

Article

Multi-Resolution Ship Association in Satellite Imagery: Integrating High-Resolution Detection with Template Matching

Yangchun Zhang ¹ , Xudong Kang ^{1,2,*}  and Puhong Duan ¹ 

¹ School of Artificial Intelligence and Robotics, Hunan University, Changsha 410082, China; b220900493@hnu.edu.cn (Y.Z.); puhong_duan@hnu.edu.cn (P.D.)

² Yuelushan Center for Industrial Innovation, Changsha 410082, China

* Correspondence: xudong_kang@163.com

Abstract

Ship association in satellite imagery is important for maritime surveillance, but existing methods are generally limited to images with similar resolutions. This paper proposes a multi-resolution ship association framework that combines high-resolution object detection with low-resolution template matching. In the high-resolution stage, ship targets are detected and their positions and patches are recorded. In the low-resolution stage, the detected positions are used to define candidate search regions, within which template matching is performed to locate the corresponding ships. Experiments on three satellite scenes with different spatial resolutions show that the proposed method consistently outperforms the baseline methods. The association accuracy, defined as the proportion of correctly associated ship pairs, reaches 0.933, 0.870 and 0.862. These results demonstrate the effectiveness of the proposed framework for multi-resolution ship association in the tested GF-1 scenes and its potential for practical maritime monitoring.

Keywords: multi-resolution; remote sensing imagery; ship association; ship detection; template matching

1. Introduction

With the development of multi-satellite collaborative observation, remote sensing imagery has made maritime surveillance increasingly feasible. Ship association aims to detect and match the same vessel across different satellite images, thereby enabling reconstruction of ship trajectories and supporting applications such as navigation safety, maritime traffic monitoring, accident rescue, and law enforcement [1–3]. In practice, satellite remote sensing data may come from different imaging modalities and acquisition conditions, including optical, SAR, and infrared imagery [4–14]. Among these, optical images are widely used because of their intuitive appearance and rich visual details. However, optical satellite images often vary significantly in spatial resolution, coverage and revisit time, which makes reliable ship association particularly challenging.

Existing ship association methods are generally designed for image pairs with comparable resolutions. A common strategy is to first detect ships in each image and then associate them according to similarity between the detected targets [9]. Such methods work well when the two images have comparable spatial resolutions, because the detected targets are similar in scale and appearance [11,15,16]. When the resolution gap is large, however, the detection results may become inconsistent: ships in the high-resolution image may be clearly visible, whereas the same targets in the low-resolution image may be



Academic Editor: Atsushi Mase

Received: 26 May 2026

Revised: 22 June 2026

Accepted: 30 June 2026

Published: 5 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

blurred, partially occluded, or even missed. This imbalance can significantly reduce the accuracy of downstream association [17]. In addition, acquiring two images of the same sea area at a suitable time interval is difficult due to the fixed orbital parameters and revisit cycles of satellites [18]. As a result, practical maritime monitoring often has to deal with multi-resolution image pairs rather than ideal same-resolution pairs.

Ship detection methods can be broadly divided into traditional handcrafted feature-based approaches and deep learning-based approaches. Traditional methods rely on handcrafted descriptors such as histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT), and color histograms [19–22]. With the development of deep learning, CNN-based detectors such as Faster R-CNN, SSD, YOLO, and DETR have become the dominant choice for ship detection in remote sensing imagery [23–27]. Nevertheless, as shown in Figure 1a, detector performance is still strongly affected by image resolution. As spatial resolution decreases, missed detections and false positives increase, which in turn affect ship association and may further lead to incorrect matching or incomplete association results.

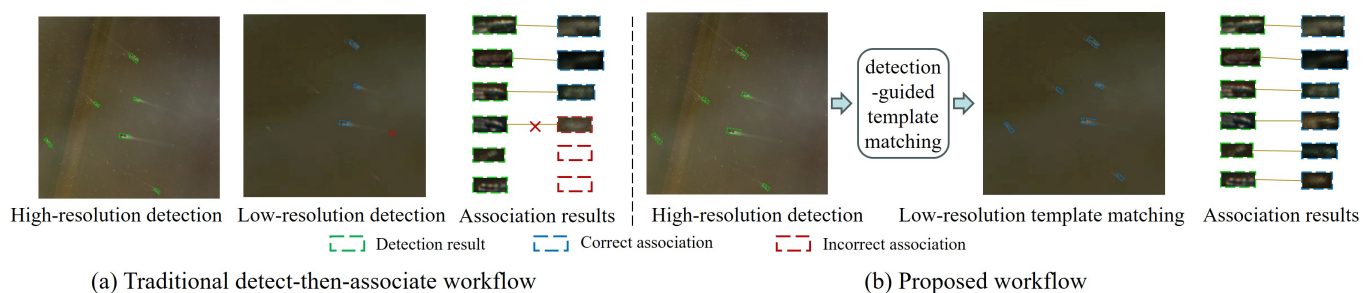


Figure 1. Comparison of traditional and proposed ship association workflows under multi-resolution scenarios.

For the ship association task, existing methods mainly rely on similarity computation between ship patches. Representative similarity measures include perceptual hash (pHash), structural similarity (SSIM), and normalized cross-correlation (NCC) [28–30]. Template matching is another common strategy, in which a target patch is compared with a candidate region to identify the best match [31]. Deep learning-based matching methods have also been explored, where learned features are used for association [32]. However, most existing association methods assume that the targets in the two images have similar scale and appearance. They are therefore not well suited for multi-resolution satellite imagery, especially when the image pair contains a large resolution difference.

As illustrated in Figure 1b, the proposed workflow addresses this limitation by using high-resolution detection results as auxiliary information to guide template matching in low-resolution images. Specifically, ship targets are first detected in the high-resolution image, and their positions, azimuths, and patches are recorded. These detections are then used to define candidate regions in the low-resolution image, where localized template matching is performed to locate the corresponding ships. In this sense, the proposed framework is a practical cross-resolution association pipeline rather than a new detector or similarity metric, thereby mitigating the adverse impact of resolution differences on association performance.

To quantitatively evaluate the proposed method, paired remote sensing images from the GaoFen-1 (GF-1) satellite over the same sea area were collected for experiments. The dataset contains two imaging modalities with different spatial resolutions: the 8 m multispectral (MS) image and the 16 m wide field (WF) image. These paired images provide a practical test case for multi-resolution ship association. We first analyzed the impact of resolution differences on ship detection performance and then conducted comparative

experiments with other association methods. The results show that the proposed method outperforms the baselines in multi-resolution ship association.

The main contributions of this study are summarized as follows:

- We propose a multi-resolution ship association framework that combines high-resolution object detection with low-resolution template matching.
- We design an association strategy that leverages high-resolution detection results to constrain the search space in low-resolution images, improving the robustness of matching under resolution differences.
- We validate the proposed framework on GF1 satellite datasets and demonstrate clear performance gains over representative baseline methods under different detector settings.

The remainder of this paper is organized as follows. Section 2 reviews related work on ship detection and ship association. Section 3 presents the proposed multi-resolution ship association framework. Section 4 describes the dataset, experimental settings, and results. Section 5 concludes the paper and discusses future work.

2. Related Work

2.1. Ship Detection in Remote Sensing Imagery

Ship detection methods in remote sensing imagery have evolved from handcrafted feature-based approaches to deep learning-based detectors. Early methods mainly relied on manually designed features such as HOG, SIFT, and texture or intensity descriptors to distinguish ships from the background [19–21]. Recently, deep learning detectors, including Faster R-CNN, SSD, YOLO, and DETR, have achieved superior performance by learning discriminative representations directly from data [23–26]. Despite these advances, detection accuracy remains sensitive to spatial resolution, especially for small or blurred ship targets in low-resolution imagery.

2.2. Ship Association Methods

Existing ship association methods generally follow a detect-then-associate paradigm. After detecting ships independently in each image, correspondence is established using similarity measures such as perceptual hash (Phash), structural similarity (SSIM), and normalized cross-correlation (NCC) [28–30]. Template matching has also been widely adopted because of its simplicity and computational efficiency [31]. In addition, learning-based matching methods have been explored to improve robustness under appearance variation [32]. However, existing ship detection and association methods mainly provide the basic components for this task; they are not explicitly designed to exploit high-resolution detections to constrain association in low-resolution images. Conventional detect-then-associate pipelines typically assume comparable scale and appearance across image pairs, while learning-based matching methods usually require more extensive cross-resolution supervision. As a result, their effectiveness is limited in multi-resolution scenarios.

2.3. Motivation

In practical satellite observation, images of the same sea area are often acquired at different spatial resolutions. Under such conditions, conventional association methods may become unreliable because target appearance, scale, and detection quality vary significantly across images. To address this issue, this study proposes a multi-resolution ship association framework that uses high-resolution detection results to guide target localization in low-resolution imagery.

3. Proposed Method

The schematic diagram of the proposed multi-resolution ship association framework is shown in Figure 2. The framework consists of two stages: high-resolution ship detection and low-resolution ship association. Its core idea is not to redesign the detector or the similarity metric, but to couple a high-resolution detector with a constrained low-resolution matching stage.

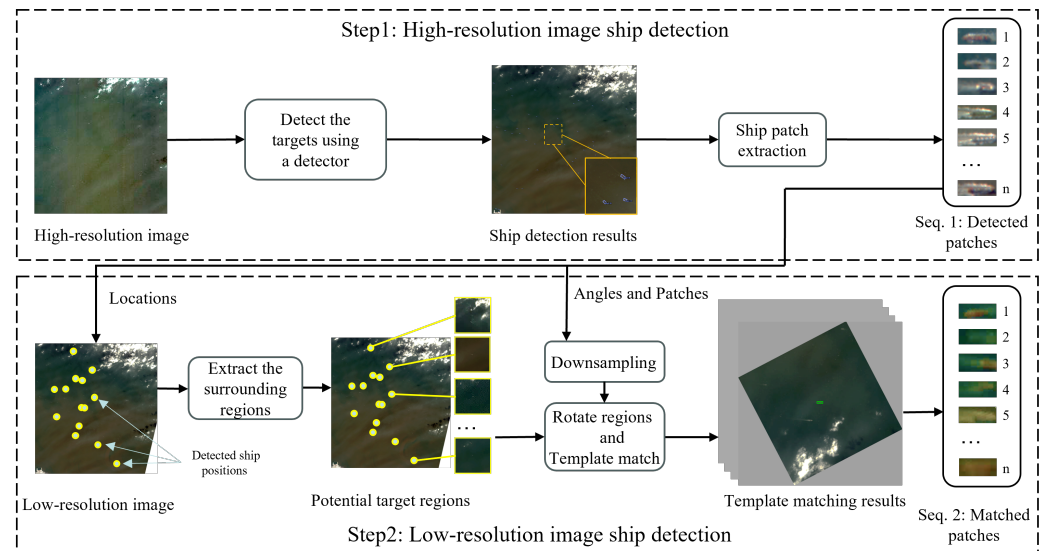


Figure 2. Framework of the proposed multi-resolution ship association method. The numbered boxes denote ship patches in Seq. 1 (detected patches) and Seq. 2 (matched patches).

In the high-resolution stage, an object detector is applied to the high-resolution image, and the detected ships are recorded as Sequence 1. In the low-resolution stage, prior information derived from Sequence 1 is used to locate the corresponding targets. Since the positional changes of ships within a short time interval are relatively small, candidate regions around the detected positions are extracted from the low-resolution image. The detection patches from the high-resolution image are then used as templates for template matching within these candidate regions, yielding the ship detections in the low-resolution image, which are recorded as Sequence 2. The overall workflow of the proposed method is summarized in Algorithm 1.

Algorithm 1 Multi-resolution Ship Association Method.

Input: High-resolution image I_h , low-resolution image I_l

Output: Two associated ship sequences, S_1 and S_2

- 1: Detect ships in I_h and record each ship's position (x_i, y_i) , azimuth θ_i , and patch I_{obj_i} to S_1 as in Equation (1)
 - 2: **for** each $((x_i, y_i), \theta_i, I_{obj_i})$ in S_1 **do**
 - 3: Extract the region of interest ROI_i centered at (x'_i, y'_i) from I_l as in Equation (2)
 - 4: Rotate ROI_i by angle θ_i to obtain I_{rot_i} as in Equation (6)
 - 5: Downsample I_{obj_i} to obtain T_i as in Equation (3)
 - 6: Perform template matching between T_i and I_{rot_i} to determine the associated ship location (x'_i, y'_i) as in Equations (5) and (8)
 - 7: Extract the ship patch I'_{obj_i} from (x'_i, y'_i) and add it to S_2
 - 8: **end for**
 - 9: **return** S_1, S_2
-

3.1. High-Resolution Image Ship Detection

In the high-resolution stage, a representative object detector, such as Faster R-CNN, SSD, YOLO, or DETR, is first applied to detect ships in the high-resolution image [23–25,33]. The detected ships, including their positions, azimuths, and patches, are then extracted and stored in Sequence 1. These results are subsequently used to guide detection and association in the low-resolution image. Sequence 1 is defined as

$$S_1 = \{((x_i, y_i), \theta_i, \mathbf{I}_{obj_i}) \mid ((x_i, y_i), \theta_i, \mathbf{I}_{obj_i}) \in D(\mathbf{I}_h)\} \quad (1)$$

where S_1 denotes Sequence 1, D represents the detector, $\mathbf{I}_h$ denotes the high-resolution image, and $((x_i, y_i), \theta_i, \mathbf{I}_{obj_i})$ represents the position, azimuth, and ship patch of the i -th ship.

3.2. Low-Resolution Image Ship Detection

Since the high-resolution and low-resolution images cover the same area within a short time interval, the same targets are likely to appear in both images at nearby locations. Therefore, the detection results from the high-resolution image can be used as auxiliary information to facilitate target detection in the low-resolution image. The low-resolution detection process consists of two steps: extracting regions of interest (ROIs) from the low-resolution image based on Sequence 1, and then applying template matching within these ROIs to detect ships.

3.2.1. Extracting Regions of Interest

Using the target positions in Sequence 1 as the center, a candidate region is extracted from the low-resolution image as the ROI. This process is defined as

$$\mathbf{ROI} = \mathbf{I}_l(x - r : x + r, y - r : y + r) \quad (2)$$

where $\mathbf{ROI}$ denotes the region of interest, $\mathbf{I}_l$ represents the low-resolution image, and (x, y) denotes the ship positions stored in Sequence 1. The variable r is the radius of the region in image pixels, with the constraint $r \geq v \cdot t / R$, where v is the ship speed, t is the time interval between satellite acquisitions, and R is the spatial resolution of the WF image (16 m spatial resolution). To determine v , we analyzed Automatic Identification System (AIS) data from approximately 20,000 ships [34], as summarized in Table 1. The AIS dataset covers 20,513 ships from 65 ship categories, with a mean speed of 1.3 m/s and a maximum speed of 19.0 m/s. In our experimental setting, the time interval t does not exceed 30 s, yielding $v \cdot t / R = 35.6$. Considering ocean currents and variations in viewing geometry, we set $r = 200$ to improve robustness.

Table 1. Statistics of Automatic Identification System (AIS) vessel data used to estimate the ROI radius.

Number of Ships	Number of Ship Categories	Speed (m/s)	
		Mean	Maximum
20,513	65	1.3	19.0

3.2.2. Downsampling

Downsampling is performed to ensure that the target patch in the high-resolution image is resized to match the pixel dimensions of the corresponding target in the low-resolution image. Bicubic interpolation, which is simple and effective, is commonly used for reconstructing remote sensing images at different resolutions [35]. The downsampling process for patches is defined as follows:

$$\mathbf{T}(x', y') = \sum_{m=-1}^2 \sum_{n=-1}^2 w(m)w(n)\mathbf{I}_{obj_i}(x + m, y + n) \quad (3)$$

$$x' = \frac{x}{d}, \quad y' = \frac{y}{d} \quad (4)$$

where $\mathbf{I}_{obj_i}(x, y)$ denotes the pixel value at (x, y) in the original high-resolution patch, $\mathbf{T}(x', y')$ denotes the pixel value in the downsampled patch, and d denotes the down-sampling factor. $w(m)$ and $w(n)$ are weighting coefficients determined by the bicubic interpolation function, and m and n denote distances between pixels.

3.2.3. Rotate and Template Matching

Template matching searches for a target template within an image by sliding the template across the test image and computing the similarity between the template and each candidate region. In this work, normalized cross-correlation (NCC) is used to measure similarity. NCC normalizes the intensity values and therefore reduces the influence of brightness and contrast variations.

In the high-resolution detection stage, a directional detector is used to generate bounding boxes at different azimuths, which reduces background interference compared with horizontal bounding boxes. However, during template matching, the template patch is horizontally oriented, which may cause angular misalignment with the target in the ROI. To address this issue, the ROI is rotated so that the target is aligned horizontally before matching.

The rotation and template matching process is defined by

$$\mathbf{C}(x, y) = \frac{\sum_{i=0}^{w-1} \sum_{j=0}^{h-1} (\mathbf{T}(i, j) - \bar{\mathbf{T}})(\mathbf{I}_{\text{rot}}(x + i, y + j) - \bar{\mathbf{I}}_{\text{rot},x,y})}{\sqrt{\sum_{i=0}^{w-1} \sum_{j=0}^{h-1} (\mathbf{T}(i, j) - \bar{\mathbf{T}})^2 \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} (\mathbf{I}_{\text{rot}}(x + i, y + j) - \bar{\mathbf{I}}_{\text{rot},x,y})^2}} \quad (5)$$

$$\mathbf{I}_{\text{rot}} = \mathbf{ROI} \cdot \mathbf{R}(\theta) \quad (6)$$

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (7)$$

where $\mathbf{C}(x, y)$ denotes the similarity at position (x, y) . $\mathbf{T}(i, j)$ denotes the pixel value at position (i, j) in the template, and $\bar{\mathbf{T}}$ is the mean intensity of the template. w and h denote the width and height of the template, respectively. $\mathbf{I}_{\text{rot}}(x + i, y + j)$ denotes the pixel value at position $(x + i, y + j)$ in the rotated ROI, and $\bar{\mathbf{I}}_{\text{rot},x,y}$ denotes the mean value of the corresponding local region. The terms $(\mathbf{T}(i, j) - \bar{\mathbf{T}})$ and $(\mathbf{I}_{\text{rot}}(x + i, y + j) - \bar{\mathbf{I}}_{\text{rot},x,y})$ suppress local brightness variations, emphasize relative intensity changes, and improve the robustness of matching. $\mathbf{R}$ is the rotation matrix, and θ is the azimuth angle.

The final matching position (x', y') of the target patch $\mathbf{I}'_{obj}$ is obtained by

$$(x', y') = \arg \max_{(x,y)} \mathbf{C}(x, y) \quad (8)$$

where (x', y') denotes the optimal matching position at which $\mathbf{C}(x, y)$ reaches its maximum value. After performing region extraction and association for all targets in Sequence 1, the corresponding targets form Sequence 2.

4. Experiments

4.1. Data Description

Figure 3 provides detailed information about the experimental dataset GF1-Bohai used in our study. This dataset was acquired from the GF1 satellite and contains images of the Bohai Sea, China, captured by both the 8-meter-resolution multispectral camera (MS) and

the 16-meter-resolution wide-field camera (WF). The dataset comprises three paired scenes, namely Scene 1, Scene 2 and Scene 3. Each scene contains one high-resolution image and one low-resolution image of the same region.

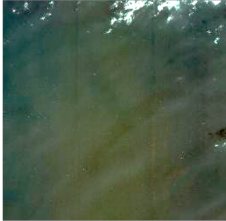
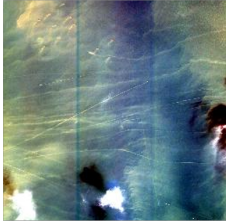
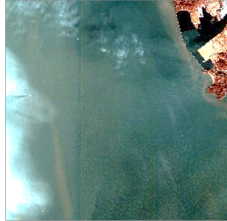
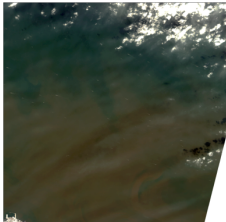
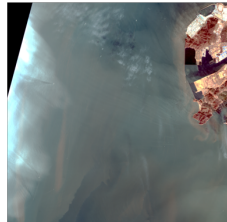
Scenes		Scene 1	Scene 2	Scene 3
Image	MS			
	WF			
Data source		GF1 MS/WF		
Shooting time		2020-12-17	2020-04-06	2020-04-14
		02:45:21 02:45:25	02:49:07 02:48:54	02:44:17 02:44:15
Latitude and longitude range		122.284E	119.123E	120.920E
		122.587E	119.432E	121.228E
		37.769N 37.453N	38.883N 38.562N	39.514N 39.196N
Resolution		8m/16m		

Figure 3. Introduction of the GF1-Bohai experimental dataset. MS denotes the 8 m image acquired by the multispectral camera, and WF denotes the 16 m image acquired by the wide-field camera.

4.2. Experiment Setup

4.2.1. Detector

Before conducting the main ship association experiments, we first evaluated the detection performance of different models on images with two spatial resolutions. To this end, three representative object detectors were selected: Faster R-CNN, DETR, and YOLOv5. Horizontal comparisons were used to examine the performance differences among the detectors, while vertical comparisons were used to assess the detection effectiveness of each detector under different resolutions.

To avoid overfitting, additional GF1 images with 8 m and 16 m spatial resolutions were used for training, while GF1-Bohai was reserved exclusively for testing. For Faster R-CNN, the main training settings were a learning rate of 0.005, and inference thresholds of 0.05 for confidence and 0.5 for non-maximum suppression. For DETR, the learning rate was set to 1×10^{-4} , and the same random crop and resize augmentation was applied during training; DETR does not use non-maximum suppression. For YOLOv5, the learning rate was set to 0.01, and the inference thresholds were 0.25 for confidence and 0.45 for intersection-over-union. The same random crop and resize augmentation was used for all detectors.

As shown in Table 2, DETR achieved the best overall recall and precision, followed by YOLOv5, whereas Faster R-CNN performed the worst. For each detector, performance also varied noticeably across resolutions, with both recall and precision generally higher on the high-resolution MS images than on the low-resolution WF images. These detection errors, including false positives and missed detections, are mainly caused by resolution differences and may reduce the accuracy of subsequent association.

Table 2. Performance of different detection models on MS and WF images.

Scenes		Ground Truth	Faster R-CNN				DETR				YOLOv5			
			TP	FP	Recall	Precision	TP	FP	Recall	Precision	TP	FP	Recall	Precision
Scene 1	MS	45	45	5	1.000	0.900	45	3	1.000	0.938	44	3	0.978	0.936
	WF	51	46	21	0.902	0.687	49	18	0.961	0.731	47	22	0.922	0.681
Scene 2	MS	23	23	1	1.000	0.958	23	0	1.000	1.000	23	1	1.000	0.958
	WF	26	22	5	0.846	0.815	26	3	1.000	0.897	24	6	0.923	0.800
Scene 3	MS	29	28	4	0.966	0.875	29	2	1.000	0.935	27	2	0.931	0.931
	WF	37	32	9	0.865	0.780	34	6	0.920	0.850	34	8	0.920	0.802

4.2.2. Comparison

To demonstrate the effectiveness of the proposed method, we compared it with representative ship association methods. In a typical association pipeline, objects are first detected independently in the two images, and then image similarity is computed to establish correspondences between matched targets.

For the detection stage, Faster R-CNN, DETR, and YOLOv5 were employed. For similarity computation, Phash, SSIM, and NCC were adopted. Phash generates a compact numerical fingerprint for an image through operations such as discrete cosine transform, feature selection, and hash encoding, and measures similarity by comparing the Hamming distance between fingerprints [28]. SSIM evaluates image similarity by comparing luminance, contrast, and structural information, producing a scalar value that reflects perceived visual similarity [29]. In addition, ResNet-50 with cross-entropy and hard triplet loss was used as a learning-based matching baseline [36]. Given ship patches extracted from the two images, the network outputs feature embeddings, and the final associations are obtained by matching feature similarities between embeddings. In the association stage, the Kuhn–Munkres (KM) algorithm was used to obtain one-to-one correspondences while maximizing the overall similarity score. Phash, SSIM, NCC, and the KM algorithm were implemented using standard off-the-shelf functions or solvers.

4.3. Experiments Results

4.3.1. Intermediate Steps in Low-Resolution Image Ship Detection

The intermediate process and results of low-resolution ship detection are illustrated in Figure 4. Panel (a) shows the detection results on the high-resolution image. Panel (b) presents the candidate regions extracted from the low-resolution image based on the positional information in (a). Panel (c) displays the red–blue heatmaps obtained after rotation and template matching on the regions in (b), where warmer colors indicate higher similarity. Panel (d) shows the final ship detections in the low-resolution image, highlighted with green boxes.

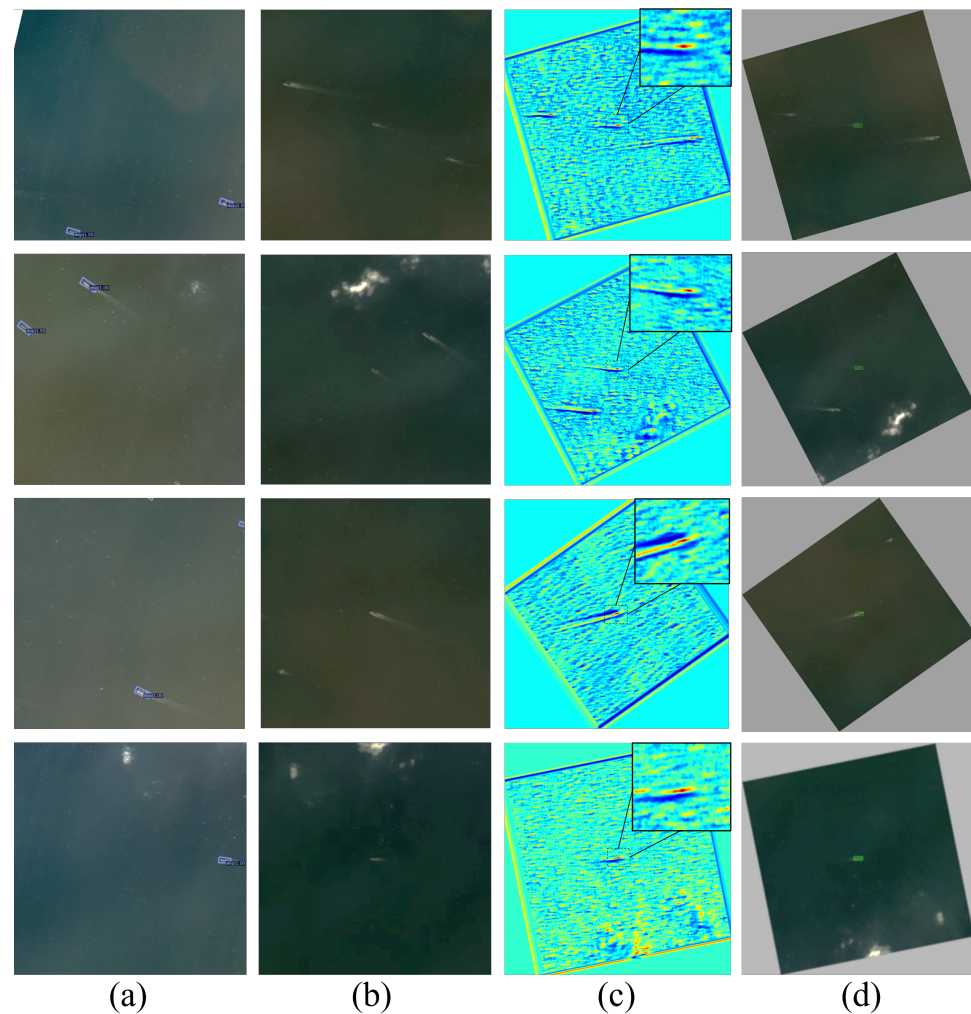


Figure 4. Intermediate process of the proposed multi-resolution ship association method. (a) Ship detection results in the high-resolution image. (b) Candidate regions extracted from the low-resolution image based on the detected ship positions. (c) NCC similarity heatmaps after ROI rotation and template matching, where warmer colors indicate higher similarity. (d) Final ship association results in the low-resolution image.

4.3.2. Association Results of Different Methods

Table 3 compares the ship association results of the proposed method with those of the baseline methods. The association accuracy is defined as

$$\text{Accuracy} = \frac{N_{\text{correct}}}{N_{\text{total}}} \quad (9)$$

where N_{correct} denotes the number of correctly associated sample pairs and N_{total} denotes the total number of sample pairs.

Overall, detectors with higher recall and precision tend to yield better association accuracy, although the improvement is limited. Under the same detector setting, the learning-based ResNet baseline performs better than the traditional similarity-based methods, while the proposed method consistently achieves the highest association accuracy. For completeness, Table 4 lists, for each scene, the total number of ground-truth ship pairs and the number of correctly associated ship pairs obtained by the proposed method under the YOLOv5 detector setting. This provides an intuitive complement to the accuracy values reported in Table 3.

Table 3. Comparison of association accuracy among different methods.

Scenes	Faster R-CNN+				
	Phash + KM	SSIM + KM	NCC + KM	ResNet + KM	Ours
Scene 1	0.222	0.356	0.444	0.511	0.867
Scene 2	0.174	0.304	0.348	0.478	0.826
Scene 3	0.103	0.207	0.241	0.448	0.621

Scene	DETR+				
	Phash + KM	SSIM + KM	NCC + KM	ResNet + KM	Ours
Scene 1	0.333	0.511	0.644	0.800	0.933
Scene 2	0.304	0.434	0.478	0.609	0.870
Scene 3	0.172	0.276	0.414	0.586	0.862

Scene	YOLOv5+				
	Phash + KM	SSIM + KM	NCC + KM	ResNet + KM	Ours
Scene 1	0.266	0.511	0.600	0.733	0.933
Scene 2	0.261	0.435	0.478	0.609	0.870
Scene 3	0.137	0.276	0.310	0.586	0.828

Table 4. Numbers of ground-truth ship pairs and correctly associated pairs achieved by the proposed method.

Scene	N_{total}	N_{correct}	Accuracy
Scene 1	45	42	0.933
Scene 2	23	20	0.870
Scene 3	29	25	0.862

Figure 5 provides a more detailed comparison of the experimental results on GF1-Bohai-Scene 1. Using Faster R-CNN as the detector, we compare NCC + KM with the proposed method. GF1-Bohai-Scene 1 contains 45 associable ship pairs. NCC + KM correctly associates 20 pairs, achieving an accuracy of 44.4%. In contrast, the proposed method correctly associates 39 pairs, achieving an accuracy of 86.7%. The superior performance of the proposed method is mainly attributed to the use of high-resolution detection results as matching templates, which alleviates the mismatch caused by resolution differences. These results demonstrate that the proposed framework is effective for ship association under multi-resolution conditions.

4.4. Error Analysis and Discussion

To further analyze the effectiveness of the proposed method, we conduct an error analysis by examining false positives (FPs) and false negatives (FNs) under different scene characteristics, including vessel scale, scene density, and imaging resolution.

Overall, most FNs are caused by missed detections in the low-resolution WF images, particularly for small vessels with weak visual saliency. This issue is more pronounced in dense scenes (e.g., Scene 1), where overlapping wakes and clutter increase detection ambiguity. In contrast, FPs mainly arise from incorrect template matches in homogeneous background regions, especially when multiple candidate regions exhibit similar texture patterns.

Compared with NCC + KM, the proposed method does not directly reduce FN rates, but mitigates their impact by leveraging high-resolution detections as spatial priors, which effectively constrains the search space and prevents missed detections from propagating to the association stage. Meanwhile, FP cases are also reduced due to ROI-based filtering and rotation alignment, which improves structural consistency during matching.

Ship pair	MS	Faster R-CNN+				Ship pair	MS	Faster R-CNN+			
		NCC+KM		Our				NCC+KM		Our	
		WF	T/F	WF	T/F			WF	T/F	WF	T/F
1			×		✓	24			×		✓
2			×		✓	25			✓		✓
3			×		✓	26			✓		✓
4			✓		✓	27			✓		✓
5			×		✓	28			×		×
6			×		✓	29			×		✓
7			×		✓	30			×		×
8			×		✓	31			✓		✓
9			✓		✓	32			×		✓
10			✓		✓	33			✓		✓
11			×		×	34			✓		✓
12			×		×	35			✓		✓
13			×		✓	36			×		✓
14			✓		✓	37			×		✓
15			×		✓	38			✓		✓
16			✓		✓	39			✓		✓
17			✓		✓	40			×		✓
18			×		✓	41			×		✓
19			✓		✓	42			×		×
20			✓		✓	43			×		✓
21			✓		✓	44			×		×
22			✓		✓	45			×		✓
23			✓		✓						

Figure 5. Pairwise comparison of association results between NCC + KM and our method on GF1-Bohai Scene 1. T/F columns indicate whether the association is correct, where “✓” denotes a correct match and “×” denotes an incorrect match.

These observations indicate that the proposed framework is particularly robust in scenarios with moderate to high vessel density, while performance degradation is mainly associated with extreme low-resolution conditions where initial detections are incomplete.

4.5. Ablation Study on ROI Radius

To investigate the impact of the ROI radius r on association performance, a sensitivity analysis is conducted by varying r while keeping all other components fixed on Scene 1 under YOLOv5. The ROI defines the spatial search range in the low-resolution image and directly affects both matching robustness and computational cost.

As reported in Table 5, a small radius ($r = 40$) leads to performance degradation due to insufficient coverage of target displacement caused by vessel motion and geo-registration uncertainty. Increasing r to 120 significantly improves accuracy, indicating that most true correspondences can be captured within a moderate search region. When r is further increased to 200, the performance saturates, suggesting that the effective displacement range is already sufficiently covered. However, an overly large radius ($r = 300$) introduces additional background interference, which degrades matching reliability.

Table 5. Sensitivity analysis of ROI radius on Scene 1 under YOLOv5.

ROI Radius r (Pixels)	Accuracy	Runtime (ms)
40	0.266	0.5
120	0.933	3
200	0.933	8
300	0.800	18

In terms of efficiency, runtime increases with r due to the enlarged search space. Overall, $r \in [120, 200]$ provides a good trade-off between accuracy and efficiency and is adopted in all experiments.

5. Conclusions

This study addresses the problem of ship association in multi-resolution satellite imagery. Conventional detect-then-associate methods often become unreliable in this setting because ships may appear with different scales and detection quality in high-resolution and low-resolution images. To alleviate this issue, we propose a multi-resolution ship association framework that uses ship detections from the high-resolution image as auxiliary information to guide template matching in the low-resolution image. The framework combines high-resolution ship detection, low-resolution ROI extraction, rotation alignment, downsampling, and NCC-based template matching to locate corresponding ships across images with different spatial resolutions.

Experiments on the GF1-Bohai dataset show that the proposed framework consistently outperforms representative baseline methods under different detector settings. The results indicate that using high-resolution detections as templates can effectively improve ship association accuracy under multi-resolution conditions and reduce the impact of missed detections and false detections caused by resolution differences.

The current evaluation is limited to paired MS and WF images from the GF1 satellite over the Bohai Sea. Future work will extend the proposed framework to additional geographic regions, larger resolution gaps, and heterogeneous sensors to further examine its generalization capability.

Author Contributions: Conceptualization and methodology, Y.Z.; software and data curation, Y.Z.; writing—original draft preparation, Y.Z.; project administration and funding acquisition, X.K.; writing—review and editing, P.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key Research and Development Program of China under Grant 2021 YFA0715203, in part by the National Natural Science Foundation of China under Grant 62525108 and Grant 62371185, in part by the Science and Technology Innovation Program of Hunan Province under Grant 2024RC1030 and Grant 2023RC3124, in part by the Project of Yuelushan Center for Industrial Innovation under Grant 2025YCII0202, and in part by the Fundamental Research Funds for the Central Universities.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Li, F.; Yu, K.; Yuan, C.; Tian, Y.; Yang, G.; Yin, K.; Li, Y. Dark ship detection via optical and SAR collaboration: An improved multi-feature association method between remote sensing images and AIS data. *Remote Sens.* **2025**, *17*, 2201. [\[CrossRef\]](#)
2. Salerno, E.; Di Paola, C.; Lo Duca, A. Remote sensing for maritime monitoring and vessel identification. *Remote Sens.* **2024**, *16*, 776. [\[CrossRef\]](#)
3. Renga, A.; Graziano, M.D.; Moccia, A. Segmentation of marine SAR images by sublook analysis and application to sea traffic monitoring. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1463–1477. [\[CrossRef\]](#)
4. Mou, F.; Fan, Z.; Ge, Y.; Wang, L.; Li, X. An efficient ship detection method based on YOLO and ship wakes using high-resolution optical Jilin1 satellite imagery. *Sensors* **2024**, *24*, 6708. [\[CrossRef\]](#) [\[PubMed\]](#)

5. Zhao, T.; Wang, Y.; Li, Z.; Gao, Y.; Chen, C.; Feng, H.; Zhao, Z. Ship detection with deep learning in optical remote-sensing images: A survey of challenges and advances. *Remote Sens.* **2024**, *16*, 1145. [\[CrossRef\]](#)
6. Wang, H.; Li, S.; Yang, J.; Liu, Y.; Lv, Y.; Zhou, Z. Cross-modal ship re-identification via optical and SAR imagery: A novel dataset and method. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 19–20 October 2025; pp. 7873–7883.
7. Galdelli, A.; Narang, G.; Pietrini, R.; Zazzarini, M.; Fiorani, A.; Tassetti, A.N. Multimodal AI-enhanced ship detection for mapping fishing vessels and informing on suspicious activities. *Pattern Recognit. Lett.* **2025**, *191*, 15–22. [\[CrossRef\]](#)
8. Leng, X.; Ji, K.; Kuang, G. Ship detection from raw SAR echo data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5207811. [\[CrossRef\]](#)
9. Liu, S.; Qu, C.; Xu, M.; Wan, J.; Sheng, H.; Zeng, Z.; Cui, J. An improved Kuhn-Munkres algorithm for ship matching in optical satellite images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4724–4738. [\[CrossRef\]](#)
10. Wang, Y.; Ning, X.; Leng, B.; Fu, H. Ship detection based on deep learning. In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2019; pp. 275–279. [\[CrossRef\]](#)
11. Yu, W.; You, H.; Lv, P.; Hu, Y.; Han, B. A moving ship detection and tracking method based on optical remote sensing images from the geostationary satellite. *Sensors* **2021**, *21*, 7547. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Chen, W.; Han, B.; Yang, Z.; Gao, X. MSSDet: Multi-scale ship-detection framework in optical remote-sensing images and new benchmark. *Remote Sens.* **2022**, *14*, 5460. [\[CrossRef\]](#)
13. Jiang, Z.; Wang, Y.; Zhou, X.; Chen, L.; Chang, Y.; Song, D.; Shi, H. Small-scale ship detection for SAR remote sensing images based on coordinate-aware mixed attention and spatial semantic joint context. *Smart Cities* **2023**, *6*, 76. [\[CrossRef\]](#)
14. Li, L.; Jiang, L.; Zhang, J.; Wang, S.Y.; Chen, F. A complete YOLO-based ship detection method for thermal infrared remote sensing images under complex backgrounds. *Remote Sens.* **2022**, *14*, 1534. [\[CrossRef\]](#)
15. Zhang, Z.; Zhang, L.; Wang, Y.; Feng, P.; He, R. ShipRSImageNet: A large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 8458–8472. [\[CrossRef\]](#)
16. Wolfe, J.D.; Speyer, J.L. Target association using detection methods. *J. Guid. Control Dyn.* **2002**, *25*, 1143–1148. [\[CrossRef\]](#)
17. Dufournaud, Y.; Schmid, C.; Horaud, R. Matching images with different resolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hilton Head Island, SC, USA, 15 June 2000; Volume 1, pp. 612–618. [\[CrossRef\]](#)
18. Huang, Z.; Han, H. Repeat-ground-track orbit design and analysis for remote sensing in specific areas. *J. Phys. Conf. Ser.* **2025**, *2977*, 012014. [\[CrossRef\]](#)
19. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
20. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
21. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3446–3456. [\[CrossRef\]](#)
22. Guiming, S.; Jidong, S. Remote sensing image edge-detection based on improved Canny operator. In Proceedings of the 8th IEEE International Conference on Communication Software and Networks (ICCSN), Beijing, China, 4–6 June 2016; pp. 652–656.
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
25. Huang, R.; Pedoeem, J.; Chen, C. YOLO-LITE: A real-time object detection algorithm optimized for non-GPU computers. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2503–2510.
26. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 213–229.
27. Yang, F.; Xu, Q.; Li, B.; Ji, Y. Ship detection from thermal remote sensing imagery through region-based deep forest. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 449–453. [\[CrossRef\]](#)
28. Laradji, I.H.; Ghouti, L.; Khiari, E.H. Perceptual hashing of color images using hypercomplex representations. In Proceedings of the 2013 IEEE International Conference on Image Processing (ICIP), Melbourne, Australia, 15–18 September 2013; pp. 4402–4406. [\[CrossRef\]](#)
29. Bakurov, I.; Buzzelli, M.; Schettini, R.; Castelli, M.; Vanneschi, L. Structural similarity index (SSIM) revisited: A data-driven approach. *Expert Syst. Appl.* **2022**, *189*, 116087. [\[CrossRef\]](#)
30. Li, D.; Zhang, Y. A fast normalized cross-correlation algorithm for InSAR image subpixel registration. In Proceedings of the 3rd International Asia-Pacific Conference on Synthetic Aperture Radar (APSAR), Seoul, Korea, 26–30 September 2011; pp. 1–4.
31. Korman, S.; Reichman, D.; Tsur, G.; Avidan, S. Fast-match: Fast affine template matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 2331–2338.

32. Jogin, M.; Madhulika, M.S.; Divya, G.D.; Meghana, R.K.; Apoorva, S. Feature extraction using convolution neural networks (CNN) and deep learning. In Proceedings of the 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 18–19 May 2018; pp. 2319–2323.
33. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable transformers for end-to-end object detection. *arXiv* **2020**, arXiv:2010.04159.
34. Bureau of Ocean Energy Management (BOEM) and National Oceanic and Atmospheric Administration (NOAA). AIS Vessel Transit Counts 2023. Available online: <http://MarineCadastre.gov> (accessed on 20 March 2025).
35. Khaledyan, D.; Amirany, A.; Jafari, K.; Moaiyeri, M.H.; Khuzani, A.Z.; Mashhadi, N. Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution. In Proceedings of the 2020 IEEE Global Humanitarian Technology Conference (GHTC), Online Conference, 29 October–1 November 2020; pp. 1–5. [[CrossRef](#)]
36. Qiao, D.; Liu, G.; Dong, F.; Jiang, S.X.; Dai, L. Marine vessel re-identification: A large-scale dataset and global-and-local fusion-based discriminative feature learning. *IEEE Access* **2020**, *8*, 27744–27756. [[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.