Applying Deep Learning with Convolutional Neural Networks to Laryngoscopic Imaging for Real-time Automated Segmentation and Classification of Vocal Cord Leukoplakia

Juanjuan Hu[1], Jiawei Luo[1], Jia Ren[1], Lan Lan[1], Ying Zhang[1], dan lu[1], Xiaobo Zhou[2], and Hui Yang[1]

[1]West China Hospital/West China School of Medicine, Sichuan University
[2]University of Texas Health Science Center at Houston

August 10, 2021

Abstract

Objectives The study was to apply deep learning (DL) with convolutional neural networks (CNNs) to laryngoscopic imaging for assisting in real-time automated segmentation and classification of vocal cord leukoplakia. Methods This was a single-center retrospective diagnostic study included 216 patients who underwent laryngoscope and pathological examination from October 1, 2018 through October 1, 2019. Lesions were classified as nonsurgical group (NSG) and surgical group (SG) according to pathology. All selected images of vocal cord leukoplakia were annotated independently by 2 expert endoscopists and divided into a training set, a validation set, and a test set in a ratio of 6:2:2 for training the model. Results Among the 260 lesions identified in 216 patients, 2220 images from narrow band imaging (NBI) and 2144 images from white light imaging (WLI) were selected. For segmentation, the average intersection-over-union (IoU) value exceeded 70%. For classification, the model was able to classify the surgical group (SG) by laryngoscope with a sensitivity of 0.93 and specificity of 0.94 in WLI, and a sensitivity of 0.99 and specificity of 0.97 in NBI. Moreover, this model achieved a mean average precision (mAP) of 0.81 in WLI and 0.92 in NBI with an IoU> 0.5. Conclusions The study found that a model developed by applying DL with CNNs to laryngoscopic imaging results in high sensitivity, specificity, and mAP for automated segmentation and classification of vocal cord leukoplakia. This finding shows promise for the application of DL with CNNs in assisting in accurate diagnosis of vocal cord leukoplakia from WLI and NBI.

ABSTRACT

Objectives

The study was to apply deep learning (DL) with convolutional neural networks (CNNs) to laryngoscopic imaging for assisting in real-time automated segmentation and classification of vocal cord leukoplakia.

Methods

This was a single-center retrospective diagnostic study included 216 patients who underwent laryngoscope and pathological examination from October 1, 2018 through October 1, 2019. Lesions were classified as nonsurgical group (NSG) and surgical group (SG) according to pathology. All selected images of vocal cord leukoplakia were annotated independently by 2 expert endoscopists and divided into a training set, a validation set, and a test set in a ratio of 6:2:2 for training the model.

Results

Among the 260 lesions identified in 216 patients, 2220 images from narrow band imaging (NBI) and 2144 images from white light imaging (WLI) were selected. For segmentation, the average intersection-over-union

(IoU) value exceeded 70%. For classification, the model was able to classify the surgical group (SG) by laryngoscope with a sensitivity of 0.93 and specificity of 0.94 in WLI, and a sensitivity of 0.99 and specificity of 0.97 in NBI. Moreover, this model achieved a mean average precision (mAP) of 0.81 in WLI and 0.92 in NBI with an IoU>0.5.

Conclusions

The study found that a model developed by applying DL with CNNs to laryngoscopic imaging results in high sensitivity, specificity, and mAP for automated segmentation and classification of vocal cord leukoplakia. This finding shows promise for the application of DL with CNNs in assisting in accurate diagnosis of vocal cord leukoplakia from WLI and NBI.

Key words

Vocal Cord Leukoplakia; Laryngoscopic Imaging; Convolutional Neural Networks; Image Segmentation; Image Classification;

Key points

Early diagnosis and treatment of vocal cord leukoplakia may prevent progression and malignancy.

Lesions of vocal cord leukoplakia were classified as nonsurgical group (NSG) and surgical group (SG) using pathology as a gold standard.

We applied deep learning (DL) with convolutional neural networks (CNNs) to segment and classify narrow band imaging (NBI) and white light imaging (WLI) of vocal cord leukoplakia .

The DL model could detect the lesions autonomously with average intersection-over-union (IoU) value exceeded 70%. For classification, the model was able to classify the surgical group (SG) by laryngoscope with a sensitivity of 0.93 and specificity of 0.94 in WLI, and a sensitivity of 0.99 and specificity of 0.97 in NBI. Moreover, this model achieved a mean average precision (mAP) of 0.81 in WLI and 0.92 in NBI with an IoU > 0.5.

The novel DL model shows promise in assisting in accurate diagnosis of vocal cord leukoplakia from WLI and NBI.

1 Introduction

Vocal cord leukoplakia is a clinical descriptor for the identification of a white plaque or patch on the vocal cords upon macroscopic examination without consideration of its histological features or prognosis. Pathologically, vocal cord leukoplakia may be accompanied by squamous hyperplasia, epithelial dysplasia, or carcinoma; and thus, it is considered a precancerous lesion within the spectrum of transformation of the laryngeal epithelium toward malignancy[1].

Laryngeal cancer is typically preceded by dysplasia, and the degree of dysplasia is directly proportional to the rate of malignant transformation of vocal cord leukoplakia[2]. While the rate of malignant transformation varies widely with estimates as low as 1.7% and as high as 46.3%[3], early diagnosis and treatment of vocal cord leukoplakia may prevent progression and malignancy[4]. The 2017 World Health Organization Classification of Head and Neck Tumors proposed a two-tier classification system for dysplasia, with reasonably clear histopathological criteria for the two groups: 1) low-grade (LG) dysplasia including squamous hyperplasia and mild dysplasia, and 2) high-grade (HG) dysplasia including moderate and severe dysplasia and carcinoma in situ[5,6]. In response to this classification, some otolaryngologists proposed that patients in the LG group of vocal cord leukoplakia with a low malignancy risk would generally require a conservative treatment or watch-and-wait policy, whereas patients in the HG group of vocal cord leukoplakia would demand both surgical treatment and close follow-up to monitor possible progression to a more aggressive pathology[7]. However, a clinical challenge in managing vocal cord leukoplakia is to assess the potential malignant transformation of the lesion, and to accordingly establish the optimal therapeutic schedule[8].

Laryngoscopy is the most important examination method for detecting vocal cord leukoplakia, but to date laryngoscopy alone cannot determine the degree or scope of vocal cord leukoplakia without biopsy. Some otolaryngologists and pathologists therefore recommend a combination of laryngoscope and random 3-spot biopsy specimens to enable early detection and follow-up. However, this procedure is invasive, time-consuming, and difficult to comply with[9]. Moreover, preoperative biopsy under laryngoscopy is unlikely to fully agree with postoperative pathology results. This discrepancy often leads to overtreatment or undertreatment even for experienced endoscopists. Another challenge in clinical practice is that not all cases of vocal cord leukoplakia need laryngoscopy with histological examination, and there is difficulty in deciding which cases indicate biopsy.

Considering the above controversies and uncertainties, further improvements in the detection of vocal cord leukoplakia, possibly using new techniques, is highly urgent for its clinical management. Currently, image-enhanced endoscopy (IEE), such as contact endoscopy (CE)[10] and narrow band imaging (NBI)[9], in addition to white light imaging (WLI) has been used for accurate diagnosis of laryngeal lesions. However, the observation procedure is time-consuming and may be biased based on the observers' experience.

Artificial intelligence (AI) using deep learning (DL) with convolutional neural networks (CNNs) has recently emerged and showed inspiring results as a method for the detection of gastrointestinal cancers[11-13]. Moreover, one single-institution study showed that an AI system for detecting pharyngeal cancers had promising performance with high sensitivity and acceptable specificity[14]. However, no study to date has applied AI for simultaneous segmentation and classification of vocal cord leukoplakia. We therefore developed an AI system that applies DL with CNNs to assist in real-time automated diagnosis of vocal cord leukoplakia and uses pathological diagnosis as the gold standard.

2 Methods

2.1 Selection of images

We retrospectively reviewed all laryngoscopic images from patients with vocal cord leukoplakia at our institution between October 2018 and October 2020. The laryngoscope used during this period was the Olympus BF-H290 (Olympus Medical Systems Corp, Tokyo, Japan). The images were originally taken by senior expert endoscopists and categorized by 2 imaging techniques: WLI and NBI. According to surgical treatment needed, we classified the images of vocal cord leukoplakia into two types based on the pathology[7, 15] (**Figure 1**): 1) a nonsurgical group (NSG) including LG dysplasia, and 2) a surgical group (SG) including HG dysplasia and invasive carcinoma. When multiple pathological types appeared in a lesion, senior type was determined.

2.2 Annotation of images

All selected images of vocal cord leukoplakia were independently annotated by 2 senior expert endoscopists with at least 10 years of laryngoscopy experience; they did not participate in manipulating the laryngoscope that took the targeted images, but were not blinded to laryngoscopic findings and pathology. The boundaries of lesions in the image were annotated with the LabelMe application (*https://github.com/wkentaro/labelme*) and used to represent the actual lesion area within the image.

2.3 Model Architecture

Mask Region-based Convolutional Neural Network (Mask R-CNN)[15] is a 2-stage object detection method: the first stage, called Region Proposal Network (RPN), proposes candidate object bounding boxes before being followed by R-CNN and a semantic segmentation model (MASK). We used Mask R-CNN to perform the image segmentation and classification. First, the input image was rescaled to a size of 512×512px and input into the region proposal network with ResNet-50 as the backbone network. With several targeted regions being generated by the region proposal network, the network cropped the corresponding area of each Region of Interest (ROI) in the feature map. Then, we performed the RoIAlign[16] operation on the cropped area, input the aligned results into the Full Convolution Network (FCN)[17] segmentation sub-network and the classification sub-network respectively, and finally output the results (**Figure 2**).

2.4 Development of the model

We fine-tuned the model of Mask R-CNN implemented by Matterport, Inc (Mountain View, CA, USA)(https://github.com/matterport/Mask-RCNN) with a ResNet-50 backbone[18]. The DL model could learn the laryngoscopic images of vocal cord leukoplakia that contained the pre-labeled regions, and then detect the lesions autonomously.

Since the training of the model requires adequate images, data augmentation were used to expand the images in the training set to 25000+ images. The specific methods included: 1) horizontal flipping; 2) vertical flipping; 3) picture rotation (45°, 90°, 135°, 180°, 225°, 270°, 315°). These methods were implemented using the imgaug library (*https://github.com/aleju/imgaug*).

The weights of the model were initialized using a pre-trained model, which was the same network as those trained with the Microsoft Common Objects in Context(MS-COCO) dataset[19]. The model was developed using Python 3.5 and a TensorFlow neural network framework. Training, validation, and testing were conducted by two Nvidia GeForce GTX 1080 GPUs with 8 GB of memory each and an Intel (R) i7-5930k 3.50 GHz CPU with a CentOS 7 operating system.

2.5 Evaluation of the Model

In each epoch, the loss value of the model on the training set and the test set were obtained simultaneously. Overfitting was determined when the training loss value decreased and the validation loss value increased. We then tested the performance of the model in the test set. Each lesion of vocal cord leukoplakia was considered as a unit for the evaluation of the performance. The subsequent DL workflow used by MaskR-CNN is demonstrated in **Figure 2.**

2.6 Statistical analysis

We used quantitative metrics to evaluate the performance of the proposed method. For the segmentation result, we used the intersection-over-union (IoU) value. For the classification result, we chose the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). For the final accuracy, we used the mean average precision (mAP). We applied a minimum threshold of IoU at 0.5 to calculate mAP, which measures the overlap between prediction and ground truth[20,21]. All statistical analyses were performed using R software version 3.6.

2.7 Ethics

The study was approved by the Ethics Committee of Blinded for review (no.:2020-978).

3 Results

124 patients with HG dysplasia and invasive carcinoma, and 92 patients with LG dysplasia were selected for this study. Because a patient may display multiple lesions, there were 168 lesions among patients with HG dysplasia and invasive carcinoma, and 92 lesions among patients with LG dysplasia. For the 2220 images taken in NBI mode, 1104 were classified as NSG, while 1100 were classified as SG. For the 2144 images in WLI mode, 1004 were classified as NSG while 1140 were SG. All images were divided into a training set, a validation set, and a test set in a 6:2:2 ratio. In NBI mode, there were 1204 images in the training set, 508 images in the validation set, and 508 images in the test set. In WLI mode, there were 1160 images in the training set, 492 images in the validation set, and 492 images in the test set .

We evaluated the performance of the DL model by the segmentation and classification of images in NBI and WLI modes. Model segmentation was compared against segmentation performed by senior expert endoscopists with at least 10 years of laryngoscopy experience. Model classification as SG or NSG was compared against classification by clinical decision using pathology as a gold standard.

For segmentation, the average IoU value exceeded 70% in WLI and NBI modes (**Table 1**). The DL model can detect 87% of vocal cord leukoplakia in WLI mode and 92% of vocal cord leukoplakia in NBI mode with an IoU$>$ 0.5. With an increased IoU criterion of $>$ 0.7, the detection rate in the two modes remained acceptable with a greater than 60% detection rate (**Table S1**). The partial segmentation results using the learned model in WLI and NBI modes are shown in **Figure 3** .

To measure the performance of pure classification of target regions, we did not initially set an IoU threshold. We abandoned images with an IoU less than or equal to zero (only 1 image was disqualified) (**Table 2**). The model's binary classification of WLI test set (616 images) into SG and NSG demonstrated a sensitivity of 93% (95% CI, 88%-98%) and a specificity of 94% (95% CI, 88%-100%). Impressively, the model's binary classification of NBI test set images (620 images) demonstrated a higher sensitivity of 99% (95% CI , 97%-101%) and a higher specificity of 97% (95% CI, 93%-101%). The model's PPV for WLI and NBI were 97% (95% CI, 94%-100%) and 98% (95% CI, 95%-101%), respectively. The model's NPV for WLI and NBI were 87% (95% CI, 78%-96%) and 98% (95% CI, 95%-101%), respectively.

The model manifested optimal performance in the segmentation and classification. However, the accuracy of the model depends on two factors: the IoU of the segmented lesions compared to manual annotation needs to be greater than the preset IoU> 0.5 criterion, and simultaneously the classification of the segmented lesion area must be accurate. Thus, we calculated the mAP of the model at different IoUs (the minimum threshold of IoU is 0.5) (**Table 3**). In our test set with an IoU>0.5, the mAP for WLI and NBI was 0.81 and 0.92, respectively. With an IoU>0.7, the mAP for WLI and NBI was acceptable.

Our video model was capable of processing at least 25 frames per second with a latency period of less than 40 ms in real-time video analysis. Video clips demonstrating classification of NSG and SG are shown for WLI in **Video 1** and NBI in **Video 2** .

4 Discussion

The management of vocal cord leukoplakia remains a challenge despite the use of IEE techniques, such as CE and NBI, for accurate diagnosis of laryngeal lesions. While surgical resection will provide a final diagnosis, LG dysplasia of vocal cord leukoplakia may not go on to be malignant, thereby resulting in potentially unnecessary surgeries. Meanwhile, the optimal opportunity for surgery may be missed if HG dysplasia and invasive carcinoma of vocal cord leukoplakia is misdiagnosed. Treatment stratification by combining laryngoscopic imaging and AI can help to alleviate this management dilemma. To the best of our knowledge, this is the first study that has applied deep learning with Mask R-CNN to laryngoscope WLI and NBI for the automated segmentation and classification of vocal cord leukoplakia.

The use of deep learning for the detection of gastrointestinal lesions has rapidly developed and has made remarkable progress in recent years[12, 22]. Presently, some studies have reported using computer-aided detection in the segmentation or classification of laryngoscopic images. In 2015, H. Irem Turkmen*et al* .[23] classified vocal fold disorders into five categories using manual extraction and Histogram of Oriented Gradients (HOG) descriptors. However, one flaw in the study is that the training set was subjective and pathology was not the gold standard of classification. Bin Ji *et al* .[24] reported a multi-scale recurrent fully convolution neural network (CNN) for laryngeal leukoplakia segmentation. Despite favorable results, their datasets included only static images taken by WLI under optimal conditions whereas NBI is crucial for the differentiation of benign from malignant lesions. In this study, we included images of WLI and NBI in the datasets, considering that the model would be used in various modalities and applied to different hospitals. Furthermore, real-time video detection is more demanding than static images because of complex conditions such as reflect light, blurring, and airway secretions. As seen in Video 1 and Video 2, our model in this study displays the extent and subtype of vocal cord leukoplakia in real-time without pausing. Encouragingly, our DL model also demonstrated a high sensitivity (93% for WLI and 99% for NBI) and specificity (94% for WLI and 97% for NBI) per lesion for binary classification into a surgical group versus a non-surgical group. While Kono M. *et al* .[14] used DL with CNNs for the real-time diagnosis of pharyngeal cancers with a sensitivity of 92%, the specificity and accuracy were 47% and 66% respectively, significantly lower than our dataset. Meanwhile, our model also detected lesions correctly with a high mAP (0.81 for WL and 0.92 for NBI, IoU>0.5). In contrast, Rintaro Hashimoto *et al* .[23] reported a study of CNNs with an IoU threshold at 0.3 for real-time detection of early esophageal neoplasia in Barrett's esophagus, and the overall mAP was 0.7533 and mAP for

NBI was 0.802, significantly lower than our dataset. More importantly, however, we combined pathological diagnosis and clinical decisions to a grouped dataset, which gives a more realistic assessment and would be conducive to clinical promotion in the future.

It is possible to implement our proposed model in an embedded decision support system for identifying patients for whom directly proceeding to surgical treatment might be advantageous. Taken together, the outcomes of this study showed promise for efficient management of vocal cord leukoplakia. First, real-time segmentation and classification would greatly shorten laryngoscopic operation time for endoscopists, especially if inexperienced. Secondly, this model can aid otolaryngologists in decision-making. Thirdly and most importantly for patients, this approach would obviate the need for unnecessary invasive procedures such as biopsy as well as mitigate medical expenses.

However, there are also some limitations to this Mask R-CNN system. First, all tested laryngoscopic images were retrospectively taken from a single-center and obtained from the same video system. A second caveat is that multiple images were extracted from a single patient's laryngoscope, so a learner bias was possible if images from the same patients existed in both the training and test sets. A third limitation is that this Mask R-CNN system could not completely exclude the influence of airway secretions and reflected light, which were major causes of pseudo-positive cases. We believe these limitations will be overcome in the future by including datasets from a multi-center setting with different hospitals and various laryngoscopic systems.

5 Conclusions

This study applied deep learning with Mask R-CNN to laryngoscopic images in WLI and NBI mode for the automated segmentation and classification of vocal cord leukoplakia with high sensitivity, specificity, and mAP. Mask R-CNN has promising potential to assist otolaryngologists in clinical treatment decisions on vocal cord leukoplakia.

References

[1] Lee DH, Yoon TM, Lee JK, Lim SC. Predictive factors of recurrence and malignant transformation in vocal cord leukoplakia. Eur Arch Otorhinolaryngol. 2015; 272(7): 1719-1724. doi:10.1007/s00405-015-3587-8.

[2] Isenberg JS, Crozier DL, Dailey SH. Institutional and comprehensive review of laryngeal leukoplakia. Ann OtolRhinolLaryngol. 2008; 117(1): 74-79. doi:10.1177/000348940811700114.

[3] Karatayli-Ozgursoy S, Pacheco-Lopez P, Hillel AT, Best SR, Bishop JA, Akst LM. Laryngeal dysplasia, demographics, and treatment: a single-institution, 20-year review. JAMA Otolaryngol Head Neck Surg. 2015; 141(4): 313-318. doi:10.1001/jamaoto.2014.3736.

[4] Weller MD, Nankivell PC, McConkey C, Paleri V, Mehanna HM. The risk and interval to malignancy of patients with laryngeal dysplasia; a systematic review of case series and meta-analysis. Clin Otolaryngol. 2010; 35(5): 364-372. doi:10.1111/j.1749-4486.2010.02181.x.

[5] Gale N, Poljak M, Zidar N. Update from the 4th Edition of the World Health Organization Classification of Head and Neck Tumours: What is New in the 2017 WHO Blue Book for Tumours of the Hypopharynx, Larynx, Trachea and Parapharyngeal Space. Head Neck Pathol. 2017; 11(1): 23-32. doi:10.1007/s12105-017-0788-z.

[6] Gale N, Cardesa A, Hernandez-Prera JC, Slootweg PJ, Wenig BM, Zidar N. Laryngeal Dysplasia: Persisting Dilemmas, Disagreements and Unsolved Problems-A Short Review. Head Neck Pathol. 2020; Dec;144(4) . doi:10.1007/s12105-020-01149-9.

[7] Ferlito A, Devaney KO, Woolgar JA, et al. Squamous epithelial changes of the larynx: diagnosis and therapy. Head Neck. 2012; 34(12): 1810-1816. doi:10.1002/hed.21862.

[8] Cui W, Xu W, Yang Q, Hu R. Clinicopathological parameters associated with histological background and recurrence after surgical intervention of vocal cord leukoplakia. Medicine (Baltimore). 2017; 96(22): e7033. doi:10.1097/MD.0000000000007033.

[9] Paleri V, Sawant R, Mehanna H, Ainsworth H, Stocken D. Laryngeal dysplasia and narrow band imaging: Secondary analysis of published data supports the role in patient follow-up. Clin Otolaryngol. 2018; 43(6): 1439-1442. doi:10.1111/coa.13182.

[10] Davaris N, Lux A, Esmaeili N, et al. Evaluation of Vascular Patterns Using Contact Endoscopy and Narrow-Band Imaging (CE-NBI) for the Diagnosis of Vocal Fold Malignancy. Cancers (Basel). 2020; 12(1). doi:10.3390/cancers12010248.

[11] de Groof AJ, Struyvenberg MR, van der Putten J, et al. Deep-Learning System Detects Neoplasia in Patients With Barrett's Esophagus With Higher Accuracy Than Endoscopists in a Multistep Training and Validation Study With Benchmarking. Gastroenterology. 2020; 158(4): 915-929.e4. doi:10.1053/j.gastro.2019.11.030.

[12] Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol. 2020; 5(4): 343-351. doi:10.1016/S2468-1253(19)30411-X.

[13] Hashimoto R, Requa J, Dao T, et al. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). GastrointestEndosc. 2020; 91(6): 1264-1271.e1. doi:10.1016/j.gie.2019.12.049.

[14] Kono M, Ishihara R, Kato Y, et al. Diagnosis of pharyngeal cancer on endoscopic video images by Mask region-based convolutional neural network. Dig Endosc. 2020; Jul 26. doi:10.1111/den.13800.

[15] Mehlum CS, Larsen SR, Kiss K, et al. Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. Laryngoscope. 2018; 128(10): 2375-2379. doi:10.1002/lary.27228.

[16] He K, Gkioxari G, Dollár P, et al. Mask r-cnn (Proceedings of the IEEE international conference on computer vision). 2017: 2961-2969.

[17] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation (Proceedings of the IEEE conference on computer vision and pattern recognition). 2015: 3431-3440.

[18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition (Proceedings of the IEEE conference on computer vision and pattern recognition). 2016: 770-778.

[19] Irem Turkmen H, ElifKarsligil M, Kocak I. Classification of laryngeal disorders based on shape and vascular defects of vocal folds. Comput Biol Med. 2015. 62: 76-85. doi:10.1016/j.compbiomed.2015.02.001.

[20] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation (Proceedings of the IEEE conference on computer vision and pattern recognition). 2014: 580-587.

[21] Gamage H, Wijesinghe W, Perera I. Instance-based segmentation for boundary detection of neuropathic ulcers through Mask-RCNN (International Conference on Artificial Neural Networks). Springer, Cham, 2019: 511-522.

[22] Sharma P, Pante A, Gross SA. Artificial intelligence in endoscopy. GastrointestEndosc. 2020; 91(4): 925-931. doi:10.1016/j.gie.2019.12.018.

[23] Irem Turkmen H, ElifKarsligil M, Kocak I. Classification of laryngeal disorders based on shape and vascular defects of vocal folds. Comput Biol Med. 2015; 62: 76-85. doi:10.1016/j.compbiomed.2015.02.001.

[24] BinJi, JianjunRen, XiujuanZheng, CongTan, RongJi, YuZhao, KaiLiu. A multi-scale recurrent fully convolution neural network for laryngeal leukoplakia segmentation. Biomed Signal Process Control. 2019; 59. doi:10.1016/j.bspc.2020.101913.

Tables

Table 1. Statistical data of the IoU in the test set

Group	WLI	WLI	NBI	NBI
	N	Mean (SD)	N	Mean (SD)
SG	404	0.67 (0.19)	372	0.72 (0.13)
NSG	212	0.75 (0.15)	248	0.78 (0.09)
All	616	0.70 (0.18)	620	0.74 (0.12)

IoU, intersection-over-union; SD, standard deviation; NBI, narrow band imaging; WLI, white light imaging; NSG, nonsurgical group; SG, surgical group.

Table 2. Results of AI binary diagnosis: per- lesion analysis in the test set

Group	Group	White light	White light	White light	NBI	NBI	NBI
		SG	NSG	All	SG	NSG	All
Pathology	SG	392	12	404	364	8	372
	NSG	28	184	212	4	244	248
	All	420	196	616	368	252	620

AI, Artificial Intelligence; SG, Surgery Group; NSG, Non-surgery Group; NBI, narrow band imaging; WLI, white light imaging.

Table 3. MAP at different IoUs in the test set

Group	IoU	IoU	IoU	IoU
	>0.5	>0.6	>0.7	>0.8
WLI	0.81	0.75	0.60	0.31
NBI	0.92	0.87	0.74	0.35

IoU, Intersection over Union; MAP, mean average precision; NBI, narrow band imaging; WLI, white light imaging.

Figure legend

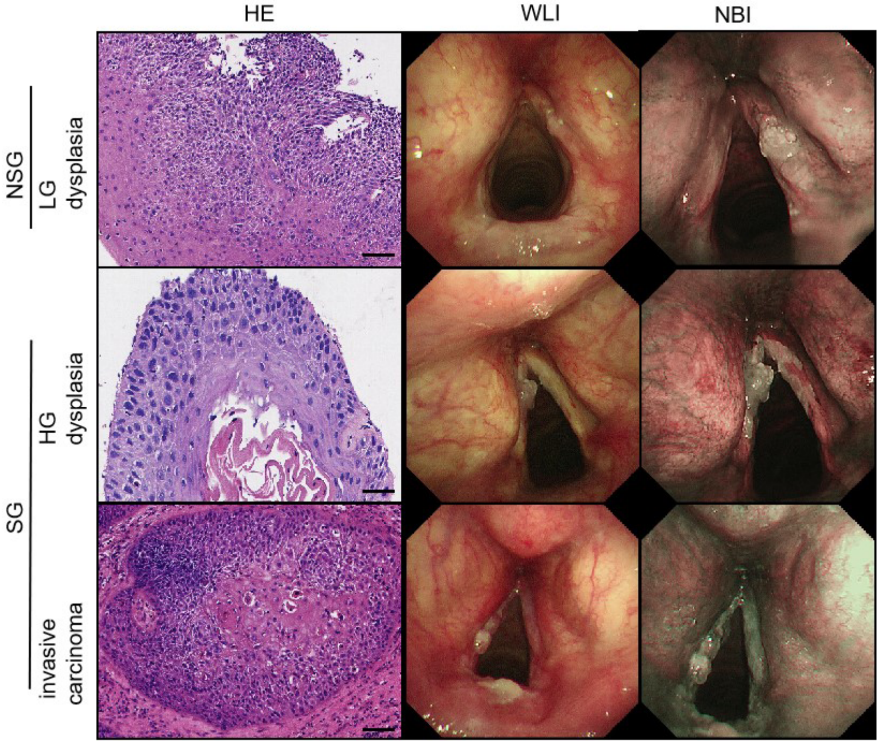
Figure 1 Classification for the laryngoscopic images of vocal cord leukoplakia, WLI, white light imaging; NBI, narrow band imaging; LG, low-grade; HG, high-grade; NSG, nonsurgical group; SG, surgical group. Scale bar: 50 μm.

Figure 2 Work flow of the model architecture.

Figure 3 Segmentation results of model prediction in WLI mode and in NBI mode.

Video 1 Classification of NSG or SG for WLI.

Video 2 Classification of NSG or SG for NBI.


HE
WLI
NBI
NSG
LG
dysplasia
SG
HG
dysplasia
invasive
carcinoma

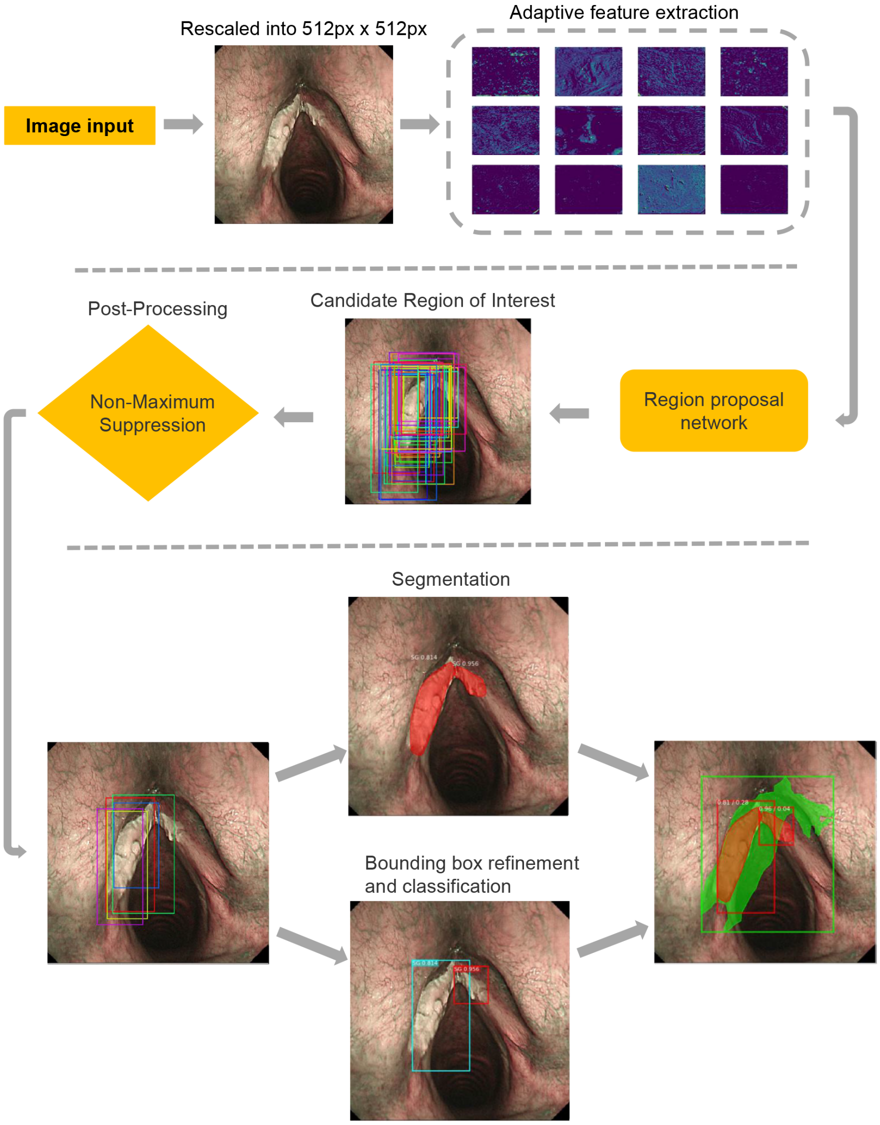
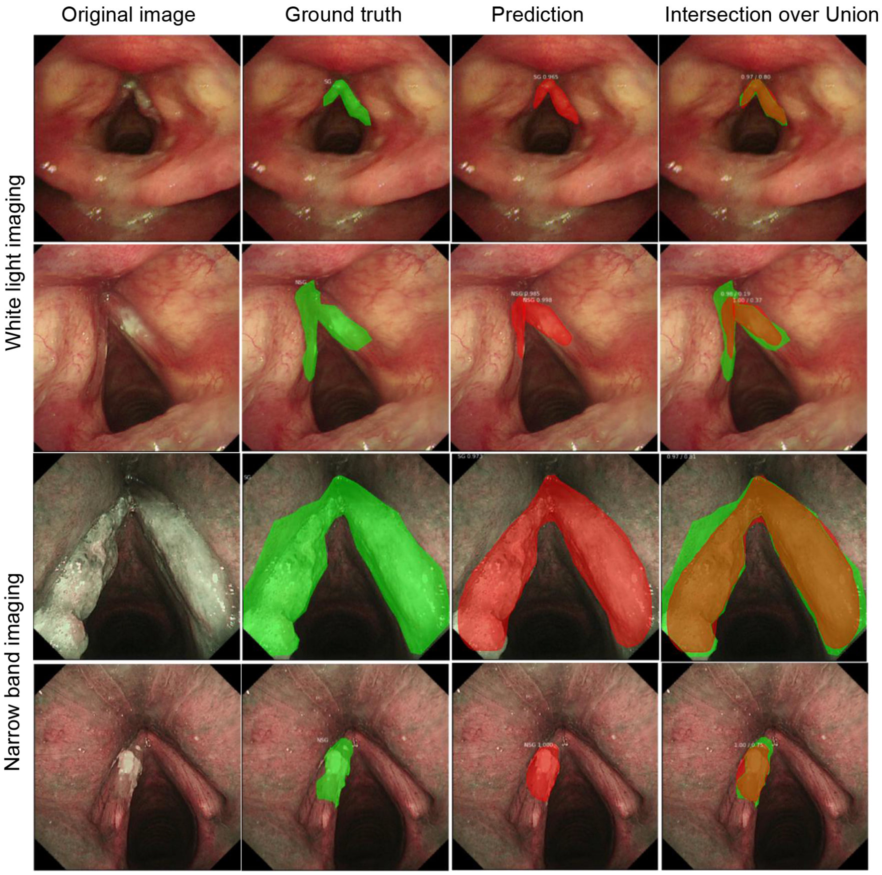


Image input
Rescaled into 512px x 512px
Adaptive feature extraction
Region proposal network
Candidate Region of Interest
Post-Processing
Non-Maximum Suppression
Segmentation
SG 0.814
SG 0.956
Bounding box refinement and classification


Original image
Ground truth
Prediction
Intersection over Union
White light imaging
Narrow band imaging
SG
SG 0.965
0.97 / 0.80
NSG
NSG 0.985
NSG 0.998
0.98 / 0.19
1.00 / 0.37
SG
SG 0.973
0.97 / 0.81
NSG
NSG 1.000
1.00 / 0.75