Behavioral Analysis of Airway Deformation during Drug Induced Sleep Endoscopy

Shane Transue, Kevin Macfarlane, and Min-Hyung Choi†
Department of Engineering, Design and Computing
University of Colorado Denver
Denver, USA
{shane.transue, kevin.macfarlane, min.choi}@ucdenver.edu

2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-0358-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/BHI50953.2021.9508571

***Abstract*—Drug Induced Sleep Endoscopy (DISE) is a procedure that simulates natural sleep which allows medical experts to obtain a dynamic assessment of the upper airway to monitor and diagnose obstructed breathing. DISE diagnosis is based on visual examination of the obtained endoscopic images that uses a subjective scoring system to diagnose reoccurring obstructive respiratory events. Due to this subjectivity, objective measurements of breathing patterns and obstructions are difficult to reliably obtain. Cine Magnetic Resonance Imaging and Computed Tomography (MRI/CT) scans are considered useful tools for both static and dynamic evaluation of OSA, but are both complex to perform and expensive. The lack of objective measurements causes difficulty for medical professionals for accurately understanding and examining the conditions of chronic airway conditions. In this work, we describe an integration of machine learning and DISE monocular video with Cine MRI data to construct dynamic 3D models, providing an objective measure of airway behaviors and conditions. Dynamic changes in airway dimensions that occur with the respiratory cycle during sleep can be reconstructed and represented by a 3D airway model for prognosis assistance. This approach for Upper Aerodigestive Tract (UAT) and laryngotracheal airway reconstruction is accomplished through image-to-surface modeling using a Generative Adversarial Network (GAN) to integrate Cine MRI and video endoscopy. Results show that we are able to reconstruct the airway into 3D point-cloud data as well as model complex dynamics of airway surfaces.**

***Index Terms*—Airway modeling, Depth GAN, DISE, Cine-MRI**

I. Introduction

Upper airway behavioral evaluation is a critical process for identifying, monitoring, and diagnosing common pulmonary conditions such as Obstructive Sleep Apnea (OSA). Advances in vision-based modeling and deep transformation networks [10] have the capability to enable novel contributions to endoscopic imagery [3] but have not been widely adopted for precision airway measurement for behavioral analysis. This is because the quantification of the degree of airway collapse or obstruction is specifically challenging to obtain from monocular video for multi-level OSA and partial collapse events. Current OSA diagnosis procedures require an overnight sleep study documenting EEG signals and breathing patterns through polysomnography (PSG), allowing medical experts to classify severity but not the specific source(s) of obstruction. The goal of DISE is to simulate natural sleep to enable a dynamic assessment of the upper airway to identify these sources.

† This work is partially supported by Children's Hospital Colorado.

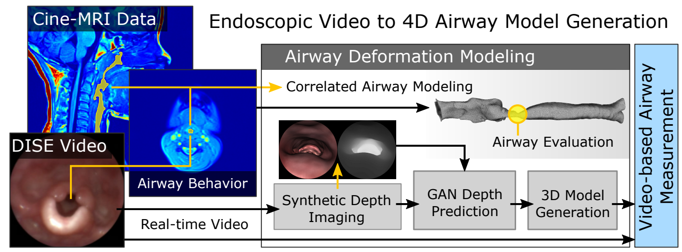

Fig. 1. Overview: Cine-MRI airway scan data is integrated with endoscopic DISE video to train a GAN network to predict depth images for 3D reconstruction. The reconstruction is then used to evaluate dynamic 3D airway restrictions and obstructions to reduce the number of required Cine-MRI scans.

However, using existing classification scoring systems such as VOTE (Velum, Oropharynx, Tongue base, and Epiglottis) [11], impose reliance on clinician judgement which is *inherently subjective in nature*, leading to inconsistent evaluation outcomes. This can lead to medical experts viewing the same video coming to different condition classifications based on their interpretation, which may hamper surgical planning and patient outcomes [1]. This indicates a diagnostic gap that needs to be addressed to allow for a truly objective, standardized, and quantitative evaluation of the upper airway during DISE. To partially address this, secondary scans (Cine-MRI) can be employed to identify obstruction sources and collapse severity. These scans provide a form of quantitative evaluation but are not widely clinically available and often limited due to expense. The objective of this work is to integrate these two forms of imaging into a trained modeling system that provides the first step towards a quantitative DISE evaluation framework that can be used with existing inexpensive devices (cameras) and procedures. As the initial link between DISE procedures, Cine-MRI and 3D reconstruction of the internal airways, challenges in obstruction identification, subjective measurements, and lacking behavioral analysis can be addressed through this form of integrative dynamic modeling.

To obtain the necessary data for dynamic airway reconstruction using DISE, we employ prior deep learning methods for estimating depth images from monocular video sequences through Generative Adversarial Networks (GANs) used to estimate 3D gastrointestinal (GI) tract models through endoscopic depth image estimation [3]. By employing synthesized data for depth-image training, these methods provide a viable direction towards the 3D reconstruction of internal structures where

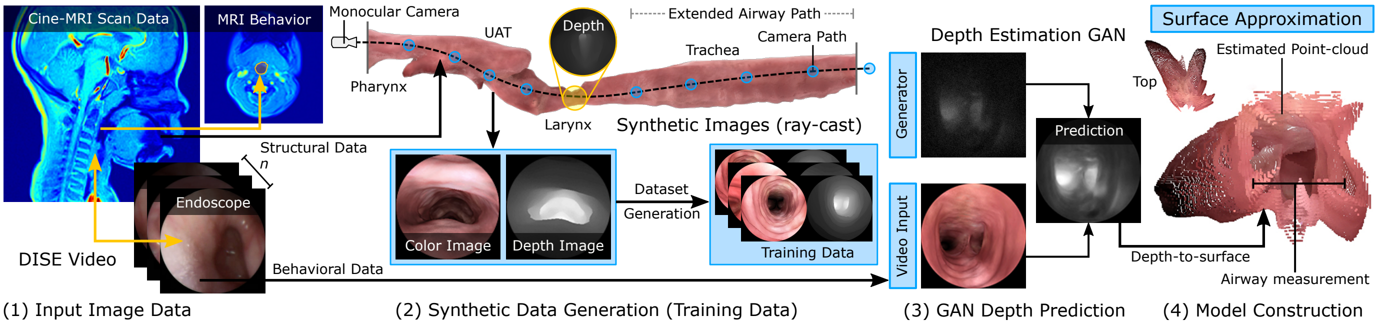

Fig. 2. Method overview. We obtain Cine-MRI data correlated with DISE video to model dynamic airway behaviors. Synthesized data is generated to train a GAN on DISE images to generate depth images of the airway for surface reconstruction. Dynamic behaviors can then be measured using the predictions.

images can only be obtained through endoscopic cameras. However, the anatomy and behavioral characteristics of airway evaluation is different than existing GI tract modeling approaches. Prior methods do not focus on the dynamic behavioral characterization of cross-sectional measurements that describe airway restrictions and are limited in terms of precise measurements of complex airway deformations. In our approach, we leverage GAN depth predictions with Cine-MRI behavioral data to provide better models for airway evaluations to obtain more accurate diagnostics. This enables us to generate models that can provide consistent quantitative measurements to address the subjective nature of current OSA related airway assessments. An overview of the imaging and reconstruction process is shown in Figure 2. Our reconstructed behavioral analysis aims to improve accuracy over subjective evaluations currently implemented by providing quantitative data through providing a 4D model which can assist in surgical planning as well as providing structural and behavioral deformation descriptors for pre-post intervention analysis.

II. Related Work

Numerous forms of surface reconstruction through general endoscopic and laryngoscopic imaging exist [4]; however, many have limited direct integration with DISE procedures for quantitative airway behavioral and deformation analysis. 3D anatomical reconstruction of the upper airway and respiratory tract are well studied including Computed Tomography (CT) scans [4], [15] with airflow evaluations through Computational Fluid Dynamics (CFD) [7]. To recreate 3D models, CT scan DICOM files are typically used to compose surface data (Marching Cubes) through 3D slices which allows for the generation of accurate *static* anatomical models. These provide the structural components of the upper airway, but are limited for *dynamic analysis*. To address the behavioral component, Cine-MRI imaging can be used for quantitative measurements [9], including airway restriction as shown in Figure 3. An advantage Cine-MRI holds over DISE is that it records multiple levels of the upper airway simultaneously and provides more objective measurements of the internal structure and behavior [6], [8]. But since DISE remains a widely available method for inspecting the upper airway, the classifications and assessments are still currently subjective [1], [5], [11] and difficult to directly tie to objective evaluation methods. To bridge the gap between subjective classification and objective measurement, we employ 3D reconstruction of internal structures through GAN-driven image transformation [3], [13], [14] with a specific focus on recording dynamic behavior to form 4D models composed from training data integrated from monocular video and Cine-MRI measurements.

III. Method

Airway measurement and UAT evaluation incorporates several different forms of image data to make quantitative assessments. In our approach, we integrate three forms of imaging to provide a multi-modal system to operate directly from DISE monocular video data. This includes CT scans for structural modeling, Cine-MRI for observing airway deformation patterns, and monocular endoscopic video sequences [2]. The aim is to utilize a trained depth-image generating GAN to predict depth images from endoscopic images to construct the internal dynamic structure of observable airways. This enables us to both reconstruct the visible geometry of the airway and perform various opening and behavioral measurements. The primary challenges in this approach include identifying ground-truth estimates of video acquired surface data, high variance in DISE videos with limited resolutions and image quality due to motion blur, and complex deformation behaviors in small enclosed areas. To overcome these problems, we employ a virtual method for generating synthesized ray-traced imaging data that generate both color and depth images for training a GAN to estimate endoscopic depth images. This provides a parallel to the real images obtained with a monocular camera that will be used by the trained GAN to estimate new depth images. We can then use the trained model to predict depth images that can be converted to point-clouds that approximate airway walls. The overall method is simplified

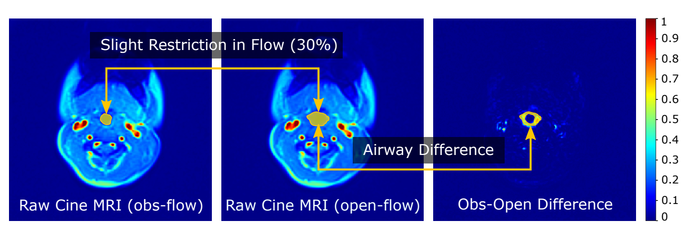

Fig. 3. Cine MRI scans illustrating the change in airway diameter demonstrating the difference between restricted (obstructed) and open airway flow. The difference in cross-section area is highlighted in the difference image.

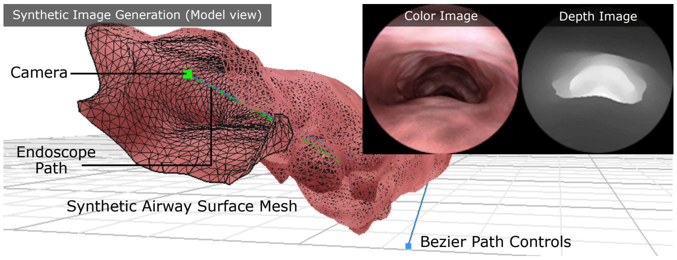

Fig. 4. Synthetic data generation setup: Ray-traced surface model with color and depth images generated for each position within an endoscope path.

into the following primary steps: (1) obtain real-world DISE and Cine-MRI data, (2) generate synthetic monocular video and depth data of virtual airway models for GAN training, (3) estimate depth maps for each frame, and (4) construct point-clouds surface states that define observable dynamic conditions with physical dimensions sourced from Cine-MRI data.

A. Synthetic DISE Data Generation

To train our model on the relationship between endoscopic and depth images for airway surface estimation, depth images are generated for the training process using a *synthetic* anatomical model [14]. We use a 3D surface model that represents the anatomy of a patient's airway to create the monocular video feed, as well as the depth images for each frame. This is achieved using a virtual camera which traverses through the 3D airway that is controlled by a Bezier curve path recording a ray-traced image for each step seen in Figure 4. Color and depth images are generated in batched parallel ray-tracing with camera attached light source with linear attenuation $(1/d)$. Depth noise is added to approximate surface uncertainty. Select color and depth training pairs are shown in Figure 5.

B. GAN Architecture and Training

The objective of the GAN is to take a monocular frame as input and output a corresponding depth prediction based on the training data. These depth images will be used to reconstruct airway surface structure. Due to complexity in tracking of surface features, handling specular reflections due to fluids, and modeling complex structures, this method reduces the problem space of ill-conditioned monocular frames by inferring structural relationships between color and depth data within our trained GAN model to estimate depth projections.

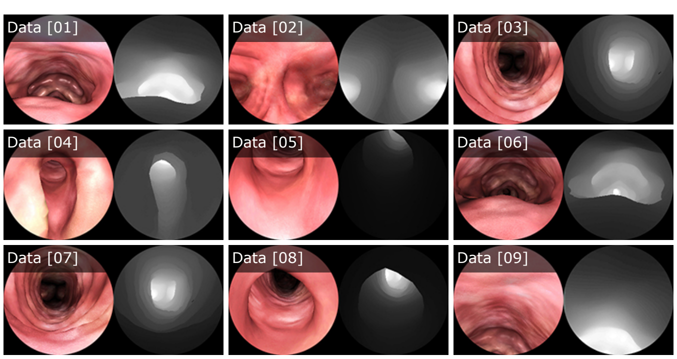

Fig. 5. Training data. Simulated endoscope data of the upper airway (256x512) for 10 datasets containing 100 images. Select samples are shown.

The GAN architecture is based on the extended Pix2Pix GAN [14]. It includes an encoder-decoder architecture with skip connections between each layer [14]. The generator consists of 8 encoder layers and 7 decoder layers in total, with the discriminator consisting of five layers. The GAN was trained on 484 images, with 126 test images. We train for 80 epochs using a batch size of 20 and a learning rate set to 0.0002. *LeakyRelu* is used in the convolution layers for the generator, with a leaky relu alpha of 0.2. Relu activation functions were used for deconvolution layers in the generator, with *tanh* being used for the output layer. We employed the ADAM optimizer [12] with $\beta_1 = 0.5$. Batch normalization originally included in Pix2Pix GAN was omitted from our model. Resulting depth predictions are shown in Figure 6.

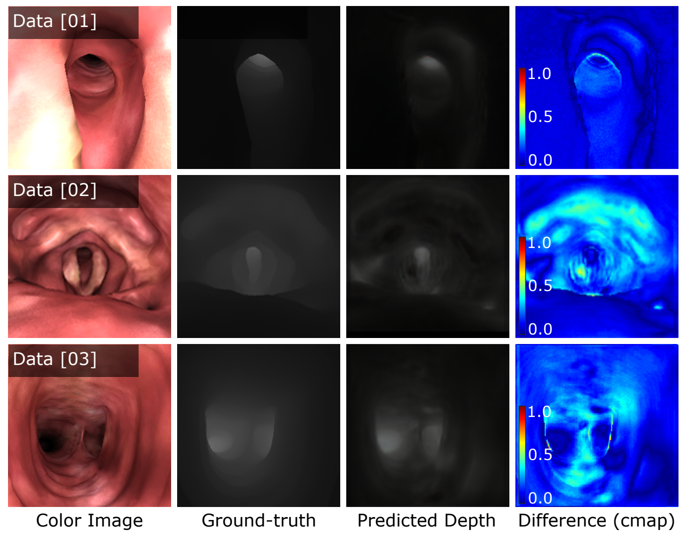

Fig. 6. GAN depth predictions on synthetic data. Difference is computed as the absolute difference between the ground-truth and predicted images.

C. Behavioral Reconstruction

DISE video provides an effective method for identifying behavioral traits associated with obstructed breathing; however, the primary disadvantage of using monocular video is the inability to make objective measurements that indicate the severity of the restricted airflow. This is typically where a Cine-MRI scan can be performed to evaluate airway restrictions accurately and quantitatively. In our approach, we leverage endoscopic cameras that accurately capture complex behavioral traits of the respiratory cycle to reconstruct the observed surfaces for measurement, but supplement this through Cine-MRI data. By using the trained GAN network, we estimate depth images for every frame collected from the monocular video calibrated using Cine-MRI data to perform quantitative measurements of restriction and airflow reduction. For the synthetic training data, we also impose surface deformations on the reference model, replicating airway restrictions.

IV. Experimental Results

Evaluation of the proposed method is based on two classes of results: (1) synthetic ray-traced images and (2) real-world DISE video data. This demonstrates the use of the trained

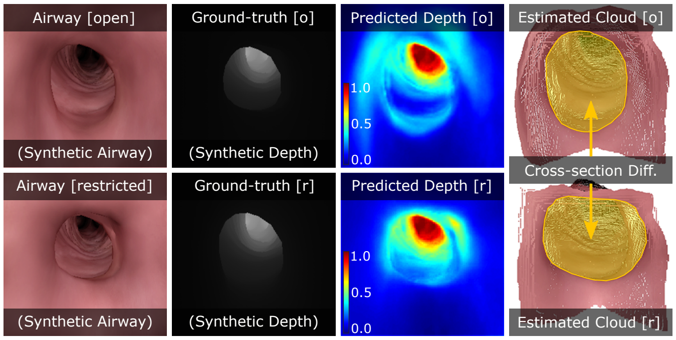

Fig. 7. Illustration of the GAN predicted depth images converted into the point-cloud data that approximates the DISE observed airway surfaces.

GAN model for predicting dynamic behaviors that are observable within the airway given Cine MRI-based deformation data (relative airway restriction) and real-world video data. Quantitatively we evaluate the absolute difference between images, but also evaluate surface predictions based on Mean Squared Error (MSE) and Structural Similarity Index Measure (SSIM) seen in Figure 6. The MSE scores are: (124.14, 745.98, 409.48) with SSIM scores: (0.89, 0.82, 0.87) Extending these results, we illustrate the use of the depth images for generating dynamic surface estimations in Figure 7. For the evaluation dataset, we obtained avg-MSE: 510.10 and avg-SSIM: 0.821. This illustrates the GAN model is effective at estimating synthesized data; however, real-world DISE video presents challenges due to the lower resolution, complex features, and occlusions. Select GAN depth estimates for the real low-resolution DISE images (R0-R2) are shown in Figure 8.

V. Discussion

Various factors contribute to the accuracy and quality of predicted depth images that define airway structural geometry. This includes the visual quality of the synthesized GAN training data, image quality of the DISE sequences, obtaining sufficient Cine-MRI data, and the integration of these data sources into a dynamic 4D model. Our objective is to enable Cine-MRI level diagnosis quality through a monocular stream of images from DISE, without the introduction of new imaging devices. Limitations of our approach with single monocular images include a wide variety of interference from close imaging proximity, internal fluids, and visual occlusions. Improving the link between synthetic depth data and endoscopic images is an ongoing process and represents the fundamental challenge of representing broader segments of patient groups. Improving the complex behavior of internal fluids and subsurface scattering required to render layered tissue is part of the ongoing research needed to improve GAN prediction quality.

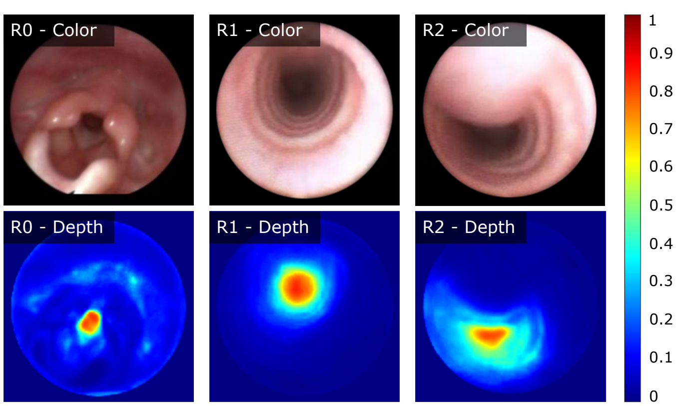

Fig. 8. Predictions on real-world DISE video color (top) and depth estimations (bottom). Each recording captures challenging dynamic surface behaviors of airway deformations/movement as well as complex camera movements.

VI. Conclusion

This work presents a novel contribution in that the anatomical reconstruction process is extended to include behavioral analysis of airways observed using DISE. Our approach utilizes the prediction of depth images through a trained GAN which provides surface estimates of airway wall deformations, including airway wall movement and opening evaluation measurements that can enable obstruction severity. Cine-MRI scans can be used in conjunction with prediction models to establish physical measurements. This provides the first step towards dynamic modeling which can lead to further developments in objective measurement solutions for DISE.

References

[1] Adrian Ong et al. Application of drug-induced sleep endoscopy in patients treated with upper airway stimulation therapy. *World J Otorhinolaryngol Head Neck Surg*, pages 92–96, Jun 2017.

[2] D. Burschka, M. Li, M. Ishii, R. H. Taylor, and G. D. Hager. Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. *Medical Image Analysis*, 9(5):413–426, 2005.

[3] R. J. Chen, T. L. Bobrow, T. Athey, F. Mahmood, and N. J. Durr. Slam endoscopy enhanced by adversarial depth prediction. 2019.

[4] X. Cheng et al. Intubot: Design and prototyping of a robotic intubation device. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1482–1487, 2018.

[5] E. Dijemeni, G. D'Amone, and I. Gbati. Drug-induced sedation endoscopy (dise) classification systems: a systematic review and meta-analysis. *Sleep breathing*, 21(4):983–994, 2017.

[6] L. F. Donnelly. Obstructive sleep apnea in pediatric patients: Evaluation with cine mr sleep studies. *Radiology*, 236(3):768–778, 2005.

[7] W. Faizal, N. Ghazali, C. Khor, I. A. Badruddin, M. Zainon, A. A. Yazid, N. B. Ibrahim, and R. M. Razi. Computational fluid dynamics modelling of human upper airway: A review. *Computer Methods and Programs in Biomedicine*, 196:105627, 2020.

[8] B. L. Fricke, M. B. Abbott, L. F. Donnelly, B. J. Dardzinski, S. A. Poe, M. Kalra, R. S. Amin, and R. T. Cotton. Upper airway volume segmentation analysis using cine mri findings in children with tracheostomy tubes. *Korean journal of radiology*, 8(6):506–511, 2007.

[9] A. Isaiah et al. Characterization of upper airway obstruction using cine mri in children with residual obstructive sleep apnea after adenotonsillectomy. *Sleep medicine*, 50:79–86, 2018.

[10] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[11] E. J. Kezirian, W. Hohenhorst, and N. de Vries. Drug-induced sleep endoscopy: the vote classification. *European Archives of Oto-Rhino-Laryngology*, 268(8):1434–4726, 2011.

[12] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014. Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

[13] A. C. Kumar et al. Monocular depth prediction using generative adversarial networks. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[14] A. Rau et al. Implicit domain adaptation with conditional generative adversarial networks for depth prediction in endoscopy. *Intl. Journal for Computer Assisted Radiology and Surgery*, 14(7):1167–1176, 2019.

[15] C. Xu et al. Inhalation airflow and ventilation efficiency in subject-specific human upper airways. *Respiratory Physiology Neurobiology*, 285:103587, 2021.