Signal Processing: Image Communication 131 (2025) 117225

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Vocal cord anomaly detection based on Local Fine-Grained Contour Features

Yuqi Fan [a,b], Han Ye [a,b], Xiaohui Yuan [c,*]

[a] School of Computer Science and Information Engineering, Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei University of Technology, Hefei, 230601, Anhui, China

[b] China and Anhui Provincial Key Laboratory of Network and Information Security, Anhui Normal University, Wuhu, 241002, Anhui, China

[c] Department of Computer Science and Engineering, the University of North Texas, 3940 N. Elm, Denton, 76207, TX, USA

ARTICLE INFO

Keywords:
Deep learning
Vocal cord disease
Laryngoscopic images
Inner contour

ABSTRACT

Laryngoscopy is a popular examination for vocal cord disease diagnosis. The conventional screening of laryngoscopic images is labor-intensive and depends heavily on the experience of the medical specialists. Automatic detection of vocal cord diseases from laryngoscopic images is highly sought to assist regular image reading. In laryngoscopic images, the symptoms of vocal cord diseases are concentrated in the inner vocal cord contour, which is often characterized as vegetation and small protuberances. The existing classification methods pay little, if any, attention to the role of vocal cord contour in the diagnosis of vocal cord diseases and fail to effectively capture the fine-grained features. In this paper, we propose a novel Local Fine-grained Contour Feature extraction method for vocal cord anomaly detection. Our proposed method consists of four stages: image segmentation to obtain the overall vocal cord contour, inner vocal cord contour isolation to obtain the inner contour curve by comparing the changes of adjacent pixel values, extraction of the latent feature in the inner vocal cord contour by taking the tangent inclination angle of each point on the contour as the latent feature, and the classification module. Our experimental results demonstrate that the proposed method improves the detection performance of vocal cord anomaly and achieves an accuracy of 97.21%.

1. Introduction

Vocal cord diseases directly affect the voice and even the ability to breathe, which may lead to dyspnea, cancer, and other serious consequences. Early diagnosis of vocal cord diseases is important for timely treatment. Electronic laryngoscopic image analysis has been developed for the detection and diagnosis of vocal cord diseases. Manually processing and analyzing laryngoscopic images is labor-intensive and depends heavily on the experience of medical specialists. A large number of patients challenge conventional human reading to meet the demand [1]. There is an urgent need for computer-aided intelligent vocal cord disease detection.

Computer-aided detection of vocal cord diseases was addressed with traditional machine learning methods, such as Naive Bayes, multilayer perceptron (MLP), k-nearest neighbors (KNN), support vector machine (SVM), random forest (RF) algorithms, etc [2,3]. Deep learning methods have been applied to many medical image classification problems, including vocal cord disease detection. These models employ deep networks such as deep convolutional neural network (DCNN) [4,5], Xception [5], ResNet [6], Inception [6], MobileNet [6], etc. Image segmentation methods, such as SegNet, U-Net, ENet, ErfNet, are used for laryngoscopic image processing [7], which improves the diagnosis of vocal cord diseases. However, these methods often fail to capture the subtle features of the vocal cord, which should play an important role in the vocal cord disease diagnosis.

laryngoscopic images have two unique properties compared to other medical images: (1) the proportion of the vocal cord is small in the image, and the non-vocal-cord area has rich textures and colors; (2) the symptoms of vocal cord diseases are mostly in the inner vocal cord contour, which are often characterized as vegetations and small protuberances. Fig. 1 shows examples of laryngoscopic images and the vocal cord areas are highlighted with a yellow polygon. Fig. 1(a) shows an example of the vocal cords of a healthy person, which are smooth, flat, and symmetrical. Fig. 1(b) shows an example of the vocal cords of a patient with a polyp, with vegetation in the inner vocal cord contour. Patients with leukoplakia of vocal cords have proliferative protuberances on the inner curves of vocal cords, as shown in Fig. 1(c). Patients with vocal nodules often show small protrusions on both sides of the inner contour of the vocal cords, as shown in Fig. 1(d).

The normal vocal cords are smooth, and the abnormal ones have protrusions in the inner side of the vocal cords. The subtle differences in the inner vocal cord regions play a critical role in vocal cord

* Corresponding author.
E-mail address: xiaohui.yuan@unt.edu (X. Yuan).

https://doi.org/10.1016/j.image.2024.117225
Received 17 September 2023; Received in revised form 15 August 2024; Accepted 27 September 2024
Available online 2 November 2024

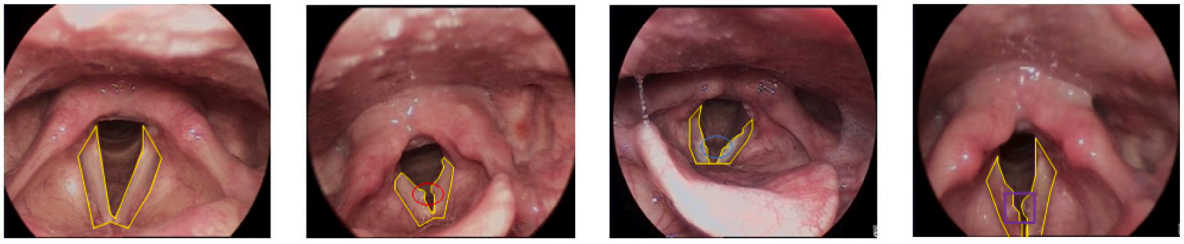

Fig. 1. Examples of laryngoscopic images. The vocal cord is highlighted with yellow polygons and the diseased regions are enclosed with polygons of different colors.

anomaly detection, although there are great differences in the non-vocal cord areas, as shown in Fig. 1. Unfortunately, the existing deep learning methods fail to highlight the local fine-grained features in the inner vocal cord region, when processing laryngoscopic images. The performance of the existing models on laryngoscopic image classification needs to be improved. This paper investigates the problem of vocal cord disease detection from laryngoscopic images and proposes a novel Local Fine-grained Contour Feature (LFCF) extraction-based laryngoscopic image classification method. The main contributions include the local fine-grained vocal cord contour feature extraction, which isolates the small part important for anomaly detection from the image and further obtains the feature of the isolated part. Our proposed method consists of a vocal cord segmentation module, a vocal cord inner contour extraction module, a potential inner vocal cord contour feature extraction module, and a classification module. We used the laryngoscopic images of the collected data set for evaluation, which is the most comprehensive to our best knowledge.

The rest of the paper is organized as follows. Section 2 reviews the related work of image classification, medical image classification, and laryngoscopic image processing. Section 3 describes the proposed method LFCF. Section 4 discusses the experimental results. Section 5 concludes this paper with a summary and introduces future research directions.

2. Related work

Early diagnosis of vocal cord diseases is crucial for timely treatment. Detecting vocal cord anomalies has been studied for improved performance. Verikas et al. [3] developed a computer-aided diagnostic system for vocal cord diseases. The system analyzes vocal cord images to extract color, texture, and geometric features and uses KNN to classify laryngeal lesions. Turkmen et al. [2] proposed a vocal cord disorder classification system based on binary decision trees. The method classifies images using the shape and vascular structural features of vocal cord lesions. The accuracy of these methods ranges from the lower to the mid-80%. There is plenty of room for improvement to make the computer-aided approach practical.

Deep learning is widely used in image classification [8]. Simonyan et al. [9] proposed VGG-Net, which uses multiple layers with 3 by 3 convolution kernels and 2 by 2 pooling kernels to improve classification performance by continuously deepening the network structure. He et al. [10] introduced ResNet to solve the problem of information loss due to the increasing number of network layers. The input directly bypasses the output, and the whole network only needs to learn the difference between the input and the output to simplify the learning objective and difficulty. Wu et al. [11] proposed a spatial recursive model for visual recognition. This model is based on bilinear pooling and supports local pairwise feature interaction between the outputs of two different convolutional neural networks (CNNs), enabling localization, extraction, and spatial encoding of relevant features. The learned deep features help distinguish different subclasses. Hu et al. [12] introduced a compression and excitation module to take advantage of the relationship between channels. In this module, global average pooling is used to collect features for channel relationships, and two fully connected layers are used to capture the relationship between channels. Woo et al. [13] proposed an attention module for a feedforward convolutional neural network, which sequentially captures the attention map in two dimensions of channel and space, and then multiplies the final attention map with the input feature map to adaptively refine the features. Wang et al. [14] proposed a channel attention module (ECA) to capture cross-channel correlation. The module has few parameters to reduce the model complexity and avoid paying little attention to learning channels due to the large number of dimensions.

To incorporate features into the medical image classification task, Yang et al. [15] proposed a depth tree training strategy (DTT), which applies multiple classifiers to the features in the DTT training process and automatically selects the optimal integration scheme from the branch classifiers based on the precision and diversity criteria, and further integrates them by using the weighting strategy, to obtain the optimal ensemble classifier. Sasikanth et al. [16] used the ANFIS classifier based on the optimal feature level fusion to classify brain magnetic resonance imaging (MRI) images. Ko et al. [17] used a CNN network for skin disease diagnosis and achieved the same results as test experts. Acharya et al. [18] extracted features by Gabor transform on cardiovascular disease images and used the featureless reduction method and probabilistic neural network for cardiovascular disease diagnosis. Chen et al. [19] used Inception V2 to identify the existence and severity of retinopathy in preterm infants. The network includes two sub-networks. The first one is used to extract high-level features from fundus images, and the second one is used for classification. Souza et al. [20] performed the initial segmentation of chest X-ray (CXR) based on pixel classification and then refined the initial segmentation based on the reconstruction steps of ResNet18 to learn the patterns of lung and non-lung patches in CXR. Antony et al. [21] presented the investigations and results of feature learning using convolutional neural networks to automatically assess knee osteoarthritis (OA) severity and the associated clinical and diagnostic features of knee OA from X-ray images and demonstrated that feature learning in a supervised manner was more effective than using conventional handcrafted features for automatic detection of knee joints and fine-grained knee OA image classification. Dwarikanath et al. [15] proposed an interpretable-driven selection framework to extract depth features from significance diagrams through an end-to-end depth learning method, which is trained to identify samples with a large amount of information by self-monitoring. Fan et al. [22] proposed model MKSC, which captures the spatial and channel attention information in parallel, for COVID-19 detection based on lung X-ray images. Gao et al. [23] reviewed the application of deep learning in the early detection of esophageal cancer, focusing on its advantages and disadvantages. Finally, the paper puts forward some future suggestions for developing real-time, clinically realizable, interpretable, and robust diagnostic support systems.

The successful application of deep learning in the diagnosis of various diseases has prompted scholars to use deep learning to process laryngoscope medical images for the detection and diagnosis of vocal cord diseases. Matava et al. [6] used ResNet, Inception and MobileNet

to classify, identify, and mark the vocal cords and trachea. Xiong et al. [4] verified the feasibility of applying DCNN in the diagnosis of laryngeal cancer. Cho et al. [5] verified the performance of CNN6, VGG16, Inception V3, and Xception in vocal cord disease diagnosis based on laryngoscopic image analysis. Won et al. [24] compared the performance of EifficientNet, Inception V3, MobileNetV2, and VGG16 in laryngeal disease classification. Laves et al. [7] compared the performance of SegNet, U-Net, ENet, and ErfNet semantic segmentation networks in laryngeal endoscopic image segmentation. Yin et al. [25] proposed a laryngeal disease classification method, which uses an attention mechanism to obtain the critical area under the supervision of image labels for laryngeal disease classification. A CNN model is trained to classify the laryngeal images. If the classification result is correct, the region with a strong response is likely to be a critical region. The regions with strong responses are used as training data to train an object localization model that can automatically locate the critical area. The located critical area is employed for image classification. Paderno et al. [26] applied the full convolutional neural network (FCNN) of video analysis to the field of head and neck oncology, with the main purpose of testing the semantic segmentation of oral and oropharyngeal squamous cell carcinoma based on the FCNN method. Azam et al. [27] used a new application of real-time detection of laryngeal squamous cell carcinoma (LSCC) based on artificial intelligence (LSCC) deep learning convolutional neural network and narrowband imaging video laryngoscope. LSCC video frames are extracted for training, verification, and testing of various YOLO models. Ren et al. [28] developed a computer-aided diagnosis system based on deep learning to distinguish laryngeal tumors (benign, precancerous lesions, and cancer), and improve the accuracy of diagnosis and evaluation based on clinical laryngoscopy results.

Image edge detection can promote precise contour location and segmentation, which further improves vocal disease detection performance [29]. Yuan [30] proposed a learning method to improve the segmentation of blurry objects from medical imagery. Wang et al. [31] proposed an interactive segmentation method based on deep learning to improve the results obtained by CNN and reduce user interaction in the refinement process. They used CNN to obtain an initial segmentation and added user interaction to indicate the wrong segmentation. They applied another CNN with the interaction between the user and the initial segmentation as the input and obtained the refined result. Chen et al. [32] proposed DeepLabv3+, which uses multiple filters and pooling operations to detect input features to encode multi-scale context information, and gradually recover spatial information for object boundary detection. Yu et al. [33] proposed a method for edge detection of cloth pattern cutting based on the holistically nested edge detection method, to extract the high-precision cloth pattern contour and clearly distinguish the pattern main body of the pattern from the background. The edge refinement and smoothing process are added, where the edge detection, edge refinement, and edge smoothing of the clothes images are carried out in sequences. Lin et al. [34] developed a deep learning model based on 3D U-Net, which is used to segment the kidney and renal mass and detect the renal mass in the corticomedullary phase of computed tomography urography (CTU). Wang et al. [35] proposed an unsupervised feature learning module to realize multi-scale feature extraction. Xie et al. [36] proposed a general framework called Segmentation–Emendation–reSegmentation–Verification (SESV) to improve the accuracy of medical image segmentation. For the location where the segmentation error is easy to occur, the error map with the image and segmentation mask is concatenated as the input of the re-segmentation network. A verification network is then introduced to determine whether to accept the refined mask produced by the re-segmentation network on a region-by-region basis. Chen et al. [37] proposed a Destruction and Construction Learning method. Destruction learning increases the recognition difficulty to guide the network to learn expert knowledge for fine-grained recognition. Construction learning can model the semantic correlation between various parts of the object. Zhuang et al. [38] integrated boundary information into a deep network to achieve a rich description of the fine details for semantic segmentation.

The existing studies show that deep learning has achieved much-improved performance. However, the characteristics of laryngoscopic images bring challenges to the application of deep learning in vocal cord disease diagnosis. The vocal cord region accounts for a small proportion of the whole laryngoscope image, whereas the non-vocal cord region has distinct texture and color properties. The symptoms of vocal cord diseases appear mostly in the middle and lower parts of the inner vocal cord, which is often depicted as vegetation and small protuberances. The existing image classification models and vocal cord disease classification methods often fail to capture the fine-grained inner vocal cord contour features. which should play an important role in the vocal cord disease diagnosis. In this paper, we use deep learning to analyze laryngoscopic images for vocal cord anomaly detection and propose a novel vocal cord disease diagnosis method based on the fine-grained features of the local contour.

3. Local fine-grained contour features for anomaly detection

The overall structure of our method is shown in Fig. 2. Our method extracts the fine-grained inner vocal cord contour features, which are fed into a classifier for laryngoscopic image classification. It consists of four stages:

Stage 1 Obtain the overall contour of the vocal cord based on DeepLabv3+ to focus on the vocal cord region in the laryngoscopic image.
Stage 2 Extract the inner contour of the vocal cord based on the image gradients.
Stage 3 Extract the latent feature based on the tangent inclination angle of each point on the inner vocal cord contour to capture the fine-grained features such as vegetation and masses.
Stage 4 Classify the laryngoscopic image using the latent feature.

3.1. Vocal cord segmentation and inner contour extraction

We extract the contour of the vocal cord in the laryngoscopic image by employing the DeepLabv3+ network [32], which removes the parts irrelevant to the vocal cord. The encoder–decoder architecture is used for image segmentation. The atrous convolution in the encoder increases the receptive field and makes the output of each convolution capture the key information, which leads to clear object boundaries. The dilated convolution with different dilation rates of 6, 12, and 18 and average pooling are performed after 1×1 convolution. The obtained 5 features are concatenated and fused with 1×1 convolution, and the result is input into the decoder. The decoder performs $4\times$ upsampling, which result is concatenated with the output of 1×1 convolution on the result of the atrous convolution. Our method performs a 3×3 convolution and then $4\times$ upsampling on the concatenated result to obtain the segmentation result.

Figs. 3 (a) and (c) depict laryngoscopic images, and Figs. 3 (b) and (d) show their corresponding segmentation results that consist of the vocal cord area (in black) and the non-vocal cord area (in white). The inner vocal cord contours of patients are not smooth as shown in Figs. 3 (a) and (b). The contour of the vocal cord includes the outlines of the inner and outer vocal cords. As diseases happen on the inner surface of the vocal cord, it is necessary to isolate it from the segmentation result. We propose an Inner Vocal Cord Contour extraction algorithm (IVCC) based on the change of adjacent pixel values. Algorithm 1 gives the detailed steps. The algorithm takes the segmented vocal cord to generate the inner contour curve.

Algorithm 1 scans each row in the overall vocal cord contour and records the number of change points in each row to extract the inner vocal cord contour. In a binary image, a change in pixel value indicates

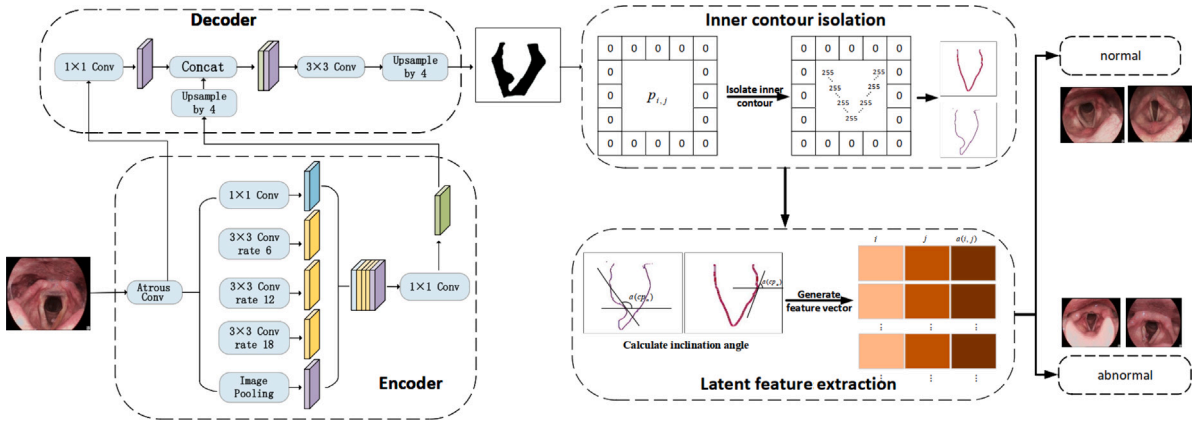

Fig. 2. Vocal cord anomaly detection based on the fine-grained feature of the local vocal cord contour.

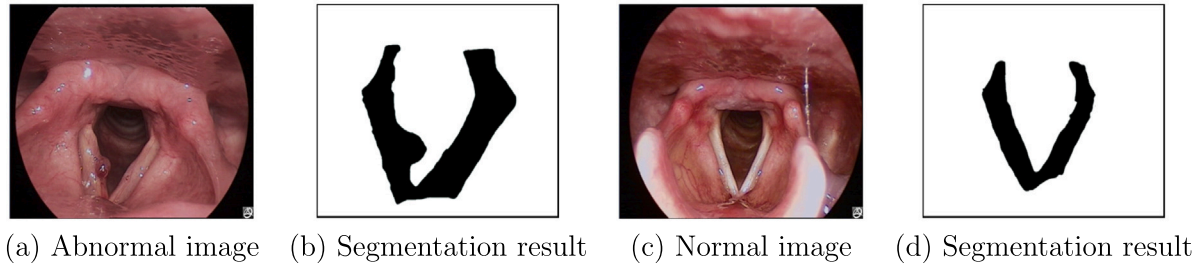

Fig. 3. (a) and (b) Abnormal laryngoscopic image and segmentation result. (c) and (d) Normal laryngoscopic image and segmentation result.

Algorithm 1 Inner Vocal Cord Contour Extraction

Input: A segmented image I of the vocal cord.
Output: Inner contour of the vocal cord I'

Initialize a zero matrix I' of the same size of I
for each row i in I **do**
  Initialize $K \leftarrow 0$, $k \leftarrow 0$, $j \leftarrow 2$
  while $j \leq J$ **do**
    if $P_{i,j} + P_{i,j-1} = 255$ **then**
      $K \leftarrow K + 1$
      $k \leftarrow k + 1$
      $Row(M_{i,k}) \leftarrow i$
      $Col(M_{i,k}) \leftarrow j$
    end if
  end while
  if $K = 2$ **then**
    $I'[Row(M_{i,2})][Col(M_{i,2})] \leftarrow 255$
    $I'[Row(M_{i,3})][Col(M_{i,3})] \leftarrow 255$
  end if
  if $K > 4$ **then**
    for each $1 < k < K$ **do**
      $I'[Row(M_{i,k})][Col(M_{i,k})] \leftarrow 255$
    end for
  end if
end for
return I'.

that the pixel is on the vocal cord contour. Each row could have more than one change point, as shown in Fig. 4. If the number of change points is four, the second and third change points are on the inner vocal cord contour. The cases where the number of change points is less than four happen at the top or the bottom of the vocal cords. In such cases, the change points are not recorded. In the case of greater than four change points, the ones other than the left-most and right-most are recorded as inner vocal cord contours.

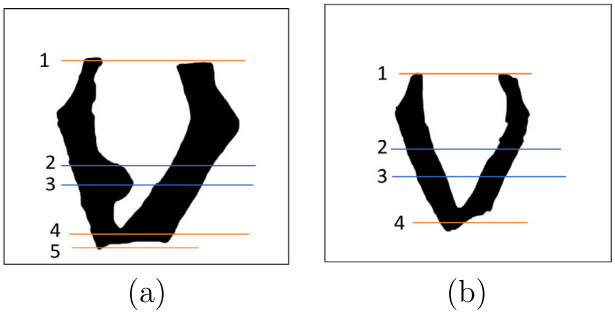

Fig. 4. Pixel value mutation points.

Fig. 5 depicts two cases of inner contour extractions. The inner contour of a normal person is mostly smooth and the tangent inclination angle of each point on the curve changes slowly. In contrast, parts of the inner contour of the vocal cords with a nodule exhibit a bump. The tangent inclination angles at the points on the bump change greatly, as shown in Fig. 5(a). Our inner contour extraction method correctly preserves the geometric features of the inner vocal cord contour.

3.2. Local contour feature extraction

It can be observed from Fig. 5(b) that the inner contour curve of a normal person is smooth and hence the tangent inclination angle of each point on the curve changes slowly. In contrast, for the patient suffering from a vocal cord disease, there are bulges in the middle and lower parts of the inner contour curve, since there are vegetations, lumps, etc. on the inner contour of the vocal cord in the laryngoscopic

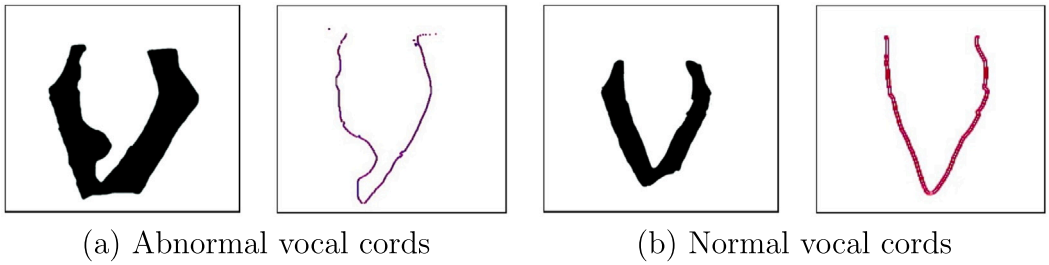

Fig. 5. Segmentation of vocal cord and inner contour (a) and normal vocal cord and inner contour (b).

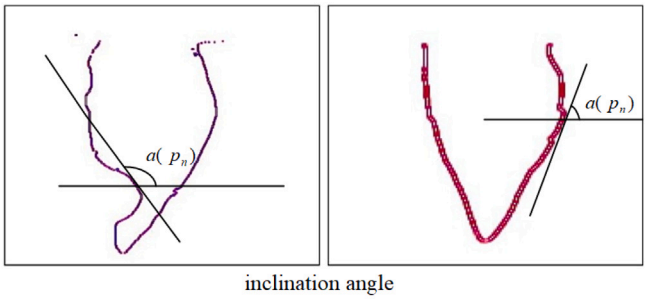

Fig. 6. Extraction of latent inner vocal cord contour feature.

images. Accordingly, the tangent inclination angles at the points on the bulges change greatly, as shown in Fig. 5(a).

We propose an Inner Vocal cord Contour Feature extraction method IVCF to calculate the tangent inclination angle of each point on the inner contour curve, as shown in Algorithm 2, and take the angles as the fine-grained feature of a laryngoscopic image.

Let $P = \{p_1, p_2, \ldots, p_N\}$ denote the set of points on the inner vocal cord contour, where N is the number of extracted pixels on the contour of the vocal cords. N is obtained by counting the number of non-zero points in the image. Different images have different vocal cord shapes, and the corresponding number of inner contour points is hence different. The slope $t(p_i)$ at point p_i on the inner vocal cord contour is computed as follows:

$$t(p_i) = \frac{1}{Col(p_i) - Col(p_{n-1})}. \tag{1}$$

$Col(p_i)$ represents the index of the column of point p_i, which is the abscissa value of the point. The tangent inclination angle (as shown in Fig. 6) $a(p_i)$ at point p_i is

$$a(p_i) = \arctan(t(p_i)). \tag{2}$$

Hence, we have a point-wise feature vector that consists of the coordinates of the point (denoted as (x, y)) and the tangent inclination angle

$$A_{p_i} = [x, y, a(p_i)] \tag{3}$$

The local contour features of a vocal cord consist of an array of point-wise feature vectors, which are used in classification. Algorithm 2 summarizes our method of feature extraction.

3.3. Classification module

In this module, we feed the local contour feature of the laryngoscopic image into a classifier to identify whether the image is abnormal. Various classifiers can be used, such as VGG, MLP, and ResNet. The classifier can be decided empirically in applications.

Algorithm 2 Contour feature extraction

Input: Inner vocal cord contour C.
Output: Feature vectors of the inner vocal cord contour.

```
for each point p_i in C do
    if i − 1 > 0 then
        Compute slop t(p_i) using Eq. (1)
        Compute tangent inclination angle a(p_i) using Eq. (2)
        Get the coordinates of point p_i: x ← Row(p_i) and y ← Col(p_i)
        A_{p_i} ← [x, y, a(p_i)]
    end if
end for
return A
```

4. Experimental results

4.1. Data and settings

Our data set includes 870 different images, among which 540 are images of vocal cord diseases and 330 are images of healthy vocal cords. All the images are labeled by doctors. We crop each image to 224 by 224 pixels, which include the vocal cord area. Ten-fold cross-validation is used to evaluate the performance of the proposed method. Each fold consists of 87 randomly selected laryngoscopic images, including 33 normal laryngoscopic images and 54 abnormal laryngoscopic images. We use 80% of the dataset for training, 10% of the dataset for validation, and the remaining 10% for testing. We repeat our experiments five times on each fold and report the average of the 50 test results. The performance metrics used in our evaluation include accuracy, sensitivity, and specificity.

Using a DeepLabv3+, our method obtains the contour of the vocal cord. We evaluate the parameter settings of DeepLabv3+ to the performance of our method, including batch size and learning rate, which determine the number of examples used in each training iteration and the speed of updating weights, respectively. Table 1 reports the accuracy of using different batch sizes and learning rates of DeepLabv3+. The best accuracy is achieved when the batch size and learning rate are at 32 and 0.001, respectively. Hence, such settings are used in the rest of our experiments.

4.2. Inner vocal cord contour extraction and classification

We evaluate the vocal cord segmentation and inner contour extraction. Fig. 7 shows four cases, where the laryngoscopic images (top row) are processed with DeepLabv3+ to segment the vocal cords. The inner vocal cord contours are delineated using our proposed method, as shown in the bottom row. The inner vocal cords of images of patients show the outline of the bumps. This contour extraction method allows us to reduce or remove the interference of the arytenoids and epiglottis, i.e., the structures that surround the glottis and can be seen during a laryngoscopy.

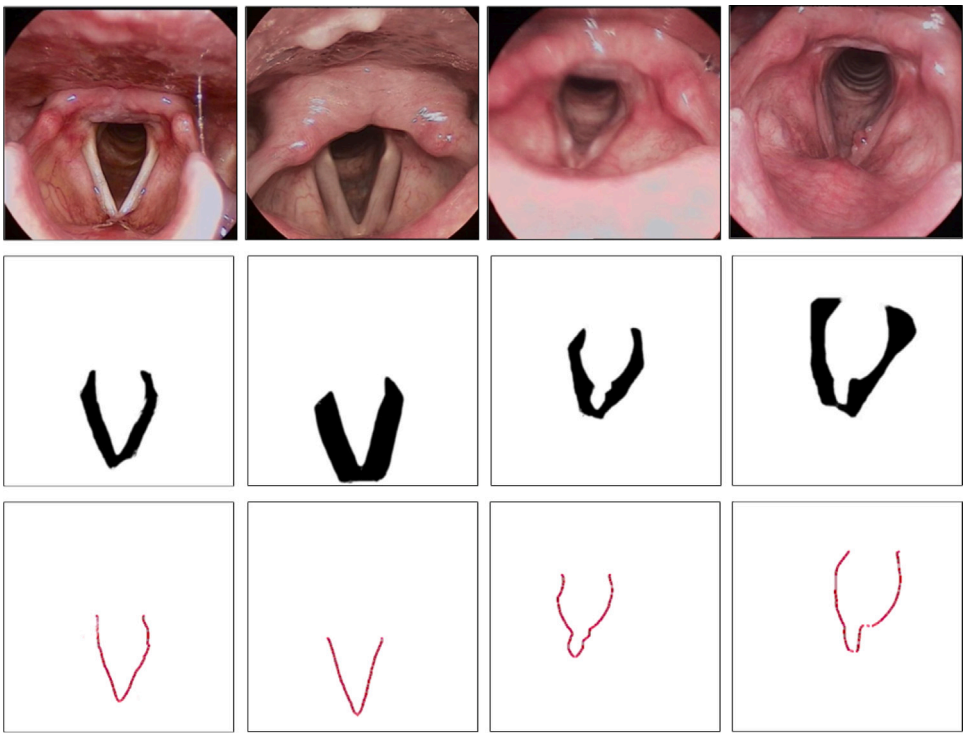

Fig. 7. Top row: the laryngoscopic images. Middle row: vocal cord segmentation using DeepLabv3+. Bottom row: inner vocal cord contour.

Table 1
Accuracy of our method using different batch sizes and learning rates of DeepLabv3+.

Batch Size	Learning Rate	Accuracy
16	0.0005	96.57
	0.001	95.32
	0.01	94.13
32	0.0005	96.95
	0.001	**97.22**
	0.01	96.32
64	0.0005	95.61
	0.001	96.39
	0.01	96.91

Table 2
Parameters of different classification modules.

	VGG-Net	MLP	ResNet
batch_size	256	128	128
learning_rate	0.05	0.01	0.001
weight_decay	0.0003	0.0003	0.00001

Table 3
Average performance of deep network classifiers. The standard deviation is in parentheses.

	VGG-Net	MLP	ResNet
Accuracy	95.7 (1.78)	96.72 (0.94)	**97.2 (0.32)**
Sensitivity	95.47 (1.29)	96.91 (1.16)	**97.04 (0.51)**
Specificity	97.29 (1.46)	96.34 (0.86)	**97.13 (0.47)**

Table 4
Accuracy of different methods.

Method	VGG-Net	CBAM	SE-Net	ECA-Net	MKSC
Accuracy	82.39	90.07	88.21	88.83	90.24
STD	1.24	1.63	2.06	1.94	1.78
Method	U-Net	PSPNet	DeepLabv3+	LFCF	
Accuracy	93.13	92.24	94.21	**97.2**	
STD	1.54	1.67	2.16	1.16	

We compare three recent deep networks, including VGG-Net [9], MLP [39], and ResNet [10]. Table 2 lists the parameters of these methods. We tested with a number of values for each parameter and the one that results in the best performance is kept. Table 3 presents their performance. Both average performance and standard deviation are reported. ResNet achieves the best performance among the three methods in terms of accuracy, sensitivity, and specificity. In addition, ResNet exhibits the minimum standard deviation, which demonstrates superior consistency. In the rest of our experiments and discussion, ResNet is used as the classification module.

4.3. Comparison with state-of-the-art methods

Our proposed method (LFCF) captures the fine-grained features of local contours to classify laryngoscopic images. A commonly used mechanism in deep networks is attention, which improves the performance of various image classification tasks. Therefore, we compare our method with the state-of-the-art algorithms using attention, including SE-Net [12], CBAM [13], ECA-Net [14], and MKSC [22]. We also compare the proposed method with the latest segmentation methods based on deep networks, including U-Net [34] and PSPNet [40]. Since these networks only perform image segmentation, ResNet is used to classify the segmentation results. In addition, we compare our method with DeepLabv3+ [32] and VGG-Net [9].

The average accuracy and standard deviation are reported in Table 4. LFCF yielded an average accuracy of 97.2%, which is the best performance among all the methods. Compared to the second-best, LFCF outperforms DeepLabv3+ by 2.99%. The results demonstrate that delineating the vocal cord and extracting the inner contour features provide a better way of characterizing vocal cords, which helps improve anomaly detection. The standard deviation of LFCF is 1.16, which demonstrates superior consistency.

LFCF obtains better results than CBAM, SE-Net, ECA-Net, and MKSC. SE-Net and ECA-Net introduce channel attention. CBAM adopts a hybrid attention mechanism to calculate the spatial attention weight and the channel attention weight in sequence. It can be seen from Table 4 that SE-Net, CBAM, and ECA-Net have better performance than VGG-Net, indicating that the attention mechanism can improve the classification performance of laryngoscopic images. However, the proposed LFCF outperforms these four algorithms using an attention mechanism, because LFCF captures the local fine-grained features of vocal cords in laryngoscopic images more directly and clearly. LFCF uses segmentation to obtain the overall contour of a vocal cord, such that the method can focus on the vocal cord region. On this basis, the change of adjacent pixel values is used to isolate the inner contour of the vocal cord, such that the LFCF method can focus more on the inner vocal cord contour. LFCF further calculates the inclination angle of the tangent line of each point on the inner vocal cord contour as the latent feature for classification. Consequently, LFCF is able to better capture the characteristics of vegetation, mass, etc. on the vocal cord for superior classification performance.

Table 4 also shows that the performance of LFCF is better than that of MKSC. MKSC introduces parallel multi-kernel-size channel attention and spatial attention to integrate different convolution kernels and cross-channel and cross-space relationships. Although MKSC achieves better performance than other comparison algorithms, it is still outperformed by LFCF. The most important features of laryngoscopic images are concentrated in the inner vocal cord contour. LFCF extracts the inner vocal cord contour and its latent features and hence can focus on the local fine-grained features of the vocal cord.

In addition, methods using segmentation outperform those using attention, which demonstrates that image segmentation enables the methods to focus on the interesting parts of the medical images for better performance. Among the three methods adopting image segmentation, the proposed method LFCF achieves the best performance. Specifically, LFCF performs better than U-Net and PSPNet by about 4.1% and 5%, respectively. U-Net continuously reduces the resolution in the process of downsampling to obtain image information of different scales. The whole network completes the extraction and combination of fine to coarse features to obtain the segmentation. PSPNet uses the spatial pyramid pooling module to realize multi-scale feature extraction to get the segmentation. U-Net and PSPNet obtain gray-scale vocal cord images which are fed into the classifier for abnormal vocal cord identification. In contrast, the proposed method LFCF uses the latent feature of inner vocal cord curves for classification, which enables the classifier to focus on the part where the disease symptoms exist. Therefore, our LFCF method is more accurate.

We also compare the sensitivity and specificity performance of different methods, and the results are shown in Fig. 8. The proposed LFCF obtains the best performance among all the methods. The sensitivity and specificity of LFCF are more than 97%. The sensitivity rate of LFCF is 4.18% higher than that of the second highest PSPNet and the specificity of LFCF is 4.35% higher than that of the second highest U-Net. The results indicate that LFCF can effectively identify vocal cord diseases. For the laryngoscopic images with masses, vegetations, etc. in the vocal cord which is a small part of the whole image, the inner vocal cord contour isolation and its latent feature extraction in LFCF can effectively learn the local fine-grained features in the vocal cord.

4.4. Sensitivity to polyp size and image quality

We divide the laryngoscopic images into three categories based on image quality: 250 good quality ones (the vocal cords are completely exposed), 380 average quality ones (a small region of the vocal cord is covered by laryngeal epiglottis), and 240 poor quality ones (the laryngeal epiglottis region is big and covers part of the vocal cord). An example image of each category is shown in Fig. 9. We test the model (trained with the methods described in Section 4.1) on the abnormal images of each category. As shown in Table 5, the quality of vocal cord images has an impact on the anomaly detection performance. The performance of all methods is at their best when dealing with the images of high quality. Table 5 reports the accuracy, sensitivity, and specificity of the compared methods on the three groups of images. LFCF achieves the best performance among all the methods regardless of the image quality. Specifically, LFCF demonstrates superior accuracy results to the other methods by at least 3.01% in comparison to the second best. The sensitivity and specificity with LFCF are improved by at least 2.39% and 1.90% over those with the other methods, respectively. When the images are of good quality, LFCF also demonstrates improved performance and advances accuracy by 3.01%–13.64% over those methods. These results demonstrate that LFCF is able to better capture the local fine-grained feature of the vocal cord for anomaly detection. In addition, LFCF has generalization ability when it is used on images with different qualities.

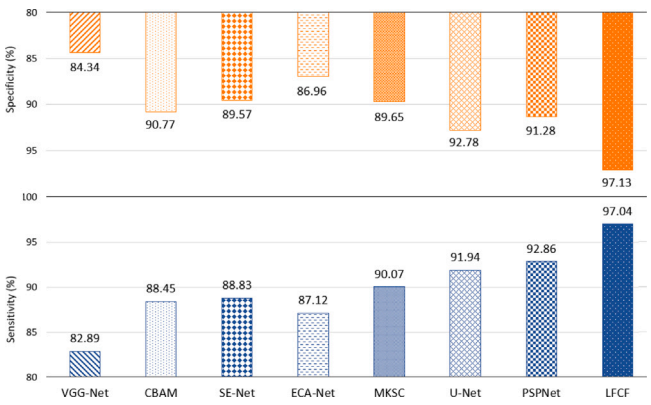

Fig. 8. Sensitivity (bottom) and specificity (top) of various methods for vocal cord disease detection.

We divide the abnormal images into three categories according to the polyp sizes: 90 images with large polyps, 270 images with medium polyps, and 180 images with small polyps. An example image of each category is shown in Fig. 10. We test the trained model on the abnormal images of each category. As shown in Table 6, the accuracy, sensitivity, and specificity of different methods increase with the increasing polyp sizes. With larger polyps, methods can better differentiate between normal and abnormal laryngoscopic images. By comparing across all methods, LFCF achieves the best performance regardless of the polyp sizes. When the images have medium or small-sized polyps in the inner vocal cord contour, LFCF improves the accuracy, sensitivity, and specificity by more than 3.66%, 1.74%, and 2.9%, respectively, in comparison to the second best. When the images have large-sized polyps in the inner vocal cord contour, LFCF increases the accuracy, sensitivity, and specificity by at least 1.75%, 1.43%, and 1.51% over the other methods, respectively. It is demonstrated that LFCF can better extract the fine-grained feature of the inner vocal cord contour from the whole laryngoscopic image for anomaly detection, and LFCF is robust to the polyp sizes.

4.5. Ablation study

Our method consists of three important stages: vocal cord segmentation, inner vocal cord contour extraction, and contour feature extraction. To study the impact of these three stages on the overall performance, we use VGG-Net, SE-Net, CBAM, ECA-Net, MKSC, and ResNet as the basic laryngoscopic image classification networks, and take the output of different stages as the input of these classification networks. Our experiments include the following four cases:

1. original laryngoscopic images as taken as the input of the classification networks;

Fig. 9. Example images with different qualities.

Table 5
Impact of image quality on the abnormal detection performance.

Metric	Method	Good	Average	Poor
Accuracy	VGG-Net	83.62 (2.21)	80.52 (3.16)	80.12 (3.11)
	CBAM	92.47 (1.65)	91.35 (2.66)	87.26 (2.16)
	SE-Net	93.27 (1.82)	88.94 (2.62)	85.66 (1.94)
	ECA-Net	92.49 (1.26)	89.67 (1.54)	86.29 (2.33)
	MKSC	93.76 (1.76)	90.82 (2.16)	86.89 (3.78)
	U-Net	93.47 (0.96)	90.12 (0.54)	88.23 (1.89)
	PSPNet	92.31 (1.15)	89.56 (1.37)	87.64 (2.31)
	DeepLabv3+	94.25 (1.32)	92.42 (1.48)	90.37 (1.25)
	LFCF	**97.26** (0.44)	**95.49** (1.21)	**94.98** (1.11)
Sensitivity	VGG-Net	84.97 (1.68)	82.54 (2.37)	81.63 (1.44)
	CBAM	93.64 (1.28)	91.23 (2.53)	86.43 (2.09)
	SE-Net	93.22 (1.85)	87.46 (1.96)	86.51 (2.38)
	ECA-Net	93.89 (1.59)	88.44 (1.28)	87.02 (2.47)
	MKSC	94.28 (0.88)	89.83 (1.78)	87.38 (2.56)
	U-Net	92.23 (1.03)	90.57 (0.84)	87.51 (2.64)
	PSPNet	91.78 (0.52)	91.63 (1.38)	86.84 (1.94)
	DeepLabv3+	93.96 (1.45)	92.46 (1.23)	90.22 (1.31)
	LFCF	**96.35** (0.57)	**95.16** (1.58)	**93.88** (1.94)
Specificity	VGG-Net	83.66 (2.38)	83.24 (3.76)	80.28 (3.22)
	CBAM	92.66 (1.77)	90.87 (2.48)	87.62 (3.12)
	SE-Net	92.78 (1.33)	86.54 (2.31)	87.68 (1.55)
	ECA-Net	94.52 (2.59)	90.26 (1.88)	88.93 (2.97)
	MKSC	93.54 (1.18)	88.54 (2.14)	86.88 (2.38)
	U-Net	92.87 (0.75)	90.66 (1.46)	86.12 (2.36)
	PSPNet	92.13 (0.78)	91.28 (1.33)	87.51 (1.72)
	DeepLabv3+	94.15 (1.12)	93.09 (1.21)	91.27 (1.55)
	LFCF	**96.16** (0.49)	**94.99** (0.94)	**94.27** (1.67)

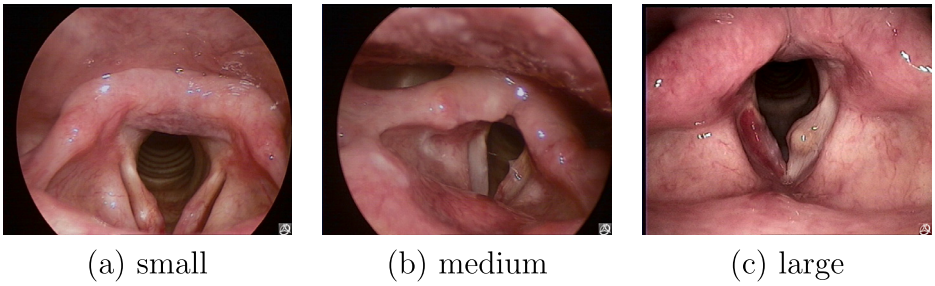

Fig. 10. Example images with polyps of different sizes.

2. segmented whole vocal cord images are fed into the classification networks to evaluate the impact of image segmentation;

3. inner vocal cord contours are used as the input of the classification networks to estimate the inner vocal cord contour isolation;

4. latent inner vocal cord contour features are input to the classification networks to investigate the impact of latent inner vocal cord contour feature extraction.

Note that the proposed method LFCF method uses the latent inner vocal cord contour feature as the input and ResNet as the classification network.

We first evaluate the performance of LFCF in different cases. The experimental results in Table 7 show that the performance of LFCF increases with more stages present in the method in terms of all the performance metrics of accuracy, sensitivity, and specificity. LFCF with the segmentation image as the input improves the accuracy by 3.8% over LFCF with the original image as the input. The proportion of the vocal cord area to the total image is small, and the non-vocal cord area has obvious texture and color characteristics. Image segmentation removes the non-vocal cord region and isolates the overall vocal cord contour which contains the information important for vocal cord disease detection. LFCF with the inner vocal cord contour as the input performs better than LFCF with the image segmentation stage

Table 6
Impact of polyp sizes on the abnormal detection performance.

Metric	Method	Small	Medium	Large
Accuracy	VGG-Net	78.56 (4.51)	81.52 (3.26)	83.12 (4.08)
	CBAM	89.48 (2.16)	91.49 (1.52)	92.67 (1.29)
	SE-Net	88.61 (2.81)	90.26 (2.18)	92.64 (1.64)
	ECA-Net	89.38 (3.16)	91.47 (1.26)	93.09 (1.59)
	MKSC	88.62 (2.77)	91.72 (1.69)	93.65 (1.27)
	U-Net	87.14 (2.46)	92.61(2.15)	95.21 (0.71)
	PSPNet	86.32 (2.71)	90.96 (1.97)	94.39 (0.65)
	DeepLabv3+	90.13 (1.43)	93.03 (0.84)	97.51 (0.55)
	LFCF	**94.52 (1.29)**	**96.69 (0.59)**	**99.26 (0.23)**
Sensitivity	VGG-Net	79.51 (3.98)	82.66 (2.88)	83.89 (2.54)
	CBAM	90.24 (2.42)	91.66 (1.89)	93.15 (1.59)
	SE-Net	89.22 (2.55)	89.84 (2.24)	92.45 (1.77)
	ECA-Net	90.19 (2.41)	91.52 (1.52)	93.22 (1.46)
	MKSC	89.47 (2.15)	92.06 (1.25)	93.88 (0.97)
	U-Net	88.34 (2.23)	93.05 (1.56)	95.21 (0.81)
	PSPNet	86.65 (2.18)	90.86 (1.81)	94.83 (0.43)
	DeepLabv3+	91.74 (1.97)	93.72 (0.66)	97.09 (0.39)
	LFCF	**93.48 (1.81)**	**96.26 (0.48)**	**98.52 (0.34)**
Specificity	VGG-Net	80.77 (4.19)	82.84 (3.11)	83.62 (2.89)
	CBAM	90.27 (1.88)	91.24 (1.47)	93.65 (1.23)
	SE-Net	88.33 (2.74)	90.25 (1.95)	91.88 (1.58)
	ECA-Net	91.07 (2.85)	91.88 (1.43)	92.63 (1.29)
	MKSC	89.55 (2.42)	92.78 (1.33)	94.18 (1.08)
	U-Net	87.46 (2.77)	93.04 (1.86)	95.42 (1.33)
	PSPNet	86.95 (2.51)	90.74 (1.46)	94.65 (1.49)
	DeepLabv3+	90.34 (1.96)	93.57 (1.23)	96.37 (0.94)
	LFCF	**96.35 (0.76)**	**96.47 (0.63)**	**97.88 (0.58)**

Table 7
Performance of LFCF with different stages.

Input	Ori. image	Seg. image	Inner contour	Local feature
Accuracy	90.83 (2.68)	94.21 (2.16)	95.93 (1.95)	**97.22 (1.16)**
Sensitivity	88.04 (2.16)	93.20 (2.35)	95.15 (1.23)	**97.04 (0.96)**
Specificity	92.79 (1.62)	95.08 (2.08)	95.78 (1.45)	**97.73 (0.45)**

by 1.72%. LFCF with the segmented image as the input performs anomaly detection based on the whole vocal cord contour. However, the symptoms of vocal cord diseases are concentrated in the inner vocal cord contour, which is often characterized as vegetations, -like mass, small protuberances, etc. The method with the inner contour isolation stage can focus on the area with disease symptoms and ignore the redundant information in the outer vocal cord contour. LFCF with the local fine-grained vocal cord contour feature extraction stages further improves the accuracy by 1.29% over LFCF taking the inner contour as the input. A healthy person has a smooth inner contour curve, and hence the tangent inclination angle of each point on the curve changes slowly. In contrast, for the patient suffering from a vocal cord disease, there are bulges in the middle and lower parts of the inner contour curve, and hence the tangent inclination angles at the points on the bulges change greatly. LFCF extracts the latent feature of the inner vocal cord contour, by calculating the tangent inclination angle of each point on the inner contour. Consequently, LFCF with all the stages available can better capture the difference between the inner contour curves of patients and healthy persons by integrating the three stages of overall vocal cord contour extraction, inner vocal cord contour isolation, and latent inner vocal cord contour feature extraction.

Table 8 presents the performance of different classification networks in various cases. We only show the accuracy performance in the following experiments for brevity, since the previous experiments demonstrate the superior performance of our method in all three performance metrics of accuracy, sensitivity, and specificity.

For all the classification networks, the methods with the segmentation stage available obtain better results than those with the original images as the input. The classification networks with segmentation images as the input improve the accuracy by 1.41%–5.03% over those directly processing the original images. The results demonstrate that image segmentation enables the methods to focus more effectively on the vocal cord region in the laryngoscopic images, which avoids the interference of the texture and color characteristics in the non-vocal cord area. Table 8 shows that the methods with the inner contours as the input outperform those with only the segmentation stage available by more than 1.35%. The methods with the local fine-grained features as the input can further improve the vocal cord anomaly detection accuracy by 1.08-1.84% over those taking the inner contour as the input. The experimental results in Table 8 illustrate that our local fine-grained vocal cord contour feature extraction can effectively improve the anomaly detection performance regardless of the classification networks, which demonstrates the robustness and generalization ability of our method.

We also investigate the impact of these three stages on the overall performance when using different segmentation algorithms of U-Net, PSPNet, and DeepLabv3+ in whole vocal contour isolation. After using different segmentation networks in Stage 1, we apply the inner vocal cord contour isolation algorithm in Stage 2 and employ our fine-grained inner vocal cord contour feature extract algorithm in Stage 3. Table 9 presents that the method using DeepLabv3+ as the segmentation algorithm achieves the best accuracy.

When only Stage 1 is available in our method, DeepLabv3+ improves the accuracy by 1.08–1.97% compared to the other two segmentation algorithms. When the inner curve isolation stage is present, the method with DeepLabv3+ improves the accuracy up to 1.61% over that with U-Net or PSPNnet. When all the stages are available, the method with DeepLabv3+ outperforms that with the other segmentation algorithms 1.09–1.88%. Table 9 also shows that the inner contour isolation and fine-grained inner curve feature extraction after image segmentation can improve the accuracy of anomaly detection. For example, the methods with the inner curve isolation outperform those with only image segmentation up to 2.8%. The methods with all three stages achieve up to 1.81% better accuracy than those with the previous

Table 8
Accuracy of different classification networks with different stages.

Input	VGG-Net	CBAM	SE-Net
Ori. image	82.39 (1.24)	90.07 (1.63)	88.2 (2.06)1
Seg. image	93.27 (2.16)	94.23 (1.59)	92.87 (1.75)
Inner contour	95.06 (1.69)	95.58 (1.34)	94.46 (2.08)
Local feature	**95.7 (2.26)**	**96.30 (1.89)**	**96.34 (1.84)**
Input	ECA-Net	MKSC	LFCF
Ori. image	88.83 (1.94)	90.24 (1.78)	90.83 (1.06)
Seg. image	93.47 (1.36)	93.86 (2.36)	94.21 (1.47)
Inner contour	95.25 (1.49)	95.66 (1.68)	95.93 (1.38)
Local feature	**96.86 (1.36)**	**97.04 (1.85)**	**97.22 (1.16)**

Table 9
Accuracy of using different segmentation algorithms in our method with different stages.

Stage	U-Net	PSPNet	DeepLabv3+
Image segmentation	93.13 (1.54)	92.24 (1.67)	94.21 (1.26)
Inner contour isolation	94.32 (0.78)	95.04 (0.92)	95.93 (0.85)
Local feature extraction	96.13 (1.27)	95.34 (1.35)	97.22 (0.98)

two stages of whole contour isolation and inner contour extraction. The experimental results illustrate that our local fine-grained vocal cord contour feature extraction can effectively improve anomaly detection performance without regard to segmentation algorithms, which demonstrates that our method is robust and has superior generalization ability.

5. Conclusion and future work

In this paper, we proposed a novel Local Fine-grained vocal cord Contour Feature extraction based vocal cord anomaly detection method (LFCF). The method is divided into four stages. The first stage is the whole vocal cord contour extraction through image segmentation. The second stage is inner vocal cord contour isolation, which obtains the inner contour of a vocal cord based on the change of adjacent pixel values. The third stage is to extract the latent feature of the inner vocal cord contour, by taking the inclination angle of adjacent tangent lines as the feature of the inner vocal cord contour. The fourth stage is the classification stage for laryngoscopic image classification.

We conducted experiments on the dataset constructed from real-world laryngoscopic images. Experimental results demonstrated that the proposed method LFCF achieved an accuracy of 97.21%, a sensitivity of 97.04%, and a specificity of 97.73%, which effectively improved the detection performance of vocal cord diseases. LFCF is superior to the existing methods in terms of accuracy by at least 4.72%. The experimental results also showed that our proposed method is robust and has favorable generalization ability in terms of different image quality, polyp sizes, segmentation algorithms, and classification networks.

Our future work will focus on the following two aspects. First, the idea of geometric features is generally applicable to other medical image classification problems, such as macular disease detection, polyp detection for colonoscopy images, etc., which can be thoroughly evaluated for the applicability and ways of generalizing to other applications. Second, limited by the available data set, our method addresses a two-class classification problem: normal or abnormal. When diagnosing various vocal cord diseases, which is a multi-class classification problem, additional features such as color and texture are needed besides geometric features. For example, leukoplakia and nodules show similar shapes but different colors. Hence, we will construct a comprehensive vocal cord disease data set, based on which we will devise an improved method and validate it on the larger-scale data set.

CRediT authorship contribution statement

Yuqi Fan: Conceptualization, Data curation, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing. **Han Ye:** Software, Visualization, Writing – review & editing. **Xiaohui Yuan:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft, Writing – review & editing, Software, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2018YFB2000505) and the open fund of the Intelligent Interconnected Systems Laboratory of Anhui Province, Hefei University of Technology (PA2021AKSK0114).

Data availability

Data will be made available on request.

References

[1] X. Yuan, B. Giritharan, J. Oh, Gradient vector flowdriven active shape for image segmentation, in: 2007 IEEE International Conference on Multimedia and Expo, 2007, pp. 2058–2061.
[2] A. Verikas, A. Gelzinis, M. Bacauskiene, V. Uloza, Towards a computer-aided diagnosis system for vocal cord diseases, Artif. Intell. Med. 36 (1) (2006) 71–84.
[3] H. Irem Turkmen, M. Elif Karsligil, I. Kocak, Classification of laryngeal disorders based on shape and vascular defects of vocal folds, Comput. Biol. Med. 62 (2015) 76–85.
[4] H. Xiong, P. Lin, J.G. Yu, J. Ye, H. Yang, Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images, EBioMedicine 48 (2019).
[5] W.K. Cho, S.H. Choi, Comparison of convolutional neural network models for determination of vocal fold normality in laryngoscopic images, J. Voice 33 (2020) 634–641.
[6] C. Matava, E. Pankiv, S. Raisbeck, M. Caldeira, F. Alam, A convolutional neural network for real time classification, identification, and labelling of vocal cord and tracheal using laryngoscopy and bronchoscopy video, J. Med. Syst. 44 (2) (2020) 1–10.
[7] M.H. Laves, J. Bicker, L.A. Kahrs, T. Ortmaier, A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation, Int. J. Comput. Assist. Radiol. Surg. 14 (3) (2019) 483–492.
[8] Y. Wang, Survey on deep multi-modal data analytics: Collaboration, rivalry, and fusion, ACM Trans. Multimed. Comput. Commun. Appl. 17 (1s) (2021) 10:1–10:25.
[9] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Int. Conf. Learn. Represent. (2014) 1–14.
[10] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[11] Y. Wang, J. Peng, H. Wang, M. Wang, Progressive learning with multi-scale attention network for cross-domain vehicle re-identification, Sci. China Inf. Sci. 65 (6) (2022) 160103:1–160103:15.
[12] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[13] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.
[14] Q. Wang, B. Wu, P. Zhu, P. Li, Q. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2020, pp. 11531–11539.
[15] Y. Yang, Y.Y. Hu, X. Zhang, S. Wang, Two-stage selective ensemble of CNN via deep tree training for medical image classification, IEEE Trans. Cybern. (2021) 1–14.
[16] S. Sasikanth, S.S. Kumar, Glioma tumor detection in brain MRI image using ANFIS-based normalized graph cut approach, Int. J. Imaging Syst. Technol. 28 (1) (2018) 64–71.
[17] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.
[18] U.R. Acharya, K.M. Meiburger, J.E.W. Koh, J. Vicnesh, E.J. Ciaccio, O.S. Lih, S.K. Tan, R.R.A.R. Aman, F. Molinari, K.H. Ng, Automated plaque classification using computed tomography angiography and Gabor transformations, Artif. Intell. Med. 100 (2019) 101724.
[19] J. Hu, Y. Chen, J. Zhong, R. Ju, Z. Yi, Automated analysis for retinopathy of prematurity by deep neural networks, IEEE Trans. Med. Imaging 38 (1) (2018) 269–279.
[20] J.C. Souza, J. Diniz, J.L. Ferreira, G. Silva, A.C. Silva, A. Paiva, An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks, Comput. Methods Programs Biomed. 177 (2019) 285–296.
[21] J. Antony, K. Mcguinness, K. Moran, N. Connor, Feature learning to automatically assess radiographic knee osteoarthritis severity, in: Deep Learners and Deep Learner Descriptors for Medical Applications, Springer International Publishing, 2020, pp. 9–93.
[22] Y. Fan, J. Liu, R. Yao, X. Yuan, COVID-19 detection from X-ray images using multi-kernel-size spatial-channel attention network, Pattern Recognit. (8) (2021) 108055.
[23] B.B. Xiaohong Gao, Artificial intelligence in endoscopy: The challenges and future directions, in: Artificial Intelligence in Gastrointestinal Endoscopy, 2020, pp. 117–126.
[24] K.C. W, J.L. Y, A.J. H, S.J. I, C. Y, Y.N. S, Diagnostic accuracies of laryngeal diseases using a convolutional neural network-based image classification system, Laryngoscope (2020) 2558–2564.
[25] L. Yin, L. Yang, M. Pei, J. Li, Y. Jia, Laryngoscope8: Laryngeal image dataset and classification of laryngeal disease based on attention mechanism, Pattern Recognit. Lett. 150 (6) (2021).
[26] A. Paderno, C. Piazza, F.D. Bon, D. Lancini, S. Moccia, Deep learning for automatic segmentation of oral and oropharyngeal cancer using narrow band imaging: Preliminary experience in a clinical perspective, in: Frontiers in Oncology, 2021, 626602.
[27] M.A. Azam, C. Sampieri, A. Ioppi, S. Africano, Deep learning applied to white light and narrow band imaging videolaryngoscopy: Toward real-time laryngeal cancer detection, in: The Laryngoscope, 2021, pp. 1798–1806.
[28] T. Lu, J. Ren, Automatic recognition of laryngoscopic images using a deep-learning technique, in: The Laryngoscope, 2020, pp. E686–E693.
[29] X. Yuan, J. Shi, L. Gu, A review of deep learning methods for semantic segmentation of remote sensing imagery, Expert Syst. Appl. 169 (2021) 114417.
[30] X. Yuan, Segmentation of blurry object by learning from examples, in: Medical Imaging 2010: Image Processing, 7623, 2010, p. 76234G.
[31] G. Wang, M.A. Zuluaga, W. Li, P. Rosalind, P.A. Patel, A. Michael, D. Tom, A.L. Divid, D. Jan, O. Sebastien, DeepIGeoS: A deep interactive geodesic framework for medical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 41 (7) (2019) 1559–1572.
[32] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: European Conference on Computer Vision, ECCV, 2018, pp. 833–851.
[33] N. Yu, Z. Zhang, Q. Xu, E. Firdaous, J. Lin, An improved method for cloth pattern cutting based on holistically-nested edge detection, in: 2021 IEEE 10th Data Driven Control and Learning Systems Conference, DDCLS, 2021, pp. 1246–1251.
[34] Z. Lin, Y. Cui, J. Liu, Z. Sun, X. Wang, Automated segmentation of kidney and renal mass and automated detection of renal mass in CT urography using 3D U-Net-based deep convolutional neural network, Eur. J. Radiol. (2021).
[35] L. Wu, Y. Wang, X. Li, J. Gao, Deep attention-based spatially recursive networks for fine-grained visual recognition, IEEE Trans. Cybern. 49 (5) (2019) 1791–1802.
[36] Y. Xie, J. Zhang, H. Lu, C. Shen, Y. Xia, SESV: Accurate medical image segmentation by predicting and correcting errors, IEEE Trans. Med. Imaging 40 (1) (2021) 286–296.
[37] Y. Chen, Y. Bai, W. Zhang, T. Mei, Destruction and construction learning for fine-grained image recognition, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, 2019, pp. 5152–5161.
[38] C. Zhuang, X. Yuan, W. Wang, Boundary enhanced network for improved semantic segmentation, in: International Conference on Urban Intelligence and Applications, 2020, pp. 172–184.
[39] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, A. Dosovitskiy, MLP-Mixer: An all-MLP architecture for vision, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2021, pp. 1–16.
[40] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2881–2890.