remote sensing

MDPI

Article

Dark Vessel Detection Agent: Evidence-Driven Screening and Verification Task Generation Using Satellite Imagery and AIS

Yuehan Wang [1,2], Fenzhen Su [1,3], Yinxia Cao [1,3], Nan Xu [4,5] and Fengqin Yan [1,2,3,*]

[1] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[2] Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China
[3] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 101408, China
[4] MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Guangdong Key Laboratory of Urban Informatics, Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China
[5] School of Architecture & Urban Planning, Shenzhen University, Shenzhen 518060, China
* Correspondence: yanfq@lreis.ac.cn

Highlights

What are the main findings?

- Geographic context and image-time AIS evidence help distinguish explainable satellite–AIS mismatches from unresolved vessel observations.
- The evidence-driven workflow reduced 825 initially AIS-unexplained targets to 247 prioritized verification tasks, a reduction of 70.06%.

What is the implication of the main finding?

- Satellite–AIS mismatch alone is insufficient to establish non-cooperative vessel activity.
- Integrating multi-source geospatial evidence enables traceable prioritization of maritime verification tasks in complex coastal waters.

Abstract

Dark vessels that deactivate their Automatic Identification System (AIS) or conceal their identities pose substantial challenges to maritime monitoring. A central scientific problem is whether satellite–AIS mismatch can be interpreted as direct evidence of non-cooperative vessel activity, or whether its evidential meaning must be resolved through spatiotemporal AIS support and maritime geographic context. To address this problem, we develop an evidence-driven Dark Vessel Detection Agent that formulates suspected dark-vessel screening as a target-level evidence-interpretation process. The framework integrates Gaofen-6 vessel-like target detection using YOLOv11 oriented bounding boxes, historical AIS-derived navigation context, rule-based geospatial reasoning, DeepSeek-V4-assisted evidence-consistency review, and image-time AIS trajectory rechecking. In the Pearl River Estuary, China, 825 AIS-unexplained targets were reduced to 361 reviewed candidates. Image-time trajectory assessment subsequently identified 114 candidates as AIS-spatiotemporally explainable, while 247 remained unresolved and were retained as prioritized verification tasks, reducing the initial candidate pool by 70.06%. Selective AIS querying reduced query volume by 32.41% relative to exhaustive rechecking and identified 65.14% of the cases meeting the closure criterion. These results show that satellite–AIS mismatch is not a self-sufficient indicator of non-cooperative vessel activity. Its evidential meaning emerges from the joint interpretation of remote-sensing observations, image-time AIS trajectory support, and maritime geographic context.

Check for updates

Academic Editor: Silvia Liberata Ullo

Received: 10 October 2026
Revised: 10 November 2026
Accepted: 14 November 2026
Published: 17 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: dark vessel screening; geospatial reasoning; maritime remote sensing; satellite–AIS mismatch; AIS explainability

1. Introduction

Maritime monitoring increasingly depends on the joint interpretation of satellite observations and cooperative vessel-tracking information. The Automatic Identification System (AIS) provides extensive records of vessel movement and supports applications ranging from fishing-effort reconstruction to regional and global analyses of maritime activity [1,2]. However, AIS records cooperatively reported vessel activity, and its coverage of vessel presence is inherently incomplete. Satellite-based studies have revealed substantial maritime activity absent from public tracking records [3], while deliberate AIS disabling, identity ambiguity, incomplete reporting, and reception limitations can further complicate the relationship between observed vessels and cooperative tracking information [4–6]. AIS trajectories themselves may also contain missing or noisy observations that affect the reconstruction of vessel movement [7]. Satellite observations and AIS therefore provide complementary but non-equivalent representations of maritime activity.

Optical remote sensing provides an independent observation source for detecting vessel-like targets and describing their surrounding spatial environment [8,9]. Deep learning has substantially improved object detection in remote sensing imagery, supported by large benchmarks and modern detection architectures [10,11]. Recent studies continue to address small-vessel detection in complex nearshore backgrounds and the exploitation of complementary information across optical and SAR observations [12,13]. For vessels, arbitrary heading and elongated geometry make oriented bounding boxes particularly suitable for preserving hull geometry while reducing irrelevant surrounding background [14,15].

Rotation-aware methods such as RoI Transformer, Oriented R-CNN, and R3Det have further demonstrated the importance of orientation-sensitive representations for aerial and remote-sensing targets [16–18]. Nevertheless, vessel detection provides only the observational layer. Interpreting the cooperative status of a detected vessel-like target requires additional tracking and contextual evidence.

Satellite–AIS association provides a means of relating these two observation systems by linking remotely sensed vessels to cooperative tracking records. Recent studies have combined AIS with independent visual or satellite observations for vessel monitoring and identification, demonstrating the value of cross-source association [19,20]. In dense coastal waters, however, this association is inherently uncertain. AIS messages may be irregularly sampled, delayed, missing, or affected by positional and identity inconsistencies, while satellite and AIS observations may also differ in acquisition time and spatial representation [21,22]. Coastal AIS records can additionally underrepresent actual vessel traffic, particularly for small vessels and non-mandatory transmitters [23]. Comparisons between satellite imagery and vessel-tracking data have consequently shown that discrepancies between the two sources can reflect both actual differences in observed activity and limitations in cooperative reporting [24]. An AIS-unexplained satellite target therefore represents an unresolved evidence state whose interpretation requires additional evidence. This interpretation depends on its spatiotemporal and geographic relations. A satellite–AIS mismatch may occur within a persistent navigation corridor, near a port approach or anchorage, along the shoreline, or in less routinely trafficked waters. Historical AIS movement patterns provide an empirical representation of cooperative navigation structures, while port, anchorage, shoreline, and scene information describe complementary dimensions of the local maritime environment. Together, these contextual variables provide relational evidence for

evaluating whether an apparent mismatch is consistent with routine maritime activity, can be explained by additional cooperative information, or remains unresolved. The central scientific problem is therefore whether satellite–AIS mismatch can be interpreted as direct evidence of non-cooperative vessel activity, or whether its evidential meaning must be resolved through the joint interpretation of remote-sensing observations, spatiotemporal AIS support, and maritime geographic context.

Resolving this problem requires a workflow that can organize heterogeneous evidence and selectively obtain additional information when the current evidence state remains insufficient. Recent artificial-intelligence agents provide a computational paradigm for structured reasoning and tool interaction across multiple information sources [25,26]. Tool-augmented approaches such as ReAct and Toolformer demonstrate how reasoning can be coupled with external information retrieval or computation [27,28].

Recent studies using vision and language models in remote sensing have also shown their potential for integrating multiple observations and supporting semantic interpretation, including fine-grained ship recognition [29]. Change-Agent combines a large language model (LLM) with a specialized model for remote sensing change interpretation, while RS-Agent integrates task planning, domain knowledge, and remote sensing tools into executable workflows [30,31]. GIS Copilot connects LLM-generated workflows with QGIS tools, and GeoAgentic-RAG integrates spatial retrieval with executable analysis of vector and raster data [32,33]. For Earth observation applications, language model reasoning is most useful when tied to observable geospatial relations. In the present problem, it is used to examine consistency across evidence sources, identify cases requiring further assessment, and trigger physically grounded queries that update the evidence state.

Accordingly, this study develops an evidence-driven Dark Vessel Detection Agent for evidence triage and verification prioritization of satellite–AIS vessel mismatches in complex coastal waters. Gaofen-6 optical observations first provide georeferenced vessel-like targets, which are associated with cooperative AIS records. Historical AIS trajectories and maritime geographic information are then used to construct target-level evidence packets describing AIS inconsistency, navigation context, and surrounding maritime conditions. A Rule Agent performs reproducible first-stage geospatial screening, after which DeepSeek-V4 reviews cross-source evidence consistency and determines whether additional image-time AIS trajectory analysis is required. For triggered candidates, AIS trajectories are reconstructed at the satellite acquisition time to test whether cooperative vessel movement provides sufficient spatiotemporal explanation. The framework integrates detection, association, contextual reasoning, and trajectory evidence within a single evidence chain to determine the final state of each candidate.

The framework's methodological contribution is a target-level agent architecture that links multisource evidence interpretation to selective AIS trajectory queries and explicit evidence-state updates. This architecture connects detection and reconstruction tools through a common decision protocol to generate traceable verification tasks. The comparative evaluation examines how prioritization and query-selection policies affect task generation.

The main contributions of this study are threefold:

- A target-level evidence formulation for satellite–AIS mismatch. Satellite detections without sufficient AIS correspondence are represented as intermediate evidence states whose evidential meaning is interpreted through spatiotemporal and geographic relations.
- An auditable multi-source evidence agent for Earth-observation interpretation. The framework integrates explicit geospatial screening, evidence consistency review with

an LLM, and selective AIS trajectory analysis at the image acquisition time; final evidence closure is anchored to reconstructed vessel positions.

- A traceable transition from unresolved observations to verification information. The framework separates AIS-spatiotemporally explainable mismatch from evidence that remains unresolved and converts the latter into spatially prioritized verification tasks while preserving target-level links among satellite observations, AIS evidence, and maritime geographic context.

2. Materials and Methods

2.1. Study Area and Datasets

The study was conducted in the Pearl River Estuary, China, a densely trafficked coastal environment characterized by intensive port operations, anchorage activity, and complex river–sea navigation. Regular shipping routes, nearshore vessel activity, and anchorage-related movement create diverse spatial contexts for examining satellite–AIS mismatch and its geographic interpretation.

GF-6 panchromatic and multispectral (PMS) imagery covering the Pearl River Estuary was used as the optical observation source. The PMS instrument provides a 2 m panchromatic band (0.45–0.90 μm) and four 8 m multispectral bands, including blue (0.45–0.52 μm), green (0.52–0.59 μm), red (0.63–0.69 μm), and near-infrared (0.77–0.89 μm), with a nominal revisit time of 4 days. The imagery underwent radiometric calibration, atmospheric correction, orthorectification, and pan-sharpening to produce 2 m multispectral imagery for vessel detection and spatial context interpretation.

AIS data provided cooperative vessel information at three temporal scales. The 10 min ship-level product supported the initial association between satellite detections and nearby vessel records. The 2 h product supported AIS trajectory reconstruction at the GF-6 acquisition time, while the one-month archive was used to characterize persistent navigation patterns. Port, anchorage, and shoreline layers provided additional information on the local maritime environment. Together, these data formed the satellite observation, AIS, and geographic context inputs of the evidence workflow summarized in Table 1.

Table 1. Data sources and evidence layers used in the maritime-agent framework.

Data Layer	Source/Product	Temporal Information	Role
Optical remote-sensing imagery	GF-6 PMS L1A	GF-6 acquisition time	Vessel-like target detection
Initial AIS evidence	10 min AIS product	Around image time	Initial AIS check
Image-time AIS evidence	2 h AIS trajectory product	Around image time	AIS trajectory recheck
Historical AIS context	One-month AIS archive	Monthly records	Navigation-context extraction
Maritime context	Port, anchorage, shoreline layers	Study area	Context constraints

2.2. Overall Framework of the Evidence-Driven Agent

As shown in Figure 1, the framework organizes satellite observations, cooperative AIS information, and maritime geographic context into a common evidence workflow. It consists of four components: remote sensing observation (A), AIS and maritime context construction (B), evidence reasoning (C), and evidence closure and verification task generation (D).

The observation component detects vessel-like targets from GF-6 PMS imagery using YOLOv11-OBB. Tile detections are transformed to study area image coordinates, geolocated with the rational polynomial coefficient model, and spatially deduplicated. Each georeferenced target then serves as a common analysis unit for integrating image observations, AIS information, and geographic context.

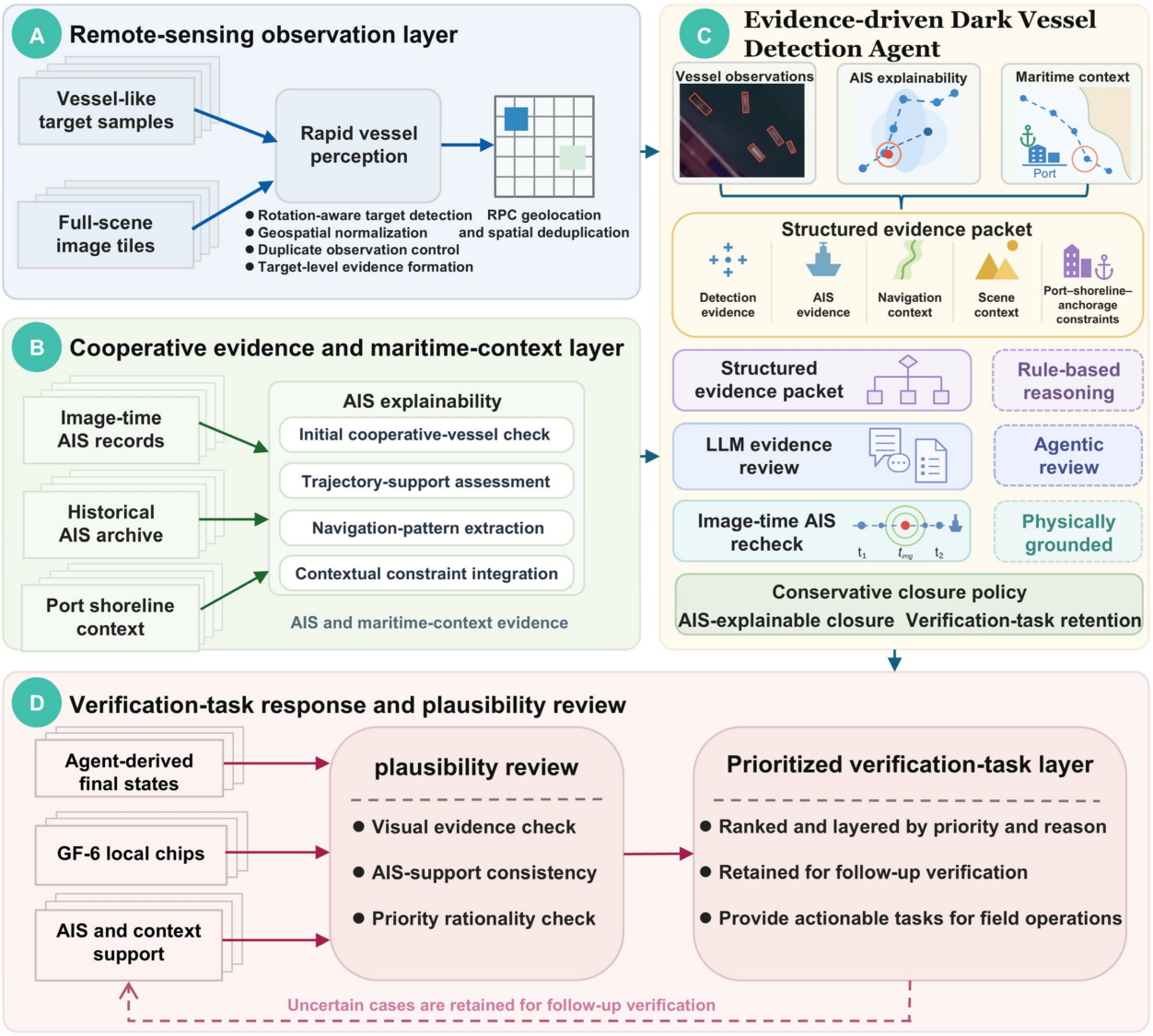


Figure 1. Evidence-driven maritime-agent framework for suspected dark-vessel candidate screening. (**A**) Remote-sensing observation; (**B**) AIS evidence and maritime context; (**C**) Evidence reasoning and closure; (**D**) Plausibility review and verification-task generation. Colored frames group workflow components, stacked rectangles represent multiple inputs, red boxes mark vessel-like targets, and circles highlight selected locations or schematic AIS matching neighborhoods.

For each target, the AIS and context component first evaluates correspondence with cooperative vessel records. Historical AIS trajectories are used to characterize navigation patterns, while port, anchorage, shoreline, and scene information describe the surrounding maritime environment. These observations are assembled into an evidence record containing target location, detection confidence, AIS association state, spatial separation, navigation context, and local maritime attributes.

The evidence record is then processed by the agent. The Rule Agent applies explicit geospatial criteria and assigns an initial verification priority. DeepSeek-V4 reviews consistency across the available evidence and identifies candidates requiring additional AIS trajectory analysis. For these candidates, vessel trajectories are reconstructed at the satellite acquisition time and compared with the corresponding remote sensing observations. The closure module combines the reviewed priority with the reconstructed AIS evidence to assign each candidate to an explainable state or a prioritized verification task. Target identifiers and evidence attributes are retained throughout the workflow, allowing each final

state to be traced across observation, association, context construction, evidence review, and trajectory assessment.

2.3. Vessel-like Target Detection and Geolocation

The GF-6 PMS imagery covering the study area was partitioned into tiles measuring 600 × 600 pixels using a sliding window with an overlap of 64 pixels and a stride of 536 pixels. The overlap reduced target truncation at tile boundaries and preserved the continuity of elongated vessels extending across adjacent tiles. Vessel-like targets were manually annotated using oriented bounding boxes in the Computer Vision Annotation Tool (CVAT; CVAT.ai Corporation, Wilmington, DE, USA; https://www.cvat.ai/) to represent target orientation and hull geometry while limiting surrounding background. The tiles were randomly divided into training, validation, and test sets containing 799, 100, and 102 images, respectively. A YOLO11n-OBB detector was implemented using Ultralytics version 8.4.30 (Ultralytics Inc., Frederick, MD, USA) and initialized with the yolo11n-obb.pt weights pretrained on DOTA-v1.0 and fine-tuned on the training set for 100 epochs using 600 × 600 pixel inputs and the AdamW optimizer. The trained detector was applied to the study area imagery through tile-wise inference [34]. The inference input size was 600 pixels, with a confidence threshold of 0.25 and an intersection over union threshold of 0.50.

Detector outputs were initially expressed in local tile coordinates. Each target was transformed to the study area image coordinate system using the corresponding tile offsets.

$$x_{global} = x_{offset} + x_{local} \quad (1)$$

$$y_{global} = y_{offset} + y_{local} \quad (2)$$

The resulting image coordinates were converted to geographic coordinates using the rational polynomial coefficient model. A 30 m spatial clustering step was then used to consolidate nearby detections. Georeferenced detections were treated as graph nodes, with edges connecting detection centers separated by no more than 30 m according to the Haversine distance. Connected components formed the clusters, including detections linked transitively through neighboring nodes. One representative was retained per cluster, prioritizing valid coordinates and successful RPC status, followed by higher detection confidence, larger OBB area, and earlier original record order. OBB overlap, tile origin, vessel length, and heading were not used as clustering constraints. The final georeferenced records retained detection confidence, oriented geometry, and geographic position for subsequent AIS association and geographic context construction.

2.4. AIS Association and Navigation Context Construction

The 10 min ship-level AIS product was used to associate georeferenced vessel detections with cooperative vessel records around the satellite acquisition time. For each target, nearby AIS records were evaluated using geodesic distance and heading consistency. A base matching radius of 500 m and a maximum heading difference of 70° were applied. An adaptive matching radius was also calculated for each target to represent its spatial association tolerance. The nearest AIS record, target–AIS distance, heading difference, adaptive radius, and association state were retained as evidence attributes. Targets without an AIS record satisfying the spatial and heading criteria were assigned an AIS-unexplained state and passed to geographic context construction. This association step preserved both the correspondence state and the magnitude of the spatial discrepancy between the satellite observation and nearby cooperative vessel records.

A one-month AIS archive was used to characterize persistent navigation patterns in the Pearl River Estuary. After spatial clipping and speed-based quality control, AIS positions were aggregated into grid cells measuring 500 × 500 m. The 0.90 density quantile defined the

navigation core, and the 0.80 quantile defined a broader influence zone. Spatial overlay and proximity analysis then assigned each target to one of four navigation contexts: channel core, channel influence, near channel, or non-channel. Port, anchorage, and shoreline information was integrated through spatial overlay and distance analysis. Each target was assigned a maritime scene category, anchorage context, shoreline distance, and nearest port distance. These attributes described the local geographic setting of each satellite–AIS relation. The AIS association attributes and geographic context variables were then assembled into a common target-level evidence record for subsequent agent screening.

2.5. Evidence Screening and Agent Review

The Rule Agent screened targets that remained unexplained after the initial AIS association. Each target was represented by an evidence record containing detection confidence, AIS association state, nearest AIS distance, adaptive matching radius, navigation context, scene category, anchorage context, shoreline distance, and port distance. This representation placed satellite observations, AIS information, and geographic context within a common target-level structure. The Rule Agent converted these attributes into an evidence score using five groups of information: AIS inconsistency, detection confidence, navigation context, maritime scene context, and port, shoreline, and anchorage constraints. Greater evidential weight was assigned to combinations associated with stronger AIS inconsistency and relevant navigation or scene conditions, while routine coastal contexts reduced the score. The resulting score was mapped to High, Medium, or Low verification priority using predefined thresholds.

A target entered the subsequent review stage when it was a 30 m spatial representative, remained unmatched under the distance criterion, had a detection confidence of at least 0.50, and satisfied the geographic screening conditions. Retained targets were ordered using verification priority, Rule Agent score, navigation context, and detection confidence. The detailed scoring and screening rules are summarized in Supplementary Table S1.

DeepSeek-V4 (configured API model identifier: deepseek-v4-flash; Hangzhou DeepSeek Artificial Intelligence Co., Ltd., Hangzhou, China; https://api.deepseek.com) then reviewed the structured evidence record together with the initial priority. The review examined consistency among the satellite observation, AIS relation, and geographic context and returned four structured fields: reviewed verification priority, evidence interpretation, review decision, and AIS trajectory recheck requirement. The model was run with a temperature of 0.0, and responses followed predefined output fields. The input variables, output structure, and review settings are summarized in Supplementary Table S2. The LLM-reviewed priority replaced the rule priority when valid; otherwise, the rule priority was retained. The reviewed evidence state also determined whether additional AIS trajectory analysis was required. Cases selected for rechecking were passed to the image-time trajectory assessment described below. AIS trajectory rechecking was triggered when the LLM-assessed risk was High or Very High, the false-alarm risk was Low or Medium, and the reviewed verification priority was High or Medium.

2.6. AIS Trajectory Rechecking and Evidence Closure

AIS trajectory rechecking was applied to candidates selected during evidence review. AIS records within two hours before and after the GF-6 acquisition time were grouped by Maritime Mobile Service Identity. When valid observations bracketed the acquisition time, vessel position was estimated by trajectory interpolation with a maximum interpolation gap of 30 min. When interpolation was unavailable, the nearest AIS observation was used if its temporal offset was within 10 min. The geodesic distance between the reconstructed vessel position and the corresponding satellite detection was then calculated at the acquisition time.

The reconstructed distance was evaluated using a target-specific adaptive evidence radius. This radius represented the spatial tolerance associated with target–AIS separation, image geolocation uncertainty, AIS reporting interval, vessel movement, and detection localization error. Based on spatial agreement and temporal support, each rechecked candidate was assigned one of three AIS evidence states: spatiotemporally explainable, explainable with low confidence, or ambiguous near the matching boundary. Candidates were assigned to the explainable closure state when their interpolated AIS positions fell within the adaptive evidence radius and the interpolation gap did not exceed 30 min.

The closure module combined the reviewed verification priority with the AIS evidence state. Candidates with sufficient spatiotemporal support were assigned an explainable state. Candidates with low confidence support, ambiguous correspondence, or no conclusive trajectory recheck retained their reviewed priority and remained in the verification queue. The outputs consisted of target-level evidence states and spatially indexed high, medium, and low-priority verification tasks. The principal parameter settings are summarized in Supplementary Table S4.

2.7. Stratified Blind Manual Plausibility Review

A stratified blind review was conducted to assess the consistency of the final evidence states. From the 361 final candidates, 120 targets were selected using stratified random sampling with a fixed random seed of 2026. The sample contained 30 targets from each of four groups: AIS-explainable closures, high-priority tasks, medium-priority tasks, and low-priority tasks. Equal sampling across the four groups enabled direct comparison of review outcomes among the final states.

The sampled targets were randomly reordered and assigned anonymized identifiers from R001 to R120. One researcher independently labeled all 120 anonymized evidence cards. During review, the original target identifier, final agent state, and verification priority were concealed. Each evidence card contained the GF-6 local image chip with the detected target marked, detection confidence and oriented box geometry, reconstructed AIS evidence, adaptive evidence radius, and navigation, port, anchorage, and shoreline information. Review considered the visual characteristics of the target, AIS correspondence at the acquisition time, and surrounding geographic context.

Five labels were defined: AIS-explainable, worth verification, routine context or low concern, uncertain, and non-vessel-like or false detection. AIS-explainable indicated a visible vessel-like target with sufficient reconstructed AIS support within the adaptive evidence radius. Worth verification indicated a visually supported target for which AIS evidence remained insufficient for closure. Routine context or low concern described targets whose surrounding maritime context supported lower verification urgency. Uncertain indicated inconclusive image, AIS, or contextual evidence, while non-vessel-like or false detection indicated insufficient visual support for a vessel-like target. After review, the anonymized identities were restored and the manual labels were compared with the corresponding agent states and priority groups. The review provided a blind plausibility check of the final evidence structure. These independently reviewed labels formed the reference for the paired ranking comparisons.

2.8. Evaluation of Task Reduction, Prioritization, and AIS Query Selection

We evaluated verification-task reduction, candidate prioritization, and AIS query selection in a retrospective comparison. Task reduction was quantified across rule screening and AIS evidence closure. The overall task-reduction rate was calculated as $(825 - 247)/825$, and the closure-stage rate as $(361 - 247)/361$. For AIS querying, closure coverage was the number of closures divided by the exhaustive-reference closure total, closure precision was

closures divided by queries, and query reduction was the decrease in query calls relative to exhaustive rechecking.

To evaluate query selection, the same image-time AIS recheck algorithm was applied to all 361 candidates, including the 117 not queried in the original workflow. The algorithm, AIS data, acquisition time, and closure criteria were held fixed. The recheck radius was the maximum of 1000 m, twice the initial adaptive radius, and the nearest AIS distance plus 500 m, bounded between 1000 and 6000 m. The search window was ±120 min, the maximum interpolation gap was 30 min, and the nearest-time threshold was 10 min. Replaying 12 archived targets reproduced their classifications and distances before the additional evaluations.

At a common budget of 244 AIS queries, we compared the LLM trigger with three baselines: descending continuous Rule Agent scores, three-level rule priority, and uniform random selection. Rechecking all 361 candidates provided the exhaustive closure reference. Ties were resolved by uniform random selection within equal scores. For tied-score selection and uniform random sampling, expected closure counts and central 95% randomization ranges were calculated analytically using hypergeometric distributions, conditional on the fixed candidate pool and AIS closure reference.

Additional query-budget comparisons evaluated retrospective LLM-priority Top-K selection and rule-first LLM tie-breaking, which used LLM priority only within equal continuous-rule scores. AIS-closure rankings were assessed using average precision (AP) and normalized discounted cumulative gain (nDCG). For nDCG, AIS-explainable targets received relevance 1 and remaining targets received 0. Relevance at rank i was discounted by $1/\log_2(i + 1)$, and the summed gain was normalized by the ideal ordering over all 361 candidates.

Verification rankings were evaluated against the human-review labels, with worth verification as the positive class. We computed design-weighted AP and precision at a selection budget of K targets (P@K) using the original stratum population-to-sample ratios as weights. Selection probabilities were determined from the complete candidate pool in each evaluation scope. Uncertain labels were excluded from the primary analysis and assigned to either class in sensitivity analyses. Differences in P@K were assessed using paired bootstrap resampling within the four original review strata (5000 replicates; random seed 20261105).

3. Results

3.1. Vessel-like Target Detection and Spatial Consolidation

The YOLOv11-OBB detector achieved a precision of 84.79%, a recall of 78.03%, an mAP50 of 84.88%, and an mAP50–95 of 61.07% on the validation set. Precision exceeded recall by 6.76 percentage points, while the 23.81 percentage point difference between mAP50 and mAP50–95 indicates greater sensitivity to localization accuracy under stricter overlap criteria.

Study area inference produced 3219 raw vessel-like detections. After geographic positioning and 30 m spatial clustering, 2544 spatial representatives remained, corresponding to a retention rate of 79.03%. This corresponded to a 20.97% reduction in detection records. Spatial consolidation provided the representative target set for subsequent AIS association and evidence analysis.

3.2. AIS-Derived Navigation Structure and Geographic Context

The one-month AIS archive contained 425,895,863 position records. After spatial clipping and speed-based quality control, 140,192,625 records (32.92%) were retained and

aggregated into 500 × 500 m grid cells. The resulting density surface represented the persistent spatial organization of cooperative vessel movement across the Pearl River Estuary.

At the 0.90 density quantile, the threshold was 4613 AIS positions per cell, producing 5600 core cells and 84 polygon components. At the 0.80 quantile, the threshold decreased to 1821 positions per cell, while the mapped area expanded to 11,204 cells and 79 polygon components. The increase in occupied cells accompanied by fewer polygon components indicates that the broader layer mainly connected and expanded adjacent traffic branches rather than introducing additional isolated activity zones.

As shown in Figure 2, vessel movement exhibited a strongly directional and hierarchical spatial structure. The navigation core followed persistent trunk routes, whereas the broader influence zone extended toward port approaches, anchorage-related movement areas, and river–sea transition zones. These layers provided a graded geographic representation of the cooperative traffic background. Navigation context was subsequently considered together with AIS separation and local maritime attributes, allowing satellite–AIS mismatch to be interpreted relative to the spatial structure in which it occurred.

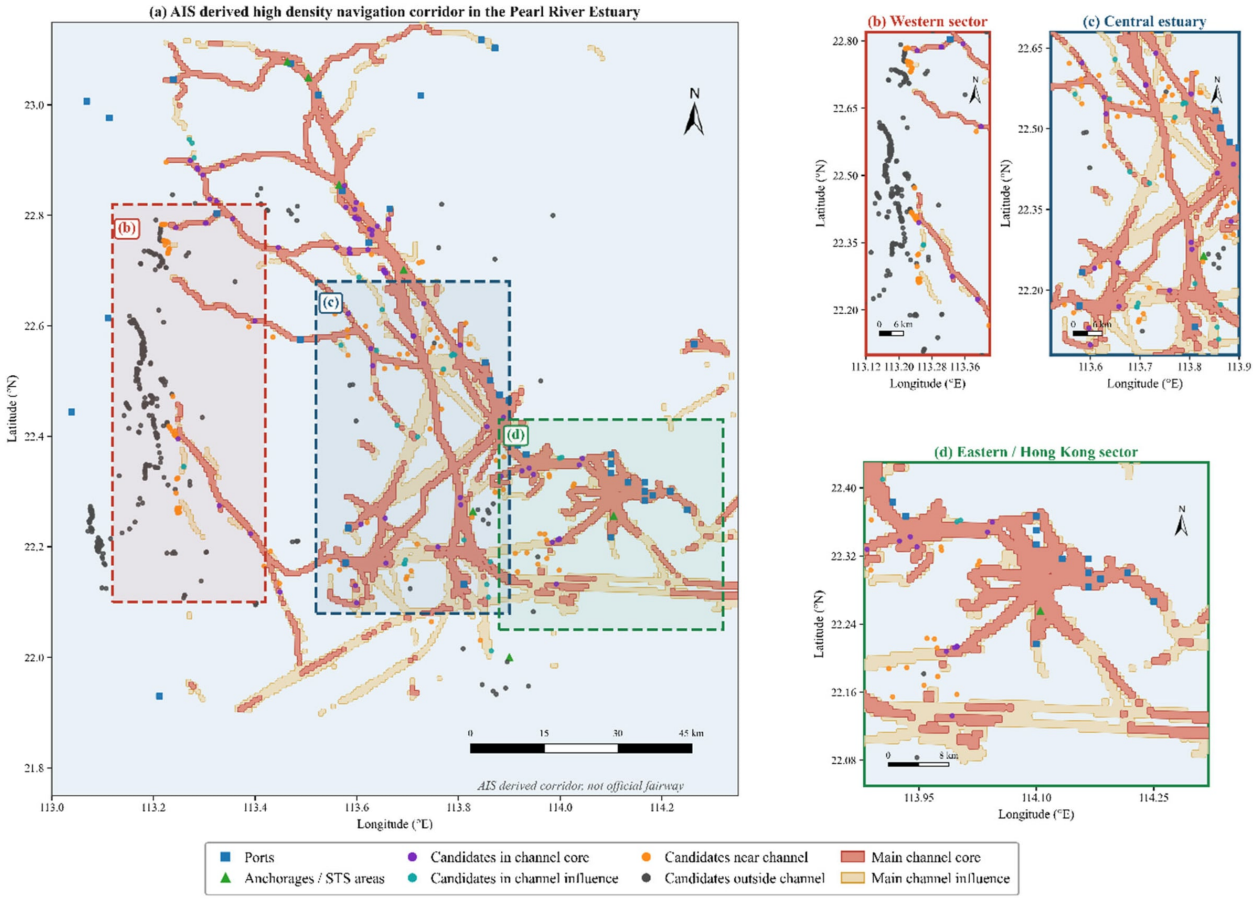


Figure 2. AIS-derived navigation-corridor evidence and candidate spatial context in the Pearl River Estuary. (**a**) Overview of the study area. (**b**) Enlarged view of the western sector. (**c**) Enlarged view of the central estuary. (**d**) Enlarged view of the eastern/Hong Kong sector. Candidate points represent the rule-screened candidate pool entering LLM review, before AIS-based evidence closure. Point colors indicate navigation context. STS, ship-to-ship transfer.

3.3. Candidate Screening Cascade and Evidence Prioritization

Figure 3 and Table 2 summarize the progressive reduction and evidence transitions across the workflow. Initial AIS association explained 67.6% of the spatial representatives and left 825 AIS-unexplained targets. Rule Agent screening subsequently retained

361 candidates, corresponding to 43.8% of the unresolved set. These stage-specific changes show that cooperative AIS correspondence accounted for the largest initial reduction, while geographic evidence was applied to differentiate the remaining mismatch observations.

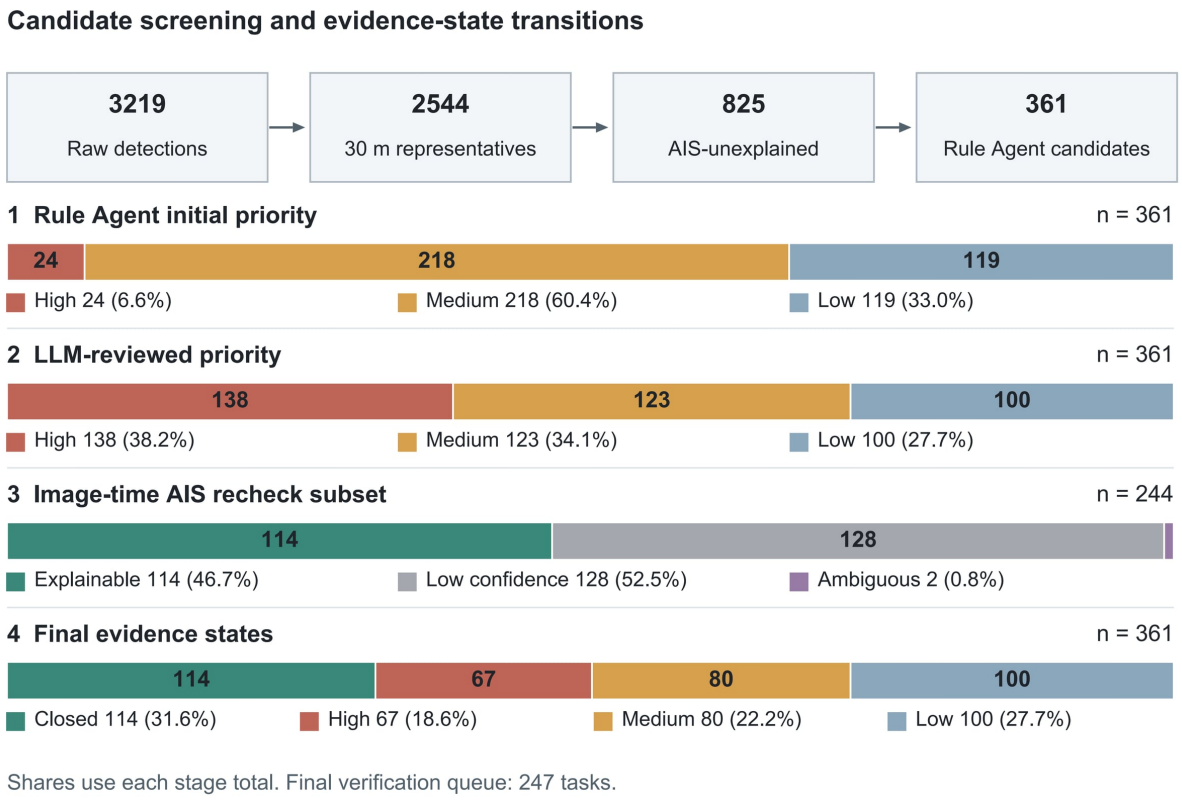


Figure 3. Multi-stage candidate screening cascade of the evidence-driven maritime agent. Percentages use the stage-specific totals shown.

Table 2. Stage contributions of the evidence-driven agent workflow.

Stage	Input	Operation	Output	Rate
Remote-sensing detection	Full scene	YOLOv11-OBB inference	3219 raw detections	—
Geographic deduplication	3219	30 m merging	2544 representatives	79.0%
Initial AIS association	2544	AIS screening	825 AIS-unexplained	32.4%
Rule Agent screening	825	Evidence scoring	361 candidates	43.8%
LLM evidence review	361	Priority review	244 recheck cases	67.6%
AIS trajectory recheck	244	Spatiotemporal test	114 closures	46.7%
Final task generation	361	Evidence closure	247 tasks	68.4%

Note: Rates describe stage-specific retention or closure relative to the input of each stage.

The Rule Agent excluded 464 targets, or 56.2% of the AIS-unexplained pool, by jointly considering detection confidence, AIS separation, navigation context, and proximity to ports, shorelines, and anchorages. This stage therefore concentrated the residual mismatch observations into a smaller set with evidence configurations requiring further review.

The 361 retained candidates comprised 24 high-priority, 218 medium-priority, and 119 low-priority targets, accounting for 6.6%, 60.4%, and 33.0% of the candidate set, respectively. Among these candidates, 331 (91.7%) had detection confidence values of at least 0.60. The dominance of the medium-priority group was therefore not driven by generally weak image detections. Instead, most candidates combined visible vessel evidence with incomplete AIS correspondence, but their overall evidence configurations did not place them in the highest priority group.

Figure 3 further separates candidate reduction from evidence-state transition. Rule Agent screening narrows and orders the unresolved set, whereas subsequent evidence review and AIS trajectory assessment update the state of individual candidates. Target identifiers and contextual attributes remain linked across these stages, allowing each final state to be traced to its satellite observation, AIS relation, and geographic context.

3.4. Agent Evidence Review and AIS Trajectory Reassessment

DeepSeek-V4 reassessed all 361 Rule Agent candidates. The priority distribution shifted from 24 high-, 218 medium-, and 119 low-priority candidates to 138 high-, 123 medium-, and 100 low-priority candidates. The high-priority share consequently increased from 6.6% to 38.2%, whereas the medium-priority share decreased from 60.4% to 34.1%. This redistribution shows that evidence review altered the initial ordering of candidates when AIS separation, navigation context, scene information, and detection confidence were considered jointly.

Priorities increased for 133 candidates and remained unchanged for 228; none were lowered. The four cases carrying the decision label downgrade-or-low-priority retained their numeric priority and were counted in the unchanged group. The increases comprised 114 changes from medium to high priority and 19 from low to medium priority.

Of the 361 reviewed candidates, 248 met the risk and false-alarm conditions; four had Low reviewed priority, leaving 244 candidates (67.6%) that met all three recheck conditions. The remaining 117 retained their reviewed priority without trajectory reconstruction. Among the 244 reassessed candidates, 114 (46.7%) were spatiotemporally explainable under the closure criterion, 128 (52.5%) had low-confidence AIS support, and two (0.8%) remained near the matching boundary. The 114 explainable cases represented 31.6% of the 361 reviewed candidates.

Sensitivity analysis showed comparable closure counts across adaptive radius scales of 0.70–1.30 and maximum interpolation gaps of 10–60 min, as shown in Figure S4. The overall evidence-state pattern therefore remained locally stable under moderate changes in spatial and temporal parameters. Together, agent review and AIS trajectory reassessment transformed the initial candidate priorities into a sequence of updated evidence states linked to reconstructed vessel movement.

3.5. Evidence Closure and Prioritized Verification Tasks

Of the 361 reviewed candidates, 114 (31.6%) were assigned to the AIS-spatiotemporally explainable state, while 247 (68.4%) remained in the verification queue. Relative to the initial 825 AIS-unexplained targets, the final verification set represented a 70.06% reduction in candidate task count. The combined evidence workflow therefore concentrated the unresolved observations into a substantially smaller set while closing cases supported by reconstructed AIS trajectories.

As shown in Table 3, the 247 retained candidates comprised 67 high-priority, 80 medium-priority, and 100 low-priority tasks, accounting for 27.13%, 32.39%, and 40.49% of the verification queue, respectively. High-priority tasks thus formed a relatively compact subset of the retained observations, while medium- and low-priority cases remained available for subsequent assessment or additional evidence collection.

Figure 4 summarizes the transition from reviewed evidence to final agent states, and Figure 5 shows the spatial distribution of explainable cases and retained verification tasks. Explainable closures were more concentrated within the core and influence zones of the AIS-derived navigation corridors than retained verification candidates (85.09% versus 71.26%). Retained candidates had a shorter median distance to the shoreline (9.08 versus 17.94 km) and to the nearest anchorage (19.74 versus 22.47 km) (Table 3b). Their final states were

differentiated by reconstructed AIS correspondence at the acquisition time together with the surrounding geographic context. The resulting task layer preserved target location, reviewed priority, AIS evidence state, and maritime context for subsequent verification.

Table 3. Final closure states and verification priorities (**a**), and geographic context by final evidence state (**b**). (**a**) Final closure states and verification priorities. (**b**) Geographic-context comparison between explainable closures and retained candidates.

(**a**)

Final State	Decision	Count	Share	Interpretation
AIS-spatiotemporal explainable	Closed	114	31.6%	Sufficient AIS evidence
Low-confidence AIS support	Retained	128	35.5%	Incomplete AIS support
No conclusive AIS recheck	Retained	117	32.4%	Unresolved evidence
Ambiguous AIS boundary	Retained	2	0.6%	Boundary uncertainty
High-priority task	Verification	67	27.1%	Strong unresolved evidence
Medium-priority task	Verification	80	32.4%	Moderate unresolved evidence
Low-priority task	Verification	100	40.5%	Lower urgency

(**b**)

Geographic Characteristic	Explainable Closures (n = 114)	Retained Candidates (n = 247)
Navigation-corridor core	76 (66.67%)	143 (57.89%)
Navigation-corridor influence	21 (18.42%)	33 (13.36%)
Near navigation corridor	16 (14.04%)	31 (12.55%)
Outside navigation-corridor context	1 (0.88%)	40 (16.19%)
Distance to nearest anchorage (km)	22.47 [10.29, 56.86]	19.74 [10.27, 39.32]
Distance to shoreline (km)	17.94 [7.65, 33.56]	9.08 [2.72, 30.62]

Note: Shares for closure states are calculated over all 361 final candidates, whereas shares for high-, medium-, and low-priority tasks are calculated over the 247 retained verification tasks. Values in panel (b) are n (%) for navigation context and median [first quartile, third quartile] for distances.

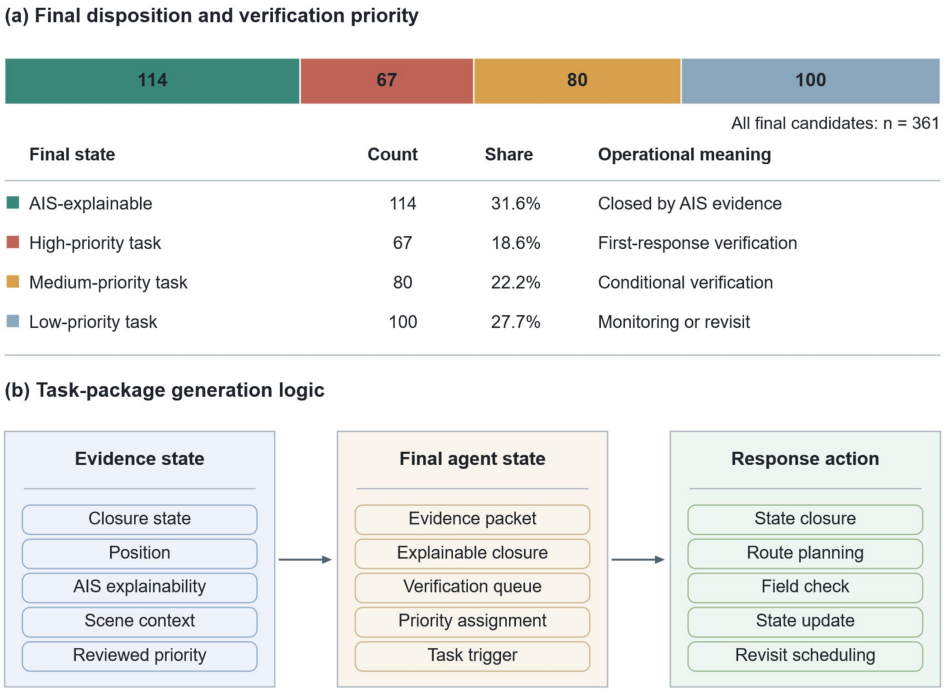


Figure 4. Final evidence closure and low-altitude or manual verification task generation. (**a**) Final disposition of all 361 candidates. (**b**) Task-package generation logic.

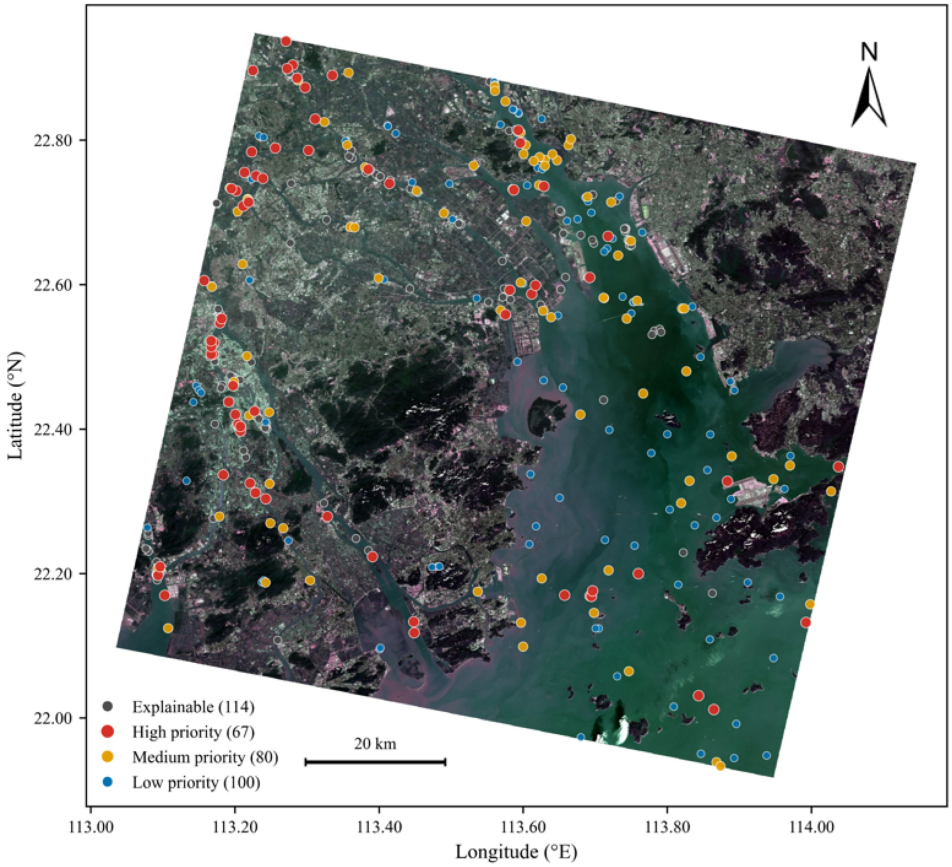


Figure 5. Spatial distribution of explainable closures and retained verification candidates within the study area.

3.6. Blind Plausibility Review of Final Evidence States

The stratified blind review showed strong consistency between the manually interpreted evidence and the final evidence states. All 30 sampled AIS-explainable closures received the AIS-explainable review label, corresponding to 100% consistency within this group. In these cases, visible vessel-like targets were accompanied by reconstructed AIS positions that provided sufficient support within the adaptive evidence radius. The Wilson 95% confidence interval for this 100% agreement was 88.6–100.0%.

The retained verification tasks showed a clear priority gradient. Among the high-priority samples, 29 of 30 were classified as worth verification and one as uncertain, giving a verification-worthy rate of 96.7%. The corresponding rates were 83.3% for medium-priority samples and 66.7% for low-priority samples. Across all 90 retained-task samples, 74 targets (82.2%) were classified as worth verification, 12 (13.3%) as routine context or low concern, and four (4.4%) as uncertain. All 120 reviewed samples retained sufficient visual support to be considered vessel-like.

As shown in Figure 6, visually supported vessel-like targets occurred across all four final-state groups, while their disposition differed according to reconstructed AIS correspondence and surrounding geographic context. Because equal numbers were sampled from each final-state group, these percentages represent within-group plausibility patterns rather than population-wide rates.

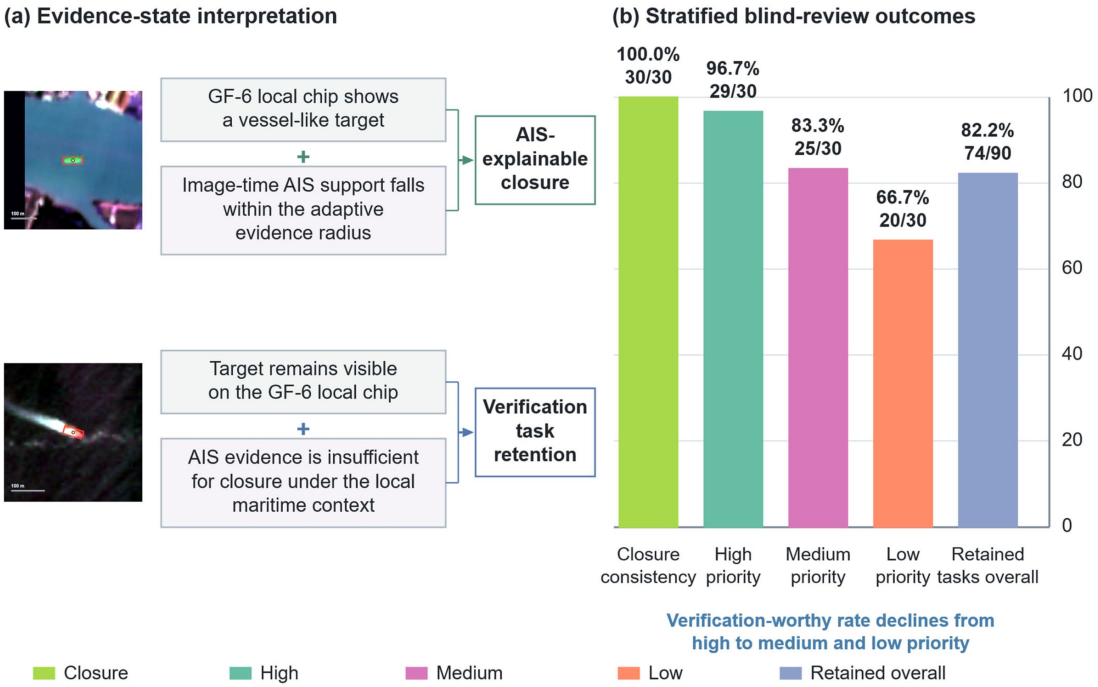


Figure 6. Blind plausibility review of final evidence states. (**a**) Representative GF-6 local chips and evidence criteria for AIS-explainable closure and verification-task retention. (**b**) Closure consistency and verification-worthy rates within the sampled groups. Each final-state group contains 30 samples; the overall retained-task rate uses the 90 sampled verification tasks.

3.7. Baseline Comparisons and AIS Query-Selection Performance

The workflow reduced 825 initially AIS-unexplained targets to 247 verification tasks, an overall reduction of 70.06%. Geographic screening excluded 464 targets, and image-time AIS evidence closed a further 114. LLM review assigned priorities and selected 244 candidates for trajectory queries. AIS closure reduced the 361 rule-retained candidates by 31.58%.

Rechecking the 117 candidates outside the original query subset identified 61 additional AIS-explainable cases. Exhaustive rechecking therefore yielded 175 closures and 186 retained tasks. The selective policy used 244 queries rather than 361, a reduction of 32.41%, and identified 114 of the 175 closures (65.14%).

At 244 queries, continuous rule scoring yielded an expected 122.9 closures, followed by uniform random selection (118.3), the original LLM trigger (114), and categorical rule priority (107.2) (Table 4). The original LLM trigger exceeded the expected yield of categorical rule priority, while continuous rule scoring produced the highest closure yield. The observed LLM closure count was above the categorical-rule range, within the uniform-random range, and below the continuous-rule range.

Table 4. Query-selection baselines under the common image-time AIS closure rule.

Query Policy	AIS Queries	Closures	Closure Range (Central 95%)	Retained Tasks	Closures/ Query
Rule priority	244	107.2 *	106–108	253.8 *	43.9% *
Continuous rule score	244	122.9 *	121–125	238.1 *	50.4% *
Uniform random	244	118.3 *	110–127	242.7 *	48.5% *
Original LLM trigger	244	114	—	247	46.7%
Exhaustive recheck	361	175	—	186	48.5%

Note: * Expected values under uniform random selection within tied scores or uniform random sampling. Ranges are central 95% randomization ranges conditional on the fixed candidate pool and AIS closure reference. Dashes denote fixed observed selections, for which randomization ranges do not apply. The four selective policies used 244 queries; the exhaustive reference used 361.

The comparison across query budgets showed a different ordering at lower budgets (Figure 7). At 72 queries, retrospective LLM-priority Top-K selection yielded 37.04 expected closures, compared with 29.53 for continuous rule scoring, an increase of 25.43%. At 120 queries, the corresponding values were 61.74 and 54.92, an increase of 12.43%. Across all 361 candidates, AIS-closure AP was 0.4908 for LLM priority and 0.4781 for continuous rule scoring; expected nDCG was 0.8470 and 0.8325, respectively (Table 5). These results describe the ordering of AIS-explainable targets across query positions, whereas Table 4 reports the original trigger at its implemented query budget.

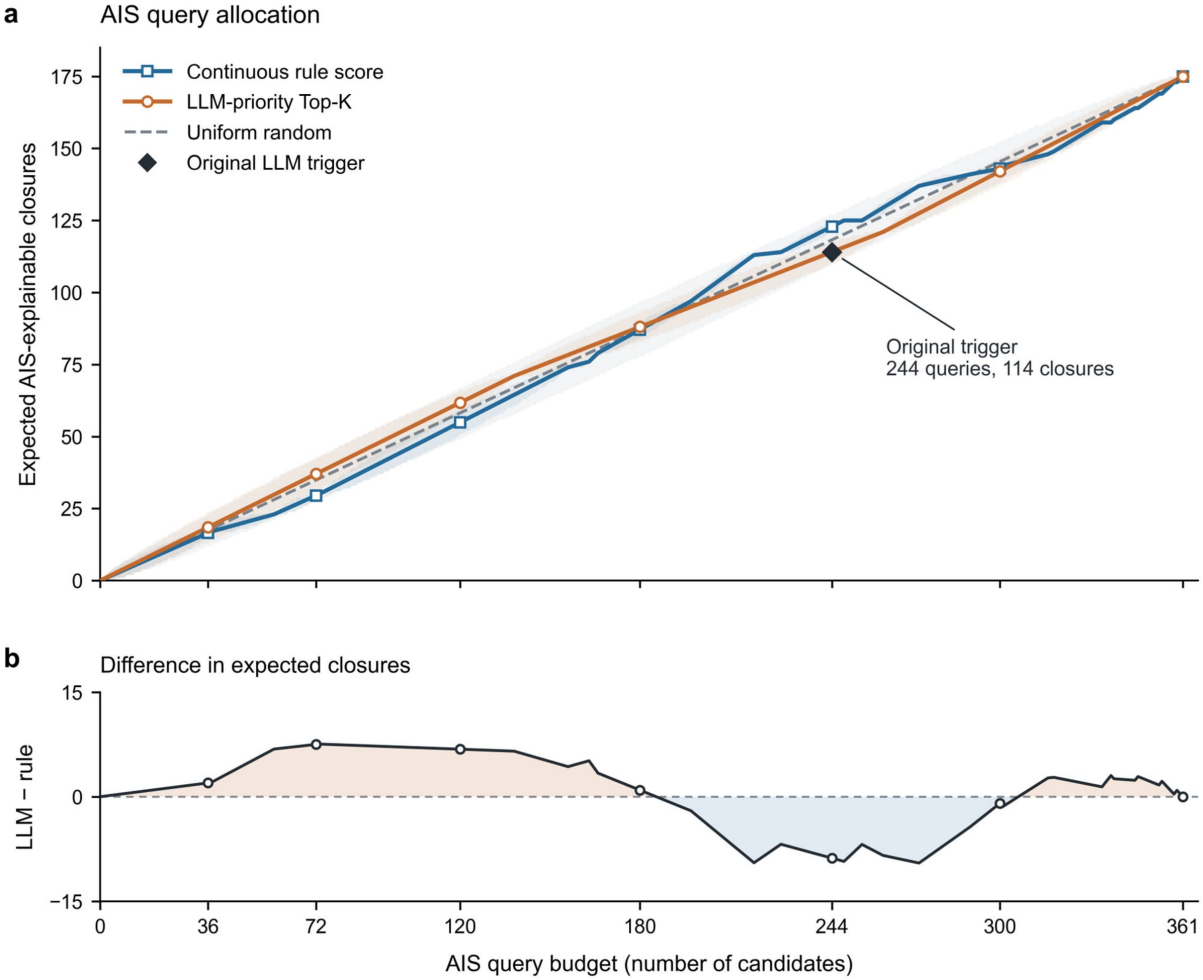


Figure 7. AIS closure yield across query budgets. (**a**) Expected closures for continuous rule scoring, LLM-priority Top-K selection, and uniform random selection using the same 361 archived candidates. Shading shows central 95% range from random selection within score ties. The diamond marks the original LLM trigger (244 queries, 114 closures); Top-K is a separate selection policy. (**b**) Expected closure difference between LLM-priority Top-K and continuous rule scoring. Positive values favor Top-K.

Before AIS closure, design-weighted human-review P@20 and P@40 were both 44.80% for LLM priority, compared with 41.34% and 43.28% for continuous rule scoring (Table 5). The differences were 3.46 percentage points (95% paired-bootstrap interval, −23.81 to 26.98) and 1.52 percentage points (−23.77 to 23.32), respectively. At P@60, continuous rule scoring was higher, at 51.22% versus 44.80%. After common original AIS closure, weighted AP was 0.9101 for categorical rule priority, 0.9864 for continuous rule scoring, 0.9265 for LLM priority, and 0.9866 for rule-first LLM tie-breaking. After exhaustive closure, the corresponding values were 0.9469, 0.9958, 0.9636, and 0.9958. LLM priority therefore showed higher point

estimates at the two smaller pre-closure selection budgets, while continuous rule scoring retained stronger overall human-review ranking performance.

Table 5. Query-selection efficiency and verification-ranking performance of the four ranking methods.

Metric and Setting	Categorical Rule	Continuous Rule	LLM Priority	Rule-First LLM Tie-Break
(a) AIS closure among 361 candidates				
Closure precision at 72 queries (%)	44.33	41.02	51.45	40.97
Closure precision at 120 queries (%)	44.03	45.76	51.45	45.28
Closure precision at 244 queries (%)	43.92	50.37	46.76	50.36
AIS closure AP (361 candidates)	0.4577	0.4781	0.4908	0.4770
AIS closure nDCG (361 candidates)	0.8277	0.8325	0.8470	0.8345
(b) Human review before closure (116 reviewed targets)				
Worth-verification P@20 (%)	54.03	41.34	44.80	41.36
Worth-verification P@40 (%)	52.68	43.28	44.80	42.40
Worth-verification P@60 (%)	51.96	51.22	44.80	50.99
(c) Human review after common AIS closure				
Weighted AP (247 retained; 86 reviewed)	0.9101	0.9864	0.9265	0.9866
Weighted AP (186 retained; 71 reviewed)	0.9469	0.9958	0.9636	0.9958

Note: AP, average precision; nDCG, normalized discounted cumulative gain. AIS metrics use the same 361 candidates. Closure precision and nDCG average uniform random orderings within tied scores. LLM priority denotes retrospective Top-K ranking, distinct from the original 244-query trigger. Human-review P@K is the design-weighted proportion judged worth verification among K selected targets. Human-review metrics use inverse sampling-fraction weights and exclude uncertain labels. Each comparison uses the same target set.

4. Discussion

4.1. Evidence Interpretation of Satellite–AIS Mismatch

Satellite–AIS mismatch indicates that the initial association did not establish sufficient correspondence; it does not by itself establish non-cooperative activity. In dense estuarine waters, route intersections, vessel maneuvering, irregular reporting, and short spatial separations can produce several plausible target–trajectory relations. Reconstructing vessel positions at the acquisition time tests these relations by considering spatial agreement together with temporal support, distinguishing well-supported AIS correspondence from weaker local proximity. This interpretation is consistent with reports of incomplete coastal AIS coverage [23] and differences between vessel distributions derived from satellite imagery and tracking data [24].

Verification priority and AIS explainability describe different dimensions of the evidence. Priority orders unresolved cases according to their strength and urgency, whereas image-time AIS reconstruction evaluates whether cooperative vessel movement supplies sufficient spatiotemporal support. A high reviewed priority can therefore coexist with an explainable closure once additional trajectory evidence becomes available. Historical navigation corridors and nearby ports, anchorages, and shorelines help interpret the local setting, but they do not substitute for image-time correspondence. Cases with low-confidence support, ambiguous matching, or no conclusive recheck remain verification tasks with their reviewed priorities. This separation preserves uncertainty while allowing the evidence state to change when additional support is obtained.

4.2. Implications of Evidence-Driven Verification Prioritization

Rule screening, LLM evidence review, and image-time AIS reassessment serve distinct roles in verification prioritization. Rule screening reproducibly narrows and orders the initial unresolved pool. LLM review examines consistency across the evidence and identifies cases requiring additional trajectory assessment, while the AIS recheck provides the spatiotemporal support used for closure. Supplementary Table S5 summarizes this methodological distinction from conventional association-based workflows rather than

a controlled performance comparison. The reported 70.06% reduction measures the decrease in candidate verification tasks relative to the initial AIS-unexplained pool; it does not quantify a measured reduction in false detections or an improvement in verification success.

The priority gradient in the stratified review suggests that the queue concentrates stronger unresolved evidence in higher-priority tasks while retaining lower-priority cases for later assessment. Because equal numbers were sampled from each final-state group, the review supports comparisons within those groups rather than population-wide estimates of task validity. The spatial task layer carries this ordering into geographically targeted verification by preserving target locations, reviewed priorities, AIS evidence states, and maritime context. It therefore provides a basis for deciding where additional observations are most useful and for tracing subsequent evidence back to the original satellite target.

The agent architecture links multisource evidence interpretation to tool use and evidence-state updates. LLM review refines verification priorities and identifies candidates for additional AIS trajectory assessment. Its contribution consequently depends on both evidence-acquisition efficiency and verification-ranking performance. The budget curves show improved expected closure yield for retrospective LLM-priority selection at lower query budgets, while continuous rule scoring performs better at the original 244-query budget. Human-review results similarly distinguish higher early-selection point estimates from stronger overall ranking performance. Selective querying reduced AIS calls by 32.41%, and exhaustive querying resolved 61 additional cases. The shared target-level evidence record supports comparison and refinement of these policies while preserving the basis for each verification task.

4.3. Limitations and Future Directions

The analysis is limited to the Pearl River Estuary and vessel-like targets observable in the analyzed GF-6 imagery. Small, weakly visible, partially occluded, or low-contrast vessels may be missed during the detection stage, and these omissions propagate to subsequent AIS association and evidence reasoning. Precise oriented box localization can also be difficult for small, elongated, and closely spaced vessels in complex coastal imagery. The resulting candidate layer therefore represents the subset of vessel observations available to the evidence workflow within the analyzed coverage. Target contrast, sea-surface conditions, and coastal background complexity remain important sources of variation in optical vessel detection [8].

AIS evidence introduces an additional source of uncertainty through missing transmissions, delayed reports, identity inconsistencies, reception gaps, and temporal differences between AIS messages and satellite acquisition. Trajectory reconstruction at the acquisition time reduces part of this uncertainty, while cases with weak temporal or spatial support remain unresolved in the verification queue. Future work should evaluate the evidence workflow using additional acquisition dates, sensors, and maritime regions, together with expert interpretation and targeted field verification. The target-level evidence schema, tool interfaces, and evidence-state logic form the reusable workflow structure. Regional adaptation involves reconstructing navigation corridors from local AIS records, updating geographic layers, and selecting density quantiles, screening rules, and priority thresholds using local traffic patterns, port and anchorage configurations, AIS coverage, and reviewed cases.

The 70.06% reduction measures the change in verification-task count, while the baselines evaluate AIS closure yield and sampled ranking performance. AIS closures and manual labels describe evidence support and verification worthiness; confirmed vessel identities, field-verification reliability, and operator time require further evaluation. The design-weighted ranking metrics account for unequal sampling fractions across the final-

state review groups. Because the 464 rule-excluded candidates were outside the manual sample, the 70.06% task reduction does not establish the retention rate of verification-worthy targets. Future evaluations should sample both retained and excluded targets, confirm vessel identities through independent evidence, and record verification success and operator effort.

5. Conclusions

This study developed an evidence-driven Dark Vessel Detection Agent to interpret satellite–AIS mismatch through the joint use of optical remote sensing, cooperative vessel trajectories, and maritime geographic context. In the Pearl River Estuary, the workflow reduced 825 AIS-unexplained targets to 361 reviewed candidates. AIS trajectory reconstruction at the satellite acquisition time explained 114 candidates, while 247 remained as prioritized verification tasks, reducing the initial candidate pool by 70.06%. The blind plausibility review further supported the resulting evidence structure, showing complete consistency for the sampled explainable cases and a clear priority gradient among the retained tasks. These results show that the evidential meaning of satellite–AIS mismatch is shaped by the joint relationship among satellite observations, AIS support at the acquisition time, and maritime geographic context.

The main methodological contribution is an agent architecture that organizes heterogeneous observations into traceable evidence states at the target level. The Rule Agent provides reproducible geographic screening, DeepSeek-V4 reviews consistency across satellite, AIS, and contextual evidence, and selective AIS trajectory reconstruction introduces additional spatiotemporal evidence when required. Candidate priorities and evidence states are updated through this sequence while target identities and evidence attributes remain linked throughout the workflow. This structure connects evidence reasoning with reconstructed vessel movement and provides a transparent path from initial satellite observation to final evidence closure.

The baseline experiments quantify how query selection affects closure coverage and verification-task volume, providing a basis for refining prioritization within the agent framework.

The resulting output is a spatially indexed verification task layer that preserves target location, reviewed priority, AIS evidence state, and maritime context. This representation converts unresolved satellite observations into structured information that can support geographically targeted verification and subsequent evidence collection. More broadly, the framework provides an approach for integrating Earth observation, cooperative tracking information, and geospatial reasoning when the meaning of a satellite–AIS mismatch depends on its surrounding evidence relations. Future work should evaluate the framework using additional acquisition dates, sensors, and maritime regions, and incorporate explicit uncertainty propagation and additional observation sources to assess its robustness and transferability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18223942/s1, Figure S1: Second-review decisions and evidence-consistency diagnostics; Figure S2: Agent-state transition from Rule Agent priority to LLM-reviewed priority and AIS-evidence state; Figure S3: Image-time AIS rechecking mechanism and evidence-closure logic; Figure S4: Sensitivity of AIS-explainable closures to radius scale and interpolation gap; Table S1: Compact Rule Agent scoring rules for suspected dark-vessel candidate screening; Table S2: Compact DeepSeek-V4 evidence-review protocol; Table S3: Compact label definitions for the manual GF-6 local-chip review; Table S4: Key parameters of detection, AIS association, and evidence reasoning; Table S5: Methodological positioning.

Author Contributions: Y.W. and F.Y.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, and Writing—original draft. F.S. and F.Y.: Supervision, Project administration, and Funding acquisition. F.Y., F.S., Y.C. and N.X.: Writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Major Bidding Project of National Social Science Fund of China (Grant Number: 24&ZD172), the National Natural Science Foundation of China (No. 42476195) and the Youth Innovation Promotion Association of Chinese Academy of Sciences (2023060). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement: The GF-6 PMS image was obtained from the Aerospace Information Research Institute, Chinese Academy of Sciences, and is subject to data-use restrictions. AIS data are not publicly available because they contain sensitive vessel-position information and third-party usage restrictions. Derived candidate-level statistics, parameter settings, and processing scripts supporting the main findings are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kroodsma, D.A.; Mayorga, J.; Hochberg, T.; Miller, N.A.; Boerder, K.; Ferretti, F.; Wilson, A.; Bergman, B.; White, T.D.; Block, B.A.; et al. Tracking the global footprint of fisheries. *Science* **2018**, *359*, 904–908. [CrossRef] [PubMed]
2. Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E.; Huang, G.-B. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1559–1582. [CrossRef]
3. Paolo, F.S.; Kroodsma, D.; Raynor, J.; Hochberg, T.; Davis, P.; Cleary, J.; Marsaglia, L.; Orofino, S.; Thomas, C.; Halpin, P. Satellite mapping reveals extensive industrial activity at sea. *Nature* **2024**, *625*, 85–91. [CrossRef] [PubMed]
4. Park, J.; Lee, J.; Seto, K.; Hochberg, T.; Wong, B.A.; Miller, N.A.; Takasaki, K.; Kubota, H.; Oozeki, Y.; Doshi, S.; et al. Illuminating dark fishing fleets in North Korea. *Sci. Adv.* **2020**, *6*, eabb1197. [CrossRef] [PubMed]
5. Park, J.; Van Osdel, J.; Turner, J.; Farthing, C.M.; Miller, N.A.; Linder, H.L.; Ortuño Crespo, G.; Carmine, G.; Kroodsma, D.A. Tracking elusive and shifting identities of the global fishing fleet. *Sci. Adv.* **2023**, *9*, eabp8200. [CrossRef] [PubMed]
6. Welch, H.; Clavelle, T.; White, T.D.; Cimino, M.A.; Van Osdel, J.; Hochberg, T.; Kroodsma, D.; Hazen, E.L. Hot spots of unseen fishing vessels. *Sci. Adv.* **2022**, *8*, eabq2109. [CrossRef] [PubMed]
7. Liang, M.; Su, J.; Liu, R.W.; Lam, J.S.L. AISClean: AIS data-driven vessel trajectory reconstruction under uncertain conditions. *Ocean Eng.* **2024**, *306*, 117987. [CrossRef]
8. Kanjir, U.; Greidanus, H.; Oštir, K. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sens. Environ.* **2018**, *207*, 1–26. [CrossRef] [PubMed]
9. Zhang, Z.; Zhang, L.; Wu, J.; Guo, W. Optical and Synthetic Aperture Radar Image Fusion for Ship Detection and Recognition: Current state, challenges, and future prospects. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 132–168. [CrossRef]
10. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
11. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
12. Dong, J.; Feng, J.; Tang, X. OptiSAR-Net: A Cross-Domain Ship Detection Method for Multisource Remote Sensing Data. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4709311. [CrossRef]
13. Zhou, G.; Wang, X.; Liu, S.; Wang, Y.; Gao, E.; Wu, J.; Lu, Y.; Yu, L.; Wang, W.; Li, K. MSS-Net: A lightweight network incorporating shifted large kernel and multi-path attention for ship detection in remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *143*, 104805. [CrossRef]
14. Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*; IEEE: New York, NY, USA, 2018; pp. 3974–3983. [CrossRef]
15. Wen, L.; Cheng, Y.; Fang, Y.; Li, X. A comprehensive survey of oriented object detection in remote sensing images. *Expert Syst. Appl.* **2023**, *224*, 119960. [CrossRef]
16. Ding, J.; Xue, N.; Long, Y.; Xia, G.-S.; Lu, Q. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019*; IEEE: New York, NY, USA, 2019; pp. 2844–2853. [CrossRef]

17. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for Object Detection. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021*; IEEE: New York, NY, USA, 2021; pp. 3500–3509. [CrossRef]
18. Yang, X.; Yan, J.; Feng, Z.; He, T. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 3163–3171. [CrossRef]
19. Qu, J.; Liu, R.W.; Guo, Y.; Lu, Y.; Su, J.; Li, P. Improving maritime traffic surveillance in inland waterways using the robust fusion of AIS and visual data. *Ocean Eng.* **2023**, *275*, 114198. [CrossRef]
20. Heiselberg, P.; Pedersen, H.B.; Sørensen, K.A.; Heiselberg, H. Identification of Ships in Satellite Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6045–6054. [CrossRef]
21. Iphar, C.; Ray, C.; Napoli, A. Data integrity assessment for maritime anomaly detection. *Expert Syst. Appl.* **2020**, *147*, 113219. [CrossRef]
22. Ribeiro, C.V.; Paes, A.; Oliveira, D.D. AIS-based maritime anomaly traffic detection: A review. *Expert Syst. Appl.* **2023**, *231*, 120561. [CrossRef]
23. Hague, E.; Walters, A.; Moscrop, A.; Steel, E.; Dyke, K.; Hartny-Mills, L.; Lomax, A.; Dudley, R.; Garrard, P.; Hampson, J.; et al. AIS data underrepresents vessel traffic around coastal Scotland. *Mar. Policy* **2025**, *178*, 106719. [CrossRef]
24. Nakayama, S.; Dong, W.; Correro, R.G.; Selig, E.R.; Wabnitz, C.C.; Hastie, T.J.; Leape, J.; Yeung, S.; Micheli, F. Comparing spatial patterns of marine vessels between vessel-tracking data and satellite imagery. *Front. Mar. Sci.* **2023**, *9*, 1076775. [CrossRef]
25. Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; Yang, H.; Zhang, J.; Chen, Z.; Tang, J.; Chen, X.; Lin, Y.; et al. A survey on large language model based autonomous agents. *Front. Comput. Sci.* **2024**, *18*, 186345. [CrossRef]
26. Xi, Z.; Chen, W.; Guo, X.; He, W.; Ding, Y.; Hong, B.; Zhang, M.; Wang, J.; Jin, S.; Zhou, E.; et al. The rise and potential of large language model based agents: A survey. *Sci. China Inf. Sci.* **2025**, *68*, 121101. [CrossRef]
27. Schick, T.; Dwivedi-Yu, J.; Dessì, R.; Raileanu, R.; Lomeli, M.; Zettlemoyer, L.; Cancedda, N.; Scialom, T. Toolformer: Language Models Can Teach Themselves to Use Tools. *arXiv* **2023**, arXiv:2302.04761. [CrossRef]
28. Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; Cao, Y. ReAct: Synergizing Reasoning and Acting in Language Models. *arXiv* **2023**, arXiv:2210.03629. [CrossRef]
29. Guo, M.; Wu, M.; Li, S.; Zhang, J.; Li, H.; Tao, C. MultiSight: A Vision-Language Model for Collaborative Understanding of Multiview Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**, *19*, 6133–6151. [CrossRef]
30. Liu, C.; Chen, K.; Zhang, H.; Qi, Z.; Zou, Z.; Shi, Z. Change-Agent: Toward Interactive Comprehensive Remote Sensing Change Interpretation and Analysis. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5635616. [CrossRef]
31. Xu, W.; Yu, Z.; Mu, B.; Wang, J.; Wei, Z.; Peng, M. RS-Agent: Automating Remote Sensing Tasks through Intelligent Agent. *Sci. China Inf. Sci.* **2026**, *69*, 180302. [CrossRef]
32. Akinboyewa, T.; Li, Z.; Ning, H.; Lessani, M.N. GIS Copilot: Towards an Autonomous GIS Agent for Spatial Analysis. *Int. J. Digit. Earth* **2025**, *18*, 2497489. [CrossRef]
33. Liang, C.; Cui, Y.; Shi, R.; Zha, G.; Yin, X.; Xiao, M.; Xu, D.; Duan, X.; Huang, B. GeoAgentic-RAG: A Multi-Agent Framework for Autonomous Geospatial Reasoning and Visual Insight Generation with LLM. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *147*, 105195. [CrossRef]
34. Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.