

Fusion of Ship Perceptual Information for Electronic Navigational Chart and Radar Images based on Deep Learning

Muzhuang Guo¹, Chen Guo¹, Chuang Zhang², Daheng Zhang² and Zongjiang Gao²

¹(College of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026 Liaoning, People's Republic of China)

²(Navigation College, Dalian Maritime University, Dalian 116026, China)
(E-mail: dmuguoc@126.com)

Superimposing Electronic Navigational Chart (ENC) data on marine radar images can enrich information for navigation. However, direct image superposition is affected by the performance of various instruments such as Global Navigation Satellite Systems (GNSS) and compasses and may undermine the effectiveness of the resulting information. We propose a data fusion algorithm based on deep learning to extract robust features from radar images. By deep learning in this context we mean employing a class of machine learning algorithms, including artificial neural networks, that use multiple layers to progressively extract higher level features from raw input. We first exploit the ability of deep learning to perform target detection for the identification of marine radar targets. Then, image processing is performed on the identified targets to determine reference points for consistent data fusion of ENC and marine radar information. Finally, a more intelligent fusion algorithm is built to merge the marine radar and electronic chart data according to the determined reference points. The proposed fusion is verified through simulations using ENC data and marine radar images from real ships in narrow waters over a continuous period. The results suggest a suitable performance for edge matching of the shoreline and real-time applicability. The fused image can provide comprehensive information to support navigation, thus enhancing important aspects such as safety.

KEYWORDS

1. Deep learning. 2. Marine radar. 3. Electronic Navigational Chart (ENC). 4. Data fusion.

Submitted: 16 October 2018. Accepted: 8 May 2019. First published online: 14 June 2019.

1. INTRODUCTION. Modern shipping requires ever-increasing accuracy and reliability from navigation systems. Marine radars offer live real-time information about targets above the water surface, but they cannot offer any comprehensive hydrographic information. The Electronic Navigational Chart (ENC), however, enhances ship navigation by displaying static information including coastlines and fixed/floating aids to navigation. Multi-sensor data fusion allows analysing and synthesising information from multiple

sources according to certain criteria to improve decision-making and control (Hall and Llinas, 2002). Radar echo images are the primary non-visual source of information for ship collision avoidance, whereas ENC provides hydrological information for the waters in which the ship is sailing. These two sources of information are complimentary in providing information to support navigation safety. Therefore, fusing ENC data and marine radar images could facilitate the identification and positioning of moving ships and improve both the automation level and navigation safety (Zhang and Fan, 2011).

Suitable marine radar performance and test standards from the International Electrotechnical Commission (IEC), IEC 62388 (2013), enable overlaying ENC on marine radar images, conforming a chart radar (IEC 62388, 2013: 115), which provides a more intuitive understanding of the ship navigation scenario and quickly illustrates the Global Navigation Satellite System (GNSS) accuracy and proper operation of other navigation instruments. Nevertheless, mismatching occurs between the objects obtained from radar and Electronic Chart Display and Information System (ECDIS) due to transmitting heading signals, shielding of electromagnetic waves, and the extension of echoes in range. Mismatching can be manually compensated if the causes of such mismatching are identified. However, the process of manual compensation is cumbersome and requires experience.

Data fusion for navigation instruments usually addresses aspects such as filtering optimisation and image fusion from different sensors. For instance, Hu et al. (2015) and Gao et al. (2018) explored filtering algorithms for GNSS/Inertial Navigation System (INS) integrated navigation. Likewise, an adaptive data-fusion fuzzy algorithm has been proposed by Al-Sharman et al. (2018) to improve estimation accuracy while an aircraft approaches the landing surface. Radar/Automatic Identification System (AIS) fusion is also a basis for navigational decision support systems, as demonstrated by Borkowski and Zwierzewicz (2011). A target tracking fusion algorithm for data from radar and AIS was proposed by Kazimierski and Stateczny (2015). Du and Gao (2017) proposed a method for image fusion of data from different navigation instruments and multi-focus image fusion using orientation information excitation on a pulse-coupled neural network. Similarly, image registration based on feature points or edges has been intensively explored (Ma et al., 2018).

Extensive research has been conducted on fusion (including superposition) of ENC data and marine radar images. Donderi et al. (2004) conducted simulations on navigation, confirming the suitability of overlaying ENC on radar images for navigation safety. Some empirical research on integrating marine radar images and ENC was conducted by Kazimierski and Stateczny (2015) by superimposing fused AIS and radar signals on ECDIS. A method for evaluating the accuracy of overlaying surveillance marine radar images has been proposed by Lubczonek (2015). Liu et al. (2005) and Yang et al. (2010) thoroughly evaluated the real-time overlapping of ECDIS and superposition of ENC and radar images. Zhang and Fan (2011) proposed automatic image fusion of radar images and ENC data based on the Harris corner to detect image features. These types of methods use different operators to detect features often affected by echo quality and target deformation. The detection results of the features are often not ideal.

High-level fusion of ENC data and marine radar images has been addressed to a limited extent. Moreover, apart from classical algorithms for estimation and recognition, recent approaches such as artificial intelligence, information theory and advanced filtering algorithms are being increasingly applied to multi-sensor data fusion (Zhao et al., 2014). Deep learning has been successfully applied in computer vision for determining high-level abstraction. Specifically, a Convolutional Neural Network (CNN) processes

inputs to maintain important features and exclude irrelevant information (Goyal et al., 2017).

In this study, we propose an algorithm employing deep learning for extracting target features from marine radar images to enable feature-level fusion of ENC data and radar images taking the characteristics of marine radar signals into consideration. We verified the performance of the proposed method and explored solutions for fusion failure due to sensor errors in ship positioning and heading. Section 2 of this paper introduces the proposed target detection method, which is based on the characteristics of marine radar images and details of the proposed fusion algorithm are presented in Section 3. Simulation results to verify the performance of the method are presented in Section 4 and the conclusions of the study are summarised in Section 5.

2. MARINE RADAR IMAGE PROCESSING. The working principles of marine radar mean that clutter pixels in the resulting radar image and irregular expansion and adhesion deformation in the echoes from some objects is inevitable. Other objects easily identified from high-quality echoes are typical targets, which include land, islands, navigation equipment, ships, navigation aids and bridges. These targets reflect electromagnetic waves with less distortion and can be used as suitable references for fusion of ENC data and marine radar images. However, some typical targets such as shoreline echoes, present variations whose shape changes with the tide and their positions are not stable. Likewise, the shapes of radar echoes from targets are affected by aspects such as meteorological conditions and ship draughts. Before fusion, we applied target detection based on deep learning to identify typical targets in marine radar images. These targets then serve as reference points for consistent radar image fusion with ENC data. In the following subsections, details of the proposed marine radar image processing are presented.

2.1. Marine radar data pre-processing. To apply deep learning to marine radar images, we first compiled the images into a dataset. This was then augmented to improve detection. Subsequently, we performed annotation of typical targets.

Before target detection, a typical radar target dataset should be constructed. We selected real-time ship radar images referenced by duty officers or pilots from vessels in different narrow channels as samples for typical target detection. After the vessels docked, we used the Voyage Data Recorder (VDR) monitor to acquire ENC data and marine radar images from the corresponding narrow channel over the previous 12 hours.

We used a data augmentation method to avoid overfitting and increase data availability during training as the obtained dataset was small for deep learning (Wu et al., 2017). This data augmentation approach consists of geometrical transformations, such as horizontal flipping, resizing, rotation and cropping to generate synthetic images.

Our team was supported by ten senior seafarers who manually annotated typical targets from marine radar images using bounding boxes and then labelled their classes. These annotations were verified with reference to the chart, and typical targets were placed at the centre of the bounding boxes.

2.2. Target detection using neural networks. For efficient and robust marine radar typical target detection, we selected two models with strong generalisation capability, namely, Faster Region-based CNN (Faster R-CNN) and Region-based Fully Convolutional Network (R-FCN) (Dai et al., 2016). In addition, we selected the Oxford Visual

Geometry Group (VGG) VGG-16 deep CNN (VGG-16) and the 101-layer residual neural network (ResNet-101) in imageNet Large Scale Visual Recognition Challenge (He et al., 2016; Simonyan and Zisserman, 2015) with fast training and testing, respectively, to be the feature extractors.

2.2.1. *Faster R-CNN*. Faster R-CNN performs detection in two stages. In the first stage, a Region Proposal Network (RPN) accepts a marine radar image as input and extracts its features to indicate the subsequent detection model of the regions it should analyse. We slid a small network over the obtained convolutional feature map of the radar image to generate region proposals. Each feature is mapped into a lower-dimensional feature, which is fed into box-regression and box-classification layers. To train the RPN, we assigned a binary class label determined from the intersection-over-union overlap compared to a threshold with any ground-truth box to each anchor indicating whether it is an object. The loss function for the image is defined as:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i + p_i^*) + \lambda \frac{1}{N} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

where i is the anchor index in a mini-batch, p_i is the probability of anchor i being an object, t_i is a vector representing the four parameterised coordinates of the predicted bounding box, t_i^* is the vector of the ground-truth box associated with a positive anchor, and λ is a balancing weight (Ren et al., 2015). The classification loss is the log loss over two classes as follows:

$$L_{cls}(p_i, p_i^*) = \log[p_i p_i^* + (1 - p_i^*)(1 - p_i)] \quad (2)$$

and the regression loss is given by:

$$L_{reg}(t_i, t_i^*) = \begin{cases} 0.5(t_i - t_i^*)^2 & \text{if } |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5 & \text{otherwise} \end{cases} \quad (3)$$

For bounding box regression, we adopted the parameterisation of the four coordinates as follows:

$$\begin{aligned} t_x &= (x_p - x_a)/w_a, t_y = (y_p - y_a)/h_a \\ t_w &= \log(w_p/w_a), \quad t_h = \log(h_p/h_a) \\ t_x^* &= (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/h_a \\ t_w^* &= \log(w^*/w_a), \quad t_h^* = \log(h^*/h_a) \end{aligned} \quad (4)$$

where x_p , x_a and x^* correspond to the predicted box, anchor box and ground-truth box, respectively (likewise for w_p , y_p , h_p). Variables x_p , y_p , w_p and h_p denote the two centre coordinates of the bounding box and its width and height. The RPN and subsequent detection are illustrated in Figure 1.

In the second stage, the box proposals are used to detect marine radar image features from the feature map. These features are fed into the remaining layers of the feature extractor for predicting the class probability and bounding box for each proposal.

The entire process is implemented on a unified network as shown in Figure 2, enabling sharing of full-image convolutional features with the detection network and significantly reducing the number of weights for simplifying and speeding-up training.

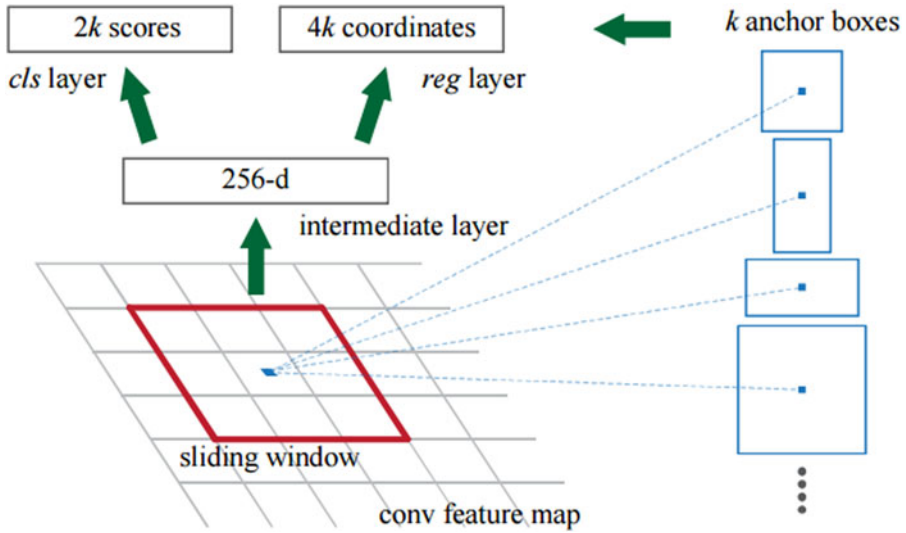


Figure 1. Diagram of the RPN architecture (left) and example of detections using RPN outcomes (right).

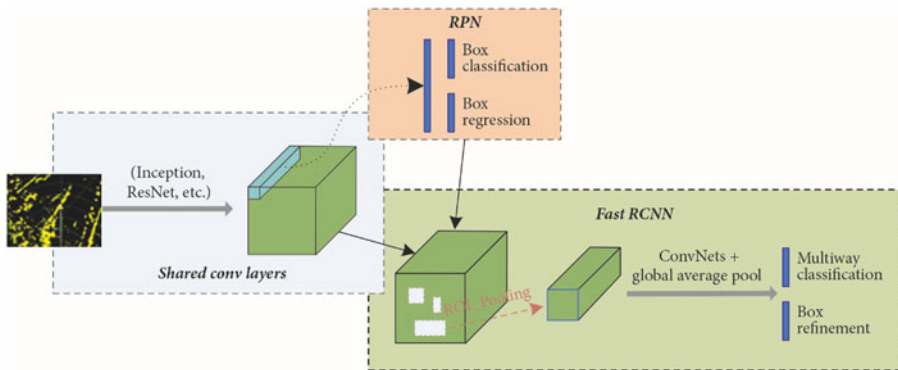


Figure 2. Block diagram of Faster R-CNN (ROI, region of interest).

2.2.2. *R-FCN*. The R-FCN illustrated in Figure 3 uses convolution for prediction and then performs region-of-interest pooling. Adding location information before pooling prevents the fully convolutional network from losing such information and specifying different score maps allows detection of different locations of the target. After pooling, the score maps obtained from different locations can be combined to reconstruct the original location information, whereas the fully convolutional structure reduces training time.

We evaluated the performance of the models using the mean average precision (Everingham et al., 2010). The average precision is the area under the precision (p)–recall (r) curve of the detection task, and its mean is computed over all the classes in a task. The average precision and mean average precision are respectively given by:

$$AP = \sum_i^{n_d} p_i \Delta r$$

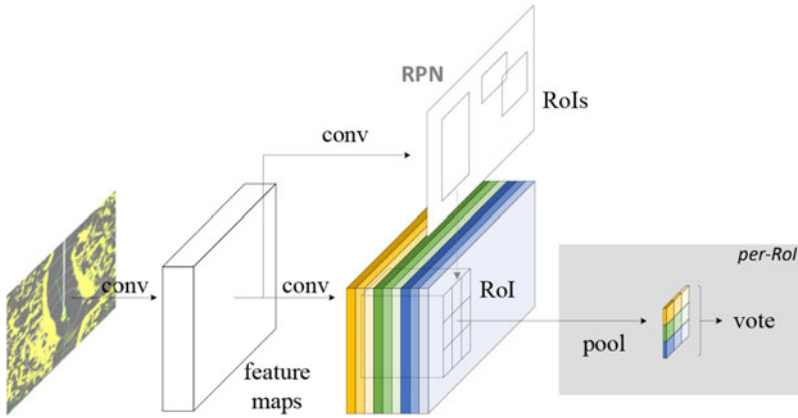


Figure 3. R-FCN with fully convolutional network.

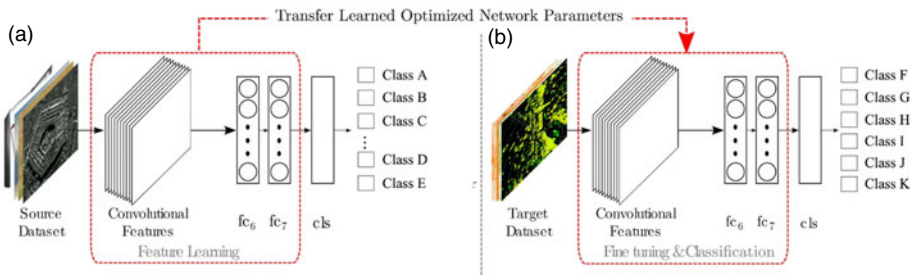


Figure 4. Diagram of transfer learning to improve target detection.

$$mAP = \frac{1}{C} \sum_{C=1}^C AP_C \tag{5}$$

The average precision AP is a dimensionless quantity and integrates both the precision and recall indicators, which are commonly used to measure the ability of neural networks to detect targets, with higher values indicating the model more suitable for the data set. Likewise, the mean average precision mAP is the average precision (including background) of each object, which is used to measure the target detection ability of the model for a category of targets.

2.3. *Transfer learning.* We used transfer learning to fine-tune the detection model as the employed dataset is small. In fact, modern object detection considers a large number of data samples and over a thousand class labels. Then, the hidden layers in a CNN provide distinct feature representations, with lower layers providing general feature extraction and higher layers providing more specific information about the intended task (Dai et al., 2016). For instance, lower layers would be edge detectors, whereas higher layers provide more specific representations belonging to the input image. Therefore, generalisation is provided by feature extraction and representation of lower layers and fine-tuning of the specific problem is provided by higher layers. The transfer learning employed in this study to improve detection is illustrated in Figure 4. In this study,

we used a dataset available online, namely UC Merced Land Use Dataset with 21 object classes (Yang and Newsam, 2010) to perform pre-training. The main purpose was to guide the lower layers of the network to extract features in the proper direction. Then, fine-tuning was performed on higher layers using the radar data set generated in the present study.

3. FUSION BASED ON DETECTED TARGETS.

3.1. *Scaling and display area of ENC.* The coordinate systems of the ENC and corresponding radar images are different, leading to superposition errors, which can be mitigated by performing real-time coordinate transformations. For the ENC, the projection mode defaults to the Mercator projection as this type of projection is simpler to calculate and yields higher accuracy in a small scanning range than similar methods. We used the Mercator projection by considering the scanning range of the radar, which is usually within six nautical miles when ships sail in confined waters.

The effect of the curvature of the Earth between the radar detection coordinate system and the Mercator coordinate system can be ignored as this value is too small compared to the radius of the Earth. When the range is small, the detection distance of the radar is approximately equal to the geographical distance considering the curvature of the earth. The conversion of the geodetic coordinate system into Mercator projection proceeds as follows (Yang et al., 2010):

$$\begin{cases} x_m = r_0 q \\ y_m = r_0 \lambda \end{cases}$$

$$r_0 = \frac{\alpha}{\sqrt{1 - e^2 \sin^2 \varphi_r}} \times \cos \varphi_r \quad (6)$$

$$q = \ln \tan \left(\frac{\pi}{4} + \frac{\varphi}{2} \right) - \frac{e}{2} \ln \frac{1 + e \sin \varphi}{1 - e \sin \varphi}$$

where x_m and y_m are the Mercator Cartesian coordinates, λ and φ are the longitude and latitude, respectively, α is the major axis of the Earth's ellipsoid, e is the first eccentricity of the Earth, φ_r is the reference dimension, q is the equivalent dimension, and r_0 is the radius of the base circle of the latitude.

According to the projection transformation of chart pixels, the latitude and longitude of the radar scan can be obtained as:

$$\begin{cases} (\lambda_{\min}, \lambda_{\min}) = \phi_{org}^{-1} (\phi (\lambda_0, \varphi_0) - (L, L)) \\ (\lambda_{\max}, \lambda_{\max}) = \phi_{org}^{-1} (\phi (\lambda_0, \varphi_0) + (L, L)) \end{cases} \quad (7)$$

where λ_0 and φ_0 are the longitude and latitude of the radar centre, respectively, $(\lambda_{\min}, \varphi_{\min})$ and $(\lambda_{\max}, \varphi_{\max})$ are the longitude and Latitude ranges of the radar scanning, respectively, $\phi(\lambda_0, \varphi_0)$ is the longitude and Latitude of the radar centre in Cartesian coordinates with ϕ being the map for Mercator projections, respectively, ϕ^{-1} is the inverse of the Mercator map for chart projection, and L is the detection range of the radar. Then, we searched the area of the ENC to be registered using the results obtained.

In terms of the registration, the ENC and marine radar scales must be consistent. Modifying the radar scale is achieved by several variation ranges, whereas ENC data can be

arbitrarily set to the desired scale. Therefore, we matched the electronic chart to radar data (Yang et al., 2010). A radar image scale and the display scale of the ENC are determined by:

$$\begin{aligned}\frac{1}{S_r} &= \frac{c \times d}{R} \\ z &= \frac{(c \times d \times s_0)}{R}\end{aligned}\quad (8)$$

where R is the detection radius of the radar, c pixels at distance d are contained in the radius scan line, and $1/s_r$ is the original scale of the radar. However, the remaining area, excluding the radar image area, is displayed in black on the radar display screen. Thus, $1/s_r$ can also be regarded as the scale of the external square of the radar circular image. $1/s_0$ is the original scale of the ENC, and z is the scaling parameter of the ENC.

According to the azimuth and distance of the corners of the bounding boxes obtained from radar target detection with respect to the registration reference point, areas outside the bounding box in the ENC are treated as background, whereas the area within the bounding box of the ENC data and radar image is the Registration Interest Area (RIA).

3.2. RIA processing.

3.2.1. *Filtering algorithm.* Unlike registration based on entire radar images, filtering and parameter selection considering only RIAs can be more accurate for extracting the edges. For enhanced applicability, we consider real-time fusion and adaptive mean filtering, which is expressed as:

$$K(x) = c_k k(\|x\|^2) \quad (9)$$

where x is a pixel in the plane, c_k is a constant greater than 0, $k(x)$ satisfies $\int k(x) dx < 0$, and $K(x)$ is a kernel function. The kernel density estimate is given by:

$$\begin{aligned}f(x) &= \frac{1}{n} \sum_{i=1}^n K_H(x - x_i) \\ K_H(x) &= |H|^{-1/2} K(H^{-1/2}x)\end{aligned}\quad (10)$$

where $H = h^2 I$. Simplifying the density function, we obtain:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n k\left(\left\|\frac{x - x_i}{h}\right\|\right) \quad (11)$$

whose gradient is given by:

$$\nabla f(x) = \frac{2C_k}{nh^{d+2}} \sum_{i=1}^n k'\left(\left\|\frac{x - x_i}{h}\right\|^2\right). \quad (12)$$

Assuming $g(x) = -k'(x)$, $G(x) = c_k g(\|x\|^2)$, and $G(x)$ is the opacification function of $K(x)$, we obtain:

$$\begin{aligned} \nabla f(x) &= \frac{2c_k}{nh^{d+2}} \left[\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|\right)^2 \right] \left[\frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|\right)^2}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|\right)^2} - x \right] \\ &= \frac{2c_k}{nh^{d+2}} \left[\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|\right)^2 \right] [m_h(x) - x] \end{aligned} \tag{13}$$

with $m_h(x)$ being the weighted average of the sample points used to update the search area with radius h and $m_h(x) - x$ being a drift vector, which terminates the iteration process when it is below a certain tolerance level.

3.2.2. *Binarization.* To avoid intensive processing of colour images, we binarized greyscale images using the method by Otsu (1979) to prevent disparity between the RIA foreground and the background size ratio. This method adaptively searches for a threshold according to the characteristics of the image and converts a greyscale image into a binary one. The transformation can be expressed as:

$$\begin{aligned} u &= w_0 * u_0 + w_1 * u_1 \\ g(t) &= w_0(u_0 - u)^2 + w_1(u_1 - u)^2 \end{aligned} \tag{14}$$

where w_0 is the ratio of foreground to image with u_0 being its average greyscale value and w_1 is the ratio of background to image with u_1 being its average greyscale value. The highest variance g maximises the foreground and background difference, and the corresponding threshold t is the optimum.

3.2.3. *Morphological processing.* Binarization may produce some isolated noise pixels besides the foreground target, affecting its integrity. We performed noise removal by applying corrosion, expansion, opening, and closing, which can be expressed using the standard set theory as:

$$\begin{aligned} A \ominus B &= \{w \in Z | w + b \in A, b \in B\} \\ A \oplus B &= \{w \in Z | w = a + b, a \in A, b \in B\} \\ A \circ B &= (A \ominus B) \oplus B \\ A \bullet B &= (A \oplus B) \ominus B \end{aligned} \tag{15}$$

where Z represents the foreground, w is a pixel at different positions in the foreground image, and a and b are members of structural element sets A and B , respectively.

We used the improved canny operator (Othman and Abdullah, 2018) to extract the edge of the radar image and boundary of the echoes in the RIAs:

$$\begin{aligned} E_x &= \frac{\partial P}{\partial x} * f(x, y) \quad E_y = \frac{\partial P}{\partial y} * f(x, y) \\ A(i, j) &= \sqrt{E_x^2(i, j) + E_y^2(i, j)} \end{aligned}$$

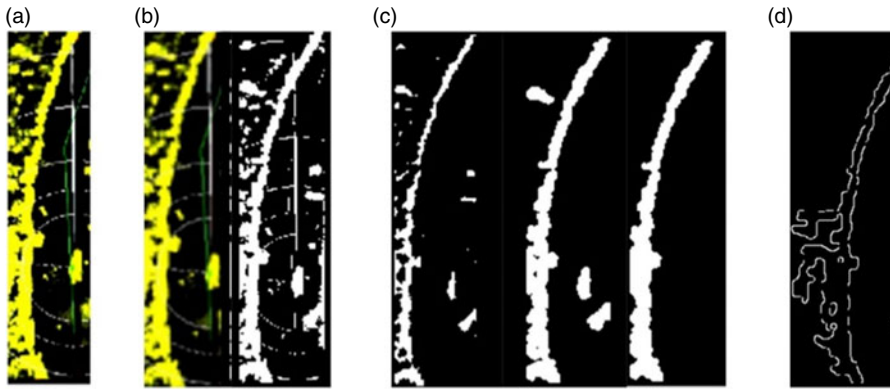


Figure 5. Image processing of RIA. (a) RIA; (b) Filtering and Binarization; (c) Morphological processing; (d) Canny.

$$a(i, j) = \arctan \left(\frac{E_x(x, y)}{E_y(x, y)} \right) \tag{16}$$

where $P(x, y)$ is a two-dimensional Gaussian function, $f(x, y)$ is the image data, $A(i, j)$ is the edge feature of point (i, j) in the image, and $a(i, j)$ is the normal vector of the image at point (i, j) . Figure 5 illustrates the complete image processing for an RIA.

3.3. *Image registration.* Image registration can be based on transformations of greyscale images, features and domain. Considering the notable divergence between marine radar images and ENC, we used a feature-based affine transformation for their fusion. Three or more reference points should be selected for the transformation, and its principle can be expressed as:

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = k' \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{17}$$

where (x_i, y_i) are the coordinates of a pixel in the ENC, k' , θ , x_0 and y_0 are parameters for the affine transformation, where θ corresponds to the image rotation, k is a scale factor and x_0 and y_0 correspond to the image shift. Thus, we obtain:

$$\begin{cases} x'_i = a_{11}x_i + a_{12}y_i + a_{13} \\ y'_i = a_{21}x_i + a_{22}y_i + a_{23} \end{cases} \tag{18}$$

with $a_{11} = a_{22} = k' \cos \theta$, $a_{13} = x_0$, $a_{12} = -a_{21} = k' \sin \theta$, and $a_{23} = y_0$.

Therefore, correlating radar images with ENC data requires at least three pairs of reference points with accurate correspondence to determine the transformation parameters. The accuracy of the reference points directly determines the fusion outcome. After the transformation, however, some deformation may occur, but it is generally negligible. On the other hand, if the deformation undermines fusion, then according to “the display of radar information shall have priority” from IEC 62388 (2013), priority should be given to the display of radar images.

To determine the reference points for registration, we considered various aspects. After target detection from the radar image, targets with small sizes and inconspicuous shapes

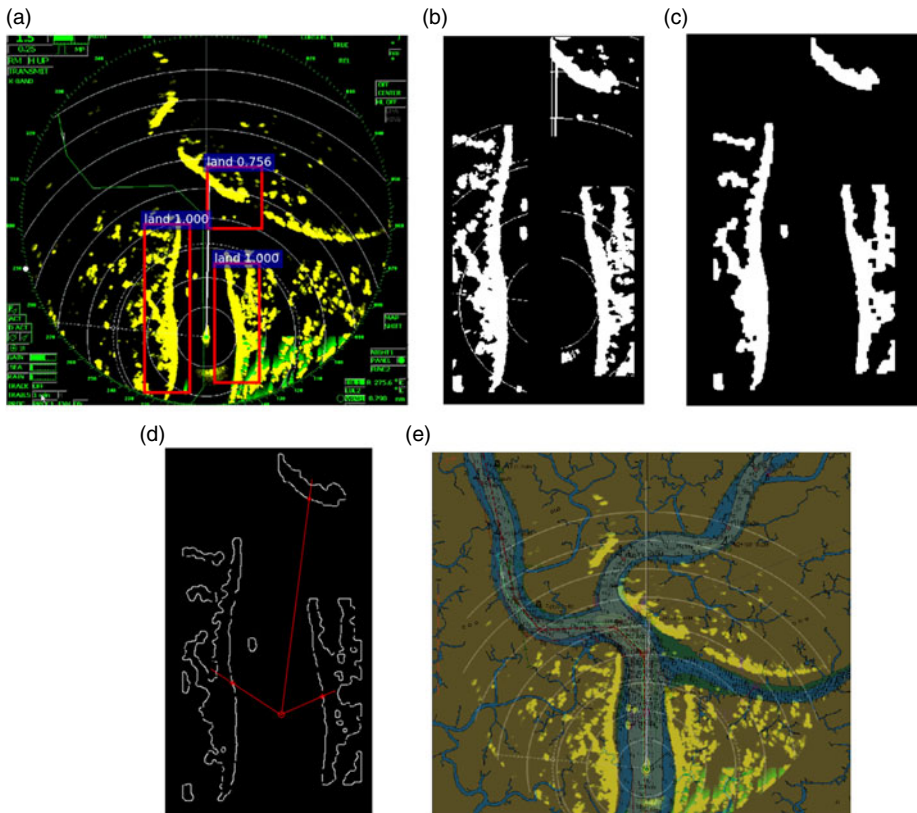


Figure 6. Complete process for automatic registration from detected typical targets. (a) Detection result; (b) Binarization; (c) Morphological processing; (d) Canny; (e) Registration.

may interfere with the detection results, as registration is less effective for small objects such as buoys. However, according to the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers, when ships are sailing along the coast in crowded waters, the crew should check the position of the ship at appropriate intervals through independent positioning methods, and reference objects are mainly used for navigation. Therefore, small reference objects that can accurately correspond to the ENC goal should not be ignored. Consequently, pixels between the radar image and ENC are suitable as registration reference points.

Depending on the target resolution of the radar, manual selection of registration reference points cannot be performed on large ranges. Therefore, we implemented the automatic registration algorithm, illustrated in Figure 6, which focused on objects that perform well in target detection as reference points for registration. The algorithm first locates a pair of accurate reference points on the ENC and radar images whose respective coordinates are (x_1, y_1) and (x'_1, y'_1) . When two or more bounding boxes are found, a ray is emitted from (x_1, y_1) and (x'_1, y'_1) to the geometric centre of the RIA in the ENC and radar images to acquire two or more groups of coordinates in front of the radar echo edge. If there is only one bounding box, a specific angle on the bow is selected, emitting two rays to the

Table 1. Detection result of Faster R-CNN and R-FCN for typical radar target.

Model	<i>mAP</i> (%)	Cape	Breakwater	Ship	Shoreline	Bridge	Racon
Faster R-CNN	0.2973	0.3856	0.3931	0.2260	0.4986	0.4356	0.065
R-FCN	0.3649	0.4017	0.3960	0.2446	0.5498	0.5032	0.064

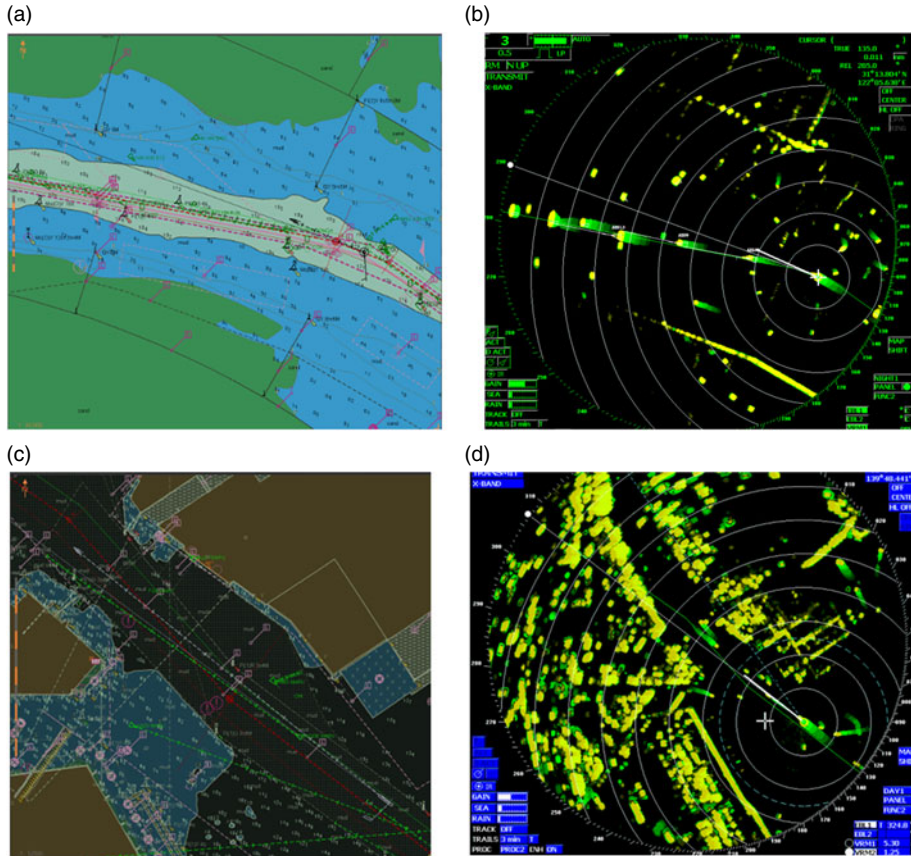


Figure 7. Typical target with undistinguishable echoes: (a) ENC of a shoal; (b) radar echoes of a shoal; (c) ENC of a breakwater; (d) Civil Marine Radar (CMR) echoes of a breakwater.

areas around (x_1, y_1) and (x'_1, y'_1) to obtain the coordinates of the two registration reference points.

4. RESULTS AND DISCUSSION.

4.1. *Typical target detection.* We leveraged the UC Merced Land Use Dataset with 21 object classes (Yang and Newsam, 2010) as a dataset sample for pre-training the network toward the problem domain of specific object classes for detection. The annotated typical target dataset (ENC data and marine radar images acquired from the voyage data recorder monitor) was used to fine-tune the pre-trained model using transfer learning. We used stochastic gradient descent with momentum and weight decay of 0.9 and 0.0005, respectively. The initial learning rate of 0.001 was divided by 10 using stepdown every

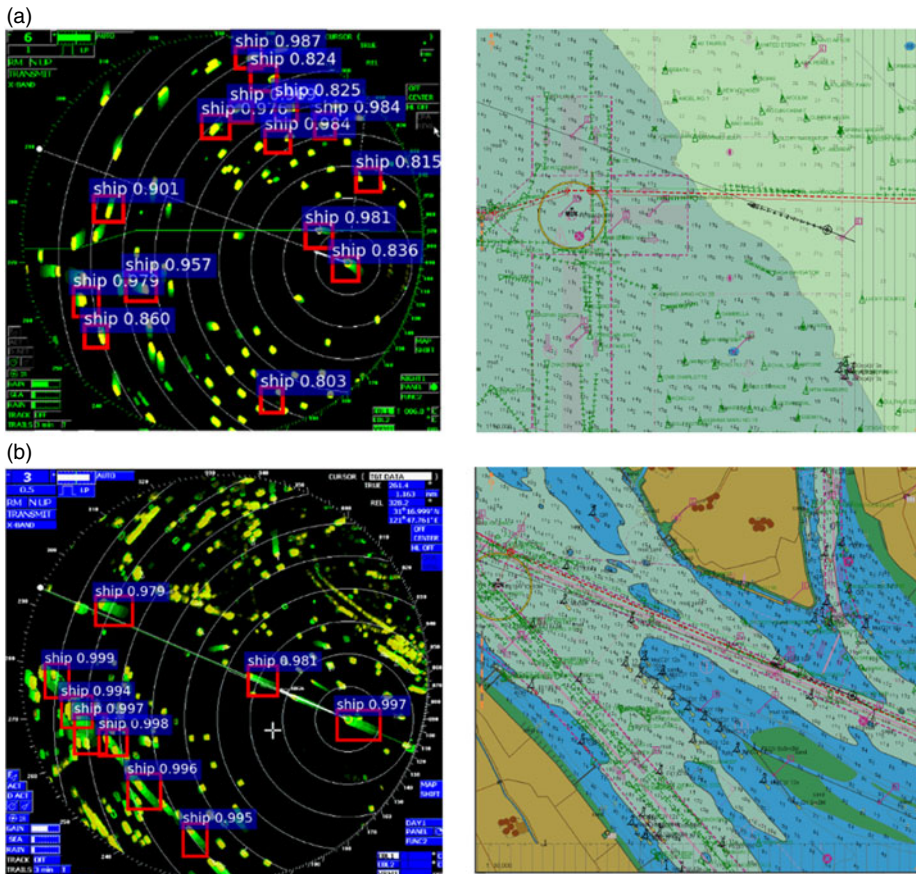


Figure 8. Target detection of ships showing detection omission (a) without and (b) with radar persistence.

12,000 iterations. For the RPN in the Faster R-CNN and R-FCN, we set the batch size to 256.

Table 1 lists the performance of the Faster R-CNN and R-FCN in the data set we constructed. Both detection performances for typical marine radar targets are not ideal. However, the R-FCN performed better than the Faster R-CNN as the *mAP* of the R-FCN is higher than that of the Faster R-CNN; a detection model with a higher *mAP* value can classify the target data set considering the precision and recall overall. Thus, the R-FCN is considered more suitable to perform the detection task in our data set efficiently. From these results, we determined that various factors should be considered for typical radar target detection. First, radar information from different target types may be very similar, thus undermining the detection accuracy. Figure 7 shows examples of indistinguishable objects, namely a shoal (Figure 7(b)) and a breakwater (Figure 7(d)).

In addition, targets with small sizes and inconspicuous shapes may interfere with the detection results. In fact, such targets produce a low average precision that undermines detection. Furthermore, although the position of the ship is stable, and its geometric shape is regular with strong capability to reflect Civil Marine Radar (CMR) echoes, the average precision of the vessel is low; furthermore, there will be leakage detection and error

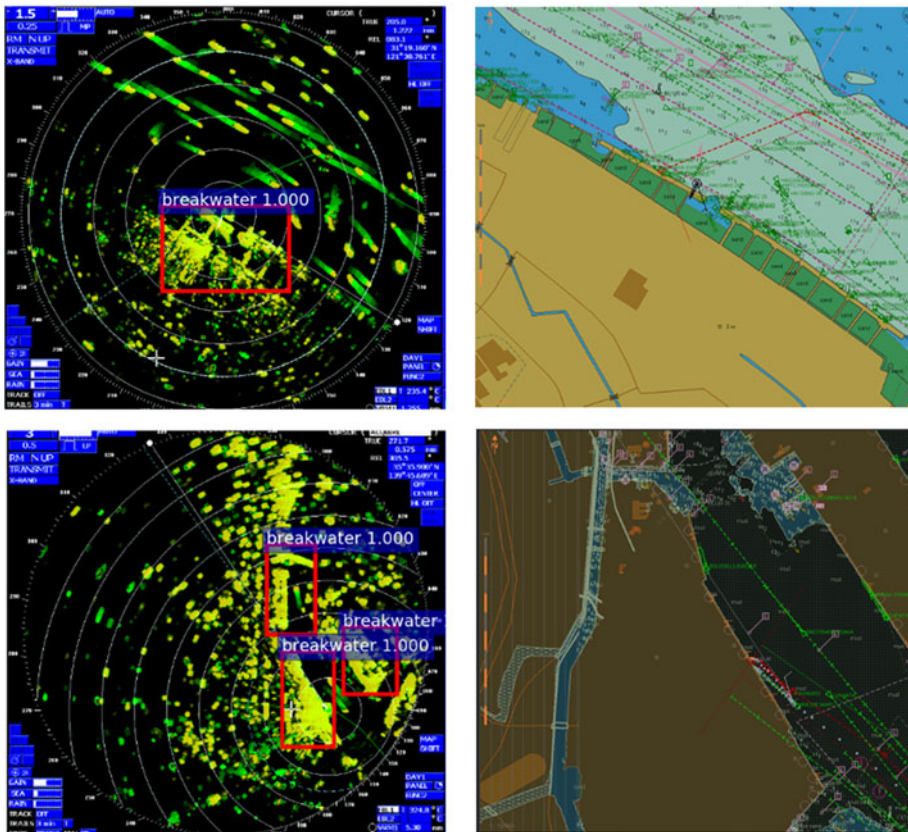


Figure 9. Breakwater selected as typical target in the area of the Yangtze river and Kobe port.

detection, as illustrated in Figure 8. Therefore, ships cannot be used as reference points for automatic registration.

To improve the detection performance, we considered typical targets in the radar images with different characteristics and trained different models to reduce the detection classes and avoid similarity among different targets. For example, we considered the waters near the Yangtze River estuary (Figure 9). According to the Shanghai Gang sailing direction, arriving ships mostly rely on lights and lightships for positioning and navigation. However, it is difficult to distinguish lighthouses and lightships from a large number of densely anchored ships. In addition, the position and shape of echoes from smooth beach shorelines and mudflats can change with tide diversification. Moreover, radar images usually do not produce notable distinctions between steep and smooth beach shorelines, which are thus not suitable as reference points for registration and navigation. In these types of waters, echo-stable buildings are preferred as typical targets such as breakwaters and quayside edges, as shown in Figure 9. When these targets were used for the R-FCN, we obtained better detection performance. Specifically, target detection in the waters of the Yangtze River and waters near the Kobe Port estuary yielded average precision for the breakwater of up to 0.4372, which is higher compared to using other targets.

Table 2. Detection of typical targets from multiple radar in waters near the Port of Caledon, Vietnam.

$mAP(\%)$	Bridge	Land	Cape
0.5178	0.5348	0.5847	0.4338

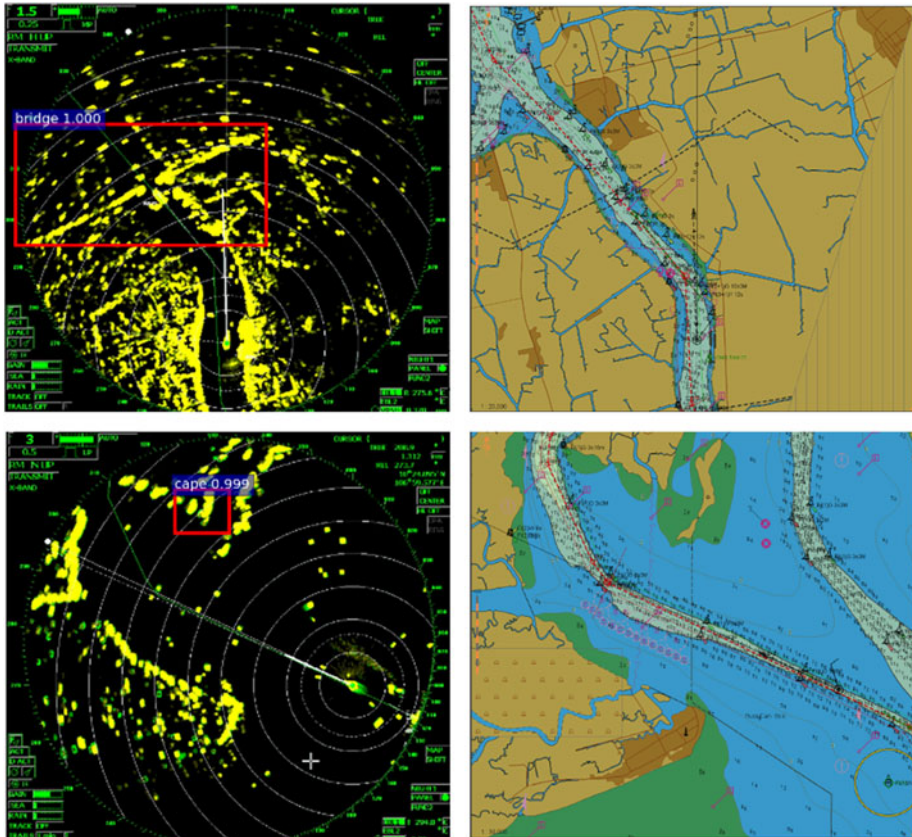


Figure 10. Detection of typical targets near the Port of Caledon, Vietnam.

We also verified the feasibility of the R-FCN in the inland waters near the Caledon port, Vietnam and the results obtained are listed in Table 2 and shown in Figure 10. The detection performance notably improved with a mAP of 0.5178, which is significantly higher than the values previously obtained (0.2973 and 0.3649). This indicates that reducing the classification category of the target helps to improve the target detection performance of the model after considering the precision and recall rate.

4.2. *Data fusion.* The affine transformation parameters, k , θ , x_o , and y_o in Equation (17) were determined to perform fusion simulations on ENC data and CMR images at intervals of 5 min. The results are shown in Figure 11.

The algorithm yields a suitable performance because fusion of the shorelines from the ENC and radar images coincide. Moreover, the shoreline and shoal are separated in

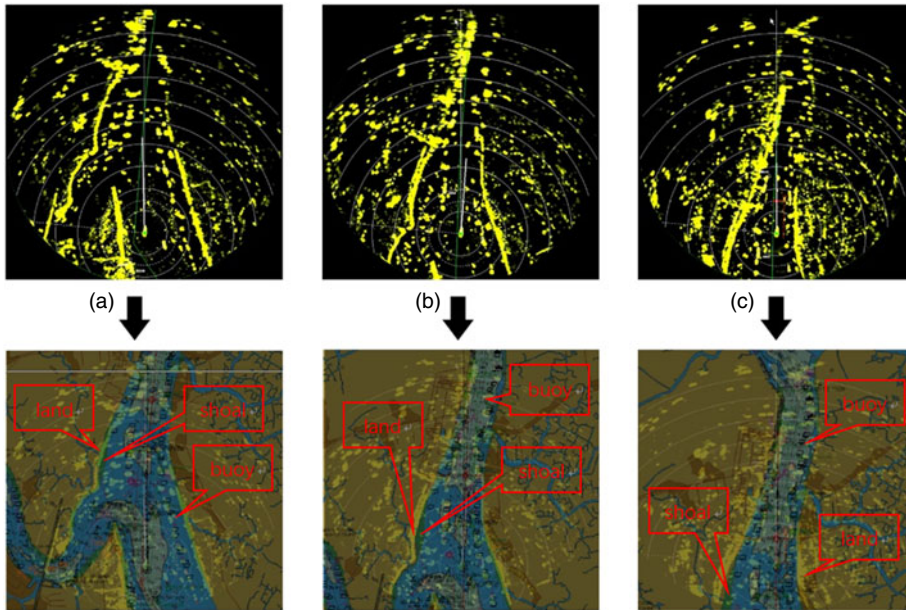


Figure 11. Using fixed parameter to perform data fusion of ENC and radar images every five minutes.

Figures 12(a) and 12(b) without loss of radar information. The image details show successful fusion by the consistent position of the land in Figure 11(a) and that of the buoys in Figures 12(a)–(c). Thus, image fusion of ENC and marine radar images can be performed even without the identification of accurate feature points.

We executed training and registration using a Nvidia GeForce GTX 1080 GPU and Intel Core i5–7 processors. With this equipment, the average time for target detection on a single test image was 0.15 s, and registration was completed within 0.5 s, which is below the maximum update rate of the marine radar antenna (48 rpm), suggesting that the proposed algorithm satisfies the requirement for real-time performance.

The success of the registration algorithm depends on the quality of marine radar image processing. Moreover, if the radar detects contiguous objects for reference points near the range, these objects will cause pixel adhesion, and the canny operator cannot extract the edge features of these objects. We performed a simulation to demonstrate this situation. As shown in Figure 12(e2), the shoreline on the left side is steep without occlusion from other objects. However, the shoreline on the right of the ship is covered by a building. Taking the shoreline on the left and right of the ship as RIAs to perform automatic registration with the corresponding reference points yields the results shown in Figure 12.

Fusion using the RIA from the left of the ship successfully matches the shoreline and the bridge, but that from the right of the ship fails. The failure is caused by limitations in the working principle and performance of a CMR. Even with human assessment, it is difficult to find pixels that can accurately correspond to the ENC position among many erratic pixels. This complicates the search for registration reference points in such areas. Therefore, registration reference points in the RIA should be searched only when no object interference occurs.

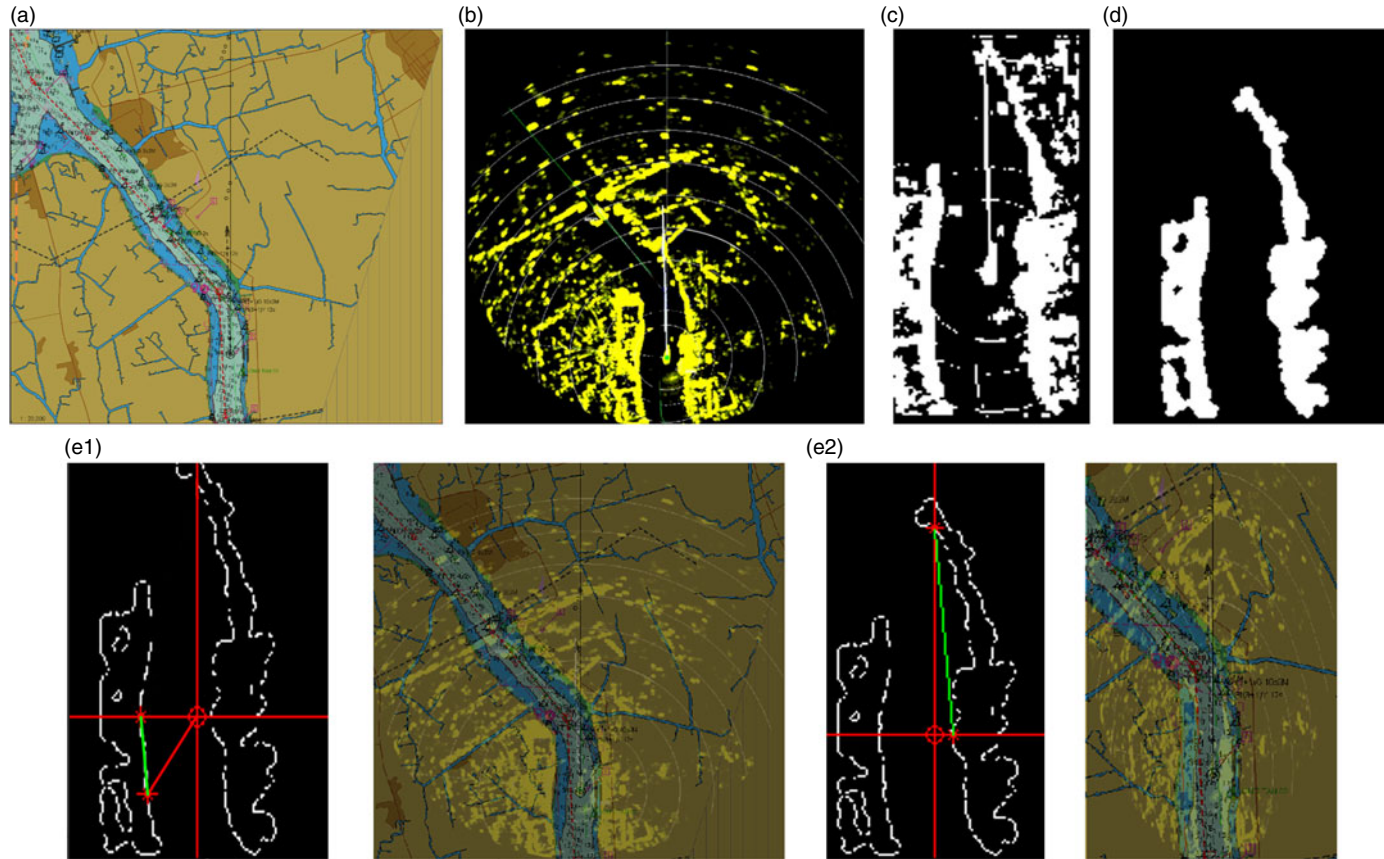


Figure 12. Image fusion using reference points on the left and right of the ship. (a) ENC; (b) CMR image; (c) Binarization; (d) Morphological; (e1) Left side of ship; (e2) Right side of ship.

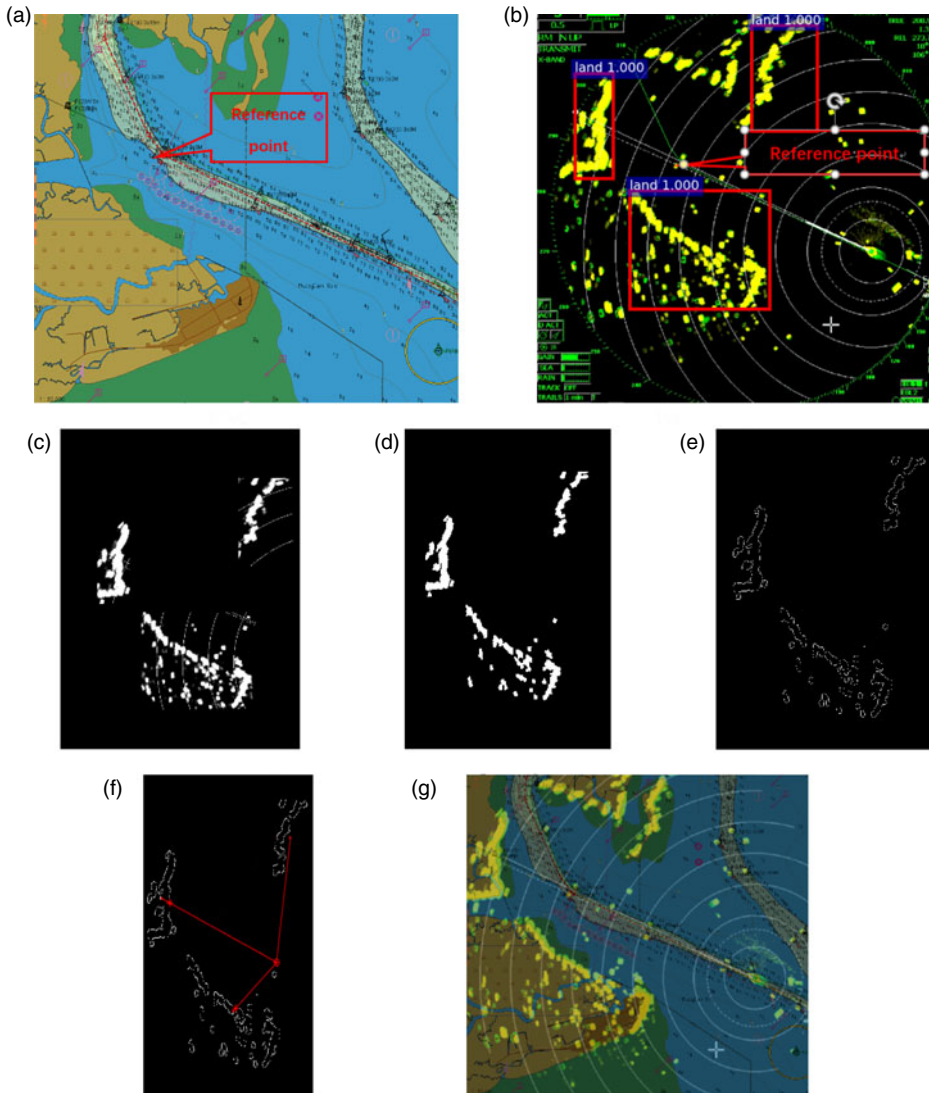


Figure 13. Data fusion using other reference points detected by radar under large error of the ship GNSS. (a) ENC; (b) Typical target detection; (c) Binarization; (d) Morphological; (e) Canny operator; (f) Auto Registration Algorithm; (g) Registration.

Now, consider the case when the ship is being used as the reference point and the GNSS is giving a large error, thus drastically undermining fusion. In this case, achieving fusion demands reference points other than the ship. For instance, if the buoy detected by radar with clear echoes and stable position on the ENC is used as a registration reference point for fusion, the algorithm yields the result shown in [Figure 13](#).

The automatic fusion of ENC data and marine radar image is realised using the buoy echo detected by the radar, disregarding information on the ship position. The fusion results show that the ship position on the radar is consistent with the GNSS signal on the ENC,

and the radar echo of the shoreline is accurately integrated with ENC data, verifying the high performance of the proposed algorithm under this scenario.

5. CONCLUSIONS. High-level data fusion is not realised by conventional chart radar, as it depends on the measurement environment and is a cumbersome process requiring much experience. Recent studies have mostly focused on target detection considering complete images and demands the adjustment of several parameters more related to image pre-processing than to RIAs. In addition, detection by different operators has the limitation of fixed types of detection features and is affected by radar echo quality and target deformation, often failing to produce ideal results. To avoid these limitations, we proposed a method based on deep learning to detect RIAs on marine radar images. The proposed method can detect different and robust image feature maps through a continuous convolution layer, which results in more accurate and robust reference points for registration obtained after contour extraction from detected targets. We employed an affine transformation to fuse corresponding ENC data and marine radar images. The proposed algorithm was verified for fusion of ENC data and marine radar images retrieved from real ships sailing in narrow channels. Fusion can be performed even under considerable errors in both the ship positioning measurements and heading signal. Fusion results demonstrate the effectiveness and consistency of the proposed algorithm. The results from data fusion between ENC data and marine radar images can help in the interpretation of navigation scenarios. Moreover, the image retrieved from fusion can provide real-time dynamic information of marine radar and suitably reflect ENC static navigation information, thus improving aspects such as safety during navigation. In a broader sense, the proposed method can serve as a guideline for future integration and data fusion of multiple navigation sensors.

ACKNOWLEDGMENTS

This work was supported by the National Nature Science Foundation of China (Nos. 51579024, 51879027, 61374114) and the Fundamental Research Funds for the Central Universities (DMU no. 3132016311)

REFERENCES

- Al-Sharman, M.K., Emran, B.J., Jaradat, M.A., Najjaran, H., Al-Husari, R. and Zweiri, Y. (2018). Precision landing using an adaptive fuzzy multi-sensor data fusion architecture. *Applied Soft Computing*, **69**, 149–164.
- Borkowski, P. and Zwierzewicz, Z. (2011). Ship course-keeping algorithm based on knowledge base. *Intelligent Automation & Soft Computing*, **17**(2), 149–163.
- Donderi, D.C., Mercer, R., Hong, M.B. and Skinner, D. (2004). Simulated navigation performance with marine electronic chart and information display systems (ECDIS). *The Journal of Navigation*, **57**(2), 189–202.
- Dai, J., Li, Y., He, K. and Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. In E. Q. Chen, Y. H. Gong, and Y. Tie (Eds.), *Advances in neural information processing systems*, 379–387. Barcelona, Spain: Springer.
- Du, C. and Gao, S. (2017). Multi-focus image fusion algorithm based on pulse coupled neural networks and modified decision map. *Optik*, **157**, 1003–1015.
- Everingham, M., van Gool, L., Williams, C.K., Winn, J. and Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, **88**(2), 303–338.
- Gao, B., Hu, G., Gao, S., Zhong, Y. and Gu, C. (2018). Multi-sensor optimal data fusion for INS/GNSS/CNS integration based on unscented Kalman filter. *International Journal of Control, Automation and Systems*, **16**(1), 129–140.

- Goyal, M., Yap, M. H., Reeves, N. D., Rajbhandari, S. and Spragg, J. (2017, October). Fully convolutional networks for diabetic foot ulcer segmentation. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 618–623. New Jersey, NJ: IEEE.
- Hall, D.L. and Llinas, J. (2002). An introduction to multisensor data fusion. *Proceedings of the IEEE*, **85**(1), 6–23.
- He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. *Computer vision and pattern recognition*, 770–778. USA: Las Vegas.
- Hu, G., Gao, S. and Zhong, Y. (2015). A derivative UKF for tightly coupled INS/GPS integrated navigation. *ISA Transactions*, **56**, 135–144.
- International Electrotechnical Commission. (IEC). (2013). *IEC 62388: Maritime navigation and radiocommunication equipment and systems—Shipborne radar-performance requirements, methods of testing and required test results*. Geneva, Switzerland. Retrieved from <https://webstore.iec.ch/publication/6967>.
- Kazimierski, W. and Stateczny, A. (2015). Radar and automatic identification system track fusion in an electronic chart display and information system. *The Journal of Navigation*, **68**(6), 1141–1154.
- Lubczonek, J. (2015). Analysis of accuracy of surveillance radar image overlay by using georeferencing method. In: *Radar Symposium*, Dresden, Germany. <http://dx.doi.org/10.1109/IRS.2015.7226230>.
- Liu, W.T., Ma, J.X. and Zhuang, X.B. (2005). Research on radar image & chart graph overlapping technique in ECDIS. *Navigation of China*, **62**(1), 59–63.
- Ma, J., Jiang, J., Zhou, H., Zhao, J. and Guo, X. (2018). Guided locality preserving feature matching for remote sensing image registration. *IEEE Transactions on Geoscience and Remote Sensing*, **56**(8), 4435–4447.
- Othman, Z. and Abdullah, A. (2017). An adaptive threshold based on multiple resolution levels for canny edge detection. In *International Conference of Reliable Information and Communication Technology*, 316–323. Germany: Springer.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, **9**(1), 62–66.
- Ren, S., He, K., Girshick, R. and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, 91–99. Tokyo, Japan.
- Simonyan, K. and Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations*, 1–14. USA: San Diego.
- Wu, S., Xu, J., Zhu, S. and Guo, H. (2017). A deep residual convolutional neural network for facial keypoint detection with missing labels. *Signal Processing*, **144**, 384–391.
- Yang, G.L., Dou, Y.B. and Zheng, R.C. (2010). Method of image overlay on radar and electronic chart. *Journal of Chinese Inertial Technology*, **18**(2), 184.
- Yang, Y. and Newsam, S. (2010). Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 270–279. New York NY: ACM.
- Zhang, C. and Fan, Z. Z. (2011). Data Fusion of Radar Image and ECDIS Based on Matching of Harris Feature Points. In *Advanced Materials Research*, **317**, 2026–2029. Stafa-Zurich, Switzerland: Trans Tech Publications.
- Zhao, Y., Gao, S.S., Zhang, J. and Sun, Q.N. (2014). Robust predictive augmented unscented Kalman filter. *International Journal of Control, Automation and Systems*, **12**(5), 996–1004.