Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.077977

ARTICLE

Dual-Stream Feature Decoupling and Temporal Variational Bayesian Inference for Ship Re-Identification with Incomplete Data

Wanhui Qiao[1], Xiaorui Zhang[1,*], Wei Sun[2], Shiyu Zhou[3] and Kaibo Wang[2]

[1]College of Computer and Information Engineering, Nanjing Tech University, Nanjing, China
[2]School of Automation, Nanjing University of Information Science & Technology, Nanjing, China
[3]School of Computer Science, Nanjing University of Information Science & Technology, Nanjing, China

*Corresponding Author: Xiaorui Zhang. Email: zxr365@126.com

Received: 21 December 2025; Accepted: 13 March 2026; Published: 08 May 2026

ABSTRACT: Ship re-identification (Re-ID) aims to match ship identities across disjoint camera views and separated time periods, which is critical for maritime target tracking and law enforcement. In real-world surveillance, variations in target distance and viewing angle frequently produce partial views and occlusions, leading to missing geometric components and fragmented appearance cues. Such incomplete observations substantially degrade the robustness and generalization of conventional single-frame methods that rely on global appearance representations. To address these challenges, this study proposes a new ship re-identification framework based on dual-stream feature decoupling and temporal variational Bayesian inference. The proposed method explicitly disentangles ship representations into appearance and structural streams, and leverages multi-frame temporal context to infer missing components and enhance discriminability under partial visibility. Specifically, a ResNet-based splitter trained adversarially against two discriminators is employed to decouple the input representation into separate feature streams. The decoupled streams are then modeled over time using a bidirectional LSTM (BiLSTM) together with a visibility-probability estimator. A graph-structured spatial prior, parameterized via a graph attention network (GAT), serves as the variational prior. Given sequential observations, the variational inference module estimates posterior distributions for missing components and performs probabilistic completion in the latent space. The framework is trained end-to-end using cross-entropy and triplet losses. Extensive experiments on the Ship-CH dataset demonstrate that our method achieves 85.67% mAP and 93.67% Rank-1 accuracy, exhibiting superior robustness under occlusion and partial visibility.

KEYWORDS: BiLSTM; bayesian inference; feature decoupling; triplet loss; cross-entropy loss; ship re-identification

1 Introduction

In the fields of maritime surveillance and maritime safety, ship re-identification (Ship Re-ID) plays a critical role. The core objective of the ship re-identification task is to retrieve all images of the same ship from a large-scale ship image gallery captured by different cameras [1]. This task is a key subproblem in the field of image retrieval. Owing to its broad application prospects and practical value in continuous ship tracking and maritime security, it has attracted significant attention from the computer vision community. However, maritime surveillance typically covers vast areas, and camera deployments are difficult to achieve full coverage. As a result, ships are often captured at excessively close distances, leading to partial visibility and incomplete feature representations, which severely constrain recognition performance. This not only reduces the uncertainty of maritime navigation but also poses challenges to maritime traffic management and rescue operations. Therefore, investigating ship re-identification under partially visible conditions is of

significant importance for enhancing feature recovery capabilities in close-range and occluded scenarios, helping to expand the application scope of ship re-identification and strengthen maritime supervision and safety control.

Most existing methods are based on single-image recognition and perform well when the input images are complete. However, when faced with missing components, the absence of cross-frame complementarity severely limits the model's ability to reconstruct missing regions, resulting in a notable degradation in recognition accuracy. Notably, ships exhibit distinct temporal coherence during navigation. For instance, in close-range scenarios, the camera may be too close to capture the entire ship within a single frame (Fig. 1g). As the ship gradually moves through the field of view, however, the bow, hull, and stern enter the scene sequentially, as illustrated in Fig. 1a–f. While individual frames may contain limited information, consecutive frame sequences preserve rich complementary features. To exploit this potential, this study proposes a framework that utilizes multi-frame image sequences to extract complementary appearance and structural information. First, a unique dual-stream feature decoupling mechanism explicitly separates appearance and structural features, providing distinct and complementary representations for subsequent processing. Building upon this, temporal modeling methods, such as Bidirectional LSTMs (BiLSTM), are employed to capture inter-frame correlations and the temporal evolution of ship components. Ultimately, by leveraging component-level spatial priors and temporal contextual information, the proposed method achieves robust completion of incomplete ship information for ship re-identification under incomplete data conditions.

Figure 1: Illustration of temporal coherence and feature complementarity in ship sequences. (**a**–**f**) Consecutive partial observations of the same ship as it gradually moves across the camera's field of view, where different structural components (e.g., bow, hull, and stern) sequentially appear; (**g**) a relatively complete observation of the ship used for comparison.

First, in terms of multi-frame image feature extraction, most existing methods directly perform unified analysis on holistic ship features; however, this strategy suffers from evident limitations. Typically, holistic ship features are encoded in a unified manner, which overlooks the inherent differences between appearance and structural characteristics. Ship appearance attributes, such as color and texture, are susceptible to environmental variations, including illumination changes and wave-induced interference [2,3]. In contrast, structural features, including contours and component layout, remain relatively invariant. Indiscriminately mixing these two types of features allows appearance-induced noise, such as reflections, ripples, and shadows, to contaminate structural representations, thereby degrading geometric accuracy and model robustness [4]. Prior studies suggest that decoupling ship features into appearance and structure streams enables more targeted attribute extraction and reduces noise interference, while also yielding clearer representation spaces for subsequent temporal modeling and missing-data completion. As shown in Fig. 2. Ultimately, this decoupling strategy facilitates the simultaneous capture of dynamic appearance changes and stable structure constraints, significantly enhancing recognition accuracy and generalization under complex environments and incomplete observations [5].

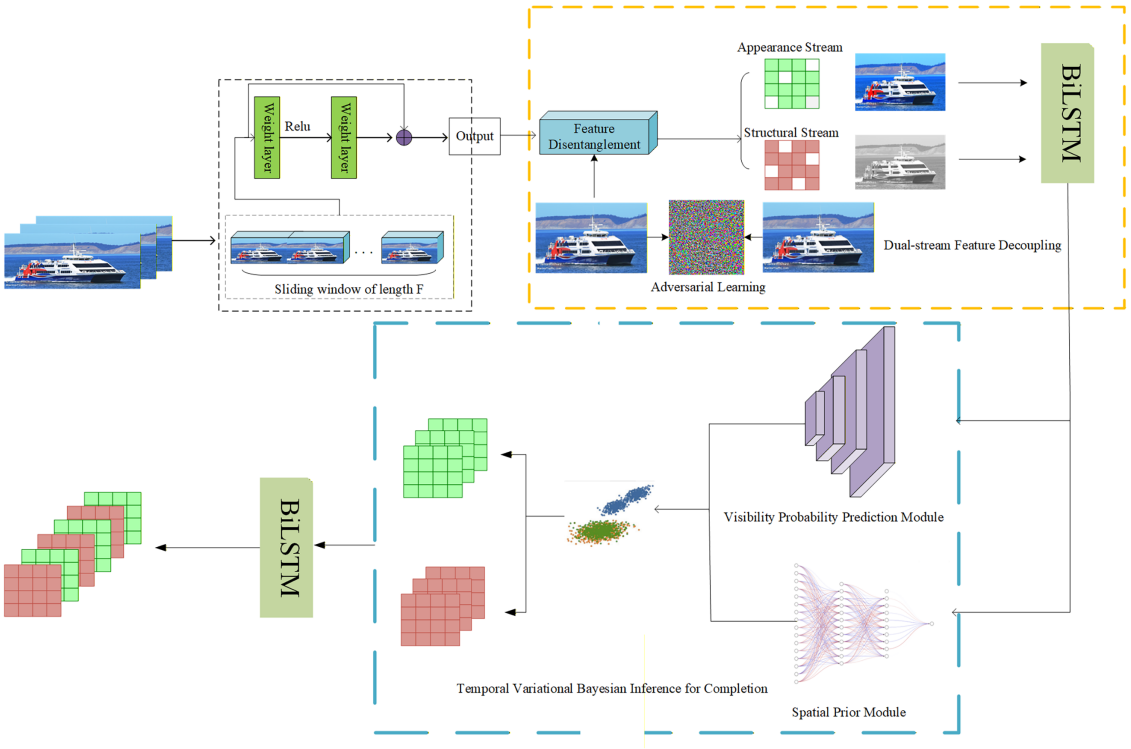

Figure 2: Overall framework for ship re-identification.

Subsequently, regarding temporal modeling, the use of image sequences necessitates that the model fully leverage complementary information from both preceding and subsequent frames. For instance, in a six-frame sequence, effectively recognizing the third frame requires simultaneously referencing historical context from the first two frames and future context from the subsequent three frames. This demand requires a model with robust capabilities for capturing bidirectional temporal dependency. Current mainstream approaches for processing ship sequence data predominantly employ models such as CNNs and TCNs (Temporal Convolutional Networks). While these methods exhibit certain advantages in extracting temporal features or local spatial features, they suffer from evident limitations. Specifically, although CNNs and TCNs are capable of capturing local features, they have limited ability to model global temporal dependencies and bidirectional information, and they adapt poorly to non-stationary ship motion sequences [6]. To address these limitations, this paper introduces a bidirectional long short-term memory network (BiLSTM), which simultaneously learns forward and backward information to comprehensively capture ship dynamic variations, thereby providing more discriminative temporal representations for the re-identification task.

Finally, regarding ship feature completion, existing methods predominantly employ global completion strategies. For instance, recent research has achieved significant progress in ship completion and recognition by utilizing pixel-level segmentation (e.g., Mask R-CNN, Faster R-CNN), and multi-scale dense feature fusion networks (e.g., FPN) [7–9]. Although these approaches can partially restore surface information, they largely disregard the intrinsic geometric constraints and spatial relationships between ship components. This oversight often leads to structurally implausible reconstructions and semantic conflicts [9]. For instance, bow features may be erroneously mapped onto the hull. To address this issue, this study constructs a component-level graph attention network. By treating visible components as nodes and establishing edge weights based on spatial proximity and hull symmetry, the network explicitly models the spatial dependencies

between components. Building upon this, the temporal representations output by the BiLSTM are utilized as conditions for variational inference, enabling the probabilistic completion of missing components. This ensures that the completed features exhibit structural coherence and consistency across both spatial and temporal dimensions.

To address these challenges, this study proposes a new ship re-identification method based on dual stream feature decoupling and temporal variational Bayesian inference. The method first employs an adversarial learning mechanism to explicitly decouple input features into distinct appearance and structure streams, effectively mitigating the interference of appearance noise on geometric representations. Subsequently, a bidirectional LSTM (BiLSTM) model is used to model the temporal evolution of both streams, capturing the dynamic context of appearance and structure across the sequence. Building upon this, a component-level graph attention network is introduced to construct spatial priors, which are then combined with temporal features for variational Bayesian inference, enabling probabilistic completion of missing components and thereby improving recognition stability and structural consistency under incomplete ship images. The main novelties of this study are summarized as follows.

(i) Dual-Stream Feature Decoupling: Most existing ship re-identification methods learn a single holistic representation, which overlooks the fundamentally different behaviors of appearance and structural cues under complex maritime environments. To address this limitation, this study introduces an adversarial learning–driven dual-stream feature decoupling strategy that explicitly separates ship representations into an appearance stream and a structural stream. The appearance stream focuses on color and texture attributes that are sensitive to environmental variations, while the structural stream captures relatively stable geometric information such as contours and component layout. By disentangling these two types of features at the representation stage, the proposed design effectively suppresses appearance-induced noise interference and provides complementary and robust representations, laying a clear foundation for subsequent temporal modeling and missing-component completion.

(ii) Temporal Variational Bayesian Inference for Missing Component Completion: Under occlusion and partial visibility, ship observations often suffer from component-level missingness, for which single-frame cues and global completion strategies are insufficient to recover a structurally consistent representation. To overcome this challenge, this paper constructs a temporal variational Bayesian completion mechanism that jointly exploits component-level spatial priors and multi-frame temporal context. Specifically, a graph attention network is employed to model spatial dependencies and symmetry constraints among components within each frame, while a BiLSTM fuses multi-frame observations within a sliding window to capture bidirectional temporal cues. Under this setting, variational Bayesian inference is adopted to probabilistically generate the feature distributions of missing components. This mechanism integrates multi-frame information to correct single-frame errors and produces reliable completion results under uncertainty, significantly improving structural consistency and recognition stability in partially visible ship scenarios.

2 Related Work

Similar to pedestrian re-identification and vehicle re-identification, ship re-identification refers to recognizing the same ship across different scenes, times, or camera perspectives. In the field of ship re-identification, existing approaches predominantly emphasize either global or localized appearance modeling to mitigate recognition challenges arising from viewpoint variation or missing structural components. Organisciak et al. attempted to conduct unified modeling for pedestrian and vehicle re-identification, leveraging mid-level features and hard-example mining to improve generalization. However, their hybrid modeling approach also encountered conflicts in feature representation across different object types.

Qian et al. further noted that while novel attention mechanisms like Transformers enhance feature expression, they remain deficient in structural constraints and dynamic adaptability [10]. In summary, existing methods predominantly focus on holistic or local representations of appearance features, struggling to maintain robustness in scenarios where appearance and structural variation patterns differ significantly. Forced hybrid modeling often leads to appearance noise contaminating structural representations, thereby undermining the effectiveness of geometric constraints [10]. To overcome these limitations, this study proposes an adversarial learning-driven dual-stream feature decoupling network. During feature extraction, the appearance stream and structure stream are completely separated, encoding dynamic surface attributes and static geometric structures, respectively. This eliminates mutual interference at the source and provides complementary and robust feature inputs for subsequent completion and recognition tasks.

On the basis of effective feature decoupling, reliably completing missing components becomes crucial for further improving recognition performance under incomplete ship observations. Here, incomplete ship observations refer to practical surveillance scenarios in which a ship cannot be fully captured within a single frame due to close-range viewpoints or occlusions, such that certain structural regions are partially or entirely invisible at different time steps. As illustrated in Fig. 1a–f, only local portions of the same ship (e.g., the bow, hull, or stern) may be visible in individual frames, whereas a complete representation can be recovered only by aggregating complementary information across the temporal sequence. In the field of incomplete ship re-identification, researchers have proposed a variety of approaches to enhance feature completion capability under local occlusion and component-missing scenarios. For instance, Zhang et al. proposed the SFCFAR method based on superpixel segmentation and feature point fusion. By dividing remote-sensing images into multiple small regions and combining texture features with background differences, it achieves effective detection of incomplete ships under cloud occlusion and low-contrast environments. However, this method primarily targets object detection and fails to deeply model spatial structural relationships between components, resulting in limited completion capabilities [11]. Li et al. proposed the LGFCT keypoint extraction and line-feature-based keypoint prediction method for incomplete 3D point cloud data. The approach infers missing component locations and combines spatiotemporal visualization to analyze ship motion states, enhancing spatial recognition capabilities under incomplete data. However, it primarily emphasizes motion state analysis and fails to achieve collaborative completion of appearance and structural features [12]. Additionally, Zeng et al. employed transfer learning and dynamic alignment networks to improve ship re-identification accuracy in complex sea conditions. Nevertheless, their approach primarily focused on overall recognition performance, with insufficient attention to partial missing parts and temporal completion issues [13]. While these methods achieved some progress in enhancing incomplete ship detection and recognition, they generally suffer from inadequate modeling of spatial dependencies among components and insufficient utilization of multi-frame temporal information [11–13]. Therefore, integrating spatial priors, temporal context, and probabilistic inference to generate structurally sound and credible reconstruction results remains a critical area in this field that necessitates breakthroughs. To address this challenge, this study employs a temporal variational Bayesian completion mechanism. This approach simultaneously integrates multi-frame information to correct single-frame errors and provides reliable completions under uncertainty, effectively handling complex conditions such as partial occlusions, incomplete viewpoints, and dynamic environmental changes. This enhances robustness and generalization capabilities for ship re-identification.

3 Proposed Method

To address the challenge of partially visible ships, this study proposes a new framework that integrates dual-stream feature decoupling with temporal variational Bayesian inference. The framework consists of two core modules: (1) a dual-stream feature decoupling module and (2) a temporal variational Bayesian

inference module. Given consecutive ship images, a residual backbone first extracts high-level features, which are then decoupled into appearance and structural streams. Using the decoupled structural features, we construct spatial priors by modeling component-level visibility and relational dependencies, which provide constraints for subsequent feature completion. By coupling sliding-window temporal encoding with variational Bayesian inference, missing components are then completed in a probabilistic manner. Finally, the completed appearance and structural features are fused to form a unified representation that captures both surface appearance details and geometric structure for ship re-identification.

In this work, components refer to structural feature units defined in the deep feature space, rather than explicitly segmented physical ship parts. Specifically, after adversarial feature decoupling, the structural feature maps are spatially partitioned into a set of feature-level components, each corresponding to a localized region with distinct geometric semantics. These components are treated as nodes in a graph and jointly modeled using a graph attention network to capture inter-component relationships and enforce structural consistency. Component-level missingness arises when the structural features associated with certain components become unreliable or unobservable due to occlusion or viewpoint variations. To address this issue, the proposed temporal variational Bayesian inference mechanism leverages both inter-component dependencies and multi-frame temporal context to robustly complete missing components.

3.1 Adversarial Learning-Driven Dual-Stream Feature Decoupling and Sequential Collaborative Modeling

Existing ship re-identification frameworks commonly encode appearance and structural attributes jointly within a single-stream architecture, which makes stable structural representations vulnerable to interference from dynamic appearance noise [14]. Consequently, it diminishes the discriminative strength of geometric constraints and degrades the robustness of downstream completion and recognition stages. To fundamentally address this issue, this paper introduces an adversarial learning mechanism that explicitly separates feature streams via dual discriminators. As shown in Fig. 3, after preliminary feature extraction from the input ship image sequence via ResNet at time t, high-level feature f_t is obtained. To explicitly separate appearance and structural information, we construct an adversarial learning-driven dual-stream decoupling network. Its core mechanism employs adversarial learning to enforce the generation of disentangled and complementary feature representations across the dual branches. The network comprises an appearance decoder E_{app} and a structural encoder E_{struct}, generating appearance features F_t^{app} and structural features F_t^{str}, respectively:

$$F_t^{app} = E_{app}(f_t) \qquad F_t^{str} = E_{struct}(f_t) \tag{1}$$

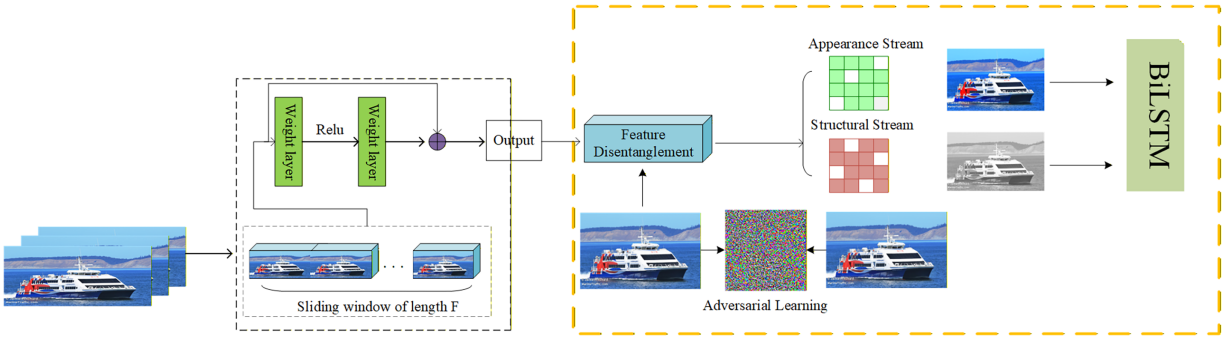

Figure 3: Adversarial learning-driven dual-stream feature decoupling and temporal co-modeling architecture.

To explicitly enforce the decoupling of appearance and structural features, an adversarial learning mechanism is adopted, consisting of two discriminators and a shared encoder–splitter. Specifically, the appearance discriminator D_{app} is designed to distinguish appearance features from structural ones. It is trained to assign high confidence to true appearance features F_t^{app} while rejecting structural features F_t^{str}. Accordingly, its objective $L_{D_{app}}$ is defined as:

$$L_{D_{app}} = E\left[\log D_{app}\left(F_t^{app}\right)\right] + E\left[\log\left(1 - D_{app}\left(F_t^{str}\right)\right)\right] \tag{2}$$

where E denotes averaging, D_{app} denotes the appearance discriminator, D_{struct} denotes the structure discriminator, F_t^{app} denotes the dynamically generated ship appearance features based on the adversarial loss function, and F_t^{str} denotes the dynamically generated ship structural features based on the adversarial loss function.

Similarly, the structure discriminator D_{struct} focuses on identifying geometrically consistent structural representations. It encourages high responses for structural features and suppresses appearance-related information, with the loss $L_{D_{struct}}$ formulated as:

$$L_{D_{struct}} = E\left[\log D_{struct}\left(F_t^{str}\right)\right] + E\left[\log\left(1 - D_{struct}\left(F_t^{app}\right)\right)\right] \tag{3}$$

While the discriminators aim to correctly classify the decoupled features, the encoder–splitter is optimized adversarially to confuse both discriminators. By minimizing the following loss L_{Enc}, the encoder is encouraged to remove domain-specific cues that can be exploited by the opposite discriminator:

$$L_{Enc} = E\left[\log\left(1 - D_{app}\left(F_t^{qpp}\right)\right)\right] + E\left[\log\left(1 - D_{struct}\left(F_t^{str}\right)\right)\right] \tag{4}$$

Through this adversarial process, appearance features are enforced to be invariant to structural information, while structural features are guided to suppress appearance-related noise, resulting in purer and more complementary representations. Overall, the adversarial learning is formulated as a standard minimax optimization problem:

$$\min_{Enc} L_{Enc}, \max_{D_{app}, D_{struct}} L_{D_{app}} + L_{D_{struct}} \tag{5}$$

which is implemented in practice via an alternating optimization strategy, where the discriminators and the encoder–splitter are updated iteratively until convergence.

To capture the motion and deformation patterns of a ship across consecutive frames, the decoupled appearance and structural features are input into independent bidirectional LSTM (BiLSTM) for temporal encoding, as shown in Fig. 4.

During ship navigation, components (e.g., bow, hull, stern) sequentially appear, become occluded, or deform due to relative motion with the camera. This paper employs a BiLSTM to perform temporal modeling on the decoupled appearance and structural feature sequences. Specifically, the BiLSTM operates in a component-wise manner. For each structural component, we construct a temporal feature sequence across consecutive frames. The forward LSTM processes the sequence in chronological order, such that the hidden state at time t depends on the hidden state at time $t - 1$. Conversely, the backward LSTM processes the sequence in reverse order, so the hidden state at time t depends on the hidden state at time $t + 1$, thereby incorporating future context. The forward and backward hidden states are concatenated at each time step to form the final temporal representation of the component. Importantly, the temporal modeling is performed independently for each component, i.e., each component constitutes its own sequence across frames. This component-wise BiLSTM is applied separately to both the appearance and structural streams.

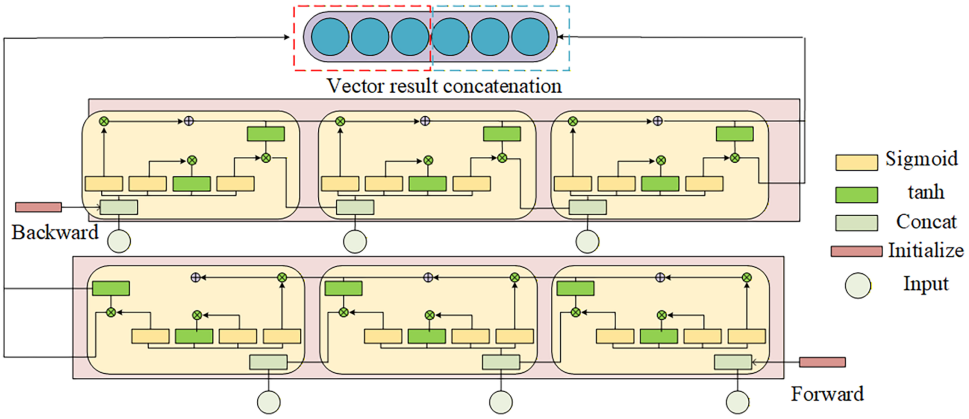

Figure 4: Temporal encoding and forward-backward information fusion based on BiLSTM.

Appearance Stream:

$$\text{Forward calculation: } h_{app,t}^{forward} = LSTM_{app}^{forward}\left(F_{app,t}, h_{app,t-1}^{forward}\right) \tag{6}$$

$$\text{Backward calculation: } h_{app,t}^{backward} = LSTM_{app}^{backward}\left(F_{app,t}, h_{app,t+1}^{backward}\right) \tag{7}$$

where: $h_{app,t}^{forward}$ denotes the appearance vector output by the forward model at time t; $h_{app,t-1}^{forward}$ denotes the appearance vector output by the forward model at time $t-1$; $LSTM_{app}^{forward}$ denotes the forward model for appearance features; $F_{app,t}$ denotes the appearance features input to the bidirectional long-short term memory network of the appearance stream at time t; $h_{app,t+1}^{backward}$ denotes the appearance vector output by the backward model at time $t+1$; $h_{app,t}^{backward}$ denotes the appearance vector output by the backward model at time t; $LSTM_{app}^{backward}$ denotes the backward model for appearance features;

Structure Stream:

$$\text{Forward calculation: } h_{struct,t}^{forward} = LSTM_{struct}^{forward}\left(F_{struct,t}, h_{struct,t-1}^{forward}\right) \tag{8}$$

$$\text{Backward calculation: } h_{struct,t}^{forward} = LSTM_{struct}^{backward}\left(F_{struct,t}, h_{struct,t+1}^{forward}\right) \tag{9}$$

where $h_{struct,t}^{forward}$ denotes the structural vector output by the forward model at time t, $h_{struct,t-1}^{forward}$ denotes the structural vector output by the forward model at time $t-1$, $LSTM_{struct}^{forward}$ denotes the forward model for structural features, $F_{struct,t}$ denotes the structural features input to the bidirectional long-short term memory network of the structural flow at time t, $h_{struct,t+1}^{backward}$ denotes the structural vector output by the backward model at time $t+1$, $h_{struct,t}^{backward}$ denotes the structural vector output by the backward model at time, and $LSTM_{struct}^{backward}$ denotes the backward model for structural features;

Finally, the forward and backward hidden states at each time step are concatenated to form a feature representation that captures the complete temporal context. This representation facilitates inference of the state of temporarily occluded components and enhances robustness single-frame incompleteness.

3.2 Feature Completion for Missing Components via Temporal Variational Bayesian Inference

To address geometric semantic discontinuities arising from structural feature gaps in scenarios with severe viewpoint deficiencies (e.g., continuous porthole absences or localized hull fractures), this paper proposes a Bayesian component completion mechanism based on variational inference. Its core innovation

lies in modeling component visibility as a latent variable, enabling generative reconstruction of missing features through multi-stage collaborative inference. As illustrated in Fig. 5 the module takes dual-stream temporal features as input and progressively completes missing parts through three core submodules: First, the visibility probability prediction module preliminarily determines the visibility status of each component based on part-level feature fusion and classification. Second, the spatial prior construction module employs a graph attention network to model spatial dependencies and symmetry constraints among visible parts, thereby establishing structured geometric priors. Finally, the temporal variational inference module generates probabilistic feature distributions for missing components via variational Bayesian inference. This process integrated outputs from the preceding modules with a multi-frame temporal context extracted by BiLSTM. The entire reconstruction process jointly optimizes reconstruction loss and KL divergence, ensuring balanced spatial plausibility and temporal consistency. Approach significantly enhances feature integrity and recognition robustness under component-missing scenarios.

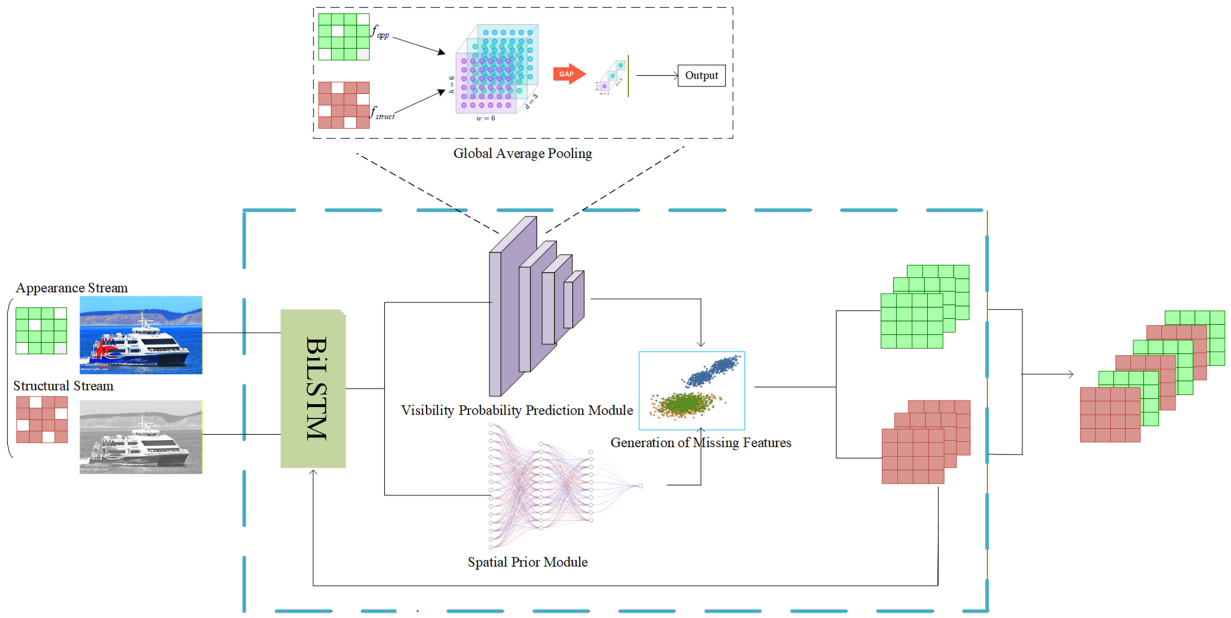

Figure 5: Feature completion for missing components based on sequential variational bayesian inference.

3.2.1 Visibility Probability Prediction Module

To estimate the visibility status of key ship components, we introduce a visibility probability prediction module. The module takes the disentangled appearance and structural features as input, allowing the assessment to benefit from both semantic cues (e.g., color, texture) and geometric patterns (e.g., contours). To integrate these modalities, we first perform local fusion and pooling on the temporal features of each component. Specifically, the appearance and structural features are concatenated along the channel dimension, followed by average pooling to generate a unified fusion vector:

$$u_{t,i} = Pool\left(h_{t,i}^{app}\right) \oplus Pool\left(h_{t,i}^{str}\right) \tag{10}$$

where $u_{t,i}$ denotes the fusion vector for frame t and component i, and $Pool\,()$ represents the pooling operation that averages the temporal vector along the channel dimension to obtain a fixed-length vector for parallel batch processing. $h_{t,i}^{app}$ is the temporal feature vector for frame t and component i in the appearance

stream, with $\oplus$ indicating the concatenation operation. $h_{t,i}^{str}$ is the temporal feature vector for frame t and component i in the structure stream.

The fused feature vector is subsequently processed through two convolutional layers to extract higher-level semantic information. Batch normalization and ReLU activation follow each convolution layer to enhance feature discriminability and suppress noise interference. Subsequently, global average pooling is applied to the encoded features to obtain a compact vector representation that balances appearance and structural information. Features for all components are concatenated and passed through a two-layer fully connected network. Finally, a Sigmoid activation function outputs the visibility probability value for the component: when the probability falls below a predefined threshold, the component is considered missing; otherwise, it is considered visible. The visibility probability is obtained using the Sigmoid activation:

$$\pi_{t,i}^{\Phi} = Sigmoid\left(W_{\Phi} u_{t,i} + b_{\Phi}\right) \tag{11}$$

where if some features are missing in a component's feature vector, that component is treated as missing, $\pi_{t,i}^{\Phi} \in [0,1]$ denotes the t frame and the i component's visibility probability. Following the probabilistic interpretation of the Sigmoid output, we use a fixed threshold of 0.5 to determine visibility: a component is regarded as missing when $\pi_{t,i}^{\Phi} < 0.5$, and visible otherwise. Φ is the parameter set of the variational inference network, W_{Φ} is the weight matrix of the variational inference network, $Sigmoid$ denotes the activation function, b_{Φ} is the bias vector of the variational inference network.

3.2.2 Spatial Prior Module

Following the prediction of component visibility probabilities, a spatial prior module is introduced to exploit the intrinsic geometric characteristics of the ship. The core innovation of this module lies in explicitly modeling the spatial relationships among components as a graph structure. By integrating spatial proximity and hull symmetry priors, it imposes geometric constraints for the subsequent completion of missing components.

Given the distinct spatial layouts inherent to ship structures, visible components are represented as nodes within a graph. The module first maps all visible components (such as bow, stern, port side, starboard side, etc.) to graph nodes. Each node corresponds to the appearance and structural characteristics of its associated component. Edge connections between nodes are established based on normalized image coordinates, with edge weights designed to account simultaneously for spatial proximity and structural symmetry:

$$w_{ij} = \exp\left(-\frac{\|p_i - p_j\|^2}{\sigma^2}\right) \cdot \left(1 + \alpha \cdot symm\left(i, j\right)\right) \tag{12}$$

where w_{ij} denotes the edge weight between nodes i and j, reflecting prior preferences for positional adjacency and symmetry. p_i and p_j denote the normalized coordinates of nodes in the image. $\|p_i - p_j\|^2$ is the squared Euclidean distance. σ is the distance scale parameter controlling decay rate. α is the symmetry weight controlling symmetry influence. $symm\left(i, j\right) \in [0,1]$ is the symmetry metric [15], where higher values indicate greater symmetry of component $iandj$.

Building upon the constructed graph structure, this paper proposes a domain information aggregation mechanism based on conditional probability. The detailed procedure is summarized in Algorithm 1. This process comprises two stages:

Algorithm 1: Spatial prior generation with graph attention and probabilistic modeling

1: Input: Appearance feature vectors H_{app,i_1} and structure feature vectors H_{struct,i_1}
2: Output: Prior distribution parameters $(\mu_{i_1}^{\theta}, \sigma_{i_1}^{\theta})$
3: Definition $\boldsymbol{\beta}$ (balance hyperparameter), σ (distance scale parameter), α (symmetry weight)
4: Initialization: w^T (weight vector), $W_a, W_g, W_\mu, W_\sigma$ (weight matrix), b_μ, b_σ (bias term)
5: If the spatial prior module is used **then**
6: Compute initial node features H_{joint,i_1} through attention fusion using Eqs. (13) and (14)
7: Compute edge weights w_{ij} using spatial proximity and symmetry via Eq. (12)
8: Compute attention scores e_{ij} for node i and its neighbors j using Eq. (15)
9: Normalize attention scores using softmax via Eq. (16) to obtain attention coefficients a_{ij}
10: Perform weighted aggregation via Eq. (17) to obtain updated node features $H_i^{'}$
11: Apply global pooling to augmented node features $\left\{H_1^{'}, H_2^{'}, \ldots, H_N^{'}\right\}$ to generate graph features H_g
12: Input H_g into two-layer MLPcalculating mean and variance of prior distribution via Eq. (18)
13: **return** $\mu_{i_i}^{\theta}, \sigma_{i_i}^{\theta}$
14: end if

In the first stage, preliminary fusion of each component's features is performed. The appearance H_{app,i_1} and structural features H_{struct,i_1} are dynamically weighted through an attention mechanism, and the joint feature representation is obtained via weighted fusion, as defined in Eqs. (13) and (14):

$$\gamma = Sigmoid\left(w^T\left[H_{app,i_1}; H_{struct,i_1}\right]\right) \tag{13}$$

$$H_{joint,i_1} = \gamma \cdot H_{app,i_1} + (1-\gamma) \cdot H_{struct,i_1} \tag{14}$$

where γ represents the dynamic weight, w^T denotes the transpose of the weight vector, H_{app,i_1} denotes the i_1 appearance feature vector of the th visible component, H_{struct,i_1} denotes the i_1 structural feature vector of the th visible component;

In the second stage, refined neighborhood information aggregation is conducted using a multi-head graph attention network (GAT). For each node, the attention scores between the node and its neighbors are computed using a learnable attention function, as formulated in, Eq. (15) and subsequently normalized via the softmax operation to obtain attention coefficients, as shown in Eq. (16):

$$e_{ij} = LeakyReLU\left(W_a\left[H_{app,i_1} \| H_{struct,i_1}\right]\right) + \beta \cdot w_{ij} \tag{15}$$

$$a_{ij} = \frac{\exp\left(e_{ij}\right)}{\sum k \in N\left(i\right) \exp\left(e_{ik}\right)} \tag{16}$$

where *Leaky*ReLU is a nonlinear activation function, W_a is the attention weight matrix, $\|$ denotes vector concatenation, β is the balance hyperparameter, and w_{ij} is the precomputed spatial edge weight. The attention scores are then normalized using the softmax function: a_{ij}. The normalized attention coefficients represent the importance weight of node j relative to node i, where $\mathcal{N}(i)$ denotes the set of neighbors for node i.

Finally, the updated node features are obtained through weighted aggregation of neighboring joint features followed by a nonlinear activation function, as expressed in Eq. (17):

$$H_i^{'} = \sigma\left(\sum_{j \in N(i)} a_{ij} W_g H_{joint,j}\right) \tag{17}$$

where: the updated feature of node i after GAT aggregation, W_g is the GAT transformation weight matrix, σ represents the linear activation function, introducing a nonlinear transformation.

To extract graph-level structural information from node-level representations, the enhanced node features $\{H_1^{'}, H_2^{'}, \ldots, H_N^{'}\}$ are used for subsequent global pooling operations, generating graph features H_g. The vector is then input into a two-layer multi-layer perceptron with dual output heads. These heads respectively predict the mean and diagonal covariance, which serve as the parameters for the prior Gaussian distribution.

$$\mu_{i_1}^{\theta} = W_{\mu} \cdot H_g + b_{\mu},\ \sigma_{i_1}^{\theta} = \exp\left(W_{\sigma} \cdot H_g + b_{\sigma}\right) \tag{18}$$

where W_{μ} and W_{σ} are weights, b_{μ} and b_{σ} are bias terms.

3.2.3 Temporal Variational Inference and Optimization Module

To address reconstruction bias caused by motion blur and occlusion transitions across consecutive frames, this paper introduces a temporal variational inference mechanism. This mechanism probabilistically reconstructs missing component features by integrating spatial prior information with multi-frame temporal cues. This module uses the spatial prior distribution parameters generated in the previous stage ($\mu_{i_1}^{\theta}$, $\sigma_{i_1}^{\theta}$) as geometric constraints for variational inference. Meanwhile, the enhanced multi-frame features—refined through component-level visibility prediction and the graph attention network—are fed into a bidirectional LSTM (BiLSTM) for temporal modeling. The concatenated bidirectional state is processed through two layers of fully connected variational networks to generate the mean and covariance parameters of the posterior Gaussian distribution, yielding the feature distribution of the missing component under temporal conditions. Although the reconstruction likelihood is evaluated only over observed components, missing components are recovered implicitly by sampling from the inferred posterior conditioned on spatial and temporal context. The overall training objective consists of a reconstruction term and a KL divergence term: the former encourages the posterior to explain the observed features, while the latter regularizes the posterior toward the spatial prior. This formulation enables structurally consistent and temporally coherent completion of missing components under uncertainty. The objective function used for model fitting is defined as: L_{ELBO}:

$$L_{ELBO} = -\sum_{i_1=1}^{N} \frac{||F_{obs,i_1} - \mu_{i_1}^{\theta}||^2}{2\left(\sigma_{i_1}^{\theta}\right)^2} - KL\left(q_{\Phi}\left(v\right)||p\left(v\right)\right) \tag{19}$$

where $-\sum_{i1=1}^{N} \frac{||F_{obs,i_1} - \mu_{i_1}^{\theta}||^2}{2\left(\sigma_{i_1}^{\theta}\right)^2}$ is the reconstruction likelihood term, F_{obs,i_1} represents the observed feature vector of the i_1-th visible component, N denotes the total number of visible components, F_{obs,i_1} constructed by concatenating the structural feature vectors and appearance feature vectors of the i_1 visible components, $\mu_{i_1}^{\theta}$ represents the mean of the θ visible components under the parameters i_1 of the variational inference network, $\sigma_{i_1}^{\theta}$ represents the variance of the θ; $KL\left(q_{\Phi}\left(v\right)||p\left(v\right)\right)$ indicates the difference between the variational $q_{\Phi}\left(v\right)$ and the prior distribution $p\left(v\right)$, $KL\left(q_{\Phi}\left(v\right)||p\left(v\right)\right)$ is calculated through the divergence of KL [16], v represents the set of visibility states for all components $v = \{v_1, v_2, \ldots, v_n\}$.

3.2.4 Loss Function

To jointly optimize ship identity classification and feature discrimination capabilities, this paper adopts a dynamically weighted multi-task loss function. The model learns different content based on distinct training phases. During the initial training phase, the model focuses on rapidly establishing fundamental

identity discrimination, thus prioritizing classification loss optimization. In the subsequent training phase, once classification capabilities have stabilized, the model emphasizes metric learning to enlarge inter-class separability and reduce intra-class variance, increasing the relative contribution of the metric learning loss [17].

$$L_{total} = \lambda(t) L_{ce} + (1 - \lambda(t)) L_{tri} \tag{20}$$

$$\lambda(t) = \max\left(0, 1 - \frac{t}{T_{\text{attenuation}}}\right) \tag{21}$$

where $\lambda(t)$ represents the dynamic weight factor, t denotes current training epoch, $T_{\text{attenuation}}$ is a predefined decay cycle hyperparameter, L_{ce} represents the cross-entropy loss function [17], L_{tri} denotes the triplet loss function [17].

4 Experiments and Analysis

This section describes the datasets and implementation details employed in the experiments and systematically evaluates the robustness and accuracy of the proposed network through comparisons with state-of-the-art methods and comprehensive ablation studies.

4.1 Experimental Settings

4.1.1 Datasets

Ship re-identification (Ship Re-ID), a critical subfield of computer vision, has attracted growing attention in recent years. However, most existing ship datasets are limited in scale, cover a narrow range of categories, and often lack designs specifically tailored for re-identification tasks. In real maritime surveillance, ships frequently exhibit "incomplete observations" (partially visible or locally missing) due to variations in viewpoint, distance, and occlusion. When a dataset contains only "complete, frontal, unobstructed" images, model performance tends to degrade significantly after deployment. To address these limitations, the Ship-CH dataset adopts a collection and annotation strategy that deliberately incorporates incomplete scenarios and explicitly models occlusion during training. Ship-CH consists of 163 ship identities and forms a large-scale dataset for ship re-identification. The dataset statistics are presented in Table 1. Compared with existing ship datasets, Ship-CH provides the following key advantages:

1. **Component-Level Incompleteness Coverage:** The images naturally contain occlusions caused by masts, shorelines, and onboard equipment, as well as boundary cropping. Each image is vertically divided into six components, and a "part-aware random erasing" strategy is applied during training to simulate local missing regions, enhancing robustness to partial invisibility.
2. **Real-World Scenarios:** All images are captured in real maritime environments, covering diverse lighting conditions, weather states, sea-surface conditions, and camera viewpoints, better reflecting the complexity of practical applications.
3. **Strict Identity Isolation:** Training, validation, and test identities are mutually exclusive to prevent data leakage. This forces the model to learn generalizable discriminative features rather than memorizing appearance details.

Table 1: Introduction to the Ship-CH dataset.

	Ship Identity	Number of Images
Training Set	97	1652
Query Set	34	237
Validation set	32	573
Image Dataset	34	381

4.1.2 Evaluation Metrics

This paper employs two evaluation metrics: Mean Average Precision (mAP) and Cumulative Matching Characteristic (CMC) to assess model performance. CMC@k denotes the probability of finding the correct result within the first k query results.

4.2 Experimental Details

Because the proposed method is sequence-based, we construct temporal inputs from consecutive frames within the same ship track. Specifically, frames are first ordered chronologically and then grouped into short sequences using a sliding-window strategy. Each window contains K (K = 4) consecutive frames, and overlapping windows are generated along each track to increase the number and diversity of training samples. These sequences are used as inputs to the BiLSTM and the temporal variational inference modules. All experiments were conducted within the PyTorch framework. During training, all input images were resized to 128 × 256 (height × width). The training batch size was set to 64, following a PK sampling strategy (P = 16 identities, K = 4 images per identity) to ensure sufficient positive and negative sample pairs within each batch. To mitigate overfitting arising from limited dataset diversity, we employed multiple data augmentation techniques: random horizontal flipping (probability 0.5), color jittering (jitter range of 0.2 for brightness, contrast, saturation, and hue), and random erasure (probability 0.5, erasure area ratio 0.02–0.4). We adopted a staged training strategy over a total of 30 epochs. During the first five epochs, we applied a linear learning-rate warm-up from 0 to the initial rate. Over the subsequent 15 epochs, the top two residual blocks of the backbone network were frozen to stabilize training. In the final 10 epochs, all parameters were unfrozen for end-to-end fine-tuning. The initial learning rate was set to 3×10^{-4}. while the backbone network used a smaller learning rate of 1×10^{-4} to maintain the stability of pre-trained features. The Adam optimizer was used for model optimization, with a weight decay of 1×10^{-4} and a cosine annealing schedule for learning rate adjustment. For the loss function, we employed a multi-loss combination, including the triplet loss for metric learning (weight $\lambda_2 = 1.0$, using a batch-hard mining strategy with margin = 0.3) and the KL divergence loss in the variational inference module (weight $\lambda_3 = 0.1$). All experiments were evaluated using Rank-1, Rank-5, and Rank-10 accuracy, as well as mean average precision (mAP), with results averaged over three independent runs to ensure statistical reliability and experimental reproducibility.

4.3 Comparison with Other Network Models

In this section, we compare our method and the proposed Ship-CH dataset with existing Re-ID approaches. The results are presented in Table 2.

Table 2: Performance comparison of existing Re-ID approaches.

Network Model	mAP	CMC@1	CMC@5	CMC@10
CoDA-Net	62.71%	75.53%	91.56%	94.89%
MCL	61.13%	75.95%	93.67%	95.78%
FGFN	56.02%	64.82%	77.78%	90.74%
VesselNet	62.67%	79.75%	94.94%	97.89%
GLF-MVFL	61.23%	72.57%	92.41%	97.89%
Proposed Method	85.67%	93.67%	98.73%	98.73%

As shown in Table 2, although CoDA-Net [18], MCL [19], FGFN [20], VesselNet [21], and GLF-MVFL [22] achieve competitive performance in conventional ship re-identification tasks, their effectiveness degrades significantly under occlusion and component-missing scenarios. For a fair comparison, we re-trained all baselines on the Ship-CH dataset using identical experimental settings, including the same ResNet-50 backbone, input resolution (256 × 128), data augmentation, optimizer, and training schedule. CoDA-Net employs collaborative attention to enhance feature representations. However, it does not explicitly model temporal dependencies and thus cannot exploit consecutive frames to infer missing information under dynamic occlusions, resulting in 62.71% mAP. MCL performs multi-level feature fusion through contrastive learning, yet its objective primarily targets static feature similarity and does not facilitate missing-component inference under partial observations, achieving 61.13% mAP. FGFN emphasizes fine-grained local feature extraction, but it lacks global contextual reasoning and temporal cues, which limit its ability to compensate for occluded components in complex maritime environments (56.02% mAP). VesselNet relies on a conventional CNN architecture without an occlusion-aware design, yielding 62.67% mAP. GLF-MVFL incorporates multi-view learning, but its view-fusion strategy is largely static and insufficient for reconstructing missing structural information, achieving 61.23% mAP. It is worth noting that these baseline methods operate in a single-frame manner (K = 1) and therefore cannot leverage across-frame temporal dependencies. In contrast, the proposed framework is explicitly sequence-based and processes short temporal clips (K = 4) to capture appearance continuity and structural evolution over time.

To further evaluate robustness to incomplete observations, we report an occlusion-level breakdown of performance on Ship-CH, as summarized in Table 3. Specifically, the original test setting (“Mixed”) contains samples with light, medium, and heavy occlusions, while we additionally evaluate the same trained models on three mutually exclusive subsets grouped by occlusion severity (Light/Medium/Heavy). As expected, mAP consistently decreases as occlusion becomes more severe due to reduced visible cues and increased ambiguity. Nevertheless, the proposed method maintains clear advantages across all occlusion levels, demonstrating strong robustness under partial visibility.

To validate the effectiveness of each branch and module in the proposed network model, we conducted multiple ablation experiments under different settings on the Ship-CH dataset, evaluating the contribution of individual branches and modules. The results are summarized in Table 4. Among these, Baseline serves as the reference, utilizing only the re-identification branch based on ResNet50 for feature extraction. +Dual-Stream denotes the addition of a dual-stream decoupling branch to the Baseline, employing adversarial learning to decompose ship features into appearance and structure streams, thereby addressing ship appearance variation issues. +BiLSTM-Temporal indicates that, on top of the dual-stream decoupling, a temporal encoding branch is added and integrated with BiLSTM to further extract the ship’s temporal feature information. +GAT-Prior adds a spatial prior branch to the temporal encoding branch, incorporating a

graph attention network to model spatial relationships among ship components. +Variational introduces a variational inference branch to the spatial prior branch, using a variational encoder to learn the posterior distribution of missing features. Finally, Full-Model represents the complete model, including the dynamic weight fusion module.

Table 3: mAP (%) comparison under different occlusion levels on Ship-CH.

Network Model	Mixed	Light	Medium	Heavy
CoDA-Net	62.71%	50.58%	45.01%	39.18%
MCL	61.13%	46.91%	41.36%	33.98%
FGFN	56.02%	47.59%	41.75%	34.44%
VesselNet	62.67%	43.91%	39.41%	31.98%
GLF-MVFL	61.23%	46.83%	40.30%	32.34%
Proposed Method	85.67%	72.70%	70.49%	67.42%

Table 4: Performance comparison of branches and modules.

Module	mAP	CMC@1	CMC@5	CMC@10
Only Baseline	37.84%	60.76%	88.61%	94.09%
+Adversarial	32.13%	59.92%	84.39%	91.98%
+Dual-Stream	33.37%	58.65%	83.12%	90.30%
+BiLSTM-Temporal	69.11%	77.09%	92.41%	94.94%
+GAT Prior	63.95%	65.82%	94.94%	98.73%
+Variational	74.74%	79.75%	98.47%	98.73%
Full-Model	85.67%	93.67%	98.73%	98.73%

As shown in Table 4, the Baseline-only model relies on a single CNN branch for feature extraction, making it vulnerable to occlusion, feature loss, and temporal inconsistency. Consequently, its performance is limited (mAP 37.84%, CMC@1 60.76%). When adversarial learning is introduced, performance drops temporarily (mAP 32.13%). This trend is expected because adversarial optimization imposes strong disentanglement constraints, which increase optimization difficulty and can reduce discriminability when sufficient temporal or structural supervision is not yet available. Adding dual-stream feature decoupling yields a modest improvement (mAP 33.37%), indicating that appearance–structure separation alone is insufficient to enhance Re-ID performance without temporal modeling. At this stage, the structural stream remains underexploited due to limited complementary supervision. A substantial gain is achieved after incorporating BiLSTM-based temporal modeling, with mAP rising to 69.11% and CMC@1 to 77.22%. Temporal aggregation compensates for information loss caused by occlusion and viewpoint changes, allowing the decoupled streams to fully exploit their complementarity. Further introducing GAT-based spatial priors slightly reduces mAP to 63.95% while significantly improving high-rank accuracy (CMC@5 and CMC@10). This pattern suggests a regularization trade-off: the GAT-enhanced spatial constraints promote structural consistency and robustness under partial occlusion, but may suppress highly discriminative yet noisy local cues. After introducing variational Bayesian feature completion, performance increases to 75.09% mAP and 82.28% CMC@1. By jointly modeling temporal context and uncertainty through KL regularization and reconstruction constraints, the model can probabilistically complete missing components, substantially

improving robustness. Finally, integrating BiLSTM temporal modeling, GAT spatial priors, and variational inference yields the best overall results, achieving 85.67% mAP and 93.67% CMC@1. The evolution of different evaluation metrics during training is illustrated in Fig. 6.

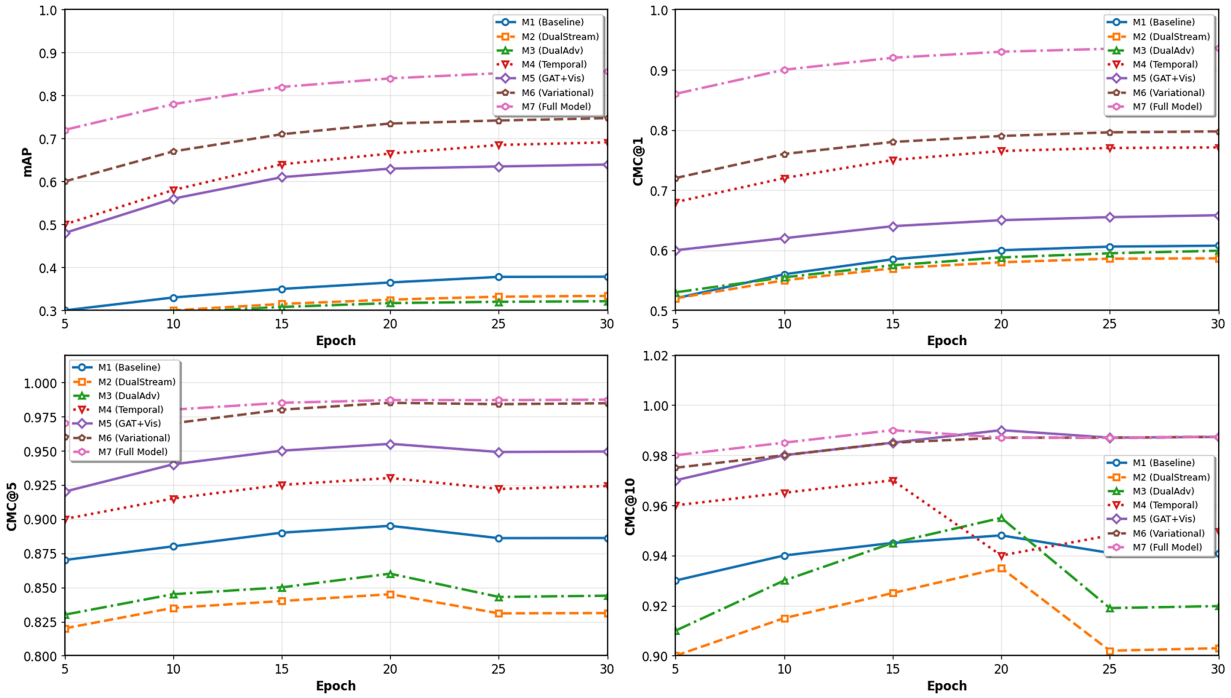

Figure 6: Evolution of metrics during training.

To further qualitatively evaluate the effectiveness of the proposed method, Fig. 7 visualizes the top-10 retrieval results of the baseline model and our method. Compared with the baseline, the proposed method retrieves more correct ship identities under occlusion and viewpoint variations, demonstrating stronger discriminative capability and robustness.

Figure 7: Visualization of top-10 results.

Although the proposed framework demonstrates strong robustness and discriminative capability under partial visibility and occlusion, several limitations should be acknowledged. First, the overall model complexity is relatively high. The framework integrates multiple components, including dual-stream feature decoupling, adversarial learning, bidirectional temporal modeling, graph-based spatial priors, and variational Bayesian inference, which jointly improve recognition accuracy and robustness but inevitably increase computational cost and inference latency. Second, the reliance on temporal sequences implies that the

method requires consecutive frames to fully realize its benefits; when only sparse observations or single frames are available, the advantage of temporal modeling may be limited. In future work, we will investigate simplification and acceleration strategies, such as adopting lightweight backbone architectures, sharing parameters across streams, and developing more efficient temporal aggregation mechanisms.

5 Conclusions

This paper addresses the challenge of ship re-identification under partially occluded and incomplete observation scenarios by proposing an innovative framework based on dual-stream feature decoupling and temporal variational Bayesian inference. The core of the framework lies in explicitly separating a ship's appearance and structural features through an adversarial learning mechanism. This separation reduces and mitigates the interference of dynamic surface noise and enhances stable geometric representations. Building upon this, we design a temporal variational Bayesian completion mechanism, which integrates component-level spatial priors—modeled by a graph attention network—with multi-frame temporal context captured by a BiLSTM, generating reliable missing component features through probabilistic inference. Experimental results demonstrate that the proposed method significantly outperforms existing mainstream approaches on occlusion-prone ship datasets, validating its robustness and superior performance in complex maritime surveillance environments.

Although this study achieved the expected results, certain limitations remain. Future research will focus on the following aspects:

(1) Domain Adaptation and Generalization Enhancement: We will investigate unsupervised or weakly supervised domain adaptation to bridge the distributional gap between synthetic data and complex real-world scenarios. For example, we plan to apply adversarial learning or feature-alignment strategies on unlabeled maritime videos to improve robustness to previously unseen occlusion patterns and environmental conditions.

(2) Model Lightweighting and Efficiency Optimization: We will pursue efficiency-oriented strategies such as network pruning, quantization, and knowledge distillation to reduce computational overhead. In parallel, we will investigate more efficient feature-fusion and sequence-modeling design as a step toward real-time ship re-identification.

(3) Multimodal Data Fusion: We will explore the integrating heterogeneous sensing modalities, such as radar/AIS information and infrared imagery, to develop multimodal fusion models. This direction aims to maintain stable recognition performance under severely degraded optical conditions, including dense fog and low-light or nighttime scenarios.

Through these enhancements, we aim to accelerate the practical deployment of this technology in critical maritime applications including regulatory surveillance, search and rescue operations, and intelligent shipping.

Acknowledgement: We are grateful to Nanjing University of Information Science and Technology and Nanjing Tech University for providing study environment and computing equipment.

Funding Statement : This study was supported, in part, by the National Nature Science Foundation of China under Grants 62272236, 62376128; in part, by the Natural Science Foundation of Jiangsu Province under Grants BK20201136, BK20191401.

Author Contributions: Study conception and design: Wanhui Qiao, Xiaorui Zhang; data collection: Kaibo Wang, Shiyu Zhou; analysis and interpretation of results: Wanhui Qiao, Xiaorui Zhang, Wei Sun; draft manuscript preparation: Wanhui Qiao, Wei Sun, Xiaorui Zhang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics Approval: This paper does not contain any studies with human participants performed by any of the authors.

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

References

1. Qiao D, Liu G, Lv T, Li W, Zhang J. Marine vision-based situational awareness using discriminative deep learning: a survey. J Mar Sci Eng. 2021;9(4):397. doi:10.3390/jmse9040397.
2. Sun W, Guan F, Zhang X, Shen X, Wang K. Ship re-identification in foggy weather: a two-branch network with dynamic feature enhancement and dual attention. Eng Appl Artif Intell. 2025;143(3):109974. doi:10.1016/j.engappai.2024.109974.
3. Wolrige SH, Howe D, Majidiyan H. Intelligent computerized video analysis for automated data extraction in wave structure interaction: a wave basin case study. J Mar Sci Eng. 2025;13(3):617. doi:10.3390/jmse13030617.
4. Ge X, Li X, Zhang C, Li J, Gao Y. Robust and real-time ship object detection method based on enhanced CNN. IEEE Access. 2024;12(22):112196–210. doi:10.1109/ACCESS.2024.3442776.
5. Yasir M, Liu S, Xu M, Wan J, Pirasteh S, Dang KB. ShipGeoNet: SAR image-based geometric feature extraction of ships using convolutional neural networks. IEEE Trans Geosci Remote Sens. 2024;62:5202613. doi:10.1109/TGRS.2024.3352150.
6. Wang Y, Tian Z, Fu H. Multivariate USV motion prediction method based on a temporal attention weighted TCN-Bi-LSTM model. J Mar Sci Eng. 2024;12(5):711. doi:10.3390/jmse12050711.
7. Zhang D, Zhan J, Tan L, Gao Y, Župan R. Comparison of two deep learning methods for ship target recognition with optical remotely sensed data. Neural Comput Appl. 2021;33(10):4639–49. doi:10.1007/s00521-020-05307-6.
8. Han Y, Yang X, Pu T, Peng Z. Fine-grained recognition for oriented ship against complex scenes in optical remote sensing images. IEEE Trans Geosci Remote Sens. 2022;60:5612318. doi:10.1109/TGRS.2021.3123666.
9. Tian Y, Meng H, Yuan F. FREGNet: ship recognition based on feature representation enhancement and GCN combiner in complex environment. IEEE Trans Intell Transp Syst. 2024;25(11):15641–53. doi:10.1109/TITS.2024.3454016.
10. Qian Y, Barthelemy J, Karuppiah E, Perez P. Identifying re-identification challenges: past, current and future trends. SN Comput Sci. 2024;5(7):937. doi:10.1007/s42979-024-03271-9.
11. Wang W, Zhang X, Sun W, Huang M. A novel method of ship detection under cloud interference for optical remote sensing images. Remote Sens. 2022;14(15):3731. doi:10.3390/rs14153731.
12. Li Y, Wang TQ. Spatial state analysis of ship during berthing and unberthing process utilizing incomplete 3D LiDAR point cloud data. J Mar Sci Eng. 2025;13(2):347. doi:10.3390/jmse13020347.
13. Zeng G, Wang R, Yu W, Lin A, Li H, Shang Y. A transfer learning-based approach to maritime warships re-identification. Eng Appl Artif Intell. 2023;125(4):106696. doi:10.1016/j.engappai.2023.106696.
14. Ma S, Wang W, Pan Z, Hu Y, Zhou G, Wang Q. A recognition model incorporating geometric relationships of ship components. Remote Sens. 2023;16(1):130. doi:10.3390/rs16010130.
15. Guo R, Cui J, Jing G, Zhang S, Xing M. Validating GEV model for reflection symmetry-based ocean ship detection with Gaofen-3 dual-polarimetric data. Remote Sens. 2020;12(7):1148. doi:10.3390/rs12071148.
16. Zhang C, Butepage J, Kjellstrom H, Mandt S. Advances in variational inference. IEEE Trans Pattern Anal Mach Intell. 2019;41(8):2008–26. doi:10.1109/TPAMI.2018.2889774.
17. He J, Wang Y, Liu H. Ship classification in medium-resolution SAR images via densely connected triplet CNNs integrating fisher discrimination regularized metric learning. IEEE Trans Geosci Remote Sens. 2021;59(4):3022–39. doi:10.1109/TGRS.2020.3009284.
18. Roy S, Jana DK, Long N. Re-identifying naval vessels using novel convolutional dynamic alignment networks algorithm. Pol Marit Res. 2024;31(1):64–76. doi:10.2478/pomr-2024-0007.

19. Zhang Q, Zhang M, Liu J, He X, Song R, Zhang W. Unsupervised maritime vessel re-identification with multi-level contrastive learning. IEEE Trans Intell Transp Syst. 2023;24(5):5406–18. doi:10.1109/TITS.2023.3243591.
20. Dou W, Zhu L, Wang Y, Wang S. Research on key technology of ship re-identification based on the USV-UAV collaboration. Drones. 2023;7(9):590. doi:10.3390/drones7090590.
21. Yu Z, Liu J, Zou S, Cao Y. VesselNet: a large-scale dataset and efficient mixed attention network for vessel re-identification. In: Proceedings of the 2023 2nd International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM); 2023 Jul 25–29; Jiuzhaigou, China. p. 437–41. doi:10.1109/MLCCIM60412.2023.00070.
22. Qiao D, Liu G, Dong F, Jiang SX, Dai L. Marine vessel re-identification: a large-scale dataset and global-and-local fusion-based discriminative feature learning. IEEE Access. 2020;8:27744–56. doi:10.1109/ACCESS.2020.2969231.