Pattern Recognition Letters 150 (2021) 207–213

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Laryngoscope8: Laryngeal image dataset and classification of laryngeal disease based on attention mechanism

Li Yin[a,1], Yang Liu[b,1], Mingtao Pei[a], Jinrang Li[b,*], Mukun Wu[b], Yuanyuan Jia[b]

[a] Beijing Laboratory of Intelligent Information Technology, Beijing Institute of Technology, Beijing, China
[b] The Department of Otorhinolaryngology Head and Neck Surgery, The Sixth Medical Center of PLA General Hospital, Beijing, China

ARTICLE INFO

Article history:
Received 18 November 2020
Revised 23 June 2021
Accepted 28 June 2021
Available online 26 July 2021

Edited by: Prof. Jiwen Lu

MSC:
41A05
41A10
65D05
65D17

Keywords:
Laryngeal image dataset
Laryngeal disease classification
Attention mechanism

ABSTRACT

Laryngeal disease is a common disease worldwide. However, currently there are no public laryngeal image datasets, which hinders the development of automatic classification of laryngeal disease. In this work, we build a new laryngeal image dataset called Laryngoscope8, which comprises 3057 images of 1950 unique individuals, and the images have been labeled with one of eight labels (including seven pathological labels and one normal label) by professional otolaryngologists. We also propose a laryngeal disease classification method, which uses attention mechanism to obtain the critical area under the supervision of image labels for laryngeal disease classification. That is, we first train a CNN model to classify the laryngeal images. If the classification result is correct, the region with strong response is most likely a critical area. The regions with strong responses are used as training data to train an object localization model that can automatically locate the critical area. Given an image for classification, the trained object localization model is employed to locate the critical area. Then, the located critical area is employed for image classification. The entire process only requires image-level labels and does not require manual labeling of the critical area. Experiment results show that the proposed method achieves promising performance in laryngeal disease classification.

1. Introduction

Laryngeal disease is a common disease worldwide. According to the China over-the-counter (OTC) Market and Media Research data from the new generation market monitoring agency, 41.6% of the residents of 18 key cities across the country have endured oral and throat discomfort in 2009. The prevalence of oral and throat diseases is only second to colds and coughs, ranking third [1]. Laryngoscope images are the main data used in the diagnosis of laryngeal disease. The automatic classification of laryngeal disease based on Laryngoscope images can reduce the workload of otolaryngologists, improve their work efficiency, and assist young doctors with less experience in making correct diagnoses. Deep learning is commonly used in medical image classification [2–4] and deep learning relies on large-scale labeled datasets [5,6]. There are currently some medical image datasets, such as chest radiographs ChestXray14 [7], CheXpert [8], COVIDGR-1.0 [9], and upper limb radiographs MURA [10]. However, as far as we know, there are no laryngeal image datasets, which hinders the development of automatic diagnosis of laryngeal diseases. To this end, we build a new laryngeal image database called Laryngoscope8, which is composed of 3057 images of 1950 unique individuals with eight labels (including seven pathological labels and one normal label), labeled by professional otolaryngologists.

Given a laryngoscope image, the common method for automatic diagnosis is using the entire image as the input to a certain deep network and the output of the network as the diagnosis of the image. However, an otolaryngologist will usually focus on certain critical areas in the image to make a diagnosis. As shown in Fig. 1, the vocal cord and nearby area is the critical area and provides important information for the diagnosis of laryngeal diseases. Therefore, automatic localization of the critical area will help in the diagnosis of laryngeal diseases.

The common method to locate critical areas is to manually label the critical area as training data to train an object localization model. However, unlike the labeling of natural images, labeling the critical area in laryngoscope images requires professional knowledge of laryngeal disease diagnosis. The person who has professional knowledge of laryngeal disease diagnosis, usually a doctor, does not have time to do the tiresome and time-consuming labeling work.

* Corresponding author.
E-mail address: entljr@sina.com (J. Li).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.patrec.2021.06.034
0167-8655/

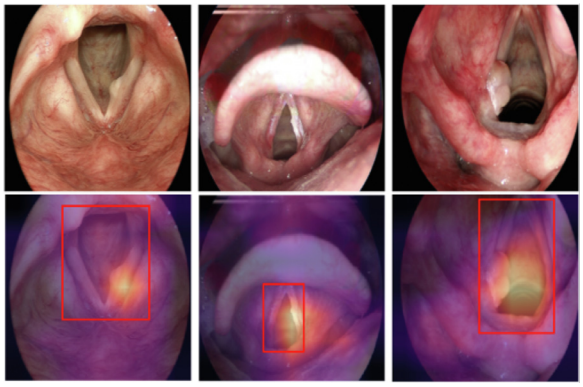

Fig. 1. Examples of laryngoscope images and their corresponding heatmaps.

Recently, attention mechanism has been introduced to CNN to improve both its performance and interpretability. The essence of attention is to learn a weight map which represents the relative importance of features. In our work, we use the attention mechanism to obtain the critical area under the supervision of image labels. That is, we first train a CNN model to classify the laryngoscope images. If the classification result is correct, then the region with strong response is most likely the critical area. The regions with strong responses are used as training data to train an object localization model that can automatically locate the critical area. Given an image for classification, the trained object localization model is employed to locate the critical area first, then the located critical area is employed for the image classification. The entire process only requires image-level labels and does not require manual labeling of the critical area.

Our contributions can be summarized as follows:

- We construct a new laryngeal image dataset called Laryngoscope8, which comprises 3057 images of 1950 individuals that have been labeled with one of eight labels (seven corresponding to laryngeal diseases and one corresponding to normal tissue) by professional otolaryngologists.
- We propose to employ the disease label as weak supervision to extract the critical area in laryngoscope images for laryngeal disease classification. The proposed method achieves better performance than using the original image directly for laryngeal disease classification.

2. Related work

2.1. Medical image dataset

Various medical image datasets have been collected and published in recent years. ChestX-Ray14 [7] is a large-scale frontal-view chest-X-ray image dataset that includes 112,120 images of 30,805 unique patients with fourteen disease labels, where each image can have multi-labels. The labels are extracted from the associated radiological reports using natural language processing. CheXpert [8] is a large dataset that contains 224,316 chest radiographs of 65,240 patients in fourteen categories. COVIDGR-1.0 [9] contains 426 positive and 426 negative PA (PosteroAnterior) CXR views. MURA [10] is a large dataset of musculoskeletal radiographs that contains 40,561 images labeled manually by radiologists as either normal or abnormal. As far as we know, there are no public datasets on laryngeal images. The dataset released in this paper, the Laryngoscope8, is the first dataset on laryngeal images. Laryngoscope8 contains 3057 images of 1950 unique individuals in eight categories (seven diseases types and one normal type). Each individual contains one or several images, and each image has only one label. All labels are manually labeled by professional otolaryngologists.

2.2. Attention mechanism for medical image classification

Deep learning has been extensively used in medical image classification [11–15]. Esteva et al.[11] demonstrated that CNN is capable of classifying skin cancer with a level of competence comparable to that of dermatologists. The CheXNet [12], which is a finetuned DenseNet-121 [16], has been reported to achieve radiologist levels of pneumonia localization. Rodrigues et al. [13] combined DenseNet201 extraction model with the KNN classifier to detect melanoma skin cancer. Polsinelli et al. [14] proposed a light CNN design, based on the model of the SqueezeNet, to discriminate COVID-19 CT image from pneumonia and healthy CT images, achieves 85.03% accuracy. Xiong et al. [15] adopted the GoogLeNet Inception v3 to detect laryngeal cancer for reducing the burden of endoscopists.

Recently, the attention mechanism has been widely employed in convolutional neural networks, improving both their performance and interpretability in medical image classification tasks [7,17–20]. Wang et al. [7] proposed an unified weakly supervised multi-label image classification and disease localization framework to detect common thoracic spine diseases. The AG-CNN proposed in [17] uses attention mechanism to obtain the local areas that needs to be paid attention, and then the obtained local areas are combined with the global image to make the final judgment. Zhang et al. [18] proposed the Attention Branch Network on the basis of CAM. Without relying on back propagation, the CAM structure is used to introduce the attention mechanism and improve the accuracy of CNN. In the COVID-19 pneumonia detection experiment [19], uniform sampling data and size-balanced sampling data were sent to attention networks to eliminate the data imbalance. In the work of [20], a decision network was proposed to select the patch on the global image, and then a soft attention network was used to classify the patch. The classification result was sent to the decision network as feedback to update the selection strategy of the decision network.

Our work is similar to AG-CNN [17], The difference is that the image label information is not utilized in AG-CNN, in which all areas with strong response are used for the classification. However, the correctly classified images usually focus on the critical area, while incorrectly classified images have a high probability of focusing on areas that are not critical areas and should not be used for the classification. Different with AG-CNN , our method uses the image label as supervision to extract the critical area automatically as training data to train a localization model in the training stage. The trained localization model is used in the testing stage to locate critical areas for classification of the image.

3. The laryngoscope8 dataset

The Laryngoscope8 dataset includes 3057 laryngeal images collected from head and neck surgery procedures in the department of otorhinolaryngology of the sixth medical center of PLA general hospital. A total of 1950 individuals are involved in our Laryngoscope8 dataset, from which 1297 individuals had only one laryngeal image, and the remaining 653 individuals collected two or more laryngeal images. The images in the dataset were classified into eight categories by the otolaryngologists in the hospital. The eight categories include: Reinke's Edema, Glottic Cancer, Granuloma, Vocal Cord Leukoplakia, Vocal Cord Cyst, Vocal Cord Nodules, Vocal Cord Polyps, and Normal. The laryngeal images in our database were taken by two laryngoscope devices: Xion Matrix HD3 and Delon HD380B. The dimensions of the laryngeal images were mostly 1920 × 1080. Table 1 lists the image distribution

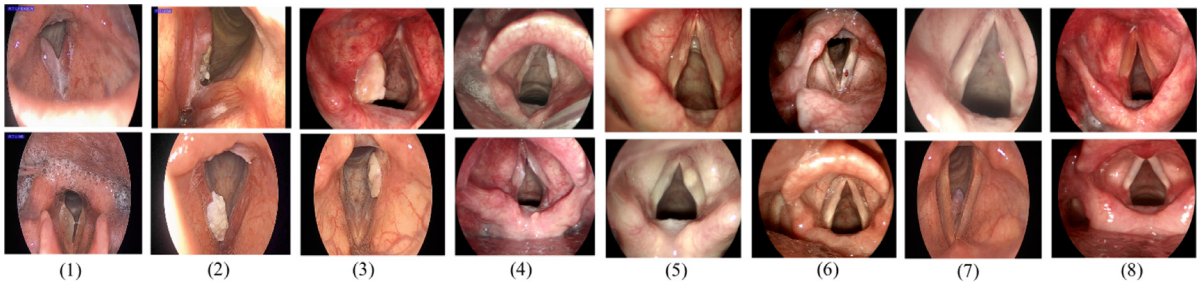

Fig. 2. Eight labels of Laryngoscope image: (1) Reinke's Edema; (2) Glottic Cancer; (3) Granuloma; (4) Vocal Cord Leukoplakia ; (5) Vocal Cord Cyst; (6) Vocal Cord Nodules; (7) Vocal Cord Polyps; (8) Normal.

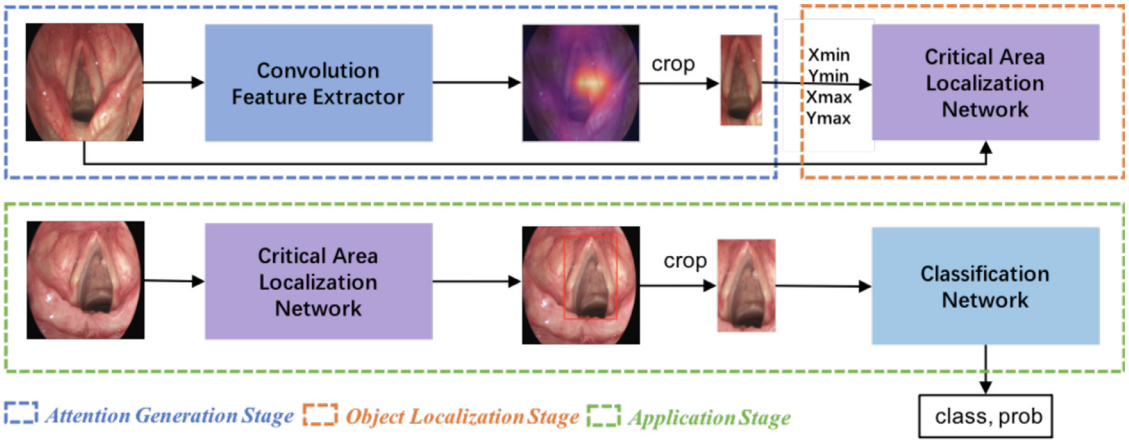

Fig. 3. Overall flowchart of our method. Feature Extractor in Attention Generation Stage specifically refers to spatial features extracted by the last convolutional layer.

Table 1
Laryngoscope image distribution.

Type	Number
Reinke's Edema	110
Glottic Cancer	22
Granuloma	604
Normal	1323
Vocal Cord Leukoplakia	191
Vocal Cord Cyst	74
Vocal Cord Nodules	172
Vocal Cord Polyps	561

Table 2
Classification results using global image and critical image.

Method	Accuracy	average AUC
global original image	71%	0.843
critical image	**77%**	**0.912**

Table 3
Classification results of different methods.

Method	Accuracy	average AUC
CheXNet [12]	71%	0.843
AG-CNN [17]	71%	0.847
Inception_v3 [15]	71%	0.870
Ours	**73%**	**0.893**

in the eight categories. Fig. 2 shows some sample images of each category.

The images are labeled during the diagnostic procedures following the regulations of the hospital. That is, for each patient, one laryngeal image is taken (more laryngeal images will be taken if necessary) by a laryngoscopist, then the image is diagnosed by the laryngoscopist. If the lesions contained in the image are diagnosed as benign lesions such as Vocal Cord Nodules, Vocal Cord Polyps, Vocal Cord Cyst, Reinke's Edema, Granuloma and normal, then the diagnostic procedure is finished and the diagnostic result is used as the image label. If the lesions contained in the image are diagnosed as Vocal Cord Leukoplakia or glottic cancer, according to the regulations of the hospital, two more ENT (Ears, Nose, and Throat) physicians including one expert are required to diagnose the image to reduce the misdiagnosis. Because these two diseases are more serious than the benign lesions, and misdiagnosis will have severe impact on the patient. If there is a disagreement among the physicians and laryngoscopist, the label given by the expert will be regarded as the final label. For images containing Glottic Cancer, the labels are further corrected by biopsy results.

4. Methodology

4.1. Framework

The overall flowchart of our method is presented in Fig. 3. The whole process is divided into three stages. The first stage is called the attention generation stage. In this stage, we use attention mechanism to obtain the critical areas under the supervision of the image label. The second stage is called the critical area localization training stage. In this stage, we use the critical areas obtained in the first stage as the training data to train a localization model that can automatically locate the critical area. The last stage is the application stage, in which the localization model obtained in the second stage is employed to locate the critical area in the test images, and the extracted critical area is used for classification.

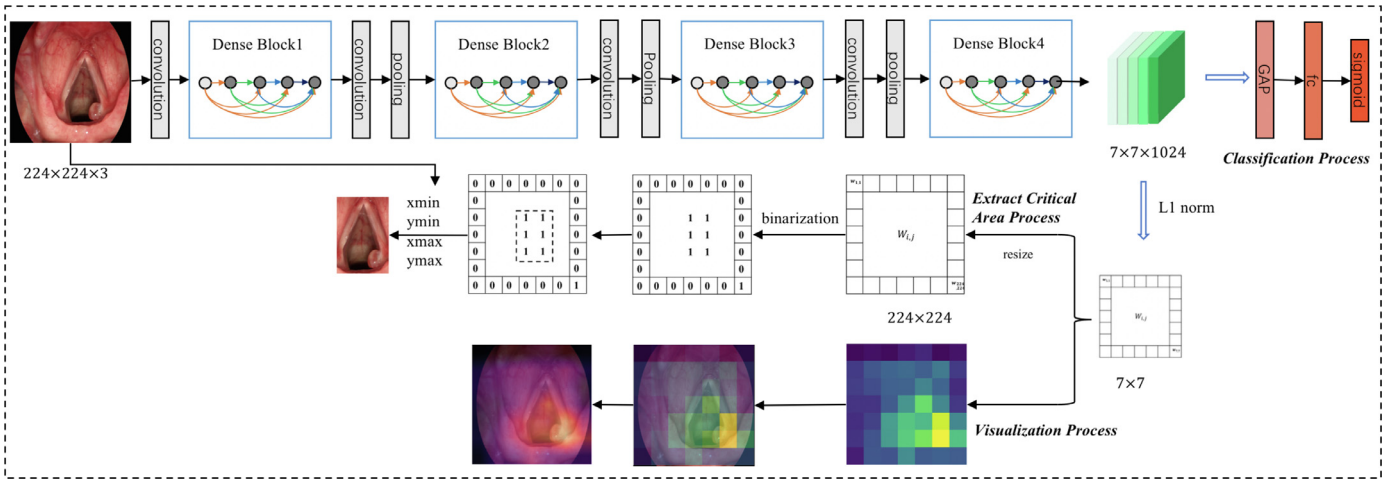

Fig. 4. Framework of attention generation stage. DenseNet-121 is used as the backbone. There are three processes here: classification, process of extracting critical areas, and visualization process that reflects critical areas.

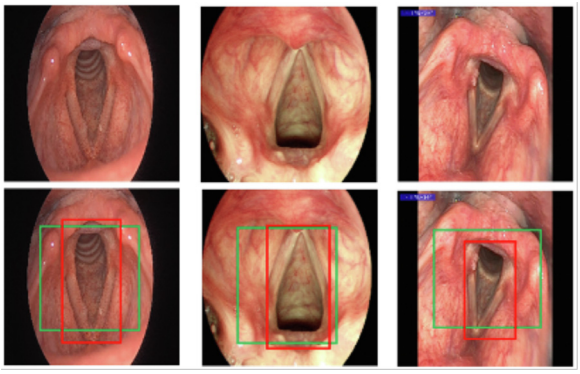

Fig. 5. Examples of critical area localization results. Critical areas obtained through the localization network are marked by green box, and manually labeled critical areas are marked by red box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Attention generation

The attention mechanism is used in this stage to extract the critical areas under the supervision of the image label. First, a classification model is trained using the whole image. We chose DenseNet-121 as the backbone, as shown in Fig. 4. Given an image, after a series of convolution and pooling operations, a spatial activation map before the global average pooling [21] (GAP) layer is obtained. This activation map contains a significant amount of spatial information. Let x, y and k represent the three dimensions of the activation map, that is (x, y) represents the spatial position and k represents the channel, $x \in \{1, \ldots, X\}$, $y \in \{1, \ldots, Y\}$ and $k \in \{1, \ldots, K\}$ where $X = 7$, $Y = 7$ and $K = 1024$ in the DenseNet-121. Let $f^k(x, y)$ epresent the activation of spatial location (x, y) on the kth channel. Heatmap H can be constructed by a statistical value across the channel dimensions. Equations (1)–3 show the Heatmap constructed by L1 norm, L2 norm and infinite norm, respectively. L1 norm is used in our method for heatmap generation, and we also compare the performance of using L2 norm and infinite norm in the experiment section.

$$H(x, y) = \sum_{k=1}^{K} |f^k(x, y)| \tag{1}$$

$$H(x, y) = \sqrt{\sum_{k=1}^{K} (f^k(x, y))^2} \tag{2}$$

$$H(x, y) = \max_k |f^k(x, y)| \tag{3}$$

$H(x, y)$ directly indicates the importance of the activations in location (x, y) for classification. In the visualization process in Fig. 4, we can see that the response of the lesion area in the image is stronger than other areas. Therefore, the area with stronger response usually is the critical area. Here, the image label is used as supervision for the critical area extraction. That is, when the image is correctly classified, the area with a stronger response is regarded as the critical area. When the image is incorrectly classified, the areas with stronger response are less likely to be critical areas and are not used to train the localization model.

The heatmap is resized to the same size as the original image and binarized by a threshold. If the value of a certain spatial position (x, y) in the heatmap is larger than a threshold τ, then the value at the corresponding position in the mask is assigned as 1.

$$M(x, y) = \begin{cases} 1, & H(x, y) > \tau \\ 0, & otherwise \end{cases} \tag{4}$$

where τ is the threshold that controls the size of the attended region. A larger τ leads to a smaller region, and vice versa. Finally, we look for the largest connection area in the binarized activation map, which corresponds to the critical area.

4.3. Critical area localization

At this stage, we use the critical area obtained in the previous stage as training data to train a localization model that can automatically locate the critical area. As previously mentioned, only critical areas in the correctly classified images are used as the training data. For correctly classified images, the areas with strong response in the heatmap usually correspond to the critical areas. For incorrectly classified images, there is a high possibility of marking the critical area incorrectly. Therefore, we only use the correctly classified images in determining the critical areas to ensure the accuracy of the localization model. Here, we choose Faster-RCNN [22] as the critical area localization model.

4.4. Application stage

Given the trained localization model, the critical areas in the training images are extracted and used as the training data to train a second classification model. Again, DenseNet-121 is used as the backbone. Then, for a testing image, the critical area is extracted

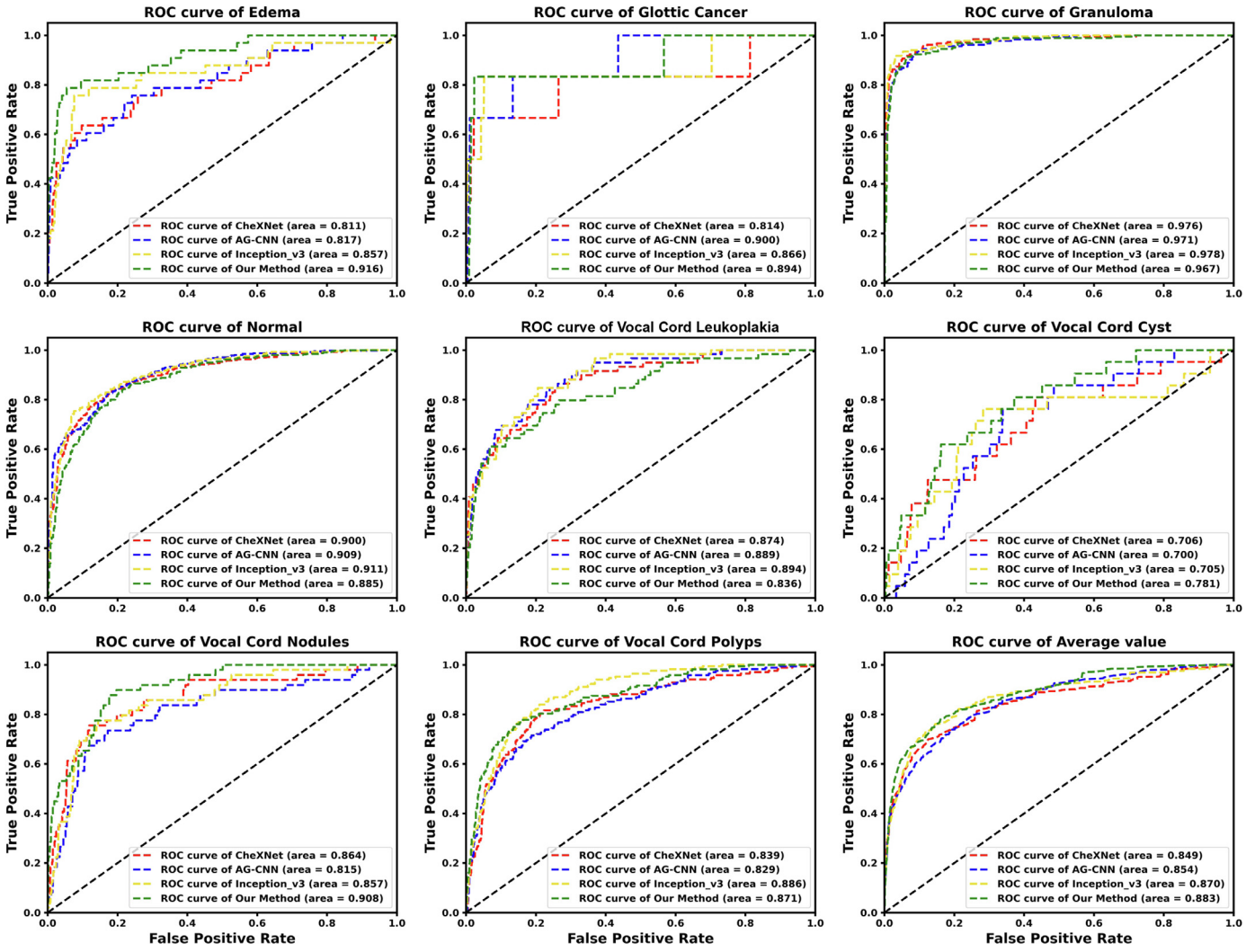

Fig. 6. Receiver operating characteristic (ROC) curves of the four methods. The corresponding Area Under Curve (AUC) values are given in Table 4.

first, and the extracted critical area is input to the classification model for classification of the testing image.

4.5. Implementation details

The attention generation module and the classification module were implemented in Python using the PyTorch library. A NVIDIA TitanXp GPU was used for training. All models are trained by the Adam optimizer [23] with initial learning rate of 10^{-4}, $\beta 1 = 0.9$, $\beta 2 = 0.999$, and weight-decay $= 10^{-4}$. The batch size is set as 32, and the models are trained for 500 epochs. The localization module was implemented in Python using TensorFlow library, and is trained by momentum optimizer, the learning rate is 10^{-3}, the weight decay and momentum are set as 5×10^{-4} and 0.9 respectively.

We use the heatmap generated by L1 norm (Eq. (1)) to evaluate our method in our experiments, and we also compare the performance of our method using different heatmaps in Section 5.4

5. Experiments and results

We test our method on the Laryngoscope8 dataset with 70% of the images chosen as training set and the remaining 30% as testing set. There is no overlap of individuals in the training set and the test set.

5.1. Evaluation metrics

We use accuracy, AUC (Area Under the Curve) for each class, and the average AUC across all the classes for performance evaluation.

Accuracy is defined as

$$Accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ predictions} \tag{5}$$

AUC for each class refers to the area under the Receiver operating characteristic (ROC) curve of that class, and the ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The average AUC is the average value of the AUC for each class.

5.2. The effectiveness of critical area

We assume that for correctly classified images, the areas with strong response in the heatmap usually correspond to the critical areas. This assumption is based on the observation that given an image, the physicians will focus on critical areas to make the diagnosis. If they make the diagnosis based on regions other than the critical areas, there will be a high probability that the diagnosis is wrong.

To justify this assumption, we classified all the 3057 images in the Laryngoscope8 dataset (2140 in the training set and 917 in the

Table 4
AUC for different methods.

Methods	Edema	Cancer	Granuloma	Normal	Leukoplakia	Cyst	Nodules	Polyps	average AUC
CheXNet [12]	0.795	0.822	0.979	0.900	0.876	0.685	0.825	0.853	0.843
AG-CNN [17]	0.805	0.879	0.972	0.895	**0.896**	0.658	0.828	0.838	0.847
Inception_v3 [15]	0.857	0.866	**0.978**	**0.911**	0.894	0.705	0.857	**0.886**	0.870
Ours	**0.900**	**0.936**	0.965	0.878	0.853	**0.849**	**0.886**	0.871	**0.893**

* We compute AUC of each class and average AUC across 8 diseases. For each column, best results are highlighted in bold.

test set) using the classification model trained on the training set, and 2769 (2109 in the training set and 660 in the test set) images were correctly classified. We check the areas with strong response in the heatmaps corresponding to these 2769 images, and find that in 2755 (2100 in the training set and 655 in the test set) images, the critical area (vocal cord part) is in accordance with the area with strong response. That is, our assumption is valid for 99.6% correctly classified images in the training set, and is valid for 99.2% correctly classified images in the test set. The experiment results justify our assumption that for correctly classified images, the areas with strong response in the heatmap usually correspond to the critical areas.

To verify the effectiveness of using critical areas for classification, we conducted two sets of experiments. In the first experiment, the original images in the training set are used to train a DenseNet-121 as the classification model. In the second experiment, the critical areas in each image in the training set are cropped out manually, and a DenseNet-121 is trained by the cropped critical areas as the classification model. Both the original images and the cropped critical areas are resized to 224 × 224 as input to the DenseNet-121. The two DenseNet-121 are trained with the same super parameters. The results in Table 2 indicate that the classification performance using the critical area is better than the performance using the original image. The accuracy is increased by 6 percentage points, and the AUC is increased by 0.069, which proves that the critical area does contain important information for the classification of laryngeal diseases.

5.3. The performance of critical area localization

The performance of critical area localization greatly affects the classification results. We chose Faster R-CNN [22] as the critical area localization model. Figure 5 shows several examples of critical area localization results. It can be seen that the localized critical areas are in accordance with the manually labeled areas, which proves that the trained localization model can localize the critical areas effectively.

5.4. The performance of laryngeal disease classification

We compare our method with AG-CNN [17], CheXNet [12], and Inception_v3 [15]. As shown in Table 3, our method outperforms the compared methods. Our method achieves a two-percentage-point increase in accuracy, also achieved varying degrees of increase in average AUC. Although the performance is not as good as when using manually labeled critical areas, it confirms the effectiveness of the attention mechanism for laryngeal disease classification. Moreover, our method only requires image-level labels and does not require manual labeling of the critical area.

Table 4 shows the AUC values of AG-CNN, CheXNet, Inception_v3 and our method for each category. Figure 6 shows the ROC curve for each class. We can see that our method achieved the best performance on Reinke's Edema, Glottic Cancer, Vocal Cord Cyst, Vocal Cord Nodules, and the average AUC.

In addition, we use the L2 norm in Eq. (2) and the infinite norm in Eq. (3) to generate heatmap as comparison experiments for laryngeal disease classification. The comparison results are listed in Table 5. We can see that using the L1 norm and L2 norm can achieve better performance than AG-CNN, CheXNet and Inception_v3, while the classification performance using infinite norm is slightly lower than that of the two compared methods. The reason maybe that when using the infinite norm, only the largest value from all channels is used, which may cause a loss of some important information, and lead to lower classification performance.

Table 5
Classification results using different heatmaps.

Method	Accuracy	average AUC
L1 norm	73%	0.893
L2 norm	72%	0.881
Infinite norm	71%	0.845

6. Conclusion and future work

In this work, we established a new laryngeal image dataset composed of 3057 images of 1950 unique individuals that have been labeled with one of eight labels by professional otolaryngologists. We also proposed a classification method that uses attention mechanism to obtain the critical area under the supervision of the image labels for laryngeal disease classification. Experiment results showed that the critical area contains important information and can help to improve the performance of laryngeal disease classification.

Currently, the classification performance of our method is not as good as using manually labeled critical areas. The reason is that the automatically located areas still contain some redundant information. In the future, we will examine how to improve the performance of critical area localization, and as a consequence, to further improve the performance of laryngeal disease classification. Currently, our dataset only contains eight different categories and is a single-label classification task. In the future, we will consider adding disease categories and consider the multi-label classification task.

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.

References

[1] N.G.M.M. Agency, Analysis of China's oral and throat disease market, Admen 000 (006) (2010). 28–28
[2] D. Shen, G. Wu, H.I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (1) (2017) 221–248.
[3] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. Van Ginneken, C.I. Snchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (9) (2017) 60–88.
[4] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.
[5] L. Oakden-Rayner, Exploring large scale public medical image datasets, Acad. Radiol. 27 (1) (2020) 106–112.

[6] C. Sun, A. Shrivastava, S. Singh, A. Gupta, Revisiting unreasonable effectiveness of data in deep learning era, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 843–852, doi:10.1109/ICCV.2017.97.
[7] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, ChestX-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3462–3471, doi:10.1109/CVPR.2017.369.
[8] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R.L. Ball, K.S. Shpanskaya, J. Seekins, D.A. Mong, S.S. Halabi, J.K. Sandberg, R. Jones, D.B. Larson, C.P. Langlotz, B.N. Patel, M.P. Lungren, A.Y. Ng, CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI Press, 2019, pp. 590–597, doi:10.1609/aaai.v33i01.3301590.
[9] S. Tabik, A. Gmez-Ros, J.L. Martn-Rodrguez, I. Sevillano-Garca, M. Rey-Area, D. Charte, E. Guirado, J.L. Surez, J. Luengo, M.A. Valero-Gonzlez, P. Garca-Villanova, E. Olmedo-Snchez, F. Herrera, COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images, IEEE J. Biomed. Health Inform. 24 (12) (2020) 3595–3605, doi:10.1109/JBHI.2020.3037127.
[10] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R.L. Ball, MURA: large dataset for abnormality detection in musculoskeletal radiographs, arXiv preprint arXiv:1712.06957 (2017).
[11] 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.
[12] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K.a. Shpanskaya, CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning, arXiv preprint arXiv:1711.05225 (2017).
[13] D. de A. Rodrigues, R.F. Ivo, S.C. Satapathy, S. Wang, J. Hemanth, P.P.R. Filho, A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system, Pattern Recognit. Lett. 136 (2020) 8–15.
[14] M. Polsinelli, L. Cinque, G. Placidi, A light CNN for detecting COVID-19 from CT scans of the chest, Pattern Recognit. Lett. 140 (2020) 95–100.
[15] 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).
[16] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269, doi:10.1109/CVPR.2017.243.
[17] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, Y. Yang, Thorax disease classification with attention guided convolutional neural network, Pattern Recognit. Lett. 131 (2020) 38–45.
[18] J. Zhang, Y. Xie, Y. Xia, C. Shen, Attention residual learning for skin lesion classification, IEEE Trans. Med. Imaging 38 (9) (2019) 2092–2103, doi:10.1109/TMI.2019.2893944.
[19] X. Ouyang, J. Huo, L. Xia, F. Shan, J. Liu, Z. Mo, F. Yan, Z. Ding, Q. Yang, B. Song, F. Shi, H. Yuan, Y. Wei, X. Cao, Y. Gao, D. Wu, Q. Wang, D. Shen, Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia, IEEE Trans Med Imaging 39 (8) (2020) 2595–2605, doi:10.1109/TMI.2020.2995508.
[20] B. Xu, J. Liu, X. Hou, B. Liu, J. Garibaldi, I.O. Ellis, A. Green, L. Shen, G. Qiu, Attention by selection: a deep selective attention approach to breast cancer classification, IEEE Trans. Med. Imaging 39 (6) (2020) 1930–1941, doi:10.1109/TMI.2019.2962013.
[21] M. Lin, Q. Chen, S. Yan, Network in network, Comput. Sci. (2013).
[22] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149, doi:10.1109/TPAMI.2016.2577031.
[23] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, 2015.