Sensors and Materials, Vol. 35, No. 5 (2023) 1613–1617
MYU Tokyo

S & M 3274

Approaches to Upgrading the Performance of Fishing Vessel Recognition Technology

Ching-Hai Lin,[1,2] Chun-Cheng Lin,[3] Yu-Cheng Zhan,[3]
Cheng-Yu Yeh,[3*] and Shaw-Hwa Hwang[4]

[1]National Museum of Marine Science & Technology, Keelung City 202010, Taiwan
[2]Department of Electrical Engineering, Lee Ming Institute of Technology, New Taipei City 243083, Taiwan
[3]Department of Electrical Engineering, National Chin-Yi University of Technology,
57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 41170, Taiwan
[4]Department of Electronics and Electrical Engineering, National Yang Ming Chiao Tung University,
Hsinchu 300093, Taiwan

(Received August 5, 2022; accepted April 13, 2023)

Keywords: fishing vessel identification, image recognition, deep learning, ArcFace

Fishing vessel recognition using face recognition has recently been addressed for the first time. This paper is actually an improved version of the original proposal, and there are two steps to improve the performance of fishing vessel recognition. In the first step, the number of recognizable fishing vessels was increased considerably from 156 to 272 and the numbers of images of different vessels were made as uniform as possible for a higher generalization ability. In the second step, an EfficientNet model was employed, input images were resized to 480 × 160 pixels to undistortedly display the side views of fishing vessels, and finally, the ArcFace loss function was used as well to train the presented model. As it turned out, the overall recognition performance was improved.

1. Introduction

As opposed to techniques to recognize hull identification numbers (HINs),(1,2) a face recognition technique(3,4) was used for fishing vessel recognition(5) for the first time. The novel recognition technique in Ref. 5 was found to outperform HIN counterparts mainly owing to the fact that most poorly maintained fishing vessels are pictured with an incomplete HIN, which led to misidentification cases.

As elsewhere, most small fishing vessels in Taiwan are not equipped with an automatic identification system (AIS). Consequently, the staff members of a fishing port administration office become heavily loaded when identifying incoming and outgoing fishing vessels repeatedly. In light of this, we developed a reliable and efficient method of identifying fishing vessels as a way to ease the personnel workload.

This paper is actually an extended work of the original proposal,(5) and there are two steps to improve the performance of fishing vessel recognition. As will be detailed later, the distribution of images in a dataset was modified, the recognition model in the original proposal was

*Corresponding author: e-mail: cy.yeh@ncut.edu.tw
https://doi.org/10.18494/SAM4077

ISSN 0914-4935

restructured, and input images were resized. Finally, a different loss function was used as well to train the presented model herein.

2. Materials and Proposed Model

2.1 Materials

First, Table 1 gives a comparison between the datasets used in the original proposal and this work. As referenced previously, the number of recognizable fishing vessels was significantly reduced from 156 in the original proposal[5] to 272 in this work. Moreover, the numbers of images of different vessels were made as uniform as possible for a higher generalization ability, and accordingly, the standard deviation (STD) of the number of images per fishing vessel was considerably reduced from 62.00 to 16.93 in this work for improved model training.

2.2 Proposed model

In contrast to the original proposal, the state-of-the-art EfficientNet-B0 model[6–9] was used to build this work. In addition, a different loss function was used for model training. An analysis of the 7308 collected images in this work, such as the side view of the fishing vessels in Fig. 1, revealed an average aspect ratio of 2.71. As opposed to the 1:1 aspect ratio commonly used in the literature, an aspect ratio of 3:1 was thus used in this work. Accordingly, all the fishing vessel images were resized from the original 160 × 160 to 480 × 160 pixels. As a consequence, the collected fishing vessel images became undistorted, and an image recognition model could be better trained using undistorted images.

Additive angular margin loss, also referred to as ArcFace,[3] was employed as a loss function for training purposes. ArcFace has been acknowledged as the commonest loss function in the field of facial recognition and definitely outperforms a triplet loss[4] as used in the original proposal. The ArcFace loss is defined as

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\left(\theta_{y_i}+m\right)}}{e^{s\cos\left(\theta_{y_i}+m\right)}+\sum_{j=1,j\neq y_i}^{n}e^{s\cos\theta_j}}, \tag{1}$$

Table 1
Comparison between datasets used in the original proposal and this work.

Database	Original proposal[5]	This work
Number of fishing vessels	156	272
Number of images	7037	7308
Minimum number of images	2	2
Maximum number of images	417	60
Mean	45.11	26.87
STD	62.00	16.93

(a) (b)

(c) (d)

Fig. 1. (Color online) Four sample images of fishing vessels, each with a unique HIN. (a) BJ3046, (b) BJ4260, (c) BK8111, and (d) BK8236.

where N and n are the batch size and class number, respectively. $\cos\theta_j = W_j^T x_i$, $\|x_i\| = 1$, and $\|W_j\| = 1$. $x_i \in \mathbb{R}^d$ denotes the embedding feature of the i-th sample, belonging to the y_i-th class. The embedding feature dimension d is set to 512. $W_j \in \mathbb{R}^d$ denotes the j-th column of the weight $W \in \mathbb{R}^{d \times n}$. s is the radius of a hypersphere on which the embedding features are distributed. m is an additive angular margin penalty between x_i and W_{y_i} to simultaneously enhance the intraclass compactness and interclass discrepancy.

A stochastic gradient descent (SGD) optimizer was used to train the model with a batch size of 16 and an epoch of 500, and the weights were recorded when the highest recognition rate was reached. The model was developed using the Python language and the PyTorch library.

As illustrated in Fig. 2, the presented model was trained using a large number of fishing vessel images, and an embedding was then extracted for each image. Subsequently, the Euclidean distance between two embeddings was taken as a way to identify whether the vessel contained in an input image existed in an image database.

3. Experimental Results

As listed in Table 2, all the collected images were divided into two groups. Group 1 is composed of the 7308 images tabulated in Table 1, 4648 of which were used as a training set and the remaining 2660 were used as a test set. None of the images contained in Group 2 was repeated in Group 1, and Group 2 was created exclusively for test purposes. An experiment in this work has two stages and was carried out in exactly the same way as in the original proposal.(5) In Stage 1, the optimal threshold was determined by optimizing the overall performance of the presented model, conducted on the training set in Table 2, with respect to the threshold, as in the Labeled Faces in the Wild database.(10) In Stage 2, a performance test was

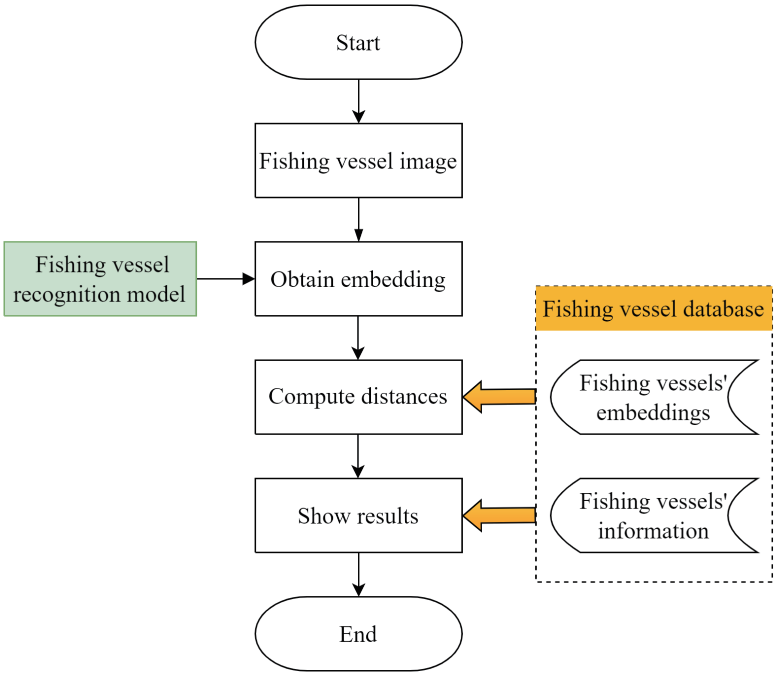

Fig. 2. (Color online) Flowchart of a fishing vessel identification task.

Table 2
Fishing vessel grouping and sizes of training and test sets in this work.

Group	Size of training set	Size of test set
7308 images (272 fishing vessels)	4648	2660
1790 images (225 fishing vessels)	–	1790
Total	4648	4450

conducted on the test set in Table 2 to obtain the performance metrics: true positive rate (TPR), false positive rate (FPR), precision, and accuracy.

The presented model and a duplicate, referred to as Models A and B hereafter, were trained to investigate the effect of aspect ratio on the model performance, and took images with sizes of 480 × 160 and 160 × 160 pixels as input, respectively. In Stage 1 of the experiments, both models performed as well as the other in terms of accuracy, that is, 100% accuracy but at a respective optimal threshold. This means that ArcFace can work well when applied to the issue of fishing vessel recognition.

For comparison purposes, Table 3 gives respective performance metrics of both models obtained in Stage 2. Both models performed comparably in every aspect. Note that Model A slightly outperformed its counterpart in terms of accuracy. It must be stressed that the data used to train and test both models are not the same as those used in the original proposal, meaning that the performance characteristics of Model B and the original proposal cannot be compared fairly.

Table 3
Performance comparison between models taking input images of different sizes.

Model name	Model A (480 × 160 pixels)	Model B (160 × 160 pixels)
Threshold	0.875176	0.824943
TPR (%)	90.66	89.61
FPR (%)	0.45	0.33
Precision (%)	99.67	99.75
Accuracy (%)	94.25	93.66

4. Conclusions

This study is conducted to develop an improved version of the original proposal that has recently been published, and there are two steps to improve the model performance. In the first step, the number of recognizable fishing vessels was increased considerably from 156 to 272, and the numbers of images of different vessels were made as uniform as possible. As a consequence, the STD fell considerably from 62.0 to 16.93 and a high generalization ability was provided. In the second step, an EfficientNet-B0 model was employed, input images were resized from 160 × 160 to 480 × 160 pixels so as to undistortedly demonstrate the side views of fishing vessels, and ArcFace was used as well to train the presented model. As expected, the overall recognition performance was improved. In other words, with this work, fishing vessels when accessing a fishing port can be recognized more accurately than before.

Acknowledgments

This research was financially supported by the Ministry of Economic Affairs under grant number 110-EC-17-A-02-S5-008, Taiwan.

References

1 J. C. Ferreira, J. Branquinho, P. C. Ferreira, and F. Piedade: Adv. Intell. Syst. Comput. **615** (2017) 9.
2 S. Huang, H. Xu, X. Xia, and Y. Zhang: Proc. IEEE Int. Symp. Comput. Intell. Des. (IEEE, 2018) 195. https://doi.org/10.1109/ISCID.2018.00051
3 J. Deng, J. Guo, N. Xue, and S. Zafeiriou: Proc. 2019 IEEE/CVF Conf. Computer Vision and Pattern Recognition (IEEE, 2019) 4685. https://doi.org/10.1109/CVPR.2019.00482
4 F. Schroff, D. Kalenichenko, and J. Philbin: arXiv (2015). https://arxiv.org/abs/1503.03832
5 C. H. Lin, C. C. Lin, R. H. Chen, C. Y. Yeh, and S. H. Hwang: IEEJ Trans. Electr. Electr. Eng. **17** (2022) 1755. https://doi.org/10.1002/tee.23686
6 M. Tan and Q. V. Le: arXiv (2020). https://arxiv.org/abs/1905.11946
7 M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen: arXiv (2019). https://arxiv.org/abs/1801.04381
8 P. Ramachandran, B. Zoph, and Q. V. Le: arXiv (2017). https://arxiv.org/abs/1710.05941
9 J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu: IEEE Trans. Pattern Anal. Mach. Intell. **42** (2020) 2011. https://doi.org/10.1109/TPAMI.2019.2913372
10 Labeled Faces in the Wild (LFW). http://vis-www.cs.umass.edu/lfw/ (accessed 22 July 2021).