**Response to Reviewers' Comments**

Dear Editor and Dear Reviewers,

We are submitting the manuscript titled "**FD-SGIL: Fraudulent Website Detection Model based on Structure Growing Incremental Learning**" to IEEE Transactions on Information Forensics and Security. This submission is related to a previously rejected manuscript—originally titled "**FD-SGIL: Fraudulent Website Detection Model based on Structure Growing Incremental Learning**" (**Manuscript ID: IoT-48797-2025**)—which was previously submitted to IEEE Internet of Things Journal and subsequently rejected.

We highly value the constructive comments from the previous review process at IEEE Internet of Things Journal, as they have been critical in enhancing the quality of our manuscript. In this submission to IEEE Transactions on Information Forensics and Security, we have thoroughly addressed all feedback from the prior reviewers, and the following content provides a detailed, point-by-point response to those comments. To ensure clarity for your review, we have formatted the previous reviewers' comments in bold red with clear headings, and marked all revised content in the manuscript in black. Moreover, for each response, we list the corresponding references. The indexes of the references cited in the manuscript correspond to Arabic number. The key revisions to the manuscript are summarized as follows:

1. First, we have supplemented the latest statistical data from the FBI Internet Crime Report [7], added a threat model, and conducted a detailed analysis of the attack methods of fraudulent websites as well as the necessity of researching them. For specific modification details, please refer to the responses to **Comments 2** and **3** in the major comments from **Reviewer #1**.
2. Second, we have reorganized and analyzed the latest key studies in the fields of phone fraud detection, phishing website detection, and incremental learning (e.g., [15], [34], and [47]). This addresses the issue that the literature review insufficiently covers important recent works. Meanwhile, we have conducted in-depth discussions on the newly added literature in the introduction and related work sections to strengthen the demonstration of research background and innovation. For specific modification details, please refer to the response to **Comment 4** in the major comments from **Reviewer #1** and the response to **Comment 7** in "**Comments on Introduction (Section I)** and **Related Work (Section II)**" from **Reviewer #2**.
3. Third, we have supplemented key details of the detection network. Specifically, we added formula descriptions in the problem definition part of the research content. Meanwhile, we have redrawn the structural diagram of the entire detection network and supplemented its key detection details. This fills the gap in the description of the detection network—where key details and mathematical modeling were missing—and makes the implementation logic of the model clearer and more complete. For specific modification details, please refer to the response to **Comment 7** in the major comments from **Reviewer #1** and the response to **Comment 5** from **Reviewer #2**.
4. Fourth, we have supplemented detailed statistical information about the dataset. A new dataset statistics table has been added, which clearly presents the number of samples, class distribution, types of features used, and feature extraction methods. Additionally, considering that the dataset includes 7 categories of fraudulent websites, we explained the process of using the "structure-growing" incremental learning strategy (each incremental task contains data from 2 categories) to simulate training for unknown attack strategies, so as to prove the generalization ability of the model for emerging fraud types. For specific modification details, please refer to the response to **Comment 7** in the major comments from **Reviewer #1** and the response to **Comment 5** from **Reviewer #2**.
5. Fifth, we have enhanced the experimental evaluation. Specifically, we introduced two new evaluation metrics—mean accuracy (mACC) and backward transfer (BWT)—to better measure the effectiveness of incremental detection. Furthermore, we supplemented comparative experiments against baseline methods from both telephone fraud detection and phishing website detection to demonstrate the broader applicability of our method. In addition, we conducted systematic comparisons between FD-SGIL and representative incremental learning baselines under the mACC and BWT metrics. Finally, we carried out ablation studies on the Effectiveness of FD-SGIL Components and the Impact of Multi-source Fused Features, which further validate the contributions of our design. For specific modification details, please refer to the responses to **Comments 3**, **4**, and **8** in the comments from **Reviewer #2**.
6. Besides, we have checked and double-checked the spelling, grammar and punctuation of the whole manuscript in the revised version. Meanwhile, we have checked the experiments for ensuring the correctness of the results, and updated some data. Last but not least, let us thank deeply AE and Reviewers for their constructive comments.

References:

[7] “2024 internet crime report,” Internet Crime Complaint Center, Federal Bureau of Investigation, USA, pp. 1–47, 2024.

[15] A. Tong, B. Chen, Z. Wang, J. Gao, and C. K. Lam, “Gdfgat: Graph attention network based on feature difference weight assignment for telecom fraud detection,” PLoS One, vol. 20, no. 5, p. e0322004, 2025. 18720–18729.

[34] F. Rashid, N. Ranaweera, B. Doyle, and S. Seneviratne, “Llms are one shot url classifiers and explainers,” Computer Networks, vol. 258, p. 111004, 2025.

[47] Z. Hu, Y. Li, J. Lyu, D. Gao, and N. Vasconcelos, “Dense network expansion for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11 858–11 867.

**Response to Major Comments from Reviewer #1**

Reviewer’s major comment 1:

The paper is outside of the scope of the journal.

We sincerely thank the reviewer for this important comment. Telecom fraud is not limited to traditional computer networks but also frequently occurs on mobile phones and smart devices, which are widely used in IoT ecosystems. When users access websites through browsers on these devices, they face the risk of being deceived. Thus, victims of telecom fraud are broadly distributed across different IoT scenarios, making fraudulent website detection highly relevant to IoT security.

In the revised version, we have clarified this connection by expanding the background discussion and explicitly linking fraudulent website detection with IoT-related risks and applications. With these improvements, the overall quality of the paper has been significantly enhanced.

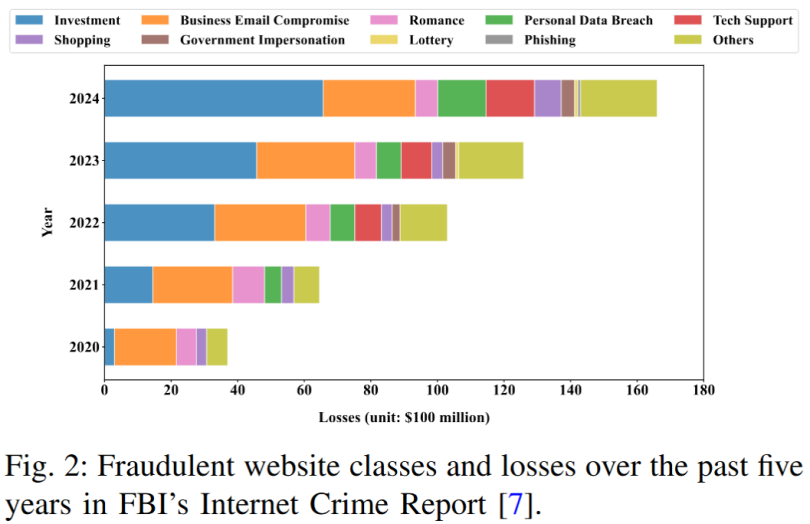
Considering the focus of the revised work, we have decided to submit it to IEEE Transactions on Information Forensics and Security (TIFS), which is more closely aligned with the scope of this study. We sincerely appreciate the reviewer’s feedback, which guided us to strengthen and improve the paper.

Reviewer’s major comment 2:

The introduction lacks proper definition of the research problem. A deeper exploration of the seriousness of the phishing URLs problem, along with some recent statistics could really help establish the need for this research.

We would like to sincerely thank the reviewer for the insightful suggestion. First, we have explicitly defined the research problem in the **section I. INTRODUCTION of Page 1**. Fraudulent websites are now clearly described as malicious web resources intentionally constructed to deceive users into disclosing sensitive information or transferring money. These fraudulent tactics extend beyond conventional phishing schemes, encompassing a broader spectrum of deceptive practices such as investment fraud and business email compromise. Furthermore, we emphasize that such fraudulent methodologies are not static but continuously emerging and evolving, which makes them particularly challenging to detect.

Second, we have reinforced the motivation by analyzing the seriousness of this threat in the **section I. INTRODUCTION of Pages 1~2**. We incorporated recent statistics from the FBI’s Internet Crime Report [7], which shows that the financial losses associated with fraudulent websites in the United States alone had exceeded 16 billion USD by 2024 (see Fig. 2 for illustration). To further illustrate the impact, we summarized the growing diversity of attack categories—including investment scams, phishing fraud, and shopping fraud—and highlighted their increasing economic and social harm. This evidence clearly demonstrates the urgent need for more effective detection techniques tailored to fraudulent websites.



References:

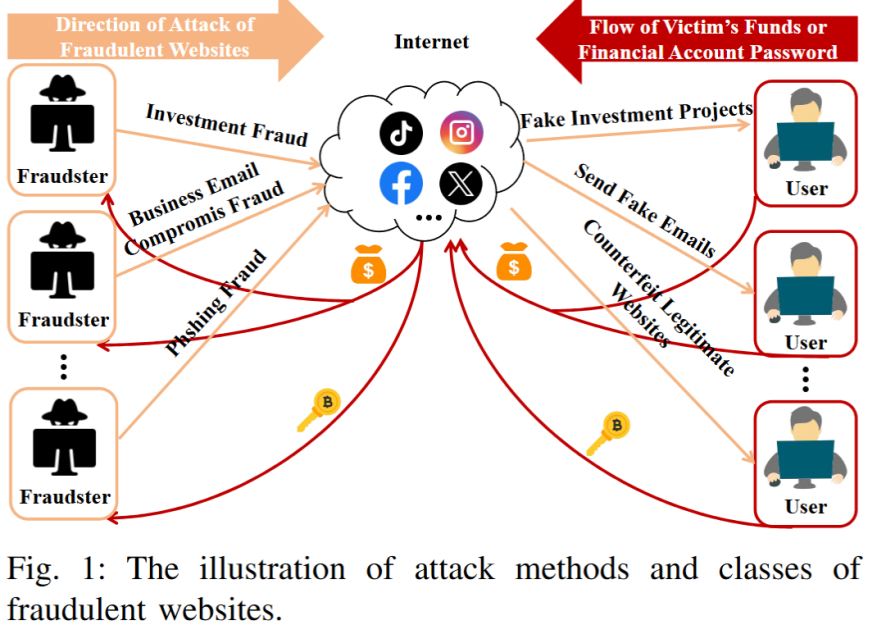
[7] “2024 internet crime report,” Internet Crime Complaint Center, Federal Bureau of Investigation, USA, pp. 1–47, 2024.

Reviewer’s major comment 3:

The paper needs to include a proper threat model either in the introduction of in a separate section.

We thank the reviewer very much for providing the instructive comment and valuable suggestion. In the introduction, we have organized and summarized the attack methods and types of fraudulent websites, and presented their threat model.

Fraudulent websites: Fraud is perpetrated by building fraudulent websites on the Internet and luring users into visiting them. As shown in Fig. 1, the fraudulent tactics employed by these websites not only encompass traditional phishing fraud but also include schemes such as investment fraud and romance fraud, with new fraud tactics continuously evolving.



The above modified content is placed in the section **I.INTRODUCTION of Page 1** in the revised version.

Reviewer’s major comment 4:

Literature review is missing a few important recent works that combine the WHOIS features with lexical features of the URL with higher accuracy and other performance metrics. This needs to be expanded.

We thank the reviewer’s helpful comment and valuable suggestion. We have reorganized the latest review studies relevant to this research, including those focusing on telephone fraud detection, phishing website detection, and incremental learning methods. Furthermore, we have conducted a detailed analysis of the shortcomings of existing review studies in addressing fraudulent website detection. Meanwhile, we have supplemented the Motivation to clarify the rationale behind this research.

III. MOTIVATION

Traditional research on telecom fraud detection has primarily focused on telephone fraud, with a lack of studies dedicated to fraudulent website detection. The data structure of fraudulent websites lacks inherent graph topological characteristics, making traditional GNN-based telephone fraud detection methods [4], [8], [12] difficult to apply directly. Existing phishing website detection methods [16], [48], [49] mainly rely on URL structures, assisted by HTML and WHOIS records for binary classification. The diverse classes and attack patterns of fraudulent websites make it difficult for traditional phishing website detection methods to be directly applicable. Most existing incremental learning methods [39], [41], [46], [47] are designed for balanced datasets, while the highly imbalanced characteristics of fraudulent website data exacerbate the model's catastrophic forgetting problem [50].

To this end, this paper proposes the FD-SGIL method. First, a multi-source feature fusion framework for URLs, webpage content, and WHOIS records is constructed based on BERT, which improves the accuracy of multi-classification detection of fraudulent websites. Second, an incremental learning strategy with dynamic structure adjustment is adopted to dynamically expand detection branches for unseen classes of fraudulent websites. Finally, a transfer loss function is designed to dynamically adjust the loss weights of misclassified samples according to sample distribution, effectively alleviating data imbalance and forgetting problems.

The above modified content is placed in section **Ⅱ. RELATED WORK of Pages 2~4 and** section **Ⅲ.MOTIVATION of Page 4** in the revised version.

References:

1. X. Hu, H. Chen, S. Liu, H. Jiang, G. Chu, and R. Li, “Btg: A bridge to graph machine learning in telecommunications fraud detection,” Future Generation Computer Systems, vol. 137, pp. 274–287, 2022.

[8] X. Hu, H. Chen, J. Zhang, H. Chen, S. Liu, X. Li, Y. Wang, and X. Xue, “Gat-cobo: Cost-sensitive graph neural network for telecom fraud detection,” IEEE Transactions on Big Data, vol. 10, no. 4, pp. 528–542, 2024.

[12] X. Hu, H. Chen, H. Chen, S. Zhang, S. Liu, and X. Li, “Telecom fraud detection via imbalanced graph learning,” in 2022 IEEE 22nd International Conference on Communication Technology (ICCT). IEEE, 2022, pp. 1312–1317.

[16] W. Wei, Q. Ke, J. Nowak, M. Korytkowski, R. Scherer, and M. Wo´zniak, “Accurate and fast url phishing detector: a convolutional neural network approach,” Computer Networks, vol. 178, p. 107275, 2020.

[39] Z. Li and D. Hoiem, “Learning without forgetting,” IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 12, pp. 2935–2947, 2017.

[41] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “icarl: Incremental classifier and representation learning,” in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2017, pp. 2001–2010.

[46] S. Yan, J. Xie, and X. He, “Der: Dynamically expandable representation for class incremental learning,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 3014–3023.

[47] Z. Hu, Y. Li, J. Lyu, D. Gao, and N. Vasconcelos, “Dense network expansion for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 11 858–11 867.

[48] B. Sabir, M. A. Babar, R. Gaire, and A. Abuadbba, “Reliability and robustness analysis of machine learning based phishing url detectors,” IEEE Transactions on Dependable and Secure Computing, 2022.

[49] A. Aljofey, Q. Jiang, A. Rasool, H. Chen, W. Liu, Q. Qu, and Y. Wang, “An effective detection approach for phishing websites using url and html features,” Scientific Reports, vol. 12, no. 1, p. 8842, 2022.

[50] Y. Yang, D. Ren, C. Peng, J. Huo, W. Li, and Y. Gao, “Dynamic replay training for class-incremental learning,” in ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024, pp. 5915–5919.

Reviewer’s major comment 5:

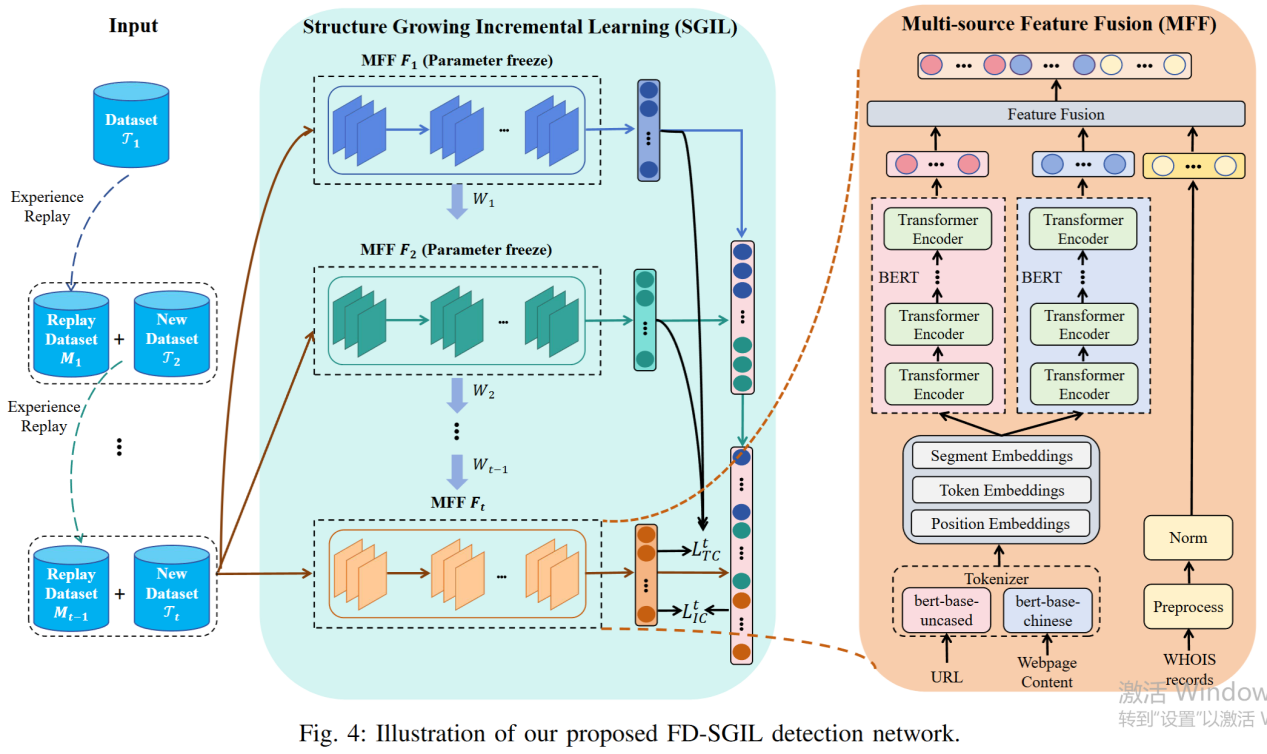
Mentioning specific countries with "weak legal oversight" without proper support of these claims is not scientifically correct. Why mention countries? The point could've been delivered without mentioning country names.

We would like to thank the reviewer very much for providing this helpful comment. We have realized that it is highly inappropriate to mention specific countries as having "weak legal oversight," so we have removed all content referring to specific countries labeled as such.

Reviewer’s major comment 6:

The description of the detection network is missing some important details. The description and mathematical modeling needs to be expanded.

We would like to thank the reviewer very much for providing this helpful comment. We have supplemented the key details of the detection network. Specifically, we have added formula descriptions in the problem definition. Meanwhile, Meanwhile, we have redrawn the structural diagram of the entire detection network and supplemented its key detection details.



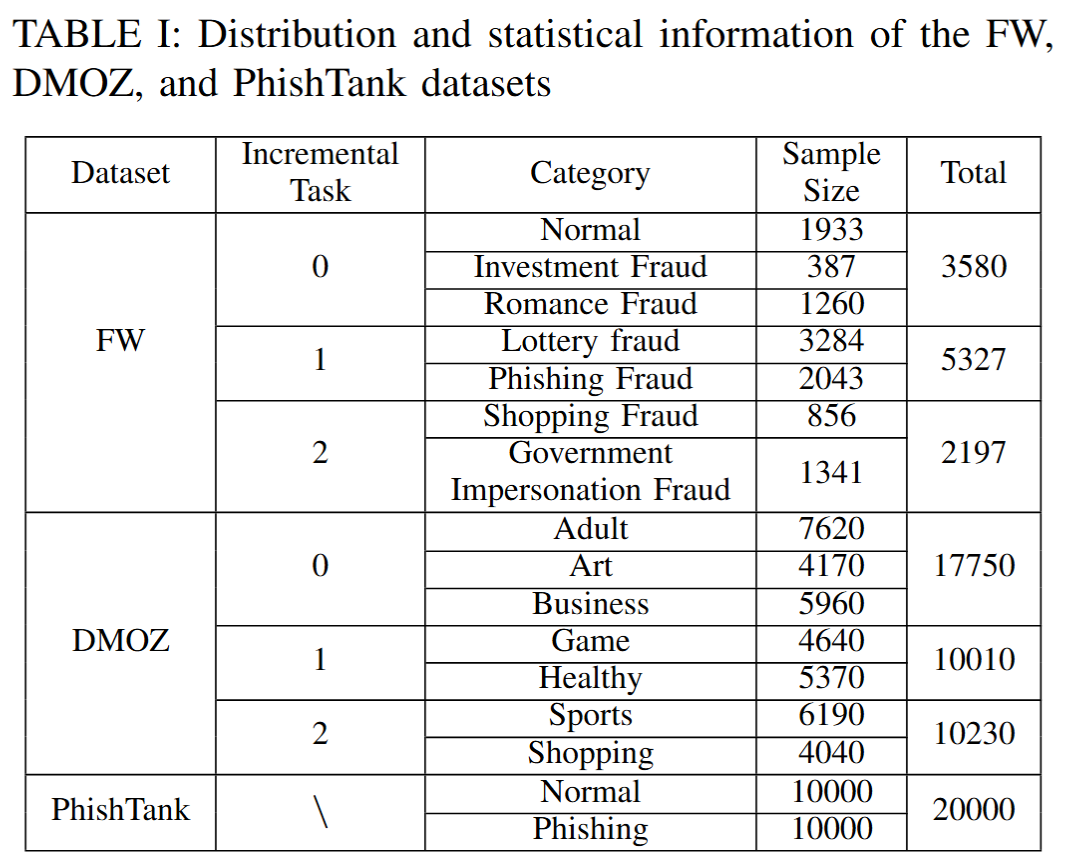
The above modified content is placed in the section **Ⅳ.METHOD of Pages 4~7** in the revised version.

Reviewer’s major comment 7:

In section IV.A there is absolutely no explanation of the features included in these datasets. No discussion of class balance or imbalance. This significantly undermines the validity of the obtained results. Did all of these datasets have only URLs? or complete featuresets? Did they contain the same features? How were these features extracted?

We would like to thank the reviewer for this constructive comment. In the revised manuscript, we have provided a detailed introduction of the three datasets, including the types of features contained in each dataset as well as the statistical distribution of the samples, as summarized in Table 1. We have also explained the feature extraction process in detail. To address class imbalance, we designed a gradient-based distillation loss, which effectively alleviates this problem.

Specifically, the PhishTank and DMOZ datasets contain only URL features, from which we extract URL representations using the BERT model. In contrast, the FW dataset includes URLs, webpage content, and WHOIS records, providing richer multi-source features. Therefore, the proposed multi-source feature fusion method (MFF) is mainly evaluated on the FW dataset.



The above modified content is placed in subsection **C. Multi-source Feature Fusion Method of Pages 5~6** and subsection **A. Experimental Setup of Page 9** in the revised version.

Reviewer’s major comment 8:

The results obtained are not validated. I'd suggest aproper validation method such as 10-fold cross-validation.

We thank the reviewer very much for this helpful comment. In the revised manuscript, we have adopted 10-fold cross-validation to ensure a more rigorous and reliable validation of the obtained results. Furthermore, in the comparative experiments with incremental learning baselines, we follow common practice in the literature by reporting two widely used evaluation metrics: mean accuracy (mACC), which measures the average performance across all previously learned tasks, and backward transfer (BWT), which assesses the influence of learning new tasks on the performance of previously learned tasks [55]. These evaluation settings are consistent with prior incremental learning studies and provide a comprehensive validation of our proposed method.

Reference:

[55] D. Lopez-Paz and M. Ranzato, “Gradient episodic memory for continual learning,” Advances in Neural Information Processing Systems, vol. 30, pp. 6470–6479, 2017.

**Response to Minor Comments from Reviewer #1**

Reviewer’s minor comment 1:

The language of the paper needs intensive review to improve readability and eliminate grammatical errors.

We would like to thank the reviewer very much for providing this helpful comment. We have carefully revised the manuscript to improve readability and eliminate grammatical errors throughout the paper.

Reviewer’s minor comment 2:

The first time the term is mentioned, it needs to be in expanded form. For example, the first work in the introduction is IoT without the long form. The same applies to many acronyms mentioned in the paper across the sections.

We would like to thank the reviewer very much for providing this helpful comment. We have ensured that all acronyms are expanded at their first occurrence across the manuscript, and consistently applied this practice in all sections.

Reviewer’s minor comment 3:

Figure 2 is too small to be legible. This is the core of the presented work. It needs to be given adequate space.

We would like to thank the reviewer very much for providing this helpful comment. We have enlarged Figure 2 in the revised manuscript and adjusted its layout to occupy adequate space, making all details clearly legible. Last but not least, let us thank deeply **Reviewer #1** for his/her constructive reviews.

**Response to Comments from Reviewer #2**

Reviewer’s comment 1:

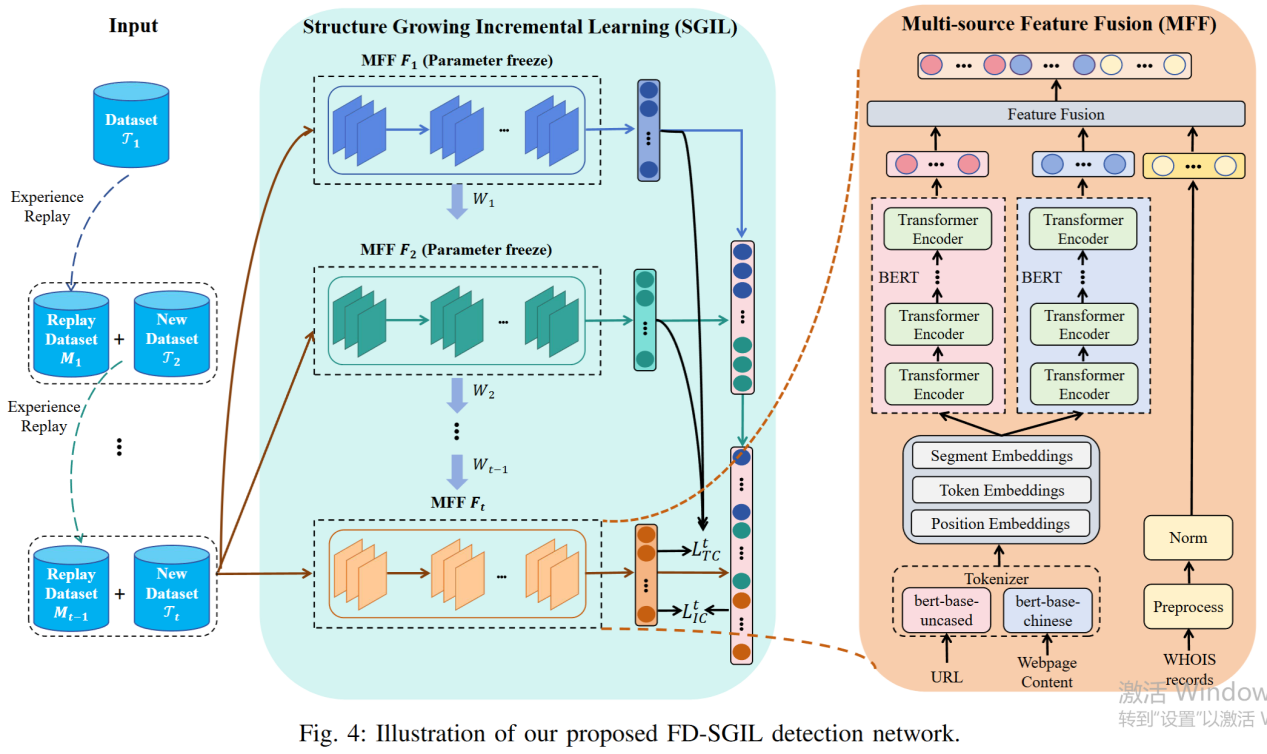
Dependence on BERT for Feature Extraction BERT, while powerful, is computationally expensive and may not be optimal for real-time detection on resource-constrained IoT devices. This could limit deployment scalability, especially for edge devices with limited processing power.

We thank the reviewer very much for this helpful comment. BERT is indeed computationally expensive, but it has strong representation ability. Our experiments demonstrate that the features extracted by BERT significantly enhance the detection performance of classifiers. For practical deployment, we envision placing BERT on the server side for feature extraction and classification, and then transmitting only the detection results to IoT devices, rather than running BERT directly on resource-constrained IoT devices.

Reviewer’s comment 2:

Structure-Growing Incremental Learning Complexity The "structure-growing" mechanism may introduce additional hyperparameters and architectural complexity. This could make the model harder to tune and deploy in practice, requiring more expertise to maintain.

We would like to thank the reviewer very much for providing this helpful comment. The structure growing incremental learning mechanism in our framework is carefully designed to balance adaptability and practicality. Considering the dynamic and evolving nature of fraudulent websites, the model must continuously adapt to new attack types while maintaining performance on previously seen ones. The SGIL strategy achieves this by selectively expanding its structure when encountering new fraud patterns, which helps preserve knowledge of earlier tasks and mitigates catastrophic forgetting. To facilitate deployment and reduce tuning difficulty, we have redrawn the overall framework diagram of the FD-SGIL detection method (see Fig. 4) and documented all hyperparameter settings and training configurations in the revised manuscript, ensuring clarity, reproducibility, and ease of maintenance.

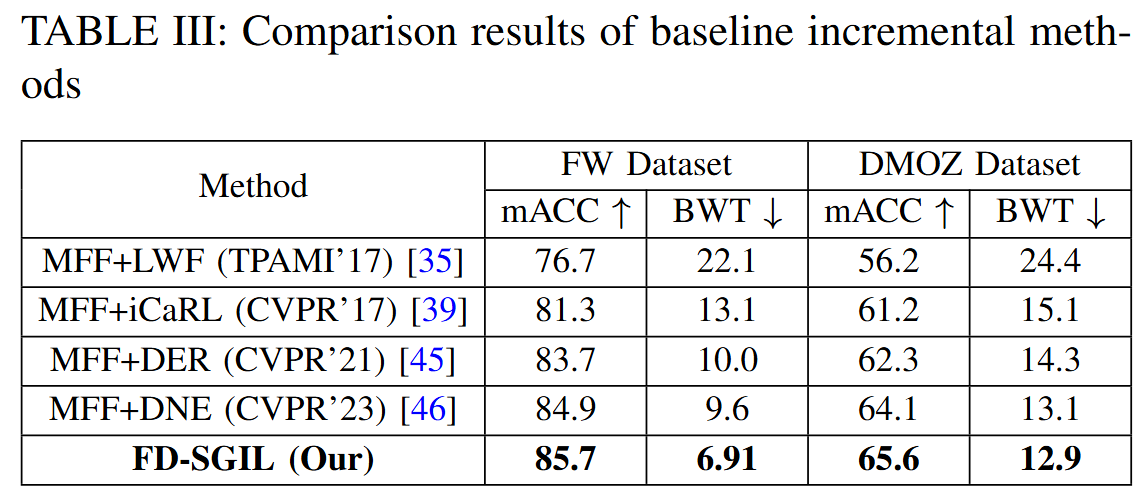


The above modified content is placed in subsection **C. Multi-source Feature Fusion Method of Pages 5~6**，subsection **D. Structure Growing Incremental Learning Method of Pages 6~7**，and **subsection A. Experimental Setup of Page 10** in the revised version.

Reviewer’s comment 3:

Catastrophic Forgetting Mitigation Claims While the paper mentions a "transfer loss mechanism," incremental learning methods often struggle with severe forgetting when new classes are drastically different from old ones. If the new fraudulent website types are highly dissimilar to previous ones, performance degradation may still occur.

We thank the reviewer very much for this helpful comment. In incremental learning, catastrophic forgetting is a common challenge: when a model is trained on new tasks that are highly dissimilar to previous ones, it tends to prioritize fitting the new data, which can degrade performance on earlier tasks. In our framework, the dynamically weighted transfer loss combined with experience replay is specifically designed to address this issue, enabling the model to learn new classes while preserving knowledge of previously learned classes. Empirical results support the effectiveness of this approach: as shown in Table III, FD-SGIL consistently achieves the highest overall performance, obtaining optimal final mACC and BWT metrics while maintaining stable mACC values across all incremental tasks.



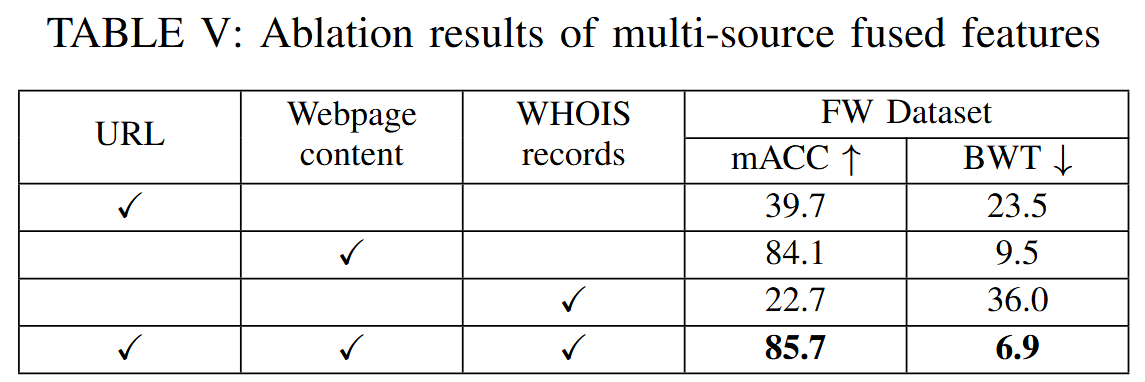
The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.

Reviewer’s comment 4:​

Multi-Source Feature Fusion Overhead Combining URL, webpage content, and WHOIS records requires significant data preprocessing and alignment. This may slow down real-time detection, particularly if WHOIS data retrieval is delayed or unavailable.

We thank the reviewer very much for this helpful comment. In our model, the most time-consuming stage is the training process, mainly due to the large number of samples and iterations. Once the model is trained, however, applying it to new samples requires only feature representation and classifier prediction, which can output detection results efficiently in real-time.

Regarding WHOIS records, its retrieval can be delayed or even fail due to network conditions. We have considered this scenario and conducted an ablation study to evaluate the impact of missing WHOIS features. As shown in Table V, even when WHOIS data is unavailable, our detection method still achieves competitive and robust performance. This demonstrates that the framework remains effective under practical conditions where WHOIS retrieval may be delayed or incomplete.



The above modified content is placed in the subsubsection **2). Impact of Multi-source Fused Features of Pages 11~12** in the revised version.

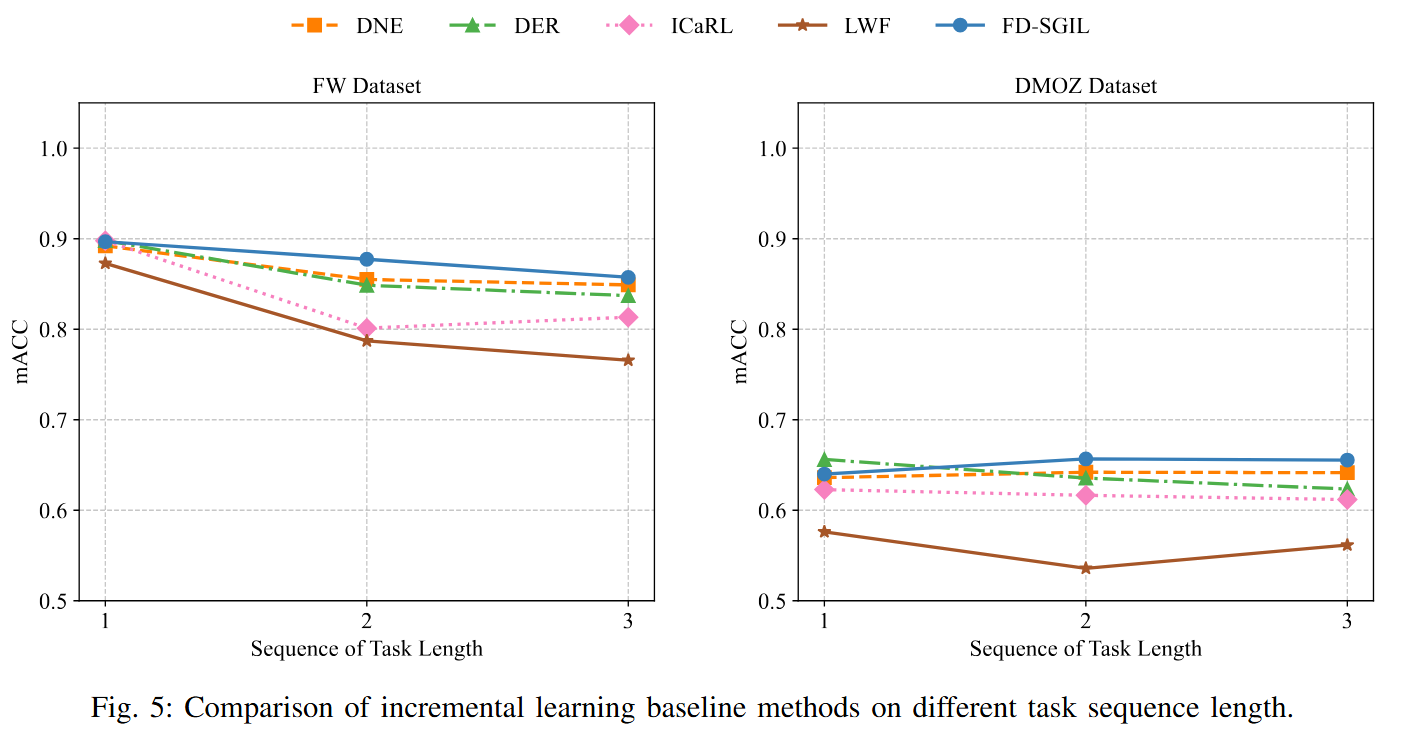
Reviewer’s comment 5:​

Dataset Generalizability The experiments are conducted on only three datasets, which may not cover all emerging fraudulent website types. The model might not generalize well to unseen or rapidly evolving attack strategies.

We thank the reviewer very much for this helpful comment. In our experiments, three datasets were used, among which the FW dataset is directly related to telecom fraud websites. To the best of our knowledge, after surveying the existing literature and available cybersecurity research resources, we did not find any publicly available dataset dedicated to telecom fraud detection. The FW dataset was therefore constructed based on real-world data provided by China Mobile, covering six major types of telecom fraud: investment fraud, romance fraud, lottery fraud, phishing fraud, shopping fraud, and government impersonation. According to the 2024 Internet Crime Report of the FBI [7], these six types already encompass the majority of the seven most common global fraud categories. Other less frequent types, such as SIM swap fraud and real estate fraud, could not be included due to the limited number of samples, which makes them unsuitable for training effective detection models.

The key motivation for our research is precisely to address the continuously emerging nature of telecom fraud. By adopting an incremental learning framework, the model can be retrained with new data as soon as sufficient samples of novel fraud types are collected, thereby retaining strong recognition capability for both previously learned and newly introduced categories.

As shown in Fig. 5, we simulated this scenario by first training the model with three fraud types, and then incrementally adding two new types in each subsequent training phase.The results demonstrate that FD-SGIL can be effectively retrained with new data while maintaining high detection performance on previously learned categories, confirming its suitability for handling diverse and evolving telecom fraud strategies.



The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison Pages 9~10** in the revised version.

References:

[7] “2024 internet crime report,” Internet Crime Complaint Center, Federal Bureau of Investigation, USA, pp. 1–47, 2024.

Reviewer’s comment 6:​

Label Scarcity in Incremental Learning Fraudulent websites evolve quickly, but obtaining labeled data for new classes in real-world scenarios is challenging. The incremental learning performance may degrade if few labeled examples are available for new attack types.

We would like to thank the reviewer very much for this insightful comment. Our research on incremental learning is precisely motivated by the difficulty of obtaining sufficient labeled data for newly emerging fraud types. When new categories appear, we first collect enough labeled samples. These samples are then used for incremental training, allowing the model to maintain high detection capability on both original and new categories.

In real-world scenarios, new fraud types may have only a few labeled samples. In such cases, oversampling can be used to expand both the quantity and diversity of the data before incremental training. This ensures the model’s recognition ability for such classes.

Moreover, our proposed FD-SGIL framework has already considered the data imbalance problem inherent in incremental learning. Specifically, we introduce a transfer loss that selects more suitable soft labels for each training sample in the distillation process, thereby alleviating catastrophic forgetting. In addition, the loss weight of misclassified samples is adaptively increased while that of correctly classified samples is decreased, which further mitigates the impact of imbalanced data during training.

Reviewer’s comment 7:​

Real-World Deployment Feasibility The abstract does not discuss computational latency or model size, which are critical for IoT deployments. If the model is too large or slow, it may not be practical for real-time use in IoT gateways.

We would like to thank the reviewer very much for providing this helpful comment. In this work, our primary focus is to design FD-SGIL as an effective framework for accurately detecting fraudulent websites and adapting to newly emerging fraud types. For real-world deployment, we envision a client–server architecture. Specifically, IoT devices such as smartphones or PCs act as clients that collect suspicious URLs with very low resource consumption. These URLs are then transmitted to a high-performance server, where the FD-SGIL model performs feature extraction, representation learning, fusion, and detection. The detection results are subsequently sent back to the client side, enabling real-time blocking of fraudulent websites.

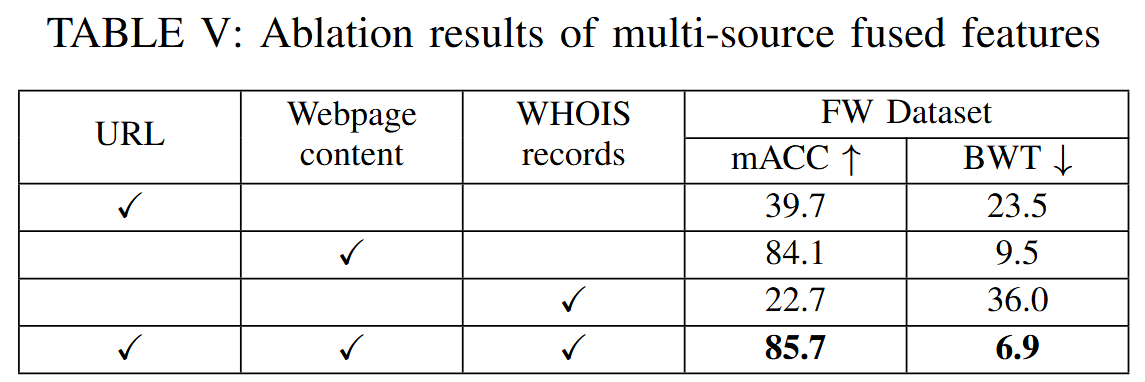
In addition, to further reduce latency on the client side, we maintain a blacklist of known fraudulent URLs. Each URL visited in the browser is first compared against this blacklist. If no match is found, the URL is then sent to the FD-SGIL server for remote detection. If classified as fraudulent, the client immediately blocks access. This design balances detection accuracy with practical deployment feasibility in IoT environments.

Reviewer’s comment 8:​

Adversarial Robustness Fraudulent websites may use adversarial techniques (e.g., URL obfuscation, dynamic content) to evade detection. The paper does not mention robustness testing against adversarial evasion attacks.

We would like to thank the reviewer very much for providing this helpful comment. Our analysis of large numbers of fraudulent websites shows that even under such obfuscation, websites within the same fraud type still exhibit strong similarities. For example, in gambling fraud websites, we observed that certain URL segments remain stable while others change in similar patterns. For instance, in two gambling websites, the prefixes such as “http://vip-xxx.com/” or “http://game-xxx.com/” were consistent, while only the suffix strings varied slightly (e.g., “/abc123” and “/xyz456”). At the same time, the webpage content, such as banners, betting tables, and login forms, as well as WHOIS registration information, were nearly identical. In other cases, such as investment fraud, the URLs and page structures may differ significantly (e.g., “http://finance-portal.net/” and “http://invest-safe.org/”), but the textual semantics in webpage content remain highly similar, both emphasizing guaranteed high returns, urgent investment opportunities, and contact via messaging apps. These observations indicate that despite adversarial obfuscation, shared patterns across multi-source features can still be leveraged for reliable detection.

Based on these observations, we proposed a multi-source feature fusion (MFF) strategy in FD-SGIL to address adversarial variations. By jointly leveraging URL, webpage content, and WHOIS features, the framework can still detect fraud effectively even when one feature source is intentionally manipulated. The ablation study on the impact of multi-source features (see Table Ⅴ) further confirms that FD-SGIL maintains strong detection performance even when certain features (e.g., WHOIS) are unavailable. This demonstrates that the framework inherently provides a degree of adaptability against adversarial techniques.



The above modified content is placed in the subsubsection **2). Impact of Multi-source Fused Features of Pages 11~12** in the revised version.

**Response to Comments from Reviewer #2 on the Introduction (Section I) and Related Work (Section II)**

Reviewer’s comment 1:

Overemphasis on Incremental Learning Without Addressing Core Challenges

While the paper highlights incremental learning (IL) as a solution for real-time detection, it does not sufficiently address: Class Imbalance: New fraudulent website types may have scarce labeled data, making IL unstable. Concept Drift: Fraudulent websites evolve dynamically (e.g., adversarial URL obfuscation), but the proposed "structure-growing" mechanism may not adapt quickly enough. Impact: The model might fail in real-world scenarios where new attack types emerge rapidly with limited labeled examples.

We would like to thank the reviewer very much for providing this helpful comment. We thank the reviewer very much for this helpful comment. Regarding class imbalance, FD-SGIL dynamically adjusts the loss weights of misclassified versus correctly classified samples according to the data distribution. This ensures that minority classes receive more attention during training, effectively mitigating the instability caused by scarce labeled data for newly emerging fraudulent website types.

Regarding concept drift, we combine experience replay with a transfer loss function to guide newly expanded branches in the structure-growing mechanism. This design helps the model adapt to new attack types while preserving knowledge of previously learned ones, reducing the risk of performance degradation when fraudulent websites evolve dynamically.

Reviewer’s comment 2:

Heavy Reliance on BERT for Multi-Source Feature Fusion BERT is used for URL, webpage content, and WHOIS record fusion, but: Computational Cost: BERT is resource-intensive, conflicting with the goal of lightweight real-time detection on IoT devices. Feature Alignment: No clear method is described for aligning heterogeneous features (e.g., textual content vs. structured WHOIS data), which could lead to suboptimal fusion. Impact: High latency and energy consumption, making deployment on edge devices impractical.

We would like to thank the reviewer very much for providing this helpful comment. Regarding computational cost, we envision a client–server architecture, where IoT devices collect URLs and transmit them to a high-performance server running FD-SGIL for feature extraction, fusion, and detection. The detection results are then returned to the client side, ensuring scalability and avoiding heavy computation on edge devices.

Regarding feature alignment, BERT is used only for encoding URL and webpage content, while WHOIS records are encoded through a lightweight feature mapping. The three representations are then concatenated for fusion, ensuring that heterogeneous features are properly integrated into a unified representation for classification. This design provides strong feature representations with only minor computational overhead compared to baseline methods, keeping inference practical for real-time detection.

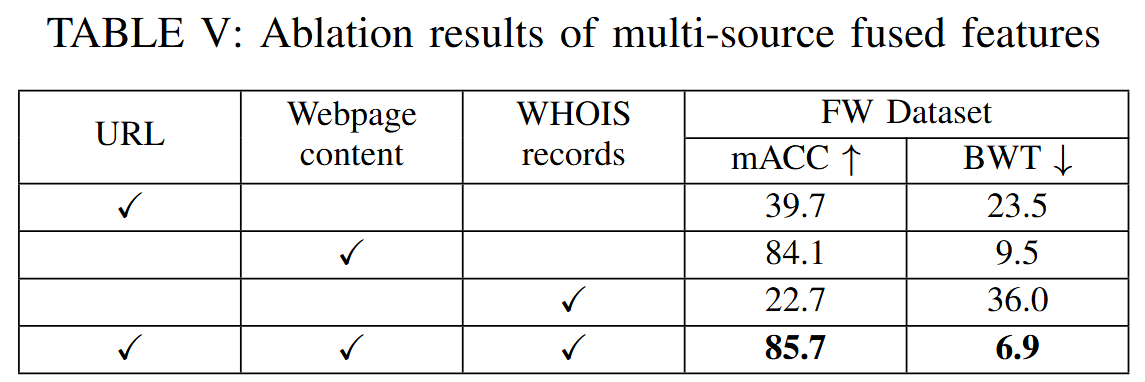
Reviewer’s comment 3:

Lack of Adversarial Robustness Considerations Fraudulent websites actively evade detection via:

Dynamic Content: AIGC-generated text or adversarial perturbations in URLs/webpages. WHOIS Spoofing: Fake registration data to bypass WHOIS-based checks. Impact: The paper does not evaluate robustness against adversarial evasion techniques, risking high false negatives in practice.

We thank the reviewer very much for this helpful comment. Regarding dynamic content and WHOIS spoofing, our analysis of real-world fraudulent websites shows that websites within the same fraud type still share stable patterns. For example, in gambling fraud websites, certain URL prefixes remain constant while suffixes vary slightly, and webpage content often remains identical. In investment fraud websites, URLs and layouts may differ, but textual semantics remain highly similar. These patterns allow FD-SGIL to capture discriminative features despite adversarial variations.

FD-SGIL mitigates reliance on any single feature source by using a multi-source feature fusion strategy. As shown in Table V, the ablation study demonstrates that FD-SGIL maintains strong detection performance even when certain features (e.g., WHOIS) are unavailable, indicating robustness to incomplete or manipulated inputs.



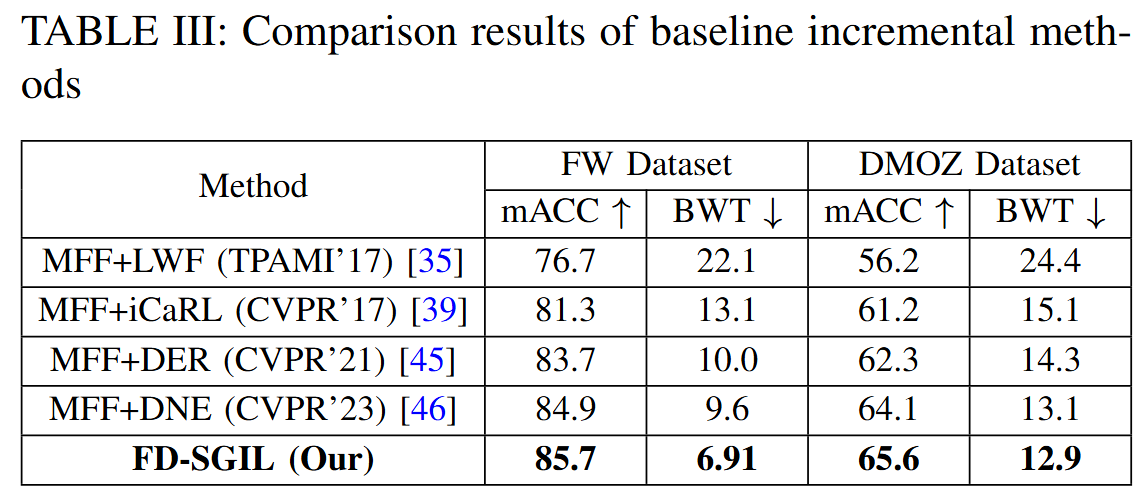
The above modified content is placed in the subsubsection **2). Impact of Multi-source Fused Features of Pages 11~12** in the revised version.

Reviewer’s comment 4:

Shallow Treatment of Multi-Class Detection The paper claims superior multi-class performance but: No Discussion of Class Overlap: Many fraudulent types (e.g., phishing vs. scam) share similar features, risking misclassification. Limited Dataset Diversity: Only three datasets are used, likely lacking coverage of rare or emerging fraud types. Impact: Overstated generalization claims; the model may struggle with fine-grained classification.

We would like to thank the reviewer very much for this helpful comment. Regarding class overlap, FD-SGIL addresses this challenge through a dynamically weighted transfer loss combined with experience replay. This design helps the model preserve knowledge of previously learned classes while learning new categories, thereby reducing misclassification among similar types.

Regarding dataset diversity, we constructed the FW dataset covering six major telecom fraud categories (investment, romance, lottery, phishing, shopping, and government impersonation). According to the 2024 FBI Internet Crime Report [7], these categories encompass the majority of common global fraud types. As shown in Table III, FD-SGIL consistently achieves superior overall performance, maintaining stable mACC values across incremental tasks. For categories not included in this study, FD-SGIL can seamlessly incorporate them through incremental training in future work.



The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.

References:

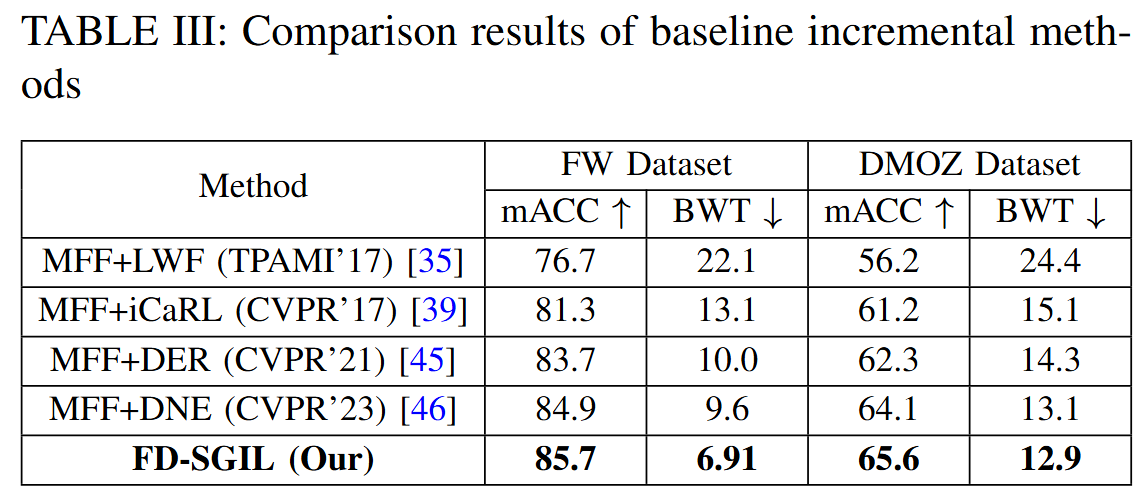
[7] “2024 internet crime report,” Internet Crime Complaint Center, Federal Bureau of Investigation, USA, pp. 1–47, 2024.

Reviewer’s comment 5:

Incremental Learning Limitations Underplayed The proposed "transfer loss" and "dynamic weight adjustment" mechanisms are not compared rigorously to state-of-the-art (SOTA) IL methods like: Replay-Based Methods: Experience replay (e.g., [30]) is more proven but may be omitted due to memory constraints. Meta-Learning: Could better handle few-shot new classes but is not explored. Impact: Catastrophic forgetting may still occur for long-tailed or dissimilar new classes.

We would like to thank the reviewer very much for this helpful comment. We thank the reviewer very much for this helpful comment. Regarding comparison with existing methods, we categorized prior approaches into regularization-based, experience replay-based, and dynamic structure adjustment-based methods [34]. FD-SGIL was compared against representative baselines from each family, namely LWF (TPAMI'17) [35], iCaRL (CVPR'17) [39], DER (CVPR'21) [45], and DNE (CVPR'23) [46]. As shown in Table III, FD-SGIL consistently outperforms these baselines in terms of mACC and BWT, demonstrating its ability to balance stability and plasticity.

Regarding meta-learning, meta-learning approaches are typically used for few-shot tasks. Our focus, however, is on long-term incremental detection of evolving fraudulent website categories. In this setting, our structure growing design combined with transfer loss and experience replay is more suitable and effective.



The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.

Reference:

[34] G. M. Van de Ven, T. Tuytelaars, and A. S. Tolias, “Three types of incremental learning,” Nature Machine Intelligence, vol. 4, no. 12, pp.1185–1197, 2022.

[35] Z. Li and D. Hoiem, “Learning without forgetting,” IEEE Transactionson Pattern Analysis and Machine Intelligence, vol. 40, no. 12, pp. 2935–2947, 2017.

[39] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “icarl: Incremental classifier and representation learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017, pp. 2001–2010.

[45] S. Yan, J. Xie, and X. He, “Der: Dynamically expandable representation for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3014–3023.

[46] Z. Hu, Y. Li, J. Lyu, D. Gao, and N. Vasconcelos, “Dense network expansion for class incremental learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp.11 858–11 867.

Reviewer’s comment 6:

Ignoring Real-World Deployment Constraints Critical IoT deployment challenges are overlooked:

Latency Requirements: No metrics on inference speed for real-time detection. Data Privacy: WHOIS records may contain sensitive user data; no privacy-preserving feature extraction is proposed. Impact: The framework may be unusable in privacy-sensitive or low-latency scenarios.

We thank the reviewer very much for this helpful comment. Regarding latency, URL and webpage content are encoded with BERT, while WHOIS features are processed using a lightweight mapping module. The fused representation introduces only minor overhead relative to baselines, allowing efficient real-time detection. To reduce latency in practice, we also envision a client–server deployment, where IoT devices perform lightweight URL collection and a server executes FD-SGIL for detection.

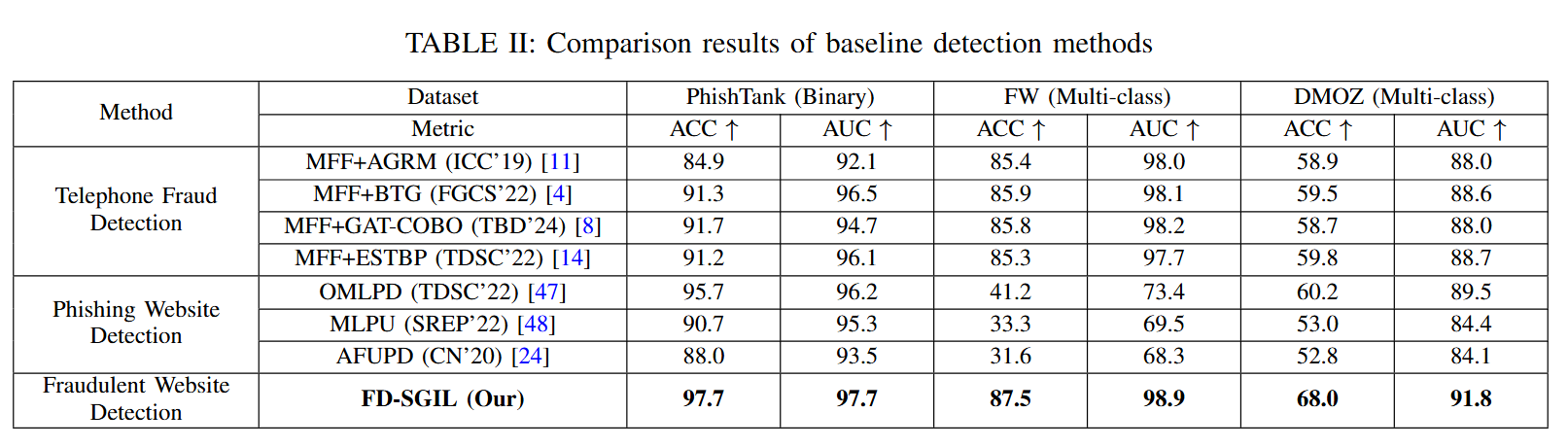
Regarding privacy, We accessed WHOIS information solely to obtain domain-specific details such as the domain's expiration date and geographic location of its registrant. Importantly, this process does not involve or risk the disclosure of any sensitive user data.

Reviewer’s comment 7:

Weak Baseline Comparison in Related Work The literature review (Section II) critiques non-ML and ML methods but: Superficial Analysis of Deep Learning Gaps: Fails to acknowledge hybrid approaches (e.g., GNNs for URL graphs) that could outperform BERT. No Mention of Zero-Shot Learning: For entirely new fraud types, zero-shot detection might be more viable than IL. Impact: The proposed method’s novelty is overstated relative to unexplored alternatives.

We would like to thank the reviewer very much for this helpful comment. Regarding deep learning gaps, we have expanded the literature review to include recent graph-based approaches such as AGRM [11], BTG [4], GAT-COBO [8], and ESTBP [14], which apply multi-source fusion for fraud detection. As shown in Table II, FD-SGIL outperforms these methods by achieving the highest ACC and AUC values.

Regarding zero-shot learning, while promising for unseen categories, it often relies on strong semantic priors or auxiliary information, which may not always be available. In contrast, FD-SGIL focuses on incremental learning, where new labeled data are gradually collected and integrated into the system without retraining from scratch, reflecting real-world fraud detection conditions.



The above modified content is placed in section **Ⅱ. RELATED WORK of Pages 2~4** and subsection **B. Comparison to Baseline Detection Methods of Page 9** in the revised version.

Reference:

[4] X. Hu, H. Chen, S. Liu, H. Jiang, G. Chu, and R. Li, “Btg: A bridge to graph machine learning in telecommunications fraud detection,” Future Generation Computer Systems, vol. 137, pp. 274–287, 2022.

[8] X. Hu, H. Chen, J. Zhang, H. Chen, S. Liu, X. Li, Y. Wang, and X. Xue, “Gat-cobo: Cost-sensitive graph neural network for telecom fraud detection,” IEEE Transactions on Big Data, vol. 10, no. 4, pp. 528–542, 2024.

[11] M. Liu, J. Liao, J. Wang, and Q. Qi, “Agrm: Attention-based graph representation model for telecom fraud detection,” in ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019, pp. 1–6.

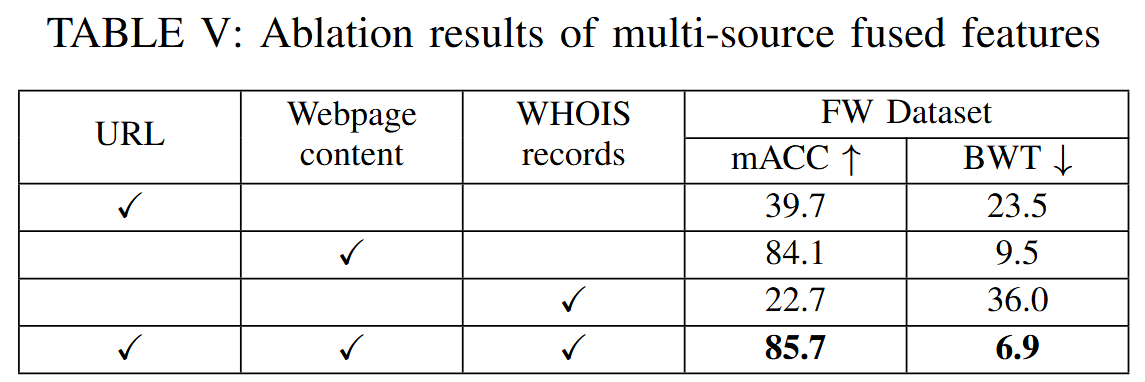
[14] G. Chu, J. Wang, Q. Qi, H. Sun, S. Tao, H. Yang, J. Liao, and Z. Han, “Exploiting spatial-temporal behavior patterns for fraud detection in telecom networks,” IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 6, pp. 4564–4577, 2022.

Reviewer’s comment 8:

Ethical and Legal Risks The paper does not discuss: Bias in WHOIS Data: Regional biases (e.g., flagging sites from "weak oversight" countries) could lead to discriminatory false positives. Misuse Potential: The model could be weaponized to flag legitimate sites as fraudulent. Impact: Unintended harm to businesses or individuals in developing regions.

We would like to thank the reviewer very much for this insightful comment. We thank the reviewer very much for this insightful comment. Regarding bias in WHOIS data, the WHOIS records in our study were obtained from real-world fraudulent websites, which ensures realistic scenarios. Furthermore, as presented in Table V, the WHOIS features have provided a certain degree of enhancement to our multi-source fused features.

Regarding misuse potential, FD-SGIL mitigates dependency on any single feature type through multi-source fusion. As shown in Table V, ablation results confirm that FD-SGIL maintains strong performance even when certain features (e.g., WHOIS) are unavailable, which reduces the risk of biased or discriminatory outcomes.



The above modified content is placed in the subsubsection **2). Impact of Multi-source Fused Features of Pages 11~12** in the revised version.

**Response to Comments from Reviewer #2 on the Introduction (Section VI)**

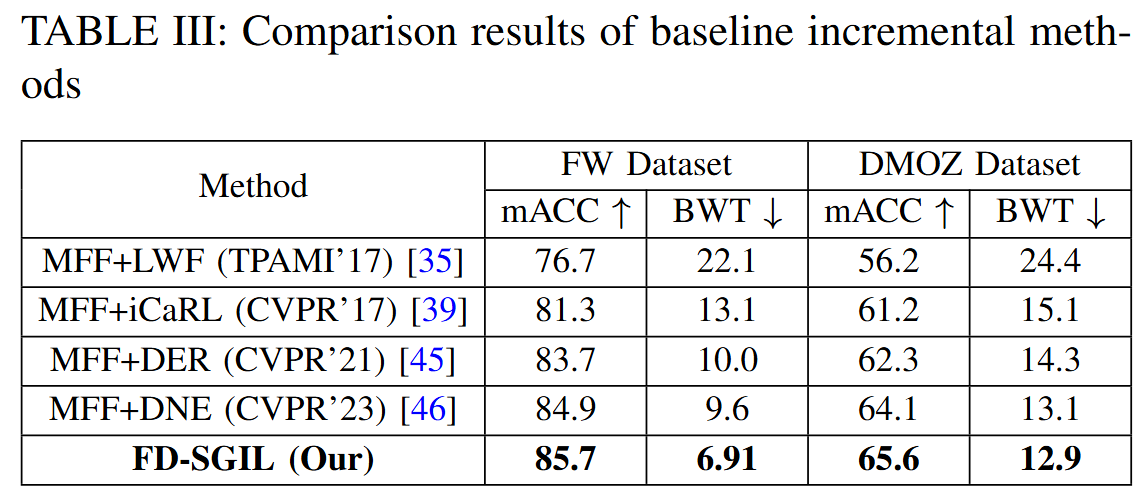
Reviewer’s comment 1:

Overstated Claims Without Caveats

Issue: The conclusion claims FD-SGIL "excels" in multi-class detection and catastrophic forgetting but omits: Limitations in the experiments: No discussion of failure cases (e.g., specific fraud types where performance drops). Real-world constraints: Computational costs of BERT or latency on edge devices are ignored.

We would like to thank the reviewer very much for this insightful comment. In incremental learning, catastrophic forgetting is a prevalent challenge: when a model is trained on new tasks that are highly dissimilar to previous ones, it tends to prioritize fitting the new data, thereby degrading its performance on earlier tasks. To better evaluate a model’s ability to alleviate such forgetting, researchers have proposed the metric of backward transfer (BWT). BWT measures the change in detection accuracy of all old tasks, before and after the model learns new tasks. A smaller BWT value means the model forgets less about old tasks. As can be seen from the comparison results with incremental learning baseline methods in Table III, FD-SGIL achieves the optimal performance under the same experimental conditions, further confirming its effectiveness in balancing new fraud detection with knowledge retention.

Regarding computational cost and latency, URL and webpage content are encoded with BERT, while WHOIS features are processed using a lightweight mapping module. The fused representation introduces only minor overhead relative to baselines, allowing efficient real-time detection. To reduce latency in practice, we also envision a client–server deployment, where IoT devices perform lightweight URL collection and a server executes FD-SGIL for detection.



The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.

Reviewer’s comment 2:

Superficial Future Work

Issue: Future directions are generic (e.g., "more superior structures") and lack specificity. Example: No mention of adversarial robustness (e.g., testing against AIGC-generated phishing sites) or privacy-aware training (critical for WHOIS data).

We thank the reviewer very much for this insightful comment. We highlight the importance of evaluating FD-SGIL against emerging threats such as AIGC-generated phishing sites, voice cloning, and deepfake-based scams. Furthermore, future work will explore novel multimodal indicators (e.g., deepfake traces in media, logical inconsistencies in user interactions) and investigate the potential of large language models (LLMs) for fraudulent website detection.

The above modified content is placed in the section **Ⅵ. CONCLUSION of Page 12** in the revised version.

Reviewer’s comment 3:

Ignored Ethical and Deployment Challenges

Issue: No reflection on: Bias risks: WHOIS-based features may disproportionately flag sites from certain regions. Scalability: Deploying BERT-based models on IoT devices remains impractical.

We would like to thank the reviewer very much for this insightful comment. Regarding bias risks, all WHOIS data used in our study were obtained from confirmed fraudulent websites, ensuring realistic fraud scenarios for evaluation.

Regarding deployment feasibility, FD-SGIL is suited for a client–server architecture, where IoT devices perform lightweight URL collection while the server executes detection, balancing accuracy with latency constraints.

Reviewer’s comment 4:

4. No Link to Broader Impact

Issue: Fails to contextualize how FD-SGIL advances cybersecurity beyond academia. Example: Could highlight potential integration with browser extensions or ISP-level detection systems.

We would like to thank the reviewer very much for this insightful comment. Regarding practical integration, FD-SGIL can be deployed in a client–server architecture, where browsers or IoT devices act as clients to collect and forward URLs, and the FD-SGIL model runs on a high-performance server to perform detection. The server then returns the results to the clients, enabling real-time blocking of fraudulent websites. The multi-source fusion design further allows adaptation to heterogeneous data sources, including URL blacklists, DNS logs, and WHOIS records. Beyond technical contributions, FD-SGIL supports broader cybersecurity goals as part of a multidisciplinary effort combining technology, law, and public education.

The above modified content is placed in the section **Ⅵ. CONCLUSION of Page 12** in the revised version.

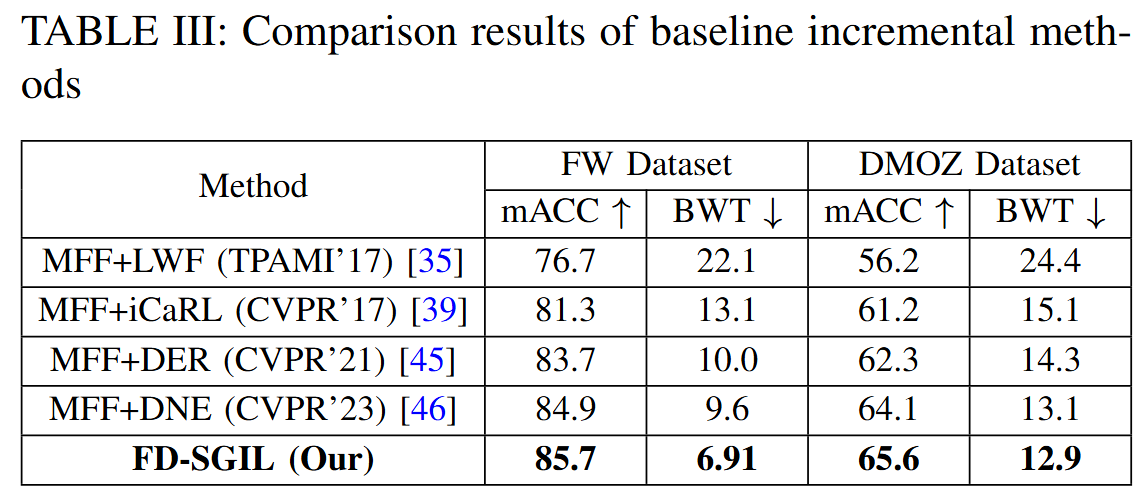
Reviewer’s comment 5:

Incremental Learning Gaps Unaddressed

Issue: While catastrophic forgetting is mentioned, the conclusion ignores: Stability-Plasticity Dilemma: Trade-offs between adapting to new classes and retaining old knowledge. Few-Shot Scenarios: Handling new fraud types with minimal labeled examples.

We would like to thank the reviewer very much for this helpful comment. Regarding stability–plasticity dilemma, FD-SGIL adopts a structure growing design in which new detection branches are expanded for emerging categories while parameters of previously learned branches are frozen. This ensures that the model can adapt to new classes (plasticity) without degrading performance on earlier ones (stability). As shown in Table III, FD-SGIL achieves superior mACC and BWT values compared with baselines, demonstrating its effectiveness in balancing adaptation and retention.

Regarding few-shot scenarios, FD-SGIL combines transfer loss with experience replay to improve learning when only limited samples are available for new categories. In addition, oversampling can be applied to increase both the quantity and diversity of scarce data before incremental updates. These strategies enable FD-SGIL to maintain recognition capability even under limited labeled data conditions. Finally, we would like to once again thank **Reviewer #2** for the constructive feedback.



The above modified content is placed in the subsection **C. Incremental Detection Performance Comparison of Pages 9~10** in the revised version.