Deep Learning Airway Structure Identification for Video Intubation

Ben Barone, BSEE, EMT-P [1], Griffin Milsap, PhD[2], and Nicholas M Dalesio, MD, MPH [3]
[1] Johns Hopkins Applied Biomedical Engineering [2] Johns Hopkins Applied Physics Laboratory
[3] Department of Anesthesiology/ Critical Care Medicine; Johns Hopkins University

Background

Endotracheal Intubation (EI) remains the gold standard for advanced airway management, however, it can have serious complications, especially when performed by inexperienced providers in non-hospital settings, such as in an ambulance or on the battlefield. Video laryngoscope (VL) systems display real-time video captured by a camera at the tip of the laryngoscope blade. The goal of this project was to collect videos of EI procedures, then apply machine learning techniques to identify airway structures in those videos.

Methods

We worked with a team of anesthesiologists at Johns Hopkins Hospital to record EI procedures. Patients presenting for surgery were consented under a protocol approved by the Johns Hopkins University IRB. During routine intra-operative care, anesthesiologists recorded EI using the Storz CMAC ® VL with age-appropriate laryngoscope blades. We trimmed each video to eliminate PII, then utilized the OpenCV Computer Vision Annotation Tool (CVAT) to label each pixel in a frame as part of a key airway structure, such as the glottic opening, vocal cords, epiglottis, and arytenoid cartilages. We used an open-source deep learning tool developed by Berkeley AI Research (BAIR) to analyze airway images. The approach, Pix2Pix[1], is an image-to-image translation technique based on a conditional Generative Adversarial Network (cGAN), which generates a target image based on a given input image. Our application used a Pix2Pix network to generate airway structure labels based on given airway images, effectively segmenting key airway structures. We trained the Pix2Pix network using 62 non-redundant image-label pairs from 4 different videos. We tested the Pix2Pix network with frames that were held out from the training process.

Results

Videos were collected from 17 patients. Patients were between 1 month and 38 years old with a mean age of 6.87 years. Patients were 70.6% male and did not have any prior diagnosis of airway comorbidities. Compared to ground truth labels, Pix2Pix generated labels had a Structural Similarity of 0.58 and a Root Mean Squared Error of 0.52. Figure 1 (left) is an image collected during a VL intubation. Figure 1 (center) displays airway structure labels generated by the Pix2Pix network. Figure 1 (right) displays ground truth labels of airway structures, which were manually labeled by our team. Figure 2 provides a map of label colors to airway structures.

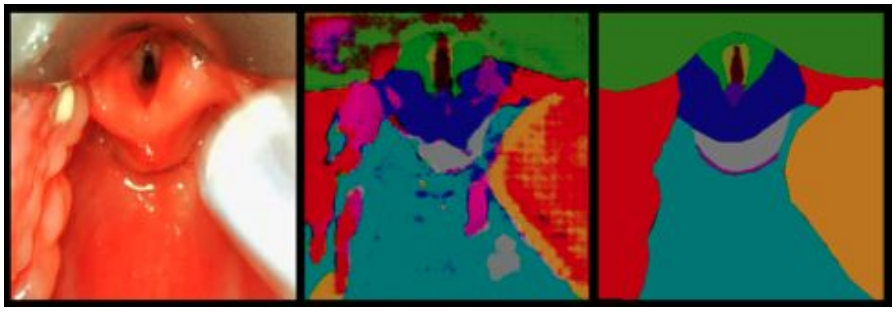

Figure 1: An original airway image (left), Pix2Pix generated labels during testing (center), and ground truth airway labels (right).

Figure 2: Map of label colors to airway structures.

Discussion

These results are promising given that there were 11 classes present in the held-out frames, many of those classes were visually similar, there was a large variation in structural appearance from patient to patient, and 62 frames was a low number of samples to effectively train a deep learning segmentation approach. We hope to expand on these results by improving our image labeling throughput – it currently takes up to 30 minutes to label a single frame. Additionally, we hope to adjust model parameters or add a post-processing step to resolve the label spattering seen in the Pix2Pix generated results.

Conclusion

Image-to-image translation using pix2pix can identify airway structures captured during VL intubation. Using this approach to label airway structures in real-time could help inexperienced healthcare providers perform successful intubations. Additionally, computer identification of airway structures could provide a framework for a video-guided autonomous or semi-autonomous intubation system in the future.

References

[1]Isola, Phillip, et al. “Image-to-Image Translation with Conditional Adversarial Networks.” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, doi:10.1109/cvpr.2017.632.