TRANSNAV
http://www.transnav.eu
the International Journal
on Marine Navigation
and Safety of Sea Transportation
Volume 20
Number 2
June 2026
DOI: 10.12716/1001.20.02.12

Designing a YOLOv8n-based Video Surveillance System for Ship Aspect Recognition

O. Pashenko, O. Pipchenko & V. Shevchenko
National University "Odesa Maritime Academy", Odesa, Ukraine

ABSTRACT: According to the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs), vessels must maintain an effective lookout by sight, hearing, and all available technical means, taking into account prevailing circumstances and environmental conditions. In this context, automated visual surveillance systems represent an important enhancement of conventional navigational tools, enabling early detection and interpretation of surrounding vessel behaviour.
This study proposes a video surveillance system based on the lightweight YOLOv8n deep learning model for the detection and classification of eight vessel aspects. The model was trained on an initial dataset of 925 images, which was further expanded to 1,742 annotated images. The dataset was designed to reflect real maritime operating conditions, including different times of day, weather scenarios, vessel categories, and geographical regions. To improve robustness, data augmentation techniques such as colour space transformations, geometric modifications, and classification-specific augmentations were applied. Class imbalance was mitigated through the use of class weighting.
The paper also describes the system architecture and camera configuration, providing effective surveillance coverage up to 6 nautical miles. The proposed approach enables not only vessel aspect recognition but also the estimation of relative and true bearings, thereby contributing to improved situational awareness and collision avoidance in maritime navigation.

1 INTRODUCTION

Automated ship aspect recognition is a critical capability for modern maritime monitoring and navigation systems, particularly in the context of autonomous and unmanned shipping [1]. Under the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs) [2], vessels are required to maintain a proper lookout by sight and hearing, as well as by all available means appropriate to the prevailing circumstances and conditions. For crewed ships, this responsibility is fulfilled by human watchkeepers; for autonomous vessels, the challenge lies in replicating this situational awareness through reliable sensing and perception systems [3].

A robust computer vision system capable of detecting and classifying ship aspects under diverse environmental conditions plays a vital role in meeting this requirement. Such a system can enhance collision avoidance, improve traffic management, and support compliance with COLREGs by enabling early and accurate identification of other vessels' aspects and likely maneuvers. This is particularly important in congested waterways, or when integrating autonomous vessels into mixed traffic with conventionally operated ships.

Recent advances in deep learning-based object detection, especially in the YOLO (You Only Look Once) family of models, have shown strong potential for high-speed, high-accuracy detection in resource-constrained environments [4-6]. Among them, YOLOv8n offers a compact architecture that balances detection precision with computational efficiency, making it well-suited for deployment on vessels with limited onboard processing capabilities [7-10].

A number of studies focus on the development of datasets and the training of dedicated computer vision models. YOLO-based models have been developed to enable the recognition of various vessel classes, floating debris, or to support search and rescue operations [8-12]. In the work [13] M. E. Lopez use both real photos and a synthetically generated dataset. However, this work is of a survey nature and does not contain specific implementations of aspect recognition or hardware integration. In addition, B. Kiefer in his work [14] presents the results of testing under real-world conditions. However, the primary objective of the study is the estimation of the distance to the object rather than the determination of its aspect.

A number of studies are devoted to bearing estimation. In their work [15], Gorad et al. employ the outdated YOLOv5 architecture in combination with thermal and optical data as well as a Kalman filter to estimate vessel bearing. M. E. López [13] does not take practical operational aspects into account.

Despite the significant volume of existing research, each study focuses on its own algorithms for solving the given task, taking into account maritime standards.

2 DATASET FOR TRAINING THE COMPUTER VISION MODEL

In the initial configuration, a dedicated image dataset was assembled to support the training of the computer vision model. The dataset covers eight ship aspect classes: forward, port bow, port side, starboard bow, starboard quarter, starboard side, stern, and port quarter (Fig. 1).

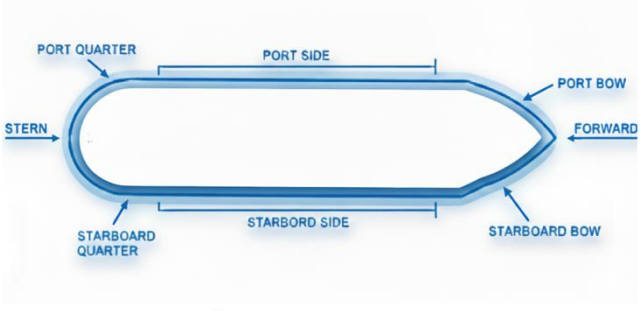

Figure 1. Ship aspects

Image selection was guided by factors essential for robust model performance, including variation in time of day (day, morning, and evening), challenging meteorological conditions (precipitation, storms), scale differences, partial occlusions, and maximum diversity in vessel types and camera perspectives (Fig. 2). The dataset encompasses imagery from a wide range of geographical regions, ensuring global representativeness. Each image has a minimum resolution of 700 × 700 pixels, providing sufficient detail and quality for accurate feature extraction and subsequent model training.

For training purposes, the objects in the dataset images were divided into three subsets: training data (70%), validation data (20%), and testing data (10%). Such proportions are standard in machine learning tasks for computer vision, as they allow allocating the majority of the data to model training. The distribution is approximate, since the training data depict real-world maritime scenes and may contain multiple vessels within a single image, making exact proportional splitting impossible. The actual proportions may slightly differ due to rounding or the need to preserve image integrity. The initial dataset contained 925 images, representing 1931 annotated objects (Table 1). This configuration was used in the early stages of model development.

Table 1. The number of objects in the updated dataset

Class №	Class name	Training	Validation	Testing
0	forward	105	26	17
1	port bow	164	42	19
2	port side	372	112	52
3	starboard bow	153	29	17
4	starboard quarter	109	33	13
5	starboard side	273	94	41
6	stern	72	17	6
7	port quarter	110	21	16
	Total	1358	374	181

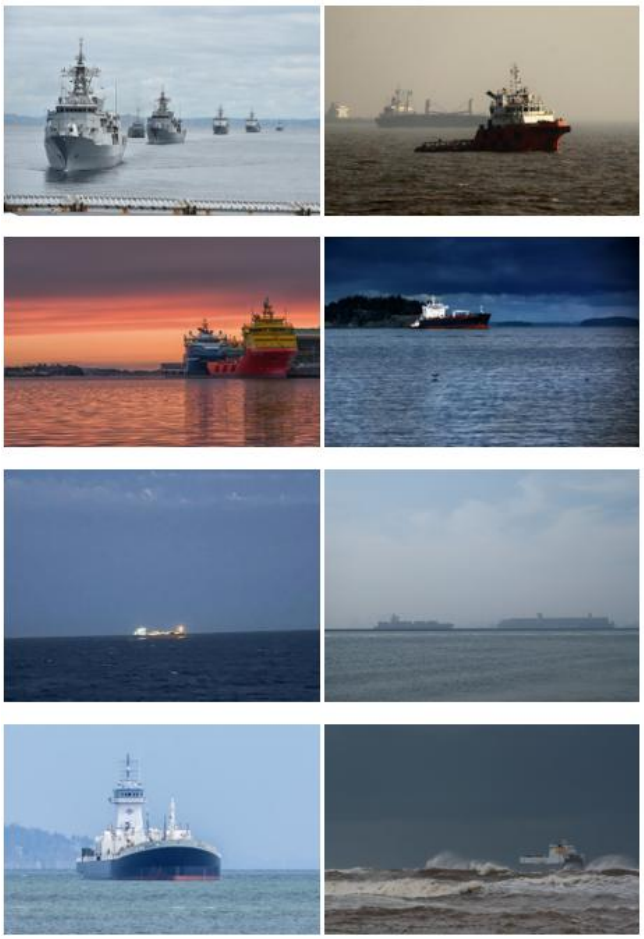

Figure 2. Example of the training dataset

However, during the initial training cycles, the model exhibited two main types of errors: false positives, where the model incorrectly detected an

object of a certain class in an image when none was actually present; and false negatives, where the model failed to detect objects that were indeed present in the image. Samples of false negatives (highlighted with yellow boxes) and false positives (highlighted with red boxes) are shown in Fig. 3. The model frequently confused the surrounding environment with actual targets, so the training dataset was subsequently augmented with negative examples representing the diversity of the environment and port infrastructure. The presence of false negatives in clear images indicates a significant class imbalance, leading to the neglect of classes represented by fewer objects (port quarter, stern, forward). The dataset in this configuration was then used from the 1 to the 4 training stages [16].

Figure 3. Examples of false negatives (highlighted with yellow boxes) and false positives (highlighted with red boxes)

To address these issues, the dataset was expanded to 1742 images, including 115 negative samples (Table 2) [17]. This enhancement increased environmental diversity and reduced the model's tendency toward misclassification.

Table 2. The number of objects in the updated dataset

Class №	Class name	Training	Validation	Testing
0	forward	266	70	29
1	port bow	291	77	64
2	port side	494	166	63
3	starboard bow	285	66	37
4	starboard quarter	223	65	33
5	starboard side	443	84	50
6	stern	150	49	22
7	port quarter	213	55	32
	negative samples	68	33	14
	Total	2433	665	344

It is worth noting that, despite the expansion of the dataset, a significant imbalance in the number of examples across certain classes remained. In particular, the most frequent class, port side (494 examples), contained more than three times as many samples as the least represented class, stern (150 examples). This imbalance tends to bias the model toward overfitting on the more prevalent classes while neglecting the underrepresented ones, which in turn leads to a deterioration in overall model performance.

To compensate for this imbalance, class weights were applied relative to the most represented class (port side), which contains the highest number of examples (494). Specifically, for each class, a weight coefficient inversely proportional to the number of examples in that class was calculated, according to the formula:

$$w_i = \frac{N_{\max}}{N_i} \quad (1)$$

where w_i – the weight for class i, N_{max} – the maximum number of objects among all classes, and N_i – the number of objects in class i.

The resulting class weight values for classes 0 through 7 were as follows: 1.86, 1.70, 1.00, 1.73, 2.22, 1.11, 3.29, 2.32. The dataset in this configuration was then used from the 5 to the 9 training stages.

3 SOFTWARE USED FOR MODEL TRAINING

As the foundation for the proposed ship aspect recognition system, the authors selected the lightweight YOLOv8n model (You Only Look Once version 8 – nano) as the primary tool for detecting and classifying objects in video streams obtained from onboard cameras. This choice was motivated by the model's high detection speed under limited computational resources, making it suitable for deployment on vessels. Despite its compact size, YOLOv8n demonstrates competitive accuracy in object detection tasks, which is critical for real-time operation [4,5]. Like other models in the YOLOv8 series, YOLOv8n supports full fine-tuning on a custom dataset. Its compact architecture also enables faster training even when hardware resources are constrained.

In this study, model training and performance evaluation were conducted in the cloud-based environment Google Colaboratory (Google Colab), which provides a Python 3 execution environment with GPU acceleration. An NVIDIA T4 GPU was used, featuring 2,560 CUDA cores, 16 GB of GDDR6 memory, and a memory bandwidth of up to 320 GB/s. These capabilities allow efficient deep learning computations, significantly reducing the time required for model training.

4 MODEL TRAINING PROCESS

To develop the final computer vision model for ship aspect recognition, the training process was conducted in several stages. After each stage, the obtained results were analysed, and a decision was made regarding their further use. Some stages were deemed unsuccessful due to a deterioration in the model's performance metrics.

During the study, various data augmentation techniques [18-20] were compared, specifically transformations of the training data aimed at increasing the complexity of the learning process. It is important to note that, despite the variety of available augmentations, the choice of a specific augmentation type depends on the recognition task. In our case, the model is designed to recognize a ship's aspect in a maritime environment. Thus, knowledge of the specific visual characteristics of maritime imagery – such as water surface wave patterns, meteorological precipitation, and variations in lighting between day and night – is essential for selecting effective transformation methods.

In this work, the authors propose three groups of augmentations: color space augmentations, geometric transformations, and classification-specific transformations [18]. Other standard augmentations – including rotation, mirroring, scaling, and compositional transformations – were excluded because they may alter the spatial structure or orientation of the object, which is critical in orientation recognition tasks. For example, mirroring an object is inappropriate for training a computer vision model to determine a ship's aspect, as it completely distorts the orientation-related features and undermines the validity of the results.

In the training process, color space augmentations were applied in the form of hue adjustment, saturation adjustment, and brightness adjustment (Fig. 4). They are particularly effective under varying natural lighting conditions, where the appearance of objects may change significantly. They enable the model to better adapt to diverse environments, including changes in weather or camera settings.

Figure 4. Examples of color space augmentations

The selected geometric and classification-specific transformation methods include translation, scale, shear, perspective, and random erasing (Fig. 5). These augmentations enable the model to effectively adapt to changes in the object's position within the frame, which is typical when filming from onboard a vessel. Such transformations simulate changes in the vessel's appearance within the camera's field of view caused by wave motion, maneuvering, or variations in the observation angle, thereby ensuring more robust recognition regardless of position.

In addition to selecting the appropriate types of augmentation, it is equally important to determine suitable parameter values. These values must be realistic and reflect the natural variations of the environment and objects [21].

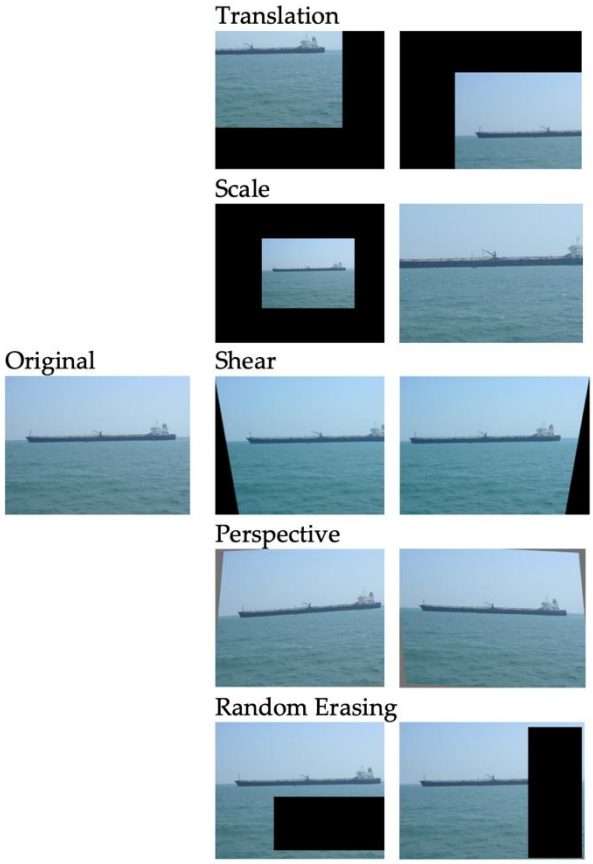

Figure 5. Examples of geometric and specific transformations

5 TRAINING RESULTS

The training results of the custom computer vision model designed to recognize a ship's aspect are presented in Table 3. It reports the results for the best training epoch, determined by the mean Average Precision (mAP) metric at different Intersection over Union (IoU) thresholds – specifically mAP@0.5 and mAP@0.5:0.95 – which are widely accepted performance indicators in object detection tasks. The training process lasted for a total of 445 epochs (Fig.6).

Table 3. Results obtained after training the computer vision model

Stage	Total epochs	Epoch	Box loss	Classification loss	Distribution focal loss	Precision	Recall	mAP50	mAP50-95
1	100	78	0.66735	0.88926	0.92138	0.28878	0.51674	0.37481	0.30066
2	20	19	0.59730	0.64540	0.88461	0.36056	0.43925	0.36487	0.29669
3	20	18	0.00428	0.67594	0.92456	0.36422	0.42528	0.35009	0.28509
4	25	25	0.00450	0.69631	0.93709	0.39640	0.41739	0.37135	0.30399
5	60	56	0.00276	1.30775	0.78449	0.59788	0.56667	0.59847	0.47747
6	60	57	0.00297	1.32399	0.81350	0.63790	0.59754	0.64198	0.51513
7	60	59	0.00293	1.01176	0.80043	0.65852	0.63970	0.66873	0.54280
8	50	28	0.00265	0.84435	0.76718	0.67918	0.64450	0.67282	0.54766
9	50	17	0.00477	0.73619	0.68824	0.70344	0.60982	0.67467	0.54036

Figure 6. Examples of ship aspect recognition using the proposed YOLOv8n computer vision model. Labels indicate predicted orientation and confidence

The results indicate a gradual improvement in model quality with an increasing number of training epochs. In particular, the parameters box loss, classification loss, and distribution focal loss exhibit a general trend of reduction or stabilization, which reflects successful learning of object localization, classification, and boundary refinement.

The precision and recall metrics also demonstrate positive dynamics. For example, precision increased from an initial value of 0.29 to 0.7, while recall rose from 0.52 to 0.61, indicating an improved ability of the model to correctly identify objects without omissions.

Particularly notable are the improvements in mAP50 and mAP50-95, which increased from 0.37 and 0.3 to 0.67 and 0.54, respectively. This growth reflects enhanced performance under both lenient and strict IoU thresholds, confirming the overall improvement in the detector's ability to recognize objects with more precise boundaries.

6 SELECTION OF SURVEILLANCE CAMERAS

Camera placement is a key stage in the design of circular video surveillance systems, as it directly affects area coverage, minimization of blind spots, and the efficiency of subsequent data processing [22]. Two camera configurations were considered for integration into the system: wide-angle cameras and PTZ cameras. Conventional wide-angle cameras (Field of View (FOV)=90°–360°) generate ultra-high-resolution images, which impose a substantial computational burden. When the input image size is reduced to 4K resolution, the model receives poorly detailed representations of distant objects, rendering their detection infeasible. Although the YOLOv8n model does not have a strictly defined maximum input image size, it performs real-time processing most efficiently on relatively low-resolution images up to 640 × 640 pixels. Consequently, wide-angle cameras produce high-resolution imagery that the model cannot efficiently interpret and process.

In contrast to wide-angle cameras, PTZ cameras, despite their lower native resolution, enable zoom-based magnification to enhance object detail within a radius of up to 6 nautical miles around the vessel. PTZ cameras provide flexible coverage of a wide surveillance area combined with dynamic zooming and orientation capabilities. They allow operators or automated systems to focus on specific targets by applying optical zoom, thereby increasing the effective image resolution and mitigating distortions inherent in fixed wide-angle lenses [23].

Based on the Nyquist sampling theorem, for surface objects moving at a speed of 15 knots (7.71666 m/s) and with an average vessel size of 20 m, the maximum scene change frequency is 0.39 frames per second. Thus, the minimum FPS (Frames Per Second) is approximately 1 frame per second for ideal lighting and sea state conditions.

In the works [24,25] successful practical tests of systems with data processing speeds of 1–10 FPS were published for calm inland waters with movement speeds up to 12.4 knots. However, to account for rapid scene changes, such as waves or maneuvers, and to ensure detection accuracy, authors use no less than 30 FPS [26,27]. Thus, based on previously conducted research, we adopt for video surveillance systems an FPS = 15–30, where the lower limit is acceptable for scenes on calm water with low dynamics, and the upper limit for maneuvers and high movement speeds, or under complicated weather conditions and strong waves.

According to the recommendations of IALA [28], video surveillance systems must ensure monitoring of the surface situation in nighttime operating conditions. According to reference data from the [29], the level of natural illumination under moonlight conditions is approximately 0.1–0.3 lux. In maritime low-visibility conditions, such as fog or sea haze, the visibility of surface objects for optical cameras can decrease to a few hundred meters. For surveillance under such visibility conditions, cameras with infrared (IR) and thermal (Thermal / EO-IR) sensors are used. These cameras register the thermal radiation of objects rather than visible light, allowing the detection of vessels and surface obstacles even in complete absence of illumination or under conditions of strong light scattering. The IR mode effectively complements standard video surveillance in daytime and nighttime spectra, ensuring stable operation of the video monitoring system under complex weather conditions [30].

Cameras intended for operation in the maritime environment must comply with marine reliability and protection standards. In particular, camera housings and mounts must withstand exposure to weather conditions, saltwater, dust, vibration, and shock, comply with the IEC 60945 marine standard [31] for mechanical and climatic resistance, and have an ingress protection rating of IP66 or higher to ensure dust and water tightness. The operational temperature range should cover −20°C to +50°C to guarantee functionality in both cold and hot climates [31,32].

7 PLACEMENT OF SURVEILLANCE CAMERAS

Camera placement constitutes a critical stage in the design of a panoramic video surveillance system, as it directly affects area coverage, the minimization of blind sectors, and the efficiency of subsequent data processing. In the absence of specific International Maritime Organization (IMO) regulations governing the placement of onboard visual surveillance equipment, radar installation standards [33,34] are adopted as a reference framework.

Structural elements of the vessel, including the bow superstructure, masts, funnels, cargo-handling equipment, and deck fittings, must not create extended or critical blind sectors within the Horizontal Field of View (HFOV) or Vertical Field of View (VFOV) of the cameras. The system must ensure continuous visibility of the sea surface in the vessel's heading direction as well as along the port and starboard sectors, which is essential for the early detection of other vessels, navigational hazards, and coastal objects.

Within the general forward-looking sector extending from the beam, the angular width of any individual blind sector caused by structural components or cargo must not exceed 10°. The total cumulative angular extent of all blind sectors within this area must not exceed 20°. Clear viewing sectors of at least 5° must be maintained between adjacent blind sectors. This requirement is of fundamental importance, as it guarantees the presence of regular visual observation sectors for monitoring the sea surface and surrounding objects, even in the presence of structural obstructions.

Resolution [33] establishes more stringent limitations for the critical field of view in the vessel's heading direction. In this case, the angular width of each individual blind sector must not exceed 5°, reflecting the heightened importance of uninterrupted monitoring of the navigational environment directly ahead of the vessel.

Within the proposed system architecture, the cameras are intended to be installed on one of the vessel's masts. A technical feature of the considered configuration is the potential co-location of the radar antenna and video cameras on the same mast. To fulfill operational requirements, the radar antenna installation height is regulated at approximately 15 m [33,34]. Mounting the camera at the same elevation is not feasible due to electromagnetic compatibility and safety requirements associated with radar antenna emissions.

To determine whether installation below the radar antenna would compromise system performance, an algorithm for estimating the minimum camera mounting height h for a specified observation range d is considered. This algorithm is based on the approximate optical horizon formula, which accounts for the Earth's curvature and ensures visibility of the sea surface without significant blind zones:

$$h = \left(\frac{d_{NM}}{1.93} \right)^2 \tag{2}$$

Considering that the operational range of the camera is limited to 6 nautical miles, the minimum required installation height is 9,66 m. Therefore, positioning the camera below the radar antenna does not reduce its effective observation range.

To enhance system robustness under dynamic maritime conditions, such as vessel roll and pitch, each PTZ camera is mounted on a gyrostabilized platform. By analogy with radar standards [33], which require equipment to withstand rolling motions up to ±10°, typical of marine environments, video surveillance systems must ensure image stability under comparable external disturbances.

The system operation in the horizontal plane is considered as follows. The camera operating mode without zoom is referred to as the wide-angle mode, in which the HFOV assumes its maximum value. Operator-initiated zoom enables the system to adapt for long-range observation without loss of detail. Given that PTZ cameras have been selected for system implementation, HFOV values range from 60° to 90° in wide-angle mode, while the minimum values range from 1.5° to 5° when optical or digital zoom is applied.

The primary coverage sector W assigned to a single camera defines the angular field of view that each individual camera must monitor to ensure continuous coverage of the designated observation area and is determined as follows:

$$W = \frac{S}{N} \tag{3}$$

where S is the angular field of view that must be covered, and N is the number of cameras. This represents the fundamental sector division; however, without overlap between adjacent sectors, the system becomes vulnerable to dynamic variations. Such overlap compensates for installation inaccuracies, lens distortion, and dynamic changes, and facilitates image stitching algorithms, particularly for large objects.

The overlap ΔW is introduced as a system robustness margin and is distributed symmetrically on both sides with respect to the image of the neighboring camera. In the absence of a dedicated regulatory framework, the overlap magnitude directly depends on the scene characteristics and the HFOV of the selected camera. These parameters are related by the following equation:

$$HFOV = W + 2 \times \left(\frac{\Delta W}{2} \right) \tag{4}$$

Another important system parameter is the nominal panning sector of each camera, i.e., the angular range within which it can rotate. Structurally, the nominal panning sector of each camera, enabled by mechanical rotation, ranges from 300° to 360°, which may lead to operational disorganization of the proposed system. To eliminate a potential failure factor associated with malfunction of horizontal rotation mechanisms, the nominal panning sector was constrained to be equal to the HFOV. Consequently, in wide-angle mode, the camera does not require mechanical rotation. However, when switching to manual operator control and zoom mode, the camera is permitted to rotate, but only within its predefined nominal sector.

The optical axis ψ_i of the i-th camera in wide-angle mode is determined by uniform circular distribution and is defined by the following relationship:

$$\psi_i = (i-1)\times W \tag{5}$$

where the camera index i=1,…, N.

The limits of the horizontal angular range D_i in wide-angle mode for each camera are described by the following equation:

$$D_i = \left[\psi_i - \frac{HFOV}{2}, \psi_i + \frac{HFOV}{2}\right] \tag{6}$$

The overall camera arrangement is illustrated in Fig. 7. To demonstrate the system configuration, let us assume it comprises six PTZ cameras with fixed nominal sectors. For clarity, each camera is numbered from 1 to N in a clockwise direction. The nominal panning sector of the first camera is indicated by a black arrow and is set equal to the HFOV. The primary coverage sector of each camera, representing its main observation area, is shown with an orange arrow. The optical axis in wide-angle mode, at rest and without focus or optical zoom, is represented by a blue dashed line. The distance between the optical axes in wide-angle mode is fixed and marked with a blue arrow.

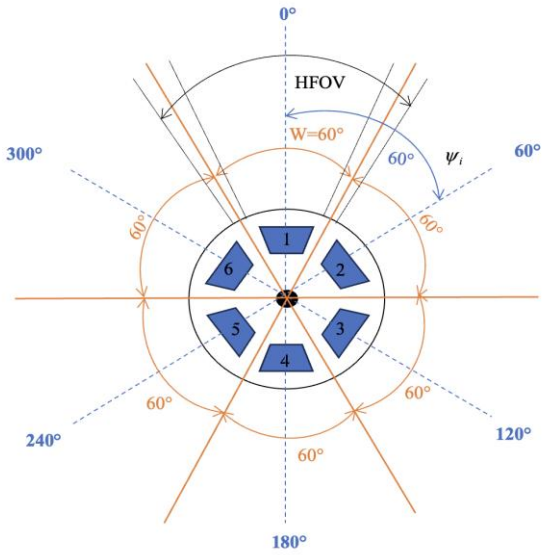

Figure 7. General layout of camera placement in the horizontal plane

In maritime surveillance systems, VFOV is a fundamental parameter defined as the angle between the upper and lower boundaries of the image captured by the camera sensor. The optimal VFOV for PTZ cameras in marine observation ranges from 30° to 60° in wide-angle mode, while during zoom it decreases to 1–5° to enable detailed focus on specific objects.

According to [33], the observation of the sea surface from the vessel's control position must not be obstructed for more than two ship lengths or 500 m ahead of the bow, whichever is smaller. This requirement applies up to 10° on either side of the bow, irrespective of the vessel's draft, trim, or the presence of deck cargo. To ensure compliance with these standards, an algorithm for determining the camera installation height and the optimal downward tilt angle is proposed. This approach allows avoidance of blind zones exceeding the prescribed limits. The lower boundary of the VFOV should intersect the sea surface at a distance not exceeding the radius of the maximum permissible blind zone. In the absence of a specific vessel, it is assumed that the radius of the maximum allowable blind zone is rmax = 500 m. For a particular vessel, this idealized value must be recalculated. Although the rmax requirement primarily applies to the forward section of the vessel, the same methodology is applied to assess blind zones around the entire vessel.

Assuming the camera is installed at a height h above sea level, then:

$$\begin{aligned} \theta_{low} &= \theta_i + \frac{VFOV}{2} \\ \theta_{up} &= \theta_i - \frac{VFOV}{2} \end{aligned} \tag{7}$$

where θ_i is the tilt angle of the optical axis, θ_{low} is the angle of the lower boundary of the field of view, and θ_{up} is the angle of the upper boundary of the field of view.

The optimal downward tilt angle θ_{low} is calculated using the following formula:

$$\theta_{low} = \arctan\left(\frac{h}{r_{\max}}\right), \tag{8}$$

where r_{max} is the maximum permissible blind zone radius, and h is the camera installation height above sea level.

The optimal angle of the upper boundary of the field of view, θ_{up} , can be determined as follows:

$$\theta_{up} = \theta_{low} - VFOV\,. \tag{9}$$

The camera must provide coverage up to the visible horizon. Excessive downward tilt limits long-range observation. Due to the curvature of the Earth, the horizon is located at a small positive angle relative to the ideal horizontal plane. For an installation height h, the angle to the visible horizon can be approximated as follows:

$$\theta_{hor} \approx \arctan\left(\frac{h}{d_{hor}}\right), \tag{10}$$

where d_{hor} is the distance to the visible horizon, determined as:

$$dhor \approx \sqrt{2\times R_e \times h}\,, \tag{11}$$

where R_e is the Earth's radius.

The condition for including the visible horizon within the camera's VFOV requires that the upper boundary of the field of view must not fall below the angle to the visible horizon, while the lower boundary must not rise above it. Mathematically, this condition is expressed as the following inequality:

$$\theta_{up} \le \theta_{hor} \le \theta_{low}\,. \tag{12}$$

For a camera installed at a height of 9.66 m, the lower field of view angle is θ_{low} =1.107° and the angle to the visible horizon is θ_{hor} =0.050°. It can be concluded that even a modest camera installation height satisfies the visibility requirements for the horizon. The downward tilt angle of the lower boundary of the field of view increases with greater camera elevation, along with the angle to the visible horizon.

8 ALGORITHM FOR DETERMINING RELATIVE AND TRUE BEARING

The relative bearing is defined as the angle between the direction of the vessel's bow and the direction toward the center of the detected object. The algorithm is based on processing the coordinates of the object's bounding box obtained from the YOLOv8n model, taking into account the camera geometry, its orientation, and the operational mode (wide-angle or zoomed). Calculations are performed in real time on a computational platform using Python.

Each camera is responsible for a specific horizontal angular range, i.e., its primary coverage sector with additional overlap on both sides. An example of the responsibility zones for each camera is shown in Fig. 7. The relative bearing to an object is measured from 0° to 360°. If the center of the bounding box falls within a camera's responsibility zone, the system assigns a relative bearing value corresponding to the horizontal angular range at that point.

The bounding box coordinates are defined by the parameters x_{min}, y_{min}, x_{max}, y_{max}; the coordinates of the central point can then be calculated as:

$$x_{center} = \frac{x_{\min} + x_{\max}}{2}, \quad y_{center} = \frac{y_{\min} + y_{\max}}{2} \tag{13}$$

Since the acquired video stream has a well-defined resolution X x Y, the coordinate xmin can vary from 0 to (X–1) along the horizontal axis, while ymin ranges from 0 to (Y–1) along the vertical axis. The maximum values, x_{max} and y_{max}, vary from 1 to X horizontally and from 1 to Y vertically, respectively.

The resulting center coordinates, x_{center} and y_{center}, represent a key point, as they indicate the center of the object. For vessels, this typically corresponds to the center of the hull. If the object is in motion, the center is updated with each frame. The camera sequentially evaluates each object according to priority, computing the relative bearing for each. If an object moves outside the dynamic range of a camera, the system transfers tracking responsibility to an adjacent camera to avoid unnecessary rotation. In cases where the object is only partially visible in the frame, the system still determines the bounding box center, even if it does not correspond to the true hull center, and calculates the bearing for the visible portion of the vessel.

To illustrate the system's operational principles, simulation modeling was performed using pre-recorded video. These videos were processed with the same algorithms applied to live video streams. Fig. 8 presents an example in which a single camera simultaneously tracks multiple vessels. Each detected vessel is enclosed within a green bounding box. The relative bearing is displayed above the top-left corner in pink in the format R-BRN: XXX.XX. Below, in green, the vessel's orientation and the probability of correct identification are indicated. Objects with an identification probability below 0.5 are not marked by the system to prevent false detections.

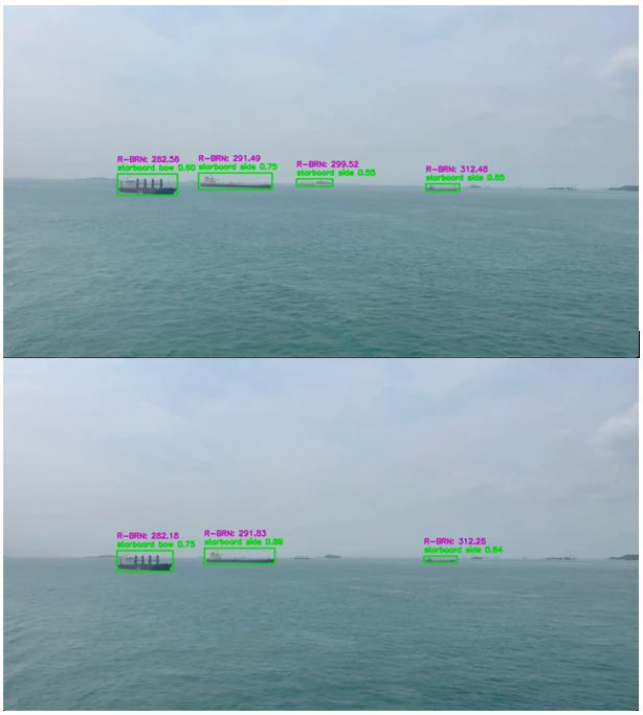

Figure 8. Vessel identification, determination of orientation and relative bearing

For comprehensive navigational analysis, particularly for monitoring surface objects, it is necessary to convert to the true bearing – i.e., the angle relative to geographic north. The true bearing T to an object can be calculated using a formula that accounts for the relative bearing R measured by the PTZ camera and the vessel's true heading H:

$$t = \begin{cases} R + H, & 0 \le R + H < 360° \\ R + H - 360° & R + H > 360° \end{cases} \tag{14}$$

The relative bearing R was obtained through image processing from the cameras, as described previously.

The vessel's true heading is not part of the video surveillance subsystem, and its integration requires synchronization with the ship's external navigation systems, specifically a gyrocompass or a magnetic compass. The gyrocompass provides the true heading independently of magnetic variations, whereas the magnetic compass provides a magnetic heading that must be corrected for deviation and variation. For system simplification, data from the gyrocompass will be used.

The gyrocompass is connected to the computational platform for automated data processing. Gyrocompass data are processed in real time using the NMEA 0183 protocol [35]. The gyrocompass provides information on the vessel's true heading (HDT) and the vessel's rate of turn (ROT). The HDT sentence is delivered in the following format: $HEHDT,123.456,T*2F, where $HE indicates that the data are from the gyrocompass, HDT specifies the sentence type indicating the true heading, 123.456 is the heading in degrees, T denotes true

heading, and *2F is the checksum used for data integrity verification [35–37]. To synchronize the information obtained from the gyrocompass with the video stream, precise timestamps are required, enabling the correlation of the two data streams.

Synchronization of timestamps between the video stream and gyrocompass data is achieved using the Network Time Protocol (NTP). Each video frame and each received NMEA message is assigned a unified system timestamp. Upon receipt of a new HDT message, the true heading value is recorded along with its reception timestamp. Similarly, each video frame receives its own timestamp at the moment of capture. Synchronization is performed by selecting the heading value closest in time to the frame being processed. The reference time source is provided by a GPS receiver, which generates high-precision timing signals.

The true bearing values are displayed using the same principle as the relative bearing. Figure 9 illustrates an example of a video stream with identified targets; each marked with a green bounding box. In the top-left corner, the upper row shows the true bearing in white, in the format T-BRN: XXX.XX, where XXX.XX represents degrees and fractional degrees. In the lower row of the same corner, the vessel's orientation and the probability of correct identification are displayed. Fig. 9 shows an example of a video stream captured by camera number 6, with the vessel moving on a true heading of 040°.

Figure 9. Vessel identification, determination of orientation, and true bearing

The described mechanism for integrating relative bearing with gyrocompass data ensures reliable real-time determination of the true bearing. This approach not only complements the video surveillance subsystem but also supports comprehensive navigational analysis. Overall, the integration of navigational data enhances the system's potential for autonomous maritime applications and contributes to improved maritime safety.

9 VIDEO STREAM PROCESSING SYSTEMS

The analysis of maritime incidents, fault diagnostics, and the subsequent improvement of the YOLOv8n model through the preservation of acquired data are integral components of the system. A network storage unit connected to the computing platform is responsible for data collection, storage, and management, thereby forming a complete cycle for the analysis and optimization of computer vision systems (Fig. 10).

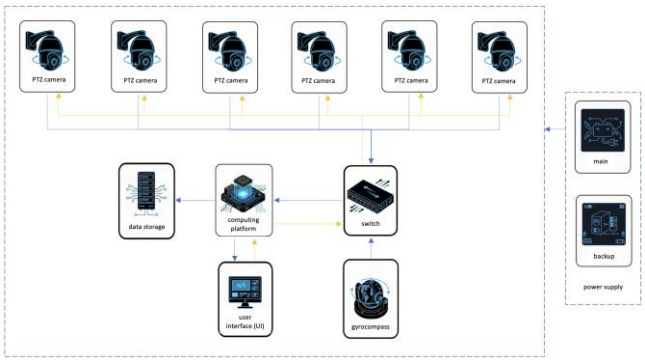

Figure 10. Block diagram of the video surveillance system for ship aspect recognition

For data visualization, all information received from the server is transmitted to the user interface, which may be implemented as a touch-screen monitor. Preference is given to models with a large diagonal of 21 inches or more and an IP65 protection rating. For more convenient manual control, a joystick may be used, allowing the operator to control the physical movement of the camera. This is particularly useful for dynamic focusing on objects where rapid adjustment of the HFOV and VFOV is required to better resolve details, for example, under conditions of dense traffic.

10 CONCLUSIONS

As a result of the conducted research, a comprehensive video surveillance system for ship aspect recognition based on the YOLOv8n model was developed. After nine training stages on a dataset comprising 1742 images with data augmentation, the system achieved stable performance metrics (precision 0.70344, recall 0.60982, mAP@0.5:0.95 0.54036), demonstrating effective adaptability to variable maritime conditions such as changes in illumination, weather, and object occlusions. The application of class weighting to address class imbalance prevented overfitting to dominant classes and improved recognition performance for underrepresented aspects, such as stern and forward.

The use of PTZ cameras with dynamic zoom and panning capabilities enables an observation range of up to 6 nautical miles. By installing the cameras on the vessel's mast at a minimum height of 9.66 m, in accordance with IMO standards, blind zones around the vessel are minimized.

Due to the fixed camera configuration, the system is able to compute the relative bearing of a target by determining the center point of the detected bounding box. Integration of gyrocompass data allows the

system to further calculate the true bearing of the detected vessels.

The proposed system not only supports compliance with COLREGs by enhancing situational awareness and traffic management in congested waterways, but also integrates with other navigational systems for autonomous vessels, enabling data storage and analysis for further model improvement. Overall, this development opens new prospects for improving maritime safety, reducing collision risks in mixed traffic environments, and expanding the application of computer vision in the maritime domain, with potential adaptability to other related tasks.

REFERENCES

[1] A. Vekinis and S. Perantonis, "Aeolus Ocean – A simulation environment for the autonomous COLREG-compliant navigation of unmanned surface vehicles using deep reinforcement learning and maritime object detection," arXiv, arXiv:2307.06688, 2023.

[2] International Maritime Organization, Convention on the International Regulations for Preventing Collisions at Sea (COLREGs), adopted Oct. 20, 1972, entered into force Jul. 15, 1977, London, U.K., 1977.

[3] D. Qiao, G. Liu, J. Zhang, and Y. Zhou, "Marine vision-based situational awareness using discriminative deep learning: A survey," J. Mar. Sci. Eng., vol. 9, no. 4, p. 397, 2021, doi:10.3390/jmse9040397.

[4] J. R. Terven and D. M. Cordova-Esparaza, "A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond," arXiv, arXiv:2304.00501, 2023.

[5] Stereolabs Inc., "Performance benchmark of YOLO v5, v7 and v8," 2023. [Online]. Available: https://www.stereolabs.com/blog/performance-of-yolo-v5-v7-and-v8

[6] Z. He, K. Wang, T. Fang, L. Su, R. Chen, and X. Fei, "Comprehensive performance evaluation of YOLOv11, YOLOv10, YOLOv9, YOLOv8 and YOLOv5 on object detection of power equipment," arXiv, arXiv:2411.18871, 2024.

[7] M. Yasir, S. Ahmed, F. Khan, Y. Zhang, A. H. Aslam, and A. J. Malik, "YOLOShipTracker: Tracking ships in SAR images using lightweight YOLOv8," Int. J. Appl. Earth Obs. Geoinf., vol. 134, p. 104137, 2024, doi:10.1016/j.jag.2024.104137.

[8] Y. Li and S. Wang, "EGM-YOLOv8: A lightweight ship detection model with efficient feature fusion and attention mechanisms," J. Mar. Sci. Eng., vol. 13, no. 4, p. 757, 2025, doi:10.3390/jmse13040757.

[9] J. Di, L. Sun, R. Zhang, and Q. Wu, "An enhanced YOLOv8 model for accurate detection of solid floating waste," Sci. Rep., vol. 15, no. 1, p. 1632, 2025, doi:10.1038/s41598-025-10163-2.

[10] B. Zhao, H. Chen, X. Liu, and J. Huang, "Modular YOLOv8 optimization for real-time UAV maritime rescue object detection," Sci. Rep., vol. 14, no. 1, p. 1158, 2024, doi:10.1038/s41598-024-75807-1.

[11] P. Liu, "A high-accuracy YOLOv8-ResAttNet framework for maritime object recognition," IET Image Process., vol. 19, no. 3, pp. 145–157, 2025, doi:10.1049/ipr2.70085.

[12] X. Zhao and Y. Song, "Improved ship detection with YOLOv8 enhanced with MobileViT and GSConv," Electronics, vol. 12, no. 22, p. 4666, 2023, doi:10.3390/electronics12224666.

[13] M. E. Lopez, C. Schönlieb, A. Aviles-Rivero, R. Mester, and E. Brekke, "Heading estimation of ships using deep learning and synthetic non-aerial images," J. Phys.: Conf. Ser., vol. 3123, no. 1, 012027, 2025, doi:10.1088/1742-6596/3123/1/012027.

[14] B. Kiefer, Y. Quan, and A. Zell, "Approximate supervised object distance estimation on unmanned surface vehicles," arXiv, arXiv:2501.05567, 2025.

[15] A. Gorad, S. S. Hassan, and S. Särkkä, "Vessel bearing estimation using visible and thermal imaging," in Lecture Notes in Computer Science, vol. 13886, Springer, 2023, pp. 373–381, doi:10.1007/978-3-031-31438-4_25.

[16] O. Pashenko and O. Pipchenko, "Design of a YOLO-based computer vision model for ships' aspect angle detection," Shipping & Navigation, no. 38, pp. 10–21, 2025, doi:10.31653/2306-5761.38.2025.10-21.

[17] Z. Yang et al., "Does negative sampling matter? A review with insights into its theory and applications," arXiv, arXiv:2402.17238, 2024.

[18] Ultralytics, "YOLO data augmentation," 2025. [Online]. Available: https://docs.ultralytics.com/guides/yolo-data-augmentation/

[19] Z. Wu, "Image data augmentation techniques based on deep learning: A survey," Math. Biosci. Eng., vol. 21, no. 6, pp. 6190–6224, 2024, doi:10.3934/mbe.2024272.

[20] B. Zoph, E. D. Cubuk, G. Ghiasi et al., "Learning data augmentation strategies for object detection," arXiv, arXiv:1906.11172, 2019.

[21] O. L. Pashenko, "Impact of data augmentation on training computer vision model for ships' aspect angle detection," Scientific Bulletin Kherson State Maritime Academy, no. 2(31), pp. 52–63, 2025, doi:10.33815/2313-4763.2025.2.31.052-063.

[22] S. Fefilatyev, D. Goldgof, and C. Lembke, "Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system," Ocean Eng., vol. 54, pp. 1–12, 2012, doi:10.1016/j.oceaneng.2012.06.028.

[23] J. Guo, L. Fan, B. Wu, J. Gu, S. Cao, and J. Ye, "PTZ-Calib: Robust Pan-Tilt-Zoom camera calibration," arXiv, arXiv:2502.09075, 2025.

[24] N. Wawrzyniak, T. Hyla, and A. Popik, "Vessel detection and tracking method based on video surveillance," Sensors, vol. 19, no. 23, 5230, 2019, doi:10.3390/s19235230.

[25] X. Liu, Y. Hu, H. Ji, M. Zhang, and Q. Yu, "A deep learning method for ship detection and traffic monitoring in an offshore wind farm area," J. Mar. Sci. Eng., vol. 11, no. 7, p. 1259, 2023, doi:10.3390/jmse11071259.

[26] M. Ud Din, A. B. Bakht, W. Akram, Y. Dong, L. Seneviratne, and I. Hussain, "Benchmarking vision-based object tracking for USVs in complex maritime environments," arXiv, arXiv:2412.07392, 2024.

[27] W. Ding, H. Li, C.-M. Chew, X. Zhang, and H. Huang, "Deep learning-based recognition of small maritime targets for obstacle avoidance in visual wave gliders," Ocean Eng., vol. 322, 120471, 2025, doi:10.1016/j.oceaneng.2025.120471.

[28] International Association of Marine Aids to Navigation and Lighthouse Authorities, Vessel Traffic Services Manual, 8th ed., Saint-Germain-en-Laye, France, 2025.

[29] Illuminating Engineering Society, IES Lighting Handbook: Reference and Application, 10th ed., New York, NY, USA, 2011.

[30] B. Wang, L. Dong, M. Zhao, H. Wu, and W. Xu, "An infrared maritime target detection algorithm applicable to heavy sea fog," Infrared Phys. Technol., vol. 71, pp. 56–62, 2015, doi:10.1016/j.infrared.2015.04.009.

[31] International Electrotechnical Commission, IEC 60945: Maritime navigation and radiocommunication equipment and systems – General requirements – Methods of testing and required test results, Geneva, Switzerland, 2002.

[32] International Electrotechnical Commission, IEC 60068: Environmental testing, Geneva, Switzerland, 2015.

[33] International Maritime Organization, Resolution A.708(17): Navigation bridge visibility and functions, London, U.K., 1991.

[34] International Maritime Organization, Resolution MSC.192(79): Adoption of the revised performance standards for radar equipment, London, U.K., 2004.

[35] National Marine Electronics Association, NMEA 0183 Standard for Interfacing Marine Electronic Devices, ver. 4.30, 2023. [Online]. Available: https://www.nmea.org
[36] International Maritime Organization, Resolution MSC.232(82): Guidelines for the use of electronic navigation systems, London, U.K., 2006.
[37] International Electrotechnical Commission, IEC 61162-1: Maritime navigation and radiocommunication equipment and systems – Digital interfaces – Part 1: Single talker and multiple listeners, 6th ed., Geneva, Switzerland, 2024.