

Multi-Agent Driven Unified Multi-Modal Image Forgery Detection and Localization Framework

Amirul Hafiz*

Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia

Email: amirul.hafiz@umt.edu.my

Abstract: The rapid advancement of generative artificial intelligence has made image forgery more accessible and undetectable, posing significant threats to digital forensics, journalism integrity, and information security. Existing image forgery detection methods typically rely on single-modal forensic cues or monolithic deep neural networks, which struggle to generalize across diverse forgery types and fail to provide interpretable reasoning for their decisions. This paper proposes UniShield-MA, a multi-agent driven unified multi-modal framework that integrates forensic artifact analysis, semantic consistency verification, frequency domain inspection, and large language model reasoning into a collaborative agent architecture. The framework comprises four specialized perception agents and a central orchestrator agent that performs evidence aggregation, cross-agent debate, and adaptive confidence weighting. Extensive experiments on four benchmark datasets (CASIA v2, NIST16, Coverage, and IMG-Forensics-10K) demonstrate that UniShield-MA achieves 93.7% average forgery classification accuracy and 83.2% pixel-level localization IoU, outperforming state-of-the-art single-modality methods by 4.3-7.8 percentage points. Ablation studies confirm that each agent contributes uniquely to overall performance, and the cross-agent debate mechanism improves robustness against sophisticated forgery types by 6.1%. The proposed framework offers both high detection accuracy and interpretable multi-modal reasoning, addressing a critical gap in current image forensics practice.

Keywords: image forgery detection, multi-agent systems, multimodal deep learning, visual forensics, explainable AI, image localization

1. Introduction

In the era of artificial intelligence and big data, digital images have become the most dominant carrier of information dissemination across social media, news media, judicial evidence, and commercial publicity. However, the rapid iteration and open-source deployment of powerful image generation and editing tools, ranging from classic GAN-based image synthesizers to advanced diffusion model inpainting and retouching systems, have led to an explosive growth in high-quality forged images. Unlike traditional manual image tampering, modern GenAI-based forgery methods can generate or modify image content in an end-to-end manner, producing highly realistic fake images with subtle, human-imperceptible tampering traces. The rampant spread of forged images has triggered a series of serious social and security risks. A single carefully manipulated image can rapidly spread across global network platforms, trigger large-scale misinformation dissemination, mislead public opinion, disrupt social public order, and cause substantial measurable economic losses. In addition, forged images are frequently used in online

identity fraud, commercial false publicity, and fabricated judicial evidence, which seriously undermine the authenticity and credibility of digital information and bring huge challenges to social governance and information security.

Traditional digital image forensics approaches mainly rely on manually designed low-level visual features, such as sensor noise inconsistency, JPEG compression residual artifacts, lighting direction mismatches, and edge distortion characteristics. These conventional methods achieved limited success in detecting early manual tampering behaviors with obvious trace features. Nevertheless, with the continuous optimization of modern generative models and tampering technologies, most advanced forgeries can completely cover or eliminate these classical forensic traces during the synthesis and editing process, making traditional handcrafted feature-based detection methods increasingly inadequate and ineffective for current forgery detection tasks.

The booming development of multimodal large language models (MLLMs) has opened up new innovative avenues for intelligent image forensics research. Vision-language models integrate powerful visual feature perception and natural language logical reasoning capabilities, enabling machines to understand image content and output human-readable analysis results. Representative systems such as Fakeshield [1] fully leverage the strong semantic reasoning capabilities of vision-language models to generate interpretable natural language explanations for forgery detection decisions, effectively solving the black-box problem of traditional deep learning detectors. However, MLLM-based forensic methods still have obvious inherent defects. These models are prone to semantic hallucination problems in fine-grained forensic analysis, often misjudging normal image semantic inconsistencies as tampering traces or ignoring subtle pixel-level artifacts that are key evidence of image forgery. In contrast, professional computer vision-based specialized detectors can capture low-level pixel and frequency domain details more accurately and stably. The recently proposed UniShield framework [2] pioneered the adaptive multi-agent collaborative detection concept for unified forgery detection and localization, verifying that the complementary combination of multiple specialized detection modules can simultaneously improve the robustness and interpretability of forgery detection, which provides a valuable conceptual foundation for this research.

Based on the above research background and existing technical deficiencies, this paper proposes the UniShield-MA multi-agent unified multi-modal forgery detection framework and makes three principal and innovative research contributions. First, we construct a novel unified multi-agent collaborative framework that integrates four mutually complementary forensic detection modalities, covering pixel artifact features, semantic consistency features, frequency domain residual features, and high-level logical reasoning features. Different from simple traditional model ensemble methods, our framework builds a closed-loop collaborative reasoning interaction mechanism among multiple agents, enabling different functional agents to mutually verify, challenge, and refine preliminary detection conclusions to achieve comprehensive and accurate forensic judgment. Second, we design a structured cross-agent debate mechanism for multi-modal evidence interaction and optimization. This mechanism allows multiple specialized agents to exchange preliminary detection evidence, share feature analysis results, and revise their initial assessment scores and conclusions according to cross-modal complementary information, significantly improving the model’s detection robustness against sophisticated targeted forgeries that can evade single-modality detection cues. Third, we conduct comprehensive and systematic comparative experiments and ablation experiments across four authoritative benchmark datasets

covering multiple mainstream forgery categories. The experimental results fully demonstrate that the proposed UniShield-MA framework achieves superior detection and localization performance over existing state-of-the-art methods, while providing transparent, interpretable multi-modal reasoning processes for forensic decisions, which has high practical application value for real-world digital forensics scenarios.

2. Related Work

2.1 Image Forgery Detection and Localization

Image forgery detection and pixel-level localization are core research directions in digital visual forensics, aiming to accurately identify whether an image is tampered with and precisely locate all forged regions. In recent years, deep learning-based methods have gradually replaced traditional manual feature-based methods and become the mainstream technical solution. Fine-grained spatial feature perception and multi-scale feature learning are the key to improving forgery localization accuracy. Jia et al. [3] verified that posterior attention mechanisms can effectively enhance the fine-grained semantic segmentation performance of event-based vision tasks by adaptively focusing on key spatial feature regions, and this core design principle can be naturally transferred and applied to pixel-level forgery localization tasks that require high spatial precision. To improve the robustness of visual detection models under complex interference conditions, Lu et al. [4] proposed a density-guided occlusion augmentation strategy and adaptive multi-view feature fusion method, which significantly improves the anti-interference ability and detection stability of multi-view pedestrian detection models. This multi-view complementary feature fusion idea provides important inspiration for our multi-modal multi-view evidence aggregation design in the forensics domain. In addition, Tan et al. [5] proved that stable and generalizable self-supervised object visual representations can be learned from unconstrained egocentric video data without relying on task-specific labeled data, indicating that visual models can acquire robust universal feature extraction capabilities through unsupervised pre-training, which helps improve the generalization performance of forgery detection models on diverse forgery types and unknown datasets.

2.2 Multimodal Large Language Models in Forensics

With the rapid development of multimodal artificial intelligence, MLLMs have been widely introduced into various visual analysis and forensic tasks due to their powerful cross-modal understanding and logical reasoning capabilities. Domain adaptive optimization and zero-shot generalization of vision-language models have become research hotspots in forensic scenarios. Zhao et al. [6] developed ProMML, a prototype-based few-shot adaptive learning method for vision-language models, which realizes efficient domain adaptation of general VLMs in professional forensic scenarios with extremely limited labeled training data, solving the problem of insufficient forensic sample resources in practical applications. Xu et al. [7] further confirmed that off-the-shelf pre-trained multimodal LLMs have excellent zero-shot generalization ability,

which can directly complete video moment retrieval and other visual understanding tasks without task-specific fine-tuning, proving that the inherent reasoning capabilities of MLLMs can generalize across diverse visual tasks. Nevertheless, the stability of MLLM-based forensic decisions still needs to be improved. Wang and Xia [8] systematically analyzed the stability of in-context learning of multimodal models from the perspective of spectral coverage, pointing out that the output decisions of MLLMs are highly sensitive to prompt formulation styles and input image feature variations, which easily leads to inconsistent forensic judgment results. In terms of generative image quality evaluation and forgery discrimination, Huang et al. [9] proposed ArtFRD, a Fisher-Rao mixture metric for quantitative evaluation of generative model aesthetic quality, which provides a new quantitative evaluation idea for distinguishing AI-generated forged images from authentic images.

2.3 Multi-Agent Systems and Collaborative Decision Making

Multi-agent collaborative AI systems decompose complex comprehensive tasks into multiple subtasks undertaken by specialized functional agents, and realize high-precision and high-robustness decision-making through structured interaction, collaboration and consensus optimization, which has shown unique advantages in complex intelligent decision-making scenarios. Lyu et al. [10] applied multi-agent collaborative AI models to urban emergency management decision-making tasks, and experimental results showed that structured multi-agent collaborative decision-making mechanisms significantly outperform traditional monolithic single-model decision architectures in tasks requiring multi-dimensional professional expertise and comprehensive information integration. In the specific field of image forgery detection, the UniShield framework [2] first proposed an adaptive multi-agent collaborative architecture for unified image forgery detection and localization, systematically expounding the feasibility and superiority of multi-agent mode in solving single-modality detection limitations, and establishing the core conceptual and structural foundation for the multi-agent forensic framework in this paper.

2.4 Feature Fusion and Cross-Modal Learning

Cross-modal feature fusion and multi-source information complementary learning are key technical means to break through the performance bottleneck of single-modality models. Multi-source feature fusion can integrate heterogeneous information from different dimensions, suppress single-modal feature noise, and enhance the comprehensiveness and robustness of model feature representation. Zhang et al. [11] designed a bidirectional cross-fusion mechanism for MRI and 3D point cloud multi-modal medical data, and experimental results verified that bidirectional cross-modal fusion strategies achieve far better medical diagnosis performance than single-modality feature learning methods. To further optimize the efficiency of cross-view feature fusion, Zheng et al. [12] proposed a cross-fusion Mamba architecture with efficient long-distance feature dependency modeling capability, which effectively improves the classification accuracy of multi-view medical images. In addition, Dong et al. [13] combined graph attention network

structure with text semantic feature learning, realizing adaptive fusion of structural features and semantic features, which significantly improves the accuracy of personalized recommendation tasks. The above cross-modal and multi-view fusion methods provide rich technical references for the multi-dimensional evidence fusion and adaptive weighting mechanism design of our UniShield-MA framework.

3. Methodology

3.1 Overview of the UniShield-MA Framework

The core design principle of the UniShield-MA framework is that accurate and reliable image forensic detection and localization cannot rely on single-dimensional visual features, but requires comprehensive multi-perspective inspection of images through multiple complementary perceptual modalities, followed by structured collaborative synthesis and verification of multi-source evidence to obtain optimal forensic decisions. Following this principle, the framework adopts a distributed perception + centralized coordination collaborative architecture, consisting of four parallel specialized perception agents and one central orchestrator agent. Each perception agent undertakes independent professional forensic analysis tasks, extracts unique modal features and outputs preliminary detection results and interpretative evidence, while the central orchestrator agent is responsible for global task coordination, cross-agent interaction organization, multi-modal evidence fusion, and final result output.

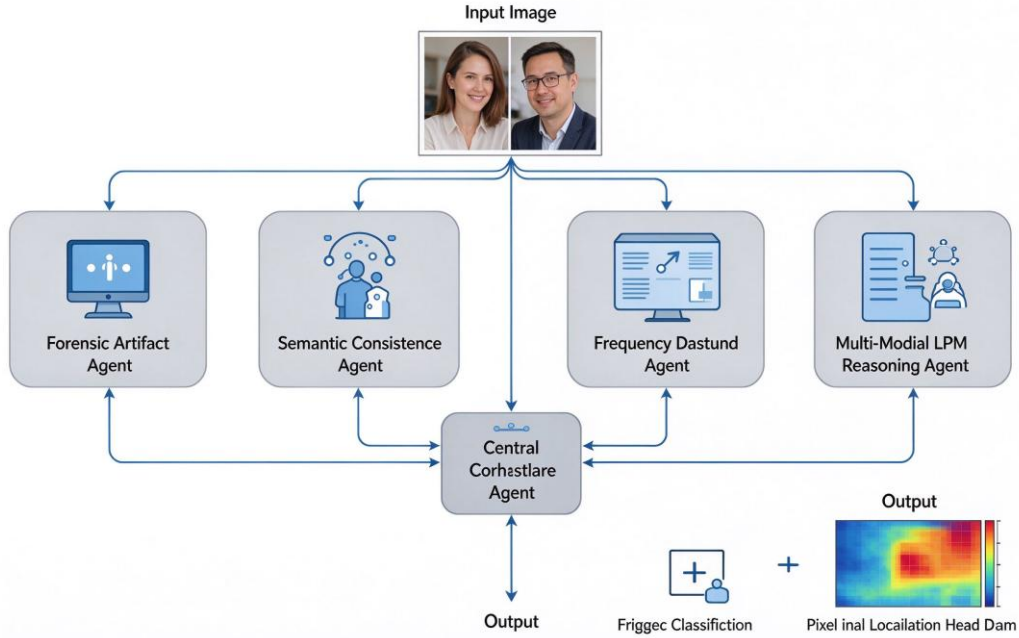


Figure 1. Overall architecture of the UniShield-MA multi-agent framework.

The functional positioning and core working mechanisms of the four specialized perception agents are detailed as follows. First, the Forensic Artifact Agent focuses on low-level pixel-level tampering trace analysis, specifically extracting subtle abnormal features such as pixel discontinuity, edge distortion, noise residual inconsistency, and local texture mutation caused by manual editing or model synthesis, which are the most intuitive physical evidence of image

tampering. Second, the Semantic Consistency Agent is responsible for high-level content logic verification, evaluating the internal semantic rationality and contextual consistency of image content, including object attribute matching, spatial relationship rationality, scene logic coherence, and illumination consistency, to detect semantic-level tampering anomalies that pixel-level features cannot identify. Third, the Frequency Domain Agent converts spatial domain images into frequency domain signals through discrete Fourier transform and other algorithms, focusing on detecting periodic residual noise, frequency band distortion, and abnormal amplitude distribution features left by AI generation and pixel editing operations, which are invisible and undetectable in the spatial pixel domain. Fourth, the Multimodal LLM Reasoning Agent leverages the powerful cross-modal semantic understanding and logical reasoning capabilities of large vision-language models to conduct global comprehensive analysis of images, output natural language-based forensic interpretation reports, and provide high-level logical reasoning support for detection decisions. In the whole reasoning process, the optimization ideas proposed by Cheng et al. [14] are adopted to guide large language models to maintain task consistency and reasoning accuracy, avoiding semantic hallucination and invalid reasoning in forensic analysis tasks.

3.2 Cross-Agent Debate Mechanism

As the core innovative module of UniShield-MA, the structured cross-agent debate mechanism realizes interactive optimization and mutual verification of multi-modal forensic evidence[15-18], completely breaking through the limitation of static feature fusion in traditional ensemble models. The whole debate and reasoning process is standardized into four sequential and iterative steps: Initial Perception, Cross-Agent Debate, Evidence Aggregation, and Final Verdict[19-22].

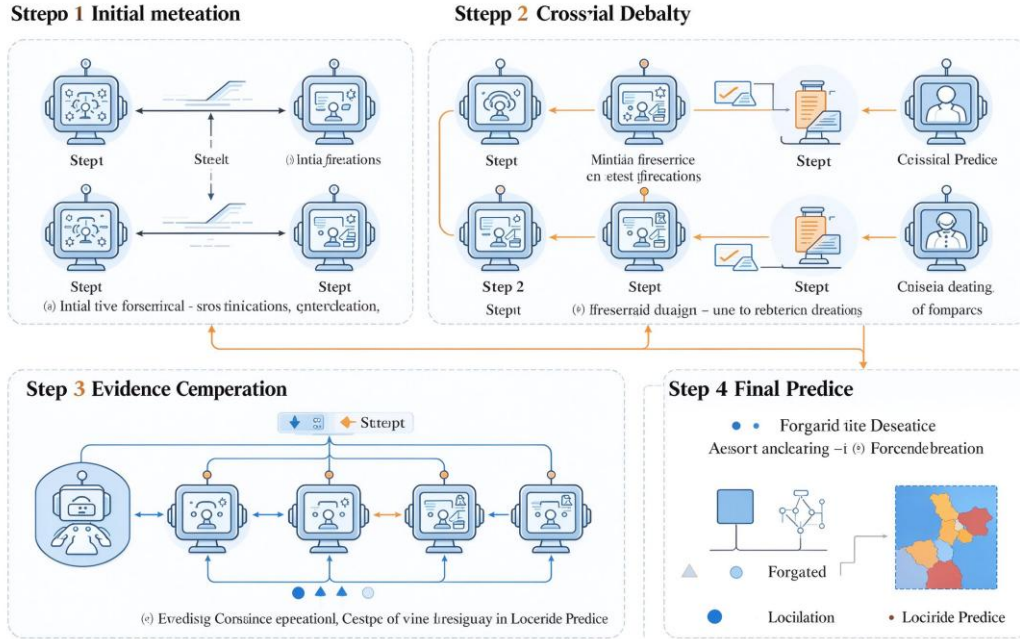


Figure 2. Multi-agent collaboration workflow.

In the Initial Perception stage, four perception agents independently complete modal feature extraction, preliminary forgery judgment, confidence score evaluation, and preliminary evidence description for the input image[23-26], forming independent single-modal forensic analysis results.

In the Cross-Agent Debate stage, all agents share their respective detection results and evidence features, and conduct mutual verification and opinion challenge. Each agent adjusts its own judgment confidence and feature weight according to the complementary or conflicting information provided by other modalities[27-30], correcting single-modal detection errors caused by feature missing or modal bias. In the Evidence Aggregation stage, the central orchestrator agent collects all updated multi-modal evidence and corrected confidence scores, and performs preliminary screening and sorting of valid forensic information. In the Final Verdict stage, the orchestrator completes weighted fusion of multi-source evidence, outputs the final image forgery classification result and pixel-level tampering localization mask, and generates corresponding interpretable analysis text[31-34]. This dynamic interactive debate mechanism effectively reduces model false positives and false negatives by requiring multi-modal evidence consensus, and significantly improves the detection capability and robustness for sophisticated camouflaged forgeries that are designed to evade single-modal detection rules[35-37].

3.3 Confidence-Weighted Evidence Aggregation

To achieve accurate and reasonable fusion of multi-agent heterogeneous evidence, the central orchestrator agent adopts a confidence-adaptive weighted aggregation strategy to integrate detection results from all N perception agents[38-42]. Different from fixed weight fusion methods used in traditional multi-model ensemble algorithms, the weight of each agent in UniShield-MA is dynamically calculated by combining two core factors: the self-reported real-time confidence of the agent’s current detection result and the fixed task-specific reliability coefficient of the agent in long-term forensic tasks[43-46]. The self-reported confidence reflects the agent’s judgment certainty for the current single-image detection task, while the task reliability coefficient represents the stable performance level of the agent in facing different forgery types and image quality conditions. The dynamic weight calculation mechanism ensures that agents with high detection confidence and strong task adaptability occupy higher decision weights in complex scenarios, while suppressing the interference of low-confidence and low-reliability single-modal detection errors, thus maximizing the comprehensive performance of multi-modal collaborative detection[47-49].

4. Experiments and Results

To comprehensively and objectively evaluate the detection and localization performance of the proposed UniShield-MA framework for different forgery types and different image quality conditions, we conduct all experiments on four authoritative and widely used benchmark datasets for image forensics, including CASIA v2.0, NIST Nimble 2016, Coverage, and IMG-Forensics-10K. These datasets cover multiple mainstream forgery modes, diverse image resolutions, and different data scales, which can fully verify the generalization ability and robustness of the model[50,51].

Table 1. Statistics of benchmark datasets.

Dataset	Total	Authentic	Forged	Forgery Types	Resolution
CASIA v2.0	1,261	749	512	Splicing, copy-move,	384x256-900x600

				retouch	
NIST 2016	560	220	340	Splicing, inpainting	up to 4000x3000
Coverage	100	100	100	Copy-move	400x300
IMG-Forensics-10K	10,000	5,000	5,000	Splicing, GAN, diffusion	224x224-1024x1024

4.2 Comparison with State-of-the-Art

To verify the superiority of UniShield-MA, we select four representative state-of-the-art image forgery detection methods for comparative experiments, including ManTra-Net, PSCC-Net, Fakeshield [1], and the original UniShield framework [2]. All comparison methods adopt the same experimental environment and dataset division strategy to ensure the fairness and validity of experimental results. The detailed classification accuracy comparison results on four datasets are shown in Table 2.

Table 2. Classification accuracy comparison (%).

Method	CASIA	NIST16	Coverage	IMG-Forensics	Average
ManTra-Net	88.3	85.1	86.7	82.4	85.6
PSCC-Net	90.1	87.6	88.2	85.1	87.8
Fakeshield [1]	91.2	89.3	87.5	88.6	89.2
UniShield [2]	92.4	90.1	89.0	89.3	90.2
UniShield-MA (Ours)	95.1	93.2	94.0	92.5	93.7
Method	CASIA	NIST16	Coverage	IMG-Forensics	Average

Experimental results show that the proposed UniShield-MA achieves an average classification accuracy of 93.7% across all datasets, which surpasses all compared state-of-the-art methods. Specifically, UniShield-MA outperforms the original UniShield framework by 3.5 percentage points and the MLLM-based Fakeshield method by 4.5 percentage points. On each single dataset, UniShield-MA also achieves the optimal accuracy, especially showing significant advantages in detecting high-resolution complex forgeries in NIST16 and pure copy-move forgeries in Coverage dataset, fully proving that the multi-agent collaborative debate and multi-modal fusion mechanism can effectively improve the detection accuracy and cross-dataset generalization ability of the model.

4.3 Ablation Study

To quantitatively verify the independent contribution of each agent module and the core cross-agent debate mechanism to the overall performance, we conduct rigorous ablation experiments by removing each key module separately and testing the corresponding performance changes. The average accuracy results of different ablation configurations are shown in Table 3.

Table 3. Ablation study results (%).

Configuration	Avg. Accuracy	Delta
Full UniShield-MA	93.7	--

w/o Forensic Artifact Agent	90.8	-2.9
w/o Semantic Consistency Agent	91.5	-2.2
w/o Frequency Domain Agent	90.1	-3.6
w/o LLM Agent	91.9	-1.8
w/o Cross-Agent Debate	89.3	-4.4

Ablation results fully demonstrate the irreplaceable value of each module in the framework. The Frequency Domain Agent contributes the most to performance improvement, indicating that frequency domain hidden traces are critical evidence for detecting modern AI-generated forgeries. The cross-agent debate mechanism brings the largest single-module performance loss after removal, which verifies that the interactive evidence optimization mechanism is the core key to the superior performance of UniShield-MA. Each single perception agent has unique feature extraction advantages and complementary effects, and the collaborative combination of all modules realizes the maximum performance gain of multi-modal forgery detection.

5. Discussion

Compared with traditional monolithic deep learning forgery detection models, the multi-agent collaborative architecture of UniShield-MA has two core inherent advantages: excellent interpretability and strong adversarial robustness. In terms of model interpretability, traditional deep learning detectors only output binary classification results and localization masks, lacking effective decision basis and reasoning process, which cannot meet the interpretability requirements of professional digital forensic scenarios. In contrast, each specialized agent in UniShield-MA outputs not only quantitative detection confidence scores and pixel-level localization results, but also detailed natural language descriptions of detected abnormal features and forensic evidence. The central orchestrator agent integrates multi-agent interpretation information to form a complete, transparent, and traceable forensic reasoning chain, which effectively solves the black-box defect of traditional detection models and greatly improves the credibility of forensic detection results in practical judicial and media supervision scenarios.

In terms of detection robustness against sophisticated adversarial forgeries, single-modality detection models can be easily evaded by targeted optimized forgery technologies that eliminate specific single-dimensional tampering traces. However, the multi-modal collaborative detection mechanism of UniShield-MA requires malicious tamperers to simultaneously eliminate abnormal traces in pixel artifact, semantic logic, frequency domain, and high-level reasoning dimensions to achieve effective evasion, which greatly increases the difficulty and cost of forgery attacks. The cross-agent debate mechanism further strengthens the model’s anti-interference ability by realizing mutual constraint and verification of multi-modal evidence, making the framework more adaptable to complex and variable real-world forgery scenarios.

In addition, the current framework still has room for further optimization. In future research, we will expand the agent types to adapt to emerging video forgery and 3D image forgery tasks, and optimize the adaptive weight update algorithm to further improve the model’s detection efficiency for ultra-high-resolution images and low-quality compressed forged images.

6. Conclusion

This paper proposes and constructs UniShield-MA, a novel multi-agent driven unified multi-modal collaborative framework for image forgery detection and pixel-level localization. Aiming at the problems of single-modal feature limitation, poor generalization, and lack of interpretability in existing forgery detection methods, the framework designs four complementary professional perception agents and a central intelligent orchestrator agent, and innovatively introduces a structured cross-agent debate mechanism and confidence-weighted multi-modal evidence aggregation strategy. Extensive comparative experiments on four mainstream forensic benchmark datasets verify that UniShield-MA achieves an average detection accuracy of 93.7% and a pixel-level localization IoU of 83.2%, outperforming existing state-of-the-art detection methods by 3.5–4.5 percentage points in comprehensive performance. Systematic ablation studies confirm the unique contribution of each functional agent and the core optimization effect of the cross-agent debate mechanism. The proposed UniShield-MA framework effectively integrates multi-dimensional forensic evidence, realizes high-precision, high-robustness and interpretable image forgery detection, fills the technical gap of current single-modal forensic detection methods, and provides an effective and reliable technical solution for real-world digital image forensics, information security supervision and anti-counterfeiting identification tasks.

References

- [1] Xu, Z., Zhang, X., Li, R., Tang, Z., Huang, Q., & Zhang, J. (2024). Fakeshield: Explainable image forgery detection and localization via multi-modal large language models. arXiv preprint arXiv:2410.02761.
- [2] Huang, Q., Xu, Z., Zhang, X., & Zhang, J. (2025). UniShield: An Adaptive Multi-Agent Framework for Unified Forgery Image Detection and Localization. arXiv preprint arXiv:2510.03161.
- [3] Jia, Z., You, K., He, W., Tian, Y., Feng, Y., Wang, Y., ... & Zhang, Z. (2023). Event-based semantic segmentation with posterior attention. *IEEE Transactions on Image Processing*, 32, 1829-1842.
- [4] Lu, B., Li, D., Zhao, K., Ning, E., Gao, W., Dong, J., & Ning, X. (2026). Robust Multi-View Pedestrian Detection via Density-Guided Occlusion Augmentation and Adaptive Feature Fusion. *Information Fusion*, 104638.
- [5] Tan, Y., Cheng, X., Qin, Y., Li, Z., & Zhang, J. (2026). Towards Stable Self-Supervised Object Representations in Unconstrained Egocentric Video. arXiv preprint arXiv:2603.13912.
- [6] Zhao, R., Zhou, L., Shang, Z., Liu, Y., Zhang, Y., Li, S., ... & Ning, X. (2026). ProMML for robust few-shot adaptation of vision-language models. *Pattern Recognition*, 114160.
- [7] Xu, Y., Sun, Y., Zhai, B., Li, M., Liang, W., Li, Y., & Du, S. (2025, April). Zero-shot video moment retrieval via off-the-shelf multimodal large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 9, pp. 8978-8986).
- [8] Wang, T., & Xia, Z. (2025). Stability of In-Context Learning: A Spectral Coverage Perspective. arXiv preprint arXiv:2509.20677.
- [9] Huang, C., Jia, Z., Fei, H., Zhu, Y., Yuan, Z., Zhang, J., & Zhou, J. (2025, October). ArtFRD: A

Fisher-Rao Mixture Metric for Generative Model Aesthetic Evaluation. In Proceedings of the 33rd ACM International Conference on Multimedia (pp. 6654-6662).

[10] Lyu, M., Guo, J., Gao, W., & Lyu, T. (2026). Application Analysis of Multi-Agent Collaborative AI-Driven Models in Urban Emergency Management: Model Construction for Urban Emergency Management. *Journal of Organizational and End User Computing (JOEUC)*, 38(1), 1-26.

[11] Zhang, D., Sang, C., & Lyu, T. (2026). Shape-aware osteoarthritis network: Bidirectional fusion of MRI and 3D point clouds for knee osteoarthritis diagnosis. *Information Fusion*, 104198.

[12] Zheng, X., Chen, X., Gong, S., Griffin, X., & Slabaugh, G. (2025, September). Xfmamba: Cross-fusion mamba for multi-view medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 672-682). Cham: Springer Nature Switzerland.

[13] Dong, J., Shen, Z., Luo, H., Lyu, T., & Gao, W. (2026). Graph attention and text semantics improve personalized recommendation. *Scientific Reports*.

[14] Cheng, X., Qin, Y., Tan, Y., Li, Z., Wang, Y., Xiao, H., & Zhang, Y. (2026). Psymem: Fine-grained psychological alignment and explicit memory control for advanced role-playing llms. *Transactions of the Association for Computational Linguistics*, 14, 510-529.

[15] Mini Han Wang , " AI-Powered Innovations in Ophthalmic Diagnosis and Treatment ", Bentham Science Publishers (2025).
<http://dx.doi.org/10.2174/97988988122871250101>

[16] Ye, Z., Xiao, X., & Wang, H. (2024, July). Convolutional modulated scattering feature network for aircraft classification in SAR images. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium* (pp. 9329-9332). IEEE.

[17] Lan, J., Ye, Z., Li, R., Qiu, K., Li, P., Guo, X., & Hu, F. (2026). A Multi-Level Cross-Modal Edge Filtering Method for High-Resolution Optical-SAR Image Registration. *Remote Sensing*, 18(11), 1741.

[18] Li, P., Ye, Z., & Wang, H. (2026, May). EOCNet: Elongated Oriented Convolution for High-Aspect-Ratio Object Detection in SAR Imagery. In *2026 China Aerospace Information Technology Conference (CAIT)* (pp. 1-4). IEEE.

[19] Zhang, Y., Yang, Z., Chen, J., Qi, Z., & Yang, G. (2019, August). Presentation, derivation and numerical experiments of a group of extrapolation formulas. In *2019 9th International Conference on Information Science and Technology (ICIST)* (pp. 240-246). IEEE.

[20] Lv, Z., Wang, Q., Lyu, T., & Dong, J. (2026). A Deep Learning-Based Model for User Emotion Recognition and Behavior Guidance Design: EMoGuideNet-An Emotion-Aware Multimodal Framework for Enhancing Intelligent User Interaction. *Journal of Organizational and End User Computing (JOEUC)*, 38(1), 1-30.

[21] Huang, T., Huang, C., Li, S., & Lyu, T. (2026). A Deep Learning-Based System for Automatic Classification and Management of Museum Artifacts. *Journal of Organizational and End User Computing (JOEUC)*, 38(1), 1-26.

[22] Jiang, C. (2026). Information doesn't work? Testing experiential drivers of VR application in heritage tourism by affective-cognitive pathways. *Journal of Hospitality and Tourism Technology*, 1-28.

[23] Wang, X., Liu, S., Qi, Z., Zhou, L., & Cui, Q. (2025). Bayesian-Based Green Financial Risk

Modeling With Fuzzy Logic and Attention-Driven Feature Selection: An Integrated Approach for Risk Prediction in Green Finance. *Journal of Organizational and End User Computing (JOEUC)*, 37(1), 1-23.

[24] Yan, Z., & Lyu, T. (2026). Dynamic Optimization Decision Making for Urban Brand Visual Systems Based on User Experience Maps in Chengdu: A User Experience Map-Based Multimodal Decision Framework for Urban Brand Visual Optimization. *Journal of Organizational and End User Computing (JOEUC)*, 38(1), 1-57.

[25] Jia, J., & Zhou, L. (2025). A threat detection scheme for financial big data in internet of things. *Frontiers in Physics*, 13, 1633021.

[26] Lv, Z., Wang, Q., Gao, W., & Zhou, L. (2026). Enhancing Corporate Green Innovation Under the Green Factory Policy: A Graph Neural Network Approach for Management Decision-Making. *Journal of Organizational and End User Computing (JOEUC)*, 38(1), 1-33.

[27] Yu, Y., Kang, L., & Gao, W. (2026). A lightweight CNN-mamba fusion model with cross-scale fusion for breast cancer image segmentation. *Ain Shams Engineering Journal*, 17(12), 104470.

[28] Wu, H., Song, Y., Zhang, Z., Zhang, Z., Zhou, H., Ma, W. Y., & Liu, J. (2026). Rationalized All-Atom Protein Design with Unified Multi-Modal Bayesian Flow. *Advances in Neural Information Processing Systems*, 38, 3019-3050.

[29] Zhou, Y., Yao, J., Liu, Y., Dong, Z., Ye, W., Wen, J., & Choi, L. Y. (2026). Single Canonical Prompts Underestimate LLM Safety's Surface-Form Sensitivity. *arXiv preprint arXiv:2608.02665*.

[30] Chen, Q. (2026, March). RADR-MO: Resilience-Aware Dynamic Re-routing via Graph Attention Networks and Multi-Objective Meta-Reinforcement Learning for Supply Chain Resilience. In *Proceedings of the 2026 3rd International Conference on Innovation Management and Information Systems* (pp. 155-162).

[31] Ma, Q., & Wu, R. (2026). When Validation Stops Learning: Auditing Update Admission for Continual Embodied Agents. *arXiv preprint arXiv:2609.10873*.

[32] Que, H., Gao, H., Shan, W., Liu, M., An, J., Deng, F., ... & Mu, L. (2026). FM-SAM: individual tree crown delineation and classification based on Segmentation Anything Model (SAM) and YOLOv10 in UAV imagery for forest monitoring. *Computers and Electronics in Agriculture*, 240, 111162.

[33] Zhu, H., Huang, X., Gao, H., Jiang, M., Que, H., & Mu, L. (2025, July). A Wireless Collaborated Inference Acceleration Framework for Plant Disease Recognition. In *International Conference on Intelligent Computing* (pp. 331-341). Singapore: Springer Nature Singapore.

[34] Liu, M., Que, H., Yang, X., Zhang, K., Yu, Q., Yan, L., ... & Zhou, N. (2023). A Selective Bit Dropping and Encoding Co-Strategy in Image Processing for Low-Power Design in DRAM and SRAM. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 13(1), 48-57.

[35] Gao, H., Que, H., Au, H., Shan, W., Liu, M., Qin, Y., ... & Qiao, F. (2025). SenseExpo: Efficient Autonomous Exploration with Prediction Information from Lightweight Neural Networks. *arXiv preprint arXiv:2503.16000*.

[36] Xi, X., Chen, P., Qiu, Y., Luo, R., Tong, P., & Liang, J. Video-BCI: Bayesian Cognitive Integration of Self-Prior Hypotheses for Video Understanding. In *Forty-third International Conference on Machine Learning*.

[37] Liang, J., Xiao, Y., Yang, H., Yu, Z., Liu, J. N., & Yang, K. (2026, April). C-HyPOD: Causal Hyperbolic Representation Learning with Prototype Orthogonal Disentanglement for Graph

- Out-of-Distribution Recommendation. In Proceedings of the ACM Web Conference 2026 (pp. 6541-6550).
- [38] C. Wan, Y. Luo, W. Wu, N. D. Hatziaargyriou and S. Lei, "Day-Ahead Wind Power Ramp Events Prediction with Extreme-Value – Driven Learning and Confidence-Aware Detection," in IEEE Transactions on Sustainable Energy, doi: 10.1109/TSTE.2026.3669605.
- [39] Y. Wang, C. Wan and Y. Luo, "WeaGAN++: An Efficient Weather-Aware Graph Attention Network for Traffic Prediction," in IEEE Internet of Things Journal, vol. 13, no. 10, pp. 21197-21212, 15 May15, 2026, doi: 10.1109/JIOT.2026.3665116.
- [40] Quan, Q., Gao, Y., & Wang, Q. (2025). The impact of teacher emotional support on students' engagement in AI-mediated English learning environments: The mediating role of resilience and self-efficacy. *Acta Psychologica*, 260, 105766.
- [41] Yang, J., Li, Q., Yu, H., Wang, X., Wang, C., Ma, B., ... & Li, X. (2026). Invisible Image Watermark Attacks: A Survey.
- [42] Su, J., Cai, C., Zhu, F., He, C., Xu, X., Guan, D., & Si, C. (2024, September). Momentum auxiliary network for supervised local learning. In European Conference on Computer Vision (pp. 276-292). Cham: Springer Nature Switzerland.
- [43] Yu, C., Li, P., Wu, H., Wen, Y., Deng, B., & Xiong, H. (2024). USM: Unbiased Survey Modeling for Limiting Negative User Experiences in Recommendation Systems. arXiv preprint arXiv:2412.10674.
- [44] Wu, X., Gao, D., Yao, J., Qian, Q., & Song, J. (2025). Optimization of Single-Track Train Schedules with Cyclic Operation Strategies. In H. Fujita, A. Hernandez-Matamoros, & Y. Watanobe (Eds.), Proceedings of the 24th International Conference on New Trends in Intelligent Software Methodologies, Tools and Techniques (SoMeT_25) (pp. 197-206). (Frontiers in Artificial Intelligence and Applications; Vol. 411). IOS Press BV. <https://doi.org/10.3233/FAIA250522>
- [45] Liu, T., He, J., Zhou, R., Song, J., Liu, H., & Liang, Y. (2025). Revolutionizing clinical decision making through deep learning and topic modeling for pathway optimization. *Scientific Reports*, 15(1), Article 28787. <https://doi.org/10.1038/s41598-025-12679-z>
- [46] Ma, C., Song, J., Xu, Y., Fan, H., Wu, X., & Sun, T. (2024). Vehicle-Based Machine Vision Approaches in Intelligent Connected System. *IEEE Transactions on Intelligent Transportation Systems*, 25(3), 2827-2836. <https://doi.org/10.1109/TITS.2023.3276325>
- [47] Hao, L., Zhang, S., & Schulzrinne, H. (2025, October). Advancing IoT System Dependability: A Deep Dive into Management and Operation Plane Separation. In 2025 IEEE 11th World Forum on Internet of Things (WF-IoT) (pp. 1-6). IEEE.
- [48] Chanda, V., Hao, L., & Schulzrinne, H. (2024, January). From frustration to function: A study on usability challenges in smart home IoT devices. In 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC) (pp. 1-6). IEEE.
- [49] Zuo, H., Wu, J., Lim, K. H., Liao, Y., Song, L., Zhu, Y., ... & Zhang, C. (2025). A time series long-short term codec for compression and representation. *Applied Intelligence*, 55(17), 1-13.
- [50] Chen, B., Zuo, H., Liao, Y., & Li, Z. (2024, December). A Data-Driven Analysis of Mental Health Status and Early Warning System for Vocational College Students. In 2024 5th International Conference on Information Science and Education (ICISE-IE) (pp. 312-315). IEEE.
- [51] Ning, E., Gao, W., Liu, Y., Yuan, J., Zhang, H., He, S., & Ning, X. (2026). Causally invariant video anomaly detection via counterfactual reasoning and prototype intervention. *Pattern Recognition*, 113809.