

Article

Hybrid Multiple-Organ Segmentation Method Using Multiple U-Nets in PET/CT Images

Yuta Suganuma ¹, Atsushi Teramoto ^{2,*} , Kuniaki Saito ¹, Hiroshi Fujita ³ , Yuki Suzuki ⁴, Noriyuki Tomiyama ⁵ and Shoji Kido ⁴

¹ Graduate School of Health Sciences, Fujita Health University, Toyoake 470-1192, Japan

² Faculty of Information Engineering, Meijo University, Nagoya 468-8502, Japan

³ Faculty of Engineering, Gifu University, Gifu 501-1193, Japan

⁴ Department of Artificial Intelligence Diagnostic Radiology, Graduate School of Medicine, Osaka University, Osaka 565-0871, Japan

⁵ Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka University, Osaka 565-0871, Japan

* Correspondence: teramoto@meijo-u.ac.jp

Abstract: PET/CT can scan low-dose computed tomography (LDCT) images with morphological information and PET images with functional information. Because the whole body is targeted for imaging, PET/CT examinations are important in cancer diagnosis. However, the several images obtained by PET/CT place a heavy burden on radiologists during diagnosis. Thus, the development of computer-aided diagnosis (CAD) and technologies assisting in diagnosis has been requested. However, because FDG accumulation in PET images differs for each organ, recognizing organ regions is essential for developing lesion detection and analysis algorithms for PET/CT images. Therefore, we developed a method for automatically extracting organ regions from PET/CT images using U-Net or DenseUNet, which are deep-learning-based segmentation networks. The proposed method is a hybrid approach combining morphological and functional information obtained from LDCT and PET images. Moreover, pre-training using ImageNet and RadImageNet was performed and compared. The best extraction accuracy was obtained by pre-training ImageNet with Dice indices of 94.1, 93.9, 91.3, and 75.1% for the liver, kidney, spleen, and pancreas, respectively. This method obtained better extraction accuracy for low-quality PET/CT images than did existing studies on PET/CT images and was comparable to existing studies on diagnostic contrast-enhanced CT images using the hybrid method and pre-training.

Keywords: organ segmentation; PET/CT; U-Net



Citation: Suganuma, Y.; Teramoto, A.; Saito, K.; Fujita, H.; Suzuki, Y.; Tomiyama, N.; Kido, S. Hybrid Multiple-Organ Segmentation Method Using Multiple U-Nets in PET/CT Images. *Appl. Sci.* **2023**, *13*, 10765. <https://doi.org/10.3390/app131910765>

Academic Editor: Athanasios Nikolaidis

Received: 30 August 2023

Revised: 24 September 2023

Accepted: 26 September 2023

Published: 27 September 2023



Copyright: © 2023 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/>).

1. Introduction

Cancer is a major cause of death in many countries and an important obstacle in extending life expectancy. According to the World Health Organization (WHO), cancer is the first or second leading cause of death in people under the age of 70 years in 112 of 183 countries [1].

In today's medical care, various diagnostic imaging examinations, such as computed tomography (CT) and magnetic resonance imaging (MRI), are used for the early detection and treatment of cancer. Notably, PET/CT examinations that scan low-dose computed tomography (LDCT) and positron emission tomography (PET) play important roles in cancer diagnosis. Specifically, LDCT images provide morphological information, such as the shape and location of organs, whereas PET images provide functional information, such as glucose metabolism in organs. ¹⁸F-Fluorodeoxyglucose (FDG), mainly used in PET/CT examinations, is a radioactive tracer that acts as a glucose analog to localize tissues with altered glucose metabolism. Importantly, the amount of FDG accumulated in each organ indicates the function of glucose metabolism, and the standard uptake value

(SUV) is widely used as a semiquantitative index of glucose metabolism in clinical settings. Furthermore, fusion images created from LDCT and PET images can be used to identify the exact anatomical location of abnormal FDG uptake observed on PET images.

Based on the morphological and functional information obtained from LDCT, PET, and fusion images, PET/CT examinations can contribute to the diagnosis of cancer and other diseases. However, PET/CT examinations require the use of a large number of medical images in a single examination, which places a heavy burden on radiologists during diagnosis. This has led to the development of computer-aided diagnosis (CAD) and other image diagnosis-assisted technologies that use artificial intelligence (AI) to reduce the burden on radiologists.

In recent years, with the development of computer technology, CAD methods for various modalities have been actively studied [2–5]; however, few of them have been developed for PET/CT examinations because PET/CT examinations target the whole body and the different patterns of FDG uptake in each organ require different algorithms for detecting and analyzing lesions in different organs. To develop such an algorithm, it is necessary to recognize each organ region. Therefore, we developed an automated organ segmentation technique for PET/CT images as a fundamental technology for developing CAD for each organ.

1.1. Related Works

Many researchers have proposed various methods for organ segmentation in diagnostic CT images. Wolz et al. and Tong et al. proposed an atlas-based organ segmentation method [6,7]. However, the accuracy of organ extraction in this model is highly dependent on the accuracy of the image registration, which requires a wide variety of atlas datasets. Moreover, Gauriau et al. and Criminisi et al. proposed machine learning-based methods [8,9], which are beneficial for detecting and estimating the location of anatomical structures; however, manually set features are important, and setting the appropriate features represents a heavy burden. Recently, segmentation methods based on deep learning that can automatically learn features obtained from image data have become the standard segmentation methods.

Hu et al. proposed an end-to-end 3D convolutional neural network (CNN)-based segmentation method [10]. In that study, the number of training cases changed, suggesting the importance of the number of training cases in deep learning-based methods. Furthermore, Zhou et al. proposed a fully convolutional network (FCN) voting method that performed organ segmentation from axial, coronal, and sagittal cross-sections and combined the results of these three cross-sections for each pixel value [11]. The training data were increased using three cross-sections, and organ segmentation was performed considering the organ information in the body's axial direction. Furthermore, Roth et al. proposed organ segmentation focusing on small organs such as the pancreas and gallbladder, which are difficult to extract, using multiple 3D FCNs in stages [12]. These existing studies have utilized CT images, contrast-enhanced CT images with a sufficient dose on a diagnostic CT scanner, or MR images, and various segmentation techniques that can extract organ regions with high accuracy have been developed.

In contrast, there are few studies on organ segmentation for LDCT images obtained using a PET/CT scanner. In an existing study on LDCT images, Wang et al. proposed a multi-atlas segmentation (MAS) framework [13]. Specifically, the organs were segmented for LDCT images using the MAS method, and the abdominal region of the PET image was automatically cropped. Next, MAS was applied to the fused PET and LDCT images to accurately extract the organs. Notably, Zhang et al. proposed an organ segmentation method using deep learning [14]. Their endeavor involved utilizing two 3DVNets in a stepwise process to achieve more accurate organ extraction from processed LDCT images. However, when compared to studies on diagnostic CT images, the accuracy of extraction in existing PET/CT image studies was inadequate. This emphasizes the need to focus on improving extraction accuracy. LDCT images are acquired without contrast agents and at

low doses; thus, they have a lower resolution than diagnostic CT images, and the margins of the organs are unclear. In addition, because LDCT images are used for attenuation correction of PET images they are captured with a wide field of view (FOV), and the proportion of organs in the FOV is small. Therefore, highly accurate organ segmentation for PET/CT images is more difficult than it is for diagnostic CT images. Therefore, improving the extraction accuracy of organ segmentation methods for PET/CT images is necessary.

1.2. Purpose

In this study, we aimed to improve the extraction accuracy of organ segmentation from PET/CT images using deep learning. We focused on the fact that although LDCT images contain abundant anatomical (morphological) information, the image contrast of soft tissue organs in the abdomen is significantly lower than that of diagnostic CT images. In addition, PET images offer higher image contrast for soft tissue organs based on FDG uptake (functional information); however, their spatial resolution is lower. To address these challenges, we proposed a hybrid organ segmentation method that focuses on morphological and functional information. We used two networks: one using only LDCT images and the other using both LDCT and PET images as inputs.

Moreover, this study used fewer cases to train the networks than did existing studies. Therefore, we aimed to achieve highly accurate organ segmentation on a small number of datasets using a pre-trained model. In many studies using deep learning, ImageNet [15], which consists of natural images, has been used as a pre-training dataset, and good results have been reported [11]. However, within a study utilizing deep learning for medical images, findings indicated that employing a pre-trained dataset (RadImageNet) containing a substantial number of labeled medical images captured across various imaging modalities led to improved accuracy when compared to the use of ImageNet. [16]. Therefore, we applied RadImageNet as a pre-training dataset and compared its accuracy with that of ImageNet.

1.3. Contribution

The contributions of this study are as follows.

We propose a hybrid method using two segmentation networks for PET/CT images, focusing on anatomical information (morphological information) and FDG uptake (functional information). This method enabled us to obtain better extraction accuracy than existing studies of organ segmentation techniques for PET/CT images and equivalent extraction accuracy compared to existing studies for high-quality diagnostic CT images.

We compared the organ extraction accuracy using two pre-trained datasets consisting of natural and medical images. The results show the pre-training's efficacy with limited training data or using medical images to enhance organ segmentation techniques.

2. Methods

2.1. Overview of the Proposed Method

Figure 1 shows a schematic of this study. LDCT images only and both LDCT and PET images were provided to the U-Net, and two types of output were obtained; these were defined as $\text{Candidate}_{\text{LDCT}}$ and $\text{Candidate}_{\text{PET/CT}}$, respectively. $\text{Candidate}_{\text{LDCT}}$ and $\text{Candidate}_{\text{PET/CT}}$ were combined and defined as $\text{Output}_{\text{proposed}}$ for the proposed output region. The accuracy of the organ extraction was evaluated for the $\text{Output}_{\text{proposed}}$.

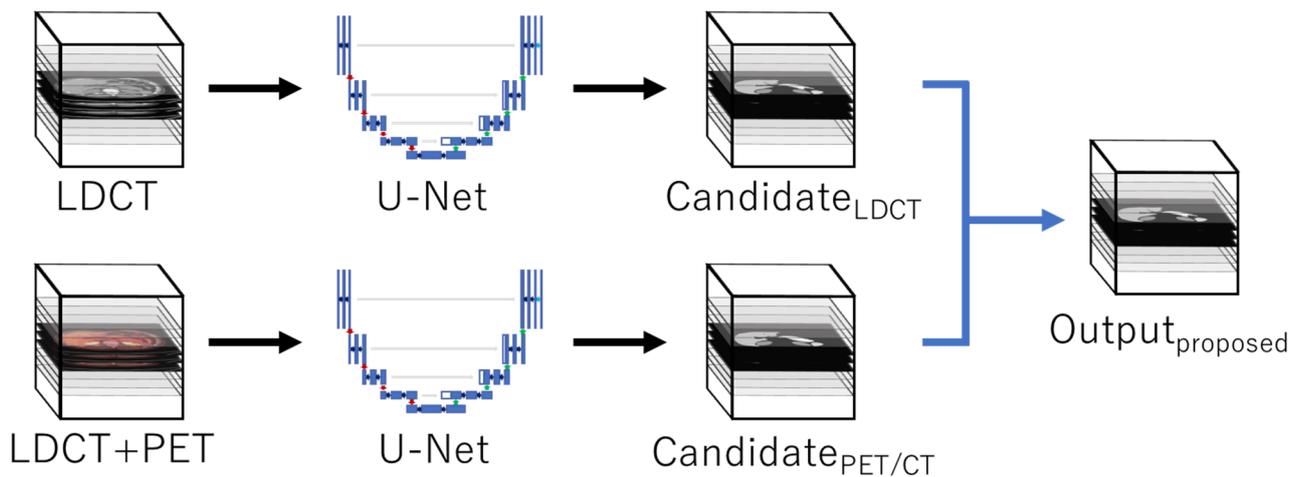


Figure 1. Overview of the proposed method.

2.2. Dataset

2.2.1. Target Data

The target data consisted of LDCT and PET images of 48,092 patients taken between 2006 and 2020 at the Jinseikai MI Clinic. From these cases, 88 healthy subjects' images taken in 2019 were randomly selected, and their data were used. Patients with a history of anatomical impacts and those who had undergone surgical treatment were excluded. The matrix size of LDCT images was 512×512 pixels, and the voxel size was $0.98\text{--}1.17 \times 0.98\text{--}1.17 \times 3.27 \text{ mm}^3$. The matrix size of PET images was 128×128 or 192×192 pixels, and the voxel size was $3.12\text{--}4.67 \times 3.12\text{--}4.67 \times 3.27 \text{ mm}^3$.

2.2.2. Data Preparation

This study used PET/CT scans of the whole body. Therefore, axial images for this study were manually selected such that four organs (liver, kidney, spleen, and pancreas) were included, and the number of slice images per case ranged from 62 to 120. LDCT and PET images saved in the DICOM format were converted to the 8-bit PNG format. Furthermore, the pixel values of the PET images were normalized such that the SUV ranged from 0 to 5. In addition, the PET images were resized to a matrix of 512×512 pixels using linear interpolation. Moreover, the labeled images of each organ were created by a qualified radiological technologist or a student on a training course who had acquired knowledge of anatomy, using a tool developed by the authors. Finally, one radiological technologist checked and modified all organ labels. An example of an organ label is shown in Figure 2.

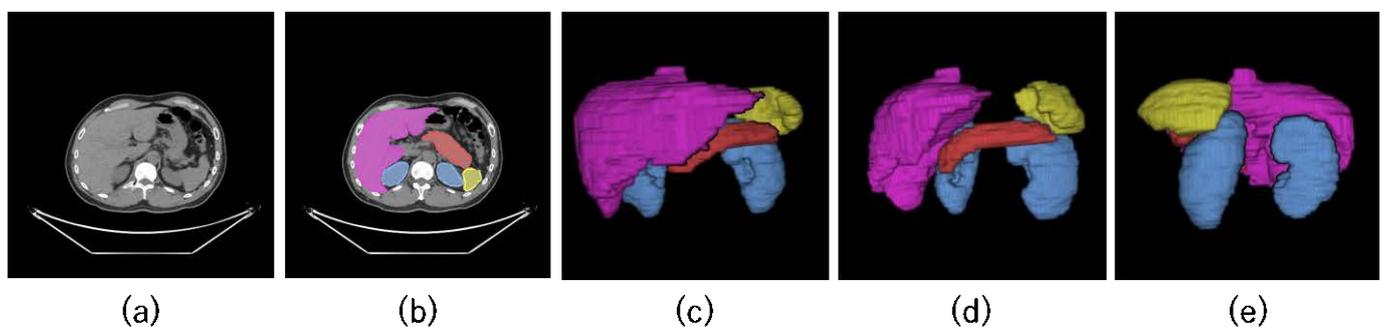


Figure 2. An example of organ labels. (a) LDCT image; (b) Image with LDCT and label overlaid; (c–e) virtual reality (VR) image of the label from multiple angles.

2.3. Organ Segmentation

In this study, we used two types of segmentation networks. The first was a custom version of U-Net [17], which was proposed by Ronneberger et al. in 2015 for segmenting medical images. U-Net consists of three blocks: encoder, bottleneck, and decoder. In general, when multiple convolutional integrations are performed in a CNN, downsampling is likely to result in a loss of spatial information. However, U-Net has a skip connection that combines the feature map obtained by the encoder with that obtained by the decoder, thereby enabling the preservation of local and spatial positional information, which is important in medical imaging. In this study, we used a customized network of the original U-Net. A regularized linear unit (ReLU) was used as the activation function. Batch regularization was also used as a regularization technique to stabilize learning and improve generalization. And maximum pooling was used for downsampling and nearest interpolation for upsampling. Furthermore, LDCT images only and LDCT and PET images were provided to U-Net, and two types of organ regions were outputted. When inputting the LDCT and PET images, each image was divided into separate channels and input into U-Net. However, this network's weights are randomly initialized, which reduces the learning stability for a small number of datasets. Notably, this study used fewer datasets to train the network than did existing studies. Therefore, we aimed to achieve highly accurate organ extraction from a small number of datasets using pre-training.

For the second network, DenseUNet with a DenseNet121 encoder was used for pre-training. DenseNet121 has been reported to perform well as a feature extractor in organ segmentation networks [18]. Moreover, the sizes of the input and output images and the input method in DenseUNet were the same as those in U-Net, as shown in the first example. Notably, in many existing studies using deep learning, ImageNet, which consists of natural images, has been used as a pre-training dataset. However, better results have been obtained for medical studies using a pre-training dataset consisting of medical images. Hence, within this study, we compared the full-scratch model, which begins with random weight initialization, to the fine-tuning model. The latter initializes network weights using both ImageNet and RadImageNet weights.

The combination loss [19] reported by Taghanaki et al. was used as the loss function. Furthermore, combo loss consists of Dice loss and weighted cross-entropy loss and has two parameters, α and β . α adjusts the weights of Dice loss and weighted entropy loss, while β affects weighted cross-entropy loss and can change the sensitivity of the output. In this study, $\alpha = 0.5$, which is similar to that of the study by Taghanaki et al. Furthermore, the combo loss is more strongly penalized for false negatives when the value of β is set to a value higher than 0.5.

PET/CT images were acquired with a wide FOV; therefore, the organ area was smaller than that of the input image. Thus, the output tends to be underestimated in cases with small body sizes and in the pancreas, which has a small organ area and large individual differences. Therefore, we performed a grid search in the range of $\beta = 0.5$ to 1.0 and selected $\beta = 0.6$, which gave the best results.

2.4. Shaping of Output Labels and Composition

LDCT images only and LDCT and PET images were given to U-Net, and two types of candidate regions were obtained. The final organ regions were obtained by combining these two regions. This section describes the combined method in detail. Specifically, two types of morphological processing were used to shape the output regions of the U-Net. The output image was then opened and closed. The $\text{Output}_{\text{proposed}}$ was obtained by combining the candidate regions, $\text{Candidate}_{\text{LDCT}}$ and $\text{Candidate}_{\text{PET/CT}}$, when the input images were LDCT images only and LDCT and PET images, respectively. In this study, $\text{Output}_{\text{proposed}}$ was defined as the region where either $\text{Candidate}_{\text{LDCT}}$ or $\text{Candidate}_{\text{PET/CT}}$ was output ($\text{Candidate}_{\text{LDCT}}$ or $\text{Candidate}_{\text{PET/CT}}$).

2.5. Evaluation Matrices

The extraction accuracy of each organ was evaluated using the three indices (Dice index, sensitivity, and false positive rate) shown in the Formulas (1)–(3) below: the number of voxels in the output region that matched the correct region was defined as a true positive (TP), and the number of voxels in the output region that were not included in the correct region (i.e., the number of voxels in the over-extracted region) was defined as a false positive (FP). Finally, the number of voxels in the correct regions that were not included in the output regions (i.e., the number of voxels in the undetected regions) was defined as a false negative (FN).

$$\text{Dice index} = \frac{2TP}{2TP + FP + FN} \times 100[\%] \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100[\%] \quad (2)$$

$$\text{False positive rate} = \frac{FP}{TP + FP} \times 100[\%] \quad (3)$$

The Dice index indicates the relative volume overlap between the output and correct regions. Furthermore, the sensitivity indicates the percentage of correct outputs among the correct regions, and the FP rate indicates the percentage of outputs that are not in the organ region out of all the outputs.

2.6. Learning Environment and Parameter

In this study, we used LDCT and PET images of 88 cases with four defined organ regions taken at the same facility, and trained and predicted them using a 5-fold cross-validation method. The segmentation models (U-Net and DenseUNet) were created using Keras and TensorFlow, and PCs equipped with NVIDIA RTX A6000 GPUs and AMD Ryzen 9 5900X CPUs were used. The learning rate was 1×10^{-4} , the number of training epochs was 100, the batch size was 8, regularization technique was batch normalization, and the Adam optimization algorithm was used to train the segmentation model. Also, weights were initialized randomly when using U-Net, and weights were initialized utilizing ImageNet or RadImageNet weights when using DenseUNet.

3. Results

The results of multiple segmentation networks confirmed different trends between candidate regions obtained using only LDCT images as inputs ($\text{Candidate}_{\text{LDCT}}$) and those obtained using LDCT and PET images as inputs ($\text{Candidate}_{\text{PET/CT}}$). Table 1 shows the extraction accuracy of $\text{Candidate}_{\text{LDCT}}$, $\text{Candidate}_{\text{PET/CT}}$, and the $\text{Output}_{\text{proposed}}$ full-scratch models trained on four target organs using U-Net with randomly initialized network weights. The mean, standard deviation (SD), median, minimum, and maximum values of the Dice are shown in Table 1. Furthermore, examples of each output are shown in Figure 3.

Table 1. Prediction results of $\text{Candidate}_{\text{LDCT}}$ and $\text{Candidate}_{\text{PET/CT}}$ and $\text{Output}_{\text{proposed}}$ from Full-Scratch.

Organ	Evaluation Metrics	$\text{Candidate}_{\text{LDCT}}$	$\text{Candidate}_{\text{PET/CT}}$	$\text{Output}_{\text{proposed}}$	
Liver	Dice [%]	Mean	94.1	94.2	94.0
		SD	4.1	4.1	4.1
		Median	94.9	94.7	94.7
		Min	59.4	59.4	59.5
		Max	97.0	96.9	96.7
	Sensitivity [%]	95.4	95.6	96.6	
	False positive rate [%]	7.1	7.2	8.4	

Table 1. Cont.

Organ	Evaluation Metrics		Candidate _{LDCT}	Candidate _{PET/CT}	Output _{proposed}
Kidney	Dice [%]	Mean	94.3	94.6	94.1
		SD	3.7	3.1	3.6
		Median	95.6	95.5	95.2
		Min	70.7	76.0	72.1
		Max	97.5	97.8	97.3
		Sensitivity [%]	96.3	96.7	97.7
		False positive rate [%]	7.5	7.3	9.1
Spleen	Dice [%]	Mean	91.2	91.2	90.9
		SD	8.3	8.4	8.3
		Median	92.9	92.8	92.6
		Min	20.0	20.0	20.7
		Max	96.6	96.5	96.4
		Sensitivity [%]	93.9	93.8	95.4
		False positive rate [%]	11.0	10.9	12.9
Pancreas	Dice [%]	Mean	69.5	70.5	71.9
		SD	16.2	15.7	14.0
		Median	73.1	75.0	75.9
		Min	0.0	22.3	21.7
		Max	89.1	89.6	88.4
		Sensitivity [%]	67.8	69.9	75.7
		False positive rate [%]	25.0	25.8	29.2

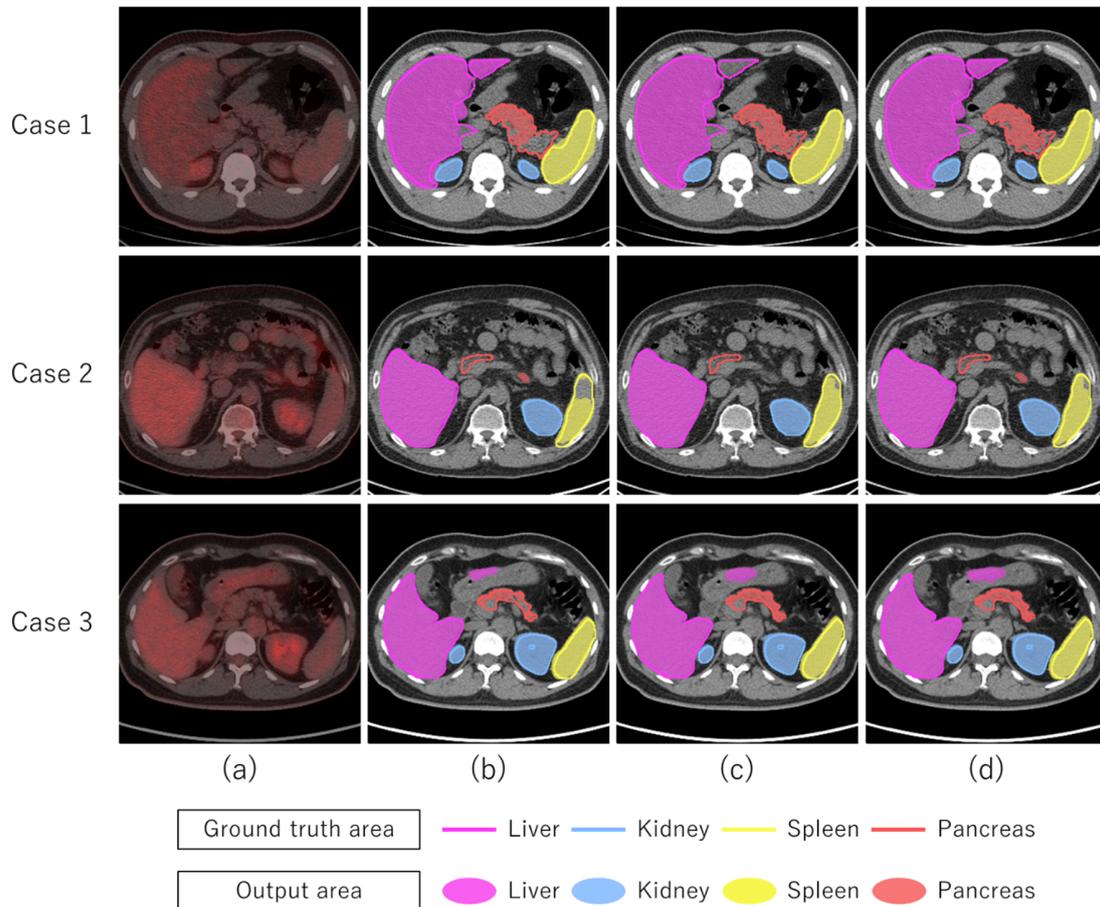


Figure 3. Output areas from full-scratch. (a) PET/CT image; (b) Candidate LDCT; (c) Candidate PET/CT; (d) Output proposed.

Additionally, we compared the training and prediction results using three segmentation models (full-scratch, ImageNet, and RadImageNet). Table 2 presents the extraction accuracies, where full-scratch is the Output_{proposed} network initialized with random weights using U-Net, and ImageNet and RadImageNet are the Output_{proposed} networks initialized with ImageNet and RadImageNet, respectively, using DenseUNet. In Table 1, the Mean, SD, Median, Min, and Max indicate the Dice index’s mean, standard deviation, median, minimum, and maximum values, respectively. Examples of the outputs of the three models are shown in Figure 4.

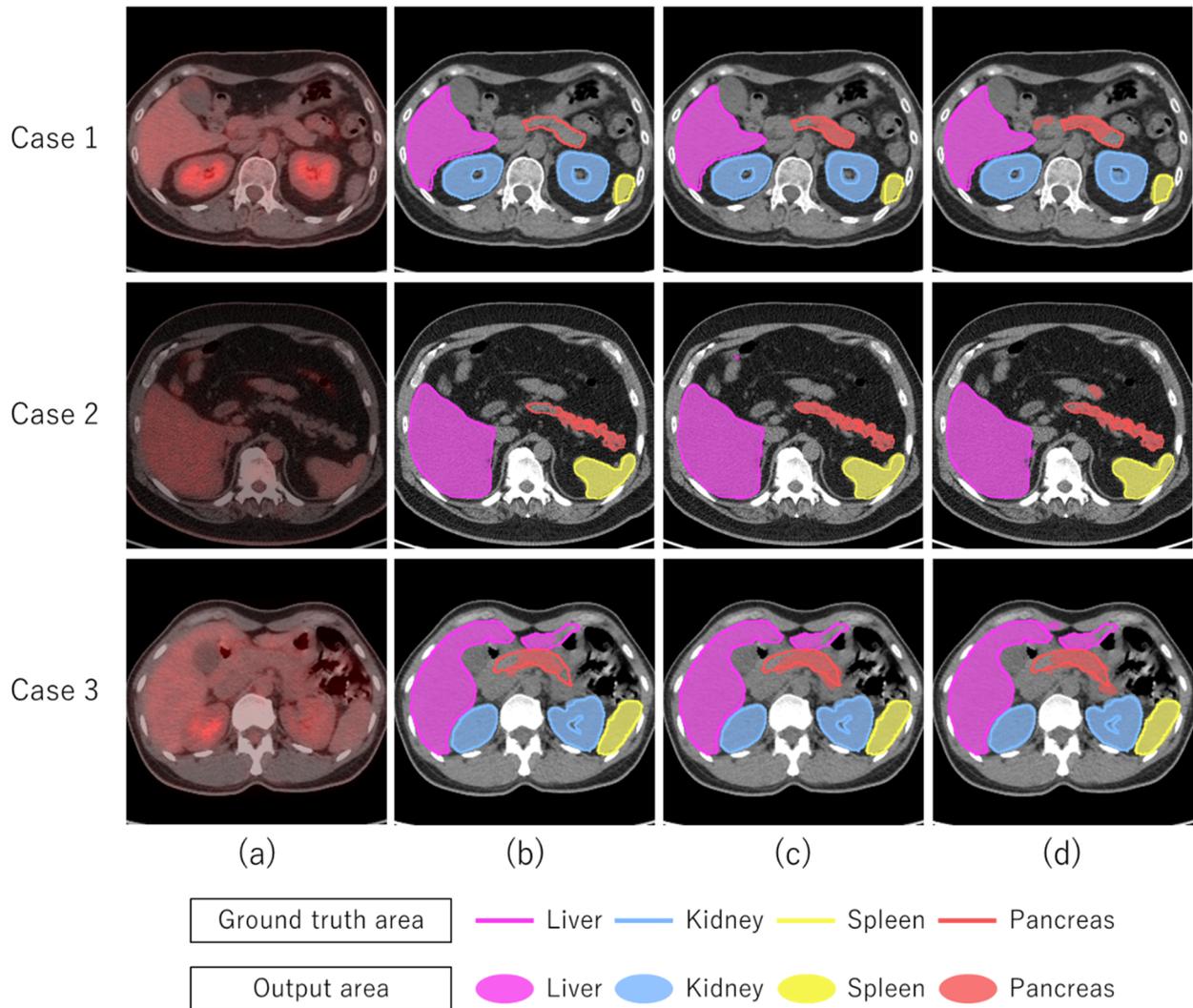


Figure 4. Output areas from three models. (a) PET/CT image; (b) Full-Scratch; (c) ImageNet; (d) RadImageNet.

Table 2. Prediction results of full-scratch, ImageNet, and RadImageNet.

Organ	Evaluation Metrics	Full-Scratch	ImageNet	RadImageNet
Liver	Mean	94.0	94.1	93.7
	SD	4.1	4.2	4.1
	Median	94.7	94.9	94.6
	Min	59.5	58.9	59.4
	Max	96.7	96.8	96.7
	Sensitivity [%]	96.6	97.0	96.8
	False positive rate [%]	8.4	8.6	9.1

Table 2. Cont.

Organ	Evaluation Metrics	Full-Scratch	ImageNet	RadImageNet	
Kidney	Dice [%]	Mean	94.1	93.9	93.1
		SD	3.6	4.1	4.1
	Sensitivity [%]	Median	95.2	95.2	94.5
		Min	72.1	71.1	72.1
		Max	97.3	97.1	97.1
	False positive rate [%]	Sensitivity [%]	97.7	98.0	97.9
		False positive rate [%]	9.1	9.6	10.9
Spleen	Dice [%]	Mean	90.9	91.3	90.3
		SD	8.3	8.2	7.9
	Sensitivity [%]	Median	92.6	92.6	92.0
		Min	20.7	20.6	24.1
		Max	96.4	96.3	95.8
	False positive rate [%]	Sensitivity [%]	95.4	95.8	95.6
		False positive rate [%]	12.9	12.6	14.2
Pancreas	Dice [%]	Mean	71.9	75.1	70.6
		SD	14.0	12.5	15.8
	Sensitivity [%]	Median	75.9	78.9	75.2
		Min	21.7	25.1	3.6
		Max	88.4	89.4	86.4
	False positive rate [%]	Sensitivity [%]	75.7	82.3	77.3
		False positive rate [%]	29.2	29.3	33.4

4. Discussion

Figure 3 shows the difference between Candidate_{LDCT} and Candidate_{PET/CT} in terms of output regions. In the axial image at the end of the organ, the organ region was correctly extracted on Candidate_{LDCT}, whereas it tended not to be extracted on Candidate_{PET/CT}. The reason for the low extraction accuracy of Candidate_{PET/CT} is thought to be the imaging mechanism of the PET images, which are less sensitive at the edges of organs, leading to a lower FDG uptake. Therefore, we believe that the organ region was not extracted from the axial image at the ends of organs in Candidate_{PET/CT}-given PET images. An example of this in the liver is shown in Case 1 of Figure 3. Here, the liver was divided into two regions in the axial image, and the liver in the smaller organ regions showed lower FDG uptake in the PET image than did that in the larger organ region. Thus, we believe that the liver in the smaller organ region was not correctly extracted in the Candidate_{PET/CT} because of this effect. Furthermore, because LDCT images were captured at low doses, streak artifacts were sometimes observed in cases with large body sizes, such as Cases 1 and 2 in Figure 3. In such cases, the spleen and pancreas were extracted using Candidate_{PET/CT}, but could not be extracted using Candidate_{LDCT}. These results suggest that morphological information, such as organ shape and CT values, were used in Candidate_{LDCT}, whereas functional information, such as glucose metabolism, was mainly used in Candidate_{PET/CT} for organ segmentation.

Moreover, in this study, the regions that were extracted in either of the two types of candidate regions were defined as Output_{proposed}. As shown above, different regions were extracted from the two types of candidate regions, and some trends in Output_{proposed} were observed. As shown in Table 1, there was no significant difference in the Dice index of Output_{proposed} for the liver, kidney, and spleen between the two candidate regions. However, there was a slight increase in sensitivity and FP rate. In contrast, the FP rate in the pancreas increased, but the Dice index increased by 2.4% and 1.4%, and the sensitivity increased by 7.9% and 5.8% compared to Candidate_{LDCT} and Candidate_{PET/CT}. These results confirmed that the extracted pancreas region was significantly increased in Output_{proposed} in comparison with the two candidate regions. Thus, we believe organ segmentation focusing on morphological and functional information was performed for the PET/CT images in Output_{proposed}, in which Candidate_{LDCT} and Candidate_{PET/CT} were combined.

However, the accuracy of organ extraction is lower in patients with smaller body sizes. Figure 5 shows an example of the extraction results for the case with small body size. The reason for the low extraction accuracy of the pancreas is thought to be the unclear margins of the organ, which is a feature of LDCT images. In addition, because the images were captured with a wide FOV, the organ regions in the images were small, and the features of the pancreas were likely to be lost. For the spleen, only the extraction accuracy of the Candidate_{PET/CT} was low. The axial image in Figure 5 shows the edge of the spleen, which was not correctly extracted because of low FDG uptake.

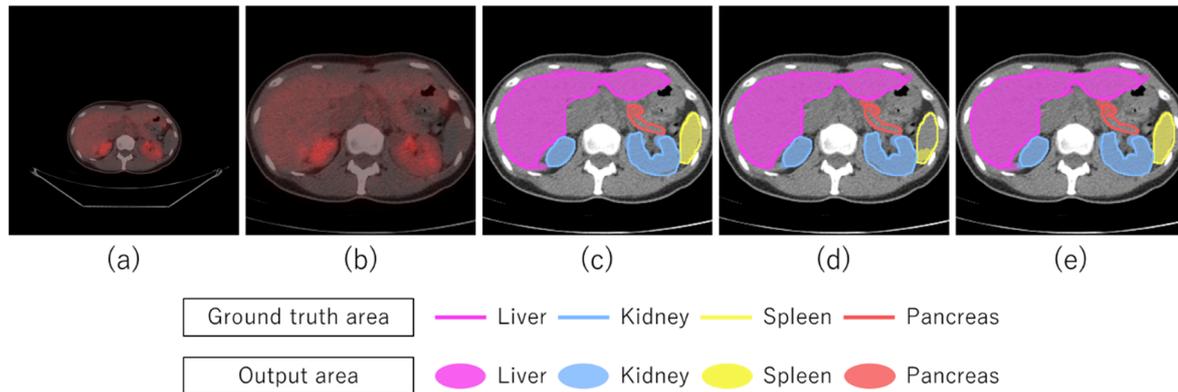


Figure 5. Output areas with low extraction accuracy. (a) Input image; (b) PET/CT image; (c) Candidate_{LDCT}; (d) Candidate_{PET/CT}; (e) Output_{proposed}.

Furthermore, this study used fewer datasets than the existing studies. Therefore, we compared the extraction accuracy using pre-trained datasets to improve the learning stability and increase the generalization ability. As shown in Table 2, when evaluated using the Dice index, there was no significant difference between full-scratch and ImageNet for the liver, kidney, and spleen; moreover, the RadImageNet was lower than that of the other two models. For the pancreas, ImageNet was 3.2% and 4.5% more accurate than full-scratch and RadImageNet, respectively; furthermore, ImageNet showed only 3.2% and 4.5% higher Dice indices than did full-scratch and RadImageNet, respectively. This is because ImageNet and RadImageNet were better at improving sensitivity than full-scratch for all organs, and were able to extract more regions of the organs. However, the FP rate increased significantly only for RadImageNet. Notably, Case 1 in Figure 4 is a small-bodied case in which the pancreas was not extracted correctly by full-scratch, whereas ImageNet and RadImageNet partially extracted it. However, FPs were observed in the pancreas using RadImageNet. In conclusion, ImageNet performs best among the three segmentation models. However, in contrast to the other organs, the pancreas had a high average of the FP rate in all three models. Nevertheless, the average FP rate was small compared to the average of the Dice index, and we believe that pancreas extraction performance has achieved acceptable accuracy.

Several researchers have proposed various methods for organ segmentation, which is an important technique in the medical field. However, it is difficult to directly compare them because the number of training datasets used in each study and the quality of the images differ depending on the target modality. Table 3 presents existing studies using various methods. The proposed method showed the best results with the highest extraction accuracy in the experiments (DenseUNet, which was pre-trained using ImageNet). Notably, the proposed method achieved better extraction accuracy for all organs than did existing methods using PET/CT images. In particular, the extraction accuracy of the proposed method was approximately 15% higher than that of the other methods for the pancreas. Furthermore, the extraction accuracy of the proposed method was comparable to those of existing studies that used diagnostic CT images of four different organs. However, there are existing studies that have achieved better extraction accuracy than the proposed

method for each of the four organs. We believe that the reason for this is that LDCT and PET images have significantly lower image contrast and spatial resolution than diagnostic contrast-enhanced CT images used in existing studies. Nevertheless, this method provided acceptable extraction accuracy because it used not only the morphological information obtained from the LDCT images but also the functional information obtained from the PET images. Moreover, soft-tissue organs neighboring other organs, such as the pancreas, have better contrast than LDCT images do, which is considered to provide useful information for organ segmentation.

Table 3. Comparison with existing studies using the Dice index.

Author	Methods			Dataset Contrast Enhance	Number of Cases	Dice Index [%]			
	Model	Validation	Modality			Liver	Kidney	Spleen	Pancreas
Tong et al. (2015) [7]	Atlas	LOO	D-CT	CE	150	94.9	93.6	92.5	71.1
Hu et al. (2017) [10]	3D CNN	5-fold CV	D-CT	Mixed	140	96.0	95.4	94.2	—
Roth et al. (2017) [12]	3D FCN	Testing	D-CT	CE	331	95.4	—	92.8	82.2
Gibson et al. (2018) [20]	DenseVNet	9-fold CV	D-CT	Mixed	90	95	93 *	95	75
Wang et al. (2019) [13]	Multi Atlas	LOO	PET/CT	NCE	69	88	79	74	—
Zhang et al. (2022) [14]	VNet	Hold out	PET/CT	NCE	175	90.7	89.9	89.1	60.3
Proposed method	DenseUNet	5-fold CV	PET/CT	NCE	88	94.1	93.9	91.3	75.1

LOO: leave-one-out; CV: cross-validation; Testing: use of cases taken at different institutions for learning and prediction; D-CT: diagnostic-CT; CE: contrast agents were used in all cases; NCE: contrast agents were not used in all cases; Mixed: a mixture of cases in which contrast agents were used and those in which they were not.
* Extracted from the left kidney alone.

The proposed method demonstrates enhanced organ segmentation capabilities on low-quality and contrast PET/CT images, outperforming existing PET/CT studies in terms of extraction accuracy. Furthermore, its extraction accuracy aligns closely with that of established studies focusing on contrast-enhanced CT images for diagnostic applications.

However, this study had three limitations. First, the data set used in this study was obtained at a single institution. Therefore, it is necessary to evaluate its robustness by conducting additional studies using data obtained at multiple institutions with different equipment and imaging protocols. However, since there is no open data set of PET/CT images for which organ labels have been created, it will be necessary to collaborate with other institutions to create image data and organ labels to evaluate the robustness and feasibility of the proposed method. Second, two-dimensional processing was used for organ segmentation. By extending the proposed method to 3D processing, information regarding the direction of the body axis can be retained, and more accurate organ extraction is expected. Third, only U-Net or DenseUNet was used in this study, thus the validation is limited. Today, convolutional neural network architectures for medical-image segmentation are being developed. Therefore, we believe that the effectiveness of the proposed method can be further proven by using those state-of-the-art architectures. Furthermore, in clinical practice, the extraction of organ regions is used in treatment planning for radiotherapy. Therefore, as a future study, we would like to compare the results of the proposed method with the accuracy of organ region extraction performed by trained workers in order to conduct a quantitative evaluation more in line with clinical practice.

5. Conclusions

We developed a hybrid method for organ segmentation of PET/CT images. Furthermore, we pre-trained using ImageNet and RadImageNet, and compared the results. The best accuracy was obtained using DenseUNet, which was pre-trained with ImageNet. Additionally, by using morphological and functional information obtained from LDCT and PET images, the extraction accuracy was better than that of existing studies using PET/CT images and was comparable to that of existing studies using contrast-enhanced CT images for diagnostic purposes. Therefore, it was suggested that the proposed method could be a fundamental technology for CAD development such as in lesion detection for PET/CT images.

Author Contributions: Conceptualization, Y.S. (Yuta Suganuma) and A.T.; formal analysis, Y.S. (Yuta Suganuma) and A.T.; methodology, Y.S. (Yuta Suganuma) and A.T.; software, Y.S. (Yuta Suganuma) and A.T.; writing—original draft preparation, Y.S. (Yuta Suganuma) and A.T.; writing—review and editing, K.S., H.F., Y.S. (Yuuki Suzuki), N.T. and S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Osaka University (HM19-549).

Informed Consent Statement: Informed consent was obtained via an opt-out process at the Osaka University and Jinseikai MI Clinic, and all data were anonymized.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to retention of patient information.

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

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Teramoto, A.; Fujita, H.; Yamamuro, O.; Tamaki, T. Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique. *Med. Phys.* **2016**, *43*, 2821–2827. [[CrossRef](#)] [[PubMed](#)]
3. Alakwaa, W.; Nassef, M.; Badr, A. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2017**, *8*, 409–417. [[CrossRef](#)]
4. Trebeschi, S.; van Griethuysen, J.J.M.; Lambregts, D.M.J.; Lahaye, M.J.; Parmar, C.; Bakers, F.C.H.; Peters, N.H.G.M.; Beets-Tan, R.G.H.; Aerts, H.J.W.L. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci. Rep.* **2017**, *7*, 5301. [[CrossRef](#)] [[PubMed](#)]
5. Salama, W.M.; Aly, M.H. Deep learning in mammography images segmentation and classification: Automated CNN approach. *Alex. Eng. J.* **2021**, *60*, 4701–4709. [[CrossRef](#)]
6. Wolz, R.; Chu, C.; Misawa, K.; Fujiwara, M.; Mori, K.; Rueckert, D. Automated abdominal multi-organ segmentation with subject-specific atlas generation. *IEEE Trans. Med. Imaging* **2013**, *32*, 1723–1730. [[CrossRef](#)] [[PubMed](#)]
7. Tong, T.; Wolz, R.; Wang, Z.; Gao, Q.; Misawa, K.; Fujiwara, M.; Mori, K.; Hajnal, J.V.; Rueckert, D. Discriminative dictionary learning for abdominal multi-organ segmentation. *Med. Image Anal.* **2015**, *23*, 92–104. [[CrossRef](#)] [[PubMed](#)]
8. Gauriau, R.; Cuingnet, R.; Lesage, D.; Bloch, I. Multi-organ localization with cascaded global-to-local regression and shape prior. *Med. Image Anal.* **2015**, *23*, 70–83. [[CrossRef](#)] [[PubMed](#)]
9. Criminisi, A.; Robertson, D.; Konukoglu, E.; Shotton, J.; Pathak, S.; White, S.; Siddiqui, K. Regression forests for efficient anatomy detection and localization in computed tomography scans. *Med. Image Anal.* **2013**, *17*, 1293–1303. [[CrossRef](#)] [[PubMed](#)]
10. Hu, P.; Wu, F.; Peng, J.; Bao, Y.; Chen, F.; Kong, D. Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 399–411. [[CrossRef](#)] [[PubMed](#)]
11. Zhou, X.; Takayama, R.; Wang, S.; Hara, T.; Fujita, H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Med. Phys.* **2017**, *44*, 5221–5233. [[CrossRef](#)]
12. Roth, H.R.; Oda, H.; Hayashi, Y.; Oda, M.; Shimizu, N.; Fujiwara, M.; Misawa, K.; Mori, K. Hierarchical 3D Fully Convolutional Networks for Multi-organ Segmentation. *arXiv* **2017**, arXiv:1704.06382.
13. Wang, H.; Zhang, N.; Huo, L.; Zhang, B. Dual-modality multi-atlas segmentation of torso organs from [¹⁸F]FDG-PET/CT images. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *14*, 473–482. [[CrossRef](#)]

14. Zhang, J.; Wang, Y.; Liu, J.; Tang, Z.; Wang, Z. Multiple organ-specific cancers classification from PET/CT images using deep learning. *Multimed. Tool Appl.* **2022**, *81*, 16133–16154. [[CrossRef](#)]
15. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
16. Mei, X.; Liu, Z.; Robson, P.M.; Marinelli, B.; Huang, M.; Doshi, A.; Jacobi, A.; Cao, C.; Link, K.E.; Yang, T.; et al. RadImageNet: An open radiologic deep learning research dataset for effective transfer learning. *Radiol. Artif. Intell.* **2022**, *4*, e210315. [[CrossRef](#)] [[PubMed](#)]
17. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci.* **2015**, *9351*, 234–241. [[CrossRef](#)]
18. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.W.; Heng, P.A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* **2018**, *37*, 2663–2674. [[CrossRef](#)] [[PubMed](#)]
19. Taghanaki, S.A.; Zheng, Y.; Kevin Zhou, S.K.; Georgescu, B.; Sharma, P.; Xu, D.; Comaniciu, D.; Hamarneh, G. Combo loss: Handling Input and Output Imbalance in multi-organ Segmentation. *Comput. Med. Imaging Graph.* **2019**, *75*, 24–33. [[CrossRef](#)] [[PubMed](#)]
20. Gibson, E.; Giganti, F.; Hu, Y.; Bonmati, E.; Bandula, S.; Gurusamy, K.; Davidson, B.; Pereira, S.P.; Clarkson, M.J.; Barratt, D.C. Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Trans. Med. Imaging* **2018**, *37*, 1822–1834. [[CrossRef](#)] [[PubMed](#)]

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.