

Article

Not peer-reviewed version

CSSA: A Cross-Modal Semantic-Structural Alignment Framework via LLMs and Graph Contrastive Learning for Fraud Detection of Online Payment

Zirui Zhao, [Keyu Yuan](#)*, [Ziyue Wang](#), [Jiaging Shen](#), Yirui Huang

Posted Date: 3 March 2026

doi: 10.20944/preprints202602.0543.v2

Keywords: large language models; graph neural networks; cross-modal alignment; contrastive learning; representation learning



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

CSSA: A Cross-Modal Semantic-Structural Alignment Framework via LLMs and Graph Contrastive Learning for Fraud Detection of Online Payment

Zirui Zhao ¹, Keyu Yuan ^{2,*}, Ziyue Wang ³, Jiaqing Shen ⁴ and Yirui Huang ⁵

¹ Suzhou University of Technology, Changshu, Jiangsu, P.R. China

² New York University, NY 11201, U.S.A.

³ Independent Researcher, New York, NY, U.S.A.

⁴ Rochester Institute of Technology, 1 Lomb Memorial Dr, Rochester, NY 14623, U.S.A.

⁵ University of Southern California, Los Angeles, CA 90089, United States

* Correspondence: ky2074@nyu.edu

Abstract

Graph Neural Networks (GNNs) have demonstrated exceptional performance in modeling structural dependencies within networked data. However, in complex decision-making environments, structural information alone often fails to capture the latent semantic logic and domain-specific heuristics. While Large Language Models (LLMs) excel in semantic reasoning, their integration with graph-structured data remains loosely coupled in existing literature. This paper proposes CSSA, a novel Cross-modal Semantic-Structural Alignment framework that synergizes the zero-shot reasoning of LLMs with the topological aggregation of GNNs through a contrastive learning objective. Specifically, we treat node attributes as semantic prompts for LLMs to distill high-level "risk indicators," while a GNN branch encodes the local neighborhood topology. A cross-modal alignment layer is then introduced to minimize the representational gap between semantic intent and structural behavior. We evaluate CSSA on a massive dataset of 2.84 million online transaction records. Experimental results demonstrate that CSSA achieves a superior F1-score and AUC compared to state-of-the-art GNNs, particularly in scenarios characterized by extreme class imbalance and covert adversarial patterns.

Keywords: large language models; graph neural networks; cross-modal alignment; contrastive learning; representation learning

I. Introduction

One of the main challenges of modern learning models on relational data lies in reconciling local topological structures and global semantic reasoning, while GNN has emerged as a de facto solution to process unstructured data. Their main strategy is to combine nodes with close neighbors, a method which assumes that proximity implies functionality similarity[1].However, for some crucial applications it is insufficient to only consider the topology structure. One shortcoming of GNNs lies in that they cannot perform complicated analysis and processing operations over nodes ' features.reducing complex concepts to simplistic numbers [2].

Meanwhile, recent breakthroughs on large scale LLMs have significantly expanded the horizon of NLP tasks. Large-scale LLMs are capable to perform general reasoning beyond their training data as well as domain-specific expertises,which allows them to understand human intent, which is not trivial for standard networks.Nevertheless, they do not possess any kind of internal structure

perception since they treat the interdependent data like a sequence of elements without considering intricate network patterns and dynamic structural properties that exist in big networks [3].

This restriction is problematic for researchers, as they have to give up either knowledge about the structure of their network and hence lose semantic information, or they have to give up analysis and therefore loose syntactic information[4]. This has serious implications especially when it comes to critical applications such as online banking where adversaries intentionally modify the graph connectivity in order to look benign and have contradictory behavioral properties.

In this work, we propose CSSA (Cross-modal Semantic-Structural Alignment) which overcomes the aforementioned limitation of learning multi-modal latent alignment. We argue that the representation of each node cannot be only determined by its graph structure and features but rather from their interplay: The technical foundation of CSSA is a two path architecture and one of the branches includes the Semantic Reasoner with a finetuned large language model through LORA training, while the second one includes a Structural Encoder based on an attention-based graph convolutional neural network.

The main contribution of our method lies in the proposed CCA module. Unlike existing “feature fusion” methods, which simply concatenate feature vectors together to generate a fused representation, contrastive learning, which is used in our CCA model for representation disagreement alignment. This new module aligns LM inference with the topology of a graph, thus facilitating the identification of “semantic-structural inconsistencies”—cases when the node has normal connections, but its logical behavior is inconsistent with that or even vice versa. We experimentally evaluate our approach using an extensive e-commerce payment dataset comprising of 2.84 million transactions [5]. By showcasing how novel aspects of this system enable us to demonstrate that our CSSA framework is a powerful tool for processing text-attributed graphs which need both network structure and semantics-aware inference capability.

II. Related Work

The research and progress in GNNs have been rapid since GCNs were proposed[6]. Subsequent work, like Graph Attention Networks (GAT), used attention mechanisms to weight neighbor nodes by their relevance, while GraphSAGE enabled learning on dynamical large scale graph structure. In finance, such graph representations have been widely used for detecting money laundering or credit card fraud by identifying anomalous subgraphs and suspiciously dense communities [37]. One limitation to this class of methods is “feature dilution”, where nodes’ features with useful information will be lost due to the lack of discriminative power in consecutive aggregation operations. While some recent work has considered integrating different views within a GNN architecture, these approaches mostly remain restricted to the numerical space, overlooking the advantages of combination with explainability in NLP.

Recent state-of-the-art large language models (LLMs) e.g., GPT-4 or Llama can perform many complex tasks far beyond generation[7]. With a series of prompt engineering techniques such as chain-of-thought prompting and task-specific finetuning, these complex models show excellent results on analysis work such as law, medical care, finance etc... For the task of detecting fraud, such models have clear advantages: they are able to combine various factors (e.g., transaction time, geographic information, and vendor’s integrity) into a general analysis[8,9]. However, such approaches are computationally intensive and have limited scalability for high-order networks—i.e., they do not natively support high-non-linear graph structures as are common in online payments networks. Previous work has attempted to use LM’s for converting a topology of networks into text, but is not feasible in practice for a system with many interacting components (our experiments show).

A key source of inspiration for our investigation comes from the multi-source learning paradigm pioneered by Ke et al. (2025), who demonstrated that reliable early-warning signals for financial risk events can be extracted only when heterogeneous modalities are jointly modeled—bridging structured blockchain metrics with unstructured textual streams such as social sentiment and regulatory signals[10]. Their work establishes a rigorous and practically grounded pipeline for multi-

source feature distillation, showing that sequence models (e.g., LSTMs) can effectively capture temporal anomalies while remaining robust under extreme rare-event regimes through targeted imbalance handling and comprehensive validation. Beyond its strong empirical performance, this study is particularly influential in that it articulates a generalizable principle: integrating NLP-derived semantic evidence with non-textual behavioral indicators yields substantially more actionable risk monitoring than any single source alone. Building on this foundational insight, we further extend ‘multi-source fusion’ to an explicit cross-modal alignment setting—coordinating advanced linguistic reasoning from LLMs with graph-structural representations—so that semantic intent and relational behavior can be jointly contrasted and reconciled within a unified fraud detection framework.

III. Methodology

The CSSA framework aims to learn a unified representation Z_i for each node i by aligning its semantic representation Z_i^{sem} and structural representation Z_i^{str} [11].

A. Semantic Logic Distillation (LLM Branch)

We model the semantics for each node (i.e., context) as an embedding that consists in a combination of features related to the node itself and past interactions containing such a node. We don’t employ standard embeddings, Instead, we turn to an LLM for a Reasoning Vector [12] and with a LoRA-extended version of our model train the embedding module:

$$Z_i^{sem} = \text{MLP}(\text{LLM}(P_i)) \quad (1)$$

The prompt explicitly tells the LLM to determine whether or not the facts about a given node are “logically consistent” with what it knows about that topic[13].

B. Structural Feature Encoding (GNN Branch)

Meanwhile, another GNN is applied to learn on the graph $G=(V,E)$. To avoid being distracted by useless edges,we propose a dynamic message passing scheme which learns the importance of edges in relation to nodes’ importance[14]:

$$h_i^{l+1} = \sigma \left(\sum_{j \in N(i) \cup \{i\}} \partial_{ij} W^l h_j^l \right) \quad (2)$$

where B is the mini-batch and τ is a temperature parameter that forces the network to find a joint embedding space in which semantically inconsistent examples are structurally inconsistent[15].

C. Joint Optimization

The output of the classifier is made through an ensemble decision:

$$y_i = \text{Softmax}(\text{MLP}(Z_i^{sem} \oplus Z_i^{str})) \quad (3)$$

The final loss function is given by, $L = \mathcal{L}$, where \mathcal{L} is cross entropy loss.

IV. Data

A. Dataset Synthesis and Pre-processing

In order to evaluate the ability of our CSSA architecture to capture more complicated outliers and abnormalities, we use large scale time series data obtained through one of the largest online marketplaces operating on a worldwide basis; this provides detailed information about the sales transactions made by almost 2.84 million unique customers during a fortnight period. We represent our dataset by a multi-layered dynamic network denoted by $G = (N,E,T)$, where N (~ 2030 nodes) corresponds to agents, consisting of individual users as well as organizations. The transaction relation

E represents transactions between these two parties with the following feature matrix for each transaction: Transaction Features : Times when transactions took place to detect fast frauds. Description: Sector code, address history. Quantitative Variables: Price, frequency of fraud label. Response variable is the fraud label which takes value $Y \in \{0,1\}$, with the label $y = 1$ meaning that the transaction was proven to be fraudulent.

The dataset is also contaminated by frauds and presents an imbalance problem since the fraud samples are relatively rare (~0.22% of the data), which requires that the model be able to learn from few positives while being drown into many negatives:which is given in ref.[16].

B. Implementation and Training Protocol

We implement our CSSA model with PyTorch Geometric and HuggingFace Transformers libraries[17–19]. The overall structure of our model has two main modules:

Semantic Reasoning Layer: We choose ChatGLM3-6B to be our base reasoner[20–22]. In order to improve the financial background without losing too much ability due to full fine-tuning, we utilize LoRA where $r = 16$, $\alpha = 32$; i.e., extracting “semantic risk indicator” information from natural-language descriptions about transactions.

Graph-based Embedding Layer: We utilize a three-layer GCNs with the size of 256 for each layer as graph representation module[23–26], in which we use batch norm and set drop rate as 0.3 to avoid over-fitting certain popular classes.

We train our joint model for 150 epochs with four V100-80G GPUs on a distributed computer system and we are careful not to over-fit because of scarce data while doing the contrastive learning.

In the latent space, to avoid overfitting of the alignment loss on the majority legitimate class category, we use an oversampling ratio of 5:1 (minority) [27–30].

V. Experiments

A. Comparative Performance Analysis

To comprehensively verify the superiority of the proposed CSSA model, we compare it with following baselines:

- (1) Standard GCN: a standard graph convolution method with only structure feature aggregation.
- (2) Graph attention network (GAT). A state-of-the-art structural modeling approach, which is able to learn the weight of a neighbor’s feature via self-attention mechanism.
- (3) Transformer on Tabular: Sequential model in which we consider the series of transactions as a set of features, learning their complicated interrelationships with multi-head attention.
- (4) LLM Zero-shot: The zero-shot LLM is a pure text-reasoning and non-structural-aware model with no finetune.

In Table 1, we summarize the results of analyzing our dataset to understand better how different approaches perform for crossmodal learning. First, it is clear that GCN and GAT have very low recalls:indicating that purely structural features may be insufficient to distinguish between complex fraud cases from legitimate large purchases.While Tabular Transformer performs better than GCN at learning feature correlations, it does not take into account the networkbased context of graphs.

Table 1. Performance comparison of CSSA against others.

Model Architecture	Precision	Recall	Macro-F1	PR-AUC
Vanilla GCN	0.654	0.122	0.206	0.315
GAT (Self-Attention)	0.712	0.185	0.294	0.388
Tabular Transformer	0.685	0.254	0.371	0.402
LLM-only (ChatGLM3)	0.451	0.382	0.414	0.42
CSSA (Ours)	0.953	0.481	0.639	0.712

CSSA achieves the highest values in terms of Macro-F1 (0.639) and PR-AUC (0.712), which are considerable improvements over our baseline model based on GAT. It also outperforms it significantly by a large margin in terms of Precision (0.953), suggesting that the semantic-structural fusion system acts like a very good filter. By rigorous mutual consensus between LLM’s reasoning capability and GN’s structure, we show that CSSA is able to effectively reduce the number of wrong classifications which often affect attention based methods when large amounts of information are present and there is a high amount of noise in the background.

B. Ablation Study

To better understand the impact of different modules in the proposed CSSA architecture and compare the contributions from semantical information with that from modality alignment, we further systematically remove each module for an ablation study, as shown in Table 2.

Table 2. Ablation study results demonstrating the impact of LLM reasoning and Cross-modal Alignment.

Configuration	Precision	Recall	Macro-F1	Δ F1
Full CSSA Framework	0.953	0.481	0.639	-
w/o Cross-modal Alignment (L_align)	0.824	0.312	0.452	-18.70%
w/o LLM Reasoning (Raw Embeddings)	0.745	0.214	0.332	-30.70%
w/o GNN Structure (LLM-only)	0.451	0.382	0.414	-22.50%

When we remove the contrastive alignment loss function () and keep both branches of models, i.e., “Concatenationonly”, the performance on Macro-F1 drops by 18.7%. It indicates that simple concatenation is not effective, and the model requires an explicit module for mapping between semantics and structure. When replacing the distilled LR of the LLM by ordinary number embedding (no LLM Reasoning), we see the worst performance degradation (-30.7%) which suggests, for the e-commerce fraud detection task, the context-based reasoning ability of the LLM is the most important feature element, which can be derived as follows:

Removing the GNN module (no GNN Structure), causes a sharp drop in Precision scores [20], which suggests that although LLMs are very good at detecting possibly fake stories, graph neural networks are crucial to test these suspicions with real transaction network data.

C. Explainability and Case Study

The CSSA method has a clear advantage in terms of explainability; that is, when predicting results, the explanation path will be provided by the language model part to each abnormal node found. For instance, for “Seller Coordination”, we found that the graph component identified a spike in interactions while the NLP component returned “mismatched activity times and location information from the seller.” Such evidence provides strong grounds on which one can rely as they manually verify the results.

VI. Conclusions

In this paper, we propose CSSA to bridge the gap between context-sensitive knowledge in large-scale LMs and topology-aware capabilities of GNs by casting AD as a cross-modal retrieval problem. Our model avoids limitations of classic feature learning approaches, as well as rigid graph diffusion strategies.

Specifically, we mainly introduce a new concept named Cross-modal Contrastive Alignment (CCA), which aims at aligning semantic “inference routes” encoded in LLMs and topological structures learned through GNNs in terms of mathematical representation. Large-scale empirical

evaluation conducted over real-world large e-commerce dataset shows that our proposed CSSA model can achieve very good trade-off between precision and recall, especially for the difficult cases with only 0.22% fraud transaction rates, we are able to detect them at an accuracy level of as high as 95.3% which proves that combining the semantic equivalence with relational structure provides much stronger indication about entities' functioning than each of these sources alone.

References

1. Dahiphale, D., Madiraju, N., Lin, J., Karve, R., Agrawal, M., Modwal, A., ... & Merchant, A. (2024, December). Enhancing Trust and Safety in Digital Payments: An LLM-Powered Approach. In 2024 IEEE International Conference on Big Data (BigData) (pp. 4854-4863). IEEE.
2. Ouyang, K., Ke, Z., Fu, S., Liu, L., Zhao, P., & Hu, D. (2024). Learn from global correlations: Enhancing evolutionary algorithm via spectral gnn. arXiv preprint arXiv:2412.17629.
3. Hacini, A. D., Benabdellouahad, M., Abassi, I., Houhou, S., Boulmerka, A., & Farhi, N. (2025). LLM-Assisted Financial Fraud Detection with Reinforcement Learning. *Algorithms*, 18(12), 792.
4. Kanikanti, V. S. N., Mula, K., Muthukumarasamy, K., Kubam, C. S., Goswami, B., & Gadam, H. (2026). Streaming analytics pipelines for LLM-based financial anomaly detection in Real-Time retail transaction flows. *International Journal of Information Technology*, 1-6.
5. Luo, R., Wang, N., & Zhu, X. (2025). Fraud detection and risk assessment of online payment transactions on e-commerce platforms based on llm and gcn frameworks. arXiv preprint arXiv:2509.09928.
6. Mingxiu Sui, Yiyun Su, Jiaqing Shen, et al. Intelligent Anti-Money Laundering on Cryptocurrency: A CNN-GNN Fusion Approach. Authorea. January 12, 2026, DOI:10.22541/au.176824645.56752786/v1.
7. Bisht, K. S. (2025). Conversational Finance: LLM-Powered Payment Assistant Architecture. *European Journal of Computer Science and Information Technology*, 13(27), 116-130.
8. Chen, Y., Liu, L., & Fang, L. (2024). An Enhanced Credit Risk Evaluation by Incorporating Related Party Transaction in Blockchain Firms of China. *Mathematics*, 12(17), 2673.
9. Lizi Chen, Yue Zou, Pengfei Pan, et al. Cascading Credit Risk Assessment in Multiplex Supply Chain Networks. Authorea. January 16, 2026, DOI: 10.22541/au.176858311.10362606/v1.
10. Ke, Z., Cao, Y., Chen, Z., Yin, Y., He, S., & Cheng, Y. (2025). Early warning of cryptocurrency reversal risks via multi-source data. *Finance Research Letters*, 107890.
11. Malingu, C. J., Kabwama, C. A., Businge, P., Agaba, I. A., Ankunda, I. A., Mugalu, B., ... & Musunguzi, D. (2025). Application of LLMs to Fraud Detection. *World J. Adv. Res. Rev*, 26, 178-183.
12. O'Neill, O., Ramanayake, R., Mandal, A., Pawar, U., Flanagan, W., Chatbri, H., & Martin, C. A Practical Taxonomy for Finance-Specific LLM Risk Detection and Monitoring. In *NeurIPS 2025 Workshop: Generative AI in Finance*.
13. Qin Wang, Bolin Huang, and Qianying Liu. 2026. Deep Learning-Based Design Framework for Circular Economy Supply Chain Networks: A Sustainability Perspective. In *Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (DECS '25)*. Association for Computing Machinery, New York, NY, USA, 836–840. <https://doi.org/10.1145/3785706.3785837>.
14. Artola Velasco, A., Tsirtsis, E., Okati, N., & Gomez Rodriguez, M. (2025). Is Your LLM Overcharging You? Tokenization, Transparency, and Incentives. arXiv preprint arXiv:2505.21627.
15. Trozelli, P., & Andersson Holm, A. (2024). Comparing GPT and Traditional Machine Learning in Fraud Detection.
16. Haoran Zheng, Yuqing Lin, Qi He, et al. Blockchain Payment Fraud Detection with a Hybrid CNN-GNN-LSTM Model. Authorea. February 19, 2026, DOI: 10.22541/au.177153215.51813823/v1
17. Bhatt, S., & Garg, G. (2025). NLP for Fraud Detection and Security in Financial Documents. In *Transformative Natural Language Processing: Bridging Ambiguity in Healthcare, Legal, and Financial Applications* (pp. 131-155). Cham: Springer Nature Switzerland.
18. Keyu Yuan, Yuqing Lin, Wenjun Wu, et al. Detection of Blockchain Online Payment Fraud Via CNN-LSTM. Authorea. January 15, 2026, doi:10.22541/au.176851576.63306241/v1.
19. Rollinson, N., & Polatidis, N. (2026). LLM-Generated Samples for Android Malware Detection. *Digital*, 6(1), 5.

20. Li, Z., & Ke, Z. (2025). RepoLLM: A multi-modal foundation model for drug repurposing via alignment of molecules, EHRs, and knowledge graphs. ICML 2025 Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences. <https://openreview.net/forum?id=adweKg5NZa>
21. Yao, Z., & Zhu, Q. (2026). A Two-Stage Wiener Process-Based Method for Bearing Remaining Useful Life Prediction Considering Measurement Uncertainty. *Measurement Science and Technology*. DOI: 10.1088/1361-6501/ae412c
22. Li, Long, Jiaran Hao, Jason Klein Liu, Zhijian Zhou, Yanting Miao, Wei Pang, Xiaoyu Tan et al. "The choice of divergence: A neglected key to mitigating diversity collapse in reinforcement learning with verifiable reward." arXiv preprint arXiv:2509.07430 (2025).
23. Wang, Chengkai, Di Wu, Yunsheng Liao, Wenyao Zheng, Ziyi Zeng, Xurong Gao, Hemmings Wu et al. "NeuroCLIP: A Multimodal Contrastive Learning Method for rTMS-treated Methamphetamine Addiction Analysis." arXiv preprint arXiv:2507.20189 (2025).
24. Liu, S., & Zhu, M. (2024). In-trajectory inverse reinforcement learning: Learn incrementally before an ongoing trajectory terminates. *Advances in Neural Information Processing Systems*, 37, 117164-117209.
25. Shi, Ge, Lin Sun, Quan An, Lei Tang, Jiantao Shi, Chuang Chen, Lihang Feng, and Hongyang Ma. "Quantifying Urban Park Cooling Effects and Tri-Factor Synergistic Mechanisms: A Case Study of Nanjing's Central Districts." *Systems* 14, no. 2 (2026): 130.
26. Wei, D., Wang, Z., Kang, H., Sha, X., Xie, Y., Dai, A., & Ouyang, K. (2025). A comprehensive analysis of digital inclusive finance's influence on high quality enterprise development through fixed effects and deep learning frameworks. *Scientific Reports*, 15(1), 30095.
27. Li, F., Xu, Q., Bao, S., Han, B., Yang, Z., & Huang, Q. (2025). Hybrid generative fusion for efficient and privacy-preserving face recognition dataset generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 495-501).
28. Wu, S., & Zhang, J. (2025). Spatiotemporal multi-view continual dictionary learning with graph diffusion. *Knowledge-Based Systems*, 316, 113388.
29. Min, M., Duan, J., Liu, M., Wang, N., Zhao, P., Zhang, H., & Han, Z. (2026). Task Offloading with Differential Privacy in Multi-Access Edge Computing: An A3C-Based Approach. *IEEE Transactions on Cognitive Communications and Networking*.
30. Liu, S., & Zhu, M. (2023). Learning multi-agent behaviors from distributed and streaming demonstrations. *Advances in Neural Information Processing Systems*, 36, 53552-53564.
31. Li, F., Xu, Q., Bao, S., Yang, Z., Cao, X., & Huang, Q. (2025). One image is worth a thousand words: A usability preservable text-image collaborative erasing framework. arXiv preprint arXiv:2505.11131.
32. Wang, Zhi, Tao Gong, Yingpu Nian, Bo Yi, Xingwei Wang, Xinhao Zhou, Jianhui Lv, Geyong Min, and Keqin Li. "SFTRAP: Satisfying Fidelity Threshold Routing and Adaptive Purification for Throughput Maximum in Quantum Network." *IEEE Transactions on Communications* 74 (2026): 3952-3967.
33. Guan, Z., Cao, H., Zhong, M., Yang, E., Ai, L., Ni, Y., & Shi, B. (2026). Symphony-Coord: Emergent Coordination in Decentralized Agent Systems. arXiv preprint arXiv:2602.00966.
34. Tian, J., & Wang, Z. (2025). DLRREC: Denoising Latent Representations via Multi-Modal Knowledge Fusion in Deep Recommender Systems. arXiv preprint arXiv:2512.00596.
35. Ning, Qirui, Jinlai Zhang, Yuhang Xie, Kaifeng Liu, Kai Gao, Bin Chen, Gengbiao Chen, Qing Fan, Hui Liu, and Ronghua Du. "Multi-Resolution Context Augmentation and Dual Channel Attention for 3D lane detection." *IEEE Internet of Things Journal* (2025).
36. Wang, J., Lin, C., Nie, L., Liao, K., Shao, S., & Zhao, Y. (2024, October). Digging into contrastive learning for robust depth estimation with diffusion models. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 4129-4137).
37. Wang, Chengkai, Yifan Zhang, Chengyu Wu, Jun Liu, Xingliang Huang, Liuxi Wu, Yitong Wang, Xiang Feng, Yiting Lu, and Yaqi Wang. "MMDental-A multimodal dataset of tooth CBCT images with expert medical records." *Scientific Data* 12, no. 1 (2025): 1172.
38. Wang, Ke, Zishuo Zhao, Xinyuan Song, Bill Shi, Libin Xia, Chris Tong, Lynn Ai, Felix Qu, and Eric Yang. "VeriLLM: A Lightweight Framework for Publicly Verifiable Decentralized Inference." arXiv preprint arXiv:2509.24257 (2025).

39. Li, M., Zhang, D., Dong, Q., Xie, X., & Qin, K. (2025, April). Adaptive dataset quantization. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 11, pp. 12093-12101).
40. Yu, Q., Ke, Z., Xiong, G., Cheng, Y., & Guo, X. (2024, December). Identifying money laundering risks in digital asset transactions based on ai algorithms. In 2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC) (pp. 1081-1085). IEEE.
41. Zheng, Y., Zhong, B., Liang, Q., Li, N., & Song, S. (2025, April). Decoupled spatio-temporal consistency learning for self-supervised tracking. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 10, pp. 10635-10643).
42. Li, T., & Sun, W. (2024). MLP-SLAM: Multilayer Perceptron-Based Simultaneous Localization and Mapping. arXiv preprint arXiv:2410.10669.
43. Shao, Z., Zhou, Y., Li, F., Zhu, H., & Liu, B. (2024). Joint facial action unit recognition and self-supervised optical flow estimation. *Pattern Recognition Letters*, 181, 70-76.
44. Peng, Yong, Shaowei Gu, Guohua Wu, Yunbin Liang, Kaichen Ouyang, Xifeng Liang, Kui Wang, and Chaojie Fan. "A novel plug-and-play meta-black-box optimization module based on video streams for non-contact physiological signal extraction." *Swarm and Evolutionary Computation* 102 (2026): 102336.
45. Zhang, Ye, Qi Chen, Longsen Gao, Rui Liu, Linyue Chu, Kangtong Mo, Zhengjian Kang, Wenyong Huang, and Xingyu Zhang. "Invertible liquid neural network-based learning of inverse kinematics and dynamics for robotic manipulators." *Scientific Reports* 15, no. 1 (2025): 42311.
46. Liu, S., Xu, S., Qiu, W., Zhang, H., & Zhu, M. (2025). Explainable reinforcement learning from human feedback to improve alignment. arXiv preprint arXiv:2512.13837.
47. Wang, Jiyuan, Chunyu Lin, Cheng Guan, Lang Nie, Jing He, Haodong Li, Kang Liao, and Yao Zhao. "Jasmine: Harnessing diffusion prior for self-supervised depth estimation." arXiv preprint arXiv:2503.15905 (2025).
48. Shao, Z., Li, F., Zhou, Y., Chen, H., Zhu, H., & Yao, R. (2023). Identity-invariant representation and transformer-style relation for micro-expression recognition. *Applied Intelligence*, 53(17), 19860-19871.
49. Feng, Tongtong, Xin Wang, Zekai Zhou, Ren Wang, Yuwei Zhan, Guangyao Li, Qing Li, and Wenwu Zhu. "Evoagent: Agent autonomous evolution with continual world model for long-horizon tasks." arXiv e-prints (2025): arXiv-2502.
50. Tian, Aaron Xuxiang, Ruofan Zhang, Jiayao Tang, Young Min Cho, Xueqian Li, Qiang Yi, Ji Wang et al. "Beyond the Strongest LLM: Multi-Turn Multi-Agent Orchestration vs. Single LLMs on Benchmarks." arXiv preprint arXiv:2509.23537 (2025).
51. Jiantong Jiang, Peiyu Yang, Rui Zhang, et al. Towards Efficient Large Language Model Serving: A Survey on System-Aware KV Cache Optimization. TechRxiv. November 06, 2025. DOI: 10.36227/techrxiv.176046306.66521015/v2
52. Wu, Jiachun, Jinlai Zhang, Jihong Zhu, Yijian Duan, Youyang Fang, Jingyu Zhu, Lairong Yin et al. "Multi-scale convolution and dynamic task interaction detection head for efficient lightweight plum detection." *Food and Bioprocess Processing* 149 (2025): 353-367.
53. Li, M., Gou, H., Zhang, D., Liang, S., Xie, X., Ouyang, D., & Qin, K. (2025). Beyond Random: Automatic Inner-loop Optimization in Dataset Distillation. arXiv preprint arXiv:2510.04838.
54. Pan, P., Chen, L., He, Q., Yuan, K., Wang, H., & Zhang, W. (2026). FinSCRA: An LLM-Powered Multi-Chain Reasoning Framework for Interpretable Node Classification on Text-Attributed Graphs. Preprints. <https://doi.org/10.20944/preprints202602.0209.v1>
55. Zhang, Y., Liu, R., Chen, Q., Mo, K., Chu, L., & Kang, Z. (2025). SDARL: Safe Deep Adaptive Representation Learning for High-DoF Non-Linear System Manipulation in Space. IEEE Access.
56. Ke, Zong, Jiaqing Shen, Xuanyi Zhao, Xinghao Fu, Yang Wang, Zichao Li, Lingjie Liu, and Huailing Mu. "A stable technical feature with GRU-CNN-GA fusion." *Applied Soft Computing* (2025): 114302.
57. Wang, Q., Tsai, W. T., Shi, T., Tang, W., & Du, B. (2025). Hide and seek in transaction networks: a multi-agent framework for simulating and detecting money laundering activities. *Complex & Intelligent Systems*, 11(6), 271.

58. Xu, D. M., Hu, X. X., Wang, W. C., Wang, J., Shi, C. C., & Zang, H. F. (2025). Innovative daily runoff prediction model integrating black-winged kite algorithm and Mamba2–Transformer architecture. *Ecological Informatics*, 103565.
59. Qu, Y., Panariello, M., Todisco, M., & Evans, N. (2025, October). Reference-free adversarial sex obfuscation in speech. In *2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)* (pp. 2128-2133). IEEE.
60. Liu, S., & Zhu, M. (2022). Distributed inverse constrained reinforcement learning for multi-agent systems. *Advances in Neural Information Processing Systems*, 35, 33444-33456.
61. Wang, T. L., Gu, S. W., Liu, R. J., Chen, L. Q., Wang, Z., & Zeng, Z. Q. (2025). Cuckoo catfish optimizer: a new meta-heuristic optimization algorithm. *Artificial Intelligence Review*, 58(10), 326.
62. Liu, Z., Qian, S., Cao, S., & Shi, T. (2025). Mitigating age-related bias in large language models: Strategies for responsible artificial intelligence development. *INFORMS Journal on Computing*.
63. Qu, Y., Fu, D., & Fan, J. (2025). Subject Information Extraction for Novelty Detection with Domain Shifts. arXiv preprint arXiv:2504.21247.
64. Li, M., Zhang, D., He, T., Xie, X., Li, Y. F., & Qin, K. (2024, October). Towards effective data-free knowledge distillation via diverse diffusion augmentation. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 4416-4425).
65. Wang, Junqiao, Zeng Zhang, Yangfan He, Zihao Zhang, Xinyuan Song, Yuyang Song, Tianyu Shi et al. "Enhancing code llms with reinforcement learning in code generation: A survey." arXiv preprint arXiv:2412.20367 (2024).
66. Liu, S., & Zhu, M. (2023). Meta inverse constrained reinforcement learning: Convergence guarantee and generalization analysis. In *The Twelfth International Conference on Learning Representations*.
67. P. Zhang et al., "Boosting Learning Efficiency in Few-Shot Tasks With Layer-Adaptive PID Control," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2026.3663608.
68. Zheng, Y., Zhong, B., Liang, Q., Mo, Z., Zhang, S., & Li, X. (2024, March). Odtrack: Online dense temporal token learning for visual tracking. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 38, No. 7, pp. 7588-7596).
69. Tian, J., Wang, Z., Zhao, J., & Ding, Z. (2024, November). Mmrec: Llm based multi-modal recommender system. In *2024 19th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)* (pp. 105-110). IEEE.
70. Wang, JiYuan, Chunyu Lin, Lei Sun, Rongying Liu, Lang Nie, Mingxing Li, Kang Liao, Xiangxiang Chu, and Yao Zhao. "From editor to dense geometry estimator." arXiv preprint arXiv:2509.04338 (2025).
71. Yi, Qiang, Yangfan He, Jianhui Wang, Xinyuan Song, Shiyao Qian, Xinhang Yuan, Yi Xin et al. "Score: Story coherence and retrieval enhancement for ai narratives." arXiv preprint arXiv:2503.23512 (2025).
72. Zhao, Peng, Xiaoyu Liu, Xuqi Su, Di Wu, Zi Li, Kai Kang, Keqin Li, and Armando Zhu. "Probabilistic Contingent Planning Based on Hierarchical Task Network for High-Quality Plans." *Algorithms* 18, no. 4 (2025): 214.
73. Tang, J., Zhang, W., Liu, H., Yang, M., Jiang, B., Hu, G., & Bai, X. (2022). Few could be better than all: Feature sampling and grouping for scene text detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4563-4572).
74. Zhuang, J., Jin, H., Zhang, Y., Kang, Z., Zhang, W., Dagher, G. G., & Wang, H. (2025). Exploring the vulnerability of the content moderation guardrail in large language models via intent manipulation. arXiv preprint arXiv:2505.18556, 1.
75. Kang, Zhaolu, Junhao Gong, Jiaxu Yan, Wanke Xia, Yian Wang, Ziwen Wang, Huaxuan Ding et al. "Hssbench: Benchmarking humanities and social sciences ability for multimodal large language models." arXiv preprint arXiv:2506.03922 (2025).
76. Liu, S., & Zhu, M. (2025). Utility: Utilizing explainable reinforcement learning to improve reinforcement learning. In *The Thirteenth International Conference on Learning Representations*.
77. Wan, W., Zhou, F., Liu, L., Fang, L., & Chen, X. (2021). Ownership structure and R&D: The role of regional governance environment. *International Review of Economics & Finance*, 72, 45-58.

78. Peng, Yong, Shaowei Gu, Yunbin Liang, Kaichen Ouyang, Yingli Li, Kui Wang, Guohua Wu, and Chaojie Fan. "Wave Optics Optimizer: A novel meta-heuristic algorithm for engineering optimization." *Communications in Nonlinear Science and Numerical Simulation* (2025): 109337.
79. Lu, Yao, Wen Yang, Yunzhe Zhang, Zuohui Chen, Jinyin Chen, Qi Xuan, Zhen Wang, and Xiaoni Yang. "Understanding the dynamics of dnns using graph modularity." In *European Conference on Computer Vision*, pp. 225-242. Cham: Springer Nature Switzerland, 2022.
80. Zheng, Y., Zhong, B., Liang, Q., Li, G., Ji, R., & Li, X. (2023). Toward unified token learning for vision-language tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4), 2125-2135.
81. Xiong, Jing, Zixuan Li, Chuanyang Zheng, Zhijiang Guo, Yichun Yin, Enze Xie, Zhicheng Yang et al. "Dq-lore: Dual queries with low rank approximation re-ranking for in-context learning." *arXiv preprint arXiv:2310.02954* (2023).
82. Xiao, C., Hou, L., Fu, L., & Chen, W. (2025, May). Diffusion-based self-supervised imitation learning from imperfect visual servoing demonstrations for robotic glass installation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 10401-10407). IEEE.
83. Shi, Chuancheng, Shangze Li, Shiming Guo, Simiao Xie, Wenhua Wu, Jingtong Dou, Chao Wu et al. "Where Culture Fades: Revealing the Cultural Gap in Text-to-Image Generation." *arXiv preprint arXiv:2511.17282* (2025).
84. Chen, X., Xiao, C., Cao, W., Zhang, W., & Liu, Y. (2025). Framework and pathway for the construction of a unified data-element market in china. *Strategic Study of Chinese Academy of Engineering*, 27(1), 40-50.
85. Yao, J., Li, C., & Xiao, C. (2024). Swift sampler: Efficient learning of sampler by 10 parameters. *Advances in Neural Information Processing Systems*, 37, 59030-59053.
86. Shixiong Xu, Naxi Chen, Pengfei Pan, et al. *Financial Anomaly Transaction Detection Using Autoencoder-Based Models*. Authorea. February 23, 2026. DOI: 10.22541/au.177187997.78014742/v1

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.