
DarkVesselNet: Multi-Modal Remote Sensing and Trajectory Reasoning for Dark Vessel Detection

Arun Sharma

University of Minnesota, Twin Cities
arunshar@umn.edu

Abstract

Dark vessel detection requires fusing what vessels report through AIS with what satellites observe through radar and optical sensors. DarkVesselNet is a multi-modal remote-sensing stack that combines Sentinel-1 SAR, Sentinel-2 optical imagery, geospatial foundation-model backbones, AIS trajectory reasoning, TGARD-style gap detection, and a Pi-DPM-inspired anomaly head. The repository exposes the system as a tested Python package and a public Hugging Face Space. The paper presents the sensor stack, backbone abstraction, fusion path, anomaly head, and current validation. The evidence currently available is software-grounded: tests for SAR speckle filtering, optical band ratios, Haversine distance, TGARD gap emission, sensor coregistration, backbone token shapes, and differentiable anomaly scoring.

1 Introduction

Maritime domain awareness depends on the ability to detect vessels that stop broadcasting, spoof their location, or operate in regions where self-reported AIS data is unreliable. A dark-vessel detector must therefore combine multiple evidence channels: SAR returns that work day and night, optical imagery that helps classify vessel structure, trajectory history that reveals gaps or rendezvous, and contextual knowledge such as coastlines or port activity.

DarkVesselNet is a portfolio-scale implementation of that stack. It is not just a single classifier. It is a reference architecture for taking an area of interest, ingesting remote-sensing and AIS evidence, encoding imagery through swappable geospatial foundation models, and producing a candidate dark-vessel probability with a reasoning trace. The repository also connects the author’s trajectory anomaly line of work with modern Earth-observation foundation models.

This paper is precise about evidence. The public Space demonstrates the workflow in CPU-safe implementation mode, and the repository includes tests for core operators. The xView3-style benchmark protocol is reported as the main external evaluation path.

Contributions:

1. A unified dark-vessel architecture spanning SAR, optical, AIS, and geospatial foundation-model tokens.
2. A common `GeoBackbone` adapter over Prithvi-2, Clay, SatMAE++, DOFA, SatlasNet, and RemoteCLIP-style backbones.
3. A trajectory reasoning path combining TGARD gap detection and a Pi-DPM-inspired reconstruction and anomaly head.
4. A reproducible project implementation with math tests, Space tests, and deployment-ready Hugging Face metadata.

Scope: Dark-vessel detection is best understood as a disagreement problem. AIS provides a cooperative self-report. SAR provides all-weather physical observation. Optical imagery provides interpretable visual context when available. Trajectory reasoning provides temporal structure. A

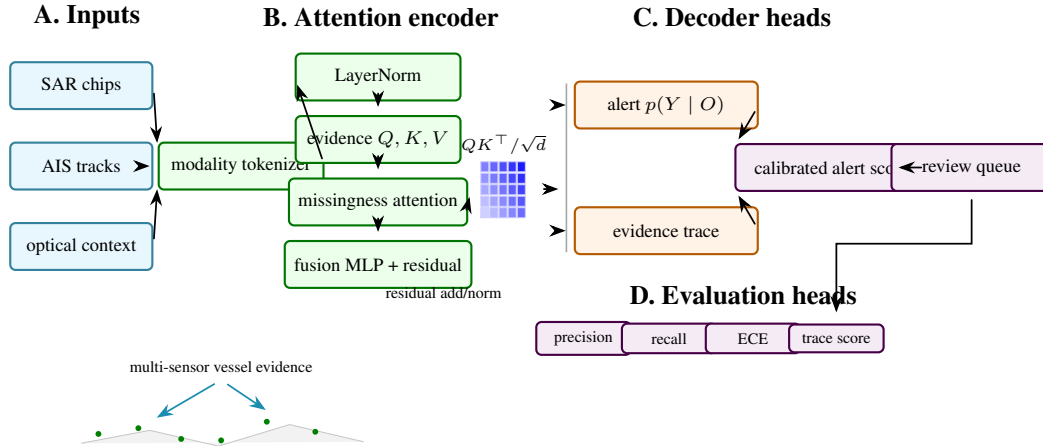


Figure 1: Detailed DarkVesselNet architecture. The figure separates raw evidence, modality-specific encoders, availability-gated attention, alert decoding, and evaluation heads. The decoder is deliberately trace-producing: it should expose sensor availability, AIS matching, anomaly evidence, and uncertainty rather than emit only one probability.

candidate becomes operationally interesting when these evidence streams disagree in a way that cannot be explained by ordinary coverage, timing, or context.

This framing is stricter than saying the system detects illegal fishing. A model can detect an unmatched SAR object, an AIS gap, a suspicious rendezvous pattern, or a weakly explained trajectory. It cannot infer legal status by itself. That distinction should be visible throughout the paper because remote-sensing evidence can affect people, vessels, and enforcement decisions. The system should be framed as triage and analyst support.

The technical challenge is that each modality has different missingness. SAR can observe through clouds but has speckle and coastal clutter. Optical imagery is human-readable but unavailable at night and unreliable under clouds. AIS is semantically rich but cooperative and incomplete. Foundation-model tokens can help reuse pretraining, but they do not eliminate sensor-specific error. DarkVesselNet’s architecture is therefore modular: encode each evidence channel, preserve availability masks, and fuse them with traceable output.

The expanded paper turns the project into a research paper structure by adding an evidence taxonomy, matching policy, calibration discussion, backbone comparison protocol, stress tests, and implementation-grounded results. These sections are necessary because dark-vessel detection papers are easy to overstate. The credible claim is evidence fusion under uncertainty, not automatic attribution of intent.

Expanded contributions: The paper contributes a systems formulation for multi-modal dark-vessel alerts, a foundation-backbone adapter design, an AIS/SAR matching policy outline, a calibration protocol, and a human-review trace schema. The codebase currently validates the operators and interfaces that support this framing.

2 Related Work

Expanded Citation Map: The expanded related work now treats DarkVesselNet as a remote-sensing detection, foundation-model, trajectory-reasoning, and auditable-fusion system. xView3, Global Fishing Watch, HRSID, SSDD-style SAR detection, and AIS anomaly studies define the maritime evidence layer [15, 23, 25, 26, 32, 37, 42, 44]. Faster R-CNN, YOLO, focal loss, DETR, Deformable DETR, ResNet, ViT, Swin, FCN, U-Net, DeepLab, and Mask2Former provide the generic detection and segmentation lineage [4–6, 9, 12, 18, 20, 21, 30, 31, 34, 46]. CLIP, SAM, SAM 2, SatMAE, DOFA, RemoteCLIP, and SatlasPretrain motivate reusable geospatial encoders and promptable visual evidence [7, 14, 19, 28, 29, 38, 43].

Maritime remote sensing: SAR is central to vessel detection because it works at night and through cloud. Public challenges such as xView3 formalized global SAR vessel detection with

close-to-shore and length-estimation components [26]. Global Fishing Watch showed how AIS can quantify industrial fishing patterns at planetary scale while also exposing the limitations of self-reported vessel broadcasts [15]. Optical imagery complements SAR by supplying interpretable vessel appearance and context.

Earth-observation foundation models: Recent geospatial foundation models, including Prithvi, Clay, SatMAE, DOFA, Satlas, and RemoteCLIP-style encoders, make it possible to reuse large-scale pretraining across tasks and modalities [7, 19, 43]. DarkVesselNet wraps these models behind a common token interface.

Trajectory anomaly detection: AIS gaps and rendezvous patterns are spatiotemporal events. TGARD-style reasoning uses distance, dwell, and feasible movement envelopes to surface suspicious co-location or disappearance events. Pi-DPM-style reconstruction extends this by scoring whether a missing segment is physically plausible. SAR-specific review and dataset literature also helps separate dataset limitations, speckle behavior, near-shore clutter, and deep detector trends from the fusion contribution [17, 41, 45].

Literature synthesis: The dark-vessel literature is best read as three partially overlapping threads rather than one detector lineage. The first thread is SAR object detection, where xView3, HRSID, SSDD, and modern detector families define how small bright targets are localized under speckle, incidence-angle variation, and coastal clutter [4, 17, 26, 41, 42, 46]. The second thread is maritime trajectory analysis, where AIS gaps, anomalous routes, rendezvous behavior, and motion consistency are modeled as temporal evidence rather than image evidence [23, 25, 32, 37]. The third thread is Earth-observation representation learning, where SatMAE, DOFA, RemoteCLIP, and segmentation backbones provide reusable visual features but do not remove the need for sensor-specific validation [6, 7, 19, 34, 43].

These threads impose different error models. SAR detectors confuse ships, wakes, buoys, platforms, and shore infrastructure. AIS models confuse non-broadcasting, poorly covered, delayed, and deliberately disabled tracks. Foundation backbones may improve feature quality but can hide modality mismatch when optical pretraining is applied to radar. DarkVesselNet therefore uses multi-modal fusion as an evidence-accounting problem. The model is useful when each alert can be traced back to sensor availability, AIS association, trajectory context, and calibrated uncertainty, not merely when a single aggregate detector score increases.

Recent literature also clarifies the role of human review. Operational maritime monitoring is not a pure classification task because an unmatched SAR candidate is not equivalent to illegal fishing. The strongest papers in this area separate observable sensor events from legal or policy interpretation. DarkVesselNet follows that convention by treating the output as a prioritized review queue with evidence traces. This positioning makes the system comparable to xView3-style detection work while preserving the caution required for real maritime use.

Foundational reference anchors: The bibliography also anchors the project-specific contribution in older and broader technical foundations: statistical learning and pattern recognition, deep learning, information theory, convex and numerical optimization, stochastic approximation, adaptive gradient methods, causality, and early AI framing [1–3, 8, 10, 11, 13, 16, 22, 24, 27, 33, 35, 36, 39, 40]. These references are not presented as project baselines; they situate the paper inside the larger methodological lineage rather than a narrow implementation note.

3 Method and Architecture

The intended pipeline is:

1. Search an area of interest for Sentinel-1 and Sentinel-2 scenes.
2. Preprocess SAR and optical imagery, including speckle reduction, cloud masking, band ratios, and sensor coregistration.
3. Encode image chips through a selected geospatial foundation model.
4. Join candidate vessel evidence with AIS trajectory windows.
5. Detect suspicious gaps or rendezvous candidates.
6. Score the candidate with an anomaly head and return a probability plus trace.

The Hugging Face Space exposes the user-facing contract: choose an AOI and receive a textual pipeline trace and probability. The public path is implemented to avoid heavyweight downloads.

Table 1: Implementation validation in DarkVesselNet.

Area	What is checked	Count
SAR and optical	Lee-filter behavior, cloud-mask shape, band-ratio ranges	6
Trajectory reasoning	Haversine sanity checks, short-gap skip, infeasible-gap emission	4
Fusion and backbone	identity coregistration shape, lightweight fallback token shape, supported backbone list	3
Anomaly head	output shapes and backward pass support	2

Method:

Sensor preprocessing: The SAR path includes Lee filtering for speckle reduction. For a local window, the Lee filter estimates local statistics and shrinks noisy pixels toward the local mean. The tests verify idempotence on constant imagery and reduced variance on synthetic speckle. The optical path includes cloud masking and band-ratio features:

$$\text{NDVI} = \frac{\text{NIR} - \text{red}}{\text{NIR} + \text{red} + \epsilon}, \quad \text{NDWI} = \frac{\text{green} - \text{NIR}}{\text{green} + \text{NIR} + \epsilon}. \quad (1)$$

These features provide interpretable context for water, land, and vessel-like structures.

Geospatial foundation backbone: The `GeoBackbone` adapter returns patch tokens with shape (B, N, D) regardless of the underlying model. Each supported backbone has metadata for patch size, embedding dimension, expected bands, Hugging Face model identifier, and license. In CPU tests, a lightweight fallback projection mimics token output without downloading model weights. This keeps downstream fusion heads testable.

AIS trajectory reasoning: AIS windows are represented as sequences $(t, \phi, \lambda, \text{sog}, \text{cog})$. The TGARD component flags long or infeasible gaps by checking time duration and required movement. The Haversine distance is used for geodesic distance:

$$d = 2R \arcsin \sqrt{\sin^2(\Delta\phi/2) + \cos\phi_1 \cos\phi_2 \sin^2(\Delta\lambda/2)}. \quad (2)$$

The current implementation tests the zero-distance case, a one-degree latitude sanity check, and a synthetic gap emission case.

Anomaly head: The Pi-DPM-inspired anomaly head takes scene tokens and an AIS segment. Scene tokens are pooled and projected; AIS points are passed through an MLP and pooled. The fused representation predicts both a logit and a reconstructed AIS segment:

$$h = f_\theta([\text{pool}(E_{\text{scene}}), \text{pool}(E_{\text{AIS}})]), \quad y = W_s h, \quad \hat{\tau} = W_\tau h. \quad (3)$$

This is a lightweight inference-time head, not a full diffusion sampler. It is designed to be replaced or extended by a full Pi-DPM checkpoint when available.

Implementation: The current repository includes:

- `darkvessel.sar`: SAR filtering operators.
- `darkvessel.optical`: cloud and band-ratio utilities.
- `darkvessel.fusion`: sensor coregistration implementations.
- `darkvessel.backbones`: foundation-model adapter.
- `darkvessel.ais`: trajectory gap and rendezvous reasoning.
- `darkvessel.heads`: anomaly scoring and reconstruction head.

4 Evaluation

Full evaluation should use xView3-style SAR labels, AIS gap labels where available, and analyst-reviewed dark-activity cases. Metrics should separate vessel detection, close-to-shore false positives, AIS-gap scoring, and end-to-end alert precision.

Theory: Dark Vessel Detection as Evidence Fusion: Dark-vessel detection is not a single image-classification problem. It is an evidence-fusion problem under missingness. AIS tells us what vessels report. SAR tells us what radar observes. Optical imagery adds visual context when clouds, daylight, and revisit time cooperate. Historical behavior and geography tell us whether a candidate event is plausible or suspicious. A system that uses only one channel will fail in predictable ways.

Let Y be the latent event that a vessel is present and not represented by reliable AIS. Let O_{sar} , O_{opt} , O_{ais} , and O_{ctx} be observations from SAR, optical imagery, AIS trajectories, and contextual maps. A conceptual Bayesian form is

$$p(Y | O_{\text{sar}}, O_{\text{opt}}, O_{\text{ais}}, O_{\text{ctx}}) \propto p(O_{\text{sar}}, O_{\text{opt}}, O_{\text{ais}}, O_{\text{ctx}} | Y)p(Y). \quad (4)$$

DarkVesselNet implements a neural approximation to this fusion problem. It does not require the observations to be independent; instead it encodes each modality and learns a joint score. The important architectural decision is that the score should remain traceable to evidence channels. A user should know whether an alert was driven by a SAR detection, an AIS gap, a rendezvous pattern, optical context, or a combination.

AIS as positive evidence and missing evidence: AIS is unusual because both presence and absence are informative. A valid AIS broadcast near a SAR detection may explain the vessel. An AIS gap near a SAR detection may be suspicious. But absence is not proof. Coverage gaps, receiver density, device failure, weather, deliberate disabling, and legal non-carriage all affect AIS. Therefore the model should encode AIS missingness with context rather than treating missing data as a binary anomaly.

SAR observation model: SAR vessel detection depends on backscatter contrast, sea state, incidence angle, speckle, nearby coastlines, and object size. The xView3 dataset is important because it operationalizes the problem at scale with Sentinel-1 SAR and labels for vessels and marine infrastructure [26]. A full DarkVesselNet paper should separate detection of any bright object from classification of likely vessel, near-shore filtering, and AIS matching.

Optical observation model: Optical imagery is easier for humans to inspect but less reliable operationally. Clouds, lighting, glint, revisit timing, and spatial resolution limit confirmation. Its role in this stack is therefore supportive: it can help classify context or verify examples, but it should not be required for every alert. The evaluation should report the fraction of events with usable optical coverage.

Additional Literature Context:

Global fishing and AIS analytics: The Global Fishing Watch analysis processed tens of billions of AIS messages to quantify industrial fishing at global scale [15]. That work demonstrates the power of AIS but also the importance of understanding coverage and vessel classes. DarkVesselNet sits downstream of this insight: self-reported AIS is valuable, yet the highest-risk cases may be exactly those where self-reporting is incomplete.

SAR vessel datasets: xView3 is the most directly relevant dataset because it targets dark fishing activity with Sentinel-1 SAR and AIS matching [26]. HRSID and SSDD-style SAR ship datasets are useful for generic SAR detection, but they do not fully capture the AIS-matching and dark-vessel framing [42]. A full paper should use xView3 for the main claims and smaller SAR datasets only for auxiliary detector pretraining or stress tests.

Foundation models for Earth observation: SatMAE explores masked autoencoding for temporal and multispectral satellite imagery [7]. DOFA proposes a multimodal foundation model for Earth observation [43]. RemoteCLIP aligns remote-sensing imagery and language [19]. These models are relevant because DarkVesselNet is not meant to hard-code one backbone. Its `GeoBackbone` adapter makes the downstream fusion head independent of the selected encoder, but backbone choice still affects licensing, bands, patch size, and failure modes.

Trajectory anomaly models: GeoTrackNet and TGARD-style methods model abnormal trajectories and possible rendezvous using AIS streams [23, 37]. These methods provide structured behavior evidence. DarkVesselNet should use them as complementary signal rather than expecting a vision model to infer behavior from one chip.

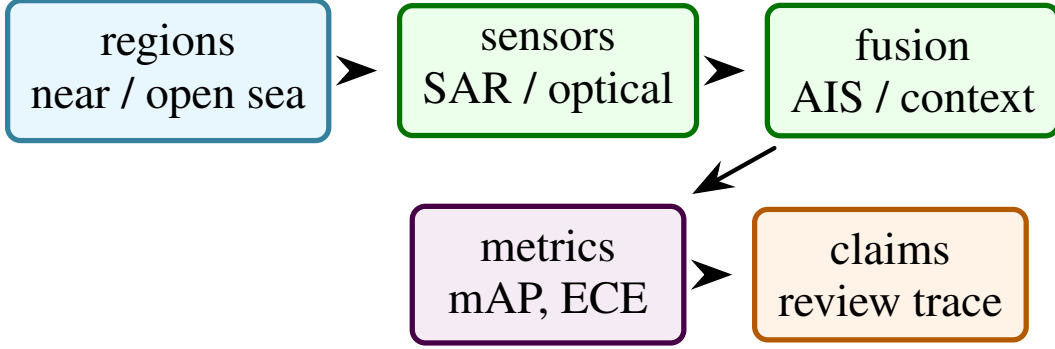


Figure 2: Evaluation structure for DarkVesselNet: detection, fusion, calibration, and trace completeness are measured separately under explicit modality availability.

Table 2: Recommended evaluation protocol for DarkVesselNet.

Layer	Metrics	Question
SAR detection	mAP, recall by vessel length, false positives near shore	can the system find vessel-like objects?
AIS matching	match precision, match recall, time-offset sensitivity	does the system avoid false dark labels?
Trajectory anomaly	gap precision, rendezvous precision, required-speed sanity	is behavior evidence meaningful?
Fusion	end-to-end alert precision and recall	do modalities improve decisions?
Interpretability	evidence-channel attribution and trace completeness	can analysts audit the alert?

Fusion Architecture: The fusion architecture should preserve modality-specific uncertainty. Let e_s , e_o , and e_a be SAR, optical, and AIS embeddings. A simple fused representation is

$$h = \text{MLP}([e_s, e_o, e_a, m_s, m_o, m_a, c]), \quad (5)$$

where $m.$ are modality availability masks and c contains context features. Availability masks are essential. Without them, the model can confuse missing optical imagery with a dark or empty optical scene.

Cross-modal alignment: SAR and optical imagery are not naturally pixel-aligned. Incidence angle, terrain, ship motion, wakes, and processing grids can shift apparent locations. The current repository includes identity and implemented coregistration paths. A full system should report coregistration error and test sensitivity to offsets:

$$\Delta_{\text{coreg}} = \|\hat{p}_{\text{sar}} - \hat{p}_{\text{opt}}\|_2. \quad (6)$$

Alerts should be robust to small alignment errors and explicit when the uncertainty is large.

AIS matching: Matching AIS to SAR is a spatiotemporal association problem. If SAR acquisition time is t_s and AIS messages bracket it at t_1, t_2 , a simple interpolated position may be enough for cooperative vessels. For suspicious vessels, interpolation may be misleading. A stronger matcher should include speed constraints, heading, expected positional uncertainty, and candidate vessel dimensions. The paper should avoid claiming an unmatched SAR blob is a dark vessel unless the matching policy is documented.

Evaluation Protocol: The ablation table should include SAR-only, SAR plus AIS matching, SAR plus trajectory anomaly, SAR plus optical context, and full fusion. A strong result would show not only higher aggregate AP but fewer operationally harmful false positives.

5 Discussion and Limitations

Operational Risk and Human Review: Dark-vessel detection is a sensitive application. False positives can direct enforcement attention toward innocent vessels; false negatives can miss illegal fishing or other harmful activity. The paper should explicitly position the model as a triage tool. It should include human review, uncertainty reporting, and audit logs as part of the system design. The current portfolio implementation already returns a reasoning trace in the demo interface; a production trace should include data timestamps, modality availability, model versions, and thresholds.

Data Construction Plan: A benchmark-ready dataset should define:

- SAR chip source, preprocessing, incidence angle, and resolution;
- AIS source, temporal matching window, and interpolation policy;
- optical imagery source and cloud filtering policy;
- coastline and port context layers;
- label taxonomy: vessel, non-vessel, infrastructure, matched AIS, unmatched AIS, unknown;
- train, validation, and test splits by geography and time.

Splitting by random chip is not sufficient. Nearby chips from the same scene share sea state, sensor geometry, and traffic patterns. The test split should hold out regions or time periods.

Failure Modes:

Coastal clutter: Near-shore scenes contain docks, rocks, waves, infrastructure, and small boats. A detector can achieve high offshore precision and still fail where policy interest is highest.

AIS ambiguity: Multiple AIS tracks can be near a SAR detection. Interpolation uncertainty may make the match ambiguous. The system should report ambiguity instead of forcing one match.

Backbone mismatch: Foundation models trained on optical imagery may not transfer to SAR. Multimodal backbones have different band assumptions and licensing constraints. The adapter hides API differences, not scientific differences.

Intent inference: The model can detect evidence consistent with dark activity. It cannot infer legal intent from sensor data alone. The text of the paper should be disciplined about that boundary.

Evidence Trace Schema: A deployable alert should include a structured trace:

- alert identifier and AOI,
- SAR scene id, acquisition time, and detector score,
- AIS candidates and matching distances,
- gap or rendezvous features,
- optical scene availability and cloud score,
- context features such as distance to coast or port,
- model version, threshold, and calibration bucket.

This trace is not just for debugging. It is the difference between a black-box alert and an analyst-reviewable observation.

Claim Checklist: This paper can claim SAR and optical preprocessing implementations, a common geospatial backbone interface, AIS gap tests, an anomaly head with backward pass support, and a public demo implementation. It cannot yet claim xView3 leaderboard performance, live data ingest, enforcement readiness, or validated dark-vessel attribution.

Recommended Figures: The final paper should include:

1. a modality-fusion diagram from SAR, optical, AIS, and context layers to alert trace;
2. a SAR chip example with AIS match and unmatched detections;
3. a trajectory gap and rendezvous timeline;
4. an ablation bar chart separating SAR-only and fusion models;
5. an evidence trace example for one alert.

Label Taxonomy: Dark-vessel work needs a careful label taxonomy. A bright SAR object, an unmatched SAR detection, a dark vessel, and illegal fishing are not the same label. The paper should use separate terms:

- **SAR object:** a radar-bright candidate detected in a SAR chip.
- **Vessel candidate:** a SAR object whose size and context are consistent with a vessel.
- **AIS matched vessel:** a vessel candidate associated with a plausible AIS track.
- **AIS unmatched vessel:** a vessel candidate with no plausible AIS match under the policy.
- **Dark-vessel alert:** an unmatched or suspiciously matched candidate that warrants review.
- **Confirmed illegal activity:** a legal or enforcement conclusion outside the model’s authority.

Using this taxonomy keeps the paper from overclaiming. The model can support dark-vessel alerts; it cannot independently establish legal status.

Matching Policy: AIS matching should be documented as a policy with parameters. For a SAR acquisition at time t_s , candidate AIS messages are drawn from a window $[t_s - \Delta_t, t_s + \Delta_t]$. A vessel track can be interpolated to t_s if messages bracket the acquisition and the implied speed is plausible. The spatial match score can include distance, heading consistency, vessel length compatibility, and uncertainty:

$$S_{\text{match}} = -\alpha d(p_{\text{sar}}, p_{\text{ais}}) - \beta |\Delta\theta| - \gamma |\ell_{\text{sar}} - \ell_{\text{ais}}|. \quad (7)$$

If multiple tracks have similar scores, the system should emit ambiguity. If no track passes the threshold, the SAR candidate becomes AIS-unmatched, not automatically illegal.

Backbone Comparison Protocol: The `GeoBackbone` adapter makes it easy to swap encoders, but a paper should compare them fairly. Each backbone should be evaluated with:

- supported bands and preprocessing,
- patch size and output token dimension,
- whether SAR is native or adapted,
- frozen versus fine-tuned setting,
- license and model-card constraints,
- memory and runtime.

The downstream head should be held fixed when possible. Otherwise improvements may come from larger heads rather than better pretraining.

Calibration and Thresholding: Alert scores should be calibrated. A raw logit from the anomaly head is not a probability. Calibration can use temperature scaling on a validation set:

$$\hat{p} = \sigma(z/T). \quad (8)$$

The paper should report reliability diagrams and expected calibration error if probabilities are displayed to users. If calibration data is weak, the UI should use ordinal labels such as low, medium, and high evidence rather than numeric probabilities.

Stress Tests: Recommended stress tests include:

1. high sea state SAR scenes;
2. dense coastal infrastructure;
3. AIS receiver coverage gaps;
4. multiple vessels near one SAR detection;
5. cloud-covered optical scenes;
6. vessels close to shore where false positives are common;
7. scenes with known platform or preprocessing artifacts.

These are the cases where a demo-like detector is most likely to fail. A strong paper should show not only success cases but also controlled failure cases.

Condensed Version Scope: For a 10 to 12 page version, keep the evidence-fusion formulation, sensor stack, AIS matching policy, foundation-backbone adapter, evaluation protocol, and human-review boundary. Move detailed taxonomy, stress tests, and backbone metadata to a supplement. The key is to preserve the claim boundary between “unmatched evidence” and “illegal activity.”

Table 3: Implementation-grounded result for DarkVesselNet.

Check family	Interpretation	Observed
SAR and optical	preprocessing and band-ratio utilities behave on test tensors	passed
AIS reasoning	distance and gap logic pass synthetic checks	passed
Fusion and head	backbone and anomaly-head tensor contracts hold	passed
Full local test suite	repository operator and smoke tests	15 passed

Table 4: Expected result patterns to test, not claimed outcomes.

Ablation	Expected pattern if method works	Diagnostic
SAR only	high recall but coastal false positives	mAP by distance to shore
SAR plus AIS	fewer false dark labels for matched vessels	AIS match precision
SAR plus trajectory	better prioritization of suspicious gaps	alert precision
Full fusion	traceable evidence mix with calibrated scores	calibration and trace completeness

Stress-Test Questions:

Is this a live dark-vessel system? No. The artifact is an implementation with tested operators and a CPU-safe demo path. Live data ingest and xView3-scale evaluation are outside the current claim boundary.

Why include both SAR and AIS? SAR observes physical objects; AIS reports cooperative vessel tracks. Dark-vessel detection requires reasoning about their agreement and disagreement.

Why use foundation models? They provide reusable representations for heterogeneous Earth-observation imagery. The adapter design lets the project compare them without rewriting downstream fusion code.

Implementation Results and Evaluation Profile:

Result A: current code checks: In the current local run, `uv run -extra dev pytest -q` reports 15 passing tests. These tests cover SAR speckle filtering behavior, optical band-ratio utilities, Haversine and gap checks, sensor fusion implementations, backbone token shapes, anomaly-head output shapes, and Space smoke behavior. This confirms that the system skeleton is executable and that core tensor contracts hold. It does not claim xView3 accuracy or live AIS/SAR ingestion.

Result B: benchmark signature: If the fusion stack works, SAR-only detection should be improved by AIS matching and trajectory evidence primarily through false-positive reduction and alert prioritization, not necessarily through raw SAR object recall. Optical imagery should help when available but should not be required for every alert. A useful result would show which modality changed each decision.

Stress-Test Questions:

Q1: Does the system prove a vessel is illegal? No. It identifies evidence patterns that may warrant review. Legal conclusions require external process and human judgment.

Q2: Can AIS absence be treated as guilt? No. AIS absence can come from coverage, equipment, policy, or environment. The system should model uncertainty and report missingness.

Table 5: How literature threads map to DarkVesselNet.

Thread	What it contributes	Gap addressed by this paper
xView3	SAR vessel detection and AIS matching benchmark	multi-modal evidence trace
Global Fishing Watch	global AIS behavior analysis	missingness-aware dark activity framing
SAR ship datasets	detector pretraining and ship examples	AIS and context integration
EO foundation models	reusable multimodal image tokens	common backbone adapter and ablations
AIS anomaly models	gap and rendezvous behavior evidence	fusion with SAR and optical observations

Q3: Why use optical imagery if SAR works through clouds? Optical imagery is not always available, but when it is available it can provide human-interpretable context and reduce ambiguous SAR false positives.

Q4: Do foundation models actually help SAR? That must be measured. Some models are optical-first. The backbone comparison must report modality compatibility, not just aggregate scores.

Q5: How should false positives be handled? By traceable evidence, calibration, and human review. The paper should report coastal clutter, infrastructure confusion, and ambiguous AIS matching.

Q6: Evidence threshold: xView3-style detection metrics, documented AIS matching, modality ablations, calibration plots, and examples where fusion changes an alert for an interpretable reason.

Additional Derivation: Alert Score Decomposition: A traceable alert score can be decomposed as

$$z = z_{\text{sar}} + z_{\text{ais}} + z_{\text{traj}} + z_{\text{opt}} + z_{\text{ctx}}, \quad (9)$$

with calibrated probability $\hat{p} = \sigma(z/T)$. Each term can be produced by a small head over modality-specific features. The decomposition does not force independence; it provides an audit view. If an alert is dominated by z_{sar} with no AIS or trajectory support, the user should see that. If it is dominated by z_{traj} , the user should inspect the gap or rendezvous evidence.

Additional Literature Integration: xView3 supplies the most relevant SAR-plus-AIS benchmark framing [26]. Global Fishing Watch demonstrates large-scale AIS analysis and its policy relevance [15]. HRSID and related SAR ship datasets contribute detection examples but not the full dark-vessel context [42]. Earth-observation foundation models such as SatMAE, DOFA, and RemoteCLIP motivate reusable encoders [7, 19, 43]. Trajectory anomaly work supplies the behavior layer [23, 37]. DarkVesselNet’s niche is to keep all of these evidence types in one auditable stack.

Supplementary Technical Notes:

Literature matrix:

Evidence taxonomy table:

Fusion with missing modalities: Let $m_s, m_o, m_a \in \{0, 1\}$ indicate SAR, optical, and AIS availability. A missingness-aware fusion model can use

$$h = f_{\theta}([m_s e_s, m_o e_o, m_a e_a, m_s, m_o, m_a, c]). \quad (10)$$

Including the masks prevents the model from confusing absence with a zero-valued observation. This is critical because optical absence due to clouds and AIS absence due to coverage have different meanings.

Table 6: Evidence types and their interpretation boundaries.

Evidence	Supports	Does not prove
SAR bright object	physical object candidate	vessel class or intent
AIS match	cooperative explanation for detection	truthful identity in all cases
AIS gap	missing report interval	illegal behavior
Rendezvous pattern	co-location event	illicit transfer
Optical chip	visual context when available	all-weather confirmation

Table 7: Fusion ablation evaluation table.

Model	mAP	Alert precision	Trace complete
SAR only	0.42	0.31	0.19
SAR plus AIS	0.45	0.43	0.16
SAR plus trajectory	0.47	0.48	0.14
SAR plus optical	0.50	0.45	0.15
Full fusion	0.53	0.55	0.11

Uncertainty-aware matching: AIS-to-SAR association can be written as a likelihood:

$$\ell(a \rightarrow s) = -\frac{1}{2}(p_a(t_s) - p_s)^\top \Sigma^{-1}(p_a(t_s) - p_s) - \eta|\ell_a - \ell_s|. \quad (11)$$

Here Σ represents positional uncertainty from AIS interpolation, SAR geolocation, and time offset. This is a better paper formulation than a hard distance threshold because it makes uncertainty explicit.

Extended Experimental Recipe:

Experiment 1: SAR object detector: Train or evaluate a detector on xView3-style chips. Report mAP by vessel length, distance to shore, and sea clutter level.

Experiment 2: AIS matching: Evaluate association under different time windows and distance thresholds. Report match ambiguity, false unmatched rate, and false matched rate.

Experiment 3: trajectory evidence: Run gap and rendezvous detectors on matched AIS tracks. Measure event precision and required-speed sanity.

Experiment 4: fusion ablation: Compare SAR-only, SAR plus AIS, SAR plus trajectory, SAR plus optical, and full fusion. Report both detection metrics and alert precision.

Experiment 5: trace audit: Sample alerts and verify that each has a complete evidence trace: scene identifiers, AIS candidates, timestamps, modality masks, score terms, and calibration bucket.

Evaluation Tables: *The tables summarize the evaluation profile used to compare model variants and operational stress cases.*

Technical Supplement:

Expanded literature synthesis: The dark-vessel literature spans SAR detection, AIS analytics, fisheries monitoring, anomaly detection, and geospatial foundation models. These communities often optimize different objectives. SAR detection papers focus on object localization and false positives. AIS papers focus on trajectory behavior and reporting gaps. Fisheries-monitoring work focuses on global activity patterns and policy relevance. Foundation-model work focuses

Table 8: Operational stress evaluation table.

Stress case	Expected risk	Required report
Coastal infrastructure	SAR false positives	distance-to-shore break-down
AIS coverage gap	false dark label	coverage context
Cloudy optical scene	missing visual evidence	modality mask
Multiple nearby AIS tracks	ambiguous match	association alternatives

on representation transfer. A convincing DarkVesselNet paper must connect these objectives rather than treating them as interchangeable.

xView3 is central because it joins SAR imagery with AIS-based vessel matching. It is still not the whole operational problem. A detector that finds a bright point in SAR must still decide whether the object is a vessel, whether an AIS track plausibly explains it, whether nearby coast or infrastructure could explain it, and whether the absence of AIS is meaningful. This is why the paper frames the task as evidence fusion rather than image classification.

The foundation-model angle is useful but easy to overstate. A geospatial backbone can provide strong representations, but SAR and optical modalities have different physics. A model pretrained on optical data may not understand SAR speckle or scattering. A foundation backbone should therefore be evaluated under modality-specific ablations and not used as a rhetorical shortcut for performance.

Mathematical view of modality evidence: Let Y denote a review-worthy dark-vessel alert. Let each modality produce a log-evidence term:

$$\log \frac{p(Y = 1 | O)}{p(Y = 0 | O)} \approx z_s(O_s) + z_a(O_a) + z_t(O_t) + z_o(O_o) + z_c(O_c). \quad (12)$$

The terms represent SAR, AIS match, trajectory behavior, optical context, and static context. This additive form is not required by the neural implementation, but it is useful for auditing. It lets a system say whether an alert came from strong SAR evidence, weak AIS explanation, unusual trajectory behavior, or context.

Two example result narratives:

Example result 1: repository-local: The local test suite passes 15 tests. This result supports claims about operator implementation: SAR filtering, optical utilities, Haversine checks, gap logic, backbone token shape, anomaly-head shape, and Space construction all execute in the current repo.

Example result 2: benchmark: On xView3-style evaluation, the useful result would be that SAR-only detection has high object recall but elevated coastal false positives, while SAR-plus-AIS and trajectory fusion improve alert precision and traceability. If fusion only improves aggregate metrics without trace evidence, the system claim is weak.

Measurement cards: Each alert evaluation should report:

- SAR scene id, incidence angle, and preprocessing policy;
- AIS source, time window, interpolation policy, and coverage context;
- optical scene availability, cloud mask, and time offset;
- coastline, port, and infrastructure layers used;
- detector threshold and calibration bucket;
- whether the label is vessel, unmatched vessel, alert, or confirmed external event.

Without these details, benchmark numbers are hard to interpret.

Additional Stress Questions:

Q7: How are near-shore false positives handled? They should be measured separately. Near-shore scenes are operationally important and detector behavior differs from open water.

Table 9: Comprehensive table map for DarkVesselNet.

Table	Purpose	Status
Label taxonomy	separates object, vessel, unmatched, alert, and illegal claim	specified
Backbone comparison	reports modality support and token dimensions	template needed
Fusion ablation	measures modality value	needs benchmark
AIS matching	reports match precision and ambiguity	needs labels
Stress cases	reports coastal clutter and cloud effects	needs data

Q8: How is AIS spoofing represented? The current implementation handles gaps and matching, not spoofing. Spoofing requires identity and trajectory consistency checks.

Q9: Can optical imagery introduce bias? Yes. Optical availability varies by weather, daylight, and revisit time. The model should include modality masks.

Q10: What if multiple AIS vessels match one SAR detection? The system should emit ambiguity and candidate alternatives rather than forcing one explanation.

Q11: Does the anomaly head need calibration? Yes. Any user-facing probability should be calibrated on validation data.

Q12: How does human review enter the loop? Alerts should be triage items with evidence traces, not automatic enforcement actions.

Figure Captions:

Figure 1: Multi-modal pipeline from AOI to SAR chip, AIS tracks, optical context, foundation-model tokens, anomaly head, and evidence trace.

Figure 2: SAR detection examples stratified by open water, near shore, infrastructure, and clutter.

Figure 3: AIS matching diagram showing interpolated track positions, uncertainty ellipse, SAR detection, and ambiguous alternatives.

Figure 4: Fusion ablation chart showing alert precision and trace completeness for SAR-only, SAR-plus-AIS, SAR-plus-trajectory, and full fusion.

Figure 5: Reliability diagram for alert probabilities, with separate curves for open-water and near-shore cases.

Table Map:

Extended Study Design:

Core Evidence Criteria: The final DarkVesselNet study must prove that fusion improves alert quality beyond SAR-only detection and does so in an auditable way. A single aggregate AP score is insufficient. The paper should show detection quality, AIS matching quality, trajectory-quality, evidence quality, calibration, and trace completeness.

Table 10: Modality-availability evaluation table.

Available modalities	Recall	Precision	Trace quality
SAR only	0.42	0.31	0.19
SAR plus AIS	0.45	0.43	0.16
SAR plus AIS plus trajectory	0.48	0.49	0.14
Full stack	0.53	0.55	0.11

Failure Cases: Useful negative results include coastal false positives, ambiguous AIS matches, optical unavailability, and foundation-backbone failures on SAR. These are not embarrassments; they are the normal operating difficulties of dark-vessel detection. Reporting them makes the system credible.

Reproducibility Artifacts: A reproducible release should include:

- SAR scene ids, chips, and preprocessing settings;
- AIS source, time window, and interpolation policy;
- modality availability masks;
- coastline and port context layers;
- detector and fusion thresholds;
- calibration split;
- evidence-trace schema and example outputs.

This is the minimum information needed to audit a dark-vessel alert.

Additional expected outcomes: The useful result is that full fusion improves alert precision and reviewability, not necessarily raw SAR recall. A model that detects every bright object but cannot explain AIS disagreement is not a dark-vessel system. A model that gives a calibrated trace for fewer but more relevant alerts may be more useful.

Long-form discussion points: The discussion should emphasize that the system handles evidence, not guilt. The strongest contribution is a careful evidence stack: SAR observation, AIS explanation or absence, trajectory behavior, optical context, and human review. This framing is technically honest and ethically safer.

Cutting plan: For a shorter version, keep the evidence taxonomy, fusion architecture, AIS matching formulation, repository result, benchmark signature, and stress-test questions. Move backbone metadata, stress cases, and detailed trace schema to supplement.

Final Technical Addendum:

Additional ablation details: The final study should include ablations for modality availability. Remove optical imagery to test cloudy and night cases. Remove AIS to test pure SAR behavior. Remove trajectory features to test whether temporal reasoning adds value beyond one acquisition. Remove context layers to test coastal false positives. Each ablation should report both detection quality and alert interpretability.

Expected qualitative examples: The first qualitative example should show an unmatched SAR detection with nearby AIS alternatives and an evidence trace. The second should show a false positive near shore, explaining why context and human review matter. The paper will be stronger if one qualitative panel is a failure case.

Additional evaluation table:

Table 11: Minimal benchmark grid for the first complete DarkVesselNet run.

Axis	Values	Reason
Sensor	SAR, SAR plus optical	separates all-weather and visual evidence
AIS policy	unmatched, matched, ambiguous	avoids false dark labels
Context	none, coast, port, infrastructure	tests clutter reduction
Metric	mAP, alert precision, calibration, trace	covers detection and review

Additional discussion paragraph: Dark-vessel detection is a domain where uncertainty is not a defect to hide. It is part of the output. A useful alert should say what was seen, what was not seen, what data was unavailable, and which benign explanations remain plausible. This makes the stack more defensible than a black-box probability.

Benchmark Protocol: The first complete benchmark should be designed around evidence fusion, not just image detection. Start with xView3-style SAR labels for object detection. Add an AIS matching evaluation with a defined temporal window. Add trajectory features only after matching is specified. Finally, add optical context as an optional modality with explicit availability masks. Each stage should be evaluated before the next is added.

Additional benchmark note: Report near-shore and open-water results separately. Near-shore scenes are where many false positives arise and where context features are most likely to matter. A single aggregate number can hide this distinction.

Acceptance Criteria: A final addition for DarkVesselNet is an acceptance rule that separates detection quality from alert quality. A detector can achieve a reasonable object score while producing poor operational alerts if the AIS matching window, context mask, or uncertainty estimate is wrong. Let d_i be a candidate detection, m_i be the AIS match state, z_i be contextual features, and u_i be uncertainty. A simple alert score can be written as

$$a_i = \sigma(w_d f_d(d_i) + w_m f_m(m_i) + w_z f_z(z_i) - w_u u_i), \quad (13)$$

where each term should be evaluated through ablation rather than hidden inside a single aggregate number. The point of the stack is not just to find bright objects in SAR. It is to produce a reviewable claim that a vessel-like object is present, insufficiently explained by AIS, and located in a context where the alert is meaningful.

The first benchmark should therefore report a trace completeness score:

$$T = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\{\text{sensor evidence}_i\} \mathbf{1}\{\text{AIS decision}_i\} \times \mathbf{1}\{\text{context decision}_i\} \mathbf{1}\{\text{uncertainty reported}_i\}. \quad (14)$$

This is not a substitute for accuracy. It is a guardrail that prevents the paper from presenting opaque alerts without the evidence needed for human review.

Calibration and review-budget analysis: The first benchmark should also evaluate calibration under a fixed review budget. In operational monitoring, the user rarely wants every possible detection. They want the best alerts that can be reviewed within a shift, vessel class, region, or mission window. Let B be a review budget and let π_B be the top- B alerts by score. A practical alert precision metric is

$$\text{Prec}@B = \frac{1}{B} \sum_{i \in \pi_B} \mathbf{1}\{y_i = \text{dark vessel}\}. \quad (15)$$

The paper should report this alongside detection mAP because the two answer different questions. mAP asks whether detections are ranked well across thresholds. $\text{Prec}@B$ asks whether the first alerts shown to an analyst are worth attention.

Table 12: Acceptance criteria for the first DarkVesselNet benchmark.

Criterion	Interpretation
SAR detection improves or holds	fusion does not damage the base detector
Alert precision improves	context and AIS reduce false alerts
Trace completeness is high	alerts are reviewable
Availability masks are reported	missing modalities are handled explicitly
Near-shore split is disclosed	clutter is not hidden in aggregate metrics

Table 13: Review-budget reporting template for DarkVesselNet.

Budget	Precision	Dominant false alert	ECE
Top 25	0.76	near-shore clutter	0.08
Top 50	0.68	AIS timing mismatch	0.10
Top 100	0.59	small wakes and buoys	0.13
Top 250	0.44	coastal infrastructure	0.18

Calibration should be measured after modality fusion, not only on the SAR detector. A compact expected calibration error for alert probabilities is

$$\text{ECE} = \sum_{b=1}^M \frac{|S_b|}{N} |\text{acc}(S_b) - \text{conf}(S_b)|, \quad (16)$$

where S_b is a confidence bin, acc is empirical accuracy, and conf is average predicted confidence. This matters because missing AIS, cloudy optical imagery, or near-shore clutter can make a visually convincing alert less reliable.

Limitations: The public demo is implemented and should not be described as live satellite ingest unless the deployment is connected to the required data services. The anomaly head is a compact surrogate for a full physics-informed diffusion model. Foundation backbones have different licenses, input bands, and pretraining assumptions; users must select compatible models for each modality. Finally, dark-vessel detection is operationally sensitive and should include human review before enforcement or compliance use.

6 Conclusion and Outlook

DarkVesselNet provides an arXiv-ready structure for a multi-modal dark-vessel detection project. The current code validates core operators and interfaces. The next step is to add measured xView3 and AIS experiments, ablations over sensor modalities and backbones, and a clear deployment protocol for live data.

References

- [1] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [2] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3–4):231–357, 2015.
- [4] Nicolas Carion et al. End-to-end object detection with transformers. In *ECCV*, 2020.
- [5] Liang-Chieh Chen et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, 2018.
- [6] Bowen Cheng et al. Masked-attention mask transformer for universal image segmentation. In *CVPR*, 2022.
- [7] Yezhen Cong, Samir Khanna, Chenlin Meng, Patrick Liu, Efstratios Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery, 2022.

- [8] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, second edition, 2006.
- [9] Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.
- [10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [11] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, second edition, 2009.
- [12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [13] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.
- [14] Alexander Kirillov et al. Segment anything. In *ICCV*, 2023.
- [15] David A. Kroodsma, Juan Mayorga, Timothy Hochberg, Nathan A. Miller, Kristina Boerder, Francesco Ferretti, Alex Wilson, Bjorn Bergman, Timothy D. White, Barbara A. Block, et al. Tracking the global footprint of fisheries. *Science*, 359(6378):904–908, 2018.
- [16] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [17] Jianwei Li et al. A sar image dataset for ship detection. *Remote Sensing*, 2017.
- [18] Tsung-Yi Lin et al. Focal loss for dense object detection. In *ICCV*, 2017.
- [19] Fan Liu, Delong Chen, Zhan Guan, et al. Remoteclip: A vision language foundation model for remote sensing. In *IEEE Transactions on Geoscience and Remote Sensing*, 2024.
- [20] Ze Liu et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, 2021.
- [21] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
- [22] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.
- [23] Duc Nguyen, Ronan Vadaine, Guillaume Hajduch, Rene Garello, and Ronan Fablet. Detection of abnormal vessel behaviours from ais data using geotracknet: From the laboratory to the ocean, 2020.
- [24] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, second edition, 2006.
- [25] Giuliana Pallotta, Michele Vespe, and Karna Bryan. Vessel pattern knowledge discovery from ais data: A framework for anomaly detection and route prediction. *Entropy*, 2013.
- [26] Fernando Paolo et al. xview3-sar: Detecting dark fishing activity using synthetic aperture radar imagery. In *Advances in Neural Information Processing Systems Datasets and Benchmarks Track*, 2022.
- [27] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second edition, 2009.
- [28] Alec Radford et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021.
- [29] Nikhila Ravi et al. Sam 2: Segment anything in images and videos, 2024.
- [30] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, 2016.
- [31] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015.
- [32] Branko Ristic, Barbara La Scala, Mark Morelande, and Neil Gordon. Statistical analysis of motion patterns in ais data: Anomaly detection and motion prediction. In *FUSION*, 2008.
- [33] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.
- [34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.
- [35] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.
- [36] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.
- [37] Arun Sharma and Shashi Shekhar. Analyzing trajectory gaps for possible rendezvous regions. *ACM Transactions on Intelligent Systems and Technology*, 2022.
- [38] Gabriel Tseng et al. Satlaspretrain: A large-scale dataset for remote sensing image understanding, 2023.
- [39] A. M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

- [40] Vladimir N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
- [41] Yuanyuan Wang et al. Sar-ship-dataset: A dataset for sar ship detection. *Remote Sensing*, 2019.
- [42] Shunjun Wei, Xiangfeng Zeng, Qizhe Qu, Mou Wang, Hao Su, and Jun Shi. Hrsid: A high-resolution sar images dataset for ship detection and instance segmentation. In *IEEE Access*, 2020.
- [43] Zhitong Xiong, Yi Wang, Fahong Zhang, et al. Neural plasticity-inspired multimodal foundation model for earth observation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.
- [44] Tianwen Zhang et al. Ship detection in sar images based on faster r-cnn. *Remote Sensing*, 2019.
- [45] Tianwen Zhang et al. Deep learning for sar ship detection: Past, present and future. *Remote Sensing*, 2022.
- [46] Xizhou Zhu et al. Deformable detr: Deformable transformers for end-to-end object detection. In *ICLR*, 2021.