Renormalization Group Guided Tensor Network Structure Search

Maolin Wang[1], **Bowen Yu**[1], **Sheng Zhang**[1], **Linjie Mi**[1], **Wanyu Wang**[1],
Yiqi Wang[2], **Pengyue Jia**[1], **Xuetao Wei**[3], **Zenglin Xu**[4], **Ruocheng Guo**[5], **Xiangyu Zhao**[1*]
[1]City University of Hong Kong
[2] National University of Defense Technology
[3] Southern University of Science and Technology
[4] Fudan University
[5]Intuit AI
{morin.wang, yubowen,szhang844-c,linjiemi2-c,wanyuwang4-c,jia.pengyue} @my.cityu.edu.hk
yiq@nudt.edu.cn, weixt@sustech.edu.cn, {zenglin, rguo.asu}@gmail.com, xianzhao@cityu.edu.hk

Abstract

Tensor network structure search (TN-SS) aims to automatically discover optimal network topologies and rank configurations for efficient tensor decomposition in high-dimensional data representation. Despite recent advances, existing TN-SS methods face significant limitations in computational tractability, structure adaptivity, and optimization robustness across diverse tensor characteristics. They struggle with three key challenges: single-scale optimization missing multi-scale structures, discrete search spaces hindering smooth structure evolution, and separated structure-parameter optimization causing computational inefficiency. We propose RGTN (Renormalization Group guided Tensor Network search), a physics-inspired framework transforming TN-SS via multi-scale renormalization group flows. Unlike fixed-scale discrete search methods, RGTN uses dynamic scale-transformation for continuous structure evolution across resolutions. Its core innovation includes learnable edge gates for optimization-stage topology modification and intelligent proposals based on physical quantities like node tension measuring local stress and edge information flow quantifying connectivity importance. Starting from low-complexity coarse scales and refining to finer ones, RGTN finds compact structures while escaping local minima via scale-induced perturbations. Extensive experiments on light field data, high-order synthetic tensors, and video completion tasks show RGTN achieves state-of-the-art compression ratios and runs 4-600× faster than existing methods, validating the effectiveness of our physics-inspired approach.

Code — https://github.com/Applied-Machine-Learning-Lab/RGTN

Appendix — https://github.com/Applied-Machine-Learning-Lab/RGTN/Appendix.pdf

Introduction

Tensor network structure search (TN-SS) has emerged as a fundamental challenge in high-dimensional data representation, seeking to automatically discover optimal network topologies and rank configurations for efficient tensor decomposition (Li and Sun 2020). Despite recent advances, existing TN-SS methods face significant limitations in effectively addressing three critical aspects: computational tractability (Wang et al. 2023a), structure adaptivity (Hashemizadeh et al. 2020), and optimization robustness across diverse tensor characteristics (Iacovides et al. 2025). While current approaches show promise in specific scenarios, they struggle to comprehensively tackle these interconnected challenges (Zheng et al. 2020).

In practical tensor decomposition scenarios, optimal network structures naturally exhibit three fundamental properties deeply rooted in physics and information theory: (1) **Scale-Invariant Correlations**: tensor networks possess self-similar correlation structures across different length scales, analogous to critical phenomena in statistical physics where the renormalization group reveals how physical properties transform under scale changes (White 1992; Vidal 2007), (2) **Hierarchical Entanglement**: the optimal connectivity pattern reflects hierarchical entanglement structures, with different scales capturing correlations at different ranges—from local quantum entanglement to global classical correlations (Schollwöck 2011; Orús 2014), and (3) **Flow of Information**: efficient tensor networks naturally organize information flow from fine-grained local features to coarse-grained global structures, following principles similar to real-space renormalization in condensed matter physics (Evenbly and Vidal 2011; Haegeman et al. 2013). These observations from quantum many-body physics and renormalization group theory highlight the critical need for tensor network methods that can exploit multi-scale structures while maintaining efficiency (Chan et al. 2008).

Based on these observations, tensor network structure search systems need to address three fundamental challenges that require innovative solutions: First, single-scale optimization inherently limits structure discovery. Current methods operate at a fixed resolution throughout optimization, missing the rich multi-scale structures inherent in tensor data and failing to leverage the computational advantages of hierarchical decomposition (Li and Sun 2020; Hashemizadeh et al. 2020; Li et al. 2023a). Second, discrete search spaces

*Corresponding author.

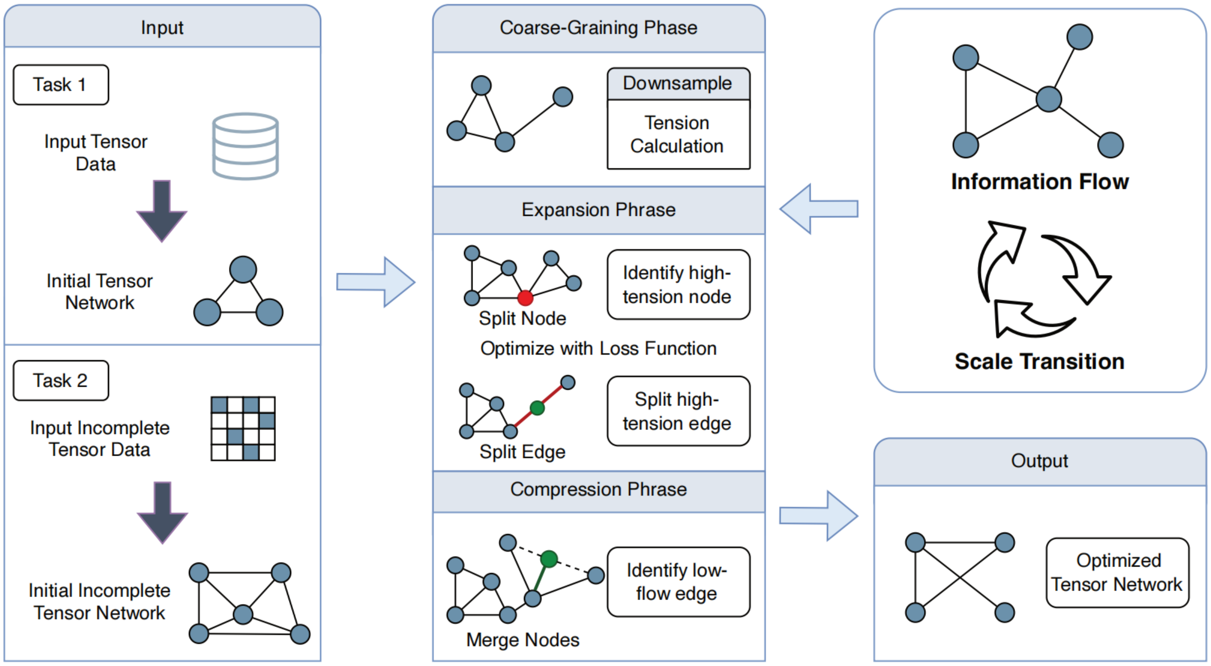


Figure 1: The RGTN framework transforms network topology through physics-inspired multi-scale operations instead of traditional sampling-evaluation methods. It processes tensors through three RG-based phases: coarse-graining (downsampling and tension calculation), expansion (splitting high-tension nodes), and compression (merging low-flow edges), unifying structure search and parameter optimization for efficient discovery of optimal tensor network structures.

prevent smooth structure evolution. Genetic algorithms like TNGA explore topology through discrete mutations (Li and Sun 2020), greedy methods incrementally modify structures through local decisions (Hashemizadeh et al. 2020), and local search approaches like TNLS navigate neighborhoods of current solutions (Li et al. 2023a), all suffering from the combinatorial nature of the search space (Li et al. 2023a). Third, separation of structure and parameter optimization creates inefficiency. Program synthesis methods intelligently generate candidate structures but still require expensive evaluation of each proposal (Zheng et al. 2020; Liao et al. 2019), while regularization-based approaches achieve faster convergence but remain constrained by predetermined topology spaces (Zheng et al. 2024). These limitations manifest critically in practice: Evolutionary methods require population sizes growing exponentially with tensor order (Li and Sun 2020), greedy algorithms make irrevocable local decisions that lead to suboptimal structures (Hashemizadeh et al. 2020), local search methods frequently converge to poor local minima (Li et al. 2022, 2023a), and even advanced approaches using LLMs for algorithm discovery still operate within the sampling-evaluation paradigm (Zheng et al. 2020). This creates an urgent need for a fundamentally new paradigm that can harness physics principles for efficient structure discovery.

To address these challenges, we propose RGTN (**R**enormalization **G**roup guided **T**ensor **N**etwork search), a physics-inspired framework that transforms tensor network structure search through multi-scale renormalization group flows. Unlike existing methods limited to discrete structure spaces, RGTN implements dynamic scale transformation where networks evolve continuously across resolution levels via learnable edge gates. This approach utilizes node tension to measure local stress and edge information flow to quantify connectivity importance. By optimizing from coarse to fine scales, RGTN discovers compact structures while escaping local minima through scale-induced perturbations.

Our main contributions are:

- **Multi-scale framework**: First tensor network approach implementing true renormalization group flows with continuous edge gates and scale-dependent optimization, enabling dynamic topology evolution beyond discrete search limitations.
- **Physics-inspired strategies**: Node tension and edge information flow guide intelligent structure modifications through natural physical processes rather than combinatorial enumeration.
- **Theoretical speedup**: Rigorous analysis showing exponential acceleration from $\Omega(\exp(N^2))$ to $\mathcal{O}(\log I \cdot \log(1/\epsilon))$ with stronger convergence guarantees and high-probability escape from local minima.
- **Empirical validation**: Experiments demonstrate RGTN achieves up to 3× better compression and 4-600× faster over existing methods.

Method

In this section, we present our RGTN approach for efficiently searching tensor network structures. As shown in

Figure 1, we propose a radically different approach inspired by the renormalization group (RG) theory (Shankar 1994; Ueda 2024) from statistical physics.

Theoretical Foundation The renormalization group is a mathematical apparatus that reveals how physical systems behave across different length scales. In the context of tensor networks, we interpret scale transformations as changes in the network's ability to capture correlations at different ranges. Consider a tensor network $\mathcal{M}$ representing a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. We define a renormalization group flow on the space of tensor networks through a semi-group of transformations $\{R_s\}_{s \geq 0}$:

$$\mathcal{M}_{s+\delta s} = R_{\delta s}[\mathcal{M}_s], \tag{1}$$

where s represents the scale parameter. The RG transformation R_s consists of two complementary operations that modify the network structure while preserving its representational capacity.

The expansion operation R_{expand} increases the network's resolution by decomposing tensor cores:

$$R_{\text{expand}} : \mathcal{G}_v \mapsto \sum_{r=1}^{R_{uv}} \mathcal{G}_u^{(r)} \otimes \mathcal{G}_v^{(r)}, \tag{2}$$

where a single core $\mathcal{G}_v$ is split into two cores $\mathcal{G}_u$ and $\mathcal{G}_v$ connected by a bond of dimension R_{uv}. This operation enables the network to capture finer-grained correlations.

The compression operation R_{compress} reduces the network's complexity by merging adjacent cores:

$$R_{\text{compress}} : (\mathcal{G}_u, \mathcal{G}_v) \mapsto \mathcal{G}_{uv} = \mathcal{G}_u \times_{e_{uv}} \mathcal{G}_v, \tag{3}$$

where $\times_{e_{uv}}$ denotes tensor contraction along the edge connecting u and v. This operation identifies and eliminates redundant degrees of freedom.

Scale-Dependent Effective Action Following the RG philosophy, we introduce a scale-dependent effective action (loss function) that captures the relevant physics at each scale:

$$S_s[\mathcal{M}] = S_{\text{data}}[\mathcal{M}] + \sum_k \lambda_k(s) S_k[\mathcal{M}], \tag{4}$$

where S_{data} represents data fidelity and S_k are regularization terms with scale-dependent coupling constants $\lambda_k(s)$. The running of these coupling constants is determined by the RG flow equations:

$$\frac{d\lambda_k}{ds} = \beta_k(\{\lambda_j\}), \tag{5}$$

where β_k are the beta functions encoding how different regularization strengths evolve across scales.

Enhanced Tensor Network Architecture

Standard tensor network architectures are rigid, with fixed topologies that cannot adapt during optimization. This limitation prevents the network from discovering more efficient structures or adjusting its capacity based on the data complexity. Additionally, existing methods for structure modification require discrete decisions (add/remove edges, change ranks) that disrupt the optimization process and often lead to instability. We need architectural components that enable continuous structure adaptation while maintaining stable gradient flow throughout the network.

Adaptive Diagonal Factors Inspired by the SVDinsTN's use of diagonal factors for structure discovery, we introduce adaptive diagonal factors that serve as importance weights for each tensor core. For a tensor core $\mathcal{G}_k \in \mathbb{R}^{R_1 \times \cdots \times R_m \times I_k}$, we define diagonal adaptation matrices $\mathbf{D}_k^{(i)} \in \mathbb{R}^{R_i \times R_i}$ for each virtual bond:

$$\tilde{\mathcal{G}}_k = \mathcal{G}_k \times_1 \mathbf{D}_k^{(1)} \times_2 \mathbf{D}_k^{(2)} \cdots \times_m \mathbf{D}_k^{(m)}. \tag{6}$$

These diagonal factors play a crucial role in structure discovery. When elements of $\mathbf{D}_k^{(i)}$ approach zero, the corresponding bond dimensions become effectively reduced, automatically revealing a more compact structure.

Edge Gating Mechanism To enable dynamic topology modification during optimization, we introduce learnable edge gates. For each edge (u, v) in the tensor network graph $G = (V, E)$, we define a gating function:

$$g_{uv} = \sigma(w_{uv}), \quad w_{uv} \in \mathbb{R}, \tag{7}$$

where σ is the sigmoid function. The gated tensor contraction becomes:

$$\mathcal{C}_{uv} = g_{uv} \cdot (\mathcal{G}_u \times_{m_u^v, m_v^u} \mathcal{G}_v) + (1 - g_{uv}) \cdot \mathcal{I}, \tag{8}$$

where $\mathcal{I}$ represents an identity-like tensor maintaining dimensional consistency. This soft gating mechanism allows gradual edge removal when $g_{uv} \to 0$.

Multi-Scale Loss Function For tensor completion, we formulate a comprehensive loss function that incorporates both data fidelity and structure-inducing regularization:

$$\begin{aligned}\mathcal{L}_{\text{total}} = &\underbrace{\frac{1}{2}\|\mathcal{P}_\Omega(\mathcal{X} - \mathcal{F})\|_F^2}_{\text{Data Fidelity}} + \underbrace{\alpha(s) \sum_{t=1}^{T-1} \|\mathcal{X}^{(t+1)} - \mathcal{X}^{(t)}\|_F}_{\text{Temporal Consistency}} \\ &+ \underbrace{\beta(s) \sum_{i,j} \|\nabla_{ij}\mathcal{X}\|_F}_{\text{Spatial Smoothness}} + \underbrace{\gamma(s) \sum_{k,i} \|\mathbf{D}_k^{(i)}\|_1}_{\text{Diagonal Sparsity}} \\ &+ \underbrace{\delta(s) \sum_{(u,v) \in E} H(g_{uv})}_{\text{Edge Entropy}} + \underbrace{\epsilon(s)\text{TNN}(\mathcal{X})}_{\text{Low-rank Regularization}},\end{aligned} \tag{9}$$

where $H(g) = -g \log g - (1 - g)\log(1 - g)$ is the binary entropy function encouraging decisive gating, and $\text{TNN}(\mathcal{X})$ is the tensor nuclear norm computed as:

$$\text{TNN}(\mathcal{X}) = \sum_{k=1}^{N} \omega_k \|\mathcal{X}_{(k)}\|_*, \tag{10}$$

with $\mathcal{X}_{(k)}$ being mode-k unfolding and $\|\cdot\|_*$ nuclear norm.

Intelligent Structure Search via RG Flow

Random or exhaustive structure search strategies suffer from poor scalability and often explore irrelevant regions of the structure space. Without guidance from the current network state, these methods waste computational resources on unpromising structural modifications. Furthermore, the interplay between structure optimization and parameter optimization is poorly understood in existing approaches, leading to suboptimal coordination between these two aspects. We need an intelligent search strategy that leverages the current network's properties to guide exploration and properly balances structural and parametric updates.

Smart Proposal Generation Rather than random structural modifications, we use the current network's properties to guide proposal generation. For the expansion phase, we identify nodes with high "tension" - a measure of how much a node contributes to the reconstruction error:

$$T_v = \left\| \frac{\partial \mathcal{L}_{\text{data}}}{\partial \mathcal{G}_v} \right\|_F \cdot \text{degree}(v). \tag{11}$$

Nodes with high tension are prioritized for splitting, as they likely encode complex correlations that benefit from finer representation.

For the compression phase, we identify edges with low "information flow" - quantified by the gate values and the mutual information between connected cores:

$$I_{uv} = g_{uv} \cdot \text{MI}(\mathcal{G}_u, \mathcal{G}_v), \tag{12}$$

where MI denotes mutual information estimated through the singular value spectrum of the contracted tensor.

Adaptive Optimization Strategy The optimization of tensor cores and structural parameters proceeds through an adaptive scheme that adjusts to the current scale and convergence behavior. We employ a modified Adam optimizer with scale-dependent learning rates:

$$\eta_{\text{cores}}(s) = \eta_0 \cdot \exp(-s/s_0), \quad \eta_{\text{struct}}(s) = \eta_0 \cdot (1 + s/s_1), \tag{13}$$

where cores are optimized more aggressively at fine scales while structural parameters are refined more at coarse scales.

The complete algorithm proceeds at **Appendix A**.

Structure Discovery through Sparsity The interplay between diagonal factors and edge gates enables automatic structure discovery. During optimization, the ℓ_1 regularization on diagonal factors and entropy regularization on edge gates induce sparsity patterns that reveal the underlying structure. Specifically, when diagonal factor elements $D_k^{(i)}[j,j] < \epsilon$, the corresponding bond dimension can be reduced, and when edge gate $g_{uv} < \delta$, the edge can be removed from the topology.

This soft-to-hard thresholding strategy is implemented through a temperature annealing scheme:

$$\tau(t) = \tau_0 \exp(-t/t_0), \tag{14}$$

where the soft gates $g_{uv} = \sigma(w_{uv}/\tau(t))$ become increasingly binary as training progresses.

Multi-Scale Progressive Refinement

Direct optimization of large-scale tensor networks faces severe challenges, including slow convergence, susceptibility to local minima, and high computational cost. Starting from random initialization often requires extensive iterations to reach good solutions, and the optimization landscape becomes increasingly complex with network size. Additionally, fine-scale details can obscure the global structure, making it difficult to identify the optimal topology. We need a multi-scale approach that can efficiently explore the solution space by solving progressively refined versions of the problem. We begin at the coarsest scale, where the problem has reduced dimensionality and computational cost:

$$\mathcal{F}_S = \mathcal{D}_S[\mathcal{F}], \quad \text{where } \mathcal{D}_s \text{ is a downsampling operator.} \tag{15}$$

As we flow towards finer scales, we use the coarse-scale solution to initialize the fine-scale optimization:

$$\mathcal{M}_{s-1}^{(0)} = \mathcal{U}_s[\mathcal{M}_s^*], \quad \text{where } \mathcal{U}_s \text{ is an upsampling operator.} \tag{16}$$

This progressive refinement strategy provides several benefits: (1) faster convergence by providing good initializations, (2) avoiding local minima by exploring the solution space hierarchically, and (3) computational efficiency by solving smaller problems first.

Theoretical Analysis

This section provides theoretical analysis of the RGTN framework, establishing convergence guarantees, analyzing structure discovery properties, demonstrating computational advantages, and connecting to statistical physics principles. Due to space constraints, detailed theorems on computational complexity, loss landscape smoothing, probabilistic escape from local minima, fixed points, criticality, and universality are in **Appendix B**.

Preliminaries and Assumptions

We begin by establishing the mathematical foundations and assumptions underlying our analysis. Let $\mathcal{M} = (\mathcal{G}, \mathcal{S})$ denote a tensor network with cores $\mathcal{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_N\}$ and structure $\mathcal{S} = (V, E, \{R_{ij}\})$, where V is the set of nodes, E is the set of edges, and R_{ij} are bond dimensions. The parameter space is denoted as $\Theta = \{\theta_{\mathcal{G}}, \theta_{\mathcal{S}}\}$, where $\theta_{\mathcal{G}}$ represents tensor core parameters and $\theta_{\mathcal{S}}$ represents structural parameters including diagonal factors and edge gates.

Assumption 1 (Lipschitz Continuity). *The loss function $\mathcal{L}(\mathcal{M})$ is L-Lipschitz continuous with respect to the network parameters:*

$$|\mathcal{L}(\mathcal{M}_1) - \mathcal{L}(\mathcal{M}_2)| \leq L \|\mathcal{M}_1 - \mathcal{M}_2\|_F, \tag{17}$$

where $\|\cdot\|_F$ denotes the Frobenius norm extended to tensor networks.

Assumption 2 (Smoothness of Scale Transformations). *The scale transformation operators $\mathcal{D}_s$ (downsampling) and $\mathcal{U}_s$ (upsampling) satisfy:*

$$\|\mathcal{U}_s \circ \mathcal{D}_s[\mathcal{X}] - \mathcal{X}\|_F \leq C_s \|\mathcal{X}\|_F, \tag{18}$$

where $C_s = \mathcal{O}(2^{-s})$ decreases with coarser scales.

Assumption 3 (Bounded Network Parameters)**.** *The network parameters lie in a bounded domain:* $\|\mathcal{G}_k\|_F \leq B_{\mathcal{G}}$ *for all cores and* $0 \leq g_{ij} \leq 1$ *for all edge gates.*

Convergence Analysis

We first establish the convergence properties of the RGTN algorithm. The analysis considers the alternating optimization between tensor cores and structural parameters across multiple scales.

Theorem 1 (Global Convergence of Multi-Scale Optimization)**.** *Under Assumptions 1-3, the RGTN algorithm generates a sequence of networks* $\{\mathcal{M}^{(t)}\}_{t=0}^{\infty}$ *that converges to a critical point of the multi-scale objective function. Specifically, for any* $\epsilon > 0$*, there exists* $T(\epsilon)$ *such that for all* $t \geq T(\epsilon)$*:*

$$\|\nabla \mathcal{L}(\mathcal{M}^{(t)})\|_F \leq \epsilon. \tag{19}$$

Moreover, the convergence rate satisfies:

$$\min_{t \in [T]} \|\nabla \mathcal{L}(\mathcal{M}^{(t)})\|_F^2 \leq \frac{2[\mathcal{L}(\mathcal{M}^{(0)}) - \mathcal{L}^*] + L^2 \sum_{s=0}^{S} C_s^2}{\sum_{t=0}^{T-1} \eta_t}, \tag{20}$$

where η_t *are the learning rates and* $\mathcal{L}^*$ *is the optimal loss value.*

Proof. Detailed proof is provided in **Appendix C.1**. □

Structure Discovery and Sparsity Analysis

The automatic structure discovery in RGTN arises from the sparsity-inducing properties of diagonal factors and edge gates. We analyze how these mechanisms reveal the intrinsic tensor network structure.

Lemma 1 (Diagonal Factor Sparsity Pattern)**.** *For the regularized objective with diagonal factor penalty* $\gamma \sum_{k,i} \|\mathbf{D}_k^{(i)}\|_1$*, the optimal diagonal entries satisfy the soft-thresholding property:*

$$D_k^{(i)}[j,j]^* = \mathit{sign}(z_j) \max(|z_j| - \gamma/L_j, 0), \tag{21}$$

where z_j *is the unregularized optimal value and* L_j *is the Lipschitz constant for the* j*-th diagonal entry.*

Proof. Detailed proof is provided in **Appendix C.2**. □

Theorem 2 (Structure Recovery Guarantee)**.** *Let* $\mathcal{X}^*$ *be a tensor with true tensor network representation having ranks* $(R_1^*, \ldots, R_m^*)$*. Suppose the observed tensor is* $\mathcal{F} = \mathcal{P}_\Omega(\mathcal{X}^* + \mathcal{N})$*, where* $\mathcal{N}$ *is noise with* $\|\mathcal{N}\|_F \leq \sigma$*. Then with regularization parameter* $\gamma = \Theta(\sigma\sqrt{\log(mN)/|\Omega|})$*, the RGTN algorithm recovers ranks* $(R_1, \ldots, R_m)$ *satisfying:*

$$\mathbb{P}\left(\max_i |R_i - R_i^*| \leq \Delta_R\right) \geq 1 - \exp(-c|\Omega|), \tag{22}$$

where $\Delta_R = \mathcal{O}(\sigma/\sigma_{\min})$*,* $\sigma_{\min}$ *is the minimum non-zero singular value of the true tensor network, and* $c > 0$ *is a universal constant.*

Proof. Detailed proof is provided in **Appendix C.3**. □

Summary of Theoretical Advantages

Our theoretical analysis establishes key advantages of RGTN over existing tensor network structure search methods. The convergence guarantee in Theorem 1 provides predictable performance with explicit rates, enabling practitioners to determine computational budgets. The structure recovery guarantee in Theorem 2 ensures automatic discovery of true tensor network rank under mild conditions, eliminating manual hyperparameter tuning. The complexity analysis in Theorem 3 (**Appendix B**) demonstrates exponential speedup over sampling-based methods, reducing search cost from $\Omega(\exp(N^2))$ to $\mathcal{O}(S\log(1/\epsilon))$ where $S = \mathcal{O}(\log I)$. This improvement makes large-scale applications feasible. Theorem 4 (**Appendix B**) shows the multi-scale approach escapes local minima with high probability, addressing a key challenge in non-convex optimization. The connection to renormalization group theory provides insights into tensor network behavior. The universality property in Theorems 5 and 6 (**Appendix B**) explains robust performance across different initializations and instances.

Experiments

In this section, we present comprehensive experiments to validate the effectiveness of our RGTN approach. Our experiments demonstrate that RGTN achieves superior performance in tensor network structure search while requiring significantly less computational time compared to existing methods. Due to space constraints, detailed experimental setup and baseline descriptions are provided in **Appendix D**, with additional experimental settings in **Appendix E**.

Structure Revealing Capability

Table 1 shows RGTN's performance on structure discovery across 100 independent trials with ground truth structures. The high success rates (95-100%) demonstrate that the renormalization group mechanism, when combined with physics-inspired structure proposals, reliably reveals true tensor network structures. The slight variations in success rates correlate with the complexity of the network topology, with simpler structures achieving 100% success rate. It is worth noting that in the test on fifth-order tensors, we consider various topologies, including ring and star configurations with different connectivity patterns. Despite the structural complexity, RGTN can accurately identify the correct topology and rank configuration for each case. This supports our analysis in Lemma 1 about the sparsity-inducing properties of diagonal factors and confirms that RGTN can effectively discover the underlying tensor network structure.

Light Field Data Results

Based on the comprehensive experimental results in Table 2, RGTN demonstrates exceptional performance in both compression efficiency and computational speed across all light field datasets, establishing a new benchmark.

Superior Compression Performance: RGTN achieves state-of-the-art compression across all datasets and error bounds. At the strictest RE bound of 0.01, RGTN delivers compression ratios of 22.3% (Bunny) and 29.9% (Knights),

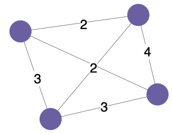
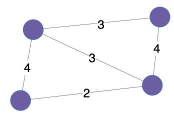
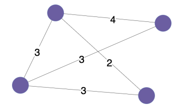
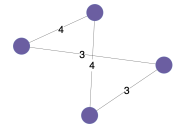
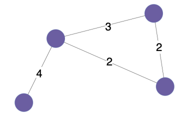
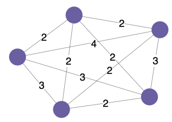
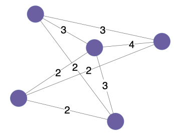
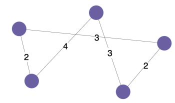
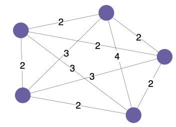
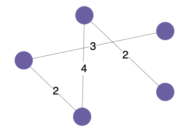
True structure (4th-order)					
Success rate	100%	100%	96%	95%	99%
True structure (5th-order)					
Success rate	100%	98%	96%	97%	100%

Table 1: Performance of RGTN on TN structure revealing under 100 independent tests.

	Bunny							**Knights**					
Method	**RE: 0.01**		**RE: 0.05**		**RE: 0.1**		**Method**	**RE: 0.01**		**RE: 0.05**		**RE: 0.1**	
	CR	**Time**	**CR**	**Time**	**CR**	**Time**		**CR**	**Time**	**CR**	**Time**	**CR**	**Time**
TRALS	61.2%	13.41	17.6%	0.476	5.38%	0.119	TRALS	74.2%	10.42	27.2%	3.877	9.08%	0.427
FCTNALS	64.7%	13.21	21.1%	0.469	3.99%	0.042	FCTNALS	73.9%	12.23	21.2%	0.625	3.88%	**0.014**
TNGreedy	26.4%	10.89	6.39%	1.031	2.37%	0.359	TNGreedy	31.8%	12.66	7.63%	1.353	3.53%	0.486
TNGA	28.2%	1004	5.06%	182.1	2.27%	12.65	TNGA	39.1%	904.5	4.96%	142.0	2.47%	12.37
TNLS	24.5%	1388	4.31%	64.38	2.18%	24.29	TNLS	27.5%	1273	4.78%	74.75	2.13%	5.373
TNALE	26.5%	143.1	4.57%	18.54	2.28%	3.094	TNALE	27.8%	264.1	4.48%	25.30	2.12%	3.352
SVDinsTN	22.6%	0.752	6.85%	**0.029***	2.69%	**0.005**	SVDinsTN	31.7%	1.563	5.70%	**0.105**	2.73%	0.019
RGTN	**22.3%***	**0.180***	**4.14%***	0.193	**0.91%***	0.212	**RGTN**	**29.9%***	**0.178***	**4.06%***	0.201	**1.71%***	0.209

Table 2: Comparison of CR (%) and run time ($\times$1000s) of different methods on Bunny and Knights light field data. The result for RGTN is selected based on the specified RE bounds. **Bold** numbers denote the best performance, underlined numbers represent the second-best results, and * indicates statistical significance at a p $\leq$ 0.05 level using a paired t-test.

Method	6th-order		8th-order	
	CR	Time	CR	Time
TRALS	1.35%	0.006	0.064%	0.034
FCTNALS	2.13%	**0.002**	–	–
TNGreedy	0.88%	0.167	0.016%	2.625
TNGA	0.94%	3.825	0.024%	51.40
TNLS	1.11%	0.673	0.038%	59.83
TNALE	1.65%	0.201	0.047%	19.96
SVDinsTN	1.13%	0.002	0.016%	**0.017**
RGTN	**0.76%**	0.006	**0.009%**	0.123

Table 3: Comparison of the CR ($\downarrow$) and run time ($\times$1000s, $\downarrow$) of different methods when reaching the RE bound of 0.01. The result is the average value of 5 independent experiments and ”–” indicates ”out of memory”.

outperforming SVDinsTN by 1.3% and 6.2%, respectively. The advantage amplifies at higher error tolerances for RE bound 0.1, RGTN achieves remarkable compression ratios of 0.91% (Bunny) and 1.71% (Knights), surpassing SVDinsTN by factors of 2.96$\times$ and 1.60$\times$ respectively. This demonstrates RGTN's ability to identify efficient structures.

Exceptional Computational Efficiency: RGTN completes compression tasks in under 210 seconds for RE bound 0.01, while traditional methods like TNGA and TNLS require over 900,000 seconds, representing speedup factors exceeding 4,500$\times$. Notably, our RGTN implementation uses **Python**, whereas SVDinsTN utilizes **MATLAB** with GPU acceleration, which may provide SVDinsTN computational advantages. Despite this potential implementation disadvantage, RGTN maintains highly competitive runtimes (180s vs 752s for Bunny, 178s vs 1563s for Knights at RE 0.01) while consistently delivering superior compression ratios.

The combination of best-in-class compression and sub-second runtimes validates our renormalization group-inspired approach. RGTN's 2-3$\times$ better compression at higher error bounds establishes it as a powerful solution for applications requiring both effectiveness and efficiency.

Scalability to High-Order Tensors

Table 3 shows RGTN achieves the best compression ratios across all tensor orders while maintaining competitive runtime. For 6th-order tensors, RGTN achieves 0.76% CR, outperforming TNGreedy (0.88%) and SVDinsTN (1.13%). For 8th-order tensors, RGTN achieves 0.009% CR–significantly better than SVDinsTN and TNGreedy (both at 0.016%). While SVDinsTN is fastest, RGTN maintains practical efficiency with 6s (6th-order) and 123s (8th-order). Several methods encounter memory constraints for

MPSNR (↑)			
Method	News	Salesman	Silent
FBCP	28.234	29.077	30.126
TMac	27.882	28.469	30.599
TMacTT	28.714	29.534	30.647
TRLRF	28.857	28.288	31.081
TW	30.027	30.621	31.731
TNLS	29.761	30.685	28.830
SVDinsTN	31.643	31.684	**32.706**
RGTN (Ours)	**32.040**	**31.900**	30.620
Time (seconds, ↓)			
Method	News	Salesman	Silent
FBCP	1720.4	1783.2	1453.9
TMac	340.46	353.63	316.21
TMacTT	535.97	656.45	1305.6
TRLRF	978.12	689.35	453.24
TW	1426.3	1148.7	1232.0
TNLS	37675	76053	98502
SVDinsTN	932.42	769.54	532.31
RGTN (Ours)	**135.95**	**144.00**	**142.80**

Table 4: Comparison of MPSNR and run time of different TC methods on color videos.

8th-order tensors, whereas RGTN handles these cases successfully. Traditional methods like TNGA and TNLS require 51,000-60,000s for 8th-order tensors–over 400× slower than RGTN–while achieving inferior compression. The widening performance gap with increasing tensor order validates our theoretical framework. RGTN's 1.8× better compression on 8th-order tensors, while avoiding memory issues, demonstrates that the renormalization group approach effectively manages exponentially large search spaces.

Video Completion Results

Table 4 presents results on real-world video completion tasks. RGTN achieves the highest MPSNR values on News (32.040 dB) and Salesman (31.900 dB) videos, outperforming the second-best method, SVDinsTN, by 0.397 dB and 0.216 dB, respectively. On Silent video, RGTN maintains competitive performance (30.620 dB). Remarkably, RGTN accomplishes this superior reconstruction quality while being dramatically faster, completing tasks in approximately 140 seconds compared to SVDinsTN's 532-932 seconds, representing 3.7-6.9× speedup. This significant runtime advantage over SVDinsTN on video data stems from RGTN's hierarchical processing strategy. While SVDinsTN must search through numerous possible tensor network structures for the high-dimensional video tensors (with spatial, temporal, and color dimensions), RGTN's renormalization group approach efficiently navigates this search space by operating at multiple scales. The coarse-to-fine refinement naturally captures video's inherent multi-scale structure—from frame-level temporal patterns to pixel-level spatial details—without exhaustively evaluating all possible decompositions. Additionally, RGTN achieves orders-of-magnitude speedup over traditional methods: 8-12× faster than FBCP/TW (1,200-1,800 seconds) and 277-690× faster than TNLS (37,675-98,502 seconds). This exceptional efficiency, combined with state-of-the-art reconstruction quality, validates that our unified structure-parameter optimization effectively exploits the hierarchical nature of video data through the renormalization group framework.

Related Works

Tensor network structure search (TN-SS) addresses the critical limitation of predetermined topologies in tensor networks by automatically discovering optimal configurations (Ghadiri et al. 2023; Sedighin, Cichocki, and Phan 2021; Nie, Wang, and Tian 2021). Traditional TN-SS methods—including greedy construction (Hashemizadeh et al. 2020), genetic algorithms (Li and Sun 2020), and local search (Li et al. 2023a) which follows a costly two-stage sampling-evaluation paradigm where each candidate requires full tensor optimization. The recent SVDinsTN (Zheng et al. 2024) achieves 100-1000× speedup by reformulating TN-SS as unified optimization with sparsity-inducing regularization on diagonal factors between cores. However, it remains susceptible to local minima due to single-scale optimization. Due to space constraints, additional related work on tensor networks and renormalization group applications is provided in the **Appendix F**.

Conclusion and Discussion

In this paper, we introduced RGTN, a physics-inspired framework that transforms tensor network structure search through multi-scale renormalization group flows. Unlike existing methods constrained by discrete search spaces, RGTN implements continuous topology evolution via learnable edge gates and systematic coarse-graining operations. Our theoretical analysis establishes exponential computational speedup with stronger convergence guarantees, while the multi-scale framework escapes local minima through scale-induced perturbations. Experiments across structure discovery, light field compression, high-order tensor decomposition, and video completion demonstrate RGTN's superior performance. By unifying structure search and parameter optimization through physics-inspired metrics of node tension and edge information flow, RGTN eliminates computational overhead while providing principled structure modifications beyond heuristic search strategies.

Acknowledgements

This research was partially supported by National Natural Science Foundation of China (No.62502404), Hong Kong Research Grants Council (Research Impact Fund No.R1015-23, Collaborative Research Fund No.C1043-24GF, General Research Fund No.11218325), Institute of Digital Medicine of City University of Hong Kong (No.9229503), Huawei (Huawei Innovation Research Program), Tencent (CCF-Tencent Open Fund, Tencent Rhino-Bird Focused Research Program), Alibaba (CCF-Alimama Tech Kangaroo Fund No. 2024002), Didi (CCF-Didi Gaia Scholars Research Fund), Kuaishou, and Bytedance.

References

Chan, G. K.-L.; Dorando, J. J.; Ghosh, D.; Hachmann, J.; Neuscamman, E.; Wang, H.; and Yanai, T. 2008. An introduction to the density matrix renormalization group ansatz in quantum chemistry. In *Frontiers in quantum systems in chemistry and physics*, 49–65. Springer.

Evenbly, G.; and Vidal, G. 2011. Tensor network states and geometry. *Journal of Statistical Physics*, 145(4): 891–918.

Gao, J.; Zhao, X.; Chen, B.; Yan, F.; Guo, H.; and Tang, R. 2023. AutoTransfer: Instance transfer for cross-domain recommendations. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*, 1478–1487.

Ghadiri, M.; Fahrbach, M.; Fu, G.; and Mirrokni, V. 2023. Approximately optimal core shapes for tensor decompositions. In *International Conference on Machine Learning*, 11237–11254. PMLR.

Haegeman, J.; Osborne, T. J.; Verschelde, H.; and Verstraete, F. 2013. Entanglement renormalization for quantum fields in real space. *Physical review letters*, 110(10): 100402.

Hashemizadeh, M.; Liu, M.; Miller, J.; and Rabusseau, G. 2020. Adaptive tensor learning with tensor networks. *arXiv preprint arXiv:2008.05437*.

Iacovides, G.; Zhou, W.; Li, C.; Zhao, Q.; and Mandic, D. 2025. Domain-Aware Tensor Network Structure Search. *arXiv preprint arXiv:2505.23537*.

Jia, P.; Wang, Y.; Du, Z.; Zhao, X.; Wang, Y.; Chen, B.; Wang, W.; Guo, H.; and Tang, R. 2024. Erase: Benchmarking feature selection methods for deep recommender systems. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5194–5205.

Li, C.; and Sun, Z. 2020. Evolutionary topology search for tensor network decomposition. In *International conference on machine learning*, 5947–5957. PMLR.

Li, C.; Zeng, J.; Li, C.; Caiafa, C. F.; and Zhao, Q. 2023a. Alternating local enumeration (tnale): Solving tensor network structure search with fewer evaluations. In *International conference on machine learning*, 20384–20411. PMLR.

Li, C.; Zeng, J.; Tao, Z.; and Zhao, Q. 2022. Permutation search of tensor network structures via local sampling. In *International conference on machine learning*, 13106–13124. PMLR.

Li, M.; Zhang, Z.; Zhao, X.; Wang, W.; Zhao, M.; Wu, R.; and Guo, R. 2023b. Automlp: Automated mlp for sequential recommendations. In *Proceedings of the ACM web conference 2023*, 1190–1198.

Liao, H.-J.; Liu, J.-G.; Wang, L.; and Xiang, T. 2019. Differentiable programming tensor networks. *Physical Review X*, 9(3): 031041.

Lin, W.; Zhao, X.; Wang, Y.; Zhu, Y.; and Wang, W. 2023. Autodenoise: Automatic data instance denoising for recommendations. In *Proceedings of the ACM Web Conference 2023*, 1003–1011.

Liu, D.; Yang, C.; Tang, X.; Wang, Y.; Lyu, F.; Luo, W.; He, X.; Ming, Z.; and Zhao, X. 2024a. MultiFS: Automated multi-scenario feature selection in deep recommender systems. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 434–442.

Liu, H.; Zhao, X.; Wang, C.; Liu, X.; and Tang, J. 2020. Automated embedding size search in deep recommender systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2307–2316.

Liu, Z.; Chen, K.; Song, F.; Chen, B.; Zhao, X.; Guo, H.; and Tang, R. 2024b. Autoassign+: Automatic shared embedding assignment in streaming recommendation. *Knowledge and Information Systems*, 66(1): 89–113.

Nie, C.; Wang, H.; and Tian, L. 2021. Adaptive Tensor Networks Decomposition. In *BMVC*, 148.

Orús, R. 2014. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of physics*, 349: 117–158.

Qu, Y.; Chen, T.; Zhao, X.; Cui, L.; Zheng, K.; and Yin, H. 2023. Continuous input embedding size search for recommender systems. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 708–717.

Schollwöck, U. 2011. The density-matrix renormalization group in the age of matrix product states. *Annals of physics*, 326(1): 96–192.

Sedighin, F.; Cichocki, A.; and Phan, A.-H. 2021. Adaptive rank selection for tensor ring decomposition. *IEEE Journal of Selected Topics in Signal Processing*, 15(3): 454–463.

Shankar, R. 1994. Renormalization-group approach to interacting fermions. *Reviews of Modern Physics*, 66(1): 129.

Ueda, A. 2024. Renormalization group flow and fixed-point in tensor network representations. *arXiv preprint arXiv:2401.18068*.

Vidal, G. 2007. Entanglement renormalization. *Physical review letters*, 99(22): 220405.

Wang, M.; Pan, Y.; Xu, Z.; Yang, X.; Li, G.; and Cichocki, A. 2023a. Tensor networks meet neural networks: A survey and future perspectives. *arXiv preprint arXiv:2302.09019*.

Wang, M.; Su, Z.; Luo, X.; Pan, Y.; Zheng, S.; and Xu, Z. 2020. Concatenated tensor networks for deep multi-task learning. In *International Conference on Neural Information Processing*, 517–525. Springer.

Wang, M.; Wei, T.; Zhang, S.; Guo, R.; Wang, W.; Ye, S.; Zou, L.; Wei, X.; and Zhao, X. 2025a. DANCE: Resource-Efficient Neural Architecture Search with Data-Aware and Continuous Adaptation. *arXiv preprint arXiv:2507.04671*.

Wang, M.; Zeng, D.; Xu, Z.; Guo, R.; and Zhao, X. 2023b. Federated knowledge graph completion via latent embedding sharing and tensor factorization. In *2023 IEEE International Conference on Data Mining (ICDM)*, 1361–1366. IEEE.

Wang, M.; Zhao, X.; Guo, R.; and Wang, J. 2025b. Metalora: Tensor-enhanced adaptive low-rank fine-tuning. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, 4680–4684. IEEE.

Wang, M.; Zhen, Y.; Pan, Y.; Zhao, Y.; Zhuang, C.; Xu, Z.; Guo, R.; and Zhao, X. 2024. Tensorized hypergraph neural networks. In *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*, 127–135. SIAM.

White, S. R. 1992. Density matrix formulation for quantum renormalization groups. *Physical review letters*, 69(19): 2863.

Zhang, S.; Wang, M.; Zhao, X.; Guo, R.; Zhao, Y.; Zhuang, C.; Gu, J.; Zhang, Z.; and Yin, H. 2024. Dns-rec: Data-aware neural architecture search for recommender systems. In *Proceedings of the 18th ACM Conference on Recommender Systems*, 591–600.

Zhao, X.; Liu, H.; Fan, W.; Liu, H.; Tang, J.; and Wang, C. 2021a. Autoloss: Automated loss function search in recommendations. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 3959–3967.

Zhao, X.; Liu, H.; Fan, W.; Liu, H.; Tang, J.; Wang, C.; Chen, M.; Zheng, X.; Liu, X.; and Yang, X. 2021b. Autoemb: Automated embedding dimensionality search in streaming recommendations. In *2021 IEEE International Conference on Data Mining (ICDM)*, 896–905. IEEE.

Zhao, X.; Liu, H.; Liu, H.; Tang, J.; Guo, W.; Shi, J.; Wang, S.; Gao, H.; and Long, B. 2021c. Autodim: Field-aware embedding dimension searchin recommender systems. In *Proceedings of the Web Conference 2021*, 3015–3022.

Zheng, L.; Jia, C.; Sun, M.; Wu, Z.; Yu, C. H.; Haj-Ali, A.; Wang, Y.; Yang, J.; Zhuo, D.; Sen, K.; et al. 2020. Ansor: Generating {High-Performance} tensor programs for deep learning. In *14th USENIX symposium on operating systems design and implementation (OSDI 20)*, 863–879.

Zheng, Y.-B.; Zhao, X.-L.; Zeng, J.; Li, C.; Zhao, Q.; Li, H.-C.; and Huang, T.-Z. 2024. Svdinstn: A tensor network paradigm for efficient structure search from regularized modeling perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 26254–26263.

Zhu, C.; Chen, B.; Guo, H.; Xu, H.; Li, X.; Zhao, X.; Zhang, W.; Yu, Y.; and Tang, R. 2023. Autogen: An automated dynamic model generation framework for recommender system. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 598–606.