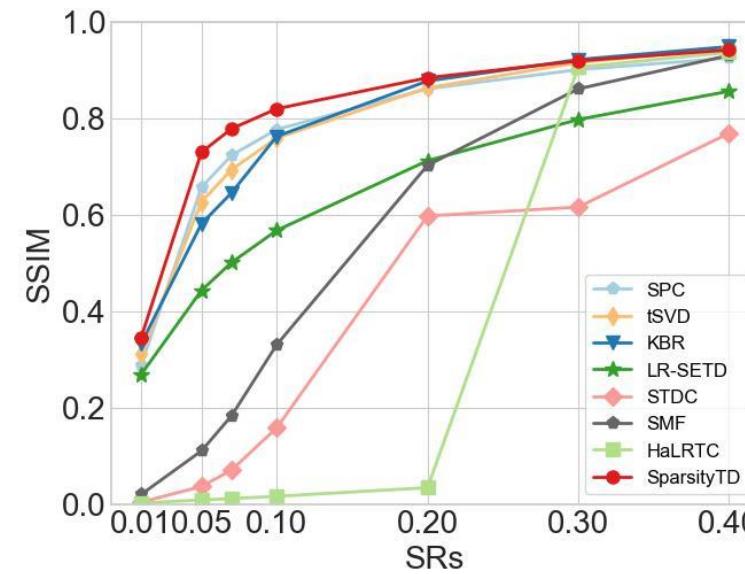
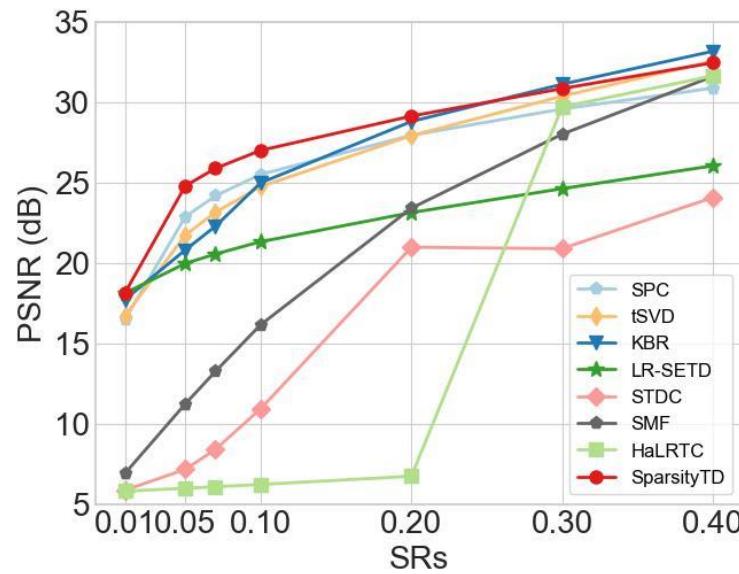
TC methods	Types of the priors			Forms
	Low rankness	Sparsity	Smoothness	
SparsityTD		✓	✓	Tucker
LR-SETD ^[46]	✓	✓	✓	Tucker
SMF ^[92]	✓		✓	Parallel Matrix
KBR ^[21]	✓	✓		Tucker
tSVD ^[34]	✓			Tubal
SPC ^[38]	✓		✓	CP
STDC ^[47]	✓		✓	Tucker
HaLRTC ^[2]	✓			Matrix

Image

Baselines	Priors			Structures
	Low-rank	Spatial	Temporal	
SparsityTD	✓	✓	✓	third-order tensor
LR-SETD ^[46]	✓		✓	third-order tensor
Hankel ^[103]	✓	✓	✓	forth-order tensor
stTT ^[42]	✓	✓	✓	third-order tensor
LATC ^[15]	✓		✓	third-order tensor
LSTC ^[89]	✓		✓	third-order tensor
SPC ^[38]	✓	✓		third-order tensor

Traffic

RM: image house random missing



HaLRTC 2013

STDC 2014

SPC 2016

tSVD 2017

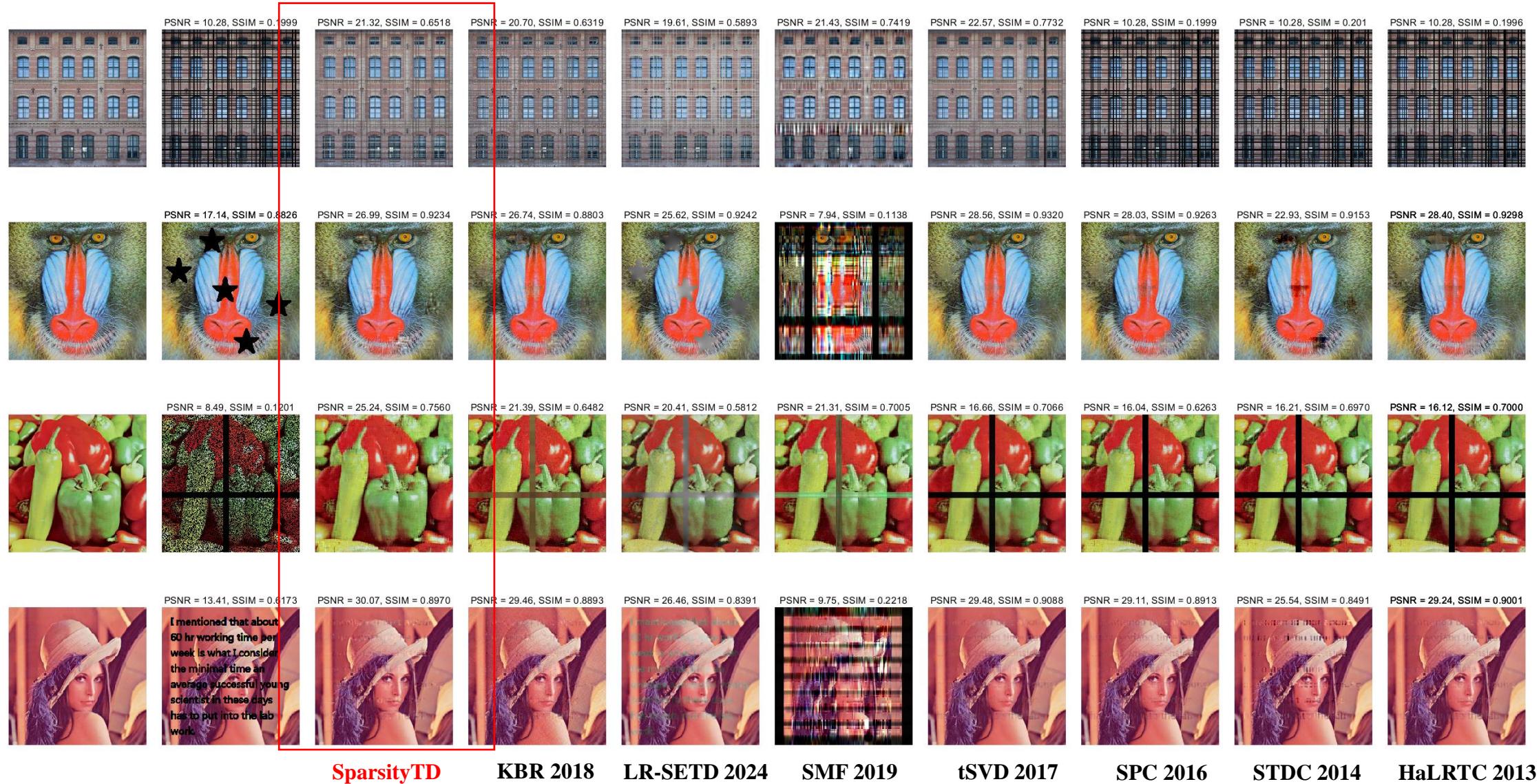
KBR 2018

SMF 2019

LR-SETD 2024

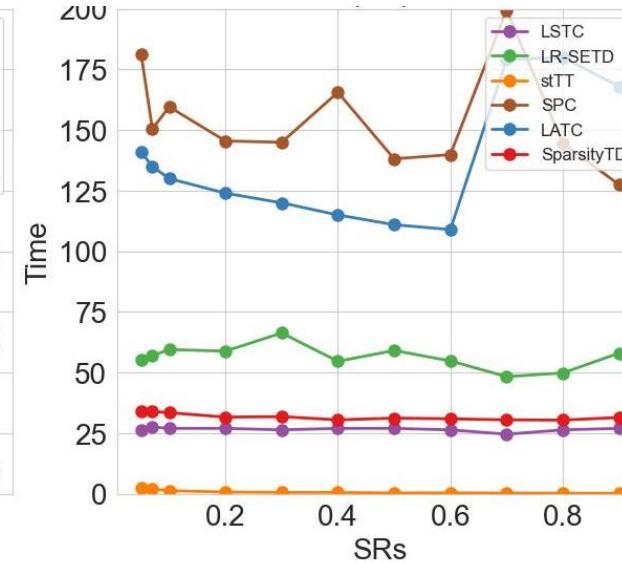
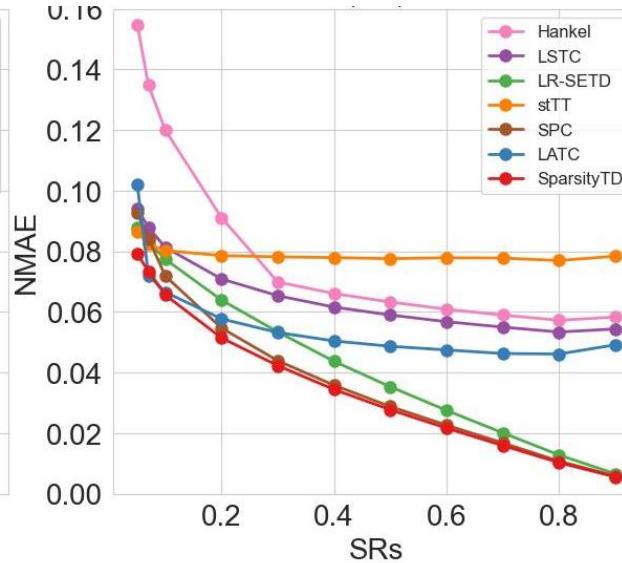
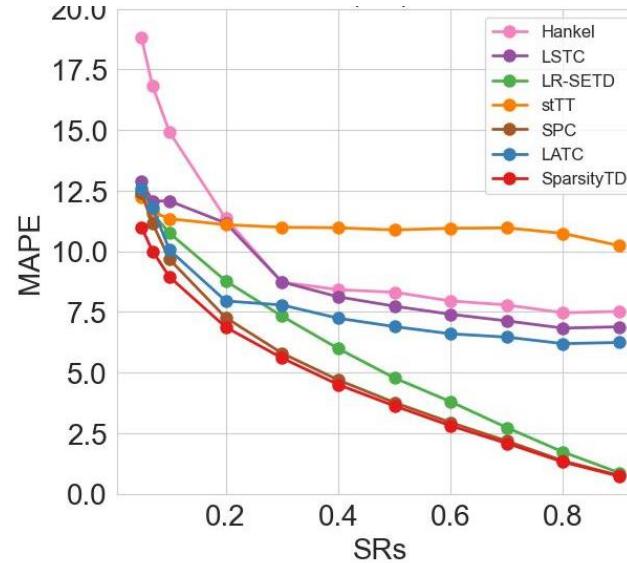


RGB image in-painting: structural missing

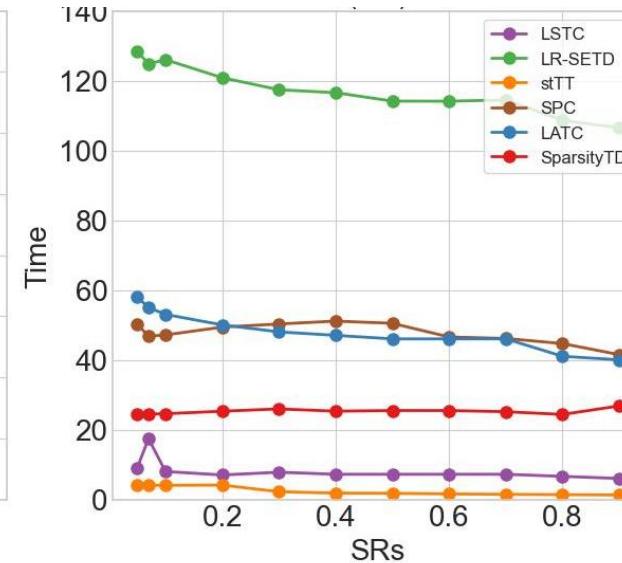
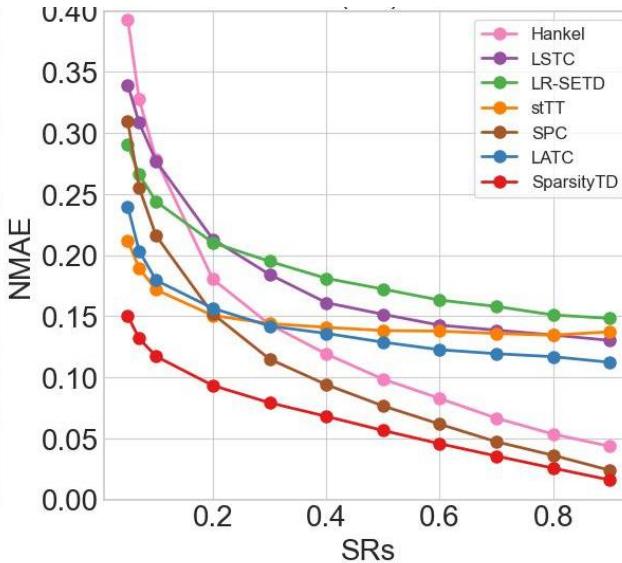
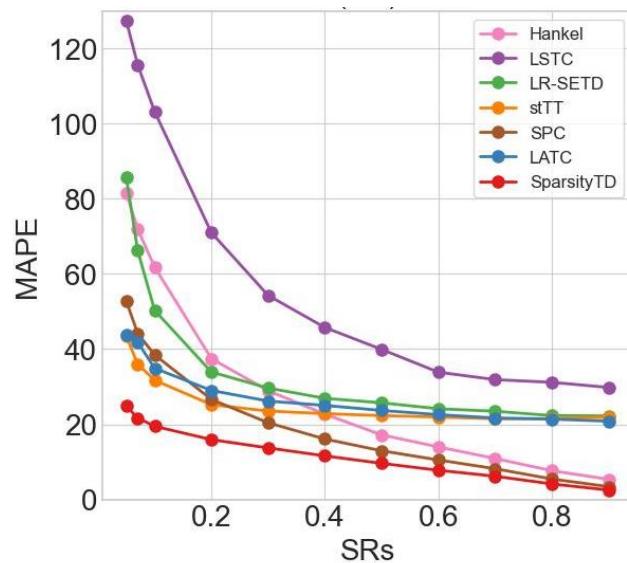


spatiotemporal traffic data imputation: random missing

Guangzhou



Hangzhou



SPC 2016

LATC 2021

LSTC 2021

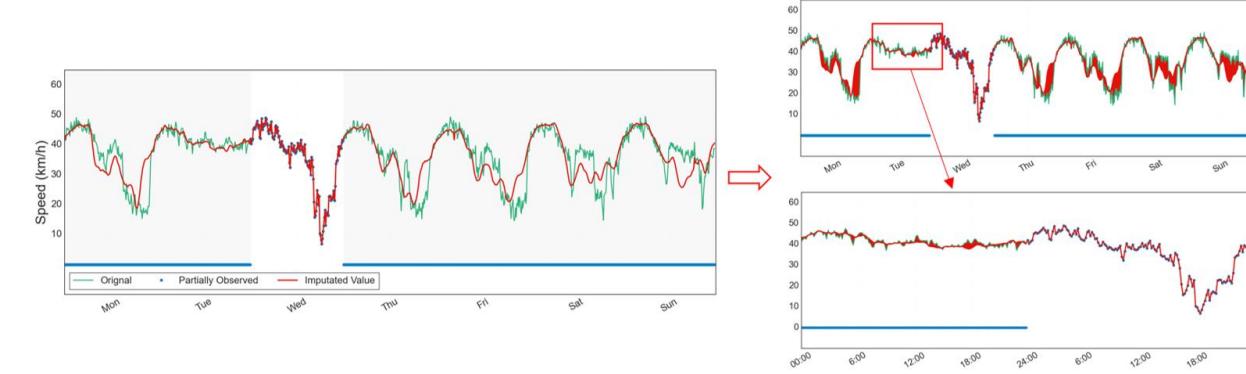
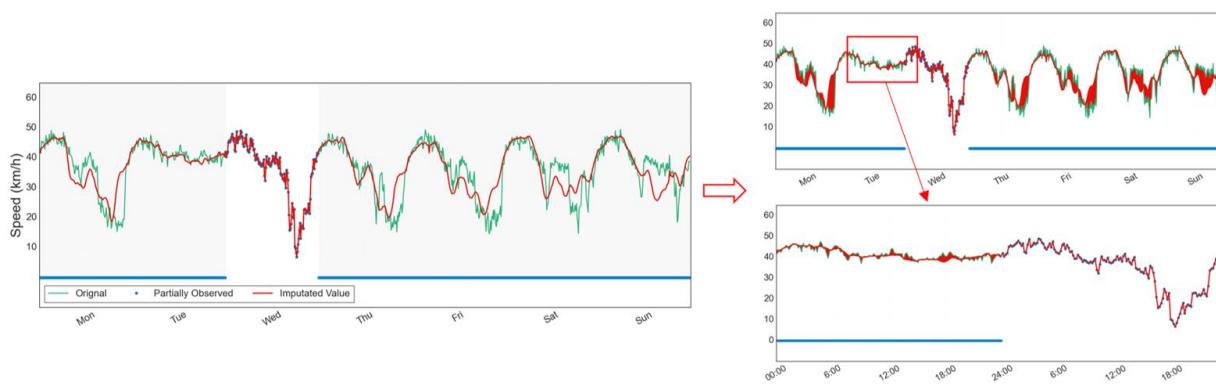
stTT 2022

Hankel 2023

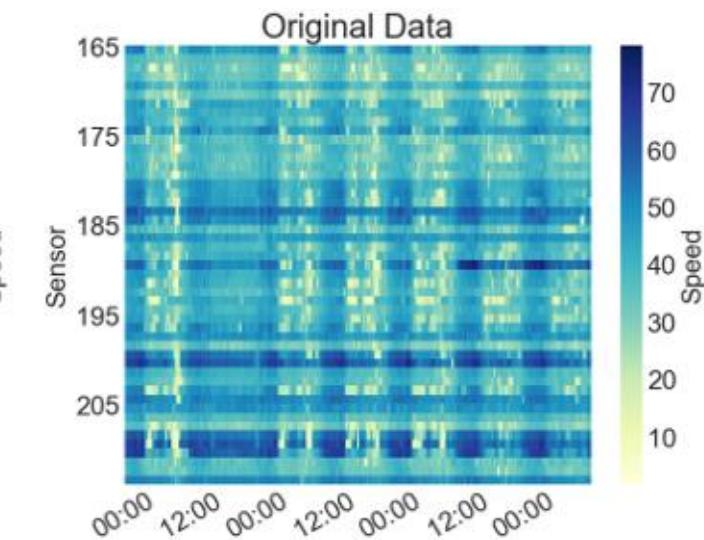
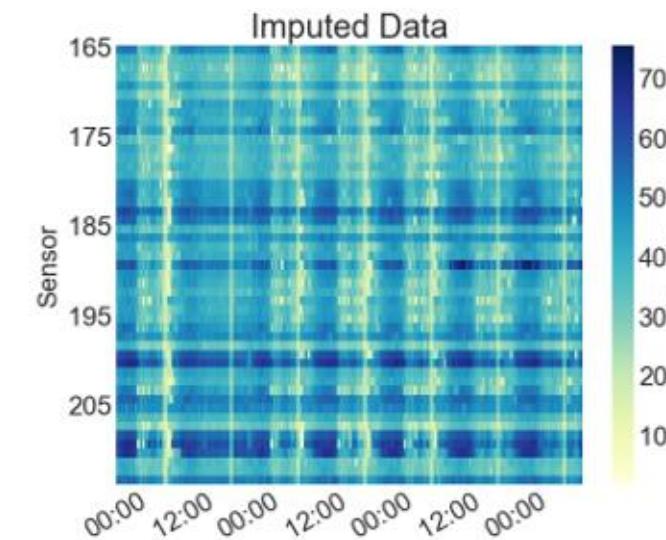
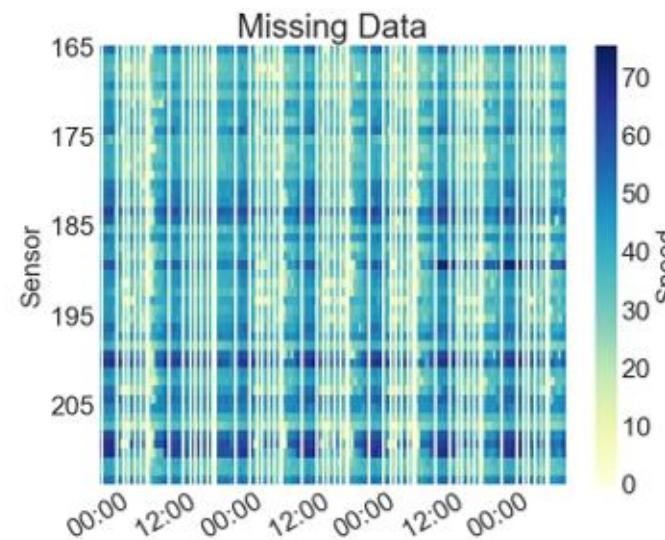
LR-SETD 2024



no-random missing

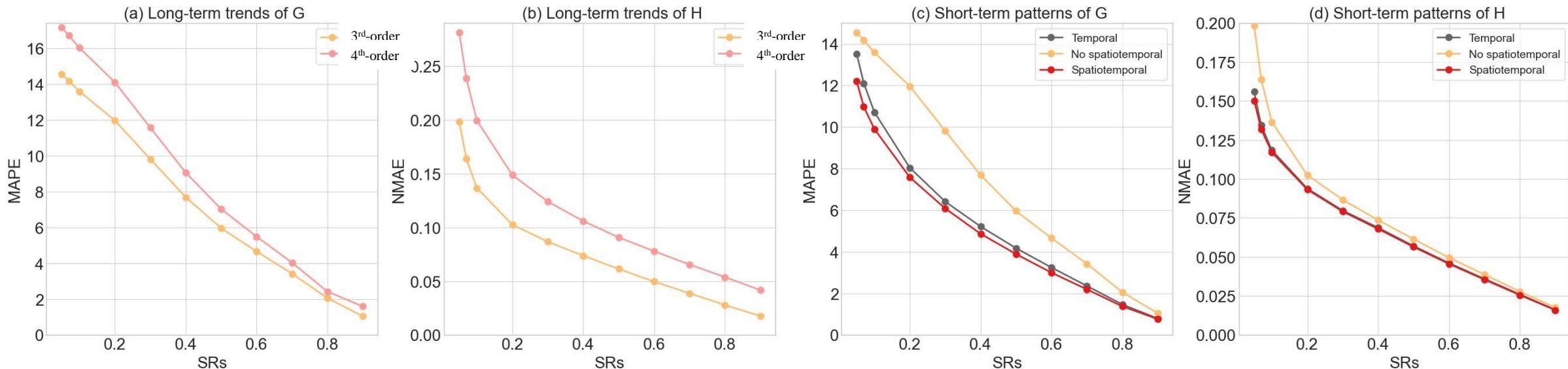


black-out missing



➤ Ablation study

SR	0.01			0.05			0.1			
	Methods	PSNR	SSIM	Time	PSNR	SSIM	Time	PSNR	SSIM	Time
NonTD	13.44	0.0865	85.54	16.20	0.2361	85.06	17.09	0.3170	85.81	
SNTD	13.68	0.1014	85.42	16.37	0.2404	85.39	17.73	0.3170	85.82	
GNTD	13.96	0.0977	86.35	16.55	0.2554	86.35	17.89	0.3436	89.89	
SparsityTD*	13.27	0.0846	32.48	16.87	0.1982	32.57	18.91	0.2936	33.13	
SparsityTD	15.18	0.1212	34.05	19.34	0.3335	35.11	21.61	0.4649	34.96	

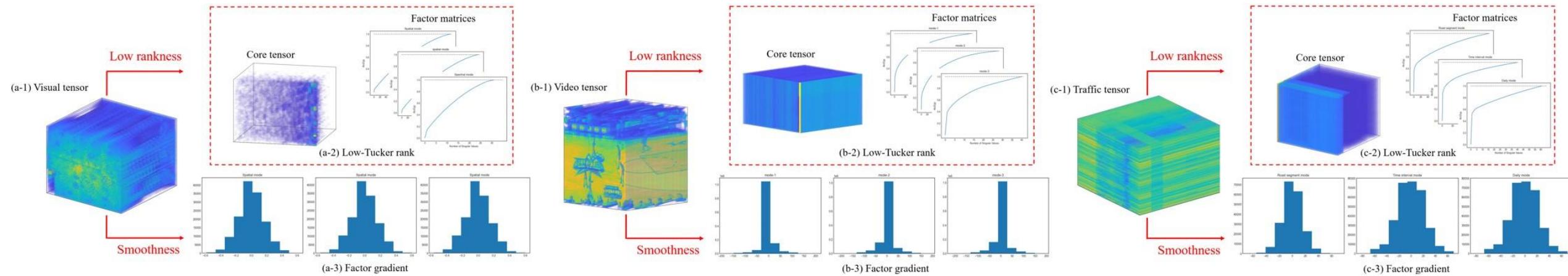




Content

- Introduction
- Literature Review
- Proposed Models
 - Sparsity-based Tucker decomposition model (**SparsityTD**)
 - Enhanced low rankness and smoothness priors Tucker decomposition (**ELST**)
 - Gradient-based Tucker decomposition model (**GradientTD**)
- Conclusion

■ Enhanced Low-rank and Smooth Tucker Decomposition



$$\begin{aligned} \min_{\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}} \quad & (1 - \alpha) \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \sum_{n \in \Gamma} \frac{\beta_n}{2} \text{tr}(\mathbf{U}_n^T \mathbf{L}_n \mathbf{U}_n) \\ \text{s.t.,} \quad & \mathcal{X} = \mathcal{G} \underset{n=1}{\overset{N}{\times}} \mathbf{U}_n, \quad \mathcal{X}_{\Omega} = \mathcal{T}_{\Omega} \end{aligned}$$

$$\omega_n = \frac{\rho_n}{\sum_{n=1}^N \rho_n}, \quad \rho_n = \prod_{i=1, i \neq n}^N \frac{1}{R_i}, \quad R_i = \sum_j \sigma_j(\mathbf{U}_i), \quad \beta_n = \frac{\sigma_1(\mathbf{X}_{(n)})}{\sigma_1(\mathbf{L}_n)}$$

Published paper: Gong, Wen Wu; Huang, Zhe Jun; Yang, Li Li. Enhanced low-rank and sparse Tucker decomposition for image completion [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Seoul, Korea, 2024, 2425-2429. (Chapter 4, CCFB, 南方科技大学认定的 A 类国际学术会议)

Gong, Wen Wu; Huang, Zhe Jun; Yang, Li Li. LSPTD: Low-rank and spatiotemporal priors enhanced Tucker decomposition for internet traffic data imputation [C]. IEEE Conference on Intelligent Transportation Systems, Bilbao, Spain, 2023, 460-465. (Chapter 4, CCFB1, 南方科技大学认定的 A 类国际学术会议)

Under Review: Gong, Wen Wu; Huang, Zhe Jun; Yang, Li Li. ELST: A Tucker-based prior modeling framework for tensor completion. SIAM Journal on Mathematics of Data Science. 2024

➤ Proposed algorithm: Proximal Alternating Linearized Minimization (PALM)

$$\min_{\mathcal{G}, \{\mathbf{U}_n\}} (1 - \alpha) \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \sum_{n \in \Gamma} \frac{\beta_n}{2} \text{tr}(\mathbf{U}_n^T \mathbf{L}_n \mathbf{U}_n) + \frac{\lambda}{2} \left\| \mathcal{G} \times_{n=1}^N \mathbf{U}_n - \mathcal{T}_\Omega \right\|_F^2$$

Algorithm 4-1 PALM-based solver for the ELST model

```

1: Input: Incomplete tensor  $\mathcal{T}$ , observed entries  $\Omega$ .
2: Output: Completion result  $\hat{\mathcal{X}}$ .
3: Initialize  $\mathcal{G}^0, \{\mathbf{U}_n^0\}$  ( $1 \leq n \leq N$ ),  $0 < \alpha < 1$ ,  $\lambda = 1$ ,  $K = 500$ ;
4:  $\mathcal{X}_\Omega^0 = \mathcal{T}_\Omega$ ,  $\mathcal{X}_{\bar{\Omega}}^0 = \text{mean}(\mathcal{T}_{\bar{\Omega}})$ ;
5: for  $k = 0$  to  $K$  do
6:   Update  $\mathcal{G}^{k+1}$  and  $\mathbf{U}_n^{k+1}$  by the Eq. (4-12);
7:   Update Tucker decomposition  $\mathcal{X}^{k+1}$  using the Eq. (4-13);
8:   if  $\Phi(\mathcal{G}^{k+1}, \{\mathbf{U}_n^{k+1}\})$  is increasing then
9:     Re-update  $\tilde{\mathcal{G}}^{k+1} = \mathcal{G}^k$  and  $\tilde{\mathbf{U}}_n^{k+1} = \mathbf{U}_n^k$ , respectively;
10:  else
11:    Re-update  $\tilde{\mathcal{G}}^{k+1}$  and  $\tilde{\mathbf{U}}_n^{k+1}$  using Eq. (4-9) and Eq. (4-10), respectively;
12:  end if
13:  until Eq. (3-19) are satisfied.
14: end for

```

$$\begin{aligned} \mathcal{G}^{k+1} &= \mathcal{S}_{\frac{\alpha}{L_{\mathcal{G}}^k}} \left(\tilde{\mathcal{G}}^k - \frac{1}{L_{\mathcal{G}}^k} \nabla_{\mathcal{G}} f(\tilde{\mathcal{G}}^k) \right) \mathbf{U}_n^{k+1} = \mathcal{D}_{\frac{(1-\alpha)\omega_n}{L_{\mathbf{U}_n}^k}} \left(\tilde{\mathbf{U}}_n^k - \frac{1}{L_{\mathbf{U}_n}^k} \nabla_{\mathbf{U}_n} H(\tilde{\mathbf{U}}_n^k) \right) \\ \mathcal{X}^{k+1} &= \mathcal{T}_\Omega + \left(\mathcal{G}^{k+1} \times_{n=1}^N \mathbf{U}_n^{k+1} \right)_\Omega \end{aligned}$$

A parameterized accelerate strategy [13]

$$\omega_k = \frac{t^{k-1} - 1}{t^k}, t^k = \frac{p + \sqrt{q(rt^{k-1})^2 + 1}}{2}, t^0 = 1 \quad p, q > 0 \quad r \in [0, 4]$$



➤ Proposed algorithm: Proximal Alternating Direction Method (ProADM)

$$\min_{\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}, \mathcal{P}^{\mathcal{X}}, \mu} (1 - \alpha) \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \sum_{n \in \Gamma} \frac{\beta_n}{2} \text{tr}(\mathbf{U}_n^T \mathbf{L}_n \mathbf{U}_n) + \frac{\mu}{2} \left\| \mathcal{X} - \mathcal{G} \times_{n=1}^N \mathbf{U}_n \right\|_F^2 + \left\langle \mathcal{P}^{\mathcal{X}}, \mathcal{X} - \mathcal{G} \times_{n=1}^N \mathbf{U}_n \right\rangle$$

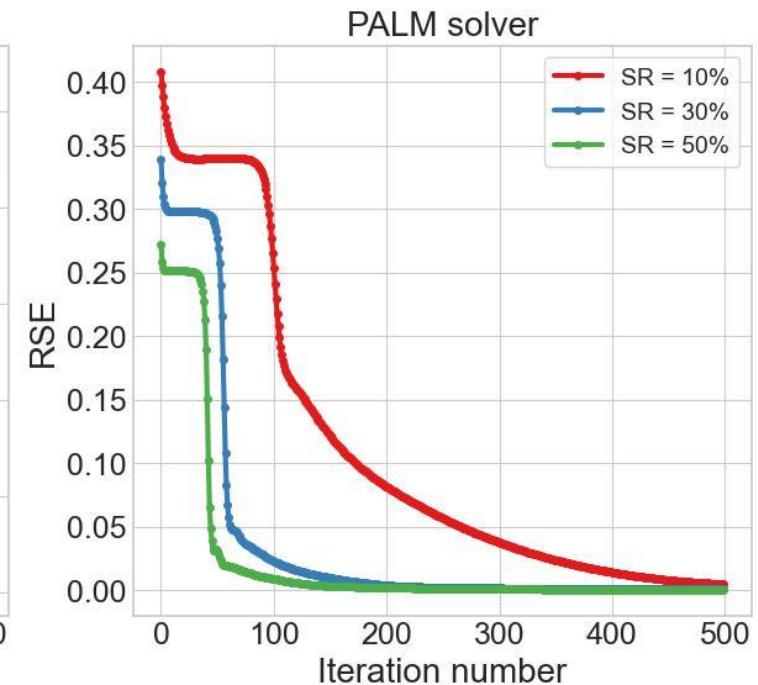
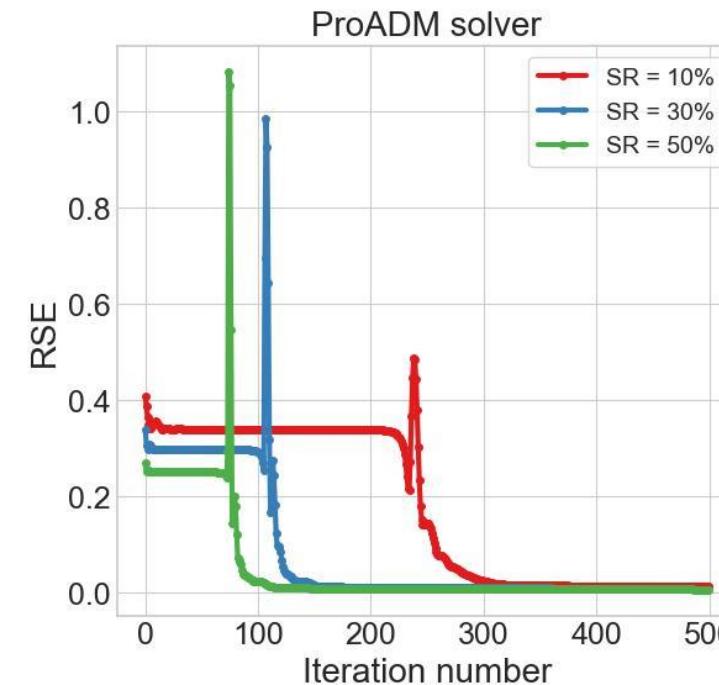
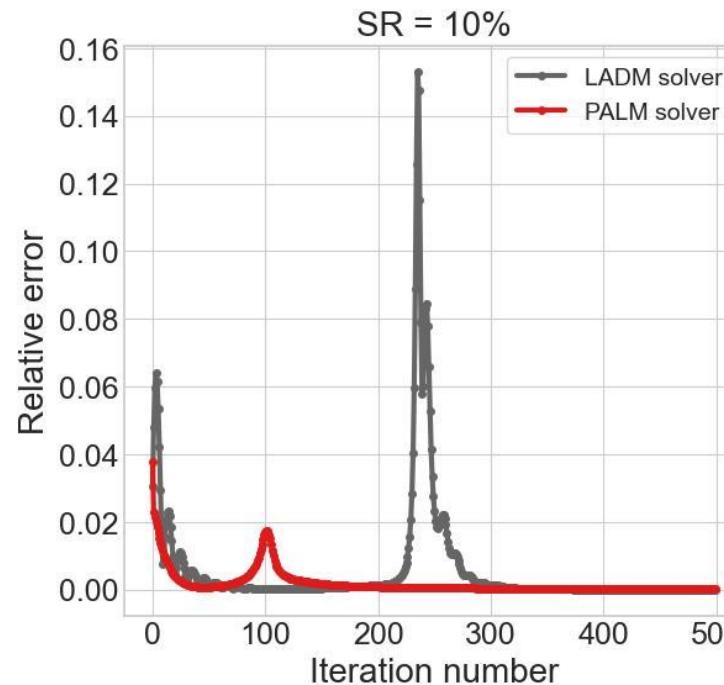
Algorithm 4-2 ProADM-based solver for the ELST model

- 1: **Input:** Incomplete tensor \mathcal{T} , observed entries Ω .
- 2: **Output:** Completion result $\hat{\mathcal{X}}$.
- 3: Initialize $\mathcal{G}^0, \{\mathbf{U}_n^0\}$ ($1 \leq n \leq N$) using HOSVD, $0 < \alpha < 1$, $\mu = 1e^{-5}$, $K = 500$; $\mathcal{G}^{k+1} = \mathcal{S}_{\frac{\alpha}{L_{\mathcal{G}}^k}} \left(\mathcal{G}^k - \frac{1}{L_{\mathcal{G}}^k} \nabla_{\mathcal{G}} f(\mathcal{G}^k) \right)$
- 4: $\mathcal{X}_{\Omega}^0 = \mathcal{T}_{\Omega}$, $\mathcal{X}_{\bar{\Omega}}^0 = \text{mean}(\mathcal{T}_{\bar{\Omega}})$;
- 5: **for** $k = 1$ to K **do**
- 6: Optimize \mathcal{G}^{k+1} via the Eq. (4-21) with other variables fixed; $\mathbf{U}_n^{k+1} = \mathcal{D}_{\frac{(1-\alpha)\omega_n}{L_{\mathbf{U}_n}^k}} \left(\mathbf{U}_n^k - \frac{1}{L_{\mathbf{U}_n^k}} \nabla_{\mathbf{U}_n} f(\mathbf{U}_n^k) \right)$
- 7: Optimize all \mathbf{U}_n^{k+1} via Eq. (4-22) with other variables fixed;
- 8: Optimize \mathcal{X}^{k+1} via the Eq. (4-23) with other variables fixed;
- 9: Update multipliers $\mathcal{P}^{\mathcal{X}}$ using Eq. (4-24); $\mathcal{X}_{\Omega}^{k+1} = \mathcal{T}_{\Omega}$, $\mathcal{X}_{\bar{\Omega}}^{k+1} = \left(\mathcal{G}^{k+1} \times_{n=1}^N \mathbf{U}_n^{k+1} - \frac{\mathcal{P}_k^{\mathcal{X}}}{\mu_k} \right)_{\bar{\Omega}}$
- 10: **until** Eq. (3-19) are satisfied.
- 11: **end for**
- 12: **return** $\hat{\mathcal{X}} = \mathcal{X}^K$. $\mathcal{P}_{k+1}^{\mathcal{X}} = \mathcal{P}_k^{\mathcal{X}} + \mu^k \left(\mathcal{X}^{k+1} - \mathcal{G}^{k+1} \times_{n=1}^N \mathbf{U}_n^{k+1} \right)$, $\mu^{k+1} = \rho \mu^k$, $\rho \in [1.1, 1.2]$

➤ Algorithm convergence

Theorem 3: Let $\Theta^k = \{\mathcal{G}^k, \{\mathbf{U}_n^k\}\}$ be the sequence generated by PALM-based algorithm. Assuming that Θ^k is bounded, then we assure that Θ^k globally converges to a stationary point $\hat{\Theta} = \{\hat{\mathcal{G}}, \{\hat{\mathbf{U}}_n\}\}$.

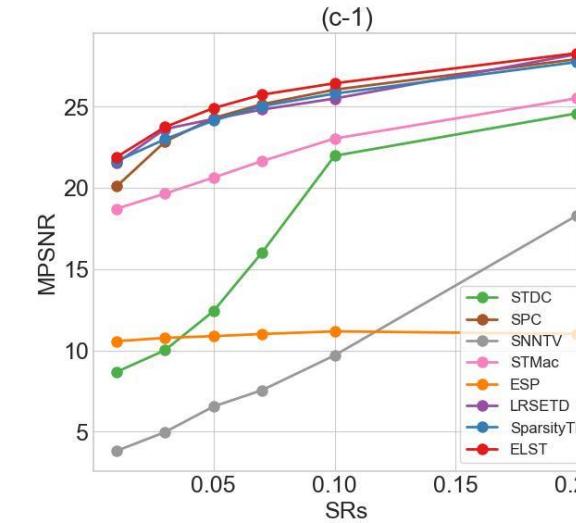
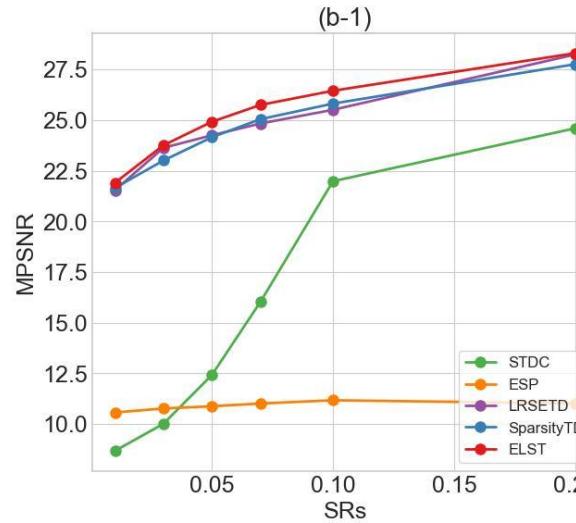
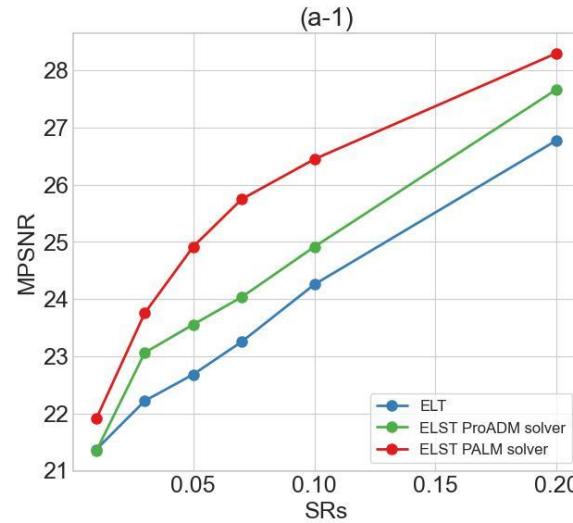
Theorem 4: For sufficiently large μ , the sequence $\Theta^k = \{\mathcal{G}^k, \{\mathbf{U}_n^k\}, \mathcal{X}^k, \mathcal{P}_k^{\mathcal{X}}\}$ produced by ProADM-based algorithm globally converges to a stationary point $\hat{\Theta} = \{\hat{\mathcal{G}}, \{\hat{\mathbf{U}}_n\}, \hat{\mathcal{X}}, \hat{\mathcal{P}}^{\mathcal{X}}\}$.





➤ Superiority performance

RGB



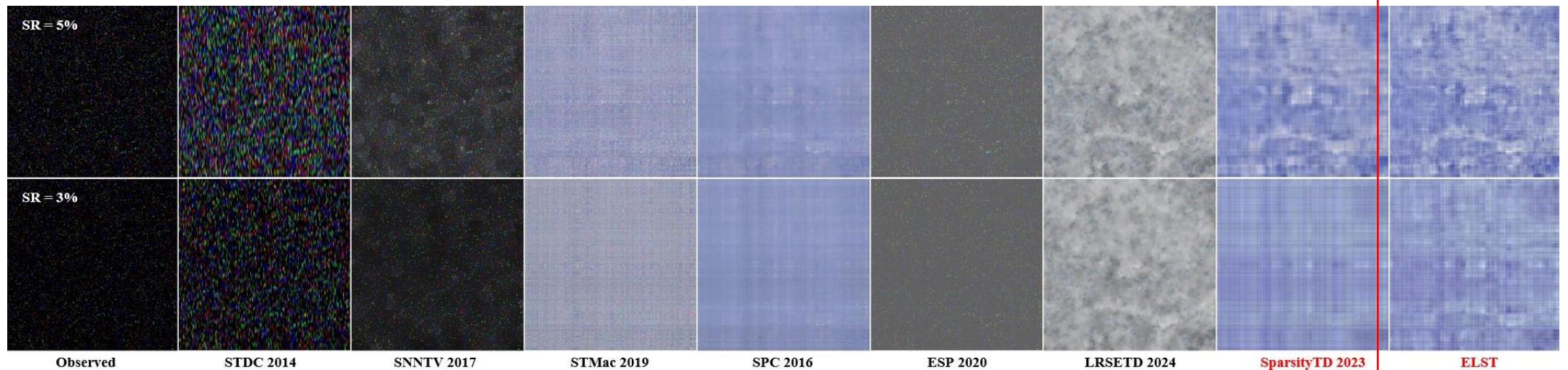
SNNTV 2017

STMac 2019

ESP 2020

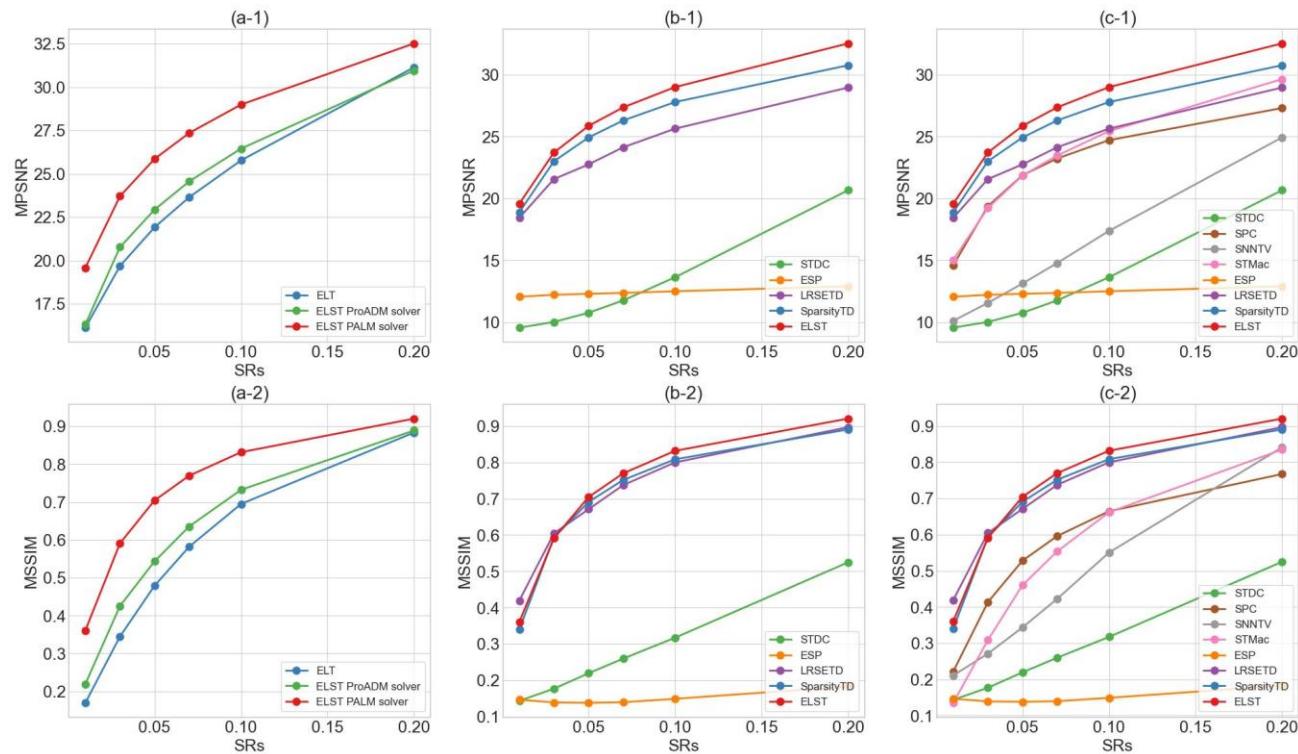
LR-SETD 2024

SparsityTD





MRI



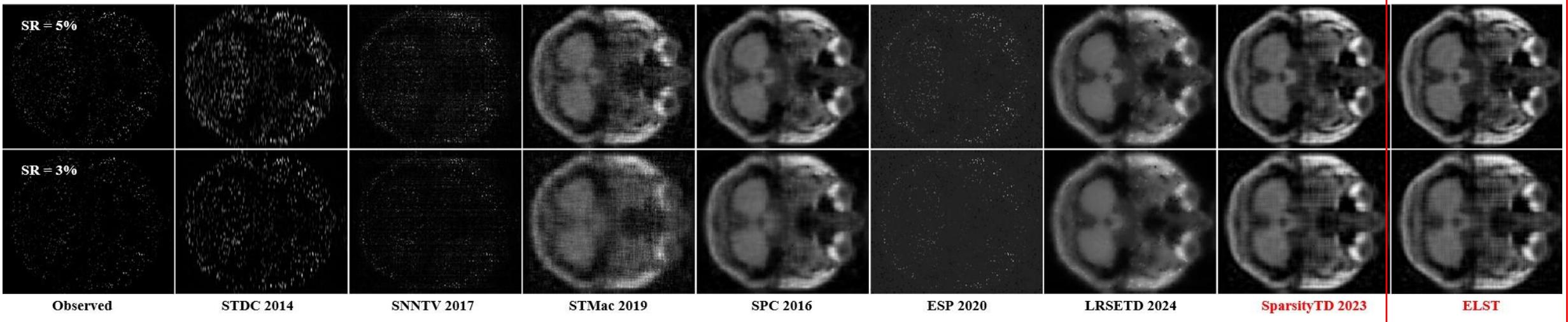
SNNTV 2017

STMac 2019

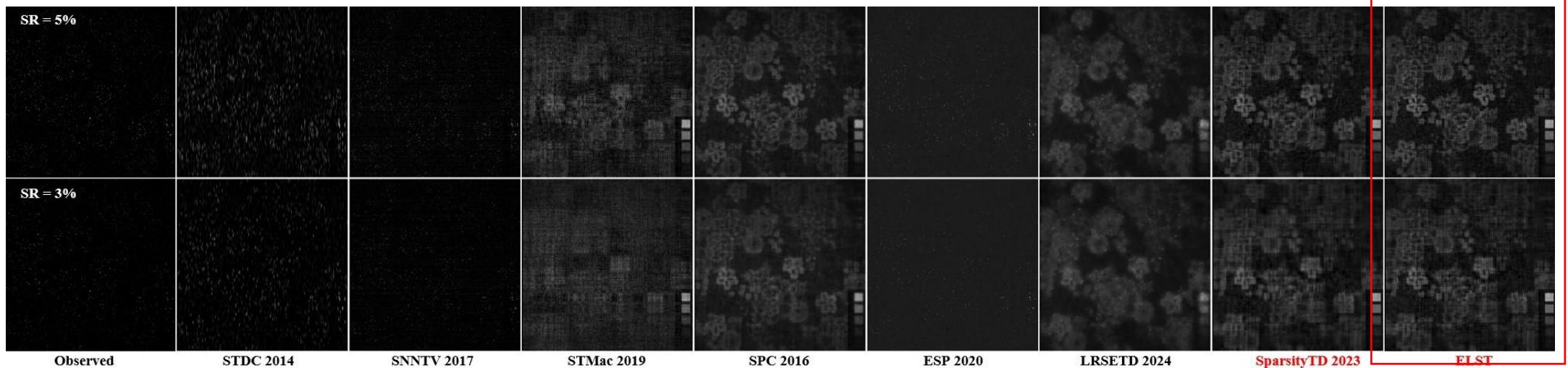
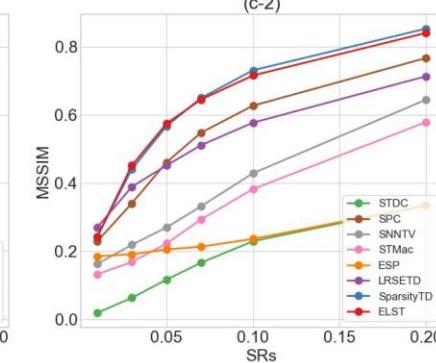
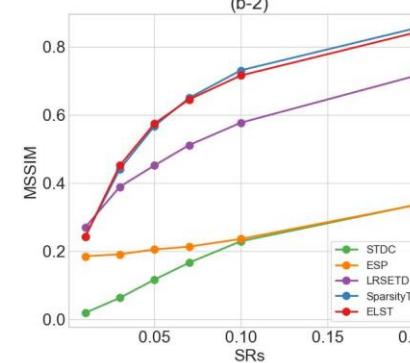
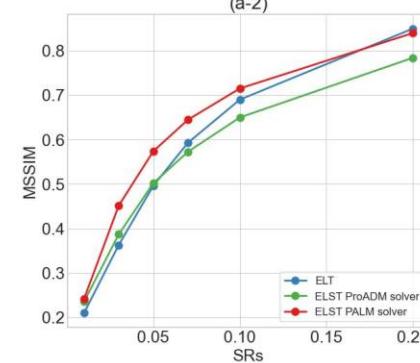
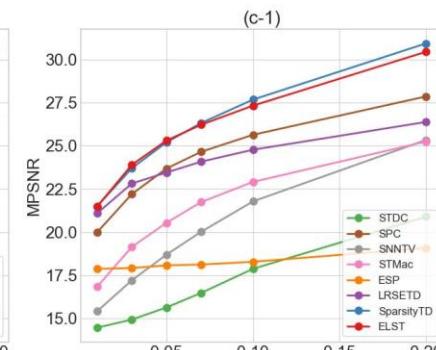
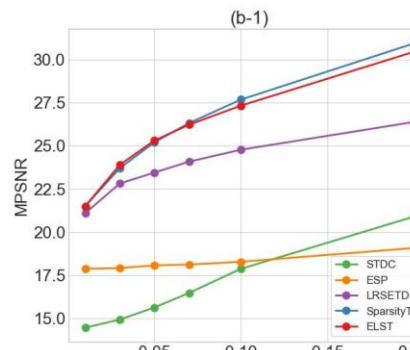
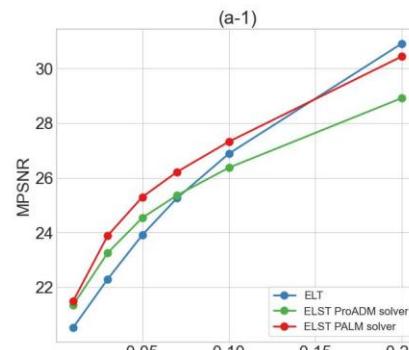
ESP 2020

LR-SETD 2024

SparsityTD

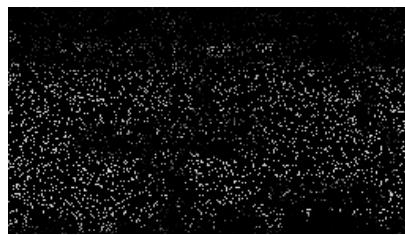
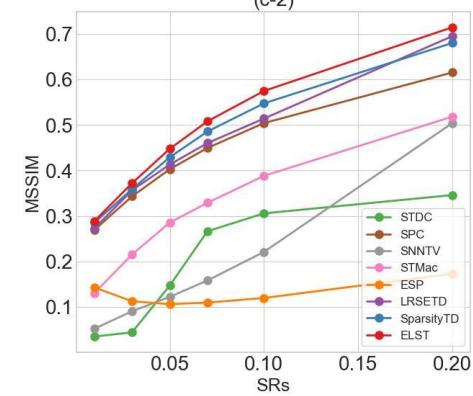
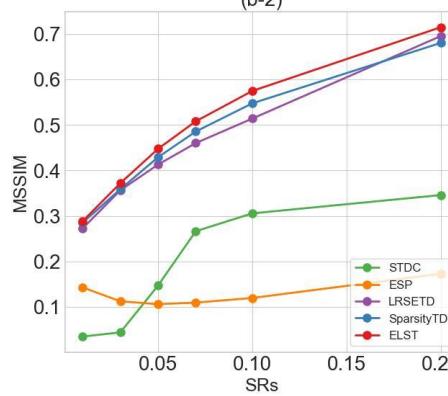
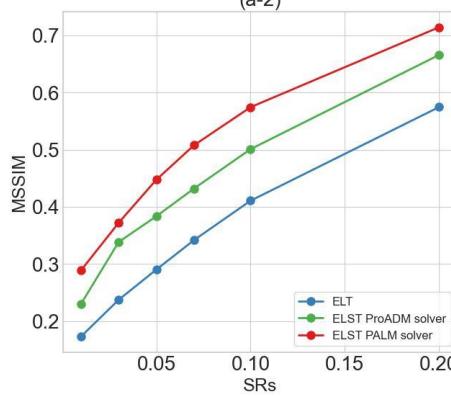
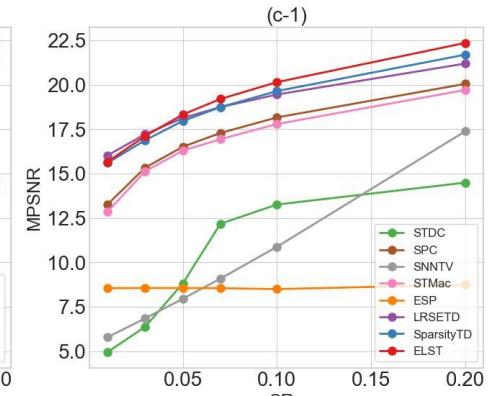
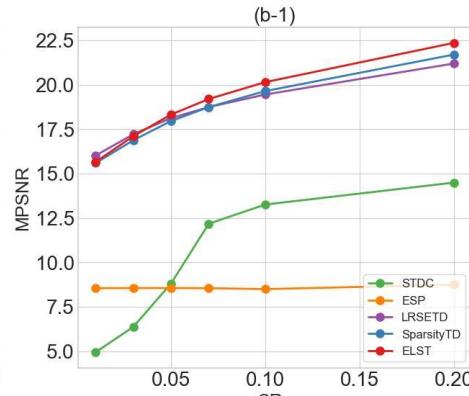
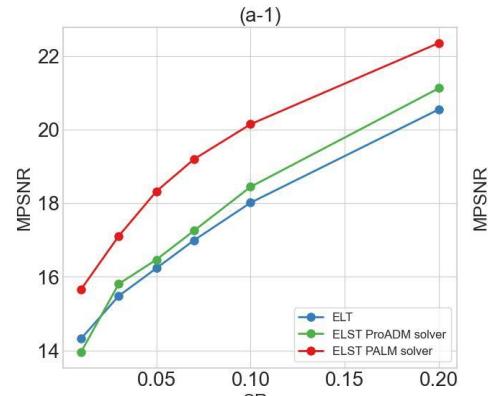


MSI

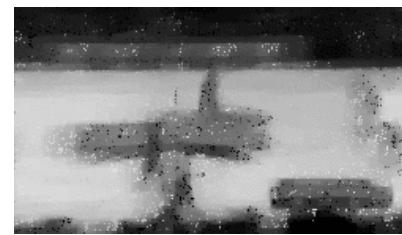




Video



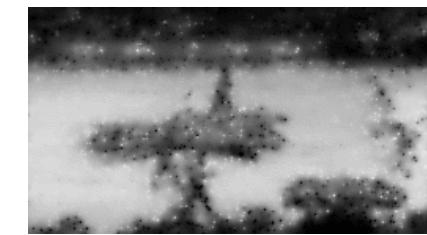
SR = 5%



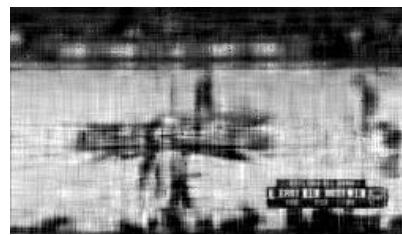
SPC 2016



STMac 2019

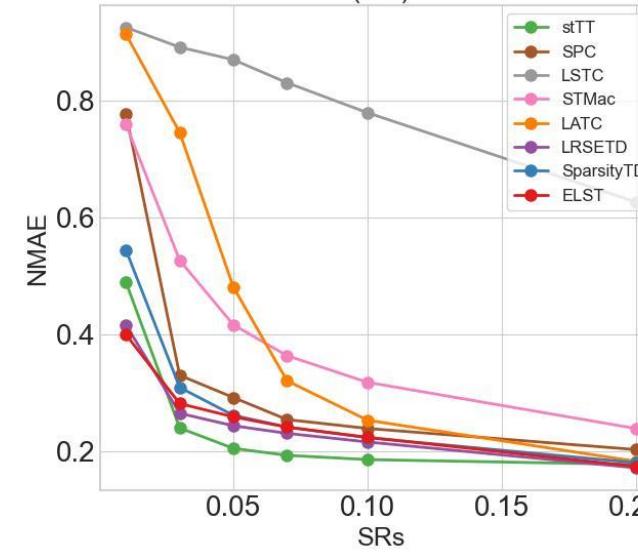
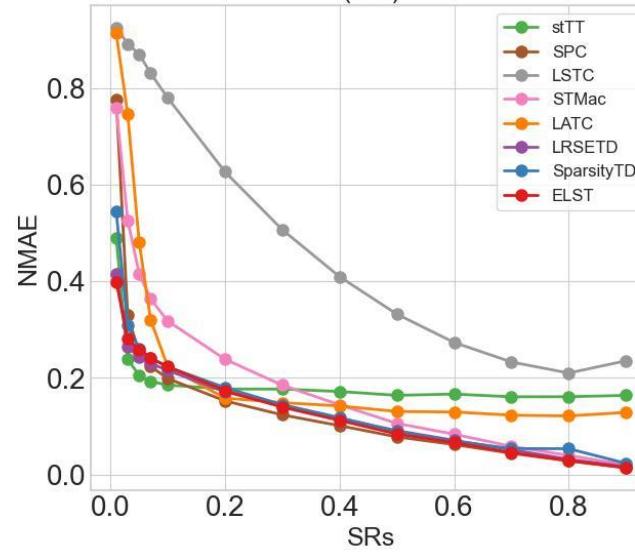
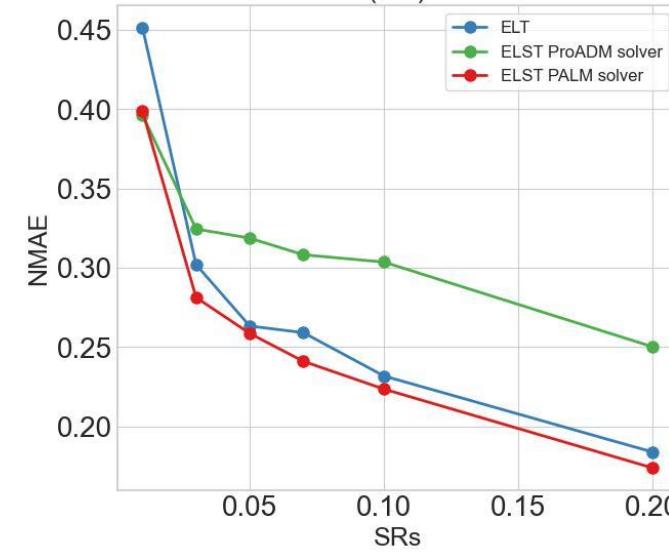
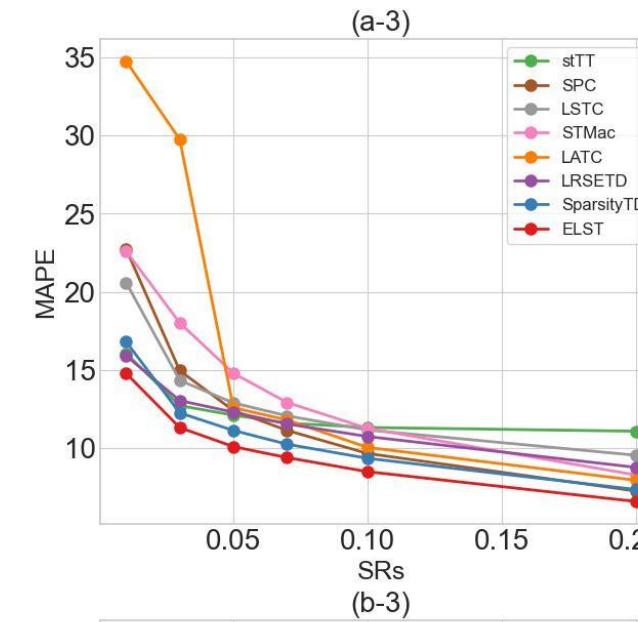
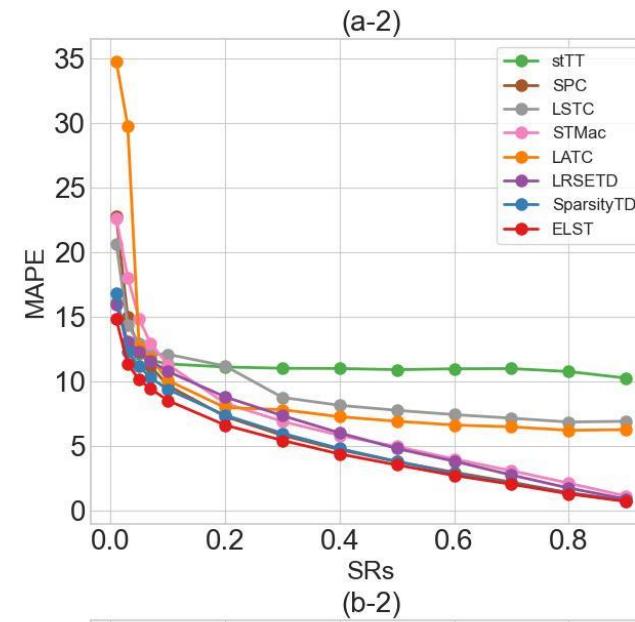
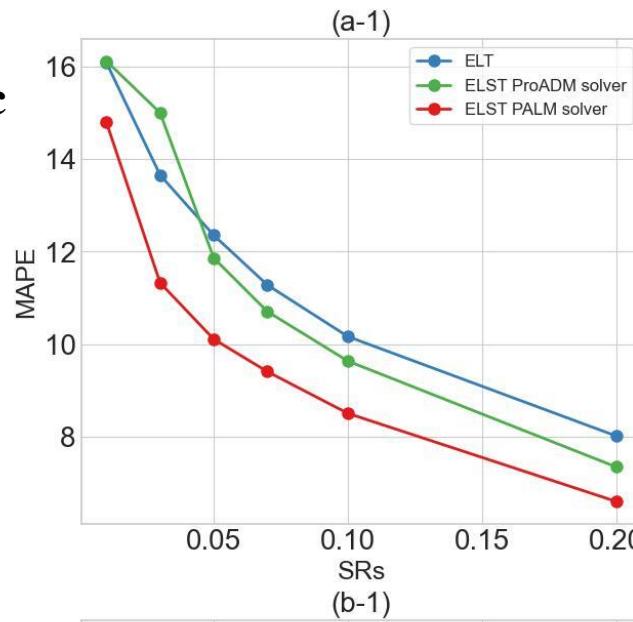


SparsityTD



ELST

Traffic



LATC 2021

LSTC 2021

stTT 2022

LR-SETD 2024

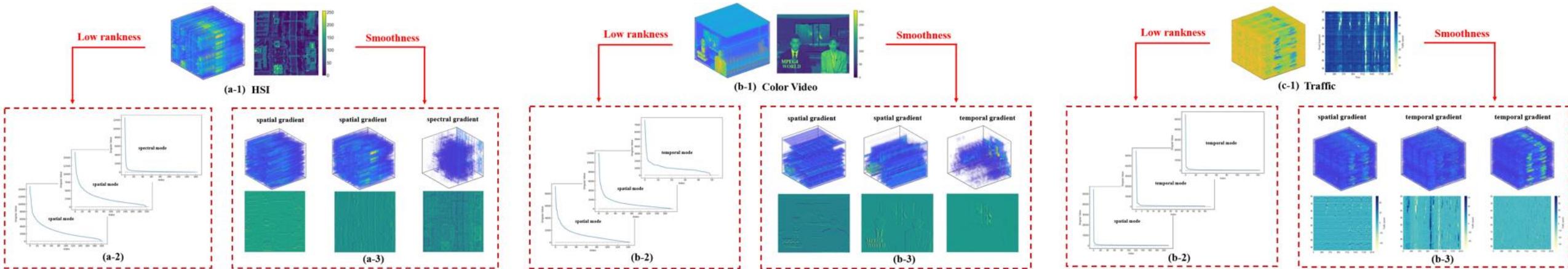
SparsityTD



Content

- Introduction
- Literature Review
- Proposed Models
 - Sparsity-based Tucker decomposition model (**SparsityTD**)
 - Enhanced low rankness and smoothness priors Tucker decomposition (**ELST**)
 - Gradient-based Tucker decomposition model (**GradientTD**)
- Conclusion

■ Gradient-based Tucker Decomposition



$$\min_{\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}} (1 - \alpha) \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \sum_{n \in \Gamma} \beta_n \|\mathbf{L}_n \mathbf{X}_{(n)}\|_p^p, \text{ s.t., } \mathcal{X} = \mathcal{G} \times_{n=1}^N \mathbf{U}_n, \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega$$

$$\begin{aligned} \|\mathcal{X} \times_n \mathbf{D}_n\|_p^p &= \|\mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_n (\mathbf{D}_n \mathbf{U}_n) \times_{n+1} \cdots \times_N \mathbf{U}_N\|_p^p \\ &= \|(\mathbf{D}_n \mathbf{U}_n) \left(\mathcal{G} \times_{p=1, p \neq n}^N \mathbf{U}_p \right)\|_p^p = \|\mathbf{D}_n \mathbf{X}_{(n)}\|_p^p \end{aligned}$$

$p = 1$ Total variation
 $p = 2$ Quadratic variation

Connection: the columns of factor gradient span the tensor gradient

$$\|\mathcal{X} \times_n \mathbf{D}\|_F^2 \leq \|\mathbf{D} \mathbf{U}_n\|_F^2 \left\| \mathcal{G} \times_{p=1, p \neq n}^N \mathbf{U}_p \right\|_F^2 = \text{const.} \|\mathbf{D} \mathbf{U}_n\|_F^2 \leq \text{tr}(\mathbf{U}_n^T \mathbf{D}^T \mathbf{D} \mathbf{U}_n)$$



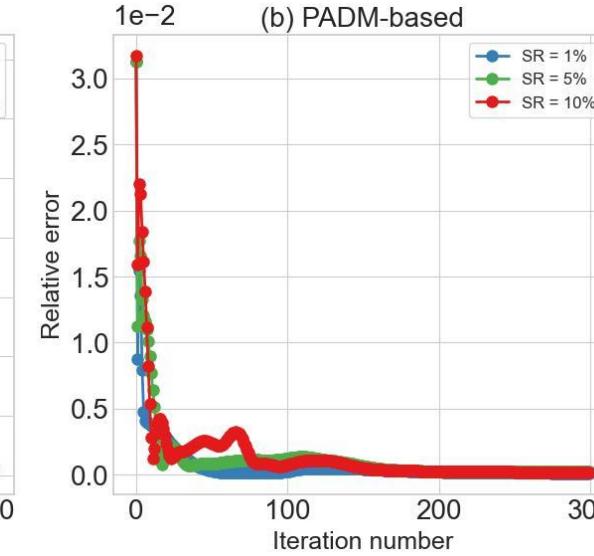
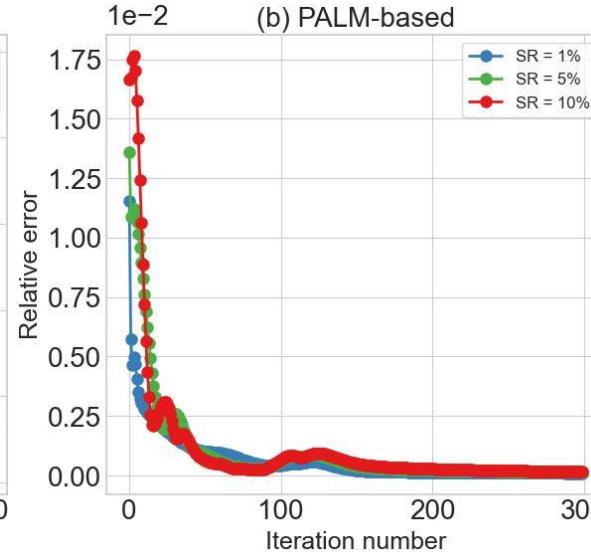
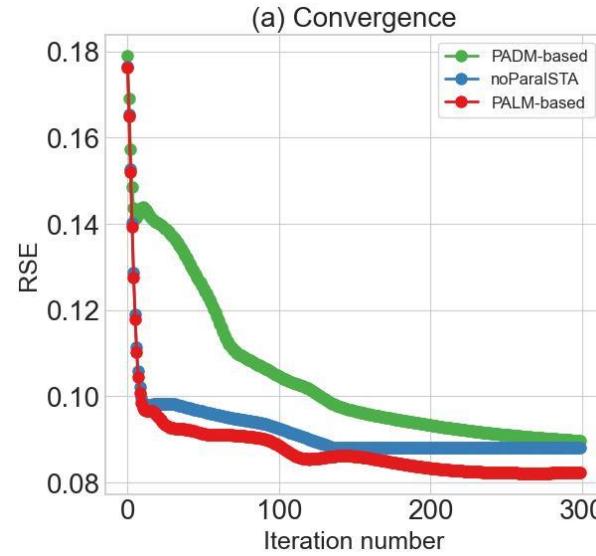
➤ Proposed algorithm

$$\begin{aligned} \min_{\mathcal{G}, \{\mathbf{U}_n\}, \mathcal{X}, \{\mathbf{Q}_n\}, \{\mathbf{Y}_n\}, \{\mathbf{P}_n^Q\}, \{\mathbf{P}_n^Y\}} & (1 - \alpha) \sum_{n=1}^N \omega_n \|\mathbf{U}_n\|_* + \alpha \|\mathcal{G}\|_1 + \sum_{n \in \Gamma} \beta_n \|\mathbf{Q}_n\|_p^p + \frac{\lambda}{2} \left\| \mathcal{X} - \mathcal{G} \times_{n=1}^N \mathbf{U}_n \right\|_F^2 \\ & + \sum_{n \in \Gamma} (\langle \mathbf{Q}_n - \mathbf{L}_n \mathbf{Y}_n, \mathbf{P}_n^Q \rangle + \langle \mathbf{Y}_n - \mathbf{X}_{(n)}, \mathbf{P}_n^Y \rangle) + \sum_{n \in \Gamma} \left(\frac{\mu_1}{2} \|\mathbf{Q}_n - \mathbf{L}_n \mathbf{Y}_n\|_F^2 + \frac{\mu_2}{2} \|\mathbf{Y}_n - \mathbf{X}_{(n)}\|_F^2 \right) \end{aligned}$$

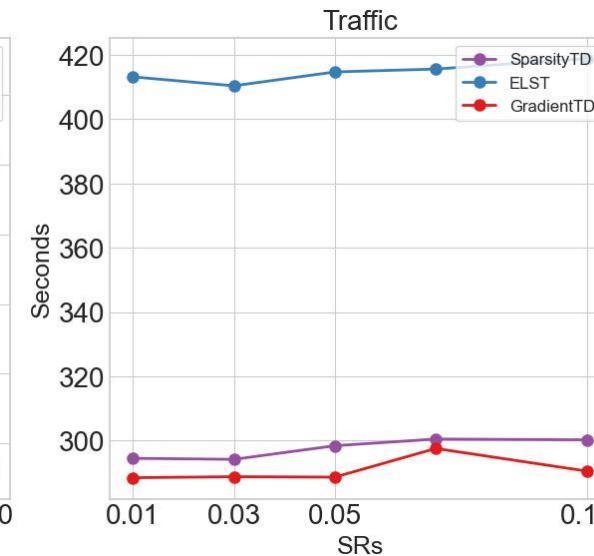
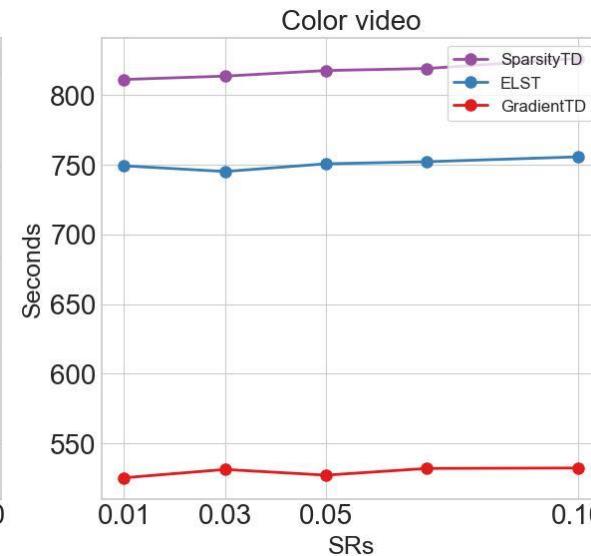
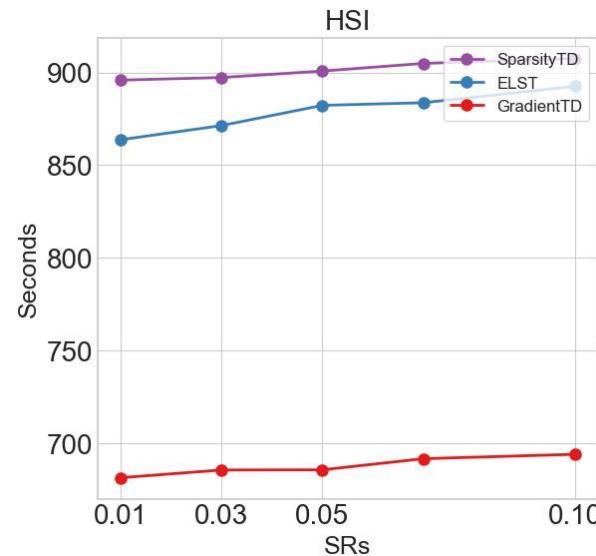
Algorithm 5-1 PALM-LRTC solver for the GradientTD model

- 1: **Input:** Incomplete tensor \mathcal{T} , observed entries Ω .
 - 2: **Output:** Completion result $\hat{\mathcal{X}}$.
 - 3: Initialize $\mathcal{G}^0, \{\mathbf{U}_n^0\}$ ($1 \leq n \leq N$), $0 < \alpha < 1$, $\lambda = \rho\mu$, $K = 300$;
 - 4: $\mathcal{X}_\Omega^0 = \mathcal{T}_\Omega$, $\mathcal{X}_{\bar{\Omega}}^0 = \text{mean}(\mathcal{T}_{\bar{\Omega}})$;
 - 5: **for** $k = 0$ to K **do**
 - 6: Update $\{\mathbf{Q}_n\}$ and $\{\mathbf{Y}_n\}$ as Eq. (5-6) and Eq. (5-8).
 - 7: Update $\{\mathbf{U}_n\}$ and \mathcal{G} simultaneously via Eq. (5-9) and Eq. (5-10);
 - 8: Update \mathcal{X}^{k+1} using Eq. (5-11);
 - 9: Update multipliers \mathbf{P}_n and penalty parameters μ via Eq. (5-12).
 - 10: **until** Eq. (3-19) are satisfied.
 - 11: **end for**
-

➤ Algorithm convergence

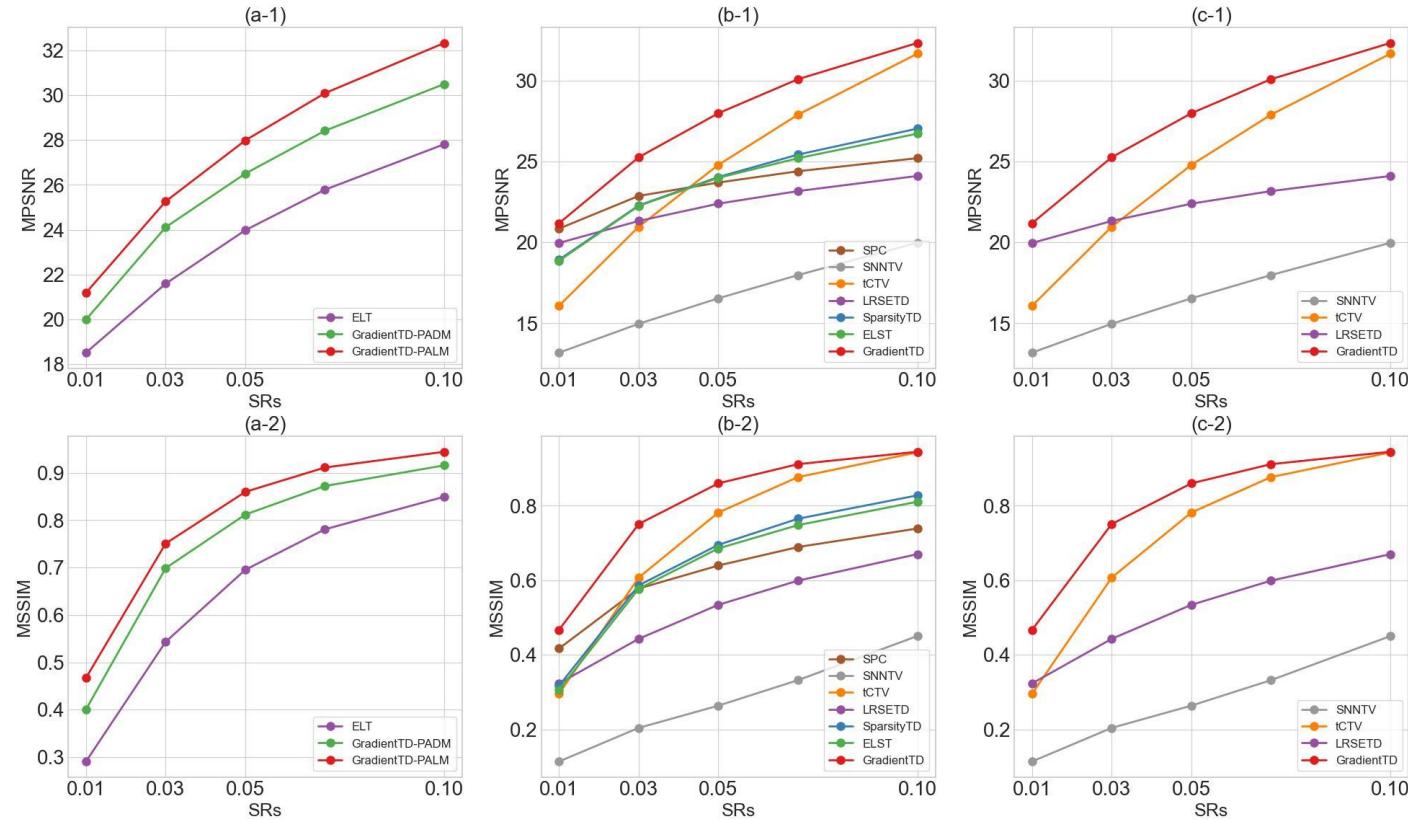


Numerical convergent



More efficient

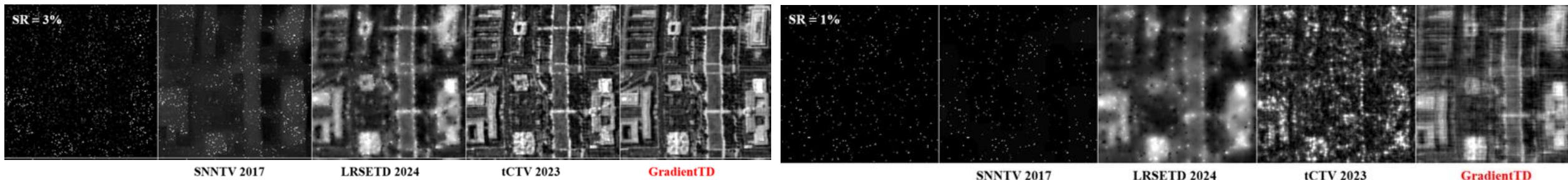
➤ HSI completion



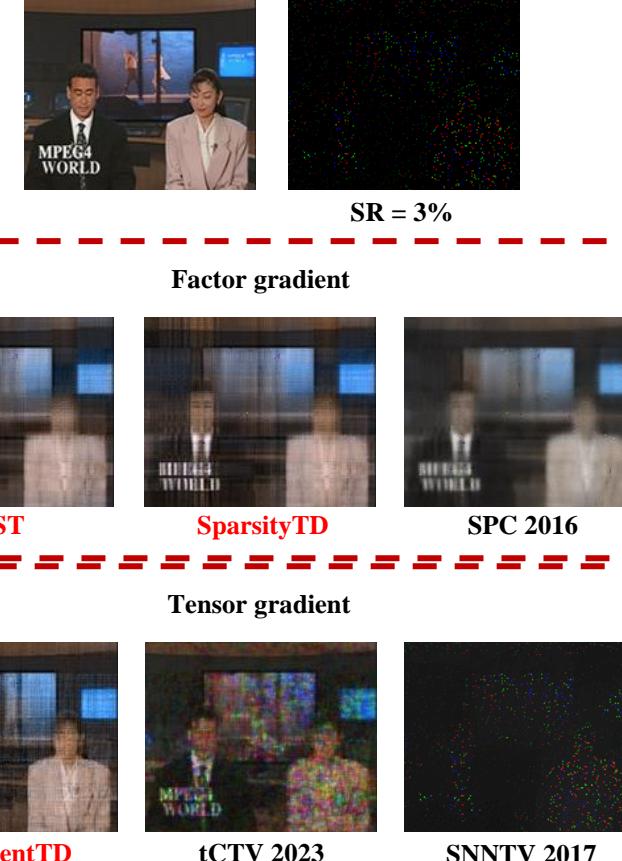
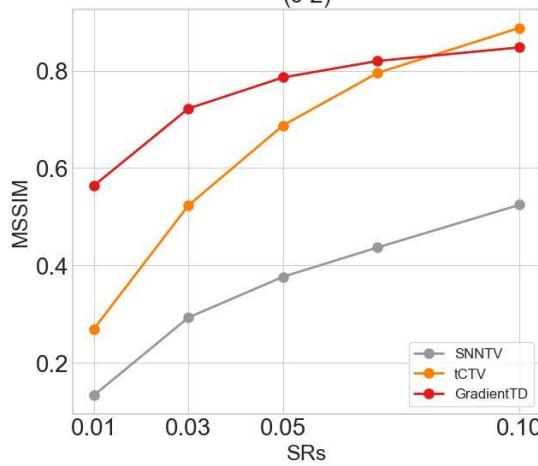
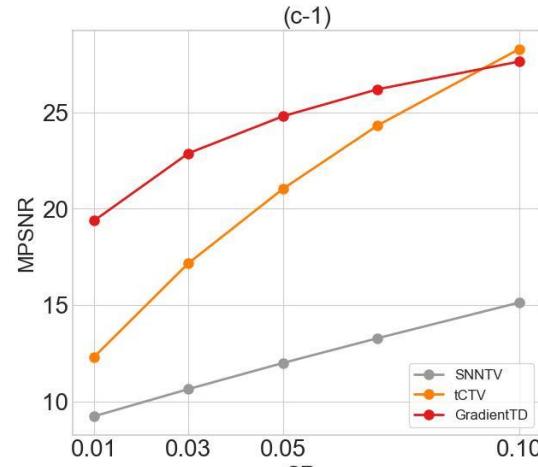
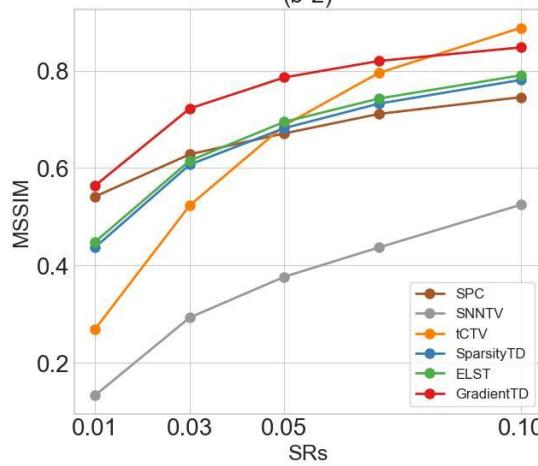
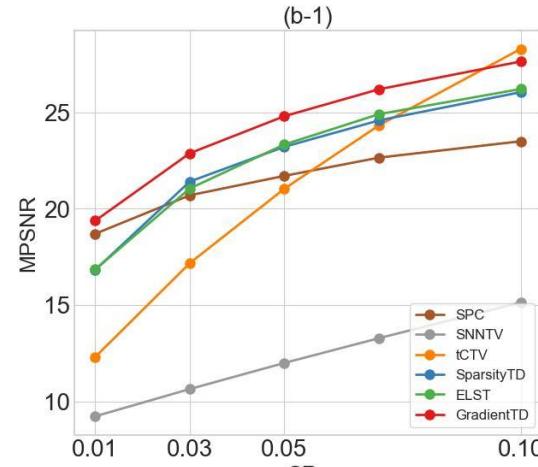
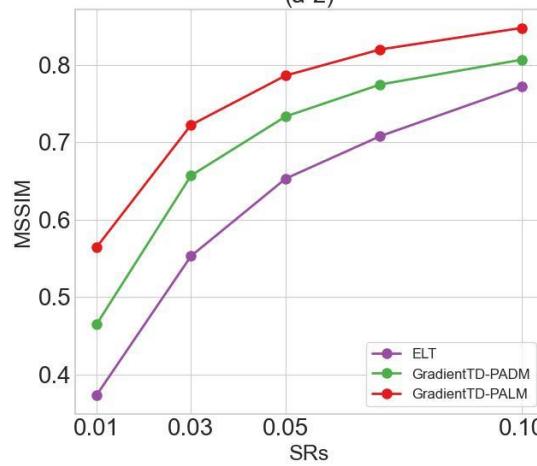
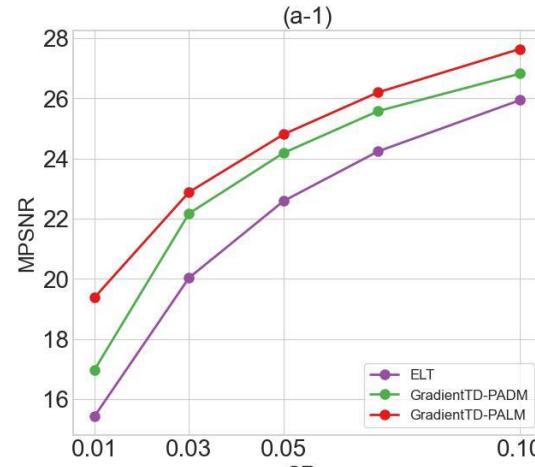
SNNTV
TG 2017

LRSETD
TG 2024

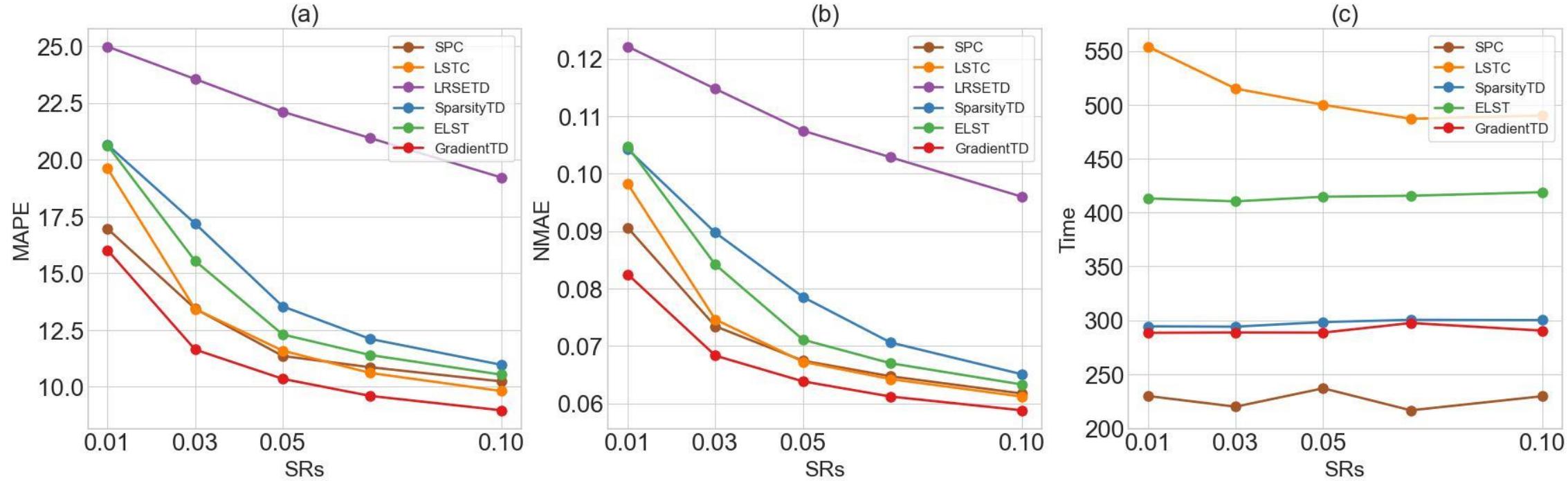
tCTV
TG 2023



➤ Color video recovery



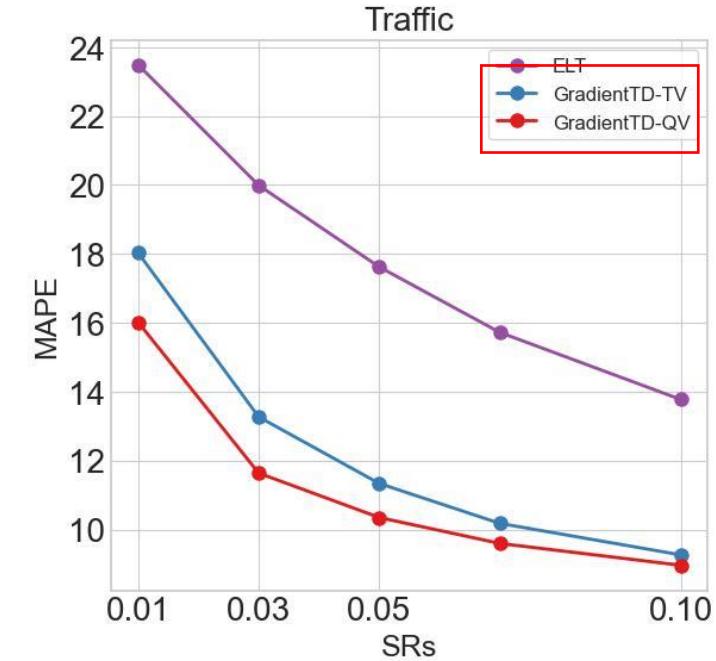
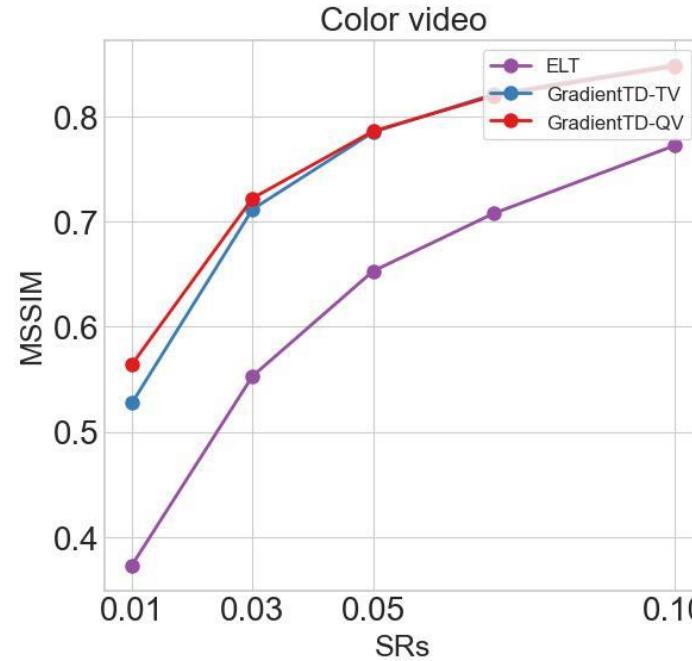
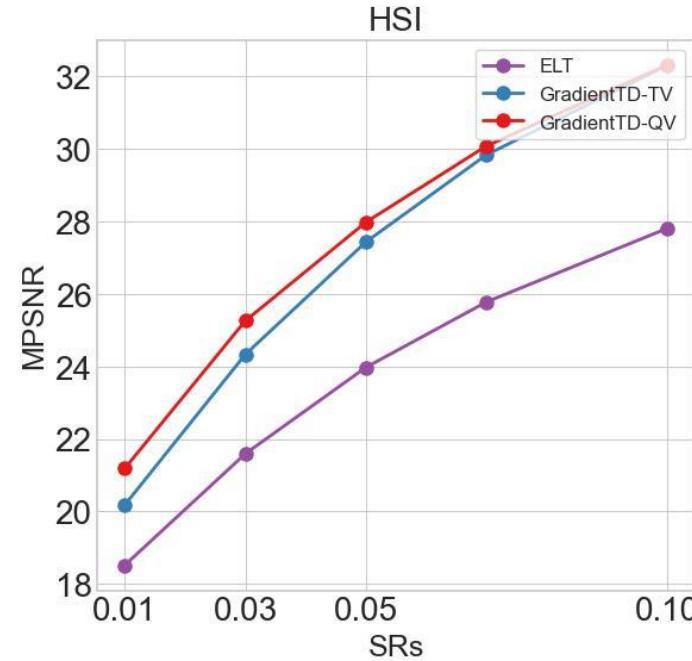
➤ Large-scale traffic data imputation



More accurate and efficient

LSTC
TG 2021
LRSETD
TG 2024

➤ Ablation study



Content

- Introduction
- Literature Review
- Proposed Models
 - Sparsity-based Tucker decomposition model (SparsityTD)
 - Enhanced low rankness and smoothness priors Tucker decomposition (ELST)
 - Gradient-based Tucker decomposition model (GradientTD)
- Conclusion



Tucker-based TC models for high-dimensional visual data completion and multi-dimensional traffic data imputation

Task A: high-dimensional visual data completion

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{T}\|_F^2, \quad \text{s. t., } \mathcal{X}_{\Omega} = \mathcal{T}_{\Omega} \\ \mathcal{X}, \mathcal{T}, \Omega \in R^{I_1 \times \dots \times I_N}$$

Task B: multi-dimensional traffic data imputation

- Sparsity-based Tucker decomposition (SparsityTD)
 - ✓ Sparsity measure
 - ✓ Graph regularization
- Enhanced Low-rankness and Smoothness priors Tucker Decomposition (ELST)
 - ✓ Enhanced Tucker decomposition
 - ✓ Factor gradient

- Sparsity-based Tucker decomposition (SparsityTD)
 - ✓ 3-rd tensor structure
 - ✓ Gradients
- Enhanced Low-rankness and Smoothness priors Tucker Decomposition (ELST)

Task C: Extreme missing tensorial data completion

- Gradient-based Tucker Decomposition (GradientTD)
 - ✓ Tensor gradient

学术成果：3篇发表，1篇接收，3篇在审

Published papers:

1. **Gong, Wen Wu**; Huang, Zhe Jun; Yang, Li Li. Accurate regularized Tucker decomposition for image restoration [J]. Applied Mathematical Modeling, 2023, 123 (11): 75-86. (**Chapter 3**, SCI, IF = 5, 中科院一区)
2. **Gong, Wen Wu**; Huang, Zhe Jun; Yang, Li Li. Enhanced low-rank and sparse Tucker decomposition for image completion [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Seoul, Korea, 2024, 2425-2429. (**Chapter 4**, EI, CCFB, 南方科技大学认定的A类国际学术会议)
3. **Gong, Wen Wu**; Huang, Zhe Jun; Yang, Li Li. LSPTD: Low-rank and spatiotemporal priors enhanced Tucker decomposition for internet traffic data imputation [C]. IEEE Conference on Intelligent Transportation Systems, Bilbao, Spain, 2023, 460-465. (**Chapter 4**, EI, 南方科技大学认定的 A 类国际学术会议)

Under review:

1. **Gong, Wen Wu**; Huang, Zhe Jun; Yang, Li Li. Spatiotemporal regularized Tucker decomposition approach for traffic data imputation. IEEE Transactions on Intelligent Transportation Systems. 2024. **Chapter 3**
2. **Gong, Wen Wu**; Huang, Zhe Jun; Yang, Li Li. ELST: A Tucker-based prior modeling framework for tensor completion. SIAM Journal on Mathematics of Data Science. 2024. **Chapter 4**
3. Lu, Jia Xin; **Gong, Wen Wu** and Yang Li Li. Lu, Jia Xin; Gong, Wen Wu and Yang Li Li. Low-rank autoregressive Tucker decomposition for traffic data imputation. Conference Paper. **Chapter 5**
4. Huang, Rong Ping; **Gong, Wen Wu**; Lu, Jia Xin and Yang Li Li. BACP: Bayesian Augmented CP factorization for traffic data imputation. Conference Paper. **Accepted**

■ Highlights



Tucker-based TC methods	Low Tucker rank		Smooth structure	
	factor matrices	core tensor	tensor gradient	factor gradient
GradientTD	√	√	√	
ELST	√	√		√
SparsityTD	√	√		√
SBCD ^[45]	√	√		√
DCT-based ^[48]		√		
LRSETD ^[46]	√	√	√	
ESP ^[44]				√
KBR ^[21]		√		
IFHST ^[49]		√		
gHOI ^[50]	√			
SNTD ^[51]	√	√	√	
STDC ^[47]	√			√
Tucker ^[22]	√			

Our proposals

Other Tucker-based models

■ Contributions

Priori modeling and optimization algorithms for tensor completion (TC) problems. The main objectives are fourfold:

- Enhanced Tucker decomposition methods are introduced

From the perspective of **tensor sparsity**, a sparsity-based TD that utilizes non-negative factor matrices and sparse core tensor is proposed to solve the issue of the traditional TD methods **requiring pre-given rank**. Furthermore, a novel low-rank TD is proposed, which solves **the imbalance of the tensor unfolding matrix and explains the low rankness of TD using low-rank factor matrices and sparse core tensor**.

- Novel Tucker-based TC models are proposed

Inspired by **joint low rankness and smoothness priori modeling**, three TC models are proposed by integrating factor and tensor gradients within the enhanced TD methods.

- High-performance and convergent algorithms are developed

Two efficient algorithms, **the proximal alternating linearized minimization and the proximal alternating direction method**, are proposed to solve the corresponding TC models. Moreover, the proposed algorithms exhibit **global convergence in theoretical and numerical analyses**.

- Tucker-based TC optimization are established for high-dimensional visual data completion and multi-dimensional traffic data imputation

Numerical results demonstrate that the proposed novel Tucker-based TC models exhibit strong generalization ability for TC problems, even in **extreme missing scenarios**.

欢迎各位评审专家
批评指正！