

BREA-Depth: Bronchoscopy Realistic Airway-geometric Depth Estimation

Francis Xiatian Zhang^{1,2}[0000–0003–0228–6359], Emile Mackute²[0009–0004–4535–4619], Mohammadreza Kasaei^{1,2}[0000–0003–4932–1457], Kevin Dhaliwal¹[0000–0002–3925–3174], Robert Thomson³[0000–0003–4978–1488], and Mohsen Khadem^{1,2}[0000–0002–6873–273X]

¹ Baillie Gifford Pandemic Science Hub, Institute of Regeneration and Repair, University of Edinburgh, Edinburgh, United Kingdom

² School of Informatics, University of Edinburgh, Edinburgh, United Kingdom
{francis.zhang, s1657385, m.kasaei, kev.dhaliwal, mohsen.khadem}@ed.ac.uk

³ Institute of Photonics and Quantum Science, Heriot-Watt University, Edinburgh, United Kingdom
r.r.thomson@hw.ac.uk

Abstract. Monocular depth estimation in bronchoscopy can significantly improve real-time navigation accuracy and enhance the safety of interventions in complex, branching airways. Recent advances in depth foundation models have shown promise for endoscopic scenarios, yet these models often lack anatomical awareness in bronchoscopy, overfitting to local textures rather than capturing the global airway structure—particularly under ambiguous depth cues and poor lighting. To address this, we propose Brea-Depth, a novel framework that integrates airway-specific geometric priors into foundation model adaptation for bronchoscopic depth estimation. Our method introduces a depth-aware CycleGAN, refining the translation between real bronchoscopic images and airway geometries from anatomical data, effectively bridging the domain gap. In addition, we introduce an airway structure awareness loss to enforce depth consistency within the airway lumen while preserving smooth transitions and structural integrity. By incorporating anatomical priors, Brea-Depth enhances model generalization and yields more robust, accurate 3D airway reconstructions. To assess anatomical realism, we introduce Airway Depth Structure Evaluation, a new metric for structural consistency. We validate BREA-Depth on a collected ex-vivo human lung dataset and an open bronchoscopic dataset, where it outperforms existing methods in anatomical depth preservation.

Keywords: Bronchoscopy · Depth Estimation · Foundation Model

1 Introduction

Accurate bronchoscopic depth estimation is essential for navigation and 3D airway reconstruction, but obtaining reliable ground truth remains challenging.

Many existing methods [24, 14, 2, 28, 15, 10] generate depth supervision from CT-derived synthetic data and often employ CycleGAN-based domain translation [14, 2, 10] to adapt to bronchoscopic images. Compared to other GANs and recent diffusion methods [4], CycleGAN remains popular for learning from unpaired data without large computational or dataset demands [18]. Yet CT-derived synthetic depth often yields oversimplified maps lacking realistic variation and generalizability, limiting performance under real-world bronchoscopic conditions.

Recent depth foundation models, such as Depth Anything [26], offer improved generalization by learning robust depth representations from large-scale, unlabeled image sets. Inspired by their success, recent works [6, 23, 5] have adapted foundation models for endoscopic applications. However, these methods still prioritize pixel-level accuracy over anatomical awareness, limiting their reliability for bronchoscopic tasks that require structurally precise depth, such as targeted biopsy or localized drug delivery.

A key challenge in adapting foundation models to bronchoscopy is their *lack of anatomical awareness*. Unlike general endoscopic or open-world scenes [22], bronchoscopy involves navigating structured yet deformable airway branches. Models relying solely on photometric cues often fail to capture the global airway geometry, particularly under ambiguous depth conditions [16]. Furthermore, large-scale models such as Depth Anything [26] have limited exposure to airway branching topologies, leading to overfitting on local textures rather than learning structural priors. Additionally, standard evaluation metrics (e.g., Abs Rel [8]) focus on pixel-level accuracy but fail to assess global anatomical consistency—an essential feature for navigation and 3D airway reconstruction.

To overcome existing limitations, we propose Bronchoscopy REalistic Airway-geometric Depth Estimation (BREA-Depth), a framework integrating airway-specific structural priors into depth estimation. Our approach employs a *Depth-aware CycleGAN* to refine synthetic-to-real translation, bridging the gap between simulated training data and real bronchoscopic footage. Additionally, we introduce an *Airway Structure Awareness Loss* to enforce depth consistency within the airway lumen while preserving smooth transitions and geometric integrity. To evaluate anatomical consistency, we propose an *Airway Depth Structure Evaluation* metric, assessing depth distribution relative to the airway lumen. For validation, We collected bronchoscopic video data from ex-vivo human lung models and annotated 3,437 bronchoscopic images with semantic segmentation across five navigation sequences.

We extensively evaluate BREA-Depth against state-of-the-art foundation models [26, 27] and recent bronchoscopy-adapted approaches [6, 23]. Our experiments on both our collected dataset and the open dataset [24] demonstrate that BREA-Depth significantly outperforms existing techniques, particularly in anatomically complex regions with limited depth cues. Source code and ex-vivo lung data are available at <https://github.com/SIRGLab/BREA-Depth>.

Our main contributions are summarized as follows:

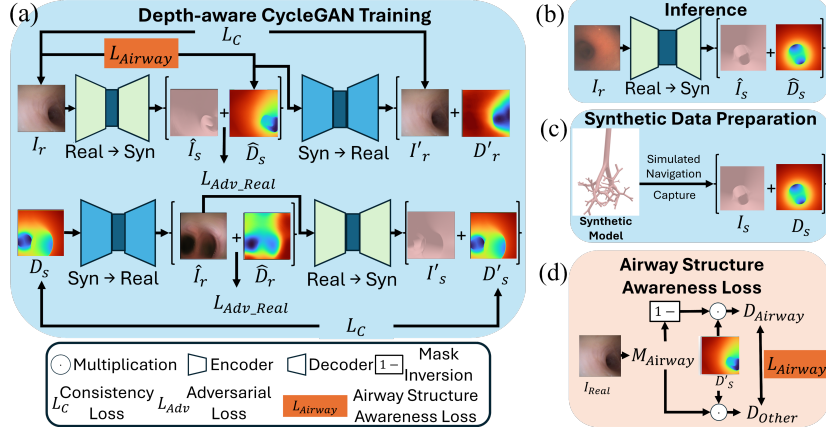


Fig. 1. (a) Overview of our bronchoscopic depth estimation framework using a cycle-consistent adversarial approach with unpaired synthetic and real bronchoscopic images through two branches: (i) *Syn-to-Real* translates simulated depth maps and frames from a synthetic airway model into a realistic style, and (ii) *Real-to-Syn* reduces texture artifacts in real images. A foundation model provides pseudo-depth supervision, while a discriminator enforces realism. (b) During inference, a bronchoscopic frame I_r generates a depth map \hat{D} and a synthetic frame \hat{I} that reduces texture noise while preserving geometric details. (c) Geometrically accurate bronchial tree generated in Blender and used to collect synthetic data. (d) Our proposed Airway Structure Awareness loss enforcing depth consistency and smooth transitions within the airway lumen.

1. BREA-Depth, a novel framework that integrates airway geometry into foundation depth estimation, ensuring anatomically consistent depth predictions while preserving the complex bronchial structure.
2. We introduce *Airway Depth Structure Evaluation*, a new metric assessing anatomical consistency in depth predictions relative to the airway lumen, complementing existing evaluation methods.
3. We open-source a human lung bronchoscopy dataset to support future research on depth estimation in bronchoscopy, facilitating further advancements in airway navigation and intervention.

2 Methodology

The BREA-Depth model (Fig. 1) processes a bronchoscopic frame $X \in \mathbb{R}^{H \times W \times 3}$ to generate a depth map $\hat{D} \in \mathbb{R}^{H \times W \times 1}$ and a synthetic-style frame $\hat{Y} \in \mathbb{R}^{H \times W \times 3}$, which reduces texture noise while preserving airway geometry. Training follows a CycleGAN approach [30], leveraging unpaired synthetic images and real bronchoscopic images through two branches: (i) *Syn-to-Real*, which refines synthetic depth maps and frames into realistic representations, and (ii) *Real-to-Syn*, which simplifies real frames while preserving geometric features. Pseudo ground truth

from a foundation model supervises real-depth predictions, while a PatchGAN discriminator [13] enforces alignment between generated depth maps and frames.

The following sections detail our simulation setup for generating synthetic airway geometry, the Depth-Aware CycleGAN for depth estimation, and a loss function that enforces depth consistency within the airway lumen.

2.1 Geometrically Accurate Model of Airways

We developed a physiologically realistic bronchoscopy simulation by generating 3D lung structures based on key geometric properties: (i) airway branching, (ii) bifurcation structure, and (iii) carinal rounding. These parameters are derived from anatomical airway models [20, 9] and airflow-based geometric models [11, 12] to ensure accurate lung morphology representation.

The bronchial tree is modeled as a series of cylindrical airways undergoing binary branching, where each airway has a length L and diameter D , with the length-to-diameter ratio L/D specified per generation. Given a parent airway of diameter D_p , daughter branches have diameters D_a and D_b , defined by $h_a = D_a/D_p$ and $h_b = D_b/D_p$, where $h_a, h_b \in (0, 1)$ follow anatomical scaling laws.

The bifurcation region smoothly transitions between the parent and daughter branches, governed by branching angles Φ_a, Φ_b , curvature radii R_a^*, R_b^* , and ring radii $R_a(\phi_{sa}), R_b(\phi_{sb})$. The branching angles are sampled within a physiologically valid range $\Phi_a, \Phi_b \in [\Phi_{\min}, 120^\circ]$, where Φ_{\min} prevents airway intersection. The curvature radii are computed as:

$$R_a^* = \frac{D_a}{2 \sin \Phi_a}, \quad R_b^* = \frac{D_b}{2 \sin \Phi_b} \quad (1)$$

A sigmoid-based transition function ensures gradual bifurcation tapering: $R_a(\phi_{sa}) = D_a \cdot f(\phi_{sa})$, $R_b(\phi_{sb}) = D_b \cdot f(\phi_{sb})$.

Carinal rounding smooths bifurcation transitions using continuous rounding circles centered at K , with radius $R_c = \min_{(x,y) \in \mathbb{R}^2} \sqrt{(x - K_x)^2 + (y - K_y)^2}$, where K_x, K_y define the circle centers, and radii are computed for sagittal angles $\phi_{sa} \in [0, R_a]$ and $\phi_{sb} \in [0, R_b]$. Unlike previous models assuming parallel circles [11], our method dynamically tilts them to match airway asymmetry.

To enhance realism, additional constraints are applied. Twist angles Θ are sampled uniformly as $\Theta \sim \mathcal{U}(0^\circ, 360^\circ)$, while airway lengths follow a Gaussian distribution $L \sim \mathcal{N}(L_{\text{mean}}, 0.3L_{\text{mean}})$, where L_{mean} is the expected length per generation. These constraints ensure anatomically plausible airway structures while introducing natural variability. The model is implemented in Blender [3] for realistic bronchoscopy simulations, as shown in Fig. 1c.

2.2 Depth-Aware CycleGAN

By using simulated RGB–depth pairs generated from our airway model, we introduce a *Depth-aware CycleGAN* that refines synthetic-to-real translation by enforcing airway-specific depth constraints. It addresses the domain gap caused

by the lack of texture and structural variation in synthetic images, even though they capture accurate lumen geometry. Unlike conventional style transfer methods that rely on feature distribution similarity [14, 28]—often failing to preserve structural geometry—our model directly incorporates depth to guide translation, resulting in more anatomically faithful adaptations. During training, synthetic images provide paired RGB–depth supervision, while real bronchoscopic images are paired with pseudo-depth maps from a foundation model [26], enabling hybrid supervision and improving depth estimation under real-world conditions.

Our framework comprises two U-Net Transformer-like [19] encoder-decoder branches: *Syn-to-Real*, which refines depth maps to enhance texture and structure, and *Real-to-Syn*, which converts bronchoscopic images into synthetic-style frames and generates depth maps. The synthetic-style output provides auxiliary supervision that guides the encoder to preserve anatomical structure and suppress texture noise. The translation process follows:

Synthetic-to-Real Translation: Given a synthetic image $X_s = \{I_s, D_s\}$, where $I_s \in \mathbb{R}^{H \times W \times 3}$ represents the synthetic RGB image and $D_s \in \mathbb{R}^{H \times W \times 1}$ is the perfect depth map, the encoder maps it to a latent representation Z_s and the decoder then generates the translated real-style RGB image \hat{I}_r and its corresponding real-style depth \hat{D}_r :

$$Z_s = E_s(D_s), \quad \hat{Y}_s = D_r(Z_s), \quad \hat{Y}_s = \{\hat{I}_r, \hat{D}_r\} \quad (2)$$

Real-to-Synthetic Translation: For a real bronchoscopic image $X_r = \{I_r\}$, the encoder extracts its feature representation Z_r , and the decoder then predicts the corresponding synthetic perfect depth \hat{D}_s while reconstructing the synthetic-style RGB image \hat{I}_s :

$$Z_r = E_r(X_r), \quad \hat{Y}_r = D_s(Z_r), \quad \hat{Y}_r = \{\hat{I}_s, \hat{D}_s\} \quad (3)$$

Similarly, in another cycle, based on the generated $\{\hat{I}_r, \hat{D}_r\}$ and $\{\hat{I}_s, \hat{D}_s\}$, respectively, our framework further generates the synthetic $\{I'_s, D'_s\}$ and real domain $\{I'_r, D'_r\}$, which are then used to compute the consistency loss.

2.3 Airway Structure Awareness Loss

Unlike general endoscopic or natural scenes, where depth cues rely on texture variations [17, 29], bronchoscopic images have a monotonous lumen texture with low saliency [14], making pixel-wise depth learning prone to errors. A key anatomical prior in bronchoscopy is that airway depth should increase as the lumen extends further into the respiratory system [1], yet standard depth estimation models fail to enforce this, leading to inconsistent predictions. To address this, we introduce the *Airway Structure Loss*, which encourages lower disparity values in the airway lumen by leveraging intensity-based segmentation.

The airway region is defined using a grayscale threshold T on the grayscale-translated image I_{gray} , forming a binary mask M_{airway} that identifies the airway

lumen $M_{\text{airway}} = (I_{\text{gray}} < T)$. The mean disparity in airway and non-airway regions is computed as:

$$D_{\text{airway}} = \frac{\sum(\hat{D} \cdot M_{\text{airway}})}{\sum M_{\text{airway}} + \epsilon}, \quad D_{\text{non-airway}} = \frac{\sum(\hat{D} \cdot (1 - M_{\text{airway}}))}{\sum(1 - M_{\text{airway}}) + \epsilon}. \quad (4)$$

To enforce anatomical consistency, we use a ReLU-based formulation to penalize cases where the airway disparity exceeds the non-airway disparity:

$$\mathcal{L}_{\text{airway}} = \mathbb{E}[\max(0, D_{\text{airway}} - D_{\text{non-airway}})]. \quad (5)$$

The total optimization objective is:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{adv}}\mathcal{L}_{\text{adv}} + \lambda_{\text{cycle}}\mathcal{L}_{\text{cycle}} + \lambda_{\text{identity}}\mathcal{L}_{\text{identity}} + \lambda_{\text{airway}}\mathcal{L}_{\text{airway}}, \quad (6)$$

where \mathcal{L}_{adv} is the adversarial loss for realistic image translation [13], $\mathcal{L}_{\text{cycle}}$ ensures bidirectional cycle consistency, $\mathcal{L}_{\text{identity}}$ preserves identity mappings, and $\mathcal{L}_{\text{airway}}$ enforces depth consistency within the airway lumen. $\mathcal{L}_{\text{cycle}}$ and $\mathcal{L}_{\text{identity}}$ are computed using L1 loss.

3 Experiment

Our model is trained on an NVIDIA RTX 3080 using PyTorch, and utilizes 9,500 synthetic image-depth pairs (Section 2.1) and 55,000 real bronchoscopic images (Fig. 2a) with pseudo-depth from DepthAnything [26]. Training runs for 30 epochs with a batch size of 2, chosen to reduce GPU memory consumption during training, and a learning rate of 0.0001. The weighting factors are set as follows: $\lambda_{\text{adv}} = 5$, $\lambda_{\text{cycle}} = 1$, $\lambda_{\text{identity}} = 1$, and $\lambda_{\text{airway}} = 0.5$. During inference, our model runs at 60 FPS, achieving online real-time performance.

We compare our framework with existing depth estimation methods, including bronchoscopy-specific approaches and general foundation models: 1) 3cGAN [2]: A CycleGAN-based method for manually rendering synthetic images. 2) DepthAnything [26]: A foundation model trained on large-scale unlabeled data for monocular depth estimation. 3) DepthAnythingV2 [27]: An improved version of DepthAnything. 4) EndoDAC [6]: An adaptation of DepthAnything for endoscopy. 5) EndoOmni [23]: An adaptation of DepthAnything leveraging annotated data, with an extended dataset for bronchoscopy. For foundation models, we use their largest versions, but for EndoOmni, we evaluate both the base (*B*) and large (*L*) versions for a comprehensive comparison.

We evaluate depth prediction using two evaluation strategies: **(i) Airway Depth Structure Evaluation:** Standard depth metrics prioritize pixel accuracy [8] but often neglect anatomical consistency in bronchoscopic navigation [2]. We propose a depth structure evaluation assuming the airway lumen is the deepest region in bronchoscopic imagery [1], assessing: 1) the localization of the lowest depth regions and 2) the depth contrast between lumen and non-lumen regions. *Lowest Depth Localization Accuracy (LocalAccu):* A well-trained model should predict the lowest depth within the lumen, not along airway walls. Given a

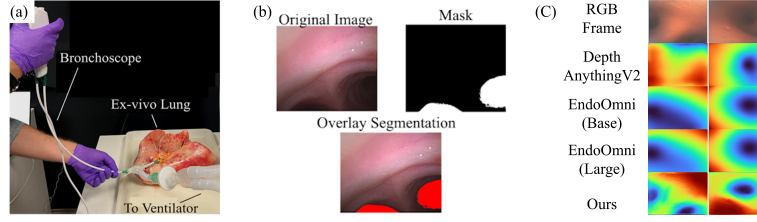


Fig. 2. (a) The experimental setup for data collection with ex-vivo human lung. A commercial bronchoscope (Ambu aScope 4 ,Ambu Ltd.) is used by an expert bronchoscopist to navigate the lung. The collection fits the appropriate institutional ethical process. (b) Airway lumen segmentation from our dataset, providing essential ground truth for *Airway Depth Structure Evaluation*. (c) Qualitative comparison of our model’s depth estimation results (disparity). The predicted depth map clearly highlights the airway lumen and captures fine anatomical details around airway branches.

predicted depth map D and airway mask M_{lumen} , the minimum depth value is $D_{\min} = \min(D)$. To assess alignment with the airway lumen, we compute the proportion of pixels where $D = D_{\min}$ within M_{lumen} :

$$R_{\text{in-lumen}} = \frac{\sum \mathbb{I}(D = D_{\min}) \odot M_{\text{lumen}}}{\sum \mathbb{I}(D = D_{\min}) + \epsilon}. \quad (7)$$

Here, M_{lumen} is a binary mask indicating the airway lumen, $\mathbb{I}(\cdot)$ is an indicator function, and ϵ prevents division by zero. If $R_{\text{in-lumen}} > 0.99$, the model is considered to have correctly localized the lowest depth inside the lumen.

Depth Contrast Consistency (DepthCon): We also assess whether the predicted depth distribution reflects airway structure by computing the mean depth inside and outside the lumen as $\bar{D}_{\text{lumen}} = \sum D \odot M_{\text{lumen}} / (\sum M_{\text{lumen}} + \epsilon)$ and $\bar{D}_{\text{outside}} = \sum D \odot (1 - M_{\text{lumen}}) / (\sum (1 - M_{\text{lumen}}) + \epsilon)$, respectively. To quantify the contrast, we compute the z-score:

$$Z_{\text{lumen-outside}} = \frac{\bar{D}_{\text{lumen}} - \bar{D}_{\text{outside}}}{\sigma_{\text{outside}} + \epsilon}, \quad (8)$$

where σ_{outside} is the depth standard deviation outside the lumen. A sufficiently negative $Z_{\text{lumen-outside}}$ (i.e., < -1.00) confirms the lumen is significantly deeper.

To address the lack of a dedicated dataset for this evaluation, we create and will open source a segmentation dataset based on five navigation sequences in an ex-vivo human lung (Fig. 2a), comprising over 20 minutes of data recorded at 2.5 fps, totaling 3,437 frames. Each frame is manually prompted and automatically segmented using SegmentAnythingV2 [21] to identify the airway lumen (Fig. 2b).

(ii) Classic Depth Performance Metric: We benchmark our framework on the Visentini-Scarzanella et al. [24] dataset, comprising 16 videos (39,599 frames) of a bronchial phantom with ground-truth depth and CT renderings. Following prior work [6, 23], we evaluate performance using median alignment post-processing between predicted depth and ground truth, followed by standard depth metrics: Abs Rel, Sq Rel, RMSE, RMSE log, and threshold accuracy (δ).

Table 1. Comparison of Airway Depth Structure Evaluation on our collected dataset and Classical Performance on the phantom dataset [24]. DA: Depth Anything. w/o CycleGAN: Supervised directly using real images with pseudo-depth and synthetic images with depth. w/o $\mathcal{L}_{\text{airway}}$: Excludes $\mathcal{L}_{\text{airway}}$ from the optimization process.

Methods	Structure Evaluation		Classical Performance				
	DepthCon \uparrow	LocalAccu \uparrow	<i>AbsRel</i> \downarrow	<i>SqRel</i> \downarrow	<i>RMSE</i> \downarrow	<i>RMSE</i> _{log} \downarrow	δ \uparrow
3cGAN [2]	99.27	57.00	0.33	7.79	15.67	0.35	57.84
DA [26]	70.55	45.64	0.24	4.21	12.98	0.27	64.25
DA-V2 [27]	60.88	51.70	<u>0.21</u>	3.72	12.02	<u>0.25</u>	68.52
EndoDAC [6]	34.58	14.18	0.27	5.56	14.13	0.30	63.31
EndoOmni _B [23]	96.53	<u>58.19</u>	0.18	<u>2.99</u>	<u>10.66</u>	0.21	<u>75.56</u>
EndoOmni _L [23]	96.09	45.79	0.18	2.77	10.27	0.21	76.54
Ours	<u>97.27</u>	62.36	0.23	4.56	12.26	<u>0.25</u>	70.64
w/o CycleGAN	68.36	25.36	0.33	7.76	16.16	0.35	56.89
w/o $\mathcal{L}_{\text{airway}}$	96.67	52.06	0.29	6.12	14.60	0.31	59.23

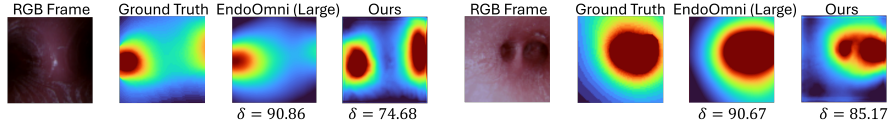


Fig. 3. Qualitative comparison of our model’s depth estimation ($1/\text{disparity}$) on phantom data [24]. Our model preserves structural consistency, accurately capturing airway depth, especially at bifurcations. However, pixel-level metrics overlook ground truth inaccuracies (e.g., misassigned depth in extended lumen regions), limiting its apparent performance (e.g., δ) despite improvements.

Airway Depth Structure Evaluation Table 1 shows that our method achieves strong performance in our proposed structure evaluation, with 97.27% *DepthCon* and 62.36% *LocalAccu*, outperforming most existing methods. Although 3cGAN [2] achieves slightly higher *DepthCon*, our method provides more balanced performance, enhancing anatomical consistency and benefiting airway reconstruction. Notably, EndoDAC [6] performs poorly, underscoring the need for airway-specific priors in foundation models. Fig. 2c illustrates our model’s performance, capturing anatomical details around the lumen and airway branches.

Classical Depth Estimation Performance Table 1 compares our method with existing approaches, showing comparable performance. The limited improvement is largely due to low-quality ground truth from phantom data, which fails to capture real-world bronchoscopic complexity. As shown in Fig. 3, the ground truth often misassigns depth in extended lumen regions or overlooks bifurcations, leading to inaccurate evaluations. Despite this, our method generalizes well, performing comparably to other foundation models.

Ablation Analysis Table 1 assesses component contributions. Removing CycleGAN (*w/o CycleGAN*) degrades performance, showing its role in domain adaptation. Excluding airway structure awareness loss ($\mathcal{L}_{\text{airway}}$) weakens structural preservation, emphasizing its importance for anatomical consistency. Both components are crucial for achieving anatomically consistent depth predictions.

4 Conclusion

We propose BREA-Depth, a bronchoscopic depth estimation framework that integrates airway-specific geometric priors into foundation model adaptation. By incorporating a *Depth-aware CycleGAN* and an *Airway Structure Awareness Loss*, our approach enhances depth consistency and anatomical realism, outperforming existing methods in structural preservation, as validated by our *Airway Depth Structure Evaluation*. Our results highlight the limitations of current evaluation metrics, which emphasize pixel accuracy over anatomical consistency, and the challenges posed by low-quality ground truth in existing datasets. To address this, we introduce new evaluation metrics tailored to bronchoscopic depth estimation. Future work includes establishing a bronchoscopic depth benchmark with our synthetic data and integrating camera pose [7] and landmark recognition [25] to refine our metrics by reducing potential bias across viewing angles.

Acknowledgments. This work was supported by the UKRI Medical Research Council (MR/T023252/1), Baillie Gifford Pandemic Science Hub, and the Engineering and Physical Sciences Research Council (EPSRC) ‘U-care’ Programme Grant (EP/T020903/1).

Disclosure of Interests. The authors declare no conflicts of interest in this paper.

References

1. Amador, C., Weber, C., Varacallo, M.A.: Anatomy, thorax, bronchial. In: StatPearls [Internet]. StatPearls Publishing (2023)
2. Banach, A., King, F., Masaki, F., Tsukada, H., Hata, N.: Visually navigated bronchoscopy using three cycle-consistent generative adversarial network for depth estimation. *Medical image analysis* **73**, 102164 (2021)
3. Blender Online Community: Blender - A 3D Modelling and Rendering Package. Stichting Blender Foundation, Amsterdam (2024), <https://www.blender.org>, version 4.3
4. Chang, Z., Koulouris, G.A., Chang, H.J., Shum, H.P.: On the design fundamentals of diffusion models: A survey. *Pattern Recognition* p. 111934 (2025)
5. Cui, B., Islam, M., Bai, L., Ren, H.: Surgical-dino: adapter learning of foundation models for depth estimation in endoscopic surgery. *International Journal of Computer Assisted Radiology and Surgery* pp. 1–8 (2024)
6. Cui, B., Islam, M., Bai, L., Wang, A., Ren, H.: Endodac: Efficient adapting foundation model for self-supervised depth estimation from any endoscopic camera. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 208–218. Springer (2024)
7. Deng, J., Li, P., Dhaliwal, K., Lu, C.X., Khadem, M.: Feature-based visual odometry for bronchoscopy: A dataset and benchmark. In: *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 6557–6564. IEEE (2023)
8. Dong, X., Garratt, M.A., Anavatti, S.G., Abbass, H.A.: Towards real-time monocular depth estimation for robotics: A survey. *IEEE Transactions on Intelligent Transportation Systems* **23**(10), 16940–16961 (2022)

9. Florens, M., Sapoval, B., Filoche, M.: An anatomical and functional model of the human tracheobronchial tree. *Journal of Applied Physiology* **110**(3), 756–763 (Mar 2011). <https://doi.org/10.1152/japplphysiol.00984.2010>
10. Guo, L., Nahm, W.: A cgan-based network for depth estimation from bronchoscopic images. *International Journal of Computer Assisted Radiology and Surgery* **19**(1), 33–36 (2024)
11. Hegedus, C., Balashazy, I., Farkas, A.: Detailed mathematical description of the geometry of airway bifurcations. *Respiratory Physiology & Neurobiology* **141**(1), 99–114 (Jul 2004). <https://doi.org/10.1016/j.resp.2004.03.004>
12. Heistracher, T., Hofmann, W.: Physiologically realistic models of bronchial airway bifurcations. *Journal of Aerosol Science* **26**(3), 497–509 (Apr 1995). [https://doi.org/10.1016/0021-8502\(94\)00113-D](https://doi.org/10.1016/0021-8502(94)00113-D)
13. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1125–1134 (2017)
14. Karaoglu, M.A., Brasch, N., Stollenga, M., Wein, W., Navab, N., Tombari, F., Ladikos, A.: Adversarial domain feature adaptation for bronchoscopic depth estimation. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*. pp. 300–310. Springer (2021)
15. Liu, H., Zhang, S., Yang, Y., Sun, L., Song, Z., Xu, S.: A robust pose optimization scheme with spatial geometry awareness for hybrid bronchoscopic navigation. *IEEE Transactions on Instrumentation and Measurement* **73**, 1–10 (2024)
16. Liu, X., Sinha, A., Ishii, M., Hager, G.D., Reiter, A., Taylor, R.H., Unberath, M.: Dense depth estimation in monocular endoscopy with self-supervised learning methods. *IEEE transactions on medical imaging* **39**(5), 1438–1447 (2019)
17. Ming, Y., Meng, X., Fan, C., Yu, H.: Deep learning for monocular depth estimation: A review. *Neurocomputing* **438**, 14–33 (2021)
18. Nayak, A.A., Venugopala, P., Ashwini, B.: A systematic review on generative adversarial network (gan): Challenges and future directions. *Archives of Computational Methods in Engineering* pp. 1–34 (2024)
19. Petit, O., Thome, N., Rambour, C., Themyr, L., Collins, T., Soler, L.: U-net transformer: Self and cross attention for medical image segmentation. In: *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. pp. 267–276. Springer (2021)
20. Raabe, O.G., Yeah, H.C., Schum, M.G., Phalen, R.F.: Tracheobronchial geometry: Human, dog, rat, hamster - a compilation of selected data from the project respiratory tract deposition models. report, University of North Texas Libraries, UNT Digital Library (Mar 1976)
21. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024)
22. Ryan, B., Yendamuri, K., Yendamuri, S.: Anatomical considerations in bronchoscopy. *Journal of Thoracic Disease* **9**(Suppl 10), S1123 (2017)
23. Tian, Q., Chen, Z., Liao, H., Huang, X., Li, L., Ourselin, S., Liu, H.: Endoomni: Zero-shot cross-dataset depth estimation in endoscopy by robust self-learning from noisy labels. *arXiv preprint arXiv:2409.05442* (2024)
24. Visentini-Scarzanella, M., Sugiura, T., Kaneko, T., Koto, S.: Deep monocular 3d reconstruction for assisted navigation in bronchoscopy. *International journal of computer assisted radiology and surgery* **12**, 1089–1099 (2017)

25. Vu, V.G., Hoang, A.D., Phan, T.P., Nguyen, N.D., Nguyen, T.T., Nguyen, D.N., Dao, N.P., Doan, T.P.L., Nguyen, T.T.H., Trinh, T.H., et al.: Bm-broncho-a rich bronchoscopy dataset for anatomical landmarks and lung cancer lesion recognition. *Scientific Data* **11**(1), 321 (2024)
26. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth anything: Unleashing the power of large-scale unlabeled data. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 10371–10381 (2024)
27. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. *arXiv preprint arXiv:2406.09414* (2024)
28. Yang, Y., Ning, G., Zhong, C., Liao, H.: Adversarial diffusion model for domain-adaptive depth estimation in bronchoscopic navigation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 46–56. Springer (2024)
29. Yang, Z., Dai, J., Pan, J.: 3d reconstruction from endoscopy images: A survey. *Computers in Biology and Medicine* p. 108546 (2024)
30. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2223–2232 (2017)