Novelty Assessment Report

Paper: Veritas: Generalizable Deepfake Detection via Pattern-Aware Reasoning
PDF URL: https://openreview.net/pdf?id=5VXJPS1HoM
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

Deepfake detection remains a formidable challenge due to the evolving nature of fake content in real-world scenarios. However, existing benchmarks suffer from severe discrepancies from industrial practice, typically featuring homogeneous training sources and low-quality testing images, which hinder the practical usage of current detectors. To mitigate this gap, we introduce **HydraFake**, a dataset that contains diversified deepfake techniques and in-the-wild forgeries, along with rigorous training and evaluation protocol, covering unseen model architectures, emerging forgery techniques and novel data domains. Building on this resource, we propose **Veritas**, a multi-modal large language model (MLLM) based deepfake detector. Different from vanilla chain-of-thought (CoT), we introduce pattern-aware reasoning that involves critical patterns such as "planning" and "self-reflection" to emulate human forensic process. We further propose a two-stage training pipeline to seamlessly internalize such deepfake reasoning capacities into current MLLMs. Experiments on HydraFake dataset reveal that although previous detectors show great generalization on cross-model scenarios, they fall short on unseen forgeries and data domains. Our Veritas achieves significant gains across different out-of-domain (OOD) scenarios, and is capable of delivering transparent and faithful detection outputs.

Core Task Landscape

This paper addresses: **Generalizable Deepfake Detection via Pattern-Aware Reasoning**

A total of **43 papers** were analyzed and organized into a taxonomy with **31 categories**.

Taxonomy Overview

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

- **Frequency and Spectral Domain Analysis**
- **Temporal and Spatiotemporal Reasoning**
- **Feature Disentanglement and Decomposition**
- **Large Pre-trained Model Adaptation**
- **Data Augmentation and Synthetic Training**
- **Domain Adaptation and Generalization Strategies**
- **Meta-Learning and Few-Shot Detection**
- **Local and Patch-Level Analysis**
- **Identity and Semantic Consistency Analysis**
- **Noise Pattern and Forensic Trace Analysis**
- ... and 3 more categories

Complete Taxonomy Tree

- Generalizable Deepfake Detection via Pattern-Aware Reasoning Survey Taxonomy
- Frequency and Spectral Domain Analysis
  - Frequency-Level Artifact Detection (2 papers)
  - [1] Frequency-aware deepfake detection: Improving generalizability through frequency space domain learning (Chuangchuang Tan, 2024) View paper
  - [3] Deepfake Detection without Deepfakes: Generalization via Synthetic Frequency Patterns Injection (Coccomini, 2024) View paper
  - Wavelet and Directional Transform Methods (1 papers)
  - [26] Dualâtree complex wavelet transformâbased direction correlation for face forgery detection (Shichao Gao, 2021) View paper
  - Spatial-Frequency Interaction and Fusion (3 papers)
  - [9] Dynamic Graph Learning with Content-guided Spatial-Frequency Relation Reasoning for Deepfake Detection (Yuan Wang, 2023) View paper
  - [16] A two-stage interaction approach for enhancing generalization of deepfake detection (Chenglong Sun, 2025) View paper
  - [36] Learning spatial-frequency interaction for generalizable deepfake detection (Tianbo Zhai, 2024) View paper
- Temporal and Spatiotemporal Reasoning
  - Temporal Inconsistency and Motion Pattern Analysis (2 papers)
  - [17] DIP: Diffusion Learning of Inconsistency Pattern for General DeepFake Detection (Fan Nie, 2024) View paper
  - [19] Hybrid Federated Deepfake Detection via Residual-Aware Temporal Modeling and Privacy-Preserving Learning (Adarsh Bhatnagar, 2025) View paper
  - Predictive Representation and Latent Pattern Learning (1 papers)
  - [29] Latent Pattern Sensing: Deepfake Video Detection via Predictive Representation Learning (Shiming Ge, 2021) View paper
  - Multimodal Audiovisual Temporal Detection (2 papers)

    - [28] HCN-TA: Hierarchical Capsule Network with Temporal Attention for a Generalizable Approach to Audio Deepfake Detection (Taiba Majid Wani, 2025) View paper
    - [30] Referee: Reference-aware Audiovisual Deepfake Detection (Lee Eunsang, 2025) View paper
- Feature Disentanglement and Decomposition
    - Forgery-Irrelevant Feature Disentanglement (1 papers)
    - [4] Ucf: Uncovering common features for generalizable deepfake detection (Zhiyuan Yan, 2023) View paper
    - Texture and Artifact Decomposition (1 papers)
    - [8] Texture and artifact decomposition for improving generalization in deep-learning-based deepfake detection (Jie Gao, 2024) View paper
    - Lighting and Physical Property Analysis (1 papers)
    - [10] LIDeepDet: Deepfake Detection via Image Decomposition and Advanced Lighting Information Analysis (Zhimao Lai, 2024) View paper
- Large Pre-trained Model Adaptation
    - Vision Transformer Fine-Tuning and Adaptation (2 papers)
    - [11] Generalized Face Forgery Detection via Adaptive Learning for Pre-trained Vision Transformer (Luo, 2023) View paper
    - [23] DeepFake-Adapter: Dual-Level Adapter for DeepFake Detection (Rui Shao, 2023) View paper
    - CLIP-Based Detection and Prompt Engineering (2 papers)
    - [5] C2P-CLIP: Injecting Category Common Prompt in CLIP to Enhance Generalization in Deepfake Detection (Tan, 2024) View paper
    - [7] AuthGuard: Generalizable Deepfake Detection via Language Guidance (Shen Guang-yu, 2025) View paper
    - Multimodal Large Language Model Reasoning ★ (3 papers)
    - [0] Veritas: Generalizable Deepfake Detection via Pattern-Aware Reasoning (Anon et al., 2026) View paper
    - [15] Skyra: AI-Generated Video Detection via Grounded Artifact Reasoning (Yifei Li, 2025) View paper
    - [32] EDVD-LLaMA: Explainable Deepfake Video Detection via Multimodal Large Language Model Reasoning (Sun Hao-Ran, 2025) View paper
- Data Augmentation and Synthetic Training
    - Adversarial Example and Forgery Configuration Synthesis (1 papers)
    - [2] Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection (Liang Chen, 2022) View paper
    - Diffusion-Based Synthetic Data Generation (1 papers)
    - [14] DiffusionFake: Enhancing Generalization in Deepfake Detection via Guided Stable Diffusion (Sun Ke, 2024) View paper
- Domain Adaptation and Generalization Strategies
    - Invariant Risk Minimization and Causal Learning (1 papers)
    - [25] Improving Deepfake Detection Generalization by Invariant Risk Minimization (Zixin Yin, 2024) View paper
    - Domain Adaptive Batch Normalization (1 papers)
    - [39] Improving Generalization of Deepfake Detection with Domain Adaptive Batch Normalization (Zixin Yin, 2021) View paper
    - Continual Learning and Generalization Preservation (1 papers)
    - [13] Generalization-Preserved Learning: Closing the Backdoor to Catastrophic Forgetting in Continual Deepfake Detection (X Zhang, 2025) View paper
- Meta-Learning and Few-Shot Detection (2 papers)
    - [22] Meta-Learning With Relation Embedding for Few-Shot Deepfake Detection (Xiaoyong Liu, 2024) View paper
    - [42] Generalization of Forgery Detection With Meta Deepfake Detection Model (Van-Nhan Tran, 2022) View paper
- Local and Patch-Level Analysis
    - Patch-Level Classification and Discontinuity Mining (2 papers)
    - [24] Loupe: A Generalizable and Adaptive Framework for Image Forgery Detection (Jiang Yuchu, 2025) View paper
    - [34] Patch-Discontinuity Mining for Generalized Deepfake Detection (Huanhuan Yuan, 2025) View paper
    - Local Focusing Mechanisms and Reconstruction (1 papers)
    - [27] A Novel Local Focusing Mechanism for Deepfake Detection Generalization (Li Mingliang, 2025) View paper
    - Cross-Domain Local Forensics (1 papers)
    - [31] Cross-Domain Local Characteristic Enhanced Deepfake Video Detection (Liu Zihan, 2022) View paper
- Identity and Semantic Consistency Analysis
    - Identity Leakage and Representation Analysis (1 papers)
    - [12] Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization (Shichao Dong, 2022) View paper
    - Semantic Forgery Detection (1 papers)
    - [33] Context-Aware Semantic Forgery Detection in Biomedical & Natural Images (Nandi, 2025) View paper
- Noise Pattern and Forensic Trace Analysis
    - Sensor Noise and PRNU-Based Detection (1 papers)
    - [20] A Pattern-based Comparative Analysis and Hybrid Model for Deepfake Image Source Attribution (Abhay Chopade, 2025) View paper
- Attention Mechanisms and Architectural Innovations
    - Cross-Modal and Spatial-Temporal Attention (1 papers)
    - [21] CSTAN: A Deepfake Detection Network with CST Attention for Superior Generalization. (Rui Yang, 2024) View paper
    - Siamese Training and Contrastive Learning (1 papers)
    - [38] AST: Generalization of Deepfake Detection with Attention Siamese Training (Taiying Peng, 2023) View paper
    - Cluster Decision and Ensemble Strategies (1 papers)
    - [37] CDNet: Cluster Decision for Deepfake Detection Generalization (Zeming Hou, 2023) View paper
- Baseline and Specialized Architectures
    - XceptionNet and Established CNN Baselines (1 papers)
    - [18] Deepfake Detection Using XceptionNet (Muskan Kumari, 2023) View paper
    - Lightweight Frequency-Aware Networks (1 papers)
    - [35] FALCON: Frequency-Aware Lightweight Convolutional Optimized Network (Himanshu Maan, 2025) View paper
    - Attribution-Based Training for Generalization (1 papers)
    - [43] Improving Generalization of Deepfake Detection by Training for Attribution (Anubhav Jain, 2021) View paper

- Surveys and Generalization Analysis (3 papers)
  - [6] Robust Deepfake Detection by Addressing Generalization and Trustworthiness Challenges: A Short Survey (Ping Liu, 2024) View paper
  - [40] On the Generalization of Deep Learning Models in Video Deepfake Detection (Davide Alessandro Coccomini, 2023) View paper
  - [41] Towards Generalization in Deepfake Detection (Luisa Verdoliva, 2022) View paper

Narrative

Core task: generalizable deepfake detection via pattern-aware reasoning. The field has evolved into a rich landscape of complementary strategies, each addressing different facets of the generalization challenge. At the highest level, the taxonomy reveals several major branches: frequency and spectral domain analysis (e.g., Frequency-aware Detection[1], Synthetic Frequency Patterns[3]) exploits artifacts in the frequency spectrum that persist across generation methods; temporal and spatiotemporal reasoning captures inconsistencies over time; feature disentanglement and decomposition (e.g., Texture Artifact Decomposition[8]) separates content from manipulation traces; large pre-trained model adaptation leverages foundation models such as CLIP or large language models; data augmentation and synthetic training strategies create diverse training signals; domain adaptation and generalization strategies (e.g., Invariant Risk Minimization[25]) explicitly optimize for cross-domain robustness; meta-learning and few-shot detection (e.g., Meta-Learning Relation Embedding[22]) enable rapid adaptation to novel forgeries; local and patch-level analysis (e.g., Patch-Discontinuity Mining[34]) focuses on fine-grained spatial cues; identity and semantic consistency analysis checks for logical coherence; noise pattern and forensic trace analysis mines low-level statistical signatures; attention mechanisms and architectural innovations introduce novel inductive biases; baseline and specialized architectures provide reference points; and surveys (e.g., Robust Detection Survey[6]) synthesize the state of the art.

A particularly active line of work centers on adapting large pre-trained models to deepfake detection, where methods such as C2P-CLIP[5] and DeepFake-Adapter[23] fine-tune vision-language or vision-only backbones to capture generalizable forgery patterns. Within this branch, a small but growing cluster explores multimodal large language model reasoning, combining visual and textual modalities to perform more interpretable, context-aware detection. Veritas[0] sits squarely in this cluster, alongside Skyra[15] and EDVD-LLaMA[32], all of which harness the reasoning capabilities of large language models to identify subtle inconsistencies that simpler architectures might miss. Compared to Skyra[15], which emphasizes cross-modal alignment, and EDVD-LLaMA[32], which integrates video-level temporal cues, Veritas[0] focuses on pattern-aware reasoning that bridges low-level forensic traces with high-level semantic understanding. This direction reflects a broader trend toward interpretable, reasoning-driven detection, contrasting with purely data-driven approaches in frequency analysis or meta-learning branches, and highlights ongoing questions about how best to combine domain-specific inductive biases with the flexibility of foundation models.

Related Works in Same Category

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

1. Skyra: AI-Generated Video Detection via Grounded Artifact Reasoning

Authors: Yifei Li, Wenzhao Zheng, Yanran Zhang, Runze Sun, Yu Zheng, et al. (8 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

The misuse of AI-driven video generation technologies has raised serious social concerns, highlighting the urgent need for reliable AI-generated video detectors. However, most existing methods are limited to binary classification and lack the necessary explanations for human interpretation. In this paper, we present Skyra, a specialized multimodal large language model (MLLM) that identifies human-perceivable visual artifacts in AI-generated videos and leverages them as grounded evidence for both...

Relationship Analysis

Both papers belong to the Multimodal Large Language Model Reasoning category, employing MLLMs with pattern-aware reasoning for explainable detection outputs. They overlap in using chain-of-thought reasoning, grounded artifact identification, and two-stage training pipelines (SFT followed by reinforcement learning) to enhance detection and explanation capabilities. The key difference is that the original paper (VERITAS) focuses on deepfake face detection with a pattern-aware reasoning framework involving five thinking patterns (fast judgment, planning, reasoning, conclusion, self-reflection), while the candidate paper (Skyra) targets AI-generated video detection with spatio-temporally grounded artifact reasoning and a hierarchical artifact taxonomy for video-specific forgeries.

2. EDVD-LLaMA: Explainable Deepfake Video Detection via Multimodal Large Language Model Reasoning

Authors: Sun Hao-Ran, Cai Chen, Hao Sun, Zhuang Huiping, Chen Cai, et al. (14 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

The rapid development of deepfake video technology has not only facilitated artistic creation but also made it easier to spread misinformation. Traditional deepfake video detection (DVD) methods face issues such as a lack of transparency in their principles and insufficient generalization capabilities to cope with evolving forgery techniques. This highlights an urgent need for detectors that can identify forged content and provide verifiable reasoning explanations. This paper proposes the explai...

Relationship Analysis

Both papers belong to the Multimodal Large Language Model Reasoning category, leveraging MLLMs with chain-of-thought reasoning for deepfake detection. They overlap in using pattern-aware or structured reasoning approaches to provide explainable detection outputs, with both emphasizing transparency and faithful decision processes. However, the original paper (Veritas) focuses on facial image detection with adaptive planning and self-reflection patterns trained via a two-stage pipeline (SFT + MiPO + P-GRPO), while the candidate paper (EDVD-LLaMA) targets deepfake video detection by incorporating spatio-temporal modeling (ST-SIT) and fine-grained multimodal chain-of-thought (Fg-MCoT) with facial landmark constraints to handle temporal inconsistencies across video frames.

Contributions Analysis

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

Contribution 1: HydraFake dataset with hierarchical evaluation protocol

Description: The authors construct a new deepfake detection dataset featuring diverse forgery techniques and in-the-wild samples. They establish a hierarchical evaluation protocol with four testing levels (in-domain, cross-model, cross-forgery, cross-domain) to simulate real-world challenges and comprehensively measure detector generalization.

This contribution was assessed against **7 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. WATCHER: Wavelet-guided texture-content hierarchical relation learning for deepfake detection

URL: View paper

Brief Assessment

WATCHER[54] focuses on wavelet-guided texture-content hierarchical relation learning for detection methods, not on dataset construction or evaluation protocols. The candidate does not present a competing dataset or hierarchical evaluation framework.

---

2. XMAD-Bench: Cross-Domain Multilingual Audio Deepfake Benchmark

URL: View paper

Brief Assessment

XMAD-Bench[59] focuses on audio deepfake detection with cross-domain multilingual evaluation, not visual deepfake detection. The candidate addresses a completely different modality (audio vs. images/video) and does not challenge the novelty of HydraFake's hierarchical visual deepfake evaluation protocol.

---

3. Multi-level distributional discrepancy enhancement for cross domain face forgery detection

URL: View paper

Brief Assessment

Multi-level Distributional Discrepancy[55] focuses on cross-domain face forgery detection methods rather than dataset construction. The candidate only mentions using existing benchmarks (FF++, CelebDF, WDF) for evaluation, not creating a new hierarchical evaluation protocol.

---

4. HiTAL: Hierarchical Thumbnail and Latent Augmentation for Deepfake Detection

URL: View paper

Brief Assessment

HiTAL[57] focuses on a hierarchical detection framework with thumbnail layouts and latent space augmentation for deepfake videos, not on constructing a dataset with hierarchical evaluation protocols for measuring detector generalization across different testing levels.

---

5. Wav2DF-TSL: Two-stage Learning with Efficient Pre-training and Hierarchical Experts Fusion for Robust Audio Deepfake Detection

URL: View paper

Brief Assessment

Wav2DF-TSL[60] focuses on audio deepfake detection using self-supervised learning and mixture of experts, not on constructing image-based deepfake datasets with hierarchical evaluation protocols.

---

6. Unmasking synthetic realities in generative ai: A comprehensive review of adversarially robust deepfake detection systems

URL: View paper

Brief Assessment

Adversarially Robust Review[56] focuses on adversarial robustness evaluation of deepfake detection systems rather than dataset construction with hierarchical protocols. The review discusses evaluation across datasets but does not propose a new dataset with the specific four-level hierarchical testing structure (in-domain, cross-model, cross-forgery, cross-domain) that characterizes HydraFake.

---

7. Multi-domain Multi-scale DeepFake Detection for Generalization

URL: View paper

Brief Assessment

Multi-domain Multi-scale[58] focuses on multi-domain multi-scale feature extraction methods for deepfake detection, not on dataset construction or hierarchical evaluation protocols. The candidate does not present a new dataset or evaluation framework that would challenge the novelty of HydraFake's hierarchical testing protocol.

Contribution 2: Pattern-aware reasoning framework for deepfake detection

Description: The authors propose a reasoning framework that incorporates five thinking patterns (fast judgement, planning, reasoning, self-reflection, conclusion) inspired by human forensic analysis. This pattern-aware approach enables logical and holistic reasoning for deepfake detection, outperforming vanilla chain-of-thought methods.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. READFake: Reflection and Environment-Aware DeepFake Detection

URL: View paper

Brief Assessment

READFake[61] focuses on physical reflection and environmental consistency analysis for deepfake detection, not on reasoning frameworks with cognitive patterns like planning and self-reflection.

Contribution 3: Two-stage training pipeline with MiPO and P-GRPO

Description: The authors develop a training pipeline consisting of pattern-guided cold-start (with SFT and Mixed Preference Optimization) and Pattern-aware Group Relative Policy Optimization. This pipeline internalizes reasoning abilities into MLLMs, enabling adaptive planning and self-reflection while delivering transparent and faithful detection outputs.

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. Towards Intrinsic Self-Correction Enhancement in Monte Carlo Tree Search Boosted Reasoning via Iterative Preference Learning

URL: View paper

Brief Assessment

The candidate focuses on arithmetic reasoning tasks using step-wise preference learning for self-correction in LLMs, not on multimodal deepfake detection with pattern-aware reasoning for MLLMs. The training objectives and application domains are fundamentally different.

---

2. Preference-Optimized Retrieval and Ranking for Efficient Multimodal Recommendation

URL: View paper

Brief Assessment

Preference-Optimized Retrieval[48] focuses on a two-stage retrieval-ranking pipeline for multimodal recommendation systems, not on training MLLMs for deepfake detection with pattern-aware reasoning. The technical domains and objectives are fundamentally different.

---

3. Step-Controlled DPO: Leveraging Stepwise Error for Enhanced Mathematical Reasoning

URL: View paper

Brief Assessment

Step-Controlled DPO[49] focuses on mathematical reasoning with stepwise error supervision using DPO variants, not on multimodal deepfake detection with pattern-aware reasoning internalization through MiPO and P-GRPO for MLLMs.

---

4. Agent Q: Advanced Reasoning and Learning for Autonomous AI Agents

URL: View paper

Brief Assessment

Agent Q[44] focuses on web navigation tasks using MCTS with DPO for agentic reasoning, not on multimodal deepfake detection with pattern-aware reasoning internalization. The technical domains and objectives are fundamentally different.

---

5. Think Twice to See More: Iterative Visual Reasoning in Medical VLMs

URL: View paper

Brief Assessment

Think Twice[47] uses a two-stage training strategy (SFT followed by RL with GRPO), but focuses on medical visual reasoning with iterative 'think-act-rethink-answer' chains rather than deepfake detection with pattern-aware reasoning and mixed preference optimization.

---

6. Advancing Multimodal Reasoning: From Optimized Cold Start to Staged Reinforcement Learning

URL: View paper

Brief Assessment

Optimized Cold Start[51] focuses on a three-stage curriculum (text cold-start, multimodal RL, text RL) for general multimodal reasoning, while the original paper targets deepfake detection with pattern-aware reasoning. The technical approaches and application domains differ fundamentally.

---

7. Iterative Tool Usage Exploration for Multimodal Agents via Step-wise Preference Tuning

URL: View paper

Brief Assessment

Iterative Tool Usage[52] focuses on multimodal agents for tool usage exploration through step-wise preference optimization, not on deepfake detection with pattern-aware reasoning internalization as in the original paper.

---

8. Progressive multimodal reasoning via active retrieval

URL: View paper

Brief Assessment

Progressive Multimodal Reasoning[46] focuses on multimodal reasoning through active retrieval and MCTS, not on deepfake detection training pipelines. The candidate uses curriculum training objectives to align a process reward model for reasoning verification, which is a different application domain than the original paper's deepfake detection with pattern-aware reasoning.

---

9. Toward Effective Tool-Integrated Reasoning via Self-Evolved Preference Learning

URL: View paper

Brief Assessment

Self-Evolved Preference[50] focuses on tool-integrated reasoning with a two-stage pipeline (SFT + self-evolved DPO), not multimodal reasoning internalization. The original paper's MiPO uses mixed non-preference data for deepfake detection, while the candidate uses entropy-guided sampling for tool-call optimization—fundamentally different domains and objectives.

---

10. Open vision reasoner: Transferring linguistic cognitive behavior for visual reasoning

URL: View paper

Brief Assessment

Open Vision Reasoner[45] uses a two-stage training pipeline (linguistic cold-start followed by multimodal RL), but does not employ Mixed Preference Optimization (MiPO) or Pattern-aware Group Relative Policy Optimization (P-GRPO). The candidate focuses on transferring linguistic cognitive behaviors to visual reasoning through standard PPO with GAE, which differs from the original paper's pattern-aware reasoning framework with specialized preference optimization strategies.

Appendix: Text Similarity Detection

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

References

- [0] Veritas: Generalizable Deepfake Detection via Pattern-Aware Reasoning View paper
- [1] Frequency-aware deepfake detection: Improving generalizability through frequency space domain learning View paper
- [2] Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection View paper
- [3] Deepfake Detection without Deepfakes: Generalization via Synthetic Frequency Patterns Injection View paper
- [4] Ucf: Uncovering common features for generalizable deepfake detection View paper
- [5] C2P-CLIP: Injecting Category Common Prompt in CLIP to Enhance Generalization in Deepfake Detection View paper
- [6] Robust Deepfake Detection by Addressing Generalization and Trustworthiness Challenges: A Short Survey View paper
- [7] AuthGuard: Generalizable Deepfake Detection via Language Guidance View paper
- [8] Texture and artifact decomposition for improving generalization in deep-learning-based deepfake detection View paper
- [9] Dynamic Graph Learning with Content-guided Spatial-Frequency Relation Reasoning for Deepfake Detection View paper
- [10] LIDeepDet: Deepfake Detection via Image Decomposition and Advanced Lighting Information Analysis View paper
- [11] Generalized Face Forgery Detection via Adaptive Learning for Pre-trained Vision Transformer View paper
- [12] Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization View paper
- [13] Generalization-Preserved Learning: Closing the Backdoor to Catastrophic Forgetting in Continual Deepfake Detection View paper

- [14] DiffusionFake: Enhancing Generalization in Deepfake Detection via Guided Stable Diffusion View paper
- [15] Skyra: AI-Generated Video Detection via Grounded Artifact Reasoning View paper
- [16] A two-stage interaction approach for enhancing generalization of deepfake detection View paper
- [17] DIP: Diffusion Learning of Inconsistency Pattern for General DeepFake Detection View paper
- [18] Deepfake Detection Using XceptionNet View paper
- [19] Hybrid Federated Deepfake Detection via Residual-Aware Temporal Modeling and Privacy-Preserving Learning View paper
- [20] A Pattern-based Comparative Analysis and Hybrid Model for Deepfake Image Source Attribution View paper
- [21] CSTAN: A Deepfake Detection Network with CST Attention for Superior Generalization. View paper
- [22] Meta-Learning With Relation Embedding for Few-Shot Deepfake Detection View paper
- [23] DeepFake-Adapter: Dual-Level Adapter for DeepFake Detection View paper
- [24] Loupe: A Generalizable and Adaptive Framework for Image Forgery Detection View paper
- [25] Improving Deepfake Detection Generalization by Invariant Risk Minimization View paper
- [26] Dualâtree complex wavelet transformâbased direction correlation for face forgery detection View paper
- [27] A Novel Local Focusing Mechanism for Deepfake Detection Generalization View paper
- [28] HCN-TA: Hierarchical Capsule Network with Temporal Attention for a Generalizable Approach to Audio Deepfake Detection View paper
- [29] Latent Pattern Sensing: Deepfake Video Detection via Predictive Representation Learning View paper
- [30] Referee: Reference-aware Audiovisual Deepfake Detection View paper
- [31] Cross-Domain Local Characteristic Enhanced Deepfake Video Detection View paper
- [32] EDVD-LLaMA: Explainable Deepfake Video Detection via Multimodal Large Language Model Reasoning View paper
- [33] Context-Aware Semantic Forgery Detection in Biomedical & Natural Images View paper
- [34] Patch-Discontinuity Mining for Generalized Deepfake Detection View paper
- [35] FALCON: Frequency-Aware Lightweight Convolutional Optimized Network View paper
- [36] Learning spatial-frequency interaction for generalizable deepfake detection View paper
- [37] CDNet: Cluster Decision for Deepfake Detection Generalization View paper
- [38] AST: Generalization of Deepfake Detection with Attention Siamese Training View paper
- [39] Improving Generalization of Deepfake Detection with Domain Adaptive Batch Normalization View paper
- [40] On the Generalization of Deep Learning Models in Video Deepfake Detection View paper
- [41] Towards Generalization in Deepfake Detection View paper
- [42] Generalization of Forgery Detection With Meta Deepfake Detection Model View paper
- [43] Improving Generalization of Deepfake Detection by Training for Attribution View paper
- [44] Agent Q: Advanced Reasoning and Learning for Autonomous AI Agents View paper
- [45] Open vision reasoner: Transferring linguistic cognitive behavior for visual reasoning View paper
- [46] Progressive multimodal reasoning via active retrieval View paper
- [47] Think Twice to See More: Iterative Visual Reasoning in Medical VLMs View paper
- [48] Preference-Optimized Retrieval and Ranking for Efficient Multimodal Recommendation View paper
- [49] Step-Controlled DPO: Leveraging Stepwise Error for Enhanced Mathematical Reasoning View paper
- [50] Toward Effective Tool-Integrated Reasoning via Self-Evolved Preference Learning View paper
- [51] Advancing Multimodal Reasoning: From Optimized Cold Start to Staged Reinforcement Learning View paper
- [52] Iterative Tool Usage Exploration for Multimodal Agents via Step-wise Preference Tuning View paper
- [53] Towards Intrinsic Self-Correction Enhancement in Monte Carlo Tree Search Boosted Reasoning via Iterative Preference Learning View paper
- [54] WATCHER: Wavelet-guided texture-content hierarchical relation learning for deepfake detection View paper
- [55] Multi-level distributional discrepancy enhancement for cross domain face forgery detection View paper
- [56] Unmasking synthetic realities in generative ai: A comprehensive review of adversarially robust deepfake detection systems View paper
- [57] HiTAL: Hierarchical Thumbnail and Latent Augmentation for Deepfake Detection View paper
- [58] Multi-domain Multi-scale DeepFake Detection for Generalization View paper
- [59] XMAD-Bench: Cross-Domain Multilingual Audio Deepfake Benchmark View paper
- [60] Wav2DF-TSL: Two-stage Learning with Efficient Pre-training and Hierarchical Experts Fusion for Robust Audio Deepfake Detection View paper
- [61] READFake: Reflection and Environment-Aware DeepFake Detection View paper