

Temporal context of lightweight network model for detecting boats approaching the tsunami early warning system

Wayan Wira Yogantara^{1,3}, Suprijanto², Anak Agung Ngurah Ananda Kusuma⁴, Yuki Istianto⁵

¹Master's Program of Instrumentation and Control, Faculty of Industrial Technology, Institut Teknologi Bandung, Bandung, Indonesia

²Instrumentation, Control and Automation Research Group, Faculty of Industrial Technology, Institut Teknologi Bandung, Bandung, Indonesia

³Research Center for Electronics, National Research and Innovation Agency, Jakarta, Indonesia

⁴Research Center for Telecommunications, National Research and Innovation Agency, Jakarta, Indonesia

⁵Research Center for Artificial Intelligence and Cyber Security, National Research and Innovation Agency, Jakarta, Indonesia

Article Info

Article history:

Received May 22, 2024

Revised Jul 4, 2025

Accepted Aug 6, 2025

Keywords:

Lightweight WaSR-T

Maritime computer vision

MobileNetV3

Network marine object detection

Tsunami early warning system

ABSTRACT

The tsunami early warning system (TEWS) is a device that detects potential tsunamis. However, a boat that approaches TEWS is a source of communication disturbance. A convolutional neural network (CNN), as part of intelligent computer vision, is one solution for detecting boats and providing a warning to move away from the TEWS area. Water segmentation and refinement-temporal (WaSR-T), as the current advanced CNN network, exhibits impressive performance in detecting object obstacles in the marine domain, although it requires a powerful computational device. In the paper, we propose a modification of WaSR-T, replacing the most computationally intensive stages with a lightweight version called lightweight WaSR-T. On the proposed lightweight WaSR-T, the previous encoder of WaSR-T was replaced with MobileNetV3, and some feature layer maps were reduced as input to the decoder. For training and validating the lightweight WaSR-T, the image dataset representing the open sea and our extended dataset from Indonesia's ocean region were used. Based on the quantitative results and evaluation of the computational load, the sensitivity to detect a boat for WaSR-T and lightweight WaSR-T is 95.71% and 90.00%, respectively. The lightweight WaSR-T required less memory at 32.57%, resulting in a 0.0761% reduction in total processing time compared to the original WaSR-T. Therefore, our proposed lightweight WaSR-T is promising for use as the central part of an intelligent maritime computer vision system in TEWS.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Suprijanto

Instrumentation, Control and Automation Research Group, Faculty of Industrial Technology

Institut Teknologi Bandung

Bandung, Indonesia

Email: supri89@itb.ac.id

1. INTRODUCTION

A tsunami early warning system (TEWS) is crucial for minimizing the potential damage and loss of life caused by natural disasters. TEWS based on tsunami buoys are deployed in the deep ocean close to a possible location of sea earthquakes and equipped with pressure sensors to detect changes in water level [1]. When a tsunami passes over the buoy, it registers the pressure change and sends this data to monitoring stations, which can then assess the potential threat from the tsunami. The quality of data transfer using acoustic link communication depends on the distance range between an ocean bottom unit (OBU) and a surface buoy, as well as noise interference due to undesired acoustic waves. The presence of unauthorized

boats in the installation area of the tsunami buoy system is a source of disturbance that must be avoided. A device that detects unauthorized boats approaching the surface buoy is required to prevent disruption to the function of this tsunami detection system. The device has not yet been installed on the existing TEWS system in the Indonesian open sea.

Various sensors, such as marine radar and optical sensor-based cameras, are commonly used for object detection on the sea surface [2]. A device with an optical camera is gaining prominence as a leading object detection method when supported by proper computer vision methods [3]–[5]. Previous classical computer vision methods for object detection rely on simple clustering of features related to specific groups of objects, which are not expressive enough for accurate detection in object detection environments [6]. A current intelligent computer vision method for object detection is based on classifying every pixel in an image using a convolutional neural network (CNN) to provide valuable information for scene understanding and object detection [7]–[9]. The state-of-the-art (SOTA) CNN network for intelligent computer vision object detection has become an established approach in autonomous ground vehicles [10]–[14]. However, the existing SOTA CNN network, primarily developed in terrestrial ground scenes, is inadequate in the maritime domain.

The water segmentation and refinement (WaSR) [15]–[17] is one of the SOTA CNN networks that has good performance for unmanned surface vehicles (USVs). The WaSR network consists of a contracting path (encoder) and an expansive path (decoder) comprising several information fusion and feature scaling blocks. The WaSR decoder was designed to fuse inertial measurement unit (IMU) and visual image information to increase the accuracy of positive object detections. The architecture upgrade of the WaSR network was proposed to extract the temporal context from a sequence of frames to differentiate objects from reflections, and it is called WaSR-T [18]. WaSR-T has been reported to reduce the number of false positive (FP) detections compared with WaSR.

We required a SOTA CNN network to detect unauthorized boats approaching tsunami buoys as part of an intelligent computer vision system. WaSR-T is one of the best-performing maritime obstacles for USVs [17]. Therefore, WaSR-T is one of the candidates for the unauthorized ship detection system. While achieving impressive object detection results for USVs, the WaSR-T network requires a powerful computational device [6].

In this work, we proposed a modification of WaSR-T with replacements for the most computationally intensive stages, called the lightweight WaSR-T. The objective of lightweight WaSR-T is to classify and label each pixel of a recorded image as an unauthorized boat with a background sea or area interface between the sea and the sky that approaches TEWS with inexpensive computational resources and good performance. For the training and validation of the lightweight WaSR-T, a unique dataset was utilized that accurately represents the open sea. This dataset was selected from the open dataset maritime semantic segmentation training [18], [19], and from datasets available on an open website. We also created an extensive dataset from the Indonesian TEWS area, which is installed in the open sea. Then, the ability of WaSR-T and lightweight WaSR-T to detect an unauthorized boat using these datasets was performed.

This paper is structured as follows: section 2 introduces the material and methods, including the concept of unauthorized boat detection for TEWS, the datasets used for developing the SOTA CNN network, and a review of the existing architecture, WaSR-T, and lightweight WaSR-T. Section 3 describes the details of the experiment procedures. The results and discussion are described in section 4. Finally, in section 5, we draw our conclusions.

2. METHOD

2.1. The concept of unauthorized boat detection for the tsunami buoy system

The typical tsunami buoy system consists of two main parts: an OBU that measures changes in sea level height at the seabed and a surface buoy that transmits measurement data to a tsunami data center as shown in Figure 1. The OBU and surface buoy exchange information using acoustic communication modems that are very susceptible to noise interference, for example, from the ship's propellers. In some cases, a surface buoy is used as a boat mooring, which can create problems in acoustic link communication due to the shifting position of the buoy system. It may also impact the communication channel between the OBU and the surface buoy [20]. An unauthorized boat approaching the TEWS installation area is a source of disturbance. Boats or other objects approaching the installation area must be warned to leave immediately and notify the data center office of the resulting disturbances. An intelligent computer vision system is one of the solutions that may be installed on the highest surface of the buoy as shown in Figure 1.

The SOTA CNN network, as the central part of an intelligent computer vision system, must be developed based on the concept of semantic segmentation [21]–[23]. Semantic segmentation is used to classify and label each pixel of a recorded image as either an unauthorized boat, a background water surface, or an area interface between the sea and sky. This SOTA CNN network must be able to detect

static and dynamic unauthorized boats of various shapes and sizes, including those not seen during training. Additionally, the SOTA CNN network must be highly adaptable to challenging and dynamic water features, allowing TEWS to operate effectively. Due to environments and conditions, an intelligent computer vision and SOTA CNN network must work for real-world small-sized energy-constrained TEWS. Therefore, a lightweight network that can be run on a device with limited memory and a small architectural device is required.

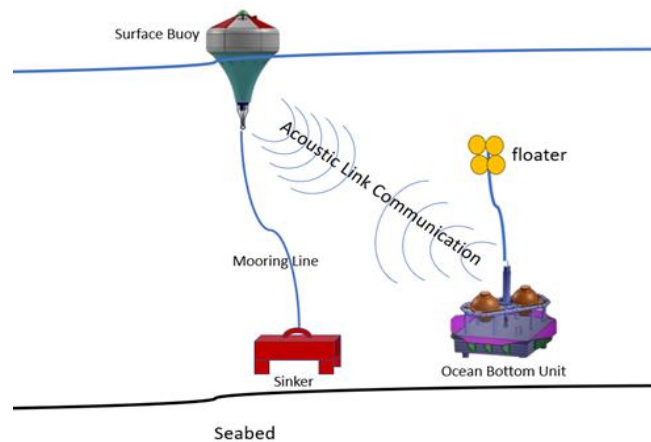


Figure 1. Configuration of the Indonesian TEWS

2.2. Open sea image datasets for the development of SOTA CNN network

Detecting objects on maritime open seas with CNNs presents several challenges, like the complexity of objects, environmental conditions, data availability, class imbalance, anomaly detection, small object detection, and model robustness. Large datasets relevant to the open sea domain are required to support the development of the SOTA CNN network as the central part of an intelligent computer vision system for TEWS. We cannot rely solely on currently available maritime datasets, such as the maritime semantic segmentation training dataset [19], because most existing images show sea conditions in calm areas such as bays, harbors, rivers, or estuary areas. In contrast, the available dataset is relatively limited for open sea conditions.

The proposed CNN network must be trained with an image dataset relevant to open-sea environments. Therefore, in the study, open-sea domain data sets were recorded directly from the location in the open sea of the Indonesian Ocean, where TEWS was currently installed, to obtain actual sea conditions. Additional datasets of marine domains relevant to conditions on the open sea were also collected from an open website.

2.2.1. Open datasets

A dataset containing image frames typical of conditions on the open sea is required to develop the SOTA-CNN algorithm for unauthorized boat detection on TEWS. Openly accessible image training datasets for the marine domain have been available for tailoring to develop obstacle detection methods in small-sized, USVs. For example, MaSTr1325 is the dataset that contains 1325 diverse images captured over two years with real USV, covering a range of realistic conditions encountered in a coastal surveillance task [19]. The extended dataset of MaSTr1325 added 153 images (including their preceding frames) and used the codename MaSTr1478 for this additional dataset. MaSTr1478 training images represent challenging objects due to mirroring, reflections, and sun glitters [18]. The entire dataset was explicitly created for training the USV; Lately, the primary purpose of this USV is to survey vessels in a coastal domain, so that sea conditions, such as those in bays, river estuaries, and harbors, dominate the main frame images.

Additional datasets of marine domains relevant to conditions in the open sea were also collected from an open-access website [24]. We selected 150 images that closely resemble the typical conditions and objects captured by the camera on the open sea. The selected images represent various shapes and sizes of boats captured within the specific field of view of the camera, under different weather conditions. Examples of selecting open-sea frame images from the MaSTr1325 and MaSTr1478 datasets, as well as from the open website, are shown in Figure 2.



Figure 2. An example of selecting open sea frame images from the MaStr1325, MaStr1478 datasets, and the open website

2.2.2. Recording image datasets from the open sea in Indonesian territory

TEWS primarily operates in the open sea of Indonesian territory, which is relatively far from the coast. New datasets were recorded from locations close to the operational area of TEWS. The Indonesian TEWS are primarily operated in the open sea of the Indian Ocean, in the southern region of Java Island, and on the west coast of Sumatra Island. The five existing TEWS locations are illustrated in Figure 3(a) and the tsunami buoy deployment process at sea is shown in Figure 3(b).

The new dataset was recorded from the camera installed on the Baruna Jaya research vessel, which is responsible for recovering and maintaining the TEWS system. We systematically recorded images representing various conditions on the open sea near the TEWS system, including morning, afternoon with peak sunlight, and evening as the sun sets. An example of recorded images used for training and testing our proposed SOTA CNN network is shown in Figure 4. These images represent various shapes and sizes of unauthorized boats captured in the specific field of view of the camera installed on the Baruna Jaya research vessel.

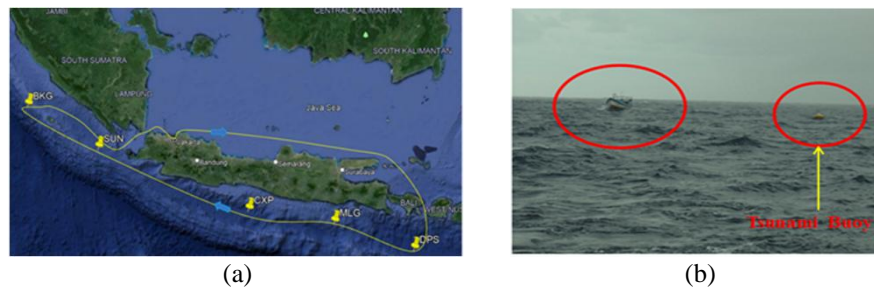


Figure 3. The five existing TEWS locations are marked with (a) BKG, SUN, CXP, MLG, and DPS, and (b) an example of the tsunami boys as part of TEWS on location SUN



Figure 4. Example of images used for training and testing our proposed network from the camera installed on the Baruna Jaya research vessel

2.3. SOTA CNN network for boat detection in the open sea

Modern CNN architectures at the forefront of maritime object detection typically employ semantic segmentation methods. These approaches process visual input at the pixel level, categorizing each pixel into distinct classes—commonly sea, sky, or obstructive elements such as vessels, buoys, or floating debris [6]. This granular classification allows for accurate identification and localization of objects within the complex and often unpredictable maritime environment.

To perform effectively in real-world conditions, these advanced algorithms must be highly adaptable to the dynamic and visually challenging nature of ocean surfaces. Variations in lighting, reflections from sunlight, shifting wave patterns, and atmospheric disturbances can drastically alter the visual characteristics of the sea and sky, complicating object recognition. Therefore, the models must be resilient to such fluctuations and capable of generalizing across diverse scenarios not encountered during training.

In addition, detection systems must be proficient in recognizing a wide array of obstructive or hazardous objects, which may differ greatly in form, scale, texture, and orientation. This includes both familiar maritime hazards and novel or previously unseen items absent from the training data. A comprehensive and diverse dataset that captures extreme and varied sea conditions is essential to support this capability.

Achieving robust generalization demands advanced feature extraction and representation learning within CNN frameworks, often enhanced through techniques such as data augmentation and transfer learning. The overarching objective is to ensure consistent and accurate detection performance across a broad spectrum of operational environments. Thereby contributing to safer navigation and more effective maritime surveillance.

2.3.1. Architecture analysis WaSR-T and lightweight WaSR-T

The WaSR [15]–[17] is one of the SOTA CNN networks that perform well for object detection in the maritime domain. The WaSR decoder was designed to fuse IMU and visual image information to increase the accuracy of positive object detections. However, WaSR still needs to improve its performance in cases where a single image is quite challenging to distinguish the reflective properties of the water surface and objects that often occur on the open sea. The modified architecture of the WaSR network, called WaSR-T, is a modification of the encoder to accommodate multiple sequence image frames as input, which has been proposed to extract spatio-temporal texture and cope with reflections [18]. The WaSR-T network architecture is shown in Figure 5(a). The encoder backbone of WaSR-T utilizes a ResNet-101 backbone featuring atrous convolutions. During network retraining, all layers are expanded, allowing the residual parts of the network to delve deeper into the feature space of the input image. In the context of an encoder for WaSR-T, ResNet-101 comprises four residual convolutional blocks (Res2, Res3, Res4, and Res5). This architecture of WaSR-T has been utilized to encode the varied appearance of open sea scenes, including elements like boats, water, and sky, and to classify each region within the target image frame $X \in \mathbb{R}^{(3 \times H \times W)}$. To improve the prediction accuracy, the role of ResNet-101 has been extended to encode discriminative temporal information about local feature appearance change of the region of the target image over T preceding context frames capped by M elements of double-struck cap R , $M \in \mathbb{R}^{(T \times 3 \times H \times W)}$ as shown in Figure 5(a).

The image input (X) and context frame (M) are first encoded with a Resnet-101 encoder network that produces per-frame feature maps frame $X_F \in \mathbb{R}^{(N \times H \times W)}$ and $M_F \in \mathbb{R}^{(N \times H \times W)}$, where N is the number of channels of feature maps. The temporal context module (TCM) extracts temporal information from the embeddings of the context and target frames. To maintain the structure and quantity of input channels to the decoder, TCM initially decreases the dimensionality of per-frame feature maps X_F and M_F into $N/2$ -dimensional per-frame representations. Finally, the output from TCM is fed to the first fusion block called the attention refinement module (ARM 1 and ARM 2 in Figure 5(a)). ARM 1 is used to adjust the weights of input feature channels based on their content within the channels. The individual weights for each channel are determined by taking the average of the input features across spatial dimensions. This process yields a 1×1 feature vector, which then undergoes a 1×1 convolution and passes through a sigmoid activation function. The second fusion block is denoted as a feature fusion module (FFM). It combines features from various network branches by concatenating them, followed by a 3×3 convolution. The third major block is referred to as atrous spatial pyramid pooling (ASPP) and SoftMax. ASPP simultaneously utilizes convolutions with varying dilation rates and combines the generated representations to effectively capture the object and image context at various scales as shown in Figure 5(a).

One of the potential problems in using WaSR-T to detect unauthorized boats approaching TEWS is its requirement for a powerful computational machine. The encoder of WaSR-T is primarily responsible for memory consumption due to its utilization of the ResNet-101 as the backbone. One strategy to address the issue is to replace the encoder with a lightweight backbone that can operate on low-power devices. MobileNets is one of a class of lightweight architectures currently used in various applications [24]–[26].

MobileNet is preferred as an encoder backbone for image segmentation networks for various applications [27]–[29]. Compared with ResNet-101, MobileNet has fewer parameters and requires less memory and storage, which can be advantageous, especially in deployment scenarios. However, limited research has been reported on the applications of MobileNet in the marine domain. The WaSR-T with an encoder using MobileNet is referred to as the lightweight WaSR-T, with its network architecture illustrated in Figure 5(b).

MobileNets [24], [30], [31] factorize convolutions into deep and precise convolutions, proposing hard-swish activation functions, squeeze-and-excite blocks [32], and future neural architecture search (NAS) to find the best architecture for mobile CPUs. MobileNets factorizes convolutions into depthwise convolutions and proposes a hard swish activation function (h-swish). The h-swish nonlinearity is employed to minimize the number of training parameters and reduce the model complexity and size. Unlike the standard use of convolution in ResNet, depthwise separable convolution splits the computation into two steps: a depthwise convolution applies a single convolutional filter to each input channel, and pointwise convolution is used to create a linear combination of the output of the depthwise convolution. The depthwise convolutional kernel is a learnable parameter applied to each input channel separately, increasing model efficiency and reducing computation costs in the MobileNets network. It is also shared across all input channels.

Furthermore, to search for the best kernel size in the depthwise convolution, NAS was employed to find the best architectures to fulfill the low-resourced hardware platforms in terms of size, performance, and latency [6], [30], [31]. On the lightweight WaSR-T, the encoder begins with a stem block that processes the input image to extract fundamental features. The shape of the resulting feature map is influenced by downsampling operations, typically involving stride-2 convolutions. For an input image of $512 \times 384 \times 3$ (width \times height \times channels), this downsampling would yield a feature map size of $256 \times 192 \times 3$ at the output stage 1. We utilize the skip connection from the first and second residual blocks of MobileNetV3, similar to the WaSR-T architecture, where the skip connection is located at stages 2 and 3. Unlike the original WaSR-T, the latent feature dimension of the FFM was reduced from 1024 to 128, and the FFM output was 96×128 to reduce the computational load in the decoder part.

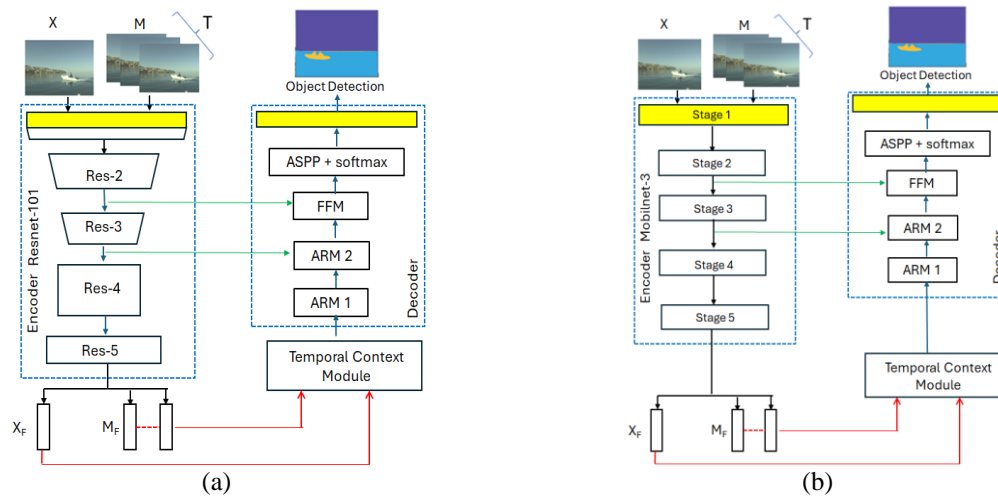


Figure 5. Two different WaSR-T network architecture (a) WaSR-T and (b) lightweight WaSR-T network architecture

3. EXPERIMENT PROCEDURES

3.1. Image dataset for training and evaluation

The datasets used for training the SOTA CNN network were selected from 126 images (including their preceding frames) chosen from MaStr1325, MaStr1478, and supplemented with 164 images (including their preceding frames) from the extended image dataset of the open sea in Indonesia's territory. The annotation image for training the proposed SOTA CNN network was available from the datasets of MaStr1325 and MaStr1478. Therefore, the extended image datasets from the open website and the extended image dataset were manually annotated per pixel for three semantic components: sea, sky, and boat. The labeling for the extended dataset is conducted using over 164 selected images for the training process of the lightweight WaSR-T network. The annotation process was carried out using the LabelMe tools [33]. An example of an annotation process using LabelMe is shown in Figure 6. Labels in ground-truth annotation

masks correspond to the following values, i.e., boat, sea, and sky are given labels 1 (Figure 6(a)), Figure 6 (b) gives 2 (value two), and Figure 6(c) gives 3 (value three), successively.

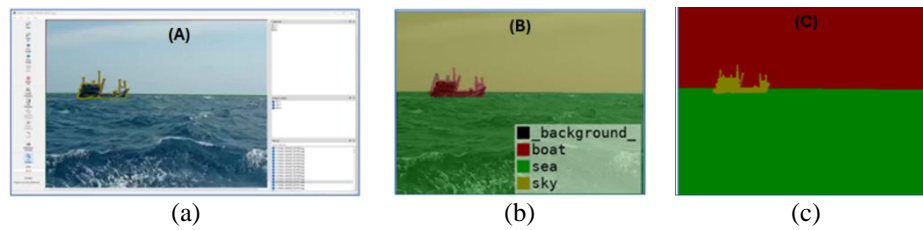


Figure 6. Example of an annotation process using LabelMe: (a) the process of a manual annotation of image, (b) manually annotated per pixel for three semantic components: sea, sky, and boat, and (c) image masking that represents the label class

3.2. Training setup lightweight WaSR-T

The proposed lightweight WaSR-T was trained using 290 image datasets, including their preceding frames with $T=3$, as well as the corresponding annotation images. The size of input images for training was $512 \times 384 \times 3$. The dataset was divided into mini-batches of 10 images to improve the efficiency of the training process. The adaptive moment estimation (Adam) optimizer utilizes a square gradient to adapt learning rates. Hyperparameter training was conducted through various configurations to achieve the best possible results, which are summarized in Table 1.

Table 1. The hyperparameter training of the Adam-optimizer

| Parameter | Value |
|---------------------|-------------|
| Learning rate | $10^{-0.6}$ |
| Learning rate decay | 0.9 |
| Weights decay | $10^{-0.6}$ |
| Epoch | 500 |
| Batch size | 6 |
| Momentum value | 0.9 |
| Patience | 50 |

The training of the lightweight WaSR-T was performed using a facility in the laboratory of high-performance computing at the Indonesian National Research and Innovation Agency, equipped with an NVIDIA DGX1. The optimized network parameters of lightweight WaSR-T, based on the criterion of minimum train/loss value, are 0.0009. A val/accuracy of 0.995 can be achieved using the hyperparameter training of Adam-optimizer with 290 epochs. After the training process, the lightweight WaSR-T was tested to detect unauthorized boats with various shapes and sizes that may have been uncovered in training datasets. The testing datasets consist of 140 images selected from an open website and datasets recorded from the Baruna Jaya research vessel in Indonesia's territory, in addition to the datasets used for data training. As a reference for evaluating detection accuracy, the pre-training original WaSR-T models are publicly available on GitHub [34]. The pre-trained original WaSR-T was also assessed to detect boats using similar datasets to those used for lightweight WaSR-T.

As mentioned in the annotation process for the image dataset, the network's target output consists of three class labels: boat, sea, and sky. The objective of the proposed network for detecting a boat approaching the TEWS is to identify the area of pixels labeled by the output network that corresponds to a class label of the boat's ground truth, with the background image serving as the label class: sea or the interface between sea and sky. The performance of lightweight WaSR-T in comparison with the original WaSR-T was evaluated using visual quality assessments [35], [36] with criteria as follows:

- If the output of the network produces a class label of a boat that perfectly overlaps with the location and an area of the pixels label of a boat is aligned to the boat ground truth, it is a subjective assessment as a true positive (TP).
- If the output of the network produces a class label of a boat with insufficient overlap with the location, and the area of the pixels labeled as a boat is slightly spread relative to the boat ground truth, it is a subjective assessment as a FP.

- If the output of the network produces a class label of a boat that lies outside the location and an area of the pixels label of a boat is spread relative to the boat ground truth, it is a subjective assessment as a false negative (FN).

Overall metrics to measure model network performance between lightweight WaSR-T and original WaSR-T were evaluated using $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$ based on overall testing of image datasets.

4. RESULTS AND DISCUSSION

Unlike previous approaches that focus on enhancing existing object detection methods by introducing new algorithms while using the same dataset, this study takes a different direction. We evaluate the performance of a pre-established object detection model by applying it to an entirely new dataset distinct from the one used during its original development. The only modification made to the model is the replacement of its backbone encoder: ResNet-101 is substituted with MobileNetV3. This change is intended to assess the impact on detection speed, particularly in the context of deployment on resource-constrained devices. Prior research has demonstrated that using a lighter backbone encoder can significantly accelerate the detection process [6], although this often comes at the cost of reduced accuracy when operating on low-power hardware.

In line with the methodology presented in [6], we conducted a comparative analysis using MobileNet V3 a more lightweight encoder backbone than ResNet-101 and evaluated both architectures on a newly introduced open ocean dataset that had not previously been available. Both networks' performance to detect boats was evaluated when testing 140 image datasets run using laptop computers with an Intel Core i7-10510U (quad-core HT 1.8 GHz, turbo 4.9 GHz) and 16 GB RAM. While running the program, we evaluate some computational load parameters. i.e., percentage of CPU resources used by a process (% CPU), percentage of physical memory used by a process (% memory), total processing time and rate segmentations per iteration (s/it) in the context produces a class label of boat to testing all image datasets. The summary of computational load parameters is shown in Table 2.

Table 2. Summary of computational load parameters between lightweight (L)-WaSR-T and WaSR-T

| Model | Testing images | CPU (%) | Memory (%) | Total processing time | Rate (s/it) |
|----------|----------------|---------|------------|-----------------------|-------------|
| WaSR-T | 140 | 190 | 13.2 | 1:13:56 | 20.07 |
| L-WaSR-T | 140 | 160 | 4.3 | 0:02:45 | 1.33 |

Furthermore, the example quantitative results of lightweight WaSR-T and original WaSR-T from training image datasets of the open sea in Indonesia's territory and an open website are sequentially shown in Figure 7. The four target frame images from each type of dataset were evaluated using visual quality assessment to determine the TP, FP, and FN rates. The summary of the visual quality assessment of the original WaSR-T and lightweight WaSR-T, which detects the class label of a boat from eight testing image datasets as shown in Table 3. The visual quality assessment of both networks to detect the class label of a boat from eight testing image datasets as shown in Figure 7 was tabulated in Table 3, which describes the output of both networks based on criteria TP, FP, and FN. The summary of qualitative results, including lightweight WaSR-T and WaSR-T for all training image datasets, is shown in Table 4.

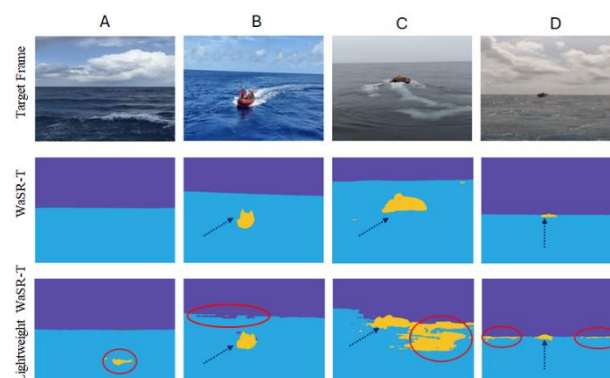


Figure 7. Example of qualitative results of lightweight WaSR-T and original WaSR-T from a testing image dataset

Based on the qualitative results in Tables 3 and 4, if the visual quality assessment is determined as TP, lightweight WaSR-T produces a class label for a boat, similar to the original WaSR-T. The area of the pixel label of the boat is completely detected. If the network output results in a visual quality assessment of FP, lightweight WaSR-T produces an area of pixel labeled as a boat that is slightly spread or smaller relative to the ground truth of the boat label. In the visual quality assessment as FP, the number of false detections produced by the WaSR-T is smaller than that of the lightweight WaSR-T. The most common source of false detections is due to the reflection of the water on the sea or the interface between the sea and the sky. In this case, the area of the pixel label of the boat is still detected. Therefore, if the class label of the boat can be detected, this information can still be used to generate an alarm. Hence, the quantitative results of lightweight WaSR-T, along with subjective assessment in terms of TP and FP, are useful for intelligent computer vision in TEWS.

Based on this condition, we also proposed an additional evaluation criterion to measure whether the network could generate an alarm due to unauthorized boats approaching TEWS. The sensitivity of generating a TEWS alarm is formulated as the ratio between the sum of TP and FP and the total training image dataset. The sensitivity of generating TEWS alarms for WaSR-T and lightweight WaSR-T performance results are 95.71% and 90.00%, respectively. Although the sensitivity of generating the TEWS alarm of WaSR-T is slightly better than that of lightweight WaSR-T, the computational load of lightweight WaSR-T is significantly lower than that of WaSR-T. Based on the testing results tabulated in Table 2, lightweight WaSR-T required less memory, at 32.57%, and the total processing time was reduced to 0.0761% compared to the original WaSR-T.

Table 3. The summary of visual quality assessment of original WaSR-T and lightweight WaSR-T

| Target frame | Assessment results WaSR-T | Assessment results L-WaSR-T | Remarks |
|--------------|---------------------------|-----------------------------|---|
| A | TP | FN | Lightweight WaSR-T produces the false class label of a boat that is a reflection of light on the sea. (marked with the red circle). |
| B | TP | FP | Lightweight WaSR-T detected the label area of the boat (marked with the dot arrow line). In this case, lightweight WaSR-T detected the class label of the sky with the sea on the interface between sea and sky (marked with the red circle). |
| C | TP | FP | Lightweight WaSR-T detected the label area of the boat (marked with the dot arrow line). However, the pixel label of a boat is slightly spread relative to the ground true label of the boat. In this case, a reflection of light on the sea is detected as the pixel label of a boat (marked with the red circle). |
| D | TP | FP | Lightweight WaSR-T detected the label area of the boat (marked with the dot arrow line). However, the pixel label of a boat is slightly spread relative to the ground true label of the boat. In this case, a reflection of light on the sea is detected as the pixel label of a boat (marked with the red circle). |

Table 4. Qualitative results lightweight WaSR-T and WaSR-T for all training image datasets

| Model | TP | FP | FN | Precision (%) | Recall (%) |
|--------------------|----|----|----|---------------|------------|
| WaSR-T | 94 | 40 | 15 | 70.15 | 86.24 |
| Lightweight WaSR-T | 77 | 49 | 14 | 61.11 | 86.14 |

5. CONCLUSION

The development and implementation of the proposed lightweight WaSR-T networks to detect unauthorized boats approaching TEWS as an integral part of an intelligent computer vision system in an open sea domain have been discussed. Based on the quantitative results and evaluation of the computational load, lightweight WaSR-T, designed as the central part of an intelligent computer vision system in TEWS, showed promise for further implementation in computational devices with a small architectural footprint. Future work will focus on real-world testing of the lightweight WaSR-T network on buoy platforms in diverse water environments. Additional experiments under extreme weather conditions should be conducted to ensure the network's robustness. Furthermore, to improve lightweight WaSR-T performance, we enhance dataset augmentation, preprocessing techniques, and the implementation of lightweight WaSR-T, which currently utilizes a large MobileNetV3 with parallel processing on a small GPU architecture, such as the NVIDIA Jetson Nano.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Mr. Andi Kurnianto for his invaluable assistance in data preprocessing and annotation during the early stages of this study. We also thank Mr. Arief

Rufiyanto for his support in proofreading the manuscript and providing helpful suggestions to improve clarity and coherence. Special thanks to Mr. Dedy Irawan for his insightful discussions on model architecture selection, which greatly influenced the direction of our experiments.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|----------------------|---|---|----|----|----|---|---|---|---|---|----|----|---|----|
| Wayan Wira Yogantara | ✓ | ✓ | ✓ | | | ✓ | ✓ | | ✓ | | ✓ | | ✓ | |
| Suprijanto | ✓ | ✓ | | | | ✓ | | | ✓ | | ✓ | ✓ | | |
| Anak Agung Ngurah | ✓ | ✓ | | | ✓ | | | | | ✓ | | ✓ | | |
| Ananda Kusuma | | | | | | | | | | | | | | |
| Yuki Istianto | | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ | | | | |

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

This paper does not involve people or animals; no investigation has involved human subjects. Therefore, the authors did not seek approval from any institutional review board.

DATA AVAILABILITY





The data that support the findings of this study are available from the corresponding author, [S], upon reasonable request.

REFERENCES





- [1] L. Zhao, F. Yu, J. Hou, P. Wang, and T. Fan, "The role of tsunami buoy played in tsunami warning and its application in South China Sea," *Theoretical and Applied Mechanics Letters*, vol. 3, no. 3, 2013, doi: 10.1063/2.1303202.
- [2] B. Bovcon and M. Kristan, "WaSR-A water segmentation and refinement maritime obstacle detection network," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 12661–12674, Dec. 2022, doi: 10.1109/TCYB.2021.3085856.
- [3] F. E. T. Schöller, M. Blanke, M. K. Plenge-Feidenhans'l, and L. Nalpantidis, "Vision-based object tracking in marine environments using features from neural network detections," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 14517–14523, 2020, doi: 10.1016/j.ifacol.2020.12.1455.
- [4] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 1993–2016, Aug. 2017, doi: 10.1109/TITS.2016.2634580.
- [5] D. Qiao, G. Liu, T. Lv, W. Li, and J. Zhang, "Marine vision-based situational awareness using discriminative deep learning: A survey," *Journal of Marine Science and Engineering*, vol. 9, no. 4, Apr. 2021, doi: 10.3390/jmse9040397.
- [6] M. Teršek, L. Žust, and M. Kristan, "eWaSR—an embedded-compute-ready maritime obstacle detection network," *Sensors*, vol. 23, no. 12, Jun. 2023, doi: 10.3390/s23125386.
- [7] A. F. Abbas, U. U. Sheikh, F. T. Al-Dhief, and M. N. H. Mohd, "A comprehensive review of vehicle detection using computer vision," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 3, pp. 838–850, Jun. 2021, doi: 10.12928/TELKOMNIKA.v19i3.12880.

- [8] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2021, doi: 10.1109/TPAMI.2021.3059968.
- [9] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, "Deep learning in computer vision: a critical review of emerging techniques and application scenarios," *Machine Learning with Applications*, vol. 6, Dec. 2021, doi: 10.1016/j.mlwa.2021.100134.
- [10] L. Peng, H. Wang, and J. Li, "Uncertainty evaluation of object detection algorithms for autonomous vehicles," *Automotive Innovation*, vol. 4, no. 3, pp. 241–252, Aug. 2021, doi: 10.1007/s42154-021-00154-0.
- [11] Y. Peng, Y. Qin, X. Tang, Z. Zhang, and L. Deng, "Survey on image and point-cloud fusion-based object detection in autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 22772–22789, Dec. 2022, doi: 10.1109/TITS.2022.3206235.
- [12] N. Gengeç, O. Eker, H. ÇeviKalp, A. Yazici, and H. S. Yavuz, "Visual object detection for autonomous transport vehicles in smart factories," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 29, no. 4, pp. 2101–2115, Jul. 2021, doi: 10.3906/ELK-2008-62.
- [13] S. A. Khalil, S. Abdul-Rahman, S. Mutalib, and N. M. A. A. Dazlee, "Object detection for autonomous vehicles with sensor-based technology using YOLO," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 1, pp. 129–134, Mar. 2022, doi: 10.18201/ijisae.2022.276.
- [14] M. R. Haque, M. M. Islam, K. S. Alam, H. Iqbal, and M. E. Shaik, "A computer vision based lane detection approach," *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 3, pp. 27–34, Mar. 2019, doi: 10.5815/ijigsp.2019.03.04.
- [15] B. Bovcon and M. Kristan, "A water-obstacle separation and refinement network for unmanned surface vehicles," in *IEEE International Conference on Robotics and Automation*, May 2020, pp. 9470–9476. doi: 10.1109/ICRA40945.2020.9197194.
- [16] Q. Cai, Q. Wang, Y. Zhang, Z. He, and Y. Zhang, "LWDNet-A lightweight water-obstacles detection network for unmanned surface vehicles," *Robotics and Autonomous Systems*, vol. 166, Aug. 2023, doi: 10.1016/j.robot.2023.104453.
- [17] B. Bovcon, J. Muhovic, D. Vranac, D. Mozetic, J. Pers, and M. Kristan, "MODS-A USV-oriented object detection and obstacle segmentation benchmark," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13403–13418, Aug. 2022, doi: 10.1109/TITS.2021.3124192.
- [18] L. Zust and M. Kristan, "Temporal context for robust maritime obstacle detection," in *IEEE International Conference on Intelligent Robots and Systems*, Oct. 2022, pp. 6340–6346. doi: 10.1109/IROS47612.2022.9982043.
- [19] B. Bovcon, J. Muhovic, J. Pers, and M. Kristan, "The MaSTr1325 dataset for training deep USV obstacle detection models," in *IEEE International Conference on Intelligent Robots and Systems*, Nov. 2019, pp. 3431–3438. doi: 10.1109/IROS40897.2019.8967909.
- [20] J. Lauterjung and H. Letz, "10 years Indonesian Tsunami early warning system : experiences, lessons learned and outlook," *Potsdam: GFZ German Research Centre for Geosciences*, 2017, doi: 10.2312/GFZ.7.1.2017.001.
- [21] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: a survey," *Applied Sciences*, vol. 12, no. 18, Sep. 2022, doi: 10.3390/app12188972.
- [22] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.
- [23] B. E. Soylu, M. S. Guzel, G. E. Bostanci, F. Ekinci, T. Asuroglu, and K. Acici, "Deep-learning-based approaches for semantic segmentation of natural scene images: a review," *Electronics*, vol. 12, no. 12, p. 2730, Jun. 2023, doi: 10.3390/electronics12122730.
- [24] Y. Harjoseputro, I. P. Yuda, and K. P. Danukusumo, "MobileNets: efficient convolutional neural network for identification of protected birds," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 10, no. 6, pp. 2290–2296, Dec. 2020, doi: 10.18517/ijaseit.10.6.10948.
- [25] Y. Yang and J. Han, "Real-time object detector based MobileNetV3 for UAV applications," *Multimedia Tools and Applications*, vol. 82, no. 12, pp. 18709–18725, May 2023, doi: 10.1007/s11042-022-14196-x.
- [26] T. Deng and Y. Wu, "Simultaneous vehicle and lane detection via MobileNetV3 in car following scene," *PLoS ONE*, vol. 17, no. 3 March, Mar. 2022, doi: 10.1371/journal.pone.0264551.
- [27] M. A. E. Alkhalisy and S. H. Abid, "Abnormal behavior detection in online exams using deep learning and data augmentation techniques," *International journal of online and biomedical engineering*, vol. 19, no. 10, pp. 33–48, Aug. 2023, doi: 10.3991/ijoe.v19i10.39583.
- [28] M. Prajapati, S. K. Baliarsingh, J. Hota, P. P. Dev, and S. Das, "Retinal and semantic segmentation of diabetic retinopathy images using MobileNetV3," in *ICCECE 2023-International Conference on Computer, Electrical and Communication Engineering*, Jan. 2023, pp. 1–6. doi: 10.1109/ICCECE51049.2023.10085191.
- [29] L. Zhao and L. Wang, "A new lightweight network based on MobileNetV3," *KSH Transactions on Internet and Information Systems*, vol. 16, no. 1, pp. 1–15, Jan. 2022, doi: 10.3837/tiis.2022.01.001.
- [30] A. Howard *et al.*, "Searching for mobileNetV3," in *Proceedings of the IEEE International Conference on Computer Vision*, Oct. 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.
- [31] A. G. Howard *et al.*, "MobileNets: efficient convolutional neural networks for mobile vision applications," *arXiv-Computer Science*, pp. 1–9, 2017.
- [32] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.
- [33] K. Wada, "Image polygonal annotation with Python (polygon, rectangle, circle, line, point and image-level flag annotation)," *Kaggle*. 2023. [Online]. Available: <https://github.com/Wkentaro/Labelme>
- [34] L. Zust, "Temporal WaSR-T model for maritime obstacle detection via semantic segmentation," *Kaggle*, 2023. [Online]. Available: <https://github.com/lojzezust/WaSR-T>
- [35] Z. Chen and H. Zhu, "Visual quality evaluation for semantic segmentation: subjective assessment database and objective assessment measure," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5785–5796, Dec. 2019, doi: 10.1109/TIP.2019.2922072.
- [36] G. Csúrká, D. Larlus, and F. Perronnin, "What is a good evaluation measure for semantic segmentation?," in *BMVC 2013-Electronic Proceedings of the British Machine Vision Conference 2013*, 2013, pp. 32.1–32.11. doi: 10.5244/C.27.32.





BIOGRAPHIES OF AUTHORS

Wayan Wira Yogantara     received Dipl.-Ing. degree in Communication Engineering 1998 from Braunschweig/Wolfenbuettel University of Applied Sciences, Germany. He holds a position as Junior Engineer at Research Center for Electronics, National Research and Innovation Agency of Indonesia. His research interest in the fields of underwater acoustic communication, maritime electronics navigation systems, and artificial intelligence. He can be contacted at email: 23821005@mahasiswa.itb.ac.id or waya001@brin.go.id.







Suprijanto     received the B.Sc. degree in Engineering Physics in 1995 and the M.Sc. degree of Technology in Instrumentation and Control in 1997 from Institut Teknologi Bandung, Indonesia. He completed Ph.D. degree in Medical Engineering from Faculty of Applied Science, University of Technology Delft, The Netherlands. He holds a position of a Professor and Head of Medical Instrumentation Laboratory at Faculty of Industrial Technology, Institut Teknologi Bandung, Indonesia. His research interest is in the fields of medical engineering, artificial intelligence, instrumentation, control, and multiphysics modelling. He can be contacted at email: supri89@itb.ac.id.



Anak Agung Ngurah Ananda Kusuma     received the B.Eng. (Hons.), M.Eng., and Ph.D. degrees in 1993, 1998, 2005, from the University of Tasmania, RMIT University, and the University of Melbourne, Australia. He holds a position as a Principal Engineer and Research Group Leader at the Research Center for Telecommunication, National Research and Innovation Agency. His research interests include routing algorithms, resource allocation problems, QoS/QoE estimation, and various problems related to network protocols. He can be contacted at email: anak001@brin.go.id.



Yuki Istianto     received his M.Sc. degree from Korea Advanced Institute of Science and Technology, South Korea. He is currently a Junior Engineer at Research Center for Artificial Intelligence and Cybersecurity, National Research and Innovation Agency of Indonesia. His current research interests include biometrics, aquaculture and artificial intelligence. He can be contacted at email: yuki001@brin.go.id.