

Convolutional Neural Networks with Template-Based Data Augmentation for Functional Lung Image Quantification

Nicholas J. Tustison¹, Brian B. Avants², Zixuan Lin¹, Xue Feng¹, Nicholas Cullen³, Jaime F. Mata¹, Lucia Flors⁴, James C. Gee³, Talissa A. Altes⁴, John P. Mugler III¹, and Kun Qing¹

¹Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA

²Cingulate, Hampton, NH

³Department of Radiology, University of Pennsylvania, Philadelphia, PA

⁴Department of Radiology, University of Missouri, Columbia, MO

Corresponding author:

Nicholas J. Tustison

ntustison@virginia.edu

Rationale and Objectives: We propose an automated segmentation pipeline based on deep learning for proton lung MRI segmentation and ventilation-based quantification which improves on our previously reported methodologies in terms of computational efficiency while demonstrating accuracy and robustness. The large data requirement for the proposed framework is made possible by a novel template-based data augmentation strategy. Supporting this work is the open-source *ANTsR-Net*—a growing repository of well-known deep learning architectures first introduced here.

Materials and Methods: Deep convolutional neural network (CNN) models were constructed and trained using a custom multilabel Dice metric loss function and a novel template-based data augmentation strategy. Training (including template generation and data augmentation) employed 205 proton MR images and 73 functional lung MRI. Evaluation was performed using data sets of size 63 and 40 images, respectively.

Results: Accuracy for CNN-based proton lung MRI segmentation (in terms of Dice overlap) was left lung: 0.93 ± 0.03 , right lung: 0.94 ± 0.02 , and whole lung: 0.94 ± 0.02 . Although slightly less accurate than our previously reported joint label fusion (JLF) approach (left lung: 0.95 ± 0.02 , right lung: 0.96 ± 0.01 , whole lung: 0.96 ± 0.01), processing time is < 1 second per subject for the proposed approach versus ~ 30 minutes per subject using JLF. Accuracy for quantifying ventilation defects was determined based on a consensus labeling where average accuracy (Dice multilabel overlap of ventilation defect regions plus normal region) was 0.94 for the CNN method; 0.92 for our previously reported method; and 0.90, 0.92, and 0.94 for expert readers.

Conclusions: The proposed framework yields accurate automated quantification in near real time. CNNs drastically reduce processing time after offline model construction and demonstrate significant future potential for facilitating quantitative analysis of functional lung MRI.

Key Words: Advanced Normalization Tools, ANTsRNet, hyperpolarized gas imaging, neural networks, proton lung MRI, U-net

INTRODUCTION

Probing lung function under a variety of conditions and/or pathologies has been significantly facilitated by the use of hyperpolarized gas imaging and corresponding quantitative image analysis methodologies. Such developments have provided direction and opportunity for current and future research trends [1]. Computational techniques targeting these imaging technologies permit spatial quantification of localized ventilation with potential for increased reproducibility, resolution, and robustness over traditional spirometry and radiological readings [2, 3].

One of the most frequently used image-based biomarkers for the study of pulmonary development and disease is based on the quantification of regions of limited ventilation, also known as *ventilation defects* [4]. These features have been shown to be particularly salient in a clinical context. For example, ventilation defect volume to total lung volume ratio has been shown to outperform other image-based features in discriminating asthmatics vs. non-asthmatics [5]. Ventilation defects have also demonstrated discriminative capabilities in chronic obstructive pulmonary disease (COPD) [6] and asthma [7]. These findings, along with related research, have motivated the development of multiple automated (and semi-automated) segmentation algorithms which have been proposed in the literature (e.g., [8–12]) and are currently used in a variety of clinical research investigations (e.g., [13]).

Despite the enormous methodological progress with existing quantification strategies, recent developments in machine learning (specifically “deep learning” [14]) have generated new possibilities for quantification with improved capabilities in terms of accuracy, robustness, and computational efficiency. Deep learning, a term denoting neural network architectures with multiple hidden layers, has seen recent renewed research development and application. In the field of image analysis and computer vision, deep learning with convolution neural networks (CNNs) has been particularly prominent in recent years due, in large part, to the annual ImageNet Large Scale Visual Recognition Challenge [15]. Specifically, one of the participating groups in the 2012 ImageNet challenge was the earliest adopter of CNNs. The resulting architecture, colloquially known as “AlexNet” [16], surpassed any approach that had been proposed previously and laid the groundwork for future CNN-based architectures for image classification such as VGG [17] and GoogLeNet [18]. The recent successes of

CNNs are historically rooted in the pioneering work of LeCun et al. [19] and Fukushima [20] and others which drew inspiration from earlier work on the complex arrangement of cells within the feline visual cortex [21]. CNNs are characterized by common components (i.e., convolution, pooling, and activation functions) which can be put together in various arrangements to perform such tasks as image classification and voxelwise segmentation. The outgrowth of research, in conjunction with advances in computational hardware, has resulted in significant developments in various image research areas including classification, segmentation, and object localization and has led to co-optation by the medical imaging analysis community [22].

In this work, we develop and evaluate a convolutional neural network segmentation framework, based on the U-net architecture [23], for functional lung imaging using hyperpolarized gas. As part of this framework we include a deep learning analog to earlier work from our group targeting segmentation of proton lung MRI [24]. This is motivated by common use case scenarios in which [proton images are used to identify regions of interest in corresponding ventilation images \[8–10\]](#), which typically contain no discernible boundaries for anatomic structures.

One of the practical constraints to adopting deep learning techniques is the large data requirement for the training process oftentimes necessitating ad hoc strategies for simulating additional data from available data—a process termed *data augmentation* [25]. While common approaches to data augmentation include the application of randomized simulated linear (e.g., translation, rotation and affine) or elastic transformations and intensity adjustments (e.g., brightness and contrast), we advocate a tailored paradigm to commonly encountered medical imaging scenarios in which data is limited but is assumed to be characterized by a population-wide spatial correspondence. In the proposed approach, an optimal shape-based template is constructed from a subset of the available data. Subsequent pairwise image registration between all training data and the resulting template permits a “pseudo-geodesic” transformation [26] of each image to every other image thus potentially converting a data set of size N to an augmented data set of size N^2 . In this way, transformations are constrained to the shape space representing the population of interest.

To enhance relevance to the research community, we showcase this work in conjunction with the introduction of *ANTsRNet*—a growing open-source repository of well-known deep learning architectures which interfaces with the Advanced Normalization Tools (ANTs) package [27] and its R

package, ANTsR [28]. This permits the public distribution of all code, data, and models for external reproducibility which can be found on the GitHub repository corresponding to this manuscript [29]. This allows other researchers to apply the developed models and software to their data and/or use the models to initialize their own model development tailored to specific studies.

In the work described below, we first provide the acquisition protocols for both the proton and ventilation MR images followed by a discussion of the analysis methodologies for the proposed segmentation framework. This is contextualized with a brief overview of existing quantification methods (including that previously proposed by our group and used for the evaluative comparison). We also summarize the key contributions of this work viz., the template-based data augmentation and the current feature set of ANTsRNet.

MATERIALS AND METHODS

Image acquisition

Both proton and ventilation images used for this study were taken from current and previous studies from our group. Ventilation images comprised both helium-3 and xenon-129 acquisitions as our current segmentation processing does not distinguish between ventilation gas acquisition protocols and we expected similar agnosticism for the proposed approach.

Hyperpolarized MR image acquisition was performed under an Institutional Review Board (IRB)-approved protocol with written informed consent obtained from each subject. In addition, all imaging studies [were](#) performed under Food and Drug Administration (FDA)-approved physician's Investigational New Drug applications for hyperpolarized gas (either helium-3 or xenon-129). MRI data were acquired on a 1.5 T whole-body MRI scanner (Siemens Sonata, Siemens Medical Solutions, Malvern, PA) with broadband capabilities and corresponding hyperpolarized-gas chest radiofrequency coils (Rapid Biomedical, Rimpar, Germany; IGC Medical Advances, Milwaukee, WI; or Clinical MR Solutions, Brookfield, WI).

Two imaging protocols were used to acquire the MR images. Both of them are combined hyperpo-

larized gas (helium-3 or xenon-129) and proton imaging acquisitions. Protocol 1 uses 3-D balanced steady-state free-precession or spoiled gradient-echo pulse sequences with isotropic resolution = 3.9 mm, TR = 1.75–1.85 ms, TE = 0.78–0.82 ms, flip angle= 9–10°, bandwidth per pixel=1050–1100 Hz/Pixel, total duration = 10–20 seconds. Protocol 2 uses a contiguous, coronal, 2-D gradient-echo pulse sequence with interleaved spiral sampling scheme, in-plane resolution = 2–4 mm, slice thickness = 15 mm, TR = 8–8.5 ms, TE = 0.8–1.0 ms, flip angle = 20° interleaves = 12–20 (plus 2 for field map), total duration = 3–8 seconds. All subjects provided written informed consent and the data were de-identified prior to analysis.

Image processing and analysis

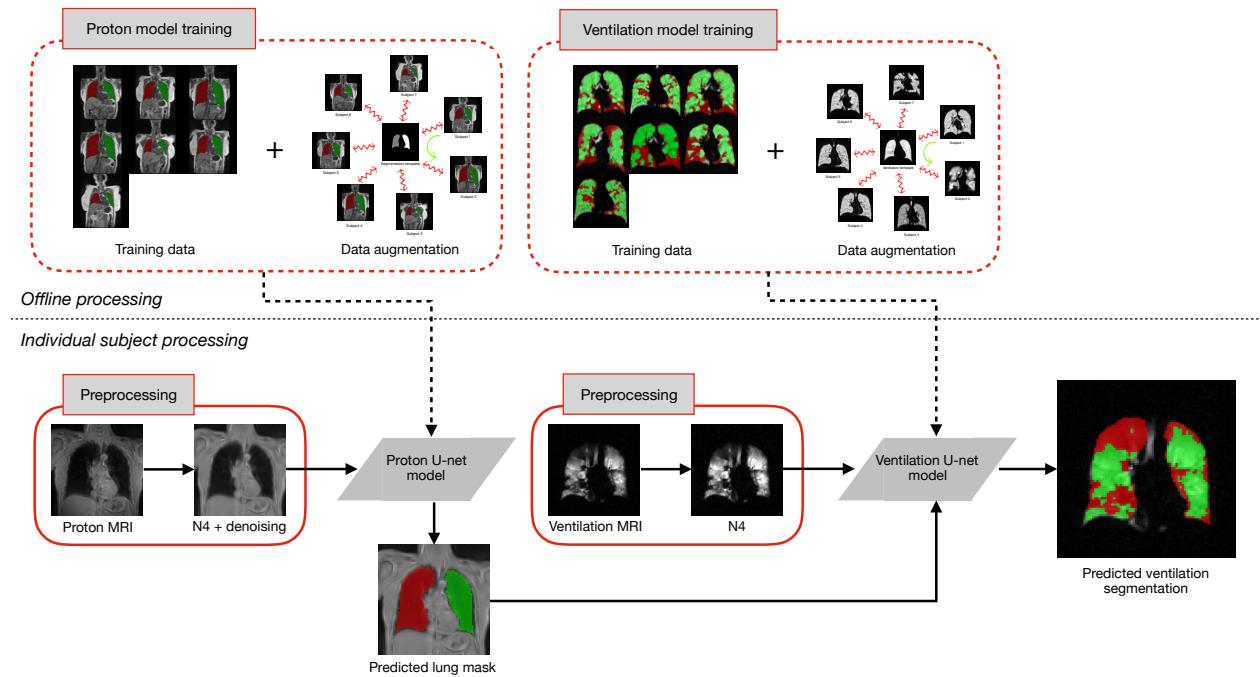


Figure 1: Illustration of the proposed workflow. Training the U-net models for both proton and ventilation imaging includes template-based data augmentation. This offline training is computationally intensive but is only performed once. Subsequent individual subject preprocessing includes MR denoising and bias correction. The proton mask determined from the proton U-net model is included as a separate channel (in deep learning software parlance) for ventilation image processing.

We first review our previous contributions to the segmentation of proton and hyperpolarized gas MR images [8, 24] as we use these previously described techniques for evaluative comparison. We then

describe the deep learning analogs (including preprocessing) extending earlier work and discuss the proposed contributions which include:

- convolution neural networks for structural/functional lung segmentation,
- template-based data augmentation, and
- open-source availability.

An overview of the resulting framework is provided in Figure 1. The most computationally intensive portion is the offline processing for model training for both structural and functional imaging. However, once that is complete, individual processing consists of a couple of preprocessing steps followed by application of the models which has minimal computational requirements.

Previous approaches from our group for lung and ventilation-based segmentation

The automated ventilation-based segmentation, described in [8], employs a Gaussian mixture model with a Markov random field (MRF) spatial prior optimized via the Expectation-Maximization algorithm. The resulting software, called Atropos, has been used in a number of clinical studies (e.g., [7, 30]). Briefly, the intensity histogram profile of the ventilation image is modeled using Gaussian functions with optimizable parameters (i.e., mean, standard deviation, and normalization factor) designed to model the intensities of the individual ventilation classes. At each iteration the resulting estimated voxelwise labels are refined based on MRF spatial regularization. The parameters of the class-specific Gaussians are then re-estimated. This iterative process continues until convergence. We augment this segmentation step by iterating the results with application of N4 bias correction [31]. [Unlike other methods that rely solely on intensity distributions, thereby discarding spatial information](#) (e.g., K-means variants [9, 11] and histogram rescaling and thresholding [10]), our technique employs both spatial and intensity information for probabilistic classification.

Because of our dual structural/functional acquisition protocol [32], we also previously formulated a joint label fusion (JLF)-based framework [33] for segmenting the left and right lungs in proton MRI as well as estimating the lobar volumes [24]. This permits us to first identify the lung mask in the proton MRI. This information is transferred to the space of the corresponding ventilation MR image via image registration. The JLF method relies on a set of atlases (proton MRI plus lung labels) which

is spatially normalized to an unlabeled image where a weighted consensus of the normalized images and labels is used to determine each voxel label. Although the method yields high quality results which are fully automated, one of the drawbacks is the time and computational resources required to perform the image registration for each member of the atlas set and the subsequent voxelwise label consensus estimation.

Note that we have provided self-contained examples for both of these segmentation algorithms using ANTs tools: lung and lobe estimation [34] and lung ventilation [35]. However, given the previously outlined benefits of deep learning approaches to these same applications, we expect that adoption by other groups will be greatly facilitated by the proposed algorithms described below.

Preprocessing

Because of the low-frequency imaging artifacts introduced by confounds such as radiofrequency coil inhomogeneity, we perform a retrospective bias correction on both proton and ventilation images using the N4 algorithm [36]. These are included in our previously proposed ventilation [8] and structural [24] segmentation frameworks. Since the initial release of these pipelines we have also adopted an adaptive, patch-based denoising algorithm specific to MR [37] which we have reimplemented in the ANTs toolkit. The effects of these data cleaning techniques on both the proton images and ventilation images are shown in Figure 2.

U-net architecture for structural/functional lung segmentation

The U-net architecture was introduced in [23] which extended the fully convolutional neural network (FCN) approach introduced by Long, Shelhamer, and Darrel [38]. U-net augments the “encoding path” (see left side of Figure 3) common to such architectures as VGG and FCN with a symmetric decoding path where the corresponding encoding/decoding layers are linked via skip paths for enhanced feature detection. The nomenclature reflects the descending/ascending aspect of its architecture. Each series in both encoding and decoding branches is characterized [by an optional dropout layer in between two convolutional layers](#). This latter modification from the original is meant to provide additional regularization for over-fitting prevention [39]. Output consists of a segmentation

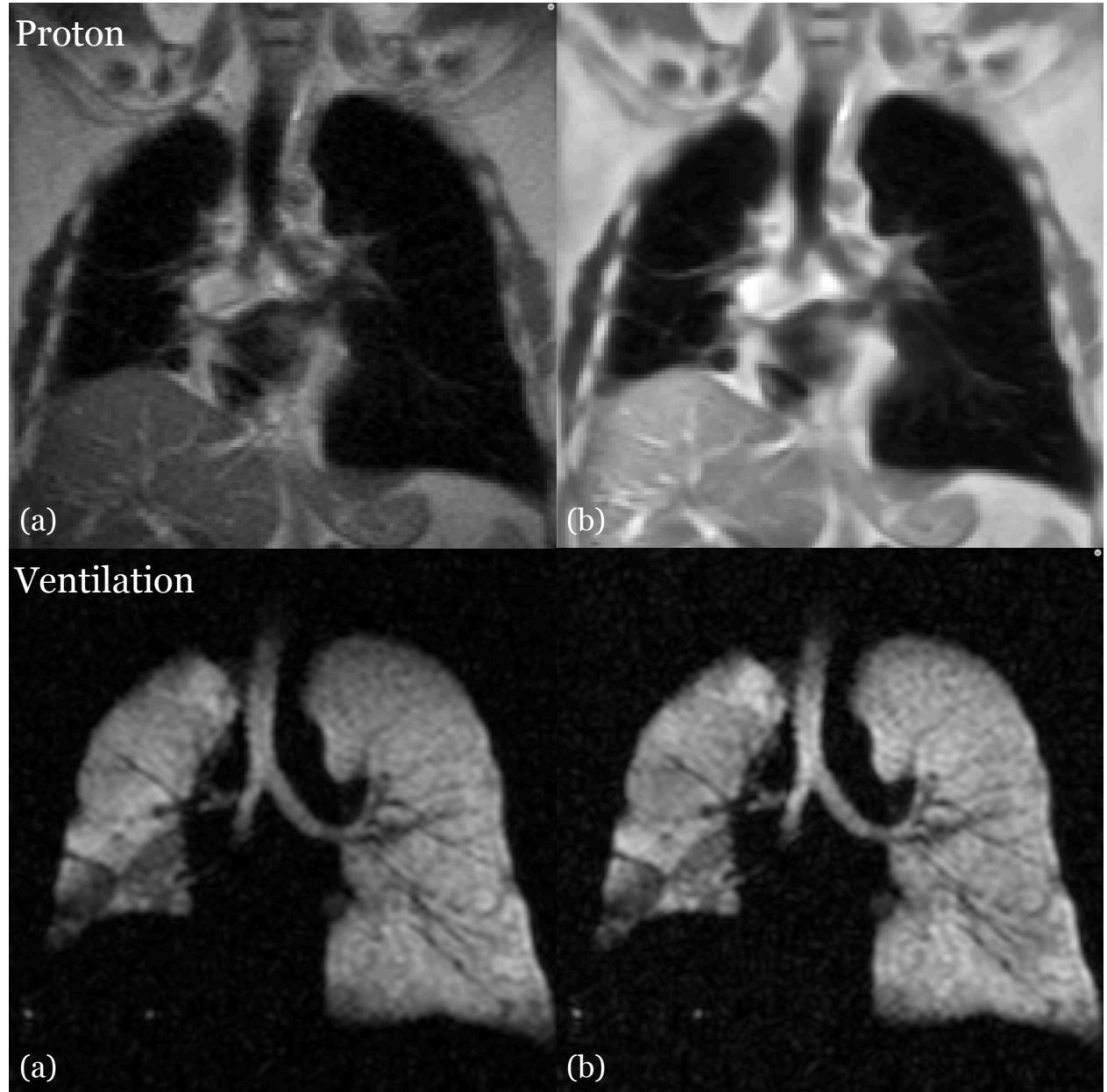


Figure 2: Side-by-side image comparison showing the effects of preprocessing on the proton (top) and ventilation (bottom) MRI. (a) Uncorrected image showing MR field inhomogeneity and noise. (b) Corresponding corrected image in which the bias effects have been ameliorated.

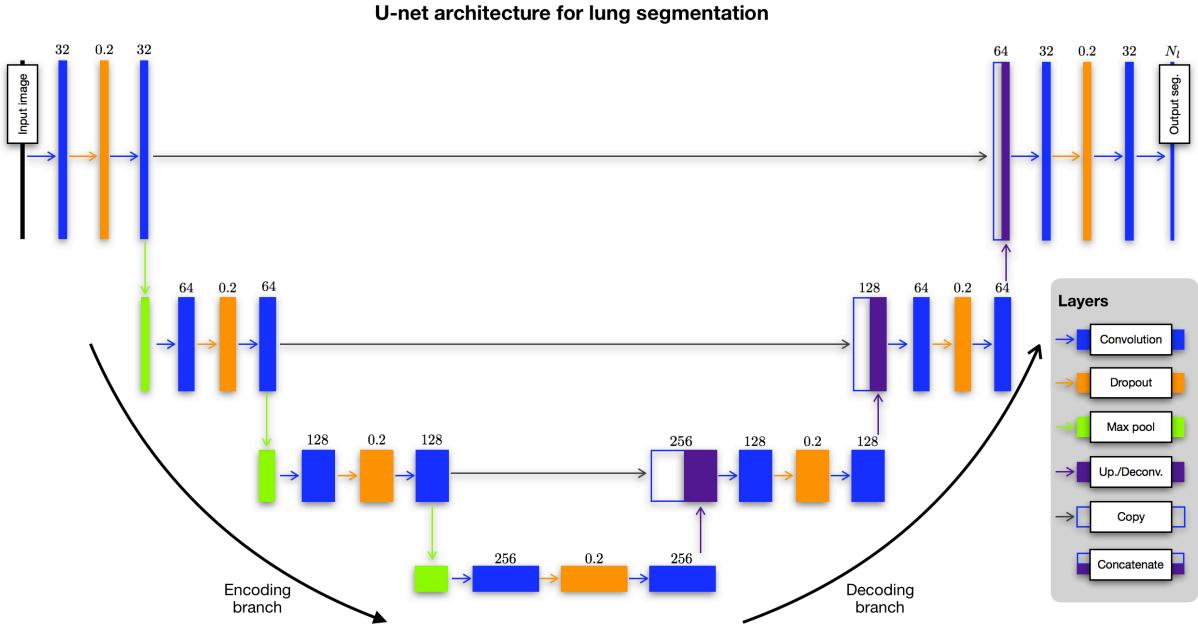


Figure 3: The modified U-net architecture for both structural and functional lung segmentation (although certain parameter values, specifically the number of filters per convolution layer, are specific to the functional case). Network layers are represented as boxes with arrows designating connections between layers. The main parameter value for each layer is provided above the corresponding box. Each layer of the descending (or “encoding”) branch of the network is characterized by two convolutional layers. Modification of the original architecture includes an intermediate dropout layer for regularization (dropout rate = 0.2). A max pooling operation produces the feature map for the next series. The ascending (or “decoding”) branch is similarly characterized. A convolutional transpose operation is used to upsample the feature map following a convolution → dropout → convolution layer series until the final convolutional operation which yields the segmentation probability maps.

probability image for each label from which a segmentation map can be generated.

We used the U-net architecture to build separate models for segmenting both structural and functional lung images. For cases where dual acquisition provides both images, we use the structural images to provide a mask for segmentation of the ventilation image. We used an open-source implementation written by our group and provided with the ANTsRNet R package [40] which is described in greater detail below. We also implemented a multi-label dice coefficient loss function along with specific batch generators for generating augmented image data on the fly.

Template-based data augmentation

The need for large training data sets is a well-known limitation of deep learning algorithms. Whereas the architectures developed for such tasks as the ImageNet competition have access to millions of annotated images for training, such data availability is atypical in medical imaging. In order to achieve data set sizes necessary for learning functional models, various data augmentation strategies have been employed [25]. These include application of intensity transformations, such as brightening and enhanced contrast. They might also include spatial transformations such as arbitrary rotations, translations, and even simulated elastic deformations. Such transformations might not be ideal if they do not represent shape variation **within the typical range exhibited by the** population under study.

We propose a template-based data augmentation approach whereby image data sampled from the population is used to construct a representative template that is optimal in terms of shape and/or intensity [41]. In addition to the representative template, this template-building process yields the transformations to/from each individual image to the template space. **This permits a propagation of individual training data to the space of every other training data as illustrated in Figure 4.** Specifically, the template building process produces an invertible, deformable mapping for each subject. For the k^{th} subject, S_k , and template, T , this mapping is denoted as ϕ_k with inverse given by ϕ_k^{-1} . During model training, each new augmented data instance, S_{new} , is produced by randomly selecting a source subject and target subject, and mapping S_{source} to the space of S_{target} according to

$$S_{new} = S_{source} \left(\phi_{target}^{-1} (\phi_{source}) \right). \quad (1)$$

Note that each S comprises all channel images and corresponding segmentation images. In the simplest case, the training data is used to construct the template and then each individual training image and corresponding labels are propagated to the space of every other image. In this way, a training data set of size N can be expanded to a data set of size N^2 . A slight variation to this would be to build a template from M data sets (where $M > N$). Transformations between the training data and the template are then used to propagate the training data to the spaces of the individual members of the template-generating data for an augmented data set size of $M \times N$.

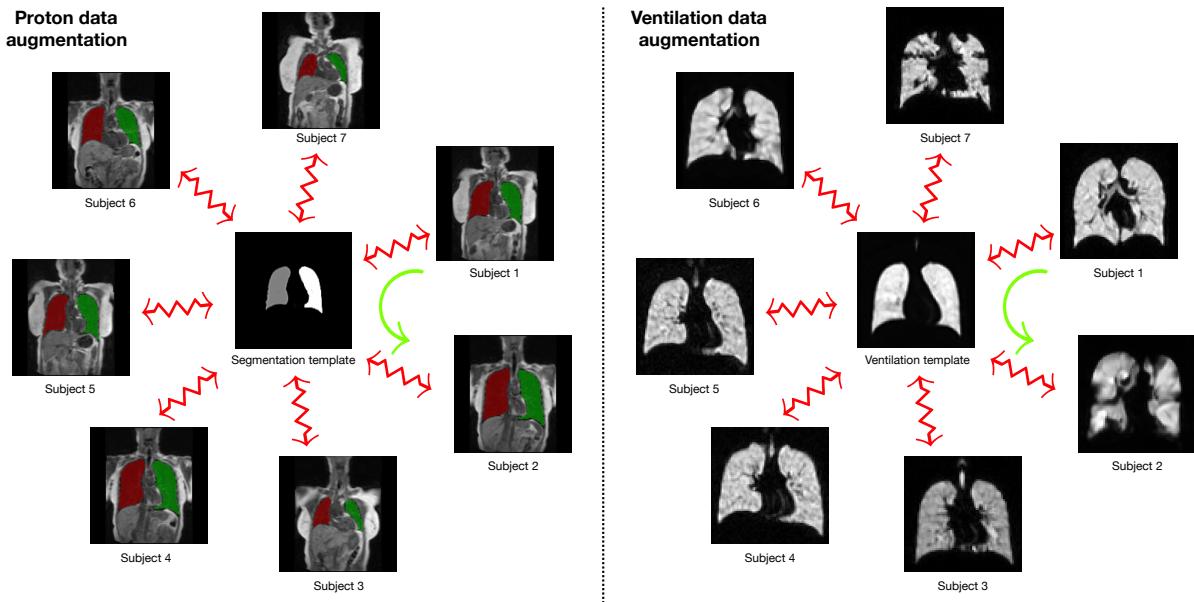


Figure 4: Template-based data augmentation for the proton (left) and ventilation (right) U-net model generation. For both cases, a template is created, or selected, to generate the transforms to and from the template. The derived deformable, invertible transform for the k^{th} subject, S_k , to the template, T , is denoted by $\phi_k : S_k \rightsquigarrow T$. These subject-specific mappings are used during model training (but not the template itself). Data augmentation occurs by randomly choosing a reference subject and a target subject during batch processing. In the illustration above, the sample mapping of Subject 1 to the space of Subject 2, represented by the green curved arrow, is defined as $\phi_2^{-1}(\phi_1)$.

Since U-net model generation is completely separate for the proton and ventilation data, template-based data augmentation is also isolated between the two protocols. The proton template is created from the right/left segmentation images of the training data denoted by the red/green labels which have voxel values of ‘1’ and ‘2’, respectively, in Figure 4. The resulting template located in the center of the left panel is an average of all the transformed label images. This whole lung approach avoids the possible lack of internal correspondence while generating plausible global shape variations when mapping between individual training data. We used 60 proton MR images thus permitting $60^2 = 3600$ possible deformable shapes which can be further augmented by more conventional strategies (e.g., brightness transformations, translations, etc.). In contrast, the ventilation template is created directly from the training ventilation images resulting in the grayscale template in the center of the right panel of Figure 4. An alternative strategy for ventilation template creation could have employed relabeling the ventilation masks to a single value and aligning the lung boundaries, somewhat similar to the proton protocol. This approach would be necessary for severe lung disease where major portions of the lung are absent in the ventilation image.

ANTsRNet

In addition to the contributions previously described, we also introduce ANTsRNet [40] to the research community which not only contains the software to perform the operations specific to structural and functional lung image segmentation but also performs a host of other deep learning tasks wrapped in a thoroughly documented and well-written R package. The recent interest in deep learning techniques and the associated successes with respect to a variety of applications have motivated adoption of such techniques. Basic image operations such as classification, object identification, segmentation, as well as more focused techniques, such as predictive image registration [42], have significant potential for facilitating basic medical research. ANTsRNet is built using the Keras neural network library (available through R) and is highly integrated with the ANTsR package, the R interface of the ANTs toolkit. Consistent with our other software offerings, ongoing development is currently carried out on GitHub using a well-commented coding style, thorough documentation, and self-contained working examples [40].

Several architectures have been implemented for both 2-D and 3-D images spanning the broad appli-

cation areas of image classification, object detection, and image segmentation (cf. Table 1). It should be noted that most reporting in the literature has dealt exclusively with 2-D implementations. This is understandable due to memory and computational speed constraints limiting practical 3-D applications on current hardware. However, given the importance that 3-D data has for medical imaging and the rapid progress in hardware, we feel it worth the investment in implementing corresponding 3-D architectures. Each architecture is accompanied by one or more self-contained examples for testing and illustrative purposes. In addition, we have made novel data augmentation strategies available to the user and illustrated them with Keras-specific batch generators.

Processing specifics

205 proton MR images each with left/right lung segmentations and 73 ventilation MR images with masks were used for the separate U-net model training. These images were denoised and bias corrected offline (as described above) and required < 1 minute for both steps per image using single-threading although both preprocessing steps are multi-threading capable. An R script was used to read in the images and segmentations (available in our GitHub [repository](#) [29]), create the model, designate model parameters, and initialize the batch generator.

For the proton data we built a 3-D U-net model to take advantage of the characteristic 3-D shape of the lungs. This limited the possible batch size as our GPU (Titan Xp) is limited to 12 GB although this can be revisited in the future with additional computational resources. [Transforms derived from the template-building process described previously](#) were passed to the batch generator where reference and source subjects were randomly assigned. During each iteration, these random pairings were used to create the augmented data according to Equation (1). These data are publicly available for download at <https://doi.org/10.6084/m9.figshare.4964915.v1>.

The U-net ventilation model was generated from 73 ventilation MRI. The smaller data set size was a result of data pruning to ensure class balance. Even though the functional images are processed as 3-D volumes and a 3-D ventilation template is created for the template-based data augmentation, the generated U-net model is 2-D. This is due to lack of any discernible anatomical signatures available for learning. This also makes model generation much faster. The basic processing strategy is that

Table 1: Current ANTsRNet capabilities comprising architectures for applications in image segmentation, image classification, object localization, and image super-resolution. Self-contained examples with data are also provided to demonstrate usage for each of the architectures. Although the majority of neural network architectures are originally described for 2-D images, we generalized the work to 3-D implementations where possible.

ANTsRNet		
Image segmentation		
U-net [15]	(2-D)	Extends fully convolutional neural networks by including an upsampling decoding path with skip connections linking corresponding encoding/decoding layers.
V-net [45]	(3-D)	3-D extension of U-net which incorporates a customized Dice loss function.
Image classification		
AlexNet [32]	(2-D, 3-D)	Convolutional neural network that precipitated renewed interest in neural networks.
VGG16/VGG19 [33]	(2-D, 3-D)	Also known as 'OxfordNet'. VGG architectures are much deeper than AlexNet. Two popular styles are implemented.
GoogLeNet [34]	(2-D)	A 22-layer network formed from <i>inception blocks</i> meant to reduce the number of parameters relative to other architectures.
ResNet [46]	(2-D, 3-D)	Characterized by specialized <i>residualized blocks</i> (and skip connections).
ResNeXt [48]	(2-D, 3-D)	A variant of ResNet distinguished by a hyper-parameter called <i>cardinality</i> defining the number of independent paths.
DenseNet [49]	(2-D, 3-D)	Based on the observation that performance is typically enhanced with shorter connections between the layers and the input.
Object localization		
SSD300/SSD512 [50]	(2-D, 3-D)	The Multibox Single-Shot Detection (SSD) algorithm for determining bounding boxes around objects of interest.
SSD7	(2-D, 3-D)	Lightweight SSD variant which increases speed by slightly sacrificing accuracy. Training size requirements are smaller.
Image super-resolution		
SRCNN/ESRCNN [51]	(2-D, 3-D)	Image super-resolution using CNNs.

any ventilation image to be segmented will be processed on a slice-by-slice basis where each slice is segmented using the 2-D model. For data augmentation, the full 3-D transforms are supplied to the batch generator. At each iteration, a set of generated 3-D augmented images are created on the fly based on Equation (1) and then a subset of slices is randomly selected for each image until the batch set is complete. For this work we randomly sampled slices in the coronal direction using a specified sampling rate ($= 0.5$). It should be noted that a whole lung mask is assumed to exist and supplied as an additional channel for processing. Prediction on the evaluation data is performed slice-by-slice and then collated into probability volumetric images.

The 3-D U-net structural model and 2-D U-net functional model that were created, as described above and included with the GitHub repository, can then be used for future processing. Using these models, the basic deep learning workflow is as follows:

- Assume simultaneous structural and functional image acquisition.
- Generate left/right lung mask from the structural image using corresponding 3-D U-net structural model.
- Convert left/right lung mask to a single label.
- Process ventilation image slice-by-slice using the 2-D U-net functional model using the ventilation image and single label structural lung mask as input.

Note that this workflow is not absolutely necessary. For example, one could use the 2-D functional model if one provides the input lung masks.

Image size was not identical across both image cohorts so we settled on a common resampled image size of $128 \times 128 \times 64$ for the proton images and 128×128 for the ventilation images. Resampling of each image and segmentation was handled internally by the batch generator after transformation to the reference image using ANTsR functions [28]. Additionally, during data augmentation for proton model optimization, a digital “coin flip” was used to randomly vary the intensity profile of the warped proton images between their original profiles and the intensity profile of the randomly selected reference image. The latter intensity transformation is the histogram matching algorithm of Nyul et al. [43] implemented in the Insight Toolkit. Specific parameters for the U-net architecture for both models are as follows (3-D parameters are included in parentheses):

- Adam optimization:
 - proton model learning rate = 0.00001
 - ventilation model learning rate = 0.0001
- Number of epochs: 150
- Training/validation data split: 80/20
- Convolution layers
 - kernel size: $5 \times 5 (\times 5)$
 - activation: rectified linear units (ReLU) [44]
 - number of filters: doubled at every layer starting with $N = 16$ (proton) and $N = 32$ (ventilation)
- Dropout layers
 - rate: 0.2
- Max pooling layers
 - size: $2 \times 2 (\times 2)$
 - stride length: $2 \times 2 (\times 2)$
- Upsampling/transposed convolution (i.e., deconvolution) layers
 - kernel size: $5 \times 5 (\times 5)$
 - stride length: $2 \times 2 (\times 2)$
 - activation: rectified linear units (ReLU) [44]

Training took approximately 10 hours for both models. After model construction, prediction per image (after preprocessing) takes < 1 second per image. Both model construction and prediction utilized a NVIDIA Titan Xp GPU.

RESULTS

Proton MRI lung segmentation

After constructing the model, we applied it to the evaluation data consisting of the same 62 proton MRI used in [8]. We performed a direct comparison with the joint label fusion (JLF) method of [8]

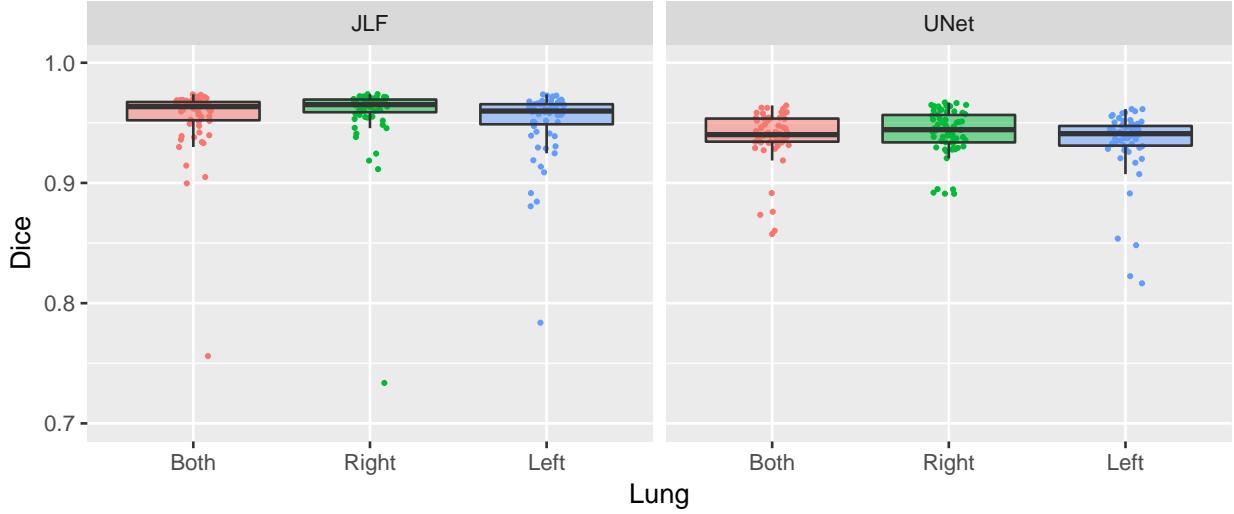


Figure 5: The Dice overlap coefficient for the left and right lungs (and their combination) between the updated joint label fusion technique (left) and our deep learning approach (right). Although slightly less accurate, the latter requires significantly less computation time.

with an adopted modification that we currently use in our studies. Instead of using the entire atlas set (which would require a large number of pairwise image registrations), we align the center of the image to be segmented with each atlas image and compute a neighborhood cross-correlation similarity metric [27]. We then select the 10 atlas images that are most similar for use in the JLF scheme. The resulting performance numbers (in terms of Dice overlap) are similar to what we obtained previously and are given in Figure 5 along with the Dice overlap numbers from the CNN-based approach. Accuracy for the latter was left lung: 0.93 ± 0.03 , right lung: 0.94 ± 0.02 , and whole lung: 0.94 ± 0.02 . The analogous JLF numbers were slightly more accurate (left lung: 0.95 ± 0.02 , right lung: 0.96 ± 0.01 , whole lung: 0.96 ± 0.01) although the processing time is significantly greater—less than 1 second per subject for the proposed approach versus ~ 25 minutes per subject using JLF using 4 CPU threads running 8 parallel pairwise registrations per evaluation image.

Ventilation MRI lung segmentation

We applied the deep learning approach described in the previous section to the evaluation data used in [8]. The resulting probability iamges were converted into a single segmentation image which were then compared with the manual segmentation results and Atropos results from our previous

work [8]. Note that the Otsu thresholding and K-means thresholding were omitted as they were the poorest performers and, as mentioned previously, discard spatial information in contrast to both computational methods and the human readers.

In the absence of ground truth, the STAPLE algorithm [45] was used to create a consensus labeling. The Dice overlap coefficient was used to quantify agreement between each of the segmentation raters and the consensus labeling as an indicator performance. The results are shown in Figure 6. Mean values (\pm standard deviation) were as follows (total, normal lung, ventilation defect):

Reader 1: 0.89 ± 0.07 , 0.91 ± 0.06 , 0.6 ± 0.3 ; Reader 2: 0.92 ± 0.05 , 0.94 ± 0.04 , 0.57 ± 0.3 ; Reader 3: 0.94 ± 0.03 , 0.96 ± 0.03 , 0.63 ± 0.3 ; Atropos: 0.92 ± 0.03 , 0.94 ± 0.03 , 0.71 ± 0.3 ; and U-net: 0.94 ± 0.03 , 0.96 ± 0.03 , 0.70 ± 0.3 . Computational time for processing was slightly less than a minute per subject for Atropos, between 30–45 for the human readers, and less than a second for the U-net model.

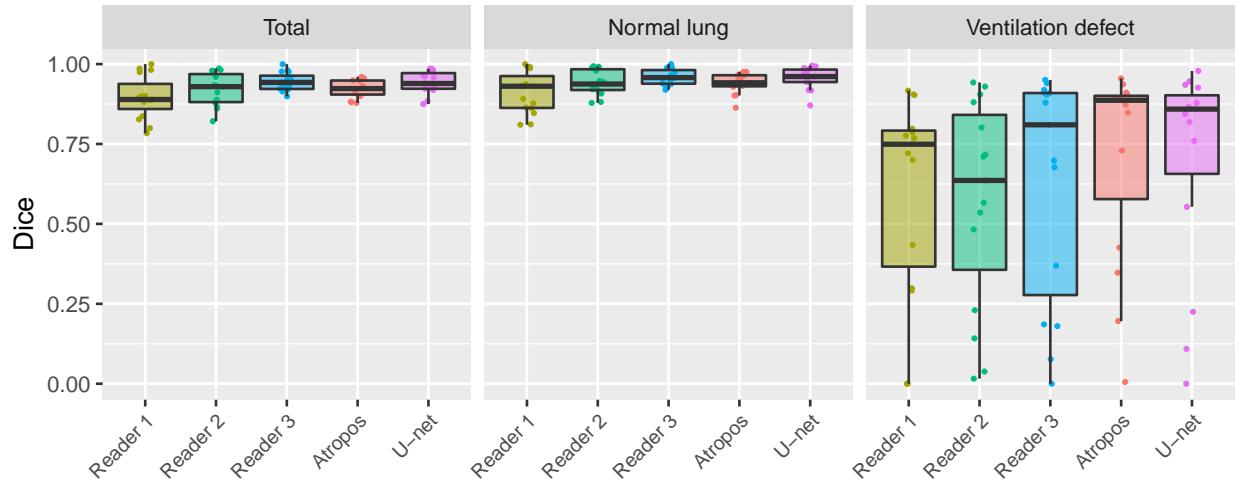


Figure 6: The Dice overlap coefficient for total, normal lung, and ventilation defect regions for segmentation of the evaluation data set.

DISCUSSION

Significant progress has been made from earlier quantification approaches in which human labelers manually identified areas of poor ventilation or applied simple thresholding techniques. More

sophisticated automated and semi-automated techniques have advanced our ability to investigate the use of hyperpolarized gas imaging as quantitative image-based biomarkers. Deep learning techniques can further enhance these methodologies by potentially increasing accuracy, generalizability, and computational efficiency. In this work, we provided a deep learning framework for segmentation of structural and functional lung MRI for quantification of ventilation. This framework is based on the U-net architecture and implemented using the Keras API available through the R statistical project.

There are several limitations to the proposed framework. The most obvious is that it only leverages the full 3-D nature of the image data collected for the proton segmentation. The trained models for ventilation image segmentation were based on 2-D coronal slices and therefore subsequent prediction is limited to those views. Even though good results were achieved in this study, even better results might be achieved by training 3-D models for the latter. Also, evaluative comparison was made using manually-refined segmentations which is certainly useful but additional evaluations using various clinical measures would also be helpful in determining the relative utility of various segmentation approaches. For example, how does the performance of the various methods translate into utility as an imaging biomarker for lung function?

Despite these limitations of the proposed framework, there are also limitations of previously reported methods. For example, in addition to the significant time requirements for JLF of lung images, shown in Figure 7 is an example where difficult pairwise image registration scenarios can cause algorithmic failures. In contrast, the trained U-net model is capable of learning features which can potentially circumvent registration failures. Similarly, the online feature capabilities of deep learning can overcome some of the drawbacks to more conventional segmentation approaches of ventilation lung images. A well-known artifact for these approaches is partial voluming effects which can confound certain intensity-based segmentation approaches (see Figure 8).

Future research will certainly look into these issues as potential improvements to the existing framework. As a surrogate for full 3-D models, we are looking into developing additional 2-D U-net models for the axial and sagittal views. Since slice-by-slice processing is computationally efficient in the deep learning paradigm, we can process 3-D images along the three canonical axes and combined the results for increased accuracy. More broadly, it would be of potential interest to investigate the use of

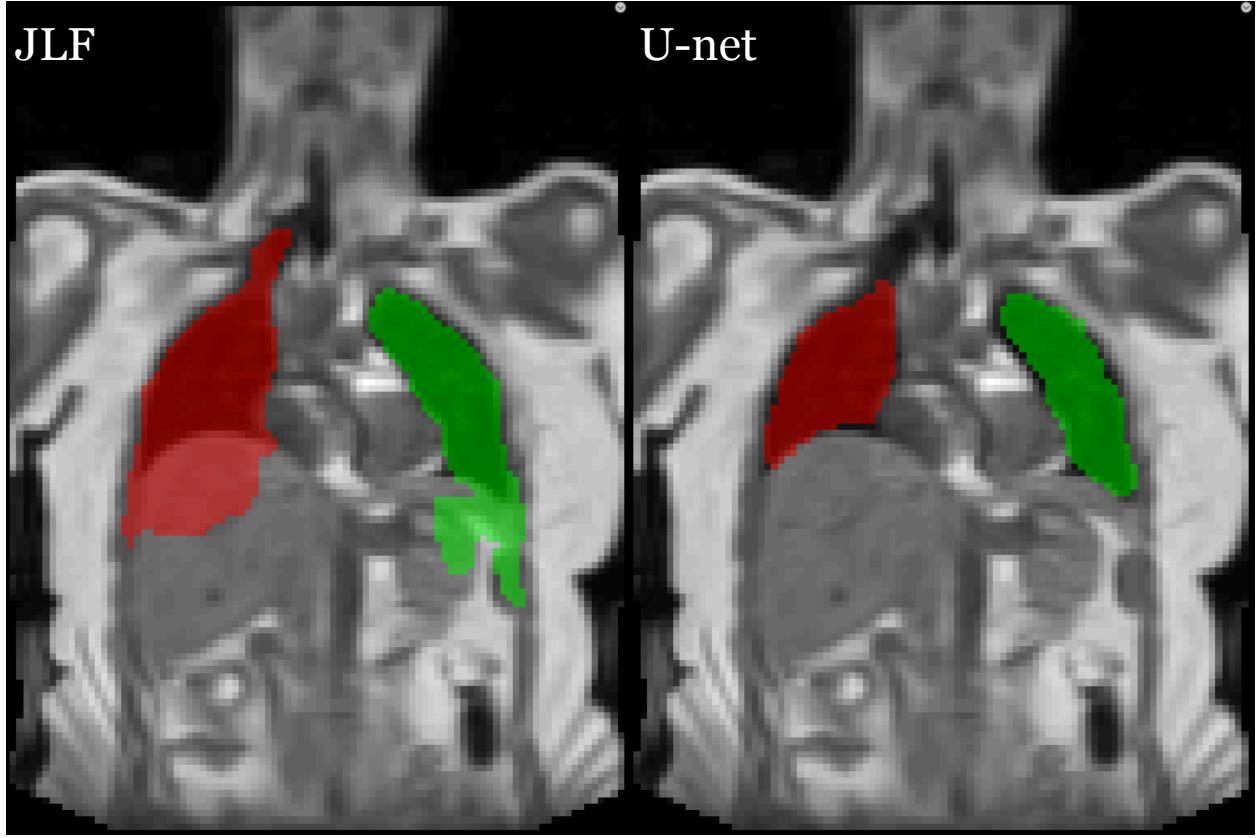


Figure 7: Problematic case showing potential issues with the JLF approach (left) for proton lung segmentation where a difficult pairwise image registration caused segmentation failure. In contrast, by learning features directly, the U-net approach (right) avoids possible registration difficulties.

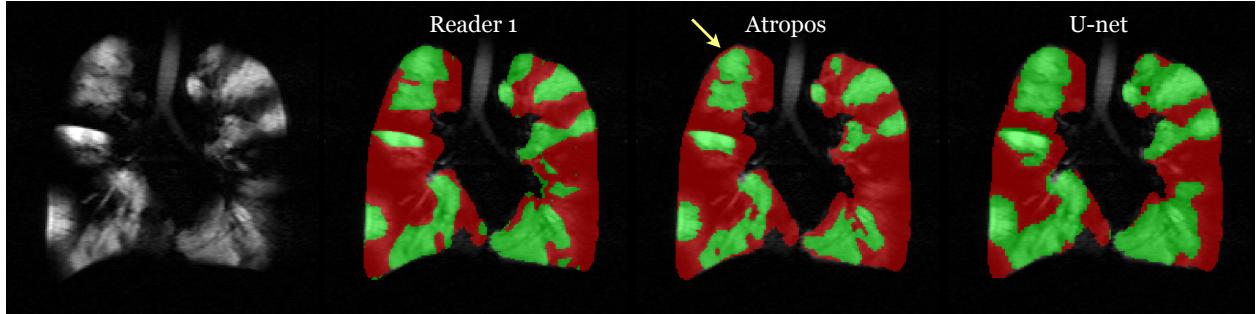


Figure 8: Ventilation segmentation comparison between a human reader and the two computational approaches. Notice the effects of the partial voluming at the apex of the lungs, indicated by the yellow arrow, which are labeled as ventilation defect by the Atropos approach whereas U-net and the human reader correctly label this region.

image classification techniques (e.g., VGG [17]) for classifying lung disease phenotype directly from the images. More immediate benefits could result from augmenting the limited, single-site data set used in this work to include data contributed from other groups which could translate into more ro-

bust models. Additionally, as the U-net architecture is application-agnostic, investigators can apply the contributions discussed in this work to their own data, such as lung CT.

ACKNOWLEDGMENTS

Research reported in this manuscript was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number R01HL133889, R01HL109618, R44 HL 087550, and R21HL129112.

We also gratefully acknowledge the support of the NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

REFERENCES

1. Liu, Z., Araki, T., Okajima, Y., Albert, M., and Hatabu, H. “**Pulmonary Hyperpolarized Noble Gas MRI: Recent Advances and Perspectives in Clinical Application**” *Eur J Radiol* 83, no. 7 (2014): 1282–1291. doi:10.1016/j.ejrad.2014.04.014
2. Roos, J. E., McAdams, H. P., Kaushik, S. S., and Driehuys, B. “**Hyperpolarized Gas MR Imaging: Technique and Applications**” *Magn Reson Imaging Clin N Am* 23, no. 2 (2015): 217–29. doi:10.1016/j.mric.2015.01.003
3. Adamson, E. B., Ludwig, K. D., Mumy, D. G., and Fain, S. B. “**Magnetic Resonance Imaging with Hyperpolarized Agents: Methods and Applications**” *Phys Med Biol* 62, no. 13 (2017): R81–R123. doi:10.1088/1361-6560/aa6be8
4. Svenningsen, S., Kirby, M., Starr, D., Coxson, H. O., Paterson, N. A. M., McCormack, D. G., and Parraga, G. “**What Are Ventilation Defects in Asthma?**” *Thorax* 69, no. 1 (2014): 63–71. doi:10.1136/thoraxjnl-2013-203711
5. Tustison, N. J., Altes, T. A., Song, G., Lange, E. E. de, Mugler, J. P., 3rd, and Gee, J. C. “**Feature Analysis of Hyperpolarized Helium-3 Pulmonary MRI: A Study of Asthmatics Versus Nonasthmatics**” *Magn Reson Med* 63, no. 6 (2010): 1448–55. doi:10.1002/mrm.22390
6. Kirby, M., Pike, D., Coxson, H. O., McCormack, D. G., and Parraga, G. “**Hyperpolarized (3)He Ventilation Defects Used to Predict Pulmonary Exacerbations in Mild to Moderate Chronic Obstructive Pulmonary Disease**” *Radiology* 273, no. 3 (2014): 887–96. doi:10.1148/radiol.14140161
7. Altes, T. A., Mugler, J. P., 3rd, Ruppert, K., Tustison, N. J., Gersbach, J., Szentpetery, S., Meyer, C. H., Lange, E. E. de, and Teague, W. G. “**Clinical Correlates of Lung Ventilation Defects in Asthmatic Children**” *J Allergy Clin Immunol* 137, no. 3 (2016): 789–96.e7. doi:10.1016/j.jaci.2015.08.045
8. Tustison, N. J., Avants, B. B., Flors, L., Altes, T. A., Lange, E. E. de, Mugler, J. P., 3rd, and Gee, J. C. “**Ventilation-Based Segmentation of the Lungs Using Hyperpolarized (3)He MRI**” *J Magn Reson Imaging* 34, no. 4 (2011): 831–41. doi:10.1002/jmri.22738

9. Kirby, M., Heydarian, M., Svenningsen, S., Wheatley, A., McCormack, D. G., Etemad-Rezai, R., and Parraga, G. “**Hyperpolarized ^3He Magnetic Resonance Functional Imaging Semiautomated Segmentation**” *Acad Radiol* 19, no. 2 (2012): 141–52. doi:10.1016/j.acra.2011.10.007
10. He, M., Kaushik, S. S., Robertson, S. H., Freeman, M. S., Virgincar, R. S., McAdams, H. P., and Driehuys, B. “**Extending Semiautomatic Ventilation Defect Analysis for Hyperpolarized $(^{129}\text{Xe}$ Ventilation MRI**” *Acad Radiol* 21, no. 12 (2014): 1530–41. doi:10.1016/j.acra.2014.07.017
11. Zha, W., Niles, D. J., Kruger, S. J., Dardzinski, B. J., Cadman, R. V., Mummy, D. G., Nagle, S. K., and Fain, S. B. “**Semiautomated Ventilation Defect Quantification in Exercise-Induced Bronchoconstriction Using Hyperpolarized Helium-3 Magnetic Resonance Imaging: A Repeatability Study**” *Acad Radiol* 23, no. 9 (2016): 1104–14. doi:10.1016/j.acra.2016.04.005
12. Hughes, P. J. C., Horn, F. C., Collier, G. J., Biancardi, A., Marshall, H., and Wild, J. M. “**Spatial Fuzzy c-Means Thresholding for Semiautomated Calculation of Percentage Lung Ventilated Volume from Hyperpolarized Gas and ^1H MRI**” *J Magn Reson Imaging* 47, no. 3 (2018): 640–646. doi:10.1002/jmri.25804
13. Trivedi, A., Hall, C., Hoffman, E. A., Woods, J. C., Gierada, D. S., and Castro, M. “**Using Imaging as a Biomarker for Asthma**” *J Allergy Clin Immunol* 139, no. 1 (2017): 1–10. doi:10.1016/j.jaci.2016.11.009
14. LeCun, Y., Bengio, Y., and Hinton, G. “**Deep Learning**” *Nature* 521, (2015): 436–444.
15. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. “**ImageNet Large Scale Visual Recognition Challenge**” *International Journal of Computer Vision* 115, no. 3 (2015): 211–252.
16. Krizhevsky, A., Sutskever, I., and Hinton, G. E. “**ImageNet Classification with Deep Convolutional Neural Networks**” *Commun. ACM* 60, no. 6 (2017): 84–90. doi:10.1145/3065386, Available at <http://doi.acm.org/10.1145/3065386>
17. Simonyan, K. and Zisserman, A. “**Very Deep Convolutional Networks for Large-Scale Image Recognition**” *CoRR* abs/1409.1556, (2014): Available at <http://arxiv.org/abs/1409.1556>

18. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. “**Rethinking the Inception Architecture for Computer Vision**” *CoRR* abs/1512.00567, (2015): Available at <http://arxiv.org/abs/1512.00567>
19. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. “**Gradient-Based Learning Applied to Document Recognition**” *Proceedings of the IEEE* 86, no. 11 (1998): 2278–2324.
20. Fukushima, K. “**Neocognitron: A Self Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position**” *Biol Cybern* 36, no. 4 (1980): 193–202.
21. Hubel, D. H. and Wiesel, T. N. “**Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex**” *J Physiol* 160, (1962): 106–54.
22. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Laak, J. A. W. M. van der, Ginneken, B. van, and Sánchez, C. I. “**A Survey on Deep Learning in Medical Image Analysis**” *Med Image Anal* 42, (2017): 60–88. doi:10.1016/j.media.2017.07.005
23. Ronneberger, O., Fischer, P., and Brox, T. “**U-Net: Convolutional Networks for Biomedical Image Segmentation**” *Proceedings of the international conference on medical image computing and computer-assisted intervention* 9351, (2015): 234–241.
24. Tustison, N. J. and Herrera, J. M. “**Two Luis Miguel Fans Walk into a Bar in Nagoya — > (Yada, Yada, Yada) —> an ITK-Implementation of a Popular Patch-Based Denoising Filter**” *Insight Journal* (2016):
25. Taylor, L. and Nitschke, G. “**Improving Deep Learning Using Generic Data Augmentation**” *CoRR* abs/1708.06020, (2017): Available at <http://arxiv.org/abs/1708.06020>
26. Tustison, N. J. and Avants, B. B. “**Explicit B-Spline Regularization in Diffeomorphic Image Registration**” *Front Neuroinform* 7, (2013): 39. doi:10.3389/fninf.2013.00039
27. Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. “**A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration**” *Neuroimage* 54, no. 3 (2011): 2033–44. doi:10.1016/j.neuroimage.2010.09.025
28. Available at <https://github.com/stnavar/ANTsR>

29. Available at <https://github.com/ntustison/DeepVentNet>
30. Altes, T. A., Johnson, M., Fidler, M., Botfield, M., Tustison, N. J., Leiva-Salinas, C., Lange, E. E. de, Froh, D., and Mugler, J. P., 3rd. “**Use of Hyperpolarized Helium-3 MRI to Assess Response to Ivacaftor Treatment in Patients with Cystic Fibrosis**” *J Cyst Fibros* 16, no. 2 (2017): 267–274. doi:10.1016/j.jcf.2016.12.004
31. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. “**N4ITK: Improved N3 Bias Correction**” *IEEE Trans Med Imaging* 29, no. 6 (2010): 1310–20. doi:10.1109/TMI.2010.2046908
32. Qing, K., Altes, T. A., Tustison, N. J., Feng, X., Chen, X., Mata, J. F., Miller, G. W., Lange, E. E. de, Tobias, W. A., Cates, G. D., Jr, Brookeman, J. R., and Mugler, J. P., 3rd. “**Rapid Acquisition of Helium-3 and Proton Three-Dimensional Image Sets of the Human Lung in a Single Breath-Hold Using Compressed Sensing**” *Magn Reson Med* 74, no. 4 (2015): 1110–5. doi:10.1002/mrm.25499
33. Wang, H., Suh, J. W., Das, S. R., Pluta, J., Craige, C., and Yushkevich, P. A. “**Multi-Atlas Segmentation with Join Label Fusion**” *IEEE Trans Pattern Analysis and Machine Intelligence* (2013):
34. Available at <https://github.com/ntustison/LungAndLobeEstimationExample>
35. Available at <https://github.com/ntustison/LungVentilationSegmentationExample>
36. Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. “**N4ITK: Improved N3 Bias Correction**” *IEEE Trans Med Imaging* 29, no. 6 (2010): 1310–20. doi:10.1109/TMI.2010.2046908
37. Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L., and Robles, M. “**Adaptive Non-Local Means Denoising of MR Images with Spatially Varying Noise Levels**” *J Magn Reson Imaging* 31, no. 1 (2010): 192–203. doi:10.1002/jmri.22003
38. Shelhamer, E., Long, J., and Darrell, T. “**Fully Convolutional Networks for Semantic Segmentation**” *IEEE Trans Pattern Anal Mach Intell* 39, no. 4 (2017): 640–651. doi:10.1109/TPAMI.2016.2572683

39. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. “**Dropout: A Simple Way to Prevent Neural Networks from Overfittin**” *Journal of Machine Learning Research* 15, no. 1 (2014): 1929–1958.
40. Available at <https://github.com/ANTsX/ANTsRNet>
41. Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., and Gee, J. C. “**The Optimal Template Effect in Hippocampus Studies of Diseased Populations**” *Neuroimage* 49, no. 3 (2010): 2457–66. doi:10.1016/j.neuroimage.2009.09.062
42. Yang, X., Kwitt, R., Styner, M., and Niethammer, M. “**Quicksilver: Fast Predictive Image Registration - a Deep Learning Approach**” *Neuroimage* 158, (2017): 378–396. doi:10.1016/j.neuroimage.2017.07.008
43. Nyúl, L. G., Udupa, J. K., and Zhang, X. “**New Variants of a Method of MRI Scale Standardization**” *IEEE Trans Med Imaging* 19, no. 2 (2000): 143–50. doi:10.1109/42.836373
44. Nair, V. and Hinton, G. E. “**Rectified Linear Units Improve Restricted Boltzmann Machines**” *Proceedings of the 27th international conference on machine learning* (2010):
45. Warfield, S. K., Zou, K. H., and Wells, W. M. “**Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation**” *IEEE Trans Med Imaging* 23, no. 7 (2004): 903–21. doi:10.1109/TMI.2004.828354
46. Milletari, F., Navab, N., and Ahmadi, S. “**V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation**” *CoRR* abs/1606.04797, (2016): Available at <http://arxiv.org/abs/1606.04797>
47. Krizhevsky, A., Sutskever, I., and Hinton, G. E. “**ImageNet Classification with Deep Convolutional Neural Networks**” *Proceedings of the 25th international conference on neural information processing systems - volume 1* (2012): 1097–1105. Available at <http://dl.acm.org/citation.cfm?id=2999134.2999257>
48. He, K., Zhang, X., Ren, S., and Sun, J. “**Deep Residual Learning for Image Recognition**” *CoRR* abs/1512.03385, (2015): Available at <http://arxiv.org/abs/1512.03385>
49. Xie, S., Girshick, R. B., Dollár, P., Tu, Z., and He, K. “**Aggregated Residual Transforma-**

tions for Deep Neural Networks” *CoRR* abs/1611.05431, (2016): Available at <http://arxiv.org/abs/1611.05431>

50. Huang, G., Liu, Z., and Weinberger, K. Q. “**Densely Connected Convolutional Networks**” *CoRR* abs/1608.06993, (2016): Available at <http://arxiv.org/abs/1608.06993>

51. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. E., Fu, C., and Berg, A. C. “**SSD: Single Shot Multibox Detector**” *CoRR* abs/1512.02325, (2015): Available at <http://arxiv.org/abs/1512.02325>