

Research Article

An End-to-End Data-Adaptive Pancreas Segmentation System with an Image Quality Control Toolbox

Yan Zhu ,¹ Peijun Hu,² Xiang Li,^{3,4} Yu Tian,¹ Xueli Bai,^{3,4} Tingbo Liang,^{3,4} and Jingsong Li ,^{1,2}

¹Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou 310027, China

²Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou 311100, China

³Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310006, China

⁴Zhejiang Provincial Key Laboratory of Pancreatic Disease, Hangzhou 310006, China

Correspondence should be addressed to Jingsong Li; ljs@zju.edu.cn

Received 6 June 2022; Revised 18 August 2022; Accepted 24 November 2022; Published 24 January 2023

Academic Editor: Weihua Yang

Copyright © 2023 Yan Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of radiology and computer technology, diagnosis by medical imaging is heading toward precision and automation. Due to complex anatomy around the pancreatic tissue and high demands for clinical experience, the assisted pancreas segmentation system will greatly promote clinical efficiency. However, the existing segmentation model suffers from poor generalization among images from multiple hospitals. In this paper, we propose an end-to-end data-adaptive pancreas segmentation system to tackle the problems of lack of annotations and model generalizability. The system employs adversarial learning to transfer features from labeled domains to unlabeled domains, seeking a dynamic balance between domain discrimination and unsupervised segmentation. The image quality control toolbox is embedded in the system, which standardizes image quality in terms of intensity, field of view, and so on, to decrease heterogeneity among image domains. In addition, the system implements a data-adaptive process end-to-end without complex operations by doctors. The experiments are conducted on an annotated public dataset and an unannotated in-hospital dataset. The results indicate that after data adaptation, the segmentation performance measured by the dice similarity coefficient on unlabeled images improves from 58.79% to 75.43%, with a gain of 16.64%. Furthermore, the system preserves quantitatively structured information such as the pancreas' size and volume, as well as objective and accurate visualized images, which assists clinicians in diagnosing and formulating treatment plans in a timely and accurate manner.

1. Introduction

Pancreatic cancer is a malignant tumor and is recognized with a low survival rate, and pancreatic diseases are characterized by rapid occurrence and continuous progression [1, 2]. Even among all types of cancer, it is one of the most dangerous and deadly. It is estimated that, in 2021, statistically, about 60,430 new cases of pancreatic cancer would be diagnosed in the US, and 48,220 people would die from this disease [3]. Because pancreatic cancer is difficult to diagnose in its early stage, the rate of diagnosis is almost

as high as the mortality rate. If not treated timely, a significant portion of patients with pancreas disease would be diagnosed with metastatic symptoms [4]. For pancreatic cancer, the common treatment options are surgical resection, neoadjuvant radiotherapy, radiotherapy, and chemotherapy [5, 6], of which surgery is acknowledged to achieve a better prognosis but with somewhat invasiveness. Regardless of the chosen treatment option, precise localization and segmentation of the pancreas are crucial for physicians to diagnose and assess the patient's condition early in the treatment phase [7].

Computed tomography (CT) and magnetic resonance imaging (MRI) are particularly important examination procedures and tools in diagnosing pancreatic diseases [8]. CT, especially contrast-enhanced CT, is the first choice for pancreatic examinations in hospitals and has the advantages of more rapid imaging and clearer strips than MRI. However, the irregularity and variability of the pancreas in morphology and low contrast in its surrounding tissues lead to a high demand for experience and prior knowledge of radiologists in the early diagnosis of pancreatic diseases [9–12]. Nowadays, precision medicine requires the clinical process upgrading from qualitative observation to quantitative analysis and diagnosis [13, 14]. Therefore, accurate automatic segmentation of pancreatic tissues in CT images can greatly accelerate the early diagnosis process for physicians and assist in the appropriate treatment process.

In recent years, with technological breakthroughs in deep learning, computer-aided pancreas segmentation has yielded a series of promising methodological studies [15–26]. Limited by the small object characteristics of the pancreas, several studies have employed the two-stage cascade approach for semantic segmentation [7, 26]. These methods first locate the pancreas in the entire CT sequence with bounding boxes and then perform pixel-level segmentation of the pancreas on the basis of box localization. nnU-Net adopts a two-stage strategy to perform segmentation on pancreatic healthy tissues and lesion tissues autonomously by configuring itself and achieves the best accuracy in the Medical Segmentation Decathlon challenge [27]. Zhu et al. [15], Fu et al. [17], and Oktay and Chen [18] introduced attention modules and spatial information to a three-dimensional (3D) segmentation structure to capture the consistency information of pancreatic tissue in CT images. More recently, a series of AI-based semantic segmentation or object detection methods has been applied to clinical practice for quantitative data measurement and morphometric analysis [28–30]. For example, pancreatic fistula prediction after pancreaticoduodenal surgery is based on quantitative pancreatic volume measurements in CT images, and gallbladder resection is guided by fat pairs from quantitative measurements of the pancreas in CT images.

However, the more challenging issue in the clinical application of computer-aided pancreas segmentation is that the heterogeneity of across domain images leads to poor generalization of models [31]. This is manifested by the fact that supervised models trained on a single dataset, even if trained with accurate expert manual annotations, are subject to significant model inference errors once they are deployed to other medical centers [32–38]. Normalized images are essential for good performance of deep learning models. The variation in medical images in terms of populations, scanning devices, scanning parameters, or imaging protocols will lead to varying quality [33, 39, 40]. This heterogeneity exists in both CT and MRI images [33, 34, 36]. Besides, pancreatic disease and pancreatic cancer dramatically change the morphology of the pancreatic tissue, specifically demonstrated by diffuse enlargement, inhomogeneous density, and

ambiguous boundaries, which make the data quality inconsistent among medical centers [41]. Liu et al. [33] and Wang et al. [36] illustrated the heterogeneity across medical sites in terms of the patient cohort and image quality. Several research studies on medical image segmentation have previously demonstrated significant performance degradation of single-site models when deployed to other sites. Lerousseau and Xiao [42] proposed a new weakly supervised multi-instance learning method as a tool for pancreas tumor segmentation which achieved promising performance by taking full advantage of less annotated data at the pixel level. However, this method yielded about 15–26% performance degradation when it was tested for other publicly available datasets. Obviously, interdomain generalization greatly limits the clinical application of automatic pancreas segmentation models. In the field of natural images, several studies have addressed this issue by using transfer learning or federation learning algorithms [43–47]. However, the effectiveness of approaches based on natural images is unsatisfactory or even worse for CT images due to the gap between the two types of images [48]. There is one research study on semisupervised segmentation pancreas tasks on the NIH-TCIA dataset and achieved a dice similarity coefficient score of 78.27% by using a portion of annotated images [49]. Moreover, to date, few research studies related to unsupervised segmentation or domain adaptation of the pancreas have been conducted.

Therefore, to tackle this serious real-clinical problem, we propose an end-to-end data-adaptive pancreas segmentation system with an image quality control toolbox in this paper. The system focuses on the pancreas segmentation model construction with heterogeneous cross-domain CT images in the presence of insufficient annotations in medical centers. The main contributions of this paper are summarized as follows: (1) The system utilizes adversarial learning to construct data-adaptive segmentation models with the assistance of domain discriminators. Besides, we employ a collaboration center to perform feature-level transfer learning without data sharing across domains. (2) A multifunctional image quality control toolbox is designed to standardize the quality of images from various medical centers in terms of the intensity range, field of view, region of interest, etc. (3) The system works in an end-to-end mode, which only requires physicians to select images, set up personalized parameters, and then wait for automatic model construction and inferences on unlabeled data. (4) The system offers a variety of pancreas-related features including textual information and imaging results, which can assist physicians in quantitative and precise clinical diagnosis of the pancreas.

We experimentally demonstrate the effectiveness of the system on a public dataset and an in-hospital dataset and validate the robustness of the system on a small dataset. The data-adaptive pancreas segmentation system we developed is able to diagnose a larger number of people quickly and effectively in the clinical practice and to obtain meaningful pancreas segmentation results.

2. Methods

Accurate segmentation of pancreatic tissue is an essential stage in clinical diagnosis of pancreatic diseases. The variability of CT images from different medical centers affects the generalizability of automatic pancreatic segmentation tools. Subtle differences in image features result in sudden decreases in segmentation accuracy for deep learning models. To address such a problem, improvements can be carried out in terms of both data alignment and segmentation model construction methods, respectively. On the one hand, some image quality control approaches for pancreas CT images are used to process different source images. On the other hand, the automatic pancreatic segmentation model should have the ability to adapt to variations in the data domain in terms of the methodology and system design. To address these issues, a novel end-to-end data-adaptive segmentation system for the pancreas with an image control toolbox is proposed for pancreatic data quality normalization and assisting in pancreas segmentation model generalization. In this chapter, the overall framework of the system and the construction and functions of each module are shown.

2.1. Framework. The overall framework of the proposed data-adaptive pancreas segmentation system embedded with a data quality control toolbox is shown in Figure 1. The system consists of two parts: the local clients of medical centers and a collaborative center on the cloud server. The primary tasks of the local client are integrating and processing data, constructing segmentation models, and visualizing segmentation results. The collaborative center is mainly responsible for transfer learning of image features among multiple medical centers.

The medical center client consists of four modules, namely, data organization module, image quality control module, segmentation module, and visualization module. The data organization module is mainly operated by imaging physicians to establish the patients' cohort to study and extract the pancreas CT images from the in-hospital database. The image quality control module performs standardized preprocessing on the previously selected pancreas CT images in order to normalize the data and reduce the variation of images from different sources. The segmentation module is mainly equipped with the predefined semantic segmentation model in the system to train annotated data and cooperates with the collaborative center to assist in the construction of the data-adaptive segmentation model for unannotated data. Besides, the segmentation module uses the well-trained model to predict the mask of pancreas tissues in input images. The visualization module presents multidimensional results of the segmented pancreas tissues in terms of visualized images and structured text.

The collaborative center mainly contains feature discriminators. The feature transmission and learning process between the medical center and the collaborative center is encapsulated as the transfer learning module. The transfer learning module mainly accomplishes adversarial learning

between image features from multiple centers. The feature discriminator optimizes the segmentation network of unlabeled images by balancing the Nash equilibrium of the domain classification loss and the segmentation loss. Thus, the segmentation network is adaptive to the new images without the need of annotation.

2.2. Image Quality Control Toolbox. The image quality control toolset provides a variety of image processing means to standardize various qualities of images from multiple sources, mainly including the intensity value cutoff, rotation augmentation, and superresolution reconstruction. The CT intensity of abdominal organs is in the range of (-160, 240) HU, and the range for the pancreas is kept in (-100, 240) HU. This scale preserves pancreatic tissue features and removes background information. Rotational augmentation refers to amplify CT images by rotating them at the axial plane with degrees in the range of ($\pm 5^\circ$, $\pm 10^\circ$). This operation is not performed in the ordinary sense of augmenting data to improve model performance but rather to attenuate the angle bias introduced by the field of view or body position during scanning so as to eliminate heterogeneity. Super-resolution reconstruction could effectively improve the image quality, thus reducing the quality inconsistency caused by scanning devices, imaging protocols, slice thickness, and so on. In this study, it is concerned that the effective abdominal region in CT images fluctuates because of different scanning fields of view, so the first step in superpixel reconstruction is framing out the region of interest. Pancreatic CT images were binarized to measure image region properties. Then, the maximum connected region containing the region of interest was found by the region-growing algorithm. The rectangular area bounded by the diagonal vertices of the maximum connected region in the image is considered the valid abdominal area. Then, the truncated 3D volume is interpolated in 3D cubic interpolation (system default settings) and reconstructed to the same resolution to feed into the network. The default size is 512×512 resolution in an axial plane and 1 mm thickness in a sagittal direction. Furthermore, the system provides various image interpolation methods to system users. The reconstruction algorithm library includes nearest neighbor interpolation, bilinear interpolation, Lanczos interpolation, and bicubic interpolation to support multiple requirements for medical studies.

2.3. Segmentation Module. The segmentation module is embedded with a deep learning semantic segmentation network applicable to the pancreas segmentation task. This module is mainly responsible for the training of labeled data models, the construction of adaptive unlabeled data models in cooperation with discriminators, and the prediction of pancreas masks for input samples.

In this paper, a 3D ResNet structure integrated with an attention mechanism is designed as the backbone model of the segmentation module. U-Net [50] is a widely used semantic segmentation network, which has the advantages of the small amount of data needed, high data utilization, and

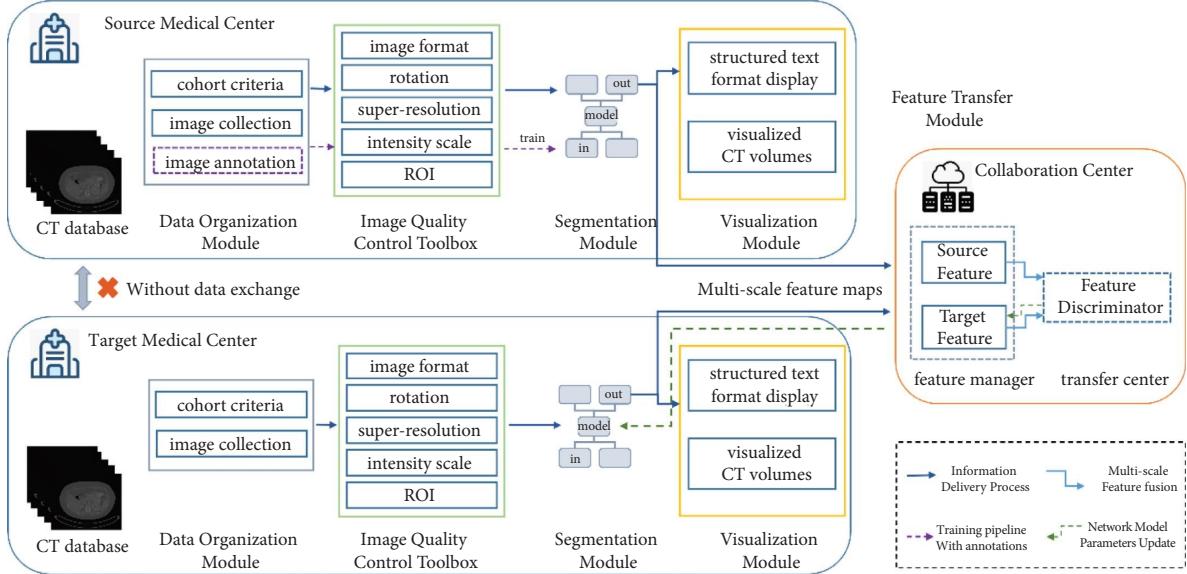


FIGURE 1: Overall framework of the data-adaptive pancreas segmentation system. The system consists of two parts: the local clients of medical centers and the collaborative center on the cloud server.

short training periods. On the basis of U-Net, the ResUNet model designed in this research introduces a deep residual structure which enables weighted interactions of image features at different scales, thus improving the segmentation performance of small targets like pancreatic tissue. The residual connection simplifies training and eliminates the degradation of the transmission of information among low and high scales, resulting in fewer parameters in networks. Besides, an attention mechanism is introduced into 3D spatial channels to enlarge the model capacity and enhance the network feature representation capability. The attention structure is inspired by the design paradigm of the squeeze and excitation (SE) network [50] and extended to 3D to handle 3D image representations. The specific structure is represented as follows: the input 3D feature map is squeezed based on pooling operations to obtain the feature vector in channels and feature activation is performed by the multilayer perception (MLP) operation of fully connected layers. After obtaining the weighting coefficients of the important channels in the feature map, the obtained coefficients are then linearly weighted into the input 3D feature map in a dot-product manner.

The feature maps at four scales (32×32 , 64×64 , 128×128 , and 256×256) of the decoder are transmitted to the collaborative center for adversarial learning. The feature maps at various scales contain not only the shallow boundary information but also the precise pancreas target information so as to ensure the effectiveness of domain adaptation. Moreover, to more strictly constrain the segmentation task on small object tasks, the model employs a linear combination of the dice loss and cross-entropy loss function as optimization criteria. The loss function is formulated as

$$\text{Loss}_{\text{all}} = \alpha \text{Loss}_{\text{CE}} + \lambda \text{Loss}_{\text{dice}}, \quad (1)$$

where α and λ are the linear coefficients that range from 0 to 1. These parameters can be customized by system users according to the needs of practical research purposes. In the experiments of this paper, $\alpha = 0.8$ and $\lambda = 0.5$.

2.4. Feature Transfer Module. The feature transfer module is mainly responsible for adversarial learning of image features among centers. Adversarial learning is performed by the discriminator in the collaborative center. The multiscale image features generated by the segmentation module of the medical center are transferred to the collaborative center as the four inputs of the discriminator. Each of the input feature maps sequentially undergoes a 3D convolutional layer and an activation function, where the step size of the 3D convolutional layer is set to 2. Therefore, the spatial scale of image features is decreased by half and thus concatenated with the next-scale feature map and fed to the next layer. Upon weighted feature fusion of multiple scales, the features are fed into the average pooling layer and the fully connected layer in turn to obtain the final domain classification results.

The workflow of feature transfer is shown in Figure 2. The labeled source-domain data are denoted as x , and the labeled dataset is defined as $S(x)$; the unlabeled target-domain data are denoted as z , and the unlabeled dataset is defined as $T(z)$. At first, the labeled dataset is trained with annotated images by the segmentation module to obtain an initial pancreas segmentation model. Then, x and z are fed into the segmentation model in pairs to generate multiscale feature maps $\mathcal{F}_S(x)$ and $\mathcal{F}_T(z)$, respectively, which are then transmitted to the collaborative center. In the collaborative center, the feature maps from the two branches are trained by the discriminator $\mathcal{H}(\cdot)$ for domain identification. Given feature maps from the source domain with label 1 and target-domain feature maps with label 0, then the optimization condition of the discriminator is to seek the weights

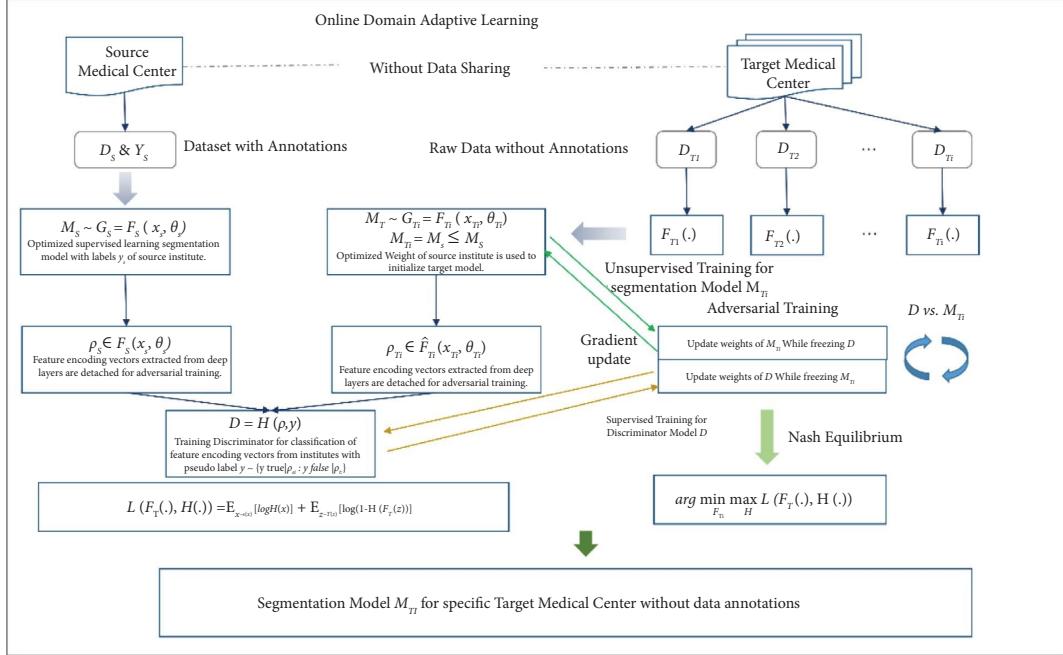


FIGURE 2: Workflow for the construction of a data-adaptive pancreas segmentation model for unlabeled data.

that maximize the difference of domain features, and the loss function is defined as

$$\begin{aligned} \mathcal{L}(\mathcal{F}_T(\cdot), \mathcal{H}(\cdot)) = & \mathbb{E}_{x \sim S(x)} [\log (\mathcal{H}(\mathcal{F}_S(x)))] \\ & + \mathbb{E}_{z \sim T(z)} [\log (1 - \mathcal{H}(\mathcal{F}_T(z)))] \end{aligned} \quad (2)$$

where $\mathbb{E}_{x \sim S(x)}$ and $\mathbb{E}_{z \sim T(z)}$ denote the mathematical expectations for the two parts.

Meanwhile, the feature maps from the target-domain images are labeled with a change to 1 and are fed into the discriminator in a single branch. The discriminator back-propagates the source-domain features with respect to the labels, thus amplifying such common image features and updating the pancreas segmentation model for the target domain. In the subsequent feature transfer process, the discriminator and the target domain segmentation model are continuously updated and frozen alternately as in the above step, thus searching for the Nash equilibrium of the two optimization functions in the adversarial process. The final optimization purpose of adversarial learning is to discover general image features between domains and utilize them to guide the segmentation of the pancreas.

2.5. Visualization Module. When the pancreas segmentation model for unannotated images is constructed by the feature transfer module, doctors could select CT images to study from the hospital local database and perform standardized preprocessing for images with custom parameters in the image quality control toolbox. Subsequently, the visualization module performs postprocessing on the pancreas mask output by the segmentation module and displays CT images and corresponding segmentation results. The presentation includes visualized images and structured text information.

Segmentation is essentially the prediction of whether each pixel in the image is a target foreground or background, so there are usually a number of isolated and noisy points in the segmentation mask. In this study, a conditional random field (CRF) model and a hole-filling algorithm are used as postprocessing operations in visualization modules to further optimize the segmentation mask to eliminate the anomalous structure and smooth the boundaries. In addition, the module offers a parallel visualization comparison between segmentation results and physician annotations in each slice-of-interest so that the physician can check annotations. The visualized image results include original CT images and segmented pancreas masks in the form of 2D slices and 3D volumes. Besides, the 3D reconstruction of surface distances between masks and annotations are also displayed. The module with visualized images also provides support for window dragging, rotating, zooming, and other operations to display images more comprehensively. Structured text information covers volume, size, and occupancy depth of pancreas tissues, surface distances between masks and annotations, etc.

2.6. Experimental Results

2.6.1. Datasets. The NIH-TCIA dataset [23] is employed as the labeled source-domain dataset in this study. The NIH-TCIA dataset is collected by the National Institutes of Health Clinical Center, which is currently the authoritative and commonly adopted public dataset for pancreas segmentation. We employed 70 cases of in-hospital CT images, collected from the First Affiliated Hospital of the Zhejiang University School of Medicine, as the unlabeled target domain dataset, noted as the Zheyi dataset. The annotations of Zheyi images were all manually outlined and cross-validated

by professional physicians. Notably, the annotations of the Zheyi dataset are not only used for the training process of the system but are also used for evaluation. The NIH-TCIA dataset contains 82 enhanced CT sequences, and the Zheyi dataset includes 70 instances. The axial resolution of CT images in the two datasets is 512×512 . Slice thickness of CT images in the NIH-TCIA dataset ranges from 0.5 mm to 1 mm, and the number of slices is in the range of 181 to 466, which are relatively high-resolution CT images. In contrast, the CT images of the Zheyi dataset range from 2.5 mm to 3 mm in layer thickness, and the slice number varies from 76 to 107.

2.6.2. Experimental Details. The system previously presented is expected to be able to cope with new sources of unlabeled CT images to construct an effective segmentation network. Taking NIH-TCIA as the labeled data center and Zheyi as the unlabeled data center, we validated the data adaptability and pancreas segmentation performance of the designed system for Zheyi images. We utilized PyTorch [51] in the Python environment to implement models and algorithms. The experiments were carried out with an NVIDIA TITAN V GPU with 12 GB memory and 2 Intel Xeon E5-2630 v4 CPUs. To guarantee the stability and reliability of the system, all trials are performed with a 5-fold cross-validation approach.

The execution time for necessary steps in the data-adaptive chain is listed in Table 1. These time data are statistically derived from the mean time of all sequences of the NIH-TCIA dataset. As can be seen in the table, the time for the data quality control module is mainly distributed over superresolution reconstruction. The duration of the complete data processing is about 16 seconds. The inference time of the segmentation model is only 2.37 seconds, and the postprocessing time consumes an average of 6.77 seconds. While 3D reconstruction takes up the majority of the time cost of the visualization module, text analysis is relatively less time consuming.

The experiments are conducted in the following steps: (1) The predefined model in the segmentation module performs supervised learning on the NIH-TCIA center to obtain the original baseline model. (2) The optimized baseline models

$$D_H(X, Y) = \{\max(\min x \in S(X) d(x, S(Y))), \max(\min y \in S(Y) d(y, S(X)))\}, \quad (4)$$

where $S(X)$ is the point set of the predicted pancreas mask and x represents the points in it. Similarly, $S(Y)$ is the point set of the ground truth mask and y represents the points in it. In the formula, $d(x, S(Y))$ indicates the distance from the point x to the surface formed by the point set $S(Y)$.

We display the average DSC, mIoU, and the Hausdorff distance on the test samples to demonstrate the average performance of the proposed pancreas segmentation system. In addition, considering the comprehensive presentation of segmentation masks, the visualization module will present the textual information and multidimensional images of the

TABLE 1: Execution time for processing steps in the system.

Processing phase	Execution time (s)
CT format conversion	0.34
Intensity rescale	0.05
Rotation in angle ranges	2.18
ROI location	0.24
Superpixel reconstruction	13.44
Segmentation interference	2.37
Postprocessing	6.77
Visualization reconstruction	5.91
Textural metric analysis	0.025

are derived as in the first step from NIH-TCIA images with various image quality control measures. (3) The original and optimized baseline models are tested on Zheyi images, which undergo the same quality control measures as NIH-TCIA images corresponding to the model, respectively. In this way, the segmentation performance without the proposed data-adaptive system can be observed. (4) Data adaptation training is carried out on NIH-TCIA and Zheyi images to get the segmentation model applicable to unannotated Zheyi data, so as to investigate the segmentation effectiveness after data adaptation. The same data adaptation trainings are also performed on images on which various processing means in the quality control toolbox have been taken.

The results are mainly evaluated by the dice similarity coefficient (DSC) and mean intersection over union (mIoU), which indicate the similarity between the pancreas mask generated by the segmentation model and ground truth. The Hausdorff distance measures the deviation between the predicted mask and ground truth mask and is calculated as the distance of points in the two masks to each other's surfaces. The DSC and mIoU are defined as follows:

$$\begin{aligned} DSC(X, Y) &= \frac{2 \times |X \cap Y|}{|X| + |Y|}, \\ mIoU(X, Y) &= \frac{|X \cap Y|}{|X \cup Y|}, \end{aligned} \quad (3)$$

where X represents the pancreas mask generated by the segmentation model and Y is the ground truth.

The Hausdorff distance is computed as

pancreas segmentation results. The textual information contains the volume difference, Hausdorff distance, center-of-mass distance, and average symmetric surface distance (ASSD) to evaluate the segmentation results from multiple perspectives.

2.6.3. Baseline Performance. Figure 3 displays CT images from NIH and Zheyi datasets processed with multiple quality control methods. As can be observed, the raw data from the two datasets exhibit differences in various aspects

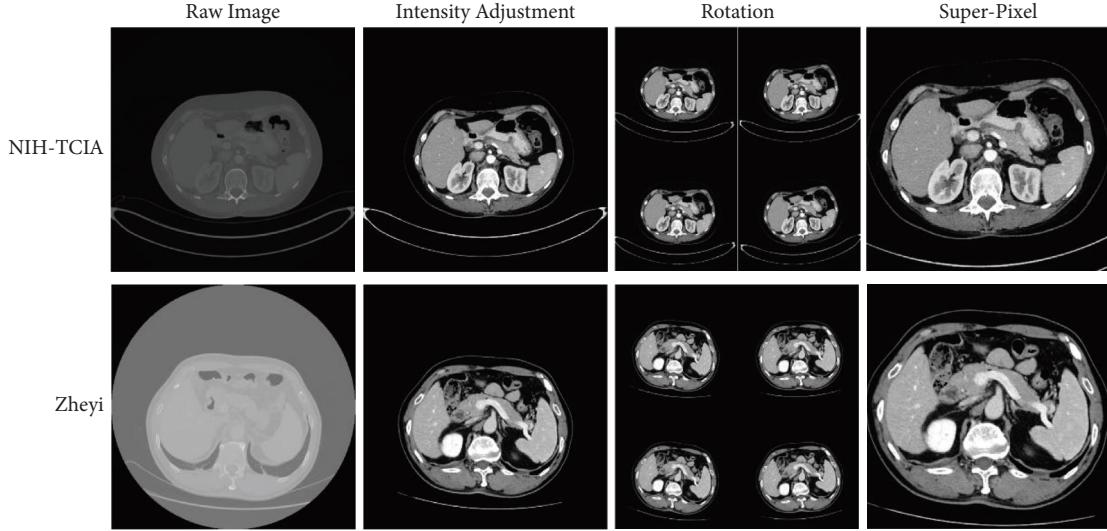


FIGURE 3: CT slice presentation with various image quality control means from two domains.

like the intensity distribution and scanning field of view (FOV) upon observation by a human. Moreover, it is also evident from the statistics shown in Table 2 that CT images of the two datasets differ significantly in terms of layer thickness. The range of intensity preserves most of information of abdominal organs and decreases the noise interference. The rotation augmentation with minor angles increases the diversity of FOV. The superresolution reconstruction algorithm frames out valid abdominal regions and eliminates redundant background information, so as to diminish heterogeneity caused by the clinical imaging process. It is observed from Figure 3 that, with the data quality control means, the visualized CT sequences narrowed the variation within and between datasets.

We first train the supervised baseline models on the NIH-TCIA dataset using raw and processed images and compare segmentation performance with existing pancreas segmentation methods. It is observed that the models trained with quality-controlled images outperform those trained on raw images, demonstrating that the image quality control toolbox designed in this paper is effective in decreasing intradomain heterogeneity. As summarized in Table 3, our model achieves a DSC of 85.45% after corresponding quality control means, which is superior to the performance of existing methods. The results show the mean value and standard deviation of the 5-fold cross-validation, which more reliably reflect the model’s performance on the whole dataset. The results indicate that the designed network in segmentation modules is of sufficient confidence to serve as the baseline models for subsequent adaptation experiments.

2.6.4. Data-Adaptive Performance. Table 4 lists the results of the unlabeled Zheyi dataset of directly tested and with data adaptation by the proposed system. When the NIH-TCIA baseline model is tested directly on the Zheyi dataset, only a DSC of 58.79% is obtained, indicating that almost half of the tissues are segmented incorrectly. Figure 4 presents the

TABLE 2: Datasets in the experiment.

Categories	Numbers	Axial resolution	Slice numbers	Slice thickness
NIH-TCIA	82	512×512	[181, 466]	[0.5, 1]
Zheyi	70	512×512	[76, 107]	[2.5, 3]

TABLE 3: Baselines segmentation performance on the NIH-TCIA dataset, where “sup-pixel” refers to superpixel reconstruction.

Model	Image options	Mean DSC (%)
Li et al. [19]	Raw images	83.06 ± 5.57
Cai et al. [20]	Raw images	82.40 ± 6.70
Yu et al. [16]	Raw images	84.50 ± 4.97
	Raw images	83.13 ± 6.93
Ours	Rotation	84.99 ± 4.86
	Sup-pixel	84.35 ± 6.10
	Rotation + sup-pixel	85.45 ± 4.51

The best performance value is presented in bold.

segmentation masks for three CT slices selected from the Zheyi dataset. As can be clearly observed, the model trained on the NIH-TCIA dataset exhibits significant degradation in pancreas segmentation on unseen images with heterogeneity. The masks in such cases do not effectively capture the pancreatic tissue information without data adaptation.

After data-adaptive training, the DSC score increases to 72.73% (a gain of 13.94%). Notably, the models trained with quality-controlled data demonstrate better performance when oriented to images from new sources. With rotation augmentation and superpixel reconstruction, the performance increases from 61.95% to 75.43% (a gain of 16.64%). To ensure the reliability of the experimental results, paired *t*-tests were performed on segmentation results of both the models with and without data adaptation. As listed in Table 4, the segmentation performance is significantly improved as observed, by the *p* value less than 0.01. Moreover,

TABLE 4: Adaptation performance of the system on the Zheyi dataset, where “sup-pixel” refers to superpixel reconstruction and “w/o” indicates without. The maximum value is presented in bold.

Labeled dataset	Unlabeled dataset	Image process options	Test w/o adaptation			Test with adaptation			<i>p</i> value of DSC
			Mean DSC (%)	mIoU (%)	D_H (mm)	Mean DSC (%)	mIoU (%)	D_H (mm)	
NIH-TCIA	Zheyi	Raw images	58.79	41.63	12.93	72.73	57.15	7.26	$3.13e-4$
		Rotation	62.73	45.70	9.37	74.91	60.00	5.57	$3.26e-7$
		Sup-pixel	61.58	44.49	11.44	73.64	58.28	6.89	$4.55e-3$
		Rotation + sup-pixel	61.95	44.88	10.16	75.43	60.55	6.34	$7.85e-6$

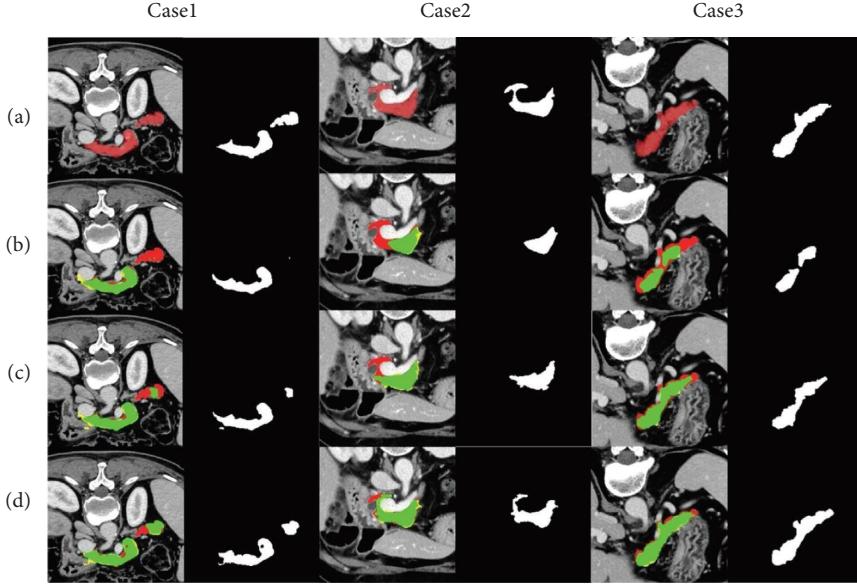


FIGURE 4: Qualitative segmentation results with different methods on several axial slices from the Zheyi dataset: manual label (red), prediction label (yellow), and overlap (green). (a) Ground truth. (b) Without adaptation. (c) With adaptation without the image quality control. (d) With adaptation with the image quality control.

after the image quality control, the difference in segmentation performance is more significant between before and after data adaptation. As can be observed in Figure 4, the pancreatic tissue can be correctly segmented out after the designed system. As expected, the model with data adaptation constructed by the system is capable of efficient segmentation for pancreas tissues and is of great significance in clinical decision-making.

2.6.5. Visualization. The visualization module provides structured textual information about the segmentation results and visualized images to assist physicians. We selected a CT sequence from the Zheyi dataset and allowed the physician to make corrections to annotations. In the designed system, the visualization results of this sample are shown in Figure 5. For the segmentation mask, statistics such as pancreas volume and size are calculated and displayed, and pancreas images in multiple dimensions are reconstructed. For physician manual corrections, the system then indicates the deviation between manual and system corrections in metrics such as the volume difference, mass distance, average symmetric surface distance (ASSD), Hausdorff

distance, and dice similarity coefficient. In addition, we calculate the 3D surface distances between two masks and reconstruct them as images.

3. Discussion

3.1. System Functionalities. In this research, a multifunctional image quality control toolbox was developed to standardize CT images from various aspects. The general rotation augmentation is employed to enhance the abundance of samples. However, the pancreas as a segmentation target is relatively small in volume in comparison with the whole abdomen. Therefore, the ordinary rotation augmentation operation is of little significance for the task. In this paper, an augmentation approach with minor rotation angles is designed to solve this problem. First, the statistical analysis of images is performed first to obtain the angle bias range of the abdomen, and then, we set a reasonable range of rotation angles with regard to data distribution characteristics. This operation reduces the discrepancy in images caused by various scanning fields of view and patient body positions with precise angle settings and enriches the samples in quantity. Furthermore, superresolution



FIGURE 5: System interface display: visualization of multidimensional segmentation results in multiple dimensions and comparison with manual annotations.

reconstruction serves two functions. On the one hand, it standardizes the images in terms of resolution, layer thickness, field of view, etc., to minimize the heterogeneity among data domains. On the other hand, more fine-grained CT images are provided in the visualization module by improving the spatial resolution of the image, which is convenient for radiologists to review and diagnose.

As can be seen in Tables 3 and 4, the images with the quality control show less heterogeneity not only in intra-domains but also in interdomains. The normalization of the images in terms of the angle, intensity scale, and region of interest lays the foundation for the subsequent transfer of the segmentation model. We performed paired *t*-tests for the ablation study in the image quality control module, and the statistical results are presented in Table 5. The results are statistically significantly improved by rotation and rotation plus superpixel reconstruction with a *p* value less than 0.01. The superpixel reconstruction only yields obviously improved results with a *p* value less than 0.05.

Moreover, in the visualization module, the reconstructed pancreas tissue displayed in Figure 5 of the visualization module demonstrates better spatial continuity, which facilitates the radiologist's diagnosis.

3.2. System Robustness. With the development of technology, the standardization and industrialization of deep learning are becoming more and more mature. As stated in the introduction, deep learning models have excellent performance in real-world applications but still face many challenges. This research improves the generalization of the segmentation model in the presence of new datasets, which addresses the potential problems of automatic pancreas

TABLE 5: Paired *t*-test results for the ablation study of raw images and images with the quality control, where “sup-pixel” refers to superpixel reconstruction.

Image options	Rotation	Sup-pixel	Rotation + sup-pixel
<i>p</i> value	1.95e - 3	3.41e - 2	7.28e - 4

segmentation systems in practical deployment and applications. However, the robustness of the deep learning model is also an issue that needs to be focused.

Robustness typically denotes the property of a system to maintain its primary performance in the presence of fluctuations in some parameters [52, 53]. Normally, robustness is used to evaluate how stable a system is against uncertain utilization environments. It is widely known that deep learning systems are driven by big data, and thus, more data lead to richer feature extraction and higher quality model construction [54–56]. Therefore, when attention is paid to the small amount of data, we investigate whether the system is affected by performance degradation.

We adopted an external dataset for the robustness experiments. The dataset was collected by Vanderbilt University and contained 30 sequences, which was marked as the BTCV dataset [57, 58]. The resolution of CT images in this dataset is 512×512 pixels, the slice number ranges in [85, 198], and the layer thickness is in the interval of [2.5, 5] mm. The experiments were performed with the NIH dataset as a labeled center and the BTCV dataset as an unlabeled center. As can be seen in Table 6, after the proposed data-adaptive pancreas segmentation system with the image quality control, the performance on the BTCV dataset reached a DSC of 74.97%, which is 14.83% improvement

TABLE 6: Robust performance of the system on the BTCV dataset with 30 images, where “sup-pixel” refers to superpixel reconstruction and “w/o” indicates without.

Labeled dataset	Unlabeled dataset	Image process options	Performance by mean DSC (%)		<i>p</i> value
			Test w/o adaptation	Test with adaptation	
NIH-TCIA	BTCV	Raw images	60.14	71.19	$7.91e - 3$
		Rotation	63.04	74.66	$1.82e - 4$
		Sup-pixel	62.35	72.63	$9.54e - 4$
		Rotation + sup-pixel	63.92	74.97	$6.34e - 6$

The best performance values are presented in bold.

compared to a DSC of 60.14%, with directly testing original images with the model trained on the NIH dataset. The paired *t*-test indicates the statistically significant improvement with a *p* value less than 0.01. It is obvious that this result is consistent with that when the Zheyi dataset served as an unlabeled dataset, which means that the system still achieves a superior data adaptation performance. This result indicates that the designed system maintains a robust performance with respect to small datasets.

3.3. System Effectiveness. In clinical diagnosis, there is a demand for modern data analysis technology to mine CT image information and assist clinicians in improving diagnosis efficiency, thus refining the medical treatment process. The data-adaptive pancreas segmentation system proposed in this study utilizes the domain-invariant features from source-domain images with annotations for transfer learning to adapt the model to CT data features from different domains, thus implementing data-adaptive pancreatic segmentation. We developed a comprehensive image quality control toolbox to rectify data quality differences among different data domains. The image quality is controlled in several dimensions, including the image intensity range, scanning field of view, CT layer thickness, and valid region of interest. In addition, the image quality control toolbox provides physicians with more discriminative CT images. In terms of system architecture, we employ an adversarial learning scheme to implement feature distribution acquisition and feature adaptation among domains. In addition, this system fully takes into consideration the difference in the contribution of semantic information at different scales for domain adaptation and adopts weighted connections to realize the stitching of multiscale information to achieve a more stable and smooth adversarial learning structure. The effectiveness of the system has been validated with public datasets and real in-hospital data, and the robustness of the system has been demonstrated on a small dataset.

In summary, the system enables the establishment of the data-adaptive segmentation model by transfer learning, both interhospital and intrahospital across time lengths. The system eliminates the need for time-consuming and tedious annotation work by radiologists, which is of significant relevance to the automation of hospital treatment processes in real-time medical scenarios. In addition, it is a meaningful research area to combine image semantic segmentation techniques with other text analysis tasks to design an automatic pancreatic disease diagnosis system applicable to richer medical scenarios. In the subsequent research, we will

combine natural language processing and image interpretability to further improve the system, optimize the pancreatic disease diagnosis process, and promote the efficiency of physicians.

4. Conclusions

In this paper, we designed an end-to-end data-adaptive pancreas segmentation system with an image quality control toolbox. The system aims to address the problem of poor generalization capability exhibited by existing pancreas segmentation networks when oriented to data from different medical centers. For the visual task of label-free semantic segmentation, this research utilizes an adversarial learning method to obtain domain-invariant supervised information and construct the data-adaptive pancreas segmentation model. In addition, a functional image quality control toolbox was designed to provide multiple image preprocessing methods. The system works in an end-to-end manner and is easy to operate by physicians. The experimental results of public datasets and in-hospital datasets demonstrated that the end-to-end data-adaptive pancreas segmentation tool proposed in this paper can effectively assist in pancreas segmentation, and the generalization of segmentation networks was enhanced when facing images from different sources. This system is of considerable relevance in medical diagnosis and treatment and greatly promotes the development of precision and automated medical processes.

Data Availability

The CT sequences in the Zheyi dataset used to support the findings of this study are restricted by the First Affiliated Hospital of the Zhejiang University School of Medicine to protect the patient privacy. The data are not publicly available due to privacy restrictions.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Yan Zhu and Peijun Hu equally contributed to this work.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Nos. 12101571, 82172069, and

81702332), the Major Scientific Project of Zhejiang Lab (No. 2020ND8AD01), the Zhejiang Provincial Natural Science Foundation of China (No. LQ20H180001), and the Zhejiang Provincial Key Research and Development Program (No. 2020C03117).

References

- [1] X. Y. Jia, P. Ku, K. Wu et al., "Pancreatic cancer mortality in China: characteristics and prediction," *Pancreas*, vol. 47, no. 2, pp. 233–237, 2018.
- [2] L. C. Chu, S. Park, S. Kawamoto et al., "Utility of CT radiomics features in differentiation of pancreatic ductal adenocarcinoma from normal pancreatic tissue," *American Journal of Roentgenology*, vol. 213, no. 2, pp. 349–357, 2019.
- [3] T. Tarver, "Cancer facts & figures 2012. American cancer society (ACS)," *Journal of Consumer Health on the Internet*, vol. 16, no. 3, pp. 366–367, 2012.
- [4] P. E. Oberstein and K. P. J. T. A. i. G. Olive, "Pancreatic cancer: why is it so hard to treat?" *Therapeutic Advances in Gastroenterology*, vol. 6, no. 4, pp. 321–337, 2013.
- [5] C. Yip, C. Dinkel, A. Mahajan, M. Siddique, G. J. R. Cook, and V. Goh, "Imaging body composition in cancer patients: visceral obesity, sarcopenia and sarcopenic obesity may impact on clinical outcome," *Insights into Imaging*, vol. 6, no. 4, pp. 489–497, 2015.
- [6] M. Ducreux, A. S. Cuhna, C. Caramella et al., "Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 26, no. 5, pp. v56–v68, Sep 2015.
- [7] Y. Zhou, L. Xie, E. K. Fishman, and A. L. Yuille, "Deep supervision for pancreatic cyst segmentation in abdominal CT scans," in *Proceedings of the International Conference on Medical Image Computing & Computer-assisted Intervention*, Beijing China, October 2017.
- [8] M. Palmowski, N. Hacke, S. Satzl et al., "Metastasis to the pancreas: characterization by morphology and contrast enhancement features on CT and MRI," *Pancreatology*, vol. 8, no. 2, pp. 199–203, 2008.
- [9] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, and R. M. J. I. T. o. I. P. Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 386–399, 2017.
- [10] M. S. Petrov, "Harnessing analytic morphomics for early detection of pancreatic cancer," *Pancreas*, vol. 47, no. 9, pp. 1051–1054, Oct 2018.
- [11] R. G. Singh, N. N. Nguyen, S. V. DeSouza, S. A. Pendharkar, and M. S. Petrov, "Comprehensive analysis of body composition and insulin traits associated with intra-pancreatic fat deposition in healthy individuals and people with new-onset prediabetes/diabetes after acute pancreatitis," *Diabetes, Obesity and Metabolism*, vol. 21, no. 2, pp. 417–423, 2019.
- [12] M. Oda, N. Shimizu, K. Karasawa, Y. Nimura, and K. Mori, "Regression forest-based atlas localization and direction specific atlas generation for pancreas segmentation," *Advanced Data Mining and Applications*, pp. 556–563, 2020.
- [13] Z. Li, J. Zhang, T. Tan et al., "Deep learning methods for lung cancer segmentation in whole-slide histopathology images -- the ACDC@LungHP challenge 2019," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 429–440, 2021.
- [14] S. Geert, *A Survey on Deep Learning in Medical Image Analysis*, Medical Image Analysis, China, 2017.
- [15] Z. Zhu, Y. Xia, W. Shen, E. K. Fishman, and A. L. Yuille, "A 3D coarse-to-fine framework for volumetric medical image segmentation," in *Proceedings of the 2018 International Conference on 3D Vision (3DV)*, Beijing China, April 2018.
- [16] Q. Yu, L. Xie, W. Yan, Y. Zhou, and A. L. Yuille, "Recurrent saliency transformation network: incorporating multi-stage visual cues for small organ segmentation," in *Proceedings of the presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach California, June 2018.
- [17] J. Fu, J. Liu, H. Tian et al., "Dual attention network for scene segmentation," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach California, June 2020.
- [18] O. Oktay and W. Chen, *Attention U-Net: Learning where to Look for the Pancreas*, Medical Image Analysis Special Issue, China, 2018.
- [19] M. Li, F. Lian, and S. Guo, "Pancreas segmentation based on an adversarial model under two-tier constraints," *Physics in Medicine and Biology*, vol. 65, no. 22, Article ID 225021, 2020.
- [20] J. Cai, L. Lu, Y. Xie, F. Xing, and L. Yang, "Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function," 2017, <https://arxiv.org/abs/1707.04912>.
- [21] C. Fang, G. Li, C. Pan, Y. Li, and Y. Yu, "Globally guided progressive fusion network for 3D pancreas segmentation," *Advanced Data Mining and Applications*, pp. 210–218, 2019.
- [22] H. R. Roth, L. Lu, N. Lay et al., "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation," *Medical Image Analysis*, vol. 45, pp. 94–107, 2018.
- [23] H. R. Roth, L. Lu, A. Farag et al., "DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation," *Lecture Notes in Computer Science*, vol. 9349, pp. 556–564, 2015.
- [24] B. Giddwani, S. Pandey, H. Tekchandani, and S. Verma, "CSTA-2P1D UNet: consecutive spatio-temporal attention for multi-scale 3D pancreas segmentation," in *Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, July 2020.
- [25] W. Wang, Q. Song, R. Feng, T. Chen, and J. Wu, "A fully 3D cascaded framework for pancreas segmentation," in *Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Iowa City, IA, USA, April 2020.
- [26] P. Hu, X. Li, Y. Tian et al., "Automatic pancreas segmentation in CT images with distance-based saliency-aware DenseASPP network," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1601–1611, 2021.
- [27] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.
- [28] Z. C. Xu, Q. Zhang, F. Xu, D. Li, and R. Zhang, "Anthropometric measurements in 126 microtia reconstructions," *Facial Plastic Surgery*, vol. 29, no. 04, pp. 321–326, Aug 2013.
- [29] U. Koc, O. J. J. o. M. I. Taydas, and R. Sciences, "Investigation of the relationship between fatty pancreas and cholecystectomy using noncontrast computed tomography," *Journal of Medical Imaging and Radiation Sciences*, vol. 50, no. 2, pp. 220–226, 2019.
- [30] Y. H. Roh, B. K. Kang, S. Y. Song, C. M. Lee, Y. K. Jung, and M. Kim, "Preoperative CT anthropometric measurements and pancreatic pathology increase risk for postoperative

- pancreatic fistula in patients following pancreaticoduodenectomy,” vol. 15, no. 12, Article ID e0243515, 2020.
- [31] S. European, “Imaging what the radiologist should know about artificial intelligence - an ESR white paper,” *Insights into Imaging*, vol. 1, 2019.
- [32] G. Campanella, M. G. Hanna, L. Geneslaw et al., “Clinical-grade computational pathology using weakly supervised deep learning on whole slide images,” *Nature Medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.
- [33] Q. Liu, Q. Dou, L. Yu, and P. A. J. I. T. o. M. I. Heng, “MS-net: multi-site network for improving prostate segmentation with heterogeneous MRI data,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2713–2724, 2020.
- [34] E. Gibson, Y. Hu, N. Ghavami, H. U. Ahmed, and D. C. Barratt, “Inter-site variability in prostate segmentation accuracy using deep learning,” in *Proceedings of the Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018*, Granada, Spain, September 2018.
- [35] K. V. Sarma, S. Harmon, T. Sanford et al., “Federated learning improves site performance in multicenter deep learning without data sharing,” *Journal of the American Medical Informatics Association*, vol. 28, no. 6, pp. 1259–1264, 2021.
- [36] Z. Wang, Q. Liu, Q. J. I. J. o. B. Dou, and H. Informatics, “Contrastive cross-site learning with redesigned net for COVID-19 CT classification,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2806–2813, 2020.
- [37] B. A. Goldstein, A. M. Navar, M. J. Pencina, and J. P. A. Ioannidis, “Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review,” *Journal of the American Medical Informatics Association*, vol. 24, no. 1, pp. 198–208, 2017.
- [38] J. A. Onofrey, D. I. Casetti-Dinescu, A. D. Lauritzen, S. Sarkar, and X. Papademetris, “Generalizable multi-site training and testing of deep neural networks using image normalization,” in *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, NY China, June 2019.
- [39] F. J. Brooks and P. W. J. B. M. I. Grigsby, “Quantification of heterogeneity observed in medical images,” *BMC Medical Imaging*, vol. 13, no. 1, pp. 7–12, 2013.
- [40] L. Gao, T. Jiao, Q. Feng, and W. J. O. I. Wang, “Application of artificial intelligence in diagnosis of osteoporosis using medical images: a systematic review and meta-analysis,” *Osteoporosis International*, vol. 32, no. 7, pp. 1279–1286, 2021.
- [41] S. Kaur, M. J. Baine, M. Jain, A. R. Sasson, and S. K. J. B. i. M. Batra, “Early diagnosis of pancreatic cancer: challenges and new developments,” *Biomarkers in Medicine*, vol. 6, no. 5, pp. 597–612, 2012.
- [42] M. Lerousseau and L. Xiao, *Weakly Supervised Pan-Cancer Segmentation Tool*, Springer, Heidelberg, Germany, 2021.
- [43] E. Ferrante, O. Oktay, B. Glocker, and D. H. Milone, “On the adaptability of unsupervised CNN-based deformable image registration to unseen image domains,” *Advanced Data Mining and Applications*, pp. 294–302, Beijing, 2018.
- [44] R. J. Xu, Z. L. Chen, W. M. Zuo, J. J. Yan, and L. Lin, “Deep cocktail network: multi-source unsupervised domain adaptation with category shift,” in *Proceedings of the presented at the 2018 Ieee/Cvpr Conference on Computer Vision and Pattern Recognition (Cvpr)*, New York, October 2018.
- [45] M. Dunnhofer, N. Martinel, and C. Micheloni, “Weakly-supervised domain adaptation of deep regression trackers via reinforced knowledge distillation,” *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5016–5023, 2021.
- [46] L. Liu, L. Yang, and B. J. K.-B. S. Zhu, “Sparse feature space representation: a unified framework for semi-supervised and domain adaptation learning,” *Knowledge-Based Systems*, vol. 156, pp. 43–61, 2018.
- [47] M. Wang and W. Deng, “Deep visual domain adaptation: a survey,” *Neurocomputing*, vol. 312, pp. 135–153, 2018.
- [48] Y. F. Zhang, Y. Wei, Q. Wu et al., “Collaborative unsupervised domain adaptation for medical image diagnosis,” *IEEE Transactions on Image Processing*, vol. 29, pp. 7834–7844, 2020.
- [49] X. Luo, J. Chen, T. Song, G. Wang, and S. Zhang, “Semi-supervised medical image segmentation through dual-task consistency,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 10, pp. 8801–8809, 2021.
- [50] O. Ronneberger, P. Fischer, and T. J. S. I. P. Brox, *U-net: Convolutional Networks for Biomedical Image Segmentation*, Springer International Publishing, Heidelberg, Germany, 2015.
- [51] A. Paszke, S. Gross, F. Massa, A. Lerer, and S. Chintala, *PyTorch: An Imperative Style*, High-Performance Deep Learning Library, China, 2019.
- [52] W. Zhou, J. S. Berrio, S. Worrall, and E. J. I. T. o. I. T. S. Nebot, “Automated evaluation of semantic segmentation robustness for autonomous driving,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 1951–1963, 2020.
- [53] H. D. Tran, N. Pal, P. Musau, D. M. Lopez, and T. T. Johnson, “Robustness verification of semantic segmentation neural networks using relaxed reachability,” *Advanced Data Mining and Applications*, pp. 263–286, NY China, 2021.
- [54] Q. Zhang, L. T. Yang, Z. Chen, and P. J. I. F. Li, “A survey on deep learning for big data,” *Information Fusion*, vol. 42, pp. 146–157, 2018.
- [55] X. W. Chen and X. J. I. A. Lin, “Big data deep learning: challenges and perspectives,” *IEEE Access*, vol. 2, no. 2, pp. 514–525, 2014.
- [56] A. M. J. I. J. o. E. R. Zain and P. Health, “Detection of COVID-19 in chest X-ray images: a big data enabled deep learning approach,” *International Journal of Environmental Research and Public Health*, vol. 18, 2021.
- [57] X. Z. Landman Ba, J. E. Iglesias, M. Styner, T. R. Langerak, and A. Klein, “MICCAI Multi-Atlas Labeling beyond the Cranial Vault - Workshop and challenge,” 2015, <https://arxiv.org/abs/1711.06853>.
- [58] Z. Xu, C. P. Lee, M. P. Heinrich et al., “Evaluation of six registration methods for the human abdomen on clinically acquired CT,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1563–1572, 2016.