SpectralSeaNet: Spectrogram and Convolutional Network-based Sea State Estimation

Xu Cheng, Guoyuan Li, Robert Skulstad, and Houxiang Zhang
Department of Ocean Operations and Civil Engineering
Norwegian University of Science and Technology
Email: xu.cheng@ntnu.no, guoyuan.li@ntnu.no,
robert.skulstad@ntnu.no, hozh@ntnu.no

Shengyong Chen
School of Computer Science and Technology,
Tianjin University of Technology.
Email: csy@tjut.edu.cn

***Abstract*—Sea State is significant to the operations on the sea. The traditional model-based approaches need lots of knowledge of vessels, which limit the real-world use. This paper proposes a spectrogram-based deep learning model for sea state estimation (SpectralNet). In this model, the ship motion data is converted to spectrogram using short time Fourier transform (STFT). Unlike other methods, the spectrogram of each sensor will be combined to a new image. And then, a 2D convolutional neural network (CNN) is built as the classifier and the sea state can be identified. The experimental results show the proposed approach can achieve higher classification accuracy compared these methods applied directly in raw time series data. Through the comparison results of the proposed approach and the combination of spectrogram of different number of sensors, the proposed approach can achieve highest classification accuracy, and the classification accuracy is growing with the number of combined sensors. The sensitivity analysis finds the classification accuracy is easily influenced by the scale factor of images.**

I. Introduction

In the era of artificial intelligence, ship intelligence has become the focus of future ship development [1]. One of the important manifestations of ship intelligence being able to become intelligent is that it needs to be able to predict wave-induced loads and response. Therefore, how to accurately and independently perceive the external environment has become a key aspect in the development of autonomous ship.

To fully understand the external environment, the traditional means is to use the external sensors, such as wave buoys, weather forecast, or satellite measurement, to measure the sea state. Those methods are widely used even though there are still some limitations. For example, the wave buoys need to be placed by humans, and its position is usually close to the shore. The information of weather forecast often lags up to several hours. As for the satellite data, the resolution is the bottleneck limiting its widely applications. Nowadays, wave radars have been installed and applied on some ships. Although the measurements of wave radar are more accurate, it is not widely applied for it still suffers from high cost and frequent calibrations.

To overcome the disadvantages of the traditional methods, there is a trend to consider the wave buoy-like approach. In those approaches, the ship is considered as a huge wave buoy and its motion data would be utilized for the estimation of sea state [2]. The reason is that the wave-induced motion provide the basis for the estimation of on-site sea state. On the basis of this idea, there are two branches for sea state estimation: model-based method and data-based method. The model-based methods are to combine the mathematical model of ship and the motion data to infer the sea state. While, the data-based approaches are directly using the sensor data to build machine learning or deep learning models on which sea state information could be inferred.

Enormous works have been done for the model-based approaches [3]. These model-based approaches are mainly focusing on the frequency domain, where the knowledge of a vessel is essential. The assumption of these methods is that there is a known transfer function mapping the sea state to the ship movements. The wave spectrum (information) can be calculated from the motion data and the transfer functions by using spectral analysis. Brodtkorb et al. proposed a novel method, which is computationally efficient and no assumptions on the wave spectrum shape, for the dynamic positioning (DP) vessels based on ship motion data [4]. Montazeri et al. proposed a shipboard wave estimation approach, in which the parameters describing the wave spectrum are optimized using the global search basin with proper constraints [5]. A network-based approach for sea state estimation is proposed in recent year, which focuses on weighting the single ship-specific wave spectrum obtained from multiple ships [6]. Obviously, the accuracy of the estimation relies inherently on availability of accurate transfer functions, and moreover, these methods are hard to be applied to the real-world use.

In the age of big data and artificial intelligence, purely data-based sea state method are ever-increasing with its biggest advantage can remarkably enhance high accuracy without requiring the knowledge of vessels. Data driven methods are using machine learning or deep learning techniques to extract features either time or frequency domains, or both. The machine learning based methods always based on the human-made features, while the deep learning based approaches can extract features purely based on their own structures without the help of humans. Although machine learning or deep learning technology has been widely used in other fields, they are rarely used in sea state estimation. A first feature-based approach is proposed by Tu et al. for identifying sea state [7]. To overcome the limitations of feature-based methods, a deep

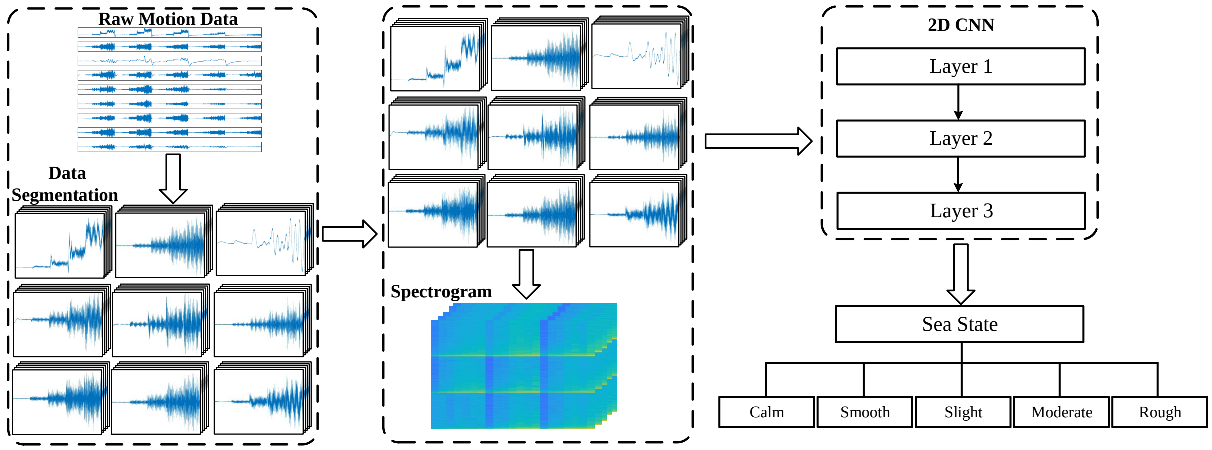


Fig. 1: Procedures for the proposed approach.

learning-based model is proposed by Cheng et al. [8]. Mak et al. studied two deep neural networks for sea state estimation on the basis of 6-DOF ship motion data [9]. Cheng et al. proposed a dense connected convolutional neural network to estimate wave height and wave direction, simultaneously [10].

These deep learning models can learn more non-linear abstract features and can use more useful information than traditional machine learning methods. Therefore, the classification accuracy of deep learning models is higher than that of traditional methods. In [8], a model was established with 3 parallel branches to extract the features in the time domain, frequency domain, and the spatial relation of motion data, aiming at improving the classification accuracy of sea states. Although the deep learning-based approach shows great potential in the application of sea state estimation, the model is lack of interpretability. In order to improve the interpretability of deep learning models, a new model is proposed. In this method, the ship motion data is converted into spectrogram-based images which contain both time and frequency information, by utilizing the short-time Fourier transform (STFT). And then, the images will be fed into a three-layer CNN, so as to obtain the sea states.

The main contributions of this paper are twofold: 1) As far as we know, this paper is the first try to estimate sea states using converted 2D time-frequency images from ship motion data. 2) the proposed method is verified on the ship motion data which is from a dynamical positioning (DP) vessel. Through the experimental results, our proposed method obtains the best estimation accuracy.

The reminder of the paper is organized as follows. Section II introduces the proposed approach. Section III presents case studies and evaluation results. Conclusion and future work are shown in Section IV.

II. Spectrogram based Sea State Estimation Model

A. Approach overview

Sea State is defined as the situation of wave and wind in the open sea for certain location and moment [11]. In each ship, there are an inertial measurement unit (IMU), which can measure the motion of the vessel. The task of this paper is how to employ the ship motion data to infer the sea condition. The overall procedure of the proposed sea state estimation model is represented in Fig. 1. The raw ship motion data would be segmented into small sequences of certain window size without overlapping. There are nine different sensors are utilized, and the detailed information can be found it in Section II-B. Afterwards, each sequence of the nine sensors is transformed into an image with the help of STFT. In this paper, we combined the nine images into a bigger image. The process of combine the nine images is introduced in Section II-C. These transformed images are the input of the proposed 2D CNN model, and the sea state can be identified and classified. There are five different sea state: calm, smooth, slight, moderate, and rough, which is defined based on the wave height.

B. Data pre-processing

In this paper, simulation data, which is from the Offshore Simulator Center AS (OSC), is utilized. The OSC is equipped with powerful physics engines that can generate almost the same wind and wave as actual environment. In this paper, there are nine parameters are utilized as the input, as shown in Table I.

Data preprocessing is usually aimed at cleaning data noise and normalizing data. Statistical estimation or median filtering are widely used methods to process the data. At the same time, ship data usually has the characteristics of discontinuity, information redundancy and so on. According to these data characteristics, the method developed in our previous paper is utilized [12].

TABLE I: Input parameter specification

Input	Unit	Description
Surge velocity	[m/s]	Velocity in surge direction
Sway velocity	[m/s]	Velocity in sway direction
Yaw velocity	[deg/s]	Velocity in yaw direction
Roll velocity	[deg/s]	Velocity in roll direction
Pitch velocity	[deg/s]	Velocity in pitch direction
Heave velocity	[deg/s]	Velocity in heave direction
Heading	[deg]	Rotation around the yaw axis
Roll	[deg]	Rotation around the roll axis
Pitch	[deg]	Rotation around the pitch axis

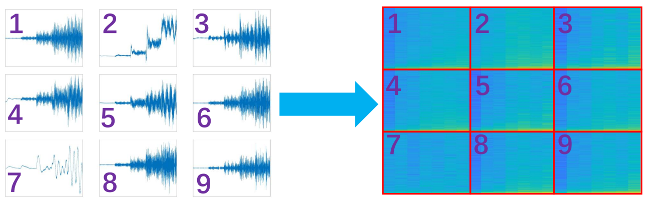


Fig. 2: Time series to image.

C. Time series to spectrogram

A spectrogram of time series can be utilized to describe both the time and frequency information. In order to calculate the spectrogram, the time series will be firstly spilt into small size with certain width (window size), and then the spectrogram function would be performed. To obtain the spectrogram, an enhanced mathematical method, STFT, which is variant of discrete Fourier transform (DFT), is utilized.

For a sensor signal, the STFT-based spectrogram can be obtained as follows:

$$\textbf{STFT}\{x[k]\}(m,\omega) = \mathbf{X}(m,\omega) = \sum_{k=-\infty}^{\infty} x[k]\omega[k-m]e^{-j\omega k} \tag{1}$$

where $x[k]$ is the time series data with sampling rate of 20 Hz. ω represents the window function. The Hanning window is employed in the this paper. Then the magnitude square of the STFT is the spectrogram:

$$spectrogram\{x(t)\}(m,\omega) = |\mathbf{X}(m,\omega)|^2 \tag{2}$$

Additionally, after the transformation from motion data of each sensor to image, these nine images is combined into a bigger image for providing more information to CNN model, which is shown in Fig. 2.

D. CNN network

The CNN is utilized as the classifier for different sea state in this paper. CNN is most widely used feature extractor and classifier, which was introduced by LeCun [13]. With the help of CNN models, various local features of images can be extracted by studying the correlation of spatially adjacent pixels. The proposed 2D CNN is illustrated in Fig. 3, as the 2D CNN is more suitable for the 2D images.

In the proposed 2D CNN network, the raw time series ship motion data is first converted to a spectrogram with a pixel size of 256×256, as described in the Section II-C. There are three hidden layers totally in the proposed CNN network. For each hidden layer, there is a Conv2D layer with kernel size of 5×5 and relu (rectified linear unit) as the activation function. After that, MaxPooling2D with a pool size of (2, 2) is used. In the proposed CNN network, the number of convolutional kernels is 8, 16, and 32, respectively. The features extracted by the three hidden layers is flattened, and then it is fed into a Dense layer with 128 nodes. Finally, the features is sent to a Softmax layer, and the sea state can be identified and classified.

TABLE II: Comparison with baselines

Window size	LSTM	CNN	SpectralNet
500	76.77	88.33	92.33
400	73.33	92.00	94.67
300	76.99	87.99	94.67
200	84.67	93.33	94.33
100	83.33	93.33	94.00
Average	79.02	91.00	94.00

III. Experiments

In order to evaluate the effectiveness of the proposed model, we conduct extensive experiments on ship motion dataset. The model is implemented by using Tensorflow [14] and trained on the colab. During the training, the batch size is set to 32, and the network is optimized using Adam with learning rate 0.001.

A. Dataset description and processing

Five sea states: calm, smooth, slight, moderate, and rough, have been generated. From [8], [11], we can know the sum of the probability of occurrence of these five sea states has accounted for more than 96%. To reflect the complexity of environmental changes, waves and winds are randomly generated, ranging from 12 minutes to 30 minutes [8]. The motion of a DP ship is collected, and the sampling frequency of the system is 20 Hz. In this experiment, over 30 hours of ship motion data were used. And the data is divided into training and testing dataset with non-overlapping 80% and 20%, respectively.

Fig. 4 represents the final combination of the nine sensors for the five sea states. From Fig. 4, we can see that different sea conditions have different characteristics in the spectrogram. like sea state 2, sea state 3, and sea state 4, there are more significant distinguishable features. At the same time, we can know that through this time-frequency map conversion, we can get the following benefits: first, we can directly see the characteristics of the time series data can not be seen directly from the time-frequency map; Second, through the conversion of time-frequency graphs, we can enhance the interpretability of machine learning models.

B. Baseline comparison

The LSTM and CNN are chosen to compare with the proposed SpectralNet on the ship motion dataset. The reason

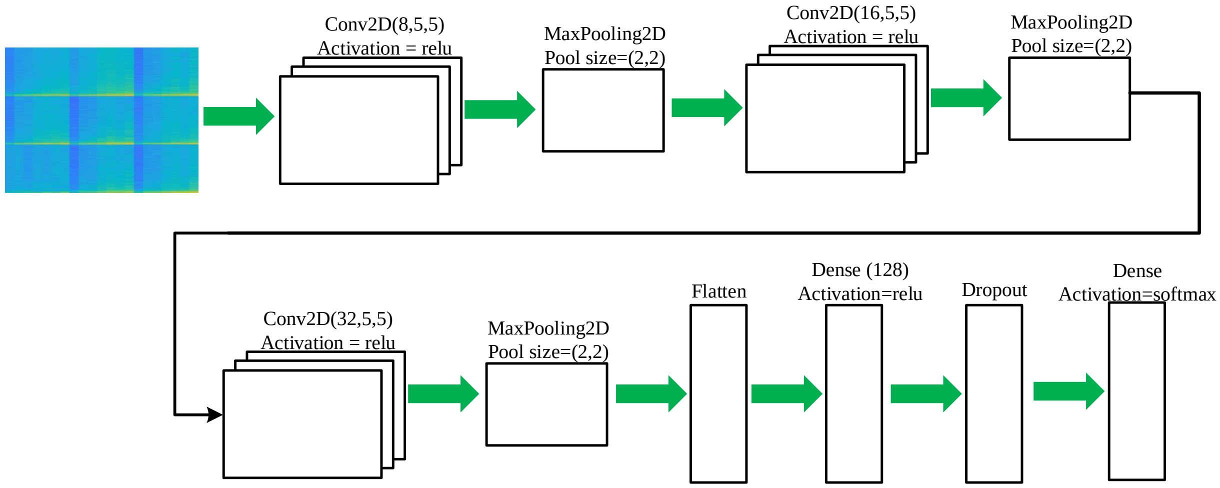


Fig. 3: Structure of the proposed CNN.

for the comparison with these two methods is that these two methods are most widely used deep learning models for time series classification. Unlike the SpectralNet which uses the spectrograms transformed from time series data as the input, the CNN and LSTM are directly employing the time series data as the input. To evaluate the performance of the three models, five datasets with different window size ranging from 100 to 500 are generated, and these three models will be tested on the five datasets. In this experiment, the kernel size of CNN is 3, and the number of filters is set to $\{64, 128, 256, 512\}$. The number of hidden nodes of LSTM is $\{8, 16, 32, 64, 128\}$. The settings of SpectralNet is shown in Fig. 3. The mini-batch size is set to 256, and the learning rate is set to 0.001. To be a fair comparison, the best models of LSTM and CNN are chosen from the settings of LSTM and CNN.

The accuracy of the three models is reported in Table II. It is clearly to know that the LSTM obtains the worst performance compared to the other two models. In addition, the over-fitting of LSTM is more serious with the growing of the length of window size. From Table II, we also can know that the CNN is better than LSTM, with almost 15.16% improvement. SpectralNet achieves the best accuracy regardless of average accuracy or each window size. The results show that the proposed SpectralNet can obtain more information from the images than these conventional methods in time series data.

C. Comparison with non-combination

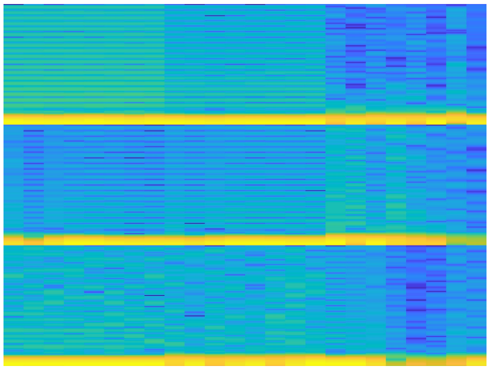
To illustrate the performance of the combination of the nine sensors, we compare the proposed method with each the spectrogram of each sensor, the combination of two sensors, and the combination of four sensors. During these comparisons, the spectrogram will be first generated, as described in Section II-C, and the CNN will be applied for classification. Fig. 5 represents the comparison of each sensor with the combination of nine sensors, and the comparison of one sensor, two sensors, four sensors, and nine sensors.

From Fig. 5a, it is obvious to know the combination of nine sensors obtains the highest classification accuracy. Among these sensors, we can see that the highest classification accuracy happens when the sway velocity is using, and heave velocity follows. To further illustrate the advantage of the combination of the nine sensors, two sensors and four sensors with the highest classification accuracy in Fig. 5a, are combined. The one sensor in Fig. 5b means the one with the highest classification accuracy, e.g. sway velocity. From Fig. 5b, it is easy to know that the classification accuracy is growing with more sensors are combined.

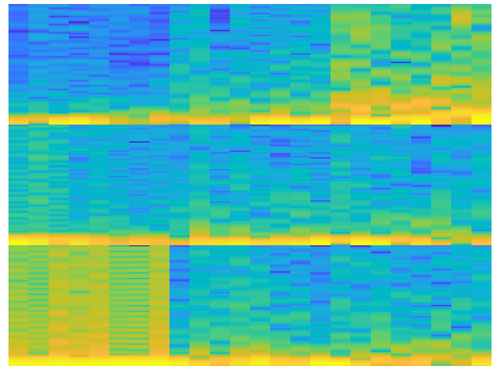
D. Sensitivity analysis

In the proposed model, there are two main parameters: scale factor and batch size. In order to obtain the best classification performance of sea state estimation, sensitivity analysis is indispensable. To illustrated the importance of scale factor and batch size in the proposed SpectralNet, several experiments with different settings were performed. The model performance is firstly tested with different scale factors when keeping the other parameters unchanged in two different window sizes. On the other hand, the batch size is varying while the other parameters are keeping unchanged.

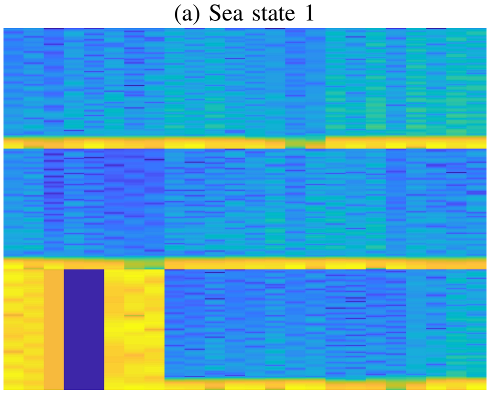
From Fig. 6a, we can know the highest classification accuracy happens when the scale factor is set to 256 in both cases. When the scale factor is 512, that is the image is set to 512×512, the worst accuracy can be obtained. More interestingly, the classification accuracy is growing when the scale factor is from 32 to 256 in both cases. It is easy to know that when the image is scaled to small, some information will lose. Thus, the performance of CNN will decreased. While, if the image is set to too big, the image might become blurred, and the CNN cannot extract suitable features, either. From the Fig. 6b, we can see that the highest classification accuracy happens when the batch size is 30, and with the growing of batch size, the accuracy is decreasing. The reason is that when the batch



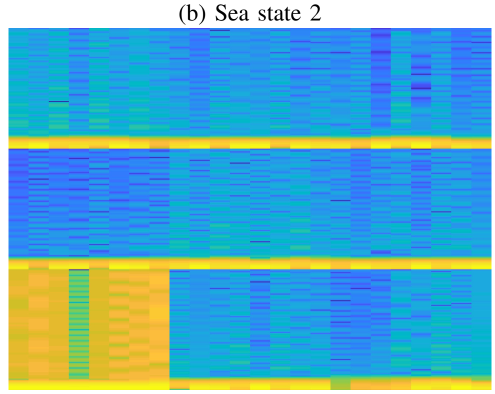
(a) Sea state 1



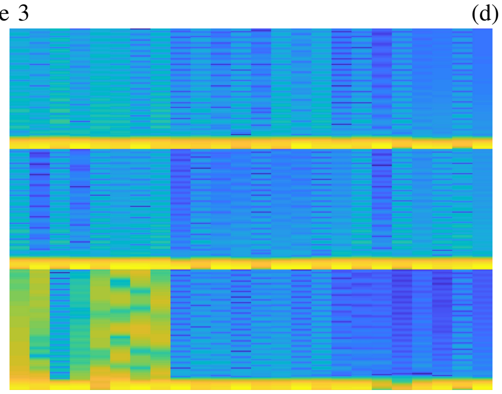
(b) Sea state 2



(c) Sea state 3



(d) Sea state 4



(e) Sea state 5

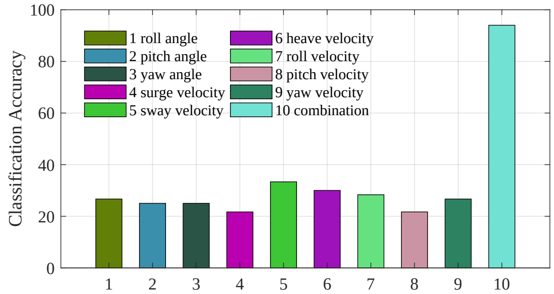
Fig. 4: Combination of the nine sensors for the five sea states.

size is small, the classifier can extract feature in a more fine fashion.

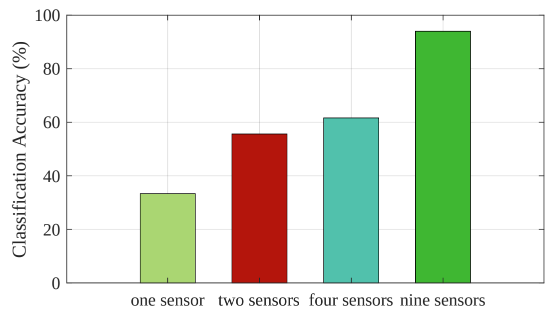
IV. Conclusion

In this paper, a sea state estimation method based on ship motion data and deep learning techniques is proposed. The ship motion data comes from the commercial simulation platform, which can simulate the sea and the ship motion of real world with high precision. To increase the interpretability of the deep learning based sea state estimation models, the raw ship motion data in the time domain is transformed into two-dimensional time-frequency spectrogram by using STFT. Different from other methods, this paper will combine all the spectrograms of all univariate signals to a new spectrogram to improve the classification accuracy. The transformed spectrograms would be used as the input of the 2D CNN models, and the sea state can be identified using the CNN.

The proposed method is evaluated and verified by experiments. From the comparison results with CNN and LSTM tested on raw time series with different window size, the proposed method achieved the highest classification accuracy in all cases. In addition, to illustrate the importance of the combination of spectrogram of each sensors, the proposed method is compared with these models which utilize the spectrogram from only one sensor, two sensors, and four sensors. From the experimental results, the proposed approach can achieve higher classification accuracy with the growing of the number of combined sensors. Finally, the sensitivity

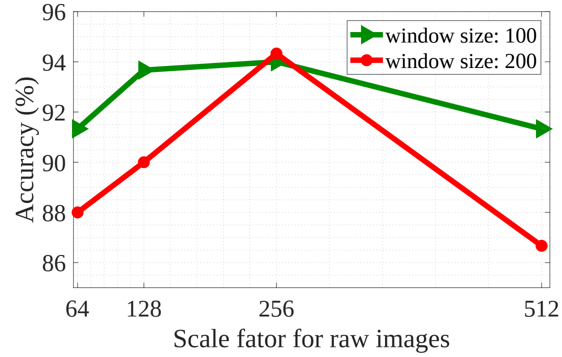


(a) Classification accuracy of each sensor and the nine combined sensors

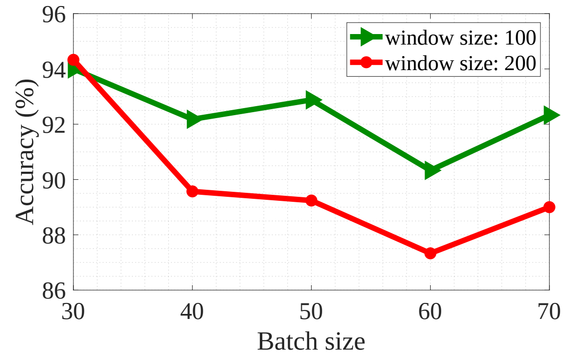


(b) Classification accuracy of one sensor, two sensors, four sensors, and nine sensors

Fig. 5: Combination of the nine sensors for the first four sea states.



(a) Classification accuracy of different scale factors



(b) Classification accuracy of different batch size

Fig. 6: Combination of the nine sensors for the first four sea states.

analysis of key parameters in the proposed model is conducted. From the experimental results, we can know the classification accuracy is easily influenced by the scale factor of images.

Future work should focus on how to further improve classification accuracy. First, a new method combining spectrograms should be proposed to make the spectrograms more easily reflect different sea conditions and sea state changes. Second, a new deep learning model should be designed so that the model can extract more subtle details regardless of the image What is the scaling.

Acknowledgment

This work is supported by the project "Digital Twins for Vessel Life Cycle Service" (Project no.: 280703) and National Natural Science Foundation of China (U1509207). A special thank is given to the Offshore Simulator Centre AS for their technical support.

References

[1] X. Cheng, G. Li, R. Skulstad, S. Chen, H. P. Hildre, and H. Zhang, "A neural-network-based sensitivity analysis approach for data-driven modeling of ship motion," *IEEE Journal of Oceanic Engineering*, 2019.

[2] U. D. Nielsen, "The wave buoy analogy—estimating high-frequency wave excitations," *Applied Ocean Research*, vol. 30, no. 2, pp. 100–106, 2008.

[3] ——, "A concise account of techniques available for shipboard sea state estimation," *Ocean Engineering*, vol. 129, pp. 352–362, 2017.

[4] A. H. Brodtkorb, U. D. Nielsen, and A. J. Sørensen, "Sea state estimation using vessel response in dynamic positioning," *Applied Ocean Research*, vol. 70, pp. 76–86, 2018.

[5] N. Montazeri, U. D. Nielsen, and J. J. Jensen, "Estimation of wind sea and swell using shipboard measurements–a refined parametric modelling approach," *Applied Ocean Research*, vol. 54, pp. 73–86, 2016.

[6] U. D. Nielsen, A. H. Brodtkorb, and A. J. Sørensen, "Sea state estimation using multiple ships simultaneously as sailing wave buoys," *Applied Ocean Research*, vol. 83, pp. 65–76, 2019.

[7] F. Tu, S. S. Ge, Y. S. Choo, and C. C. Hang, "Sea state identification based on vessel motion response learning via multi-layer classifiers," *Ocean Engineering*, vol. 147, pp. 318–332, 2018.

[8] X. Cheng, G. Li, R. Skulstad, S. Chen, H. P. Hildre, and H. Zhang, "Modeling and analysis of motion data from dynamically positioned vessels for sea state estimation," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6644–6650.

[9] B. D. Bart Mak, "Ship as a wave buoy - estimating relative wave direction from in-service ship motion measurements using machine learning," in *38th international conference on ocean, offshore and arctic engineering, OMAE*, 2019.

[10] X. Cheng, G. Li, A. L. Ellefsen, S. Chen, H. P. Hildre, and H. Zhang, "A novel densely connected convolutional neural network for sea state estimation using ship motion data," *IEEE Transactions on Instrumentation and Measurement*, pp. 1–1, 2020.

[11] T. I. Fossen, *Handbook of marine craft hydrodynamics and motion control*. John Wiley & Sons, 2011.

[12] G. Li, H. Zhang, B. Kawan, H. Wang, O. L. Osen, and A. Styve, "Analysis and modeling of sensor data for ship motion prediction," in *OCEANS 2016-Shanghai*. IEEE, 2016, pp. 1–7.

[13] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[14] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.