Journal of Zhejiang University-SCIENCE A 2026 27(1):1-11
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2500277

Check for updates

Digital twin-assisted automatic ship size measurement for ship–bridge collision early warning systems

Ruixuan LIAO[1], Yiming ZHANG[1✉], Hao WANG[1✉], Jianxiao MAO[1], Aoyang LI[2], Zhengyi CHEN[1,3]

[1]*Key Laboratory of Concrete & Prestressed Concrete Structures of Ministry of Education, Southeast University, Nanjing 211189, China*
[2]*Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana IL 61801, USA*
[3]*Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*

Abstract: Long-span bridges are usually constructed over waterways that involve substantial ship traffic, resulting in a risk of collisions between the bridge girders and over-height ships. The consequences of this can be severe structural damage or even collapse. Accurate measurement of ship dimensions is an effective way to monitor approaching over-height ships and avoid collisions. However, the performance of current techniques for estimating the size of moving objects can be undermined by large sensor-to-object distance, limiting their applicability. In this study, we propose a digital twin-assisted ship size measurement framework that can overcome such limitations through a predictive model and virtual-to-real-world transfer learning. Specifically, a 3D synthetic environment is first established to generate a synthetic dataset, which includes ship images, positions, and dimensions. Then the pixel information and spatial coordinates of ships are adopted as regressors, and ship dimensions are selected as the output variables to pre-train deep learning models using the generated dataset. Coordinate system transformations are applied to address dataset bias between the simulated world and real-world, as well as improve the model's generalization. The pre-trained models are compared using supervised virtual-to-real-world transfer learning to select the version with optimal real-world performance. The mean absolute percentage error is only 3.74% across varying camera-to-ship distances, which demonstrates that the proposed method is effective for over-limit ship monitoring.

Key words: Ship–bridge collision early warning; Over-height ship monitoring; Ship size measurement; Digital twins; Computer vision; Transfer learning

1 Introduction

Collisions between bridges and over-height ships are becoming more frequent because of the rapid development of waterborne transport and the increasing number of bridges that span busy waterways (Wang et al., 2008). These accidents often lead to severe structural damage, substantial economic losses, and even human casualties (Fan et al., 2020). To mitigate the consequences of ship–bridge collisions, great emphasis has been placed on improving the impact resistance of bridge sub-structures such as piers and pylons (Fan et al., 2008; Guo and He, 2020). However, the ship collision design of bridge super-structures, such as bridge girders, is commonly neglected (Sha et al., 2019). The high kinetic energy of passing ships whose height exceeds the clearance could therefore lead to catastrophic bridge collapse, highlighting the need to monitor and warn over-height ships in waterways with bridges (Pedersen et al., 2020; Zhang et al., 2022).

Typical methods for monitoring over-height ships adopt a predefined height threshold, which is established by projecting laser beams across the navigation channel (Urazghildiiev et al., 2007). Ships that intersect these beams are identified as oversized. In practice, such laser-based systems are costly and the beams are easily attenuated or scattered by atmospheric conditions, leading to frequent false alarms (Sazonov, 2011). Direct measurement of ship dimensions is expected to support early detection of over-height ships, which in turn would enhance the reliability of ship–bridge

✉ Yiming ZHANG, yiming.zhang@seu.edu.cn
Hao WANG, wanghao1980@seu.edu.cn
Yiming ZHANG, https://orcid.org/0009-0007-9261-5510
Hao WANG, https://orcid.org/0000-0002-1187-0824

Received June 28, 2025; Revision accepted Sept. 18, 2025; Crosschecked Dec. 11, 2025

collision warnings (Pallotta et al., 2013; Inazu et al., 2016). Nevertheless, studies focusing on ship size measurement in this context remain scarce. Current methods for estimating the size of moving objects mainly include light detection and ranging (LIDAR) scanning, and computer vision (CV)-based methods (Chan and Lee, 2013; Yurdakul et al., 2021; Shao et al., 2024). Although LIDAR can be used to extract object sizes through point cloud analysis, its associated equipment is expensive and its effective range is typically limited to a few hundred metres (Yurdakul et al., 2021). In practical ship size calculation scenarios, the LIDAR-to-object distance usually exceeds this range. As a result, the measurement performance may be undermined due to reduced resolution and incomplete data (Gargoum et al., 2018). CV techniques, such as monocular vision, stereo vision, view geometry, and depth camera techniques, provide cost-effective measurement solutions (Lu and Dai, 2023; Liao et al., 2024, 2025). These approaches are often integrated with deep learning (DL) to enhance the ability to extract the dimensions of objects (Bian et al., 2010). For instance, the width and length of bridge cracks can be determined by combining monocular vision and object detection algorithms (Ni et al., 2019), while vehicle height can be calculated by constructing 3D bounding boxes and estimating the vanishing points (Lu and Dai, 2023). These CV-based methods exhibit high precision when operating at relatively short camera-to-object distances (Ye et al., 2021). Since these methods rely on camera parameter estimation to calculate object dimensions, their accuracy can be significantly influenced by variations in ship-to-camera distances (Bian et al., 2021). Consequently, CV-based methods may have limited applicability to ship size measurement (Hou et al., 2025).

Regression modeling is an efficient data-driven approach that exploits the intrinsic relationship between predictors and outputs (Chen et al., 2010), being used for measuring grain sizes (Gajalakshmi et al., 2017), footprint sizes (Fascione et al., 2014), and body weights (Ibrahim et al., 2021), among many other applications. Current regression methodologies for ship dimension measurement involve constructing regression datasets using Sentinel-1 synthetic aperture radar (SAR) images and automatic identification system (AIS) data, followed by training DL models to capture the nonlinear relationships between the input variables and ship size parameters (Li et al., 2018; Ren et al., 2022). These approaches do not require camera parameter estimation, effectively mitigating the influence of object distance on the accuracy of ship size measurement. However, SAR images usually contain few 2D pixel features of objects, and AIS data only includes ship length and width information (Li et al., 2018; Zhang et al., 2022). Such factors reduce the ability of these approaches to estimate ship height, thus impacting the monitoring efficacy. Additionally, such approaches require large training datasets, which often involve time-consuming collection efforts (Gajalakshmi et al., 2017).

Digital twins (DT), with their capacity to depict physical and structural conditions, virtual space, and their interrelations, present a promising solution for 3D simulations and synthetic data generation (Zhai et al., 2025). Based on these simulated data, pre-trained models can be developed and applied to various real-world cases through virtual-to-real-world transfer learning (Gaidon et al., 2016). For instance, Iuzzolino et al. (2018) collected images from the Unity environment to train models for classifying trail directions, later applying them in real-world scenarios. Similarly, Huang et al. (2024) generated synthetic rail surface images to train defect detection models, which were subsequently transferred to real-world applications. With this context, it is reasonable to construct a DT scenario to generate the required data and pre-train models for predicting ship sizes. Nevertheless, a substantial domain gap often exists between synthetic and real-world data, leading to dataset bias in transfer learning and potentially undermining the generalization ability of pre-trained models (Iuzzolino et al., 2018; Liao et al., 2025). To mitigate such impact, various domain adaptation methods have been proposed, which are typically classified into three categories: discrepancy-based adaptation, adversarial-based adaptation, and reconstruction-based adaptation (Lu et al., 2023; Huang et al., 2024). While these methods have commonly been used to minimize cross-domain discrepancies in image data, they have not been explored for the specific task of ship size prediction.

In this study, we propose a DT-assisted measurement framework for over-height ship monitoring to mitigate the effects of camera-to-ship distance and address data collection challenges; this is accomplished through a combination of regression modeling and virtual-to-real-world transfer learning. The major contributions of this work can be summarized as follows:

(1) A 3D virtual environment for ship–bridge collision monitoring is established to supplement the dataset. This environment generates multi-source data—including ship images, ship positions, and ship sizes—which are then integrated to construct a regression dataset for measuring ship dimensions.

(2) Domain mapping relationships between simulated and real data are established to reduce the domain gap. The pixel points and spatial coordinates of ships in the real-world data are transformed into the simulated space, which helps reduce the bias between the synthetic and real data.

(3) DL models are employed to exploit nonlinear relationships between the input variables (the spatial positions and 2D pixel information of the ships) and ship size parameters. The optimal DL model, as trained on the synthetic datasets, is generalized to predict ship sizes in real-world scenarios through supervised transfer learning.

2 Framework for ship size measurement

In this work, we aim to employ a DT-assisted framework for automatic prediction of ship sizes, so as to monitor for collisions between oversized ships and bridge girders. This framework is developed based on similarities between real-world and virtual spaces, as depicted in Fig. 1. In particular, the virtual space serves as a digital representation of the real-world space, emphasizing its role in simulating ship navigational environments across various dimensions and time scales.

As shown in Fig. 1, the framework can be broken down into the following steps:

I. Virtual modeling: A 3D digital scene depicting the bridge navigational environment is established based on building information modeling (BIM) (Volk et al., 2014) and Unity3D simulation (Liao et al., 2026). Synthetic images are generated through the virtual camera. Ship models with various positions and sizes are also created in the simulated space.

II. Ship monitoring: The You Only Look Once version 8 (YOLOv8) (Mu et al., 2023) machine learning algorithm is employed to detect ships on the synthetic images, extracting 2D pixel information of ships within the 3D simulated world.

III. Object matching: An object matching method is developed based on the mapping between the 3D scene and 2D images. Key pixel points of ships are subsequently extracted in order to match detected ships with synthetic ships in the scene.

IV. Regression modeling: The spatial positioning, dimensions, and 2D pixel information of ships in the images are fused to establish a synthetic regression dataset. This dataset is used to pre-train DL models, with the optimal model selected based on prediction accuracy.

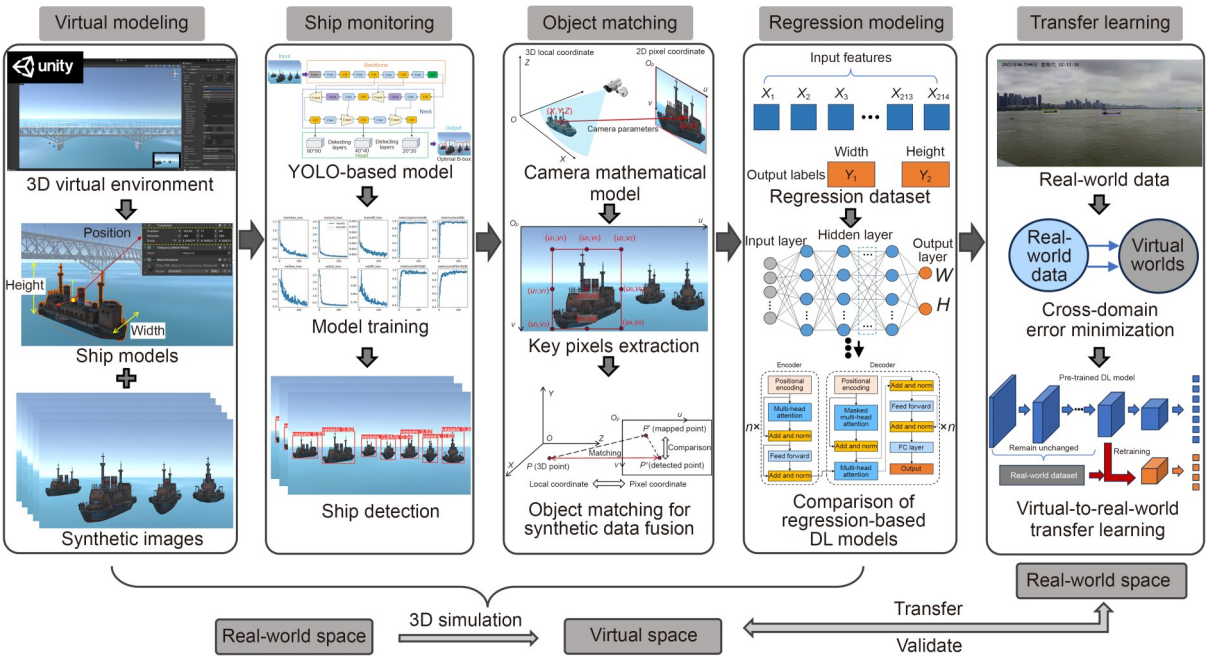

Fig. 1 Flowchart of the proposed framework. *W*: ship width; *H*: ship height

V. Transfer learning: A real-world dataset is constructed to match the regression dataset. The data bias between the synthetic space and the real-world is mitigated through coordinate system transformations. Then, virtual-to-real-world transfer learning is applied to predict ship sizes in the real-world.

Upon accurately measuring ship dimensions, bridge managers can then promptly identify ships that exceed the height or width limits, and issue remote warnings to prevent collisions.

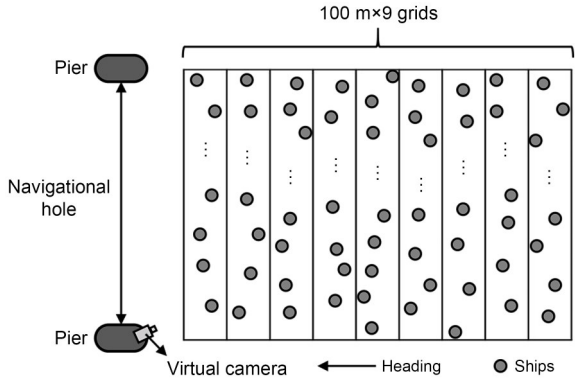

Fig. 2 Grid-based ship distribution

3 Methods

3.1 Synthetic environment establishment

The first step of this process is to acquire image datasets. The fully integrated professional game engine Unity3D has shown impressive performance in simulating 3D scenarios (Bynum et al., 2013). Hence, we employed it to build a reliable simulation environment for generating ship images. The bridge environment and camera were modeled first. The bridge environment is exported from a BIM model to Unity3D in '.fbx' format, with format transferring accomplished by an industry foundation classes-authoring software (Revit) (Liao et al., 2026). With the geometric and semantic resources imported, the physical properties required for the virtual camera and ship navigation modeling are established. This process is illustrated in Section S1 of the electronic supplementary materials (ESM).

For bridge collision early warning, it is considered sufficient to issue alerts for ships within 1000 m of the bridge (Zhang et al., 2022). Therefore, in the 3D synthetic scenario, ship models are systematically arranged in a grid between 100 and 1000 m from the bridge to generate a substantial number of samples. Most ships passing through bridge waterways have lengths under 100 m (Wu et al., 2019); thus, each grid has a length of 100 m and corresponds to the width of the navigational opening, resulting in nine grids total. A schematic diagram of the grid-based ship distribution is illustrated in Fig. 2.

Ship models are densely distributed within each grid, with their positions and sizes being randomized. The ships are permitted to move vertically within [0 m, 1 m] to simulate wave-induced motion, and their orientation angles are varied within [−15°, 15°] to reflect more realistic trajectories (Liao et al., 2025). The location and size of each ship are recorded by a C# script. Each grid is populated individually to minimise occlusion between ships in the simulated images. The random distribution of ships can be repeated multiple times until enough ship samples are obtained within each grid interval.

3.2 Fusion of synthetic data via ship detection and object matching

YOLOv8 is a widely used object detection model that has demonstrated excellent detection performance across various large-scale datasets (Wang et al., 2025). To illustrate the generalization of the proposed framework, YOLOv8 is adopted to extract the 2D pixel information of the ships (Mu et al., 2023). The structure of the YOLOv8 network is displayed in Fig. S3 of the ESM (Ma et al., 2024).

The alignment between a ship's 3D spatial and 2D pixel information can be established by mapping the relationship between the local and pixel coordinate systems (Yoon et al., 2018), as illustrated in Fig. S4 of the ESM.

The local coordinates of a ship represent the centroid of the ship model, denoted as $P(X, Y, Z)$, while the corresponding pixel coordinates (u, v) are located within the ship bounding box. The pinhole camera model is utilized to map between pixel and local coordinates (Xu et al., 2025), which is defined by:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{K}\begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{1}$$

where s is an arbitrary scale factor, $\boldsymbol{K}$ is the camera intrinsic matrix, $\boldsymbol{R}$ is the 3×3 rotation matrix, and $\boldsymbol{t}$ is

the 1×3 translation vector. The formulas for ***K***, ***R***, and ***t*** are provided in Section S3 of the ESM.

The bounding box of a ship contains numerous pixels. To enhance computational efficiency, it is essential to select a representative pixel to denote the pixel coordinates of the ship. Here, nine pixels are extracted from the bounding box to determine the optimal representative point, denoted as (u_p, v_p) (p=0, 1, 2, ⋯, 8), as shown in Fig. 3.

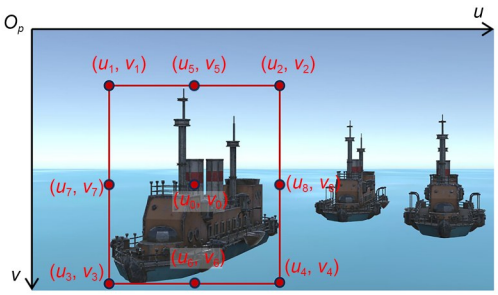

Fig. 3 Nine pixels extracted from a bounding box. References to color refer to the online version of this figure

The local coordinates of a ship in the computer-simulated scene are projected onto the image plane via the pinhole camera model to obtain $P^*(u^*, v^*)$. The YOLOv8 detector provides 2D bounding boxes, from which nine candidate pixels are extracted. Among these, the optimal representative point is selected and denoted as $P^o(u^o, v^o)$. The Euclidean distance between P^* and P^o is then used to associate 3D ships with their 2D detections. This matching process is illustrated in Fig. 4.

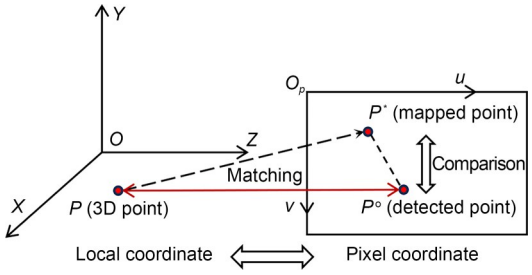

Fig. 4 Matching between *P* and P^o

P^* is solved to minimize the Euclidean distance from P^o, as the following objective function:

$$F(u, v) = \min\left[\sqrt{(u_m^o - u_n^*)^2 + (v_m^o - v_n^*)^2}\right], \quad (2)$$

where m and n represent the indices of the ship models, and $m \neq n$. It is assumed that the corresponding P at this point matches P^o. Subsequently, the fusion of ship sizes, positions, and 2D pixel information can be achieved.

3.3 Regression modeling for ship size

After ship detection and object matching, each ship is associated with corresponding 3D spatial information (spatial coordinates, width, and height) and 2D pixel information (pixel coordinates, pixel height, and pixel width). A regression dataset is then constructed, using the 2D pixel information and spatial coordinates of the ships as input factors, and their sizes as output variables. The composition of the regression dataset is illustrated in Fig. 5.

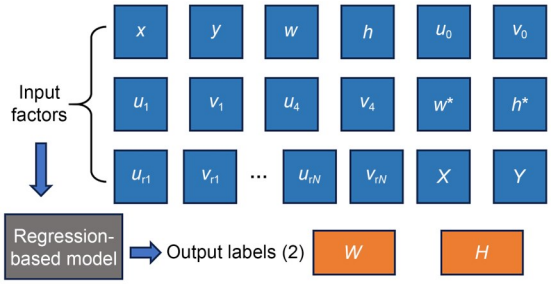

Fig. 5 Composition of the regression dataset

The selected inputs include the bounding box features of the ship (12 factors), a set of adaptively sampled pixels within each bounding box (depending on the box size), as well as the X and Y coordinates of the ship in the digital environment. In Fig. 5, (x, y) are the normalized center coordinates of the bounding box, (w, h) are the normalized width and height of the bounding box, (w^*, h^*) are the pixel width and height of the bounding box, (W, H) are the ship width and height, and (u_{ri}, v_{ri}) (i=1, 2, ⋯, N) represent pixels sampled within each bounding box. Given that farther-away ships occupy fewer pixels than closer ones in the same field of view, ships appear at different scales in the image. To maintain approximately uniform sampling density across these varying scales, the number of sampled pixels N is adaptively determined by:

$$N = \max\left(N_{\min}, \min\left(N_{\max}, \lceil whN_{\text{ref}} \rceil\right)\right), \quad (3)$$

where $N_{\min}$ is the minimum number of sampled pixels (set to 32) to ensure sufficient features for small-scale ships at far range, $N_{\max}$ is the maximum number of sampled pixels (set as 256) to prevent excessive sampling from large-scale ships and to control computational cost

(Gargoum et al., 2018), N_{ref} is the reference sampling number (set to 100), and $\lceil \cdot \rceil$ is the ceiling operator. This adaptive sampling strategy helps mitigate unfair density biases in the regression model. Both the real-world and synthetic data share the same format as this regression dataset.

DL algorithms, such as the multi-layer perceptron (MLP), convolutional neural network (CNN), and long short-term memory (LSTM), have been extensively applied for regression analysis in various fields (Ye et al., 2023; Zhang et al., 2025). Seven benchmark models, including MLP, CNN, LSTM, Bidirectional LSTM (BiLSTM), CNN-LSTM, CNN-BiLSTM, and Transformer, are selected for comparison (Zhai et al., 2025) (refer to Fig. S8 in Section S7 of the ESM). The model that performs the best on the synthetic dataset will be selected to make predictions of ship dimensions.

3.4 Ship size measurement through transfer learning

3.4.1 Cross-domain discrepancy minimization

Due to the challenge of ensuring identical intrinsic and extrinsic parameters between virtual and real cameras, as well as differences in coordinate systems between the virtual and real-world, domain discrepancies will exist between the real-world and virtual regression datasets. To mitigate this dataset bias, the pixel points and spatial coordinates of ships in the real-world data are transformed into the virtual space through domain mapping relationships. The transformation of pixel points is achieved through a cross-camera coordinate transformation. This process is defined by

$$\begin{bmatrix} u_{a\to s} \\ v_{a\to s} \\ 1 \end{bmatrix} = Z_s^{-1} \boldsymbol{K}_s \begin{bmatrix} \boldsymbol{R}_s & \boldsymbol{t}_s \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{R}_a & \boldsymbol{t}_a \\ 0 & 1 \end{bmatrix} \left(Z_a \boldsymbol{K}_a^{-1} \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} \right), \quad (4)$$

where $(u_{a\to s}, v_{a\to s})$ are coordinates transformed from the pixel coordinates (u_a, v_a) of the ship's bounding box in the real-world camera images to the pixel coordinate system of the virtual camera, Z_s and Z_a represent the depth values of the objects captured by the virtual and real-world cameras, respectively, $\boldsymbol{K}_s$ and $\boldsymbol{K}_a$ denote the intrinsic matrices, $\boldsymbol{R}_s$ and $\boldsymbol{R}_a$ correspond to the rotation matrices, and $\boldsymbol{t}_s$ and $\boldsymbol{t}_a$ represent the translation vectors of the virtual and real-world cameras, respectively.

Ship images are typically captured using bridge-mounted cameras. It is assumed that the positional coordinates of the real-world camera can be transformed into the local coordinate system of the synthetic space. In this case, the coordinates of the real-world camera are expected to coincide with those of the virtual camera, establishing a reference point pair for transforming data between the real and simulated worlds. The derivation of the formula for mapping real-world ship coordinates to the virtual environment is presented in Section S4 of the ESM.

3.4.2 Virtual-to-real-world transfer learning

Upon completing the cross-domain feature transformation, a real-world dataset can be constructed to match the regression dataset (Fig. 6) using 2D pixel information, spatial positions, and real-world ship sizes. Nevertheless, gathering such data is a labor-intensive and time-consuming process, and lack of adequate data could undermine the effectiveness of the DL models (Zhang et al., 2024). Transfer learning has emerged to mitigate the considerable dependency of DL models on data size and maximize the utilization of available data (Shen et al., 2020). Essentially, a well-trained model can serve as a starting point for a new task, especially when the amount of data required for this new task is small.

The optimal model trained on the synthetic regression dataset is selected to predict ship size in real-world scenarios. A portion of the real-world data is used to fine-tune the pre-trained model, while the remaining data is utilized as a test set to demonstrate the feasibility of virtual-to-real-world transfer learning. The transfer process is depicted in Fig. 6.

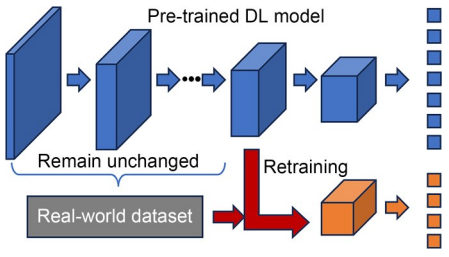

Fig. 6 Process of virtual-to-real-world transfer learning

4 Case study

4.1 Synthetic data generation and integration

A BIM model of a long-span bridge is constructed and imported into the Unity3D engine, with various scenes prepared for time progression of the simulation.

Within each scene, the objects, algorithms, 3D models, and lighting are designed to be adequately realistic (Bynum et al., 2013). Fig. 7 presents the 3D simulated environment, which was developed to simulate a bridge, ships, cameras, channels, and aquatic areas.

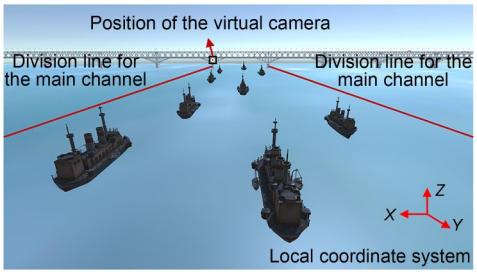

Fig. 7 3D simulated environment

In the real-world, factors such as wiring connections and installation space limitations make the area above bridge piers the most suitable for camera installation (Liao et al., 2026). Consequently, in the simulated environment, the virtual camera is positioned above the bridge piers. The parameters of the virtual camera are defined in Table 1.

To balance data quantity, computational resources, and processing time, 300 synthetic images were generated, which contained 1350 ships of various dimensions and positions for virtual pre-training. Image samples of synthetic ships at varying camera-to-ship distances, together with the corresponding YOLOv8 detection results, are provided in Section S5 of the ESM.

According to the virtual camera parameters in the Unity3D game engine, $\boldsymbol{K}$, $\boldsymbol{R}$, and $\boldsymbol{t}$ can be solved for. The mapping relationship between the 3D world coordinates and 2D pixel coordinates is written as:

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}2482 & 0 & 960\\0 & -7784 & 540\\0 & 0 & 1\end{bmatrix}\times\begin{bmatrix}-0.961 & 0 & 0.276 & 33.9054\\0 & 1 & 0 & -22\\-0.276 & 0 & 0.961 & -21.5\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix}, \quad (5)$$

from which the pixel point P^* of each ship can be calculated.

After this, using Eq. (5), points P^* with the minimum Euclidean distance from P^o are determined to match with P. This calculation is performed individually on the nine extracted pixels. Fig. S7 of the ESM presents a scatter plot of the Euclidean distance distribution for the nine pixels following target matching. Among them, (u_4, v_4) emerges as the optimal representative 2D pixel point of a ship. A simple comparison of the Euclidean distance between (u_4, v_4) and P^* for each ship facilitates the alignment of its 3D spatial and 2D pixel information. Each ship is manually inspected in the Unity3D settings, and the method is found to be effective in matching detected ships with synthetic ships in the scene. Then, multiple data sources can be integrated, including ship images, dimensions, and positions. As a result, each ship is associated with its spatial position, 2D pixel information, and size, which collectively establish a regression dataset comprising 1350 ship samples, following the format presented in Fig. 5.

4.2 Pre-training of regression models using the synthetic dataset

As mentioned previously, seven benchmark DL models are selected for comparison (Zhang et al., 2025); note that their corresponding model parameter configurations are detailed in Section S7 of the ESM. Four criteria, mean absolute error (MAE), root mean square error (RMSE), mean absolute percent error (MAPE), and coefficient of determination (R^2)—are utilised to evaluate the performance of the different models (Nguyen et al., 2019). Smaller MAE, RMSE, and MAPE values indicate higher measurement accuracy of the ship sizes. The R^2 index, ranging between 0 and 1, measures the goodness of fit of the regression model. If $R^2 \geqslant 0.90$, the measurement is considered very acceptable; if $0.60 < R^2 < 0.90$, it is acceptable; if $R^2 \leqslant 0.60$, it is unacceptable (Li et al., 2018). The model that exhibits the best performance on the regression

Table 1 Parameters of the virtual camera

x_c (m)	y_c (m)	z_c (m)	V_H (°)	V_V (°)	f_x (pixels)	f_y (pixels)	c_x (pixels)	c_y (pixels)
−33.905	22	21.5	40.598	30.372	2482	−7784	960	540

(x_c, y_c, z_c) denote the local coordinates of the installation position for the virtual camera, V_H represents the horizontal field of view of the virtual camera, V_V represents the vertical field of view, f_x and f_y are the focal lengths of the camera in the horizontal and vertical directions, respectively, and (c_x, c_y) is the principal point, which is the point where the optical axis intersects the image plane

dataset will be selected. Formulas for these metrics are provided in Section S7 of the ESM. Fig. S9 of the ESM describes the distribution of actual values, predicted values, and their relative errors for ship sizes across the test dataset for the seven models. The MAE, RMSE, MAPE, and R^2 of these models are compared in Fig. S10 of the ESM.

4.3 Real-world ship size measurement

In this study, the open-source FVessel dataset (Guo et al., 2023), which provides known ship sizes, positions, and 2D pixel information through the fusion of camera and AIS data, is used to validate the effectiveness of the proposed framework. This dataset integrates ship videos captured from bridge-mounted cameras and AIS data, providing the 2D pixel information of ships in the videos, along with their spatial positions (Guo et al., 2023). The maritime mobile service identification (MMSI) from the AIS data allows querying of ship width information registered with maritime authorities (Li et al., 2018; Ren et al., 2022). As a result, this open-source dataset enables rapid acquisition of influential factors and ship size parameters required for the regression task. Three sets of videos (Video 1: 607 s in length, Video 2: 1192 s, Video 3: 677 s), along with their corresponding AIS data, are used to supplement the real-world dataset. A detailed description of the real-world dataset is provided in Section S8 of the ESM.

The number of input variables remains unchanged, while the output labels are set to one, representing the actual width of the ships in the videos. Consequently, the final fully connected layer of the CNN-LSTM is adjusted to output a single dimension. The initial layers of the pre-trained CNN-LSTM are frozen, with only the final layers being updated. Additionally, the learning rate is reduced to one-tenth of the original (i.e., 0.001), and the model is fine-tuned using the stochastic gradient descent (SGD) optimizer (Liao et al., 2025). To compare the performance of the fine-tuned model with the original CNN-LSTM, the real-world data is also used to train the latter. Both sets of models are trained for 200 epochs, employing the same training and testing data split, with 80% of the data for training and 20% for testing. The MAE, RMSE, MAPE, and R^2 of these two models are compared in Fig. S11 of the ESM. Measurement results of the transfer learning model at varying camera-to-ship distances are provided in Table S4 of the ESM.

Even at far camera-to-ship distances, ranging from 300 to 1000 m, the MAPE is 3.74%. This indicates that the proposed regression-based approach effectively mitigates the adverse impact of long camera-to-ship distances on accuracy. For waterway ships, whose widths and heights typically exceed several tens of metres, this level of error is considerably smaller than the horizontal and vertical clearances (often 1–2 m or more) commonly adopted by bridge authorities (Wu et al., 2019; Zhang et al., 2022). This indicates that the proposed method shows potential for real-world collision-prevention applications. The relative error distributions between the predicted and actual values on the test data, for both the fine-tuned model and the original CNN-LSTM, are compared in Fig. 8.

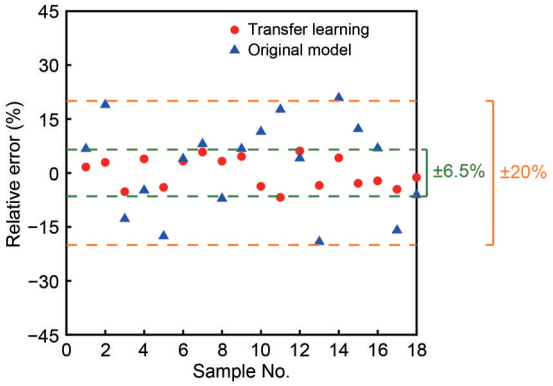

Fig. 8 Distribution of relative error between the predicted and actual values

Fig. 8 illustrates that the relative error in ship width predicted by the fine-tuned model is significantly smaller than that predicted by the original CNN-LSTM. Specifically, the former's relative error mostly falls within −6.5% to 6.5%, whereas the latter predominantly falls between −20% and 20%. Overall, the transfer learning method effectively overcomes the domain gap between the virtual and real-world environments, enabling accurate ship size estimations even with limited real-world data.

5 Conclusions

In this work, we proposed a DT-assisted automatic framework for estimation of ship dimensions, so as to monitor for potential collisions between oversized ships and bridge girders. Specifically, a 3D virtual environment was first established for generating multi-data

sources, including ship images, dimensions, and positions. Afterwards, ship detection was executed on the images, followed by matching of ships in images with synthetic ships in the scene. The spatial positions, 2D pixel information, and dimensions of the ships were then fused to create a regression dataset for pre-training DL models. The pre-trained models were compared in performance to identify the optimal model type. Coordinate system transformations were applied to address dataset bias between the simulated and real-world. A portion of the real-world data was subsequently used to fine-tune the optimal pre-trained model, while the remaining data served as a test set to demonstrate the feasibility of virtual-to-real-world transfer learning.

The YOLOv8 model effectively extracts 2D pixel information of ships from the images, and this information is matched with their 3D spatial coordinates to serve as predictors for regression modeling of ship size. The DL models are found to accurately predict ship sizes, with the best-performing model selected from comparative experiments on the synthetic dataset, and then applied to real-world ship size measurement through transfer learning. The virtual-to-real-world transfer learning framework reduces the data discrepancy between the simulated and real environments and boosts the measurement accuracy across varying camera-to-ship distances, even with limited real-world data.

Despite the success of the proposed prediction framework, a few limitations remain and could be addressed in future work. For one, open-source datasets on ship height have been rarely reported. As a result, the size prediction framework has only been validated using ship width data, while its height prediction capabilities have not yet been fully validated. Future efforts will focus on obtaining reliable field measurement data for ship height using total stations and laser scanning devices. These measurements will be incorporated into regression models to enable comprehensive validation of the height estimation. Moreover, the current 3D proxy environment does not account for a broad range of ship categories and environmental variations. To make the synthetic data more realistic and better reduce the gap between the simulated and real-world, more environmental factors will be integrated, promoting the interaction of digital twins with real-world scenarios.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 52338011 and 52108274) and the Start-up Research Fund of Southeast University (No. RF1028624058), China. The first author would also like to acknowledge the support from the SEU Innovation Capability Enhancement Plan for Doctoral Students (No. CXJH_SEU 26112), China.

Author contributions

Ruixuan LIAO: writing–original draft, visualization, validation, methodology, and conceptualization. Yiming ZHANG: writing–review & editing, writing–original draft, funding acquisition, and supervision. Hao WANG: writing–review & editing, funding acquisition, and formal analysis. Jianxiao MAO: investigation and funding acquisition. Aoyang LI: software and validation. Zhengyi CHEN: formal analysis.

Conflict of interest

Ruixuan LIAO, Yiming ZHANG, Hao WANG, Jianxiao MAO, Aoyang LI, and Zhengyi CHEN declare that they have no conflict of interest.

References

Bian XC, Jiang HG, Chen YM, 2010. Accumulative deformation in railway track induced by high-speed traffic loading of the trains. *Earthquake Engineering and Engineering Vibration*, 9(3):319-326. https://doi.org/10.1007/s11803-010-0016-2

Bian XC, Shi KH, Li W, et al., 2021. Quantification of railway ballast degradation by abrasion testing and computer-aided morphology analysis. *Journal of Materials in Civil Engineering*, 33(1):04020411. https://doi.org/10.1061/(ASCE)MT.1943-5533.0003519

Bynum P, Issa RRA, Olbina S, 2013. Building information modeling in support of sustainable design and construction. *Journal of Construction Engineering and Management*, 139(1):24-34. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000560

Chan PW, Lee YF, 2013. Performance of LIDAR- and radar-based turbulence intensity measurement in comparison with anemometer-based turbulence intensity estimation based on aircraft data for a typical case of terrain-induced turbulence in association with a typhoon. *Journal of Zhejiang University-SCIENCE A*, 14(7):469-481. https://doi.org/10.1631/jzus.A1200236

Chen YM, Ke H, Fredlund DG, et al., 2010. Secondary compression of municipal solid wastes and a compression model for predicting settlement of municipal solid waste landfills. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(5):706-717. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000273

Fan W, Yuan WC, Fan QW, 2008. Calculation method of ship collision force on bridge using artificial neural network. *Journal of Zhejiang University-SCIENCE A*, 9(5):614-623. https://doi.org/10.1631/jzus.A071556

Fan W, Sun Y, Yang CC, et al., 2020. Assessing the response and fragility of concrete bridges under multi-hazard effect of vessel impact and corrosion. *Engineering Structures*, 225:111279.

https://doi.org/10.1016/j.engstruct.2020.111279

Fascione JM, Crews RT, Wrobel JS, 2014. Association of footprint measurements with plantar kinetics: a linear regression model. *Journal of the American Podiatric Medical Association*, 104(2):125-133.
https://doi.org/10.7547/0003-0538-104.2.125

Gaidon A, Wang Q, Cabon Y, et al., 2016. VirtualWorlds as proxy for multi-object tracking analysis. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, p.4340-4349.
https://doi.org/10.1109/CVPR.2016.470

Gajalakshmi K, Palanivel S, Nalini NJ, et al., 2017. Grain size measurement in optical microstructure using support vector regression. *Optik*, 138:320-327.
https://doi.org/10.1016/j.ijleo.2017.03.052

Gargoum SA, Karsten L, El-Basyouny K, et al., 2018. Automated assessment of vertical clearance on highways scanned using mobile LiDAR technology. *Automation in Construction*, 95:260-274.
https://doi.org/10.1016/j.autcon.2018.08.015

Guo J, He JX, 2020. Dynamic response analysis of ship-bridge collisions experiment. *Journal of Zhejiang University-SCIENCE A*, 21(7):525-534.
https://doi.org/10.1631/jzus.A1900382

Guo Y, Liu RW, Qu JX, et al., 2023. Asynchronous trajectory matching-based multimodal maritime data fusion for vessel traffic surveillance in inland waterways. *IEEE Transactions on Intelligent Transportation Systems*, 24(11):12779-12792.
https://doi.org/10.1109/TITS.2023.3285415

Hou CC, Wang H, Guan W, et al., 2025. Road pavement performance prediction using a time series long short-term memory (LSTM) model. *Journal of Zhejiang University-SCIENCE A*, 26(5):424-437.
https://doi.org/10.1631/jzus.A2300643

Huang QL, Wang JZ, Song YX, et al., 2024. Synthetic-to-realistic domain adaptation for cold-start of rail inspection systems. *Computer-Aided Civil and Infrastructure Engineering*, 39(3): 424-437.
https://doi.org/10.1111/mice.13087

Ibrahim A, Artama WT, Budisatria IGS, et al., 2021. Regression model analysis for prediction of body weight from body measurements in female Batur sheep of Banjarnegara District, Indonesia. *Biodiversitas Journal of Biological Diversity*, 22(7):2723-2730.
https://doi.org/10.13057/biodiv/d220721

Inazu D, Waseda T, Hibiya T, et al., 2016. Assessment of GNSS-based height data of multiple ships for measuring and forecasting great tsunamis. *Geoscience Letters*, 3(1):25.
https://doi.org/10.1186/s40562-016-0059-y

Iuzzolino ML, Walker ME, Szafir D, 2018. Virtual-to-real-world transfer learning for robots on wilderness trails. IEEE/RSJ International Conference on Intelligent Robots and Systems, p.576-582.
https://doi.org/10.1109/IROS.2018.8593883

Li BY, Liu B, Guo WW, et al., 2018. Ship size extraction for Sentinel-1 images based on dual-polarization fusion and nonlinear regression: push error under one pixel. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8): 4887-4905.
https://doi.org/10.1109/TGRS.2018.2841882

Liao RX, Wu T, Zhang YM, et al., 2024. Vision-based vessel detection for vessel-bridge collision warnings under complex scenes. *Journal of Southeast University (English Edition)*, 40(1):33-40.
https://doi.org/10.3969/j.issn.1003-7985.2024.01.004

Liao RX, Zhang YM, Wang H, et al., 2025. An effective ship detection approach combining lightweight networks with supervised simulation-to-reality domain adaptation. *Computer-Aided Civil and Infrastructure Engineering*, 40(27): 4732-4757.
https://doi.org/10.1111/mice.13501

Liao RX, Zhang YM, Wang H, et al., 2026. Multi-objective optimisation of surveillance camera placement for bridge-ship collision early-warning using an improved non-dominated sorting genetic algorithm. *Advanced Engineering Informatics*, 69:103918.
https://doi.org/10.1016/j.aei.2025.103918

Lu LJ, Dai F, 2023. Automated visual surveying of vehicle heights to help measure the risk of overheight collisions using deep learning and view geometry. *Computer-Aided Civil and Infrastructure Engineering*, 38(2):194-210.
https://doi.org/10.1111/mice.12842

Lu XT, Zhang WX, Xu L, et al., 2023. A lateral pressure prediction model for bottom-up pumping of SCC in large-diameter steel tubes based on Bernoulli's Principle. *Case Studies in Construction Materials*, 19:e02470.
https://doi.org/10.1016/j.cscm.2023.e02470

Ma HP, Zhang YJ, Sun SY, et al., 2024. Weighted multi-error information entropy based you only look once network for underwater object detection. *Engineering Applications of Artificial Intelligence*, 130:107766.
https://doi.org/10.1016/j.engappai.2023.107766

Mu ZH, Qin Y, Yu CC, et al., 2023. Adaptive cropping shallow attention network for defect detection of bridge girder steel using unmanned aerial vehicle images. *Journal of Zhejiang University-SCIENCE A*, 24(3):243-256.
https://doi.org/10.1631/jzus.A2200175

Nguyen T, Kashani A, Ngo T, et al., 2019. Deep neural network with high-order neuron for the prediction of foamed concrete strength. *Computer-Aided Civil and Infrastructure Engineering*, 34(4):316-332.
https://doi.org/10.1111/mice.12422

Ni FT, Zhang J, Chen ZQ, 2019. Zernike-moment measurement of thin-crack width in images enabled by dual-scale deep learning. *Computer-Aided Civil and Infrastructure Engineering*, 34(5):367-384.
https://doi.org/10.1111/mice.12421

Pallotta G, Vespe M, Bryan K, 2013. Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction. *Entropy*, 15(6):2218-2245.
https://doi.org/10.3390/e15062218

Pedersen PT, Chen J, Zhu L, 2020. Design of bridges against ship collisions. *Marine Structures*, 74:102810.
https://doi.org/10.1016/j.marstruc.2020.102810

Ren YB, Li XF, Xu H, 2022. A deep learning model to extract ship size from Sentinel-1 SAR images. *IEEE Transactions*

on Geoscience and Remote Sensing, 60:5203414. https://doi.org/10.1109/TGRS.2021.3063216

Sazonov KE, 2011. Navigation challenges for large-size ships in ice conditions. *Ships and Offshore Structures*, 6(3):231-238. https://doi.org/10.1080/17445302.2010.548123

Sha YY, Amdahl J, Liu K, 2019. Design of steel bridge girders against ship forecastle collisions. *Engineering Structures*, 196:109277. https://doi.org/10.1016/j.engstruct.2019.109277

Shao YC, Jin YB, Huang ZL, et al., 2024. A learning-based control pipeline for generic motor skills for quadruped robots. *Journal of Zhejiang University-SCIENCE A*, 25(6): 443-454. https://doi.org/10.1631/jzus.A2300128

Shen S, Sadoughi M, Li M, et al., 2020. Deep convolutional neural networks with ensemble learning and transfer learning for capacity estimation of lithium-ion batteries. *Applied Energy*, 260:114296. https://doi.org/10.1016/j.apenergy.2019.114296

Urazghildiiev I, Ragnarsson R, Ridderstrom P, et al., 2007. Vehicle classification based on the radar measurement of height profiles. *IEEE Transactions on Intelligent Transportation Systems*, 8(2):245-253. https://doi.org/10.1109/TITS.2006.890071

Volk R, Stengel J, Schultmann F, 2014. Building information modeling (BIM) for existing buildings—literature review and future needs. *Automation in Construction*, 38:109-127. https://doi.org/10.1016/j.autcon.2013.10.023

Wang JD, Ma XL, Zhu XH, et al., 2025. Kinematic modeling and stability analysis for a wind turbine blade inspection robot. *Journal of Zhejiang University-SCIENCE A*, 26(2): 121-137. https://doi.org/10.1631/jzus.A2300619

Wang LL, Yang LM, Huang DJ, et al., 2008. An impact dynamics analysis on a new crashworthy device against ship-bridge collision. *International Journal of Impact Engineering*, 35(8):895-904. https://doi.org/10.1016/j.ijimpeng.2007.12.005

Wu B, Yip TL, Yan XP, et al., 2019. Fuzzy logic based approach for ship-bridge collision alert system. *Ocean Engineering*, 187:106152. https://doi.org/10.1016/j.oceaneng.2019.106152

Xu HR, Yin JN, Zhang N, 2025. Transformer-based deformation measurement of underground structures from a single-camera video. *Automation in Construction*, 172:106070. https://doi.org/10.1016/j.autcon.2025.106070

Ye XW, Jin T, Ang PP, et al., 2021. Computer vision-based monitoring of the 3-D structural deformation of an ancient structure induced by shield tunneling construction. *Structural Control and Health Monitoring*, 28(4):e2702. https://doi.org/10.1002/stc.2702

Ye XW, Zhang XL, Zhang HQ, et al., 2023. Prediction of lining upward movement during shield tunneling using machine learning algorithms and field monitoring data. *Transportation Geotechnics*, 41:101002. https://doi.org/10.1016/j.trgeo.2023.101002

Yoon H, Shin J, Spencer Jr BF, 2018. Structural displacement measurement using an unmanned aerial system. *Computer-Aided Civil and Infrastructure Engineering*, 33(3):183-192. https://doi.org/10.1111/mice.12338

Yurdakul O, Küçüksu GN, Saydam AZ, et al., 2021. A decision-making process for the selection of better ship main dimensions by a Pareto frontier solution. *Ocean Engineering*, 239:109908. https://doi.org/10.1016/j.oceaneng.2021.109908

Zhai GH, Xu YJ, Spencer BF, 2025. Bidirectional graphics-based digital twin framework for quantifying seismic damage of structures using deep learning networks. *Structural Health Monitoring*, 24(1):86-110. https://doi.org/10.1177/14759217241231129

Zhang L, Chen PF, Li MX, et al., 2022. A data-driven approach for ship-bridge collision candidate detection in bridge waterway. *Ocean Engineering*, 266:113137. https://doi.org/10.1016/j.oceaneng.2022.113137

Zhang YM, d'Avigneau AM, Hadjidemetriou GM, et al., 2024. Bayesian dynamic modelling for probabilistic prediction of pavement condition. *Engineering Applications of Artificial Intelligence*, 133:108637. https://doi.org/10.1016/j.engappai.2024.108637

Zhang YM, Li HQ, Wang H, 2025. Data-driven wind-induced response prediction for slender civil infrastructure: progress, challenges and opportunities. *Structures*, 74:108650. https://doi.org/10.1016/j.istruc.2025.108650

Electronic supplementary materials

Sections S1–S9, Tables S1–S4, Figs. S1–S12