Qeios

Peer Review

Review of: "TagFog: Textual Anchor Guidance and Fake Outlier Generation for Visual Out-of-Distribution Detection"

Shuwen Pan[1]

1. Independent researcher

1. Summary and Strengths

The manuscript presents a novel framework, TagFog, which integrates textual guidance from ChatGPT-generated semantic embeddings and Jigsaw-based fake OOD data generation to enhance visual OOD detection. The approach demonstrates state-of-the-art performance across multiple benchmarks (CIFAR, ImageNet100), with extensive ablation studies and sensitivity analyses supporting the efficacy of its components. Key strengths include:

2. Innovation

The integration of ChatGPT-generated textual anchors with CLIP embeddings is novel and addresses the limitation of sparse class-name-based representations. The Jigsaw-based fake OOD generation strategy is simple yet effective compared to GAN-based alternatives.

3. Empirical Rigor

Comprehensive experiments on diverse benchmarks (Tables 2–4) and ablation studies (Table 5, Figure 3) validate the contributions of each component (fake OOD data, contrastive losses).

4. Practical Flexibility

The framework's compatibility with existing post-hoc methods (Tables 6–7) highlights its adaptability for real-world deployment.

5. Areas for Improvement

(1) Methodological Clarity

doi.org/10.32388/JRJ15A

1）Section 3.2 (Fake Outlier Generation) While the Jigsaw strategy is described, details about patch size selection (e.g., why 4×4 for CIFAR vs. 8×8 for ImageNet) and the rationale behind generating 1–4 fake OOD samples per ID image are not thoroughly justified. A brief discussion on how these parameters affect robustness would strengthen reproducibility.

2）Section 3.3 (Textual Anchor Guidance) The process of generating ChatGPT descriptions (e.g., prompt design, handling class ambiguity) is not detailed. Including example prompts and a discussion of potential biases in ChatGPT outputs (e.g., over/under-describing certain classes) would enhance transparency.

（2）Experimental Design

Baseline Comparisons——While the manuscript compares TagFog to methods like VOS and CIDER, recent SOTA works such as POEM (ICML 2022) and WOOD (2023) are omitted. Including these would better contextualize the claimed advancements.

Real-World Applicability——The experiments focus on synthetic OOD datasets (e.g., LSUN, Textures). Testing on real-world OOD scenarios (e.g., adversarial examples, domain shifts) would better demonstrate practical utility.

（3）Limitations and Broader Impact

1）Dependency on CLIP——The framework relies on CLIP's frozen text encoder. A discussion of how performance might degrade with weaker pre-trained models (e.g., smaller CLIP variants) is needed.

2）Computational Cost—— The computational overhead of generating ChatGPT descriptions and training with multiple loss terms should be quantified (e.g., training time vs. baseline methods).

（4）Writing and Presentation

1）Figure 2——The diagram is informative but lacks labels for key components (e.g., projection module gg, temperature parameters). Annotating these would improve readability.

2）Table 2——The OOD dataset abbreviations (e.g., LSUN-R, LSUN-C) should be expanded in a footnote for clarity.

6. Recommendations

（1）Expand Method Details——Clarify the design choices for Jigsaw parameters and ChatGPT prompts.

（2）Update Baseline Comparisons——Include recent SOTA methods (e.g., POEM, WOOD) and discuss their relevance.

（3）Strengthen Real-World Evaluation——Add experiments on adversarial or domain-shifted OOD data.

（4）Discuss Limitations ——Address dependencies on CLIP and computational costs.

7. Conclusion

The manuscript presents a compelling and innovative framework for OOD detection, supported by rigorous experimentation. Addressing the above points will enhance its technical depth, reproducibility, and impact. The work has strong potential for publication pending minor revisions.

8. Rating

Novelty 4/5

Technical Soundness 4/5

Clarity 3/5

Impact 4/5

Declarations

Potential competing interests: No potential competing interests to declare.