

Novelty Assessment Report

Paper: Guiding Mixture-of-Experts with Temporal Multimodal Interactions

PDF URL: <https://openreview.net/pdf?id=qF9WJxvHX8>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-01

Abstract

Mixture-of-Experts (MoE) architectures have become pivotal for large-scale multimodal models. However, their routing mechanisms typically overlook the informative, time-varying interaction dynamics between modalities. This limitation hinders expert specialization, as the model cannot explicitly leverage intrinsic modality relationships for effective reasoning. To address this, we propose a novel framework that guides MoE routing using quantified temporal interaction. A multimodal interaction-aware router learns to dispatch tokens to experts based on the nature of their interactions. This dynamic routing encourages experts to acquire generalizable interaction-processing skills rather than merely learning task-specific features. Our framework builds on a new formulation of temporal multimodal interaction dynamics, which are used to guide expert routing. We first demonstrate that these temporal multimodal interactions reveal meaningful patterns across applications, and then show how they can be leveraged to improve both the design and performance of MoE-based models. Comprehensive experiments on challenging multimodal benchmarks validate our approach, demonstrating both enhanced performance and improved interpretability.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Guiding Mixture-of-Experts Routing with Temporal Multimodal Interaction Dynamics**

A total of **28 papers** were analyzed and organized into a taxonomy with **12 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Multimodal Fusion and Routing Mechanisms**
- **Spatiotemporal Forecasting with MoE**
- **Domain-Specific MoE Applications**
- **Omnimodal and Large-Scale Multimodal Models**

Complete Taxonomy Tree

- Guiding Mixture-of-Experts Routing with Temporal Multimodal Interaction Dynamics Survey Taxonomy
- Multimodal Fusion and Routing Mechanisms
 - Temporal Interaction-Guided Routing ★ (3 papers)
 - [0] Guiding Mixture-of-Experts with Temporal Multimodal Interactions (Anon et al., 2026) [View paper](#)
 - [1] Fusemoe: Mixture-of-experts transformers for fleximodal fusion (Xing Han, 2024) [View paper](#)
 - [7] Dual Routing Mixture-of-Experts for Multi-Scale Representation Learning in Multimodal Emotion Recognition (C Da-Eun, 2025) [View paper](#)
 - Adaptive Modality Handling with MoE (3 papers)
 - [3] CAG-MoE: Multimodal Emotion Recognition with Cross-Attention Gated Mixture of Experts (Axel Gedeon Mengara Mengara, 2025) [View paper](#)
 - [6] Hierarchical moe: Continuous multimodal emotion recognition with incomplete and asynchronous inputs (Zhu Yitong, 2025) [View paper](#)
 - [13] MAESTRO: Adaptive Sparse Attention and Robust Learning for Multimodal Dynamic Time Series (Payal Mohapatra, 2025) [View paper](#)
 - Graph-Augmented and Hierarchical Routing (3 papers)
 - [2] Graph mixture of experts and memory-augmented routers for multivariate time series anomaly detection (Xiaoyu Huang, 2025) [View paper](#)
 - [10] H3M-SSMoEs: Hypergraph-based Multimodal Learning with LLM Reasoning and Style-Structured Mixture of Experts (Xie Liang, 2025) [View paper](#)
 - [28] Architect Thyself: Neural Darwinism and Self-Evolving Multimodal Networks (A Sar, n.d.) [View paper](#)
- Spatiotemporal Forecasting with MoE
 - Traffic and Urban Prediction (6 papers)
 - [11] STMMoE: A Spatio-Temporal Multimodal Mixture-of-Experts Model for Urban Traffic Prediction (Kenan Kang, 2025) [View paper](#)
 - [18] STMMOE: A Spatio-Temporal Multimodal Mixture-of-Experts Model for Urban (K Kang, 2025) [View paper](#)
 - [22] Spatiotemporal Graph Mixture of Experts for Highway Traffic Flow Prediction (Sheng Hao, 2025) [View paper](#)
 - [23] STAMImputer: Spatio-Temporal Attention MoE for Traffic Data Imputation (Wang Yi-ming, 2024) [View paper](#)
 - [24] ST-MoE: Spatio-Temporal Mixture-of-Experts for Debiasing in Traffic Prediction (Shuhao Li, 2023) [View paper](#)
 - [25] Spatial-Temporal Graph Multi-Gate Mixture-of-Expert Model for Traffic Prediction (Tong Guan, 2023) [View paper](#)
 - General Multivariate Time Series Forecasting (1 papers)
 - [26] ST-MoE: Spatio-Temporal Mixture of Experts for Multivariate Time Series Forecasting (Hanwen, 2023) [View paper](#)

- Domain-Specific MoE Applications
 - Autonomous Driving and Vehicle Behavior (4 papers)
 - [4] A knowledge-driven diffusion policy for end-to-end autonomous driving based on expert routing (Xu Chengkai, 2025) [View paper](#)
 - [19] RAST-MoE-RL: A Regime-Aware Spatio-Temporal MoE Framework for Deep Reinforcement Learning in Ride-Hailing (Yuhan Tang, 2025) [View paper](#)
 - [20] A Temporal Multi-Gate Mixture-of-Experts Approach for Vehicle Trajectory and Driving Intention Prediction (Renteng Yuan, 2023) [View paper](#)
 - [21] A Novel Temporal Multi-Gate Mixture-of-Experts Approach for Vehicle Trajectory and Driving Intention Prediction (Yuan, 2023) [View paper](#)
 - Recommendation Systems (2 papers)
 - [5] Hierarchical Time-Aware Mixture of Experts for Multi-Modal Sequential Recommendation (Zhang Sheng-zhe, 2025) [View paper](#)
 - [8] Multimodal Temporal Knowledge Graph Embedding Method Based on Mixture of Experts for Recommendation (Bingchen Liu, 2025) [View paper](#)
 - Video Understanding and Question Answering (1 papers)
 - [9] Temporal-guided Mixture-of-Experts for Zero-Shot Video Question Answering (Yixin Qin, 2025) [View paper](#)
 - Human Interaction and Motion Generation (1 papers)
 - [15] InterMoE: Individual-Specific 3D Human Interaction Generation via Dynamic Temporal-Selective MoE (Lipeng Wang, 2025) [View paper](#)
 - Survival Analysis and Biomedical Applications (2 papers)
 - [12] Covariate-guided Bayesian mixture of spline experts for the analysis of multivariate high-density longitudinal data. (Haoyi Fu, 2024) [View paper](#)
 - [17] Dual Mixture-of-Experts Framework for Discrete-Time Survival Analysis (Shin, 2025) [View paper](#)
 - Battery and Energy Systems (1 papers)
 - [27] A Novel Spatiotemporal Graph Topology Method for Serialized Charging Feature Reduction in Lithium-Ion Battery Remaining Useful Life Prediction (Liu Teng, n.d.) [View paper](#)
- Omnimodal and Large-Scale Multimodal Models (2 papers)
 - [14] From Perception to Reasoning and Interaction: A Comprehensive Survey of Multimodal Intelligence in Large Language Models (W Qian, 2025) [View paper](#)
 - [16] Uni-MoE-2.0-Omni: Scaling Language-Centric Omnimodal Large Model with Advanced MoE, Training and Data (Yunxin Li, 2025) [View paper](#)

Narrative

Core task: Guiding mixture-of-experts routing with temporal multimodal interaction dynamics. The field explores how to effectively route information through specialized expert networks when dealing with multiple modalities that evolve over time. The taxonomy reveals four main branches: Multimodal Fusion and Routing Mechanisms focuses on designing routing strategies that leverage cross-modal interactions and temporal dependencies, with works like Fusemoe[1] and Dual Routing MoE[7] developing sophisticated gating mechanisms; Spatiotemporal Forecasting with MoE applies these architectures to prediction tasks in urban computing and time-series domains, exemplified by ST-MoE Forecasting[26] and STMMOE Urban[18]; Domain-Specific MoE Applications tailors mixture-of-experts to specialized settings such as emotion recognition, survival analysis, and reinforcement learning; and Omnimodal and Large-Scale Multimodal Models scales these ideas to handle diverse input types simultaneously, as seen in Uni-MoE-2.0-Omni[16]. These branches collectively address the challenge of dynamically selecting and combining expert knowledge based on both modality characteristics and temporal context.

A particularly active line of work centers on temporal interaction-guided routing, where the key question is how to make routing decisions sensitive to evolving multimodal relationships rather than treating modalities as static inputs. Temporal Multimodal MoE[0] sits squarely within this cluster, emphasizing how temporal dynamics between modalities should inform which experts are activated at each time step. This contrasts with approaches like Hierarchical Time MoE[5], which structures experts hierarchically across temporal scales but may not explicitly model cross-modal interaction patterns, and Temporal MoE VideoQA[9], which applies temporal routing primarily to video question-answering tasks with a narrower scope. The trade-off across these works involves balancing routing complexity against computational efficiency: more sophisticated temporal and cross-modal routing can improve specialization but risks increased overhead and training instability, a challenge that remains central to advancing mixture-of-experts architectures in multimodal temporal settings.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Fusemoe: Mixture-of-experts transformers for fleximodal fusion

Authors: Xing Han, Carl Harris, Nhat Ho, Huy Nguyen, Suchi Saria | **Year/Venue:** 2024 | **URL:** [View paper](#)

Abstract

and irregular temporal dynamics present in multimodal data; the Top-K router for effective multimodal fusion, considering , which, in conjunction with entropy regularization, guide Z

Relationship Analysis

Both papers belong to the Temporal Interaction-Guided Routing category, focusing on leveraging time-varying multimodal dynamics to guide expert selection in MoE architectures. FuseMoE overlaps with the original paper in addressing temporal irregularity and multimodal fusion through MoE routing mechanisms, particularly in handling missing modalities and irregular sampling. However, FuseMoE differs by primarily focusing on 'FlexiModal' data scenarios with a novel Laplace gating function and theoretical convergence guarantees, while the original paper emphasizes explicit quantification of temporal RUS (redundancy, uniqueness, synergy) interactions to directly inform routing decisions through an interaction-aware router with RUS-specific auxiliary losses.

2. Dual Routing Mixture-of-Experts for Multi-Scale Representation Learning in Multimodal Emotion Recognition

Authors: C Da-Eun, L Seok-Pil | **Year/Venue:** 2025 | **URL:** [View paper](#)

Abstract

introduces a Dual Routing MoE framework that adaptively MoE-based models primarily focus on expert routing based mechanism that considers the temporal scale inherent in

Relationship Analysis

Both papers belong to the Temporal Interaction-Guided Routing category, leveraging time-varying multimodal dynamics to guide MoE expert selection. They overlap in using temporal information to inform routing decisions in multimodal contexts. However, the original

paper focuses on quantifying temporal multimodal interactions through information-theoretic RUS (Redundancy, Uniqueness, Synergy) decomposition to guide routing, while the candidate paper emphasizes dual routing mechanisms for multi-scale representation learning specifically in emotion recognition tasks.

Contributions Analysis

Overall novelty summary. The paper proposes a framework that guides mixture-of-experts routing using quantified temporal multimodal interaction dynamics, formulated through directed information decomposition. It resides in the 'Temporal Interaction-Guided Routing' leaf, which contains only three papers total, including this work. This leaf sits within the broader 'Multimodal Fusion and Routing Mechanisms' branch, indicating a relatively sparse research direction focused specifically on leveraging time-varying cross-modal relationships for expert selection, rather than static fusion or domain-specific applications.

The taxonomy reveals that neighboring leaves address related but distinct challenges: 'Adaptive Modality Handling with MoE' focuses on incomplete or asynchronous modalities through dynamic activation, while 'Graph-Augmented and Hierarchical Routing' integrates relational structures and multi-scale representations. The paper's emphasis on temporal interaction dynamics distinguishes it from these directions, which either handle modality availability issues or impose structural priors without explicitly modeling evolving cross-modal relationships. The broader 'Spatiotemporal Forecasting with MoE' branch applies similar architectures to prediction tasks, but excludes non-forecasting multimodal fusion scenarios like the one addressed here.

Among the three contributions analyzed, the temporal multimodal interaction framework examined zero candidates, while the multi-scale BATCH estimator and RUS-aware router examined six and ten candidates respectively, with none identified as clearly refutable. The literature search scope covered sixteen candidates total, drawn from top-K semantic search and citation expansion. This limited examination suggests that within the accessible prior work, no direct overlaps were detected for the specific combination of temporal interaction quantification and interaction-guided routing losses, though the small search scale means substantial related work may exist beyond these candidates.

Given the sparse taxonomy leaf and limited search scope, the work appears to occupy a relatively underexplored niche at the intersection of temporal multimodal interaction modeling and mixture-of-experts routing. However, the analysis is constrained by examining only sixteen candidates and does not constitute an exhaustive literature review. The absence of refutable pairs within this scope suggests potential novelty in the specific technical approach, but broader field coverage would be necessary to assess whether similar interaction-based routing strategies exist in adjacent research communities.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Temporal multimodal interaction framework using directed information decomposition

Description: The authors introduce a formulation of temporal multimodal interactions based on directed information that decomposes multi-source information flow into redundancy, uniqueness, and synergy (RUS) components across multiple time lags. This framework captures time-varying interaction dynamics between modalities with respect to target outcomes.

This contribution was assessed against **0 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

Contribution 2: Multi-scale BATCH estimator for efficient temporal RUS computation

Description: The authors develop an efficient computational method that extends the BATCH estimator to handle high-dimensional temporal data by training a single model to predict temporal RUS values at multiple time lags simultaneously, achieving significant speedup while maintaining accuracy.

This contribution was assessed against **6 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Cohort-Individual Cooperative Learning for Multimodal Cancer Survival Analysis

URL: [View paper](#)

Brief Assessment

Cohort-Individual Cancer Survival[32] applies RUS decomposition to static multimodal cancer data (pathology images and genomic profiles) without temporal dynamics. The original paper's contribution focuses specifically on efficient computation of temporal RUS across multiple time lags using a multi-scale approach, which is fundamentally different from the static multimodal decomposition in the candidate.

2. SI: Score-based O-INFORMATION Estimation

URL: [View paper](#)

Brief Assessment

Score-based O-INFORMATION[33] focuses on continuous variable information decomposition using score-based methods, not on temporal RUS computation or multi-scale BATCH estimation for high-dimensional temporal data.

3. Quantifying & modeling multimodal interactions: An information decomposition framework

URL: [View paper](#)

Brief Assessment

Multimodal Interaction Decomposition[30] proposes the BATCH estimator for static multimodal interactions, not temporal RUS computation. The original paper extends BATCH to handle temporal dynamics with multi-scale lag embeddings, which is a distinct contribution not addressed in the candidate.

4. Information-Theoretic Sequential Framework to Elicit Dynamic High-Order Interactions in High-Dimensional Network Processes

URL: [View paper](#)

Brief Assessment

High-Order Network Interactions[34] focuses on extending information-theoretic measures (II, RSI, OI) to dynamic network processes using stepwise strategies for identifying multipliants, not on developing efficient neural network-based estimators for computing temporal RUS values in high-dimensional multimodal data as the original paper does.

5. Coded aperture design for temporal compressive imaging in a color-polarized video

URL: [View paper](#)

Brief Assessment

Coded Aperture Imaging[29] focuses on temporal compressive imaging for color-polarized video reconstruction using coded apertures and ADMM algorithms. It does not address temporal redundancy-uniqueness-synergy (RUS) computation or information-theoretic multimodal interaction estimation, which are central to the original paper's contribution.

6. Fast-Vid2Vid++: Spatial-Temporal Distillation for Real-Time Video-to-Video Synthesis

URL: [View paper](#)

Brief Assessment

Fast-Vid2Vid++[31] focuses on video-to-video synthesis using spatial-temporal distillation for generative models, not on computing temporal redundancy-uniqueness-synergy (RUS) values for multimodal interaction analysis in high-dimensional data.

Contribution 3: RUS-aware MoE router with interaction-guided auxiliary losses

Description: The authors design an interaction-aware routing mechanism that incorporates temporal RUS sequences through attention and recurrent modules, combined with auxiliary loss functions that enforce routing strategies aligned with redundancy, uniqueness, and synergy principles to improve expert specialization.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Advancing Expert Specialization for Better MoE

URL: [View paper](#)

Brief Assessment

Expert Specialization MoE[35] focuses on orthogonality and variance losses to prevent expert overlap in general token routing, not on temporal multimodal interaction dynamics (RUS) that guide routing based on redundancy, uniqueness, and synergy patterns across time-delayed modalities.

2. Routing Matters in MoE: Scaling Diffusion Transformers with Explicit Routing Guidance

URL: [View paper](#)

Brief Assessment

Routing Diffusion Transformers[39] focuses on visual token routing in diffusion models using conditional/unconditional partitioning and prototypical routing, not temporal multimodal interaction dynamics (RUS) with auxiliary losses for enforcing redundancy, uniqueness, and synergy principles.

3. TA-MoE: Topology-Aware Large Scale Mixture-of-Expert Training

URL: [View paper](#)

Brief Assessment

TA-MoE[41] focuses on topology-aware routing for distributed training optimization across heterogeneous networks, not on interaction-aware routing using temporal multimodal interactions or auxiliary losses based on redundancy, uniqueness, and synergy principles.

4. Coupling Experts and Routers in Mixture-of-Experts via an Auxiliary Loss

URL: [View paper](#)

Brief Assessment

Coupling Experts Routers[38] focuses on coupling router embeddings with expert capabilities through internal activations, not on temporal multimodal interactions or RUS-based routing strategies. The auxiliary loss mechanisms serve fundamentally different purposes.

5. EvidenceMoE: A Physics-Guided Mixture-of-Experts with Evidential Critics for Advancing Fluorescence Light Detection and Ranging in Scattering Media

URL: [View paper](#)

Brief Assessment

EvidenceMoE[42] focuses on physics-guided expert models for fluorescence lidar signal processing in scattering media, using evidence-based critics for quality assessment. This is fundamentally different from the original paper's information-theoretic routing mechanism based on temporal multimodal interactions (redundancy, uniqueness, synergy).

6. Adamoe: Token-adaptive routing with null experts for mixture-of-experts language models

URL: [View paper](#)

Brief Assessment

Adamoe[36] focuses on token-adaptive routing with null experts for efficiency in language models, not on multimodal interaction dynamics or RUS-based routing mechanisms.

7. Fusemoe: Mixture-of-experts transformers for fleximodal fusion

URL: [View paper](#)

Brief Assessment

Fusemoe[1] uses entropy regularization for load balancing across experts and does not incorporate temporal multimodal interaction dynamics (RUS) into its routing mechanism. The candidate focuses on handling missing modalities and irregularity through per-modality routers, not on interaction-aware routing guided by redundancy, uniqueness, and synergy principles.

8. Guiding the Experts: Semantic Priors for Efficient and Focused MoE Routing

URL: [View paper](#)

Brief Assessment

Semantic Priors Routing[40] focuses on spatial semantic alignment in vision models using foreground-guided auxiliary losses, while the original paper addresses temporal multimodal interaction dynamics across different modalities using information-theoretic RUS decomposition. The technical approaches and application domains are fundamentally different.

9. SMAR: Soft Modality-Aware Routing Strategy for MoE-based Multimodal Large Language Models Preserving Language Capabilities

URL: [View paper](#)

Brief Assessment

SMAR[37] focuses on modality-aware routing using KL divergence regularization to preserve language capabilities in multimodal LLMs, not on temporal interaction dynamics or RUS-based auxiliary losses for expert specialization.

10. Mixture-of-Experts Meets In-Context Reinforcement Learning

URL: [View paper](#)

Brief Assessment

MoE In-Context RL[43] focuses on token-wise and task-wise routing for in-context reinforcement learning with contrastive learning for task routing, not on temporal multimodal interaction dynamics (RUS) with information-theoretic auxiliary losses for guiding expert specialization.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Guiding Mixture-of-Experts with Temporal Multimodal Interactions [View paper](#)
- [1] Fusemoe: Mixture-of-experts transformers for fleximodal fusion [View paper](#)
- [2] Graph mixture of experts and memory-augmented routers for multivariate time series anomaly detection [View paper](#)
- [3] CAG-MoE: Multimodal Emotion Recognition with Cross-Attention Gated Mixture of Experts [View paper](#)
- [4] A knowledge-driven diffusion policy for end-to-end autonomous driving based on expert routing [View paper](#)
- [5] Hierarchical Time-Aware Mixture of Experts for Multi-Modal Sequential Recommendation [View paper](#)
- [6] Hierarchical moe: Continuous multimodal emotion recognition with incomplete and asynchronous inputs [View paper](#)
- [7] Dual Routing Mixture-of-Experts for Multi-Scale Representation Learning in Multimodal Emotion Recognition [View paper](#)
- [8] Multimodal Temporal Knowledge Graph Embedding Method Based on Mixture of Experts for Recommendation [View paper](#)
- [9] Temporal-guided Mixture-of-Experts for Zero-Shot Video Question Answering [View paper](#)
- [10] H3M-SSMoEs: Hypergraph-based Multimodal Learning with LLM Reasoning and Style-Structured Mixture of Experts [View paper](#)
- [11] STMMoE: A Spatio-Temporal Multimodal Mixture-of-Experts Model for Urban Traffic Prediction [View paper](#)
- [12] Covariate-guided Bayesian mixture of spline experts for the analysis of multivariate high-density longitudinal data. [View paper](#)
- [13] MAESTRO: Adaptive Sparse Attention and Robust Learning for Multimodal Dynamic Time Series [View paper](#)
- [14] From Perception to Reasoning and Interaction: A Comprehensive Survey of Multimodal Intelligence in Large Language Models [View paper](#)
- [15] InterMoE: Individual-Specific 3D Human Interaction Generation via Dynamic Temporal-Selective MoE [View paper](#)
- [16] Uni-MoE-2.0-Omni: Scaling Language-Centric Omnimodal Large Model with Advanced MoE, Training and Data [View paper](#)
- [17] Dual Mixture-of-Experts Framework for Discrete-Time Survival Analysis [View paper](#)
- [18] STMMOE: A Spatio-Temporal Multimodal Mixture-of-Experts Model for Urban [View paper](#)
- [19] RAST-MoE-RL: A Regime-Aware Spatio-Temporal MoE Framework for Deep Reinforcement Learning in Ride-Hailing [View paper](#)
- [20] A Temporal Multi-Gate Mixture-of-Experts Approach for Vehicle Trajectory and Driving Intention Prediction [View paper](#)
- [21] A Novel Temporal Multi-Gate Mixture-of-Experts Approach for Vehicle Trajectory and Driving Intention Prediction [View paper](#)
- [22] Spatiotemporal Graph Mixture of Experts for Highway Traffic Flow Prediction [View paper](#)
- [23] STAMImputer: Spatio-Temporal Attention MoE for Traffic Data Imputation [View paper](#)
- [24] ST-MoE: Spatio-Temporal Mixture-of-Experts for Debiasing in Traffic Prediction [View paper](#)
- [25] Spatial-Temporal Graph Multi-Gate Mixture-of-Expert Model for Traffic Prediction [View paper](#)
- [26] ST-MoE: Spatio-Temporal Mixture of Experts for Multivariate Time Series Forecasting [View paper](#)
- [27] A Novel Spatiotemporal Graph Topology Method for Serialized Charging Feature Reduction in Lithium-Ion Battery Remaining Useful Life Prediction [View paper](#)
- [28] Architect Thyself: Neural Darwinism and Self-Evolving Multimodal Networks [View paper](#)
- [29] Coded aperture design for temporal compressive imaging in a color-polarized video [View paper](#)
- [30] Quantifying & modeling multimodal interactions: An information decomposition framework [View paper](#)
- [31] Fast-Vid2Vid++: Spatial-Temporal Distillation for Real-Time Video-to-Video Synthesis [View paper](#)
- [32] Cohort-Individual Cooperative Learning for Multimodal Cancer Survival Analysis [View paper](#)
- [33] SI: Score-based O-INFORMATION Estimation [View paper](#)
- [34] Information-Theoretic Sequential Framework to Elicit Dynamic High-Order Interactions in High-Dimensional Network Processes [View paper](#)
- [35] Advancing Expert Specialization for Better MoE [View paper](#)
- [36] Adamoe: Token-adaptive routing with null experts for mixture-of-experts language models [View paper](#)
- [37] SMAR: Soft Modality-Aware Routing Strategy for MoE-based Multimodal Large Language Models Preserving Language Capabilities [View paper](#)
- [38] Coupling Experts and Routers in Mixture-of-Experts via an Auxiliary Loss [View paper](#)
- [39] Routing Matters in MoE: Scaling Diffusion Transformers with Explicit Routing Guidance [View paper](#)
- [40] Guiding the Experts: Semantic Priors for Efficient and Focused MoE Routing [View paper](#)
- [41] TA-MoE: Topology-Aware Large Scale Mixture-of-Expert Training [View paper](#)
- [42] EvidenceMoE: A Physics-Guided Mixture-of-Experts with Evidential Critics for Advancing Fluorescence Light Detection and Ranging in Scattering Media [View paper](#)
- [43] Mixture-of-Experts Meets In-Context Reinforcement Learning [View paper](#)