

## Article

# Development of a Deep Learning-Based Epiglottis Obstruction Ratio Calculation System

Hsing-Hao Su <sup>1,2,3</sup>  and Chuan-Pin Lu <sup>4,\*</sup>

<sup>1</sup> Department of Otorhinolaryngology-Head and Neck Surgery, Kaohsiung Veterans General Hospital, Kaohsiung 81362, Taiwan; shsu@vghks.gov.tw

<sup>2</sup> Department of Physical Therapy, Shu-Zen Junior College of Medicine and Management, Kaohsiung 82144, Taiwan

<sup>3</sup> Department of Pharmacy and Master Program, College of Pharmacy & Health Care, Tajen University, Pingtung 90741, Taiwan

<sup>4</sup> Department of Information and Communication Engineering, Chaoyang University of Technology, Taichung 41349, Taiwan

\* Correspondence: cplu@cyut.edu.tw; Tel.: +886-4-2332-3000 (ext. 4680); Fax: +886-4-2332-9898

**Abstract:** Surgeons determine the treatment method for patients with epiglottis obstruction based on its severity, often by estimating the obstruction severity (using three obstruction degrees) from the examination of drug-induced sleep endoscopy images. However, the use of obstruction degrees is inadequate and fails to correspond to changes in respiratory airflow. Current artificial intelligence image technologies can effectively address this issue. To enhance the accuracy of epiglottis obstruction assessment and replace obstruction degrees with obstruction ratios, this study developed a computer vision system with a deep learning-based method for calculating epiglottis obstruction ratios. The system employs a convolutional neural network, the YOLOv4 model, for epiglottis cartilage localization, a color quantization method to transform pixels into regions, and a region puzzle algorithm to calculate the range of a patient's epiglottis airway. This information is then utilized to compute the obstruction ratio of the patient's epiglottis site. Additionally, this system integrates web-based and PC-based programming technologies to realize its functionalities. Through experimental validation, this system was found to autonomously calculate obstruction ratios with a precision of 0.1% (ranging from 0% to 100%). It presents epiglottis obstruction levels as continuous data, providing crucial diagnostic insight for surgeons to assess the severity of epiglottis obstruction in patients.



**Citation:** Su, H.-H.; Lu, C.-P.

Development of a Deep Learning-Based Epiglottis Obstruction Ratio Calculation System. *Sensors* **2023**, *23*, 7669. <https://doi.org/10.3390/s23187669>

Academic Editor: Michael Beigl

Received: 27 July 2023

Revised: 31 August 2023

Accepted: 1 September 2023

Published: 5 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** obstructive sleep apnea; epiglottis obstruction; deep learning; computer vision; region puzzle algorithm; drug-induced sleep endoscopy

## 1. Introduction

According to existing research, 2 out of 10 individuals in Taiwan [1] may be suffering from obstructive sleep apnea (OSA) without knowing it, and at least 800,000 people may be suffering from sleep apnea. OSA symptoms include excessive daytime sleepiness, fatigue, and even lethargy [2]. This condition can also lead to depression in the long run and increase the risk and severity of chronic diseases [3]. To check a patient's condition and provide the best treatment, it is necessary to know the level and severity of the collapse/obstruction of the patient's upper airway. Surgeons use endoscopy to examine the soft palate in the supra-glottic area, observe the position of the collapsed soft tissue around the respiratory tract, roughly judge the severity of the collapse, and determine the surgical correction needed for their patient's condition and the degree of correction. This method is known as drug-induced sleep endoscopy (DISE) [4].

The causes of OSA symptoms are intermittent relaxation of patients' throat muscles, tonsil hypertrophy, redundancy of the soft palate and uvula, anterior-to-posterior collapse of the tongue base, and epiglottic collapse [5–7]. These factors can cause airway

obstruction during deep sleep. Common treatments for OSA include continuous positive airway pressure, oral appliance therapy, behavioral modification, and surgical procedures, among others [8]. The treatment method is determined based on the type and degree of obstruction in the affected site. For mild cases, non-invasive treatments should be the first-line management, while invasive surgical treatments are the primary method for severe cases [8–10].

The severity of OSA is represented by the apnea–hypopnea index [11], which is defined by the number of apnea and hypopnea events per hour of sleep. Hypopnea refers to a partial airway blockage (a reduction of  $\geq 30\%$  in airflow that lasts for more than 10 s) where blood oxygen saturation or sleep arousal decreases by at least 3%. The current commonly used three obstruction degrees [12] (0–49% collapse (degree 0: no obstruction), 50–75% collapse (degree 1: partial obstruction), and 76–100% collapse (degree 2: complete obstruction)) do not intuitively correspond to the reduction in airflow. Moreover, the difference between 49% “no obstruction” and 50% “partial obstruction” is not significant (the same applies to 75% “partial obstruction” and 76% “complete obstruction”). For instance, if the actual collapse values are “30%, 49%, 50%, 75%, 76%, 100%”, then it would become “0, 0, 1, 1, 2, 2” when represented by the three obstruction degrees. Additionally, if the actual collapses are “50%, 60%, 70%, 75%”, then the obstruction degrees would all be “1,1,1,1”. Collapse values (obstruction ratios) can directly correspond to possible airflow changes, but obstruction degrees are not easily mapped precisely to airflow changes.

Medical information systems based on artificial intelligence have been extensively applied in the medical field in recent years. In situations such as the COVID-19 pandemic, AI healthcare systems can generate multidimensional strategies according to the limited pharmacological resources available to control and minimize casualties from major outbreaks [13]. This is because AI algorithms can generate specific policies based on the importance assigned to casualties or different policies in each sub-model. In recent years, deep learning has been regarded as a significant benchmark in the field of artificial intelligence. It has achieved remarkable results in various complex cognitive tasks, particularly surpassing human performance in tasks such as image matching. Deep learning neural network algorithms that were developed based on the core architecture of convolutional neural networks (CNNs) [14,15] include AlexNet [16], GoogLeNet [17], SSD [18], and YOLOv1–v3 [19–21], as well as YOLOv4 and v7 [22–24]. CNNs offer a non-parametric approach to prediction and recognition. Recently, advanced parametric, multi-dimensional, constrained, and probabilistic machine learning algorithms have been developed to provide long-term, stable, and optimal predictions and strategic formulations. Currently, common programming languages that can implement CNNs include C++ and Python, with corresponding development tools such as Microsoft Visual Studio, Anaconda, and Jupyter Notebooks. EKEN introduced medical data analysis notebooks for collaborative and reproducible research across different data types in a 2020 study [25]. This information is immensely beneficial for scholars who are just starting to explore deep learning. Numerous studies on the application of information technology in OSA detection have been conducted. For example, Ferrer-Lluis et al. used smartphones and PSG signals to monitor patients’ sleep status and found a correlation between sleep posture and OSA [26]. Kou et al. proposed a deep learning-based unsupervised model that was applied to esophageal manometry to learn the distinctive patterns of clinical manifestations associated with esophageal motility disorders [27]. In 2022, Liu et al. applied snoring sounds to classify obstruction sites in OSA [28]. Snoring sounds were recorded during the DISE examination of OSA patients. Hanif et al. used a deep learning approach for the automatic scoring of drug-induced sleep endoscopy in obstructive sleep apnea [29]. In their study, the research objective was to identify the degree of obstruction in the diseased sites (velum, oropharynx, tongue base, and epiglottis). The obstruction levels were categorized into three degrees. ResNet18 was used for feature extraction of DISE video frames, and all feature vectors were concatenated and input into a bidirectional long short-term memory network for temporal analysis.

These neural networks were then used to recognize the three degrees of obstruction for each diseased site [12].

Evaluating the degree of obstruction in an affected area manually using DISE videos is not only inefficient but also lacks precision. Therefore, Hanif et al.'s study [29] utilized deep learning to automatically score the degree of OSA obstruction. Their method was based on the grading criteria introduced by Kezirian in 2011 [12], using images of three degrees of OSA obstruction as the target for deep-learning image training. This research treated images showing the three degrees of obstruction from DISE videos as three classes for training the neural network. The experimental results indicated that the average recognition rate (F1 score) for epiglottic obstruction was 65%. However, when looking only at the presence or absence of obstruction, the accuracy increased to 91%. Upon observing the confusion matrix, the recognition effectiveness for the three obstruction degrees was not good. The representation of the three obstruction degrees was neither precise nor intuitive. Compared to obstruction degrees, an obstruction ratio can offer surgeons a more accurate diagnosis. To expand the calculation of the obstruction ratio from the three degrees of obstruction to a numerical ratio, this study proposed a new method for analyzing epiglottic obstruction images to calculate the obstruction ratio and integrate it into a web system for ease of use. The method proposed in this research can intuitively and accurately calculate the obstruction ratio of an epiglottic region (ranging from 0% to 100%).

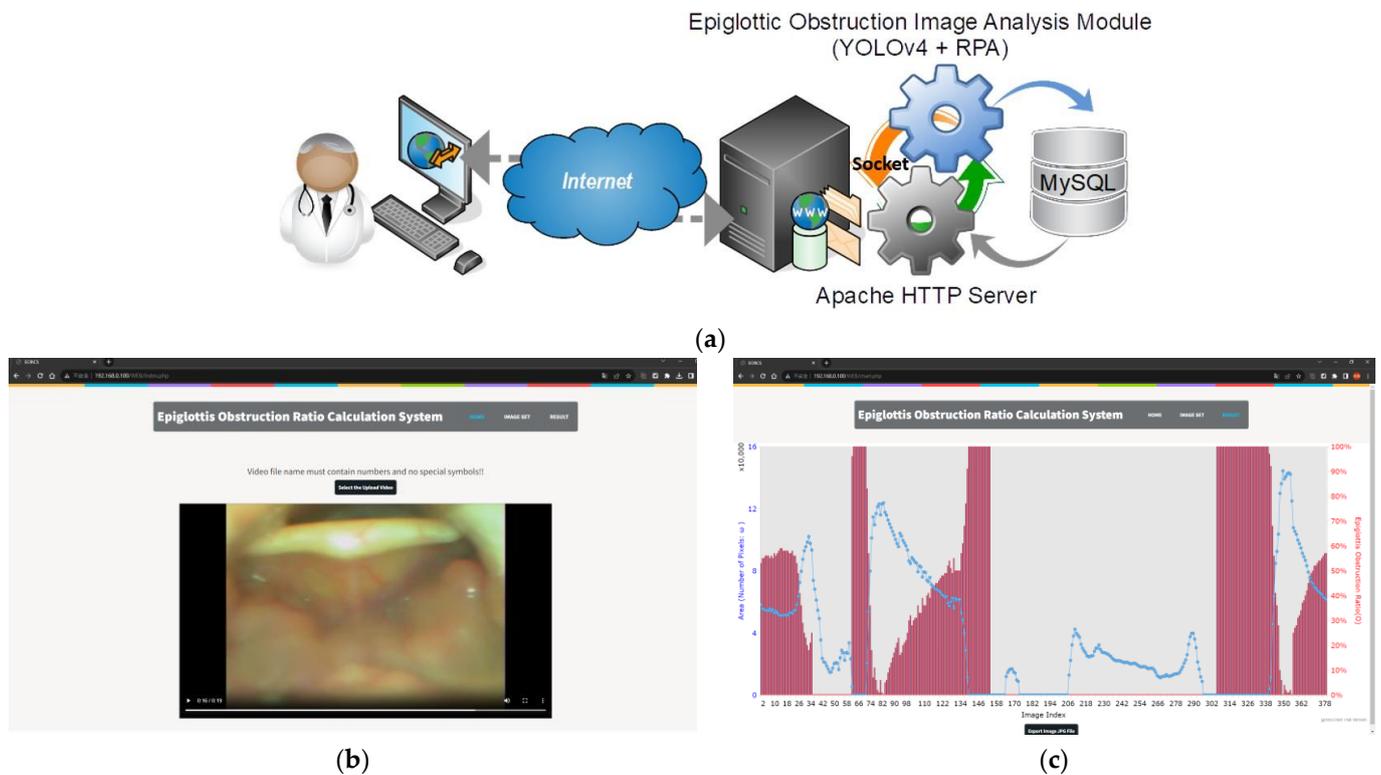
This research developed a computer vision system for the automatic calculation of the epiglottic obstruction ratio. This system utilizes a client–server network architecture and integrates both web-based and PC-based technologies, thus enabling cross-platform usability. The image analysis method used in this study involves utilizing video frames recorded during endoscopic inspections to detect the range of the airway (termed field of view (FOV); the larger the FOV, the lower the obstruction ratio), presenting the airway range data as continuous values, and calculating the time and proportion of epiglottic obstruction from different endoscopic views. The detection method for airway range uses the YOLOv4 model to locate the epiglottic cartilage in an image and then employs the median-cut color quantization method [30] to transform the image into color regions. Subsequently, the proposed region puzzle algorithm (RPA) is used to obtain the image regions of the epiglottic cartilage and the airway range (FOV). The RPA is a region-merging algorithm based on region location and geometric shape. The merging of regions is performed by taking the color regions as the basic units, and the airway region is obtained through the merging process. Finally, the obstruction ratio is calculated based on the continuous data of the FOV and the location of the endoscopy in the throat. The algorithm's effectiveness and practicality were verified by comparing its results with clinical judgment results. The detailed structure and process of this system are introduced in the following sections.

## 2. Methods

### 2.1. System Architecture and Processing Procedures

The proposed system adopts a client–server architecture and uses a web platform to provide remote access for users, thus offering the convenience of cross-platform compatibility (see Figure 1a). The front-end web interface was developed using HTML5 (see Figure 1b,c), while the back-end website system was developed to operate on Apache HTTP Server, utilizing PHP programming to implement the system workflow (including upload of endoscopic videos, transmission of socket messages to the Image Analysis Module, generation of an epiglottis obstruction rate chart, and report generation). Figure 1b shows the web interface that allows users to upload DISE videos, while Figure 1c shows the obstruction rate and the visible airway area curve obtained after the system processes the DISE videos. After a system user completes the upload of endoscopy videos and initiates the image analysis execution, the Apache Server sends a message to the Image Analysis Module (PC-based) via a socket. The Epiglottic Obstruction Image Analysis (EOIA) Module reads the video files in the directory, conducts frame capturing from the videos, and carries out the YOLOv4, RPA, and other required image processing algorithms. The execution

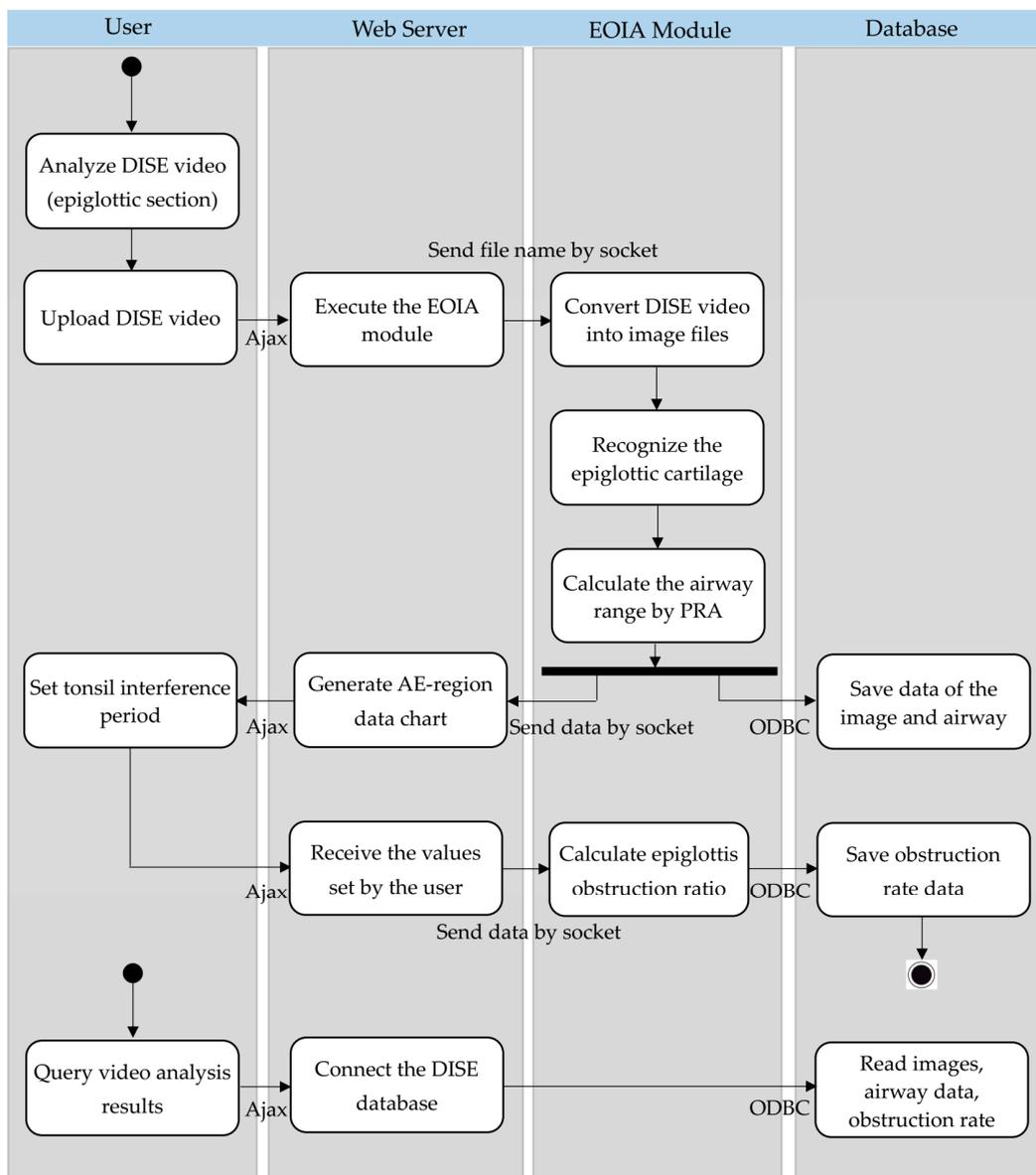
results are stored in a MySQL database. A socket message is then sent back to the Apache Server, notifying it that the image analysis task is completed. In addition, this system uses Asynchronous JavaScript and XML (Ajax) for front-end and back-end communication. Both the Apache Server and EOIA connect to the MySQL database using Open Database Connectivity (ODBC). The system's flowchart (activity diagram) is shown in Figure 2.



**Figure 1.** The proposed system: (a) architecture; (b) the user interface of the web page for the upload of DISE videos; and (c) web page showing the generated chart with epiglottis obstruction ratios.

## 2.2. FOV Detection of Epiglottis Airway

When patients with OSA, who have epiglottic collapse, fall asleep, the epiglottis is impelled to shift against the posterior pharyngeal wall by the negative pressure airflow during inspiration, and obstruction happens. In the case that the endoscopy is placed in front of the epiglottis and the airway is not completely obstructed, the obtained endoscopy image shows the arytenoid cartilage, aryepiglottic folds, hypopharynx, and vocal cords. This is called the airway's FOV. It may show the patient's lax epiglottis collapse at the entrance of the larynx or the tissues above the epiglottis collapse (especially the palatine tonsils on both sides) [31]. In such a case, the epiglottis collapse narrows or completely obstructs the airflow channel, thus reducing the airway's FOV. Based on the above phenomenon, the proposed research method aims to find the FOV of the airway. For ease of explanation, we call the FOV of the airway the "AE-region". The endoscopy, which is placed in front of a patient's epiglottis cartilage, shows the epiglottis, arytenoid, aryepiglottic fold, vocal fold, and glottis. In addition, it shows the posterior and lateral pharyngeal walls. The AE-region must exclude the epiglottis and the site of the posterior and lateral pharyngeal walls. Therefore, the proposed RPA is used to analyze the AE-region from an image and use it as the range of the airflow path.

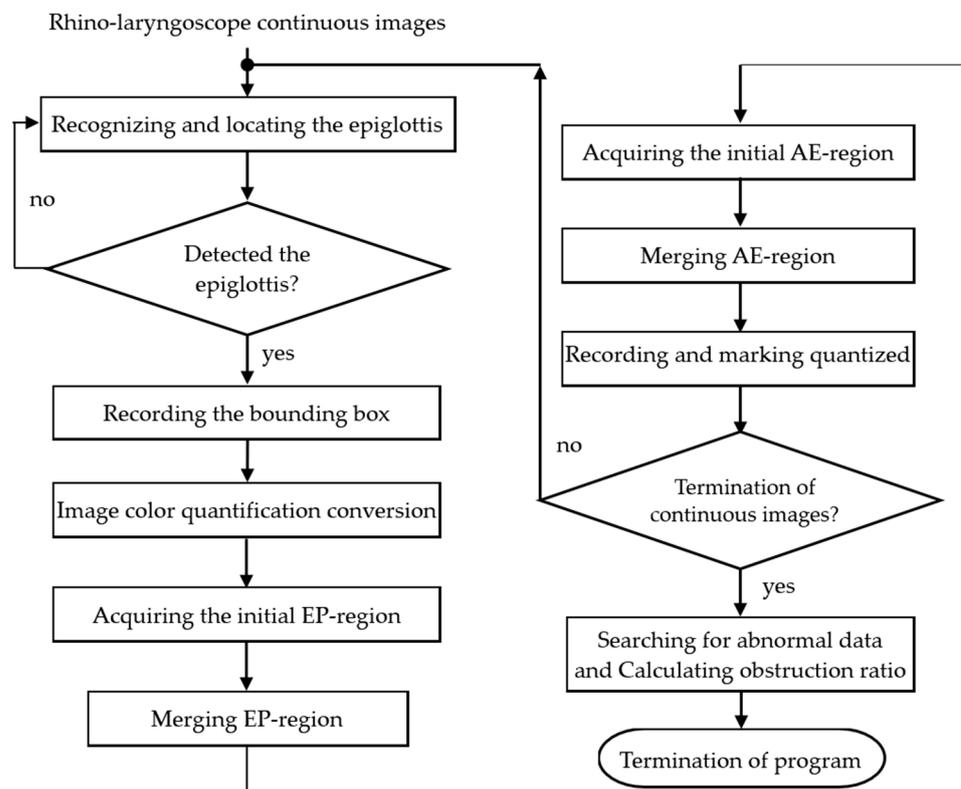


**Figure 2.** The flowchart (activity diagram) of the proposed system.

We use the YOLOv4 model to locate the epiglottis for two main reasons. The first reason is to mark the images of the endoscopy's FOV and acquire the maximum AE-region of the period as a reference value for the unobstructed airway range. The image where the epiglottic cartilage is recognized for the first time is regarded as the starting point of the new period, and the period number  $m$  is set to 1. The first image where the epiglottic cartilage cannot be recognized indicates that the FOV of the endoscopy has changed and is regarded as the last image of the period. When the epiglottis is recognized again, the period number is updated to  $m + 1$ , and another period begins. The program repeats the detection of the start and end points repeatedly. The second reason is to find the AE-region of an epiglottis image. The bounding box of the epiglottis can be obtained using the YOLOv4 model. Then, the 24-bit image is converted into color regions using the median-cut color quantization method [30], and all regions are subsequently labeled [32]. The region located in the center of the bounding box is used as the initial region of the epiglottis. Within the bounding box range, we use the RPA to merge the surrounding adjacent regions one by one. After merging several times, the epiglottis region (EP-region) is obtained. Then, we use the coordinates at a specific point in the EP-region as the basis for finding the starting region of

the AE-region, and we execute the RPA to obtain the AE-region of each image. If an image detects the epiglottis and not the AE-region, then it indicates that the epiglottis is completely obstructed. Simultaneously, we draw the AE-region area of the continuous image into a curve, search for abnormal values in the data, and mark these values (not counting the obstruction ratio). There are two ways to determine whether a value is abnormal. The first is that the detection result of the front image and the back image does not contain the epiglottis or the epiglottis is completely obstructed but the current image presents a different result. Then, the detection result of the image is regarded as abnormal. The second way is to use backward difference calculation and compare the before and after values of the data group with the threshold value. If the data group meets the conditions, the value that is higher than the threshold value ( $\epsilon$ ) is regarded as abnormal. Finally, the largest AE-region in each period is used as the denominator to calculate the obstruction ratio of the period. The calculation of the obstruction ratio ( $O_m^n$ ) (percentage) is shown in Equation (1), where  $A$  denotes the AE-region,  $|A|$  denotes the number of elements in the set (number of pixels), and  $n$  is the image index of continuous frames. The process of AE-region detection is shown in Figure 3 (the endoscopy used in this study is a rhino-laryngoscope):

$$O_m^n (\%) = 100 \times (\max\{|A_m^n|\} - |A_m^n|) / \max\{|A_m^n|\}, \quad (1)$$



**Figure 3.** The process of AE-region detection.

### 2.3. You Only Look Once Version 4 Model

The advantages of the YOLO model for object positioning have been confirmed in previous research [33–35]. Bochkovski, Wang, and MarkLiao innovatively introduced YOLOv4 in April 2020 [22]. To further enhance the performance of YOLOv4, they also proposed a strategy known as network scaling [23]. Compared to YOLOv3, significant improvements were made to the architecture of YOLOv4. Firstly, in the backbone part of the original Darknet53 network, the authors incorporated Cross Stage Partial Connections (CSP), which effectively reduced the network parameters and computational load while maintaining excellent accuracy. This new architecture is called CSPDarknet53 [36]. Secondly, in the neck part of the network, the original Feature Pyramid Network was replaced by

the PANet (Path Aggregation Network). The PANet aggregates the backbone parameters corresponding to detectors at different levels. This section also employs the Spatial Pyramid Pooling strategy, which strengthens the effect of the receptive field by separating the context. In the head part used for predicting object classes and drawing bounding boxes, the one-stage Dense Prediction strategy of YOLOv3 is retained. In addition, YOLOv4 introduces a non-monotonic and continuously differentiable Mish activation function  $M(x)$  [37] (see Equation (2)), which allows gradients to be more effectively propagated without causing a significant increase in computational cost:

$$M(x) = x \tanh(\ln(1 + e^x)) \quad (2)$$

For the bounding box regression loss part, YOLOv4 uses the Complete IoU (CIoU) Loss to replace the original mean squared error loss. The function of the IoU Loss is to calculate bounding box coordinate regression, source prediction, and class score prediction. The CIoU Loss ( $\varphi$ ) is formulated as follows [38]:

$$\varphi = 1 - \text{IoU} + \Psi, \quad \Psi = \frac{\vartheta^2(ed, ed^t)}{c^2} + \alpha v, \quad (3)$$

$$\alpha = \frac{4}{\pi^2} \left( \tan^{-1} \left( \frac{bw^t}{bh^t} \right) - \tan^{-1} \left( \frac{bw}{bh} \right) \right)^2, \quad (4)$$

$$v = \frac{\alpha}{(1 - \text{IoU}) + \alpha}, \quad (5)$$

where the parameter  $\Psi$  is the penalty term for the predicted box and the target box;  $\vartheta^2(ed, ed^t)$  is the Euclidean distance between the predicted box and the target box;  $c$  is the diagonal length of the smallest enclosing box covering the two boxes;  $\alpha$  evaluates the consistency in the aspect ratio between the predicted bounding box and the ground truth; and  $v$  serves as a trade-off parameter to balance the scale.  $\alpha$  and  $v$  are calculated using Equation (4) and Equation (5), respectively. In Equation (4),  $bw^t$  and  $bh^t$  denote the width and height of the target box, while  $bw$  and  $bh$  denote the width and height of the predicted box. When the differences in width, height, and distance between the predicted box and the target box are significant, this penalty term will increase the loss value; otherwise, it will not affect the loss value.

In terms of data augmentation strategies, YOLOv4 adopts the Mosaic method aiming at reducing GPU (Graphics Processing Unit) computational load and memory requirements. During the training process, a DropBlock strategy is used to prevent the model from overfitting, and a Modified Spatial Attention Module method is used to further enhance the network's representation power. This method applies YOLOv4 to locate the epiglottis due to its fast calculation speed, good mAP, and good identification results. The C++ version currently only supports YOLOv4, while YOLOv7 is supported in the Python (Pytorch Framework) version. As this research aims at practical application in hospitals, YOLOv4 was chosen as the object detection model for the system.

#### 2.4. Region Puzzle Algorithm

The concept of this algorithm is similar to a jigsaw puzzle. The edges of the puzzle board are more curved at the beginning, but as the puzzle gets closer to completion, the edges smooth out. The RPA is used based on the concept of puzzles, and the setting with a low color quantity quantifies the image setting. The converted color region (denoted  $R_j$ ) is combined within the analysis region and merged into the target region (denoted  $T$ , the EP- and AE-regions).  $R_j$  is known as the "puzzle pieces." In the  $k$ th merge procedure ( $k \geq 1$ ), we check whether  $R_j$  fulfills two conditions around  $T^{(k-1)}$ : (1) a single region  $R_j$  is completely adjacent to the surrounding region  $T^{(k-1)}$ , and (2)  $R_j$  does not exceed the scope

of the analysis region. If both conditions are met, then  $T^{(k-1)}$  will merge with  $R_j$  to form  $T^{(k)}$ . To speed up the merging operation, regions with holes are filled after each merge. In the  $k + 1$  merge, the puzzle pieces  $R_j$  are based on the total number of points on edge after region merging (defined as the puzzle function  $P_j^{(k+1)}$ ). This is used to determine whether  $R_j$  should include  $T^{(k)}$ . This algorithm calculates the edge line  $C_j^{(k+1)} (= Edge(R_j \cup T^{(k)}))$  of the region when merging  $R_j$  with  $T^{(k)}$ . The  $k$ th merge of the puzzle function  $P^{(k)}$  is defined as shown in Equation (6). The  $k + 1$ th merge puzzle function  $P_j^{(k+1)}$  is defined as shown in Equation (7); the edge of  $T^{(k)}$  is defined as  $C^{(k)}$ , with  $C^{(k)}$  being obtained through the gradient operation processing [14] of  $T^{(k)}$ , and  $w_j$  is the weight value of  $R_j$ . The algorithm performs the exclusive-OR (XOR) operation on the edge  $D_j$  of  $R_j$  and the edge  $C^{(k)}$  of  $T^{(k)}$ . The XOR operation can subtract the overlapped part of  $D_j$  and  $C^{(k)}$ . The result of the operation is  $P_j^{(k+1)}$ . If  $P_j^{(k+1)} > P^{(k)}$ , then  $R_j$  is not merged with  $T^{(k)}$ ; if  $P_j^{(k+1)} \leq P^{(k)}$ , then  $R_j$  is merged with  $T^{(k)}$ , and the result is updated to  $T^{(k+1)}$ ,  $C^{(k+1)}$ , and  $P^{(k+1)}$ . The merging process is repeated until  $P^{(k+i)} = P^{(k+i+1)}$ , when the best result of the puzzle of the target region  $T$  is acquired.

The puzzle function for the  $k$ th iteration:

$$P^{(k)} = |C^{(k)}| \tag{6}$$

The puzzle function for the  $k + 1$ th iteration:

$$P_j^{(k+1)} = |C_j^{(k+1)}| - w_j = |(D_j \oplus C^{(k)})| - w_j, w_j = Y_j - Y_e/2 \tag{7}$$

Among them,  $P^{(k)}$  is the puzzle function for the  $k$ th iteration of  $T^{(k)}$ ,  $P_j^{(k+1)}$  is the puzzle function under the  $k + 1$ th execution of the program, and when  $R_j$  and  $T^{(k)}$  are merged,  $|C|$  denotes the number of elements in the set  $C$ , while  $w_j$  is the weight of  $R_j$ . The closer  $R_j$  is to the EP-region, the bigger the value of  $w_j$ , and the further away  $R_j$  is from the EP-region, the smaller the value of  $w_j$ .  $Y_j$  is the vertical distance from the  $R_j$  region.  $Y_e$  is the vertical distance from the edge point  $\tau(x_\rho, Y_e)$  on the EP-region (Figure 4). In Figure 4,  $A$  represents the AE-region;  $B$  indicates the Bounding Box's region;  $C_1, C_2,$  and  $C_3$  are the edge lines of the puzzle pieces.

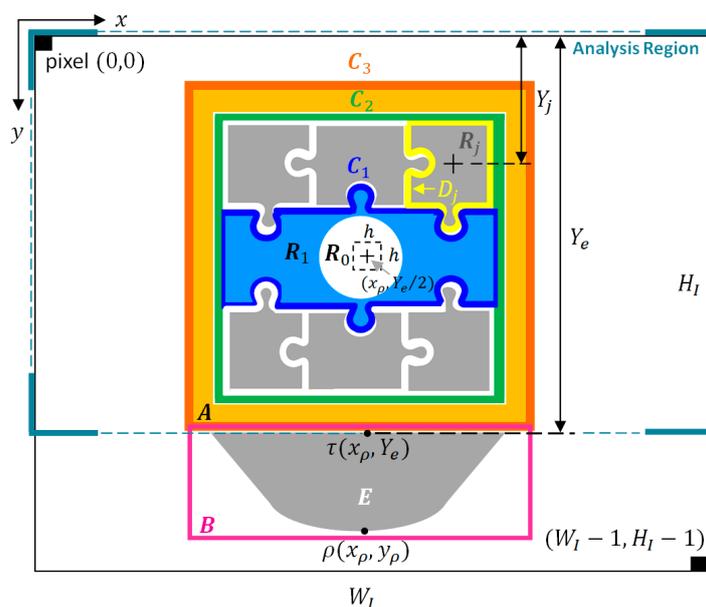
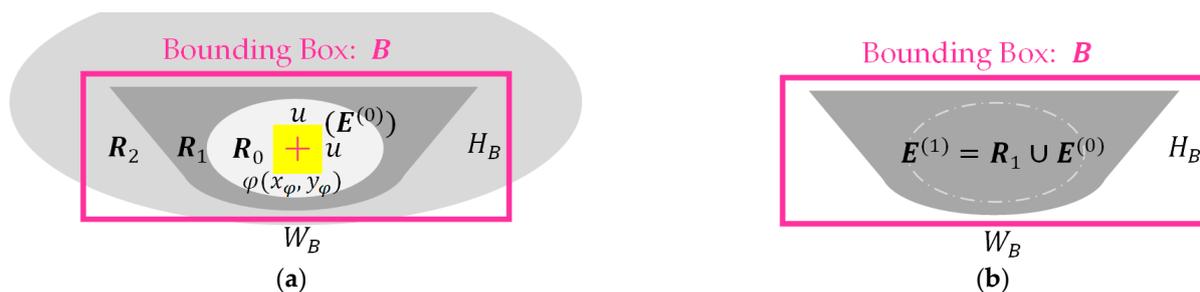


Figure 4. A diagram of AE-region merging using RPA.

The following shows how the RPA merges the EP-region and the AE-region in practice. There are two differences between the two regions: one is the selection of the initial region and the other is the analysis region range.

#### 2.4.1. EP-Region Merging

After the YOLOv4 model finds the bounding box (denoted  $B$ ) of the epiglottis, we start to merge the regions defined by  $B$ , thereby yielding the target region (EP-region, denoted  $E$ ). First, in the  $u \times u$  range (center of  $\varphi(x_\varphi, y_\varphi)$ ), we find the brightest region  $R_0$  (with the largest RGB color value, such as  $R_0$ , as shown in Figure 5) as the puzzle board (the initial region  $E^{(0)}$  of the epiglottic cartilage,  $k = 0$ ). Then, we find the surrounding regions of  $R_0$  (such as  $R_l$ , with  $l = 1, 2, 3, \dots$ ), and merge  $E^{(0)}$  with  $R_l$  repeatedly. If  $R_l$  exceeds the range of  $B$ , it will not be merged, and the result of the  $E^{(k+i)}$  merge will be defined as  $E^{(1)}$ . The above steps are repeated. If  $P^{(k+i)}$  is the same as  $P^{(k+i+1)}$ , merging will be stopped, and  $E^{(k+i)}$  is the merged result of the EP-region. As shown in Figure 5 as an example,  $H_B$  is the height of  $B$ , and  $W_B$  is the width of  $B$ . In Figure 5a,  $E^{(0)}$  is the initial region of the epiglottis, and  $u$  is set to a value between  $H_B/4$  and  $H_B/2$ , which means that the surrounding region  $R_1$  of the initial region  $E^{(0)}$  does not exceed the analysis range  $B$ , and is completely adjacent to the surrounding region  $E^{(0)}$ . The range of  $R_2$  that exceeds  $B$  will not be merged; therefore,  $E^{(1)} (= R_1 \cup E^{(0)})$  is the final merged result (Figure 5b). The common EP-region is similar to the type shown in Figure 5. In addition, if the range of  $R_0$  exceeds the bounding box, we use it as the epiglottis region and stop the merging procedure.



**Figure 5.** A diagram of EP-region merging: (a) the initial region  $E^{(0)} (= R_0)$  and (b) the final merged result  $E^{(1)} (= R_1 \cup E^{(0)})$ .

#### 2.4.2. AE-Region Merging

The RPA merges the  $R_j$  in the analysis region to obtain the target region (AE-region, denoted  $A$ ). We demonstrate this in Figure 4. The common AE-region is similar to Figure 4.  $H_I$  and  $W_I$  in Figure 4 are the height and width of the image, respectively, and the pixel  $(0, 0)$  is the origin of the image, while  $(W_I - 1, H_I - 1)$  is the last point of the image. Before performing this algorithm, we first find the highest point  $\rho(x_\rho, y_\rho)$  and the edge point  $\tau(x_\tau, y_\tau)$  of the EP-region  $E$ . The analysis region is the range within  $Y_e$  corresponding to the  $h \times h$  range at  $(x_\rho, Y_e/2)$  everywhere ( $h$  is between  $Y_e/4$  and  $Y_e/2$ ). Next, we find the darkest color in this range  $R_0$  (where the RGB color value is the smallest) and use it as a puzzle board and as the initial region  $A^{(0)}$  of the AE-region ( $A^{(k)}$ ,  $k = 0$ ).  $A^{(k)}$  is the AE-region after the  $k$ th merge. After detecting  $R_0$ , we set it to  $A^{(0)}$  and execute the RPA for region-merging analysis. The algorithm analyzes the puzzle functions of the  $k$ th merging process of  $R_j$  and  $A^{(k-1)}$  one by one with the puzzle function  $P_j^{(k)} (= |C_j^{(k)}| - w_j)$ . The merging process is repeated until  $P^{(k+i)} = P^{(k+i+1)}$ , upon which the program is stopped, outputting the best result of the AE-region  $A$ . As shown in Figure 4, we describe this in a sequence of steps for AE-region merging using the RPA:

Step 1. Find the region covered by the  $h \times h$  range at the location  $(x_\rho, Y_\tau/2)$  as the initial region  $A^{(0)}$  and fill any holes within  $A^{(0)}$ .

- Step 2. In the  $k$ th merging procedure ( $k \geq 1$ ), check whether the region surrounding  $A^{(k-1)}$  and region  $R_j$  meets two conditions: (1)  $A^{(k-1)}$  is surrounded by a single region  $R_j$ , and (2) the range of  $R_j$  does not exceed  $Y_e$ . If the two conditions are satisfied,  $A^{(k-1)}$  will be directly merged with  $R_j$  and the AE-region is updated to be  $A^{(k)}$ . Fill in the holes in  $A^{(k)}$ , and proceed to Step 3.
- Step 3. Perform the edge operation processing on  $A^{(k)}$  to obtain the edge line  $C^{(k)}$  and calculate the puzzle function  $P^{(k)}$ .
- Step 4. Search for puzzle pieces  $R_j$  adjacent to  $A^{(k)}$ , and calculate the edge lines  $D_j$  of  $R_j$ .
- Step 5. Perform an XOR operation on each edge  $D_j$  of  $R_j$  with  $C^{(k)}$  and calculate  $P_j^{(k+1)}$ . If  $P_j^{(k+1)} \leq P^{(k)}$ , merge  $R_j$  with  $A^{(k)}$ , and vice versa. After completing all the surrounding analysis of  $R_j$ , fill in the holes of  $R_j$ .
- Step 6. Compare the values of  $P^{(k)}$  and  $P^{(k+1)}$ . If  $P^{(k)} \leq P^{(k+1)}$ , stop the algorithm.  $A^{(k)}$  is the best result. If  $P^{(k)} > P^{(k+1)}$ , update the AE-region result to be  $A^{(k+1)}$ , and repeat steps 2 through 6.

### 3. Results

Six experiments were performed to verify the effectiveness of the method proposed in this study, namely, epiglottis image positioning, EP-region merging, AE-region merging, the effect of color quantification on the AE-region, analysis of continuous image AE-region change, and comparison of the proposed method with the clinical judgment results. This study's experimental equipment and development tools are as follows: Epiglottis images were captured from the epiglottis of OSA patients and recorded during endoscopy (frame rate: 30 frames/s, image resolution:  $640 \times 480$  (pixels)). Endoscopy was performed using a KARL STORZ 3.7 mm CMOS video rhino-laryngoscope (11101CMK was manufactured by KARL STORZ in Tuttlingen, Germany; Figure 6a) connected with an 8402 ZX monitor (8402 ZX monitor was manufactured by KARL STORZ in Tuttlingen, Germany; Figure 6b), and the video recording file format was Mpeg4. The patients underwent the DISE method at the Kaohsiung Veterans General Hospital to evaluate the obstruction sites during sleep. After the patients fell asleep, the surgeon placed the rhino-laryngoscope in front of the epiglottis and recorded the collapse/obstruction of the epiglottis during the inspiration phase. The IRB of the authors' institution approved this research, with approval no. IRB-KSVGH20-CT2-02. The image analysis algorithm and YOLOv4 implemented in this research were written and executed in the C++ programming language. The computer hardware included an Intel(R) Core (TM) i7 2.8 GHz Central Processing and NVIDIA RTX A4500 GPU graphics card for epiglottis image training and executing the proposed method.



(a)



(b)

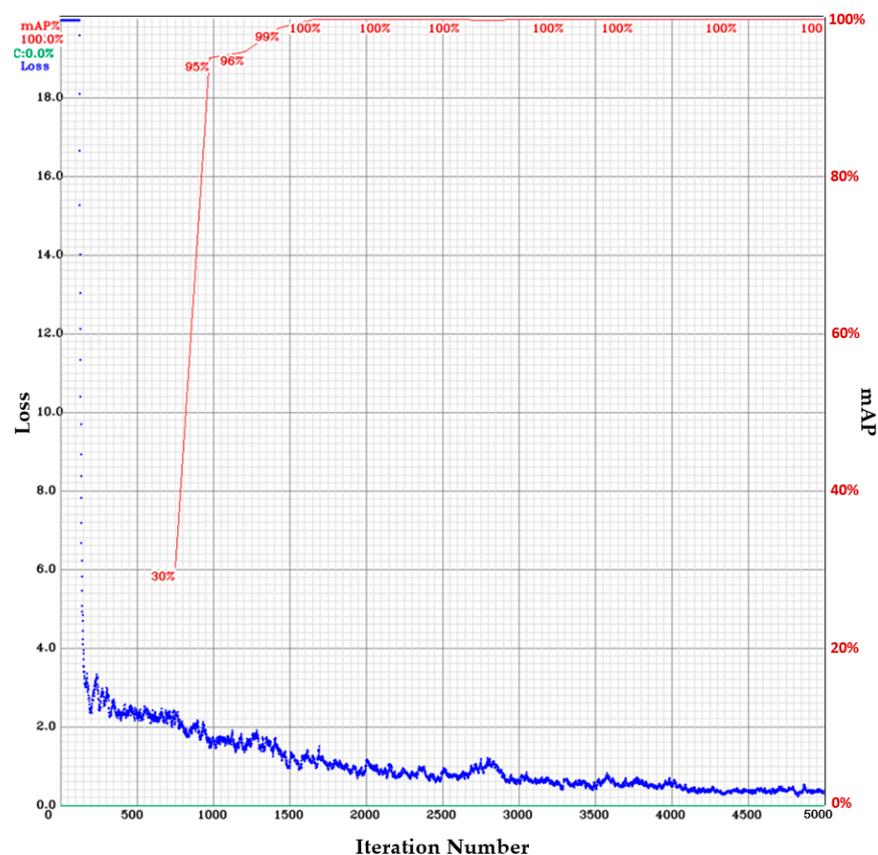
**Figure 6.** This study's experimental endoscopy: (a) CMOS video rhino-laryngoscope and (b) 8402 ZX monitor.

### 3.1. Epiglottis Image Positioning by the YOLOv4

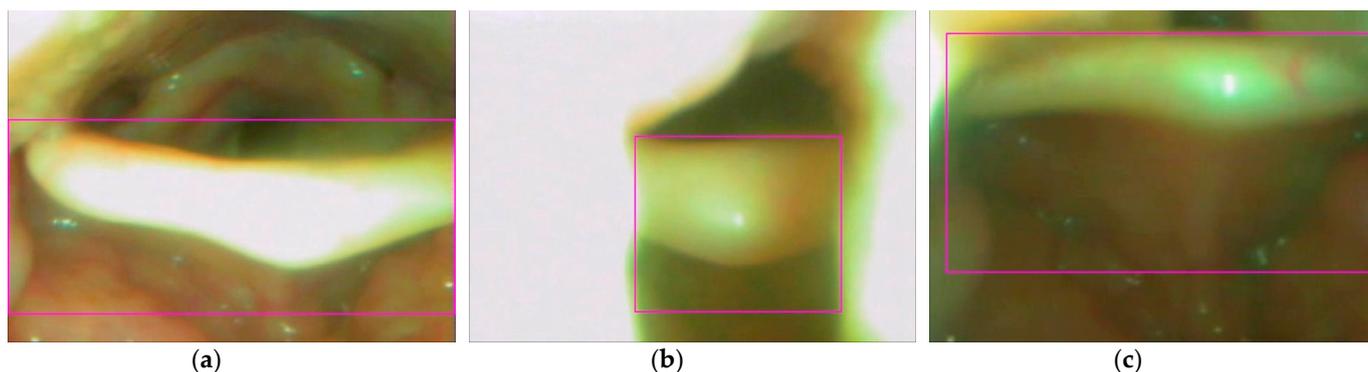
We collected 2230 epiglottis images for training the neural network (1784 images were used for model training (80%), and 446 images were used for model validation (20%)) and labeled the epiglottis images using Labellmg [39]. During the training of the sample, the parameters set by YOLO4 were used in this study. The configuration file was set as follows:

1. The batch size and the mini-batch size are 32 and 8, respectively.
2. The image input to the network is 24-bit color images with  $416 \times 416$  pixels.
3. The activation function is Mish.
4. The regional comparison rate is 0.9 (Momentum = 0.949).
5. The weight reduction ratio is 0.0005 (Decay = 0.0005).
6. The sample image diversification parameters are set to Angle = 0, Saturation = 1.5, Exposure = 1.5, and Hue = 0.1.
7. The network learning rate is 0.001 (Learning Rate = 0.001).
8. The maximum number of iterations is 5000 (Max Batches = 5000).
9. The learning policy is "Step".
10. The number of feature generation filters is 32 (Filter = 32).
11. The number of object classes is 1 (epiglottis) (Classes = 1).

The YOLOv4 model was used to perform 5000 iterations of training of the obtained epiglottis images. A chart of the loss value and mAP during training is shown in Figure 7 (for the training samples: the average loss is 0.3186; for the validation samples: IoU\_threshold is set as 0.75, mAP is 87.92%, Precision is 0.91, Recall is 0.91, and F1 score is 0.91). The horizontal axis of Figure 7 represents the iteration number, the left vertical axis indicates the loss value, and the right vertical axis shows the mAP value. For this application, the epiglottis was the only recognized object. A rectangular box (bounding box) is used to mark the epiglottis image. Figure 8 shows the recognized results of three cases.



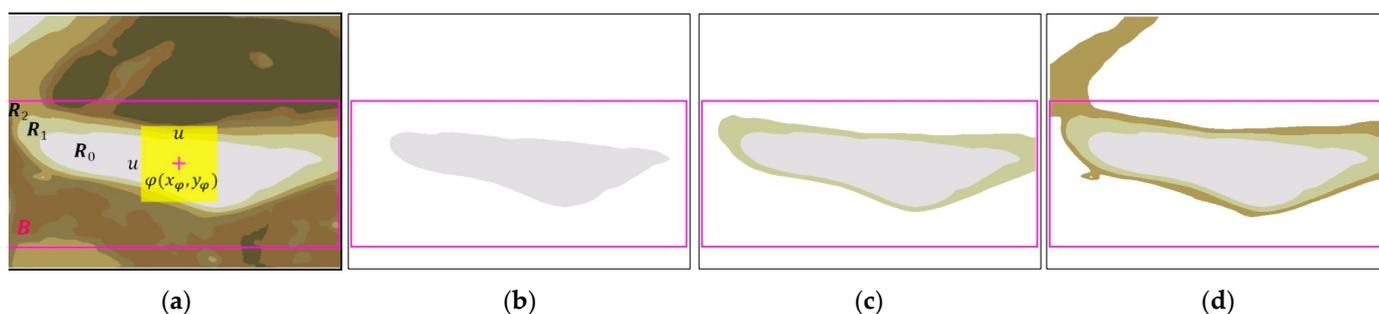
**Figure 7.** A chart of the loss value and mAP in the epiglottis images trained using YOLOv4 (the blue curve represents the loss value, while the red curve represents the mAP value).



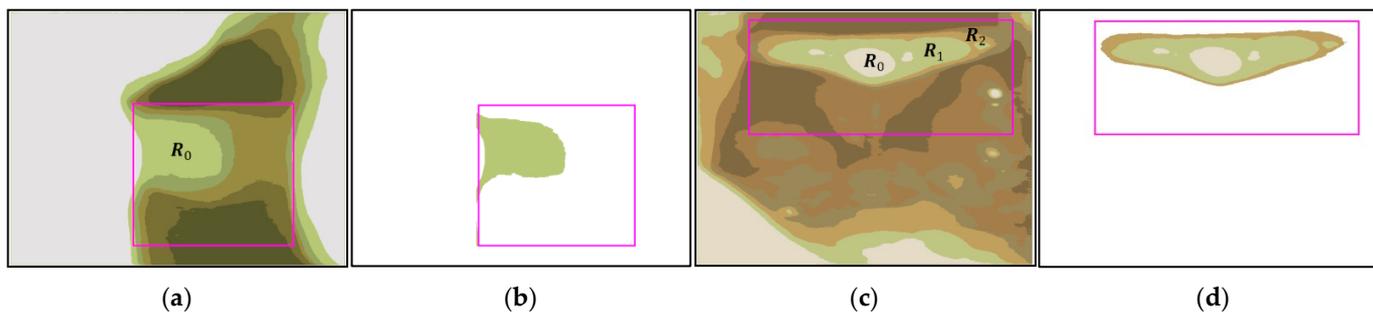
**Figure 8.** Three epiglottis images' positioning (shown in bounding boxes of fuchsia color) based on YOLOv4: (a) case 1; (b) case 2; and (c) case 3.

### 3.2. EP-Region Merging

After obtaining each image's bounding box ( $B$ ) of the epiglottis, we used the median-cut method to perform color quantization (the number of quantized colors,  $q = 6$ ). The color-quantized image is shown in Figure 9a, and all regions are marked.  $\varphi(x_\varphi, y_\varphi)$  in Figure 9a is the center of the region  $B$ , and the  $h \times h$  range in the experiment is  $60 \times 60$  pixels. Figure 9a shows the image of Figure 8a case 1 after color quantization. We found and denoted the brightest region in the  $60 \times 60$  range as  $R_0$  and used it as the initial region  $E^{(0)}$  (Figure 9b). Then, we detected region  $R_1$  that is adjacent to the initial region  $E^{(0)}$ .  $R_1$  meets the condition that it is within the range of  $B$ , so  $R_1$  and the initial region  $E^{(0)}$  were merged to become  $E^{(1)}$ . As shown in Figure 9c, the program continued to detect region  $R_2$  that surrounds  $E^{(1)}$ . Since  $R_2$  exceeds the range  $B$  (Figure 9d),  $E^{(1)}$  was not merged with  $R_2$ , and the EP-region is  $E^{(1)}$  ( $(R_1 \cup R_0) \subseteq B$ ). Figure 10 shows the EP-region merging results of case 2 and case 3 in Figure 8. Case 2 in Figure 10a is directly used as the EP-region (Figure 10b) because the range of  $R_0$  exceeds  $B$ ; Figure 10c, which shows case 3, combines the three regions ( $(R_2 \cup R_1 \cup R_0) \subseteq B$ ) as the EP-region  $E^{(2)}$ , and the result is shown in Figure 10d.



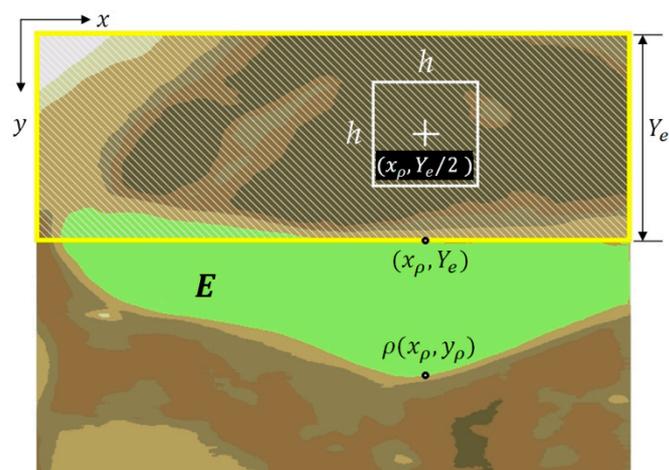
**Figure 9.** EP-region merging of case 1 ( $E^{(1)}$ ) within the bounding box: (a) color-quantized image; (b)  $E^{(0)}$  ( $= R_0$ ); (c)  $E^{(1)}$ ; and (d)  $R_2$  exceeds the range  $B$ .



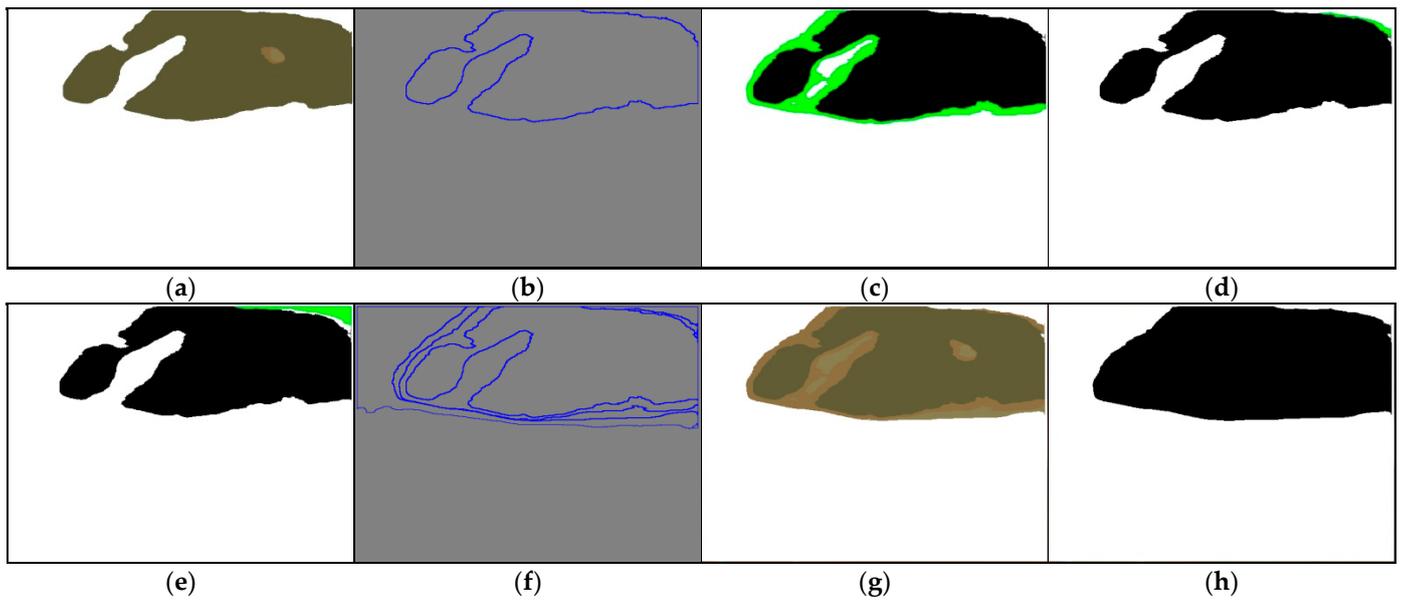
**Figure 10.** EP-region merging of case 2 and case 3: (a) case 2; (b) EP-region  $E^{(0)}$ ; (c) case 3; and (d) EP-region  $E^{(2)} (= R_2 \cup E^{(1)} = R_2 \cup (R_1 \cup E^{(0)}) = R_2 \cup R_1 \cup R_0)$ .

3.3. AE-Region Merging

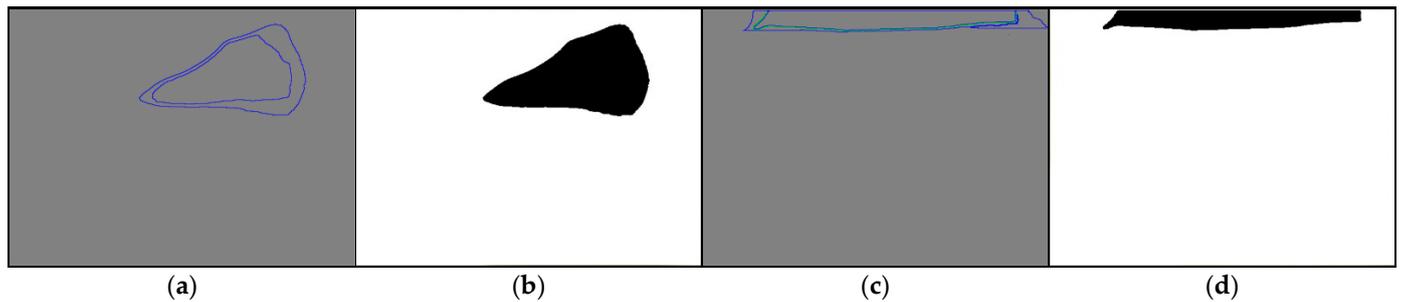
As shown in Figure 11, the RPA comprises six steps. In Step 1, we acquired the highest point  $\rho(418, 371)$  of the EP-region. Then, we calculated the  $56 \times 56$  ( $h \times h$ ) range at coordinate  $(x_\rho, Y_e/2) = (418, 112)$ . We chose the region  $R_0$  with the smallest RGB value and used it as the initial region  $A^{(0)}$  of the AE-region. We then filled in the holes of  $A^{(0)}$  (Figure 12a). In Step 2, we checked whether  $A^{(0)}$  is surrounded by a single region. In Step 3, we used edge calculations [14] on  $A^{(0)}$  to acquire the puzzle function  $P^{(0)} (= |C^{(0)}|)$  with edge line  $C^{(0)}$ . In Step 4, we found nearby puzzle pieces  $R_j = \{R_1, R_2, R_3\}$  along  $C^{(0)}$  (Figure 12b).  $R_1 - R_3$  are shown in Figure 12c–e, respectively. The edge lines of these regions are  $D_1, D_2,$  and  $D_3$ , respectively. In Step 5, we performed XOR calculation on  $C^{(0)}$  and  $D_1, D_2,$  and  $D_3$ . The puzzle function of  $C^{(0)}$  and  $D_1$  and  $C^{(0)}$  and  $D_2$  are  $P_1^{(1)} (< P^{(0)})$  and  $P_2^{(1)} (< P^{(0)})$ , so we merged  $A^{(0)}$  with  $R_1$  and  $R_2$  to acquire  $A^{(1)} = (R_2 \cup R_1 \cup A^{(0)})$ . The puzzle function of  $C^{(0)}$  and  $D_3$  is  $P_3^{(1)}$ , and as  $P_3^{(1)} > P_0$ , no merging was performed. In Step 6, we updated the puzzle function to be  $P^{(1)}$  and the AE-region to be  $A^{(1)}$ , and repeated steps 2 through 6. Figure 12f shows the edge line  $D_j$ , Figure 12g shows the merging of  $R_1$  and  $R_2$  with  $R_0$ . Figure 12h shows  $A^{(1)}$  (the best result). Figure 13 shows the analyzed results of the AE-regions of case 2 and case 3 in Figure 8. These three cases are common forms of AE-regions.



**Figure 11.** Analysis of the AE-region (the yellow box with diagonal brush style).



**Figure 12.** AE-region merging of case 1: (a)  $A^{(0)}$ ; (b)  $C^{(0)}$ ; (c)  $R_1$ ; (d)  $R_2$ ; (e)  $R_3$ ; (f)  $C^{(1)}$ ; (g)  $R_2 \cup R_1 \cup R_0$ ; and (h)  $A^{(1)}$ .



**Figure 13.** AE-region merging of case 2 and case 3: (a)  $C^{(1)}$  of case 2; (b)  $A^{(1)}$ ; (c)  $C^{(0)}$  of case 3; and (d)  $A^{(0)}$ .

### 3.4. The Effect of Color Quantification on the AE-Region

The AE-region detection method that we proposed is mainly based on setting a low number of quantized colors. The quantized color number of each experimental image was less than 12, and to understand the influence of quantized color number on AE-region detection, 10 continuous images ( $t_1$ – $t_{10}$ , 1/3 s) were tested in this experiment for the AE-region, with the number of quantized colors ranging from 3 to 12. Figure 14 lists five AE-regions with the number of quantized colors ( $q = 4, 6, 8, 10, 12$ ). As shown in Figure 14, the greater the number of colors, the greater the number of  $R_j$  in the AE-region, and the greater the change in the shape of the region. To obtain a better understanding of the change in the AE-regions, we quantified the number of AE-region pixels ( $\omega_n = |A^n|$ ) for the number of colors ( $q$  ranging from 3 to 12) by 10, and the results are shown in Figure 15. To demonstrate subtle differences, we show  $q = 3 - 8$  in Figure 15a and  $q = 9 - 12$  in Figure 15b. Figure 15a shows that when  $q = 6 - 8$ , the AE-region areas of the three color quantization numbers are closer, and when  $q = 9 - 12$ , the areas are even closer, but the results are quite different. If the 10 continuous images were judged by humans, the AE-regions behind the epiglottis would gradually become smaller; thus, the number of colors during the detection of the AE-regions is more in line with the results of human observation.

### 3.5. Continuous Image AE-Region Change Analysis

We used a 12.7 s video (total image number was 380) recorded using the rhino-laryngoscope to perform the AE-region detection experiment. The color quantization number was set to  $q = 6$  (Figure 16). These 380 images include partial obstruction, complete obstruction, and rhino-laryngoscope sliding. The movement of the rhino-laryngoscope was used as the basis for updating the period number. When the epiglottis could not be detected, the rhino-laryngoscope was considered to be slipping. The two periods of t175–t191 and t298–t305 in the 380 images were regarded as the movement of the rhino-laryngoscope. Therefore, three periods ( $m = 1, 2, 3$ ) were detected in the 380 images; among them, the period  $m = 1$  is t1–t174, the period  $m = 2$  is t192–t29, and the period  $m = 3$  is t307–t341. The fully blocked part contains images t64–t72 (for 0.30 s), t141–t165 (for 0.83 s), and t307–t341 (for 1.16 s), while the rest are partially blocked and unblocked. During the analysis process, abnormal numerical values of the AE-region in images t306 and t357 (Figure 17(a1,b1), respectively) were detected. The color-quantized images are shown in Figure 17(a2,b2). The detection results of the AE-region in images t306 and t357 ( $q = 6$ ) are shown in Figure 17(a3,b3), respectively. In addition, the detection results of the color quantization number ( $q = 4$ ) are shown in Figure 17(a4,b4). The differences among Figure 17(b3,b4) are explained in Section 4.

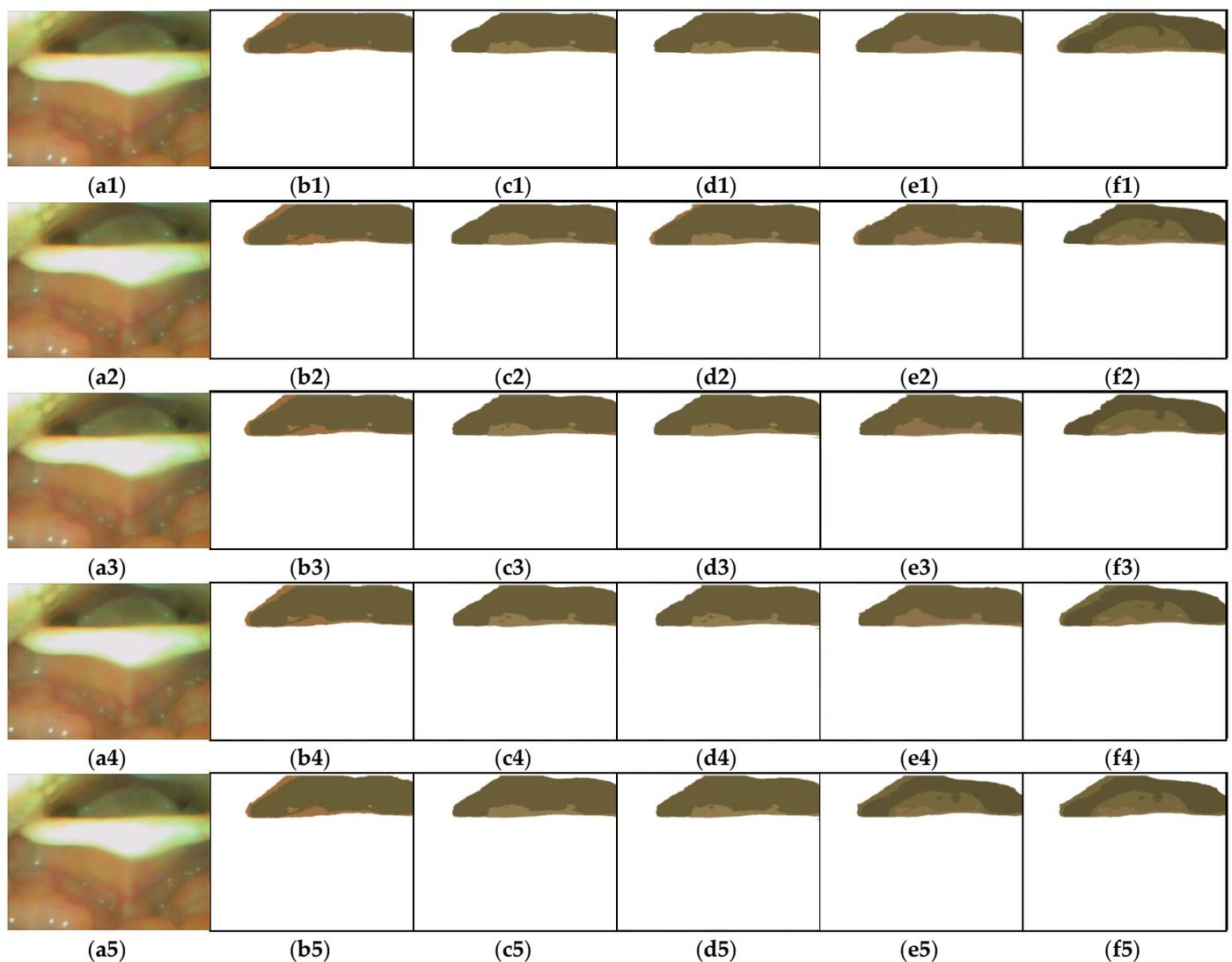


Figure 14. Cont.

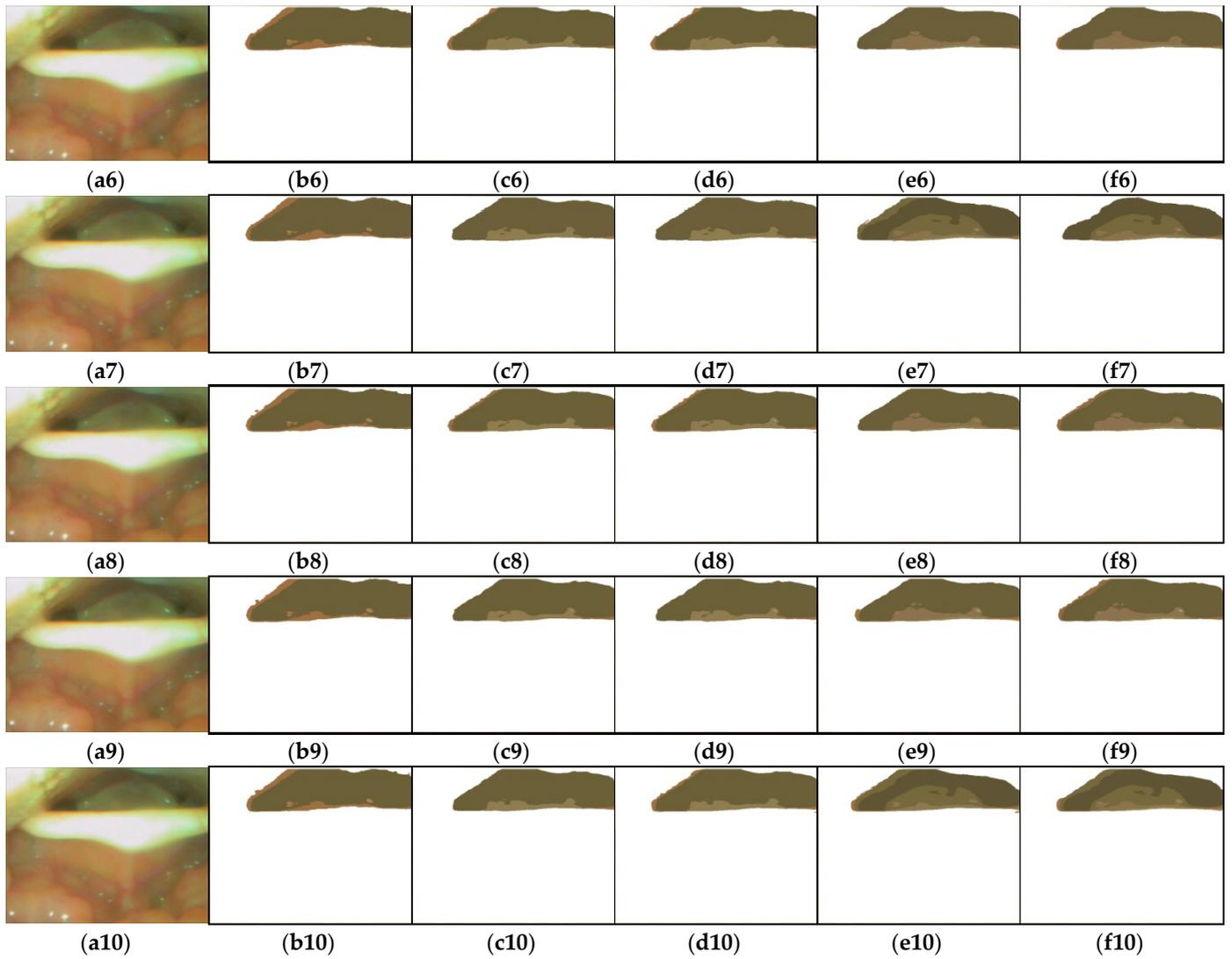
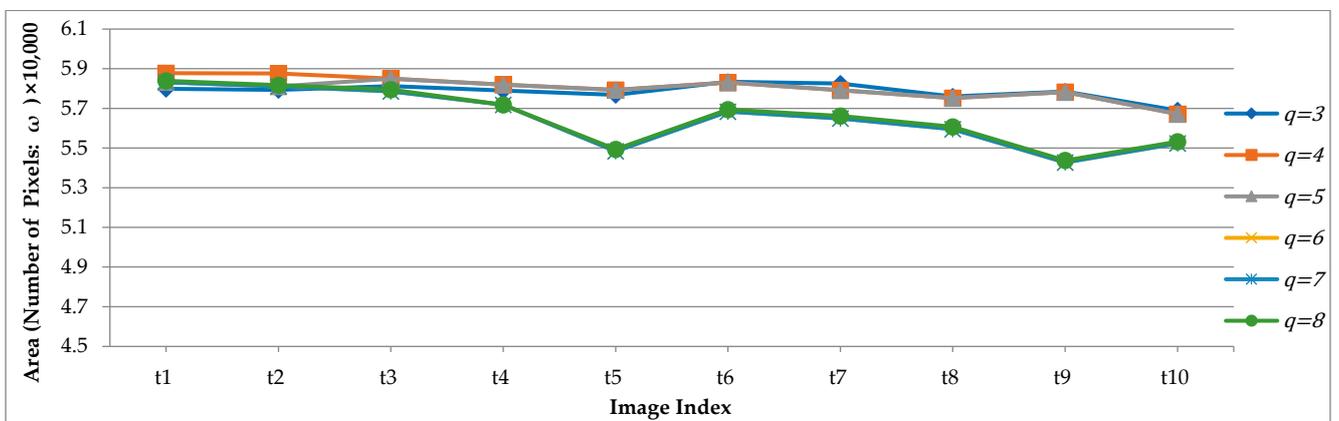
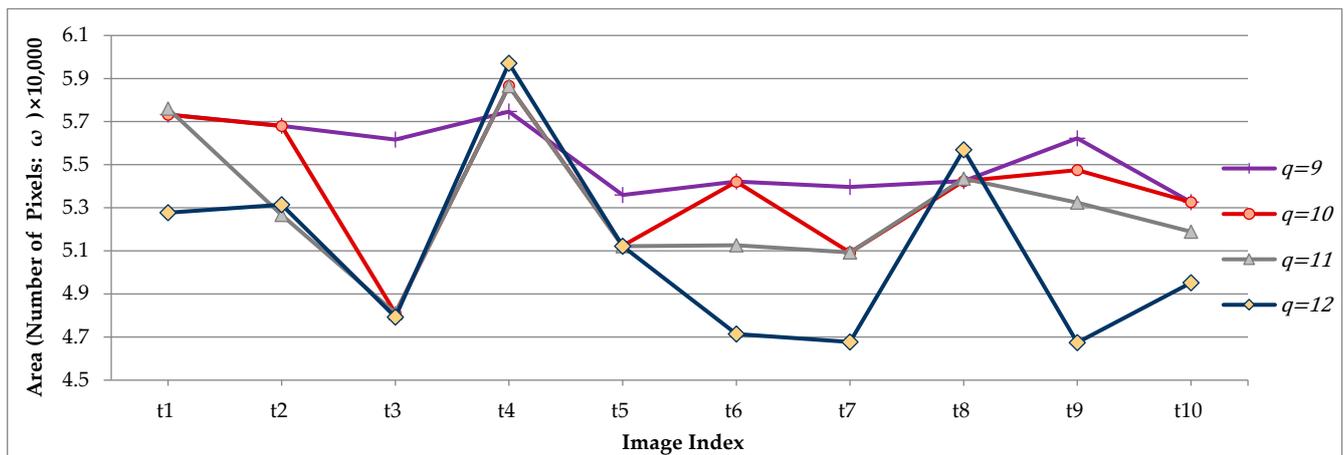


Figure 14. The AE-regions with the number of quantified colors ( $q = 4, 6, 8, 10, 12$ ): (a1–a10)  $t_1$ – $t_{10}$ ; (b1–b10)  $q = 4$ ; (c1–c10)  $q = 6$ ; (d1–d10)  $q = 8$ ; (e1–e10)  $q = 10$ ; (f1–f10)  $q = 12$ .



(a)

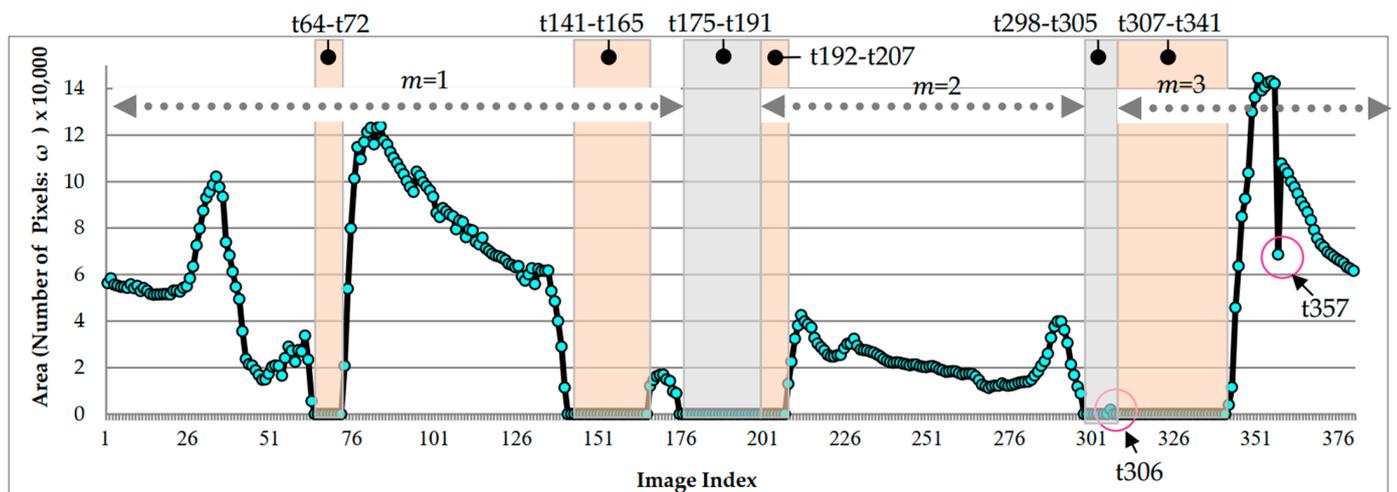
Figure 15. Cont.



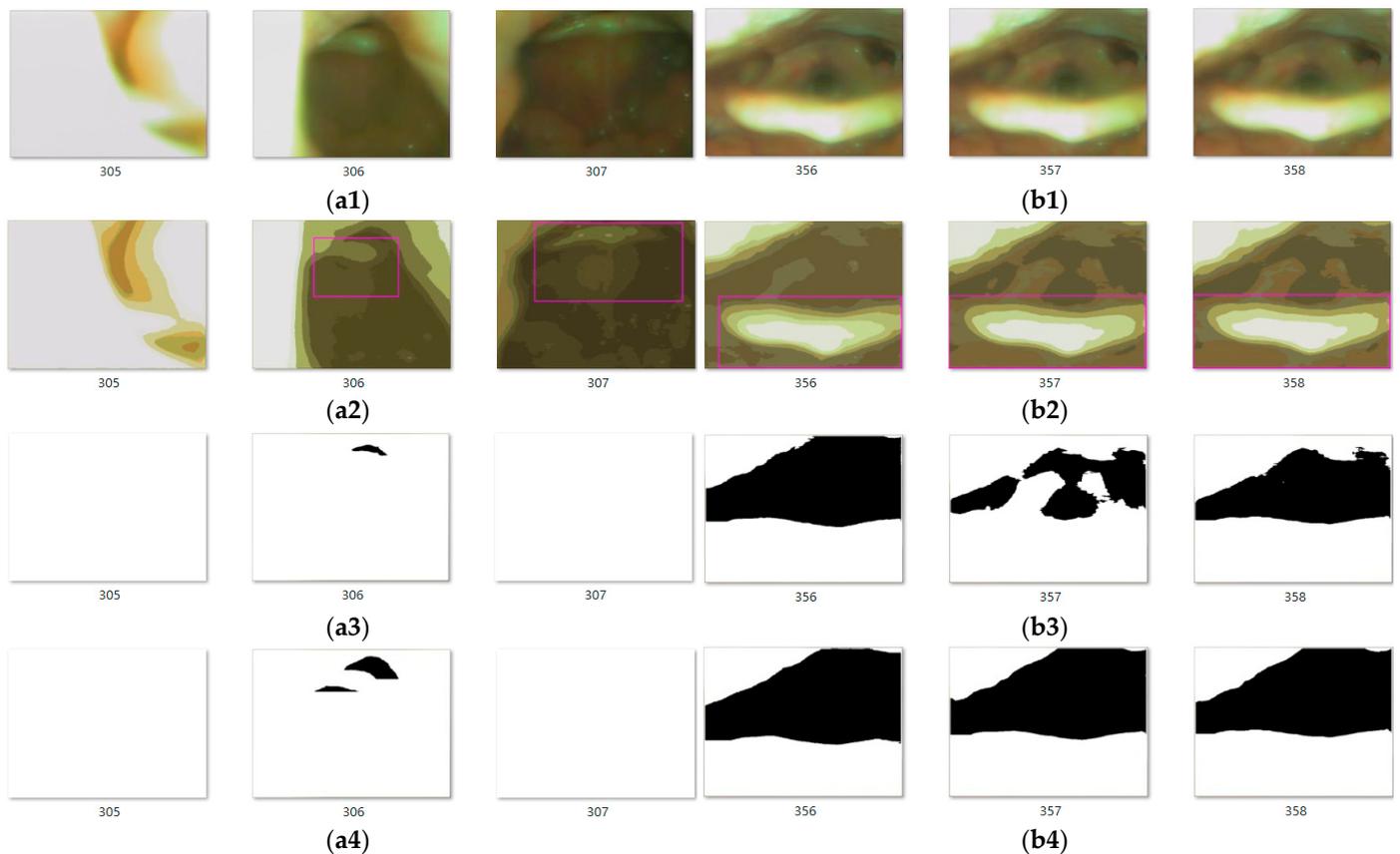
(b)

**Figure 15.** The areas of AE-region ( $\omega$ ) with the number of quantified colors: (a)  $q = 3, 4, 5, 6, 7,$  and  $8,$  and (b)  $q = 9, 10, 11,$  and  $12.$

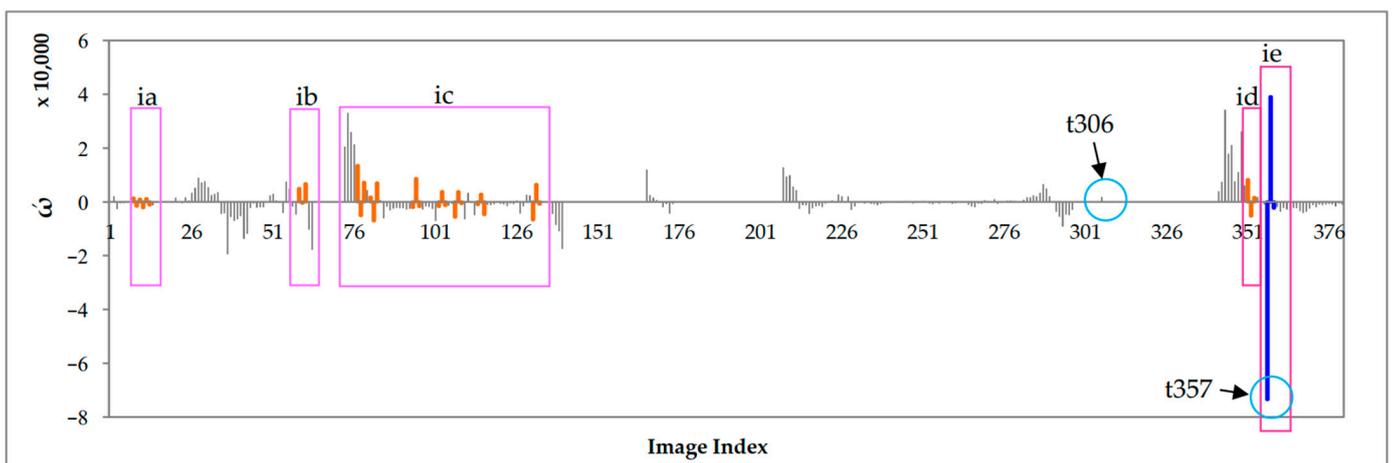
Among the images, the t306 image was detected to have an AE-region; however, the before and after images of t305 and t307 do not detect the epiglottis nor complete obstruction. Thus, we consider t306 to be abnormal data based on the results of the before and after images. In addition, there is a large difference between the detection results of images t357, t356, and t358 (Figure 17(b3)). We used inverse difference calculation ( $\omega'_{n+1} = \omega_{n+1} - \omega_n$ ) to obtain the results ( $\omega'$ ), which are shown in Figure 18, and examined the  $\omega'$  dataset for “positive and negative positive” and “negative positive and negative” values (as shown in Figure 18 in boxes ia–ie). If the difference value is higher than the threshold ( $|\varepsilon| > 30,000$ , where the symbol  $|\cdot|$  is the absolute value operation), then it is regarded as abnormal. Therefore, the AE-region data of t357 were judged as abnormal, and we did not perform obstruction ratio calculations. Moreover, the computation time for the 380 images was 269.8 s, with an average computation time of 0.71 s per image.



**Figure 16.** The AE-region areas  $\omega$  in the 12.7 s video (total image number is 380).



**Figure 17.** The AE-regions for images t306 and t357: (a1,b1) the original images; (a2,b2) the color-quantized images with the bounding box of epiglottis localization; (a3,b3) AE-regions ( $q = 6$ ); and (a4,b4) AE-regions ( $q = 4$ ).



**Figure 18.** The backward difference ( $\omega'$ ) of the AE-region areas  $\omega$ .

### 3.6. Comparing the Proposed Method with the Clinical Judgment Results

This experiment used the same 380 images as the judgment targets, compared the results of the proposed method with the results obtained based on clinical judgment, and evaluated the appropriateness and practicability of our method through the comparison. During the comparison process, the period of tonsil interference was artificially determined. The mark was not included in the calculation of the obstruction ratio. The images of these two periods were t37–t62 and t156–t306. This experiment compares results based on current clinical diagnostic obstruction degrees, which were 0–49%, 50–75%, and 76–100%. The

following are the results of the clinical judgment for the 380 continuous images (the clinical judgment was determined by the Chief of the Department of Otorhinolaryngology-Head and Neck Surgery at Kaohsiung Veterans General Hospital):

- In t1–t36, the epiglottis shrinks at least 50%;
- In t37–t62, there is tonsil interference (not included in the calculation of obstruction ratio);
- In t63–t73, there is no posterior pharyngeal wall in the image, but the epiglottic cartilage is sucked to the posterior pharyngeal wall, which is clinically judged to be 76–100% obstructed;
- In t74–t137, the epiglottis shrinks less than 50%;
- In t138–t155, there is no posterior pharyngeal wall in the image, but the epiglottic cartilage is sucked to the posterior pharyngeal wall, which is clinically judged to be 76–100% obstructed;
- In t156–t306, there is no posterior pharyngeal wall in the image, but there is tonsil interference (not included in the calculation of obstruction ratio);
- In t307–t343, the epiglottis cartilage is sucked to the direction of the posterior pharyngeal wall, which is clinically judged to be 76–100% obstructed;
- In t345–t380, the epiglottis shrinks less than 50%.

There were three periods detected by our method ( $m = 1, 2, 3$ ). The two periods disturbed by the tonsils (t37–t62 and t156–t306) were excluded, and period 2 (t192–t297) fell within the period of tonsil interference, so we only counted periods 1 and 3. We estimated the maximum value of the AE-region ( $\max\{|A_m^n|\}$ ) and calculated the obstruction ratio  $O_m^n$  according to Equation (1), and the results are shown in Figure 19 (comparison of epiglottis obstruction ratio results). CI denotes the clinical judgment result,  $\bar{\omega}$  is the average obstruction ratio in each period, the right vertical axis shows the epiglottis obstruction ratio ( $O$ ), and the left vertical axis shows the number of pixels in the AE-region. The reddish-brown color histogram ( $O_m^n$ ) is the epiglottis obstruction ratio of the image. We averaged the obstruction ratios of periods  $m = 1$  and  $m = 3$  to obtain the average value  $\bar{\omega}$ . The results are as follows:

- $m = 1$  (t1–t36):  $\bar{\omega}$  is 48.8% (clinical judgment CI = 0–49%).
- $m = 1$  (t63–t73):  $\bar{\omega}$  is 98.1% (clinical judgment CI = 76–100%);
- $m = 1$  (t74–t137):  $\bar{\omega}$  is 31.3% (clinical judgment CI = 0–49%);
- $m = 1$  (t138–t155):  $\bar{\omega}$  is 96.4% (clinical judgment CI = 76–100%);
- $m = 3$  (t307–t343):  $\bar{\omega}$  is 99.7% (clinical judgment CI = 76–100%);
- $m = 3$  (t345–t380):  $\bar{\omega}$  is 34.5% (clinical judgment CI = 0–49%).

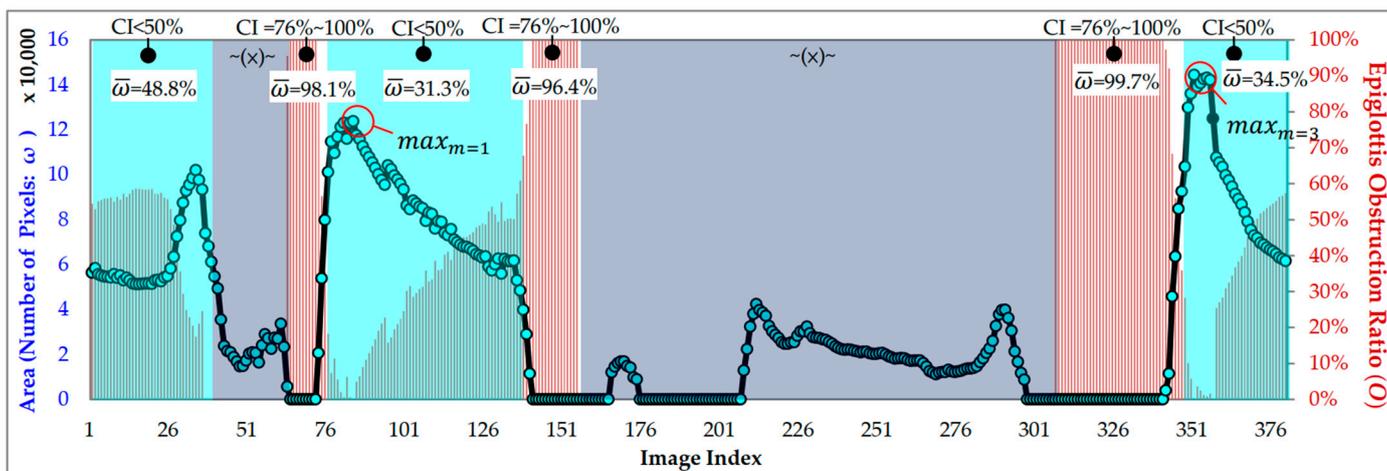


Figure 19. The results of epiglottis obstruction ratio ( $O$ ).

After excluding the two periods of tonsil interference (t37–t62 and t156–t306), the average values obtained for the two periods are consistent with the results of the clinical judgment.

## 4. Discussion

### 4.1. Method Limitations

The proposed image processing method of the system is based on the YOLOv4 for the identification of the epiglottis and the low color quantization number of regions. Together with our proposed RPA to obtain the AE-region, the AE-region area data are plotted as a curve, and abnormal data are filtered out by comparing the image data of the front and back images with the threshold value without calculating the obstruction ratio. The quantitative data obtained by the proposed method can provide surgeons during diagnosis with the severity of obstruction of the epiglottis of patients, solve the problem that surgeons can only roughly judge the severity of obstruction, and help surgeons improve the accuracy of the diagnosis. In addition, surgeons can also explain to patients their situation with continuous data. Patients can better understand the time and trend of airway obstruction in their epiglottis. We use color quantization to convert the image into color regions. The distribution of color regions affects the result of merging AE-regions. A more complete region can be obtained with low color quantization, but larger meandering regions are also prone to appear. Unnecessary regions might be merged into the AE-region. Still, the AE-region changes between consecutive images are relatively gentle and the trend of continuous data changes is closer to the surgeon identification (Figure 15a). Regions with a higher color quantization number are smaller, fragmented, incomplete, and require more calculation time. The AE-region between consecutive images varies greatly, which is quite different from the surgeon identification (Figure 15b,  $q = 12$ ). Take Figure 17(a4,b4) as examples, if  $q = 4$ , although t306 is also an abnormal value, t357 is a normal value. After comparing related experiments, we think that merging AE-regions with a low color quantization number is a better choice ( $q = 4 - 8$ ), so in the whole experiment, we take the middle-value  $q = 6$  as the color quantization number. In addition, we find the abnormal data by comparing the AE-region data of the before and after images with the threshold. We do not calculate the obstruction ratio from the abnormal data. From the experimental results, it can be observed that the results of our proposed method are consistent with the clinical judgment, which verifies the appropriateness and the practicability of this research method.

### 4.2. Work Limitations

This study introduces a novel method of epiglottic obstruction image analysis to calculate the obstruction ratio. The calculation of the obstruction ratio has been expanded from three degrees to a numerical ratio. Compared to obstruction degrees, the obstruction ratio can provide surgeons with more accurate diagnoses, allowing patients to receive more suitable treatment options. However, the method proposed in this study and the experiments conducted still have several limitations, as follows:

1. The study specifically calculates the precise extent of airway obstruction in the epiglottic region only. The method designed for the study cannot calculate the obstruction ratio for other obstruction sites.
2. After calculating the AE-Region, a surgeon needs to determine the period affected by the tonsils, as this method cannot identify the tonsils.
3. This research did not conduct a control group analysis, meaning the study did not analyze the images of the epiglottic region in healthy individuals who do not have OSA. However, theoretically, the research method is equally applicable.

## 5. Conclusions

Surgeons determine the type of treatment a patient needs based on the severity of epiglottic obstruction. Traditional manual detection cannot accurately calculate the extent and duration of epiglottic obstruction, and previous automated methods for scoring epiglottic obstruction have primarily identified only three obstruction degrees. Moreover, these methods are better suited for distinguishing between obstructed and non-obstructed states. However, the accuracy of the three-level obstruction identification is insufficient,

and its correlation with the apnea–hypopnea index (AHI) is not direct. The obstruction ratio provides a more precise diagnosis for surgeons. To accurately calculate the obstruction ratio, this study developed a system for calculating the epiglottic obstruction ratio. This system can perform in-depth image analysis of DISE videos and precisely calculate the epiglottis obstruction rate and duration. The system can assist surgeons in obtaining an accurate epiglottis obstruction ratio, enabling them to devise the best treatment plan for patients. This system is presented in a web format, offering cross-platform accessibility. Furthermore, the image analysis method proposed in this study integrates the YOLOv4, RPA, and other essential image processing algorithms. The proposed method only identifies a single epiglottic cartilage object. The YOLOv4 model’s training demonstrated a high mAP performance, with an overall high recognition rate (IoU threshold: 0.75 and F1 score: 91%). During the experiment of merging the EP-regions and AE-regions, it was also validated that the RPA could accurately merge the laryngeal cartilage region and the airway region. Additionally, this study explored the rationality of the number of color quantization areas. The last experiment utilized continuous data to compare the system’s calculated results with the clinical diagnostic outcomes determined by a surgeon. The system developed in this study could calculate the airway collapse’s percentage accuracy to 0.1%, and the average calculation time for a single image was only 0.71 s. Through experimental validation, the efficacy and practical value of this system’s image analysis method were proven. In future research, we will continue to develop image analysis methods to calculate the obstruction rate for the velum, oropharynx, and tongue areas, incorporate AHI index-related experiments, and enhance the functionality of this system.

**Author Contributions:** Supervision, H.-H.S.; Recommendations, H.-H.S.; Investigation, C.-P.L.; Methodology, C.-P.L.; Writing—Original draft, C.-P.L.; Writing—Review and Editing, C.-P.L.; Experiment, C.-P.L. and H.-H.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Kaohsiung Veterans General Hospital, Taiwan, and the Ministry of Science and Technology, Taiwan, grant numbers KSVMHU110-003, VGHMHU111-002, and MOST 111-2221-E-324-022.

**Institutional Review Board Statement:** This study was approved by the Institutional Review Board of Kaohsiung Veterans General Hospital, Taiwan (IRB: KSVGH20-CT2-02). The IRB approval was on 6 January 2020.

**Informed Consent Statement:** Patient consent was waived due to this study being a retrospective analysis of medical records status post de-identification.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Chen, Y.-T.; Chang, R.; Chiu, L.-T.; Hung, Y.-M.; Wei, C.-C. Obstructive sleep apnea and influenza infection: A nationwide population-based cohort study. *Sleep Med.* **2021**, *81*, 202–209. [[CrossRef](#)] [[PubMed](#)]
2. Evans, E.C.; Sulyman, O.; Froymovich, O. The goals of treating obstructive sleep apnea. *Otolaryngol. Clin. N. Am.* **2020**, *53*, 319–328. [[CrossRef](#)]
3. Lin, C.-H.; Chin, W.-C.; Huang, Y.-S.; Wang, P.-F.; Li, K.K.; Pirelli, P.; Chen, Y.-H.; Guilleminault, C. Objective and subjective long term outcome of maxillomandibular advancement in obstructive sleep apnea. *Sleep Med.* **2020**, *74*, 289–296. [[CrossRef](#)]
4. Vanderveken, O.M.; Maurer, J.T.; Hohenhorst, W.; Hamans, E.; Lin, H.-S.; Vroegop, A.V.; Anders, C.; Nico, V.; Heyning, P.H.V. Evaluation of drug-induced sleep endoscopy as a patient selection tool for implanted upper airway stimulation for obstructive sleep apnea. *J. Clin. Sleep Med.* **2013**, *9*, 433–438. [[CrossRef](#)] [[PubMed](#)]
5. Hsu, Y.-B.; Lan, M.-Y.; Huang, Y.-C.; Huang, T.-T.; Lan, M.-C. The correlation between drug-induced sleep endoscopy findings and severity of obstructive sleep apnea. *Auris Nasus Larynx* **2021**, *48*, 434–440. [[CrossRef](#)]
6. Kwon, O.E.; Jung, S.Y.; Al-Dilajjan, K.; Min, J.-Y.; Lee, K.H.; Kim, S.W. Is epiglottis surgery necessary for obstructive sleep apnea patients with epiglottis obstruction? *Laryngoscope* **2019**, *129*, 2658–2662. [[CrossRef](#)]
7. Li, H.-Y.; Lo, Y.-L.; Wang, C.-J.; Hsin, L.-J.; Lin, W.-N.; Fang, T.-J.; Lee, L.-A. Dynamic drug-induced sleep computed tomography in adults with obstructive sleep apnea. *Sci. Rep.* **2016**, *6*, 35849. [[CrossRef](#)]

8. Xia, F.; Sawan, M. Clinical and Research Solutions to Manage Obstructive Sleep Apnea: A Review. *Sensors* **2021**, *21*, 1784. [[CrossRef](#)] [[PubMed](#)]
9. Edwards, C.; Almeida, O.P.; Ford, A.H. Obstructive sleep apnea and depression: A systematic review and meta-analysis. *Maturitas* **2020**, *142*, 45–54. [[CrossRef](#)] [[PubMed](#)]
10. Sunter, G.; Omercikoglu, O.H.; Vural, E.; Gunal, D.I.; Agan, K. Risk assessment of obstructive sleep apnea syndrome and other sleep disorders in multiple sclerosis patients. *Clin. Neurol. Neurosurg.* **2021**, *207*, 106749. [[CrossRef](#)] [[PubMed](#)]
11. Jonas, D.E.; Amick, H.R.; Feltner, C.; Weber, R.P.; Arvanitis, M.; Stine, A.; Lux, L.; Middleton, J.C.; Voisin, C.; Harris, R.P. *Screening for Obstructive Sleep Apnea in Adults Us Preventive Services Task Force Recommendation Statement*; Report No.: 14-05216-EF-1; Agency for Healthcare Research and Quality (US): Rockville, MD, USA, January 2017.
12. Kezirian, E.J.; Hohenhorst, W.; de Vries, N. Drug-induced sleep endoscopy: The VOTE classification. *Eur. Arch. Oto-Rhino-Laryngol.* **2011**, *268*, 1233–1236. [[CrossRef](#)]
13. Tutsoy, O. Pharmacological, non-pharmacological policies and mutation: An artificial intelligence based multi-dimensional policy making algorithm for controlling the casualties of the pandemic diseases. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 9477–9488. [[CrossRef](#)]
14. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson: Harlow, UK, 2017; pp. 184–187.
15. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: London, UK, 2016; pp. 330–372.
16. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *1*, 1097–1105. [[CrossRef](#)]
17. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7 June 2015. [[CrossRef](#)]
18. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the 14th European Conference on Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Berlin/Heidelberg, Germany, 2016; Volume 9905, pp. 21–37. [[CrossRef](#)]
19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016. [[CrossRef](#)]
20. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
21. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement (Tech Report) 2018. Available online: <https://pjreddie.com/media/files/papers/YOLOv3.pdf> (accessed on 20 January 2021).
22. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2022**, arXiv:2004.10934. [[CrossRef](#)]
23. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. Scaled-YOLOv4: Scaling cross stage partial network. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2021, Nashville, TN, USA, 20–25 June 2021. [[CrossRef](#)]
24. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696. [[CrossRef](#)]
25. EKEN, S. Medical data analysis for different data types. *Int. J. Comput. Exp. Sci. Eng.* **2020**, *6*, 138–144. [[CrossRef](#)]
26. Ferrer-Lluis, I.; Castillo-Escario, Y.; Montserrat, J.M.; Jané, R. Enhanced monitoring of sleep position in sleep apnea patients: Smartphone triaxial accelerometry compared with video-validated position from polysomnography. *Sensors* **2021**, *21*, 3689. [[CrossRef](#)] [[PubMed](#)]
27. Kou, W.; Carlson, D.A.; Baumann, A.J.; Donnan, E.; Luo, Y.; Pandolfino, J.E.; Etemadi, M. A deep-learning-based unsupervised model on esophageal manometry using variational autoencoder. *Artif. Intell. Med.* **2021**, *112*, 102006. [[CrossRef](#)]
28. Liu, Y.; Feng, Y.; Li, Y.; Xu, W.; Wang, X.; Han, D. Automatic classification of the obstruction site in obstructive sleep apnea based on snoring sounds. *Am. J. Otolaryngol.* **2022**, *43*, 103584. [[CrossRef](#)]
29. Hanif, U.; Kiaer, E.K.; Capasso, R.; Liu, S.Y.; Mignot, E.J.M.; Sorensen, H.B.D.; Jennum, P. Automatic scoring of drug-induced sleep endoscopy for obstructive sleep apnea using deep learning. *Sleep Med.* **2023**, *102*, 19–29. [[CrossRef](#)]
30. Heckbert, P.S. Color image quantization for frame buffer display. *Comput. Graph.* **1982**, *16*, 297–307. [[CrossRef](#)]
31. Torre, C.; Camacho, M.; Liu, S.-Y.; Huon, L.-K.; Capasso, R. Epiglottis collapse in adult obstructive sleep apnea: A systematic review. *Laryngoscope* **2016**, *126*, 515–523. [[CrossRef](#)] [[PubMed](#)]
32. He, L.; Chao, Y.; Suzuki, K.; Wu, K. Fast connected-component labeling. *Pattern Recognit.* **2009**, *42*, 1977–1987. [[CrossRef](#)]
33. Morera, Á.; Sánchez, Á.; Moreno, A.B.; Sappa, Á.D.; Vélez, J.F. SSD vs. YOLO for detection of outdoor urban advertising panels under multiple variabilities. *Sensors* **2020**, *20*, 4587. [[CrossRef](#)] [[PubMed](#)]
34. Lu, C.-P.; Liaw, J.-J. A novel image measurement algorithm for common mushroom caps based on convolutional neural network. *Comput. Electron. Agric.* **2020**, *171*, 105336. [[CrossRef](#)]
35. Su, H.-H.; Pan, H.-W.; Lu, C.-P.; Chuang, J.-J.; Yang, T. Automatic detection method for cancer cell nucleus image based on deep-learning analysis and color layer signature analysis algorithm. *Sensors* **2020**, *20*, 4409. [[CrossRef](#)]

36. Wang, C.-Y.; Mark Liao, H.-Y.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPR Workshop), Seattle, WA, USA, 14–19 June 2020. [[CrossRef](#)]
37. Misra, D. Mish: A self regularized non-monotonic activation function. *arXiv* **2020**, arXiv:1908.08681. [[CrossRef](#)]
38. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-iou loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence 2020, New York, NY, USA, 7–12 February 2020. [[CrossRef](#)]
39. LabelImg. Available online: <https://github.com/tzutalin/labelImg> (accessed on 27 July 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.