IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 29, 2019, accepted December 5, 2019, date of publication December 13, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959312

Vision-Based Real-Time Obstacle Segmentation Algorithm for Autonomous Surface Vehicle

HANGUEN KIM[1], (Member, IEEE), JUNGMO KOO[2], (Student Member, IEEE), DONGHOON KIM[1], (Member, IEEE), BYEOLTEO PARK[1], (Member, IEEE), YONGGIL JO[1], HYUN MYUNG[2], (Senior Member, IEEE), AND DONGHWA LEE[3]

[1]Seadronix Corporation, Daejeon 34051, South Korea
[2]School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, South Korea
[3]Division of Computer and Communication Engineering, Daegu University, Gyeongsan 38453, South Korea

Corresponding authors: Hyun Myung (hmyung@kaist.ac.kr) and Donghwa Lee (leedonghwa@daegu.ac.kr)

This work was supported in part by the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Ministry of Science and Information and Communication Technology (MSIT), Korean Government, (Research on Adaptive Machine Learning Technology Development for Intelligent Autonomous Digital Companion) under Grant 2016-0-00563, and in part by the Korea Ministry of Land, Infrastructure and Transport (MOLIT) as ''Innovative Talent Education Program for Smart City'' and BK21+.

ABSTRACT Among various sensors used to recognize obstacles in marine environments, vision sensors are the most basic. Vision sensors are significantly affected by the surrounding environment and cannot recognize distant objects. However, despite these drawbacks, they can detect objects that radars cannot detect in nearby regions. They can also recognize small obstacles such as boats that are not equipped with an automatic identification system (AIS) or buoys. Thus, vision sensors and radar can be used in a complementary manner. This paper proposes a vision sensor-based model, called Skip-ENet, for recognizing obstacles in real time. Compared with ENet, the amount of computation is not significantly higher. Further, Skip-ENet can segment complex marine obstacles effectively by increasing the values for the class accuracy and mean Intersection of Union (mIoU). Moreover, this model enables even low-cost embedded systems to compute 10 or more frames per second (fps). The superiority of the proposed model was verified by comparing its performance with that of the conventional segmentation models, MobileNet, ENet, and DeeplabV3+.

INDEX TERMS Autonomous surface vehicle, computer vision, deep learning, obstacle segmentation, ship navigation.

I. INTRODUCTION

In recent years, major shipbuilding companies such as Rolls-Royce and Kongsberg have announced their roadmaps for developing autonomous surface vehicles, and are presently actively engaged in the associated research and development. In order to autonomize surface vehicles just as in autonomous vehicles, a combination of various sensors, such as radar and LiDAR (Light Detection And Ranging) sensor, is necessary. A drawback for radar is that there are areas known as radar shadows, where radar is not effective. Conversely, in marine environments, LiDAR has the disadvantage of having more noise than in road environments owing to diffused reflection from water. Hence, most researchers use a vision sensor as the fundamental sensor to overcome the drawbacks of other sensors. Vision sensors have the disadvantages of being significantly affected by the surroundings in a marine environment and inability to recognize distant objects. However, they can detect objects in nearby areas that radars cannot discover. Objects equipped with an automatic identification system (AIS) can be easily recognized as obstacles. However, even without the AIS, the vision sensor can recognize a buoy or small ship as an obstacle. Many researchers have proposed a variety of algorithms using vision sensors. So we developed a special device to collect vision data. This device can collect data simultaneously including GPS, IMU as well as vision sensor. It is made of waterproof material that can be used by the boat, and it is also possible to record for a long time. Using the device, we collected data from a variety of marine environments. This data includes diverse

The associate editor coordinating the review of this manuscript and approving it for publication was Habib Ullah.

seasons and includes both daytime and nighttime data. It has been collected since February 2017 and has collected a total of 2.5 million data. Of these data, we segmented the image into 5 classes and completed about 4000 images for training data. In this paper, we propose a system that recognizes obstacles using a real-time image segmentation algorithm based on deep learning. Unlike other obstacle detection/recognition algorithms, the proposed algorithm can classify obstacles in a relatively high-resolution input image at the pixel level. Further, this algorithm can robustly estimate obstacles even when sea clutters or sun glinting from the ocean surface are present in the input images. We propose an improved architecture called Skip-ENet with ENet [21] as a baseline. Skip-ENet has a good performance for recognizing small objects even when using a simple architecture. In addition, Skip-ENet can be operated at speeds of 10 fps or higher on NVIDIA TX2 embedded board, so it can be installed in the unmanned surface vehicle and used in real time. In summary, our contributions are as follows:

- Various marine environmental data were collected. This data spans over diverse seasons and also daytimes and nighttimes.
- We designed a special equipment for data collection. With the device, color image, IR image, GPS coordinates, and IMU data can be saved at the same time.
- We designed a novel network architecture for image segmentation. We proposed an improved ENet architecture called Skip-ENet.
- The proposed network is implemented on the embedded board for actual operation of an unmanned surface vehicle. It runs at speeds of 10 fps or more on the TX2 board.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 describes the data collection device and the collected data used to study the system; it also outlines the proposed system. Section 4 compares the performance and computation speed of the proposed algorithm with those of other segmentation models. Finally, Section 5 presents concluding remarks and outlines future research and development plans.

II. RELATED WORK

Roads for cars have rules that can be recognized visually, such as lanes and traffic lights. In contrast, there are no such rules in marine environments. Vision sensor-based obstacle recognition has to be performed in such open spaces without any rules. Therefore, studies have been conducted on limiting the detection area in order to improve performance while maintaining a certain computation level. For example, Wang *et al.* located the horizon by connecting the points that had the largest changes in gradient based on a uniform grid and excluded the sky from the computation. Afterward, the obstacles were detected using saliency values [1], [2]. Furthermore, Wang *et al.* extended a single vision sensor into a stereo vision sensor to enhance the obstacle detection performance [3] and adopted an HD high-resolution vision sensor so that obstacles more than 200 m away could be recognized [4], [5]. Likewise, Oren used the canny edge detection method to find the horizon in order to reduce the amount of computation and detected interest points of obstacles through the thinning computation of areas where the pixel values changed [6]. However, the methods cited above could not detect obstacles above the horizon. In addition, the solutions lacked robustness because they could not be utilized when it was difficult to locate the horizon owing to the camera installation angle or the marine environment. Kristan *et al.* solved this problem by adopting an obstacle image map based on the Gaussian mixture model, categorizing what used to be separated into two areas (sky and sea) by the horizon into three regions (sky, terrain, sea), and applying the Markov random field technique to segment the obstacles [7]. Bovcon *et al.* fused the algorithm proposed by Kristan *et al.* with the positional information of the unmanned surface vehicle, which is measured by an inertial measurement unit (IMU), to improve the segmentation of the obstacle image map [8]. In addition, there are algorithms to remove sea clutters or sun glints from the sea surface [9]. Kim *et al.* [10] and Koo *et al.* [11] used a vision sensor in an UAV to recognize jellyfish in the marine environment and developed a deep learning based algorithm to accurately recognize jellyfish. In the case of marine environment, synthetic data generation method is sometimes used to overcome the problem of lack of real image data [12].

Neural networks are used in various fields, such as in pixel grouping or object tracking. Rota *et al.* used a multilayer perceptron (MLP) for grouping particles [14]. Likewise, Ullah *et al.* used a combination of HMM and deep features for multi-target tracking [15]. In these two cases, feature extraction proceeds individually and neural networks were used for classification. Unlike this method, Numerous studies have been conducted on deep learning-based segmentation, beginning with fully convolutional networks (FCNs) [16]. Very small objects can be recognized based on the predicted heat map from the network using FCNs [13]. Currently, the encoder-decoder combination architecture provides the best performance, with examples such as the Pyramid Scene Parsing Network (PSPNet) [17] and DeepLabV3+ [18]. These algorithms can segment even small objects accurately using multi-scale input images. However, these network models are all based on the residual network (ResNet; thus, they cannot perform real-time computations [19]. To rectify this issue, Zhao *et al.* proposed a cascade architecture-based ICNet [20]. However, although their solution can process high-resolution images in real time, it still cannot be used in embedded systems such as NVIDIA Jetson. A network architecture for mobile application processor (AP) has also been proposed. Paszke *et al.* proposed ENet, which is now one of the most widely used network architectures [21]. ENet significantly reduces the input images at the early stage to enhance computation speed. Further, ENet adopts an architecture similar to that of ResNet in order to maintain the performance. Compared with SegNet [22], ENet exhibits similar

FIGURE 1. Prototype of the marine data collection device. Data such as visible light, infrared radiation (IR), inertial measurement unit (IMU), and GPS are saved simultaneously.

performance and significantly faster computation speed. Howard *et al.* also proposed MobileNet [23], which utilizes the depthwise separable convolution method to significantly reduce the amount of computation of the convolution layer without much degradation in performance.

Among various network models that can be used from a mobile AP, the ENet architecture was used as the basis for this study. Further, the network architecture was modified, and the features of SkipNet added [24] to enhance the performance even more. Similar to the features of ResNet, SkipNet is designed such that a layer can be skipped depending on the complexity of the object being classified. Hence, loss of data due to the encoder is minimized without significantly increasing the amount of computation, and the Intersection over Union (IoU) value can be improved efficiently. The performance and computation speed of the proposed network model, called Skip-ENet (or simply, Skip-ENet), were evaluated by comparing them with those of MobileNet, which utilizes the U-Net architecture [25], as well as with ENet and DeeplabV3+.

III. METHODS

A. DATA COLLECTION AND MANAGEMENT

In order to design an image-based obstacle recognition algorithm, we first fabricated the marine data collection device shown in Figure 1. The data collection device was designed such that images can be recorded both in the daytime and nighttime and high-resolution visible light and infrared radiation (IR) images can be saved simultaneously. Further, both the location where the image was collected and the six degrees of freedom positional information of the image are saved at the same time. Since February 2017, data have been collected along the western and southern coasts of the Republic of Korea, where islands and aquafarms are predominantly distributed. To date, approximately 2.5 million valid images with position and orientation information have been obtained. A valid image is one in which there are variations in the information owing to environmental changes—such as changing from day to night or from sunny to rain—or a change in information, such as when a boat or buoy moves in the same place. Examples of collected valid images are shown in Figure 2.

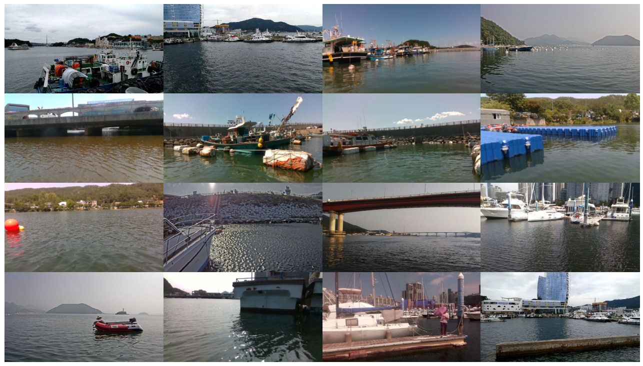


FIGURE 2. Examples of the images collected primarily from the coastal areas of the western and southern seas of the Republic of Korea. Only valid images containing various environmental information were used for learning.

Image labeling and augmentation needed to be performed to segment the collected images. We categorized the images into the classes shown in Table 1, so that they could be used for obstacle recognition and other applications. As explained in Table 1, information that is not needed for navigating the unmanned surface vehicle, such as the sky, birds, and airplanes, was not included in the learning process. Further, a separate class was provided for better segmentation of small obstacles. The results of the labeling task are shown in Figure 3. Augmentation of the images used for learning can be very effective even with a small number of images. Unlike cars, vessels can move, while being at the same location, in three dimensions and six degrees of freedom because of waves or tidal current. Hence, data that reflect the Affine transformation effect of the images can be obtained. In addition, we were able to double the amount of data, by reversing the labeled images along the y-axis, and utilized them for learning.

B. SKIP-ENET

When Paszke *et al.* proposed the ENet architecture, they applied various techniques to increase the computation speed while maintaining the performance. We analyzed the initial block and encoder-decoder bottleneck techniques proposed, and modified the architecture such that obstacles can be well segmented in the marine environment with the following characteristics:

1) COUNTERMEASURES FOR THE BLUR EFFECT DUE TO "HAEMOO" (SEA FOG)

"Haemoo" is a type of sea fog that occurs when the surrounding air is cooled as the air over a meteorologically warm

TABLE 1. Classification of marine obstacles.

Class	ID	Description
Sky and other (birds, airplanes, etc.)	0	Not used for learning.
Sea surface	1	Sun glints from the sea surface and sea clutters included
Vessels	2	Fishing boats, yachts, small boats, etc.
Small obstacles such as buoys	3	Small-sized fixed obstacles
Terrain	4	Coastal terrains such as island and wharf

Classification for the segmentation of the sea surface obstacle information. A separate class was used to categorize the small-sized obstacles so they can be well segmented.

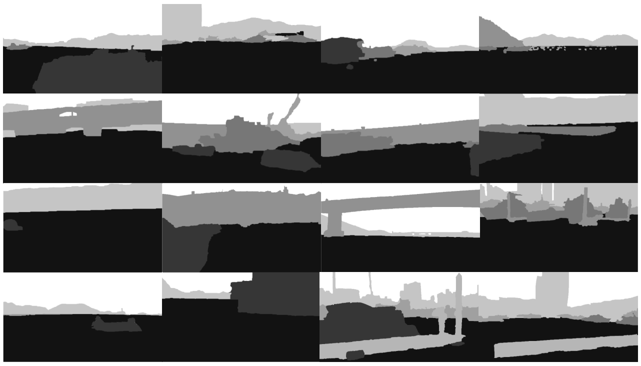


FIGURE 3. Results of the labeling task according to Table 1. Only the objects needed for autonomous navigation were labeled.

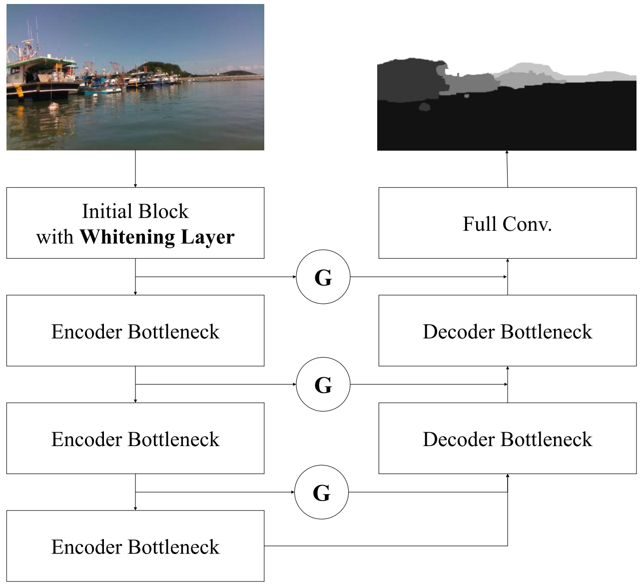


FIGURE 4. Architecture of the proposed Skip-ENet. The whitening layer is added to the initial block to remove the blur caused by sea fog. Further, the gate structure of SkipNet is adopted for segmentation of small obstacles, so the inference performance is improved without increasing the amount of computation significantly.

surface of the sea moves over a cold sea surface. Sea fogs often occur between April and October in all the seas of Korea. Images collected during this period are blurry, as if a blur effect had been applied to them. Hence, additional computation is needed to deblur the image before it is fed into the convolution layer, so that the encoder block can robustly segment the characteristics. We added a layer that performs whitening transformation [26] to the initial block; thus, this layer along with the max pooling layer produce a deblur effect.

2) METHOD FOR SEGMENTING SMALL OBSTACLES SUCH AS BUOY AND SIMPLE STRUCTURES

The probability of objects, such as fishing nets or marine structures that can cause accidents, existing underneath buoys or simple structures floating on the sea is high. Hence, these objects must be recognized in advance and avoided when navigating vessels. In order to segment small objects without increasing the amount of computation, we adjusted the size of the receptive field in the dilated convolution computation [27]. With a large receptive field of dilated convolution, compression of the feature can proceed quickly, even if it is not a very deep network. Also, we adopted the SkipNet architecture. The SkipNet architecture was applied to the ENet architecture because, as mentioned previously, the loss of information due to the encoder can be reduced without increasing the computational amount significantly, thereby increasing the IoU value. Furthermore, the shape information lost can be supplemented by restoring the size of the results using the upsampling computation at the decoder block. For ENet, the encoder-decoder structure is not symmetric, so they were connected using gates, as shown in Figure 4. Unlike SkipNet, the gates were constructed to perform only the resizing computation, so the increase in the amount of computation is minimized. When one looks at the proportions of the objects defined in Table 1 in the images of marine environments, the proportions of the sea surface and the sky are high. Therefore, during the learning process, small obstacles that have relatively small proportions are given additional arbitrary weight when calculating the class weight, and the errors are updated. The formula for calculating the class weight is as follows:

$$w_n = \frac{1}{\ln(c_n + p_n)}. \tag{1}$$

This method is different from that used by ENet in that the hyper parameter c can be assigned to each class. The value of the w_n has the range [1–50] and n means class ID. The c_n

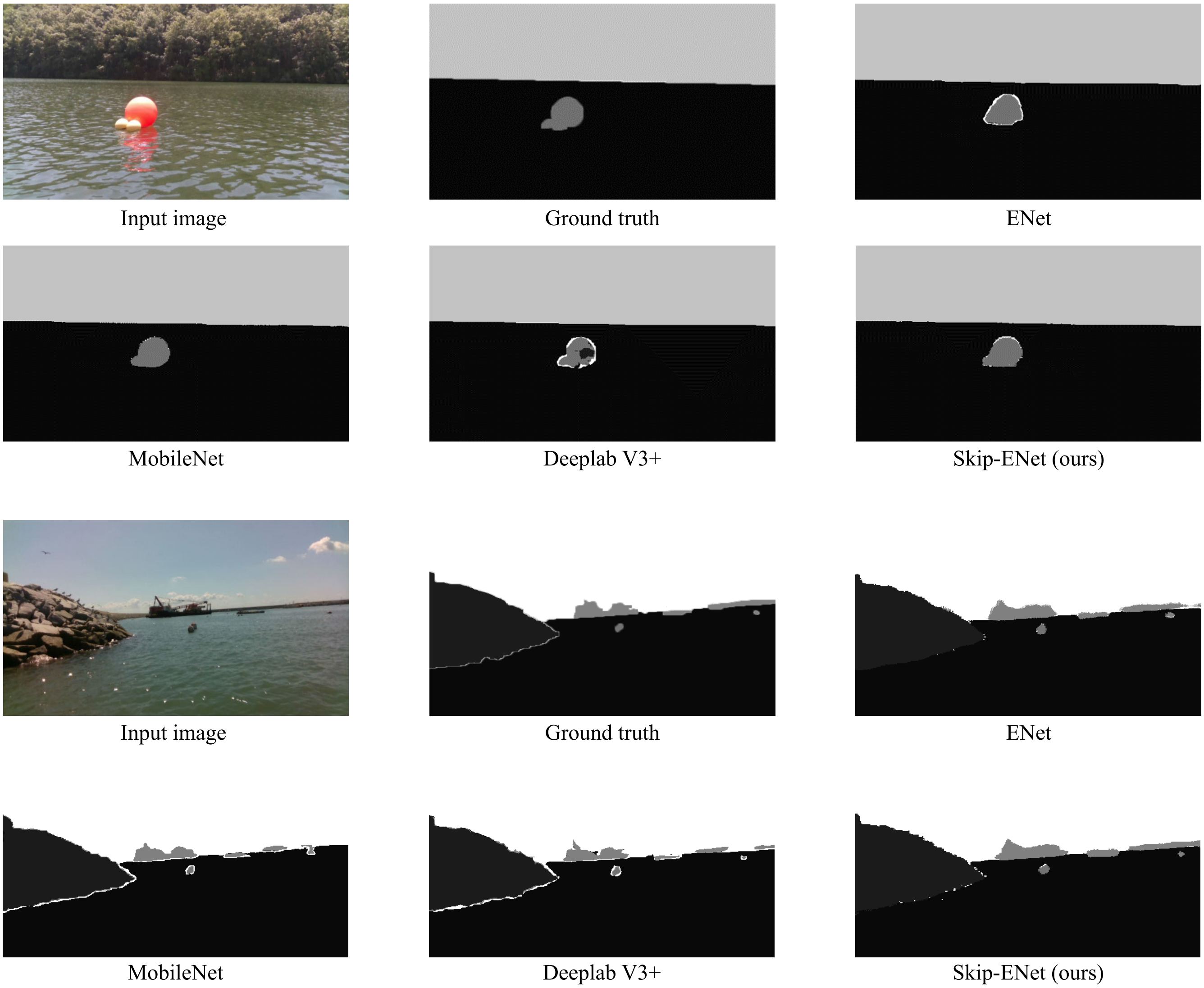


FIGURE 5. Comparison results. Segmented images from ENet, MobileNet, Deeplab V3+, and Skip-ENet compared to the ground truth for two input images.

is a constant value of each class and the p_n is obtained by calculating the average number of pixels each class occupies in the input image.

3) NETWORK DETAILS

Encoder block is composed of three sub blocks. The first is a regular block consisting of a basic residual architecture. This block compresses the feature using a normal convolution. The second is an asymmetric block. Using two convolution filters asymmetrically, $n \times n$ convolution can be factorized into $n \times 1$ and $1 \times n$ convolution filters. This block increases the computational efficiency of the network. The third is a dilated block. It spreads the compressed information in the feature map. Each encoder exploits three blocks sequentially and repeatedly. The decoder is similar to a regular encoder block except that the upsampling process is added. A spatial dropout is applied between all blocks to prevent divergence that can occur when learning. Because, when using Adam optimizer, the network can converge to local optima, the spatial dropout also has the effect of preventing convergence to local optima.

IV. EXPERIMENTS

We compared the segmentation performance and computation amount of the proposed Skip-ENet to those of ENet, MobileNet (which uses U-Net as the base network), and the DeeplabV3+ algorithms. Of all the images collected in-house, 4000 images were used to learn all the networks, and 500 were used for evaluation. All the learning networks utilized the Adam optimization algorithm [28], and the loss function was defined using proposed Equation (1) as follows:

$$-\sum_{n=1}^{M_{class}} w_n y_n \log(y'_n). \tag{2}$$

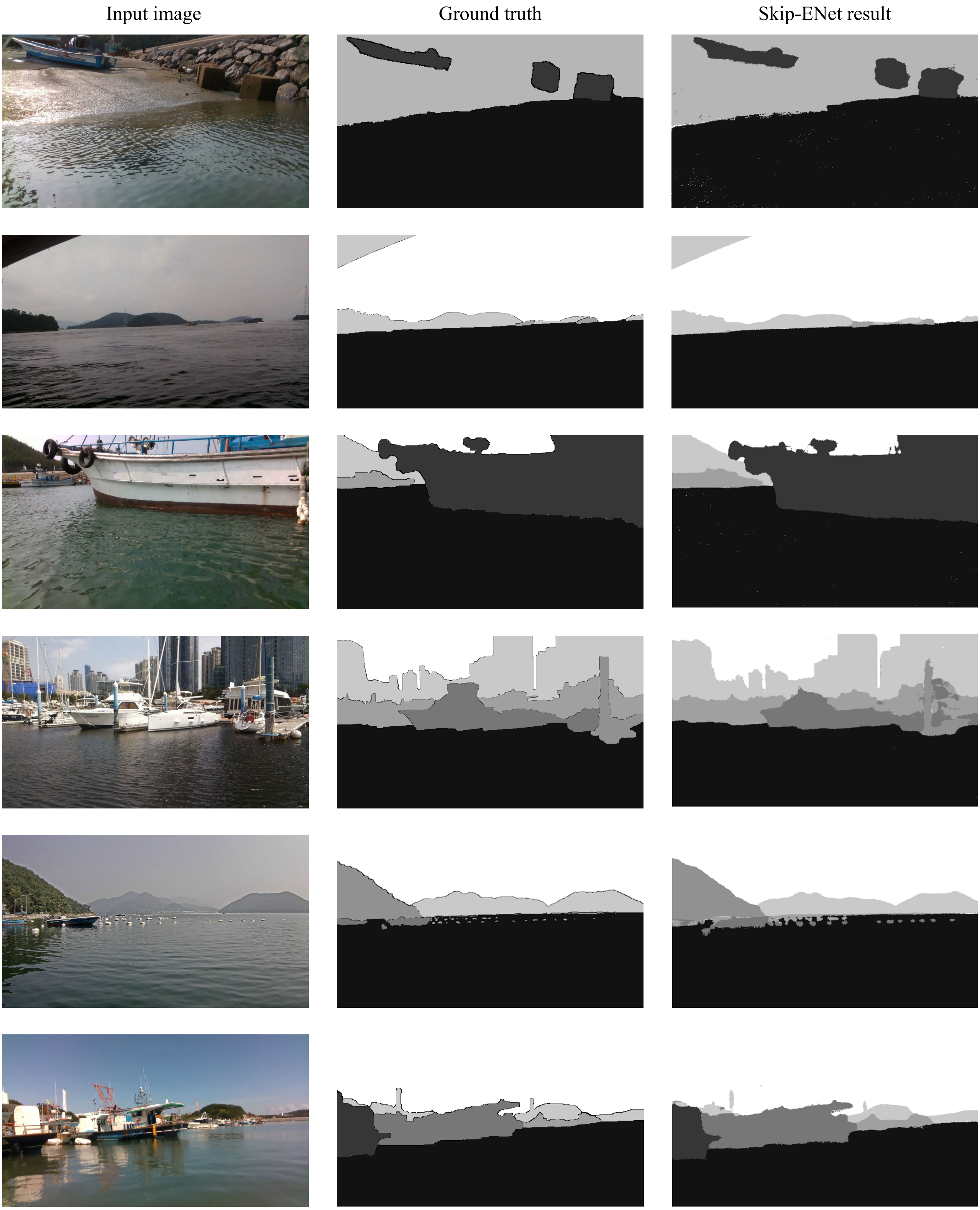


FIGURE 6. Skip-ENet results and ground truths comparison according to input images.

where M is the class value, y is the value of the true label, and y' is the label obtained by inference. Equation 2 is a cross-entropy loss function that reflects w_n. Learning, performance evaluation, and measurement of the computation amount for all the models were conducted using the TensorFlow [29] library, which supports CUDA. The average

TABLE 2. Results of computation time and amount of operations.

Model	Time(ms)	FLOPs(G)
ENet	18.9	2.5
MobileNet	88.7	72.0
DeeplabV3+	134.2	396.0
Skip-ENet (ours)	21.4	2.8

Results of computation time and amount of operations on NVIDIA GTX 1070 for a resolution of 600 × 336. NVIDIA Jetson TX2 can run at a speed of over 10 FPS using ENet or Skip-ENet, and even MobileNet based on U-Net can be run if the target frame rate is reduced.

TABLE 3. Results of average class accuracy and mean Intersection of Union (mIoU) evaluation.

Model	Avg. Class Accuracy	mIoU
ENet	58.19	13.98
MobileNet	58.29	15.88
DeeplabV3+	62.33	16.32
Skip-ENet (ours)	60.08	16.06

Results of class accuracy and IoU evaluation for a resolution of 600 × 336. Compared to ENet, both the average class accuracy and IoU value have increased for Skip-ENet

class accuracy and IoU were used as the evaluation indices. An NVIDIA GTX 1070 was used to evaluate and compare the performance between DeeplabV3+ and the proposed model. Further, embedded usability was assessed by comparing the Giga floating point operations per seconds (GFLOPS) needed for each model.

A. ANALYSIS OF THE AMOUNT OF COMPUTATION

The resolution of the input image used for learning and performance evaluation was 600 × 336. The NVIDIA Jetson TX2 ran at 10 frames per second (FPS) using Skip-ENet. It was installed in a small unmanned surface vehicle and was set to a minimum resolution, so small obstacles of BLANK could be recognized. The evaluation results of the computation speed and computation amount using the evaluation data are shown in Table 2. The table shows that the amount of computation for Skip-ENet rose by about 13% compared to ENet, but the actual computation speed increased by around 2 ms, implying that Skip-ENet could be used in embedded systems. For MobileNet, a significant gain could not be verified because learning was performed based on U-Net; however, if the target frame rate (FPS) could be lowered, MobileNet could be run on an embedded system. DeeplabV3+ required a very large GFLOPs and could not run on the NVIDIA GTX 1070 at 10 FPS. In order to run DeeplabV3+ on the NVIDIA GTX 1070, either a high-performance GPU needs to be used, or optimization is required.

B. PERFORMANCE EVALUATION

Similar to the analysis of the computational amount, the average class accuracy and IoU were calculated for the 500 evaluation images. The results are presented in Table 3. The results in Table 3 show that the average class accuracy and mIoU value for Skip-ENet increased by 1.89% and 2.03%, respectively, when compared to ENet. These values are better than those for MobileNet, which is based on the U-Net, and they are close to the values for DeeplabV3+. The mIoU value is lower than the average class accuracy for all the evaluated models. This phenomenon occurs because the sea surface and terrain objects are very diverse and their shapes are very complex. This makes it difficult to label the pixel-level shape information using the ground truth data, which results in a certain level of error being included. Therefore, considering that we are working with marine data, these results can be viewed as good. The segmentation results of each algorithm that can be compared is shown in Figure 5. The ground truth labeling and segmentation results of each input image for the proposed Skip-ENet are shown in Figure 6.

V. CONCLUSION

In this paper, we proposed the vision sensor-based Skip-ENet model to recognize marine obstacles effectively. Skip-ENet's architecture is geared towards marine environments and their associated types of obstacles. Further, the amount of computation is not significantly increased compared with the ENet. In addition, the class accuracy and mIoU value showed increases, indicating that the performance of Skip-ENet is close to that of DeeplabV3+, which currently has one of the best performing architectures. Hence, complex marine obstacles can be segmented effectively, and computation of over 10 FPS can be performed on a low-cost embedded system. Thus, a low-cost vision sensor-based obstacle recognition system can be developed. In the future, the proposed model will be improved for commercialization. Further, an image enhancement technique will be incorporated to pre-process images so that the model can be used both day and night in all weather conditions. The model will also be enhanced such that the distances to obstacles can be estimated to track and avoid them. In addition, various technologies will be developed to increase the power ratio and efficiency of the obstacle recognition system.

ACKNOWLEDGMENT

(Hanguen Kim and Jungmo Koo contributed equally to this work.)

REFERENCES

[1] H. Wang, Z. Wei, S. Wang, C. S. Ow, H. T. Ho, and B. Feng, "A vision-based obstacle detection system for unmanned surface vehicle," in *Proc. IEEE Conf. Robot., Autom. Mechatronics (RAM)*, Sep. 2011, pp. 364–369.

[2] H. Wang, Z. Wei, C. S. Ow, K. T. Ho, B. Feng, and J. Huang, "Improvement in real-time obstacle detection system for USV," in *Proc. Int. Conf. Control Autom. Robot. Vis. (ICARCV)*, Dec. 2012, pp. 1317–1322.

[3] H. Wang and Z. Wei, "Stereovision based obstacle detection system for unmanned surface vehicle," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2013, pp. 917–921.

[4] H. Wang, X. Mou, W. Mou, S. Yuan, S. Ulun, S. Yang, and B.-S. Shin, "Vision based long range object detection and tracking for unmanned surface vehicle," in *Proc. IEEE Int. Conf. Cybern. Intell. Syst. (CIS) IEEE Conf. Robot., Automat. Mechatronics (RAM)*, Sep. 2015, pp. 101–105.

[5] X. Mou and H. Wang, "Wide-baseline stereo-based obstacle mapping for unmanned surface vehicles," *Sensors*, vol. 18, no. 4, pp. 1–14, 2018.

[6] G. Oren, "Automatic obstacle detection for USV's navigation using vision sensors," in *Robotic Sailing*. Berlin, Germany: Springer, 2011, pp. 127–140.

[7] M. Kristan, V. S. Kenk, S. Kovačič, and J. Perš, "Fast image-based obstacle detection from unmanned surface vehicles," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 641–654, Mar. 2016.

[8] B. Bovcon, R. Mandeljc, J. Perš, and M. Kristan, "Improving vision-based obstacle detection on USV using inertial sensor," in *Proc. 10th Int. Symp. Image Signal Process. Anal.*, Sep. 2017, pp. 1–6.

[9] P. Paccaud and D. A. Barry, "Obstacle detection for lake-deployed autonomous surface vehicles using RGB imagery," *PLoS ONE*, vol. 13, no. 10, pp. 1–24, 2018.

[10] H. Kim, J. Koo, S. Jung, J. U. Shin, S. Lee, and H. Myung, "Image-based monitoring of jellyfish using deep learning architecture," *IEEE Sensors J.*, vol. 16, no. 8, pp. 2215–2216, Apr. 2016.

[11] J. Koo, S. Jung, and H. Myung, "A jellyfish distribution management system using an unmanned aerial vehicle and unmanned surface vehicles," in *Proc. IEEE Underwater Technol.*, Feb. 2017, pp. 1–5.

[12] K. Kim and H. Myung, "Autoencoder-combined generative adversarial networks for synthetic image data generation and detection of jellyfish swarm," *IEEE Access*, vol. 6, pp. 54207–54214, 2018.

[13] K. Kim, J. Hyun, H. Kim, H. Lim, and H. Myung, "A deep learning-based automatic mosquito sensing and control system for urban mosquito habitats," *Sensors*, vol. 19, no. 12, pp. 2785–2795, 2019.

[14] P. Rota, H. Ullah, N. Conci, N. Sebe, and F. G. B. De Natale, "Particles cross-influence for entity grouping," in *Proc. 21st Eur. Signal Process. Conf.*, 2013, pp. 1–5.

[15] M. Ullah and F. A. Cheikh, "A directed sparse graphical model for multi-target tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 1897–1904.

[16] J. Long, E. Shelhamer, and T. Darrel, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[17] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 2881–2890.

[18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schoff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018, pp. 1–18, *arXiv:1802.02611*. [Online]. Available: https://arxiv.org/abs/1802.02611

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[20] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "ICNet for real-time semantic segmentation on high-resolution images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 405–420.

[21] A. Pazke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2017, pp. 1–10, *arXiv:1606.02147*. [Online]. Available: https://arxiv.org/abs/1606.02147

[22] V. Badrinarayanan, A. Handa, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," 2015, pp. 1–10, *arXiv:1505.07293*. [Online]. Available: https://arxiv.org/abs/1505.07293

[23] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, pp. 1–9, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861

[24] X. Wang, F. Yu, Z.-Y. Dou, T. Darrell, and J. E. Gonzalez, "SkipNet: Learning dynamic routing in convolutional networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 420–436.

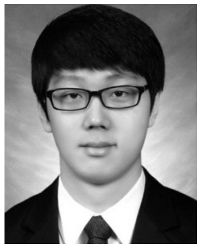
[25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Oct. 2015, pp. 234–241.

[26] A. C. Koivunen and A. B. Kostinski, "The feasibility of data whitening to improve performance of weather radar," *J. Appl. Meteorol.*, vol. 38, no. 6, pp. 741–749, Jun. 1999.

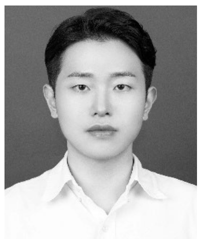
[27] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, pp. 1–13, *arXiv:1511.07122*. [Online]. Available: https://arxiv.org/abs/1511.07122

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, pp. 1–15, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[29] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement. (OSDI)*, Nov. 2016, pp. 265–283.



HANGUEN KIM (M'15) received the Ph.D. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2016, the B.S. and M.S. degrees from Kyunghee University, Suwon, South Korea, in 2009 and 2011, respectively. From 2015 to March 2017, he worked as a Managing Researcher at Samsung S1 Corporation. He is currently the Director of Seadronix Corporation, Daejeon, South Korea. His research interests are deep learning (deep neural networks), computer vision, human–computer interaction, and robot navigation.



JUNGMO KOO received the B.S. degrees from Kyunghee University, Suwon, South Korea, in 2015, and the M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2017. His research interests include computer vision, machine learning, deep learning (deep neural networks), and robot intelligence.

DONGHOON KIM (M'19) received the B.S. degree from the University of Seoul, Seoul, South Korea, in 2009, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2011 and 2016, respectively. He is currently the Director of the Seadronix Corporation, Daejeon, South Korea. His research interests are deep learning (deep neural networks), computer vision, surface robots, and robot navigation.

BYEOLTEO PARK (M'19) received the dual B.S. degrees in electrical engineering and civil engineering and the M.S. and Ph.D. degrees in civil engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2010, 2011, and 2016, respectively. He is currently the CEO of the Seadronix Corporation, Daejeon. His research interests include the areas of SLAM (simultaneous localization and mapping), robot navigation, underground localization, indoor localization, artificial intelligence, and unmanned surface vehicle.

YONGGIL JO received the B.S. degrees from Kyunghee University, Suwon, South Korea, in 2014. He is currently a Researcher with Seadronix Corporation, Daejeon, South Korea. His research interests are deep learning (deep neural networks), computer vision, and surface robots.

HYUN MYUNG (SM'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1992, 1994, and 1998, respectively. He was a Senior Researcher with the Electronics and Telecommunications Research Institute (ETRI), from 1998 to 2002, the CTO and the Director of the Digital Contents Research Laboratory, Emersys Corporation, from 2002 to 2003, and a Principal Researcher with the Samsung Advanced Institute of Technology, Samsung Electronics Company, Ltd., from 2003 to 2008. Since 2008, he has been a Professor with the Department of Civil and Environmental Engineering, KAIST, and he is also the Chief of the KAIST Robotics Program. Since 2019, he has been a Professor at the School of Electrical Engineering. His current research interests include autonomous robot navigation, SLAM (simultaneous localization and mapping), SHM (structural health monitoring), machine learning, AI, and swarm robots.

DONGHWA LEE received the B.S. and M.S. degrees from Kyungpook National University, Daegu, South Korea, in 2010 and 2005, respectively, and the Ph.D. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2015. From December 2014 to January 2016, he worked as a Research Staff Member at the Samsung Advanced Institute of Technology (SAIT), Samsung Electronics. He is currently working as an Assistant Professor of the Division of Computer and Communication Engineering, Daegu University. His research interests are deep learning (deep neural networks), robot navigation, computer vision, probabilistic robotics, underwater robots, and sonar signal processing.