

A Comparative Study of 2D and 3D U-Net-Based Architectures for Pancreas and Tumor Segmentation in CT Imaging

İbrahim Şen¹, Azer Çelikten^{2*}, Ece Bingöl³, Muhammed Kerem Demir⁴, Andaç Akpulat⁵, Semih Demirel⁶, Abdulkadir Budak⁷, Hakan Karataş⁸

¹Manisa Celal Bayar University, Department of Software Engineering, 45140 Manisa, Türkiye.
^{2,3,4,5,6,7,8} Akgun R&D Center, Artificial Intelligence and Image Processing Group, 06790 Ankara, Türkiye.

* azer.celikten@akgun.com.tr

* Orcid No: 0000-0002-6804-737X

Received: June 30, 2025

Accepted: March 9, 2026

DOI: [10.18466/cbayarfb.1808561](https://doi.org/10.18466/cbayarfb.1808561)

Abstract

Pancreatic cancer ranks among the deadliest forms of cancer, largely due to its asymptomatic progression, late-stage diagnosis, and the pancreas's deep-seated and anatomically complex location within the abdominal cavity. These factors collectively prevent early detection, which is critical for improving patient outcomes. Deep learning-based segmentation models, particularly those based on the U-Net architecture, have demonstrated strong potential for the automatic identification of pancreas and pancreatic tumors in computed tomography (CT) images. This study presents a comparative analysis of 2D and 3D U-Net architectures for the segmentation of the pancreas and its tumors. The results show that the 3D Res-U-Net achieved the best performance, with a Dice score of 0.81 for pancreas segmentation and 0.51 for tumor segmentation, outperforming both the 2D U-Net and the 3D U-Net. Additionally, inference time analysis for real-time clinical settings showed that the 2D U-Net enables the fastest predictions, whereas the 3D U-Net and 3D Res-U-Net require significantly higher computation, underscoring key trade-offs for clinical deployment. These findings offer valuable insights into model selection and optimization strategies for real-world clinical applications. The findings underscore the feasibility of developing efficient deep learning-based systems to support early diagnosis and personalized treatment planning for pancreatic cancer.

Keywords: 3D U-Net, 3D-Res-U-Net, CT imaging, deep learning, medical image segmentation, pancreatic cancer, U-Net

1. Introduction

The pancreas is a vital organ located in the abdominal region, playing critical roles in both endocrine and exocrine functions within the human body. Pancreatic cancer ranks among the deadliest types of cancer worldwide, with a notably poor prognosis, primarily due to the fact that most cases are diagnosed at advanced stages. According to Cancer Research UK, the five-year survival rate for pancreatic cancer is approximately 10%, largely due to delayed diagnosis and the pancreas's complex anatomical positioning [1]. While pancreatic cancer can be detected using medical imaging techniques such as CT and PET (Positron Emission Tomography), accurate identification remains challenging due to considerable variability in the shape, size, and anatomical position of the pancreas among individuals. The

pancreas's small size and low contrast with surrounding organs make its boundaries difficult to identify in CT, hindering early cancer detection. Therefore, early diagnosis is crucial for effective treatment planning and the implementation of personalized medicine approaches.

To address these challenges, deep learning-based models have been increasingly utilized for the accurate segmentation of the pancreas and pancreatic tumors. Among these models, U-Net and its variants have demonstrated outstanding performance across a wide range of medical image segmentation tasks. For instance, U-Net-based models have been successfully applied to brain tumor segmentation [2], liver and kidney delineation [3], aorta segmentation [4], polyp detection in colonoscopy images [5], and segmentation of pancreas in CT scans [6]. These applications highlight the

flexibility, versatility, and robustness of the U-Net architecture in handling diverse anatomical structures and imaging modalities, as well as in the segmentation of the pancreas and its associated tumors in CT scans. Furthermore, U-Net has proven to be effective even when trained on limited datasets, achieving high levels of accuracy and robustness. Over time, various extensions of the U-Net architecture have been developed. While 2D U-Net operates on two-dimensional images, the 3D U-Net model has demonstrated superior performance on volumetric data, such as CT scans. In addition, variants such as Attention U-Net, which incorporate attention mechanisms, enable more accurate localization of target regions. Other extensions like Residual U-Net enhance the architecture by introducing residual connections to deepen the learning process and reduce information loss.

Accurate pancreas segmentation remains a challenging task due to the organ's small size and high anatomical variability. In recent years, numerous deep learning-based approaches have been developed to address this problem using contrast-enhanced abdominal CT images.

Meiyu Li et al. proposed ADAU-Net, which integrates attention mechanisms and pairwise adversarial learning into a U-Net-based architecture. Their model achieved a Dice score of 83.76% on the widely recognized NIH dataset, which includes 82 CT volumes, demonstrating improved segmentation performance through adversarial training [7].

Tirumalasetti Devendhar et al. employed a standard U-Net to address pancreas segmentation on the TCIA dataset, investigating the impact of training conditions on performance. Their model achieved a Dice score of 81.38%, IoU of 79.62%, and boundary F1 score of 80.36%, showing that even the vanilla U-Net is effective for pancreas segmentation tasks [8].

Himali Ghorpade et al. introduced a hybrid two-stage U-Net enhanced with the GWBCO optimization algorithm, combining Grey Wolf and Border Collie Optimization. The proposed method achieved 93.33% Dice for pancreas and 91.46% Dice for tumor segmentation, making it a strong candidate for early-stage clinical applications [9].

Tualombo et al. improved segmentation performance by combining U-Net with Super-Resolution GANs (SRGAN). Their model reported a Dice score of 92.11% using U-Net alone and 53.54% with SRGAN+U-Net, along with recall of 63.51% and precision of 64.22%, demonstrating promise for real-time decision support systems due to its low training time [10].

Another comparative study evaluated several state-of-the-art models—U-Net, LinkNet, SegNet, SQ-Net, DABNet, EDANet, and ESNet—for both pancreas and

pancreatic tumor segmentation. SQ-Net outperformed all others, achieving a Dice score of 0.917, IoU of 0.847, sensitivity of 0.920, specificity of 1.000, precision of 0.914, and accuracy of 0.999, leading to the development of an effective AI-based decision support system [11].

In another study, an extensive comparison of U-Net variants—including Attention U-Net, Residual U-Net, Attention Residual U-Net, and Residual U-Net++ was performed using the NIH-CT82 dataset. The Residual U-Net model achieved the best performance with a Dice score of 0.903, IoU of 0.823, sensitivity of 0.898, specificity of 1.000, precision of 0.908, and accuracy of 0.999, supporting the creation of an AI-assisted segmentation system [12].

Chen et al. proposed CTUNet, a novel architecture that combines Transformer modules with a 3D U-Net backbone. The model introduces a Pancreas Attention module tailored to the organ's anatomical complexity and incorporates Transformer-based skip connections for improved global context modeling. A Feature Link module was also integrated to strengthen multi-level feature fusion. Trained and validated with 4-fold cross-validation on the NIH dataset, CTUNet achieved superior performance compared to previous methods across multiple metrics including Dice Similarity Coefficient (0.868 ± 0.041), Jaccard Index (0.769 ± 0.061), Precision (0.862 ± 0.065), and Recall (0.880 ± 0.060), thus advancing the state of the art in pancreas segmentation [13].

In their study, Azim et al. conducted a comparative analysis of U-Net and V-Net architectures for pancreas segmentation using the NIH-CT-82 dataset, which contains contrast-enhanced abdominal CT scans. The models were evaluated based on validation sensitivity and validation loss. Results from the ablation experiments indicated that U-Net outperformed V-Net, achieving a validation sensitivity of 0.9978 and validation loss of 0.041, compared to 0.9975 sensitivity and 0.046 loss obtained by V-Net. These findings highlight the effectiveness of U-Net in handling the complex anatomical structure of the pancreas and reinforce its suitability for medical image segmentation tasks [14].

In another study, Ghorpade et al. explored automated methods for the segmentation of the pancreas and pancreatic tumors, leveraging the widely adopted NIH-TCIA dataset, which consists of 82 contrast-enhanced CT scans. Utilizing deep learning architectures such as U-Net and V-Net, the study reported Dice values exceeding 80%, indicating reliable segmentation performance. Moreover, the authors investigated the use of GAN-based approaches to synthesize contrast-enhanced CT images from non-contrast CT scans, which helped reduce the need for manual annotations and further improved segmentation accuracy. These advances offer meaningful

contributions toward addressing the inherent challenges of pancreas segmentation and enhancing clinical decision support systems [15].

Unlike previous studies that often prioritize model complexity while overlooking data preparation, this study emphasizes the importance of dataset-agnostic preprocessing pipelines and utilizes five publicly available CT datasets with heterogeneous imaging characteristics to ensure robust model training and evaluation. Beyond segmentation accuracy, inference time analyses were systematically conducted across CPU and GPU configurations to assess real-time clinical feasibility, accompanied by clinically oriented discussions that link technical outcomes to practical applications. Collectively, these contributions underscore the value of a comprehensive, system-level benchmark that balances model performance, computational efficiency, and clinical usability. The contribution of this study can be summarized as follow:

- Development of efficient deep learning-based segmentation models to support the early detection of pancreatic cancer.
- A comprehensive evaluation of 2D and 3D U-Net architectures for pancreas and pancreatic tumor segmentation on CT images from 5 different datasets.
- Investigation of the impact of advanced preprocessing strategies on the performance and robustness of deep learning models in medical image segmentation.
- Comparative analysis of training and inference times across models to inform real-world clinical deployment.

2. Materials and Methods

2.1 Datasets

In this study, five publicly available abdominal CT datasets were utilized to develop and evaluate the performance of deep learning models for pancreas and pancreatic tumor segmentation. These datasets are characterized by varying modalities, resolutions, patient counts, and labeling standards, providing a comprehensive experimental setting for model training, validation, and generalization assessment.

MSD Pancreas Dataset: The Task07 dataset from the Medical Segmentation Decathlon (MSD) consists of contrast-enhanced abdominal CT scans from 281 patients [16]. It is specifically designed for pancreas and pancreatic tumor segmentation and contains high-resolution 3D volumes acquired during the portal-venous phase. The pancreas's proximity to other abdominal organs and its low contrast with surrounding tissues pose significant challenges for accurate segmentation. Additionally, tumors exhibit considerable variation in size, shape, and intensity. These factors make the Task07 dataset a widely recognized and challenging benchmark

for evaluating the performance of deep learning models in pancreas segmentation.

AMOS Dataset: AMOS (Abdominal Multi-Organ Segmentation) dataset includes CT and MRI scans from 198 patients, annotated for 15 abdominal organs such as the liver, spleen, pancreas, kidneys, stomach, and intestines [17]. This multi-modal and multi-center dataset is particularly valuable for testing model generalizability across imaging modalities. Its heterogeneous composition—derived from different scanners and institutions—makes it a robust benchmark for evaluating both single-organ and multi-organ segmentation algorithms.

The TotalSegmentator-Small Dataset: This dataset is a lightweight version of the original TotalSegmentator dataset [18] comprising abdominal CT scans from 102 patients. It provides highly detailed segmentation masks for 104 anatomical structures, including organs, bones, vessels, and soft tissues. Despite its relatively smaller size, it offers high-quality annotations and anatomical diversity, making it ideal for rapid prototyping and evaluation in computationally constrained environments.

5510650 Pancreas Dataset: This dataset consists of abdominal CT scans from 91 patients, with expert-labeled segmentations for the pancreas, liver, kidneys, and pancreatic adenocarcinoma [19]. The data were collected from various publicly available sources for AI-based pancreas research. Each scan is provided in NIfTI format, along with rich metadata including tumor location, T stage, largest tumor diameter, and presence or absence of pancreatic atrophy or ductal dilatation. Despite the modest patient count, its pathological diversity and expert annotations make it a valuable dataset for developing tumor-focused segmentation algorithms.

FLARE 2021: This dataset contains abdominal CT scans from 361 patients and was designed for low-resource and semi-supervised learning scenarios [20]. It includes both labeled and unlabeled data, supporting the development of advanced learning strategies. The labeled portion consists of annotations for four organs: liver (label=1), kidney (label=2), spleen (label=3), and pancreas (label=4). The dataset combines data adapted from MSD (Liver, Spleen, Pancreas), NIH Pancreas, and KiTS, as well as 50 additional CT cases from Nanjing University. Its large size, institutional diversity, and semi-supervised potential make it highly suitable for evaluating models intended for real-world clinical deployment.

A combination of five publicly available datasets was utilized for training and evaluating the segmentation models. The details of the datasets are presented in Table 1. For pancreas segmentation tasks, the AMOS, TotalSegmentator, and FLARE 2021 datasets were employed, while the MSD and the 5510650 Pancreas dataset were used for pancreatic tumor segmentation.

Each dataset was split on a per-patient basis into 80% for training, 10% for validation, and 10% for testing. In total, 661 patients were included for pancreas segmentation and 372 patients for pancreatic tumor segmentation. The number of axial CT slices per patient varied between 30 and 400.

Table 1. The details of pancreas and pancreatic tumor datasets.

Dataset Name	Labeled Region	#Patients
MSD Pancreas Cancer	Pancreatic tumor	281
5510650 Pancreas	Pancreatic tumor	91
AMOS	Pancreas	198
Total Segmentor	Pancreas	102
FLARE 2021	Pancreas	361

2.2. Pre-Processing

In this study, distinct preprocessing pipelines were implemented for 2D and 3D model due to model architecture requirements to ensure optimal data standardization and compatibility.

2.2.1. 2D U-Net Preprocessing

In order to achieve high accuracy in image-based analysis of pancreatic cancer, it is essential to convert raw CT data into a model-compatible format. Medical images may vary in size and orientation due to differences in scanners, clinical centers, or acquisition protocols; therefore, eliminating such inconsistencies is critical for both model performance and anatomical fidelity. Accordingly, the 3D CT images in all datasets were subjected to a comprehensive preprocessing pipeline. The workflow includes 3D volume slicing, HU normalization, image resizing, mask orientation correction, mask resizing, and data type conversion for the masks.

Firstly, the original 3D CT scans in NIfTI format are decomposed into individual 2D axial slices to facilitate slice-wise analysis and model training. This process is illustrated in Figure 1, which provides a visual comparison between the original axial slice and its standardized version following resizing and orientation adjustment.

Next, intensity values are clipped to a predefined HU range (-150 to 250) to suppress irrelevant structures and enhance contrast within the anatomical region of interest. These values are then normalized to the range [0, 1]. Figure 2 demonstrates the enhancement achieved through HU normalization by comparing the same axial slice before and after the transformation.

Following this, all axial slices are resized to a fixed target resolution to maintain uniform input dimensions across the dataset. This resizing step contributes to the standardization process depicted in Figure 1.

To ensure compatibility with deep learning architectures and to maintain uniformity in model inputs, the masks are rescaled to a fixed resolution. Figure 3 illustrates the transformation of a raw mask slice into its rescaled version during the preprocessing pipeline.

As a result, all CT images and corresponding masks were transformed into a standardized and machine-readable format, optimized for subsequent deep learning-based segmentation tasks using 2D U-Net architectures.

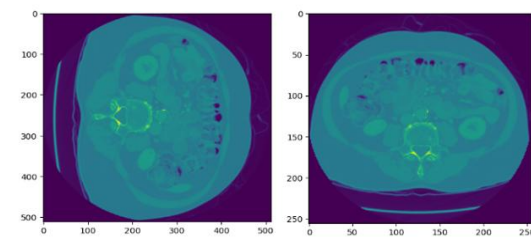


Figure 1. Comparison of original axial image and standardized (resized and oriented) version

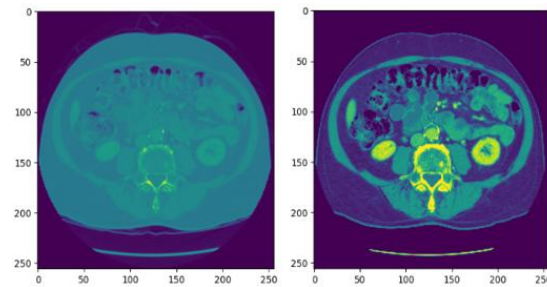


Figure 2. Comparison of the axial slice image obtained before and after HU normalization

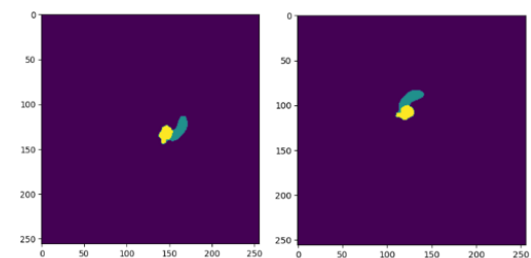


Figure 3. Comparison of raw mask slice and rescaled version

2.2.2. 3D U-Net Preprocessing

In medical image analysis with volumetric CT data, raw images often require extensive preprocessing before they can be used in deep learning models. In this study, a systematic preprocessing pipeline was developed using the MONAI framework for 3D pancreas segmentation based on the Task07_Pancreas dataset. MONAI provides a modular structure that facilitates a wide range of medical image transformations.

First, the volumetric CT images and their corresponding segmentation masks are loaded into memory. All volumes are resampled to a uniform voxel spacing. After this, intensity values of the image volumes are clipped within a predefined HU range (-150 to 250) and scaled to the normalized range [0, 1]. This normalization enhances the visibility of relevant anatomical structures and ensures contrast uniformity across the dataset.

The next step involves cropping the foreground region to eliminate empty or irrelevant areas and reduce computational load. Each cropped volume is then resized to a fixed spatial dimension, ensuring uniformity for batch-based model training.

To further enhance the visibility of low-contrast regions such as the pancreas borders, a Contrast Limited Adaptive Histogram Equalization method is applied to the image volumes. This step improves local contrast and detail representation. As a result, the raw 3D CT data is transformed into a consistent and model-compatible format that preserves anatomical accuracy, improves contrast, and standardizes input dimensions for reliable and reproducible segmentation tasks.

2.3. Proposed Methods

2.3.1. 2D U-Net

2D U-Net is a deep learning model with a U-shaped architecture that performs pixel-level segmentation by processing each 2D CT slice in the axial plane separately [21]. The convolution and pooling layers in the lower part of the model extract meaningful structural features from the image, while the decoder layers and jump links in the upper part transform this information into detailed segmentation maps. In this study, volumetric 3D CT images are segmented into 2D slices in the axial plane, normalized to 256x256 pixels and then presented as input to a 2D U-Net model. In the preprocessing stage of the images, HU normalization, orientation correction and resizing were applied to ensure that the model works with standard and compatible data. In addition, in some patients, completely black slices with no anatomical content were filtered and removed from the dataset to prevent negative effects on model training. While 2D U-Net offers a suitable option for rapid prototyping and limited hardware infrastructures due to its computationally less resource requirements, it does not take into account the spatial connectivity between slices, which is an important limitation. Therefore, while 2D U-Net provides a lightweight and accessible framework, it is inherently limited in modeling volumetric continuity, which is critical in accurately segmenting organs with complex 3D morphology. The architecture of 2D U-Net is presented in Figure 4.

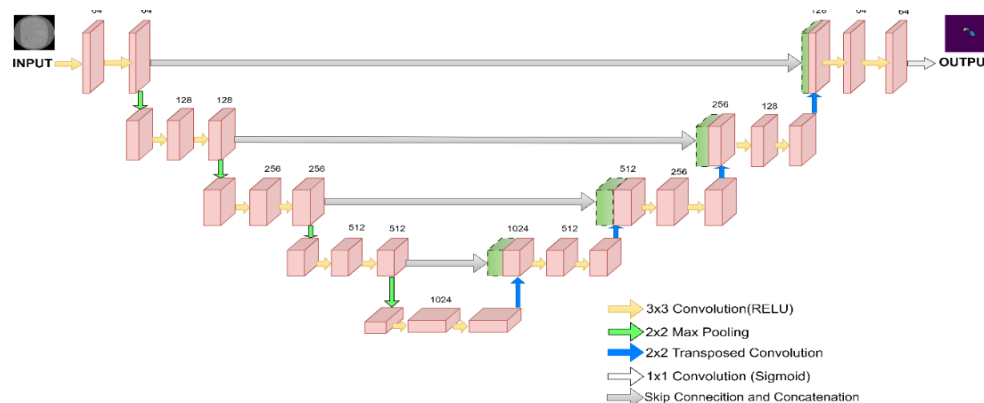


Figure 4. 2D U-Net Architecture

2.3.2. 3D U-Net

3D U-Net [22] is an extended version of the traditional 2D U-Net architecture to work with volumetric data. It takes as input an entire 3D CT volume at a time - in this case 128x128x32 and 128x128x64 data blocks. It uses three-dimensional convolution and pooling layers to simultaneously analyze spatial correlations along the three spatial axes (anterior-posterior, left-right, and

superior-inferior) for a more holistic and contextual segmentation performance. The inputs are pre-processed before the model through various pre-processing steps such as orientation standardization and density normalization based on HU range. 3D U-Net has the potential to provide high accuracy, especially in accurately parsing complex three-dimensional

anatomical structures such as the liver. However, it requires much more memory and computational power, making the training process only possible with powerful hardware. It offers clear advantages over 2D models, especially when axial context is important. Compared to

its 2D counterpart, 3D U-Net excels in scenarios where volumetric coherence plays a central role in segmentation performance. The architecture of 3D U-Net is presented in Figure 5.

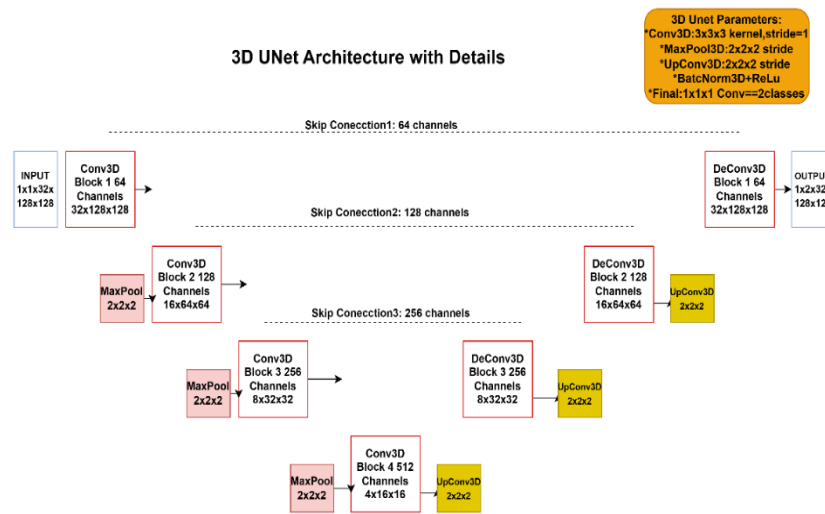


Figure 5. 3D U-Net Architecture

2.3.3. 3D Res-U-Net

3D ResU-Net extends the classical U-Net architecture by integrating residual learning mechanisms, which facilitate deeper network training and improved gradient propagation. Based on the encoder-decoder structure of U-Net, residual blocks are integrated to transmit the features learned at each stage to deeper layers. These residual connections allow the model to learn more stable and faster, while reducing the gradient fading problem that can occur especially in deep networks. Since 3D ResU-Net works with 3D convolution layers, it enables detailed analysis of volumetric medical data, especially CT and MRI which has been used for medical image

segmentation tasks [23,24,25]. This structure effectively combines both local and global contextual information, enabling high accuracy in complex tasks such as organ or lesion segmentation. This model takes as input an entire 3D CT volume at a time - in this case, data blocks prepared in 128x128x32 and 128x128x64 dimensions. Thus, 3D ResU-Net represents a powerful alternative in volumetric medical segmentation, particularly when structural complexity and contextual accuracy are critical. The architecture of 3D-ResUNet is presented in Figure 6.

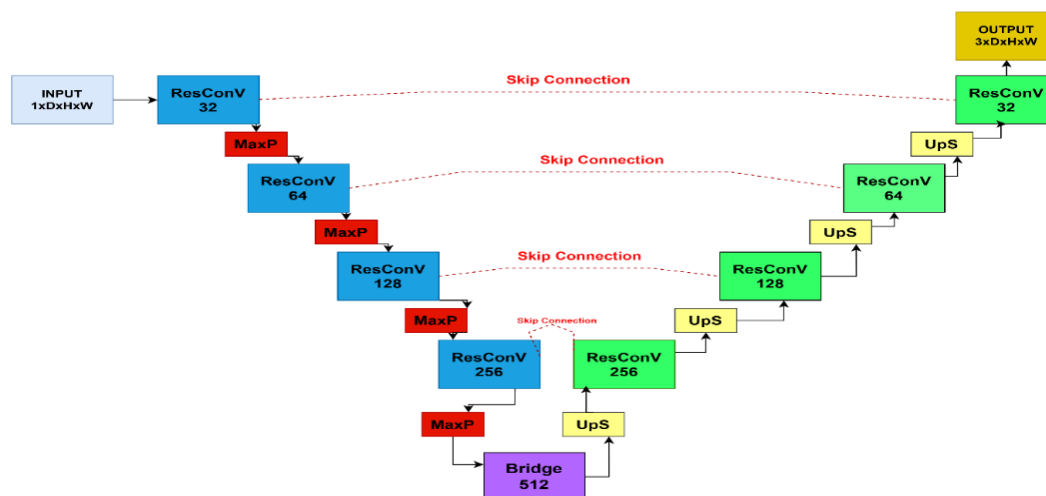


Figure 6. 3D Res-U-Net Architecture

2.3.4. Experimental Settings

In this study, U-Net architectures were utilized with a dropout rate of 0.2 and without the use of pretrained weights. The multi-class Dice Loss function was adopted to address class imbalance and ensure accurate segmentation. Model optimization was performed using the Adam optimizer, configured with a learning rate of 1×10^{-4} , a weight decay of 1×10^{-5} , and the AMSGrad variant enabled to improve convergence stability. To further refine the training process, a learning rate scheduling strategy based on ReduceLROnPlateau was employed, which adaptively reduces the learning rate when the validation metric shows no improvement. In this setup, the learning rate was halved after 10 consecutive stagnant epochs, allowing for more effective exploration of the optimization landscape and mitigating the risk of convergence to suboptimal local minima. The training process was conducted over 200 epochs with a batch size of 4.

To complement the training setup, inference time analyses were also carried out on different hardware configurations (CPU: Intel i7 10th Gen; GPU: RTX 3050). Average inference times were measured separately for pancreas and pancreatic tumor segmentation across all models, providing an additional benchmark for evaluating computational efficiency in clinical environments.

2.3.5. Evaluation Metrics

Dice, IoU, sensitivity and specificity are used to evaluate the proposed models.

Dice: It is a commonly used metric in medical image segmentation that quantifies the overlap between the predicted segmentation and the ground truth. It can be formulated as follows:

$$\text{Dice} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \quad (2.1)$$

IoU (Intersection over Union): Also known as the Jaccard Index, IoU measures the ratio of the intersection to the union between the predicted and ground truth regions. It can be formulated as follows:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (2.2)$$

Sensitivity: Sensitivity measures the proportion of true positives correctly identified by the model, reflecting its ability to detect relevant regions. It can be formulated as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.3)$$

Specificity: Specificity measures the proportion of true negatives correctly identified by the model, indicating its

ability to avoid false positives. It can be formulated as follows:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.4)$$

3. Results and Discussion

In this study, the segmentation performances of three different U-Net-based architectures—2D U-Net, 3D U-Net, and Residual 3D U-Net—were evaluated on pancreas and tumor segmentation tasks. Quantitative comparisons were conducted using standard metrics including Dice score, IoU, sensitivity and specificity. As shown in Table 2.

For pancreas segmentation, the 2D U-Net achieved a Dice score of 0.77, IoU of 0.63, sensitivity of 0.82, and specificity of 0.97. The 3D U-Net obtained slightly higher performance with a Dice score of 0.78, IoU of 0.64, sensitivity of 0.76, and specificity of 0.98. The best pancreas segmentation results were observed with the 3D

Res-U-Net, yielding a Dice score of 0.81, IoU of 0.69, sensitivity of 0.83, and specificity of 0.96.

In the tumor segmentation task, the 2D U-Net reported a Dice score of 0.40, IoU of 0.25, sensitivity of 0.62, and specificity of 0.98. The 3D U-Net improved upon these results, achieving a Dice score of 0.51, IoU of 0.34, sensitivity of 0.55, and specificity of 0.97. Comparable performance was achieved by the 3D Res-U-Net, with a Dice score of 0.51, IoU of 0.40, sensitivity of 0.78, and specificity of 0.95. These results indicate that while 3D architectures generally offer advantages in pancreas segmentation, their performance in tumor segmentation remains challenging.

Figure 7 shows that the 2D U-Net captures the pancreas location and shape reasonably well, though boundary details are less precise. Figure 8 demonstrates that the model can localize tumors, but segmentation remains coarse, reflecting the limitations of 2D context for small and low-contrast lesions. Figure 9 demonstrates pancreas segmentation with the 3D U-Net. The prediction closely matches the ground truth, showing smoother boundaries and better structural continuity compared to the 2D model, highlighting the advantage of volumetric context. Figure 10 presents tumor segmentation results. The 3D U-Net successfully localizes the tumor and provides a more compact and anatomically consistent mask than the 2D model. However, minor discrepancies in size and shape remain, reflecting the inherent difficulty of small lesion delineation. Figure 11 illustrates pancreas segmentation using the 3D Res-U-Net, where the predicted mask shows close agreement with the ground truth. Figure 12 presents tumor segmentation with the same model, demonstrating strong overlap with the ground truth and providing a more precise delineation compared to both the 2D U-Net and the standard 3D U-Net.

Table 2. Pancreas and Pancreatic Tumor Segmentation Results of U-Net Based Model

Model	Target	Dice	IOU	Sensitivity	Specificity
2D U-Net	Pancreas	0.77	0.63	0.82	0.97
	Pancreatic Tumor	0.40	0.25	0.62	0.98
3D U-Net	Pancreas	0.78	0.64	0.76	0.98
	Pancreatic Tumor	0.51	0.34	0.55	0.97
3D Res-U-Net	Pancreas	0.81	0.69	0.83	0.96
	Pancreatic Tumor	0.51	0.40	0.78	0.95

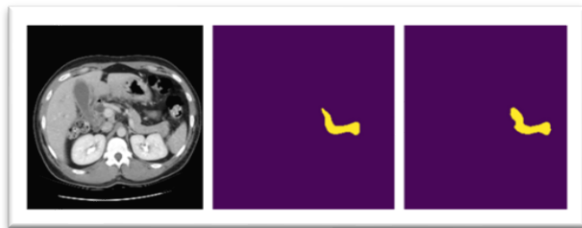


Figure 7. 2D U-Net pancreas segmentation: CT, ground truth, and prediction.

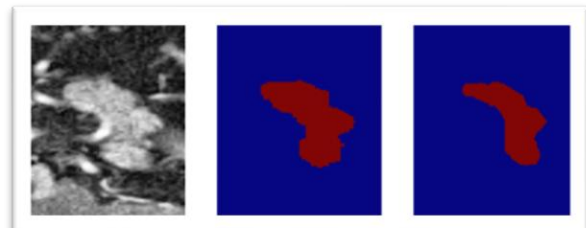


Figure 11. 3D Res-U-Net pancreas segmentation: CT, ground truth, and prediction.

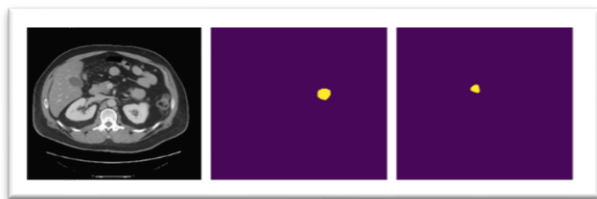


Figure 8. 2D U-Net tumor segmentation: CT, ground truth, and prediction.

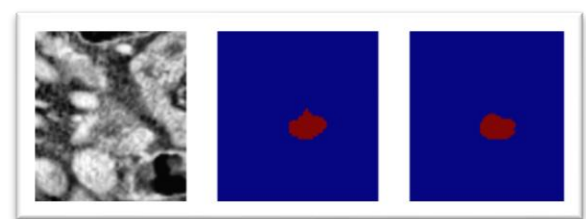


Figure 12. 3D-Res-U-Net tumor segmentation: CT, ground truth, and prediction.

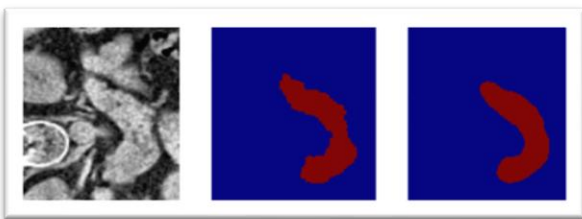


Figure 9. 3D U-Net pancreas segmentation: CT, ground truth, and prediction.

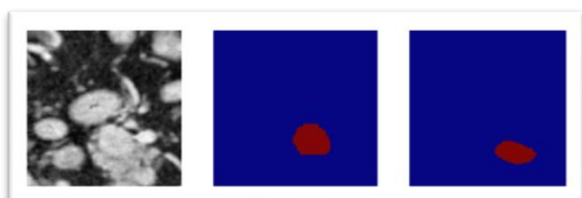


Figure 10. 3D U-Net tumor segmentation: CT, ground truth, and prediction.

Table 3 presents the average inference times per CT scan for each U-Net model under both CPU (Intel i7 10th Gen) and GPU (RTX 3050) environments. The findings highlight significant differences in computational efficiency across model architectures. Among all models, the 2D U-Net demonstrated the fastest inference times, requiring approximately 90–92 seconds on CPU and 25–26 seconds on GPU per CT volume for both pancreas and tumor segmentation tasks. These results emphasize the computational efficiency of 2D models, making them highly suitable for real-time or resource-constrained clinical applications. In contrast, the 3D U-Net exhibited significantly longer inference times, particularly under CPU conditions—up to 344 seconds per volume for tumor segmentation. Even with GPU acceleration, inference took 155–165 seconds, highlighting the model's higher computational complexity and latency. This performance limitation may restrict its use in time-sensitive diagnostic workflows.

The Residual 3D U-Net, being the most complex architecture, recorded the longest inference times overall,

with up to 420 seconds on CPU and 180 seconds on GPU for tumor segmentation. Although this model achieved a more balanced trade-off between sensitivity and precision, as previously discussed, the high computational cost indicates that further optimization or hardware resources may be necessary for practical deployment.

Table 3. Average Inference Time Across Different Hardware Configurations (GPU: 3050 RTX, CPU: Intel i7 10th Gen)

U-Net Models	Target	Avg. CPU Inference Time(s)	Avg. GPU Inference Time (s)
2D U-Net	Pancreas	90.0	25.0
2D U-Net	Pancreatic Tumor	92.0	26.0
3D U-Net	Pancreas	298.0	165.0
3D U-Net	Pancreatic Tumor	344.0	155.0
3D-Res-U-Net	Pancreas	371.0	171.0
3D-Res-U-Net	Pancreatic Tumor	420.0	180.0

The analysis confirms that model selection should consider not only segmentation accuracy but also inference efficiency. While 3D architectures may offer advantages in volumetric context and clinical accuracy, 2D models remain a viable and efficient alternative, especially when speed and resource usage are critical factors.

Table 4. Comparison of pancreas and pancreatic tumor segmentation performance across related studies

Study	Model	Datas et	Target	Dice	IOU
[7]	U-Net+attention	NIH	Pancreas	0.84	-
[8]	Vanilla U-Net	TCIA	Pancreas	0.81	0.80
[26]	SRS GS	MSD	Pancreas /Tumor	0.89/0.54	-
[12]	Res-U-Net	NIH	Pancreas	0.903	0.82
[16]	nn-U-Net	MSD	Pancreas /Tumor	0.69/0.37	0.53/0.35
[13]	3D U-Net	NIH	Pancreas	0.87	0.77
Ours	3D Res-U-Net	5 Datasets	Pancreas /Tumor	0.81/0.51	0.69/0.40

Table 4 shows the comparison across related works. The studies on single datasets such as NIH reported high pancreas Dice scores, up to 0.903 with Residual U-Net [12], while tumor segmentation remained considerably lower, as in SRS GS (0.54) [26] and nnU-Net on MSD (0.37) [16]. In contrast, our 3D Res-U-Net achieved 0.81 for pancreas and 0.51 dice scores for tumors across five heterogeneous datasets, demonstrating competitive pancreas performance and notably stronger tumor segmentation under more challenging and generalizable conditions. These results underscore the clinical relevance of our approach, as robust tumor delineation across diverse datasets is essential for early diagnosis and treatment planning in pancreatic cancer.

3.1. Clinical Implications and Future Directions

The findings of this study carry important implications for the clinical management of pancreatic cancer. Accurate pancreas and tumor segmentation is essential for volumetric analysis, radiotherapy planning, surgical navigation, and longitudinal monitoring of treatment response. While 3D models such as the 3D Res-U-Net demonstrated stronger performance in pancreas segmentation, the superior computational efficiency of 2D models positions them as highly practical for time-sensitive and resource-constrained environments such as routine diagnostic radiology. This underscores the need for context-specific model selection that balances accuracy with feasibility in clinical workflows.

By leveraging preprocessing strategies that generalize across heterogeneous datasets, our approach advances the development of segmentation systems with broader clinical reliability. Nevertheless, further validation in multi-center prospective studies is required to confirm performance in diverse patient populations, scanner types, and clinical settings. Future work should focus on seamless integration of these models into PACS/RIS platforms, extending segmentation outputs toward predictive tasks such as treatment outcome estimation and survival modeling. Finally, hybrid approaches that combine the efficiency of 2D models with volumetric refinements from 3D architectures may represent an optimal path toward clinically deployable solutions.

3.2. Limitation

Despite the promising outcomes, this study has several limitations. First, the heterogeneity of the datasets, while enhancing generalizability, may also introduce potential biases arising from differences in imaging protocols, scanners, and annotation standards across centers. Second, the relatively small number of tumor cases compared to pancreas cases increases the risk of class imbalance, which may reduce robustness in tumor segmentation. Third, although regularization strategies such as dropout and learning rate scheduling were applied, the potential for overfitting remains, as models may capture dataset-specific features that hinder

generalization to unseen data. Finally, the high computational cost of 3D architectures poses practical challenges for deployment in time-sensitive or resource-constrained clinical environments.

4. Conclusion

In this study, we systematically evaluated 2D, 3D, and residual U-Net architectures for pancreas and tumor segmentation. Unlike many prior works that rely on single datasets, our approach integrated multiple heterogeneous datasets and implemented dataset-agnostic preprocessing pipelines to ensure robust training and generalizability. In addition to segmentation accuracy, we also investigated inference times across CPU and GPU configurations, providing insights into the clinical feasibility of different architectures. The findings indicate that while 2D U-Net offers an efficient and resource-friendly solution, residual 3D models better capture volumetric context, particularly for anatomically complex structures. Importantly, the study highlights that model complexity alone does not ensure superior outcomes, and that preprocessing strategies play a pivotal role in enhancing volumetric segmentation performance. Collectively, these contributions provide a comprehensive benchmark that balances accuracy, efficiency, and clinical usability, offering practical guidance for model selection in real-world applications.

Acknowledgement

We gratefully acknowledge the AKGÜN Technology for their contributions and financial support.

Author's Contributions

İbrahim Şen: Performed the experiments and result analysis and drafted the manuscript,

Azer Çelikten: Assisted in analytical analysis, supervision, result interpretation, drafted and wrote the manuscript.

Ece Bingöl: Assisted in analytical analysis, supervised the experiment's progress, result interpretation.

Muhammed Kerem Demir: Assisted in analytical analysis, supervised the experiment's progress, result interpretation.

Andaç Akpulat: Assisted in analytical analysis, supervised the experiment's progress, result interpretation.

Semih Demirel: Assisted in analytical analysis and critical review of the manuscript.

Abdülkadir Budak: Project administration, supervision, result interpretation and critical review of the manuscript.

Hakan Karataş: Project administration, supervision.

Ethics

There are no ethical issues after the publication of this manuscript.

References

- [1]. Cancer Research UK. Pancreatic cancer survival. <https://www.cancerresearchuk.org/about-cancer/pancreatic-cancer/survival> (accessed at 30.04.2025).
- [2]. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211. (<https://doi.org/10.1038/s41592-020-01008-z>)
- [3]. Christ, P. F., Ettliger, F., Grün, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S., & Menze, B. (2017). Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. *arXiv preprint*, arXiv:1702.05970.
- [4]. Bozkır, Ö. F., Urfalı, A., Çelikten, A., Demirel, S., Budak, A., Karataş, H., & Ceylan, M. (2024). Enhancing aorta segmentation in contrast CT images: A novel deep architectural approach. *Osmaniye Korkut Ata University Journal of The Institute of Science and Technology*, 7(5), 2284–2303. (<https://doi.org/10.47495/okufbed.1457674>)
- [5]. Zhang, R., Zheng, Y., Poon, C. C. Y., Shen, D., & Lau, J. Y. W. (2018). Polyp detection during colonoscopy using a regression-based convolutional neural network with a tracker. *Pattern Recognition*, 83, 209–219. (<https://doi.org/10.1016/j.patcog.2018.05.026>)
- [6]. Jain, S., Sikka, G., & Dhir, R. (2024). A systematic literature review on pancreas segmentation from traditional to non-supervised techniques in abdominal medical images. *Artificial Intelligence Review*, 57(12), 317. (<https://doi.org/10.1007/s10462-024-10966-1>)
- [7]. Li, M., Lian, F., Li, Y., & Guo, S. (2022). Attention-guided duplex adversarial U-Net for pancreatic segmentation from computed tomography images. *Journal of Applied Clinical Medical Physics*, 23(4), e13537. (<https://doi.org/10.1002/acm2.13537>)
- [8]. Devendhar, T., Priya, L., & Jayalakshmy, S. (2024). U-Net based pancreas segmentation from computed tomography images. In *Proceedings of the 2024 Third International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, pp. 1–5. IEEE. (<https://doi.org/10.1109/ICSTSN61422.2024.10671042>)
- [9]. Ghorpade, H., Kolhar, S., Jagtap, J., & Chakraborty, J. (2024). An optimized two stage U-Net approach for segmentation of pancreas and pancreatic tumor. *MethodsX*, 13, 102995. (<https://doi.org/10.1016/j.mex.2024.102995>)
- [10]. Tualombo, M. E., Reyes, I., Vizcaino-Imacaña, P., & Morochocayamcela, M. E. (2024). Pancreas segmentation using SRGAN combined with U-Net neural network. In *Conference on Information and Communication Technologies of Ecuador*, pp. 21–38. Springer Nature Switzerland. (https://doi.org/10.1007/978-3-031-75431-9_2)
- [11]. Bayram, A. F., Gürkan, C., Budak, A., & Karataş, H. (2022). Fully automatic end-to-end convolutional neural networks-based pancreatic tumor segmentation on CT modality. *Turkish Journal of Forecasting*, 6(2), 67–72. (<https://doi.org/10.34110/forecasting.1190299>)
- [12]. Derin, A., Gürkan, C., Budak, A., & Karataş, H. (2022). Pancreas segmentation using U-Net based segmentation networks in CT

modality: A comparative analysis. *European Journal of Science and Technology*, 40, 94–98. (<https://doi.org/10.31590/ejosat.1171803>)

Biomedical Signal Processing and Control, 87, 105562. (<https://doi.org/10.1016/j.bspc.2023.105562>)

[13]. Chen, L., & Wan, L. (2023). CTUNet: Automatic pancreas segmentation using a channel-wise transformer and 3D U-Net. *The Visual Computer*, 39(11), 5229–5243. (<https://doi.org/10.1007/s00371-022-02656-2>)

[14]. Uslucuk, A., & Öcal, H. (2023). Comparative analysis of baseline VNet and UNet architectures on pancreas segmentation. *Orclever Proceedings of Research and Development*, 3(1), 146–157. (<https://doi.org/10.56038/oprd.v3i1.309>)

[15]. Ghorpade, H., Jagtap, J., Patil, S., Kotecha, K., Abraham, A., Horvat, N., & Chakraborty, J. (2023). Automatic segmentation of pancreas and pancreatic tumor: A review of a decade of research. *IEEE Access*, 11, 108727–108745. (<https://doi.org/10.1109/ACCESS.2023.3320570>)

[16]. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., & Cardoso, M. J. (2022). The medical segmentation decathlon. *Nature Communications*, 13(1), 4128. (<https://doi.org/10.1038/s41467-022-30695-9>)

[17]. Ji, Y., Bai, H., Yang, J., Ge, C., Zhu, Y., Zhang, R., Li, Z., Zhang, L., Ma, W., Wan, X., & Luo, P. (2022). AMOS: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *Advances in Neural Information Processing Systems*, 35, 36722–36732. (<https://arxiv.org/abs/2206.08023>)

[18]. Wasserthal, J., Breit, H. C., Meyer, M. T., Pradella, M., Hinck, D., Sauter, A. W., Heye, T., Boll, D., Cyriac, J., Yang, S., Bach, M., & Segeroth, M. (2023). TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*, 5(5), e230024. (<https://doi.org/10.1148/ryai.230024>)

[19]. Zhang, U. (2021). Pancreas cancer segmentation on portal-venous phase CTs [Dataset]. Zenodo. (<https://doi.org/10.5281/zenodo.5510650>)

[20]. Ma, J., et al. (2021). FLARE 2021: Fast and low-resource abdominal organ segmentation challenge [Dataset]. Grand Challenge. <https://flare.grand-challenge.org/> (accessed at 23.06.2025).

[21]. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241. Springer International Publishing. (https://doi.org/10.1007/978-3-319-24574-4_28)

[22]. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 424–432. Springer International Publishing. (https://doi.org/10.1007/978-3-319-46723-8_49)

[23]. Fu, W., Ma, Z., Yang, Z., Yu, S., Zhang, Y., Zhang, X., Mei, B., Meng, Y., Ma, C., & Gong, X. (2025). Fully automated image quality assessment based on deep learning for carotid computed tomography angiography: A multicenter study. *Physica Medica*, 134, 104990. (<https://doi.org/10.1016/j.ejmp.2025.104990>)

[24]. Wang, J., Zhang, J. H., Zhang, J. L., Lu, F. M., Meng, R. G., & Wang, Z. (2021). Research on fault recognition method combining 3D Res-UNet and knowledge distillation. *Applied Geophysics*, 18(2), 199–212. (<https://doi.org/10.1007/s11770-021-0894-2>)

[25]. Liu, J., Yildirim, O., Akin, O., & Tian, Y. (2023). AI-driven robust kidney and renal mass segmentation and classification on 3D CT images. *Bioengineering*, 10(1), 116. (<https://doi.org/10.3390/bioengineering10010116>)

[26]. Cao, L., & Li, J. (2024). Strongly representative semantic-guided segmentation network for pancreatic and pancreatic tumors.