Airway measurement by refinement of synthetic images improves mortality prediction in idiopathic pulmonary fibrosis

Ashkan Pakzad[1], Mou-Cheng Xu[1], Wing Keung Cheung[1], Marie Vermant[2,3], Tinne Goos[2,3], Laurens J De Sadeleer[2,3], Stijn E Verleden[2,4], Wim A Wuyts[2,3], John R Hurst[5], and Joseph Jacob[1,5]

[1] Centre for Medical Image Computing, University College London, UK
[2] BREATHE, Department of Chronic Diseases and Metabolism, KU Leuven, Leuven, Belgium
[3] Department of Respiratory Diseases, Unit for interstitial lung diseases, University Hospitals Leuven, Leuven, Belgium
[4] Antwerp Surgical Training, Anatomy and Research Centre (ASTARC), Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium.
[5] UCL Respiratory, University College London, UK
a.pakzad@cs.ucl.ac.uk
https://ashkanpakzad.github.io

Abstract. Several chronic lung diseases, like idiopathic pulmonary fibrosis (IPF) are characterised by abnormal dilatation of the airways. Quantification of airway features on computed tomography (CT) can help characterise disease progression. Physics based airway measurement algorithms have been developed, but have met with limited success in part due to the sheer diversity of airway morphology seen in clinical practice. Supervised learning methods are also not feasible due to the high cost of obtaining precise airway annotations. We propose synthesising airways by style transfer using perceptual losses to train our model, Airway Transfer Network (ATN). We compare our ATN model with a state-of-the-art GAN-based network (simGAN) using a) qualitative assessment; b) assessment of the ability of ATN and simGAN based CT airway metrics to predict mortality in a population of 113 patients with IPF. ATN was shown to be quicker and easier to train than simGAN. ATN-based airway measurements were also found to be consistently stronger predictors of mortality than simGAN-derived airway metrics on IPF CTs. Airway synthesis by a transformation network that refines synthetic data using perceptual losses is a realistic alternative to GAN-based methods for clinical CT analyses of idiopathic pulmonary fibrosis. Our source code can be found at https://github.com/ashkanpakzad/ATN that is compatible with the existing open-source airway analysis framework, AirQuant.

Keywords: Generative model evaluation · Style transfer · Computed tomography · Airway measurement · Bronchiectasis · Idiopathic pulmonary fibrosis

1   Introduction

Chronic lung disease is one of the leading causes of morbidity and mortality across the world. As smoking rates in the developing world increase, the prevalence of chronic lung disease is set to rise. Interstitial lung diseases (ILD) are characterised by inflammation and scarring of the lung and the incidence of ILD continues to increase [24].

A subset of ILDs are characterised by lung fibrosis, with idiopathic pulmonary fibrosis (IPF) having the worst prognosis of all the fibrosing ILDs [3]. In IPF the airways are pulled open by fibrotic contraction of the surrounding connective tissue. Computed tomography (CT) imaging is used to visualise airway structure. In IPF the presence of dilated airways in the lung periphery on CT, termed traction bronchiectasis, is a disease hallmark.

When assessing disease severity in IPF, physiologic measurements are typically used. However these are associated with a degree of measurement variability. It has been postulated that combining imaging measures of airway abnormality with lung function measurements could help improve estimation of disease severity in IPF [17]. Importantly, better measures of disease severity would benefit cohort enrichment of subjects into therapeutic trials.

Lung damage in IPF progresses from the distal lung towards the centre of the lung [14]. As a result, the earliest signs of lung damage are seen in the smaller airways. Yet these airways are typically the most challenging to quantify. Airway measurement is complicated by partial volume effects that result in smaller airways having a blurred contour to their walls. Measurement challenges are compounded by variations in CT image acquisition including different reconstruction kernels, scan parameters and scanner models as well as the underlying pathology affecting the lung.

Physics based airway measurement algorithms tend to perform sub optimally when measuring the lumens of small airways [11,2]. Identifying airway walls can also be challenging. Airway paths often run in tandem with those of the pulmonary artery. Consequently, in regions when the pulmonary artery abuts the airway wall, identification of the contour of the outer airway wall is compromised.

1.1   Related work

Deep learning frameworks have been applied to the measurement of airways in the lung in a bid to improve measurement accuracy. However, these machine learning methods are extremely data hungry and can be biased towards the training data sample [9]. Synthetic data by way of generative models has been employed to improve the training of deep learning models. This helps overcome the data limitations that are ubiquitous to medical imaging studies [23].

A state of the art method in measuring airway lumen radius and wall thickness on CT imaging, simGAN [15,20], takes labelled simplistic representations of airway patches (synthetic images) and aims to transforms them in to the emulations of real airways by generative adversarial training (GAN) [5]. These are then used for supervised training of a convolutional neural regressor (CNR)

which learns to measure airway radius and wall thickness and ultimately to run inference on real CT images.

The driving loss for realism in simGAN is cross-entropy loss computed on the classifications of the discriminator. For successful synthetic refinement by image transformation, the synthetic and refined images must have good correspondence in their shared label. To this end, a per-pixel $\|l\|_1$ regularisation loss is applied between input and output of the refiner.

GAN training is inherently unstable with mode collapse complicating and lengthening training times. As an alternative strategy, in this paper we propose the first use of perceptual losses to generate labelled synthetic airway images. Perceptual loss functions have been applied to image style transfer and super-resolution tasks[10]. We explore the clinical benefits of learning from perceptual loss generated synthetic data in mortality prediction.

2   Methods

In the first part of our study we generate synthetic airway patches that demonstrate realistic airway characteristics. In tandem, we segment the airways on clinical CT scans of a cohort of IPF patients. We train our Airway Transfer Network (ATN) to transform our synthetic images to refined images across our synthetic and real datasets by optimising for perceptual losses. We then compare the results of ATN with simGAN. A CNR is trained on the resultant refined datasets for the purpose of inference on real CT airways.

We compare the two refiner models qualitatively. We compare ATN and simGAN against the full width at half maximum edgecued segmentation limited (FWHMesl) technique as implemented in [19], originally by [11]. The FWHMesl technique is widely used in the literature as the reference for comparison of previous airway measurement methods [15,25,6]. In our clinical comparison, we examine which of the three methods of airway measurement provides the best and most consistent association with mortality on CT scans of patients with IPF.

Airway segmentation was performed using a 2D dilated U-Net [26] trained on CT scans in 25 IPF and healthy individuals [16]. We extract orthogonal airway patches for all segmented airways. We parameterise airway labels as two ellipses that share centre and rotation, resulting in 7 parameters for each patch: inner airway wall major and minor axis radii R_A and R_B; outer airway wall major and minor axis radii W_A and W_B; centre coordinates C_x and C_y; and rotation θ. Due to the phase in θ, for the purposes of CNR training the rotation angle is converted into a double angle representation [12].

Once the refiner model has been trained, its output is used to train a CNR by supervised learning to regress to target airway labels. The inner and outer airway wall measures are then derived. All deep learning methods were implemented in pytorch [18] and CT image processing was done using the open source airway analysis framework known as AirQuant [16]. We release our code open source[6].

---

[6] https://github.com/ashkanpakzad/ATN

2.1   Airway Synthesis

Details of airway parameters and synthesis pipeline have been previously described [15]. Airway characteristics are sampled from a set of distribution parameters informed by [22]. We deviate from these parameters in two ways. First, we use an airway lumen radius (LR) interval of [0.3, 6] to permit measurement of smaller airways. Second, we use an airway wall thickness $[0.1 \cdot LR + 0.2, 0.3 \cdot LR + 0.8]$ mm to reflect the lack of airway wall thickening in IPF. We add four further parameters: (i) parameters for the airway centre determined by a normal distribution $X \sim N(0, 1)$ mm to account for airway skeletons that are not perfectly positioned within the centre of the airway lumen. (ii) $p = 0.4$ that an adjacent airway of similar diameter is randomly added. This is performed to accommodate airway patches close to airway bifurcations and to train the CNR to correctly identify the airway in the centre of the patch. (iii) We model our airways as ellipsoids, we achieve this by an ellipsoidness characteristic, sampled from a uniform distribution, $X \sim U(0.9, 1)$ which determines the ratio in major and minor radii of the ellipse. (iv) Uniformly random rotation applied to the airway in the horizontal axis. We include our synthetic airway generator and configuration parameters in the open-source code repository.

2.2   Perceptual Losses

We implement perceptual losses for computing high level perceptual differences between synthetic and real images as described by [10]. These losses are computed by comparing the activations in particular layers, j of a pretrained convolutional neural network (CNN), ϕ between a pair of images. Different activation layers of a trained CNN learn to represent different image features on the same sampled patch. In minimising for perceptual losses we are looking to reduce the differences in the activation of these layers between the refiner output and some objective image. For each calculation of perceptual losses on a synthetic input image, x we have a refiner prediction, $\hat{y}$. As a modification of the original style transfer implementation [10], a randomly chosen real image is selected as the style target, y_s. Perceptual losses are then calculated and summed for different layers ϕ_j.

We utilise feature reconstruction loss. This is defined as the mean euclidean distance between activations of the input and output images of the refiner, where C, H, and W are the number of channels, height and width of layer j respectively. We use a VGG-16 [21] network pretrained on the ImageNet dataset [1] in our calculations of style and feature losses.

$$l_{feat}^{\phi,j}(\hat{y}, x) = \frac{1}{C_j H_j W_j} \|\phi_j(\hat{y}) - \phi_j(x)\|_1 \tag{1}$$

We also employ style reconstruction loss, which considers those features that tend to be activated together between the refiner output and the given style target image, a random real airway, where G_j^ϕ is the gram matrix for a given layer j of ϕ as described in [4].

$$l_{style}^{\phi,j}(\hat{y}, y_s) = \frac{1}{C_j H_j W_j} \|G_j^\phi(\hat{y}) - G_j^\phi(y_s)\|_1 \qquad (2)$$

2.3 Clinical Data

We examined CT images from 113 IPF patients diagnosed at the University Hospitals Leuven, Belgium. CTs were evaluated by an experienced chest radiologist (author JJ) for quality i.e. absence of breathing artefacts and infection. The quality of the automated segmentation was also visually inspected to ensure contiguous airway segmentations without oversegmentation blowouts. Airway segmentations were also required to reach the sixth airway generation in the upper and lower lobes to be selected for analysis. Pulmonary function tests were considered if they occurred within 90 days of the CT scan: Forced Vital Capacity (FVC, n=111)); diffusing capacity of the lung for carbon monoxide (DLco, n=103).

The trachea and first generation bronchi were excluded from analysis. We define an airway segment as the length of airway that runs between airway branching points or an airway endpoint. All airway segments were pruned by 1 mm at either end to avoid bifurcating patches. 80×80 pixel size orthogonal airway patches were linearly interpolated with a pixel size of 0.5×0.5 mm from the CT at 0.5 mm intervals along each segment. This resulted in a final set of 546,790 real CT-derived airway patches. A synthetic dataset of 375,000 patches was generated to train our refiner.

27% of patients were female. 74% of patients had smoked previously. The median patient age was 71, with 57% of patients having died. All patients had received antifibrotic drug treatment.

Measures of intertapering, intratapering [13] and absolute airway volume were derived from the airway measurements for each airway segment. **Segmental intertapering** represents the relative difference in diameter of an airway segment when compared to its parent segment. Segmental intertapering is calculated as the difference in mean diameter, $\bar{d}$ of an airway segment and its parent segment, $\bar{d}_p$, divided by the mean diameter of the parent segment. **Segmental intratapering** is the gradient of change in diameter of the airway segment relative to the diameter of the origin of the segment[7]. Segmental intratapering is computed by dividing the gradient, m by the zero-intercept, c of a line $y = mx+c$ fitted to the diameter measurements of an airway segment. **Segmental volume** is computed by summing area measurements along an airway segment, and multiplying this value by the measurement interval, i.e. an integration of area along the segment's length.

$$intertapering = \frac{\bar{d}_p - \bar{d}}{\bar{d}_p} \qquad (3)$$

---

[7] Segments are considered to be oriented from the centre of the lung to the periphery. Accordingly, measurement of the airway origin beings at the end closest to the trachea

$$intratapering = \frac{-m}{c} \tag{4}$$

Univariable and multivariable Cox proportional hazards models were used to examine patient survival. Multivariable models included patient age (years), gender, smoking status (never vs ever) and either FVC or DLco (as measures of disease severity) as covariates. The goodness of fit of the model was denoted by the concordance index [7]. A p-value of <0.05 was considered statistically significant.

2.4   Implementation details

We use the same refiner architecture as in [20,15], the refiner is a purely convolutional network with four repeating 3x3, 64 feature ResNet blocks [8]. The measurement CNR, described in [15], is a convolutional network that feeds into two fully connected layers to learn the airway ellipse parameters. Instead of the custom CNR loss described in [15], we implemented a mean square error (MSE) loss for regressing to the airway ellipse parameters.

Synthetic images were generated to 0.5×0.5 mm pixel size making 80×80 pixel patches, corresponding to the real patch generation noted in section 2.3. All images were standardised and augmented on the fly, adding random Gaussian noise [25,25] Hounsfield units, random levels of Gaussian blurring with standard deviation scaled in the interval [0.5, 0.875] and random flipping ($p = 0.2$). We apply random scaling on real images only, in the interval [0.75,1.25] to increase diversity in airway size. Finally, a centre crop was applied to make a 32×32 pixel input patch.

Both simGAN and ATN models were trained for 10000 steps, where the simGAN refiner had 50 training iterations and the discriminator 1 iteration for every 1 step. The simGAN discriminator was implemented as described in the original method, with a memory buffer and local patch discrimination [20].

3   Results

We implemented all training on an NVIDIA GeForce RTX 2070 graphical processing unit with a batch size of 256, learning rate of 0.001, $\|l\|_1$ regularisation factor in range of [0.0001, 0.1]. Figure 1 shows training convergence for different learning rates where simGAN and ATN took 14 and 0.6 hours respectively to train. We qualitatively found that both simGAN and ATN produced refined images of optimal quality with a $\|l\|_1$ regularisation factor of 0.01.

Style-transfer from paintings to natural images show that larger-scale structure is transferred from the target image when training on losses of higher layers [10]. In order to maintain label correspondence between refiner input and output, we similarly only use the feature loss using the `relu3_3` activation layer. Style loss is computed from the two lower `relu1_2`, `relu2_2` activation layers only [8]. Figure 2 demonstrates qualitative results of our airway refinement method.

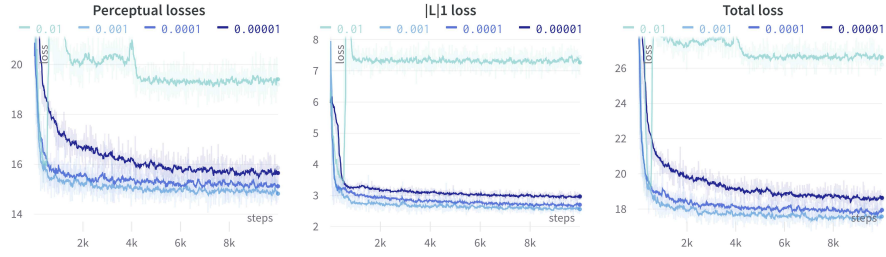[8] higher activation layers are considered in the supplementary material

Fig. 1. Comparison of loss minimisation for training the airway transformation network with different learning rates, demonstrating best model convergence at 0.001.
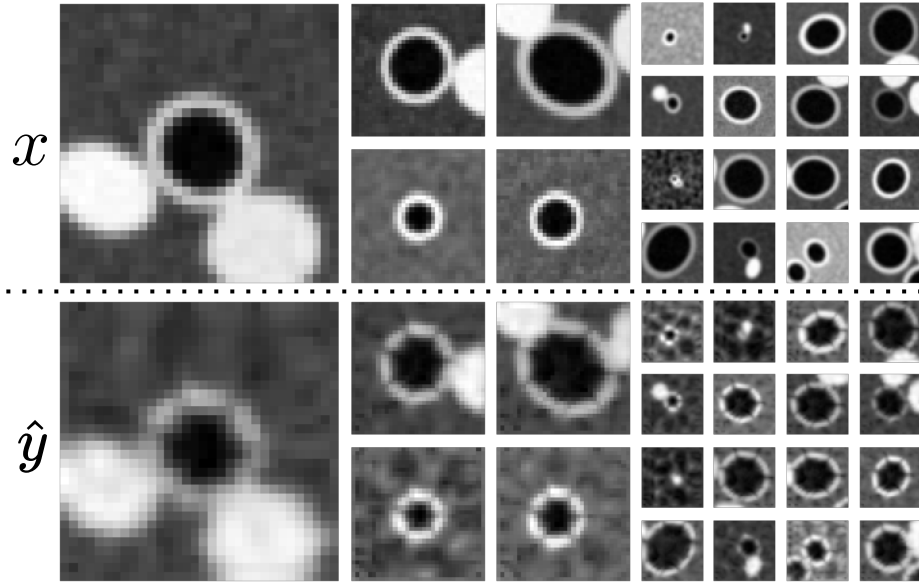

Fig. 2. Uncurated set of synthetic images x and output $\hat{y}$ of our airway transformation network in the same relative position below. Our model was trained to minimise perceptual losses. Airways are all represented at different scales.

The CNR was trained with batch size in the interval [256,2000] and learning rate of 0.001. Batch size of 2000 was chosen for its speed, and converged at around 40 epochs within one hour. The CNR achieves comparable results on ATN and simGAN refined images. Figure 3 demonstrates qualitative results of our ATN method on real CT data. Table 1 shows results of the Cox regression survival analyses.

Table 1. Cox proportional hazards results comparing mortality prediction of airway biomarkers derived by different measurement methods.

Method	Univariable (n=113)		DLCo (n=103)		FVC (n=111)	
	C index	p-value	C index	p-value	C index	p-value
Volume						
FWHMesl[19]	0.61	0.00190	0.67	0.03031	0.68	0.03965
simGAN[15]	0.65	0.00006	0.68	0.00233	0.70	0.00086
ATN(ours)	**0.67**	**0.00001**	**0.69**	**0.00013**	**0.71**	**<0.00001**
Intertapering						
FWHMesl[19]	0.55	0.07009	0.66	0.14999	0.68	0.08744
simGAN[15]	0.60	0.00925	0.67	0.03460	0.69	0.04764
ATN(ours)	**0.62**	**0.00084**	**0.69**	**0.00062**	**0.70**	**0.00052**
Intratapering						
FWHMesl[19]	0.55	0.33623	0.66	0.93103	0.69	0.63837
simGAN[15]	0.59	0.09232	0.67	0.35460	0.69	0.48513
ATN(ours)	**0.63**	**0.00026**	**0.68**	**0.00208**	0.69	**0.00192**
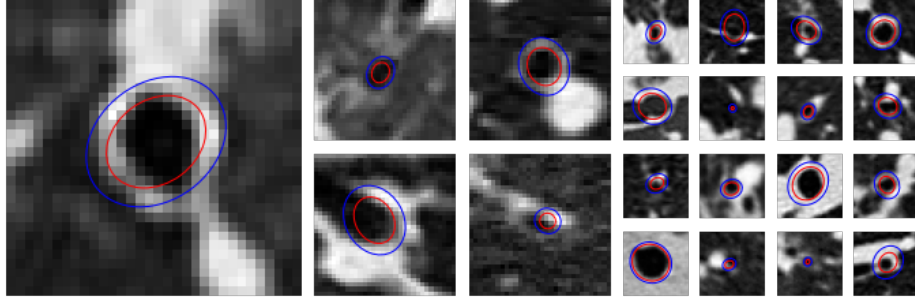
Fig. 3. Uncurated inference on real airway patches performed by our airway measurement regressor network. The network was trained on refined synthetic data from our proposed airway transformation network, which minimises perceptual losses. The inner red ellipse delineates the inner airway wall and the outer blue ellipse, the outer airway wall. Airways are all presented at different scales.

4   Conclusion

We present a learning based airway measurement method trained on a transformation network that refines synthetic data using perceptual losses. Our model ATN was compared with a state-of-the-art model simGAN [15] and a physics based method FWHMesl. When assessing the clinical utility of ATN, we found that it was the strongest predictor of survival across all three airway biomarkers. We found that our method trains faster and with minimal complications, unlike a GAN framework. We expect future work to consider the versatility of such a method, for example examining airways in patients with different pathologies, different scanner parameters and potentially on higher scale imaging such as micro-CT studies of the lungs.

Acknowledgements

References

1. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (Jun 2009). https://doi.org/10.1109/CVPR.2009.5206848
2. Estépar, R.S.J., Washko, G.G., Silverman, E.K., Reilly, J.J., Kikinis, R., Westin, C.F.: Accurate Airway Wall Estimation Using Phase Congruency. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006. vol. 4191, pp. 125–134. Springer Berlin Heidelberg (2006). https://doi.org/10.1007/11866763_16
3. Flaherty, K.R., Andrei, A.C., Murray, S., Fraley, C., Colby, T.V., Travis, W.D., Lama, V., Kazerooni, E.A., Gross, B.H., Toews, G.B., Martinez, F.J.: Idiopathic Pulmonary Fibrosis. American Journal of Respiratory and Critical Care Medicine **174**(7), 803–809 (Oct 2006). https://doi.org/10.1164/rccm.200604-488OC
4. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style (2015). https://doi.org/10.48550/ARXIV.1508.06576, https://arxiv.org/abs/1508.06576
5. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. arXiv:1406.2661 [cs, stat] (Jun 2014). https://doi.org/http://arxiv.org/abs/1406.2661
6. Gu, S., Fuhrman, C., Meng, X., Siegfried, J.M., Gur, D., Leader, J.K., Sciurba, F.C., Pu, J.: Computerized identification of airway wall in CT examinations using a 3D active surface evolution approach. Medical Image Analysis **17**(3), 283–296 (Apr 2013). https://doi.org/10.1016/j.media.2012.11.003
7. HARRELL Jr., F.E., LEE, K.L., MARK, D.B.: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine **15**(4), 361–387 (1996). https://doi.org/https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4¡361::AID-SIM168¿3.0.CO;2-4
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition (Dec 2015). https://doi.org/10.48550/arXiv.1512.03385
9. Hofmanninger, J., Prayer, F., Pan, J., Rohrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem. arXiv:2001.11767 [physics, stat] (Jan 2020). https://doi.org/10.1186/s41747-020-00173-2
10. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual Losses for Real-Time Style Transfer and Super-Resolution (Mar 2016). https://doi.org/10.48550/arXiv.1603.08155
11. Kiraly, A.P., Reinhardt, J.M., Hoffman, E.A., McLennan, G., Higgins, W.E.: Virtual bronchoscopy for quantitative airway analysis. In: Amini, A.A., Manduca, A.

(eds.) Medical Imaging 2005: Physiology, Function, and Structure from Medical Images. vol. 5746, p. 369. International Society for Optics and Photonics (Apr 2005). https://doi.org/10.1117/12.595283

12. Kluvanec, D., Phillips, T.B., McCaffrey, K.J.W., Moubayed, N.A.: Using Orientation to Distinguish Overlapping Chromosomes (Mar 2022). https://doi.org/10.48550/arXiv.2203.13004

13. Kuo, W., Perez-Rovira, A., Tiddens, H., de Bruijne, M.: Airway tapering: an objective image biomarker for bronchiectasis. European Radiology **30**(5), 2703–2711 (May 2020). https://doi.org/10.1007/s00330-019-06606-w

14. Lederer, D.J., Martinez, F.J.: Idiopathic Pulmonary Fibrosis. New England Journal of Medicine **378**(19), 1811–1823 (May 2018). https://doi.org/10.1056/NEJMra1705751

15. Nardelli, P., Ross, J.C., San José Estépar, R.: Generative-based airway and vessel morphology quantification on chest CT images. Medical Image Analysis **63**, 101691 (Jul 2020). https://doi.org/10.1016/j.media.2020.101691

16. Pakzad, A., Cheung, W.K., Quan, K., Mogulkoc, N., Van Moorsel, C.H.M., Bartholmai, B.J., Van Es, H.W., Ezircan, A., Van Beek, F., Veltkamp, M., Karwoski, R., Peikert, T., Clay, R.D., Foley, F., Braun, C., Savas, R., Sudre, C., Doel, T., Alexander, D.C., Wijeratne, P., Hawkes, D., Hu, Y., Hurst, J.R., Jacob, J.: Evaluation of automated airway morphological quantification for assessing fibrosing lung disease. Tech. Rep. arXiv:2111.10443, arXiv (Nov 2021). https://doi.org/10.48550/ARXIV.2111.10443

17. Pakzad, A., Jacob, J.: Radiology of Bronchiectasis. Clinics in Chest Medicine **43**(1), 47–60 (Mar 2022). https://doi.org/10.1016/j.ccm.2021.11.004

18. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

19. Quan, K., Shipley, R.J., Tanno, R., McPhillips, G., Vavourakis, V., Edwards, D., Jacob, J., Hurst, J.R., Hawkes, D.J.: Tapering analysis of airways with bronchiectasis. In: Angelini, E.D., Landman, B.A. (eds.) Medical Imaging 2018: Image Processing. vol. 10574, p. 87. SPIE (Mar 2018). https://doi.org/10.1117/12.2292306

20. Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R.: Learning from simulated and unsupervised images through adversarial training (2017). https://doi.org/10.48550/ARXIV.1612.07828, https://arxiv.org/abs/1612.07828

21. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs] (Apr 2015). https://doi.org/10.48550/ARXIV.1409.1556

22. Weibel, E.R.: Morphometry of the Human Lung. Springer, Berlin, Heidelberg (1963). https://doi.org/10.1007/978-3-642-87553-3_6

23. Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L.R., Summers, R.M., Rubin, D.L., Lungren, M.P.: Preparing Medical Imaging Data for Machine Learning. Radiology **295**(1), 4–15 (Apr 2020). https://doi.org/10.1148/radiol.2020192224

24. Xie, M., Liu, X., Cao, X., Guo, M., Li, X.: Trends in prevalence and incidence of chronic respiratory diseases from 1990 to 2017. Respiratory Research **21**(1), 1–13 (Dec 2020). https://doi.org/10.1186/s12931-020-1291-8

25. Xu, Z., Bagci, U., Foster, B., Mansoor, A., Udupa, J.K., Mollura, D.J.: A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT. Medical Image Analysis **24**(1), 1–17 (Aug 2015). https://doi.org/10.1016/j.media.2015.05.003
26. Yu, F., Koltun, V.: Multi-Scale Context Aggregation by Dilated Convolutions. arXiv:1511.07122 [cs] (Apr 2016). https://doi.org/10.48550/ARXIV.1511.07122

A Supplementary Material

A.1 Style loss ablation study

We consider the effects of minimising style loss with higher layers of ϕ in Figure 4. The pretrained VGG-16 layers of imagenet from low to high was `relu1_2`, `relu2_2`, `relu3_3` and `relu4_3`. The style reconstruction losses are summed consecutively for each higher layer considered. This results in the final loss computed for `relu4_3` being the previous 3 activation layers' losses.

We find the most significant qualitative difference occurs when moving from `relu1_2` to `relu2_2`, where `relu1_2` appears quite smooth. The addition of each higher layer appears to add higher frequency details. The final layers introduce visually discernible larger-scale spatial changes, particularly in the smaller airways. Informed by this experiment and previous style transfer experiments by [10], we chose to only train on style reconstruction losses computed from `relu1_2` and `relu2_2`.
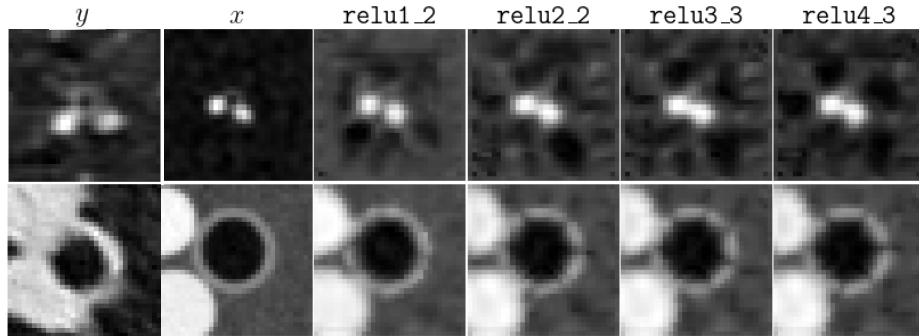

Fig. 4. Similar to [10], for two example synthetic inputs x we demonstrate the effect of minimising for style reconstruction loss using different activation layers of the pretrained VGG-16 loss network. Losses are accumulated with each higher layer. y taken from a real clinical CT scan is similar in appearance to the synthetic images.