Deep Learning-Based Computed Tomography Perfusion Mapping (DL-CTPM) for Pulmonary CT-to-Perfusion Translation

Ge Ren[1], Jiang Zhang[1], Tian Li[1], Hao Nan Xiao[1], Andy Lai Yin Cheung[1,2], Wai Yin Ho[2], Jing Qin[3], and Jing Cai[1]

[1] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong

[2] Department of Nuclear Medicine, Queen Mary Hospital, Hong Kong

[3] School of Nursing, The Hong Kong Polytechnic University, Hong Kong

Corresponding author:

Jing Cai, PhD

Department of Health Technology and Informatics,

The Hong Kong Polytechnic University

Hung Hom, Kowloon

Hong Kong

Tel: (852) 3400-8645

Email: jing.cai@polyu.edu.hk

Statistical analyses author:

Ge Ren, MSc

Department of Health Technology and Informatics,

The Hong Kong Polytechnic University

Hung Hom, Kowloon

Hong Kong

Tel: (852) 5749-9195

Email: ge2018.ren@connect.polyu.hk

Abstract

Purpose: To develop a deep learning-based computed tomography (CT) perfusion mapping (DL-CTPM) method that synthesizes lung perfusion images from CT images.

Methods and Materials: This paper presents a retrospective analysis of the pulmonary technetium-99m-labeled macro-aggregated albumin (MAA) single-photon emission computed tomography (SPECT)/CT scans obtained from 73 patients at Queen Mary Hospital in Hong Kong in 2019. The left and right lung scans were separated to double the size of the dataset to 146. A three-dimensional attention residual neural network (ARNN) was constructed to extract textural features from the CT images and reconstruct corresponding functional images. Eighty-four samples were randomly selected for training and cross-validation, and the remaining 62 were used for model testing in terms of voxel-wise agreement and function-wise concordance. To assess the voxel-wise agreement, the Spearman's correlation coefficient (R) and structural similarity index measure ($SSIM$) between the images predicted by the DL-CTPM and the corresponding SPECT perfusion images were computed to assess the statistical and perceptual image similarities, respectively. To assess the function-wise concordance, the Dice similarity coefficient (DSC) was computed to determine the similarity of the low/high functional lung volumes.

Results: The evaluation of the voxel-wise agreement showed a moderate-to-high voxel value correlation (0.6733 ± 0.1728) and high structural similarity (0.7635 ± 0.0697) between the SPECT and DL-CTPM predicted perfusions. The evaluation of the function-wise concordance obtained an average DSC value of 0.8183 ± 0.0752 for high-functional lungs, ranging from 0.5819 to 0.9255, and 0.6501 ± 0.1061 for low-functional lungs, ranging from 0.2405 to 0.8212. Ninety-four percent of the test cases demonstrated high concordance ($DSC > 0.7$) between the high functional volumes contoured from the predicted and ground-truth perfusions.

Conclusions: We developed a novel DL-CTPM method for estimating perfusion-based lung functional images from the CT domain using a 3D ARNN, which yielded moderate-to-high voxel-wise

approximations of lung perfusion. To further contextualize these results toward future clinical application, a multi-institutional large-cohort study is warranted.

Introduction

Lung-function imaging is used to measure the circulation of air (ventilation) and blood (perfusion) within the lung, and is commonly used to detect pulmonary emboli or abnormal blood flow [1]. Impaired lung function generally indicates the presence of respiratory diseases, such as pulmonary embolism, pulmonary hypertension, or chronic obstructive disease [2]. Lung function imaging is the primary imaging modality used for the diagnosis of these diseases [3]. The imaging of lung function is also used to guide functional lung avoidance radiation therapy (FLART) for lung cancer patients. FLART is an emerging technique that selectively minimizes dose delivery to the high-functional lung area and favors dose deposition to the low-functional lung area [4], based on the regional information obtained from pulmonary function imaging. Thus, the aim of FLART is to reduce radiation-induced lung complications and improve tumor control via dose escalation [5].

The majority of methods that are currently used for pulmonary function imaging are based on either nuclear medicine or deformable image registration (DIR). Nuclear medicine imaging methods identify lung function with the aid of imaging contrast agents. These methods include single-photon emission computed tomography (SPECT) with technetium-99m-labeled macroaggregated albumin (MAA) [5, 6], positron emission tomography (PET) with gallium-68 [7], magnetic resonance imaging (MRI) with hyperpolarized gas (helium-3 or xenon-129) [8, 9], and contrast-enhanced MRI or CT [10]. The recently developed DIR-based function imaging method uses surrogates of the regional pulmonary ventilation from lung images obtained from four-dimensional computed tomography (4D-CT), which are subjected to sophisticated image-processing and analysis [11-13].

Each pulmonary function imaging method has its merits and drawbacks. SPECT ventilation/perfusion is the most widely used method, but involves the use of ionizing radiation and suffers from relatively low spatial and temporal resolution. PET ventilation provides more quantitative information, but requires extra

ionizing radiation and a long imaging time. Hyperpolarized gas MRI ventilation produces exquisitely detailed images of pulmonary ventilation with superior spatial resolution, without the need for ionizing radiation. However, the limited availability of helium-3 gas and the need for specialized equipment hampers its clinical implementation. Contrast-enhanced MRI/CT perfusion requires the injection of imaging contrast agents into the patient, and its data-processing procedures can be complicated. 4D-CT-based ventilation is prone to error, due to its non-robust DIR algorithms [13]. Furthermore, additional scans are often required for all of these methods, which makes them resource-intensive, inconvenient, and technically challenging, thereby reducing the availability of personalized lung function imaging for diagnosis and treatment.

To overcome the drawbacks associated with the above-described pulmonary function imaging techniques, we synthesized functional images based on textural information obtained from anatomical CT images. Specifically, we developed a deep learning-based CT perfusion mapping (DL-CTPM) method for pulmonary CT-to-perfusion translation. In clinical practice, lung perfusion imaging is used as a clinical pulmonary function test to measure blood circulation within the lungs. It is routinely used to detect abnormal blood flow, to assist in the diagnosis of respiratory diseases characterized by impaired lung perfusion function [14]. Lung perfusion imaging can also guide intensity-modulated radiation therapy and effectively reduce the mean radiation dose deposited in the high-functional lung areas of locally advanced-stage non-small cell lung cancer patients [15]. Deep learning-based convolutional neural networks (DL-CNNs), also known as deep convolutional neural networks, are a class of machine learning algorithms that use several layers of convolutional processing units for feature extraction and signal transformation [16, 17]. This technique has gained increasing popularity in medical imaging research and has demonstrated satisfactory performance in many applications of medical imaging, such as low-dose CT image correction [18, 19], MR-to-CT image synthesis [20], and image segmentation [21, 22]. The results of our preliminary study also indicated the feasibility of deriving meaningful perfusion images from the CT domain [23] using deep neural network-based modeling. These applications suggest that a deep neural network may be used

to extract textural information from CT images, from which corresponding perfusion-based pulmonary functional images may be generated.

In this study, we developed a deep learning-based CT perfusion mapping (DL-CTPM) method for realizing pulmonary CT-to-perfusion translation. We constructed and trained a three-dimensional (3D) attention residual neural network (ARNN) to extract textural features and reconstruct perfusion images. This method was then evaluated using Spearman's correlation coefficient (R) and the structural similarity index measure ($SSIM$) to determine the voxel-wise agreement between predicted and ground-truth perfusion images. The function-wise concordance of the low- and high-functional lung regions was examined using the Dice similarity coefficient (DSC).

Methods and Materials

Image acquisition and study design

SPECT/CT images were collected from 75 patients who underwent lung MAA SPECT/CT scans for suspected lung diseases and treatment evaluation in 2019 at Queen Mary Hospital in Hong Kong. Two patients were excluded due to the incompleteness of their lung images. Among the 73 patients, 12 were diagnosed with lung cancer and four were diagnosed as having no disease. The remaining patients had different types of lung diseases, such as pulmonary hypertension, chronic pulmonary embolism, or pulmonary arteriovenous malformation. Table 1 lists details of the patient characteristics.

The effectiveness of our method was evaluated using a training–validation–testing scheme (Fig. 1). To ensure the reliability of the test performance in a small dataset, a relatively large percentage of the test set (~40%) was randomly selected from the entire patient cohort. The remaining patients were used as the

training set (~60%), and were randomly grouped into four folds for cross-validation. In each validation, we used three folds of training samples to derive the network weights, and the remaining fold for validation. The cross-validation results were used to tune the hyperparameters of the model. It was then fitted to the entire training set to afford the final model, and the performance was evaluated on the test set.

This study was approved by the Institutional Review Boards of the University of Hong Kong/Hospital Authority Hong Kong West Cluster. Patients were injected with 3 mCi (111 MBq) of the radiopharmaceutical MAA before imaging, and were then immobilized in a supine position with normal resting breathing during image acquisition. The 3D SPECT/CT scans were acquired across 360 degrees to cover the entire volume of the lung, using a GE Discovery 670 SPECT/CT scanner at a frame rate of 30 s/f and a total of 60 frames. Each CT image was reconstructed into a 512×512 matrix with 0.977×0.977 mm^2 pixel spacing and 1.25-mm slice spacing, and each SPECT image was reconstructed into a $128 \times 128 \times 128$ matrix with a voxel size of $4.42 \times 4.42 \times 4.42$ mm^3.

Image preprocessing and deep neural network

Figure 2(a) shows the flowchart of the image preprocessing procedures in this study. The SPECT images were registered to the CT coordinates using 3D Slicer software (10.4.2) [24]. Lung parenchyma segments were automatically generated from the CT images using the 3D Slicer extension Chest Imaging Platform (2019) [25]. To increase the size of the dataset and reduce the computational cost, the image of the left and right lungs was separated. To further accelerate the training process and reduce the computational memory cost, the separated parenchyma volumes were cropped to include only the lung, and then down-sampled into $128 \times 64 \times 64$-sized matrices. As convolutional neural networks (CNNs) receive and generate values ranging between 0 and 1, the down-sampled CT and SPECT images were standardized using CT enhancement and SPECT standardization, respectively, as described below.

CT enhancement: This is achieved through histogram equalization and the application of two filters. The histogram equalization has been demonstrated to be effective in displaying different regions in CT images in grayscale [26]. To enhance the contrast in the parenchymal region, the histograms of the down-sampled CT images were equalized to within the range of −1000 to −300 Hounsfield unit (HU) values. The use of filters has been proven to improve the correlation between lung CT and lung functional images [27]. Thus, a median filter with a kernel size of 10 voxels was subsequently applied to enhance the signal in low-function regions and reduce the degree of noise fluctuation. This was followed by the application of a uniform filter with a kernel size of five voxels to realize further noise reduction. These CT enhancement procedures, including histogram equalization and filtering, are summarized as follows (Equation (1)):

$$CT(x,y,z)' = filtering(\frac{chf[CT(x,y,z)] - chf_{-1000}}{chf_{-300} - chf_{-1000}}), \tag{1}$$

where $CT(x,y,z)$ is the original HU value at voxel (x,y,z). $CT(x,y,z)'$ indicates the enhanced CT image. The cumulative histogram function (chf) provides the cumulative count for a single HU bin. chf_{-1000} is the chf for a voxel intensity of −1000 and chf_{-300} is that for −300. Outlier voxel values were replaced with the threshold values. The term $filtering$ indicates the deployed filters.

SPECT standardization: A pulmonary SPECT value may be influenced by many factors that depend on a patient's condition, such as respiratory capacity, breathing frequency, and lung diseases. Thus, SPECT normalization is commonly used to reduce the effect of individual patient conditions during the inter-patient assessment of lung function [27-29]. In this study, a SPECT image was first normalized by the SPECT intensity value in the high intensity region (HIR). The normalized image was then used as ground truth for model evaluation. This image was subsequently discretized into 11 uniformly spaced values between 0 and 1 (0, 0.1, 0.2, … 1) for modeling. Image discretization decreases the noise-induced variance of the extracted features and accelerates modeling convergence. This approach has been previously adopted for PET

ventilation images to mitigate the effect of small noise fluctuations [30]. The SPECT standardization

procedures, including normalization and discretization, are formulated as follows (Equation (2)):

$$SPECT(x, y, z)' = discretize\left(\frac{SPECT(x, y, z)}{HIR}\right),$$
$$HIR = P_{75}(SPECT) \tag{2}$$

where $SPECT(x, y, z)$ is the original SPECT value at the voxel (x, y, z). $SPECT(x, y, z)'$ indicates the

standardized SPECT image. HIR is the perfusion value of the high intensity region. In this study, HIR was

set as the 75th percentile value for each SPECT ($P_{75}(SPECT)$), as this is close to the perfusion value of

normal lung tissue [31]. The term $discretize$ indicates the deployed discretization.

Proposed CNN architecture: CT-to-perfusion translation requires a full exploration of high-level features

in 3D images, which can only be achieved by a 3D CNN [32]. However, the training process for a

conventional 3D CNN requires significant memory and computational power. As such, we used an ARNN

with a reduced number of parameters to realize CT-to-perfusion translation. This model uses a 3D

encoding–decoding structure to capture the hierarchical textural features of the input, as shown in Figure

2(b). The processed CT image as input is propagated through the network along all possible paths from left

to right. In the end, the perfusion image is generated as output. The maximum number of channels is set to

128 to reduce the computational memory cost. Like other CNN models, ARNN comprises a number of

different modules. Two skip attention (Skip Att) modules similar to the attention U-Net [33] are used for

target structure attention and original shape recovery [34]. Similar to ResNet, the residual block [35]

(ResConv) is used to speed up the convergence. Strided convolution (StriConv) and transposed convolution

(ConvT) are used to halve the size of the feature map (S1 to S4) and recover the size of the feature map,

respectively. To increase the size of the receptive field of the 3D architecture, we used three $2 \times 2 \times 2$

strided convolution (StriConv) modules and set all convolutional filters with a size of $5 \times 5 \times 5$. The 5×5

$\times 5$-sized filter increases the volume of the receptive field by a factor of approximately $5^3/3^3$ (4.6) compared

to that of commonly used $3 \times 3 \times 3$ convolution filters. An ROI attention module (ROI Att) was also designed for foreground region attention (Fig. 2(c)). To obtain the ROI attention feature map, the input feature map is convoluted to increase the number of feature channels to four, followed by thresholding, whereby each voxel is set to 1 if it is larger than the drop threshold 0. After a subsequent sigmoid layer, this binary mask is applied to the convoluted feature map by spatial-wise multiplication to obtain the attention feature map. In the last layer, a sigmoid function (Sigmoid) sums the results of the previous layers and scales the predicted values to a range of [0,1]. Each convolution is followed by batch normalization, a parametric rectified linear unit with a rectification parameter of 0.2, and a dropout layer with a dropout rate of 0.1.

The proposed method learns the optimal parameter values by minimizing the loss function between the ground truth y and the network output p. We used binary cross-entropy (BCE) loss as the loss function to increase the sensitivity to the small low-functional regions, as follows (Equation (3)):

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \cdot \log(p_i) + (1 - y_i) \cdot log(1 - p_i)], \tag{3}$$

where p_i and y_i denote the predicted and ground-truth perfusion values at voxel i, respectively. N denotes the total number of non-zero voxels. Each layer is updated using error backpropagation with an adaptive moment estimation optimizer. The number of training epochs was set as 500, where the loss function changed less than 0.2% for continued five epochs in the pilot study. We implemented our network using the PyTorch 1.1 framework. All of the preprocessing steps were coded in Python. All of the experiments were performed using a workstation with central-processing Intel Core i7-8700 @ 3.2GHz, a graphics-processing unit NVIDIA GTX 2080 TI with 11 GB of memory and 32 GB of RAM.

Performance evaluation

To evaluate the performance of the proposed DL-CTPM method, we used Spearman's correlation coefficient (R) to assess the intensity correlation (Equation (4)) and the structural similarity index measure

(*SSIM*) to determine the perceptual similarity (Equation 5). The generated DL-CTPM images were up-sampled to the original SPECT resolution and compared with the normalized SPECT perfusion. R is defined by the following equation (Equation (4)):

$$R = \frac{\sum_{i=1}^{N}[(y_i - \bar{y}) \cdot (p_i - \bar{p})]}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(p_i - \bar{p})^2}}, \tag{4}$$

where $\bar{p}$, $\bar{y}$, p_i, and y_i denote the average value and value at voxel i for the predicted and ground-truth perfusions, respectively. N denotes the total number of non-zero voxels.

The *SSIM* contains three terms that compare three measurements of the generated perfusion and ground truth: the luminance term ($L(y, p)$), the contrast term ($C(y, p)$), and the structural term ($S(y, p)$) [36]. The overall *SSIM* is a multiplicative combination of these terms, as follows (Equation (5)):

$$SSIM = L(y, p) \cdot C(y, p) \cdot S(y, p) = \frac{2\mu_y\mu_p + C_1}{\mu_y{}^2 + \mu_p{}^2 + C_1} \cdot \frac{2\sigma_y\sigma_p + C_2}{\sigma_y^2 + \sigma_p^2 + C_2} \cdot \frac{2\sigma_{yp} + C_2}{2\sigma_y\sigma_p + C_2}, \tag{5}$$

where μ_y, μ_p, σ_y, σ_p, and σ_{yp} are the respective local means, standard deviations, and cross-variance for the ground-truth image y and predicted image p. $C_1 = (k_1 L)^2$, $C_2 = (k_2 L)^2$ are two variables that stabilize the divisions when the denominators are too small. L is the dynamic range of the voxel values. k_1 and k_2 are scalar constants, which are by default set to 0.01 and 0.03, respectively.

Function-wise concordance evaluation

To validate the potential applicability of DL-CTPM in FLART, we examined the function-wise concordance between the ground-truth SPECT images and the DL-CTPM generated images. The synthesized SPECT perfusion image was directly compared with the normalized original SPECT perfusion

image (normalization process was described in SPECT standardization section). An image intensity threshold of 0.66 was then used to separate the high-functional lung volumes from the low-functional lung volumes. Similar approaches were used in a lung ventilation study [37], and were suggested for use in FLART treatment-plan optimization [38]. Subsequently, we calculated the dice similarity coefficient (*DSC*) for the low- and high-functional volumes obtained using the threshold value between the predicted and ground-truth perfusion images, as follows (Equation (6)):

$$DSC = \frac{2 * |p \cap y|}{|p| + |y|},$$

(6)

where p is the low- and high-functional volume in the predicted perfusion images, and y is the corresponding volume in the ground-truth SPECT perfusion images. The overall concordance is inferred as the mean *DSC* value of all the test cases.

Rationale for CT-to-perfusion translation

The HU value is a function of the air/tissue ratio, which determines the attenuation coefficient in a specified location. Specific to the lung parenchyma, the fractional air/tissue can serve as a surrogate of the blood–gas exchange rate [11]. Therefore, pulmonary diseases that involve abnormal blood–gas exchange generally manifest as textural alterations of the lung parenchyma. As shown in Figure 3, the low-functional region (blue contour) contoured on the SPECT perfusion cannot be observed on the grayscale CT image. After adjusting the window width and visualizing the CT image in a dense rainbow colormap, different patterns can be observed between the low- and high-functional regions. These patterns could hardly be visualized in grayscale because of the small intensity difference within each pattern and the limited grayscale differentiability by human eyes. Another study also shows that CT images can provide pulmonary functional information in the form of distinctive textures, such as reticulation, honeycombing, ground-glass opacity, consolidation, and micronodule patterns [39]. Based on these textural features, a classification

model of interstitial lung diseases achieved a high accuracy rate of 85.5% using a six-layer neural network [40]. The high predictive power of CT textures further supports the feasibility of our CT-to-function translation.

Results

Figures 4(a) and 4(b) show the evolution of the R and $SSIM$ evaluation metrics, respectively, with an increasing number of epochs in the four validation folds and test set. For both the R and $SSIM$ metrics, the model performance on the test set is within the variance range of the validation groups. The respective performances of the final model are presented in the box plots in Figures 4(c) and 4(d). For the validation folds, the DL-CTPM perfusion images show a moderate-to-high correlation ($R = 0.6906 \pm 0.1425$) and high structural similarity ($SSIM = 0.7704 \pm 0.0703$) with the corresponding reference perfusion images. The testing performances are similar to those for the validation folds, achieving an R value of 0.6733 ± 0.1728 and an $SSIM$ value of 0.7635 ± 0.0697. Of the test cases, 85% have correlations greater than 0.5, and 56% have correlations greater than 0.7. For both R and $SSIM$, the results of an unpaired two-tailed t-test showed no statistically significant difference ($p = 0.2212$ for R, $p = 0.5037$ for $SSIM$) in their validation and test results. Of the 62 cases in the test set, we observed only three outliers and one outlier (outside the 1.5 interquartile range) for R and $SSIM$, respectively.

Figure 5 shows three representative testing cases, categorized as low-performance, average-performance, and high-performance. For the low-performance case, the low-functional region in the middle of the lung was predicted to be a high-functional region in the corresponding DL-CTPM predicted perfusion image. The low-functional regions indicated by the arrows were successfully predicted in the average-performance and high-performance cases.

Figure 6 shows the *DSC* distributions for low- and high-functional lung volumes in the test set. Of the 62 test cases, 98% have a high-functional lung *DSC* value greater than 0.6, and 71% for low-functional lung. 94% showed high concordance (*DSC* > 0.7) between the high-functional lung volumes contoured from the predicted and ground-truth perfusion images, with 35% showing high concordance for the low-functional lung volumes. The mean *DSC* value for high-functional lungs was 0.8183 ± 0.0752, ranging from 0.5819 to 0.9255; the mean *DSC* value for low-functional lungs was 0.6501 ± 0.1061, ranging from 0.2405 to 0.8212. Most testing cases demonstrated high concordances (DSC > 0.7) for high-functional volumes (94%), determined from the generated and ground-truth perfusion images.

Discussion

In this study, we successfully generated perfusion-based functional lung images from the CT domain using a 3D ARNN. To the best of our knowledge, this is the first application of CT-to-perfusion translation using a deep-learning approach. A majority of existing papers present CT-based function imaging in the form of ventilation. For instance, a multi-institution study on CT ventilation imaging showed a median correlation of 0.49 (range: 0.27–0.73) [12]. In our test set, the perfusion images generated using our DL-CTPM method achieved a moderate-to-high correlation (0.6733 ± 0.1728) with the ground-truth perfusion images. In terms of function-wise concordance, our findings revealed that 94% of test cases had high-functional lung *DSC* values of >0.7, and 71% had low-functional lung *DSC* values of >0.6. This suggests that the functional lung volume determined by the DL-CTPM method was in good agreement with those shown in the SPECT perfusion images. Here, we note that although our results were based on lung perfusion rather than ventilation, both studies provide insights and have implications for radiotherapy, especially as both ventilation and perfusion imaging are correlated with the radiotherapy side-effects, such as radiation pneumonitis [41]. The results from this novel study are encouraging and may serve as a stepping stone toward the broad clinical application of synthesized functional lung images.

The evolution of the evaluation metrics R (Figure 4(a)) and $SSIM$ (Figure 4(b)) along the training epochs are consistent in the four validation cohorts and one test set. This suggests that the DL-CTPM method was highly robust for our dataset. We also observed consistently higher R and $SSIM$ values for the second validation fold, regardless of the number of training epochs. This may be due to random effects arising from the small sample size (n = 84) and the wide range of lung abnormalities (Table 1). The similar performances of the test set and validation cohorts, as shown by the similar bar levels in Figure 4(c) and 4(d), indicate the minimum overfitting and high internal generalizability of the trained model. The values of the evaluation metric R have overall higher variances than those of the $SSIM$ throughout the training, which suggests that R is more sensitive to changes in the performance of the generated perfusion image.

Distinctive SPECT image characteristics have been observed for several lung diseases. For instance, the original SPECT perfusion values of patients with pulmonary hypertension (ranging from 1,000 to 12,000) were much higher than those of patients with other lung conditions (typically less than 5,000). This abnormal distribution could diminish the learning ability of the ARNN model. To assess whether this had led to the degradation of model performance, we performed an additional experiment in which we excluded all patients with pulmonary hypertension. The results showed an increase (~4%) in R to 0.6983 ± 0.1392 and a negligible reduction (~1%) in $SSIM$ to 0.7462 ± 0.071. This suggests that the heterogeneous distribution of lung abnormalities among patients had a mild effect on the CT-to-perfusion translation.

During the development of our methodologies, we continually optimized the preprocessing procedures until the best performance was achieved. For CT images, histogram equalization and filtering were applied to enhance the textural difference. The recommended kernel sizes of the median and uniform filters for ARNN modeling were $10 \times 10 \times 10$ and $5 \times 5 \times 5$, respectively. For SPECT images, discretization and normalization were used to increase model robustness and generalizability. In this study, each SPECT image was discretized into 11 uniformly spaced bins to preserve useful information during the removal of noise signals. For SPECT normalization, different upper thresholds have been reported [11, 27-29]. We

adopted a relatively low normalization threshold (75 percentile) to reduce the learned features from high-functional lung, due to large volumes of high-intensity regions in the SPECT images. Further investigation of the effects of these preprocessing parameters on model performance is warranted.

The computational cost of 3D CNNs and data scarcity are commonly encountered challenges in deep learning research. Although 3D CNNs can fully exploit 3-dimensional features, this incurs computational demand due to the significantly increased parameters. This leads to prolonged training time and reduced learning ability [32]. To overcome these challenges, we designed an ARNN with a reduced number of parameters. However, an inadequate number of training samples still may incur a higher risk of model overfitting. Thus, we separated the left and right lungs to double the size of the dataset. This separation also reduced computational memory cost. In view of the anatomical difference between the left and right lungs, such as their respective number of lobes, we conducted a two-sample paired t-test to examine whether their anatomical differences affected the CT-to-perfusion translation. There was no statistically significant difference between the left and right lungs ($p = 0.876$ for R, $p = 0.4047$ for *SSIM*). This suggests that the performance of the ARNN is not influenced by anatomical differences, as CNNs mainly capture texture-based features [42-44] or shape-based features [45-47]. However, there are physiological correlations between the left and right lungs that are difficult to capture by CNN models. Such correlations should be considered in future clinical applications, such as FLART for lung cancer patients.

Despite the encouraging findings presented above, our study has several limitations. First, several low-performance test cases had significantly lower voxel-wise concordance than others. In the low-performance case shown in Figure 5, the low-functional region in the middle of the lung was predicted to be a high-functional region, which may be partly attributable to the low occurrence and small volume of the low-functional regions. Both reasons reduced the effects of low-functional regions on the network loss function. To enhance the overall accuracy and utility of the deep-learning model, further investigation into these low-performance cases is warranted.

Second, although SPECT is generally favored for diagnosis and radiation treatment assessment [3, 14], it does not quantify the concentration of radioactivity in absolute units [27], due to factors such as attenuation, scatter, and partial volume effects. This leads to uncertainties regarding the actual perfusion conditions, which increases the heterogeneity of the values obtained for patients. Besides, the perfusion condition is only one of the indicators for assessing lung function. Future studies that translate CT images into both pulmonary perfusion and ventilation information may be able to enhance the value of the deep-learning model for assessing overall lung function.

Third, the potential applicability of DL-CTPM in FLART was only evaluated with respect to structural concordance with real SPECT images. Direct contextualization of the synthesized SPECT for application to FLART requires validations of the dosimetry and clinical outcomes. Nevertheless, our findings highlight the potential of CT-to-perfusion translation and will serve as a valuable stepping-stone toward the clinical application of FLART, without the need for additional function imaging. We encourage researchers in the field to conduct a large-cohort study specifically on lung cancer patients, and incorporate information regarding dosimetry and radiation-induced toxicities whenever possible, to fully contextualize the results of our DL-CTPM approach in the context of FLART application.

Last, similar to other deep learning studies, the generalizability of our model may have been influenced by heterogeneities in the scanners (CT and SPECT), imaging protocols (such as image acquisition and reconstruction settings), and image preprocessing steps (such as normalization and discretization). To achieve broader model deployment, a standardized protocol and/or a multi-institutional study is warranted to ensure model generalizability across various clinical settings.

Conclusion

In this study, we developed a novel DL-CTPM method for estimating perfusion-based lung functional images from the CT domain using a 3D ARNN. Statistical and structural similarity analyses showed that the results provided a moderate-to-high voxel-wise approximation of lung function mapping. To further contextualize these results for future clinical application, a multi-institutional large-cohort study is warranted.

References

1. Elojeimy S, Cruite I, Bowen S, et al. Overview of the novel and improved pulmonary ventilation-perfusion imaging applications in the era of SPECT/CT. *Am J Roentgenol* 2016;207:1307-1315.

2. Mistry NN, Pollaro J, Song JY, et al. Pulmonary perfusion imaging in the rodent lung using dynamic contrast-enhanced MRI. *Magnet Reson Med* 2008;59:289-297.

3. Rieger B, Zimmer F, Zapp J, et al. Magnetic resonance fingerprinting using echo-planar imaging: Joint quantification of T1 and T2 * relaxation times. *Magn Reson Med* 2017;78:1724-1733.

4. Hoover DA, Capaldi DP, Sheikh K, et al. Functional lung avoidance for individualized radiotherapy (FLAIR): Study protocol for a randomized, double-blind clinical trial. *BMC Cancer* 2014;14:934.

5. Eslick EM, Stevens MJ, Bailey DL. SPECT V/Q in lung cancer radiotherapy planning. *Semin Nucl Med* 2019;49:31-36.

6. Matuszak MM, Matrosic C, Jarema D, et al. Priority-driven plan optimization in locally advanced lung patients based on perfusion SPECT imaging. *Adv Radiat Oncol* 2016;1:281-289.

7. Le Roux PY, Hicks RJ, Siva S, et al. PET/CT lung ventilation and perfusion scanning using Galligas and Gallium-68-MAA. *Semin Nucl Med* 2019;49:71-81.

8. Mathew L, Wheatley A, Castillo R, et al. Hyperpolarized (3)He magnetic resonance imaging: Comparison with four-dimensional x-ray computed tomography imaging in lung cancer. *Acad Radiol* 2012;19:1546-53.

9. Tahir BA, Hughes PJC, Robinson SD, et al. Spatial comparison of CT-based surrogates of lung ventilation with hyperpolarized Helium-3 and Xenon-129 gas MRI in patients undergoing radiation therapy. *Int J Radiat Oncol Biol Phys* 2018;102:1276-1286.

10. Eichinger M, Puderbach M, Fink C, et al. Contrast-enhanced 3D MRI of lung perfusion in children with cystic fibrosis--initial results. *Eur Radiol* 2006;16:2147-52.

11. Kipritidis J, Hofman MS, Siva S, et al. Estimating lung ventilation directly from 4D CT Hounsfield unit values. *Med Phys* 2016;43:33.

12. Kipritidis J, Tahir BA, Cazoulat G, et al. The vampire challenge: A multi-institutional validation study of CT ventilation imaging. *Med Phys* 2019;46:1198-1217.

13. Yamamoto T, Kabus S, Klinder T, et al. Four-dimensional computed tomography pulmonary ventilation images vary with deformable image registration algorithms and metrics. *Med Phys* 2011;38:1348-1358.

14. Gefter WB, Hatabu H. Functional lung imaging: Emerging methods to visualize regional pulmonary physiology. *Acad Radiol* 2003;10:1085-9.

15. Shioyama Y, Jang SY, Liu HH, et al. Preserving functional lung using perfusion imaging and intensity-modulated radiation therapy for advanced-stage non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 2007;68:1349-1358.

16. Akkus Z, Galimzianova A, Hoogi A, et al. Deep learning for brain MRI segmentation: State of the art and future directions. *J Digit Imaging* 2017;30:449-459.

17. Valueva MV, Nagornov NN, Lyakhov PA, et al. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Math Comput Simulat* 2020;177:232-243.

18. Chen H, Zhang Y, Kalra MK, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging* 2017;36:2524-2535.

19. Wolterink JM, Leiner T, Viergever MA, et al. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging* 2017;36:2536-2545.

20. Jin CB, Kim H, Liu M, et al. Deep CT to MR synthesis using paired and unpaired data. *Sensors-Basel* 2019;19.

21. Leung K, Marashdeh W, Wray R, et al. A deep-learning-based fully automated segmentation approach to delineate tumors in FDG-PET images of patients with lung cancer. *J Nucl Med* 2018;59.

22. Shin HC, Roth HR, Gao MC, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;35:1285-1298.

23. Ren G, Ho WY, Qin J, et al. Deriving lung perfusion directly from CT image using deep convolutional neural network: A preliminary study. In: Nguyen D, Xing L, Jiang S, eds. *Artificial Intelligence in Radiation Therapy*. New York, NY: Springer International Publishing; 2019:102-109.

24. Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: A platform for subject-specific image analysis, visualization, and clinical support. In: Jolesz FA, eds. *Intraoperative imaging and image-guided therapy*. New York, NY: Springer New York; 2014:277-289.

25. Estepar RS, Ross JC, Harmouche R, et al. Chest imaging platform: An open-source library and workstation for quantitative chest imaging. *Am J Resp Crit Care* 2015;191.

26. Lehr JL, Capek P. Histogram equalization of CT images. *Radiology* 1985;154:163-9.

27. Kipritidis J, Siva S, Hofman MS, et al. Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using 68Ga-labeled nanoparticles. *Med Phys* 2014;41:011910.

28. Hegi-Johnson F, Keall P, Barber J, et al. Evaluating the accuracy of 4D-CT ventilation imaging: First comparison with Technegas SPECT ventilation. *Med Phys* 2017;44:4045-4055.

29. Yamamoto T, Kabus S, Lorenz C, et al. Pulmonary ventilation imaging based on 4-dimensional computed tomography: Comparison with pulmonary function tests and SPECT ventilation images. *Int J Radiat Oncol Biol Phys* 2014;90:414-22.

30. Kipritidis J, Siva S, Hofman MS, et al. Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using GA-68-labeled nanoparticles. *Med Phys* 2014;41.

31. Bucknell NW, Hardcastle N, Bressel M, et al. Functional lung imaging in radiation therapy for lung cancer: A systematic review and meta-analysis. *Radiother Oncol* 2018;129:196-208.

32. Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal* 2017;36:61-78.

33. Oktay O, Schlemper J, Le Folgoc L, et al. Attention u-net: Learning where to look for the pancreas; 2018. arXiv:1804.03999.

34. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention, Pt Iii* 2015;9351:234-241.

35. He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016:770-778.

36. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* 2004;13:600-12.

37. Rankine LJ, Wang Z, Driehuys B, et al. Correlation of regional lung ventilation and gas transfer to red blood cells: Implications for functional-avoidance radiation therapy planning. *Int J Radiat Oncol Biol Phys* 2018;101:1113-1122.

38. Yamamoto T, Kabus S, von Berg J, et al. Impact of four-dimensional computed tomography pulmonary ventilation imaging-based functional avoidance for lung cancer radiotherapy. *Int J Radiat Oncol Biol Phys* 2011;79:279-88.

39. Humphries SM, Notary AM, Centeno JP, et al. Deep learning enables automatic classification of emphysema pattern at CT. *Radiology* 2020;294:434-444.

40. Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 2016;35:1207-1216.

41. Kimura T, Doi Y, Nakashima T, et al. Combined ventilation and perfusion imaging correlates with the dosimetric parameters of radiation pneumonitis in radiation therapy planning for lung cancer. *Int J Radiat Oncol Biol Phys* 2015;93:778-87.

42. Geirhos R, Rubisch P, Michaelis C, et al. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness; 2018. arXiv:1811.12231.

43. Brendel W, Bethge MJae-p. Approximating CNNs with bag-of-local-features models works surprisingly well on ImageNet; 2019. arXiv:1904.00760.

44. Gatys LA, Ecker AS, Bethge M. Texture synthesis using convolutional neural networks. *Adv Neural Inform Process Syst 28 (Nips 2015)* 2015;28.

45. Ritter S, Barrett DGT, Santoro A, et al. Cognitive psychology for deep neural networks: A shape bias case study; 2017. arXiv:1706.08606.

46. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-444.

47. Kriegeskorte N. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Perception* 2016;45:73-73.

Figure Captions

Fig. 1. Flowchart of the patient-selection process for model building and testing. Eligible patients were randomly split into training and test sets. Samples in the training set were randomly partitioned into four folds. In each validation, three folds retained as training folds were directly fitted to the model, and the remaining fold was used for validation. The cross-validation results were used to tune the hyperparameters of the model. The generalizability of the final constructed model was evaluated on the test set.

Fig. 2. (a) Flowchart (right) and images (left) from the preprocessing pipeline. (b) Design of the attention residual neural network (ARNN). The blue bars represent the three-dimensional (3D) residual convolutional (ResConv) layers. The purple and green bars indicate 3D strided convolution (StriConv) and transposed convolution (ConvT), respectively. The red blocks represent the skip-attention connection (SkipAtt). The numbers under the boxes indicate the image size and number of channels. Sequentially, from S1 to S4, the size of each image is half that of the previous feature map. (c) Region-of-interest (ROI) attention block of ARNN.

Fig. 3. Illustration of healthy lung tissue and typical low-functional lung region: (a) SPECT perfusion image in a gray colormap. The blue contour indicates the clinically significant low-functional region. (b) CT image in a gray colormap. (c) CT image in a rainbow colormap. The black arrow indicates a defective region. (d) CT image in a 256 red-green-blue (RGB) colormap, which contains a random selection of 256 RGB colors that distinguish a small number of labeled regions. The local magnification in (d) reveals typical textural patterns in a low-functional lung region (1) and in normal lung tissue (2).

Fig. 4. Results of voxel-wise agreement (correlation coefficient R and structural similarity index measure $SSIM$) evaluation for the four validation folds and the test set. (a) and (b) show the evolution of R and $SSIM$ in the predicted and ground-truth perfusion images, respectively, with an increasing number of epochs during modeling; (c) and (d) are box plots of the voxel-wise agreement evaluation results of the final model.

Fig. 5. Comparisons of SPECT perfusion and deep learning-based CT perfusion (DL-CTPM) mapping images for three representative cases. Each case is presented in three slices in either the coronal or sagittal view. The red arrow indicates the main low functional region. The two left-side columns show a low-performance case, the middle columns show an average case, and the right-side columns show a high-performance case. The original SPECT images were normalized to the 75th percentile value without intensity discretization.

Fig. 6. Histogram of function-wise concordance for the test set.

Table 1 Patient characteristics (n = 73)

		Number	Percent
Sex	Male	24	32.9%
	Female	49	67.1%
Age	Mean ± SD	67 ± 14	
Disease	Pulmonary hypertension	25	34.2%
	Pulmonary embolism	12	16.4%
	Lung cancer	12	16.4%
	Chest pain	2	2.7%
	Chronic obstructive airway disease	2	2.7%
	Systemic lupus erythematosus	2	2.7%
	Others*	14	19.2%
Non-Disease	Healthy lung	4	5.5%

*Others: Pulmonary fibrosis, hypoxia, congestive heart failure, aortic dissection, pulmonary arteriovenous malformation, chronic renal failure, hepatopulmonary syndrome, fever, post COVID-19, shortness of breath, hypoplastic right ling, tuberculosis, pneumonia, deep venous thrombosis.

```
┌─────────────────────────────────────────────────┐
│  Patients who underwent pulmonary SPECT/CT       │
│  perfusion scans at Queen Mary Hospital in 2019  │
└─────────────────────────────────────────────────┘
                      │
                      │          ┌──────────────┐
                      │─────────▶│  Incomplete  │
                      │          │  CT volume   │
                      │          │  n = 2       │
                      │          └──────────────┘
         ┌────────────┴────────────────┐
         ▼                             ▼
┌─────────────────┐           ┌─────────────────┐
│  Cohort 1       │           │  Cohort 2       │
│  total patients │           │  total patients │
│  n = 42         │           │  n = 31         │
└─────────────────┘           └─────────────────┘
         │                             │
         ▼                             ▼
┌─────────────────┐           ┌─────────────────┐
│  Lung           │           │  Lung           │
│  separation     │           │  separation     │
│  n = 84         │           │  n = 62         │
└─────────────────┘           └─────────────────┘
         │                             │
         ▼                             ▼
┌─────────────────────────┐   ┌─────────────────┐
│ Split 1   ███████ n = 22│   │  Testing set    │
│ Split 2   ███ n = 22 ███│   │  n = 62         │
│ Split 3   █ n = 20 █████│   └─────────────────┘
│ Split 4   n = 20 ███████│
│ Training set ███         │
│ Validation set ░░░       │
└─────────────────────────┘
```

(a)

SPECT/CT scan

Registration

Parenchyma segmentation

Left/right lung separation

Resize

CT enhancement

SPECT standardization

(b)

StriConv   ResConv   ConvT   Sigmoid   Skip Att   ROI Att

Input

Output

S1 × 8   S2 × 16   S3 × 32   S4 × 64   S4 × 128   S4 × 64   S3 × 32   S2 × 16

(c)

ROI Att

Input Feature

ConV   Threshold   Sigmoid

ConV

⊗

Output Feature

Low-performance case   Average-performance case   High-performance case
R = 0.37, SSIM = 0.69   R = 0.69, SSIM = 0.70   R = 0.88, SSIM = 0.82
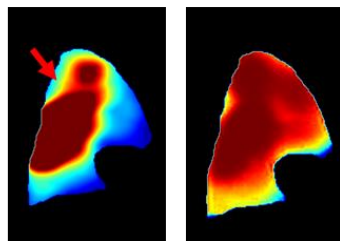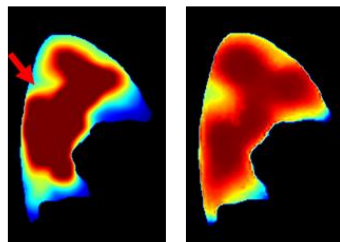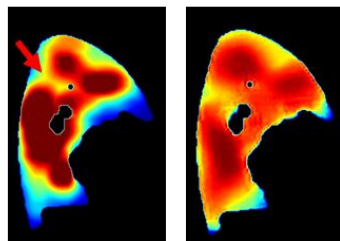
SPECT   DL-CTPM   SPECT   DL-CTPM   SPECT   DL-CTPM