Holistic segmentation of the lung in cine MRI

William Kovacs
Nathan Hsieh
Holger Roth
Chioma Nnamdi-Emeratom
W. Patricia Bandettini
Andrew Arai
Ami Mankodi
Ronald M. Summers
Jianhua Yao

Holistic segmentation of the lung in cine MRI

William Kovacs,^a Nathan Hsieh,^a Holger Roth,^a Chioma Nnamdi-Emeratom,^b W. Patricia Bandettini,^c Andrew Arai,^c Ami Mankodi,^b Ronald M. Summers,^a and Jianhua Yao^{a,*}

^aNational Institutes of Health, Radiology and Imaging Sciences, Clinical Center, Clinical Image Processing Services, Bethesda, Maryland, United States

^bNational Institutes of Health, National Institute of Neurological Disorders and Stroke, Neurogenetics Branch, Bethesda, Maryland, United States

^cNational Institutes of Health, National Heart, Lung and Blood Institute, Advanced Cardiovascular Imaging, Bethesda, Maryland, United States

Abstract. Duchenne muscular dystrophy (DMD) is a childhood-onset neuromuscular disease that results in the degeneration of muscle, starting in the extremities, before progressing to more vital areas, such as the lungs. Respiratory failure and pneumonia due to respiratory muscle weakness lead to hospitalization and early mortality. However, tracking the disease in this region can be difficult, as current methods are based on breathing tests and are incapable of distinguishing between muscle involvements. Cine MRI scans give insight into respiratory muscle movements, but the images suffer due to low spatial resolution and poor signal-to-noise ratio. Thus, a robust lung segmentation method is required for accurate analysis of the lung and respiratory muscle movement. We deployed a deep learning approach that utilizes sequence-specific prior information to assist the segmentation of lung in cine MRI. More specifically, we adopt a holistically nested network to conduct image-to-image holistic training and prediction. One frame of the cine MRI is used in the training and applied to the remainder of the sequence (>200 frames). We applied this method to cine MRIs of the lung in the axial, sagittal, and coronal planes. Characteristic lung motion patterns during the breathing cycle were then derived from the segmentations and used for diagnosis. Our data set consisted of 31 young boys, age 9.6 ± 2.3 years, 15 of whom suffered from DMD. The remaining 16 subjects were age-matched healthy volunteers. For validation, slices from inspiratory and expiratory cycles were manually segmented and compared with results obtained from our method. The Dice similarity coefficient for the deep learning-based method was 97.2 ± 1.3 for the sagittal view, 96.1 ± 3.8 for the axial view, and 96.6 ± 1.7 for the coronal view. The holistic neural network approach was compared with an approach using Demon's registration and showed superior performance. These results suggest that the deep learning-based method reliably and accurately segments the lung across the breathing cycle. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.4.4.041310](https://doi.org/10.1117/1.JMI.4.4.041310)]

Keywords: muscular dystrophy; cine MRI; lung segmentation; deep learning.

Paper 17134SSRR received May 9, 2017; accepted for publication Nov. 6, 2017; published online Nov. 30, 2017.

1 Introduction

Duchenne muscular dystrophy (DMD) is the most common childhood-onset muscular dystrophy. It results in the degeneration of muscle, beginning with those of the extremities and eventually progressing toward more vital systems, such as those involved with pulmonary function. Current methods to track respiratory function in these patients involve spirometry tests to obtain values, such as forced vital capacity. However, such measures give only a general sense of respiratory muscle failure and do not distinguish between the different muscle regions, such as the diaphragm versus intercostal muscles.

One way to distinguish better between these muscle regions during breathing is to have the patient undergo imaging. MRI is the preferred imaging modality because DMD patients tend to be children or young adults, who are thought to be more susceptible to the effects of CTs ionizing radiation. This is especially important when visualization of the entire breathing cycle is desired because it requires repeated imaging of the chest area, further increasing the risks of CT radiation. Meanwhile, cine MRI can generate images of the lung as the patient is breathing, allowing for visualization of the entire breathing cycle without the risk of ionizing radiation. Cine MRI has been shown to have

potential in evaluating the pediatric airways¹ and other lung diseases.^{2–6} By imaging the breathing cycle of a patient, measurements that are more detailed than common spirometry measurements can be used to characterize the disease state. However, to obtain the temporal resolution required for fine analysis of muscle involvement, the spatial resolution suffers, resulting in poor quality images and making the image analysis more challenging. Figure 1 shows the examples of lung cine MRI in axial, sagittal, and coronal planes obtained during a deep breathing cycle in an 11-year-old patient from our study.

To extract the measurements from a cine MRI sequence, some form of segmentation needs to be performed. Manual segmentations are tedious, especially if the full breathing cycle of 100 to 400 images is to be analyzed. Furthermore, artifacts and noise in some series can be severe. This makes not only the manual segmentation even more difficult to perform, but also makes common automatic methods, such as active contour algorithms, produce poor segmentations. While segmentation of lung MRI images has not garnered much attention in the past, there has been increasing interest in utilizing cine MRI to analyze certain diseases, including Pompe's disease⁵ and chronic obstructive pulmonary disorder.⁷ In some of these papers, the authors resort to manual segmentation to make their measurements.

*Address all correspondence to: Jianhua Yao, E-mail: jyao@cc.nih.gov

In other papers, authors try other methods so that such tedious manual measures are not necessary. Some prefer a semiautomated method,⁸ using toolkits such as medical imaging interaction toolkit, as well as plug-ins for the CHILI radiology system.^{9,10} Böttger et al.¹¹ also developed another semiautomated method that utilizes interactive simplex meshes.

Automated methods to achieve the same ends have become increasingly popular, as they allow for even further reduction in human effort. Middleton and Damper¹² first attempted to use a neural network, the multilayer perceptron, to obtain the segmentations. However, when this method alone resulted in poor results on unseen slices, they added another step that used snakes to refine the network results.¹³ Ray et al.¹⁴ also used snake algorithms to identify the lung cavity, where they ran multiple gradient vector flow snakes independently of each other within the lung and took the final result to be the union of the regions covered by the snakes. Because not all lungs are perfectly healthy, Sensakovic et al.¹⁵ designed a method that would be able to generate lung contours in MR images of unhealthy lungs or images corrupted by artifacts. To this end, they utilized standard image processing methods, including thresholding, erosion, rolling ball, and shape descriptors to some success, especially when considering the quality of their images.

Over time, the snake algorithms appeared to be a popular choice to accomplish this task, with Osareh and Shadgar¹⁶ using region-aided geometric snakes, and Yang et al.¹⁷ using the snake algorithm provided ITK-SNAP. Even when there was an approach via the Hough transform,¹⁸ Tavares et al.¹⁹ realized the necessity of snakes in their follow-up. However, not all recent methods have relied on this versatile algorithm. When Kohlmann et al.⁷ required an automatic segmentation system for these images before registering them to perfusion images, they relied on more standard segmentation methods, such as region growing and closing operations with some postprocessing, to ensure that the lungs were separated. Also, Guo et al.²⁰ developed a Potts model using the volume proportion of right to left lungs as a *prior* constraint via the use of a max-flow model.

While automated methods do provide a time-saving alternative to other procedures, semiautomated methods provide similar advantages, albeit to a lesser extent, while ensuring that the quality of the final contours is higher. In a recent paper, Mogalle et al.² relied on a registration-based approach to segment dynamic 3-D MRI. From a static MRI scan, they manually obtained lung masks before registering them to the corresponding slices in the dynamic scan. From there, they were able to register these slices with the remaining images to ultimately obtain measurements that correlated well with various spirometry values.

We present a method that is capable of handling segmentation of the lungs in poor quality cine MR images. The method uses deep learning, which is more robust than previous automatic methods and much quicker than performing segmentations by hand. While deep learning has been introduced into other medical imaging contexts, we proved its usefulness for lung segmentation in cine MRI by comparing it to a registration-based method commonly used for CT image segmentation.^{21,22} A new, deep learning architecture known as a holistically nested network (HNN), initially introduced as holistically nested edge detection, was demonstrated to provide efficient and accurate results for the identification of edges.²³ This network aimed to predict edges on an image-to-image basis and to utilize information from the images in a multilevel fashion. More details of this network are presented in Sec. 2. Recently,

this method was shown not only to be effective at identifying edges but also at addressing broader medical imaging problems.²⁴ We adopted HNN for an efficient and powerful method for lung segmentation in cine MRI. The high temporal resolution of cine MRI allows us to image dynamic breathing cycles, but it comes with the cost of sacrificing spatial resolution and makes the quantitative analysis very challenging. Thus, our application of HNN to lung cine MRI segmentation is a fresh approach to solving a very difficult issue. Our proposed framework not only tracks the respiratory lung movements, but also allows us to derive characteristic lung functions, such as diaphragm movement and chest wall movement area (CWMA). Understanding how the diaphragm and CWMAs are affected by disease informs clinicians on how to intervene for the patient. In medical imaging, the availability of large numbers of data sets is limited so we use a slightly modified approach that trains the model using one slice from each cine MRI sequence for the segmentation of the entire sequence. Our main contributions have two parts. First, we propose a method to build deep models using both population data and patient-specific data. Second, we derive characteristic lung functions such as chest wall movement and diaphragm movement from the cine MRI segmentations.

2 Methods

2.1 Data Acquisition

Our study was compliant with the Health Insurance Portability and Accountability Act and was conducted with an Institutional Review Board approval. Our patient population included 31 young boys (age 5 to 14, average 9.6 ± 2.3 years). About 15 patients had DMD, while the remaining 16 boys were age-matched healthy volunteers. 2-D cine MRI scans of each boy's chest were obtained using one of two 1.5T MR scanners by Siemens: the MAGNETOM Aera (16 patients) or the MAGNETOM Avanto (15 patients). This type of scan involves generating images along a single plane with roughly 0.065 s between images. For each subject, the scanning was performed along two axial planes (upper and lower cross sections), two sagittal planes (left and right lung cross sections along the midclavicular line), and a coronal plane (Fig. 1). To reduce the impact of the heart on the axial scans, we utilized only the upper axial plane scans. For the sagittal scans, we analyzed the right lung scans to ensure that the heart would not interfere with the segmentations. The patients were scanned at two different breathing patterns: normal and deep. During normal breathing, the scan was taken during the patient's tidal breathing, while during deep breathing, the scan was taken as the patient was encouraged to take deep breaths to measure forced vital capacity. The normal breathing scans tended to include more breathing cycles than the corresponding deep breathing scans. During normal breathing, axial series had an average of 298 ± 81 images, coronal series had 269 ± 70 images, and sagittal series had 283 ± 78 images. During deep breathing, axial series had 168 ± 35 images, coronal series had 154 ± 54 images, and sagittal series had 159 ± 32 images. For these scans, a T1-weighted gradient echo sequence was used. Detailed information about the scanner settings is included in Table 1.

2.2 Registration-Based Approach

For the registration-based method, we required a slice from the cine MRI to be manually segmented first and used as an atlas,

Fig. 1 Examples of lung cine MRI. Two images each for (first column) axial, (second column) coronal, and (third column) sagittal planes. First row at the minimum lung volume (expiration), and second row at the maximum lung volume (inspiration). The images were acquired during the patient's deep breathing cycle.

Table 1 Cine MRI scanning parameters.

	Axial plane	Coronal plane	Sagittal plane
Thickness (mm)	6.39 ± 0.8 , 6 to 8	6.45 ± 0.9 , 6 to 8	6.39 ± 0.8 , 6 to 8
Pixel size (mm)	2.0 ± 0.2 , 1.6 to 2.3	2.2 ± 0.3 , 1.8 to 2.8	2.0 ± 0.2 , 1.6 to 2.3
Flip angle	49.7 ± 1.2 , 45 to 50	49.7 ± 1.2 , 45 to 50	49.7 ± 1.2 , 45 to 50
TR	64.3 ± 12.8 , 51.7 to 107.6	73.6 ± 14.8 , 52.2 to 125.4	68.3 ± 13.5 , 50.6 to 107.6
TE	1.2 ± 0.2 , 0.99 to 1.48	1.2 ± 0.2 , 0.97 to 1.43	1.2 ± 0.2 , 0.99 to 1.48
SAR	1.7 ± 0.2 , 1.3 to 2.3	1.7 ± 0.3 , 1.3 to 2.5	1.7 ± 0.2 , 1.3 to 2.3
Number of images in normal breathing	168 ± 35 , 115 to 310	154 ± 54 , 79 to 383	159 ± 32 , 92 to 235
Number of images in deep breathing	298 ± 81 , 117 to 387	269 ± 70 , 117 to 383	283 ± 78 , 117 to 395

regardless of location. We then applied Demon's registration algorithm (ITK)²⁵ based on mutual information from this slice to every other slice. We adopted a multiresolution approach to better capture the large movements that are possible during deep breathing. The deformation field obtained from the registration was applied to the segmentation mask of the atlas. 3-D Gaussian smoothing is applied to the resulting masks to induce a smooth transition between adjacent slices.

Sometimes drastic differences between new data and the atlas caused either over-segmentation or under-segmentation along the diaphragm. A postprocessing step was conducted to correct the missegmentation. First, we fit a piecewise third-degree Bernstein spline using the location of the diaphragm in the current segmentation and then removed any points from the diaphragm that were too far away from the estimation

(beyond two standard deviations from the mean). Based on the remaining diaphragm points, another Bernstein spline estimation was applied. Once again, the points that were too far from the estimation were removed. This process was repeated until no additional points were removed. The new diaphragm was constructed using all of the remaining original diaphragm points as well the points estimated from the final spline.

2.3 Deep Learning-Based Approach

Despite demonstrating moderate success, registration-based methods still contained weaknesses. They tended to over-smooth the segmentation and had trouble accurately segmenting the extremities of the lung even after postprocessing. They were also sensitive to the selection of the reference atlas. If the

difference between the target image and the reference image (e.g., inspiration and expiration) was too big, the registration tended to fail. Thus, we turned to a deep learning-based approach for more accurate segmentations that required no more human labor than the registration-based approach used previously. Holistically nested Net, HNN, proposed by Xie and Tu,²³ is a deep learning model that leverages fully convolutional networks and deeply supervised nets. This type of neural network was initially used for edge detection, but has been shown to be effective at organ segmentation as well.²⁴

To apply HNN to our lung cine MRI sequences, we first created one training set for each scanning plane (axial, sagittal, and coronal). Each training set consisted of one manually segmented slice per sequence in the respective plane. Therefore, there were three separate training sets each containing 62 slices from 62 cine MRI series. The remainder of the images that were not used for the training sets were then grouped together to be testing sets. Once the training and testing sets were completed, all of the images in both sets were converted to 8-bit portable network graphics images to be made compatible with the HNN model. To do so, we used histogram equalization to reduce the 16-bit DICOM images to 8-bit while maintaining good global contrast in the images. The training images were then fed into the HNN architecture to train a deep learning model.

The HNN framework is shown in Fig. 2. The major difference between HNN and a standard convolutional neural network is that in addition to the standard layers, HNN has several side output layers, each of which is a fully convolutional layer generating intermediate segmentation results at each stage of the neural network. The side outputs are ultimately fused together for the final output. If there are M side-output layers, we can define their corresponding weights as $[w^{(1)}, \dots, w^{(M)}]$, and for simplicity, we can define the standard network layer as W . These can be used to formulate the objective function as

$$L_{\text{side}}(W, w) = \sum_{m=1}^M \alpha_m \ell_{\text{side}}^{(m)}[W, w^{(m)}], \quad (1)$$

where ℓ_{side} is the image-level loss function for side outputs, which is computed over all pixels in the training image with its corresponding ground truth mask. α_m is the loss weight for side output layer m . Equal weights are currently used.

HNN also provides a strategy to balance the positive and negative samples: the introduction of a class-balancing weight β , which refers to the proportion of negative sample pixels. The loss functions from negative and positive samples were adjusted inverse-proportionally, i.e., the fewer the samples, the higher the weight. Originally, Xie et al. calculated β for each 2-D image independently. However, in our method, this weight was calculated across the entire training set, which was shown to be helpful with another medical imaging problem when there are no positive samples on certain slices.²⁴

To use the side-output layer information, a “weighted-fusion” layer is introduced, which can be learned during training. The purpose of the fusion layer is to combine the side outputs into the final fused output, and its loss function is

$$L_{\text{fuse}}(W, w, h) = \text{Dist}(Y, \hat{Y}_{\text{fuse}}), \quad (2)$$

where $\hat{Y}_{\text{fuse}}$ is the fused prediction, Y is the ground truth, and h is the fusion weight. $\text{Dist}(Y, \hat{Y}_{\text{fuse}})$ refers to the distance between the fused predictions and the ground truth, and here, this is the cross-entropy loss. Finally, we can minimize this objective function using back propagation and gradient descent. For testing, we can input an image X and retrieve the lung segmentation prediction from the side-output layers as well as the fusion layer

$$[\hat{Y}_{\text{fuse}}, \hat{Y}_{\text{side}}^{(1)}, \dots, \hat{Y}_{\text{side}}^{(M)}] = \text{CNN}[X, (W, w, h)^*]. \quad (3)$$

The lung prediction can then be generated for each side output layer by aggregating the sigmoid of its activation at each point.

The initial network structure is based on a VGGNet model²⁶ pretrained on ImageNet²⁷ before being fine-tuned²⁸ on our data set with a learning rate of 10^{-6} . The VGGNet offers a deep architecture in five stages (a stage consisting of some number of convolutional layers and a max pooling layer), with the final

Fig. 2 HNN framework. Image-to-image training and prediction mechanism with side outputs and fused output.

Table 2 HNN network layers.

Layer	Conv1_1	Conv1_2	Conv2_1	Conv2_2	Conv3_1	Conv3_2	Conv3_3
Learning/decay rate multipliers #1	10/1	10/1	10/1	10/1	10/1	10/1	10/1
Learning/decay rate multipliers #2	20/0	20/0	20/0	20/0	20/0	20/0	20/0
Layer	Conv4_1	Conv4_2	Conv4_3	Conv5_1	Conv5_2	Conv5_3	
Learning/decay rate multipliers #1	10/1	10/1	10/1	100/1	100/1	100/1	
Learning/decay rate multipliers #2	20/0	20/0	20/0	200/0	200/0	200/0	

Table 3 Side-output layers.

Layer	DSN conv 1	DSN conv 2	DSN conv 3	DSN conv 4	DSN conv 5
Linked to	Conv1_2	Conv2_2	Conv3_3	Conv4_3	Conv5_3
Convolution learning/decay rate multiplier #1	0.01/1	0.01/1	0.01/1	0.01/1	0.01/1
Convolution learning/decay rate multiplier #2	0.02/0	0.02/0	0.02/0	0.02/0	0.02/0
Deconvolution learning/decay rate multiplier #1	N/A	0/1	0/1	0/1	0/1
Deconvolution learning/decay rate multiplier #2	N/A	0/0	0/0	0/0	0/0

convolutional layer in the stage being associated with a side-output layer. These stages have strides of 1, 2, 4, 8, and 16. Meanwhile, the pretraining on a general image classification task such as ImageNet, as mentioned in Ref. 23, has been shown to help with low-level tasks. Details about hyperparameters in the network layers are included in Tables 2, 3, and 4.

The kernel size for each layer of the main HNN network was 3, the kernel size for each side layer was 1, and the kernel size for the concatenation layer was 1. The HNN model generated from the training data was then applied to the test set to obtain segmentations for the remainder of the cine MRI scans. Each sequence ranged from 200 to 400 images.

2.4 Postprocessing

After segmentations, all of the cine MRI scans were obtained, and then processed further to clean up the segmentation before quantification. First, the segmentation probability map was binarized with values 1 assigned to the lung pixels and 0 to all other pixels. Small islands were then removed by identifying the largest objects representing the lungs in the binary masks and deleting all other blobs in the segmentation. For each sagittal case, the single largest blob corresponded to the lung, while for each axial and coronal case the two largest blobs represented the lungs.

Table 4 Concatenation layer.

Layer	Concatenation and multiscale weight
Learning/decay rate multiplier #1	0.001/1
Learning/decay rate multiplier #2	0.002/0

Once the islands were removed, holes in the lung segmentations were filled. To do so, inverted binary masks were created to represent the segmentations, where values 0 were assigned to the pixels representing the lungs and 1 to all other pixels. Then, the location of the holes was determined by finding in the masks where the nonlargest blobs were. Once these holes were found, the lung segmentations were filled in at these locations.

2.5 Lung Characteristic Measurements

The final goal of the segmentation was to be able to quantify lung characteristic measurements on the cine MRI because these measurements give insight on how disease affects lung movement. One characteristic is the lung area in one plane, which gives measurements that correlate well with spirometry values.⁸ To obtain more regional information from the lungs, other measurements were available as well.^{29,30} For diaphragm-specific information in the sagittal plane, the distances from the apex of the lung to the diaphragm at the anterior (ANT) and posterior (POST) costophrenic angles and at the cranio-caudal diameter ($C - C$ diameter, the distance of a vertical line dropped from the apex of the lung to the diaphragm) were found.³¹ The diaphragm movement area (DMA) quantifies the relative expansion of the diaphragm and was found by calculating the difference in area between the diaphragm positions in two different slices. Furthermore, to check the involvement of intercostal muscles, the CWMA was derived in a similar way to the DMA.³² The anterior-posterior diameter ($A - P$ diameter), which is defined as the distance between the anterior-posterior chest walls at the level of the apex of the visible diaphragm, was also measured. Various transverse, AP and vertical diameters were also determined from the lung segmentations in the axial and coronal planes. Figure 3 shows some of the measurements. Ultimately, our proposed method allowed for accurate lung

Fig. 3 Lung characteristic measurements from automatic segmentation. (a) The A–P diameter and C–C diameters and the DMA (green shading), (b) CWMA (yellow shading), lung area and transverse diameter in (c) axial and (d) coronal planes (red shading).

segmentation on large cine MRI datasets, which gave insight into lung movements via these lung characteristic measurements and increased the understanding of respiratory pathophysiology.

2.6 Validation Experiments

For validation, we compared the computer segmentation results to corresponding manual tracings. Using the manual segmentations as the gold standard, we examined the deviation of the HNN segmentations from the manual segmentation to judge our method's performance. The manual tracings were performed at the maximum inspiration and the maximum expiration for every sequence is obtained in every plane. In the axial and coronal planes, we obtained one set of manual segmentations that included the heart as well as another set that did not include the heart. This allowed us to evaluate the effect of the heart on lung segmentation. For the sagittal plane, two observers (one medical fellow and one research fellow) conducted independent manual segmentation to assess interobserver variability. For other planes, only one observer (research fellow) conducted the manual segmentation. Additionally, we manually segmented every image for a single patient and compared the breathing pattern obtained by manual and computer segmentations (research fellow) for the entire breathing cycles. The breathing pattern can be illustrated by plotting the lung characteristic measurements over time.

3 Results

Figure 4 shows the examples of manual, registration-based and HNN-based segmentations. The two manual segmentations in the sagittal plane come from two different observers. Qualitatively, it can be seen that the registration method fails to accurately delineate the lung borders, while the HNN method is able to capture a much smoother boundary. The quantitative comparisons of manual and computer segmentation in different planes are shown in Table 5. We show four different values: the area percentage difference, Dice similarity coefficient, average surface error, and maximum surface error. These metrics indicate the degree of agreement or disagreement between two different methods, and we use them to analyze the performance of both the registration and the HNN methods against the gold standard manual segmentation method. When comparing the values for the registration and HNN methods, there are significant differences for almost all of the metrics. Most importantly, there are significant differences in the Dice scores and average distance measurements between the two different methods, suggesting that the HNN method is a significant improvement from the registration method. In the axial and coronal planes, HNN segmentation on images that included the heart generated better results than on images without the heart. Overall, our results indicate a higher degree of agreement between the HNN

Fig. 4 Segmentation result from registration-based method and HNN method. (a) Manual segmentation (axial, coronal, and two segmentations on sagittal plane from two different observers), (b) segmentation result of registration-based method, and (c) segmentation result of the HNN method.

Table 5 Comparison of manual and computer segmentation on different planes.

	Area difference (%)	Dice	Avg dist error (mm)	Max dist error (mm)
Sagittal				
Manual versus registration	12.2 ± 3.1	93.1 ± 1.6	3.2 ± 0.6	13.1 ± 3.9
Manual versus HNN	-1.1 ± 3.7	97.2 ± 1.3	1.3 ± 0.4	5.6 ± 2.9
Obs 1 versus Obs 2	5.0 ± 3.2	95.9 ± 1.8	1.7 ± 0.6	6.0 ± 3.6
Coronal				
Manual versus registration w/heart	8.3 ± 10.1	93.5 ± 4.2	3.9 ± 1.9	16.6 ± 10.3
Manual versus HNN w/heart	-4.2 ± 10.8	96.1 ± 3.8	2.3 ± 1.8	12.2 ± 11.3
Manual versus HNN w/o heart	-5.2 ± 7.9	93.3 ± 2.7	2.1 ± 1.0	13.3 ± 6.6
Axial				
Manual versus registration w/heart	3.8 ± 3.7	95.6 ± 1.8	2.3 ± 0.8	9.2 ± 4.0
Manual versus HNN w/heart	-2.7 ± 3.9	96.6 ± 1.7	1.8 ± 0.7	6.5 ± 3.2
Manual versus HNN w/o heart	-2.7 ± 8.1	93.3 ± 3.1	2.2 ± 0.8	11.7 ± 5.7

Note: Bold values are the best performer in each category.

Table 6 Statistical comparison of manual and computer segmentation on different planes.*P*-values from paired two sample for means *t*-test

Plane	Comparison	Area difference (%)	Dice	Avg dist error (mm)	Max dist error (mm)
Sagittal	Manual versus registration versus manual versus HNN	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p < 0.05$
Coronal	Manual versus registration w/heart versus manual versus HNN w/heart	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p < 0.05$
Axial	Manual versus registration w/heart versus manual versus HNN w/heart	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p < 0.05$

computer segmentation and the gold standard than between the registration segmentation and the gold standard. It is also important to note that there was little inter-observer variability between the two manual observers.

For the cases that include the heart, the total segmentation area refers to the lung plus the heart. Evaluating cases with and without the heart allowed us to observe the effect of the heart on lung segmentation. After comparing the mean performance metrics between the registration and HNN methods via two-sample *t* tests, we concluded that the superiority of the HNN method is statistically significant. *T* test results are presented in Table 6.

4 Discussion

Deep learning has been recently recognized as an important tool to improve the efficacy of not only general computer vision problems, but for medical imaging as well. One of the core tenants of deep learning is the utilization of “big data” to train robust models that perform effectively on new data sets. However, in the field of medical imaging, a major problem that exists is the lack of large data sets available. While there are attempts to remedy this situation through the use of challenges and the encouragement of publicizing data, many areas still suffer from this paucity of data. Over time, more data will inevitably become publicly available, though it may be delayed due to issues including patient privacy. However, our results demonstrate deep learning’s usefulness in medical imaging despite this lack of data: using only a single slice from each sequence to build a training set, we are able to train a model that works for the entire sequence of hundreds of images.

2-D cine MRI has been found to be a successful method for imaging dynamic breathing cycles, but with the increased temporal resolution of the acquisition method also comes decreased spatial resolution. Thus, performing automated segmentations of the lungs on 2-D cine MRI is extremely difficult. As can be seen in Fig. 1, the contrast at the lung boundary can be very poor, making intensity-based methods, such as snakes, over-segment at these blurry boundaries. Clearly, some prior knowledge is necessary to improve this method. Both HNN- and registration-based methods try to incorporate the prior knowledge via prior segmentations.

The registration-based method showed some moderate success at this task, though there were a few issues that limited this method. For example, in some of our deep breathing sagittal cases, the lung expanded considerably more than average. In these cases, the Demon’s registration algorithm would either over- or under-segment the lung region, depending on the cycle phase that was being registered: inspiration to expiration could result in over-segmentation, while expiration to inspiration could result in under-segmentation. A large deviation between the registered slice and target slice was necessary for

this issue to occur, and though such a large deviation was rare, it necessitated the inclusion of a postprocessing step. Because we did not require the manually segmented slice to be either the maximum inspiration or maximum expiration, both over- and under-segmentation could occur. Furthermore, vessels could interfere with the registration if they were present in some of the slices but not others. A common issue with the lung segmentation is the difficulty of ensuring that the extremities of the lung are properly segmented. Because their shape and position can vary drastically, the registration-based approach had difficulty in fully capturing the extremities, necessitating the postprocessing step. This could be caused by the kernel size of the Gaussian smoothing. Despite these issues with the Demon’s method, we still choose to use the algorithm as the baseline method for comparison because of its availability and success in other tasks. In this investigation, we demonstrate that registration-based method is sensitive to the selection of reference models and so it does not work well for large motion.

Meanwhile, our deep learning-based method, which utilized information derived from the same patient and other similar lung images, performed very well. The deep learning method produced the best segmentations in the sagittal plane, and so in these sagittal cases the only postprocessing step necessary was the Gaussian smoothing to ensure that the slices transitioned smoothly. The other issues present in the registration method, such as the difficulty in the inclusion of the lung corners, were not as prevalent when using the HNN method. Overall, our method of segmentation via HNN most successfully overcame the resolution limits in cine MRI, and it can be readily applied to both new and existing data.

Furthermore, the axial and coronal plane analyses were complicated by the inclusion of the heart in these scans (the sagittal plane was taken in the right lung, so the heart would not be included). This made the sagittal plane the most likely to work with both methods, while the axial plane had more trouble than the coronal plane with the HNN-based method. The deep learning method performed much better on the axial sequences that omitted the heart than on the axial sequences that included it. Meanwhile, the lung itself with its relatively simple shape and intensity distribution was very accurately segmented by the deep learning method. These accurate segmentations proved to be crucial for extracting more lung characteristics measurements, such as DMA and CWMA. It is almost infeasible to measure these dynamic areas by hand, and so obtaining accurate segmentations was a necessary step that led to the ability to make these measurements. In the end, lung segmentations as well as the derived measurements are all useful for clinical and research purposes.

One of the useful properties of the deep learning method is that as more patients are processed, the more robust the method becomes. Once there is enough data, the model will be able to

run effectively without the manual segmentation, though it may be difficult to determine when this switch will be possible. Most likely a softer transition would be required: from creation of a new model for each new dataset, to the fine-tuning of a more robust model, to the completion of the desired model that requires no further training.

Another advantage of the HNN method is that once the initial model is created for a data set, generating the segmentations of unknown slices is extremely quick, making it possible to generate data for all of the slices, rather than just the inspiration and expiration slices. This would allow for the analysis of the breathing pattern in addition to the absolute differences between the start and end of a breathing cycle. While it may be difficult to generate meaningful data due to the voluntary nature of the breathing in this case, it has proven to be useful to identify patients who had an abnormal breath within a full breathing cycle.

In this paper, we propose a deep learning-based method to segment lung MRI cine images and find that it outperformed the traditional registration-based method, potentially leading to a more accurate assessment of lung ventilation in young boys with DMD and the age-matched healthy volunteers.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This research was funded by the Intramural Research Program of the National Institutes of Health Clinical Center and National Institute of Neurological Disorders and Stroke. The authors thank the study participants and their families; Donovan Stock, Elizabeth Hartnett, and Alice Schindler (NINDS) for help with coordinating study visits and scheduling MRI studies; Christine Mancini and Drs. Laura Olivieri, and Sujata Shanbhag (National Heart, Lung and Blood Institute) for help with acquiring MRI data; Hirity Shimellis for assistance with database audit; and the Principal Investigators of the referring sites (Drs. Barry Russman, Portland, OR; Benjamin Renfro, Gulf Breeze, FL; Brenda Wong, Cincinnati, OH; Douglas Sproule, New York, NY; Edward Smith, Durham, NC; and Kathryn Wagner, Baltimore, MD) for sharing patients with DMD. This work utilized the computational resources of the NIH HPC Biowulf cluster. (<http://hpc.nih.gov>)

References

1. R. A. Faust, K. B. Remley, and F. L. Rimell, "Real-time, cine magnetic resonance imaging for evaluation of the pediatric airway," *Laryngoscope* **111**, 2187–2190 (2001).
2. K. Mogalle et al., "Quantification of diaphragm mechanics in Pompe disease using dynamic 3D MRI," *PLoS One* **11**, e0158912 (2016).
3. C. Kolb et al., "Regional lung ventilation analysis using temporally resolved magnetic resonance imaging," *J. Comput. Assist. Tomogr.* **40**, 899–906 (2016).
4. T. Batinic et al., "Dynamic diaphragmatic MRI during apnea struggle phase in breath-hold divers," *Respir. Physiol. Neurobiol.* **222**, 55–62 (2016).
5. P. Ciet et al., "Spirometer-controlled cine magnetic resonance imaging used to diagnose tracheobronchomalacia in paediatric patients," *Eur. Respir. J.* **43**, 115–124 (2014).
6. K. Suga et al., "Impaired respiratory mechanics in pulmonary emphysema: evaluation with dynamic breathing MRI," *J. Magn. Reson. Imaging* **10**, 510–520 (1999).
7. P. Kohlmann et al., "Automatic lung segmentation method for MRI-based lung perfusion studies of patients with chronic obstructive pulmonary disease," *Int. J. Comput. Assist. Radiol. Surg.* **10**, 403–417 (2015).
8. R. Tetzlaff et al., "Lung function measurement of single lungs by lung area segmentation on 2D dynamic MRI," *Acad. Radiol.* **17**, 496–503 (2010).
9. T. Böttger et al., "Implementation and evaluation of a new workflow for registration and segmentation of pulmonary MRI data for regional lung perfusion assessment," *Phys. Med. Biol.* **52**, 1261–1275 (2007).
10. T. Kunert et al., "An interactive system for volume segmentation in computer-assisted surgery," *Proc. SPIE* **5367**, 799 (2004).
11. T. Böttger et al., "Application of a new segmentation tool based on interactive simplex meshes to cardiac images and pulmonary MRI data," *Acad. Radiol.* **14**, 319–329 (2007).
12. I. Middleton and R. Damber, "Segmentation of magnetic resonance images of the thorax by backpropagation," in *IEEE Int. Conf. on Neural Networks*, pp. 2490–2494 (1995).
13. I. Middleton and R. Damber, "Segmentation of magnetic resonance images using a combination of neural networks and active contour models," *Med. Eng. Phys.* **26**, 71–86 (2004).
14. N. Ray et al., "Merging parametric active contours within homogeneous image regions for MRI-based lung segmentation," *IEEE Trans. Med. Imaging* **22**, 189–199 (2003).
15. W. F. Sensakovic et al., "Automated lung segmentation of diseased and artifact-corrupted magnetic resonance sections," *Med. Phys.* **33**, 3085–3093 (2006).
16. A. Osareh and B. Shadgar, "A segmentation method of lung cavities using region aided geometric snakes," *J. Med. Syst.* **34**, 419–433 (2010).
17. Y. Yang et al., "A spatiotemporal-based scheme for efficient registration-based segmentation of thoracic 4-D MRI," *IEEE J. Biomed. Health Inform.* **18**, 969–977 (2014).
18. R. S. Tavares et al., "Temporal segmentation of lung region MR image sequences using Hough transform," in *Conf. Proc. IEEE Engineering in Medicine and Biological Society*, pp. 4789–4792 (2010).
19. R. S. Tavares et al., "Temporal segmentation of lung region from MRI sequences using multiple active contours," in *Conf. Proc. IEEE Engineering in Medicine and Biological Society*, pp. 7985–7988 (2011).
20. F. Guo et al., "Anatomical pulmonary magnetic resonance imaging segmentation for regional structure-function measurements of asthma," *Med. Phys.* **43**, 2911–2926 (2016).
21. R. M. Summers, "Progress in fully automated abdominal CT interpretation," *Am. J. Roentgenol.* **207**, 67–79 (2016).
22. H. Greenspan, B. v. Ginneken, and R. M. Summers, "Deep learning in medical imaging: overview and future promise of an exciting new technique," *IEEE Trans. Med. Imaging* **35**, 1153–1159 (2016).
23. S. Xie and Z. Tu, "Holistically-nested edge detection," in *2015 IEEE Int. Conf. on Computer Vision*, pp. 1395–1403 (2015).
24. H. Roth et al., "Spatial aggregation of holistically-nested networks for automated pancreas segmentation," in *19th Int. Conf. on Medical Image Computing and Computer Assisted Intervention*, Athens, Greece (2016).
25. L. Ibanez and W. Schroeder, *ITK Software Guide*, Kitware, Inc., USA (2003).
26. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. on Learning Representations* (2014).
27. J. Deng et al., "ImageNet: a large-scale hierarchical image database," in *IEEE Computer Vision and Pattern Recognition* (2009).
28. H.-C. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging* **35**, 1285–1298 (2016).
29. T. Kondo et al., "A dynamic analysis of chest wall motions with MRI in healthy young subjects," *Respirology* **5**, 19–25 (2000).
30. M. Gaeta et al., "Clinical and pathophysiological clues of respiratory dysfunction in late-onset Pompe disease: new insights from a comparative study by MRI and respiratory function assessment," *Neuromuscul. Disord.* **25**, 852–858 (2015).
31. T. Kondo et al., "A dynamic analysis of chest wall motions with MRI in healthy young subjects," *Respirology* **5**, 19–25 (2000).
32. M. Gaeta et al., "Clinical and pathophysiological clues of respiratory dysfunction in late-onset Pompe disease: new insights from a comparative study by MRI and respiratory function assessment," *Neuromuscul. Disord.* **25**, 852–858 (2015).

Biographies for the authors are not available.