Computer vision with deep learning for ship draft reading

Article *in* Optical Engineering · February 2021

DOI: 10.1117/1.OE.60.2.024105

Zhiming Liu
University of Nevada, Reno

Computer vision with deep learning for ship draft reading

Bangping Wang,[a,*] **Zhiming Liu**,[b] **and Haoran Wang**[c]
[a]Chengdu University of Information Technology, School of Software Engineering, Chengdu, China
[b]Chengdu Tianren Civil Engineering, Chengdu, China
[c]Qingdao Certification and Inspection Co. Ltd., Qingdao, China

Abstract. The draft is a measurement of the vertical distance between the waterline and the bottom of the ship hull. The displacement tonnage of a ship then can be calculated by the observed draft. The current draft survey is done by surveyors, which is subject to human errors. We propose to use computer vision with deep learning for draft reading from images. First, mask R-CNN is used to segment the region of interest—draft marks and water—from images. Then UNet is used to refine the waterline detection. The detection of marks is based on the computer vision methods and the content of marks is recognized by ResNet. Finally, we can infer the draft of a ship based on the extracted visual information. Experimental results on a realistic dataset have shown that the proposed method can perform the task of draft reading on a par with humans. [DOI: 10.1117/1.OE.60.2.024105]

Keywords: measurement; draft reading; computer vision; deep learning.

Paper 20201469 received Dec. 18, 2020; accepted for publication Feb. 8, 2021; published online Feb. 27, 2021.

1 Introduction

The transportation of commodities plays a critical role in business and the economy. It is estimated that up to 90% of world trade is carried by the international shipping industry, and cargo shipping is more important than ever before. In the maritime shipping, the displacement or displacement tonnage is the basis for safety rules, manning regulations, registration fees, calculation of port dues, etc. The displacement of a ship is measured by calculating the volume of water displaced by the ship based on Archimedes' principle, then converting that value into weight. The process of calculation begins with averaging the draft marks of bow (forward part of the ship) and stern (back part of the ship) on both sides, which show the average vertical distance from the waterline to the keel. An illustration of draft marks on ship is shown in Fig. 1.

The draft is observed by surveyors at each set of marks and the average is calculated to find a mean draft. Although this procedure sounds simple, it is frequently error-prone and complicated in practice. Many factors can influence the draft reading of ships. First of all, the draft reading by the human eye is subject to human errors and dependent upon the subjective opinions of independent surveyors. Other limiting factors include adverse weather conditions and the presence of waves on the water surface, which may cause a danger to surveyors.

While computer vision has become practical and successful, a few automated methods[1–3] have been proposed as alternative to the draft reading by eye. In general, these existing methods use some traditional image-processing methods to estimate draft. For example, the Otsu's binarization and morphological operations for the detection of draft marks, the Canny edge for the waterline detection, etc. In this paper, we apply the deep learning methods for ship draft reading. First, mask R-CNN[4] is used to segment the area of draft marks on ship hull in images. Then, this area is divided into two regions by some traditional computer vision methods. That is, the area of draft marks, which contains horizontal and vertical text lines, can be separated by morphological

*Address all correspondence to Bangping Wang, wbp@cuit.edu.cn

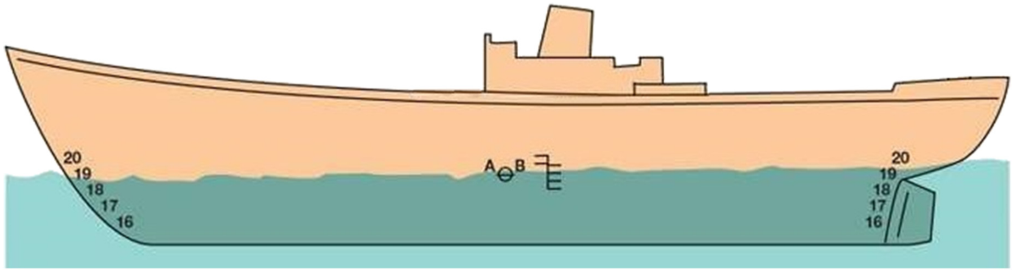


Fig. 1 Measurement of displacement using draft marks on a ship's hull.

operations. The maximally stable extremal regions (MSER) features[5] are used to further detect the positions of text. After that, a residual neural network (ResNet)[6] is used to recognize the draft numbers from text. Also, an encoder–decoder network named UNet,[7] which is the model for image segmentation, is used to detect the fine waterline from image regions inferred from mask R-CNN. As a result, we can calculate the draft of a ship using the extracted visual information. Experimental results on a realistic dataset with small sample size have shown that the proposed method can perform the task of draft reading on a par with humans.

2 Methodology

Deep convolutional neural networks (CNNs), which currently are the preferred approach when the inputs of system are digital images, dominate the field of computer vision. We have applied three CNNs-based methods toward the understanding of ship-water scene in images and then recognizing draft marks and the waterline.

2.1 *Semantic Segmentation*

The first step of our method is to detect the area of draft marks above the waterline in images. In the context of computer vision, this belongs to the task of semantic segmentation. Some popular methods used for semantic segmentation include FCN,[8] DeepLab family,[9,10] mask R-CNN,[4] etc. Mask R-CNN has been the state-of-the-art due to its simplicity and effectiveness since 2017. As such, we first use mask R-CNN to segment the area of draft marks and water from background. Figure 2(a) is the example of annotating draft marks and water in a training image. Figures 2(b)–2(d) are some sample results on test images using mask R-CNN. The results in Fig. 2 show that the area of draft marks can be detected well by mask R-CNN. But the results of water detection are by no means perfect, due to the large feature variation of water in images. Therefore, mask R-CNN is insufficient to generate the fine-grained waterline.

2.2 *Detection of the Waterline*

Further improvements for detecting the waterline can be achieved in a smaller region below the area of draft marks [see red rectangle in Fig. 2(a)]. The complexity and variability of the visual features in such a region should be less than those contained in the whole image. We believe that such input facilitates the network layers to learn representations of data for better differentiating the water region from background, i.e., the ship's hull. Given this, we have applied UNet for water segmentation, because of its original design for image segmentation, especially for two-class problems, and its smaller computational overhead than mask R-CNN. Figure 3 shows the results of water segmentation on test images.

2.3 *Detection of Draft Marks*

Draft numbers consist of integer and decimal. They have different layout orientations. The integer numbers are horizontal text and conversely the decimal numbers are vertical text. Specifically, the decimal numbers arranged in vertical line are even numbers, which occur in

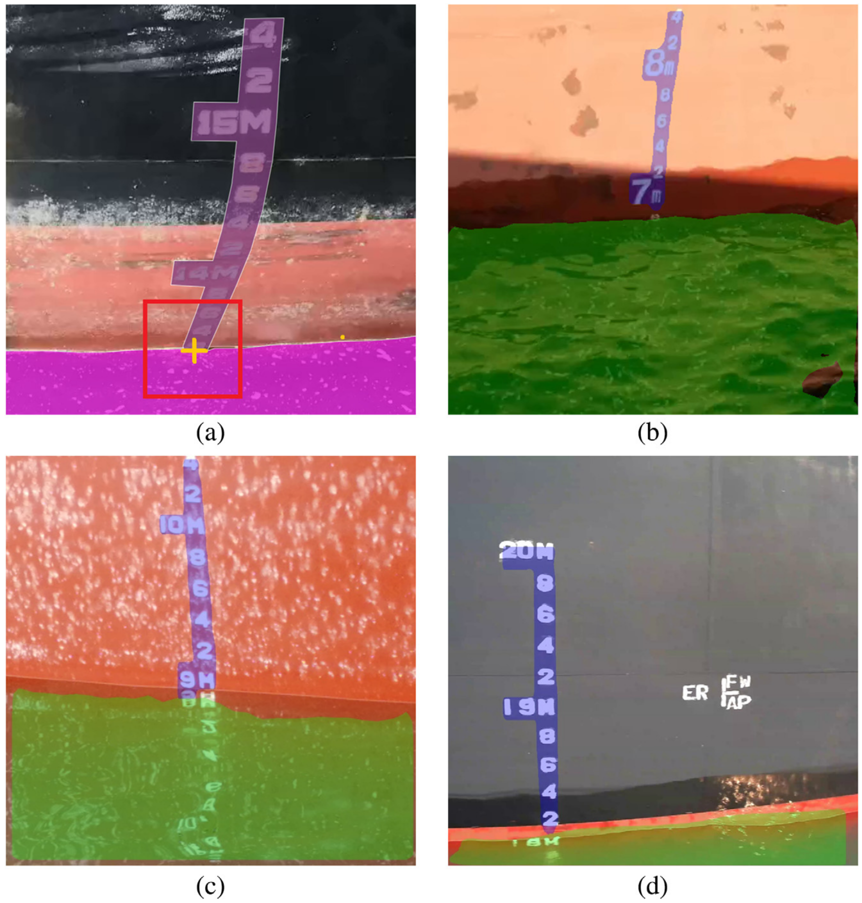


Fig. 2 Results of detecting draft marks and water using Mask R-CNN. (a) Example of annotating draft marks and water. (b)–(d) Some results of Mask R-CNN.

Fig. 3 UNet for water segmentation. Gray: water, black: background.

the repeated pattern, "M, 8, 6, 4, 2." Based on this prior knowledge, we use a preprocessing method to divide the area of marks into two regions.

Figure 4 is the diagram of the preprocessing method. We have a mask image that indicates the area of the draft marks generated from mask R-CNN, as shown in Fig. 4(a). Our task is to separate it into a trunk containing vertical text and several blobs containing horizontal text, as shown in Figs. 4(d) and 4(f).

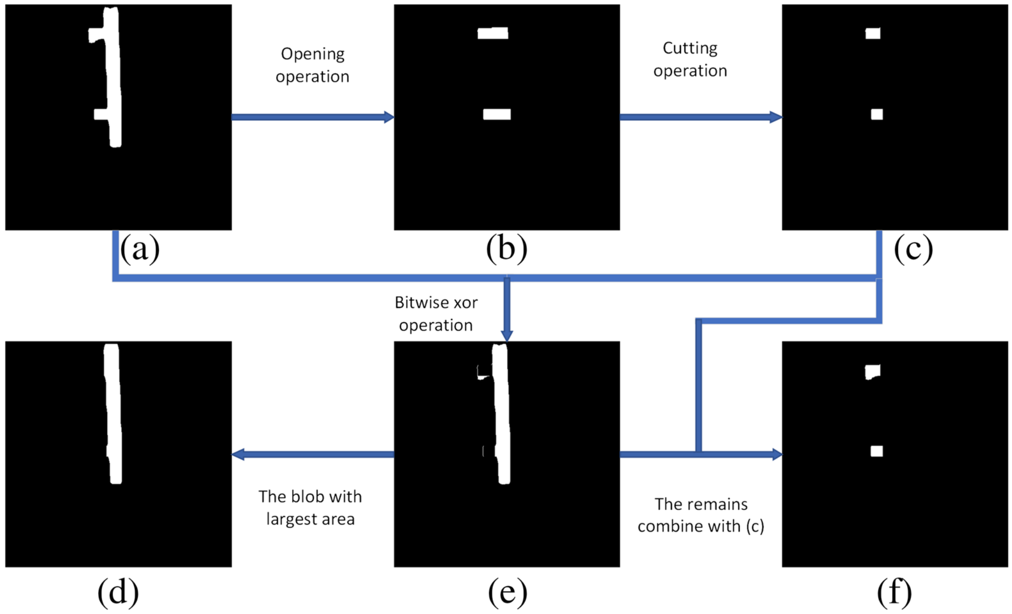


Fig. 4 Preprocessing to divide the area of draft marks into two regions: vertical and horizontal regions. (a) input mask image, (b) horizontal blobs, (c) result of cutting operation, (d) vertical trunk, (e) per-element bitwise XOR of (a) and (c), and (f) combination of (c) and the small blobs in (e).

First, we obtain the average width of the mask by scanning each row of the image. The value of width is indicated as w_m. The morphological opening operation is performed on the mask with a kernel of the size ($h = 1$, $w = w_m$). This process removes the vertical trunk and retains the horizontal blobs [see Fig. 4(b)]. Then, the right side of blobs are cut off the length of w_m [see Fig. 4(c)]. After that, the operation of the per-element bitwise XOR is executed on both the remained blobs and the mask. This leads to some disconnected blobs [see Fig. 4(e)]. Among them, the blob with the largest area is the vertical trunk, and the other smaller ones combine with Fig. 4(c) to form the blobs as shown in Fig. 4(f).

In order to detect the waterline meeting draft marks, the vertical trunk is first thinned as shown in Fig. 5(a), and then the resulted points are used to fit a curve with quadratic polynomials,

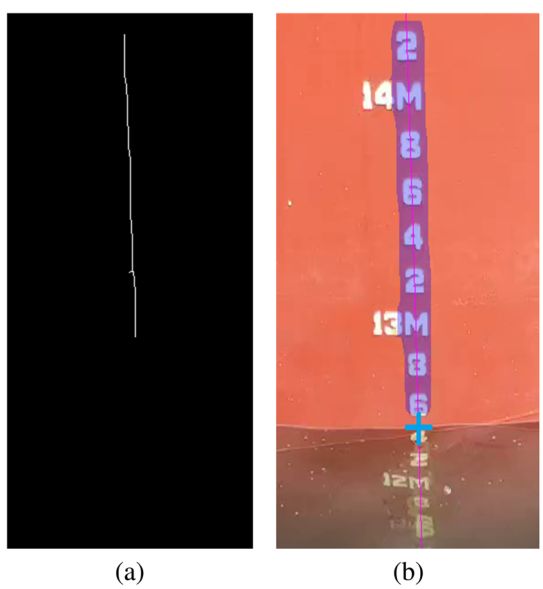


Fig. 5 Detection of the waterline meeting draft marks. (a) Thinning operation. (b) A curve intersecting with the waterline.

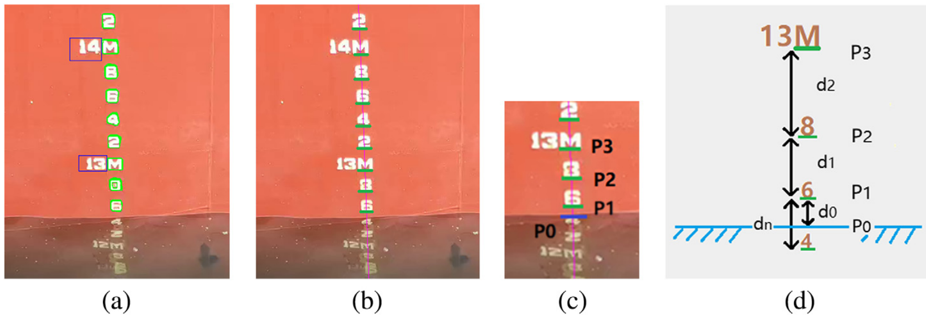


Fig. 6 Draft estimation. (a) The MSER detections shown in green. (b) Positions of marks. (c) Draft calculation. (d) Illustration of (c).

see Fig. 5(b). The point, where this curve intersects with the waterline, is the estimation of water hitting draft marks.

2.4 *Draft Estimation*

The positions of the draft marks above the waterline need to be located accurately in images. The MSER detector[5] is used to detect the marks in the vertical trunk. Then, the bottom position of each character is obtained by structure analysis after thresholding. Results are shown in Fig. 6. While the horizontal text does not need to be located accurately, their content needs to be identified. A ResNet model has been trained to recognize these marks.

After text detection and identification, we proceed to estimate the value of draft. As Fig. 6(c) shows, four positions need to be determined. The first position, p_0, is the point of the waterline crossing the curve. The other three, p_1, p_2, p_3, are positions where three continuous characters have been identified. We then calculate three distances, $d_0 = \|p_0 - p_1\|$, $d_1 = \|p_1 - p_2\|$, and $d_2 = \|p_2 - p_3\|$. When we observe the draft marks in different views, perspective distortion may be seen in images, for example, Fig. 9(b), which changes the distances among marks, i.e., d_1 and d_2. But modeling such perspective transformation is not easy, because the ship's hull may not be planar. As a simple approximation of such transformation, we calculate the ratio between d_1 and d_2, $r = d_1/d_2$. Then, the ratio r indicates a trend of distance changing between adjacent marks around the waterline. In Fig. 6(d), that is the illustration of Fig. 6(c), we first estimate the distance d_n between the mark 6 and the mark 4 that is below the waterline. We have $d_n = r \cdot d_1$. Given the interval distance between two marks is 0.2 m, the estimated distance between p_0 and p_1 is $\frac{d_0}{d_n} \times 0.2$. Taken all together, the value of draft at p_0 can be expressed in the form:

$$v_0 = v_1 - \frac{d_0}{r \cdot d_1} \times 0.2. \tag{1}$$

In Eq. (1), v_1 is the value of draft at p_1, which is 12.6 m in Fig. 6(c). In the end, the estimated value of draft at p_0 is 12.53 m in Fig. 6(c).

3 Experiments

3.1 *Dataset*

We have collected 50 videos from the Qingdao Port, China, for experiments. The average elapsed time is 20 s. Some screenshots are shown in Fig. 7. These videos are randomly split into training (40 videos) and test (10 videos) sets.

In the training video set, the total of 514 images are sampled from videos. Be aware that the region of interest processed by our method locates in the center of the image, we have applied center crop to get $d \times d$ square, where d is the shorter side of the image. Then, square images are labeled as containing the area of marks and water used for the Mask RCNN and UNet models. Also, draft marks are cropped from images to generate a data set used by the ResNet model.

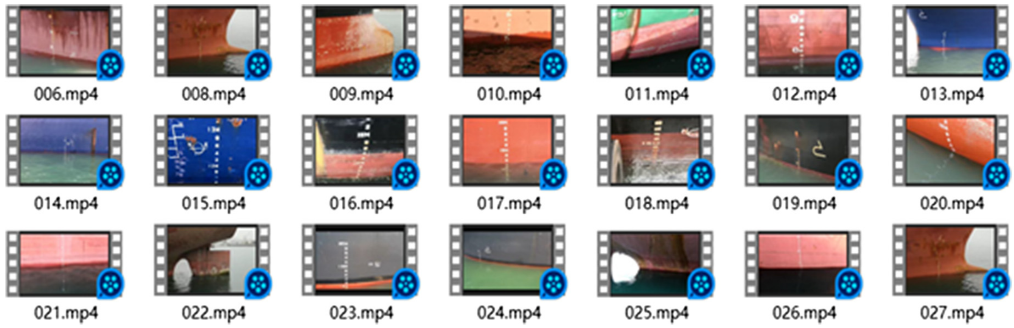


Fig. 7 Examples of video screenshots.

Table 1 Backbones and hyperparameters used in our experiments.

Method	Backbone	Input size	Epoch #
Mask R-CNN	ResNet-101-FPN	1024×1024	80
UNet	ResNet-50	128×128	70
ResNet	ResNet-18	64×64	200

For training the three CNN models, the annotated images are further split into training and validation sets based on an 80:20 split.

3.2 *Results*

Our method applies transfer learning to train three CNN models. In particular, the mask R-CNN model uses pretrained weights on MS COCO,[11] and the UNet and ResNet models use pretrained weights on ImageNet.[12] Table 1 lists the backbones and hyperparameters used in our experiments. Other parameters and training procedures, including batch size, learning rate, optimizer, etc., are set to follow the existing works.[4,6,7] After training, we perform inferences on the test videos. Note that the aforementioned center-cropping operation has been applied to the test videos. Our method is implemented on a PC with 32 GB RAM and a single NVIDIA GTX 1080Ti GPU. The average time for inference is 0.36 s per image.

We first conduct experiments to show the quantitative results for the waterline detection. The evaluation is confined in the rectangle area as shown in Fig. 2(a). We report the standard mask AP (average precision) metric for measuring the accuracy of semantic segmentation. The COCO API[11] is used to compute mask AP. The results in Table 2 show that UNet based on mask R-CNN can perform much better than mask R-CNN alone in the small regions, which validates the effectiveness of our method for the waterline detection.

We then conduct a set of experiments for the detection and recognition of draft marks. The experiments include mask R-CNN for the semantic segmentation of the draft marks, MSER for the detection of the draft marks, and ResNet for the recognition of the draft marks. The experimental results are shown in Table 3. Mask R-CNN is trained to segment the regions of water and draft marks simultaneously. Although the mask AP of draft marks is much lower than that of

Table 2 Quantitative results of the waterline detection.

Method	Mask AP
Mask R-CNN	81.4
UNet	93.0

Table 3 Quantitative results of the detection and recognition of draft marks.

Method	Mask AP	Accuracy
Mask R-CNN	61.8	N/A
MSER	N/A	95.7
ResNet	N/A	98.5

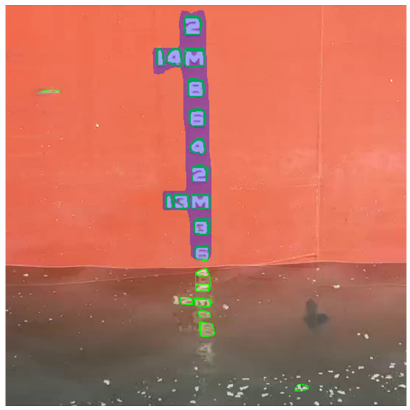
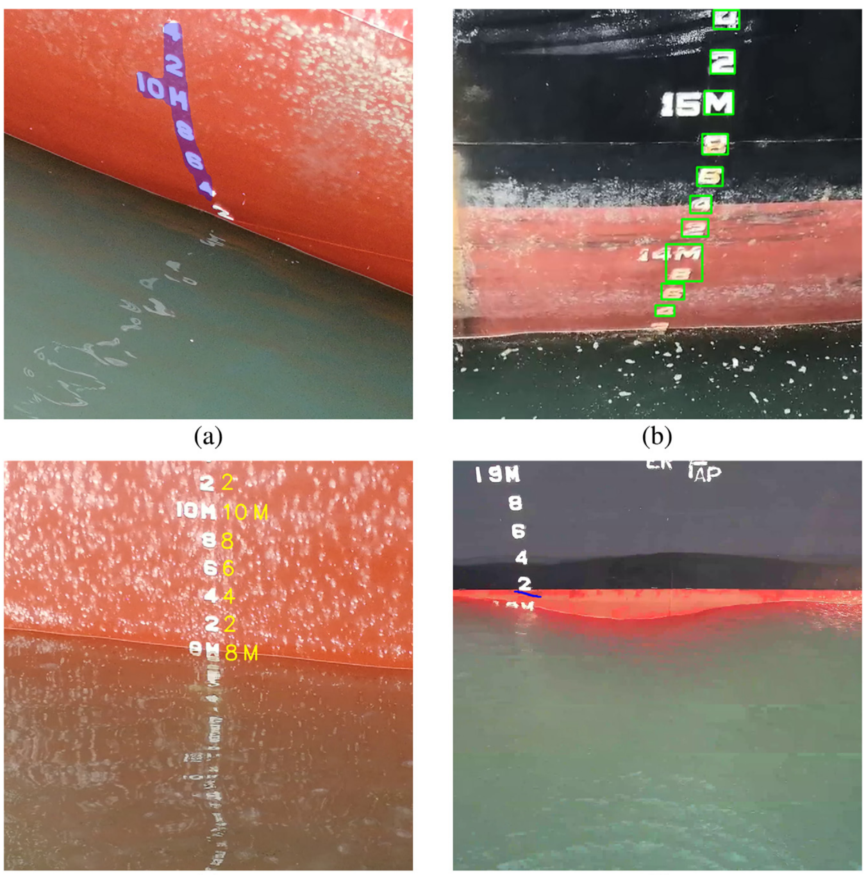


Fig. 8 Semantic segmentation helps MSER in reducing the false positives of draft marks.



(a) (b) (c) (d)

Fig. 9 Some failure cases. (a) The segmentation of draft marks, (b) the MSER detector, (c) the recognition of draft marks, and (d) the waterline detection.

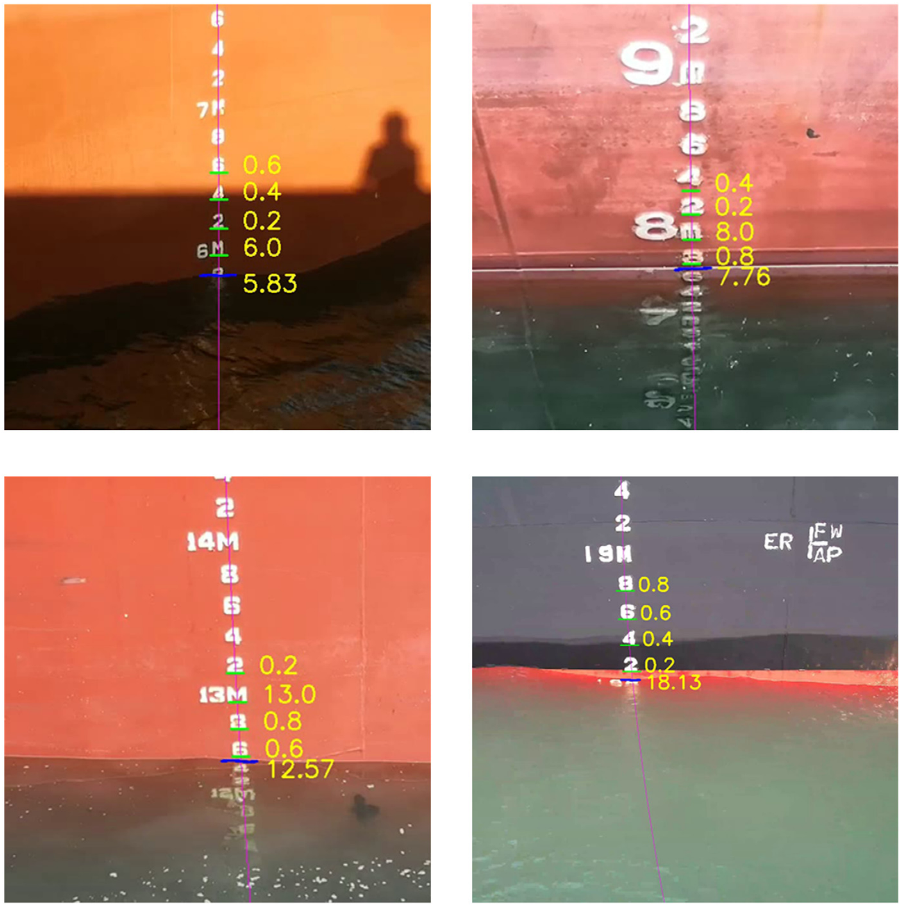


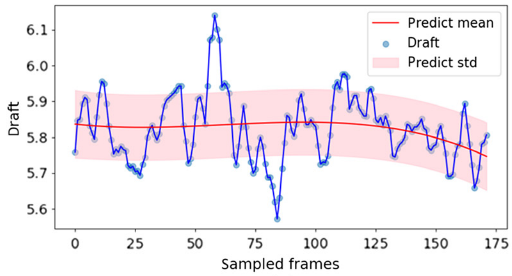
Fig. 10 Visualization of some examples of draft reading on the test videos.

water, the region of draft marks can be segmented properly without the need for precise boundary. Also, the semantic segmentation helps the MSER detector in reducing false positives and obtaining the true positive rate of 95.7%. Figure 8 shows an example in which mask R-CNN can eliminate the false draft marks that are irrelevant to our task. The image regions selected by MSER are used to train a ResNet to recognize the draft marks, and the accuracy in the test set is 98.5%. Figure 9 shows some failure cases in the experiments conducted above.

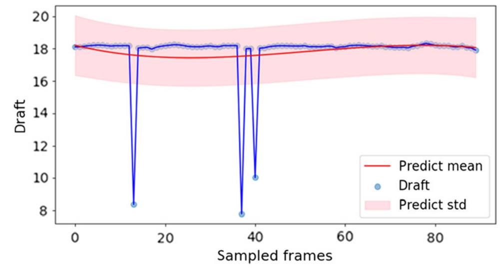
Once we have the results generated by the above experiments, Eq. (1) becomes our vehicle for converting all the input into a draft value. Figure 10 shows the visualization of some draft readings successfully inferred by the proposed method on the cropped images from the set of test videos. The draft for each video is then calculated as the mean of those inferred from frames. Due to errors in the detection of the waterline and marks, Eq. (1) produces some outliers in the distribution of draft values. Figure 11 shows the examples of the distribution of draft values in videos. If the data of draft reading is considered as a polynomial curve, we can apply the curve fitting with Bayesian ridge regression[13] to remove outliers. The fitting process returns the mean and standard deviation of the predictive distribution of data, as shown in Fig. 11. The points of data within the standard deviation of distribution can be considered as inliers. Finally, the draft is the mean of these inliers.

We quantitatively evaluate the ability of the proposed method compared to humans. Table 4 presents the results of our method and the measurements by humans (four surveyors) on 10 test videos. We use Z-scores as a numerical measure to describe the relationship of our results to the mean of a group of results by humans. In statistics, standard scores are most commonly called Z-scores, which are defined as

$$z = \frac{x - \mu}{\sigma}, \tag{2}$$



(a)



(b)

Fig. 11 Examples of draft distribution for videos. (a) distribution without detection errors and (b) distribution with some outliers due to the false detection of the waterline and marks.

Table 4 Quantitative results (meters) of draft reading on the test videos.

Video #	1	2	3	4	5	6	7	8	9	10
P1	5.80	7.66	12.56	13.23	18.12	5.60	5.19	5.22	9.02	7.41
P2	5.83	7.71	12.58	13.25	18.12	5.58	5.19	5.23	9.02	7.42
P3	5.81	7.69	12.57	13.25	18.15	5.56	5.20	5.22	9.00	7.42
P4	5.78	7.69	12.56	13.26	18.15	5.53	5.22	5.21	9.01	7.41
Mean	5.81	7.69	12.57	13.25	18.14	5.57	5.20	5.22	9.01	7.42
Std	0.018	0.018	0.008	0.011	0.015	0.026	0.012	0.007	0.008	0.005
Ours	5.82	7.70	12.55	13.24	18.15	5.55	5.17	5.20	9.05	7.41
Z-scores	0.83	0.70	−2.11	−0.68	1	−0.67	−2.44	−2.82	4.52	−1

where x is a raw score, μ and σ are the mean and standard deviation of the population, respectively. Simply, a Z-score gives us an idea of how far from the mean of population a data point x is. In a normal distribution, 99.7% of samples in population fall within three standard deviations from the mean. Table 4 shows 90% of Z-scores for our results are within ± 3 standard deviations from the mean of humans' measurements, so we can conclude that the ability of our method is on a par with humans on this dataset with small sample size.

4 Conclusions

We have proposed a system of vision-based measurement boosted with deep learning for the draft reading of ships. Three CNN models are used to detect the waterline and draft marks from images. The draft then can be inferred from the extracted information. Experiments on a dataset with small sample size have shown that our method can perform similarly to humans. Therefore, there is a potential that our method can be used in camera-equipped unmanned aerial vehicles for developing the automated system of ship draft reading. In the future work, we will extend the size of our dataset and develop more powerful models for increasing the accuracy of detection for the waterline and draft marks.

References

1. T. Tsujii, H. Yoshida, and Y. Iiguni, "Automatic draft reading based on image processing," *Opt. Eng.* **55**(10), 104104 (2016).
2. A. Okamoto et al., "A draught reading method by image processing with the robustness of measurement distance," *J. Jpn. Inst. Navig.* **130**, 135–140 (2014).

3. X. Ran et al., "Draft line detection based on image processing for ship draft survey," in *Proc. 2011 2nd Int. Congr. Comput. Appl. and Comput. Sci.*, Vol. 2, pp. 39–44 (2012).
4. K. He et al., "Mask R-CNN," in *IEEE Int. Conf. Comput. Vision (ICCV)*, Venice, pp. 2980–2988 (2017).
5. J. Matas et al., "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vision Comput.* **22**(10), 761–767 (2004).
6. K. He et al., "Deep residual learning for image recognition," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, Las Vegas, Nevada, pp. 770–778 (2016).
7. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
8. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, Boston, Massachusetts, pp. 3431–3440 (2015).
9. L. Chen et al., "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2018).
10. L. Chen et al., "Encoder-decoder with atrous separable convolution for semantic segmentation," *Lect. Notes Comput. Sci.* **11211**, 833–851 (2018).
11. T.-Y. Lin et al., "Microsoft COCO: common objects in context," *Lect. Notes Comput. Sci.* **8693**, 740–755 (2014).
12. O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.* **115**, 211–252 (2015).
13. R. M. Neal, *Bayesian Learning for Neural Networks*, Springer-Verlag, Berlin, Heidelberg (1996).

Bangping Wang received his PhD in computer application from Sichuan University in 2009. Since 2010, he has been with the College of Software Engineering, Chengdu University of Information Technology, China. His research interests include visual surveillance and machine vision.

Zhiming Liu received his PhD in computer science from New Jersey Institute of Technology, Newark, New Jersey, USA, in 2011. His research interests include computer vision and machine learning.

Haoran Wang received his master's degree in business administration from Northwest University, Shanxi, China, in 2010. Since 2010, he has been with Qingdao Certification and Inspection Co., Ltd., China. His research interests include inspection and testing technology.