

Article Adipose Tissue Segmentation after Lung Slice Localization in Chest CT Images Based on ConvBiGRU and Multi-Module UNet

Pengyu Lei¹, Jie Li^{1,*}, Jizheng Yi^{1,2,*} and Wenjie Chen¹

- ¹ College of Computer and Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China; 20221100369@csuft.edu.cn (P.L.); ckeiko@163.com (W.C.)
- ² Yuelushan Laboratory Carbon Sinks Forests Variety Innovation Center, Changsha 410000, China
- * Correspondence: jieli.jsj@csuft.edu.cn (J.L.); t20152279@csuft.edu.cn (J.Y.)

Abstract: The distribution of adipose tissue in the lungs is intricately linked to a variety of lung diseases, including asthma, chronic obstructive pulmonary disease (COPD), and lung cancer. Accurate detection and quantitative analysis of subcutaneous and visceral adipose tissue surrounding the lungs are essential for effectively diagnosing and managing these diseases. However, there remains a noticeable scarcity of studies focusing on adipose tissue within the lungs on a global scale. Thus, this paper introduces a ConvBiGRU model for localizing lung slices and a multi-module UNet-based model for segmenting subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT), contributing to the analysis of lung adipose tissue and the auxiliary diagnosis of lung diseases. In this study, we propose a bidirectional gated recurrent unit (BiGRU) structure for precise lung slice localization and a modified multi-module UNet model for accurate SAT and VAT segmentations, incorporating an additive weight penalty term for model refinement. For segmentation, we integrate attention, competition, and multi-resolution mechanisms within the UNet architecture to optimize performance and conduct a comparative analysis of its impact on SAT and VAT. The proposed model achieves satisfactory results across multiple performance metrics, including the Dice Score (92.0% for SAT and 82.7% for VAT), F1 Score (82.2% for SAT and 78.8% for VAT), Precision (96.7% for SAT and 78.9% for VAT), and Recall (75.8% for SAT and 79.1% for VAT). Overall, the proposed localization and segmentation framework exhibits high accuracy and reliability, validating its potential application in computer-aided diagnosis (CAD) for medical tasks in this domain.

Keywords: deep learning (DL); localization; segmentation; subcutaneous adipose tissue (SAT); visceral adipose tissue (VAT)

1. Introduction

Human health has always been a widely concerned problem. Excessive body fat is a worldwide public health problem, and extreme fat content will lead to obesity, metabolic disorders, and a decrease in life quality [1]. Human metabolism speed is closely related to adipose tissue content [2–4]. According to [5–9], excessive fat may lead to fatty liver and liver inflammation and other diseases. Kwon et al. [10] showed that most cardiopulmonary risk factors are strongly and positively correlated with SAT and VAT. The works in [11–13] have shown that adipose tissue content affects the success rate of organ transplantation. Heart and lung transplant recipients are more likely to have a higher risk of cardiovascular disease due to high breast fat content. Anderson et al. [14] have shown that excess fat is associated with an increased risk of primary graft dysfunction (PGD) after lung transplantation. Cho et al. [15] explored the contents of muscle and SAT in lung CT images and probed the relationship between these two contents and the survival rate of patients who have received lung transplantation. The experiments in the paper have shown that excess fat also increases mortality in lung transplant patients. The studies showed that excessive fat content in the body is harsh to human health. Traditional adipose tissue estimation



Citation: Lei, P.; Li, J.; Yi, J.; Chen, W. Adipose Tissue Segmentation after Lung Slice Localization in Chest CT Images Based on ConvBiGRU and Multi-Module UNet. *Biomedicines* 2024, *12*, 1061. https://doi.org/ 10.3390/biomedicines12051061

Academic Editors: Moritz Wildgruber and Taesup Lee

Received: 29 March 2024 Revised: 23 April 2024 Accepted: 6 May 2024 Published: 10 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). methods calculate the fatty tissue by the weight value or CT images and other clinical parameters such as Body Mass Index (BMI). The literature [16,17] has shown that BMI does not accurately and precisely reflect the mass of fat tissue in the body. Therefore, we need a more intelligent and accurate way to estimate human adipose tissue accurately.

Computer-aided diagnosis (CAD) has been extensively studied and applied in the medical field [18,19]. In medical image research, many researchers conducted analyses based on X-ray images. Bottigli et al. [20], when studying breast cancer, used the distributed database to analyze mammography images and classified fat and other tissues through the obtained texture analysis results. In recent years, the rapid development of artificial intelligence (AI) has also made significant progress in applying DL in the medical domain. Deep learning (DL) methods have recently drastically affected medical imaging, revolutionizing this field. Mina et al. [21] proposed a Dense Residual UNet (DRU), which made an improvement on the Residual Network (ResNet) and the Dense Network (DenseNet) to complete the medical image segmentation task. This work has superior performance in labeled datasets with few pixels and few training samples, but its utility in datasets with more examples needs to be proved. Zhang et al. [22] proposed a weak medical image segmentation supervision method. They achieved the segmentation results via the scale constraint method but compared it with the supervised model. Usman et al. [23] conducted a three-dimensional segmentation of lung CT using an adaptive region of interest (ROI) algorithm to study the exact quantification of pulmonary nodules based on the deep residual UNet structure. Still, it has yet to be reliably demonstrated for the fat segmentation task, which is more complex, irregular in shape, and more difficult to distinguish from adjacent tissues. Peng et al. [24] conducted a texture analysis on CT images. They introduced a multi-perception model to analyze further the five parameters of the lowest density value, energy, kurtosis, low-density enhancement value, and hara variance to improve the accuracy of diagnosis and reduce the misdiagnosis rate.

Adipose tissue is characterized by a complex structure, significant individual differences, and random shape [25]; image-based fat quantification, therefore, remains a considerable challenge. Cao et al. [26] proposed an unsupervised clustering segmentation method of fat based on MRI images of the human abdomen. However, due to the limitations of the unsupervised method, the distinction between SAT and VAT becomes more difficult without the guidance of labels during the training process, and there is still room for improvement in the segmentation accuracy. Hussein et al. [27] used gradient points to provide the initial boundary of SAT and VAT and then refined the edge based on appearance and geometric calculation. They finally obtained the final boundary and segmentation results through the context label fusion based on the conditional random field (CRF), which replaced the function of supervised methods with unsupervised methods. Irmak et al. [28] proposed an affinity propagation clustering method based on fuzzy connectivity (FC) image segmentation for automatic segmentation. This is considering the heterogeneity of adipose tissue but might be a bit subjective as the automated selection approach evades the need for filtering systems. Amer et al. [29] proposed a patch integration algorithm that fuses the contextual information of adjacent pixels for the irregular problem of adjpose tissue and applied the UNet structure and K-Means algorithm of embedded parts to achieve the segmentation task. It has not been demonstrated whether the results obtained by this work in leg muscle and adipose tissue segmentation are equally applicable to the structurally complex task of adipose tissue segmentation in lung slices.

Estrada et al. [25] implemented image segmentation through a two-dimensional, complete convolutional network based on UNet. They competed for the output of each pooling layer and the input of its peer layer in the upsampling process after a fusion via a convolutional layer, retained the optimal one, and transmitted it to the next layer to obtain more accurate segmentation results. This network structure has yet to be demonstrated in adipose tissue segmentation in lung slices. A Squeeze-and-Excitation (SE) block is proposed for zonal prostate segmentation [30]. Koitka et al. [31] added multi-resolution blocks combined with the convolution results of adjacent layers. They obtained segmentation predictions with more comprehensive features. Langner et al. [32] compared two complete convolutional neural network structures, UNet and VNet, in a study of abdominal adipose tissue segmentation and verified the advantages of UNet. Our paper adopts the UNet structure as the baseline model and explores its performance in adipose tissue 2D segmentation. Masoudi et al. [33] used the conditional generated network (c-GAN) to generate the synthesized CT images (s-CT) of adipose tissue in the study on abdominal adipose tissue and then used the UNet structure to refine SAT and VAT s-CT. The effect of the model varies with the data, but the advantages of UNet are clear.

The fusion of spatial and learning information with contextual information can help to localize pulmonary tissues. The gated recurrent units (GRUs) [34] model works based on the recursive updating of the features in the system. Bidirectional GRU (BiGRU) is a modified GRU. It can not only preserve the learned information but also provide feedback on the future learned information and is implemented in a two-way information fusion mechanism.

In this paper, we propose a model inspired by BiGRU and a multi-module combined approach for lung slice localization and adipose tissue segmentation using the UNet structure. Compared to conventional methods in terms of accuracy and specificity, our proposed methods demonstrate superiority. The model utilizes multi-modules and attention mechanisms [35] to segment VAT and SAT in CT slices, showing a significantly better performance than feature fusion across multiple components. We explore different segmentation results of various modules (the competitive module, multi-resolution module, and attention mechanism) and the segmentation outcomes of combined modules. The results indicate more accurate segmentation of fat content values and reduced measurement errors, demonstrating a more comprehensive characterization of adipose tissue.

2. Materials and Methods

In this section, we introduce the overall framework and detailed design of this work, which is primarily divided into three parts: the preprocessing module, the lung slice localization model, and the adipose tissue segmentation model.

2.1. Overall Framework of the Proposed Methodology

The overall workflow and the connection to each module are shown in Figure 1. We set the continuous lung CT slice sequences as the input in this paper.

In our approach, every five slices (this sequence size is discussed in the comparison experiment) are grouped, taking into account their positional relevance and order in the body. For each sample, the continuous CT image sequences' information is fused to automatically position the lung in CT images based on BiGRU. The localization task essentially involves a triple-classification process. The resulting three categories (the first slice of the lung, the last portion of the lung, and other pieces, i.e., FL, LL, and OS) are used to locate the entire lung, thereby determining the spatial location of the CT sequence of the lung. Lung slice localization serves as a selection step to exclude unnecessary slices in the whole chest CT for the subsequent segmentation process.

The input of the segmentation model is chosen from FL to LL (including the pieces in between) and ultimately detects the SAT and VAT areas using the multi-module UNet model. Additionally, in this paper, we test the combination of multi-module mechanisms (i.e., multi-resolution, the competitive module, and the attention module) used to modify the original UNet, which leads to improvements in the segmentation accuracy in the localized lung CT slices.



Figure 1. Workflow of the proposed methodology. The upper part of the diagram represents the localization model, while the lower part depicts the segmentation model. The localization model identifies the first and last slices of the lung in the CT sequence through multi-class classification, establishing the spatial location of the lung region within the CT sequence. In the results of the upper part, '00', '01', and '10' correspond to the top slice, the bottom slice, and other slices between these boundaries, respectively. Subsequently, SAT and VAT are segmented and identified by integrating multi-modules into the layers of the UNet network. This enhancement results in improved segmentation accuracy compared to traditional UNet methods.

2.2. Lung Slices Localization Model in Chest CT Imaging

Observing the CT slice sequence, the lung is consistently surrounded by intricate tissues, with variations observed from case to case. Distinguishing the targeted slice from its neighboring pieces presents a challenge, even for seasoned experts, due to the overall variability among organs and imaging conditions. This challenge is further compounded in cases where differentiating nodules from tumors is necessary. This study underscores the enhancement of pulmonary localization results through the amalgamation of structure and texture information from lung CT slices. The localization task is tackled through the fusion of contextual information within the CT slice sequence. To accomplish this, the ConvBiGRU model, depicted in Figure 2, performs the fusion of information from multiple slices along with their spatial correlation.



Figure 2. ConvBiGRU structure. The input is the CT slice sequence with five slices in each group, and the output is the specific class of the slices. There are overall three class results of the output in localization. The $\overrightarrow{h_{t-1}}$ and $\overleftarrow{h_{t-1}}$ represent the forward and backward information, respectively, and \tilde{x}_t is the obtained results at time *t*.

In this model, the original linear layers are replaced by convolution layers to extract the spatial image information. The input of ConvBiGRU requires over three CT slices due to the space contextual information need, which will be discussed in the comparison experiment

part. The model extracts characteristic information of these slices through forwarding and backpropagation. The presented model is inspired by the ConvLSTM network [36], combining forward and backward spatial information with temporal information. The convolution part reduces the spatial redundancy while the BiGRU function extracts more comprehensive information than LSTM and finally precisely identifies the slices with the lung. The update gate, z_t , at the time t in a GRU block is as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{1}$$

where h_{t-1} and σ are the information at t-1 and activation functions, respectively. The update gate determines the amount of information retained from t-1. The reset gates at time t in the model, r_t , are as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{2}$$

The reset gate is considered to be a linear transform of the combined information at the time t - 1 for the new input with a sigmoid activation function.

$$h_t = \tanh \sigma \left(W \cdot [r_t \cdot h_{t-1}, x_t] \right) \tag{3}$$

Let \tilde{h}_t be the updating value of the gate, where \tilde{h}_t represents the newly generated information from the reset gate, r_t , while activated by a hyperbolic tangent function. Adding the results from two entrances, the output of the model obtained by the following recursive function at time *t* is as follows:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h_t \tag{4}$$

Like the GRU model, the BiGRU operates using the same logic but multiple times in two different directions, as presented in Figure 3.



The model fuses forward information and receives information from the back. To perform a lung localization, we measure the interval distance between the classification



predictions and the ground truth, as well as the evaluation index of the model. As the model takes information fusion from multiple CT slices, the feature learning in the model can be affected by the length of the input sequences. To alleviate that, we performed a cross-comparative analysis of the model with different sizes of input sequences and other normalization methods. Different lengths of lines (three slices, five slices, and seven slices) were tested to compare the results to determine the optimal one.

At the same time, to minimize the loss and increase the learning effect of the model through feature intensity, we performed two different normalizations. For the first normalization, called *MaxIntensity*, we rescale the image's pixels by the maximum power of the image, $f_n(x, y) = f_n(x, y)/4095$, where $f_n(x, y)$ means the pixel intensity of the image at (x, y) coordinate and 4095 represents the maximum intensity of the pixels in the image. For the second normalization way, called *MaxMinNorm*, we rescale the pixels' intensity in the range of [0, 1], as shown below:

$$f_n(X,Y) = \frac{f(X,Y) - \min(f(X,Y))}{\max(f(X,Y)) - \min(f(X,Y))}$$
(5)

where $min(\cdot)$ and $max(\cdot)$ denote the same slice's minimum and maximum pixel intensity, respectively.

Because the original dataset has massive medical images of coarse types of the lesion including bone, abdomen, lung, and five more classes, the localization part of this paper was trained on 1483 images containing the target lung region selected from 2184 CT images of 100 patients, and 252 images were randomly chosen as the test set. The validation set was separated individually from the training set with a ratio of 2:8 in the localization process to improve the model.

2.3. Adipose Tissue Segmentation Using Multi-Module UNet

A UNet-based multi-module model is used to segment SAT and VAT. We conduct the cross-comparison among the attention, multi-module, and multi-resolution blocks. The best combination improves the accuracy and robustness of the model to perform segmentation. The segmentation structure is shown in Figure 4.



Figure 4. Segmentation model structure. Adding the blocks and converting the original model to the multi-module structure through the cross-test. It should be noted that attention modules only set the upsampling at the peer layer as the input.

Three modules, including a multi-resolution module, competitive module, and attention mechanism module, are added to the original UNet structure and compared to their performance. The output of the multi-resolution module is combined with the production of different convolutional layers and then transferred to the next layer as the input. The competitive module maxes out the result by comparing the outputs of the multi-resolution module and the peer layer. The attention mechanism reduces the influence of noise and low-intensity contrast between VAT and SAT. Through the different combinations of these modules, we conduct a comparative analysis of the segmentation. Figure 5 presents the variety of the multi-resolution module and the competitive module.



Figure 5. Combination structure of competitive mechanism block and multi-resolution block. The steps before the 'Add' operation represent the multi-resolution block, and the operations below are the competitive mechanism part. All different modules will eventually go through Maxout, using its strong fitting ability to obtain more accurate results.

In addition to obtaining the optimal outcome between each upsampling layer and their corresponding pooling layer, the multi-resolution results are also added to the competition to construct a two-step selection mechanism. Applying the sigmoid function in the model, we achieved a binary classification for predictions, which maps the data into (0, 1). The sigmoid function is defined as follows:

$$S(p(x,y)) = 1/(\exp(-p(x,y)))$$
(6)

where p(x, y) represents the probability value of the corresponding (x, y) coordinates. The optimizer is Stochastic gradient descent (SGD) used in the model. This algorithm makes a more effective outcome while having some level of redundancy. To maintain the training speed while the number of samples increases, each weight update in SGD updates once during its iteration. The updating process is shown as follows:

$$\theta_j = \theta_j + \left(y^i - h_\theta \left(x^i \right) \right) x_j^i \tag{7}$$

where h_{θ} is the prediction to x^i , y^i as input samples and the output of the layer, respectively, and x_j^i is the *j*th variable in the *i*th sample. To accelerate SGD in the relevant direction, we use the momentum parameter beta β set to 0.9 to calculate the exponentially weighted averages as follows:

$$v_{dW} = \beta v_{dW} + (1 - \beta) d_W \tag{8}$$

$$v_{db} = \beta v_{db} + (1 - \beta) d_b \tag{9}$$

where d_W and d_b is the weight and bias, respectively. Having noise in the results is seemingly inevitable due to the local randomness during the descent. The loss is measured by the Dice coefficient distance/loss (DCL) and measured as follows:

$$DC(y_{ture}, y_{pred}) = \frac{\kappa + 2\sum y_{ture} \cdot y_{pred}}{\kappa + \sum y_{ture} \cdot y_{true} + \sum y_{pred} \cdot y_{pred}}$$
(10)

$$DCL(y_{ture}, y_{pred}) = 1 - DC(y_{ture}, y_{pred})$$
(11)

We use the weight penalty in our analysis to tackle the imbalance input sample problem for the localization model, which improves the learning convergence speed. In other words, we calculated the ratio of each sample through a logarithmic function to proportionally and effectively control the growth rate. The formula to calculate the sample weight is as follows:

$$weight = weight_i / \sum_{i=1}^{i=n} weight_i$$
(12)

$$W = 1/\log(\alpha + weight) \tag{13}$$

where *weight*_{*i*} is the times of occurrences of each category and *n* and *α* are the number of sample categories and the factor controlling the sample proportion, respectively. The constant *α* is set to 1.1 based on experience in the experiment to avoid negative numbers and zeros. Thus, $W \in (\frac{1}{\log 2.1}, \frac{1}{\log 1.1})$.

Finally, the sample weight is added to the loss as follows:

$$DCL_{W}(y_{true}, y_{pred}) = DCL(y_{true}, y_{pred}) + W$$
(14)

The segmentation part of this paper was trained on 2351 CT images, and 975 images were selected as the verification set each time, which was verified by the four-fold cross-verification method. Besides the images from the localization, we added the remaining images to the dataset to ensure sufficient training. Then, 572 and 573 images were selected as the test set for SAT and VAT segmentation, respectively. The validation set was separated individually from the training set with a ratio of 2:8 in the segmentation process to improve the model.

3. Results

Here we describe the experiment in detail and analyze the experimental results. This study selected 100 patients with lung CT images from the DeepLesion dataset by the National Institutes of Health Clinical Center (NIH CC) [37]. The CT images from the original dataset were all 16-bit unenhanced grayscale images ($512 \times 512 \times 1$) in png format. Slice spacing among these input CT images ranged from 0.25 mm, 0.4 mm, 0.6 mm, 1 mm, 1.25 mm, 2 mm, 2.5 mm, and 3 mm to 5 mm.

3.1. Data Preprocessing

For our study, we meticulously selected lung slice images from the DeepLesion dataset based on specific criteria. We focused on CT scan sequences that comprehensively captured the entire range of lung anatomy from initial appearance to disappearance. Within these sequences, we categorized the slices into distinct lung regions: apex, midsection, and base. From the original DeepLesion dataset, which encompasses diverse medical images covering various lesion types, we specifically extracted 1483 CT images containing lung regions out of a total of 2148 images from 100 patients. Our selection process aimed to ensure the representation of typical lung anatomy and pathology encountered in clinical practice. The label annotation tool [38] was used to manually label the images. Our group annotated the labels with the guidance of an expert and verified them by a professional experienced radiologist. Labels are stored and read following the serial number of the cases. Segmentation labels are recorded in JSON format, and images and their labels are saved accordingly in eight-bit *png* format.

After reading the images and their labels, we translated them into array format and split them into training and testing sets. The images and labels were augmented in 16-bit format to improve the contrast between different tissues, organs, and other components. The augmentation solely enhances contrast without cropping or rotating the images. This facilitates the manual labeling of the targeted slice and segmentation of their labels. Converting the images to eight-bit format resulted in missing information, leading to greater disparity in prediction outcomes. Therefore, we only convert them for annotation work, not for the actual training input

We conducted a multi-class classification to localize the lung slices and labeled them based on their position in the CT image as '0', '1', or '2' (Figure 6). '0' signifies the first slice of the lung, '2' denotes the last slice of the lung, and any other slices outside of these boundaries are labeled as '1' (as depicted in Figure 6).

With these three categories, we could distinguish the different locations of the lung in slices. The label is processed by one-hot encoding into binary format and then input into the network.

All experiments in this study were conducted in a Python 3.6 environment, running on a computer equipped with an Intel Core[™] i7-9750H CPU, 8GB RAM, and a GTX 1660TI GPU. Additionally, the computational tasks were also performed on four high-performance servers in our laboratory. These servers are equipped with NVIDIA Titan Xp series GPUs and Intel i9 processors, providing robust support for our experiments.





(b) last slices labeled with "2"



(c) other slices labeled with "1"

Figure 6. Examples of three classes of localization labels. There are large differences in the size of body tissues between different cases, which brings difficulties to the training and testing effects of the model.

3.2. Results of Lung Slice Localization

When using the entire dataset for training and testing subsets, sample imbalance problems in localization would affect the results, i.e., much more irrelevant slices than lung slices (the first and the last slices). This will result in an imbalanced classification of the model as the features learned by the model may be more biased toward negative samples, which weakens the correct model prediction. To solve this problem, we divided the training set into three subsets with 700 slices in each subset. The total number of annotated lung slices in each subgroup was kept at $320(\pm 10)$ to balance the proportion of positive and negative samples. Moreover, we add a penalty during the backward propagation according

to different types of incorrect outputs in the classification model. Every error will add a weight of -1 to the misclassification in the results. We obtained more balanced results in the localization model based on this mechanism. We also conducted comparative analyses by the different number of input sequence slices (three, five, or seven slices) and normalization methods (*MaxIntensity* or *MaxMinNorm*). These experiments were conducted based on three subdatasets from the segmentation datasets on 2315 CT slices and tested on 975 CT slices. The results are presented in Table 1.

As shown in Table 1, the results of *MaxMinNorm* are better than those of the *Max-Intensity* normalization method. Also, the pulmonary texture information obtained by *MaxMinNorm* is preserved better than other normalization techniques. The results of the different number of slices indicated that five slices of input yielded the best localization outcome. At the same time, three slices of input reduced the proportionality of positive and negative samples, negatively affecting localization prediction. Also, the number of negative examples in the slice sequence increased for the seven-slice input while the proportion of positive models decreased. Boxplots show the difference between the two normalization methods and input slices in Figure 7.

Table 1. Results of the localization model for lung slices.

Number of flices in a Securate	Mean Distance (mm)		
Number of Sinces in a Sequence —	MaxIntensity	MaxMinNorm	
Three	13.2 ± 6.3	12.7 ± 5.7	
Five	12.5 ± 4.6	10.2 ± 6.3	
Seven	13.4 ± 4.4	12.5 ± 8.6	



Figure 7. The localization model visualization results of two scheme combinations. (**a**) is the boxplot of scheme 1; (**b**) is the boxplot of scheme 2. The horizontal lines in each box in the chart represent maximum, Upper Quartile (*Q*3), median, Lower Quartile (*Q*1), and minimum, respectively, which indicate the stability and distribution of data. The statistical data in the figure come from the test results of a set of 200 slice images of the test set randomly selected, which do not completely overlap with the data selected in Table 1, which further supports the stability of the results.

In Figure 7, we plotted the boxplots with the different numbers of slices per sequence under the two normalization techniques. It shows that the slice spacing obtained in *MaxMinNorm* is generally lower, and the variation range is more extensive than the *MaxIntensity* normalization technique. However, the median value of *MaxMinNorm* is more stable, and its distribution of *Q*1 and *Q*3 is more uniform. To calculate the spacing distance for the localization analysis, we used the following equation:

$$d_t = \begin{vmatrix} y_{pred} - y_{true} \end{vmatrix} \cdot sp_t \tag{15}$$

where sp_t represents the distance between adjacent slices in this case, and d_t represents the slice distance between the predicted result and its targeted position in the ground truth (in millimeters—mm). In the localization experiment, the average distance between the ground truth and the predicted target is shown by the Mean Distance (MD) and is calculated as follows:

$$MD = \sum_{i=1}^{N} \left(\left| No_{pred} - No_{true} \right| \cdot sp_t \right) / N$$
(16)

where No_{pred} is the number of the predicted slice with the highest score in the tested volume. No_{true} is the number of the first or last slices (localizing the lung area). N is the total number of testing samples, and sp_t is as mentioned before.

3.3. Results of Adipose Tissue Segmentation

Several comparative experiments were carried out to evaluate the performance of the segmentation model. The segmentation model can consider adipose tissues, VAT, and SAT as binary classification problems for each task. We conducted four experiments to segment both types of adipose tissue automatically. The experimental results of UNet and different modules based on UNet are shown in Table 2.

The results indicated that the overall segmentation of the SAT tissue has higher accuracy and better performance than VAT tissues using the proposed models. The model with the multi-resolution (MR) module has the best performance, while it is close to the results of the model with the attention module (ATTN). Both these two results exhibited better performance than the competitive module (CP), while the overall difference is negligible and all better than the baseline UNet model. For the segmentation of VAT, the segmentation model with MR showed the best performance, followed by the ATTN and CP models with a small gap.

Model	Adipose Type	Dice Score	F1 Score	Precision	Recall
UNet	SAT VAT	$\begin{array}{c} 90.3\pm0.8\\ 79.8\pm1.6\end{array}$	$\begin{array}{c} 81.2\pm0.7\\ 74.1\pm1.5\end{array}$	$95.9 \pm 0.9 \\ 77.3 \pm 1.4$	$\begin{array}{c} 74.6 \pm 0.9 \\ 78.6 \pm 1.9 \end{array}$
UNet + MR	SAT VAT	$\begin{array}{c} \textbf{90.7} \pm \textbf{0.8} \\ \textbf{81.4} \pm \textbf{1.6} \end{array}$	$\begin{array}{c} \textbf{81.6} \pm \textbf{0.8} \\ \textbf{77.1} \pm \textbf{1.8} \end{array}$	$\begin{array}{c} \textbf{96.4} \pm \textbf{0.4} \\ \textbf{78.2} \pm \textbf{1.8} \end{array}$	$\begin{array}{c} \textbf{75.0} \pm \textbf{0.8} \\ \textbf{77.1} \pm \textbf{2.1} \end{array}$
UNet + ATTN	SAT VAT	$90.5 \pm 1.2 \\ 80.8 \pm 2.1$	$\begin{array}{c} 81.7\pm0.9\\ 76.0\pm2.0\end{array}$	95.6 ± 0.4 76.6 ± 1.4	$\begin{array}{c} 74.3\pm0.8\\ 75.2\pm2.4\end{array}$
UNet + CP	SAT VAT	$\begin{array}{c} 90.3 \pm 0.9 \\ 80.5 \pm 1.7 \end{array}$	$\frac{81.3 \pm 0.8}{75.6 \pm 1.7}$	96.0 ± 0.6 76.6 ± 1.4	$74.5 \pm 1.2 \\ 74.8 \pm 1.8$

Table 2. Comparative results of segmentation using single-module modified UNet structures.

We conducted a deep analysis on the comparative experiments of each module with baseline UNet. We also aggregated every two-module group to compare with the baseline UNet. Table 3 shows the experimental results of such a segmentation SAT and VAT combination.

Table 3. Comparative results of segmentation using multi-module modified UNet structure.

Model	Adipose Type	Dice Score	F1 Score	Precision	Recall
UNet + MR + ATTN	SAT VAT	$\begin{array}{c}\textbf{92.0}\pm\textbf{1.7}\\\textbf{82.6}\pm\textbf{1.4}\end{array}$	$\begin{array}{c} \textbf{83.5}\pm\textbf{1.6}\\ \textbf{78.6}\pm\textbf{2.4} \end{array}$	$\begin{array}{c}\textbf{96.4}\pm\textbf{2.5}\\\textbf{78.9}\pm\textbf{2.9}\end{array}$	$\begin{array}{c} \textbf{77.5} \pm \textbf{1.2} \\ \textbf{79.3} \pm \textbf{3.1} \end{array}$
UNet + MR + CP	SAT VAT	$\begin{array}{c} 91.7 \pm 1.1 \\ 81.1 \pm 1.6 \end{array}$	$\begin{array}{c} 83.9\pm1.3\\ 76.4\pm2.2 \end{array}$	$\begin{array}{c} 95.9 \pm 2.3 \\ 77.4 \pm 2.3 \end{array}$	$\begin{array}{c} 75.2 \pm 2.4 \\ 77.3 \pm 2.3 \end{array}$
UNet + ATTN + CP	SAT VAT	$91.9 \pm 0.7 \\ 80.6 \pm 1.2$	$\frac{83.5 \pm 2.6}{76.0 \pm 3.0}$	95.7 ± 1.7 76.0 ± 2.6	$74.8 \pm 2.3 \\ 76.9 \pm 3.1$

MR: multi-resolution module; ATTN: attention module; CP: competitive module.

The results indicate that the combination of the modified UNet model with MR and ATTN has the best performance in this segmentation task. Then, the modified UNet with ATTN and CP showed better accuracy than other additive modules of MR and CP. They are adding an ATTN block for the segmentation of VAT-inhibited redundancy, which brings the information of other larger thoracic tissues.

The lung CT slices showed that the adipose tissue had non-uniformity and irregularity, i.e., heterogeneity, while the adipose tissue has a similar morphological profile to fluid in the body. Comparing the prediction and ground truth visually in segmentation demonstrates the significant performance while missing some tiny fat tissues at the edge of the predicted results. The misclassification in the mentioned regions might be because of the blurred border between the pixel intensity of the adjacent area. During the experiment, we adjusted the label of the training set to clean the misclassified parts and screened the CT slices with big noises. This is aimed to ensure the relative purity of the training dataset and exclude the negative influence during the learning process that is not supposed to be there.

The model applied a four-fold cross-validation to segment SAT and VAT to ensure the results' reliability and robustness. For that, the averaged overall Precision and Recall rates were computed along with the overall averaged F1 Score and averaged Dice Score for the segmentation accuracy. Some examples of segmented parts and their corresponding ground truth for SAT and VAT are visualized in Figure 8.



(a) SAT Slices

(b) VAT Slices

Figure 8. Visualizations of segmentation results. Both the segmented parts and ground truth for SAT and VAT are visualized. Green, yellow, and red regions represent the ground truth, true predictions, and false predictions, respectively.

The proposed segmentation model showed better accuracy in segmenting lung adipose tissue than other models [21,39–42]. The proposed model outperformed other approaches in SAT or VAT segmentation tasks (see Table 4). From Table 4, it can be observed that there are significant differences in the performance of the Dice Score, F1 Score, and Precision, while the performance of Recall shows little variation. Recall represents the proportion of samples that are truly positive and are correctly predicted as positive by the model, calculated as Recall = TP/(TP + FN). The similar Recall values across all methods in Table 4 are due to the balanced nature of our dataset and the design of our experiments. We carefully balanced the dataset during preprocessing to ensure an equal representation of positive and negative samples. Additionally, our experiments were conducted under similar conditions, including the same evaluation metrics and testing procedures. As a result, the Recall values reflect the ability of each method to correctly identify positive instances relative to the total number of positive instances in the dataset, leading to a comparable performance across methods.

Model	Adipose Type	Dice Score	F1 Score	Precision	Recall
DRUNet [21]	SAT VAT	$91.1 \pm 1.1 \\ 80.1 \pm 2.5$	$\begin{array}{c} 82.0\pm2.5\\ 78.0\pm2.7\end{array}$	$\begin{array}{c} 95.9 \pm 1.7 \\ 77.3 \pm 2.7 \end{array}$	$\begin{array}{c} 72.5 \pm 4.5 \\ 75.4 \pm 3.7 \end{array}$
UNet [39]	SAT VAT	$\begin{array}{c} 87.8\pm2.7\\ 81.2\pm1.3\end{array}$	$\begin{array}{c} 81.5\pm2.5\\ 77.3\pm3.4\end{array}$	$\begin{array}{c}91.7\pm0.9\\76.1\pm2.2\end{array}$	$\begin{array}{c} 74.6\pm1.5\\ 78.6\pm2.1 \end{array}$
UNetTransformer [40]] SAT VAT	$\begin{array}{c} 88.6\pm1.5\\ 81.4\pm2.0 \end{array}$	$\begin{array}{c} 80.9\pm1.8\\ 78.3\pm4.9\end{array}$	$\begin{array}{c} 91.5 \pm 1.9 \\ 75.7 \pm 3.6 \end{array}$	$\begin{array}{c} 76.9\pm5.5\\ 78.1\pm4.2\end{array}$
FedDG [41]	SAT VAT	$\begin{array}{c}90.7\pm2.1\\80.1\pm2.5\end{array}$	$\begin{array}{c} 83.3\pm3.5\\ 73.4\pm3.3\end{array}$	$\begin{array}{c}92.2\pm2.4\\75.3\pm6.4\end{array}$	$\begin{array}{c} 75.0\pm1.8\\ 77.1\pm2.1\end{array}$
Transfuse [42]	SAT VAT	$\begin{array}{c} 91.3 \pm 1.6 \\ 80.6 \pm 2.9 \end{array}$	$\begin{array}{c} 82.1\pm2.0\\ 77.4\pm4.7\end{array}$	$\begin{array}{c}95.2\pm4.6\\76.4\pm2.4\end{array}$	$\begin{array}{c} 76.7\pm2.7\\ 77.6\pm3.3\end{array}$
Ours	SAT VAT	$\begin{array}{c} \textbf{92.0} \pm \textbf{1.7} \\ \textbf{82.6} \pm \textbf{1.4} \end{array}$	$\begin{array}{c} \textbf{83.5} \pm \textbf{1.6} \\ \textbf{78.6} \pm \textbf{2.4} \end{array}$	$\begin{array}{c}\textbf{96.4}\pm\textbf{2.5}\\\textbf{78.9}\pm\textbf{2.9}\end{array}$	$\begin{array}{c} {\bf 77.5 \pm 1.2} \\ {\bf 79.3 \pm 3.1} \end{array}$

Table 4. Comparison results of segmentation with other methods on the same dataset.

4. Discussion

This study proposes an automated model for lung localization and adipose tissue segmentation in CT images using deep learning. The ConvBiGRU structure, incorporating contextual spatial features, is utilized for lung tissue localization. Subsequently, a UNet structure modified by a multi-resolution module and attention mechanism is employed for segmenting subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT). Contextual information, such as forward and backward spatial information, is utilized for lung slice localization. In this study, the dataset is divided into multiple subsets to address sample imbalance issues. Furthermore, comparative analyses are conducted on different image normalization methods, input numbers, and multi-module combinations. According to Table 4, it can be observed that different network methods exhibit certain differences in the segmentation performance of SAT and VAT. For example, DRUNet shows a slightly better performance than other methods in SAT's Dice Score, F1 Score, and Precision, while lagging slightly behind in VAT's Recall. In contrast, FedDG performs the best in SAT's Recall but slightly lower in other metrics compared to other methods. Our method overall outperforms other methods, not only showing the best performance in SAT and VAT's Dice Score and F1 Score but also maintaining high levels of Precision and Recall, indicating the comprehensive advantage of our method in localization and segmentation accuracy. We employ a ConvBiGRU structure for lung slice localization, combined with a multi-module UNet structure for SAT and VAT segmentation. This combination not only helps in accurately localizing lung slices but also accurately identifying SAT and VAT in segmentation tasks, reflecting the advancedness of our method in model design and structure optimization. Our method exhibits good stability in performance metrics with small deviation ranges, indicating its high robustness.

However, we also acknowledge some limitations in this study. Firstly, positive and negative classifications exist within the boundaries of adipose tissues due to the fuzzy areas surrounding other tissues. Additionally, although our model achieves satisfactory results on the current dataset, we must explicitly state that these results cannot be directly generalized to other datasets. Because different datasets may have different features and distributions, our model needs to be validated and adjusted on different datasets to ensure its generalization ability and reliability. Most importantly, our research results do not imply that our model can replace the diagnosis and judgment of professional radiologists. Although our model performs well in lung localization and adipose tissue segmentation, the expertise and experience of doctors are still indispensable in clinical applications. Our research aims to provide auxiliary diagnostic tools for doctors rather than replace their decisions.

5. Conclusions

To address the significant challenges associated with adipose tissue segmentation in lung CT imaging, this paper introduces an automated lung localization and adipose tissue segmentation model employing deep learning techniques. The proposed framework consists of a bidirectional gated recurrent unit (BiGRU) structure for lung slice localization and a multi-module UNet model for segmenting subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT). Furthermore, the BiGRU Net model is enhanced through the incorporation of an additive weight penalty term. In the segmentation phase, we investigate three combined mechanisms—attention, competition, and multi-resolution—within the UNet structure to assess their impact on SAT and VAT segmentation, facilitating a comparative evaluation. Finally, extensive experiments conducted on a publicly available dataset demonstrate the efficiency and effectiveness of the proposed framework.

In our future research, firstly, we plan to further optimize the model performance, especially in positive and negative classification. Although our model performs well in localization and segmentation tasks, there still exists a certain proportion of positive and negative classifications. Future work will focus on how to improve the model to reduce the occurrence of this situation, with possible methods including introducing more contextual information or exploring more complex post-processing techniques. Secondly, we will attempt to apply the model to more diverse datasets. While we conducted experiments on one publicly available dataset, we are aware that different datasets may have differences. Therefore, we plan to apply the model to more datasets from different sources and with different characteristics, further validating its robustness and universality. Lastly, we will focus on the application of the model in clinical practice. We will collaborate with radiologists to compare our model with the results manually segmented by them, further validating its feasibility and effectiveness in actual clinical settings. Through the aforementioned improvements and explorations, we believe that our research will provide a deeper understanding of adipose tissue segmentation in lung CT images and offer valuable references and insights for the development of future related studies.

Author Contributions: Conceptualization, P.L., J.L. and J.Y.; methodology, P.L. and J.L.; software, W.C.; validation, P.L. and J.L.; formal analysis, P.L.; investigation, W.C.; resources, J.Y. and W.C.; data curation, P.L.; writing—original draft preparation, P.L. and J.L.; writing—review and editing, J.L. and J.Y.; visualization, P.L.; supervision, W.C.; project administration, J.L.; funding acquisition, J.L. and J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Hunan Provincial Natural Science Foundation of China (grant no. 2022JJ31022) and the National Natural Science Foundation of China (grant no. 62202505).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The DeepLesion datasets, which were used and analyzed during the current study, are accessible at the following URL: https://nihcc.app.box.com/v/DeepLesion, accessed on 4 April 2023.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Padwal, R.; Leslie, W.D.; Lix, L.M.; Majumdar, S.R. Relationship Among Body Fat Percentage, Body Mass Index, and All-Cause Mortality: A Cohort Study. Ann. Intern. Med. 2016, 164, 532–541. [CrossRef] [PubMed]
- Kuda, O.; Rossmeisl, M.; Kopecky, J. Omega-3 fatty acids and adipose tissue biology. Mol. Asp. Med. 2018, 64, 147–160. [CrossRef] [PubMed]
- Lee, M.J.; Wu, Y.; Fried, S.K. Adipose Tissue Heterogeneity: Implication of Depot Differences in Adipose Tissue for Obesity Complications. *Mol. Asp. Med.* 2013, 34, 1–11. [CrossRef] [PubMed]
- He, L.; Ai, Z.; Xiang, Z. The application value of human component analysis technique in the evaluation of nonalcoholic fatty liver disease. *Mod. Hosp.* 2019, 19, 1157–1160, 1165. (In Chinese, with English Abstract)

- 5. Su, B.; Tian, S.; Wang, H. Correlation study of CT images of fatty liver and abdominal fat distribution. *J. Aerosp. Med.* **2017**, *28*, 1043–1047. (In Chinese, with English Abstract)
- 6. Zhang, J.; Qi, Y. Correlation between visceral obesity index and pathology of chronic hepatitis C. *Hebei Pharm.* **2012**, *34*, 369–370. (In Chinese)
- Castro, O. Systemic fat embolism and pulmonary hypertension in sickle cell disease. *Hematol.-Oncol. Clin. N. Am.* 1996, 10, 1289–1303. [CrossRef] [PubMed]
- 8. Despres, J. Body Fat Distribution and Risk of Cardiovascular Disease. *Circulation* **2012**, *126*, 1301–1313. [CrossRef] [PubMed]
- 9. Ohlson, L.O.; Larsson, B.; Svardsudd, K. The Influence of Body Fat Distribution on the Incidence of Diabetes Mellitus: 13.5 Years of Follow-up of the Participants in the Study of Men Born in 1913. *Diabetes* **1985**, *34*, 1055–1058. [CrossRef]
- 10. Kwon, H.W.; Lee, S.M.; Lee, J.W. Association between volume and glucose metabolism of abdominal adipose tissue in healthy population. *Obes. Res. Clin. Pract.* 2017, *11* (Suppl. S1), 133–143. [CrossRef]
- 11. Tong, Y.; Udupa, J.K.; Torigian, D.A. Chest Fat Quantification via CT Based on Standardized Anatomy Space in Adult Lung Transplant Candidates. *PLoS ONE* **2017**, *12*, e0168932. [CrossRef] [PubMed]
- 12. Entwistle, T.R.; Green, A.C.; Fildes, J.E. Adherence to Mediterranean and low-fat diets among heart and lung transplant recipients: A randomized feasibility study. *Nutr. J.* **2018**, *17*, 22. [CrossRef] [PubMed]
- Tong, Y.; Udupa, J.K.; Torigian, D.A. Fat quantification and analysis of lung transplant patients on unenhanced chest CT images based on standardized anatomic space. In Proceedings of the Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging, San Diego, CA, USA, 29 March 2016.
- 14. Anderson, M.R.; Udupa, J.K.; Edwin, E. Adipose tissue quantification and primary graft dysfunction after lung transplantation: The Lung Transplant Body Composition study. *J. Heart Lung Transplant.* **2019**, *38*, 1246–1256. [CrossRef] [PubMed]
- Cho, Y.H.; Do, K.H.; Chae, E.J.; Choi, S.H.; Jo, K.W.; Lee, S.O.; Hong, S.B. Association of Chest CT-Based Quantitative Measures of Muscle and Fat with Post-Lung Transplant Survival and Morbidity: A Single Institutional Retrospective Cohort Study in Korean Population. *Korean J. Radiol.* 2019, 20, 522–530. [CrossRef] [PubMed]
- Dudeja, V.; Misra, A.; Pandey, R.M. BMI does not accurately predict overweight in Asian Indians in northern India. *Br. J. Nutr.* 2001, *86*, 105–112. [CrossRef] [PubMed]
- 17. Burkhauser, R.V.; Cawley, J. Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *J. Health Econ.* **2008**, *27*, 519–529. [CrossRef]
- 18. Heker, M.; Greenspan, H. Joint liver lesion segmentation and classification via transfer learning. arXiv 2020, arXiv:2004.12352.
- 19. Ferdian, E.; Suinesiaputra, A.; Dubowitz, D.J.; Zhao, D.; Wang, A.; Cowan, B.; Young, A.A. 4DFlowNet: Super-Resolution 4D Flow MRI Using Deep Learning and Computational Fluid Dynamics. *Front. Phys.* **2020**, *8*, 138. [CrossRef]
- 20. Bottigli, U.; Cerello, P.; Cheran, S.C.; Delogu, P.; Fantacci, M.E.; Fauci, F.; Golosio, B.; Lauria, A.; Torres, E.L.; Magro, R.; et al. GPCALMA: A Tool For Mammography with A GRID-Connected Distributed Database. *Med. Phys.* **2003**, *682*, 67–72.
- Jafari, M.; Auer, D.P.; Francis, S.T.; Garibaldi, J.M.; Chen, X. DRU-Net: An Efficient Deep Convolutional Neural Network for Medical Image Segmentation. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 1144–1148.
- 22. Zhang, P.; Zhong, Y.; Li, X. ACCL: Adversarial constrained-CNN loss for weakly supervised medical image segmentation. *arXiv* **2020**, arXiv:2005.00328.
- Usman, M.; Lee, B.D.; Byon, S.S.; Kim, S.H.; Lee, B.I.; Shin, Y.G. Volumetric lung nodule segmentation using adaptive ROI with multi-view residual learning. *Sci. Rep.* 2020, 10, 12839. [CrossRef] [PubMed]
- 24. Peng, Y.; Wang, M.; Liu, C.; Jia, H. A preliminary study on the value of CT texture analysis combined with machine learning in auxiliary diagnosis of vertebral occult fractures. *J. Jinan Univ. (Nat. Sci. Med.)* **2020**, *3*, 1–8. (In Chinese, with English Abstract)
- Estrada, S.; Lu, R.; Conjeti, S.; Orozco-Ruiz, X.; Panos-Willuhn, J.; Breteler, M.M.B.; Reuter, M. FatSegNet: A fully automated deep learning pipeline for adipose tissue segmentation on abdominal dixon MRI. *Magn. Reson. Med.* 2020, 83, 1471–1483. [CrossRef]
- 26. Cao, H.; Sheng, B.; Wu, W. Automatic quantitative detection algorithm of abdominal fat based on improved K-Means. *J. Comput. Aided Des. Graph.* **2017**, *29*, 575–583. (In Chinese, with English Abstract)
- 27. Hussein, S.; Green, A.; Watane, A.; Papadakis, G.Z.; Osman, M.M.; Bagci, U. Context Driven Label Fusion for Segmentation of Subcutaneous and Visceral Fat in CT Volumes. *arXiv* 2015, arXiv:1512.04958.
- Irmakci, I.; Hussein, S.; Savran, A.; Kalyani, R.R.; Reiter, D.; Chia, C.W.; Fishbein, K.W.; Spencer, R.G.; Ferrucci, L.; Bagci, U. A Novel Extension to Fuzzy Connectivity for Body Composition Analysis: Applications in Thigh, Brain, and Whole Body Tissue Segmentation. *IEEE Trans. Biomed. Eng.* 2019, 66, 1069–1081. [CrossRef] [PubMed]
- Amer, R.; Nassar, J.; Bendahan, D.; Greenspan, H.; Ben-Eliezer, N. Automatic Segmentation of Muscle Tissue and Inter-muscular Fat in Thigh and Calf MRI Images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019.
- Rundo, L.; Han, C.; Nagano, Y.; Zhang, J.; Hataya, R.; Militello, C.; Tangherloni, A.; Nobile, M.S.; Ferretti, C.; Besozzi, D.; et al. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing* 2019, 365, 31–43. [CrossRef]
- Koitka, S.; Kroll, L.; Malamutmann, E.; Oezcelik, A.; Nensa, F. Fully automated body composition analysis in routine CT imaging using 3D semantic segmentation convolutional neural networks. *Eur. Radiol.* 2021, 31, 1795–1804. [CrossRef] [PubMed]

- 32. Langner, T.; Hedstrom, A.; Morwald, K. Fully convolutional networks for automated segmentation of abdominal adipose tissue depots in multicenter water-fat MRI. *Magn. Reson. Med.* **2019**, *81*, 2736–2745. [CrossRef]
- 33. Masoudi, S.; Anwar, S.M.; Harmon, S.A. Adipose Tissue Segmentation in Unlabeled Abdomen MRI using Cross Modality Domain Adaptation. *arXiv* 2020, arXiv:2005.05761.
- 34. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
- 35. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention gated networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* 2019, *53*, 197–207. [CrossRef] [PubMed]
- Tan, C.; Feng, X.; Long, J.; Geng, L. FORECAST-CLSTM: A New Convolutional LSTM Network for Cloudage Nowcasting. In Proceedings of the IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 9–12 December 2018.
- 37. Yan, K.; Wang, X.; Lu, L.; Summers, R.M. DeepLesion: Automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J. Med. Imaging* **2018**, *5*, 036501. [CrossRef] [PubMed]
- Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. Int. J. Comput. Vis. 2008, 77, 157–173. [CrossRef]
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III 18; Springer: Cham, Switzerland, 2015; pp. 234–241.
- Petit, O.; Thome, N.; Rambour, C.; Themyr, L.; Collins, T.; Soler, L. U-net transformer: Self and cross attention for medical image segmentation. In *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, 27 September 2021, Proceedings 12*; Springer International Publishing: Cham, Switzerland, 2021; pp. 267–276.
- Liu, Q.; Chen, C.; Qin, J.; Dou, Q.; Heng, P. FedDG: Federated Domain Generalization on Medical Image Segmentation via Episodic Learning in Continuous Frequency Space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 1013–1023.
- Zhang, Y.; Liu, H.; Hu, Q. TransFuse: Fusing Transformers and CNNs for Medical Image Segmentation. In Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, 27 September–1 October 2021, Proceedings, Part I 24; Springer International Publishing: Cham, Switzerland, 2021; pp. 14–24.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.