Target Vessel Identification in Aerial Search Imagery via MLLM-Based Attribute Extraction and Geolocation Fusion

Jeonghyo Oh[1], Youngon Oh[2], Impyeong Lee[3,*]

[1] Dept. of Geoinformatics, University of Seoul, Seoul, Republic of Korea - ohrora98@uos.ac.kr
[2] Dept. of Geoinformatics, University of Seoul, Seoul, Republic of Korea - ohrgon3446@uos.ac.kr
[3] Dept. of Geoinformatics, University of Seoul, Seoul, Republic of Korea - iplee@uos.ac.kr*

Keywords: Target Vessel Identification, Aerial Search Imagery, MLLM-based Attribute Extraction, Semantic–Geolocation Fusion, Embedding Similarity, Maritime Search and Rescue

Abstract

Identifying a distressed vessel among many ships detected in wide-area aerial imagery is a critical challenge in maritime Search and Rescue (SAR) operations. Conventional methods cannot determine which vessel matches the incident description, especially when Automatic Identification System (AIS) reports are uncertain. This study proposes an integrated framework that combines MLLM-based semantic attribute extraction with geolocation fusion to prioritize candidate vessels according to their consistency with Situation Report (SITREP) based scenarios. The method detects vessels using YOLOv8, tracks them with Deep Simple Online and Real-time Tracking (DeepSORT), and performs image-based georeferencing using onboard metadata. A Multi-modal Large Language Model (MLLM) extracts appearance/status attributes from representative vessel images, while scenario descriptions are also converted to attributes. Both sets are encoded using MiniLM embeddings. Finally, semantic similarity is fused with geolocation proximity within an Support Vector Machine (SVM) classifier to produce a probability-ranked list of candidates. Experiments using real aerial search footage demonstrate robust identification performance across a range of scenario quality levels. The correct vessel appears within the top three candidates in more than 73% of cases and within the top five in more than 91%, even when attribute extraction is affected by low resolution, illumination effects, or missing scenario information. These results show that coarse semantic cues, when combined with approximate geolocation, provide a resilient basis for identifying target vessels under high uncertainty. The proposed framework offers a practical foundation for automated SAR decision support, enabling faster and more reliable prioritization during wide-area maritime search operations.

1. Introduction

In maritime SAR operations, rapidly locating a distressed vessel is critical, yet the task is challenging due to high meteorological variability, wide-area drift, and uncertainty in last-known AIS positions (Martinez-Esteso et al., 2025). As search areas expand, aerial platforms increasingly acquire large volumes of high-resolution video, but manual inspection is slow, error-prone, and operationally unsustainable (Oh et al., 2023, Spagnolo et al., 2019).

Recent deep learning-based ship detection models—particularly the YOLO series have improved real-time detection performance in maritime imagery. YOLO variants have demonstrated strong accuracy and speed compared to traditional detectors such as Faster R-CNN or RetinaNet in UAV-based maritime settings (Zhao et al., 2024). Several domain-adapted improvements have been proposed, including GGT-YOLO for complex backgrounds (Li et al., 2022), attention-enhanced YOLOv7 variants (Zhang et al., 2025), and dynamic-convolution-based architectures for challenging sea conditions (Li et al., 2024). However, despite improved detection, these models classify all vessels into a single “ship” category and therefore cannot distinguish the single distressed vessel from numerous non-targets.

Multi-object Tracking (MOT) has been used to maintain vessel identities across frames. Approaches such as DeepSORT integrate appearance embeddings to reduce ID switching, achieving high Multiple Object Tracking Accuracy (MOTA)/Multiple Object Tracking Precision (MOTP) in UAV-based multi-ship tracking (Zhang et al., 2025). Nevertheless, MOT provides only relative pixel-space trajectories; abrupt zoom changes, viewpoint transitions, or partial occlusion still lead to frequent ID switches, making tracking alone insufficient for target identification during SAR missions. Alternative strategies have attempted to identify vessels using Optical Character Recognition (OCR), focusing on hull registration numbers (Fabijanic et al., 2025). Yet OCR performance in aerial imagery is highly sensitive to illumination, viewing angle, resolution, and sea-surface reflections, resulting in inconsistent recognition. Appearance-based vessel Re-ID approaches have also been explored (Sun et al., 2025, Zhang et al., 2023, Qiao et al., 2020, Chen et al., 2020, Carrillo-Perez et al., 2022), but these methods rely solely on visual similarity and do not incorporate contextual cues such as vessel status or geolocation. Most importantly, they cannot leverage SITREP-based semantic descriptions—such as hull color, vessel type, operational status, or last-reported coordinates—which are essential for real SAR decision-making.

Thus, despite progress in detection, tracking, OCR, and Re-ID, no existing work jointly integrates (1) semantic attributes, (2) vessel status, (3) real-world geolocation, and (4) SITREP-based scenario information for identifying a target vessel among many candidates. Addressing this gap requires a method capable of extracting rich semantic cues from aerial imagery while aligning them with scenario descriptions and geographic proximity. To address these limitations, this study proposes an integrated framework for target vessel identification that fuses MLLM-based semantic attribute extraction with image-based geolocation estimation and scenario matching. The contributions of this work are fourfold:

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-503-2026

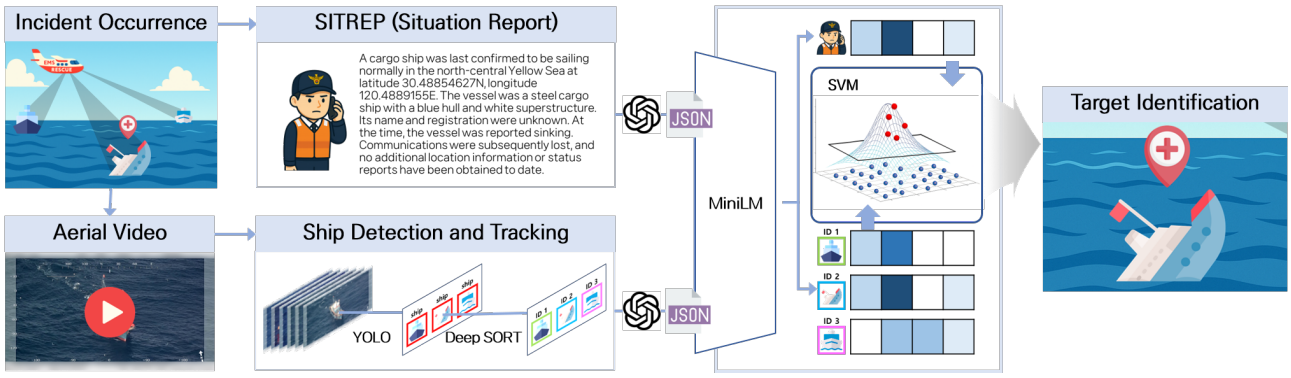

Figure 1. Overview of the proposed target vessel identification methodology.

- **MLLM-based attribute extraction:** We extract vessel appearance and status attributes from aerial imagery using a Multi-modal Large Language Model, enabling semantic reasoning beyond conventional detection or Re-ID.
- **Geolocation estimation:** We compute each vessel’s latitude–longitude position via image-based georeferencing, providing real-world spatial context unavailable in prior studies.
- **Scenario–imagery alignment:** We convert both imagery-derived attributes and SITREP-like scenarios into embedding vectors to compute semantic similarity under varying information quality.
- **Probability-based identification:** We integrate semantic and geographic features using an SVM classifier to generate probability-ranked candidate vessels, supporting practical SAR decision-making under uncertainty.

Through this multi-modal fusion of semantic attributes and geolocation information, the proposed approach overcomes key limitations of existing ship identification pipelines and provides a robust foundation for automated prioritization in real SAR environments.

2. Methodology

Based on SITREP information generated during maritime search, target vessels must be automatically identified among the numerous vessels detected through aerial search. To achieve this, vessels are first detected and tracked in the image, each assigned a unique ID, and georeferencing is applied to calculate their actual locations. MLLM is then used to extract the visual attributes of the vessels in text format. Next, SITREP information and the attribute information of the detected vessels are converted into vectors using an embedding model to verify the similarity between the attributes. Finally, a method was designed to identify the target vessels using an SVM classifier, as shown in Figure 1.

2.1 Dataset Construction

This study utilized aerial video footage provided by the Korea Coast Guard, acquired over the West Sea approximately 100 km offshore. The original video lasted 16 minutes and 54 seconds, from which frames were extracted at 5 fps, resulting in a total of 5,077 images with a resolution of 3840×2160 pixels. All frames were used without any geometric or radiometric preprocessing to preserve real operational conditions. The dataset includes 12 vessels of varying types—primarily fishing boats, one cargo vessel, and two coast guard vessels—as well as scenes containing no vessels. The imagery exhibits wide variability in visual conditions that commonly arise during search missions, including: Multiple zoom levels (five distinct levels), ranging from wide-area views to close-up shots; Varying illumination and sun–glint effects due to changing time and camera orientation; Different viewing angles, including oblique and near-vertical perspectives; and Cases where vessels appear extremely small or partially occluded. These characteristics make the dataset well suited for evaluating robust vessel identification, as the imagery reflects the uncertainties and visual challenges typical of aerial SAR operations. The diversity in scale, angle, and environmental conditions also enables comprehensive assessment of detection, attribute extraction, and geolocation performance under realistic maritime scenarios.

2.2 Attribute Extraction from Aerial Imagery

Vessels detected in the aerial imagery were processed through a multi-stage pipeline to extract semantic and spatial attributes required for target identification. First, ship instances were detected using a YOLOv8 model fine-tuned on 2,791 maritime aerial images, achieving robust performance for the single “Ship” class. To maintain object identity across frames, DeepSORT was applied, producing consistent tracking IDs except in cases of abrupt zoom or viewpoint changes.

Since each vessel appears under varying scales and conditions throughout the video, a representative image was selected for each tracking ID to maximize the reliability of attribute inference. Among all frames associated with a vessel ID, the frame with the largest bounding box was chosen, as it provides the clearest depiction of hull shape, color, and superstructure details.

The geolocation of each vessel was estimated by converting the pixel coordinates of the representative bounding box center into real-world latitude and longitude. Embedded sensor metadata—such as the aircraft’s GPS position, altitude, and camera parameters—were extracted using OCR and used to compute the camera’s external orientation. Assuming a flat sea surface, a homography transformation mapped image coordinates to geographic coordinates. Validation using land-based control points yielded an Root Mean Squared Error (RMSE) of approximately 13.7 m in X and 53.4 m in Y, which is sufficient for coarse-level localization in wide-area maritime search scenarios.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-503-2026 | © Author(s) 2026. CC BY 4.0 License.

Level	Pos. Var. (Lat/Lon)	Color/Material	Missing Attr.
High	within ±30 km	similar / visually close	0
Medium	30–60 km deviation	similar / visually close	2
Low	60–90 km deviation	similar / visually close	4

Table 1. Scenario quality level definition.

Semantic attribute extraction was performed using a MLLM. The representative vessel image was passed to the model with a structured prompting scheme that requested both appearance attributes (vessel type, hull color, superstructure color, material, identifiable markings) and status attributes (moving, drifting, stationary, capsized, presence of smoke or distress indicators). The MLLM leveraged its joint visual–linguistic reasoning capabilities to infer attributes even under partially occluded or low-contrast conditions, although extreme illumination and small object size introduced occasional errors.

All extracted attributes—appearance, status, and estimated geolocation—were stored in a unified JSON schema. These structured descriptions form the basis for downstream scenario matching, embedding-based similarity computation, and probability-ranked target vessel identification.

2.3 SITREP-based Accident Scenario Generation

In real maritime SAR missions, responders rely on SITREP information that describes the distressed vessel’s appearance, status, and last-known position. However, actual SITREP datasets are difficult to obtain due to confidentiality and event rarity. To emulate realistic search conditions, a set of synthetic accident scenarios was generated based on the International Aeronautical and Maritime Search and Rescue (IAMSAR) manual and the ground-truth attributes of the 12 vessels observed in the aerial footage.

For each vessel, a baseline scenario was constructed using its ground-truth attributes, including vessel type, hull color, superstructure color, material, status, and last-known geographic position. To simulate varying degrees of uncertainty commonly encountered in SAR operations, additional scenarios were produced using an MLLM to introduce controlled variations and omissions. Three scenario quality levels were defined:

High quality: The reported position was perturbed within ±30 km, and appearance attributes were kept identical or visually similar to the ground truth.

Medium quality: The position was perturbed within ±30 ∼ 60 km, and two appearance or status attributes were randomly omitted to reflect partial information loss over time.

Low quality: The position deviation was increased to ±60 ∼ 90 km, and four attributes were omitted to represent severe uncertainty in vessel appearance and operational state.

The criteria for each scenario level are summarized in Table 1. For each vessel, three variations were generated per quality level, resulting in 18 scenarios per vessel and 216 scenarios in total. To enable consistent comparison with imagery-derived attributes, all scenarios were converted into a unified JSON schema containing fields corresponding to appearance attributes, status indicators, and reported coordinates.

This scenario construction process provides a systematic way to evaluate target vessel identification performance under controlled uncertainty conditions, ranging from accurate SITREPs to highly incomplete, ambiguous, or drifted reports conditions that frequently arise in practical wide-area maritime search operations.

2.4 Target Vessel Identification

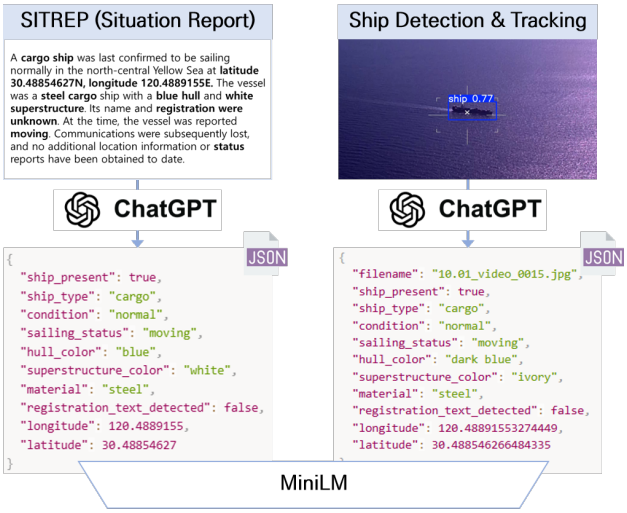

Figure 2. Comparison structure between the scenario JSON and the detected vessel JSON.

To identify the distressed vessel among multiple candidates detected in the aerial imagery, the semantic attributes extracted from each vessel and the attributes described in the corresponding scenario JSON must be jointly compared, as shown in Figure 2. Because textual descriptions may differ in wording, level of detail, or recording style, simple string matching is insufficient. Instead, both imagery-derived attributes and scenario descriptions were encoded into vector embeddings to measure semantic similarity in a continuous feature space. Each attribute set was converted into a 384-dimensional embedding vector using a MiniLM sentence-transformer. Let v_{det} denote the embedding of a detected vessel and v_{tar} the embedding of the scenario. Their cosine similarity:

$$s_{\cos} = \frac{\mathbf{v}_{\mathrm{tar}} \cdot \mathbf{v}_{\mathrm{det}}}{\|\mathbf{v}_{\mathrm{tar}}\| \, \|\mathbf{v}_{\mathrm{det}}\|} \tag{1}$$

measures the semantic consistency between the two descriptions, capturing similarity even when different but related terms (e.g., “sky blue” vs. “light blue”) are used. To incorporate spatial information, the geographic distance between the scenario-reported coordinates and the georeferenced vessel location was computed using the Haversine formula. A proximity weight was then defined as:

$$p_{\mathrm{prox}} = \exp\left(-\frac{d_{\mathrm{geo}}}{\tau}\right), \qquad \tau > 0. \tag{2}$$

where d_{geo} is the great-circle distance and $\tau > 0$ a scale factor. Because distressed vessels may drift far from their initially re-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-503-2026 | © Author(s) 2026. CC BY 4.0 License.

ported position, an inverse-distance term:

$$\frac{1}{1+d_{\text{geo}}} \tag{3}$$

was additionally included to maintain a residual influence of location even for large positional deviations. The final feature vector for each candidate vessel is constructed as:

$$\mathbf{x} = \left[\mathbf{v}_{\text{det}},\ p_{\text{prox}},\ \frac{1}{1+d_{\text{geo}}},\ s_{\cos}\right] \in \mathbb{R}^{387} \tag{4}$$

An SVM classifier with an RBF kernel was trained to estimate the probability that a detected vessel corresponds to the target described in the scenario. The classifier outputs a probability score for each candidate, and vessels are ranked in descending order to form a prioritized search list. This probability-based ranking approach provides flexible decision boundaries and robust performance under attribute uncertainty, incomplete scenario information, and large geolocation deviations. By jointly leveraging semantic similarity and geographic coherence, the method effectively narrows down potential distressed vessels in wide-area aerial search operations.

3. Results and Analysis

3.1 Evaluation Metrics

To evaluate how effectively the proposed method prioritizes the correct distressed vessel among multiple candidates, two ranking-based metrics were employed: Hit@k and Normalized Discounted Cumulative Gain (NDCG@k). These metrics reflect not only whether the correct vessel appears within the top-ranked candidates but also how highly it is positioned in the ranking, which is crucial for practical decision-making in maritime search operations. Hit@k measures whether the target vessel is included within the top k positions of the ranked list:

$$\text{Hit@}k = \begin{cases} 1, & \text{if correct ship,} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

A higher Hit@k indicates that the system can quickly narrow down the candidate set that search teams must inspect, thus reducing the initial response time. However, Hit@k only provides a binary assessment and does not distinguish whether the correct vessel is ranked first or barely included as the k-th item. To assess ranking quality, NDCG@k was additionally used:

$$\text{NDCG@}k = \frac{\text{DCG@}k}{\text{IDCG@}k} \tag{6}$$

Here, Discounted Cumulative Gain (DCG@k) reflects the cumulative relevance score of the top k ranked items, discounted by rank position, while Ideal DCG (IDCG@k) denotes the ideal ordering. NDCG@k therefore captures how effectively the model places the correct vessel near the top of the list, assigning higher weight to higher-ranked positions. Together, Hit@k and NDCG@k provide a comprehensive evaluation of the identification performance, indicating both the likelihood of correctly including the target in early candidates and the model’s ability to prioritize it at the highest ranks.

3.2 Detection and Tracking Performance

Vessel detection was performed on the 5,077 extracted frames using the fine-tuned YOLOv8 model. As shown in Figure 3, the detector exhibited stable performance across a wide range of imaging conditions, including close-range views, partially occluded vessels, and distant targets appearing only a few pixels in size. This ensured that all visible vessels were consistently captured as detection candidates.

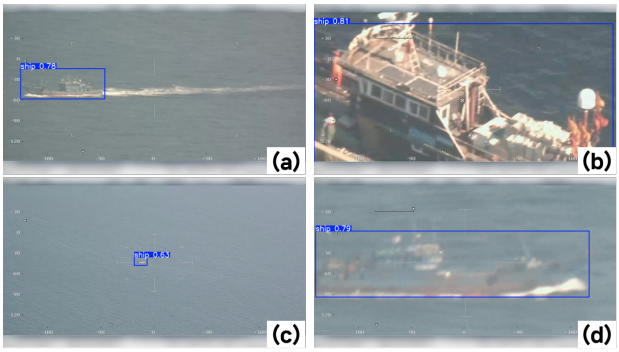

Figure 3. YOLOv8 ship detection results.

To maintain object identity across frames, DeepSORT was applied to the detection results. When camera motion and zoom transitions were smooth, the tracker assigned stable IDs to the same vessel over time, as illustrated in Figure 4. However, abrupt zoom changes, rapid viewpoint transitions, and sea-surface reflections occasionally caused ID switching (IDSW), where a single vessel was assigned multiple IDs (Figure 5). As a result, 153 tracking IDs were generated from only 12 physical vessels.

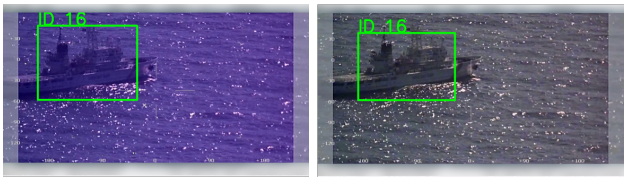

Figure 4. Example of consistent vessel tracking using DeepSORT.

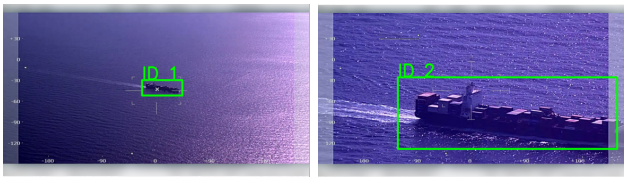

Figure 5. Example of IDSW in DeepSORT tracking.

Importantly, the presence of IDSW did not adversely affect the subsequent identification stage. The proposed method relies on representative images extracted per tracking ID rather than on long-term trajectory consistency. Selecting the frame with the largest bounding box for each ID ensured that even fragmented trajectories still provided a clear visual sample for attribute extraction. Thus, despite frequent ID switches in challenging aerial footage, detection and tracking results were sufficient to support reliable downstream semantic and geolocation fusion.

3.3 MLLM-Based Attribute Extraction

The representative images selected for each tracking ID were analyzed using the MLLM to derive appearance and status at-

Level	Hit@3	Hit@5	NDCG@3	NDCG@5
High	0.77027	0.918919	0.609618	0.650697
Medium	0.753623	0.916667	0.606024	0.636285
Low	0.736111	0.913043	0.567738	0.635452

Table 2. Target identification performance across different scenario quality levels.

tributes required for scenario matching. When vessels were captured with sufficient resolution and clear structural visibility, the MLLM reliably inferred major attributes such as vessel type, hull color, superstructure color, and material. As illustrated in Figure 6, the model effectively interpreted complex visual cues without task-specific fine-tuning, producing semantically consistent descriptions that aligned well with the vessel’s true characteristics.

Figure 6. Reliable attribute extraction results using MLLM.

However, attribute inference performance varied depending on imaging conditions, and several recurring error patterns were observed. First, strong sunlight, sea-surface reflections, and high contrast frequently led to color misinterpretation. In such cases, bright reflections caused blue or red hull components to be inferred as white or orange, as shown in Figure 7(a). Second, when vessels appeared extremely small or blurred due to long-distance capture or low zoom levels, the model often produced null or highly uncertain attribute values (Figure 7(b)), as fine structural cues were indistinguishable. Third, variations in wake visibility and shallow viewing angles occasionally led to incorrect status interpretation, particularly when motion blur obscured the vessel’s movement.

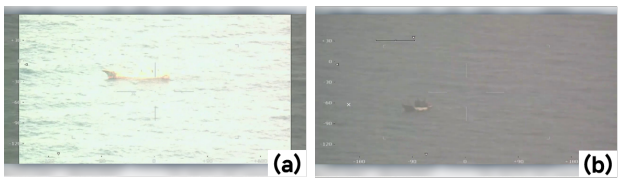

Figure 7. Attribute extraction errors caused by adverse lighting and low resolution.

Text-based attributes, such as registration numbers, exhibited the lowest reliability. As shown in Figure 8, successful recognition occurred only when character shapes were sharply defined and high in contrast. In most aerial frames, text appeared blurred, oblique, or partially submerged in reflections, leading to frequent detection failures. These limitations are consistent with known challenges in OCR under maritime aerial imaging conditions.

Despite these uncertainties, the imperfect attribute predictions did not critically degrade overall identification performance. The subsequent semantic–geolocation fusion mitigates missing or inaccurate attributes by combining embedding similarity with geographic proximity. As demonstrated in Section 3.4, vessels were correctly ranked even when multiple visual attributes were incomplete or misclassified. This indicates that the MLLM need not produce perfectly accurate descriptions; rather, coarse but semantically meaningful cues are sufficient for downstream target identification.

Figure 8. MLLM-based detection of vessel registration text.

3.4 Target Vessel Identification Performance

The performance of the proposed identification framework was evaluated across the three scenario quality levels described in Section 2.3. Table 2 summarizes the Hit@k and NDCG@k results. Overall, the method demonstrated strong and stable performance even under substantial uncertainty in scenario information. For High-quality scenarios, the Hit@3 score reached 0.770, indicating that in 77% of cases the true target vessel appeared within the top three ranked candidates. Performance remained similarly robust for Medium- and Low-quality scenarios (0.754 and 0.736, respectively), showing that moderate attribute omissions or positional deviations had limited impact on candidate prioritization.

Hit@5 exceeded 0.91 across all scenario conditions, confirming that the correct vessel was almost always included within the top five candidates even when several attributes were missing or inaccurately described. This is a practical advantage for real SAR missions, where operators typically review a small number of highest-priority candidates rather than a single definitive prediction. The NDCG@3 and NDCG@5 scores further indicate that the ranking quality remained stable as scenario quality decreased. Although minor degradation was observed in Low-quality cases, the correct vessel generally remained near the top of the ranked list, reflecting effective semantic–geolocation integration.

Figure 9 presents example matching results for the cargo vessel used in the evaluation. In Figures 9(a)–9(c), the extracted attributes from aerial imagery aligned well with the scenario descriptions despite slight wording or categorization differences. Even in the challenging case shown in Figure 9(d), where the vessel appeared extremely small and attribute extraction was largely unsuccessful, the proposed method still ranked the correct vessel within the top ten among 153 candidates. This demonstrates that geographic cues can compensate for incomplete semantic information when visual details are limited. Conversely, in scenarios with large positional uncertainty, semantic similarity guided the identification by aligning scenario descriptions with MLLM-extracted attributes.

These results collectively show that the complementary roles of semantic embeddings and geographic proximity enable the proposed framework to maintain reliable identification performance under a wide range of uncertainty conditions. The ability to consistently surface the correct vessel among the top-ranked

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-503-2026 | © Author(s) 2026. CC BY 4.0 License.

candidates is essential in wide-area maritime search operations, where rapid triage of potential targets directly influences response effectiveness.

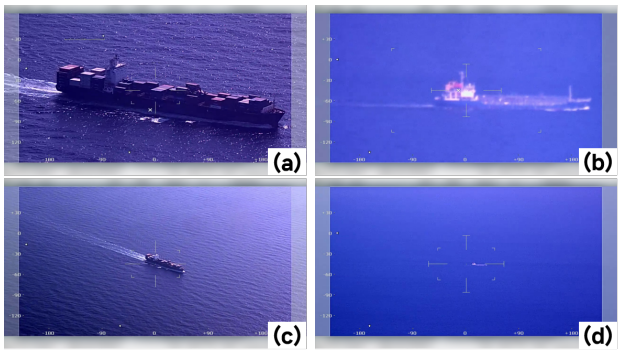

Figure 9. Scenario-based matching results for the target vessel.

3.5 Discussion

The experimental results demonstrate that the proposed framework maintains reliable identification performance even under significant uncertainty in both imagery-derived attributes and scenario information. This robustness stems from the complementary roles of semantic embeddings and geographic proximity. When attribute extraction was accurate—typically under high-resolution and stable illumination conditions—semantic similarity strongly guided the ranking process. Conversely, when attributes were missing or misclassified due to low resolution, glare, or motion blur, geolocation information compensated by enforcing spatial coherence with the scenario’s last-known position.

The frequent ID switches observed during tracking, as well as the limited reliability of registration text recognition, highlight the challenges of aerial maritime imagery. Nevertheless, these factors did not critically impact identification performance because the method relies on representative images rather than long-term tracking consistency or OCR. This suggests that, in wide-area SAR missions, high-level semantic cues may be more valuable than precise object tracking or text recognition.

The stable Hit@5 performance across all scenario quality levels indicates that the model effectively narrows the candidate set even when the scenario contains substantial omissions or position drift. This behavior is particularly important in operational contexts, where search teams must rapidly prioritize a small number of vessels rather than rely on a single definitive prediction.

Overall, the results imply that reliable target identification does not require perfect attribute extraction or exact positional accuracy. Instead, the integration of coarse semantic cues with approximate geolocation forms a resilient decision basis under real-world uncertainty. This finding underscores the potential of semantic–geolocation fusion as a practical component in future automated SAR decision-support systems.

4. Conclusions

This study presented an integrated framework for identifying a target vessel in wide-area aerial search imagery by combining MLLM-based attribute extraction with geolocation estimation and scenario matching. The approach jointly utilizes semantic cues derived from representative vessel images and spatial proximity computed through image-based georeferencing, enabling robust target identification under uncertain and incomplete information.

Experiments using real aerial footage demonstrated that the method consistently ranks the correct vessel among the top candidates, achieving Hit@3 values above 73% and Hit@5 values above 91% across all scenario quality levels. Even when attribute extraction was impaired by low resolution or strong illumination effects, or when scenario information was partially missing, the complementary fusion of semantic similarity and geographic distance maintained stable performance. These results indicate that reliable identification does not require perfect attribute inference or precise positional accuracy; instead, coarse semantic cues combined with approximate geolocation provide a resilient basis for decision support.

The proposed framework offers a practical foundation for automated prioritization of vessels in maritime Search and Rescue operations, where rapid triage of potential targets is essential. Future work will focus on improving robustness under extreme imaging conditions, expanding attribute categories, and integrating temporal reasoning to further support real-time operational deployment.

5. Acknowledgements

This research was supported by Korea Institute of Marine Science & Technology Promotion(KIMST) funded by the Ministry of Oceans and Fisheries, Korea(RS-2022-KS221629).

References

Carrillo-Perez, B., Barnes, S., Stephan, M., 2022. Ship segmentation and georeferencing from static oblique view images. *Sensors*, 22(7), 2713. doi.org/10.3390/s22072713.

Chen, X., Sui, H., Fang, J., Feng, W., Zhou, M., 2020. Vehicle re-identification using distance-based global and partial multi-regional feature learning. *IEEE Trans. Intell. Transp. Syst.*, 22(2), 1276–1286. doi.org/10.1109/TITS.2020.2972418.

Fabijanic, M., Ferreira, F., Magdalenic, M., Obradovic, J., Kapetanovic, N., Miskovic, N., 2025. Vessel registration number detection and recognition system. *Proc. WACV Workshops*, 1535–1541.

Li, Y., Yuan, H. R., Wang, Y., Xiao, C., 2022. GGT-YOLO: A novel object detection algorithm for drone-based maritime cruising. *Drones*, 6(11), 335. doi.org/10.3390/drones6110335.

Li, Z., Deng, Z., Hao, K., Zhao, X., Jin, Z., 2024. A ship detection model based on dynamic convolution and an adaptive fusion network for complex maritime conditions. *Sensors*, 24(3), 859. doi.org/10.3390/s24030859.

Martinez-Esteso, J. P., Castellanos, F. J., Calvo-Zaragoza, J., Gallego, A. J., 2025. Maritime search and rescue missions with aerial images: A survey. *Comput. Sci. Rev.*, 57, 100736. doi.org/10.1016/j.cosrev.2024.100736.

Oh, J., Lee, J., Jeon, E., Lee, I., 2023. Development of a deep-learning model with maritime environment simulation for detection of distress ships from drone images. *Korean Journal of Remote Sensing*, 39(6-1), 1451–1466.

Qiao, D., Liu, G., Dong, F., Jiang, S.-X., Dai, L., 2020. Marine vessel re-identification: A large-scale dataset and global-and-local fusion-based discriminative feature learning. *IEEE Access*, 8, 27744–27756. doi.org/10.1109/ACCESS.2020.2971550.

Spagnolo, P., Filieri, F., Distante, C., Mazzeo, P. L., D’Ambrosio, P., 2019. A new annotated dataset for boat detection and re-identification. *Proc. AVSS*, 1–7.

Sun, W., Guan, F., Zhang, X., Shen, X., Wang, K., 2025. Ship re-identification in foggy weather: A two-branch network with dynamic feature enhancement and dual attention. *Eng. Appl. Artif. Intell.*, 143, 109974. doi.org/10.1016/j.engappai.2024.109974.

Zhang, G., Liu, J., Zhao, Y., Luo, W., Mei, K., Wang, P., Song, Y., Li, X., 2025. A reliable unmanned aerial vehicle multi-ship tracking method. *PLOS ONE*, 20(1), e0316933. doi.org/10.1371/journal.pone.0316933.

Zhang, Q., Zhang, M., Liu, J., He, X., Song, R., Zhang, W., 2023. Unsupervised maritime vessel re-identification with multi-level contrastive learning. *IEEE Trans. Intell. Transp. Syst.*, 24(5), 5406–5418. doi.org/10.1109/TITS.2023.3245700.

Zhao, C., Liu, R. W., Qu, J., Gao, R., 2024. Deep learning-based object detection in maritime unmanned aerial vehicle imagery: Review and experimental comparisons. *Eng. Appl. Artif. Intell.*, 128, 107513. doi.org/10.1016/j.engappai.2023.107513.