Title
High throughput image labeling on chest computed tomography by deep learning.

Journal
Journal of Medical Imaging, 7(2)

ISSN
2329-4302

Authors
Wang, Xiaoyong
Teng, Pangyu
Ontiveros, Ashley
et al.

Publication Date
2020-03-01

DOI
10.1117/1.JMI.7.2.024501

Peer reviewed

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

High throughput image labeling on chest computed tomography by deep learning

Xiaoyong Wang
Pangyu Teng
Ashley Ontiveros
Jonathan G. Goldin
Matthew S. Brown

SPIE.

Xiaoyong Wang, Pangyu Teng, Ashley Ontiveros, Jonathan G. Goldin, Matthew S. Brown, "High throughput image labeling on chest computed tomography by deep learning," *J. Med. Imag.* **7**(2), 024501 (2020), doi: 10.1117/1.JMI.7.2.024501

High throughput image labeling on chest computed tomography by deep learning

Xiaoyong Wang,[a,b,*] Pangyu Teng,[a,b] Ashley Ontiveros,[a,b] Jonathan G. Goldin,[a,b] and Matthew S. Brown[a,b]

[a]University of California, Los Angeles, Center for Computer Vision and Imaging Biomarkers, Los Angeles, California, United States

[b]University of California, Los Angeles, Department of Radiological Sciences, Los Angeles, California, United States

Abstract. When mining image data from PACs or clinical trials or processing large volumes of data without curation, the relevant scans must be identified among irrelevant or redundant data. Only images acquired with appropriate technical factors, patient positioning, and physiological conditions may be applicable to a particular image processing or machine learning task. Automatic labeling is important to make big data mining practical by replacing conventional manual review of every single-image series. Digital imaging and communications in medicine headers usually do not provide all the necessary labels and are sometimes incorrect. We propose an image-based high throughput labeling pipeline using deep learning, aimed at identifying scan direction, scan posture, lung coverage, contrast usage, and breath-hold types. They were posed as different classification problems and some of them involved further segmentation and identification of anatomic landmarks. Images of different view planes were used depending on the specific classification problem. All of our models achieved accuracy >99% on test set across different tasks using a research database from multicenter clinical trials. © *2020 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JMI.7.2.024501]

Keywords: computed tomography; image labeling; convolutional neural network; clinical trials.

Paper 19189RR received Jul. 30, 2019; accepted for publication Feb. 26, 2020; published online Mar. 20, 2020.

1 Introduction

Image data curation, involving identification and labeling of relevant types of scans with consistent acquisition parameters, is important for machine learning with big data and clinical trials. Labeling of images meeting standardization requirements is vital in acquiring reliable research findings from multicenter clinical trials[1–4] since images collected from different sites are heterogeneous and variable in terms of the types of scans received. For example, each patient may include image series from multiple time points and each time point could have multiple series that can be redundant, irrelevant, or unusable. The traditional approach is to manually assign the labels by reviewing each image series and selecting the best series to process that meets standardization requirements. It is very time-consuming and prone to human errors. As such, an efficient automatic image labeling method will be beneficial for large-scale clinical research.

Digital imaging and communications in medicine (DICOM) is the standard format in medical imaging and it contains a variety of scan parameters and other metadata. DICOM headers, e.g., series description, are extensively used to extract for labeling information. Nevertheless, DICOM tags are often insufficient or unreliable due to manual entry or vendor discrepancies.[5–9] To overcome this limitation, image-based identification is crucial and different techniques have been introduced.[10–13] In recent years, deep learning has been widely used in medical imaging, especially for classification[14–17] and segmentation.[18–22] A five-layer convolutional neural network (CNN) was used to classify the anatomical region scanned in computed tomography (CT) including brain, neck, chest, abdomen, and pelvis.[23,24] Yan et al.[25] used a multistage framework

*Address all correspondence to Xiaoyong Wang, E-mail: xiaoyongw@ucla.edu

to recognize 12 different body parts by approximating bounding box of local patches. Similar approach was applied to anatomical classification in MRI.[26] Images of orthogonal planes (axial, sagittal, and coronal) were used to build three independent CNNs and results were combined to produce 3-D bounding box for each organ, such as heart, lung, kidney, and liver.[27,28] Different organs and tissue were segmented on abdominal CT by a 2-D U-net like network to quantify body composition.[29]

For a chest CT scan, we want to identify the scan direction (head first or feet first), scan posture (prone or supine), whole lung anatomical coverage, contrast agent usage, and breath-hold level [residual volume (RV) or total lung capacity (TLC)]. They are either not consistently recorded in DICOM header or unreliable in clinical practice. These labeling tasks can be posed as classification problems. Although there has been extensive image processing, classification, and segmentation work applied to chest CT,[30–32] to our knowledge, there is no related works on this specific problem of classification for comprehensive scan labeling. Although these labels seem basic for chest CT, they are labor intensive to assign for large data sets, and effectively and correctly extracting them is crucial for data mining in clinical research.

A fully automated high throughput labeling method is proposed using deep learning to create classification models for each. Our hypothesis is that each classifier can achieve >95% accuracy individually in assigning its label.

2 Materials and Methods

In total, there are five labeling tasks to accomplish. All of them are posed as classification by CNN and contrast detection and breath-hold identification will further require segmentation of anatomic landmarks. For efficient processing of the 3-D CT data sets, 2-D images of different view planes will be used.

2.1 *Scan Direction*

Patients can be scanned head first or feet first. Examples of head-first and feet-first coronal images are shown in Fig. 1. Most chest CT analysis algorithms are developed expecting head-first ordering of images.

Previous studies[23–25] used axial slices to identify different anatomies, including brain, shoulder, chest, abdomen, and pelvis; and thus scanning direction can be inferred by the aggregated anatomical order. One challenge of this approach is the requirement of annotation on every axial slice. Our approach rather uses the entire scan to determine the scan direction. A total of 1000 chest CT scans of different subjects from a research database comprising six clinical trials were used, including two idiopathic pulmonary fibrosis (IPF) trials, two chronic obstructive pulmonary disease (COPD) trials, one scleroderma trial, and one lung cancer screening trial. Scanners from Siemens, GE, Philips, and Toshiba were used with varied reconstruction kernels. Slice thickness and spacing is within (0.6, 3 mm). For each scan, a middle coronal slice plus 2 adjacent slices 10 pixels apart were selected for input to the classifier. All scans were confirmed manually as being head-first and corresponding feet-first scans were generated by vertically

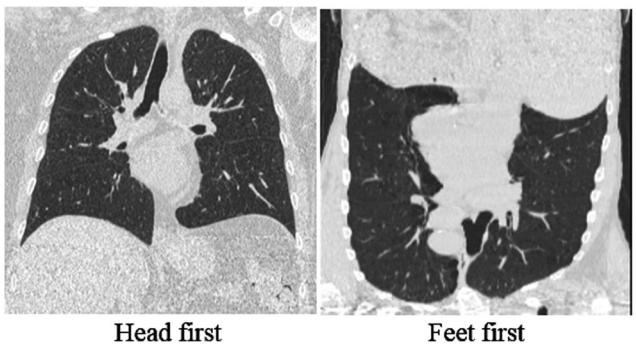


Fig. 1 Example coronal images of head-first and feet-first chest scans.

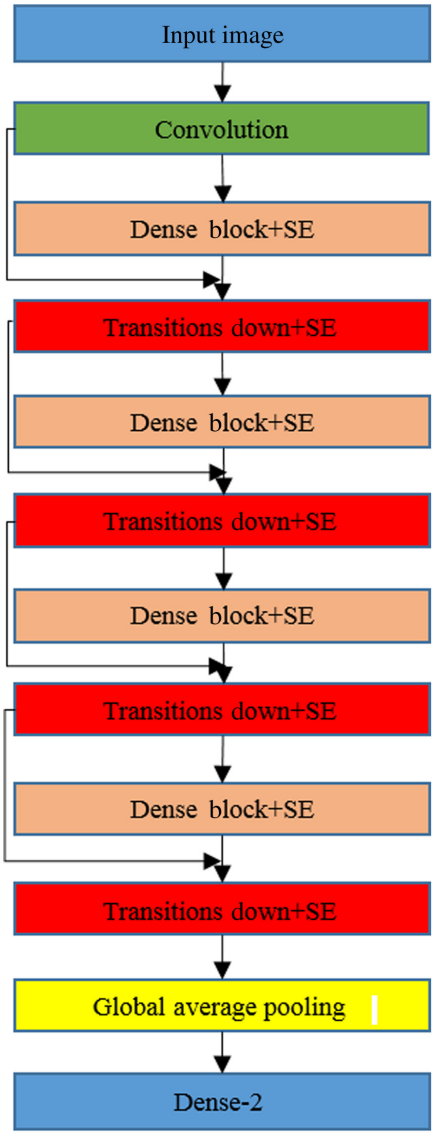


Fig. 2 SE-DenseNet121 classification architecture.

flipping the original scans to form a balanced dataset (so each scan generated six samples, two classes with three coronal slices each). In total, 6000 slices were used and they were split into training and test set by 4:1.

As shown in Fig. 2, squeeze and excitation (SE)[33] embedded DenseNet121 (SE-DenseNet121) was used as the classification architecture. In DenseNet, each layer obtains additional inputs from all preceding layers and passes on its own feature maps to all subsequent layers. It contributes to strengthening feature propagation, encouraging feature reuse, and substantially reducing the number of parameters. The SE architecture unit is a content aware mechanism that aims to adaptively recalibrate or weight channel-wise feature responses by explicitly modeling interdependencies between channels. It contributes to performance improvement with little additional computation cost.

2.2 *Scan Posture*

Scan posture, face up (supine) or face down (prone), dependent on placement of the patient in the scanner. Example sagittal images of prone and supine scan are shown in Fig. 3 with the vertical lines representing the scanner table. Supine imaging is most commonly performed, for example, in lung cancer screening; however, prone imaging is often performed in evaluating interstitial

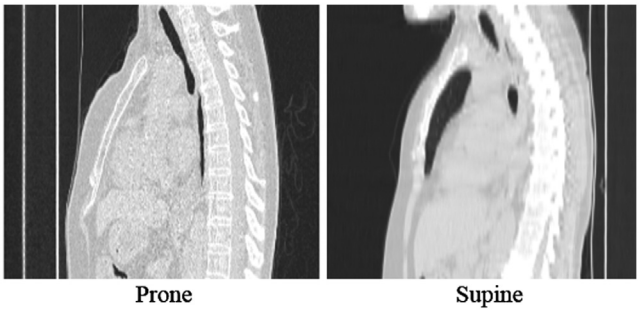


Fig. 3 Example sagittal images of prone and supine scans.

lung disease.[34–36] When there are suspicious dependent opacities on a supine scan, it is highly recommended to also capture corresponding image in the pone position since these dependent opacities may resolve. Both axial and sagittal images can be used to determine if a scan is prone or supine. We will use sagittal slices since it clearly depicts the relationship between entire spine and table. The middle sagittal slice is an ideal representational image and two other sagittal slices at ¼ and ¾ way through the body in the sagittal direction are also used in case the table is not present in the image center. A total of 5000 scans (15,000 slices) of different patients from a research database comprising 16 clinical trials (five IPF, four scleroderma, three COPD, two lung cancer screening, one tuberculosis, and one lymphangioleiomyomatosis) were used in the training and test sets, formed with a ratio of 4:1. Prone and supine scans account for 50% in both sets and the reference is from manual labeling. Scanners of different manufacturers were used and slice thickness/spacing is within (0.5, 3 mm). The same SE-DenseNet121 was used as classification architecture in this task.

2.3 *Lung Coverage*

A chest scan can be defined as complete lung coverage if it includes the lung apex, intermediate lung, and lung base. Example scans with complete and incomplete lung coverage are shown in Fig. 4. Scans with only partial lung are not suitable for subsequent processing, such as lung and lobar segmentation. Representative coronal slices, e.g., middle coronal slice, could be used to identify the lung coverage; however, for scans where lung coverage is slightly incomplete, a few coronal slices may appear complete. Therefore, we decided to use axial images to tackle this problem. First, a deep learning mode was built to detect the presence of lung on each axial image. After that aggregating individual slice labels to infer whole scan label. A complete lung coverage scan should begin and end with "no-lung" slices and have "lung" slices in the middle, whereas superior and/or inferior slices would be classified as "lung" for an incomplete lung coverage scan. To build the model recognizing lung presence, 20,000 axial slices from 210 different scans/patients were used with equal number of lung and nonlung slices. They are a collection from two IPF trials and two COPD trials and all the scans have radiologist approved lung segmentations that were used to determine individual slice labels. Scanners of various

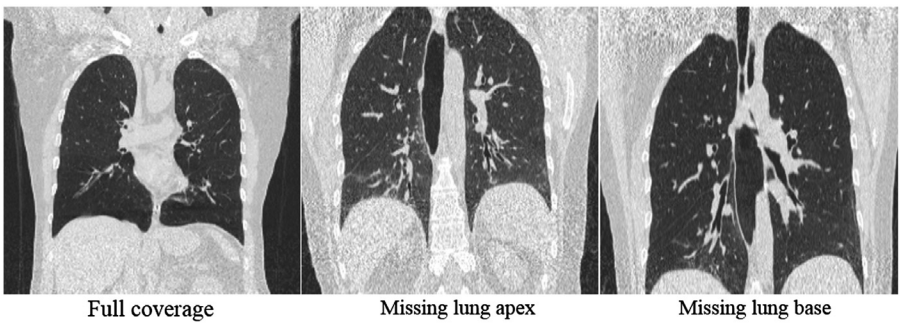


Fig. 4 Example images with complete lung coverage, missing lung apex, and lung base.

manufacturers were used and slice thickness/spacing is within (0.6, 3 mm). Training and test sets were formed by 4:1 and the same classification architecture used in scan direction and posture was employed again.

2.4 *Contrast Usage*

CT can be ordered either with or without contrast. Contrast CT is able to enhance the intensity of target tissue, e.g., blood vessels and tumor and assist physicians by providing better structural and functional information. The presence of enhancement has a considerable impact on subsequent image analysis results. To identify the presence of contrast, the aorta is an ideal anatomic location since enhancement can be observed persistently within the vessels during early arterial phase and late arterial phase.[37,38] For contrast scans, the intensity within aorta is over 90 Hounsfield unit (HU), whereas noncontrast scan intensity is <50 HU.[39] Two example contrast and noncontrast scans are shown in Fig. 5. For contrast detection, we begin with aorta segmentation. Descending aorta is preferred since it has a relatively consistent circular shape beyond the aortic arch. From the aorta segmentation, the mean intensity is computed to determine presence or absence of contrast. In total, 116 scans (8447 slices) of 70 patients from one lung cancer screening trial were used, comprising 46 contrast scans and 70 noncontrast scans. They were acquired by scanners of multiple manufactures and slice thickness/spacing is within (1, 3 mm). Independent 500 slices (half contrast and half noncontrast) were used as a test set to evaluate the segmentation performance. An additional independent 500 scans with only scan level label (250 contrast and 250 noncontrast scan) were used as test set for contrast detection and they are a collection from five clinical trials (two IPF, two COPD, and one scleroderma). More noncontrast scans, in which the aorta has bad contrast with surrounding tissues, were used in building the segmentation model to reinforce it. The reference segmentation of the training set was done by simple threshold plus manual editing.

The previous classification architecture SE-DenseNet121 was extended to perform segmentation by adding a decoder section to recover original resolution and achieve pixel-wise segmentation.

As shown in Fig. 6, the segmentation architecture is composed of encoder and decoder stagers. The down-sampling path includes four transitions down and four dense blocks (256×256 to 16×16). Corresponding up-sampling path includes four transitions up and four dense blocks (16×16 to 256×256). Each dense block includes four densely connected layers and an SE block. Transition down includes (batch normalization, Convolution2D, maxpooling, and SE block). Transition up includes (transposed Convolution2D, batch normalization, SE block).

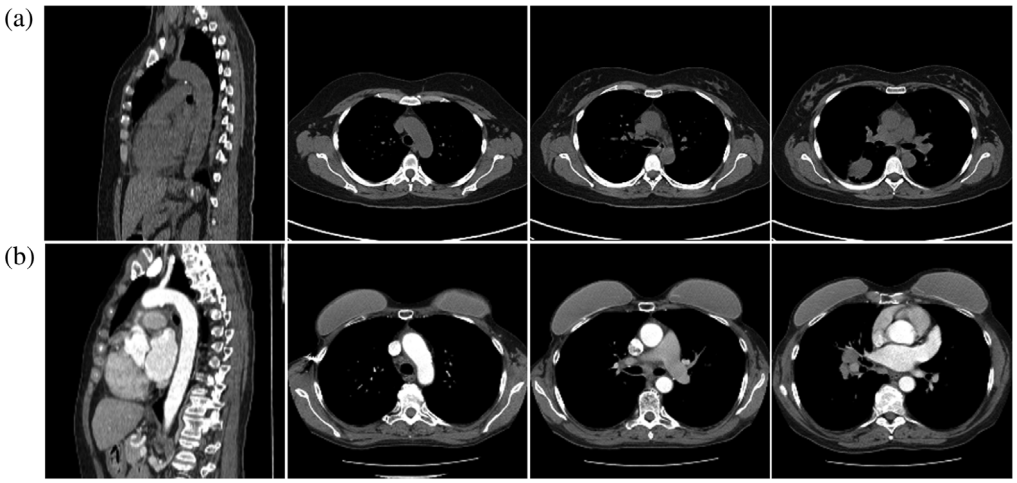


Fig. 5 Typical (a) noncontrast and (b) contrast enhanced scans.

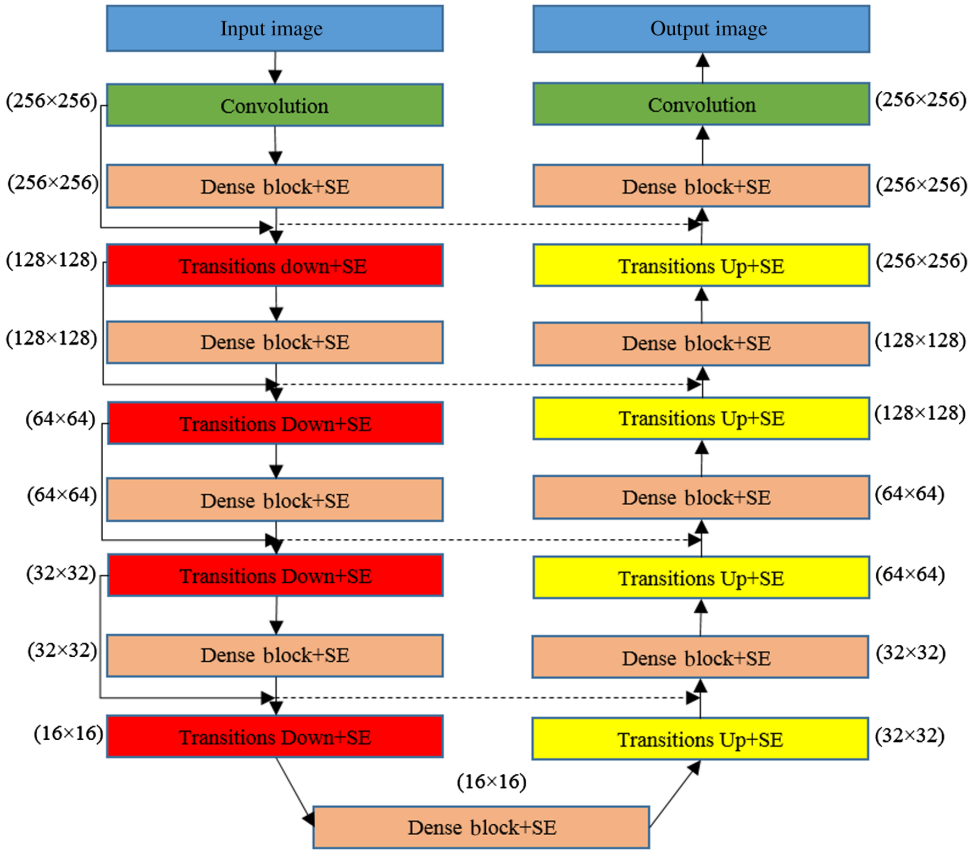


Fig. 6 SE-DenseNet121-based segmentation network. Dashed lines indicate skip connections from down-sampling to up-sampling.

2.5 *Breath-Hold Types*

Patients are usually asked to hold their breath during scanning at full inspiration, i.e., TLC, or at full expiration, i.e., at RV. Typically, inspiratory chest CT scans are the preferred imaging method in patients with pulmonary disease, whereas expiratory scans sometimes are superior to demonstrate some pathophysiological alterations. For example, TLC scan is commonly used in evaluation of most diffuse lung diseases, including fibrosis and emphysema. On the other hand, RV scan is ideal for the assessment of air trapping.[40–42] Lung volume alone is not sufficient to differentiate between TLC and RV since it is influenced by a variety factors, such as gender difference, adult versus pediatrics, and pulmonary disease. Clinically, the compression of the trachea is often used by radiologists to identify RV scans. It is characterized by the collapse of carina and posterior wall of trachea bows forward.[43–48] Some examples of TLC and RV axial images were shown in Fig. 7. As such, breath-hold identification can be divided into three steps:

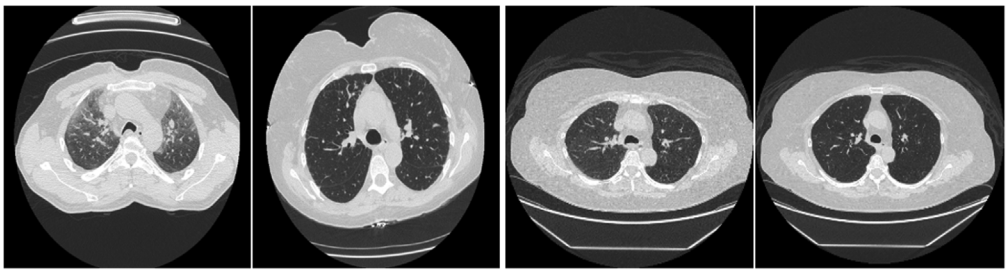


Fig. 7 Example RV and TLC scans shown in axial slice and the compression of trachea carina is a characteristic of RV.

Table 1 Number of training, validation, and test set by slice/scans.

Classification task	Training set	Validation set	Test set
Scanning direction (head or feet first) (scans)	640	160	200
Scanning posture (prone or supine) (scans)	3200	800	1000
Lung presence detection (slices)	12,800	3200	4000
Aorta segmentation/contrast detection	6757 slices	1690 slices	500 slices/independent 500 scans
Trachea segmentation (slices)	25,838	6460	1000
Breath-hold types (RV or TLC) (slices)	15,297	3824	4780

(1) segmentation of trachea, (2) identification of carina, and (3) identification of RV or TLC based on the shape of carina.

A total of 356 scans (32,298 slices) of 92 patients from two lung cancer screening trials were used to build the trachea segmentation model. Of these, 176 were RV scans and 180 were TLC scans. Images were acquired from scanners of different manufacturers and slice thickness/spacing is within (0.6, 1.5 mm). Independent 1000 slices were saved for trachea segmentation evaluation. The segmentation of the training set was done by intensity thresholding plus manual editing, covering the main bronchi and left/right bronchus beyond carina. Because the ultimate goal of trachea segmentation is to find the carina, it is not necessary to segment the whole airway tree. The same SE-Denset21-based segmentation network was applied to build the model.

After segmenting the trachea, the carina is identified by searching slice by slice until the bifurcation. In addition to the segmentation model, a classification model is required to differentiate RV and TLC based on carina shape. To train the classification model with more samples, multiple slices above the bifurcation will be used and this strategy also applies in the inference stage using majority voting for final output. As a result, 1633 scans of 495 patients that is independent of dataset of trachea segmentation were used, including 11,948 RV slices and 11,953 TLC slices and 20% of cases were used as the test set. They were collected from 11 clinical trials (four IPF, three COPD, one scleroderma, one non-small cell lung cancer, one lung cancer screening, and one tuberculosis). The previous SE-DenseNet12 classification architecture was used to differentiate RV and TLC.

Table 1 shows the number of training, validation, and test set on each specific task.

3 Data Preparation

Fivefold cross validation was applied in evaluation for all of the classification tasks. The same image normalization and down-sampling was used in preprocessing. Every slice was normalized by linear remapping of [−1000, 1000 HU] to [0.0, 1.0] and resized to 256×256.

Data augmentation was used, including rotation, translation, horizontal and vertical flipping, rescaling, and options are varied across different tasks. For example, vertical flipping was dropped in training the head- or feet-first model using coronal slices.

In the training of aorta and trachea segmentation, dice loss was used as loss function since they are very small compared to the background. The optimizer was Adam[49] with learning rate of 0.001 and decay rate of 1×10^{-4}. Training of models was performed using a NVIDIA TITAN X with 12 GB memory and implemented using Keras[50] with tensorflow[51] backend.

4 Results

The specific accuracy on each classification task as shown in Table 2. The number of classes is balanced in test sets of different classification tasks and accuracy was used as metric. The last columns show the computational time for a chest scan with 300 slices.

Table 2 Test accuracy and speed of each task.

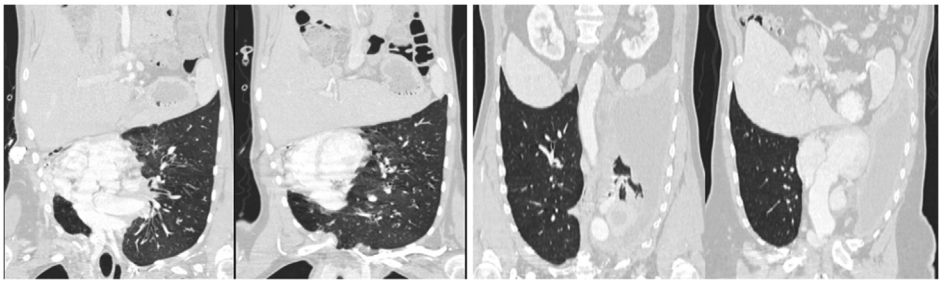
Classification task	# cases	Accuracy	Scan of 300 slices
Scanning direction (head or feet first)	1200 slices/400 scans	100%/100%	10 ms
Scanning posture (prone or supine)	3000 slices/1,000 scans	99.5%/100%	10 ms
Lung presence/coverage completeness	4000 slices/42 scans	99.1%/100%	5 s
Contrast detection	500 scans	100%	5 s
Breath-hold types (RV or TLC)	4780 slices/326 scans	98.4%/99.3%	10 ms

4.1 *Scan Direction*

The scanning direction model achieved 100% accuracy on test set and works even when the lung coverage is incomplete in clinical practice. Figure 8 shows two cases from the public dataset Lobe and Lung Analysis 2011 (LOLA11).[52] One of them has partial right lung and the other has only right lung. Figure 9 contains examples applying the model on low-resolution scans (spacing = 20 mm) and it still succeeded in differentiating the scan directions.

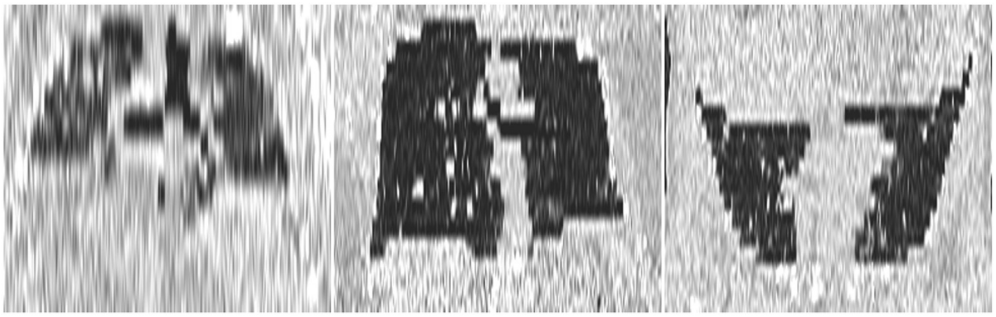
4.2 *Scan Posture*

The scan posture model achieved 99.5% accuracy on test set. Figs. 10(a) and (b) are the cases where the CNN correctly identifies the posture even though the information in the DICOM header tag is incorrect. There is a scan posture related DICOM tag called "PatientPosition," which is a relatively reliable, but as shown in Fig. 10 it is not always correct when the image is flipped. Our CNN model could consistently recognize the posture despite of bed position.



(a) CNN: feet-first (b) CNN: feet-first

Fig. 8 CNN inference on (a) a case with partial right lung and (b) a case with only right lung.



# of slice = 14
CNN: head-first

# of slice = 32
CNN: head-first

# of slice = 36
CNN: feet-first

Fig. 9 Application on low-resolution scans (spacing = 20 mm).

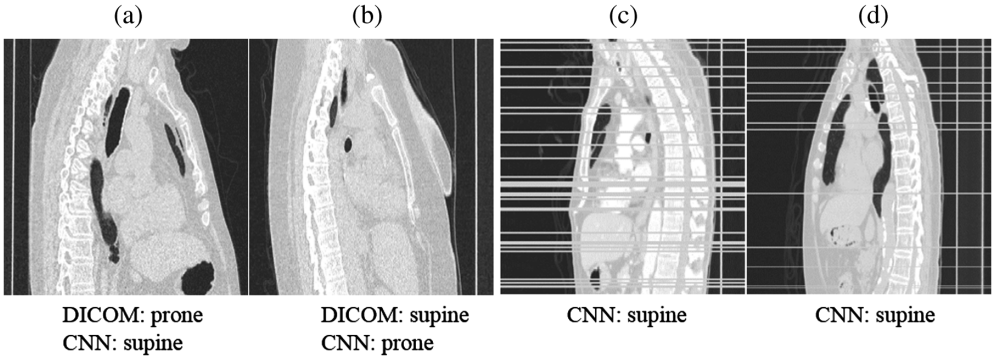


Fig. 10 Example scan posture detection by the proposed CNN. (a) and (b) Examples with incorrect DICOM header information. (c) and (d) Examples with intermediate blank slices.

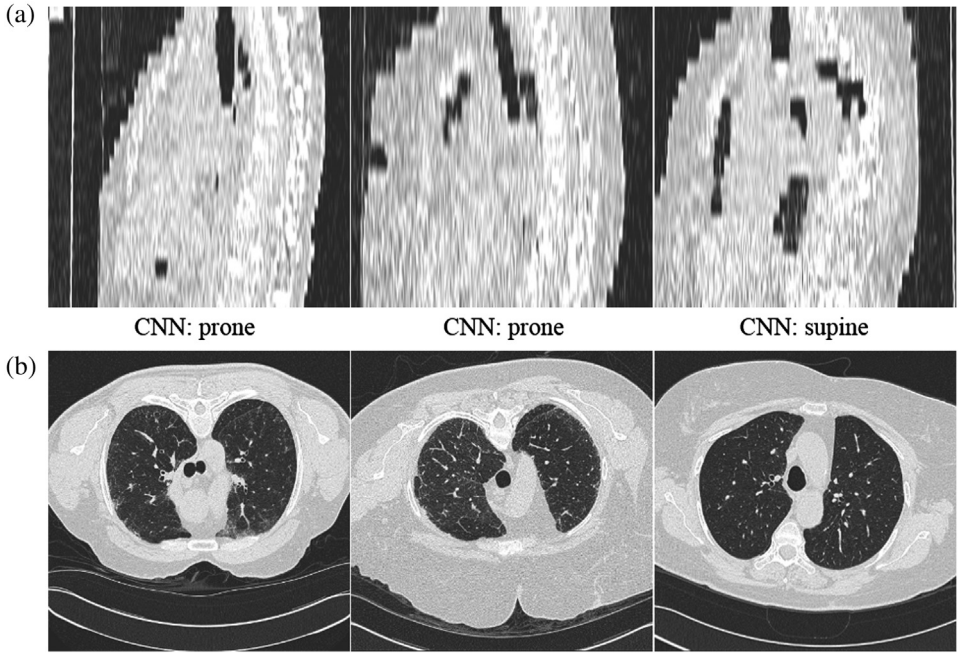


Fig. 11 Scan posture model applied to low-resolution scans (spacing = 20 mm). (a) Sagittal slices as input to the model and (b) corresponding axial slices to confirm that the CNN inference is correct.

Figs. 10(c) and (d) show two cases with intermediate blank slices and the model is still able to correctly identify the posture.

Figure 11 shows the application of scan posture model to three low-resolution scans (slices pacing = 20 mm) and their postures were all correctly identified.

4.3 *Lung Coverage*

The lung coverage model achieved 99.4% accuracy on identifying slice-based lung presence and 100% accuracy on lung completeness recognition. Figure 12 shows two incomplete lung coverage cases. The first row is an example with incomplete coverage of the lung apex (most superior axial slices contains lung) and the second row corresponds to a case with incomplete coverage of the lung base (most inferior axial slices contains lung).

The model is able to correctly classify slices as "no lung" containing only gut or intestines instead of lung (central image of first row). Additionally, it is capable of correctly identifying slices as "lung" with small amount of basal (central image of second row). Indicated by the

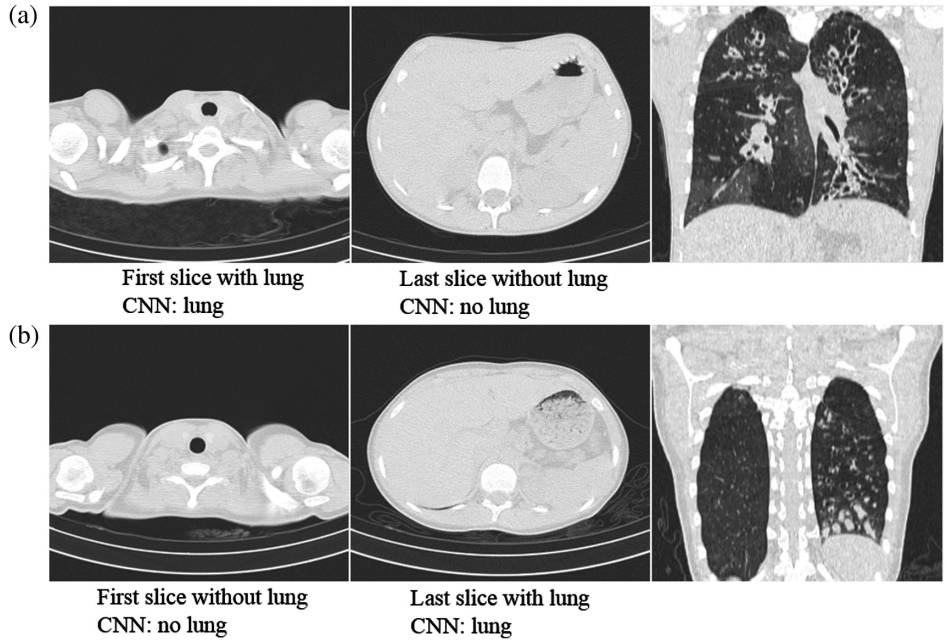


Fig. 12 (a) A case with incomplete coverage of lung apex. (b) A case with incomplete coverage of lung base.

coronal view of last column, using representative coronal slice, e.g., middle coronal slice, is sometimes risky.

Although the lung completeness model achieved 100% accuracy, there are some misclassifications by the lung presence mode. They are thin slices either from lung apex and lung base are shown in Fig. 13. All of them were identified without lung presence (false negative, if 0 = no lung and 1 = with lung) by the model, which is in contrast with reference as indicated by red circle in each case. Their slice thickness and spacing is 0.625 mm. It is challenging to the

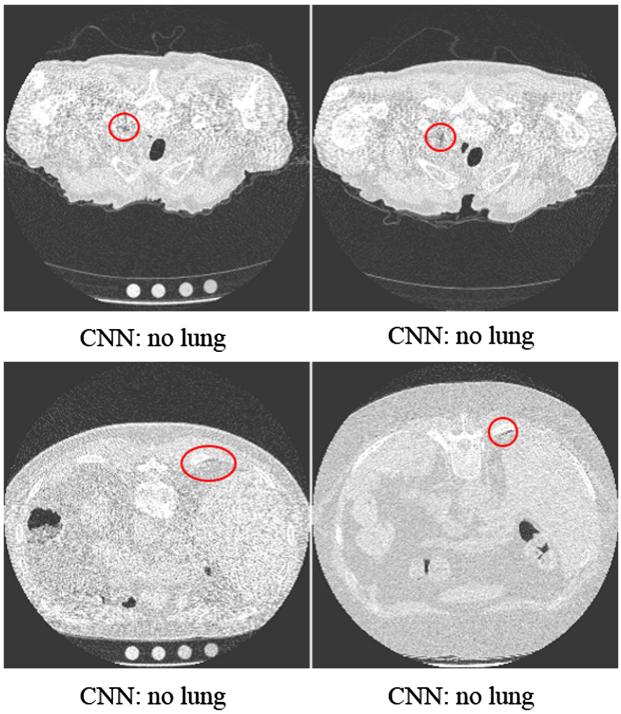


Fig. 13 Examples of misclassified slices from lung apex and base by CNN.

model since those slices only contain very small area of lung, especially when the image is noisy. However, such misclassified 1 or 2 slice is unlikely to impact the overall lung coverage inference unless an incomplete scan happens to start and end with that thin slice. On the other hand, even they are identified with complete lung coverage, it will not significantly affect the subsequent quantitative image analysis, e.g., lung segmentation and fibrosis scoring.

4.4 *Contrast Usage*

Since our ground truth does not cover the whole aorta, we evaluated the aorta segmentation based on 500 slices and achieved a dice coefficient of 0.938 ± 0.12. In terms of contrast detection, the system achieved an accuracy of 100% on the test set of 500 scans using 80 HU as threshold. Examples of aorta segmentation on contrast and noncontrast scan are shown in Figs. 14 and 15.

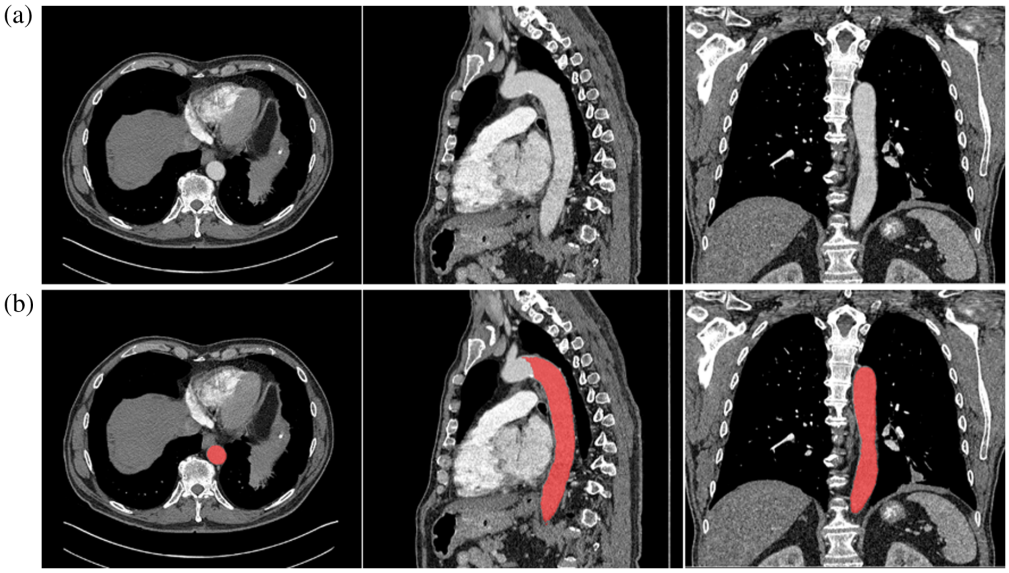


Fig. 14 Aorta segmentation on a contrast scan and overlay image demonstrated by axial, sagittal, and coronal views. (a) Raw image and (b) overlay with segmentation.

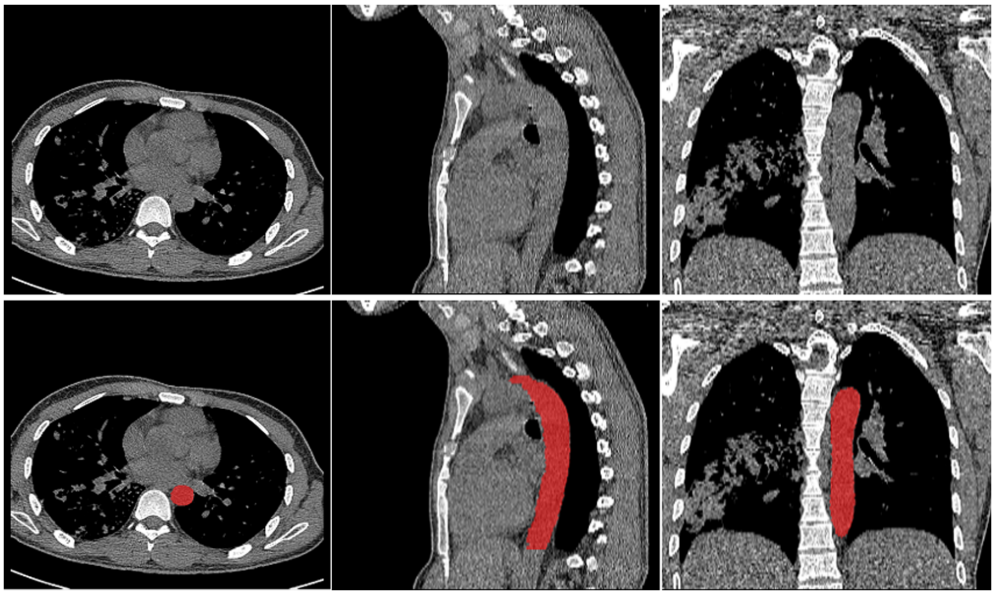


Fig. 15 Aorta segmentation on a noncontrast scan by the model.

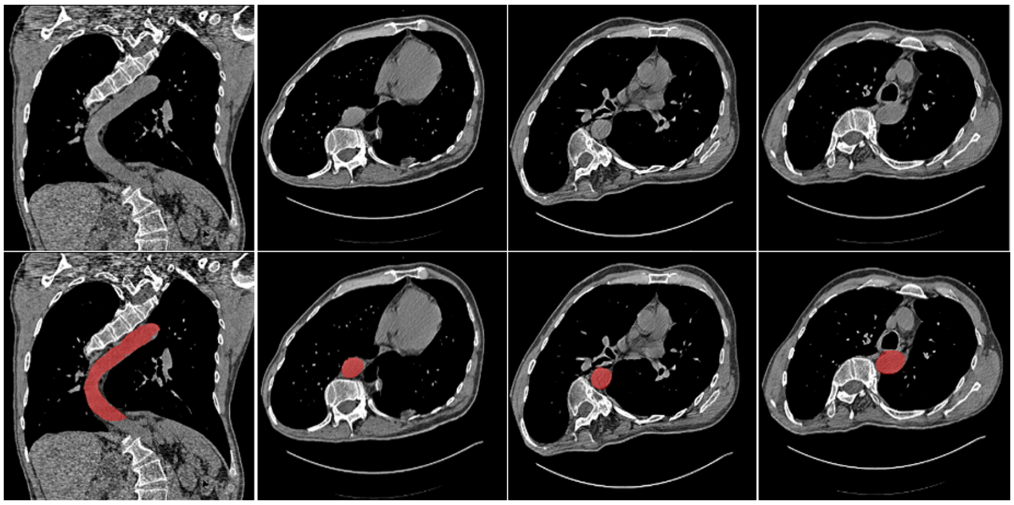


Fig. 16 Segmentation on a noncontrast scan from LOLA 11 with warped aorta.

Figure 16 shows a noncontrast scan from LOLA11[52] with a warped aorta that was successfully segmented by the model.

4.5 *Breath-Hold Types*

1000 slices were used in trachea segmentation evaluation and achieved dice coefficient of 0.948 ± 0.08. Figure 17 shows an example of trachea segmentation on a scan with spacing of 1 mm.

Figure 18 shows trachea segmentation applied to a challenging case from LOLA11 with tilted body.

After trachea segmentation, the carina is localized by searching slice by slice until bifurcation. Using the label from previous scanning direction model, the search direction can be corrected if a scan is feet-first. Figure 19 shows the identification of carina based on mask images.

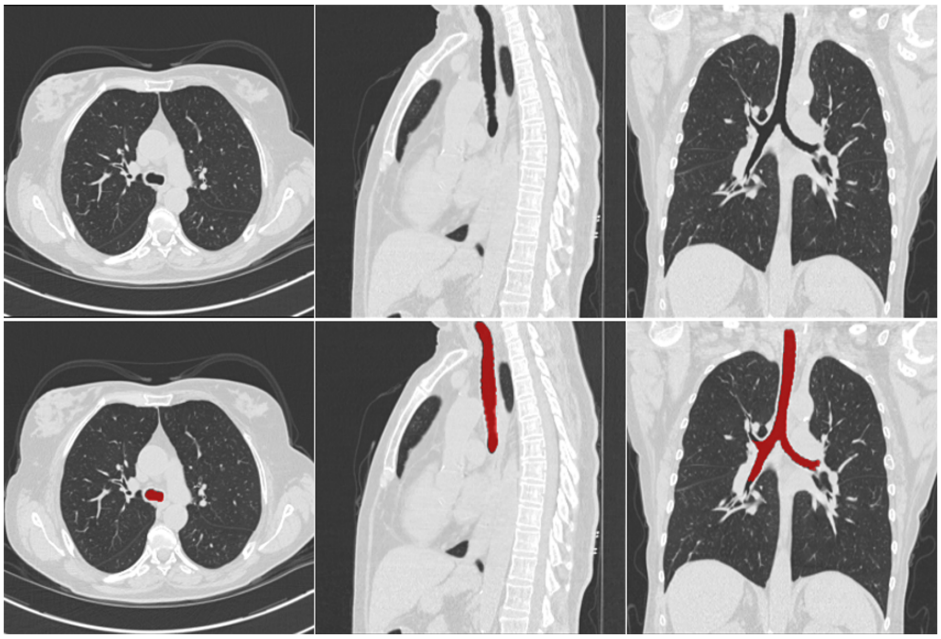


Fig. 17 Example trachea segmentation by the model.

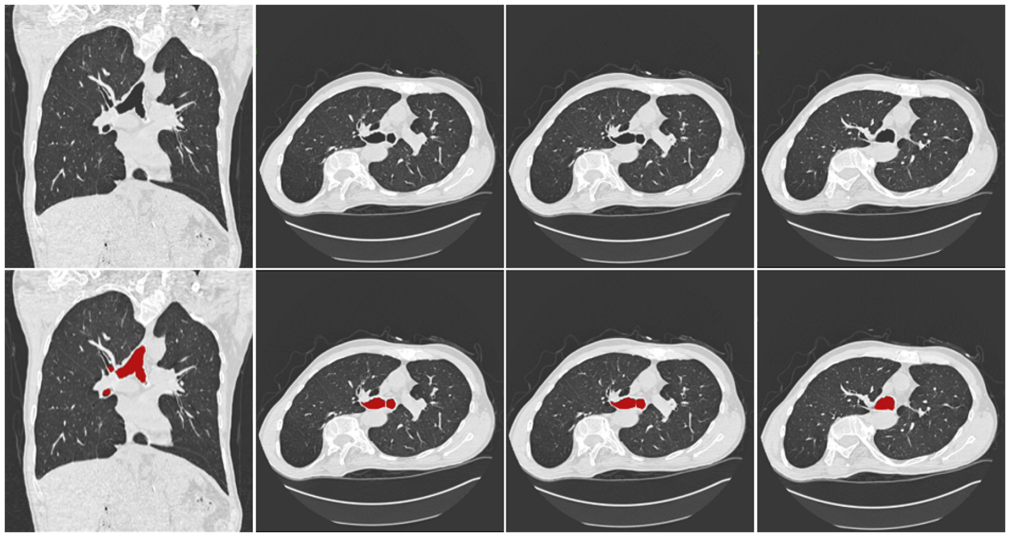


Fig. 18 Segmentation of trachea on tilted chest scan from LOLA11.

Fig. 19 Process to localize the carina through searching slice by slice.

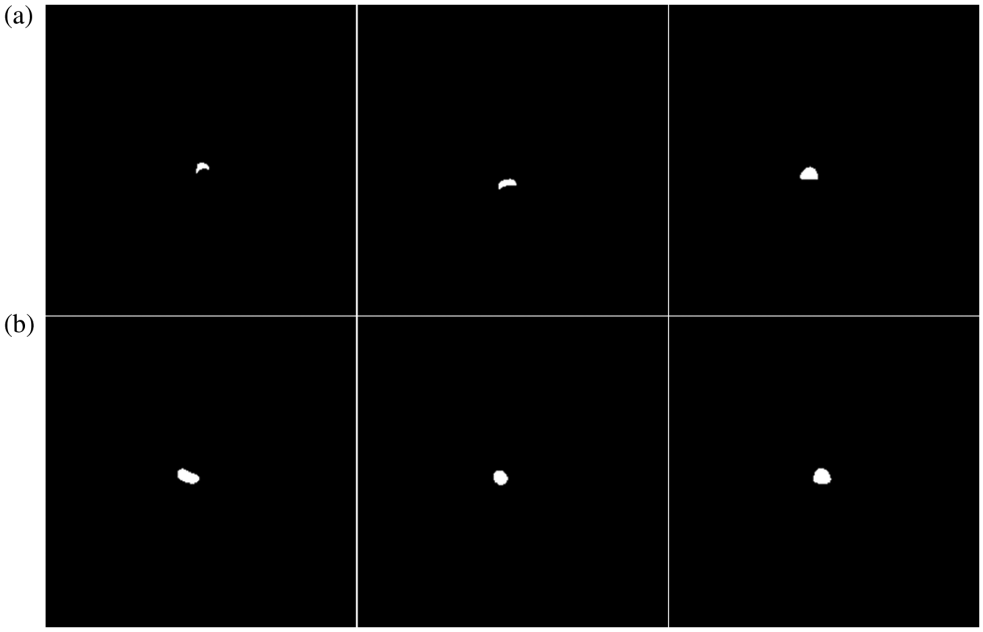


Fig. 20 Classified mask images of (a) RV and (b) TLC scans.

The breath-hold classification based on the shape of carina achieved an accuracy of 99.3% on the test set. Example classified RV and TLC masks are shown in Fig. 20. The first row contains three example masks of carina from RV scans and the second row contains three masks of carina at TLC.

5 Discussions

The labeling tasks were able to be posed as classification problems and deep learning was shown to be an ideal method to solve them. For individual labeling tasks, images of different view planes were exploited to achieve both reliability and efficiency. This helped to mitigate the difficulty of annotating a large training set. Specifically for contrast usage and breath-hold detection, identification of anatomic landmarks (descending aorta and carina, respectively), allowed us to tackle the problem using segmentation followed by classification, rather than by applying deep learning to the entire image. As such, the contrast detection accuracy is largely dependent on how accurate the aorta segmentation is and trachea/carina segmentation is critical for breath-hold identification. Although a small annotated training set was used for aorta segmentation, it demonstrated good performance on both contrast and noncontrast images. Data augmentation techniques mentioned before were very helpful in preventing overfitting and it demonstrated feasibility to build a robust model with limited annotation in medical imaging.

We did not assess 3-D quantitative segmentation with a large test set because we our ultimate goal is not to segment the whole aorta or airway tree perfectly. In practice, annotating the entire target for a large data set is very challenging and not necessary in this setting.

All models built with the SE-DenseNet121 architecture are generalizable and robust, working on incremental (large slice spacing, e.g., >10 mm) and volumetric (high-resolution) scans, and various lung pathologies. We used to be concerned about the scan direction and posture model when applied to low-resolution scans since they were built using coronal and sagittal images, respectively, and all training samples were high resolution (spacing < 3 mm). However, those two models demonstrated reliable performance even when the slice spacing was 20 mm.

We observed that the aorta and trachea segmentation were robust even in very noisy scans although they were trained only using diagnostic scans ($\sim$15 mGy). Figures 21 and 22 show application of aorta and trachea segmentation on an ultralow-dose scan ($\sim$0.2 mGy, 1/75 of training set dose) simulated by a CT reconstruction pipeline[53] in our group.

Such robustness is not unique to segmentation model. Fig. 23 shows example application of scan direction and scan posture model to ultralow-dose scans. Figure 24 corresponds to results of lung coverage identification.

There are a few potential reasons for the models' immunity to noise. First, the features extracted by the deep network are very rigorous to noise perturbation. Second, the diversity of the training set contributes to robustness of models, including various slice thicknesses, reconstruction kernels, and cohorts from various trials. Finally, the extra noise introduced does not change the original HU range of different tissues. Some limitations of this labeling pipeline

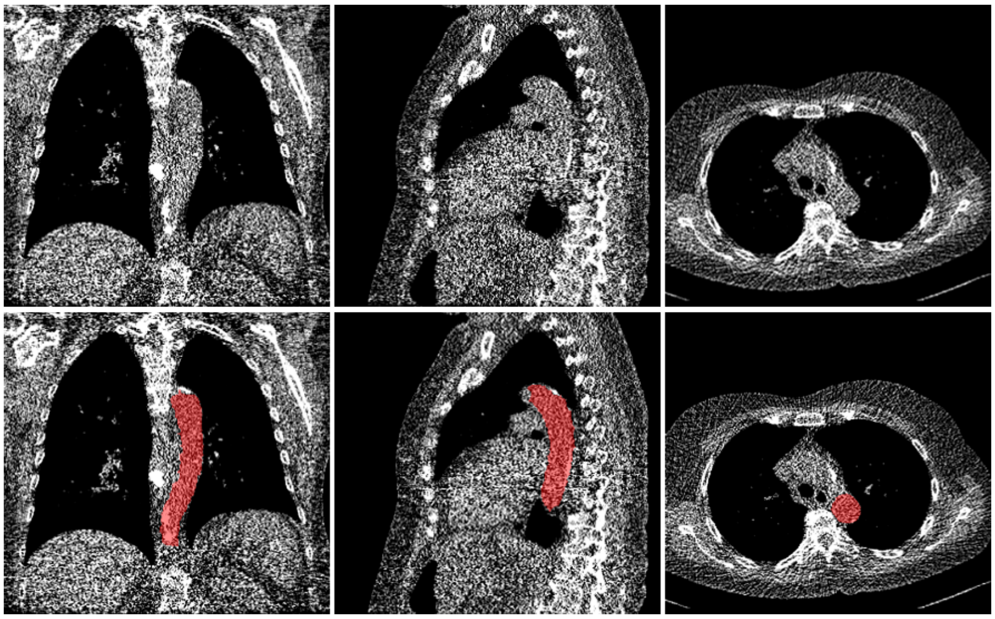


Fig. 21 Aorta segmentation on an ultralow-dose scan.

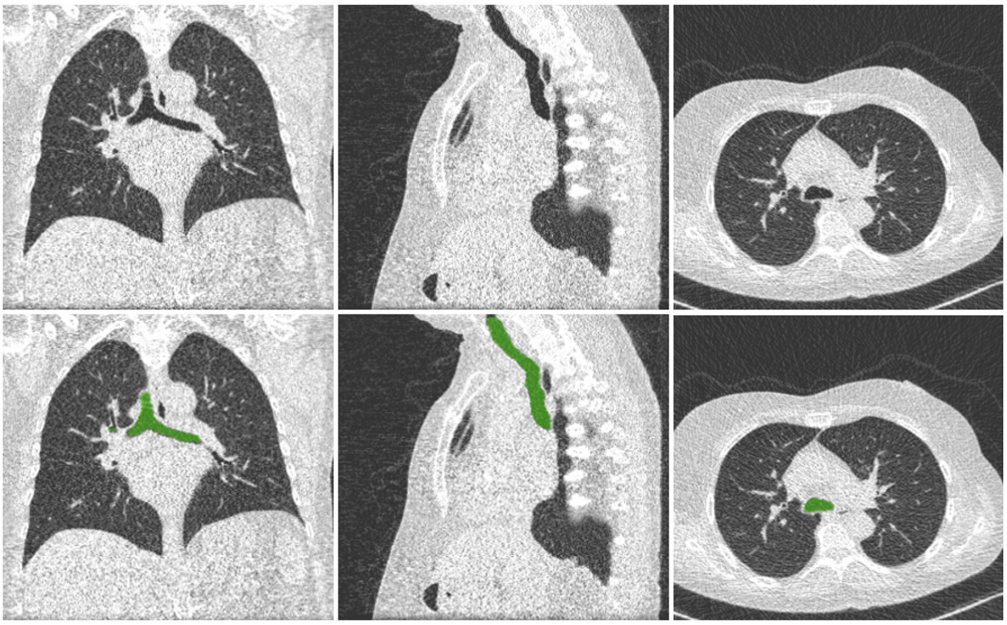


Fig. 22 Trachea segmentation on an ultralow-dose scan.

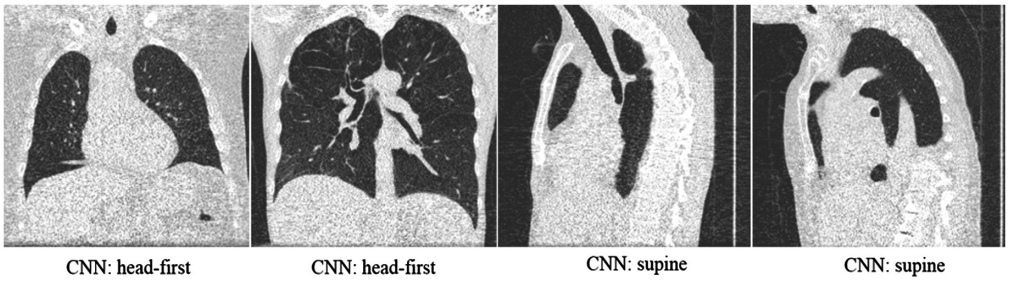


Fig. 23 Scan direction and posture detection on ultralow-dose cases.

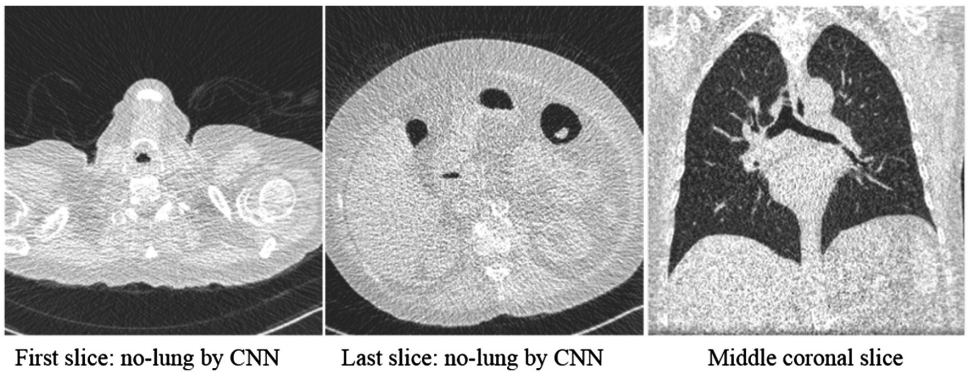


Fig. 24 Lung coverage identification on an ultralow-dose scan.

should be noted: (1) It is built for chest CT, not for other modality, e.g., MR and PET. (2) The proposed contrast detection method is based on the segmentation of aorta. As such, it is not applicable to other vascular CT studies aimed contrast identification in pulmonary arteries. (3) The identification of breath-hold is dependent on the shape of carina that is acquired from trachea segmentation. For extreme cases with one lung resected, it is not possible to find the carina.

Automatic labeling enables further quantitative image analysis appropriate for the type of scan identified. For example, we and others have developed deep learning models for lung

and lobe segmentation,[54] quantitative analysis of emphysema,[55] lung fibrosis,[56] and nodule detection.[57] Future studies may involve testing the utility of the automated labeling as a driver for such analysis systems.

6 Conclusion

The proposed automatic chest CT labeling pipeline was successfully applied to identify scanning direction, scanning posture, lung coverage completeness, contrast usage, and breath-hold types. Labeling accuracy from the classification models is sufficient for use in big data mining and high throughput processing.

Disclosures

There is no conflict of interest to disclose.

Acknowledgments

The authors acknowledge the support of MedQIA, LLC in collaboration and data sharing, and Ms. Eloisa Rodriguez-Mena for the article preparation and formatting.

References

1. B. Krishnankutty et al., "Data management in clinical research: an overview," *Indian J. Pharmacol.* **44**(2), 168–172 (2012).
2. J. R. Gaddale, "Clinical data acquisition standards harmonization importance and benefits in clinical data management," *Perspect. Clin. Res.* **6**(4), 179–183 (2015).
3. L. Houston et al., "Exploring data quality management within clinical trials," *Appl. Clin. Inf.* **9**(1), 72–81 (2018).
4. C. Ducar et al., "Benefits of a comprehensive quality program for cryopreserved PBMC covering 28 clinical trials sites utilizing an integrated, analytical web-based portal," *J. Immunol. Methods* **409**, 9–20 (2014).
5. P. M. Kuzmak and R. E. Dayhoff, "Minimizing digital imaging and communications in medicine (DICOM) modality worklist patient/study selection errors," *J. Digital Imaging* **14**(2 Suppl. 1), 153–157 (2001).
6. M. O. Gueld et al., "Quality of DICOM header information for image categorization," *Proc. SPIE* **4685**, 280–287 (2002).
7. P. Mongkolwat et al., "Validating DICOM content in a remote storage model," *J. Digital Imaging* **18**(1), 37–41 (2005).
8. A. G. Selvarani and S. Annadurai, "Medical image retrieval by combining low level features and DICOM features," in *Int. Conf. Comput. Intell. and Multimedia Appl.*, Vol. 1, pp. 587–589 (2007).
9. D. S. Lituiev et al., "Automatic labeling of special diagnostic mammography views from images and DICOM headers," *J. Digital Imaging* **32**, 228–233 (2019).
10. X. Wang et al., "A machine learning approach for classification of anatomical coverage in CT," *Proc. SPIE* **9785**, 97853P (2016).
11. N. Sharma et al., "Segmentation and classification of medical images using texture-primitive features: application of BAM-type artificial neural network," *J. Med. Phys.* **33**(3), 119–126 (2008).
12. J. Hong et al., "Non-Euclidean classification of medically imaged objects via s-reps," *Med. Image Anal.* **31**, 37–45 (2016).
13. E. Miranda, M. Aryuni, and E. Irwansyah, "A survey of medical image classification techniques," in *Int. Conf. Inf. Manage. and Technol.*, pp. 56–61 (2016).
14. J. Antony et al., "Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks," in *23rd Int. Conf. Pattern Recognit.* (2016).
15. E. Kim, M. Corte-Real, and Z. Baloch, "A deep semantic mobile application for thyroid cytopathology," *Proc. SPIE* **9789**, 97890A (2016).

16. H. Il Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," *Lect. Notes Comput. Sci.* **8150**, 583–590 (2013).
17. S. M. Plis et al., "Deep learning for neuroimaging: a validation study," *Front. Neurosci.* **8**, 229 (2014).
18. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
19. Ö. Çiçek et al., "3D U-net: learning dense volumetric segmentation from sparse annotation," *Lect. Notes Comput. Sci.* **9901**, 424–432 (2016).
20. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: fully convolutional neural networks for volumetric medical image segmentation," in *Fourth Int. Conf. 3D Vision*, pp. 1–11 (2016).
21. A. Kalinovsky, A. Kalinovsky, and V. Kovalev, "Lung image segmentation using deep learning methods and convolutional neural networks," in *XIII Int. Conf. Pattern Recognit. Inf. Process. (PRIP-2016)* (2016).
22. V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017).
23. H. R. Roth et al., "Anatomy-specific classification of medical images using deep convolutional nets," in *IEEE 12th Int. Symp. Biomed. Imaging*, pp. 101–104 (2015).
24. H. Sugimori, "Classification of computed tomography images in different slice positions using deep learning," *J. Healthcare Eng.* **2018**, 1–9 (2018).
25. Z. Yan et al., "Bodypart recognition using multi-stage deep learning," *Lect. Notes Comput. Sci.* **9123**, 449–461 (2015).
26. H. Bertrand, M. Perrot, and R. Ardon, "Classification of MRI data using deep learning and Gaussian process-based model selection," in *IEEE 14th Int. Symp. Biomed. Imaging* (2017).
27. B. D. de Vos et al., "2D image classification for 3D anatomy localization: employing deep convolutional neural networks," *Proc. SPIE* **9784**, 97841Y (2016).
28. G. E. Humpire-Mamani et al., "Efficient organ localization using multi-label convolutional neural networks in thorax-abdomen CT scans," *Phys. Med. Biol.* **63**(8), 085003 (2018).
29. A. D. Weston et al., "Automated abdominal segmentation of CT scans for body composition analysis using deep learning," *Radiology* **290**(3), 181432 (2018).
30. I. Sluimer et al., "Computer analysis of computed tomography scans of the lung: a survey," *IEEE Trans. Med. Imaging* **25**(4), 385–405 (2006).
31. M. Haas, B. Hamm, and S. M. Niehues, "Automated lung volumetry from routine thoracic CT scans," *Acad. Radiol.* **21**(5), 633–638 (2014).
32. A. El-Baz et al., "Lung imaging data analysis," *Int. J. Biomed. Imaging* **2013**, 1–2 (2013).
33. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, Salt Lake, Utah, pp. 7132–7141 (2018).
34. J. A. Verschakelen et al., "Differences in CT density between dependent and nondependent portions of the lung: influence of lung volume," *AJR Am. J. Roentgenol.* **161**(4), 713–717 (1993).
35. J. R. Mayo, "CT evaluation of diffuse infiltrative lung disease," *J. Thorac. Imaging* **24**(4), 252–259 (2009).
36. J. A. Verschakelen, "The role of high-resolution computed tomography in the work-up of interstitial lung disease," *Curr. Opin. Pulm. Med.* **16**(5), 503–510 (2010).
37. "ACR LI-RADS," https://cortex.acr.org/RadsPreview/CaseView.aspx?CaseId=naC0fAazxPg= (accessed 14 February 2019).
38. R. Iezzi et al., "Tailoring protocols for chest CT applications: when and how?" *Diagn. Interventional Radiol.* **23**(6), 420–427 (2017).
39. R. Bibb, D. Eggbeer, and A. Paterson, *Medical Modelling: The Application of Advanced Design and Rapid Prototyping Techniques in Medicine*, 2nd ed., Elsevier, Amsterdam, Netherlands (2015).
40. T. B. Grydeland et al., "Quantitative CT measures of emphysema and airway wall thickness are related to DLCO," *Respir. Med.* **105**(3), 343–351 (2011).
41. D. M. Hansell et al., "CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: a position paper from the Fleischner society," *Lancet Respir. Med.* **3**(6), 483–496 (2015).

42. D. A. Lynch et al., "Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper," *Lancet Respir. Med.* **6**(2), 138–153 (2018).
43. M. Loeve et al., "Cystic fibrosis: are volumetric ultra-low-dose expiratory CT scans sufficient for monitoring related lung disease?" *Radiology* **253**(1), 223–229 (2009).
44. O. M. Mets et al., "Identification of chronic obstructive pulmonary disease in lung cancer screening computed tomographic scans," *JAMA* **306**(16), 1775–1781 (2011).
45. E. Exhibit et al., "What every radiologist should know about expiratory and in the prone position chest computed tomography" (2015).
46. M. Gaeta et al., "Expiratory CT scan in patients with normal inspiratory CT scan: a finding of obliterative bronchiolitis and other causes of bronchiolar obstruction," *Multidiscip. Respir. Med.* **8**(1), 44 (2013).
47. C. P. Hersh et al., "Paired inspiratory-expiratory chest CT scans to assess for small airways disease in COPD," *Respir. Res.* **14**(1), 42 (2013).
48. J. Park et al., "Inspiratory lung expansion in patients with interstitial lung disease: CT histogram analyses," *Sci. Rep.* **8**(1), 15265 (2018).
49. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd Int. Conf. Learn. Represent. (ICLR)*, Y. Bengio and Y. LeCun, Eds., San Diego, California, (2015).
50. F. Chollet, "Keras," GitHub (2015).
51. M. Abadi et al., "TensorFlow: a system for large-scale machine learning," in *12th USENIX Symp. Oper. Syst. Des. and Implementation*, pp. 265–283 (2016).
52. "LOLA11—Home," https://lola11.grand-challenge.org/ (accessed 5 February 2019).
53. J. Hoffman et al., "Technical Note: FreeCT_wFBP: a robust, efficient, open-source implementation of weighted filtered backprojection for helical, fan-beam CT," *Med. Phys.* **43**(3), 1411–1420 (2016).
54. X. Wang et al., "High throughput lung and lobar segmentation by 2D and 3D CNN on chest CT with diffuse lung disease," *Lect. Notes Comput. Sci.* **11040**, 202–214 (2018).
55. D. Chong et al., "Reproducibility of volume and densitometric measures of emphysema on repeat computed tomography with an interval of 1 week," *Eur. Radiol.* **22**(2), 287–294 (2012).
56. H. J. Kim et al., "Comparison of the quantitative CT imaging biomarkers of idiopathic pulmonary fibrosis at baseline and early change with an interval of 7 months," *Acad. Radiol.* **22**(1), 70–80 (2015).
57. M. S. Brown et al., "Toward clinically usable CAD for lung cancer screening with computed tomography," *Eur. Radiol.* **24**(11), 2719–2728 (2014).

Xiaoyong Wang is a PhD student of bioengineering at the University of California, Los Angeles. His research interests include computer vision and machine learning in the field of medical imaging.

Pangyu Teng received his PhD in bioengineering from the University of Illinois at Chicago in 2014. He is a staff scientist at the Center for Computer Vision and Imaging Biomarkers (CVIB), University of California, Los Angeles. His research interests include medical imaging analysis and processing, machine learning, and development of imaging biomarkers.

Ashley Ontiveros received her bachelor's degree in biology from the University of California, Los Angeles, in 2010. She is a research associate at the CVIB, University of California, Los Angeles.

Jonathan G. Goldin received his MD and PhD degrees from the Faculty of Medicine at the University of Cape Town. Currently, he is an executive chief of Clinical Care, a chief of Radiology Department, a professor of Radiology and Biomedical Physics Program at the University of California, Los Angeles, as well as a director of Santa Monica Multi-Specialty Radiology.

Matthew S. Brown received his PhD in computer science in 1997 from the University of New South Wales, Sydney, Australia. Currently, he is a director of the CVIB and a professor of radiological sciences at the University of California, Los Angeles.