



Deep invariant texture features for water image classification



Minglong Xue¹ · Palaiahnakote Shivakumara² · Xuerong Wu¹ · Tong Lu¹  · Umapada Pal³ · Michael Blumenstein⁴ · Daniel Lopresti⁵

Received: 24 June 2020 / Accepted: 12 November 2020 / Published online: 24 November 2020
© Springer Nature Switzerland AG 2020

Abstract

Detecting potential issues in naturally captured images of water is a challenging task due to visual similarities between clean and polluted water, as well as causes posed by image acquisition with different camera angles and placements. This paper presents novel deep invariant texture features along with a deep network for detecting clean and polluted water images. The proposed method first divides an input image into H, S and V components to extract finer details. For each of the color spaces, the proposed approach generates two directional coherence images based on Eigen value analysis and gradient distribution, which results in enhanced images. Then the proposed method extracts scale invariant gradient orientations based on Gaussian first order derivative filters on different standard deviations to study texture of each smoothed image. To strengthen the above features, we explore the combination of Gabor-wavelet-binary pattern for extracting texture of the input water image. The proposed method integrates merits of aforementioned features and the features extracted by VGG16 deep learning model to obtain a single feature vector. Furthermore, the extracted feature is fed to a gradient boosting decision tree for water image detection. A variety of experimental results on a large dataset containing different types of clean and stagnant water images show that the proposed method outperforms the existing methods in terms of classification rate and accuracy.

Keywords Gradient direction · Eigen values · Texture · Gabor wavelet features · VGG model · Gradient boosting decision tree · Clean water · Polluted water · Water image classification · Hydrology

1 Introduction

Water image classification has received special attention for researchers because it plays a vital role in analyzing surface water for agriculture, food production, domestic water consumption, classification of rain water and monitoring river water quality [4, 5, 15]. Apart from that, there are other surveillance applications, where water image analysis is essential, such as monitoring floods to prevent disasters, detecting water hazards, building aerial water

maps e.g. for safe zone detection to land drones, and wildlife surveillance to detect animals [28, 32]. For all the above applications, water image analysis and classification of images of different water types help to improve the performance of the systems significantly. There are several methods proposed for analyzing water reflection and depth of water and underwater image restoration in the literature. Zong et al. [32] developed an approach for water reflection recognition, Yang et al. [28] proposed a method for analyzing depth from water reflections, Peng

✉ Tong Lu, lutong@nju.edu.cn; Minglong Xue, xueml@smail.nju.edu.cn; Palaiahnakote Shivakumara, shiva@um.edu.my; Xuerong Wu, 403054326@qq.com; Umapada Pal, umapada@isical.ac.in; Michael Blumenstein, michael.blumenstein@uts.edu.au; Daniel Lopresti, lopresti@cse.lehigh.edu | ¹National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China. ²Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. ³Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India. ⁴Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia. ⁵Computer Science & Engineering, Lehigh University, Bethlehem, PA, USA.



et al. [17] proposed underwater image restoration based on image blurriness and light absorption. The primary goal of these methods is to detect water reflection and understanding underwater images, but not the classification of different water images as in the proposed work. Similarly, methods have been proposed in the past for classification of images containing water. Shi and Pun [23] proposed super pixel-based 3D deep neural networks for hyperspectral image classification. Galvis et al. [2] proposed remote sensing image analysis by aggregation of segmentation-classification collaborative agents. These methods usually target classification of remote sensing images but not the images captured by normal cameras. However, we scarcely find methods for the classification of multiple clean and polluted water images. Besides, these methods usually focus on a particular type of water, which may include the water of a river, pond, ocean, fountain lake, etc., but not different types of polluted waters, such as waters with algae, animals, fungi, oil and rubbish.

According to the literature [28, 32], the classification of different types of clean water images is still considered to be challenging because the surfaces of such water images

may share similar properties. For example, when we look at sample images of clean water, namely, fountains, lakes or ocean, and polluted water, namely, algae, animals, fungi, industrial pollution, oil and rubbish, respectively as shown in Fig. 1, where one can see the common information in different water type images. Thus, we can assert that the classification of clean and polluted water images of different types is much more challenging. Hence, there is a scope for proposing a new imaging system for the classification of images with different types of clean and polluted water.

This work focuses on developing a method combining handcrafted and deep features with a gradient boost decision tree for classification of water images. It is noted that color, gradient, gradient orientation, texture and spatial information are the key features to represent different types of water images. For instance, color and gradient information are the salient features for representing different clean water images while texture, color and spatial information are the significant features for representing different polluted water images. These observations motivated us to propose the following features. We propose

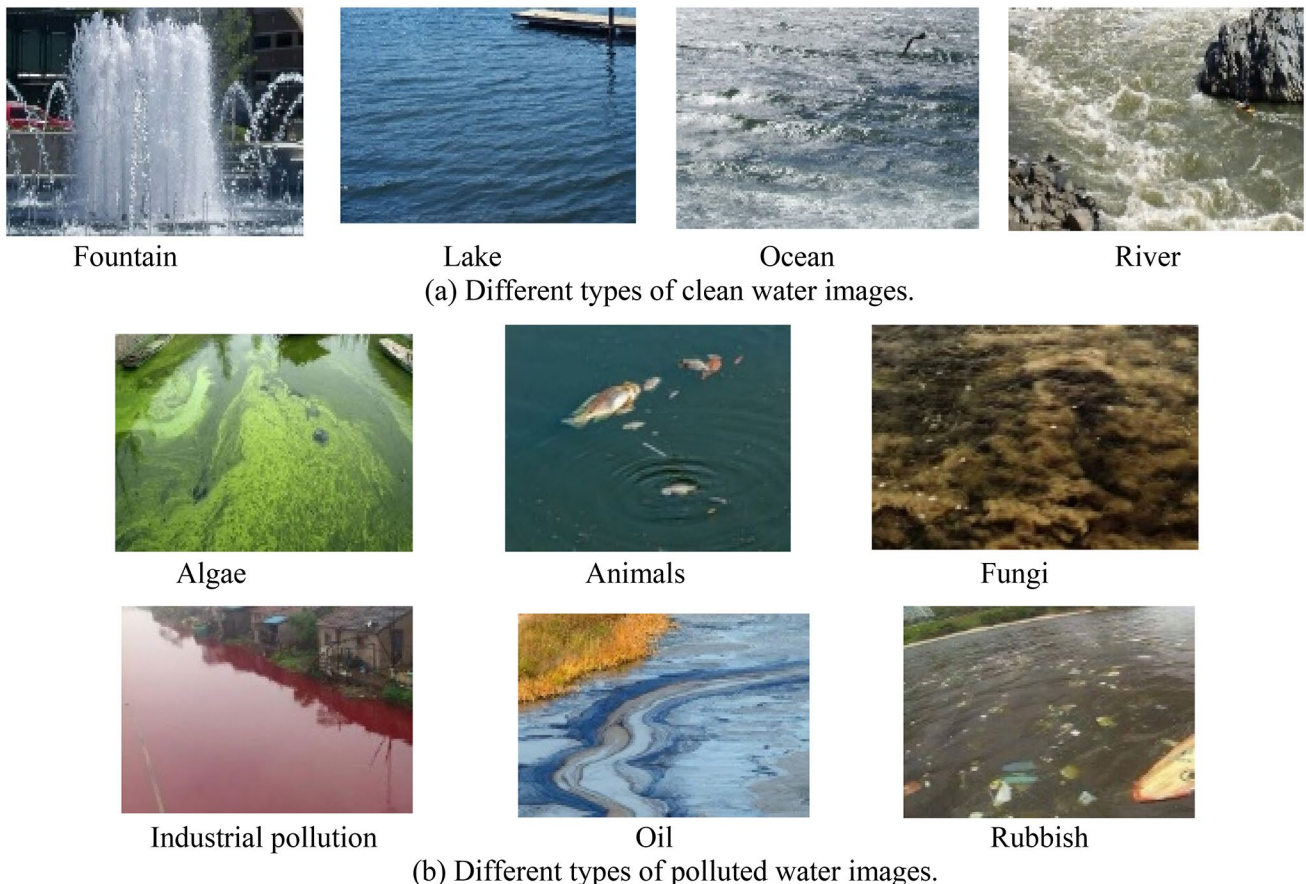


Fig. 1 Examples of different types of clean and polluted water images

to explore scale-invariant gradient orientation features to study the gradient information, and Gabor wavelet binary pattern to study the texture property in the images. In addition, to take advantage of deep learning and pixel values, we explore the VGG-16 model to extract features from the input image directly. The way the proposed approach integrates the merits of each concept to solve complex clean and polluted water image classification is the main contribution of the proposed work. To the best of our knowledge, this is the first work that integrates features as mentioned above for classifying different clean and polluted water images.

The key contributions of the present work are as follows. (1) Exploring color, gradient and Eigen information for smoothing different water type images. (2) Exploring gradient with a Gaussian first order derivative filter and the combination of Gabor with wavelet binary patterns for extracting texture features which are invariant to geometrical transformation from the smoothed images; this is new for classification of water images. (3) The way the proposed method combines the extracted features with deep learning is new for classification of different water type images.

The organization of the rest of the paper is as follows. The review of the existing methods on image scene classification and water image classification is presented in Sect. 2. Section 3 presents scale-invariant gradient orientation features, the Gabor wavelet binary pattern feature and features extracted using the VGG-16 model with gradient boost decision tree for classification. To validate the proposed method, Sect. 4 discusses experimental analysis of the proposed method and comparison with the existing methods. Conclusions and future work are described in Sect. 5.

2 Review of related work

We review the methods on general image classification and the methods on water image classification here. Liu et al. [11] proposed a method for scene classification based on ResNet and an augmentation approach. The method adapts multilayer feature fusion by taking advantage of inter-layer discriminating features. However, the scope of the method is to classify general scene images but not water images. Liu et al. [10] explored a deep learning kernel function for image classification. The main ideas of the method are to use sparse representation to design a deep learning network. In addition, the optimized kernel function is used to replace the classifier in the deep learning model, which improves the performance of the method. Li et al. [9] proposed deep multiple instance convolutional neural networks for learning robust scene representation.

The aim of the approach is to extract local information and spatial transformation for classification unlike most existing methods, which use global features. The method obtains patches with labels to train the proposed network to study local information in the images. Li et al. [8] proposed a method for image scene classification based on an error-tolerant deep learning approach. The method identifies correct labels of the data and it proposes an iterative procedure to correct the error caused by incorrect labels. To achieve this, the approach adapts multiple features of CNNs to correct the labels of uncertain samples. Nanni et al. [14] proposed a method for bio image classification based on neural networks. The approach combines multiple CNNs as a single network and it includes handcrafted features for training the network. The method shows that the combination of handcrafted features and deep features extracted by multiple CNNs is better than individual networks and features.

In summary, it is noted from the above methods that the approaches introduced deep learning models in different ways for learning and solving the classification problem. From the experimental results of the methods, one can infer that the performance of the methods depend on the number of samples with correct labels. Therefore, when a dataset does not have enough samples and it is hard to find relevant samples, the methods may not perform well. In the case of classifying polluted water type images, it is hard to predict the nature of the contamination. Therefore, the scope of the method is limited to scene images but not water type images.

Similarly, we review the methods developed for water image classification as follows. The methods proposed in the past use color, spatial and texture features for water image detection. Rankin and Matthies [21] proposed a method for water image detection using color features. Water body detection is undertaken by studying the combination of color features. Rankin et al. [20] proposed to use sky reflections for water image classification. The method estimates similarities between pixel values. The above two approaches perform well for detecting large water bodies but not small water bodies. Zhang et al. [29] proposed a flip invariant shape descriptor for water image detection. The method uses edge features to trace contours of reflections. Prasad et al. [18] proposed a method based on the use of quadcopters for stagnant water image detection. The method explores color and directional features for water image detection.

Santana et al. [22] proposed an approach for water image classification based on segmentation and texture features analysis. For extracting texture features, the method explores entropy. It exploits water flow and directional features to study the ripples. However, the performance of the method degrades for the polluted water

image type. Qi et al. [19] explore deep learning models for feature extraction and analyze texture features of water images. The main objective of the method is to classify scene images, and water images reconsidered a type of scene image for classification. The method requires a large number of labeled samples for training the proposed model. Mettes et al.'s approach (2017) explores spatio-temporal information for water image classification. However, the method expects clear object shapes in images for successful water image classification. In addition, the method is limited to video but not still images.

Zhuang et al. [33] proposed a method for water body extraction based on the tasseled cap transformation from remote sensing images. The method explores tasseled transformation and spectrum photometric methods for water body and non-water body classification. The method is good for two class classification but not classification of different water type images. Patel et al. [16] proposed a survey on river water pollution analysis using high-resolution satellite images. The work discusses the quality of the water images based on machine learning concepts. The methods discussed in this work are limited to the quality of the water images but not for classifying different water type images. Wang et al. [25] proposed water quality analysis for remote sensing images based on an inversion model. The method proposes spectral reflectance and water quality parameters for analyzing the quality of the images. The focus of the method is not to classify the water type of images, rather to analyze the quality of water images. In addition, the methods are developed for remote sensing images. Zhao et al. [31] proposed a discriminant deep belief network for high-resolution SAR image classification. The method explores deep learning model for learning features at a high level by combining ensemble learning with deep belief networks in an unsupervised way. However, the method was developed for images captured by synthetic aperture radar but not the images captured by normal cameras as in the proposed work. In addition, the method was considered to be computationally expensive.

In light of the above discussions, one can understand that most of the methods are confined to specific water type images and expected video information. Therefore, when we input different water images, including polluted water images and different clean water images, these methods may not perform well. Thus there is a need to propose a new approach that can cope with the challenges of both clean and polluted water images. Furthermore, the features, namely, color and texture, are good for images of clean water but not polluted water images, where unpredictable water surfaces are expected due to the presence of objects. However, recently, Wu et al. [26] proposed a method for the classification of clean and

polluted water images by exploring the Fourier transform [26]. The approach divides the Fourier spectrum into sub-regions to extract statistical features, such as mean and variances. The extracted features are passed to an SVM for the classification of water type images. It is noted from the experimental results that the method achieves better results for two classes and reports poor results for multi classes. This is because the proposed features are not sufficiently robust and inadequate to cope with the challenges of multiple classes. In contrast to this work, the proposed work considers 10 classes for classification and achieves better results. To overcome the limitations of the above-mentioned methods, Wu et al. [27] proposed a method for clean and polluted water image classification by exploring an attention neural network. The method extracts local and global features through a hierarchical attention neural network approach. The main limitation of this work is that if any one of the stages introduces an error, the subsequent stages fail to extract the expected information because in the case of the hierarchical approach, it ensures that each stage should deliver correct results. Otherwise, the hierarchical system does not work well. In addition, the method is too expensive and it lacks generalization ability.

Inspired by the method in [12], which stated that the HSV color space could mimic human color perceptions well, we explore the same observations for different situations in this work. It is evident from the literature that color features are considered prominent for water image detection. We noted that gradient direction is insensitive to poor quality and blur (Lee and Kim 2015), thus we propose to explore the dominant direction given by the gradient to generate Directional Coherence (DC) features based on Eigen value analysis for color components, which results in enhanced images. However, when an image is scaled up or down, the gradient may not give consistent features [30]. Therefore, we propose Gaussian first order derivative filters to obtain stable features for different scaled images, which are named as Scale Invariant Gradient Orientations (SIGO). Since the problem under consideration is complex, as it involves intra and inter-class variations, we further propose some features to investigate the texture of water images in a different way to strengthen the above-features. Inspired by the success of LBP and Gabor wavelets for texture description [3], we propose the combination of Gabor wavelets and LBP to extract texture features for DC images, which are called Gabor Wavelet Binary Patterns (GWBP). Finally, to utilize the strengths of deep learning, we use a VGG16 model for feature extraction [24]. As a result, the proposed method combines SIGO, GWBP and the features of VGG16 to obtain a single feature matrix, which is subjected to a Gradient Boosting Decision Tree (GBDT) for the classification of water images [7].

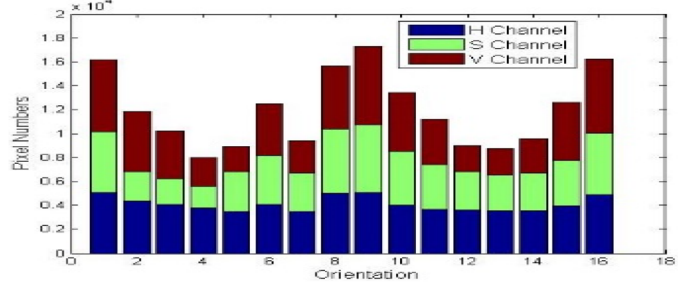
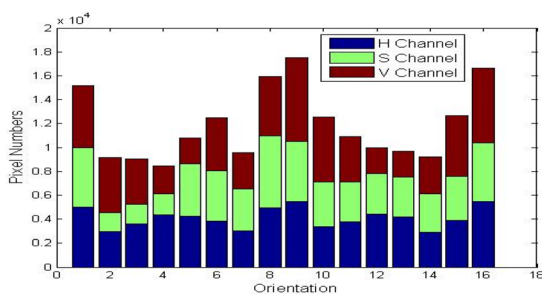
3 Proposed method

As discussed in the previous section, for each input image, the proposed method obtains HSV color components. Then color components are used for obtaining DC images based on Eigen value analysis and gradient distributions, which results in two Eigen images for each color component. This process outputs enhanced images as it combines the advantages of gradient and color information. We believe that the insights made based on observations from the images are as effective as a theoretical justification. The features are extracted based on observations and insights in this work. For instance, the brightness and gradient information are good features for representing different clean water images, whilst color and texture information are good for representing different polluted water images as shown in

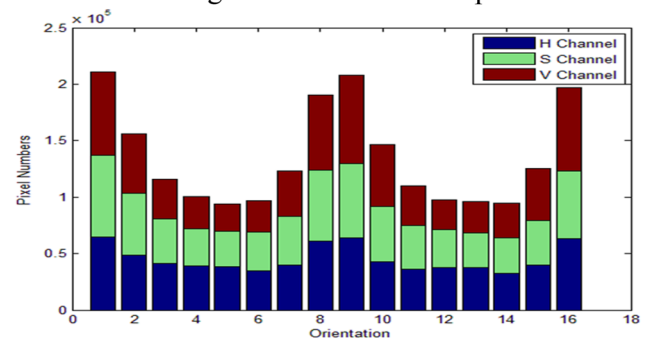
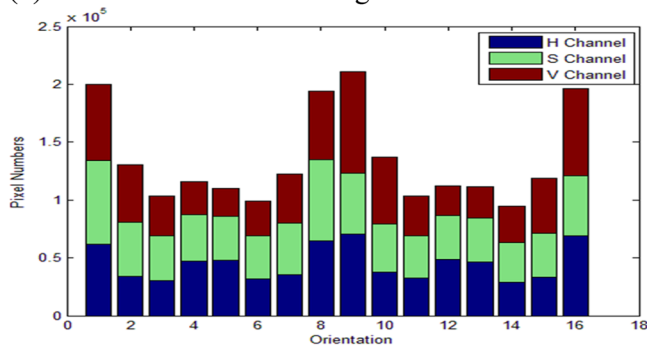
Fig. 2a. It is illustrated in Fig. 2, where it is noted that the behavior of the histograms of gradient with Gaussian filters is better than the behavior of the histograms of gradient without Gaussian filters in representing clean and polluted water images. For the sample images of clean and polluted water as shown in Fig. 2a, we perform histogram operations for the values of gradient orientations without Gaussian filters over HSV components by quantizing orientations into 16 bins as shown in Fig. 2b. At the same time, we perform histogram operations for the values of gradient orientations with Gaussian filters as shown in Fig. 2c. It is observed from Fig. 2b, c that the behaviors of the histograms in Fig. 2b appear almost the same, while the behaviors in Fig. 2c appear different. This is because the gradient helps us to enhance pixel values, while Gaussian filters remove noise created during gradient operations. This observation motivated us to explore the combination of gradient orientations and Gaussian



(a) Original images for clean water and polluted water.



(b) Gradient orientation histogram without Gaussian filters for water images of three-color components.



(c) Gradient orientation histogram with Gaussian filters for water images of three-color components.

Fig.2 Clues from gradient, Gaussian filters for extracting features

filters. With this notion, for each DC image, we propose SIGO based on different standard deviations of the derivatives of Gaussian filters for studying texture properties of water images. In the same way, we propose to explore the combination of Gabor-wavelets with binary patterns for DC images to study the texture properties of water images, namely, GWBP. In addition, the proposed method extracts features using VGG16 Deep Learning to take advantage of its inherent properties. Furthermore, the proposed method combines the features of SIGO, GWBP and VGG16, which generates the final feature matrix. The feature matrix is fed to a GBDT for the classification of water images. The reason to explore GBDT is that the GBDT is an efficient classifier, which does not require a large number of samples for training in contrast to deep learning models. In addition, GBDT has the ability to balance the features from imbalanced features through optimization.

The framework of the proposed method is shown in Fig. 3. In this work, if the input image contains too little pollution, the method may not perform well due to inadequate information for the proposed method. Therefore, the scope of the proposed work is limited to the images, which contain a certain amount of pollution to extract distinct features, as shown in the sample images in Fig. 2b.

3.1 Directional Coherence Images (DC) Detection

For each input image of clean and polluted water, as shown in Fig. 2a, the proposed method obtains color components, H, S, V as shown in Fig. 4a, where it is noted that the H, S and V of clean and stagnant water images appear differently. Specifically, the H of clean water images preserves fine details compared to that of

stagnant (polluted) water images, the S of clean water images loses brightness compared to that of polluted water images, while the V of clean water images lose sharpness compared to that of polluted water images. This shows that the above-mentioned color components provide clues for classifying different types of clean and polluted water images. In order to extract such observations, we define structure tensor [6], as in Eq. (1), for each patch p of the color components, which extracts the predominant direction of gradient in neighboring regions of a pixel. Besides, it summarizes the dominant direction and coherence of directions on the patch.

$$LST(p) = \begin{bmatrix} \sum_{i \in p} I_x^2(i) & \sum_{i \in p} I_x(i)I_y(i) \\ \sum_{i \in p} I_x(i)I_y(i) & \sum_{i \in p} I_y^2(i) \end{bmatrix} \tag{1}$$

where I_x, I_y denote the gradient of a pixel in the horizontal and vertical directions, respectively, and p denotes a patch of size 16×16 . Based on the above discussion, we define directional coherence as structure tensor coherence as defined in Eq. (2), where Eigen images, λ_1 and λ_2 are computed from Eigenvalue decomposition of a matrix $LST(p)$. The effect of Eigen value decomposition, λ_1 and λ_2 is illustrated in Fig. 4b, c respectively for clean and polluted water images. In Eq. (2), λ_1 and λ_2 denote the relative magnitude of the dominant orientation of the gradients in the patch and its perpendicular direction, respectively. It is noted from Fig. 4b, c that the dominant information is enhanced compared to the images in Fig. 4a.

$$DC(i) = \left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 \tag{2}$$

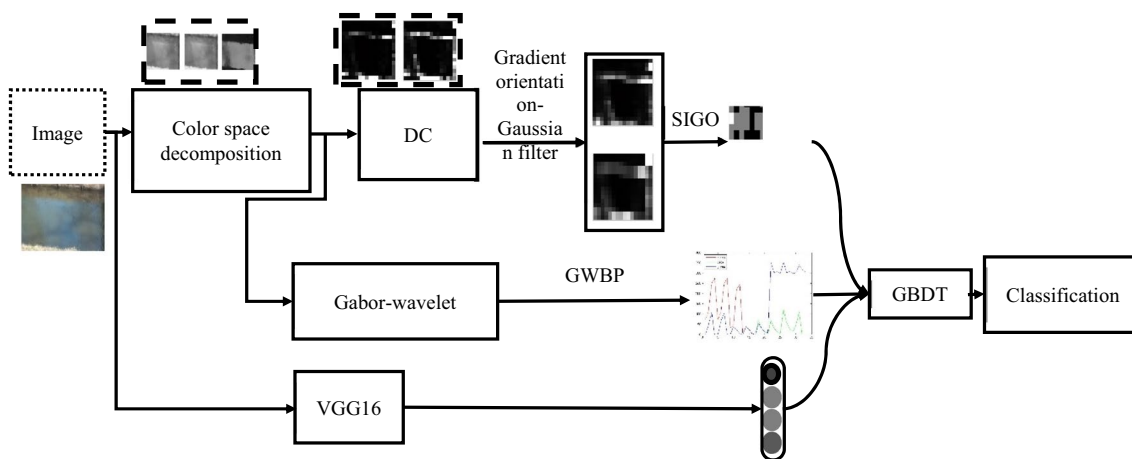


Fig.3 The framework of the proposed method

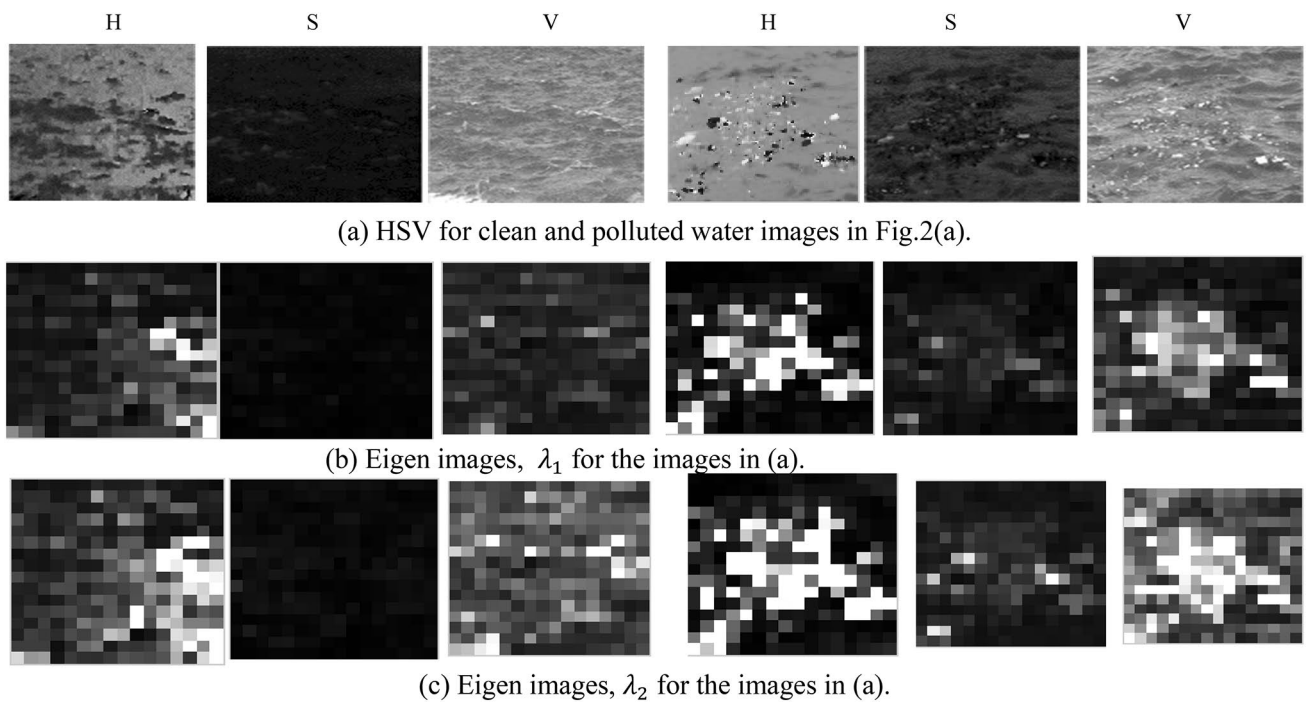


Fig.4 Directional coherence images for HSV based on Eigen value and gradient distribution

3.2 Scale invariant gradient orientation (SIGO) features

The gradient information is inconsistent for different scales due to the calculation of a partial differentiation for each pixel in the patch. Therefore, in this work, we propose gradient orientations given by the first order derivative along with different standard deviations of Gaussian filters. In other words, the proposed method uses the Gaussian first order derivative filters to calculate the derivatives and select the scale invariant gradient orientations in a new way. To achieve this, we quantize the value into 16 bins and perform histogram operations. The bin which gives the highest values in the histogram is considered as the stable scale-invariant orientation.

Specifically, the steps for obtaining SIGO features for the two DC images are as follows. For each patch of λ_1 and λ_2 images, we obtain the Gaussian first order derivative filters to calculate the derivatives as defined in Eq. (3) and Eq. (4):

$$G_x(\sigma_i) = \frac{1}{2\pi\sigma_i^2} e^{-\frac{x^2+y^2}{2\sigma_i^2}} \cdot -\frac{x}{\sigma_i^2} \tag{3}$$

$$G_y(\sigma_i) = \frac{1}{2\pi\sigma_i^2} e^{-\frac{x^2+y^2}{2\sigma_i^2}} \cdot -\frac{y}{\sigma_i^2} \tag{4}$$

where σ_i denotes the standard deviation of the Gaussian filter, $i \in \{1, 2, \dots, T\}$, and T is the number of standard deviations. With Eq. (3) and Eq. (4), $l_x(\sigma_i) = \lambda_1 * G_x(\sigma_i)$ and $l_y(\sigma_i) = \lambda_2 * G_y(\sigma_i)$ are calculated. Then, the gradient orientation for each patch of different standard deviations can be calculated as defined in Eq. (5) and Eq. (6):

$$\beta(\sigma_i) = \arctan 2 \left(\frac{l_x(\sigma_i)}{l_y(\sigma_i)} \right) + \pi \tag{5}$$

where,

$$\arctan 2 \left(\frac{l_x(\sigma_i)}{l_y(\sigma_i)} \right) = \begin{cases} \arctan \left(\frac{l_x(\sigma_i)}{l_y(\sigma_i)} \right), l_x(\sigma_i) > 0, l_y(\sigma_i) > 0 \\ \arctan \left(\frac{l_x(\sigma_i)}{l_y(\sigma_i)} \right) + \pi, l_x(\sigma_i) < 0, l_y(\sigma_i) > 0 \\ \arctan \left(\frac{l_x(\sigma_i)}{l_y(\sigma_i)} \right) - \pi, l_x(\sigma_i) < 0, l_y(\sigma_i) < 0 \\ \arctan \left(\frac{l_x(\sigma_i)}{l_y(\sigma_i)} \right), l_x(\sigma_i) > 0, l_y(\sigma_i) < 0 \end{cases} \tag{6}$$

For each patch p , the gradient orientation $\beta(p, \sigma)$ is divided into 16 states and a histogram is generated as defined in Eq. (7):

$$hist(\beta(p, \sigma)) = \{h_1, h_2, \dots, h_{16}\} \tag{7}$$

where h_d denotes the distribution of gradient orientations and h_d can be calculated as defined in Eq. (8):

$$h_d = \sum_{i=1}^T \delta(\beta(p, \sigma_i) == d) \tag{8}$$

where $\delta(x)$ is a function defined as Eq. (9):

$$\delta(x) = \begin{cases} 1, & x \text{ is true} \\ 0, & x \text{ is false} \end{cases} \tag{9}$$

Therefore, the stable scale-invariant gradient orientation for each patch is calculated by a histogram as defined in Eq. (10):

$$GO(p) = \underset{d \in \{1, \dots, 16\}}{\text{arg max}} h_d \tag{10}$$

Note: In this work, we set the standard deviation of the Gaussian filter $\sigma_i = 1.2^i$, where $i \in \{1, 2, \dots, T\}$. The value of 1.2 is determined empirically in this work.

3.3 Gabor wavelet binary patterns (GWBP) features

As discussed in the proposed method section, the features extracted in the previous section alone are not sufficient for achieving better results. Therefore, we propose a new combination for extracting texture features to strengthen the extracted features. The proposed method performs LBP for Gabor wavelet responses. To make LBP robust to noise, we propose to perform LBP over images filtered by Gabor wavelets. In other words, the proposed method first utilizes a Gabor wavelet filter bank to filter the input texture image at different resolutions and orientations. Then, the proposed method computes several binary patterns based on filter responses. It results in GWBP features.

The formal steps of the method are as follows. For each color component of the input image, the proposed method divides the image into different patches of the same size. Let GW_{s, μ_k} be the complex Gabor filter at scale s and orientation $\mu_k = k\pi/K$ in the spatial domain. Here, we empirically determine $K = 8$. Since the Gabor filter is complex, the real and the imaginary parts of GW_{s, μ_k} are denoted as GW_{s, μ_k}^r and GW_{s, μ_k}^i , respectively. For each pixel in patch p , multiplying pixel value $I(p)$ by each Gabor filter in a point-wise manner results in the response for patch p as defined in Eq. (11)-Eq. (13):

$$Res_p^{s, \mu_k, r} = I(p) \cdot GW_{s, \mu_k}^r \tag{11}$$

$$Res_p^{s, \mu_k, i} = I(p) \cdot GW_{s, \mu_k}^i \tag{12}$$

$$Res_p^{s, \mu_k} = \sqrt{(Res_p^{s, \mu_k, r})^2 + (Res_p^{s, \mu_k, i})^2} \tag{13}$$

The proposed method performs LBP operations for each pixel in each patch of the input image. Let $Res_j^s = [Res_j^{s, \mu_0}, Res_j^{s, \mu_1}, \dots, Res_j^{s, \mu_{K-1}}]$ be the vector of the magnitude of Gabor responses, and $Res^s = \frac{1}{P} \sum_{j=1}^P Res_j^s$ be the mean magnitude of Gabor responses of all the pixels in patch p , where P is the number of pixels in patch p . Then a rotation-invariant binary code α_j^s is computed for the pixel j as defined in Eq. (14):

$$\alpha_j^s = \max\{ROR(\alpha^{all}, k) | k = 0, \dots, K - 1\} \tag{14}$$

where

$$\alpha^{all} = \sum_{k=1}^K \text{sgn}(Res_j^s(k) - \overline{Res^s(k)}) 2^{k-1} \tag{15}$$

and $\text{sgn}(x)$ is a sign function as defined in Eq. (16):

$$\text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{16}$$

To consider the amount of the deviation of Gabor filter magnitude from the mean magnitude of Gabor responses, the proposed method defines the deviation as.

$Dev_j^s = Res_j^s - Res^s$ and $Dev^s = \frac{1}{P} \sum_{j=1}^P Dev_j^s$. Here, the binary code β_j^s is defined in Eq. (17):

$$\beta_j^s = \max\{\beta^{all}, k | k = 0, 1, \dots, K - 1\} \tag{17}$$

where

$$\beta^{all} = \sum_{k=1}^K \text{sgn}(Dev_j^s(k) - \overline{Dev^s(k)}) 2^{k-1} \tag{18}$$

Considering the differences of the real and imaginary parts of the Gabor filter, the proposed method defines the third Gabor responses as $Gdiff_j^s = Res_j^{s, r} - Res_j^{s, i}$,

where

$$Res_j^{s, r} = [Res_j^{s, \mu_0, r}, Res_j^{s, \mu_1, r}, \dots, Res_j^{s, \mu_{K-1}, r}] \text{ and}$$

$$Res_j^{s, i} = [Res_j^{s, \mu_0, i}, Res_j^{s, \mu_1, i}, \dots, Res_j^{s, \mu_{K-1}, i}]$$

Then the third binary code γ_j^s is computed as defined in Eq. (19):

$$\gamma_j^s = \max\{\gamma^{all}, k | k = 0, 1, \dots, K - 1\} \tag{19}$$

where

$$\gamma^{all} = \sum_{k=1}^K \text{sgn}(Gdiff_j^s(k) - \overline{Gdiff^s(k)}) 2^{k-1} \tag{20}$$

The above three binary codes are illustrated in Fig. 5, where we can see the clear discrimination for clean and

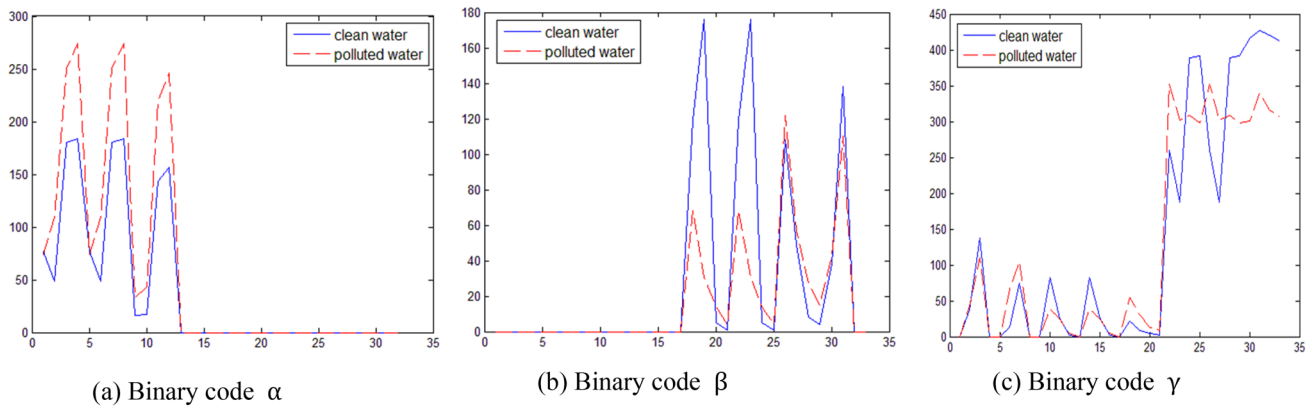


Fig.5 Histogram of three binary codes for clean and polluted water images

polluted water images. In Fig. 5, the feature values are divided into 8 bins on the X axis, and the frequencies of the binary code are on the Y axis. In this work, we consider 4 levels of Gabor wavelets at 8 orientations based on experimentation. Furthermore, the proposed method performs histogram operations for each binary code discussed in the above and concatenates all the three histograms, which results in GWBP.

Convolutional Neural Networks (CNNs) have been widely used for image classification and feature extraction [24]. One such CNN is VGG16, which consists of two parts, namely, feature extraction and a classifier. The feature extraction can translate an image into feature vectors. However, we use the former for feature extraction from input images. The complete architecture of the feature extractor can be seen in Fig. 6. In this set up, we use a trained VGG16 for initialization instead of training from the very beginning. For each input image, the VGG16 model extracts 1000 dimensions.

A deep convolutional neural network (VGG-16) is simple and efficient compared to other networks, such as ResNet. Since the main objective of the proposed work is to explore the combination of hand-crafted and deep features for achieving better classification results, we propose to use a simple architecture rather than a heavy/complex architecture such as ResNet. The reason is that collecting a large number of samples, which represent different polluted water images, is a difficult task. This is because the nature of polluted water images is unpredictable. Therefore, it is necessary to propose a method that can withstand the challenges caused by adverse effects of contamination. To overcome this problem, rather than developing end-to-end deep learning models, which depend heavily on a number of samples and labels, the proposed method focuses on feature extraction such that the method can work well with a smaller number of samples. This work uses a simple VGG-16 model for feature extraction but not for classification. In contrast, the heavy network like ResNet compared to VGG-16 requires a large number of

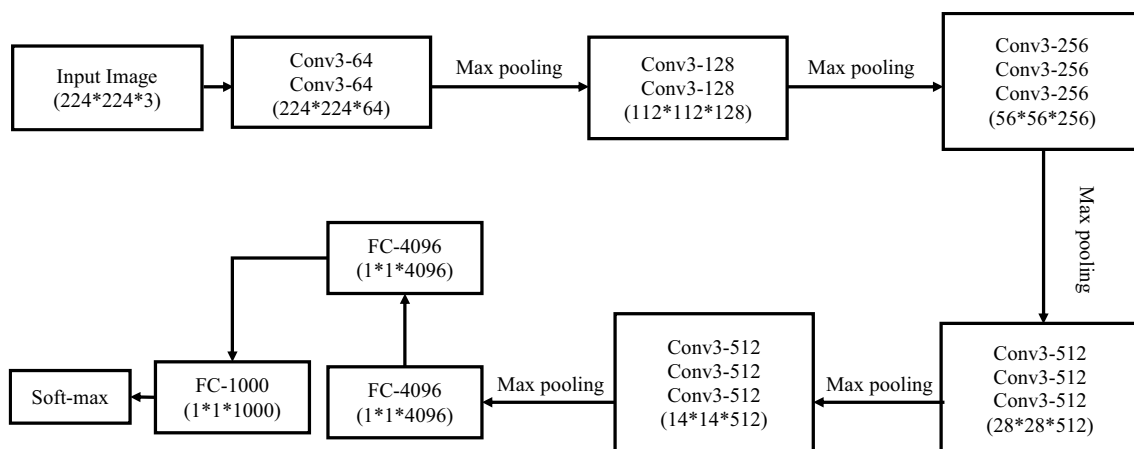


Fig.6 The architecture of the VGG16

labeled samples to obtain accurate results for complex problems. Otherwise, the network may not perform well due to the problem of overfitting.

3.4 Gradient boosting decision tree (GBDT) for water image classification

As discussed in the previous section, the dataset does not provide a large number of samples, and the variations in intra-class and inter-classes are unpredictable. Hence, there is a need to propose a method, which does not depend much on having a large number of samples to achieve better results. Since the proposed method extracts features that have the ability to differentiate different water images, a simple classifier is enough to extract such differences to obtain good classification results. Furthermore, it is also noted that GBDT is a well-known optimized and efficient classifier. It has the ability to balance the features when the extracted features are imbalanced. Therefore, in order to avoid the problems of overfitting with end-to-end deep learning models, the proposed work uses GBDT for classification in this work.

GBDTs are popular for achieving high accuracy, and are one of the effective methods of statistical learning in classification [7]. Therefore, we use GBDT for the classification of different types of water images in this work. Before applying GBDT, the proposed method uses augmentation for balancing training samples for unbalanced sub-classes. For the GBDT, the proposed method fuses the features extracted using SIGO, GWBP and VGG-16 as a single feature matrix, as shown in Fig. 7 where it can be seen that three different features are fused. For fusion, the proposed approach uses simple concatenation operations to obtain a single feature matrix. The GBDT model is illustrated as follows. It is the sum of the products of several basis functions with their weights as defined in Eq. (21):

$$f(x) = \sum_{n=1}^N W_n b(x; \theta_n) \tag{21}$$

where b is a basis function, and W is the weight of the basis function. The objective of the algorithm is to minimize the expected value of the loss function as defined in Eq. (22).

$$\min_{W_n, \theta_n} \sum_{i=1}^M L \left(y_i, \sum_{n=1}^N W_n b(x; \theta_n) \right) \tag{22}$$

where L is the loss function. It is not optimal to use N classifiers in parallel at the same time. Therefore, the proposed method uses one function with their coefficient sequentially. This process results in minimum values, gradually as defined in Eq. (23):

$$\min_{W_n, \theta_n} \sum_{i=1}^M L(y_i, f_{n-1} + W_n b(x; \theta_n)) \tag{23}$$

To obtain the minimum value of the loss function, we set the basis function as defined in Eq. (23):

$$W_n b(x; \theta_n) = -\lambda \frac{\partial L(y_i, f_{n-1})}{\partial f} \tag{24}$$

where λ means step size. When the process reaches n , we calculate the residual as defined in Eq. (25).

$$r_{i,n} = \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F=F_{n-1}}, i = 1, 2, \dots, m \tag{25}$$

Based on this, the proposed method fixes the n -th basis function with $(x_i, r_{i,n})$ by assuming the decision tree divides the input space into j spaces, namely, $R_{1n}, R_{2n}, \dots, R_{jn}$, and its output of each space is denoted by τ_{jn} . The n -th tree is defined as Eq. (26).

$$t_n(x) = \sum_{j=1}^J \tau_{jn} I(x \in R_{jn}) \tag{26}$$

Then, the proposed method searches for the best step size in each region of the decision tree linearly. The step size is combined with τ_{jm} as mentioned above, and the n -th objective function is defined as in Eq. (27):

$$F_n(x) = F_{n-1}(x) + \sum_{j=1}^J \alpha_{jn} I(x \in R_{jn}) \tag{27}$$

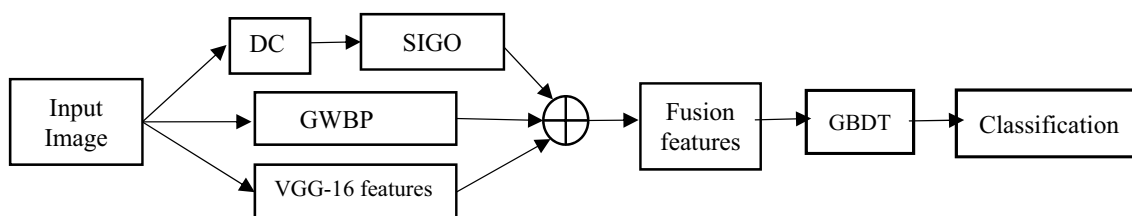


Fig. 7 Fusing features with the feature extracted by the VGG-16 model for classification with GBDT

where α_{jn} denotes the item, which combines the step size with τ_{jn} . Finally, the objective function is selected as the softmax function, as defined in Eq. (28):

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (28)$$

where k denotes the dimension of the data.

The distributions of features for two classes and multiple classes are shown in Fig. 8a, b, respectively. It is observed from Fig. 8a, b that there is a clear discrimination for two classes, but slightly poorer discrimination for multiple classes compared to two classes. Figure 8 shows that the X and Y axes indicate the first and second dimensions containing the largest variance of features, respectively. The proposed approach estimates a covariance matrix for the feature matrix using Principal Component Analysis. Furthermore, the proposed approach finds Eigen vectors corresponding to the maximum two Eigenvalues, and those values are multiplied by the feature matrix. This outputs the 2-dimensional space as illustrated in Fig. 8.

The parameters used in the GBDT classifier are `max_depth=5`, `objective='multi:softmax'`, `learning_rate=0.01`, `gamma=0.1`. The values of the parameters are determined empirically by conducting experiments on 500 random samples chosen across classes. Therefore, the values of the parameters do not have a significant effect on the overall performance of the proposed method.

4 Experimental results

Since there is no benchmark dataset for water images of different types, especially for polluted water images, we collected images from Google, Bing and Baidu, as well as our own resources. In addition, we also collected images

from [13], which provides clean water images of different types. The dataset includes images of different sizes or resolutions, and images captured from different height distances, images with complex backgrounds and poor quality, etc. In total, the dataset consists of 1000 images, which includes 500 for clean water image classes and 500 for polluted water image classes for evaluating the proposed and existing methods. The clean water image classes are Fountains, Lakes, Oceans and Rivers, as shown in Fig. 9a, where we can see water with different backgrounds. In case of ocean and lake images, one expects the presence of tides and waves to make visible differences as shown in Fig. 9a. Whereas the polluted water image classes include Algae, Animals (which may be alive or dead), Fungi, Industrial Pollution, Oils and Rubbish, as shown in Fig. 9b, where it can be seen the images are complex compared to those of clean water due to background and foreground variations. In total, 10 classes are considered for experimentation. The reason to choose more than 10 classes in this work is that as per our knowledge, those classes commonly arise in locations of interest, and present significant health risks to certain segments of society. Note that for all the experiments in this work, we consider 75% for training samples and 25% for testing. To support reproducibility of research, our dataset and code of the proposed method will be released publicly¹.

To evaluate the performance of the proposed and existing methods, we use standard measures, namely, Recall, Precision, F-measure as defined in Eqs. (29–31), respectively. The definitions for the above measure are as follows. True Positive (TP) is the number of images detected correctly in the positive class, True Negative (TN) is the number of images detected correctly in the negative class; False Positive (FP) and False Negative (FN) are the numbers of images detected incorrectly in positive and negative classes, respectively.

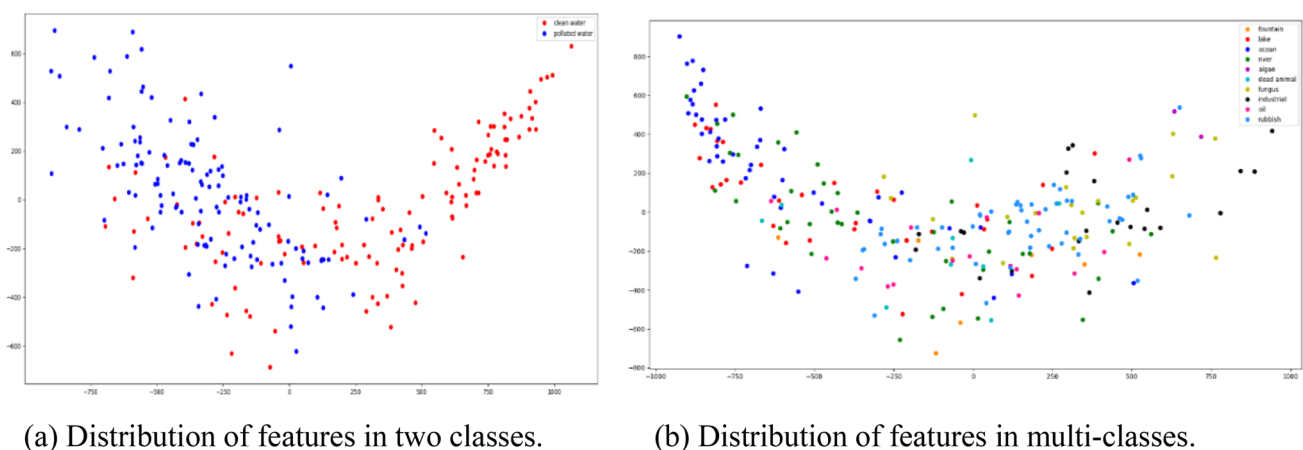


Fig. 8 Distribution of features after fusion in different classes

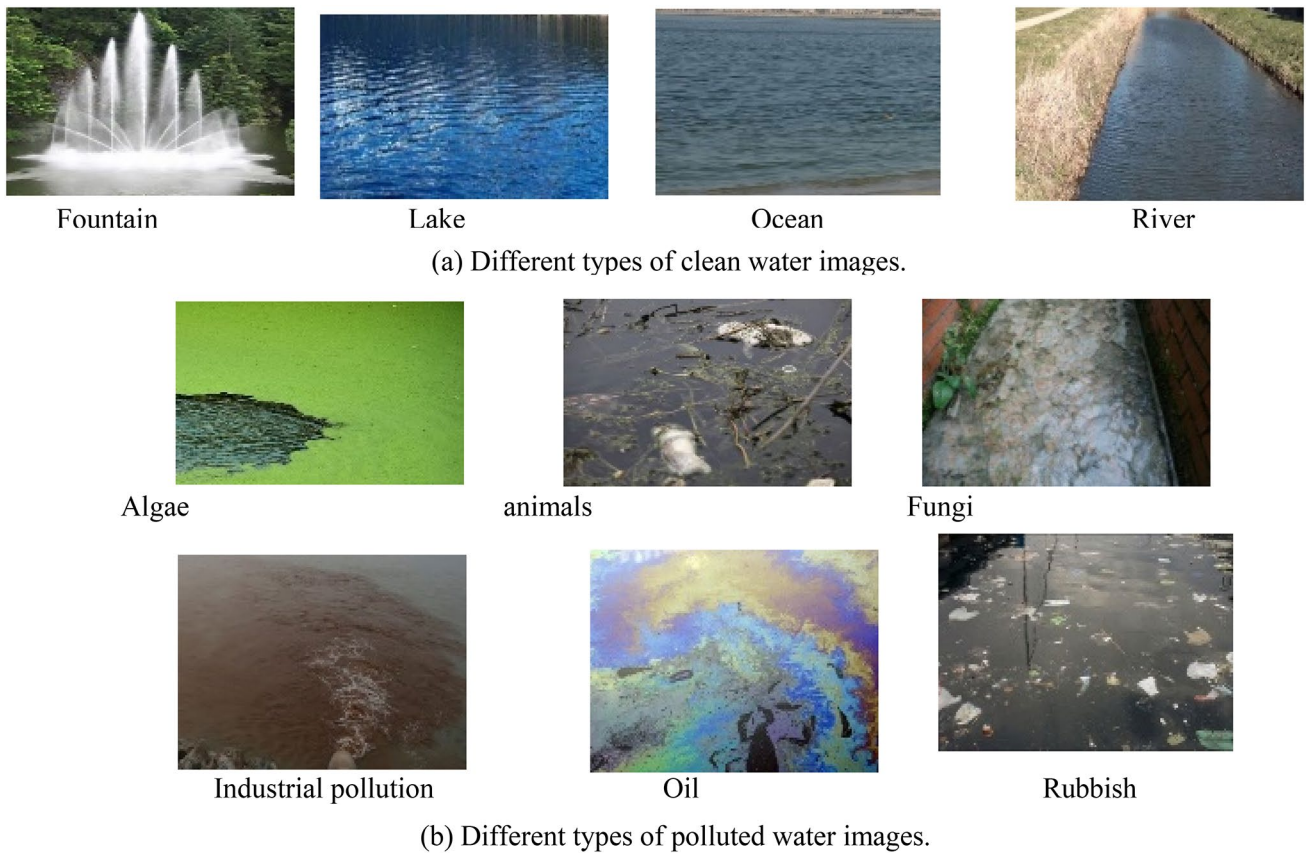


Fig. 9 Examples of different types of clean and polluted water images

$$Precision(P) = \frac{TP}{TP + FP} \tag{29}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{30}$$

$$F1 - score (F) = \frac{2 * Precision * Recall}{Precision + Recall} \tag{31}$$

There are a few methods for the classification of both clean and polluted water images. However, we choose relevant and state-of-the-art methods for a comparative study with the proposed method to demonstrate its effectiveness. Mettes et al. [13] proposed water detection through spatio-temporal invariant descriptors. The method focuses on video for clean water image classification by exploring motion properties of water. Qi et al. [19] proposed dynamic textures and scene classification by transferring deep image features. The method explores deep learning for feature extraction to detect water images. It is noted from the above two methods that their main objective is to detect clean water images but not polluted water images. In addition, the methods

are developed for video but not still images. However, for experimentation on our dataset, we considered each image as a key frame and created duplicate frames for the existing methods. The method [31] explores deep learning for extracting high level features for classification of images captured by radar containing water. The method [26] explores the Fourier spectrum for extracting features and it classifies water images from the polluted water images. The scope of the former method is limited to radar images and the latter method is limited to two classes. The reason to choose the method [31] is to show that the deep learning model developed for radar images may not work well for the images captured by a normal camera. Similarly, we selected the method due to [26] to demonstrate that the extracted features are not sufficient to achieve better results for multi-classes. We also implemented a method [27] which proposes an attention neural network for classification of clean and polluted water images. Since the objective of the method is the same as the proposed method, and to show that the deep neural network may not be sufficient to achieve consistent results for different experiments, the proposed method is compared with this method. Furthermore, to show that conventional features,

such as color histogram-based features do not have the ability to classify accurately, we extracted color histogram-based features as presented in [1] to undertake comparative study with the proposed method.

The proposed method requires approximately 6.18 min for training and 0.3 s for testing with the following system configuration. 2.4 GHz 24-core CPU, 62G RAM, no GPU device. However, it is noted that the processing time depends on several other factors also, such as coding, platform and programming, operating system etc. Since the scope of the proposed work is to classify water images, we do not focus on developing an efficient method.

4.1 Ablation study

The proposed method comprises three key steps, namely, Scale Invariant Gradient Orientation (SIGO) features, the Gabor Wavelet Binary Pattern (GWBP), and feature extraction using the VGG-16 model for the classification of clean and polluted water images. To validate the effectiveness of each step, we conducted experiments on clean, polluted water images and all the classes to compute the measures as reported in Table 1. In addition, to

test VGG-16 model against RestNet-50 when the dataset is small, we calculated the measures using only RestNet without hand-crafted features as reported in Table 1. Note that in this work, we use pre-trained VGG-16 and ResNet models for experimentation. The main reason is the lack of labeled samples, and the proposed method does not require a deep learning models. When we look at the average precision, recall and F-measure of all the classes over three experiments, the proposed method is the best at all the three measures compared to the other experiments. At the same time, the results of SIGO and GWBP are almost the same for 4, 6 and 10 class classification. This shows that both SIGO and GWBP are effective in achieving the best results by the proposed method. When we compare the result of ResNet and VGG-16, the results of VGG-16 are better than ResNet for the three experiments. Therefore, one can infer that for a small dataset, ResNet does not work well because of overfitting. On the other hand, the VGG-16 model reports better results than SIGO and GWBP. Therefore, the VGG-16 model is also effective in achieving the best result for the classification by the proposed method. In summary,

Table 1 Analyzing the effectiveness of the key steps and the proposed method for classifying 4, 6 and 10-class classification (Bold indicates the best results). Here P, R and F represent Precision, Recall and F-measure, respectively

Methods		SIGO			GWBP			ResNet50			VGG-16			Proposed		
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Clean water (4 classes)	Fountain	0.86	0.86	0.86	0.83	0.71	0.77	0.82	0.81	0.81	0.84	0.82	0.83	0.86	0.86	0.86
	Lake	0.59	0.52	0.55	0.61	0.58	0.59	0.6	0.62	0.61	0.61	0.63	0.62	0.62	0.64	0.63
	Ocean	0.69	0.74	0.71	0.65	0.7	0.67	0.71	0.7	0.7	0.7	0.7	0.7	0.71	0.74	0.72
	River	0.58	0.59	0.58	0.56	0.56	0.56	0.61	0.59	0.6	0.63	0.6	0.61	0.67	0.62	0.64
	Average	0.68	0.68	0.68	0.66	0.64	0.65	0.69	0.68	0.68	0.70	0.69	0.69	0.72	0.72	0.71
Polluted water (6 classes)	Algae	0.2	0.5	0.29	0.33	0.5	0.4	0.46	0.6	0.52	0.38	1.00	0.55	0.74	0.76	0.75
	Animal	0.75	0.33	0.46	0.75	0.33	0.46	0.71	0.42	0.51	0.75	0.41	0.53	0.62	0.64	0.63
	Funguses	0.62	0.36	0.46	0.53	0.36	0.43	0.64	0.4	0.5	0.69	0.41	0.51	0.71	0.74	0.72
	Industrial	0.46	0.73	0.56	0.47	0.73	0.57	0.5	0.87	0.63	0.52	0.9	0.65	0.67	0.62	0.64
	Oil	0.75	0.18	0.29	0.67	0.24	0.35	0.58	0.23	0.35	0.61	0.26	0.36	0.84	0.84	0.84
	Rubbish	0.69	0.83	0.75	0.75	0.89	0.81	0.76	0.8	0.78	0.78	0.82	0.8	0.62	0.64	0.63
	Average	0.58	0.49	0.47	0.58	0.51	0.50	0.61	0.55	0.55	0.62	0.63	0.57	0.70	0.71	0.70
Clean + Polluted (10 classes)	Fountain	0.86	0.86	0.86	0.63	0.71	0.67	0.7	0.9	0.78	0.72	0.92	0.81	0.78	1	0.88
	Lake	0.59	0.52	0.55	0.53	0.55	0.54	0.54	0.53	0.54	0.56	0.53	0.55	0.59	0.58	0.58
	Ocean	0.6	0.7	0.65	0.6	0.65	0.62	0.64	0.66	0.65	0.63	0.64	0.63	0.65	0.67	0.66
	River	0.49	0.56	0.52	0.54	0.56	0.55	0.5	0.55	0.53	0.52	0.56	0.54	0.5	0.46	0.48
	Algae	0.33	0.5	0.4	0.00	0.00	0.00	0.25	0.4	0.3	0.25	0.4	0.3	0.25	0.5	0.33
	Animal	0.75	0.33	0.46	0.75	0.33	0.46	0.71	0.4	0.51	0.74	0.43	0.53	0.8	0.44	0.57
	Funguses	0.56	0.23	0.33	0.56	0.45	0.5	0.52	0.41	0.46	0.57	0.44	0.49	0.59	0.45	0.51
	Industrial	0.43	0.59	0.5	0.43	0.59	0.5	0.41	0.61	0.48	0.41	0.66	0.53	0.47	0.73	0.57
	Oil	0.33	0.12	0.18	1.00	0.12	0.21	0.6	0.22	0.34	0.61	0.23	0.34	0.67	0.24	0.35
	Rubbish	0.69	0.83	0.75	0.74	0.79	0.76	0.72	0.78	0.75	0.76	0.8	0.78	0.78	0.85	0.81
Average	0.56	0.52	0.52	0.58	0.48	0.48	0.56	0.55	0.53	0.58	0.56	0.55	0.61	0.59	0.57	

the steps used in the proposed method are effective and contribute equally for achieving the best results.

4.2 Experiments on two-class classification

Sample results of the proposed method for clean and polluted water image detection are shown in Fig. 10a,

b, respectively, where it can be seen that the proposed method classifies images with different backgrounds, successfully.

Quantitative results of the proposed and existing methods are reported in Table 2, where it is noted that the proposed method is the best at F-measure compared to existing methods. When we compare the results of the existing methods [13, 19, 26, 27, 31] and Color based features of

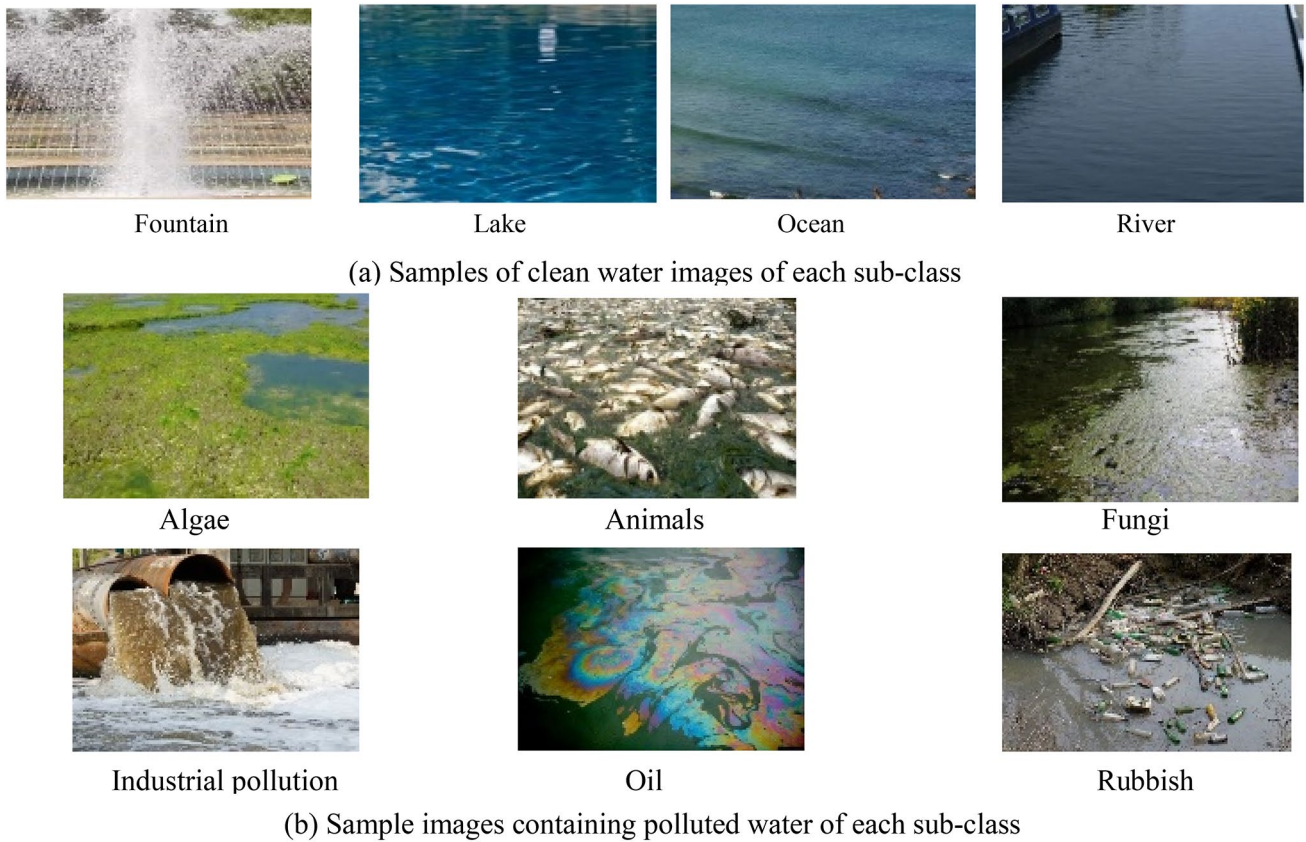


Fig. 10 The proposed method classifies clean and polluted water images successfully

Table 2 Performance of the proposed and existing methods for clean and polluted water image classification (two-class classification)

Method		Proposed	Mettes et al. [12]	Qi et al. [18]	Wu et al. [25]	Zhao et al. [30]	Color Histogram Features	Wu et al. [26]
Clean water	P	0.95	0.62	0.88	0.91	0.87	0.28	0.96
	R	0.98	0.59	0.93	0.89	0.84	0.35	0.97
	F	0.96	0.60	0.90	0.90	0.85	0.31	0.96
Polluted water	P	0.98	0.61	0.92	0.88	0.90	0.35	0.96
	R	0.94	0.63	0.87	0.91	0.88	0.32	0.95
	F	0.96	0.62	0.89	0.89	0.89	0.33	0.95

(Bold indicates the best results)

Here P, R and F represent Precision, Recall and F-measure, respectively

[1], the method [27] is better than all other existing methods. This is because of the advantage of the attention-based deep network model, which combines both local and global information in the images for classification, while most of the existing methods extract global information for classification. However, the results of Wu et al. [27] are lower than the proposed method. This is because of the combination of hand-crafted features and deep features, which do not depend heavily on the number of samples unlike Wu et al.'s [27] method. It is observed from Table 2 that the color-based features and the method in [13] report poor results compared to the proposed and other existing methods. The main reason is that the methods extract conventional features, which may not be as robust as those features extracted by deep learning models. Although other existing methods use deep learning models for classification, the methods report poor results compared to the proposed method. This is because of the inherent limitations of the existing methods. In addition, the models are not robust for obtaining good results on small datasets. On the other hand, the proposed method involves hand-crafted features (which are invariant to rotation, scaling) and deep learning-based features, which enhances the robustness and generalization ability. Hence the proposed method is best for classification compared to the existing methods.

4.3 Evaluation on multi-class classification

To test the effectiveness of the proposed method on multi-class classification, we conducted experiments on multiple classes of clean, polluted water images and together 10 classes of water images. Quantitative results of the proposed and existing methods for 4 classes of clean water, 6 classes of polluted water and 10 classes of both clean and polluted water images are reported in Table 3. It is observed from Table 3 that the proposed method is the best at average precision, recall and F-measure for 4 and 6 classes while it is the best at average recall for 10-class classification. As mentioned in the previous section, the method due to Wu et al. [27] outperforms all other existing methods. The reason is that the method is developed for clean and polluted water image classification, as in the case of the proposed method. However, this method does not consider the advantages of hand-crafted features for classification, and hence it reports poor results compared to the proposed method especially for 4- and 6-class classification. For 10-class classification, Wu et al. [27] reports almost the same results as the proposed method. This shows that the complex deep network proposed in Wu et al. [27] is effective when the dataset has a large number of samples for training. Since the deep network proposed in Wu et al. [27] is complex compared to the VGG-16 model

used in our method, it is considered too computationally expensive. Therefore, one can conclude that the proposed method is accurate as well as effective for classifying 4, 6 and 10 classes compared to the existing methods. The reason for the poor results of the existing methods is the same as discussed in the previous section.

To show the proposed method is invariant to rotation, scaling and is to some extent robust to noise and blur, which are common causes introduced by open environments in real-time situations, we conducted experiments on 10 classes to compute the measures as reported in Table 4. In this experiment, the Gaussian noise (mean 0 and the variance varies from 0.01 to 0.1), blur (kernel of size 5×5 and the value sigma varies from 1 to 5) at different levels is added to the input images. In addition, the images in the dataset are scaled up and down and rotated randomly to validate the invariance property of the proposed features. It is noted from the average precision, recall and F-measure reported in Table 4 that for the images affected by different scenarios, the proposed method reports poor results compared to the unaffected images. However, for the rotated and scaled images, the results are better than the images affected by noise, blur and the results are almost the same as the results of unaffected images. This shows that the proposed features have the ability to handle different rotation and scaling of the images. For noise and blurred images, the proposed method reports poor result compared to normal images. This is because the features proposed in the method are sensitive to noise and blur, which is a limitation of the proposed work and it is beyond the scope of this research. Thus, there is a scope for improvement in the near future.

However, sometimes, the proposed method misclassifies images such as those shown in Fig. 11a, b, where it is seen that when images include water with other content, such as objects, the proposed method misclassifies clean water images as polluted water ones and vice versa. This is understandable because it is hard to define shapes of background objects in water images since they are unpredictable.

5 Conclusions and future work

In this work, we have proposed a new method for classifying water images of clean and polluted water. The proposed method explores Eigen value analysis and gradient distributions for enhancing fine details in the images. The contributions of the proposed method can be concluded as follows. (i) The proposed method adapts the concepts of SIGO, GWBP and VGG16 for feature extraction in new way to solve the complex problem of multiple-class classification of clean and polluted water images. (ii) The way

Table 3 Performance of the proposed and existing methods for multi-class classification

Classes	Methods	Proposed			Mettes et al. [12]			Qi et al. [18]			Wu et al. [25]			Zhao et al. [30]			Color Histogram Features			Wu et al. [26]		
		P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Clean water (4 classes)	Fountain	0.86	0.14	0.24	0	0	0	0.71	0.09	0.16	0.69	0.10	0.18	0.69	0.09	0.16	0	0	0	0.85	0.85	0.85
	Lake	0.65	0.45	0.53	0.09	0.07	0.09	0.50	0.34	0.40	0.48	0.33	0.39	0.53	0.35	0.42	0.07	0.07	0.07	0.66	0.65	0.65
	Ocean	0.76	0.55	0.64	0.34	0.27	0.34	0.69	0.46	0.55	0.68	0.43	0.52	0.70	0.43	0.53	0.24	0.2	0.22	0.65	0.55	0.6
	River	0.62	0.50	0.55	0.32	0.24	0.32	0.56	0.40	0.47	0.57	0.39	0.48	0.55	0.42	0.48	0.25	0.25	0.25	0.68	0.61	0.64
Polluted water (6 classes)	Average	0.72	0.72	0.71	0.19	0.15	0.19	0.19	0.15	0.19	0.62	0.32	0.40	0.61	0.31	0.39	0.14	0.13	0.14	0.71	0.67	0.69
	Algae	1.00	0.05	0.11	0	0	0	0.50	0.02	0.04	0.52	0.03	0.6	0.19	0.04	0.07	0	0	0	0.4	1.00	0.57
	Animal	0.44	0.10	0.16	0	0	0	0.44	0.10	0.16	0.47	0.09	0.15	0.45	0.10	0.16	0	0	0	0.78	0.45	0.57
	Fungi	0.45	0.26	0.33	0	0	0	0.45	0.26	0.33	0.44	0.27	0.33	0.43	0.29	0.35	0	0	0	0.75	0.66	0.64
Clean + Polluted (10 classes)	Industrial Pollution	0.91	0.34	0.50	0	0	0	0.73	0.31	0.44	0.77	0.32	0.45	0.75	0.32	0.45	0	0	0	0.75	0.72	0.73
	Oil	0.29	0.15	0.20	0	0	0	0.35	0.17	0.23	0.35	0.18	0.24	0.34	0.19	0.24	0	0	0	0.51	0.34	0.4
	Rubbish	0.83	0.52	0.64	0.98	0.42	0.59	0.89	0.57	0.69	0.85	0.57	0.70	0.88	0.58	0.68	0.75	0.23	0.38	0.86	0.8	0.83
	Average	0.70	0.71	0.70	0.16	0.07	0.10	0.16	0.07	0.10	0.51	0.25	0.30	0.51	0.25	0.30	0.13	0.04	0.06	0.68	0.66	0.62
Clean + Polluted (10 classes)	Fountain	1.00	0.07	0.15	0	0	0	0.71	0.05	0.09	0.72	0.04	0.9	0.70	0.05	0.08	0	0	0	0.75	0.75	0.75
	Lake	0.59	0.20	0.30	0	0	0	0.50	0.17	0.25	0.55	0.16	0.26	0.51	0.17	0.25	0	0	0	0.58	0.63	0.61
	Ocean	0.69	0.28	0.40	0.31	0.08	0.13	0.67	0.26	0.37	0.66	0.27	0.40	0.66	0.29	0.38	0.07	0.06	0.06	0.6	0.52	0.56
	River	0.46	0.20	0.28	0.15	0.03	0.05	0.59	0.21	0.31	0.59	0.22	0.32	0.60	0.22	0.32	0.05	0.04	0.05	0.58	0.61	0.59
	Algae	0.50	0.01	0.02	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0.4	0.7	0.5
	Animal	0.44	0.04	0.07	0	0	0	0.44	0.04	0.07	0.42	0.04	0.06	0.45	0.03	0.06	0	0	0	0.73	0.41	0.52
	Fungi	0.45	0.11	0.18	0	0	0	0.36	0.08	0.13	0.37	0.09	0.13	0.35	0.08	0.13	0	0	0	0.61	0.6	0.6
	Industrial Pollution	0.73	0.15	0.25	0	0	0	0.55	0.12	0.20	0.59	0.13	0.20	0.57	0.12	0.20	0	0	0	0.55	0.71	0.62
	Oil	0.24	0.05	0.08	0	0	0	0.35	0.06	0.10	0.37	0.05	0.9	0.35	0.06	0.10	0	0	0	0.51	0.3	0.38
	Rubbish	0.85	0.34	0.49	0.57	0.14	0.22	0.83	0.32	0.46	0.83	0.31	0.45	0.81	0.30	0.44	0.41	0.07	0.12	0.76	0.83	0.79
Average	0.61	0.59	0.57	0.10	0.03	0.04	0.10	0.03	0.04	0.64	0.13	0.20	0.64	0.13	0.20	0.05	0.02	0.02	0.61	0.61	0.59	

(Bold indicates the best results)

Here P, R and F represent Precision, Recall and F-measure, respectively

Table 4 The performance of the proposed method for the images affected by noise, blur, scaling and rotation

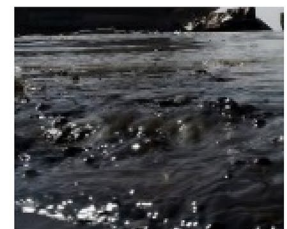
Classes	Methods	Proposed			Different scaled and rotated images			Gaussian blur			Gaussian white noise		
		P	R	F	P	R	F	P	R	F	P	R	F
Clean + Polluted (10 classes)	Fountain	0.78	1	0.88	0.72	0.85	0.78	0.69	0.7	0.69	0.68	0.72	0.70
	Lake	0.59	0.58	0.58	0.56	0.55	0.55	0.5	0.51	0.50	0.51	0.5	0.50
	Ocean	0.65	0.67	0.66	0.62	0.63	0.62	0.57	0.59	0.58	0.58	0.6	0.59
	River	0.5	0.46	0.48	0.49	0.45	0.47	0.45	0.4	0.42	0.44	0.41	0.42
	Algae	0.25	0.5	0.33	0.25	0.48	0.33	0.22	0.43	0.29	0.21	0.43	0.28
	Animal	0.8	0.44	0.57	0.81	0.46	0.59	0.72	0.43	0.54	0.73	0.41	0.53
	Fungj	0.59	0.45	0.51	0.53	0.42	0.47	0.5	0.41	0.45	0.5	0.39	0.44
	Industrial Pollution	0.47	0.73	0.57	0.45	0.71	0.55	0.42	0.68	0.52	0.41	0.67	0.51
	Oil	0.67	0.24	0.35	0.65	0.22	0.33	0.55	0.21	0.30	0.57	0.2	0.30
	Rubbish	0.78	0.85	0.81	0.75	0.83	0.79	0.72	0.84	0.78	0.73	0.85	0.79
average	0.61	0.59	0.58	0.58	0.56	0.55	0.53	0.52	0.51	0.54	0.52	0.51	
Average		0.60	0.59	0.57	0.58	0.56	0.54	0.53	0.52	0.50	0.53	0.51	0.50

(Bold indicates the best results)

Here P, R and F represent Precision, Recall and F-measure, respectively



(a) Examples for misclassification of clean water images as polluted water images



(b) Examples of misclassifying polluted water images as clean water images

Fig.11 The limitations of the proposed method

the proposed work integrates the features extracted from the above concepts with the help of GBDT for classification is something new compared to the existing work. (iii) The

combination of hand-crafted features and deep features are better than deep learning-based methods alone. (iv) Experimental results show that the proposed method is

better than existing methods for both two and multiple class classification. (v) It is also noted from experimental results that the proposed method is robust to rotation and scaling and to some extent to noise and blur at different levels. (vi) Sometimes, when the images share the extracted features, the proposed method misclassifies the images as shown in the sample misclassification results in the experimental section. (vii) When we observed the classification results on multiple-class classification, the results are not very high. In order to improve the results, we need to further investigate robust deep learning architectures with different features.

As mentioned in the Introduction Section, real time applications, such as monitoring stagnant water is also one of the key applications and hence the proposed work can be considered as a reference or base work for investigating new ideas, such as tackling the challenges caused by drone images at different angles and height distances in different weather conditions. This is very challenging because when angle and height distance changes, the complexity increases in terms of quality, contrast, size, resolution and distortion. Developing a method to cope with such challenges requires a new direction for researchers. In order to support the reproducibility of the research, we have a plan to release the dataset publicly along with the code after acceptance. The link can be found on Page 23 at footnote.

Acknowledgements The work is supported by the Natural Science Foundation of China under Grant No. 61672273, the program B for Outstanding PhD candidate of Nanjing University No.202002B101, and the Faculty Grant: GPF014D-2019, University of Malaya, Malaysia.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Alnihoud J (2012) Content-based image retrieval system based on self organizing map, fuzzy color histogram and subtractive fuzzy clustering. *Int Arab J Inf Technol* 9(5):452–458
- Galvis AT, Gancarski P, Equille LB (2018) Remote sensing image analysis by aggregation of segmentation-classification collaborative agents. *Pattern Recogn* 73:259–274
- Hadizadeh H (2015) Multi-resolution local Gabor wavelets binary patterns for gray-scale texture description. *Pattern Recogn Lett* 65:163–169
- Kar S, Rathore VS, Ray PKC, Sharma R, Swain SK (2016) Classification of river water pollution using Hyperion data. *J Hydrol* 537:221–233
- Lagrange M, Andrieu H, Emmanuel I, Busquets G, Loubrie S (2018) Classification of rainfall radar images using the scattering transform. *J Hydrol* 556:972–979
- Lee H, Kim C (2014) Blurred image region detection and segmentation. In: *Proceedings ICIP*, pp 4427–4431
- Li L, Wu Y, Ye M (2014) Multi class image classification based on fast stochastic gradient boosting. *Informatics* 38:145–153
- Li Z, Xu K, Xie J, Bi Q, Qin K (2020) Deep multiple instance convolutional neural networks for learning robust scene representations. *IEEE Trans Geosci Remote Sens* 58(5):3685–3702
- Li Y, Zhang Y, Zhu Z (2020) Error-tolerant deep learning for remote sensing image scene classification. *IEEE Trans, Cybernetics*
- Liu J, An FP (2020) Image classification algorithm based on deep learning-kernel function. *Sci Programming* 3:1–14
- Liu S, Tian G, Xu Y (2019) A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. *Neurocomputing* 338:191–206
- Liu GH, Yang JY, Li ZY (2015) Content based image retrieval using computational visual attention model. *Pattern Recogn* 48:2554–2566
- Mettes P, Tan RT, Veltakamp RC (2017) Water detection through spatio-temporal invariant descriptors. *Comput Vis Image Underst* 154:182–191
- Nanni L, Ghidoni S, Brahmam S (2020) Ensemble of convolutional neural networks for biomaage classification. *Appl Comput Inf*. <https://doi.org/10.1016/j.aci.2018.06.002>
- Nguyen KA, Stewart RA, Zhang H (2014) An autonomous and intelligent expert system for residential water end-use classification. *Expert Syst Appl* 41:342–356
- Patel HJ, Dabhi VK, Prajapati HB (2019) River water pollution analysis using high resolution satellite images: A survey. In: *Proceedings ICACCS*, pp 520–525
- Peng YT, Cosman PC (2017) Underwater image restoration based on image blurriness and light absorption. *IEEE Trans Image Process* 26:1579–1594
- Prasad MG, Chakraborty A, Chalasani R, Chandran S (2015) Quadcopter based stagnant water identification. In: *Proceedings NCVPRIPG*, pp 1–5
- Qi X, Li CG, Zhao G, Hong X, Pietikainen M (2016) Dynamic texture and scene classification by transferring deep image features. *Neurocomputing* 171:1230–1241
- Rankin AL, Mathies LH, Bellutta P (2011) Daytime water detection based on sky reflections. In: *Proceedings ICRA*, pp 5329–5336
- Rankin A, Matthies L (2010) Daytime water detection based on color variation. In: *Proceedings ICIRS*, pp 215–221
- Santana P, Mendonca R, Barata J (2012) Water detection with segmentation guided dynamic texture recognition. In: *Proceedings ICRB*, pp 1836–1841
- Shi C, Pun CM (2018) Superpixel-based 3D deep neural networks for hyperspectral image classification. *Pattern Recogn* 74:600–616
- Wang Y, Song B, Zhang P, Xin N, Cao G (2017) A fast feature fusion algorithm in image classification for cyber physical systems. *IEEE Access* 5:9089–9098
- Wang J, Zhang J, Wang X (2018) Water quality analysis of remote sensing images based on inversion model. In: *Proceedings IGARSS*, pp 4861–4864
- Wu X, Shivakumara P, Zhu L, Zhang H, Shi J, Lu T, Pal U, Blumenstein M (2018) Fourier transform based features for clean and polluted water image classification. In: *Proceedings ICPR*, pp 1707–1712
- Wu Y, Zhang X, Xiao Y, Feng J (2020) Attention neural network for water image classification under IoT environment. *Appl Sci* 10(3):909
- Yang L, Liu J, Tang X (2015) Depth from water reflection. *IEEE Trans Image Process* 24:1235–1243

29. Zhang H, Guo X, Cao X (2010) Water reflection detection using a flip invariant shape detector. In: Proceedings ICPR, pp 633–636
30. Zhang J, Liang J, Zhang C, Zhao H (2014) Scale invariant texture representation based on frequency decomposition and gradient orientation. *Pattern Recogn Lett* 51:57–62
31. Zhao Z, Jiao L, Zhao J, Gu J, Zhao J (2017) Discriminant deep belief network for high-resolution SAR image classification. *Pattern Recogn* 61:686–701
32. Zhong SH, Liu Y, Liu Y, Li CS (2013) Water reflection recognition based on motion blur invariant moments in curvelet space. *IEEE Trans Image Process* 22(11):4301–4313
33. Zhuang Y, Chen C (2018) A method for water body extraction based on the tasseled cap transformation from remote sensing images. In Proc, IWEORSA

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.