Learning Accurate Representation to Nonstandard Tensors via a Mode-Aware Tucker Network

Hao Wu, *Member, IEEE*, Qu Wang, Xin Luo, *Fellow, IEEE*, and Zidong Wang, *Fellow, IEEE*

I. INTRODUCTION

THIS is the supplementary file for paper entitled *Learning Accurate Representation to Nonstandard Tensors via a Mode-Aware Tucker Network*. Supplementary equations and experimental results are put into this file.

II. SUPPLEMENTARY PROOF OF CONVERGENCE

First, we review the formula for MSGD and rewrite it to separate the deterministic gradient part from the random noise part.

$$\begin{aligned}
&v_{n+1} = \gamma v_n + \eta \nabla_{\omega_n} J\left(\omega_n\right) + \eta\left(\nabla_{\omega_n} J\left(\omega_n, \xi_n\right) - \nabla_{\omega_n} J\left(\omega_n\right)\right), \\
&\omega_{n+1} = \omega_n - v_{n+1}.
\end{aligned} \tag{S1}$$

This transformation borrows the idea of random approximation. Then calculate the following Taylor expansion and do some processing to get:

$$J\left(\omega_{n+1}\right) - J\left(\omega_n\right) = -\nabla_\omega J(\omega_n)^T v_n + \tfrac{1}{2} v_n^T H_{\omega\omega}\left(\zeta_n\right) v_n. \tag{S2}$$

Using the above Taylor formula, we can make the following calculation:

$$\begin{aligned}
&\nabla_{\omega_n} J(\omega_n)^T v_n \\
&= \left(\nabla_{\omega_{n-1}-v_{n-1}} J\left(\omega_{n-1} - v_{n-1}\right)\right)^T \left(\gamma v_{n-1} + \eta_n \nabla_{\omega_n} J\left(\omega_n, \xi_n\right)\right) \\
&= \gamma\left(\nabla_{\omega_{n-1}} J\left(\omega_{n-1}\right) - H_{\omega\omega}\left(\zeta_{n-1}\right) v_{n-1}\right)^T v_{n-1} + \eta_n \nabla_{\omega_n} J\left(\omega_n\right) \nabla_{\omega_n} J\left(\omega_n, \xi_n\right) \\
&= \gamma \nabla_{\omega_{n-1}} J(\omega_{n-1})^T v_{n-1} - \gamma v_{n-1}^T H_{\omega\omega}\left(\zeta_{n-1}\right) v_{n-1} + \eta_n \nabla_{\omega_n} J\left(\omega_n\right) \nabla_{\omega_n} J\left(\omega_n, \xi_n\right).
\end{aligned} \tag{S3}$$

Next, processing the above recursive formula, we get:

$$\nabla_{\omega_n} J(\omega_n)^T v_n = \gamma \nabla_{\omega_1} J(\omega_1)^T v_1 - \sum_{i=1}^{n-1} \gamma v_i^T H_{\omega\omega}\left(\zeta_i\right) v_i + \sum_{i=2}^{t} \gamma \eta_i \nabla_{\omega_i} J\left(\omega_i\right) \nabla_{\omega_i} J\left(\omega_i, \zeta_i\right). \tag{S4}$$

Then, bring (S4) back to (S2),

$$J\left(\omega_{n+1}\right) - J\left(\omega_n\right) = -\gamma \nabla_{\omega_1} J(\omega_1)^T v_1 - \sum_{i=2}^{t} \gamma \eta_i \nabla_{\omega_i} J\left(\omega_i\right) \nabla_{\omega_i} J\left(\omega_i, \xi_i\right) + \sum_{i=1}^{n-1} \gamma v_i^T H_{\omega\omega}\left(\zeta_i\right) v_i + \tfrac{1}{2} v_n^T H_{\omega\omega}\left(\zeta_n\right) v_n. \tag{S5}$$

Next, we will show that $J(\omega_n + 1)$ is almost certainly convergent. We do this by recursing on (S5) and dividing it into three parts:

$$
\begin{aligned}
J\left(\omega_{t+1}\right) = & \underbrace{J\left(\omega_1\right) - \frac{\gamma-\gamma^{t+1}}{1-\gamma} \nabla_{\omega_1} J(\omega_1)^T v_1 + \frac{1-\gamma^t}{1-\gamma} \eta_1 \nabla_{\omega_1} J\left(\omega_1\right) \nabla_{\omega_1} J\left(\omega_1, \xi_1\right)}_{A} \\
& \underbrace{- \sum_{n=1}^{t} \frac{1-\gamma^{t-n+1}}{1-\gamma} \eta_n \nabla_{\omega_n} J(\omega_n)^T \nabla_{\omega_n} J\left(\omega_n, \xi_n\right)}_{B} \\
& \underbrace{- \frac{1}{2} \sum_{n=1}^{t} v_n^T H_{\omega\omega}\left(\zeta_n\right) v_n + \sum_{n=1}^{t} \frac{1-\gamma^{t-n+1}}{1-\gamma} v_n^T H_{\omega\omega}\left(\zeta_n\right) v_n}_{C} .
\end{aligned}
\tag{S6}
$$

Because $a < 1$, an is convergent, which ensures that part A is converged. Then, Lemma 3 ensures that part C converges almost everywhere. For Part B, we have the following proof:

$$
\begin{aligned}
B = & \sum_{n=1}^{t} \tfrac{1-\gamma^{t-n+1}}{1-\gamma} \eta_n \nabla_{\omega_n} J(\omega_n)^T \nabla_{\omega_n} J\left(\omega_n, \xi_n\right) \\
= & \sum_{n=1}^{t} \tfrac{1-\gamma^{t-n+1}}{1-\gamma} \eta_n \left\| \nabla_{\omega_n} J\left(\omega_n\right) \right\|^2 \\
+ & \sum_{n=1}^{t} \tfrac{1-\gamma^{t-n+1}}{1-\gamma} \eta_n \nabla_{\omega_n} J(\omega_n)^T \left(\nabla_{\omega_n} J\left(\omega_n, \xi_n\right) - \nabla_{\omega_n} J\left(\omega_n, \xi_n\right) \right).
\end{aligned}
\tag{S7}
$$

By Lemma 4, we have:

$$
\sum_{n=1}^{t} \tfrac{1-\gamma^{t-n+1}}{1-\gamma} \eta_n \left\| \nabla_{\omega_n} J\left(\omega_n\right) \right\|^2 < +\infty.
\tag{S8}
$$

According to Lemma 1 and Lemma 4, we can conclude

$$
\sum_{n=1}^{t-1} \tfrac{1-\gamma^{t-n}}{1-\gamma} \eta_{n+1} \nabla_{\omega_n} J(\omega_n)^T \left(\nabla_{\omega_n} J\left(\omega_n, \xi_n\right) - \nabla_{\omega_n} J\left(\omega_n, \xi_n\right) \right).
\tag{S9}
$$

The above formula is convergent. So B is convergent, and $g(\omega_t + 1)$ is also convergent. Substituting (S7) into (S6) we get:

$$
J\left(\omega_{t+1}\right) = \zeta_t - \sum_{n=1}^{t} \eta_n \left\| \nabla_{\omega_n} J\left(\omega_n\right) \right\|^2,
\tag{S10}
$$

where $\{\zeta_t\}$ is defined as follows:

$$
\begin{aligned}
\zeta_t = & J\left(\omega_1\right) - \tfrac{\gamma-\gamma^{t+1}}{1-\gamma} \nabla_{\omega_1} J(\omega_1)^T v_1 + \tfrac{1-\gamma^t}{1-\gamma} \eta_1 \nabla_{\omega_1} J\left(\omega_1\right) \nabla_{\omega_1} J\left(\omega_1, \xi_1\right) \\
+ & \sum_{n=1}^{t} \tfrac{1-\gamma^{t-n+1}}{1-\gamma} \eta_n \nabla_{\omega_n} J(\omega_n)^T \left(\nabla_{\omega_n} J\left(\omega_n, \xi_n\right) - \nabla_{\omega_n} J\left(\omega_n\right) \right) \\
- & \tfrac{1}{2} \sum_{n=1}^{t} v_n^T H_{\omega\omega}\left(\zeta_n\right) v_n + \sum_{n=1}^{t} \tfrac{1-\gamma^{t-n+1}}{1-\gamma} v_n^T H_{\omega\omega}\left(\zeta_n\right) v_n.
\end{aligned}
\tag{S11}
$$

Since $\{\zeta_t\}$ almost certainly converges, Lemma 2 shows that,

$$
\omega_t \to \omega^*.
\tag{S12}
$$

III. Supplementary Tables and Figures

Here are some supplementary tables and figures in the Experiments section.

TABLE S1
HYPER-PARAMETERS SETTINGS OF M1-10 ON D1-8

Datasets	Hyper-parameter settings				
D1	M1: Adaptive	M2: η=2^{-7}, λ=2^{-6}	M3: η=2^{-10}, λ=2^{-8}	M4: η=2^{-7}, λ=2^{-8}	M5: P=1.9, λ=2^{-6}
	M6: λ_1=2^{-10}, λ_2=2^{-9}	M7: λ_1=2^{-6}, λ_2=2^{-2}	M8: λ_1=2^{-3}, λ_2=2^{-4}	M9: η=2^{-9}, λ=2^{-10}	M10: α=2^{-0}, β=2^{-8}, λ=2^{-7}
D2	M1: Adaptive	M2: η=2^{-6}, λ=2^{-6}	M3: η=2^{-10}, λ=2^{-10}	M4: η=2^{-8}, λ=2^{-6}	M5: P=1.9, λ=2^{-5}
	M6: λ_1=2^{-8}, λ_2=2^{-11}	M7: λ_1=2^{-3}, λ_2=2^{-2}	M8: λ_1=2^{-1}, λ_2=2^{-3}	M9: η=2^{-8}, λ=2^{-6}	M10: α=2^{-0}, β=2^{-9}, λ=2^{-6}
D3	M1: Adaptive	M2: η=2^{-7}, λ=2^{-6}	M3: η=2^{-9}, λ=2^{-12}	M4: η=2^{-7}, λ=2^{-8}	M5: P=1.9, λ=2^{-5}
	M6: λ_1=2^{-7}, λ_2=2^{-10}	M7: λ_1=2^{-7}, λ_2=2^{-1}	M8: λ_1=2^{-9}, λ_2=2^{-3}	M9: η=2^{-8}, λ=2^{-2}	M10: α=2^{-1}, β=2^{-7}, λ=2^{-6}
D4	M1: Adaptive	M2: η=2^{-7}, λ=2^{-6}	M3: η=2^{-9}, λ=2^{-12}	M4: η=2^{-8}, λ=2^{-5}	M5: P=1.9, λ=2^{-5}
	M6: λ_1=2^{-6}, λ_2=2^{-8}	M7: λ_1=2^{-5}, λ_2=2^{-2}	M8: λ_1=2^{-0}, λ_2=2^{-4}	M9: η=2^{-7}, λ=2^{-7}	M10: α=2^{-0}, β=2^{-4}, λ=2^{-5}
D5	M1: Adaptive	M2: η=2^{-7}, λ=2^{-6}	M3: η=2^{-10}, λ=2^{-11}	M4: η=2^{-6}, λ=2^{-9}	M5: P=1.7, λ=2^{-6}
	M6: λ_1=2^{-7}, λ_2=2^{-9}	M7: λ_1=2^{-12}, λ_2=2^{-6}	M8: λ_1=2^{-2}, λ_2=2^{-4}	M9: η=2^{-9}, λ=2^{-11}	M10: α=2^{-0}, β=2^{-3}, λ=2^{-8}
D6	M1: Adaptive	M2: η=2^{-7}, λ=2^{-7}	M3: η=2^{-9}, λ=2^{-12}	M4: η=2^{-5}, λ=2^{-10}	M5: P=1.9, λ=2^{-5}
	M6: λ_1=2^{-8}, λ_2=2^{-11}	M7: λ_1=2^{-9}, λ_2=2^{-6}	M8: λ_1=2^{-1}, λ_2=2^{-4}	M9: η=2^{-9}, λ=2^{-10}	M10: α=2^{-2}, β=2^{-1}, λ=2^{-8}
D7	M1: Adaptive	M2: η=2^{-10}, λ=2^{-7}	M3: η=2^{-10}, λ=2^{-12}	M4: η=2^{-5}, λ=2^{-7}	M5: P=1.9, λ=2^{-5}
	M6: λ_1=2^{-9}, λ_2=2^{-12}	M7: λ_1=2^{-12}, λ_2=2^{-6}	M8: λ_1=2^{-0}, λ_2=2^{-5}	M9: η=2^{-9}, λ=2^{-9}	M10: α=2^{-2}, β=2^{-2}, λ=2^{-8}
D8	M1: Adaptive	M2: η=2^{-7}, λ=2^{-7}	M3: η=2^{-8}, λ=2^{-12}	M4: η=2^{-5}, λ=2^{-9}	M5: P=1.9, λ=2^{-5}
	M6: λ_1=2^{-8}, λ_2=2^{-11}	M7: λ_1=2^{-5}, λ_2=2^{-6}	M8: λ_1=2^{-0}, λ_2=2^{-4}	M9: η=2^{-8}, λ=2^{-10}	M10: α=2^{-2}, β=2^{-3}, λ=2^{-11}

TABLE S2
TOTAL TIME OF ALL TEST MODELS ON D1-8. (UNIT: MINUTES. NOTE THAT THE TIME INCLUDES THE TIME TO ADJUST THE HYPER-PARAMETERS.)

Models	D1	D2	D3	D4	D5	D6	D7	D8	Win/Loss
				Time cost in RMSE↓					
M1	5.8$_{\pm 3.3E-01}$	**3.3**$_{\pm 6.2E-01}$	2.0$_{\pm 2.4E-01}$	**0.7**$_{\pm 6.1E-02}$	28.0$_{\pm 3.1E-00}$	16.4$_{\pm 7.2E-01}$	8.1$_{\pm 4.5E-01}$	5.6$_{\pm 1.0E-00}$	–
M2	308.1$_{\pm 1.3E+01}$	71.7$_{\pm 1.0E+00}$	99.0$_{\pm 2.5E+00}$	32.0$_{\pm 1.0E+00}$	74.2$_{\pm 1.1E+00}$	74.4$_{\pm 1.9E+01}$	295.0$_{\pm 2.6E+00}$	22.3$_{\pm 7.4E-01}$	8/0
M3	306.7$_{\pm 2.1E+01}$	82.7$_{\pm 3.7E-01}$	23.7$_{\pm 1.7E-01}$	11.0$_{\pm 2.1E-01}$	28.8$_{\pm 1.5E-00}$	**13.3**$_{\pm 7.3E-02}$	18.6$_{\pm 1.7E-01}$	1.8$_{\pm 5.4E-02}$	6/2
M4	50.5$_{\pm 4.5E-00}$	80.6$_{\pm 5.6E-00}$	22.8$_{\pm 5.6E-01}$	20.5$_{\pm 3.9E-01}$	48.9$_{\pm 7.8E-01}$	38.1$_{\pm 2.1E-00}$	5.2$_{\pm 1.8E-00}$	11.1$_{\pm 5.3E-01}$	7/1
M5	40.3$_{\pm 5.2E-00}$	43.0$_{\pm 4.7E-00}$	33.2$_{\pm 3.2E-01}$	11.5$_{\pm 4.8E-01}$	267.4$_{\pm 6.3E-00}$	125.2$_{\pm 1.0E-01}$	115.5$_{\pm 5.6E-01}$	86.1$_{\pm 6.5E-00}$	8/0
M6	**2.2**$_{\pm 6.8E-01}$	5.9$_{\pm 6.4E-01}$	**1.3**$_{\pm 4.8E-02}$	0.8$_{\pm 7.2E-02}$	**3.4**$_{\pm 2.7E-02}$	17.6$_{\pm 3.4E-00}$	38.6$_{\pm 2.2E-01}$	10.5$_{\pm 3.4E-01}$	5/3
M7	67.4$_{\pm 1.4E+01}$	56.2$_{\pm 6.4E-00}$	17.4$_{\pm 6.8E-00}$	20.3$_{\pm 5.5E-00}$	139.9$_{\pm 1.1E-+01}$	224.2$_{\pm 6.6E-00}$	123.9$_{\pm 8.5E-00}$	246.6$_{\pm 2.1E-00}$	8/0
M8	5.1$_{\pm 1.3E-00}$	16.6$_{\pm 2.6E-00}$	10.9$_{\pm 6.0E-02}$	1.3$_{\pm 1.9E-01}$	55.1$_{\pm 1.2E+01}$	39.0$_{\pm 6.4E-00}$	5.2$_{\pm 1.3E-01}$	**1.6**$_{\pm 1.0E-01}$	5/3
M9	131.5$_{\pm 2.8E+01}$	45.5$_{\pm 1.2E-00}$	36.6$_{\pm 9.3E-01}$	9.7$_{\pm 4.3E-01}$	19.1$_{\pm 5.1E-01}$	16.3$_{\pm 4.8E-01}$	63.9$_{\pm 1.6E-00}$	8.8$_{\pm 5.9E-01}$	6/2
M10	222.1$_{\pm 8.8E-00}$	202.2$_{\pm 5.9E-00}$	42.8$_{\pm 4.3E-00}$	1.9$_{\pm 7.6E-02}$	127.9$_{\pm 3.1E-00}$	109.7$_{\pm 5.2E-01}$	95.2$_{\pm 1.6E-00}$	32.0$_{\pm 9.7E-01}$	8/0
				Time cost in MAE↓					
M1	8.1$_{\pm 2.0E-01}$	**4.1**$_{\pm 7.8E-01}$	2.8$_{\pm 8.9E-02}$	**0.6**$_{\pm 5.8E-02}$	48.1$_{\pm 1.2E-00}$	26.7$_{\pm 4.2E-00}$	**15.0**$_{\pm 2.2E-00}$	8.3$_{\pm 1.0E-00}$	–
M2	515.8$_{\pm 2.2E+01}$	133.6$_{\pm 4.6E+00}$	190.6$_{\pm 5.5E+00}$	34.7$_{\pm 1.5E+00}$	252.7$_{\pm 1.4E+00}$	192.6$_{\pm 3.5E+01}$	294.7$_{\pm 3.2E+00}$	46.3$_{\pm 1.4E+00}$	8/0
M3	328.2$_{\pm 2.0E+01}$	106.2$_{\pm 1.8E-00}$	41.9$_{\pm 4.2E-01}$	13.2$_{\pm 1.9E-01}$	117.3$_{\pm 2.9E-00}$	63.5$_{\pm 1.6E-00}$	62.7$_{\pm 1.0E-00}$	**4.0**$_{\pm 4.4E-02}$	7/1
M4	70.2$_{\pm 6.4E-00}$	110.8$_{\pm 4.6E-00}$	34.0$_{\pm 1.2E-00}$	18.4$_{\pm 4.5E-01}$	117.4$_{\pm 2.6E-00}$	94.1$_{\pm 3.1E-00}$	22.9$_{\pm 7.3E-01}$	18.7$_{\pm 6.4E-01}$	8/0
M5	40.0$_{\pm 3.7E-00}$	35.3$_{\pm 1.9E-00}$	27.5$_{\pm 8.1E-01}$	10.9$_{\pm 4.8E-01}$	197.8$_{\pm 9.0E-00}$	91.8$_{\pm 4.7E-00}$	82.1$_{\pm 5.5E-01}$	53.2$_{\pm 3.6E-00}$	8/0
M6	**2.2**$_{\pm 6.8E-01}$	5.3$_{\pm 2.6E-02}$	**1.2**$_{\pm 1.5E-01}$	0.9$_{\pm 7.2E-02}$	**3.5**$_{\pm 3.2E-02}$	17.9$_{\pm 3.8E-00}$	36.8$_{\pm 7.4E-01}$	10.6$_{\pm 2.2E-01}$	4/4
M7	153.2$_{\pm 3.6E-00}$	71.2$_{\pm 2.9E-00}$	29.7$_{\pm 1.1E-00}$	20.3$_{\pm 2.4E-00}$	145.6$_{\pm 1.0E+01}$	205.9$_{\pm 1.1E+01}$	146.9$_{\pm 1.0E+01}$	235.0$_{\pm 5.5E-00}$	8/0
M8	46.8$_{\pm 6.6E-00}$	19.2$_{\pm 1.1E-00}$	11.8$_{\pm 5.0E-02}$	7.3$_{\pm 2.7E-01}$	55.8$_{\pm 7.5E-00}$	41.6$_{\pm 3.1E-00}$	40.8$_{\pm 9.0E-00}$	12.0$_{\pm 2.7E-00}$	8/0
M9	131.0$_{\pm 2.8E+01}$	52.8$_{\pm 2.6E-00}$	50.0$_{\pm 1.9E-00}$	11.0$_{\pm 3.8E-01}$	19.1$_{\pm 5.1E-01}$	**16.3**$_{\pm 4.8E-01}$	64.7$_{\pm 1.4E-00}$	12.9$_{\pm 7.1E-01}$	6/2
M10	389.5$_{\pm 2.0E+01}$	242.0$_{\pm 2.7E-00}$	66.7$_{\pm 3.2E-00}$	3.6$_{\pm 3.0E-01}$	127.6$_{\pm 3.1E-00}$	109.7$_{\pm 5.2E-01}$	95.2$_{\pm 1.6E-00}$	32.0$_{\pm 9.7E-01}$	8/0
				Time cost in R^2↓					
M1	5.8$_{\pm 3.3E-01}$	**3.2**$_{\pm 8.2E-01}$	2.3$_{\pm 1.5E-01}$	**0.7**$_{\pm 6.1E-02}$	33.0$_{\pm 3.1E-00}$	16.6$_{\pm 5.1E-01}$	8.1$_{\pm 3.6E-01}$	5.6$_{\pm 1.0E-00}$	–
M2	366.2$_{\pm 1.8E+01}$	78.2$_{\pm 1.1E+00}$	123.1$_{\pm 2.0E+00}$	34.6$_{\pm 1.2E+00}$	78.7$_{\pm 1.4E+00}$	83.2$_{\pm 2.3E+01}$	294.7$_{\pm 3.2E+00}$	22.5$_{\pm 8.1E-01}$	8/0
M3	322.1$_{\pm 2.3E+01}$	101.6$_{\pm 1.1E-00}$	25.5$_{\pm 2.9E-01}$	11.6$_{\pm 2.1E-01}$	30.6$_{\pm 1.3E-00}$	**13.9**$_{\pm 3.1E-01}$	20.3$_{\pm 1.7E-01}$	1.8$_{\pm 6.6E-02}$	5/3
M4	53.0$_{\pm 3.3E-00}$	110.7$_{\pm 6.2E-00}$	24.5$_{\pm 5.6E-01}$	20.7$_{\pm 3.0E-01}$	65.8$_{\pm 7.9E-01}$	55.0$_{\pm 2.1E-00}$	**5.1**$_{\pm 2.8E-01}$	18.5$_{\pm 6.6E-01}$	7/1
M5	40.3$_{\pm 1.2E-00}$	42.1$_{\pm 3.7E-00}$	33.6$_{\pm 6.6E-01}$	12.0$_{\pm 9.7E-01}$	266.5$_{\pm 4.7E-00}$	124.7$_{\pm 8.1E-00}$	117.3$_{\pm 3.1E-00}$	84.4$_{\pm 5.4E-00}$	8/0
M6	**2.2**$_{\pm 6.8E-01}$	5.9$_{\pm 6.4E-01}$	**1.3**$_{\pm 4.8E-02}$	0.8$_{\pm 7.2E-02}$	**3.4**$_{\pm 4.4E-02}$	17.6$_{\pm 3.4E-00}$	38.6$_{\pm 1.8E-01}$	10.4$_{\pm 1.7E-01}$	5/3
M7	72.2$_{\pm 2.4E+01}$	73.3$_{\pm 4.9E-00}$	21.0$_{\pm 9.3E-00}$	23.8$_{\pm 3.6E-00}$	139.9$_{\pm 1.1E+01}$	229.8$_{\pm 8.2E-00}$	131.8$_{\pm 7.2E-00}$	249.1$_{\pm 2.8E-00}$	8/0
M8	5.1$_{\pm 1.3E-00}$	16.6$_{\pm 2.6E-00}$	10.9$_{\pm 6.0E-02}$	1.3$_{\pm 1.9E-01}$	56.9$_{\pm 1.0E+01}$	39.0$_{\pm 6.4E-00}$	5.2$_{\pm 6.4E-02}$	**1.6**$_{\pm 1.0E-01}$	5/3
M9	131.5$_{\pm 2.8E+01}$	46.8$_{\pm 1.1E-00}$	39.3$_{\pm 3.4E-00}$	9.9$_{\pm 3.5E-01}$	19.1$_{\pm 5.1E-01}$	16.3$_{\pm 4.8E-01}$	64.7$_{\pm 1.4E-00}$	9.3$_{\pm 5.9E-01}$	6/2
M10	233.1$_{\pm 4.3E-01}$	203.7$_{\pm 5.7E-00}$	43.5$_{\pm 4.7E-00}$	1.9$_{\pm 7.6E-02}$	127.9$_{\pm 3.1E-00}$	109.7$_{\pm 5.2E-01}$	95.2$_{\pm 1.6E-00}$	32.0$_{\pm 9.7E-01}$	8/0

TABLE S3
EXPERIMENTAL RESULTS ON D5 OF THE PARALLEL DESIGN OF MTN-TRL

Model	Accuracy			Time Cost		
	RMSE↓	MAE↓	R^2↑	Time in RMSE	Time in MAE	Time in R^2
MTN-TRL	$0.2662_{\pm 1.6\text{E-}04}$	$0.1905_{\pm 2.1\text{E-}04}$	$0.2913_{\pm 8.3\text{E-}04}$	$28.0_{\pm 3.1\text{E-}00}$	$48.1_{\pm 1.2\text{E-}00}$	$33.0_{\pm 3.1\text{E-}00}$
MTN-TRL-P	$0.2663_{\pm 2.1\text{E-}04}$	$0.1906_{\pm 1.7\text{E-}04}$	$0.2911_{\pm 7.5\text{E-}04}$	$13.7_{\pm 1.2\text{E-}00}$	$23.6_{\pm 8.5\text{E-}01}$	$15.2_{\pm 1.2\text{E-}00}$

TABLE S4
FRIEDMAN TEST RESULTS ON ACCURACY (RMSE, MAE, R^2) AND EFFICIENCY (TIME COST IN RMSE, TIME COST IN MAE, TIME COST IN R^2)

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Accuracy	**1.00**	2.42	4.96	3.04	6.46	8.17	5.67	6.08	7.63	9.58
Efficiency	**2.25**	8.83	5.21	5.77	6.79	2.38	7.88	3.31	4.88	7.71

TABLE S5
WILCOXON SIGNED-RANKS TEST RESULTS

Comparison	Accuracy			Efficiency		
	R+	*R-*	*p*-value*	*R+*	*R-*	*p*-value*
M2 vs M1	300	0	6.0E-08	300	0	6.0E-08
M3 vs M1	300	0	6.0E-08	273	27	7.5E-05
M4 vs M1	300	0	6.0E-08	297	3	3.0E-07
M5 vs M1	300	0	6.0E-08	300	0	6.0E-08
M6 vs M1	300	0	6.0E-08	155	144	4.5E-01
M7 vs M1	300	0	6.0E-08	300	0	6.0E-08
M8 vs M1	300	0	6.0E-08	265	35	2.5E-04
M9 vs M1	300	0	6.0E-08	259	41	5.5E-04
M10 vs M1	300	0	6.0E-08	300	0	6.0E-08

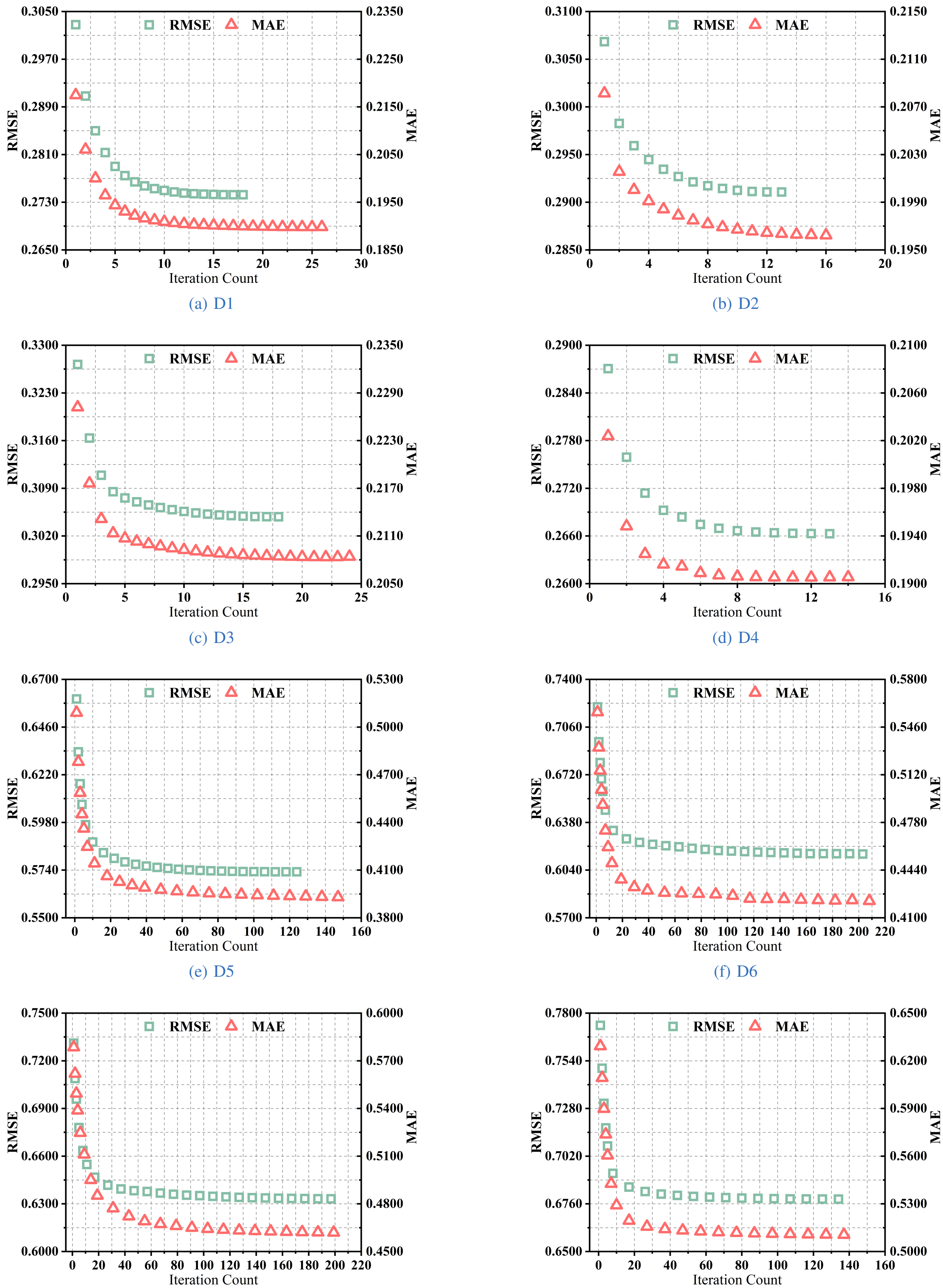
*Accept the hypothesis if the *p*-value < 0.1.

TABLE S6
PERFORMANCE OF MTN-TRL WITH AND WITHOUT HYPER-PARAMETERS ADAPTATION (TIME UNIT: MINUTE)

Datasets	RMSE↓		MAE↓		R^2↑		Time in RMSE↓		Time in MAE↓		Time in R^2↓	
	Adaptive	Manual	Adaptive	Manual	Adaptive	Manual	Adaptive	Manual	Adaptive	Manual	Adaptive	Manual
D1	$\mathbf{0.2745_{\pm 2E\text{-}4}}$	$0.2749_{\pm 2\text{E-}5}$	$\mathbf{0.1900_{\pm 2E\text{-}4}}$	$0.1905_{\pm 6\text{E-}6}$	$\mathbf{0.3717_{\pm 1E\text{-}3}}$	$0.3701_{\pm 4\text{E-}5}$	$\mathbf{5.8_{\pm 3E\text{-}1}}$	$165.8_{\pm 2\text{E-}0}$	$\mathbf{8.1_{\pm 2E\text{-}1}}$	$218.3_{\pm 2\text{E-}0}$	$\mathbf{5.8_{\pm 3E\text{-}1}}$	$168.0_{\pm 1\text{E-}0}$
D2	$\mathbf{0.2904_{\pm 6E\text{-}5}}$	$0.2915_{\pm 1\text{E-}4}$	$\mathbf{0.1957_{\pm 2E\text{-}4}}$	$0.1965_{\pm 6\text{E-}5}$	$\mathbf{0.2764_{\pm 4E\text{-}4}}$	$0.2681_{\pm 7\text{E-}4}$	$\mathbf{3.3_{\pm 6E\text{-}1}}$	$82.3_{\pm 3\text{E-}0}$	$\mathbf{4.1_{\pm 8E\text{-}1}}$	$96.8_{\pm 4\text{E-}0}$	$\mathbf{3.2_{\pm 8E\text{-}1}}$	$84.0_{\pm 3\text{E-}0}$
D3	$\mathbf{0.3041_{\pm 2E\text{-}4}}$	$0.3042_{\pm 4\text{E-}5}$	$\mathbf{0.2077_{\pm 1E\text{-}4}}$	$0.2080_{\pm 6\text{E-}5}$	$\mathbf{0.2660_{\pm 8E\text{-}4}}$	$0.2655_{\pm 2\text{E-}4}$	$\mathbf{2.0_{\pm 2E\text{-}1}}$	$60.6_{\pm 1\text{E-}0}$	$\mathbf{2.8_{\pm 9E\text{-}2}}$	$70.5_{\pm 9\text{E-}1}$	$\mathbf{2.3_{\pm 2E\text{-}1}}$	$67.9_{\pm 2\text{E-}0}$
D4	$0.2662_{\pm 2\text{E-}4}$	$\mathbf{0.2660_{\pm 8E\text{-}5}}$	$\mathbf{0.1905_{\pm 2E\text{-}4}}$	$0.1908_{\pm 7\text{E-}5}$	$0.2913_{\pm 8\text{E-}4}$	$\mathbf{0.2918_{\pm 4E\text{-}4}}$	$\mathbf{0.7_{\pm 6E\text{-}2}}$	$19.7_{\pm 7\text{E-}1}$	$\mathbf{0.6_{\pm 6E\text{-}2}}$	$15.3_{\pm 7\text{E-}1}$	$\mathbf{0.7_{\pm 6E\text{-}2}}$	$19.9_{\pm 7\text{E-}1}$
D5	$\mathbf{0.5745_{\pm 3E\text{-}5}}$	$0.5766_{\pm 4\text{E-}4}$	$\mathbf{0.3929_{\pm 5E\text{-}4}}$	$0.3946_{\pm 2\text{E-}4}$	$\mathbf{0.4323_{\pm 1E\text{-}4}}$	$0.4281_{\pm 6\text{E-}4}$	$\mathbf{28.0_{\pm 3E\text{-}0}}$	$217.1_{\pm 2\text{E+}1}$	$\mathbf{48.1_{\pm 1E\text{-}0}}$	$385.5_{\pm 3\text{E+}1}$	$\mathbf{33.0_{\pm 3E\text{-}0}}$	$269.5_{\pm 2\text{E+}1}$
D6	$\mathbf{0.6207_{\pm 2E\text{-}3}}$	$0.6219_{\pm 1\text{E-}4}$	$\mathbf{0.4194_{\pm 1E\text{-}3}}$	$0.4204_{\pm 5\text{E-}5}$	$\mathbf{0.3837_{\pm 5E\text{-}3}}$	$0.3801_{\pm 2\text{E-}4}$	$\mathbf{16.4_{\pm 7E\text{-}1}}$	$281.8_{\pm 5\text{E-}0}$	$\mathbf{26.7_{\pm 4E\text{-}0}}$	$444.7_{\pm 6\text{E-}0}$	$\mathbf{16.6_{\pm 5E\text{-}1}}$	$446.7_{\pm 6\text{E-}0}$
D7	$0.6350_{\pm 8\text{E-}4}$	$\mathbf{0.6324_{\pm 5E\text{-}3}}$	$0.4605_{\pm 1\text{E-}3}$	$\mathbf{0.4596_{\pm 5E\text{-}3}}$	$0.3305_{\pm 2\text{E-}3}$	$\mathbf{0.3360_{\pm 1E\text{-}2}}$	$\mathbf{8.1_{\pm 5E\text{-}1}}$	$37.2_{\pm 4\text{E-}0}$	$\mathbf{15.0_{\pm 2E\text{-}0}}$	$69.5_{\pm 2\text{E+}1}$	$\mathbf{8.1_{\pm 4E\text{-}1}}$	$38.3_{\pm 5\text{E-}0}$
D8	$0.6806_{\pm 6\text{E-}3}$	$\mathbf{0.6748_{\pm 2E\text{-}4}}$	$\mathbf{0.5124_{\pm 3E\text{-}3}}$	$0.5125_{\pm 2\text{E-}4}$	$0.3352_{\pm 1\text{E-}2}$	$\mathbf{0.3470_{\pm 3E\text{-}4}}$	$\mathbf{5.6_{\pm 1E\text{-}0}}$	$76.5_{\pm 6\text{E+}1}$	$\mathbf{8.3_{\pm 1E\text{-}0}}$	$150.3_{\pm 4\text{E-}0}$	$\mathbf{5.6_{\pm 1E\text{-}0}}$	$127.4_{\pm 5\text{E-}0}$

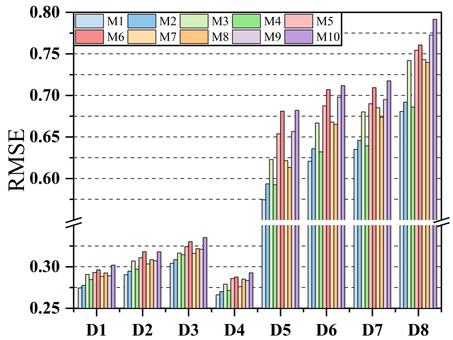
TABLE S7
EXPERIMENTAL RESULTS ON FOURTH-ORDER TENSORS

Model	Tensor1			Tensor2		
	RMSE↓	MAE↓	R^2↑	RMSE↓	MAE↓	R^2↑
MTN	$\mathbf{0.8389_{\pm 5.3E\text{-}03}}$	$\mathbf{0.5909_{\pm 9.5E\text{-}04}}$	$\mathbf{0.8746_{\pm 4.2E\text{-}03}}$	$\mathbf{0.0850_{\pm 1.2E\text{-}04}}$	$\mathbf{0.0520_{\pm 3.5E\text{-}04}}$	$\mathbf{0.6248_{\pm 5.2E\text{-}03}}$
Tucker	$0.9113_{\pm 6.2\text{E-}03}$	$0.6721_{\pm 3.2\text{E-}03}$	$0.8520_{\pm 5.8\text{E-}03}$	$0.0864_{\pm 2.8\text{E-}04}$	$0.0531_{\pm 1.3\text{E-}04}$	$0.6126_{\pm 6.7\text{E-}03}$
TR	$0.9103_{\pm 1.8\text{E-}03}$	$0.5974_{\pm 1.6\text{E-}03}$	$0.8524_{\pm 2.6\text{E-}03}$	$0.0875_{\pm 5.8\text{E-}04}$	$0.0542_{\pm 4.5\text{E-}04}$	$0.6052_{\pm 6.1\text{E-}03}$
TW	$0.8712_{\pm 2.5\text{E-}03}$	$0.6155_{\pm 2.4\text{E-}03}$	$0.8647_{\pm 3.7\text{E-}03}$	$0.0860_{\pm 7.6\text{E-}04}$	$0.0531_{\pm 2.2\text{E-}04}$	$0.6176_{\pm 8.2\text{E-}03}$

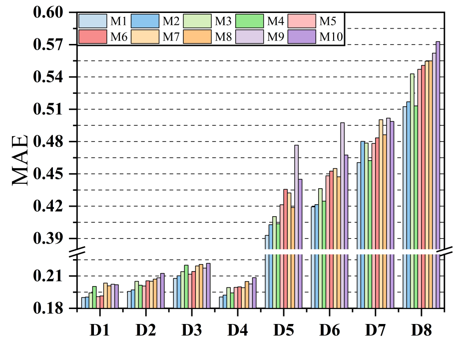


(a) D1 (b) D2 (c) D3 (d) D4 (e) D5 (f) D6 (g) D7 (h) D8

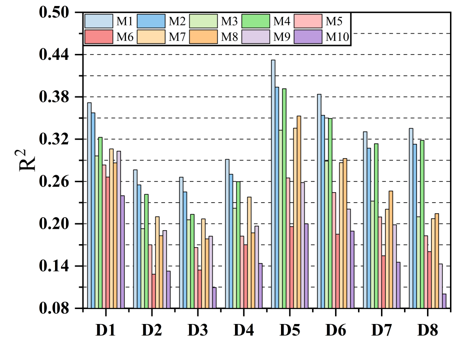
Fig. S1. MTN-TRL model's iterative curves of RMSE and MAE on D1-D8.



(a) RMSE of M1-10 on D1-8

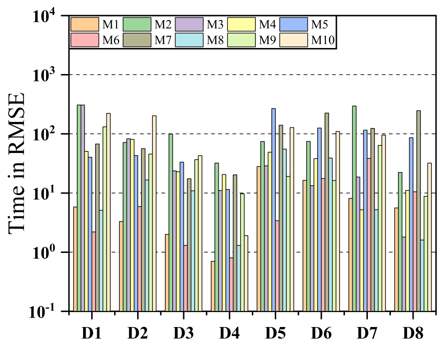


(b) MAE of M1-10 on D1-8

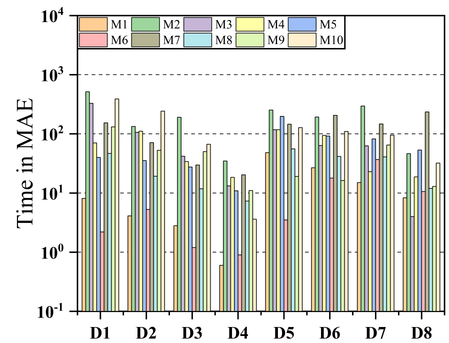


(c) R^2 of M1-10 on D1-8

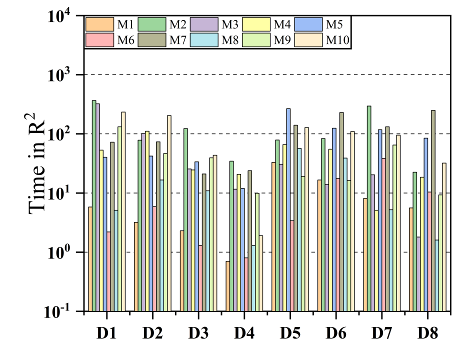
Fig. S2. RMSE, MAE, and R^2 for all tested models on D1-8.



(a) Time in RMSE of M1-10 on D1-8



(b) Time in MAE of M1-10 on D1-8



(c) Time in R^2 of M1-10 on D1-8

Fig. S3. Total time cost in RMSE, MAE, R^2 for all tested models on D1-8.

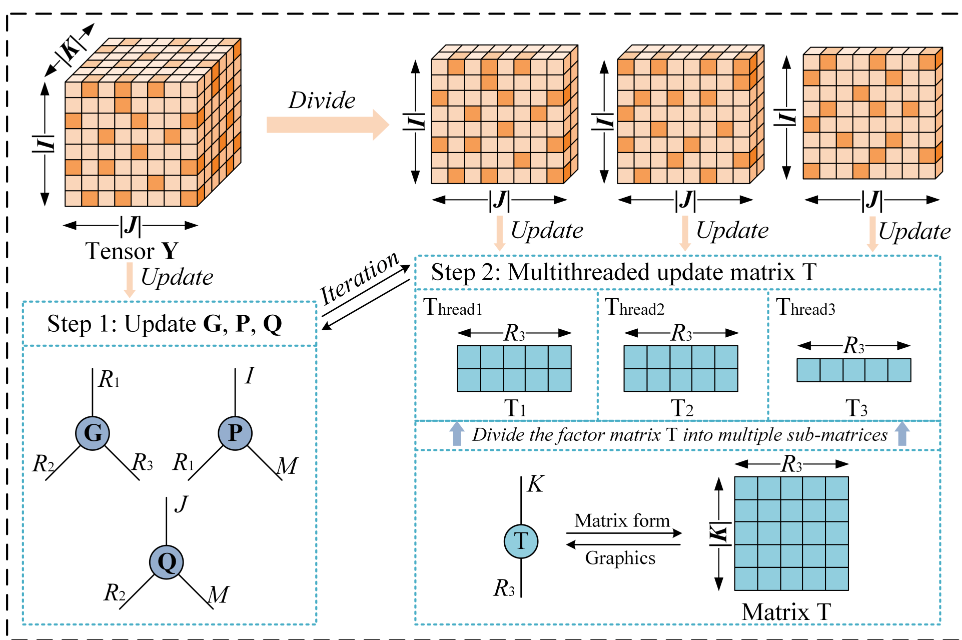
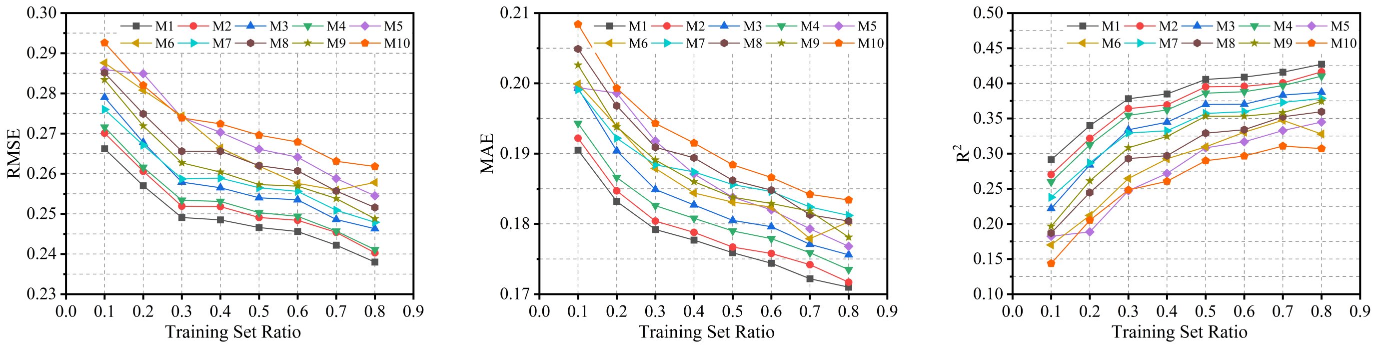
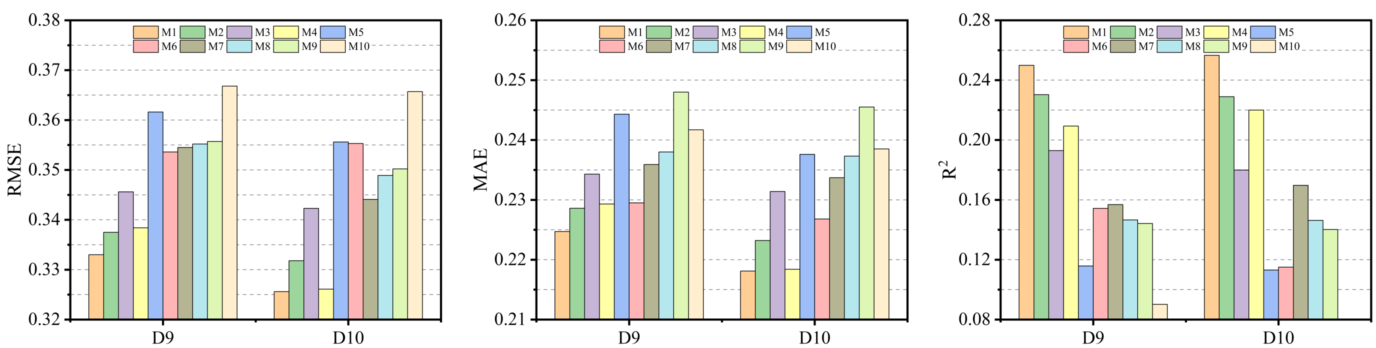


Fig. S4. Parallel computing design of MTN-TRL model.



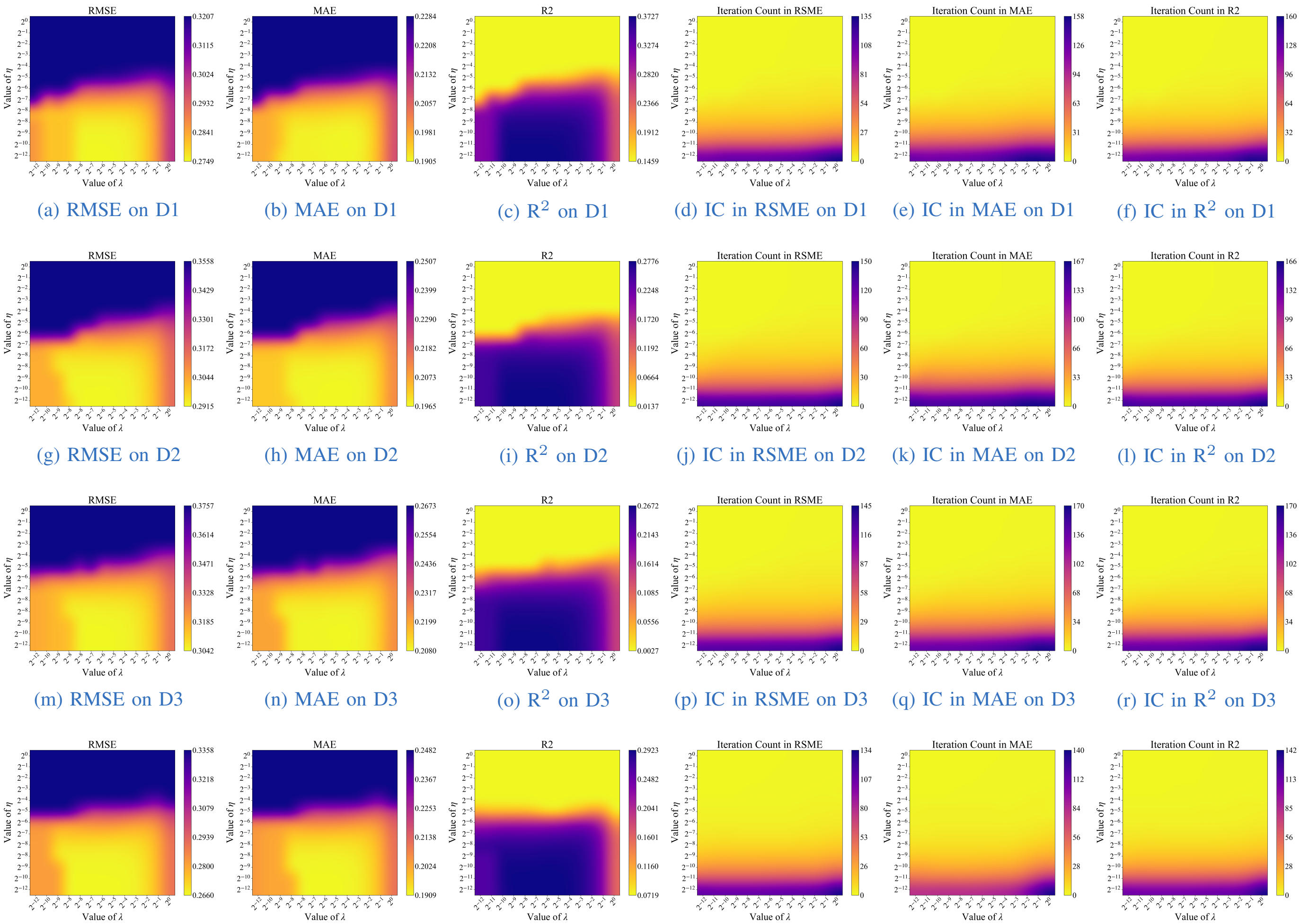
(a) RMSE on D4 (b) MAE on D4 (c) R^2 on D4

Fig. S5. Performance of all models on D4 with different training set ratios.



(a) RMSE of M1-10 on D9-10 (b) MAE of M1-10 on D9-10 (c) R^2 of M1-10 on D9-10

Fig. S6. Performance of MTN-TRL and other tested models on D9 and D10.



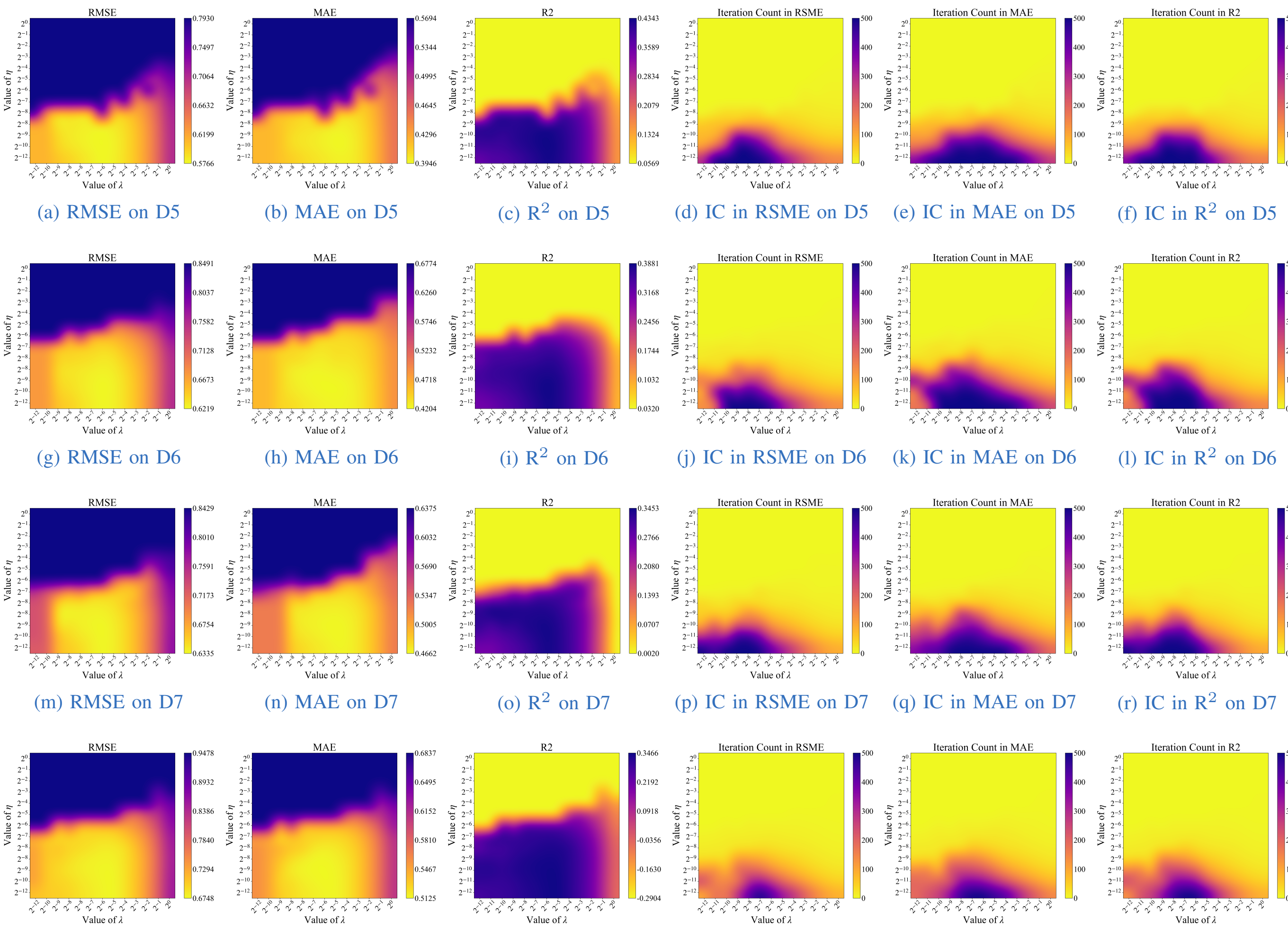
(a) RMSE on D1 (b) MAE on D1 (c) R^2 on D1 (d) IC in RSME on D1 (e) IC in MAE on D1 (f) IC in R^2 on D1

(g) RMSE on D2 (h) MAE on D2 (i) R^2 on D2 (j) IC in RSME on D2 (k) IC in MAE on D2 (l) IC in R^2 on D2

(m) RMSE on D3 (n) MAE on D3 (o) R^2 on D3 (p) IC in RSME on D3 (q) IC in MAE on D3 (r) IC in R^2 on D3

(s) RMSE on D4 (t) MAE on D4 (u) R^2 on D4 (v) IC in RSME on D4 (w) IC in MAE on D4 (x) IC in R^2 on D4

Fig. S7. MTTN-TRL's sensitivity test of manually tuned hyper-parameters η and λ on D1-D4. (*IC represents the iteration Count)



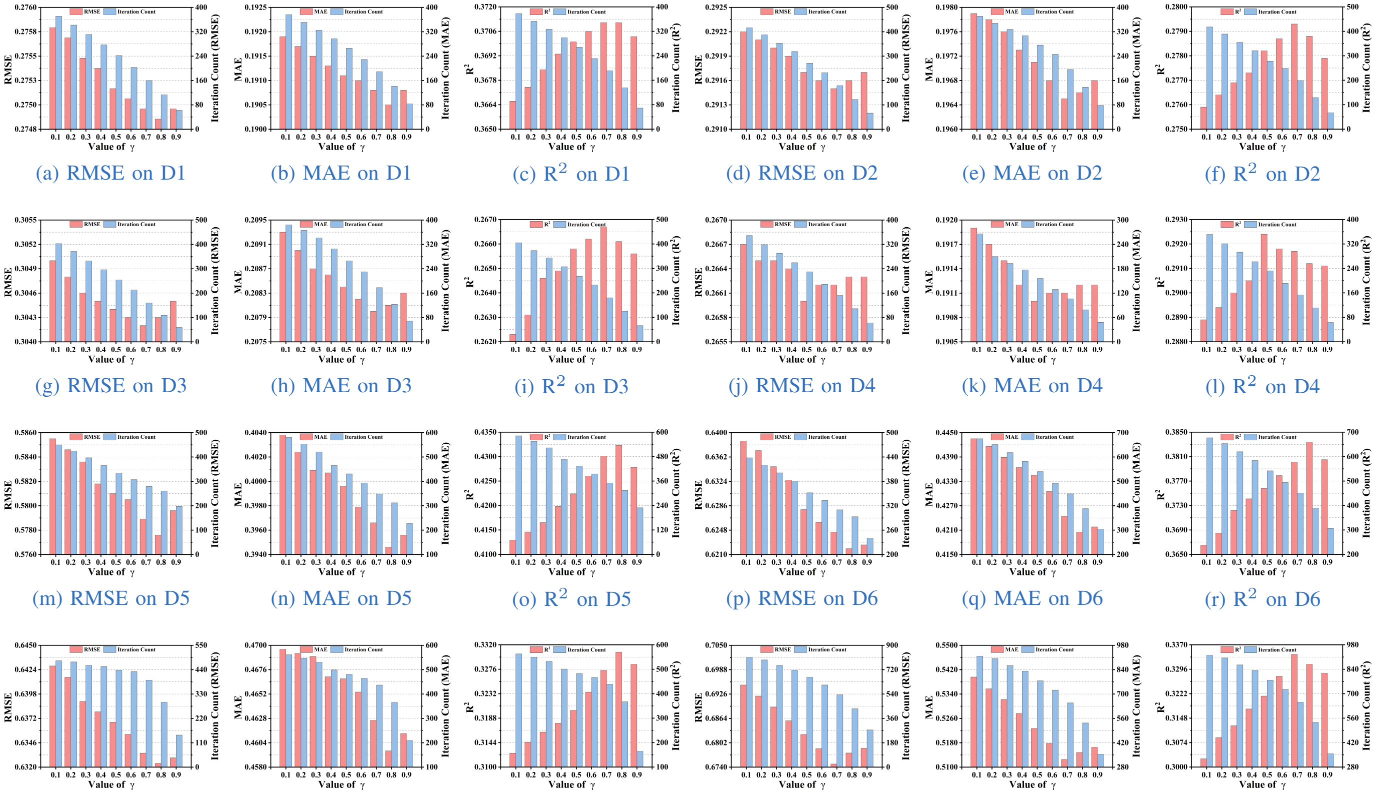
(a) RMSE on D5 (b) MAE on D5 (c) R^2 on D5 (d) IC in RSME on D5 (e) IC in MAE on D5 (f) IC in R^2 on D5

(g) RMSE on D6 (h) MAE on D6 (i) R^2 on D6 (j) IC in RSME on D6 (k) IC in MAE on D6 (l) IC in R^2 on D6

(m) RMSE on D7 (n) MAE on D7 (o) R^2 on D7 (p) IC in RSME on D7 (q) IC in MAE on D7 (r) IC in R^2 on D7

(s) RMSE on D8 (t) MAE on D8 (u) R^2 on D8 (v) IC in RSME on D8 (w) IC in MAE on D8 (x) IC in R^2 on D8

Fig. S8. MTTN-TRL's sensitivity test of manually tuned hyper-parameters η and λ on D5-D8. (*IC represents the iteration Count)



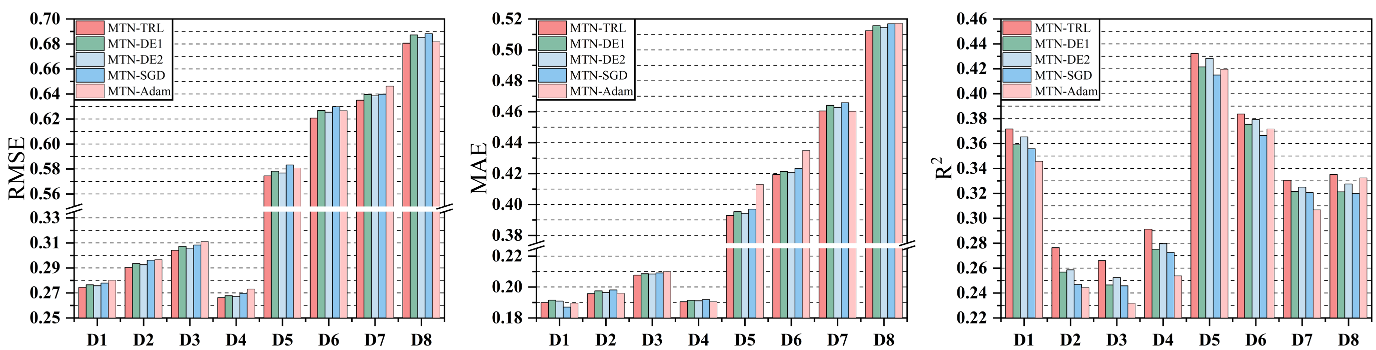
(a) RMSE on D1 (b) MAE on D1 (c) R^2 on D1 (d) RMSE on D2 (e) MAE on D2 (f) R^2 on D2

(g) RMSE on D3 (h) MAE on D3 (i) R^2 on D3 (j) RMSE on D4 (k) MAE on D4 (l) R^2 on D4

(m) RMSE on D5 (n) MAE on D5 (o) R^2 on D5 (p) RMSE on D6 (q) MAE on D6 (r) R^2 on D6

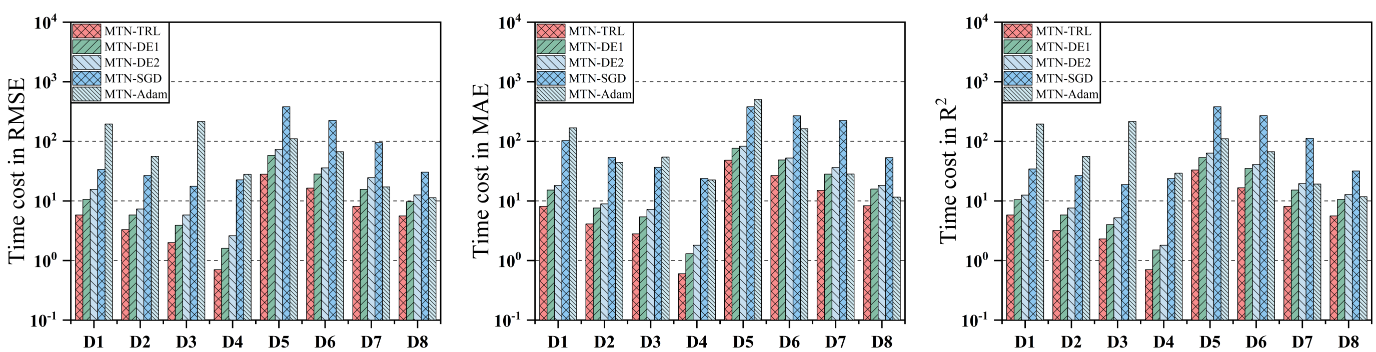
(s) RMSE on D7 (t) MAE on D7 (u) R^2 on D7 (v) RMSE on D8 (w) MAE on D8 (x) R^2 on D8

Fig. S9. MTTN-TRL's sensitivity test of manually tuned hyper-parameters γ on D1-D8.



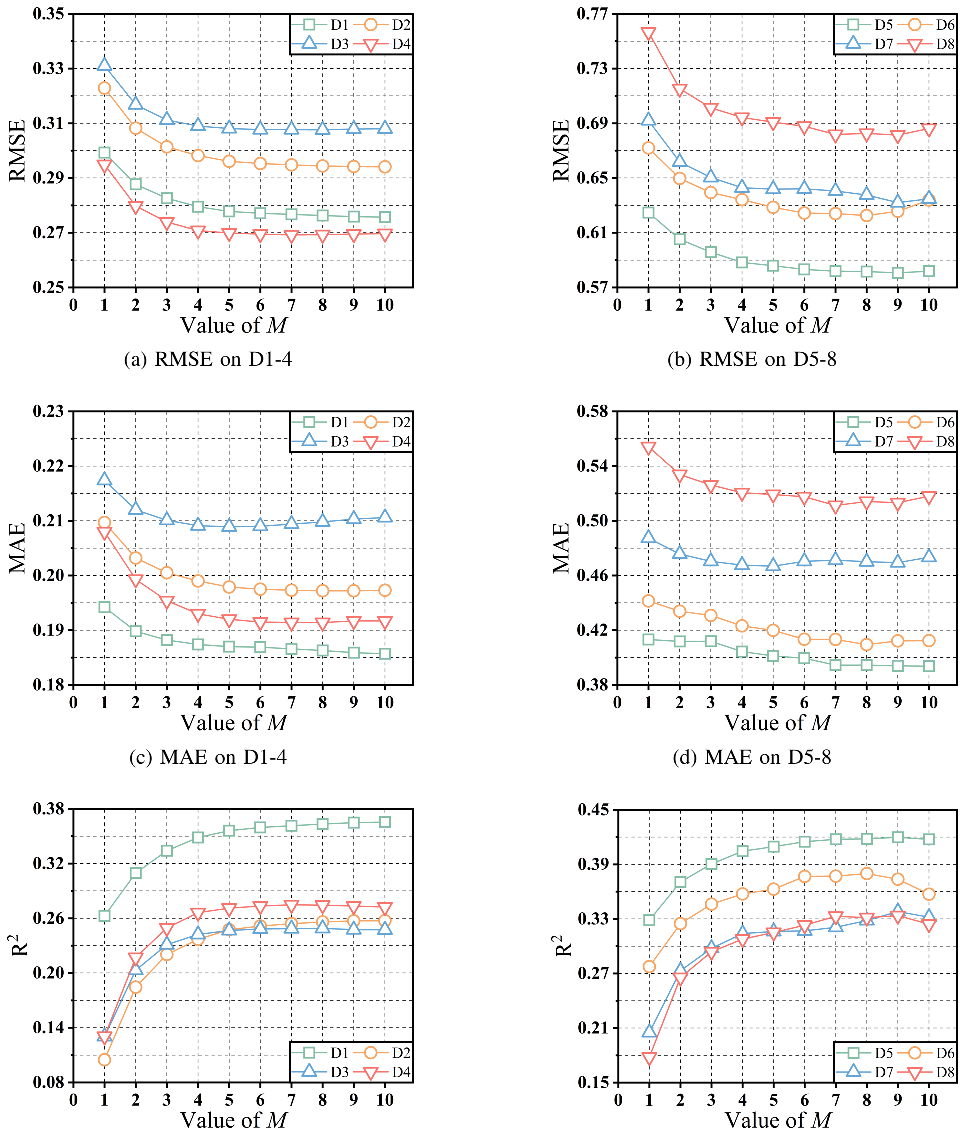
(a) RMSE on D1-D8 (b) MAE on D1-D8 (c) R^2 on D1-D8

Fig. S10. RMSE, MAE, and R^2 of MTN with different parameters learning schemes on D1-D8.



(a) Time cost in RMSE on D1-D8 (b) Time cost in MAE on D1-D8 (c) Time cost in R^2 on D1-D8

Fig. S11. Time cost in RMSE, MAE, and R^2 of MTN with different parameters learning schemes on D1-D8.



(a) RMSE on D1-4
(b) RMSE on D5-8
(c) MAE on D1-4
(d) MAE on D5-8
(e) R^2 on D1-4
(f) R^2 on D5-8

Fig. S12. The impact of dimension M to MTN-TRL on D1-D8.

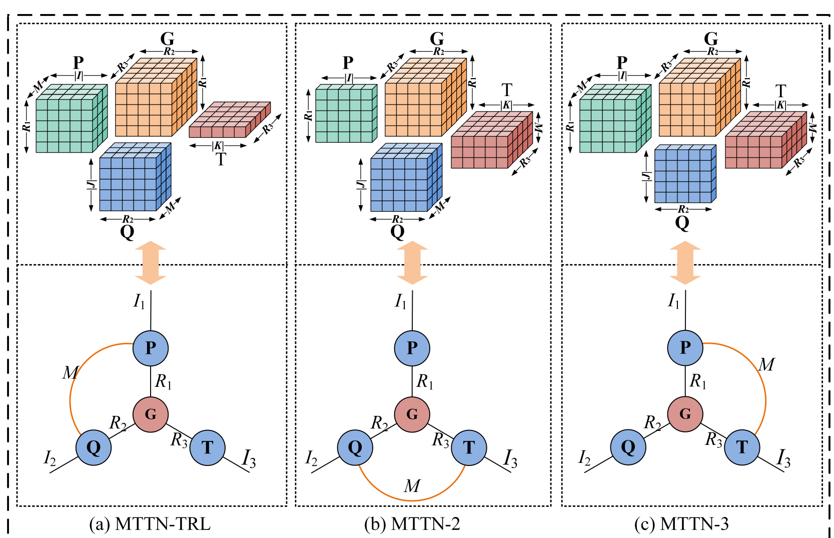
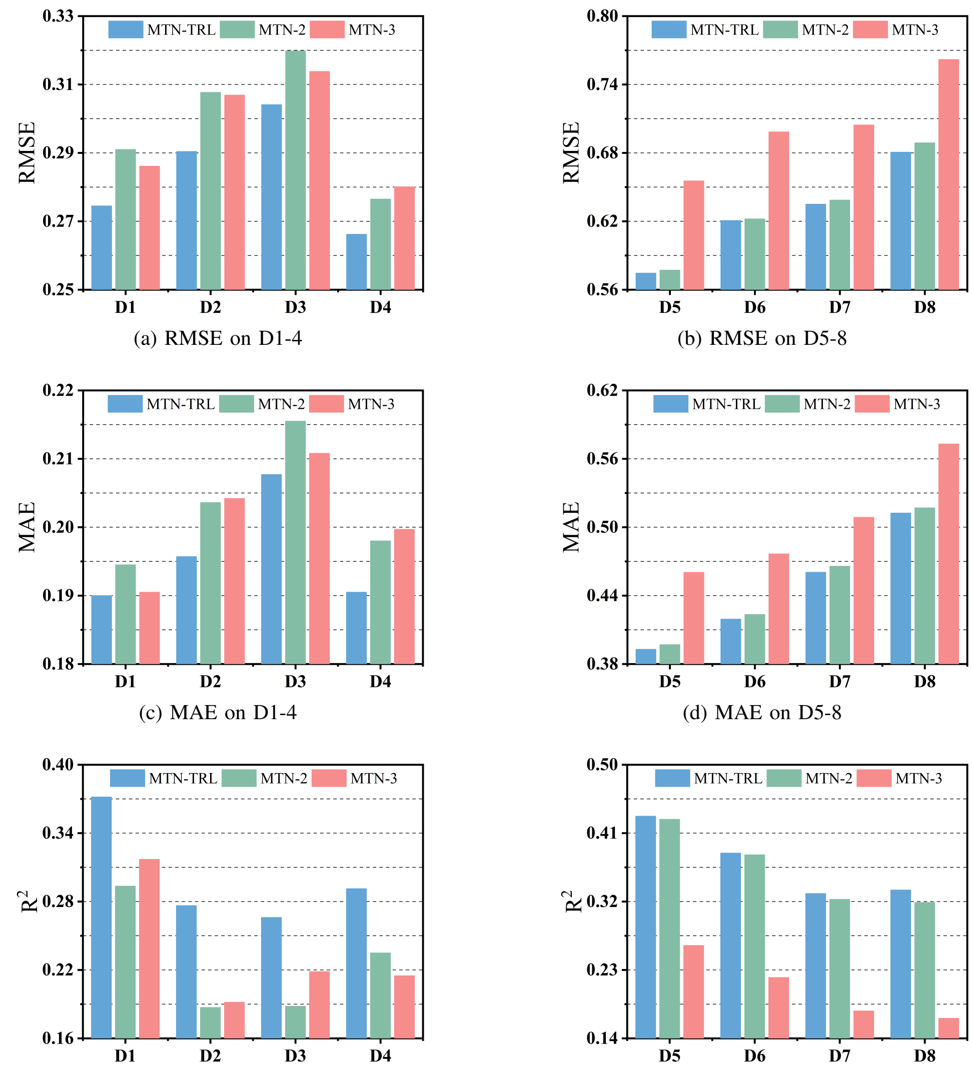


Fig. S13. Three different structures of MTN.



(a) RMSE on D1-4

(b) RMSE on D5-8

(c) MAE on D1-4

(d) MAE on D5-8

(e) R^2 on D1-4

(f) R^2 on D5-8

Fig. S14. Ablation study of MTN-TRL on D1-D8.

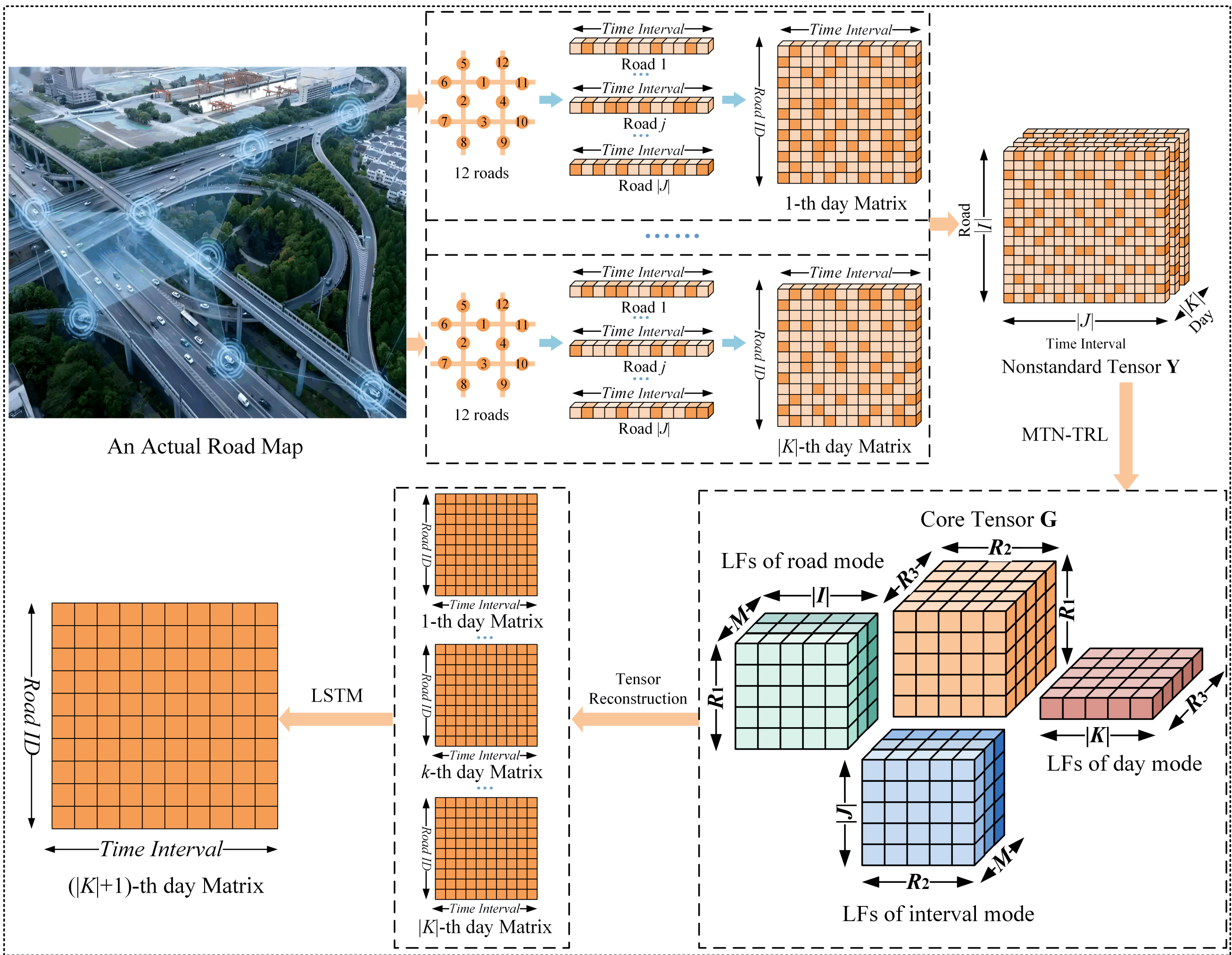


Fig. S15. Application of the MTN-TRL model in a traffic network.