Self-Attention Driven Tensor Representation for High-Order Data Recovery

Zhi-Wei Shi[1], Yu-Bang Zheng[1,*], Heng-Chao Li[1]
[1]Southwest Jiaotong University, China
E-mail: szw@my.swjtu.edu.cn

Highlight

CVPR JUNE 3-7, 2026 DENVER COLORADO

Background: Tensor Representation

★ Low-rank tensor representation (LRTR) is a technique for characterizing low-rank properties by reconstructing an accurate low-rank structure from a small number of latent factors through multilinear multiplication between elements.

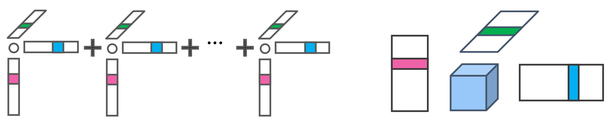


Figure 1. A graphical illustration of Classic LRTR

Motivations

★ Self-attention mechanisms can adaptively model the long- and short-range dependencies between factors and the global structural relationship, driving tensor decomposition to obtain more accurate and structurally adaptive low-rank representations.

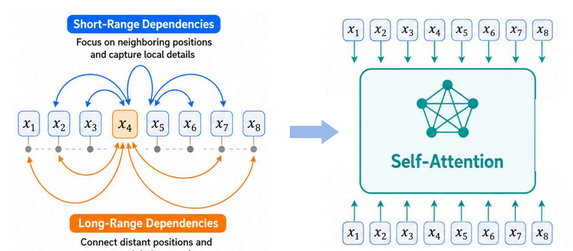
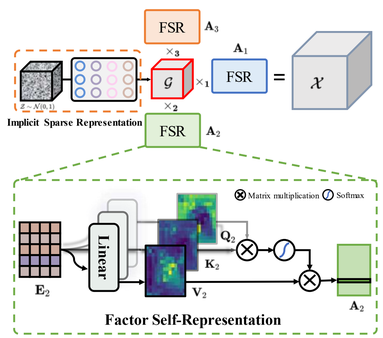


Figure 2. A graphical illustration of self-attention mechanism modeling long- and short-range dependencies.

Self-Attention Driven Tensor Representation

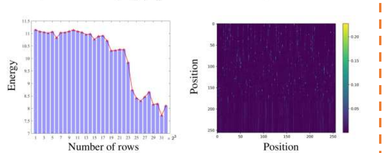
★ We construct a novel paradigm called Self-Attention Driven Tensor Representation (SADTR), which is the first framework that models nonlinearity from the perspective of self-attention. Specifically, we design a factor self-representation to capture both local and non-local nonlinear dependencies in the factor space. Moreover, we introduce an implicit sparse representation to impose sparsity constraint.



Definition 1 (SADTR). *Let $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ be any n-order tensor. Then, we define the self-attention driven tensor representation (SADTR) as follows:*

$$\mathcal{X} = \mathcal{H}_\Phi(\mathcal{Z}) \times_1 \mathrm{FSR}_{\Theta_1}(\mathbf{E}_1) \times_2 \mathrm{FSR}_{\Theta_2}(\mathbf{E}_2) \times_3 \cdots \times_n \mathrm{FSR}_{\Theta_n}(\mathbf{E}_n), \quad (1)$$

where $\mathrm{FSR}_{\Theta_i}(\cdot)$ denotes the factor-self representation for each mode, $\mathcal{H}_\Phi(\cdot)$ denotes the implicit sparse representation, and Φ and $\{\Theta_i\}_{i=1}^n$ are the learnable parameters.



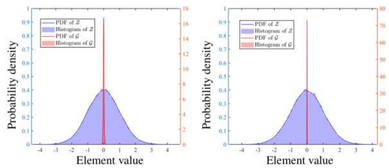
(a) Energy distribution of $\mathbf{Q}_1$ (b) Attention weight map $\mathbf{W}_1^a$

Figure 3. A graphical illustration of SADTR

☐ We theoretically expose the implicit sparse constraint on the tensor core.

Theorem 1 (Implicit Sparsity). *Let $\mathcal{G} = \mathcal{H}_\Phi(\mathcal{Z})$ be the core tensor parameterized by MLPs with parameters Φ, where $\mathcal{Z} \sim \mathcal{N}(0,1)$ sampled from the continuous Gaussian distribution. Assume that the weight matrices of MLPs satisfy $\|\mathbf{W}_i\|_1 \leq \gamma$ for $i = 1, 2, \cdots, l$, and the activation function $\sigma(\cdot)$ is δ-Lipschitz continuous. Then, for any $a, b \in \mathcal{Z}$, the following inequality holds:*

$$\|\mathcal{H}_\Phi(a) - \mathcal{H}_\Phi(b)\|_1 \leq \gamma^l \delta^{l-1} \|a - b\|_1,$$



☐ In theory, we provide an analysis to show the recoverability of SADTR.

Theorem 2 (Recoverability). *Under the tensor completion model (7), assume that $\hat{\mathcal{X}}^*$ such that $\hat{\mathcal{X}}^* = \hat{\mathcal{G}}^* \times_1 \hat{\mathbf{A}}_1^* \times_2 \hat{\mathbf{A}}_2^* \times_3 \hat{\mathbf{A}}_3^*$ is an optimal solution and satisfies $\hat{\mathcal{G}}^* = \mathcal{H}_\Phi(\mathcal{Z})$ and $\hat{\mathbf{A}}_n^* = \mathrm{FSR}_{\Theta_n}(\mathbf{Z}_n)$ for $n = 1, 2, 3$. Then, with probability at least $1 - \delta$, the following inequality holds:*

$$\frac{1}{\sqrt{N}}\|\hat{\mathcal{X}}^* - \mathcal{X}_\oplus\|_F \leq \frac{1}{\sqrt{|\Omega|}}\|\mathcal{X}^* - \mathcal{X}_\oplus\|_F + \mathrm{Gap}^*(\Omega) + \frac{1}{\sqrt{|\Omega|}}\|\mathcal{M} * \mathcal{N}\|_F + \frac{1}{\sqrt{N}}\|\mathcal{N}\|_F$$

Experiments and Discussions

★ We conduct a lot of experiments on various high-order data recovery tasks to verify the effectiveness of the proposed SADTR.

High-Order Data Inpainting

Methods	MR=95%		MR=90%		MR=80%	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3D MSIs: *Feathers*, *Flowers*, *Toys* (256 × 256 × 31)						
Observed	12.467	0.205	12.701	0.245	13.212	0.318
FTNN	30.558	0.891	35.063	0.949	40.499	0.981
FCTN	30.323	0.814	35.625	0.920	39.129	0.956
HTNN	23.686	0.687	31.679	0.880	36.775	0.952
HLRTF	31.298	0.869	37.219	0.957	44.296	0.988
LRTFR	33.934	0.926	39.030	0.974	43.441	0.987
OTLRM	34.667	0.946	39.219	0.975	43.561	0.988
SADTR*	35.935	0.957	40.720	0.982	46.936	0.986
SADTR	**36.759**	**0.974**	**41.580**	**0.990**	**48.018**	**0.996**
4D Color Videos: *News*, *Suzie*, *Foreman* (144 × 176 × 3 × 100)						
Observed	6.733	0.013	6.969	0.021	7.480	0.036
FTNN	25.337	0.759	28.542	0.848	32.234	0.920
FCTN	26.896	0.713	29.820	0.820	32.580	0.891
HTNN	24.710	0.701	27.994	0.829	31.752	0.922
HLRTF	26.215	0.720	28.718	0.836	33.577	0.915
LRTFR	26.771	0.754	30.195	0.856	32.517	0.899
OTLRM	28.273	0.806	31.032	0.877	33.610	0.925
SADTR*	28.357	0.843	32.168	0.907	35.162	0.938
SADTR	**29.732**	**0.856**	**32.668**	**0.913**	**35.903**	**0.953**

High-Order Data Denoising

Methods	Case 1		Case 2		Case 3	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MSIs: *Imgb2*, *Beers* (256 × 256 × 31)						
Observed	13.980	0.078	10.450	0.047	10.585	0.047
LRTDTV	32.830	0.870	29.779	0.835	27.883	0.804
E3DTV	31.514	0.875	29.852	0.821	27.555	0.795
FGSLR	32.715	0.894	30.762	0.876	29.441	0.853
HLRTF	34.112	0.889	32.110	0.870	29.732	0.790
LRTFR	34.021	0.894	32.014	0.874	29.533	0.798
HIR-Diff	35.001	0.907	25.124	0.790	25.099	0.790
SADTR*	35.726	0.911	33.143	0.871	30.483	0.867
SADTR	**36.251**	**0.929**	**33.586**	**0.899**	**31.104**	**0.872**
HSIs: *WDC mall* (256 × 256 × 150), *Pavia* (256 × 256 × 80)						
Observed	13.976	0.116	10.459	0.070	10.601	0.069
LRTDTV	31.923	0.825	30.370	0.804	27.536	0.737
E3DTV	33.022	0.871	31.336	0.839	27.444	0.782
FGSLR	31.988	0.872	30.554	0.837	27.842	0.797
HLRTF	32.767	0.855	31.106	0.836	28.108	0.744
LRTFR	31.537	0.831	30.433	0.817	26.725	0.650
HIR-Diff	**33.897**	**0.900**	25.018	0.786	24.134	0.756
SADTR*	33.168	0.868	31.259	0.832	29.047	0.803
SADTR	33.681	0.882	**31.809**	**0.850**	**29.427**	**0.810**