2024 IEEE International Conference on Image Processing (ICIP) | 979-8-3503-4939-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIP51287.2024.10648237

ENHANCING INTUBATION ACCURACY: ADVANCED TRACHEAL SEGMENTATION TECHNIQUES IN VIDEO ENDOSCOPY

Adel Oulefki[1], *Abbes Amira*[2], *Fatih Kurugollu*[2], *Thaweesak Trongtirakul*[3], *Sos Agaian*[4], Menen Kassim Mohammed[2], Mohammad Alshoweky[2]

[1]Research Institute of Sciences & Engineering (RISE), University of Sharjah (UoS), Sharjah, UAE
[2]Computer Science Department, University of Sharjah (UoS), Sharjah, UAE
[3]Rajamangala University of Technology Phra Nakhon (RMUTP), Bangkok, Thailand
[4]Computer Science Department, College of Staten Island, CUNY, USA

ABSTRACT

Tracheal intubation is a critical medical procedure involving the insertion of a tube into the trachea to maintain an open airway. While essential, this procedure carries significant risks, such as incorrect tube placement. Advances in visually guided intubation methods, like video laryngoscopy, have enhanced safety by enabling precise tracheal segmentation from endoscopic images. Our study introduces an innovative image enhancement technique for video endoscopy that significantly improves tracheal visibility and segmentation accuracy. This novel approach not only facilitates safer and more accurate intubation but also minimizes patient discomfort and procedural risks. Tested against the UoS Dataset and real patient data from thyroidectomy procedures, our method demonstrated superior performance, achieving a segmentation accuracy of 97%, a precision of 94%, and a recall of 99%. Our tailored method is computationally efficient, making it suitable for implementation on edge devices like Arduino, thereby enhancing intubation safety and efficiency in various medical settings.

Index Terms— Video Endoscopy, Medical Imaging Enhancement, Trachea Segmentation, Intubation.

1. INTRODUCTION

The human upper airway comprises the oral cavity, pharynx, and its subdivisions, including the nasopharynx, oropharynx, hypopharynx, and larynx, all of which play a crucial role in respiratory health [1]. These structures are responsible for humidifying and warming inhaled air and are vascularized by the external and internal carotid arteries.

Sensory innervation of the nasopharynx and oropharynx is provided by the trigeminal, facial, and glossopharyngeal nerves. The trachea's unique structure, including its size and cartilaginous rings, is essential for distinguishing it during intubation, as depicted in Figure 1. Tracheal intubation is a critical emergency procedure to manage the airway and aid in ventilation, but it carries risks such as asphyxia or pulmonary aspiration. The procedure's success is crucial, and visually guided methods like video laryngoscopy have been introduced to reduce complications associated with emergency intubations. These technologies, especially with the latest developments in computer vision and artificial intelligence, have revolutionized medical imaging, including tracheal segmentation during intubation [2]. Tracheal segmentation identifies and isolates the tracheal region from endoscopic images, facilitating accurate and safe intubation.

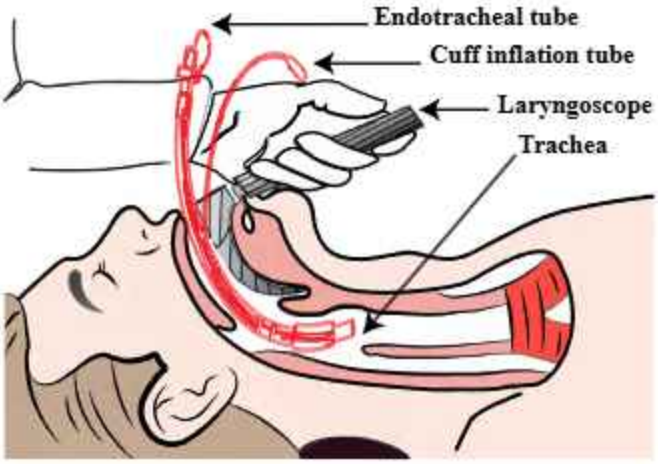

Fig. 1: Tracheal intubation process

Various computational techniques have been proposed for enhancing laryngeal and tracheal image analysis using deep learning and convolutional neural networks. Table 1 provides a summary of these studies, highlighting their methodologies, results, and potential limitations. While these methods have contributed significantly to the field, there remains a need for techniques that are both highly adaptable to diverse imaging conditions and capable of achieving high accuracy in complex scenarios. Thus, the critical contributions of our study include:

- Image enhancement technique as a pre-processing step, significantly improving the descriptors and clarity and definition of tracheal structures for more accurate segmentation.
- The proposed method effectively addresses challenges of variability and sub-optimal quality in endoscopic images, such as noise, blur, and illumination.
- tube insertion pathways post-segmentation, crucial for video endoscopy precision. The enhancement algorithm sharpens boundary definition, significantly improving segmentation. Additional video frame analyses are detailed is available in our GitHub repository Link.

These techniques cover various aspects of tracheal segmentation, such as datasets, preprocessing methods, and performance metrics. However, despite their innovations, they have significant limita-

Table 1: Summary of Studies on Tracheal Segmentation and Detection Methods

Study	Method Used	Dataset	Results Obtained	Weaknesses
Hamad et al. [3]	Deep learning-based fully automated video segmentation	Laryngeal endoscopy videos	Promising performance on human subjects despite image variations. Aimed for high-throughput, quantitative Vocal Fold (VF) motion behavior analysis.	Potential limitations in handling diverse pathological conditions or extreme anatomical variations.
Adamian et al. [4]	Computer vision tool for automatic quantitative tracking	Video laryngoscope	Demonstrated high accuracy in glottic opening angle estimation and its potential in diagnosing unilateral vocal fold paralysis.	Performance may be suboptimal under varied recording conditions or in presence of laryngeal lesions; not designed for stroboscopic examinations.
Matava et al. [5]	Convolutional Neural Networks (CNNs) - ResNet, Inception, MobileNet	775 laryngoscope and bronchoscopy videos	Inception CNN showed high performance in simulated live video. Aims to assist in airway management procedures.	Potential issues in adapting to a larger or more diverse dataset; may generate more false positives with confidence adjustments.
Ren et al. [6]	Deep-learning-based CAD system	24,667 laryngoscopy images	Achieved 96.24% overall accuracy for classifying laryngeal conditions (leukoplakia, benign, malignancy, normal, vocal nodule).	Limited to five major laryngeal conditions, not tested on rare diagnostic entities. Small vocal nodules were not biopsied or surgically verified for database inclusion.
Cho et al. [7]	CNN-based Image Classification System	4106 images of various laryngeal diseases	Demonstrated high accuracy and consistency in diagnoses, with potential as a supplementary diagnostic tool	Limited to the interpretation of still images, which might not capture the dynamic nature of clinical assessments. Potential for misdiagnosis due to overlapping morphologic features among laryngeal diseases.
Bandyopadhyay et al. [8]	Automated methods for blood vessel segmentation in Narrow Band Imaging (NBI) bronchoscopic video	NBI bronchoscopic video	Achieved superior results compared to existing methods with higher specificity and accuracy but lower sensitivity	Deep learning approach showed lower sensitivity compared to other methods, indicating potential challenges in detecting all relevant features in complex imaging conditions.

tions, particularly in adaptability to diverse image conditions and accuracy in complex imaging scenarios. Our paper introduces a novel approach to tracheal imaging in video endoscopy, characterized by an advanced image enhancement technique applied before segmentation. This technique significantly improves the visibility and definition of tracheal structures, enabling more precise segmentation.

The paper is organized as follows: Section 1 presents a comprehensive review of related work in tracheal segmentation and detection methods. Section 2 describes the proposed methodology, and Section 3 discusses the results and performance evaluation. Section 4 concludes with a summary of key findings and potential future research directions.

2. PROPOSED METHOD

In medical image processing, especially for tracheal segmentation in video endoscopy, enhancing image quality is crucial for effective segmentation. To address this, we introduce a novel image enhancement technique called Bi-Logarithmic Histogram Equalization with Quasi Correction Functions (BLHEQCFs), inspired by recent paper in the field [9]. This method significantly improves the quality and distinctiveness of visual information in endoscopic images, leading to more accurate and reliable segmentation results.

Figure 2 presents a flowchart of the proposed method. The process begins with the input of imagery, which is then enhanced to highlight specific features. This is followed by applying the segmentation phase, where a mask is generated to identify distinct regions in the image. The output is a set of resulting images after overlaying the segmented mask and generating the optimal path. Lastly, we compare the segmented results against a ground truth.

2.1. Enhancement method using Bi-Logarithmic Histogram Equalization with Quasi Correction Functions

Image enhancement is crucial for image segmentation as it significantly improves the quality and distinctiveness of visual information, leading to more accurate and reliable segmentation results. Enhanced images exhibit better contrast, reduced noise, and improved clarity, all of which contribute to the creation of feature-rich representations. These enhanced features are essential for segmentation algorithms to effectively differentiate between object boundaries, textures, and shapes. The heightened visibility of relevant details in enhanced images enables segmentation models to make informed decisions, resulting in improved precision and recall in identifying and delineating objects of interest. In this sub-section, we propose the Bi-Logarithmic Histogram Equalization with Quasi Correction Functions (BLHEQCFs), inspired by [9].

Let $I_{(i,j)}$ denote a given image, which is composed of x discrete intensity levels referred to as $\{x_0, x_1, \ldots, x_{(L-1)}\}$. The given image can be separated by an image-dependent threshold, x_t, and assume that $x_t \subset \{x_0, x_1, \ldots, x_{(L-1)}\}$. The transformation function can be described as:

$$f(x) = \begin{cases} x_0 + \alpha_L (x_t - x_0) e^{c(x)\log((\alpha_L+\varepsilon)/\alpha_L)}, \\ \text{if } x < x_t \\ x_t + \alpha_U (x_{(L-1)} - x_t) e^{c(x)\log((\alpha_U+\varepsilon)/\alpha_U)}, \\ \text{if } x \geq x_t \end{cases} \tag{1}$$

where x_0 refers to the minimum intensity level of an image in a permitted range, $x_{(L-1)}$ represents the maximum intensity level of an image in a permitted range, $c(x)$ is a cumulative density function of an image, α_L and α_U are constants, and ε is an offset value.

Quasi-symmetric correction functions play a significant role against over-enhanced and artifacts in enhanced images. By strategically mitigating the excursiveness of enhancements, these correction functions contribute to maintaining the integrity and visual fidelity of images, ensuring a balanced and realistic representation of the enhanced content. The Quasi-symmetric correction functions can be expressed as:

$$Q_L(x) = \begin{cases} (f(x))^{\gamma}, & \text{if } x < x_t \\ f(x), & \text{if } x \geq x_t \end{cases} \tag{2}$$

$$Q_U(x) = \begin{cases} 1 - (1 - f(x))^{\gamma^{-1}}, & \text{if } x < x_t \\ f(x), & \text{if } x \geq x_t \end{cases} \tag{3}$$

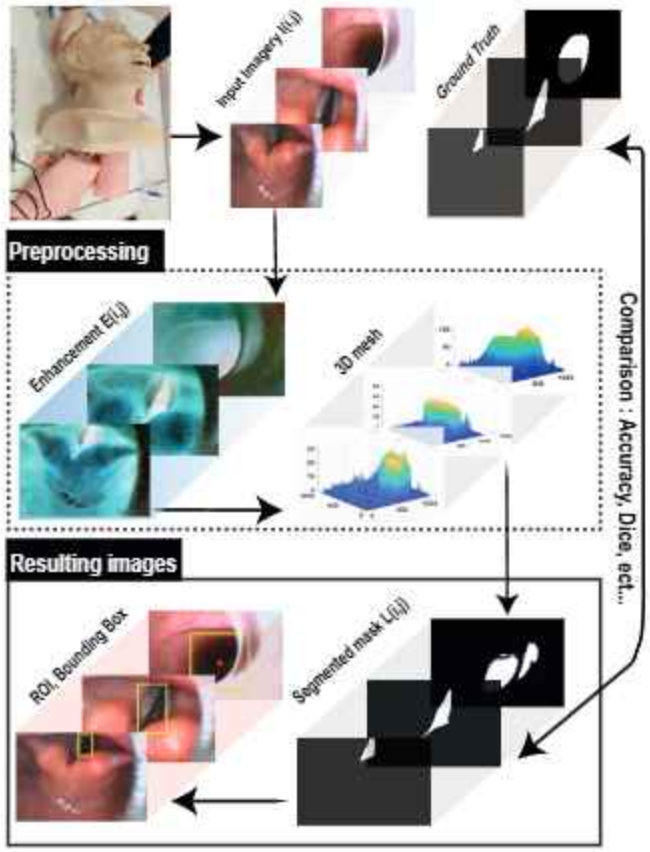

Fig. 2: Flowchart of the proposed enhancement and segmentation method

where γ denotes a gamma parameter.

2.2. Image decomposition using Agaian-Trongtirakul Entropy (KL2 Entropy) [9, 10]

In this subsection, the decomposition of an image into dark and bright regions is performed using the KL2 Entropy. The extended Kapur Entropy is defined as follows:

$$x_\gamma =_x \{a(x) + b(x)\} \tag{4}$$

The image-dependent threshold is represented by x_γ. It is assumed that $x_\gamma \subset \{x_0, x_1, \ldots, x_{L-1}\}$ and $L = 2^B$. B denotes the depth bit image per channel. For RGB-images, B can be set as 8. The input image is then decomposed into two sub-images $I_L(x)$ and $I_U(x)$ separated by x_γ as:

$$\begin{aligned} a(x) &= (I_L(x) + \varepsilon)^\gamma \cos(\log(\log(I_L(x) + \varepsilon))); \\ I_L(x) &= \{I_{i,j} | I_{i,j} \leq x_\gamma\} \end{aligned} \tag{5}$$

$$\begin{aligned} b(x) &= (I_U(x) + \varepsilon)^\gamma \cos(\log(\log(I_U(x) + \varepsilon))); \\ I_U(x) &= \{I_{i,j} | I_{i,j} > x_\gamma\} \end{aligned} \tag{6}$$

It should be noted that $I_L(x)$ consisted of $\{x_0, x_1, \ldots, x_\gamma\}$ and $I_U(x)$ consists of $\{x_{\gamma+1}, x_{\gamma+2}, \ldots, x_{L-1}\}$, $I_{i,j}$ refers to a given image, and ε denotes an offset value to avoid a logarithmic calculation error.

The combination of the Quasi-symmetric correction functions provides a complete transformation function, providing a refined and balanced approach to address unpleasant artifacts and enhance overall performance. The complete Quasi-symmetric correction function can be expressed as:

$$Q(x) = \omega Q_L(x) + (1 - \omega) Q_U(x) \tag{7}$$

where ω denotes an adaptive weight, $Q_L(x)$ and $Q_U(x)$ represent the Quasi-symmetric correction functions of a dark component, $X_L = \{\forall x | x_0, x_1, \ldots, x_t\}$, and a bright component, $X_U = \{\forall x | x_{(t+1)}, x_{(t+2)}, \ldots, x_{(L-1)}\}$. In this paper, the adaptive weight, ω, is set to the value of 0.80. The proposed algorithm can be written in Algorithm 1.

Algorithm 1 Bi-Logarithmic Histogram Equalization with Quasi Correction Functions (BLHEQCFs)

Require: $I_{(i,j)}$ is a given image.
Initialize $\alpha_L, \alpha_U \leftarrow 0.95, \varepsilon \leftarrow 1.0, \omega \leftarrow 0.80$.
Read a given image, $I_{(i,j)}$.
Generate the cumulative density function, $c(x)$.
Calculate an image-dependent threshold, x_t.
Generate the bi-logarithmic transformation function:
if $x < x_t$ **then**
$f(x) \leftarrow x_0 + \alpha_L (x_t - x_0) e^{c(x)\log((\alpha_L+\varepsilon)/\alpha_L)}$
else
$f(x) \leftarrow x_t + \alpha_U (x_{(L-1)} - x_t) e^{c(x)\log((\alpha_U+\varepsilon)/\alpha_U)}$
end if
Calculate the Quasi-symmetric correction function for a dark component, $X_L \leftarrow \{\forall x | x_0, x_1, \ldots, x_t\}$:
if $x < x_t$ **then**
$Q_L(x) \leftarrow (f(x))^\gamma$
else
$Q_L(x) \leftarrow f(x)$
end if
Calculate the Quasi-symmetric correction function for a bright component, $X_U \leftarrow \{\forall x | x_{(t+1)}, x_{(t+2)}, \ldots, x_{(L-1)}\}$:
if $x < x_t$ **then**
$Q_U(x) \leftarrow 1 - (1 - f(x))^{\gamma^{-1}}$
else
$Q_U(x) \leftarrow f(x)$
end if
Generate the combination of the Quasi-symmetric correction functions:
$Q(x) \leftarrow \omega Q_L(x) + (1 - \omega) Q_U(x)$
Apply the combination function to the given image, $I_{(i,j)}$:
$E_{(i,j)} \leftarrow 0.5(I_{(i,j)} + Q(I_{(i,j)}))$
Ensure: $E_{(i,j)}$ is an enhanced image.

The 0.8 value is chosen to balance computational efficiency with accuracy, effectively minimizing noise impact while maintaining image processing robustness. Where I_L and I_H represent the intensity values corresponding to the 10^{th} and 90^{th} percentiles of the entire pixel distribution within the given image, respectively. Let N represent the maximum intensity value. Then, we can define:

$$I_H = 0.9N$$

$$I_L = 0.1N$$

Therefore, we calculate ω as follows:

$$\omega = \frac{0.9N - 0.1N}{0.9N + 0.1N} = 0.8$$

We can fix $\omega = 0.8$ to reduce the computational complexity. It is commonly used in practical applications.

2.3. Segmentation

Watershed image segmentation [11] is a technique used in image processing to partition an image into regions based on the topology of the image intensity surface. The basic idea is to treat the pixel intensity values as a topographical surface, where pixels with similar intensity values form regions separated by boundaries. The modified algorithm for segmentation is described in Algorithm 2.

Algorithm 2 Segmentation using Watershed Transformation

Require: $E_{(i,j)}$ is an enhanced image.
Initialize $r_1 \leftarrow 3, r_2 \leftarrow 2$.
Read an input image, $E_{(i,j)}$.
Compute the gradient of the image to highlight the edges and boundaries:
$G_{(i,j)} \leftarrow \sqrt{G^2_{x(i,j)} + G^2_{y(i,j)}}$; G_x is a horizontal gradient, and G_y is a vertical gradient.
Apply morphological operations:
$M_{1(i,j)} \leftarrow \text{Erosion}(E_{(i,j)}, r_1)$
$M_{3(i,j)} \leftarrow \text{Binarization}(M_{1(i,j)})$
Compute the distance transform of the image, $\delta_{(i,j)}$.
Apply the watershed transformation to the distance transformed image, $\delta_{(i,j)}$.
$M_{4(i,j)} \leftarrow \text{Watershed}(\delta_{(i,j)})$
$M_{5(i,j)} \leftarrow \text{RegionalMinima}(G_{(i,j)}, M_{4(i,j)} | M_{3(i,j)})$
Apply the watershed transformation to the regional minima image, $M_{5(i,j)}$.
$w_{(i,j)} \leftarrow \text{Watershed}(M_{5(i,j)})$
Calculate the regional mean, μ, of $w_{(i,j)}$.
Calculate the total number of the segmented image, $w_{(i,j)}$.
for $a = 1$ to N **do**
  if $\mu_a > t$ **then**
    $\chi \leftarrow a$
    $l_{(i,j)}(w_{(i,j)} = \chi) \leftarrow E_{(i,j)}(w_{(i,j)} = \chi)$
  end if
end for
$L_{(i,j)}(l_{(i,j)} > \max(l_{(i,j)})/2) \leftarrow 1$
Apply the regional filling algorithm to the segmented image, $L_{(i,j)}$.
$L_{(i,j)} \leftarrow \text{Fill}(L_{(i,j)})$
Ensure: $L_{(i,j)}$ is a segmented image.

3. RESULTS

This section presents a comparative analysis of enhancement and segmentation approaches both statistically and visually, alongside showcasing the optimal pathway for tracheal intubation. We begin with a statistical evaluation of our proposed enhancement method against histogram equalization (HE) and CLAHE methods [12] using BRISQUE and NIQE metrics, demonstrating its superior performance. A similar statistical comparison is conducted for segmentation performance, employing metrics such as Accuracy, Precision, Specificity, F1 Score, and Dice metric (DSC), where our method outperforms existing approaches.

Visual assessments further validate the effectiveness of the enhancement and segmentation techniques. The comparative visual performance of image enhancement methods is illustrated, highlighting the improvements in image quality. Additionally, visual segmentation assessment using samples from diverse databases showcases the precision of our proposed segmentation method.

Finally, we detail how our segmentation approach assists in the tracheal intubation process, guiding medical professionals by identifying the optimal path for tube insertion. This is achieved through the use of Region of Interest (ROI) and Bounding Box techniques, facilitating a more accurate and expedient intubation procedure.

3.1. Statistical Enhancement Performance Comparison

The proposed segmentation approach was rigorously evaluated and compared with HE and CLAHE methods. We employed BRISQUE and NIQE metrics for a comprehensive statistical comparison. The results, as shown in Table 2, demonstrate that our proposed method outperforms the existing approaches in terms of statistical quality metrics.

Table 2: Statistical comparative performance metrics using BRISQUE and NIQE of the enhanced images from UoS data and Styletubation data by Pan et al [13]. Best results are highlighted in red, indicating optimal performance, and worst results in blue. Lower scores are better.

Metric	Original	HE	CLAHE	Enhanced
BRISQUE [14]	56.087	73.886	62.382	25.918
NIQE	2.6078	3.5577	6.1293	1.6288

3.2. Statistical Segmentation Performance Comparison

The proposed segmentation approach was rigorously evaluated and compared with state-of-the-art methods. We employed several metrics for a comprehensive statistical comparison, including Accuracy, Precision, Specificity, F1 Score, and Dice metric (DSC). The results, as shown in Table 3, demonstrate that our proposed method outperforms the existing approaches in terms of accuracy and other significant metrics.

Table 3: Comparative Performance Metrics of Segmentation: Best results are highlighted in red, and second-best results in blue. The results of the proposed method are derived from the aggregate analysis of both datasets. Methods used: M1 - EfficientNet-B5 with DeepLabv3+ [15]; M2 - EfficientNet-B5 with U-Net [16]; M3 - Configured Mask R-CNN [15]; M4 - Trainable Weka Segmentation [17]; M5 - Segment Anything Model (SAM) [18]; Proposed - Proposed Segmentation Method. Metrics included are Accuracy (Acc.), Precision (Prec.), Specificity (Spec.), F1 Score (F1), and Dice metric (DSC).

Model	Acc.	Prec.	Spec.	F-1.	DSC.
M1	0.8283	0.9219	0.5896	0.8852	0.7660
M2	0.7991	0.9057	0.5333	0.8655	0.7395
M3	0.807	0.8011	0.8281	0.8669	0.7207
M4	0.6869	0.9660	N/A	0.4921	N/A
M5	0.6212	0.9791	N/A	0.7568	N/A
Proposed	0.9774	0.9488	0.9938	0.8227	0.8227

3.3. Visual Enhancement Comparative Performance

The effectiveness of the proposed image enhancement techniques can be qualitatively evaluated through visual comparison. In this subsection, we present the visual comparative performance metrics of image enhancement methods. The best results are observed when comparing the original image as a reference to the enhanced versions, where improvements in quality are evident. The results presented in figure 3 are derived from three random images selected from both Styletubation data and UoS data. The methods utilized for enhancement include: Histogram Equalization (HE) in the second row. Contrast Limited Adaptive Histogram Equalization (CLAHE) in the third row. Proposed Enhancement Method last row. Upon visual inspection, it is apparent that the proposed enhancement method significantly improves the visual features of the trachea. This enhancement contributes to better segmentation.

Fig. 3: Visual comparison of enhanced images: original image in the first column, He in the second row, CLAHE in the third row, and proposed method in the last four rows.

3.4. Visual Segmentation Assessment

To further validate the effectiveness of the proposed segmentation method, we conducted a visual assessment using samples from two distinct databases: one from manikin-based imagery (UoS data) and the other from a public dataset collected from the Styletubation study by Pan et al [13]. Three samples of visual segmentation results were generated from each data to demonstrate the effectiveness of the proposed. As illustrated in Figure 4, the proposed segmentation method demonstrates superior performance, where the segmented areas (marked in green) almost perfectly intersect with the Ground Truth (GT) images (marked in red). This visual alignment reaffirms the accuracy and reliability of our proposed method in practical scenarios.

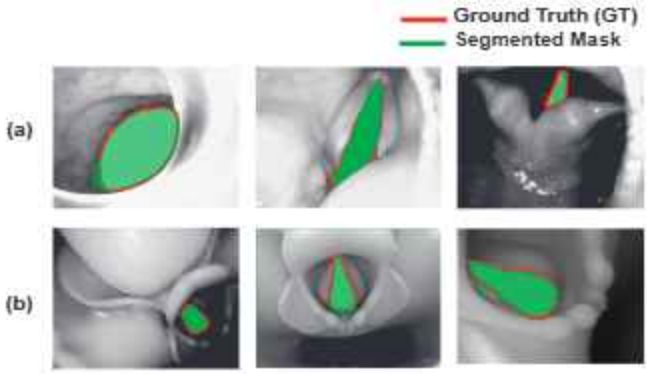

Fig. 4: Visual inspection; (a) Styletubation Data, (b) UoS Data

The implementation of the proposed scheme was carried out on a system with the following specifications: a $13th$ Gen Intel(R) Core(TM) $i7 - 1365U$ processor operating at 1.80 GHz, 16.0 GB of RAM $(15.7GB$ usable), and a 64-bit operating system with an $x64$-based processor. The system runs on Windows 10 Enterprise Edition (Version $22H2$). The development and computational tasks were performed using the latest MATLAB 2024, provided by the University of Sharjah (UoS).

3.5. Assistance in Tracheal Intubation Pathway

The implementation of our segmentation approach facilitates the tracheal intubation process by assisting medical professionals in identifying the optimal path for tube insertion. Our method utilizes the Region of Interest (ROI) and Bounding Box techniques to pinpoint key areas in the trachea that are critical for successful intubation. Once the segmentation is complete, the algorithm calculates the centroid of the largest connected component in the segmented image. This centroid, marked by a red point ('drawpoint' function) in Matlab on the original image, suggests an optimal entry point or pathway for the intubation tube. Additionally, a bounding box is drawn around this largest component, providing a visual guide for the area of focus during intubation. As shown in Figure 5

Fig. 5: ROI and bounding box; (a) Styletubation study Data, (b) UoS Dataset

This guidance is particularly beneficial in complex or emergency scenarios where precise and rapid intubation is crucial. By highlighting the key regions and suggesting an entry point, our system aims to enhance the accuracy and speed of the tracheal intubation procedure.

4. CONCLUSION

This study has successfully developed an advanced method for tracheal segmentation in Video Endoscopy, a critical advancement for improving intubation techniques. By effectively segmenting both the full larynx and the glottis, the algorithm has demonstrated a significant enhancement in tracheal identification accuracy and efficiency. This innovation is pivotal in guiding the endoscope tip precisely, thereby aiming to minimize patient discomfort and reduce procedural complications. Employing a dataset of videoendoscopy frames from a UoS Dataset and supplemented by the Styletubation data, the algorithm has achieved decent performance metrics. These results underscore the method effectiveness across various data sources and its substantial potential for clinical application. Furthermore, the integration of video laryngoscopy into this framework underscores its value in clinical settings. As a cost-effective and efficient solution, video laryngoscopy enhances the management of difficult airways, shortens intubation times, and improves success rates, either as a standalone tool or in conjunction with other devices.

References

[1] Qiwei Xiao, Neil J Stewart, Matthew M Willmering, Chamindu C Gunatilaka, Robert P Thomen, Andreas Schuh, Guruprasad Krishnamoorthy, Hui Wang, Raouf S Amin, Charles L Dumoulin, et al., "Human upper-airway respiratory airflow: In vivo comparison of computational fluid dynamics simulations and hyperpolarized 129xe phase contrast mri velocimetry," *Plos one*, vol. 16, no. 8, pp. e0256460, 2021.

[2] Pablo Gómez, Andreas M Kist, Patrick Schlegel, David A Berry, Dinesh K Chhetri, Stephan Dürr, Matthias Echternach, Aaron M Johnson, Stefan Kniesburges, Melda Kunduk, et al., "Bagls, a multihospital benchmark for automatic glottis segmentation," *Scientific data*, vol. 7, no. 1, pp. 186, 2020.

[3] Ali Hamad, Megan Haney, Teresa E Lever, and Filiz Bunyak, "Automated segmentation of the vocal folds in laryngeal endoscopy videos using deep convolutional regression networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[4] Nat Adamian, Matthew R Naunheim, and Nate Jowett, "An open-source computer vision tool for automated vocal fold tracking from videoendoscopy," *The Laryngoscope*, vol. 131, no. 1, pp. E219–E225, 2021.

[5] Clyde Matava, Evelina Pankiv, Sam Raisbeck, Monica Caldeira, and Fahad Alam, "A convolutional neural network for real time classification, identification, and labelling of vocal cord and tracheal using laryngoscopy and bronchoscopy video," *Journal of medical systems*, vol. 44, pp. 1–10, 2020.

[6] Jianjun Ren, Xueping Jing, Jing Wang, Xue Ren, Yang Xu, Qiuyun Yang, Lanzhi Ma, Yi Sun, Wei Xu, Ning Yang, et al., "Automatic recognition of laryngoscopic images using a deep-learning technique," *The Laryngoscope*, vol. 130, no. 11, pp. E686–E693, 2020.

[7] Won Ki Cho, Yeong Ju Lee, Hye Ah Joo, In Seong Jeong, Yeonjoo Choi, Soon Yuhl Nam, Sang Yoon Kim, and Seung-Ho Choi, "Diagnostic accuracies of laryngeal diseases using a convolutional neural network-based image classification system," *The Laryngoscope*, vol. 131, no. 11, pp. 2558–2566, 2021.

[8] Saptarashmi Bandyopadhyay, Vahid Daneshpajooh, Danish Ahmad, Jennifer Toth, Rebecca Bascom, and William E Higgins, "Blood vessel segmentation in narrow band imaging bronchoscopic video," in *Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging*. SPIE, 2021, vol. 11600, pp. 415–423.

[9] Thaweesak Trongtirakul and Sos Agaian, "New retinex model-based infrared image enhancement," in *Multimodal Image Exploitation and Learning 2023*. SPIE, 2023, vol. 12526, pp. 34–45.

[10] Samir Benbelkacem, Adel Oulefki, Sos Agaian, Nadia Zenati-Henda, Thaweesak Trongtirakul, Djamel Aouam, Mostefa Masmoudi, and Mohamed Zemmouri, "Covi3d: Automatic covid-19 ct image-based classification and visualization platform utilizing virtual and augmented reality technologies," *Diagnostics*, vol. 12, no. 3, pp. 649, 2022.

[11] Jun Zhou, Yue Yin, and Shuai Wang, "Image segmentation based on watershed algorithm," in *2021 International Conference on Intelligent Computing, Automation and Applications (ICAA)*. IEEE, 2021, pp. 10–13.

[12] Krishna Gopal Dhal, Arunita Das, Swarnajit Ray, Jorge Gálvez, and Sanjoy Das, "Histogram equalization variants as optimization problems: a review," *Archives of Computational Methods in Engineering*, vol. 28, pp. 1471–1496, 2021.

[13] Hui-Shan Pan, Tiffany Corey, Hsiang-Ning Luk, Jason Zhensheng Qu, and Alan Shikani, "Combined styletubation with videolaryngoscopy for tracheal intubation in patients undergoing thyroidectomy with intraoperative neuromonitoring," *Anesthesia Research*, vol. 1, no. 1, pp. 8–23, 2023.

[14] Ziga Babnik and Vitomir Štruc, "Assessing bias in face image quality assessment," in *2022 30th European Signal Processing Conference (EUSIPCO)*. IEEE, 2022, pp. 1037–1041.

[15] Seung Jae Choi, Dae Kon Kim, Byeong Soo Kim, Minwoo Cho, Joo Jeong, You Hwan Jo, Kyoung Jun Song, Yu Jin Kim, and Sungwan Kim, "Mask r-cnn based multiclass segmentation model for endotracheal intubation using video laryngoscope," *Digital Health*, vol. 9, pp. 20552076231211547, 2023.

[16] MT Vyshnav, V Sowmya, EA Gopalakrishnan, Sajith Variyar VV, Vijay Krishna Menon, et al., "Deep learning based approach for multiple myeloma detection," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2020, pp. 1–7.

[17] Mr Nitin Kanuri, Ahmed Abdelkarim, and Sonali A Rathore, "Trainable weka segmentation tool machine learning-enabled segmentation on features of orthopantomograms," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 134, no. 3, pp. e77, 2022.

[18] Lv Tang, Haoke Xiao, and Bo Li, "Can sam segment anything? when sam meets camouflaged object detection," *arXiv preprint arXiv:2304.04709*, 2023.