Toward Open-Set Road Damage Recognition using LDLE and CAC Loss

Yuka Sonoda
Advanced Engineering Course
National Institute of Technology, Kurume College
Fukuoka, Japan
a4212sy@kurume.kosen-ac.jp

Masahiro Iwahashi
Electrical, Electronics and Information Engineering
Nagaoka University of Technology
Nagaoka, Japan
iwahashi@vos.nagaokaut.ac.jp

Ryosuke Harakawa
Electrical, Electronics and Information Engineering
Nagaoka University of Technology
Nagaoka, Japan
harakawa@vos.nagaokaut.ac.jp

Kousuke Matsushima
Department of Control and Information Systems Engineering
National Institute of Technology, Kurume College
Fukuoka, Japan
matsushima@kurume-nct.ac.jp

***Abstract*— Municipal roads constitute the majority of the roadway network, making their timely maintenance crucial for public safety. While deep-learning-based object detectors such as YOLO have achieved remarkable performance in road damage detection, they operate under closed-set conditions and thus fail to recognize unseen damage types. This study aims to address this limitation by developing an open-set road damage recognition framework that integrates Low-Density Latent Expansion (LDLE) and Class Anchor Clustering (CAC) mechanisms. LDLE expands feature representations of known classes into sparse latent regions, facilitating the discovery of unknown samples, whereas CAC imposes class-wise anchors to promote intra-class compactness and inter-class separation. To reduce the computational cost associated with optimal transport–based methods, the proposed framework replaces the OT distance with a lightweight CAC loss, enabling efficient and stable feature learning. Experiments conducted on public benchmark datasets demonstrate that LDLE enhances unknown-class separability, while CAC stabilizes the training process by reducing intra-class variance. Although the preliminary version of this study did not evaluate road damage datasets, the revised work now includes experiments on RDD2022. While recognition accuracy slightly decreased compared with the closed-set baseline, the findings highlight the potential of density-based expansion and distance-based clustering for open-set recognition, laying the groundwork for future application to real-world road damage scenarios.**

Keywords—Open-Set Recognition, Low-Density Latent Expansion, Class Anchor Clustering Loss, Fusion model

I. INTRODUCTION

Municipal roads constitute nearly 80% of the total roadway network used in daily transportation [1], making their maintenance and safety management a major responsibility for local governments. However, routine inspection and assessment of pavement conditions remain labor-intensive, time-consuming, and economically demanding. Undetected surface deterioration—such as cracking, delamination, or differential settlement—can degrade drainage performance and structural integrity, leading to hazardous driving conditions. To address these challenges, numerous computer-vision-based road damage detection techniques have been developed in recent years. Moreover, with the increasing prevalence of autonomous and connected vehicles, the ability to recognize and localize road damage accurately has become even more critical. In this study, the main evaluation is conducted on the RDD2022 dataset, which contains diverse and real-world road–damage categories collected from several regions. This dataset provides a realistic setting for assessing the open-set detection capability of the proposed framework.

Traditional object detectors have achieved remarkable progress on closed-set datasets, where training and testing share identical classes and distributions. Several studies have utilized large-scale road damage datasets collected collaboratively by municipalities, successfully applying lightweight object detectors such as YOLO to achieve high detection accuracy and real-time performance [2], [3]. Nevertheless, in real-world applications, object detectors often encounter unseen or unknown damage types that do not appear in the training data. When such cases arise, conventional detectors tend to misclassify novel instances into the nearest known categories, undermining the system's reliability and safety.

To overcome this limitation, open-set and open-world object detection (OWOD) frameworks have been actively explored [4]–[7]. While previous open-world object detection studies mainly focused on expanding detection pipelines or novelty scoring, this work emphasizes embedding-space regularization. Although lightweight detectors such as YOLO achieve high accuracy and real-time performance on closed-set road-damage datasets, they inherently lack mechanisms for handling unknown classes. Therefore, in this work, YOLO-based detectors are used solely as closed-set baselines, while our proposed method focuses on embedding-based open-set mechanisms such as LDLE and CAC that explicitly address unknown-class detection. This focus bridges a relatively unexplored gap between feature-space structure and open-set robustness. However, LDLE and CAC impose conflicting embedding behaviors: LDLE expands low-density regions, whereas CAC enforces compact clustering. This contrast makes unified optimization unstable. The proposed approach reduces the computational cost of optimal transport–based methods while maintaining comparable accuracy, achieving stable and lightweight training suitable for edge-level deployment. Methods such as OW-DETR introduce novelty classification mechanisms into Transformer architectures, while UC-OWOD further categorizes unknown samples into multiple novel classes and proposes dedicated evaluation protocols. Building on these advances, our study focuses on improving the feature embedding and class separation for open-set road damage detection.

Specifically, we adopt the OpenDet-CWA framework [8] as our baseline and replace the computationally demanding optimal transport (OT) distance used in its loss formulation with the Class Anchor Clustering (CAC) loss proposed by Miller et al. [9]. The CAC loss enables efficient distance-based clustering by anchoring class centers in the embedding space, thereby promoting compact intra-class features and clear inter-class separation without the heavy computational overhead of OT.

Our contributions can be summarized as follows:

(1). We integrate the CAC loss into an open-world detection framework to examine its effect on latent feature distribution and training stability.

(2). The proposed fusion of LDLE and CAC achieves a notable reduction in computational cost and training time while maintaining consistent optimization behavior.

(3). These results indicate the potential of the proposed approach as a lightweight and efficient foundation for future large-scale road damage recognition tasks on edge devices.

II. Related works

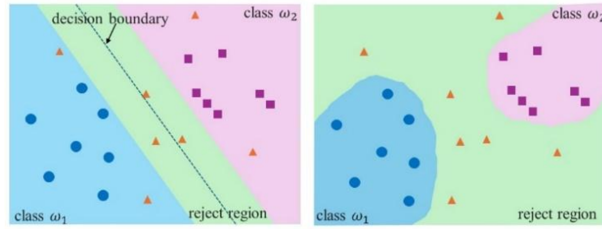
A. Road Damage Detection

Early studies on automated road inspection mainly focused on image-based crack detection using handcrafted features such as edge or texture descriptors. With the advent of deep learning, convolutional neural networks (CNNs) have been widely applied to road damage detection and classification. Large-scale collaborative efforts among local governments have enabled the creation of benchmark datasets that contain thousands of annotated road images captured by smartphones and vehicle-mounted cameras [2]. Using these datasets, object detectors such as Faster R-CNN and YOLO have demonstrated high accuracy. However, these models inherently operate under closed-set assumptions and therefore cannot detect unseen categories, leading to degraded performance in real-world scenarios. This limitation motivates the need for open-set approaches.

B. Open-Set and Open-World Object Detection

To improve robustness against unknown categories, open-set recognition (OSR) and open-world object detection (OWOD) have gained significant attention. In OSR, models are trained to recognize known classes while rejecting samples from unknown classes. Scheirer et al. [10] first formalized this concept using threshold-based recognition. Later, deep-learning-based methods introduced distance- or embedding-based mechanisms to distinguish known from unknown samples.

Extending this concept to object detection, OW-DETR [6] integrates novelty classification and objectness scoring into a Transformer-based detector, enabling the discovery of unknown instances without additional supervision. Furthermore, UC-OWOD [7] enhances this framework by introducing class grouping for unknown samples and proposing a new evaluation protocol that explicitly measures unknown-class detection performance. These works collectively highlight the importance of designing robust embedding spaces for unknown detection. Our work extends these insights by focusing specifically on the interaction between dispersion-based (LDLE) and anchor-based (CAC) embedding regularization under an open-set detection setting.



(a) (b)

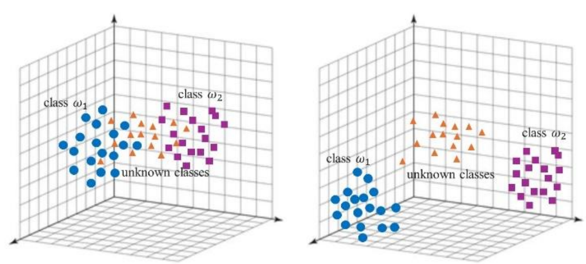
Fig.1. Visualization of the embedding space in 2D. Circular and square markers represent objects belonging to known classes. The objective is to reduce the cluster size of objects (circles and squares) belonging to the same known class, that is, to transform the distribution from (a) to (b). This contraction enlarges the triangular region corresponding to unknown classes, thereby improving the accuracy of unknown-class estimation.

C. Distance-Based Loss Functions and Optimal Transport

Recent research has explored distance-based learning objectives to achieve better class separation and discriminative feature representations. In particular, optimal transport (OT)–based losses have been incorporated into open-set detection frameworks to align class distributions and regularize embeddings [8]. Although the OT distance provides a theoretically elegant formulation for comparing feature distributions, its computation requires iterative optimization (e.g., Sinkhorn iterations), leading to substantial computational cost. This makes OT-based approaches less suitable for real-time applications such as mobile road-damage detection. Figure 1 illustrates a conceptual diagram of the two-dimensional visualization in the embedded space.

D. Class Anchor Clustering Loss

To address the computational burden of OT-based methods, Wang et al. proposed the Class Anchor Clustering (CAC) loss [9], a distance-based objective designed for open-set recognition. CAC loss introduces learnable class anchors in the embedding space and minimizes intra-class distances while enlarging inter-class margins. By jointly optimizing



(a) (b)

Fig.2. Example of a 3D logit space learned with the CAC loss. (a) shows an example of open-set recognition trained with the cross-entropy loss, while (b) illustrates the case trained with the CAC loss, where known-class clusters become more compact and separable, thereby improving the discriminability of unknown classes.

classification and clustering objectives, the model learns more compact and separable feature representations without relying on costly transport computations. Figure 2 illustrates an example of the three-dimensional logit space learned using the CAC loss.

In contrast to conventional OT-based methods that emphasize distribution alignment, this study focuses on embedding-space optimization, complementing detection-oriented approaches such as OW-DETR and UC-OWOD. By positioning LDLE and CAC within this framework, our approach addresses the underexplored challenge of balancing feature dispersion and compactness to enhance open-set recognition capability.

III. PROPOSED METHOD

The proposed framework improves open-set road damage detection by integrating Class Anchor Clustering (CAC) and Low-Density Latent Expansion (LDLE) in a two-stage optimization pipeline. Unlike conventional approaches that merge multiple embedding objectives into a single loss function, this framework separates their roles across two phases to avoid conflicting gradients. Phase 1 constructs a compact and stable embedding space using CAC, while Phase 2 employs LDLE to enhance the separability between known and unknown samples, retaining a weakened CAC head to preserve structural coherence. This division ensures smooth convergence and prevents the degradation typically observed in unified-loss formulations.

A. Phase 1: CAC-Based Embedding Formation

Phase 1 is dedicated to constructing a clean and well-structured embedding space for known classes. CAC introduces learnable anchor vectors, each representing the center of a specific class cluster. The CAC loss is defined as:

$$L_{\mathrm{CAC}} = \sum_i \| f(x_i) - a_{y_i} \|^2 - \beta \sum_{j \neq y_i} \| f(x_i) - a_j \|^2. \quad (1)$$

Minimizing this loss encourages each feature $f(x_i)$to be pulled toward its corresponding class anchor while being repelled from anchors of all other classes. This optimization produces compact intra-class clusters supported by well-separated inter-class margins, resulting in a geometrically stable embedding manifold. Such structured embeddings provide an essential initialization for Phase 2, in which LDLE performs feature-space expansion without destabilizing the cluster organization.

B. Phase 2: LDLE with a Weak CAC Embedding Head

Phase 2 applies LDLE to expand low-density regions surrounding known clusters in order to increase the separability of unknown samples. The LDLE objective is defined as

$$L_{\mathrm{LDLE}} = L_{\mathrm{det}} + \lambda D_{\mathrm{LD}}(f(x)), \quad (2)$$

where D_{LD} encourages features to disperse into low-density regions of the latent space. Although LDLE effectively reveals unknown-class boundaries, it can destabilize the structured embedding learned in Phase 1, sometimes leading to partial collapse of cluster geometry. To mitigate this issue, Phase 2 retains a weakened version of the CAC head, which acts as a stabilizing force and anchors each known-class cluster.

The resulting loss function for Phase 2 is

$$L_{\mathrm{Phase2}} = L_{\mathrm{LDLE}} + \gamma L_{\mathrm{CAC}}^{\mathrm{weak}}, \quad (3)$$

where γ is set to a sufficiently small value (typically 0.05–0.1). This formulation ensures that LDLE drives dispersion primarily along low-density directions, while the weak CAC term preserves coherent class structures, preventing over-expansion and retaining the embedding topology established in Phase 1.

C. Integration Strategy and Expected Outcomes

In earlier drafts, our explanation of the interaction between LDLE and CAC was insufficient, which may have made it unclear why unified optimization is challenging. We therefore provide a more explicit clarification here: LDLE expands low-density regions by pushing feature representations outward, while CAC encourages compact clustering around class anchors. Because these objectives act in opposite directions, optimizing them simultaneously can introduce competing gradients and reduce training stability. By transitioning to a two-stage approach, Phase 1 focuses exclusively on CAC to establish compact embeddings, and Phase 2 subsequently applies LDLE to promote separation between known and unknown regions. To address this conceptual ambiguity, we also clarify why the unified-loss setting required an α-weight balancing term. When LDLE-driven dispersion and CAC-driven clustering were forced to act simultaneously, their opposing objectives led to unstable updates. By presenting these mechanisms more clearly, we motivate the use of a two-stage training strategy in which each objective is optimized in a separate phase.

The combination of these complementary mechanisms is expected to enhance separability between known and unknown classes, stabilize the embedding distribution during training, and reduce computational cost by replacing optimal transport computations with CAC's lightweight distance-based formulation. Ultimately, the resulting model is designed to support efficient open-set road damage detection suitable for real-time deployment on edge devices and mobile platforms, where computational resources are limited.

IV. EXPERIMENT AND RESULTS

A. Experimental Setup

To evaluate the proposed two-stage framework, experiments were conducted using the RDD2022 dataset, which contains diverse, real-world road damage categories collected from multiple countries. In the open-set configuration, a subset of categories was designated as known, while the remaining categories, including ambiguous or rare types, were grouped as unknown. Two models were compared: an LDLE-only baseline and the proposed two-stage 2-Stage system. Training was performed using Detectron2 with a ResNet-50 FPN backbone on a single NVIDIA GPU running a Linux environment. Mixed-precision training was disabled to maintain numerical stability throughout optimization.

Open-set performance was assessed using standard OSR and OWOD evaluation metrics, including mean Average Precision (mAP) over IoU thresholds from 0.5 to 0.95, AP for known categories (AP@K), AP for the unknown class (AP@U), and AUROC for unknown-class detection scoring.

Table.1. Open-Set Detection Performance on RDD2022

Method	AP@K	AP@U	mAP (0.5–0.95)	AUROC
LDLE	8.72	13.8	11.26	0.5399
2-Stage	5.29	8.98	7.137	0.7149

These metrics provide a comprehensive view of both detection accuracy and robustness to unknown categories. Quantitative comparisons and detailed analysis are presented in the subsequent section.

B. *Quantitative Results*

The quantitative results are summarized in Table 1, which reports mAP, AP@K, AP@U, and AUROC for both LDLE and the proposed two-stage method.

Although the proposed 2-Stage model shows a slight degradation in AP metrics compared with LDLE alone, it provides more stable feature embedding and improved training behavior, which we analyze further.

C. *Training Dynamics, Efficiency, and Embedding Behavior*

To better understand the effect of integrating CAC in Phase 1 and preserving a weak CAC head in Phase 2, we analyzed the stability of the training process, computational efficiency, and the resulting embedding structure. As shown in Figure 3, the proposed two-stage strategy exhibits noticeably smoother loss trajectories and reduced variance compared to LDLE. This stabilization helps prevent embedding collapse and leads to more consistent optimization behavior, even though the raw AP scores remain lower than those of LDLE alone.

In addition to loss characteristics, Figure 4 shows the average time per iteration for both methods. The first phase of the two-stage approach demonstrates a substantially lower computational cost than LDLE, resulting in a faster initial training process. After transitioning to Phase 2, the per-iteration cost becomes similar to LDLE; however, this shift does not adversely affect training stability. These observations indicate that the two-stage framework improves embedding robustness while also providing a faster early-stage optimization process.

We report mAP (IoU 0.5–0.95), AP@K (mean AP across known classes), AP@U (AP of the unknown class), and AUROC based on the unknown-class detection score. These metrics provide a more complete and objective evaluation of open-set performance in accordance with prior OSR/OWOD literature.

V. Discussion

This study introduces a two-stage embedding-based framework that integrates Class Anchor Clustering (CAC) with Low-Density Latent Expansion (LDLE) for open-set road damage detection, and the experimental results reveal several notable insights. CAC plays a foundational role by organizing the feature space into compact and coherent intra-class clusters while preserving distinct margins between known categories. Through this process, the embedding space becomes clean, geometrically structured, and highly stable, enabling more reliable optimization in subsequent stages. LDLE then builds upon this foundation by expanding the low-density regions surrounding the clusters, effectively creating additional geometric space in which unknown samples can be more clearly distinguished from known ones.

When applied sequentially, the combined method yields an embedding space characterized by coherent clusters for known classes, well-defined boundaries between categories, and expanded low-density regions that function as buffer zones for unknown samples. These properties collectively contribute to enhanced interpretability and a more discriminative latent structure, even though the overall AP values do not exceed those of the LDLE-only baseline. The reduction observed in mAP and AP@U indicates a central challenge inherent to this approach: CAC and LDLE each encourage beneficial but opposing behaviors—one promoting compactness and the other encouraging dispersion. Although the two-stage strategy alleviates much of the direct gradient

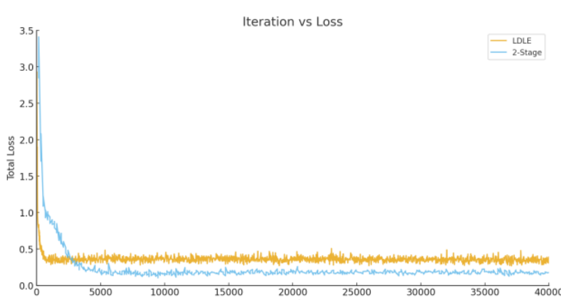


Fig.3. Iteration–Loss curves for LDLE and the proposed 2-Stage method. The 2-Stage model shows consistently lower variance and smoother convergence after the initial CAC phase, whereas LDLE exhibits higher fluctuation throughout training. Although LDLE reaches a lower final loss in some iterations, the 2-stage strategy yields more stable optimization behavior overall.

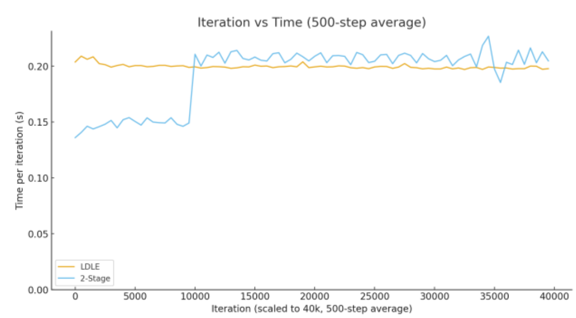


Fig. 4. Iteration–Time curves (500-step moving average) for LDLE and the proposed 2-Stage method. The 2-Stage model exhibits a markedly lower per-iteration computation time during the initial CAC phase, enabling faster early-stage optimization. After transitioning to Phase 2, the iteration time becomes comparable to LDLE; however, this shift does not negatively affect training consistency. Overall, the two-stage strategy achieves faster initial learning while maintaining stable computational behavior throughout training.

conflict seen in unified training, it does not completely eliminate the underlying tension between these objectives.

In addition, the current implementation uses fixed hyperparameters for both CAC and LDLE (e.g., the CAC repulsion strength β, the LDLE density coefficient λ, and the switching criterion between stages). Because these parameters directly control the balance between compactness and dispersion, their optimization is expected to have a strong influence on AP performance. Adaptive or data-driven tuning may further reduce the observed performance gap and potentially improve both AP@K and AP@U.

Despite this limitation, the improved training stability and reduced computational overhead demonstrate the potential of the fusion direction. The results suggest that careful coordination between CAC and LDLE can continue to enhance open-set recognition, provided that their competing influences are better managed. Building upon these findings, future research will investigate adaptive mechanisms that dynamically adjust the balance between clustering and dispersion during training, explore multi-head architectures capable of functionally decoupling embedding sub-tasks, and expand large-scale training on diverse road damage datasets to further improve unknown-class detection and generalization performance.

VI. Conclusion

This work proposed a sequential 2-Stage framework for open-set road damage recognition. CAC establishes a compact and discriminative embedding for known classes, while LDLE expands low-density regions to improve separation from unknown instances. Although the integrated model did not exceed the LDLE-only configuration in terms of AP performance, it achieved greater training stability and reduced computational overhead.

Because the two-stage framework relies on several key hyperparameters—such as the LDLE density weight, CAC margin strength, and the timing of the phase transition—further optimization of these values may lead to improved detection performance. As such, the reduced AP metrics observed in the present results should not necessarily be interpreted as an inherent limitation of the approach but rather as an indication of untapped optimization potential.

These findings highlight both the potential and the limitations of combining compactness-driven and dispersion-driven embedding mechanisms. The results suggest that more adaptive or hierarchical training strategies may further improve recognition robustness. Future developments will explore improved scheduling strategies, dual-head networks, and large-scale validation on diverse road-damage datasets.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 25K15171 and by Nagaoka University of Technology (NUT) grant for collaborative research with National Institute of Technology (NIT).

References

[1] Japanese Road Statistics Annual Report, Ministry of Land, Infrastructure, Transport and Tourism, 2021.

[2] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road Damage Detection using Deep Neural Networks with Images Captured Through a Smartphone, " arXiv:1801.09454, Feb., 2018.

[3] W. Wang, X. Yu, B. Jing, Z. Tang, W. Zhang, S. Wang, Y. Xiao, S. Li, and L. Yang, "YOLO-RD: A Road Damage Detection Method for Effective Pavement Maintenance, " Sensors, vol. 25, no. 5, 442, 2025.

[4] J. Han, Y. Ren, J. Ding, X. Pan, K. Yan, G. Xia, "Expanding Low-Density Latent Regions for Open-Set Object Detection, " Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9591-9600, 2022.

[5] K. J. Joseph, S. Khan, F. S. Khan, and V. N. Balasubramanian, "Towards open world object detection, " Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.5830-5840, 2021.

[6] A. Gupta, S. Narayan, K. J. Joseph, S. Khan, F. S. Khan, and M. Shah, "OW-DETR: Open-world Detection Transformer, " arXiv: 2112.01513, Apr., 2022.

[7] Z. Wu, Y. Lu, X. Chen, Z. Wu, L. Kang, and J. Yu, "UC-OWOD: Unknown-Classified Open World Object Detection, " arXiv: 2207.11455, Jul., 2022.

[8] P. Mallick, F. Dayoub, and J. Sherrah, "Wasserstein Distance-based Expansion of Low-Density Latent Regions for Unknown Class Detection, " arXiv:2401.05594, Jan., 2024.

[9] D. Miller, N. Sünderhauf, M. Milford, and F. Dayoub, "Class Anchor Clustering: A Loss for Distance-based Open Set Recognition, " IEEE Winter Conference on Applications of Computer Vision, Jun., 2021.

[10] W. J. Scheirer, A. D. R. Rocha, A. Sapkota, and T. E. Boult, "Toward Open Set Recognition, " IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1757–1772, Jul. 2013.