Environmental Pollution 258 (2020) 113688

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Single spectral imagery and faster R-CNN to identify hazardous and noxious substances spills☆

Hui Huang [a, c], Chao Wang [a], Shuchang Liu [b], Zehao Sun [a], Dejun Zhang [a], Caicai Liu [c], Yang Jiang [d], Shuyue Zhan [a, *], Haofei Zhang [c], Ren Xu [c]

[a] Ocean College, Zhejiang University, Zhoushan, Zhejiang, 316021, China
[b] College of Science, Zhejiang University of Technology, Hangzhou, Zhejiang, 310014, China
[c] East China Sea Environmental Monitoring Center, Shanghai, 310058, China
[d] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, 130033, China

ARTICLE INFO

Article history:
Received 25 July 2019
Received in revised form
25 November 2019
Accepted 27 November 2019
Available online 2 December 2019

Keywords:
Hazardous and noxious substances
Spectral imagery
Faster R-CNN
Spill response
Hyperspectral imaging

ABSTRACT

The automatic identification (location, segmentation, and classification) by UAV- based optical imaging of spills of transparent floating Hazardous and Noxious Substances (HNS) benefits the on-site response to spill incidents, but it is also challenging. With a focus on the on-site optical imaging of HNS, this study explores the potential of single spectral imaging for HNS identification using the Faster R-CNN architecture. Images at 365 nm (narrow UV band), blue channel images (visible broadband of ~400–600 nm), and RGB images of typical HNS (benzene, xylene, and palm oil) in different scenarios were studied with and without Faster R-CNN. Faster R-CNN was applied to locate and classify the HNS spills. The segmentation using Faster R-CNN-based methods and the original masking methods, including Otsu, Max entropy, and the local fuzzy thresholding method (LFTM), were investigated to explore the optimal wavelength and corresponding image processing method for the optical imaging of HNS. We also compared the classification and segmentation results of this study with our previously published studies on multispectral and whole spectral images. The results demonstrated that single spectral UV imaging at 365 nm combined with Faster R-CNN has great potential for the automatic identification of transparent HNS floating on the surface of the water. RGB images and images using Faster R-CNN in the blue channel are capable of HNS segmentation.

1. Introduction

With the demand for chemicals increasing worldwide, the number of chemical spill incidents has increased because more chemical tanks are travelling by ocean (Cunha et al., 2015). Hazardous and noxious substances (HNS) are defined by the International Maritime Organization (IMO) as any substance other than crude oil, which, if spilled into the marine environment, is likely to create a hazard to human health, harm living resources and other marine life, damage amenities, or interfere with other legitimate uses of the sea (IMO, 2000). There are many types of HNS with different physical and chemical properties that are shipped together (Harold et al., 2014). Different types of HNS often have different levels of toxicity and require different emergency measures (Mackay et al., 2006). Hence, to deal with spill incidents, instant identification of the HNS location, area, and category help initiate emergency countermeasures to control the damage (Purnell, 2009).

Frequently-shipped HNS include petrochemicals (Gérin et al., 1998) such as benzene and xylene, and vegetable oils (Cunha et al., 2015). There have been many laboratory methods developed to detect types of HNS, such as chromatography (Koeber et al., 1999) and mass spectrometric (MS) (Li et al., 2000) and electrochemical methods (Hilmi and Luong, 2000), which are limited to on-site area detection. Other on-site methods, such as RADAR and synthetic aperture radar (SAR) imaging, have shown effectiveness for area detection of remote oil spills with high cost and complex signals (Alpers et al., 2017). HNS monitoring has shown that three chemicals (toluene, heptane, and methanol) are undetectable from SAR images (Singha et al., 2016). This seems to be caused by the high transparency and volatility of the chemicals, and the relatively long time lag between discharge and observation.

☆ This paper has been recommended for acceptance by Baoshan Xing.
* Corresponding author.
E-mail address: shuyue_zhan@zju.edu.cn (S. Zhan).

https://doi.org/10.1016/j.envpol.2019.113688
0269-7491/

Compared with RADAR and SAR images, optical images allow for swath width monitoring and relatively low cost, and provide more frequent information (Taravat and Del Frate, 2012; Zhao et al., 2014). However, unlike crude oil, transparent HNS is quite challenging to detect due to its lack of colour and thin layer. Limited work has been published on HNS identification. Studies on the segmentation and classification of chemicals indicate that spectral imaging, especially UV band imaging, supported by proper data analysis has great potential for on-site HNS identification using unmanned aerial vehicle (UAV) (Huang et al., 2019; Zhan et al., 2019). In the context of transparent HNS spill segmentation, an improved local fuzzy thresholding methodology (LFTM) combining target enhancement and the original LFTM (Aja-Fernández et al., 2015) outperformed segmentation algorithms such as the original LFTM, Otsu, (1979), and maximum entropy (Max entropy) (Gull and Skilling, 1984), while it performed well on UV images to distinguish HNS and look-alike objects (Zhan et al., 2019). However, the improved LFTM does not classify different HNS. For HNS classification, partial least squares discriminant analysis (PLS-DA) and least squares support vector machine (LS-SVM) have been compared on multispectral images (Huang et al., 2019). Four spectral bands were suggested for the classification of benzene, xylene, and palm oil. It would benefit the system design of UAV imaging if spectral bands could be optimised to a small number, or even a single band. Proper data processing may reveal the most efficient information, thus assisting in the reduction of imaging spectral bands.

In recent years, numerous approaches based on deep convolutional neural networks (DCNNs), which were derived from convolutional neural networks (CNNs), have shown impressive performance on object classification and detection in natural images. AlexNet, proposed by Krizhevsky et al., (2012) in 2010, was the first DCNN capable of achieving a promising result on a historically difficult ImageNet dataset. The network was used for classification with 1000 possible categories. Ren et al., (2015) proposed a faster region-based CNN (Faster R-CNN) using a region proposal network (RPN) to replace the selective search in R-CNN (Girshick et al., 2014) and Fast R-CNN (Girshick, 2015), and combining CNN to classify targets in the proposals generated by the RPN. This improvement reduced the computation required for the generation of proposals.

No application of DCNN has yet been used for HNS identification on the surface of water. However, there have been studies using CNN to detect oil spill locations in optical images. Reports have been made on the detection of other sea targets, such as ships (Liu et al., 2017; Nieto-Hidalgo et al., 2018; Yao et al., 2017). Meanwhile, research on the use of CNNs to detect oil spills in SAR images has also been reported (Nieto-Hidalgo et al., 2018; Yu et al., 2018). As an improvement to the CNN, it is interesting to explore the potential of the DCNN, especially the Faster R-CNN, for the identification of HNS in optical images.

This study explores the potential for classification and segmentation of a floating HNS spill in a single spectral band image, combining state-of-the art object detection networks based on Faster R-CNN architecture. RGB, 365 nm, and blue channel spill images were collected and processed with Faster R-CNN using different segmentation methods. The location bounding boxes and classification results for the spill area were generated by Faster R-CNN, while the segmentation results were generated by applying different segmentation methods on the original images or location bounding boxes using Faster R-CNN. Both the classification and segmentation performance were compared with the related findings in (Huang et al., 2019; Zhan et al., 2019). Finally, the possibility of HNS identification using a single spectral band image was evaluated.

2. Materials and methods

This paper proposes a framework for HNS spill identification using both 365 nm and RGB images by combining Faster R-CNN and a segmentation process to extract the pixel-level spill area. The flowchart of this framework is illustrated in Fig. 1.

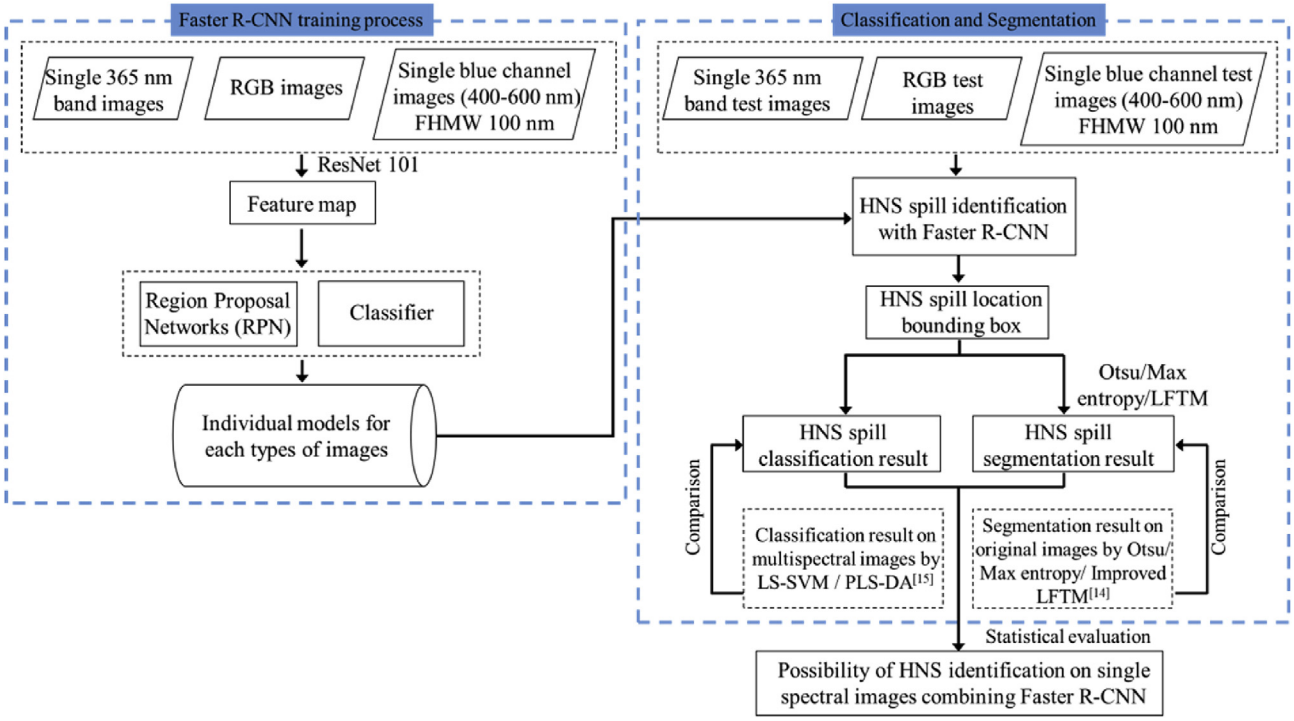

Fig. 1. The flowchart of this framework.

2.1. Experimental system

Three colourless HNS were selected as typical samples to evaluate the capability of the identification method: benzene (Analytical Reagent, AR, Aladdin, Shanghai), xylene (Analytical Reagent, AR, Aladdin, Shanghai), and palm oil (purchased from a supermarket in Zhoushan). In order to obtain spill images of different scenes and create a dataset for model training, the spill experiments were conducted in three different environments: a plastic swimming pool, a lake, and an artificial channel, and on the campus of Ocean College, Zhejiang University, Zhoushan, Zhejiang, China. The UV images were captured using a lab-made multispectral imaging system consisting of a UVTEC-L000 camera body (Indigo, China), narrow band pass filters (only 365 nm was used here based on our previous work), and a 75 mm optical lens that generated an 8 bit grey-level image with a resolution of 2016×1296. The exposure time was set to 1/50 s. UV images were captured approximately 30 s after the sample was released, allowing enough time for spill stabilisation. Broadband RGB images were captured using a digital camera (a6000, Sony, Japan) with a 16–50 mm Sony lens to generate 8 bit grey-level images with a resolution of 3008×2000. The experiments were conducted under protected conditions, and all pouring chemicals were cleaned according to the *Regulation on the Safety Management of Hazardous Chemicals* (General Office of the State Council 2011).

The computational process was implemented with the TensorFlow deep learning framework and was executed on a computer with an Intel Core i5-6500 processor running at 3.2 GHz, 16 GB DDR4 RAM, and an Nvidia GTX 1080 graphics card with 6 GB GDDR5 on the Windows 10 operating system. Due to the limited computing resources, all images were downsampled to a resolution of approximately 350×250.

2.2. Imagery collection and augmentation

Normally, there are thousands of images in a dataset for CNN training to reduce overfitting on the models, such as the open datasets ImageNet and COCO. However, to the best of our knowledge, there are no open datasets for HNS spill detection. Hence, we created an HNS spill dataset for model training using spectral and digital imaging systems. To simulate the location and angle changes in airborne imaging, we captured HNS images from different locations and angles (20–40°). To simulate scale changes caused by the flying altitude in airborne imaging, we recorded images from different distances, ranging from 1.5 to 10.0 m. These setups represented the diversity of airborne images and helped to improve the generalisability of the built HNS identification model. For this step, the number of training and testing RGB images were 468 and 60, respectively. The training and testing datasets of UV images contained 387 and 60 images, respectively. The number of images for each HNS category in the test set was equal (20 images for each category).

Because the number of images was not sufficient for training, we conducted data augmentation, including flip, rotation, scale, and affine transformations, to increase the volume of the training dataset (Perez and Wang, 2017). After data augmentation, the number of RGB and UV training images increased to 1096 and 958, respectively. The exact location (bounding) boxes of all spill samples were annotated for model training.

2.3. Spill location detection and classification based on faster R-CNN

Faster R-CNN is a state-of-the-art, real-time detection architecture used in multiclass object detection, which can extract the implicit features of an image. In this study, the Faster R-CNN model was trained on three kinds of HNS samples (benzene, xylene, and palm oil) to extract features at both the local and global scales, which was then input into the RPN and CNN classifier. Recently, the residual networks (ResNet) architecture proposed by He obtained successful results in an ImageNet and MS-COCO competition due to its deeper convolution network (He et al., 2016). ResNet solves the vanishing gradient problem caused by the increment in network depth. In this work, ResNet-101 (He et al., 2016) was selected as the backbone for feature extraction. The network was 101 layers deep with additional shortcut connections and a bottleneck design. The feature map extracted by ResNet-101 was shared to an RPN and classifiers.

For the RPN of the Faster R-CNN, the key step is calculating the loss of classification and location. The RPN training is intended to minimise the multi-task loss function for an image, defined as (Girshick, 2015):

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

where i is the index of an anchor and p_i is the predicted probability of anchor i being a spill target. The ground-truth labelled p_i^* equals 1 if the anchor is positive, otherwise it equals 0. t_i is a vector containing the four parameterised coordinates of the predicted bounding box, and t_i^* is the vector of the ground-box labelled as a positive anchor. The classification loss L_{cls} is a cross-entropy loss over two classes. The regression loss L_{reg} is smooth $L1$ defined in (Dai et al., 2015). These two parts are normalised by N_{cls} and N_{reg} with a balancing parameter λ. The default settings of N_{cls}, N_{reg}, and λ are 254, 2400, and 10, respectively (Ren et al., 2015).

After the above steps, the classifier in Faster R-CNN is the same as Fast R-CNN, using two fully connected layers to produce a class score for each location bounding box.

Transfer learning provides a time-saving way to build accurate models for specific detection tasks, and it is a popular method in computer vision (Rawat and Wang, 2017). By using pre-trained models which have been previously trained on large datasets instead of starting the learning process from scratch, we can directly use the trained architecture and obtained weights and apply the learning on HNS spill detection. The pre-trained ResNet-101 based on the COCO dataset was adopted to accelerate the training process for the HNS detection models. To adapt the pre-trained ResNet-101 to the HNS spill detection task, we fine-tuned the network weights using the augmented dataset. Because the pre-trained model was well-trained, the learning rate could not be set too large to reduce the risk of losing previous knowledge. In this work, the multi-level learning rates are listed in Table 1.

2.4. Spill area segmentation based on faster R-CNN

The image segmentation algorithms, including LFTM, Otsu, and Max entropy, extracted the HNS area at pixel level from the location bounding box generated by the Faster R-CNN. The segmentation

Table 1
Learning rate settings.

Steps	Learning rate
0-50,000	0.002
50,000–120,000	0.001
120,000–160,000	0.0006
160,000–180,000	0.0003
160,000–200,000	0.00006

results were compared with each other and the results by applying improved LFTM, Otsu, and Max entropy algorithms to the original images, as proposed by Zhan et al. (2019).

2.5. Evaluation indicator

To evaluate the location detection of the Faster R-CNN model, *intersection-over-union* (*IoU*, Eq. (2)) was adopted to calculate the overlap of the detected bounding box with the ground-truth bounding box.

$$\mathrm{IoU} = \frac{area(Dbox \cap GT)}{area(Dbox \cup GT)} \tag{2}$$

where $Dbox \cup GT$ and $Dbox \cap GT$ denote the union and intersection of the detected bounding box (*Dbox*) and ground-truth (*GT*) bounding box, respectively, and *area* is the area/total pixels of the box.

To evaluate the classification results of Faster R-CNN and the results of Faster R-CNN-based segmentation, we chose the accuracy (*AC*), precision (*PR*), recall (*RE*), and F1 score (*F1*) as typical indicators at pixel level, and these are defined as:

$$AC = \frac{TP + TN}{TN + TP + FN + FP} \tag{3}$$

$$\mathrm{PR} = \frac{TP}{TP + FP} \tag{4}$$

$$\mathrm{RE} = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = \frac{2PR \times RE}{PR + RE} \tag{6}$$

where true positive (*TP*) is the number of pixels being classified into the correct target category, and true negative (*TN*) is the number of pixels that are correctly recognised as non-target categories. False negative (*FN*) is the number of target pixels being classified into the wrong category, and false positive (*FP*) is the number of pixels of a non-target category being incorrectly classified as a target. *AC*, *PR*, *RE*, and *F1* range from 0 (worst) to 1 (best), respectively. The ground-truth of HNS segmentation refers to the mask in the original images instead of the HNS mask in the location bounding box.

3. Results and discussion

3.1. RGB and blue channel images

3.1.1. Location bounding box

The test RGB images were detected by the RGB trained model, and the average processing time was approximately 0.6 s per image. For RGB images, the location bounding box produced by Faster R-CNN is shown in Fig. 2, where the green bounding box indicates the ground-truth location, the blue bounding box indicates the bounding box detected by Faster R-CNN, and the red bounding box indicates the intersection of the above two bounding boxes.

In Fig. 2, most locations of the transparent HNS targets could be detected by Faster R-CNN, even when there were no complete HNS profiles in Fig. 2(d) and (f), which suggested that with the help of the proper method, colourless HNS could be detectable, although they may be indistinguishable by common methods, especially in images containing a background with different water colours, surface waves, or illumination conditions. Similar objects like sun reflectance and clouds in Fig. 2(g) could also be distinguished as

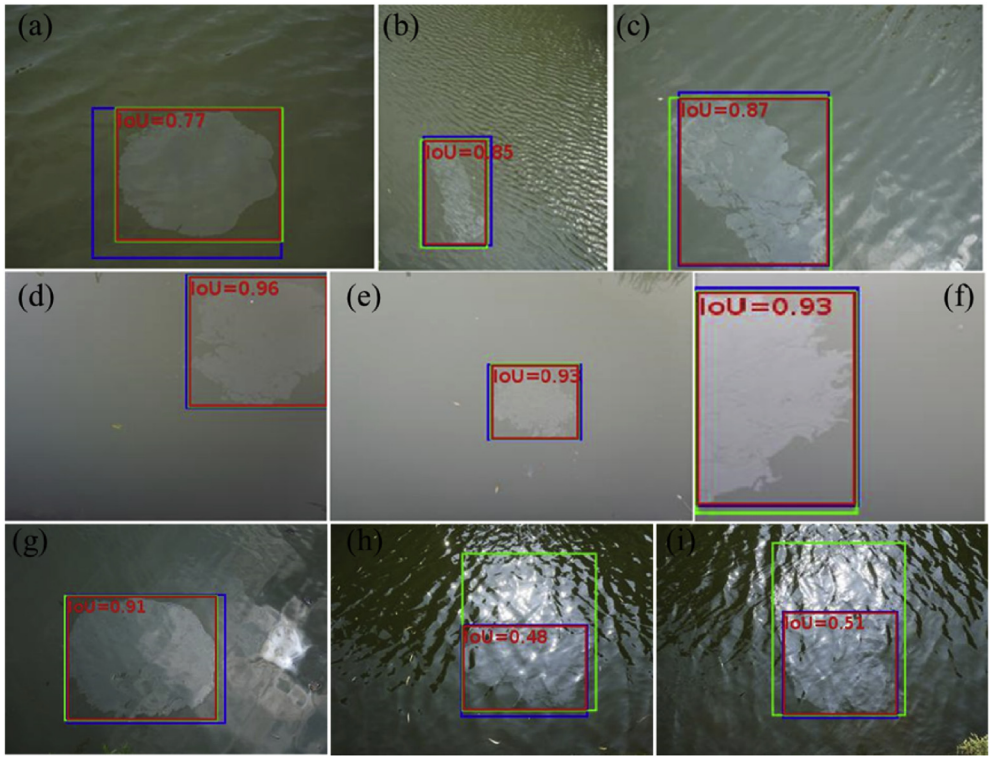

Fig. 2. Examples of *IoU* in RGB images.

background successfully, due to the powerful feature extraction capacity of deep neural networks. In Fig. 2(h) and (i), the reflectance of the sun coincides with the spill area; the spill area is highly overlapped with the reflectance area and cannot be accurately distinguished from the sun reflectance, resulting low *IoU*. However, part of the bounded region can still be recognised, even though it has a strong similarity to the water background.

Fig. 3 presents the *IoU* statistics for 60 test images. Most of the detected bounding boxes for the three kinds of HNS were correct, except that the *IoU* values of two xylene samples are close to 0.5, as shown in Fig. 2(h) and (i). The average *IoU* values for benzene, xylene, and palm oil were 0.821, 0.817, and 0.926 respectively. The Faster R-CNN model yielded better location detection results for palm oil than for benzene or xylene, which may be associated with the more significant reflectance of palm oil at the 500–700 nm wavelengths (Huang et al., 2019). Successful location detection of HNS in a mosaicked image combined with airborne GPS information could assist in instantaneously locating HNS after an incident.

3.1.2. Classification based on faster R-CNN

The classification result of Faster R-CNN-processed RGB is shown in Fig. 4. The number of misclassified images was 10, 9, and 0 out of 20 samples for benzene, xylene, and palm oil, respectively. The statistical results of classification are listed in Table 2. Consistent with location detection, the classification results of palm oil

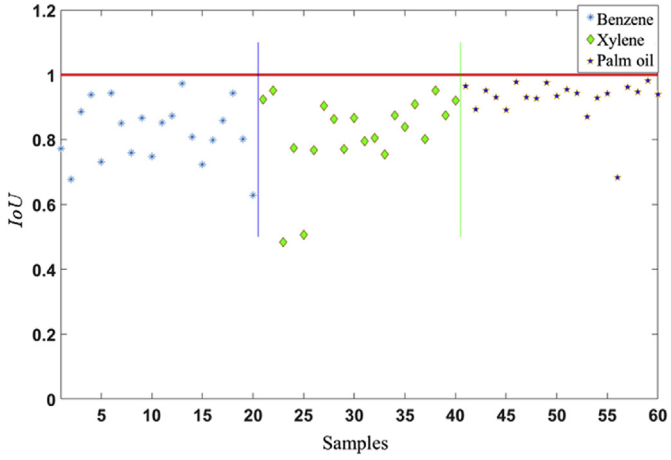

Fig. 3. *IoU* results of RGB test images.

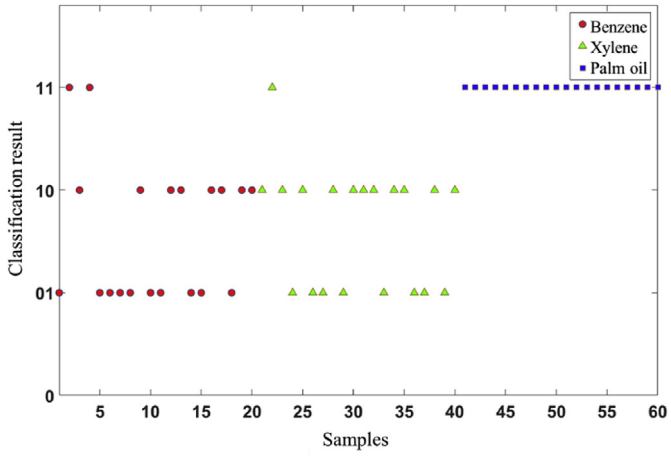

Fig. 4. Classification result of RGB test images.

Table 2
Statistical classification performance of RGB test images.

Category	*AC*	*PR*	*RE*	*F1*
Benzene	0.700	0.555	0.500	0.526
Xylene	0.716	0.578	0.550	0.564
Palm oil	0.950	0.869	1.000	0.930

achieved an F1 score of 0.930, higher than the 0.526 of benzene or the 0.564 of xylene. The *AC* and *PR* of palm oil were slightly lower than 1, as only two samples were misclassified as benzene and only one sample was misclassified as xylene. Both the chemical composition and reflectance characteristics of the three HNS under visible light were similar, which increased the difficulty of classification.

In an incident, there are different response measures for spills with different toxicity. Therefore, BTEX chemicals (which includes benzene and xylene) should be identified correctly as BTEX chemicals are much more harmful to both the environment and humans than palm oil. In a binary classification between BTEX chemicals (benzene and xylene) and palm oil, the classification resulted in an F1 score of 0.762. It may be possible to classify BTEX and palm oil using RGB imaging and Faster R-CNN.

3.1.3. Segmentation based on faster R-CNN

Different segmentation methods were used on images with bounding boxes to extract the exact HNS area (region of interest) at the pixel level. Otsu and Max entropy are fully automatic methods without parameter settings. The LFTM requires an initial number of clusters, which was set to 2 clusters here, i.e. background and HNS, due to the comparatively homogeneous background in the location bounding box.

To analyse the efficiency of Faster R-CNN based segmentation, we also tested the masking on 60 original test images (without Faster R-CNN) using Otsu, Max entropy, and the improved LFTM proposed in Zhan et al. (2019), respectively. Based on the recommended parameter settings of the improved LFTM, three important parameters, $const_{th}$, th_1, and th_2, were selected as the tuning object, and a total of 45 parameter combinations were traversed to select the optimal set of parameters. The $const_{th}$, th_1, and th_2 of the improved LFTM were set to 85, 5, and 0.25, respectively. The cluster number of the improved LFTM was adapted to the images.

The mean values of the four quantitative evaluation measures (*AC*, *PR*, *RE*, *F1*) for segmentation with and without Faster R-CNN are listed in Fig. 5 and Table 3. All three Faster R-CNN-based segmentation methods performed well. The highest computational time was approximately 0.647 s per image using Faster R-CNN + LFTM. The results proved that Faster R-CNN-based segmentation is suitable for RGB images. The segmentation on the original RGB images using Otsu, Max entropy, and the improved LFTM failed with low *PR* and *F1*. The high recall (*RE*) of Otsu and Max entropy was 'fake' due to the misrecognition of many non-target images as target images. The improved LFTM failed to extract the HNS mask from the RGB images, possibly because it was proposed in UV images [14]. The *AC*, *PR*, *RE*, and *F1* of the spill segmentation in the RGB images was significantly enhanced after the implementation of Faster R-CNN. The bounding box which was generated by Faster R-CNN determined the approximate range of the HNS against a complicated background, thereby simplifying the background and highlighting the differences between it and the HNS, which resulted in the successful masking of the HNS spill.

3.1.4. Single blue channel images

According to the results in [15], the four wavelengths 365, 410, 450, and 850 nm were suggested for HNS classification, where 410

Fig. 5. Segmentation examples of RGB images. (a) Original images (the blue box is a location bounding box), (b) Faster R-CNN + Otsu, (c) Faster R-CNN + Max entropy, (d) Faster R-CNN + LFTM with $N_{cluster} = 2$, (e) Otsu, (f) Max entropy, (g) improved LFTM (h) ground-truth. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Results of different segmentation methods on RGB test images.

Method	AC	PR	RE	F1	Average time(s)
Faster R-CNN + Otsu	0.933	0.771	0.833	0.792	0.606
Faster R-CNN + Max entropy	0.921	0.703	0.734	0.696	0.608
Faster R-CNN + LFTM with $N_{cluster} = 2$	0.943	0.807	0.860	0.826	0.647
Otsu	0.545	0.252	0.852	0.361	0.009
Max entropy	0.606	0.222	0.575	0.271	0.012
Improved LFTM	0.815	0.292	0.224	0.236	0.551

and 450 nm were covered by the broad blue channel. In order to verify whether RGB images using a single blue channel can be used for location detection, classification, and segmentation, a Faster R-CNN model using blue channel images was trained. The spectral range of the single blue channel was approximately 400–600 nm, and the full width at half maximum (FWHM) was approximately 100 nm.

Consistent with the RGB images, the single blue channel images obtained a satisfactory location detection result, where the average *IoU*s of benzene, xylene, and palm oil were 0.756, 0.773, and 0.919,

Table 4
Statistical classification performance of single blue channel test images.

Category	*AC*	*PR*	*RE*	*F1*
Benzene	0.700	0.555	0.500	0.526
Xylene	0.683	0.529	0.450	0.486
Palm oil	0.916	0.800	1.000	0.888

respectively. The classification and segmentation results are shown in Tables 4 and 5, respectively, and are consistent with the results of the RGB images. Palm oil can be distinguished from benzene and xylene, while it is difficult to distinguish benzene and xylene from the blue channel images.

After tuning the model as described in Section 3.1.3, the parameters of the improved LFTM for segmentation in the original blue channel images were set to $cons_{th} = 80$, $th_1 = 25$, and $th_2 = 0.25$. Otsu, Max entropy, and the improved LFTM did not perform well on the blue channel images, but performed slightly better than on the RGB images, suggesting that there was redundant information in the RGB images for HNS spill segmentation. It was indicated that segmentation of the HNS spill was a challenging task for both RGB images and blue channel images.

However, implementing the Faster R-CNN improved the segmentation performance significantly. The HNS masks were successfully extracted from the blue channel images by different Faster R-CNN-based segmentation methods with the assistance of the accurate location bounding boxes that were generated by the Faster R-CNN. In Tables 3 and 5, the HNS in both the RGB and single blue channel images were successfully segmented by using Faster R-CNN, where the computational complexity could be further improved.

3.2. Spectral images at 365 nm

3.2.1. Location bounding box

UV images captured at 365 nm in different scenarios, such as a swimming pool, an artificial channel, or a lake, were adopted to train and build a Faster R-CNN model. The different scenarios included different interference, lighting conditions, shadows, and sun reflectance for HNS identification. The test images were processed by the UV images trained by the Faster R-CNN model, with an average processing time of 0.6 s per image.

Examples of the bounding box of the detected location in the 365 nm test images by Faster R-CNN are shown in Fig. 6. In Fig. 6, various geometric shapes of HNS, such as a long line, an oval, and a circle, are detectable in different environments, which indicates the generalisability of the Faster R-CNN-based detection for images recorded by UV sensors.

Fig. 6(a)–(c) shows the detection results under different water environments (lake, swimming pool, and artificial channel) with different surface texture features. Noise like sun glitter partly crosses the target (Fig. 6(c)) or beside the target (Fig. 6(h) and (i)),

Table 5
Results of the different segmentation methods on single blue channel test images.

Method	AC	PR	RE	F1	Average time(s)
Faster R-CNN + Otsu	0.931	0.761	0.812	0.776	0.606
Faster R-CNN + Max entropy	0.923	0.737	0.771	0.722	0.608
Faster R-CNN + LFTM with $N_{cluster} = 2$	0.932	0.773	0.808	0.781	0.647
Otsu	0.563	0.274	0.903	0.388	0.009
Max	0.599	0.233	0.631	0.284	0.012
Improved LFTM	0.829	0.384	0.293	0.309	0.666

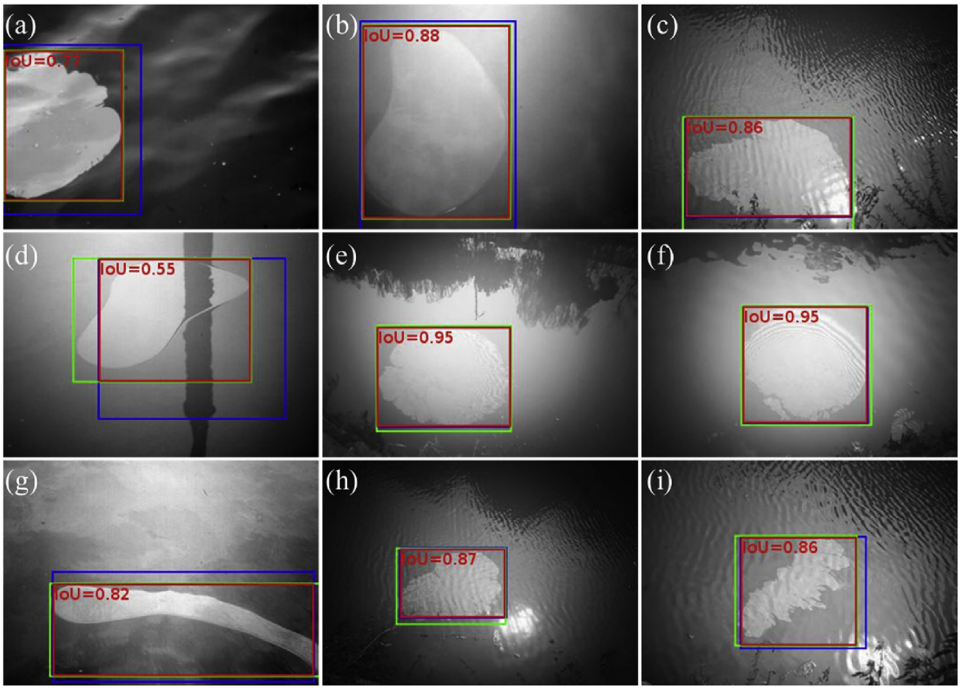

Fig. 6. Examples of *IoUs* of 365 nm spectral images.

and the shadows (Fig. 6(d), (e) and (g)) were successfully detected as the correct category. Fig. 6(g) shows that the model can detect spill targets in an image with uneven illumination. The detection result shows that the generalisability of the model is satisfactory for spills floating on water.

The average *IoU*s of benzene, xylene, and palm oil were 0.849, 0.818, and 0.940, respectively. By examining Figs. 3–7, the 365 nm spectral images yield better location results for the test samples than the RGB images. The location detection performance for palm oil in UV images was still the best of all three types of HNS, as palm oil also outperformed the other HNS in RGB images.

3.2.2. Classification using faster R-CNN

The number of misclassified images was 5, 3, and 0 for benzene, xylene, and palm oil, respectively (Fig. 8). Statistical classification results for each category at the pixel level are listed in Table 6. The mean *F1* score for the classification of spectral images at 365 nm increased to 0.864, which was better than the 0.673 of RGB images, demonstrating significant improvement to the classification by UV images. All of the evaluation indicators were greater than or equal to 0.750, with the contribution to accurate selection by the bounding box and the different absorption features of HNS at 365 nm (Huang et al., 2019; Zhan et al., 2019). The classification of xylene and benzene could be improved further. Compared with the classification accuracy (≥95%) by PLS-DA and LS-SVM of the multispectral images at four wavelengths by Huang et al. (2019), the classification results by Faster R-CNN are acceptable but could be improved. These results indicate that it is quite possible to classify floating HNS in a single spectral UV image using Faster R-CNN.

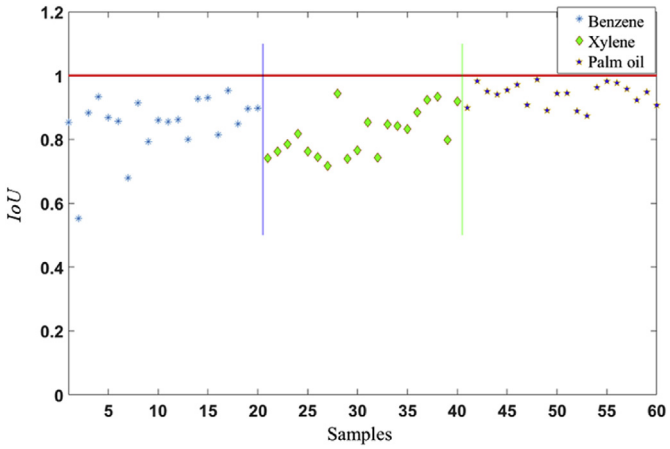

Fig. 7. *IoU* results of 365 nm spectral test images.

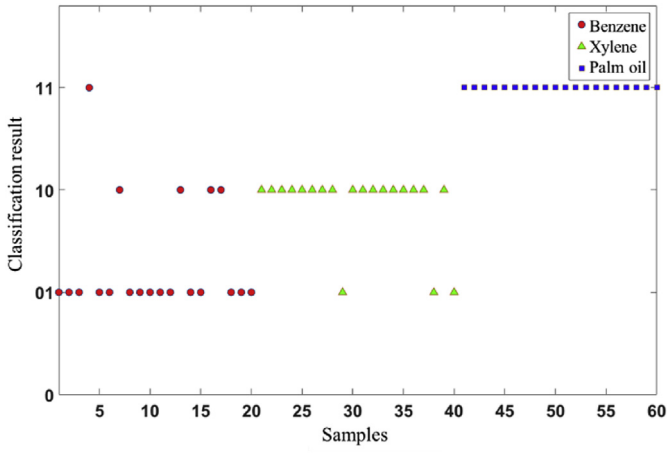

Fig. 8. Classification results of 365 nm spectral test images.

Table 6
Statistical classification performance of 365 nm spectral test images.

Category	*AC*	*PR*	*RE*	*F1*
Benzene	0.866	0.833	0.750	0.789
Xylene	0.883	0.809	0.850	0.829
Palm oil	0.983	0.952	1.000	0.975

3.2.3. Segmentation results based on faster R-CNN

As was done with the RGB and blue channel images, we used three segmentation methods on the location bounding box in the UV images captured at 365 nm by Faster R-CNN, and we used Otsu, Max entropy, and an improved LFTM (Zhan et al., 2019) on the original UV images. The $const_{th}$, th_1, and th_2 of the improved LFTM were set to 80, 5, and 0.45, respectively.

As shown in Fig. 9 and Tables 3, 5 and 7, with similar computational complexity, the segmentation on single spectral UV images generally outperforms the RGB and blue channel images, where Faster R-CNN + Max entropy achieves a large improvement in *F1* from 0.696 to 0.791. The high contrast between HNS and the background in UV images contributed to the HNS segmentation.

In Table 7, the comparison between the segmentation on UV images with and without Faster R-CNN proves that the efficiency of the HNS segmentation was improved significantly by Faster R-CNN, especially in vivid scenarios where the segmentation can be affected by the complex background, which is mostly outside the location bounding box.

Similar to the results obtained by Zhan et al. (2019), and different from the RGB and blue channel images, the improved LFTM significantly improved the segmentation results for the original UV images compared with Otsu and Max entropy, as shown in Fig. 9 and Table 7. The *RE* and *F1* of the improved LFTM on 60 test UV images did not achieve a result as good as the one by Zhan et al. (2019), where the UV images were only collected in an artificial channel instead of the vivid scenarios of swimming pool, lake and artificial channel in this study. This would affect the enhancement and clustering by the improved LFTM, and image details in a swimming pool scenario would be lost due to the low shooting distance. Many target pixels were falsely recognised as non-target, which indicated that further training of the improved LFTM was necessary, as the parameter tuning could help improve the model generalisation.

The *F1* score of the improved LFTM on the original images is lower than that of the Faster R-CNN + Otsu/Max entropy/LFTM. The *AC* of the improved LFTM outperformed the Faster R-CNN-based methods, indicating a high accuracy of segmentation. The results indicated that spectral images captured at 365 nm used with Faster R-CNN have great potential for segmentation in multiple scenarios.

3.3. Possibility of single spectral band for HNS identification

Three models using the same Faster R-CNN architecture were trained on the spectral dataset for the 365 nm, RGB, and blue channel images, respectively. All models performed well on location detection with an average *IoU* greater than 0.756, while the models performed differently on classification and segmentation. The processing time per image was approximately 0.6 s, which means that the spill area could be segmented from an HNS image in

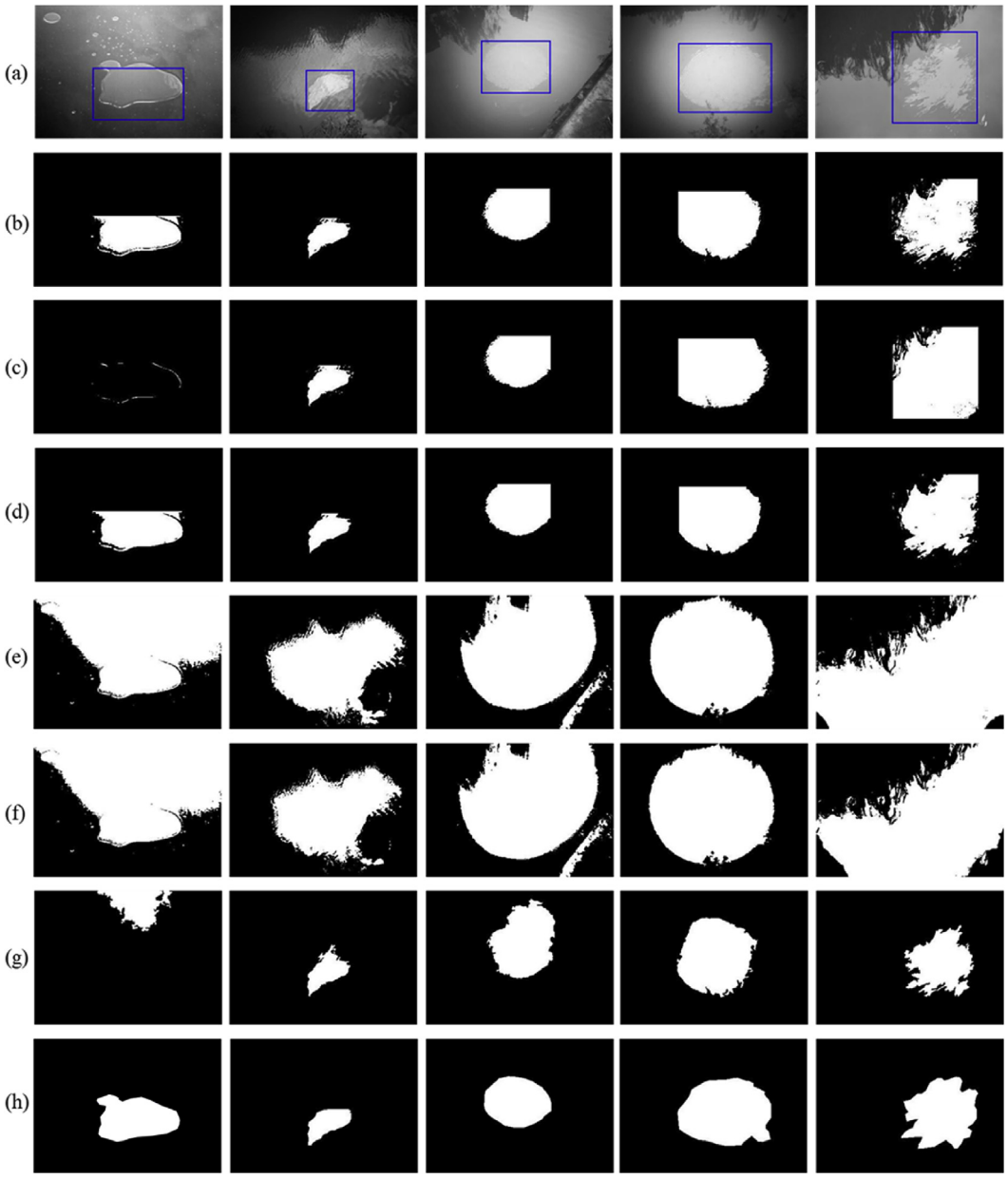

Fig. 9. Segmentation examples of UV images. (a) Original images (the blue box is the location bounding box), (b) Faster R-CNN + Otsu, (c) Faster R-CNN + Max entropy, (d) Faster R-CNN + LFTM with $N_{cluster} = 2$, (e) Otsu, (f) Max entropy, (g) Improved LFTM (h) ground-truth. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 7
Results of segmentation on 365 nm spectral test set by different methods.

Method	AC	PR	RE	F1	Average time(s)
Faster R-CNN + Otsu	0.934	0.800	0.862	0.822	0.607
Faster R-CNN + Max entropy	0.928	0.761	0.861	0.791	0.610
Faster R-CNN + LFTM with $N_{cluster} = 2$	0.935	0.807	0.856	0.823	0.652
Otsu	0.631	0.313	0.951	0.443	0.009
Max entropy	0.644	0.329	0.946	0.459	0.012
Improved LFTM	0.888	0.765	0.590	0.649	0.666

less than a second in common detection applications.

The classification of single spectral images at 365 nm demonstrated better classification capability of UV images compared to RGB and blue channel images. UV images were more stable than RGB and blue channel images, with better *AC*, *PR*, *RE*, and *F1*. Although existing studies have shown that the LS-SVM model can achieve an accuracy greater than 95% on spectral images at four optimal spectral bands (Huang et al., 2019), the cost and difficulty of

acquiring multi-spectral spill images is higher than the cost of acquiring single band images. With an increase in training data and the optimisation of deep learning frameworks, HNS classification on single spectral images combined with Faster R-CNN may be further improved. The results also indicate that the key factor affecting the classification accuracy may be the imaging band, as the classification of the RGB and blue channel images requires more improvement than the UV images at 365 nm.

Compared with the floating HNS segmentation using UV images of one scenario in Zhan et al. (2019), HNS segmentation based on a location bounding box by Faster R-CNN, and Otsu, Max entropy, and LFTM yielded satisfactory results for single spectral band images at 365 nm, single blue channel images, and RGB images, with the best performance on 365 nm UV images. Comparing the masking result on original images to the one in bounding box, the segmentation of HNS in the RGB images and the blue channel images became possible after the adoption of Faster R-CNN. In addition, the floating HNS segmentation in different scenarios would be much more efficient if Faster R-CNN were used.

Single spectral images captured at 365 nm used with Faster R-CNN demonstrated great potential for the automatic location detection, classification, and segmentation of floating HNS spills. More training would help to improve the Faster R-CNN-based method.

3.4. Analysis for practical implementation

Speed is one of the main concerns for the practical implementation of the proposed algorithm in rapid response. There would be technological limitations for both the hardware and the algorithm at the current stage of implementation. In an UAV application, hardware affecting speed mainly involves camera, wireless transmitter, and computer on the ground.

The camera affects the contrast between the spill target, the background, and interference in UV images. A camera with high sensitivity is required for the proposed method to guarantee the accuracy or even reduce the complexity of the algorithm and enhance the speed. The detection time of the proposed method is acceptable (currently approximately 0.6 s per image), but it should be improved further as there are many images used in real-time detection. The images could be processed on board or transmitted to a ground station followed by processing by the proposed method. Assuming stable and fast wireless data transfer between a UAV platform and a ground station, images could be analysed on any computer, where computers on the ground could be expanded. Therefore, HNS detection using a Faster R-CNN framework on a high-performance workstation (e.g. equipped with a graphic processing unit) on the ground is preferable to on board DCNN detection. Limitation of computer may slow down the response speed in real time detection in sea water case.

For the algorithm aspect, the limitation of Faster R-CNN is the high computational hardware demand and the massive amount of data required for long-term training, which mainly results from the convolution in the architecture. Research has been conducted to reduce the training time by means of separable convolution (Chen et al., 2018; Chollet, 2017). Compared with the standard convolution of ResNet-101 in our study, separable convolution decomposes a standard convolution into depth-wise convolution and pointwise convolution to reduce the total number of parameters in the network and accelerate convergence, thus generating a model of lesser volume. A low-volume model reduces the hardware requirement and is more suitable for deployment on a variety of terminals (e.g. PC, smartphone, even industrial computer on board) to improve the efficiency of real-time monitoring.

In a realistic sea scenario, various conditions may lead to incorrect results, i.e. errors or low accuracy. There are several possible errors and causes.

1) Over-fitting problems. This happens when the model learns the features and noise in the training data so deeply that it may not generalise well to unseen features and noise. In this study, images were collected under different lighting conditions, most with sufficient lighting. To build a large database, we applied data augmentation, the effectiveness of which has been proved in a large number of applications (Ding et al., 2016; Yu et al., 2017). However, data augmentation cannot introduce unseen image features or noises in the augmented images, thus resulting in over-fitting problems or limited generalisability of the network. Suggested solutions to this problem are to collect more data containing various scenarios for training, such as various lighting conditions (Mikołajczyk and Grochowski, 2018).
2) Low contrast between the spill area and the background or the blurred edge of the spill area on seawater. This is mainly due to the low viscosities and low surface tension forces of the studied liquid chemicals, which makes them form a thin liquid film on the surface of the water. When the illumination is insufficient, the UV image quality may be poor. Images in various scenarios and efficient image pre-processing may be explored, whereas the recommended solution is a sensitive imaging device.
3) Complex backgrounds, including interferences such as reflections, increase the difficulty of HNS detection. Some reflections had features similar to those of a spill in the selected UV images. Our current dataset included images with reflections on fresh and saltwater, on which the Faster R-CNN performed well. However, reflections on seawater may have their own pattern. Images in different realistic scenarios should be complemented in the training set, and image pre-processing methods could be explored to assist the proposed method.

4. Conclusion

In this paper, we proposed a Faster R-CNN architecture-based method for automatic identification, i.e. the location, classification, and mask of transparent floating HNS. Spectral images at 365 nm, RGB, and blue channel images were studied to explore the optimal imaging method for HNS identification. The results demonstrated that single spectral images at 365 nm used with Faster R-CNN were suitable for HNS spill location detection, classification, and segmentation. RGB and blue channel images used with Faster R-CNN were suitable for the location and mask detection of HNS. Instead of SAR and RADAR, successful HNS identification of single spectral images at 365 nm would significantly reduce the complexity and cost of an HNS imaging system, and thus enhance the efficiency of HNS imaging.

Possible errors may occur in the practical implementation of the proposed method, with interference from overfitting, low contrast, and complex backgrounds. It is reasonable to believe that further improvement can be achieved for the proposed Faster R-CNN-based practical identification with single spectral images by inputting more images in various scenarios as the training set. Future research work will focus on improving the speed and accuracy of the proposed method, especially for practical application.

CRediT authorship contribution statement

Hui Huang: Conceptualization, Methodology, Software. **Chao Wang:** Data curation, Software, Writing - original draft. **Shuchang Liu:** Data curation, Investigation. **Zehao Sun:** Software, Investigation. **Dejun Zhang:** Software, Validation. **Caicai Liu:** Resources, Writing - review & editing. **Yang Jiang:** Writing - review & editing.

Shuyue Zhan: Supervision. **Haofei Zhang:** Resources, Writing - review & editing. **Ren Xu:** Resources, Funding acquisition.

Acknowledgements

Funding This work was financially supported by the National Natural Science Foundation of China (Grant number: 31801619), Natural Science Foundation of Zhejiang Province (Grant number: LY18F050002), and the National Key Research and Development Program of China (Grant number: 2016YFC1402403).

References

Aja-Fernández, S., Curiale, Ariel Hernán, Vegas-Sánchez-Ferrero, Gonzalo, 2015. A local fuzzy thresholding methodology for multiregion image segmentation. Knowl. Based Syst. 83, 1–12.

Alpers, W., Holt, B., Zeng, K., 2017. Oil spill detection by imaging radars: challenges and pitfalls. Remote Sens. Environ. 201, 133–147.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.

Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258.

Cunha, I., Moreira, S., Santos, M.M., 2015. Review on hazardous and noxious substances (HNS) involved in marine spill incidents-An online database. J. Hazard Mater. 285, 509–516.

Dai, J., He, K., Sun, J., 2015. Convolutional feature masking for joint object and stuff segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3992–4000.

Ding, J., Chen, B., Liu, H., Huang, M., 2016. Convolutional neural network with data augmentation for SAR target recognition. IEEE. Geosci. Rem. Sens. Lett. 13 (3), 364–368.

General Office of the State Council, P, 2011. Regulation on the Safety Management of Hazardous Chemicals(2011 Revision).

Gérin, M., Siemiatycki, J., Désy, M., Krewski, D., 1998. Associations between several sites of cancer and occupational exposure to benzene, toluene, xylene, and styrene: results of a case-control study in Montreal. Am. J. Ind. Med. 34 (2), 144–156.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587.

Gull, S.F., Skilling, J., 1984. Maximum entropy method in image processing. IEE Proc. F (Commun. Radar Signal Process.) 131 (6), 646–659.

Harold, P., De Souza, A., Louchart, P., Russell, D., Brunt, H., 2014. Development of a risk-based prioritisation methodology to inform public health emergency planning and preparedness in case of accidental spill at sea of hazardous and noxious substances (HNS). Environ. Int. 72, 157–163.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hilmi, A., Luong, J.H., 2000. Micromachined electrophoresis chips with electrochemical detectors for analysis of explosive compounds in soil and groundwater. Environ. Sci. Technol. 34 (14), 3046–3050.

Huang, H., Liu, S., Wang, C., Xia, K., Zhang, D., Wang, H., Zhan, S., Huang, H., He, S., Liu, C., Li, X., 2019. On-site visualized classification of transparent hazards and noxious substances on a water surface by multispectral techniques. Appl. Optic. 58 (16), 4458–4466.

IMO, 2000. Protocol on Preparedness, Response and Co-operation to Pollution Incidents by Hazardous and Noxious Substances (OPRC-HNS Protocol).

Koeber, R., Bayona, J.M., Niessner, R., 1999. Determination of benzo [a] pyrene diones in air particulate matter with liquid chromatography mass spectrometry. Environ. Sci. Technol. 33 (10), 1552–1558.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Neural Inf. Process. Syst. 1097–1105.

Li, C.W., Benjamin, M.M., Korshin, G.V., 2000. Use of UV spectroscopy to characterize the reaction between NOM and free chlorine. Environ. Sci. Technol. 34 (12), 2570–2575.

Liu, Z., Hu, J., Weng, L., Yang, Y., 2017. Rotated region based CNN for ship detection. In: 2017 IEEE International Conference on Image Processing. ICIP, pp. 900–904.

Mackay, D., Shiu, W.Y., Ma, K.C., Lee, S.C., 2006. Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals.

Mikołajczyk, A., Grochowski, M., 2018. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinoujście, pp. 117–122.

Nieto-Hidalgo, M., Gallego, A.J., Gil, P., Pertusa, A., 2018. Two-stage convolutional neural network for ship and spill detection using SLAR images. IEEE. Trans. Geosci. Remote. Sens. 56 (9), 5217–5230.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst, Man, Cybern. 9 (1), 62–66.

Perez, L., Wang, J., 2017. The Effectiveness of Data Augmentation in Image Classification Using Deep Learning. arXiv:1712.04621.

Purnell, K., 2009. Are HNS spills more dangerous than oil spills. In: Proceedings of the A White Paper for the Interspill Conference & the 4th IMO R&D Forum, pp. 12–14.

Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 29 (9), 2352–2449.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. Neural Inf. Process. Syst. 91–99.

Singha, S., Ressel, R., Lehner, S., 2016. Multi-frequency and multi-polarization analysis of oil slicks using TerraSAR-X and RADARSAT-2 data, 2016. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4035–4038.

Taravat, A., Del Frate, F., 2012. Development of band ratioing algorithms and neural networks to detection of oil spills using Landsat ETM+ data. EURASIP. J. Adv. Sig. Process. 2012 (1), 1–8.

Yao, Y., Jiang, Z., Zhang, H., Zhao, D., Cai, B., 2017. Ship detection in optical remote sensing images based on deep convolutional neural networks. J. Appl. Remote Sens. 11 (4), 042611.

Yu, X., Wu, X., Luo, C., Ren, P., 2017. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. GISci. Remote Sens. 54 (5), 741–758.

Yu, X., Zhang, H., Luo, C., Qi, H., Ren, P., 2018. Oil spill segmentation via adversarial f-divergence learning. IEEE. Trans. Geosci. Remote. Sens. 56 (9), 4973–4988.

Zhan, S., Wang, C., Liu, S., Xia, K., Huang, H., Li, X., Liu, C., Xu, R., 2019. Floating xylene spill segmentation from ultraviolet images via target enhancement. Remote Sens. 11 (9), 1142. Basel.

Zhao, J., Temimi, M., Ghedira, H., Hu, C., 2014. Exploring the potential of optical remote sensing for oil spill detection in shallow coastal waters-a case study in the Arabian Gulf. Opt. Express 22 (11), 13755–13772.