Deep Transform Framework with Adaptive Temporal Smoothing

(DTF-ATS) Continuous Low-Rank Priors for Tube-Wise Missing Tensor Completion

Hongyu Chen[1] Beining Bao[1] Zichun Liu[1] Tianshuo Wei[1] Maolin Wang[2] Xiaopeng Li[3] Hing Cheung So[4]

[1]Dept. of Data Science, City University of Hong Kong ▪ [2]HKAI-Sci, City University of Hong Kong ▪ [3]State Key Lab of RF Heterogeneous Integration, Shenzhen University ▪ [4]Dept. of Electrical Engineering, City University of Hong Kong

Problem: Tube-Wise Missing Data

High-dimensional visual data (illumination-varying faces, volumetric MRI) are tensors whose **third mode is a continuous physical coordinate** t (illumination angle, anatomical depth).

Under **tube-wise missing**, entire fibers along that mode are lost. Classical low-rank / deep-unfolding methods treat the third mode as a *discrete index*, so they:

- interpolate poorly over large missing intervals;
- and uniform TV regularization **over-smooths** structural edges.

Contributions

- A **dual-stream DTF-ATS** framework for tube-wise missing tensor completion.
- A problem-specific coupling of **CC-CP** (continuous low-rank prior) and **ATS** (adaptive regularization).
- Quantitative, qualitative, and ablation evidence supporting the design.

Continuous-Coordinate CP (CC-CP)

Replace the discrete third-mode factor with a coordinate-conditioned generator $\mathbf{h}(t) = g_\theta(t)$ (time encoder + MLP):

$$\mathcal{X}_{\mathrm{cp}}(i,j,t) = \sum_{r=1}^{R} \mathbf{A}_{ir}\,\mathbf{B}_{jr}\,\mathbf{h}_r(t)$$

Queryable at *any* $t \in [0,1]$ $\Rightarrow$ robust extrapolation across missing tubes.

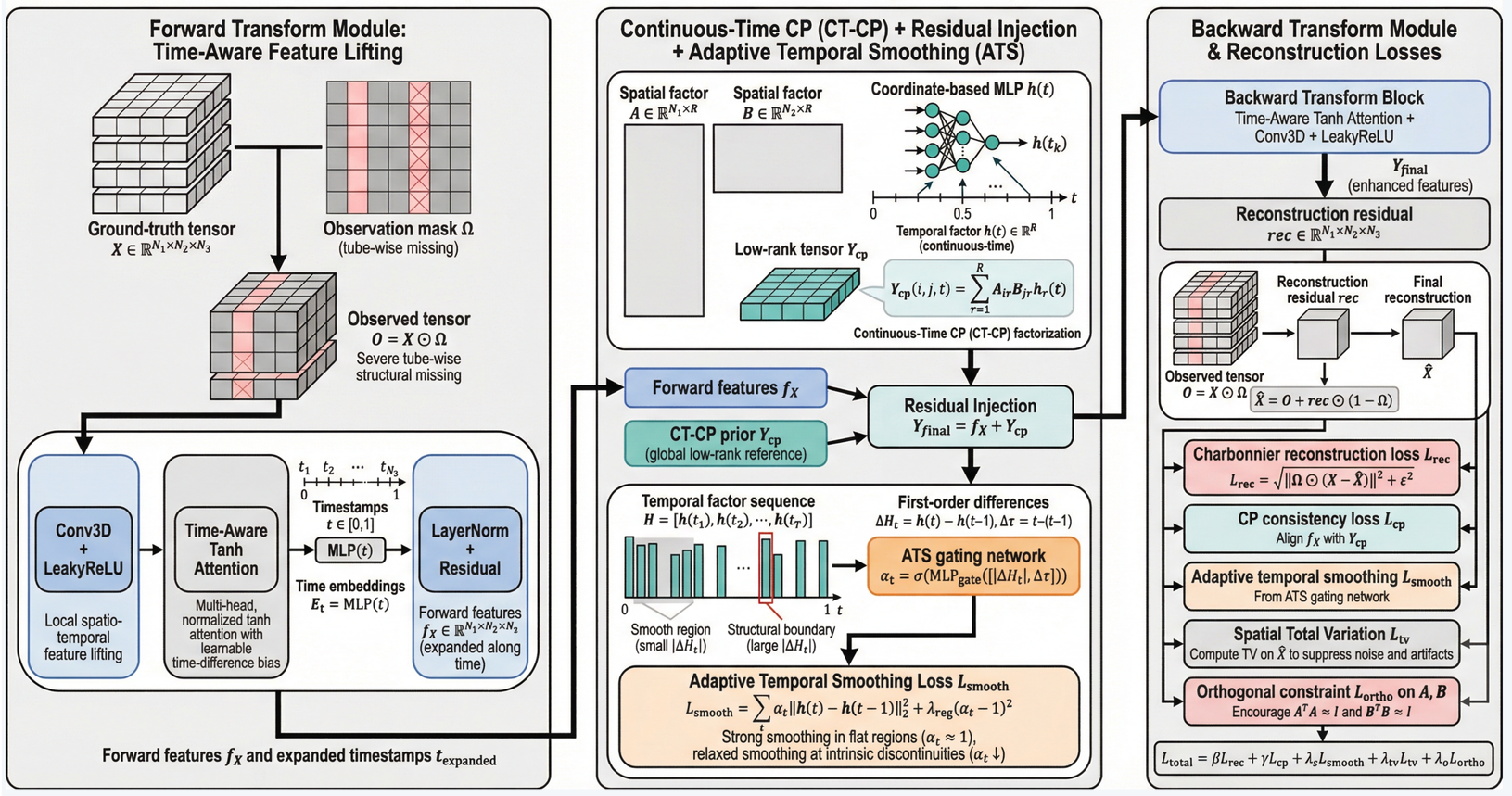
Adaptive Temporal Smoothing (ATS)

A learnable gate sets the smoothing strength from local factor variation:

$$\alpha_{t,r} = \sigma\big(\mathcal{G}_\phi([\,|\Delta h|,\ \Delta\tau\,])\big)$$
$$\mathcal{L}_{\mathrm{ats}} = \textstyle\sum_{t,r} \alpha_{t,r}(\Delta h_{t,r})^2 + \lambda_{\mathrm{reg}}(\alpha_{t,r}-1)^2$$

Flat regions $\alpha \to 1$ (strong smoothing); structural edges $\alpha \to 0$ (detail preserved).

Method: DTF-ATS Dual-Stream Framework



A **time-aware Main Stream** (Conv3D $1\times3\times3$ + LeakyReLU, time-aware tanh attention, residual LayerNorm; forward block $N_3 \to 4N_3$, backward block $4N_3 \to N_3$) captures local non-linear detail. The **Prior Stream (CC-CP)** injects a global continuous low-rank reference via residual fusion, and **data consistency** keeps observed entries exact:

$$\mathcal{X}_{\mathrm{rec}} = \mathcal{B}\big(\mathcal{F}(\mathcal{Z}_{\mathrm{latent}}, \mathbf{t}) + \mathcal{X}_{\mathrm{cp}},\ \mathbf{t}\big), \qquad \hat{\mathcal{X}} = \mathcal{O} + \mathrm{rec} \odot (1 - \Omega).$$

Training objective (jointly optimizing network + latent tensor):

$$\mathcal{L}_{\mathrm{total}} = \beta\mathcal{L}_{\mathrm{fid}} + \gamma\mathcal{L}_{\mathrm{cp}} + \lambda_s\mathcal{L}_{\mathrm{ats}} + \lambda_{\mathrm{tv}}\mathcal{L}_{\mathrm{tv}} + \lambda_{\mathrm{ortho}}\mathcal{L}_{\mathrm{ortho}}$$

($\mathcal{L}_{\mathrm{fid}}$: Charbonnier fidelity; $\mathcal{L}_{\mathrm{cp}}$: stream consistency; $\mathcal{L}_{\mathrm{tv}}$: spatial TV; $\mathcal{L}_{\mathrm{ortho}}$: factor orthogonality.)

Datasets & Implementation

- **Extended Yale B** $48\times100\times42$: 100 illumination frames; $\mathrm{SR} \in \{0.1, 0.3, 0.5\}$; rank $R \in \{18, 24, 32\}$; $\sim$1.69M params.
- **MRI** $181\times217\times181$: T1 brain, third mode = anatomical depth; $\mathrm{SR} \in \{0.1, 0.2, 0.3\}$; rank $R \in \{56, 80, 96\}$; $\sim$2.52M params.
- **Training:** full-tensor opt. (batch 1); Adam lr $2\times10^{-3}/2\times10^{-2}$ (net/latent); wd 10^{-5}; cosine annealing; 3100 iters.
- **Rank rule:** lower SR $\to$ smaller R (stronger bias); higher SR $\to$ larger R (more capacity).

Quantitative — Extended Yale B

Illumination dimension $\Rightarrow$ clear, uniform gains

Method	SR=0.1		SR=0.3		SR=0.5	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
HaLRTC	9.88	.118	18.20	.538	23.17	.790
TNN	15.66	.252	21.26	.652	26.09	.854
HLRTF	19.59	.510	23.27	.746	26.12	.852
KBR	15.98	.248	21.92	.686	25.57	.853
DTF-ATS	**19.79**	**.542**	**24.58**	**.810**	**28.99**	**.925**

+1.31 dB at SR=0.3 and +2.87 **dB at SR=0.5** over the strongest baseline; best PSNR *and* SSIM at every rate.

Quantitative — MRI (T1, $181^2\times181$)

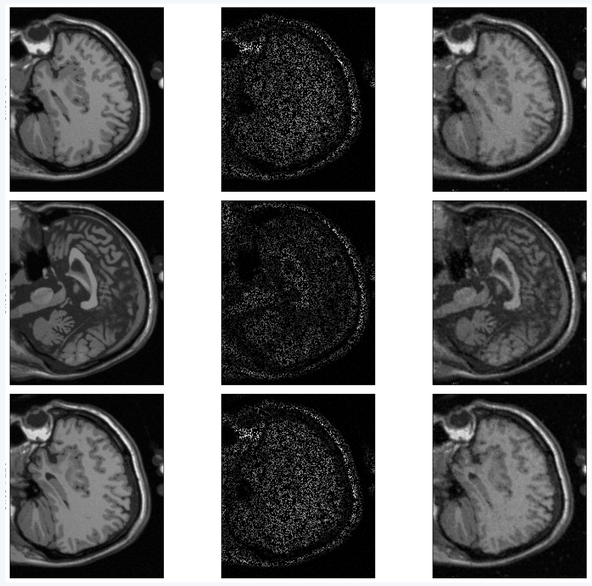
Anatomical depth $\Rightarrow$ competitive

Method	SR=0.1		SR=0.2		SR=0.3	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
TNN	18.96	.278	22.49	.460	25.52	.615
HLRTF	17.88	.378	24.87	**.688**	26.43	.763
KBR	17.76	.240	**25.53**	.611	29.73	.781
DTF-ATS	**19.03**	.334	25.42	.660	**29.84**	**.821**

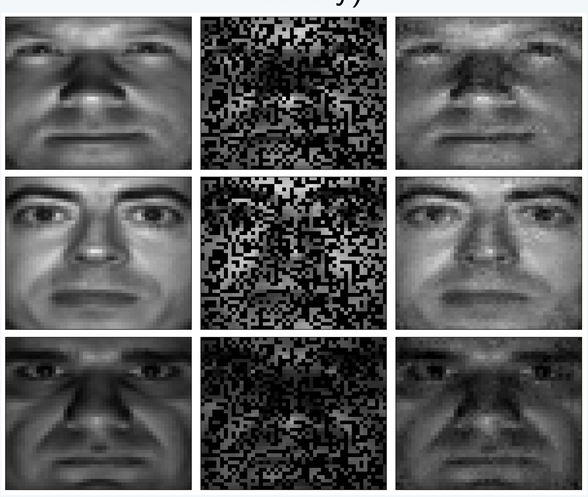
Best PSNR at SR=0.1 / 0.3; competitive at SR=0.2. Gains are strongest when the third mode follows a smooth physical trajectory.

Qualitative Recovery

MRI, SR=0.3 (GT ▪ Observed ▪ Recovery)



Extended Yale B, SR=0.5 (GT ▪ Observed ▪ Recovery)



Under severe tube-wise missing, DTF-ATS restores anatomical structures and shadow transitions while keeping sharp boundaries intact.

Ablation (Yale B, SR=0.1)

Component	Setting	PSNR	SSIM
Full Model	**CC-CP + ATS**	**19.79**	**.542**
CC-CP	Discrete CP	19.75	.540
	Static CP	19.52	.529
ATS	w/o ATS	19.01	.502
	Standard TV	19.18	.498
Streams	Main only	19.69	.541
	Prior only	15.62	.284

Removing ATS: -0.78 **dB**; ATS beats global TV by +0.61 **dB**. Prior-only collapses to 15.62 dB $\Rightarrow$ the two streams are **complementary**.

Key Takeaways

- Model the third mode as a **continuous coordinate**, not a discrete index.
- CC-CP + ATS = **continuous low-rank prior** with **content-aware** smoothing.
- Largest gains where the coordinate is physically smooth (illumination); competitive elsewhere (depth).

Conclusion & Future Work

DTF-ATS couples continuous low-rank priors with deep non-linear transforms for tube-wise missing recovery, with the clearest gains on Yale B and competitive MRI results. **Limitation:** extra compute from the unfolding backbone, attention, and coordinate prior. **Next:** efficiency profiling and extension to dynamic video completion.