IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 February 2024, accepted 4 April 2024, date of publication 23 April 2024, date of current version 10 June 2024.
*Digital Object Identifier 10.1109/ACCESS.2024.3392595*

SURVEY

# State-of-the-Art and Challenges in Pancreatic CT Segmentation: A Systematic Review of U-Net and Its Variants

**CHAOHUI ZHANG[1,2], (Student Member, IEEE), ANUSHA ACHUTHAN[1], (Member, IEEE), AND GALIB MUHAMMAD SHAHRIAR HIMEL[1]**
[1]School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Malaysia
[2]Center for Technology Enhanced Learning, Guangdong Maoming Health Vocational College, Maoming 525400, China
Corresponding author: Anusha Achuthan (anusha@usm.my)

This work was supported in part by Guangdong University Featured Innovation Program Project 2021, Grant number: 2021KTSCX326.

**ABSTRACT** In medical image analysis, segmenting pancreatic CT images presents a significant challenge due to the complex anatomy of the pancreas and the generally low contrast of these images. Accurate pancreas segmentation is crucial in clinical scenarios, particularly for the diagnosis and treatment of pancreatic cancer. The U-Net architecture and its variations have achieved significant progress in deep learning-based image segmentation, especially in the context of pancreatic CT image segmentation. However, there is a noticeable gap in the comprehensive evaluation of their performance, limitations, and potential improvements specifically in this area. This systematic review aims to address this gap in the literature, focusing particularly on U-Net and its variants in pancreatic CT image segmentation. Adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, this review includes relevant studies published since 2019 in the field of pancreatic segmentation. The findings illuminate the current limitations of these methods and establish a theoretical foundation for future research directions.

**INDEX TERMS** Pancreas segmentation, U-Net, CT imaging, systematic review, deep learning.

## I. INTRODUCTION

Over the last decade, deep learning has propelled groundbreaking developments across various domains such as computer vision and natural language processing [1] and uncovered extensive potential in the area of medical image analysis [2]. Nevertheless, deep learning faces substantial obstacles when dealing with the diversity and intricacy of medical data, especially organs like the pancreas, which exhibit complex anatomy and typically low image contrast [3].

The pancreas is indispensable for endocrine and exocrine functions. Alarmingly, the incidence of pancreatic cancer is escalating, ranking it among the most lethal cancers worldwide [4]. This serious situation highlights the necessity of accurate pancreatic segmentation in medical diagnostics and treatment, which is vital for the advancement of personalized medicine and tracking treatment efficacy [5].

In the realm of biomedical image analysis, U-Net, a neural network engineered explicitly for image segmentation tasks [6], has proven to be highly effective and adaptable for segmenting pancreatic Computed Tomography (CT) images [7]. Although numerous investigations have explored a variety of approaches for pancreatic CT segmentation [8], [9], [10], there is a considerable disparity in the focus of these studies. For instance, Bowen et al. not only focus on pancreatic segmentation but also emphasize the importance of introducing human interaction to intervene and optimize segmentation outcomes, especially when initial automated segmentation is unsatisfactory [7]. Conversely, Kumar et al. [9] and Poce et al. [11] have conducted exhaustive discussions of diverse segmentation techniques, while

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

Xiao et al. have exclusively investigated the applications of transformers in pancreatic segmentation [10].

Despite the wide range of topics addressed, there is an absence of a comprehensive evaluation of U-Net and its derivatives in the particular context of pancreatic CT image segmentation in the existing literature. This systematic review is intended to fill this gap, providing a complete assessment of the capabilities and limitations of U-Net and its variants in the application of pancreatic CT image segmentation, as well as potential avenues for improvement. This review delineates the boundaries of current technologies with clarity, thus laying a solid theoretical foundation for future work. Additionally, future research directions suited to this specific application scenario are proposed. By adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, the review includes pivotal studies published since 2019 [12].

The subsequent structure of this paper is as follows: Section II details the research methodology foundational to the analysis; Section III introduces U-Net and its variants, preparing the ground for discussions on their enhancements; Section IV provides a review and analysis of segmentation methods with an emphasis on component-level changes in U-Net, such as the introduction of attention mechanisms and the optimization of skip connections; Section V discusses network-level adjustments, including strategies for multi-input configurations and progressive U-Net models; Section VI concentrates on evaluation methodologies, popular datasets like the NIH pancreas dataset, and current trends. The paper concludes with a comprehensive discussion in Section VII, exploring future research directions and challenges, followed by a final conclusion in Section VIII.

## II. SEARCH METHODOLOGY

This review aims to summarise and identify relevant research articles from the past five years that employ computer vision techniques for segmenting pancreatic CT images. Adhering to the PRISMA framework, specific inclusion and exclusion criteria were applied, and papers and journal articles published on eight databases were screened: ACM Digital Library, IEEE Xplore, Springer Link, Web of Science, MDPI, Scopus, PubMed, and Science Direct. The time frame for the published research was limited to 2019-2023.

For literature retrieval, the following keywords were utilized: (pancreas OR "pancreatic cancer" OR "pancreatic tumour") AND ("automated segmentation" OR "automatic segmentation" OR "segmentation" OR "semi-automatic segmentation" OR "semi-automated segmentation"). Literature collection and management were facilitated using EndNote 20 to ensure data integrity and accuracy.

The initial search yielded a total of 1641 studies. After the removal of duplicates, 620 studies were retained. Titles, abstracts, and methods were extracted as the foundational data for preliminary screening, which eliminated 448 studies not based on CT modalities, leaving 172 studies for full-text review. Studies were excluded if they met the following conditions: non-English articles, repeated articles, articles without full text available, or articles published before 2019. Inclusion criteria were studies employing CT imaging, using U-Net-like networks for segmentation, and focusing on pancreatic segmentation. The full-text review indicated that 44 studies met the inclusion and exclusion criteria. Of these, four focused on pancreatic cancer segmentation, one on pancreatic duct segmentation, and one on Pancreatic ductal adenocarcinoma (PDAC) segmentation. A forward and backward citation search methodology proposed by Webster and Watson [13] was also employed to guarantee a more comprehensive literature assessment, thus enhancing the review's comprehensiveness and accuracy. This search method added one more study. Following all screenings and searches, 39 studies were ultimately selected for inclusion in this review. These works' network models, datasets, and performance metrics were extracted for comparison. The detailed filtering and search strategy is illustrated in Figure 1.

It should be noted that focus is placed on using U-Net and U-Net-like architectures for pancreatic segmentation without covering other network architectures or delving into the segmentation of pancreatic cancer or pancreatic ducts.

## III. U-NET AND ITS VARIANTS

The U-Net model, introduced by Ronneberger et al. in 2015, is a model with a symmetric encoder-decoder network structure [6]. Its distinctive "U-shaped" architecture consists of an encoder, a decoder, and skip connections between them, illustrated in Figure 2. The encoder, also known as the "contracting path," employs convolution and pooling operations to effectively extract both low-level and high-level features from the input image, substantiating its critical role in medical image segmentation [2], [14]. Concurrently, the downsampling process lowers the computational complexity for subsequent layers but may sacrifice some spatial information [15], [16]. The decoder, or "expansive path," reconstructs the image details and dimensions through upsampling [17]. The skip connections merge low-level and high-level features to further enhance segmentation accuracy [6], [18]. While these designs elevate the model's performance, they also increase its computational complexity, especially when dealing with large or high-resolution images.

It's worth noting that there are variants of U-Net tailored for different application scenarios and requirements. For example, V-Net and 3D U-Net are designed for three-dimensional images [16], [17].Unet++ introduces nested and dense skip-over connections to bolster feature learning and utilizes a deeper U-shaped structure to broaden the receptive field, aiming to augment segmentation precision [14]. In contrast, nnU-Net combines a pre-trained encoder to boost generalization capabilities and employs a configurable architecture to adapt to diverse data types, striving to improve performance generalization across multiple medical imaging tasks [19].

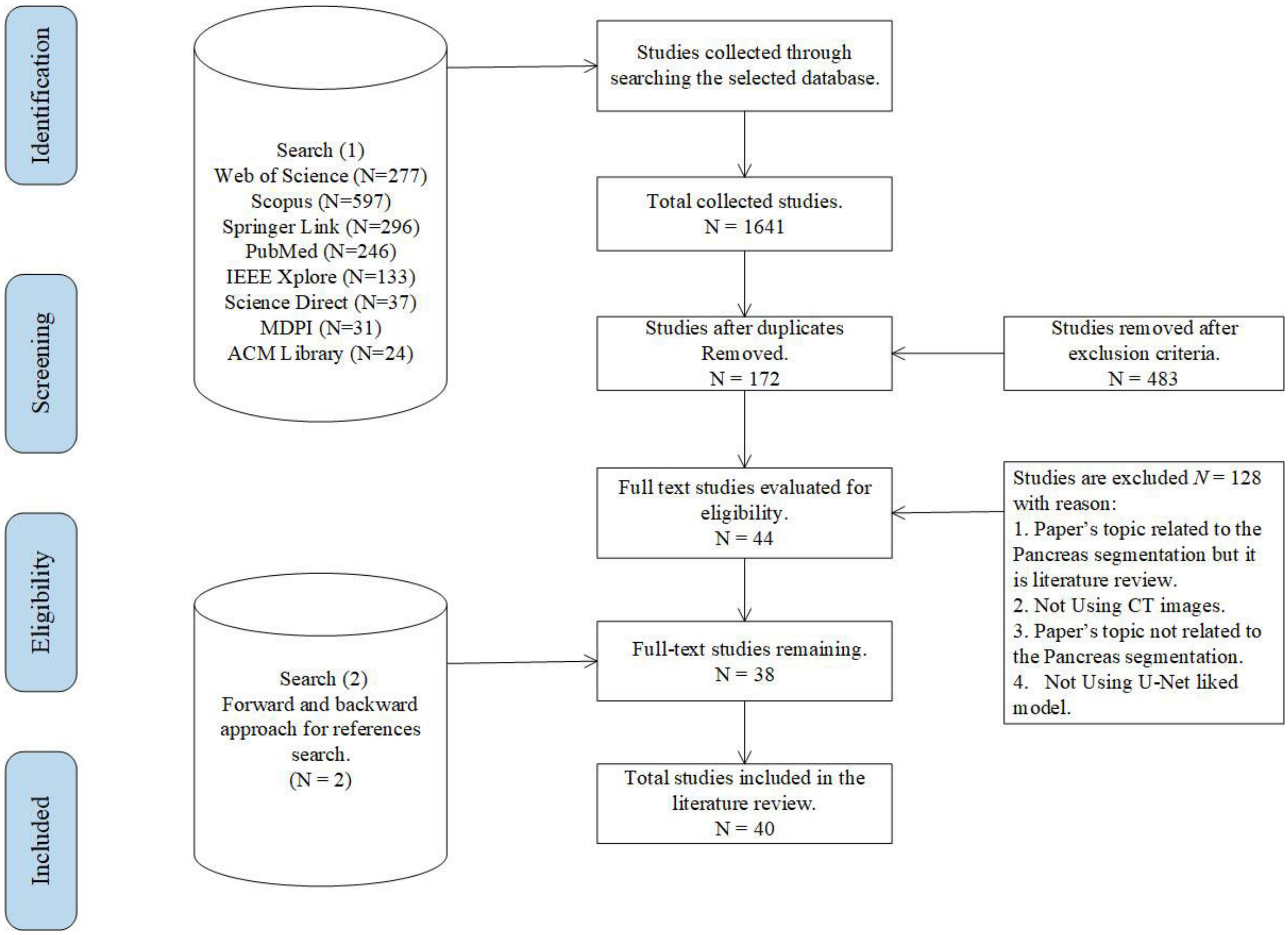


**FIGURE 1. PRISMA-guided search and selection strategy [12].**

The choice between 2D and 3D U-Net hinges on application requirements and computational resources. For instance, 3D U-Net captures more complex volumetric relationships at a higher computational cost, while 2D U-Net is more efficient but has limitations in utilizing anatomical context. This study aims to investigate how various network architectures and mechanisms can enhance the performance of pancreatic CT image segmentation. In this context, dimensionality (2D or 3D) is not the main distinguishing factor. Rather, the focus is on effectively incorporating attention mechanisms or other structural enhancements to achieve superior segmentation precision and robustness. The study considers U-Net and its variants as a collective, presenting a comprehensive overview of the latest advancements and challenges in pancreatic CT segmentation.

## IV. MODIFICATIONS TO U-NET COMPONENTS

In the continuous quest to enhance the U-Net architecture for more nuanced and robust medical image processing, a substantial body of research has concentrated on modifications at the component level [20], [21]. These adjustments range from data augmentation strategies that enrich the input data spectrum, attention mechanisms that guide the model focus, and alterations in convolutional layers for more refined feature extraction to adaptations in pooling layers, skip connections, and comprehensive integrations of various modifications. Each component-level change targets a specific aspect of the U-Net's functionality, aiming to boost the model's performance metrics and adaptability to diverse imaging challenges. Figure 3 delineates these component-level subdivisions, offering insights into the multifaceted approaches researchers have employed to optimise the individual elements of the U-Net architecture.

### A. INTEGRATION OF ATTENTION MECHANISMS

The importance of attention mechanisms in image segmentation has become increasingly apparent. Mnih et al. first introduced this concept, using Recurrent Neural Networks (RNNs) to simulate the focusing aspect of visual attention, thus enhancing performance in tasks such as image classification and object detection [22]. Further studies, including Woo et al.'s Convolutional Block Attention Module

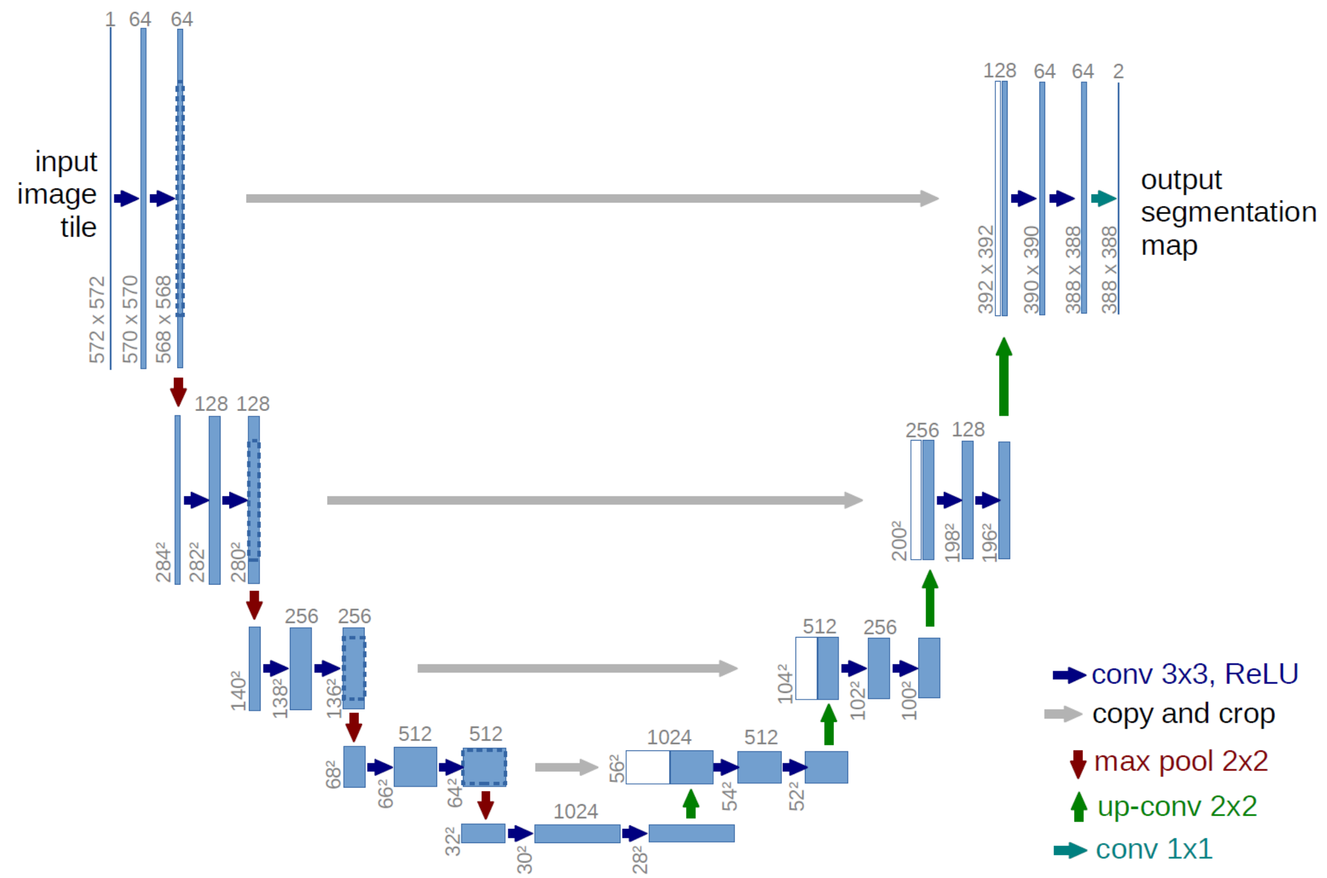


**FIGURE 2.** **Basic U-net architecture.**

(CBAM) and Wang et al.'s Non-local Neural Networks, have delved deeper into spatial attention mechanisms, enabling neural networks to grasp long-range dependencies and intricate spatial structures within images, proving particularly beneficial in various image segmentation tasks [23], [24].

Similarly, channel attention mechanisms, exemplified by Squeeze-and-Excitation Networks (SENet) and Efficient Channel Attention Networks (ECA-Net), adjust the weights across different channels, refining feature representations and, consequently, bolstering performance across numerous visual tasks [25], [26]. Remarkably, feed-forward attention networks like the Residual Attention Network and Dual Attention Network provide end-to-end solutions capable of processing granular and high-level visual information, offering extensive support for diverse visual tasks [27], [28].

Dosovitskiy et al. have showcased the productive application of self-attention mechanisms in visual tasks, specifically emphasising pancreatic segmentation [29]. A groundbreaking development in the literature is the Attention Gate model (AG), introduced in [30]. This innovative method, incorporated into the conventional U-Net architecture, significantly amplifies the model's proficiency in identifying nuanced tissue characteristics, particularly in low-contrast pancreas tissue images with varied morphology. Despite its tendency to generate false positives with small or morphologically diverse tissues, this method markedly enhances segmentation accuracy by zeroing in on pertinent features. Furthermore, this study is a pivotal guide for future endeavours aimed at refining attention gate models' training protocols, investigating higher-resolution inputs, and bettering connection strategies.

Subsequent literature introduces more advancements. For instance, Li et al. outline a Multi-scale Attention Dense Residual U-Net (MAD-UNet), which bolsters the model's capacity for pinpointing pancreatic nuances by amalgamating dense residual blocks and multi-scale convolution kernels [31]. This approach augments contextual information via the attention mechanism, thus sharpening the precision of segmentation boundaries. Yan and Zhang present a 2.5D U-Net architecture that merges 2D and 3D convolution layers, moderating computational demands while preserving substantial spatial data [32]. Incorporating spatial and channel attention mechanisms further uplifts the calibre of feature representation. Wang et al. feature a network known as V-mesh, which heightens sensitivity to minute details and fosters spatial feature transformation and fusion by instilling an attention-oriented control mechanism within V-Net's skip connections [33]. Chen and Wan suggest a

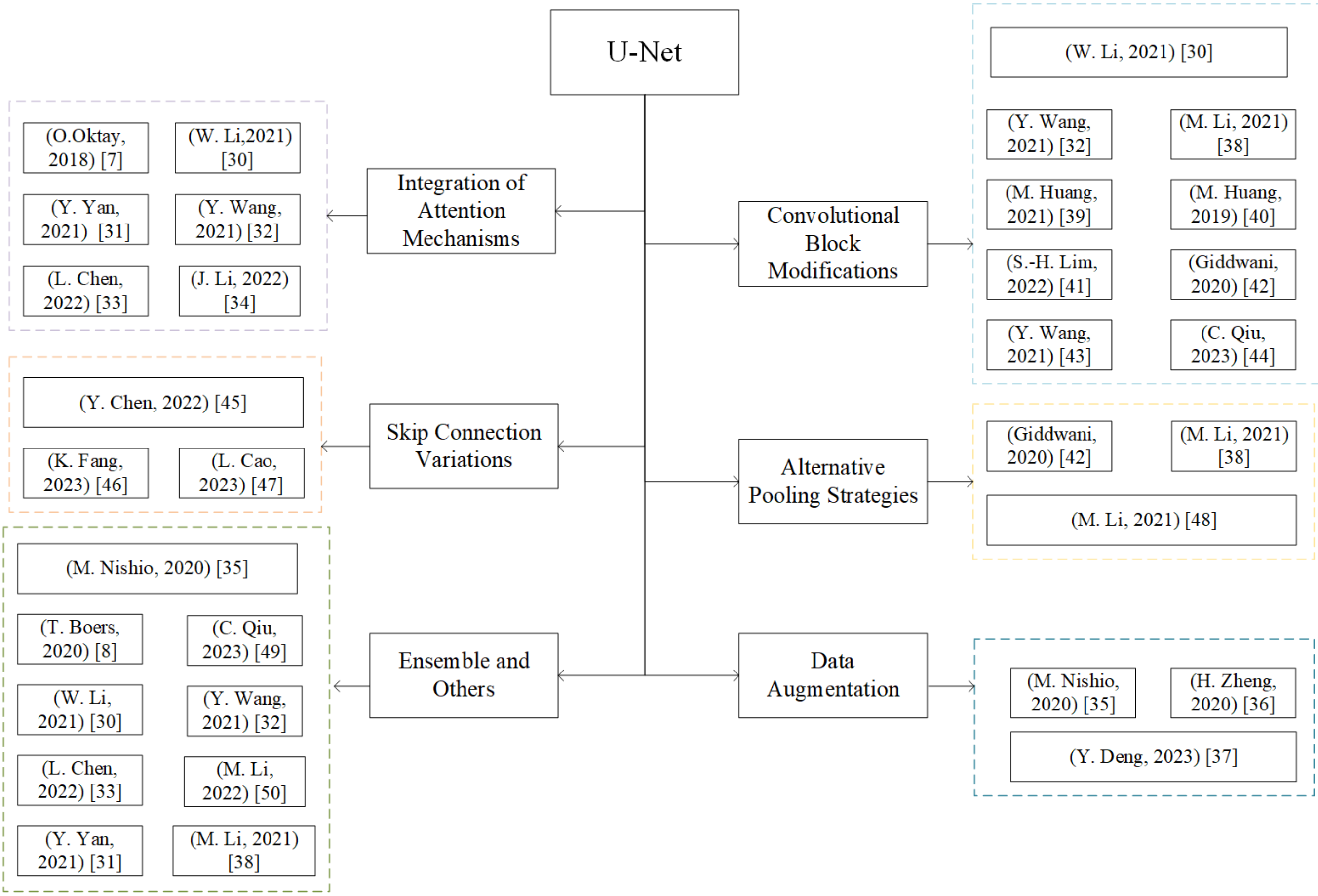


**FIGURE 3.** An overview of U-Net component level modification.

pioneering network that fuses Transformer and 3D U-Net, dubbed CTUNet, achieving harmonisation and instructional oversight of global features by implementing a Transformer within the skip connections [34]. Lastly, Li et al. explore the feasibility of amalgamating several attention mechanisms within a U-Net framework, specifically targeting lightweight 3D voxels [35]. This method underscores the criticality of both global and local features by embedding multiple attention blocks within the skip connections and employing a global context feature (GCF).

Despite the significant performance improvements achieved by integrating attention mechanisms into U-Net networks, issues such as reliance on extensive data and lack of methodological universality remain in the context of complex medical image analysis tasks. Future research should focus on developing algorithms that can effectively process limited data and on optimizing network structures to reduce computational resource dependence. Such harmonisation will aid in extending the application of attention mechanism technologies in medical image analysis.

### *B. DATA AUGMENTATION*

In the domain of complex medical image segmentation, particularly for pancreatic segmentation, the integration of innovative data augmentation strategies is essential to optimize the performance of the U-Net architecture. Nishio et al. marked a significant advancement by integrating data augmentation with the segmentation framework, attaining a Dice coefficient of 0.789 [36]. This study introduced two novel data augmentation techniques, mixup and Random Image Cropping and Patching (RICAP), targeting the prevalent overfitting issues in deep learning models. Furthermore, the study validated the benefits of mixup and RICAP in generating novel training samples alongside traditional data augmentation optimizing rotation, translation, and scaling. The combination of mixup and RICAP led to a 5% improvement in performance over the conventional U-Net architecture.

In contrast, Zheng et al. propose a "shadowed sets" mechanism to create uncertainty regions, which were iteratively fine-tuned using a weight parameter [37]. Utilising U-Net as the primary segmentation architecture to optimize, this study applied the "shadowed sets" mechanism to progressively refine uncertainty weights during training, resulting in improved segmentation accuracy. Significantly, this study expanded the dataset nearly tenfold by extensive data augmentation, encompassing rotation, translation, and scaling.

From another perspective, Deng et al. highlight the critical role of data augmentation in overcoming the inherent limitations of medical imaging datasets, particularly in cases like acute pancreatitis with large slice intervals and scarce slices per patient [38]. By employing data augmentation, this study effectively expanded the training data, thereby enhancing performance in pancreatic segmentation tasks using the U-Net architecture.

These studies illustrate the transformative impact of innovative data augmentation and model refinement strategies, charting new avenues for elevating U-Net's efficacy in complex medical image segmentation tasks. However, a notable gap remains, as most current data augmentation techniques primarily cater to natural images, often neglecting the specific needs of medical imaging. This is particularly true for CT imaging, where data augmentation techniques remain underexplored, revealing a critical area for future research. Specialized data augmentation methods, tailored for the distinct characteristics of CT imaging, promise significant advancements in feature extraction efficiency and segmentation accuracy.

### C. CONVOLUTION BLOCK MODIFICATIONS

Recent advancements in research indicate that researchers have made multiple attempts to further optimise the performance of U-Net, particularly in the design and optimisation of convolutional layers. For instance, Multi-scale Attention and Dense Residual Block (MAD-UNet) incorporated dense connections and specially designed attention modules within its convolutional layers [31]. This integrated design enhances the model's sensitivity towards the target area, particularly the edges. The advantage is that it effectively reduces segmentation errors caused by class imbalance; however, the downside is that it may increase the computational complexity of the model. This method has successfully elevated the Dice coefficient to 0.861 [31] in pancreatic CT image segmentation.

Adversarial loss combined with multi-level pyramid pooling modules [39] optimises the multi-scale representation of features and strengthens the model's discriminative ability for different structures and textures. The merit lies in its excellent performance in handling class imbalance and irregular target segmentation, which may lead to model overfitting [39].

Deformable convolution dramatically improves the model's adaptability to irregular and non-rigid structures by adding adaptive offsets to each sampling position of the 2D convolutional kernel [40], [41]. In complex and highly irregular applications like pancreatic segmentation, this method elevated the Dice coefficient to 0.8725 [40], although it requires additional training time.

Lim et al. [42] and Giddwani et al. [43] have recently replaced the traditional convolutional layers with residual dense blocks, optimising the transfer of information during the down-sampling phase and further enhancing segmentation accuracy. The advantage here is the increased information transfer efficiency, but the drawback is the increased parameter count in the model [42].

Additionally, studies have been aimed at enhancing model capabilities by introducing innovative modifications within the convolutional blocks. For instance, view adaptive 3D U-Net (VA-3DUNet) employs residual blocks instead of standard convolutional layers to augment the model's understanding of the three-dimensional context [44]. The V-mesh network enhances the model's feature extraction capacity by replacing traditional convolutional layers with deformable residual convolutional layers and incorporating attention mechanisms [33]. Meanwhile, the residual transformer U-Net (RTUNet) utilises residual blocks combined with transformer blocks instead of standard convolution in the U-Net encoder part, effectively capturing the pancreas' multi-scale features and high variability [45]. These methods demonstrate the potential to improve segmentation accuracy and the model's contextual understanding, though they also face challenges related to computational complexity, resource demands, and data sensitivity.

### D. SKIP CONNECTION VARIATIONS

In recent years, improvements in the skip connections of the U-Net architecture have garnered significant attention. The primary aim is to convey richer spatial information during the down-sampling and up-sampling processes. One model worth noting is the target-sensitive U-Net (tU-Net), which purposefully optimises the skip connections in traditional U-Net. Instead of duplicating low-resolution features, tU-Net introduces a fuzzy skip connection module, building more advanced semantic features. This not only suppresses irrelevant background features but also enhances features relevant to the target. This modification leads to an approximate 5% improvement in segmentation accuracy [46].

Similarly, MRFormer optimises capturing the target and its surrounding context by embedding multi-head attention mechanisms and residual depth-wise convolutional networks within the skip connections. Fang et al. demonstrate that MRFormer outperforms the basic U-Net model in pancreatic CT image segmentation [47]. Additionally, MDAG-Net incorporates a novel multi-dimensional attention module (MDAG) to bolster skip connections, capturing context information and precisely defining target features more effectively [48].

There are also other intriguing studies worthy of mention. Li et al. improve model focus on areas of interest by adding spatial and channel attention to skip connections [31]. Yan and Zhang employ an adaptive weight allocation mechanism and use adaptive convolution kernels to capture the target region more precisely [32]. Wang et al. utilise dense connections within skip connections, providing more paths for higher-level feature transmission [33]. Lastly, Chen and Wan introduce a "Feedback Fusion" strategy, effectively integrating low-level and high-level features [34].

In summary, the strategies for improving skip connections are diverse and complex. These enhancements have been experimentally proven to address various pancreatic CT image segmentation challenges, such as enhancing segmentation accuracy and reducing mis-segmentation.

### E. ALTERNATIVE POOLING STRATEGY

Given the complexities of pancreatic anatomy and the small size of the pancreas, segmentation of pancreatic CT images is a challenging task crucial for precision medicine. In this context, the role of pooling layers becomes crucial, particularly regarding information compression and feature selection [43].

Multi-Rate Depthwise Dilated Network (MR-DDN) is an improvement over V-Net [43]. In this architecture, conventional pooling layers are replaced with convolutions and depthwise dilated convolutions are introduced to expand the receptive field. Although this enhances the recognition of small targets and reduces the number of parameters, the increased complexity makes it unsuitable for real-time or resource-constrained applications.

Multi-Level Pyramid Pooling Residual U-Net improves performance by incorporating a Multi-Level Pyramid Pooling (MLPP) module [39]. MLPP employs pooling at various scales to capture multi-scale information, combined with residual learning and adversarial training. This allows for stronger feature extraction and generalisation capabilities but at the cost of increased computational demands.

Dual Adversarial U-Net introduces innovative design elements [49]. In addition to replacing standard pooling, the model integrates attention mechanisms to optimise target recognition. This resolves the issue of information loss and improves focus on the task-specific region, although computational complexity is increased as a result.

In summary, pooling layers are critical in U-Net and its variants. Various pooling strategies have advantages and disadvantages, offering opportunities for further optimisation in future research.

### F. ENSEMBLE AND OTHERS

U-Net and its various derivative models have achieved noteworthy improvements in pancreatic CT image segmentation. These improvements predominantly focus on data augmentation, network architecture, interactive segmentation, and multi-scale feature fusion.

Firstly, Nishao et al. introduce a novel approach that combines data augmentation with the segmentation framework [36]. The study specifically implemented data augmentation techniques like Mixup and random image cropping, demonstrating their superiority over conventional methods. Moreover, the study deepened the traditional U-Net model, especially in terms of down-sampling and up-sampling, by increasing the original four layers to 6 layers to enhance segmentation performance.

Secondly, Boers et al. employe an interactive segmentation approach [8]. In this study, certain layers in the U-Net model were unfrozen and renamed as iU-Net layers. If a user is unsatisfied with the generated label map, manual intervention is possible, followed by fine-tuning through retraining the Fully Convolutional Network (FCN). Regarding multi-scale feature fusion, two studies have made significant contributions. Qiu et al. adopt a coarse-to-fine strategy and performed vertical calibration of features at each scale [50]. In contrast, Li et al. introduce multi-level pyramid pooling modules and adversarial mechanisms to capture richer multi-scale features [39].

Furthermore, some studies have also incorporated attention mechanisms to bolster model performance. To address the issues of insufficient sensitivity to pancreatic details and difficulty in distinguishing contextual information between the pancreas and surrounding tissues, Li et al. proposed an enhanced model known as the Multi-scale Attention Dense Residual U-Net (MAD-UNet) [31]. This model bolsters the network's ability to capture image details through the introduction of dense residual blocks and further enhances the model's focus on the pancreatic region via attention mechanisms, thereby improving segmentation accuracy. Additionally, MAD-UNet employs multi-scale feature fusion techniques, enabling the network to concurrently learn contextual information at various scales and strengthening the overall understanding of pancreatic structures. These innovative strategies have endowed MAD-UNet with not only improved accuracy and robustness in pancreatic segmentation tasks but also heightened sensitivity to pancreatic details and a richer representation of contextual information. MAD-UNet holds significant practical value for pancreatic image segmentation in clinical medicine. Advancing further, two studies have explored 3D data and boundary information, respectively. Chen and Wan proposed CTUNet, which integrates the Transformer with 3D U-Net and deploys Transformers at skip-connections to enhance the coherence of global features [34]. A differentiated graph-based visual saliency (GBVS) algorithm was introduced and extracted more useful features through a V-mesh network [33].

Li et al. introduced a segmentation model called MSC-DUNet, which addresses the issue of spatial information loss due to frequent reddening phenomena [51]. It utilises adversarial learning and incorporates multi-scale field selection and multi-channel fusion modules to integrate multi-level features. Lastly, Yan and Zhang introduce a 2.5D U-Net model that includes attention mechanisms within the skip connections and combines 2D and 3D convolutional layers to capture more comprehensive spatial information [32]. This architecture reduces computational complexity while enhancing segmentation accuracy, although it lacks precision in boundary region segmentation. Table 1 provides an overview of the strengths and limitations of each component-level modification method within the U-Net architecture.

## V. NETWORK-LEVEL MODIFICATIONS TO U-NET

Beyond intricate adjustments at the component level, an extensive body of research is dedicated to redefining U-Net at the network level [20]. These extensive

**TABLE 1.** An overview of advantages and limitations of component-level modifications to U-Net.

| Modification Type | Advantages | Limitations |
|---|---|---|
| Integration of Attention Mechanisms | 1. Improve segmentation accuracy.<br>2. Enhance focus on important regions. | 1. Increase computational burden.<br>2. Might prolong training time.<br>3. Correct implementation is key for performance improvement. |
| Data Augmentation | 1. Enhance model generalization.<br>2. Increase dataset diversity.<br>3. Reduce overfitting. | 1. Might lead to incorrect feature learning.<br>2. Needs to be consistent with the application scenario. |
| Convolution Block Modifications | 1. Reduce the number of parameters.<br>2. Mitigate overfitting.<br>3. Increase receptive field. | 1. Might lead to performance degradation.<br>2. Might increase training complexity. |
| Skip Connection Variations | 1. Reduce gradient vanishing.<br>2. Promote feature reuse. | 1. Might make the network overly complex.<br>2. Increase computational cost.<br>3. Might not always improve performance. |
| Alternative Pooling Strategy | 1. Capture different types of image features.<br>2. Reduce feature loss.<br>3. Capture global information. | 1. Might not be suitable for all tasks.<br>2. Might result in the loss of crucial information. |
| Ensemble and Others | 1. Improve performance and stability.<br>2. Reduce biases of individual models. | 1. Increase computational cost and resource demand.<br>2. Need to find the optimal combination strategy. |

modifications encompass a holistic reimagining of the structural and functional paradigms of the network, including the integration of multiple inputs for complex imaging data interpretation, the creation of parallel U-Net architectures for varied feature processing, and the use of cascaded U-Net strategies for iterative refinement. Additionally, the combination of different network modules aims to harness synergistic strengths, marking a shift toward more collaborative learning processes tuned for the rising complexities in medical imaging tasks. To elucidate this concept, Figure 4 depicts these network-level modifications to elucidate this concept, emphasizing their divergent facets.

### A. MULTI-INPUTS AND MULTI-MODAL U-NET

For pancreatic CT image segmentation using U-Net and its variants, the strategic incorporation of multi-inputs represents a significant research trajectory, vital for bringing forth enhanced models. Multi-inputs provide an array of contextual details beyond single-image inputs, thus enriching the models' analytical capabilities. Innovative methodologies include positioning object detection algorithms ahead of the U-Net to refine background noise reduction and sharpen the focus on regions of interest. A notable technique is the integration of a Region Proposal tasked with initial object location, before merging these focused areas with the original image for U-Net processing, enhancing segmentation precision [38].

Uncertainty plays a pivotal role in influencing segmentation outcomes, prompting researchers to incorporate uncertainty weights directly into the baseline U-Net architecture, generated from methodologies like shadowed sets which define regions of uncertainty. During the training process, these weights undergo iterative adjustments, enabling the model to steadily refine and calibrate the segmentation results for increased accuracy [37].

Apart from image information, other types of data can also be integrated into the model to enhance its performance. An innovative approach supplements image features with clinical attributes from case data. These features are processed through a phenotype embedding model and serve as inputs to U-Net, along with imaging data. Such multi-modal inputs give the model a more comprehensive understanding, thus optimising segmentation results [52].

In recent years, the integration of 3D information has also been gaining attention. In an architecture called PBR-UNet, the model generates probability maps representing the likelihood of each pixel belonging to the pancreas, and these maps are then fused with the original image to form multi-channel data. This multi-channel input supplies the model with richer local 3D contextual information, thereby improving segmentation quality [53].

Attention mechanisms have also been widely employed in multi-input schemes. In a method aimed at lightweight 3D voxels using 2.5D segmentation, researchers proposed a Multi-Attention Context Network (MADC-Net) based on U-Net. This network utilises multiple attention blocks and Global Context Features (GCF) to capture meaningful features and performs feature fusion through a module called the Context Feature Fusion Model (CFFM) [35].

Additionally, some research has attempted to enhance input data using more advanced image processing algorithms like GBVS for more refined edge information processing. These enhanced data are combined with the original image as multi-inputs to further improve the model's segmentation performance [33].

Finally, an approach known as MDS-Net has been introduced, utilising a unique stack-based fully convolutional network structure and sliding window fusion technique. This method captures 2D and 3D contextual information by segmenting the input into multiple "stacks," each comprising several slices. These slices, processed through U-Net, yield multiple segmentation outcomes, consolidated via a specific fusion algorithm to produce the final image segmentation. This strategy effectively leverages deep learning and prior structural knowledge, showing exceptional performance, especially in handling variability between different slices. However, it also faces challenges, such as highly inconsistent slices [54].

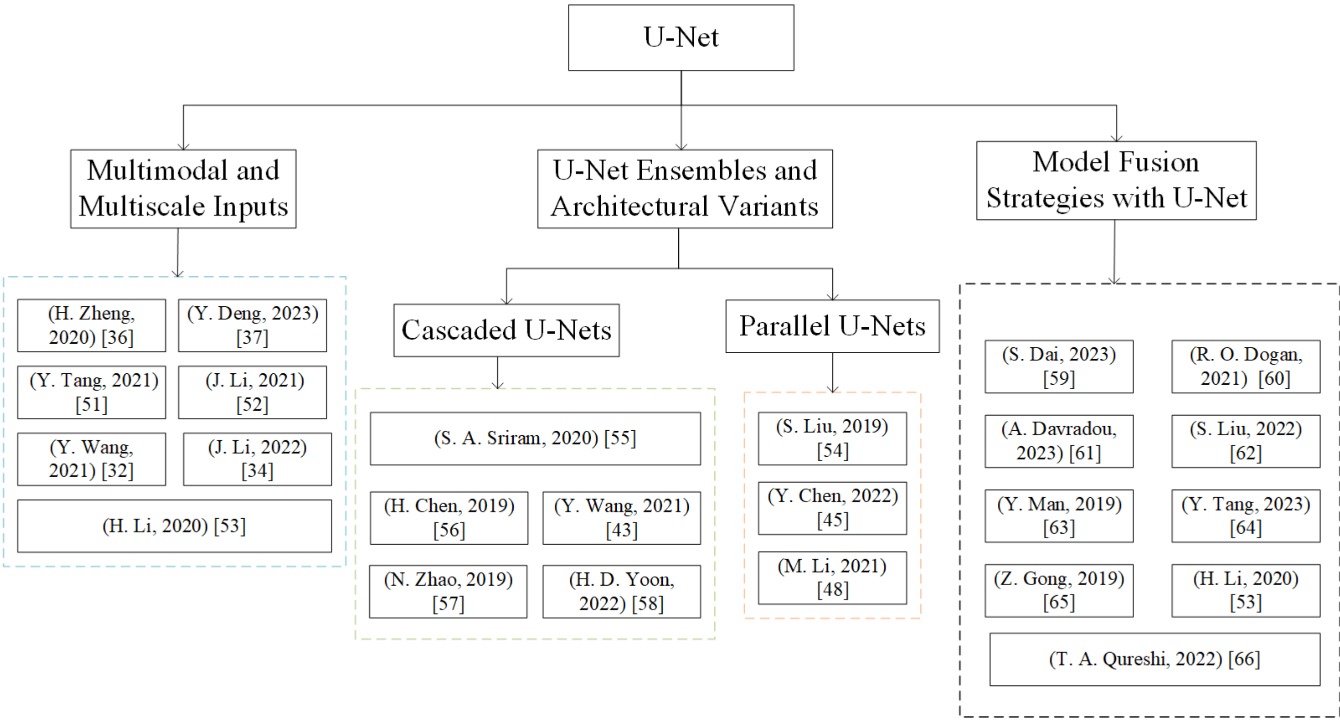


**FIGURE 4.** An overview of network-level modification to U-Net.

In summary, multi-inputs can fuse information of various types or scales and further improve segmentation accuracy by integrating different algorithms or models. These innovative fusion schemes have demonstrated immense potential and applicability in enhancing the quality of pancreatic CT image segmentation.

### B. PARALLEL U-NETS

The evolution of pancreatic CT image segmentation methods, particularly those based on U-Net and its derivatives, has been significant. The strategic implementation of parallel U-Net models, alongside the adoption of ensemble approaches, has proven effective. These methods specifically address the complex nuances involved in pancreatic segmentation, leading to enhanced outcomes.

Liu et al. employ a unique multi-U-Net architecture by deploying five Fully Convolutional Networks (FCNs) based on U-Net in parallel [55]. Each network is trained with different objective functions to optimise for various issues or characteristics. The ensemble of five different segmentation maps is then used to achieve more precise and robust segmentation outcomes. This approach allows for multi-angle information capture and effectively minimises background noise through a superpixel labeling strategy.

Upon closer analysis, the study by Chen et al. introduce a Target-Aware U-Net (tU-Net) distinguished by two innovative modules: fuzzy skip connections and target attention mechanisms [46]. These two modules run in parallel to the standard U-Net and aim to enhance the accuracy of pancreatic target segmentation. Combining fuzzy skip connections and target-aware attention mechanisms boosts the model's ability to recognise small, variable pancreatic structures.

Li et al. also introduce an innovative Dual Adversarial U-Net incorporating Generative Adversarial Networks (GANs) and pyramid pooling modules [49]. These combined technological elements improve the synergistic performance between the segment and the discriminator, allowing the model to capture key information at different scales. This multi-scale and multi-angle information fusion strategy further validates the utility of U-Net parallelisation and ensemble strategies in enhancing segmentation accuracy.

In summary, the parallelisation and ensemble of multiple U-Nets allow for more precise localisation of Regions of Interest (ROI) and multi-scale and multi-angle information fusion. Such optimisation strategies have also increased the model's sensitivity to small and complex target structures. These parallel and ensemble U-Net strategies undoubtedly offer a range of promising optimisation pathways and directions worthy of further exploration in pancreatic CT image segmentation.

### C. CASCADED U-NETS

The Cascaded U-Net architecture has attracted widespread interest in the medical image segmentation domain, particularly for pancreatic CT images. The research community has continuously put forth innovative modifications and practical applications tailored to the U-Net architecture, tackling the inherently challenging features in pancreatic CT imaging, such as inconsistent contrast and structural variability.

Cascaded U-Net demonstrates immense potential, especially in contrast-enhanced and non-enhanced CT images. For instance, Sriram et al. employ domain adaptation techniques and a multi-stage 3D U-Net architecture to narrow the gap in segmentation performance between these two types of scans by utilising synthetic non-enhanced CT images generated from intravenous contrast (IVC) scans [56]. The key to this approach lies in its success in overcoming the issue of indistinct pancreatic borders in non-contrast CT images, showcasing Cascaded U-Net's capability in handling complex medical imaging tasks.

Further research strengthens U-Net's application by combining information across dimensions and leveraging view adaptiveness. Chen et al. highlight the importance of dimensional adaptiveness through a fusion of 2D and 3D networks, employing an innovative Dimensional Adaptation Module (DAM) for better capturing 3D structural information [57]. Concurrently, the View Adaptive 3D U-Net in [44] accentuates the importance of spatial awareness and ROI localisation through view-adaptive training approaches and cascading methods.

Notably, the Two-Stage 3D U-Net architecture proposed by Zhao et al. adopts a dual-stage approach [58]. Initially, a 3D U-Net is used for rough pancreatic localisation and initial segmentation to determine the ROI; subsequently, another more refined 3D U-Net is employed for further precise segmentation, especially within the ROI. This two-stage method effectively balances speed and accuracy, showcasing the possibility of precise segmentation in complex anatomical structures, albeit requiring higher computational resources and more refined data annotation.

Moreover, facing the intrinsic uncertainty in pancreatic segmentation, the Multi-Scale Prediction Network with Pancreatic Uncertainty (MP-Net) introduced by Yoon et al. utilises a cascading strategy that considers pancreatic uncertainty, applying a specialised 2D MP-Net following preliminary 2D U-Net segmentation [59]. This method enhances segmentation accuracy and reduces variability among different patients.

In summary, through innovative network design, dimensional adaptiveness, view-adaptive training, and consideration of uncertainty, these studies demonstrate the powerful potential of Cascaded U-Net in pancreatic CT image segmentation. Despite challenges in dataset size, computational complexity, and generalizability, these innovative approaches provide a solid foundation for enhancing segmentation accuracy, handling the complexity of medical images, and propelling future research in this field.

### D. INTEGRATION WITH OTHER MODELS

With the U-Net architecture as a cornerstone, a broad spectrum of approaches is being explored to advance its capabilities. The enhancement strategies are diverse, focusing on module optimisation, applying forward-thinking training methods, and amalgamation with complex algorithms to refine segmentation accuracy.

Regarding module optimisation, Dai et al. specifically design a Trans-deformer module to address the nonlinear deformations in the pancreas [60]. In conjunction with the Scale Inter-Active Fusion (SIF) module, this module efficiently fuses local and global features. Furthermore, the study employs wavelet analysis to tackle the issue of blurred edges in the pancreas, thereby enhancing the model's capability for detail recognition.

Dogan et al. introduce a two-stage method, initially employing Mask R-CNN for the coarse localisation of the pancreas on 2D CT slices, followed by 3D U-Net to further refine these candidate regions, thereby producing segmentation results [61]. The advantages of this method include higher accuracy, as it can learn compelling features from image data through deep learning models and reduced computational cost. Although 3D networks require more computational power and memory, the study achieves good segmentation performance under lower GPU capabilities through a two-stage approach and memory processing optimisation.

In the study by Davradou et al., a streamlined two-part method is presented. Initially, YOLO v4 is employed to generate an estimated probability map for rough localisation of the pancreas. YOLO stands for "You Only Look Once" which is a fast and popular method for detecting objects in images by making just one pass. It's great for quick tasks like finding where the pancreas might be in a scan. Following this, a modified U-Net performs the segmentation on the localised regions. Post-segmentation, a morphological active contour algorithm is applied as a post-processing technique, refining the segmentation results [62]. This method effectively combines deep learning and traditional image processing techniques, proving efficient, especially in limited data scenarios or where computational resources are scarce.

Liu et al. address the challenge of obtaining a large number of annotated voxel images for training by proposing a Graph-Enhanced Partitioning Segmentation network (GEPS-Net) that incorporates an iterative uncertainty-guided pseudo-label refinement semi-supervised learning framework [63]. GEPS-Net integrates a graph enhancement module into U-Net, focusing on handling spatial relationship information. The study utilises a 3D U-Net to extract high-level features. It employs a Graph Convolutional Network (GCN) to leverage detailed information, such as spatial relationships modelled by regional adjacency graphs. Although this method enhances segmentation accuracy, it also has limitations, such as the inherent constraints of GCN and the stability of spatial relationship information.

To address challenges posed by class imbalance, background noise, and the non-rigid geometric characteristics of the pancreas, recent research has introduced innovative training strategies. A significant development among these is the use of a deformable U-Net model enhanced by Deep Q-Networks (DQN) [64]. This approach trains the model to learn context-adaptive localization strategies, enabling precise identification of the pancreas against complex backgrounds. The deformable U-Net is tailored to accommodate the geometric variations of the pancreas, utilizing geometrically deformable filters that improve feature extraction. By integrating deep reinforcement learning with deformable convolutional networks, this method not only increases sensitivity to detailed pancreatic features but also enhances the model's ability to process contextual information, as evidenced by its performance on the NIH dataset. Tang et al. introduce the Curriculum Knowledge Switching (CKS) framework to address limited samples [65]. This framework enhances the model's adaptability by incrementally introducing tasks of different difficulty levels.

As for algorithmic integration, Gong et al. propose an optimised scheme that combines the level set method [66]. This scheme enables pixel-level probability maps generated by U-Net to serve as initial conditions for the level set method, achieving finer image segmentation. On the other hand, Liu et al. employ a joint approach of semi-supervised learning and graph augmentation [63]. The study significantly improves segmentation accuracy by generating pseudo-labels and applying graph augmentation algorithms.

In addition, some innovative network architectures and fusion strategies, such as multi-stage morphological guidance methods, were introduced in [67] and the MDS-Net network fusion strategy in [54].

In summary, the ongoing adaptation and application of U-Net and its variants in pancreatic CT segmentation paved the way for integrated and multifaceted research. By applying novel integration and optimization strategies, the fundamental performance of the U-Net model is elevated and directed towards surmounting the inherently multifarious challenges present in medical image segmentation. The culmination of these advancements is represented in the dynamic landscape of current research, as systematically outlined in Table 2, which underscores the strengths and potential limitations of each network-level modification within the overarching schema of the U-Net architecture.

## VI. DISCUSSION AND ANALYSIS

### A. DATASETS

In pancreatic computed tomography (CT) image segmentation, several key public datasets have emerged as substantial drivers advancing the field. These datasets not only underpin and validate innovative algorithms but also foster integration of machine learning and deep learning in medical image analysis, thanks to diverse samples and high-quality annotations. Highlighted below are primary datasets instrumental in this field:

NIH-TCIA Pancreas-CT Dataset: Furnished by The Cancer Imaging Archive (TCIA), under the auspices of the National Institutes of Health (NIH), this dataset comprises 82 enhanced 3D CT scans of the abdomen, capturing a wide demographic spectrum across various age groups and genders and encompassing both healthy volunteers and individuals afflicted with pancreatic cancer [68]. The dataset's high resolution and demographic diversity render it an invaluable resource for inquiries into pancreatic structures.

Medical Segmentation Decathlon (MSD) Challenge: This seminal competition furnishes 421 cases of portal venous phase CT scans, each targeting pancreatic tumours [69]. The presence of imbalanced labels markedly impinges on the performance and precision of segmentation algorithms, thereby galvanising researchers to devise novel strategies adept at navigating the inherent diversity and complexity.

"Multi-Atlas Labeling Beyond the Cranial Vault" (BTCV) Challenge 2015: Stemming from multi-centre clinical investigations, this compendium features 50 superior-quality abdominal CT scans, augmenting sample heterogeneity and representativeness [70]. It is versatile, catering to various tasks, including meticulous segmentation and annotation of the pancreas. It paves the way for a more profound comprehension of intricate abdominal anatomical structures and facilitates cross-dataset algorithm validation.

International Symposium on Image Computing and Digital Medicine (ISICDM) 2018 Pancreatic Segmentation Challenge: This challenge zeroes in on the accurate diagnosis of pancreatic cancer, proffering exhaustively annotated data crafted manually. The employment of anonymous test data bolsters the transparency and reproducibility of investigative methodologies, spurring the genesis of avant-garde technologies and methodologies [66].

Clinical Proteomic Tumor Analysis Consortium-Pancreatic Ductal Adenocarcinoma (CPTAC-PDA) Dataset: Endorsed by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), this dataset is dedicated to an exhaustive exploration of pancreatic ductal adenocarcinoma (PDA), showcasing 91 meticulously annotated abdominal CT scans [71]. This wealth of information lays a robust groundwork for strides in the diagnosis and therapeutic approaches pertaining to PDA.

Upon systematic review, it has been discerned that the NIH-TCIA dataset reigns supreme in application in pertinent research, commanding a usage frequency of 68.5%. In contrast, the MSD and proprietary datasets are utilised at a rate of 12.96%, with other datasets trailing in adoption. Despite the NIH-TCIA dataset's constrained sample size, potentially impinging on the generalizability of models, its stringent standards and practicality have garnered widespread acclaim. This recognition has catalysed researchers to initiate preliminary training on proprietary datasets, subsequently capitalising on these public datasets for augmented validation and refinement, thereby fortifying the robustness and reliability of the resulting models.

### B. PERFORMANCE METRIC ANALYSIS

In medical image segmentation, relying solely on a single metric often inadequately captures a model's multifaceted

**TABLE 2.** An overview of advantages and limitations of network-level modifications to U-Net.

| Modification Type | Advantages | Limitations |
|---|---|---|
| Multi-input or Multimodal | 1. Provides a richer, more comprehensive data representation.<br>2. Improves segmentation or classification accuracy.<br>3. Especially valuable in medical image processing. | 1. Requires more data preprocessing.<br>2. Increases model complexity and computational cost.<br>3. Requires more data input, which may not always be feasible or available. |
| Parallel U-Nets | 1. Allows for multi-task or multi-view data processing.<br>2. Improves efficiency and performance.<br>3. Enhances model's generalization capability | 1. Training is more complex and computationally intensive.<br>2. Requires a large amount of labelled data. |
| Cascaded U-Nets | 1. Refines predictions incrementally.<br>2. Potential to enhance segmentation accuracy. | 1. Increases computational burden.<br>2. Errors from previous stages may be amplified.<br>3. Requires training and inference across multiple networks. |
| Integration with Other Models | 1. Combines the strengths of different networks.<br>2. Better feature extraction, deeper network structures, or more efficient training processes. | 1. Models may become overly complex and difficult to train.<br>2. Increases the challenge of finding suitable training strategies and hyperparameters. |

performance. For example, the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) are highly sensitive to small-region segmentation but may not precisely indicate overall model effectiveness. Similarly, accuracy can be misleading in imbalanced classes, inherently favoring the majority class. The Average Surface Distance (ASD) primarily evaluates average edge alignment but lacks sensitivity to general border agreement. The Hausdorff Distance (HD) is highly sensitive to outliers, where a single aberrant point can drastically distort evaluation outcomes. Additionally, while reducing false-negatives, a high Recall rate can increase false-positives, affecting Precision.

Therefore, a multi-metric assessment approach is commonly adopted to evaluate the performance of segmentation models comprehensively. The reviewed literature indicates that the DSC, Recall, and Jaccard index are the most frequently applied metrics, with usage rates of 97%, 38%, and 24%, respectively. This underscores the research community's appraisement of these metrics' overall coverage and reliability. It is important to note that despite the limitations discussed, each metric has its indispensable value and relevance in specific circumstances. Table 3 compiles the studies demonstrating the best performance within various metric categories, providing a collective and balanced perspective on model performance through the consideration of the metrics as mentioned above.

The analysis for the performance evaluation is grounded in an in-depth study of the data presented in Table 3. It is evident that a significant number of studies preferentially utilized the NIH-TCIA pancreas dataset for model training and testing, attributed to its extensive sample coverage and precise annotations. The range of performance scores for the DSC metric oscillates between 66.82% and 89.89%, with one study, [60], achieving the highest score of 89.89%±1.82. Although Recall is less widely employed compared to DSC, [60] also leads with a superior score of 91.13%±1.48 in this metric. Utilization of the Jaccard index is relatively infrequent; nonetheless, [46] scores 78.52%±4.14, demonstrating robust performance. As for processing time, significant variations are present across different studies, governed by the variation in hardware and software configurations used; notably, [30] reports a processing time of just 0.179 seconds, while [50] extends up to 1.26 minutes.

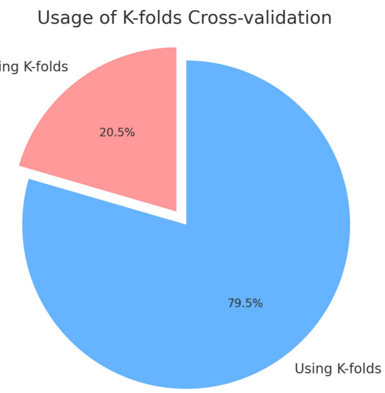


**FIGURE 5.** The percentage of K-Fold cross-validation utilization in model validation.

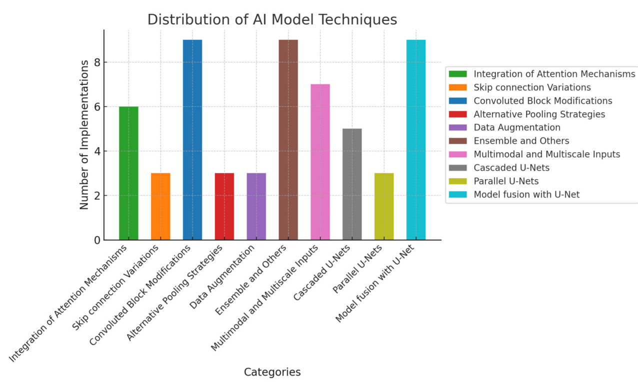


**FIGURE 6.** Distribution of categories within the total population.

### *C. EMERGING TRENDS IN PANCREATIC CT IMAGE SEGMENTATION*

The rapid advancements in deep learning and computer vision technologies have profoundly impacted the healthcare industry, particularly in tasks related to medical imaging, such as classification, detection, and segmentation. U-Net has become a pivotal biomedical image segmentation architecture within this domain, offering crucial support for diagnostic processes. In the past few years, pancreatic CT image segmentation research has undergone significant changes and developments. Below are some of the major trends:

**TABLE 3.** Performance evaluation of each study.

| Ref | Datasets | Data Split | DSC (%) | Recall (%) | Jaccard (%) | Processing Time |
|---|---|---|---|---|---|---|
| [8] | Private dataset, NIH-TCIA Pancreas, BTCV Challenge 2015 | 5 folds | 78.1 ± 8.7 | / | / | / |
| [30] | Private dataset, NIH-TCIA Pancreas | 5 folds | 76.7±13.2 | 76.2±14.5 | / | 0.179 s |
| [31] | NIH-TCIA Pancreas, MSD pancreas | 4 folds | NIH: 86.10± 3.52 MSD: 88.50± 3.70 | / | 75.55±5.42 | / |
| [32] | NIH-TCIA Pancreas | 4 folds | 86.61±3.47 | / | / | / |
| [33] | NIH-TCIA Pancreas | 4 folds | 87.4 ± 6.8 | / | / | / |
| [34] | NIH-TCIA Pancreas | 4 folds | 86.8 ± 4.1 | 88.0 ± 6.0 | 76.9 ± 6.1 | / |
| [35] | Private dataset, NIH-TCIA Pancreas, | / | 89.77 ± 7.84 | / | / | 0.19 s/2.5D slice 3.11 s/3D slice |
| [36] | NIH-TCIA Pancreas | 4 folds | 78.9 ± 8.3 | / | / | / |
| [37] | NIH-TCIA Pancreas, Private MRI dataset | 4 folds | CT: 84.37 MRI: 73.88 | / | / | / |
| [38] | NIH-TCIA Pancreas, Private dataset | 4 folds | 66.82 ± 16.44 | / | / | 32G V100, 20 seconds |
| [39] | NIH-TCIA Pancreas | 4 folds | 81.36 | / | / | / |
| [40] | NIH-TCIA Pancreas | 4 folds | 87.25 ± 3.27 | 89.97 | / | / |
| [41] | NIH-TCIA Pancreas | 4 folds | 87.25±3.27 | / | / | / |
| [42] | NIH-TCIA Pancreas, Private dataset | 4 folds | 84.2 ± 12.8 | 84.2 ± 15.6 | / | / |
| [43] | NIH-TCIA Pancreas | / | 83.31 | / | / | / |
| [44] | NIH-TCIA Pancreas | 4 folds | 86.19 | / | / | GTX1080Ti, 0.14 s |
| [45] | NIH-TCIA Pancreas | 4 folds | 86.25±4.25 | / | / | / |
| [46] | NIH-TCIA Pancreas, MSD pancreas | 4 folds | 87.91 ± 2.65 | 85.77 ± 4.61 | 78.52 ± 4.14 | / |
| [47] | MSD pancreas, CPTAC-PDA Pancreas | / | 85.54±0.04 | / | / | / |
| [48] | NIH-TCIA Pancreas, MSD pancreas | / | 75.56 | 81.03 | / | 324 slices, about 1 minute |
| [49] | NIH-TCIA Pancreas | 4 folds | 82.38± 5.46 | 82.30±7.45 | 70.39± 7.58 | / |
| [50] | NIH-TCIA Pancreas | 4 folds | 86.30±4.03 | / | 76.26±5.01 | 1.26 Min |
| [51] | NIH-TCIA Pancreas | 4 folds | 82.5 | / | / | 0.335 second/slice |
| [52] | BTCV Challenge 2015, NIH-TCIA Pancreas | 4 folds | mean 79.1 | / | / | / |
| [53] | NIH-TCIA Pancreas | 4 folds | 85.35±4.13 | / | / | total 1.12s for 3 stages |
| [54] | NIH-TCIA Pancreas | 4 folds | 85.7±4.1 | 84.8±7.5 | 75.3±6.1 | / |
| [55] | NIH-TCIA Pancreas | 4 folds | 84.10±4.91 | 85.33±8.24 | 72.86±6.89 | TITAN X, 1.99s |
| [56] | NIH-TCIA Pancreas, MSD pancreas | / | 69.9 ± 14 | / | / | / |
| [57] | NIH-TCIA Pancreas | 4 folds | 85.22± 4.07 | / | / | 0.44 min |
| [58] | NIH-TCIA Pancreas | 4 folds | 85.99 ± 4.51 | / | / | / |
| [59] | NIH-TCIA Pancreas | / | / | 82.49 (± 08.09) | / | / |
| [60] | NIH-TCIA Pancreas, MSD pancreas | 4 folds | 89.89 ± 1.82 | 91.13 ± 1.48 | / | / |
| [61] | NIH-TCIA Pancreas | 4 folds | 86.15±4.45 | 86.27±2.73 | 75.93±6.46 | no actual time |
| [62] | NIH-TCIA Pancreas, MSD pancreas | / | 67.67% ± 8.62 | / | / | / |
| [63] | NIH-TCIA Pancreas | 4 folds | 84.22 ± 5.24 | / | 73.10 ± 7.36 | 20 s |
| [64] | NIH-TCIA Pancreas | 4 folds | 86.93±4.92 | / | / | / |
| [65] | NIH-TCIA Pancreas | 4 folds | 85.42 ± 4.39 | / | / | / |
| [66] | ISICDM2018 | / | 82 | / | / | / |
| [67] | NIH-TCIA Pancreas | 4 folds | 88.53 | / | / | / |

- Evolution in Handling CT Images: With the rapid advancement of Graphics Processing Unit (GPU) technology, the methods for processing CT images are evolving from 2D to a mixture of 2D and 3D, and eventually leaning towards pure 3D processing. This trend reflects the increased computational capabilities that allow researchers to handle more complex three-dimensional data, thereby achieving more accurate image segmentation.
- Extraction and Fusion of Multi-Scale Features: Researchers are increasingly applying the extraction and fusion of multi-scale features to gain a more comprehensive understanding of images. This strategy helps the model capture information at different scales

and regions in the image, thereby improving the segmentation accuracy.

- Transformer-based U-Net Variants: Inspired by self-attention mechanisms, researchers have explored Transformer-based U-Net variants. These variants enhance the model's ability to capture global and local contextual information, albeit at the cost of increased computational burden.
- Multi-Modal Fusion: Researchers are investigating how to extract valuable features further through multi-modal fusion. This approach aids in synthesising information from different imaging modalities, enhancing model performance.
- Coarse-to-Fine Strategy: This strategy initially locates the region of interest (ROI) through a preliminary network model and then performs fine segmentation to improve model performance.
- Additionally, the analysis of Figures 5 and 6 reveals further insights: Data from Figure 5 shows that approximately 80% of the studies use K-fold cross-validation, which enhances the robustness and generalizability of the training process, ensuring that the model is not overly fitted to a particular subset of the data. According to Figure 6, other models are being combined with U-Net, and modifications are being made to the convolutional layers, including the addition of attention mechanisms and integration of residual networks. Simultaneous modifications to multiple aspects of U-Net can lead to better segmentation outcomes.

## VII. FUTURE DIRECTIONS AND CHALLENGES

Having delineated the key trends shaping the realm of pancreatic CT image segmentation, it becomes imperative to pivot our discussion toward the uncharted territories of future research and the hurdles that lie ahead. Despite the remarkable strides already made, persistent challenges and untapped avenues could further elevate the performance of U-Net and its variants in medical imaging analytics, especially when grappling with the intricate structures of the pancreas.

### A. MODEL REFINEMENT AND ROBUSTNESS

Although U-Net excels in various medical image segmentation scenarios, it has certain limitations when dealing with the pancreas, which may have diverse morphologies, varying sizes, and complex pathological conditions. Future research should focus on developing advanced models specifically designed for various pancreas morphologies and physiological states. This could include the integration of more advanced edge-detection algorithms and denoising modules to increase the model's robustness for specific tasks.

### B. DATA IMBALANCE AND ANNOTATION COSTS

In pancreatic segmentation, the target area (i.e., the pancreas) is often much smaller than the entire image, exacerbating the issue of data imbalance. To tackle this problem, future research might consider adopting novel loss functions, multitask learning, and semi-supervised learning techniques to handle data imbalance and costs more efficiently.

### C. MULTI-MODALITY AND TEMPORAL SEQUENCE ANALYSIS

Cross-modal analysis, such as integrating CT, Magnetic Resonance Imaging (MRI), and X-ray images, along with temporal sequence data, may offer more comprehensive and detailed information for future research. This improves the model's generalization capabilities and provides additional clinically relevant information.

### D. INCORPORATION OF PRIOR KNOWLEDGE AND UNCERTAINTY QUANTIFICATION

Given that existing models are largely data-driven, incorporating prior knowledge from anatomy and physiology into the models could enhance their robustness and accuracy. In addition, quantifying model uncertainty is another area deserving further study, as it can offer more reliable support for medical decision-making.

### E. ALGORITHM INTERPRETABILITY AND CLINICAL APPLICATION

Enhancing the algorithms' interpretability is crucial before U-Net and its variants can be widely adopted in clinical settings. This will strengthen healthcare professionals' trust in the models and serve as a foundation for their application in clinical environments.

### F. DEVELOPMENT OF EVALUATION METRICS

Current evaluation metrics such as the Dice coefficient and Intersection over Union (IoU) may only partially capture the complexities and clinically relevant details of segmentation tasks. Therefore, developing more clinically relevant evaluation metrics, such as specific indicators for the early diagnosis of pancreatitis or pancreatic cancer, is an important direction for future research.

These research directions and challenges highlight the need for in-depth theoretical research and empirical validation to achieve higher accuracy and reliability in future clinical applications. Through these focused methodologies and practical insights, the enormous potential and applicability of U-Net and its variants in pancreatic CT image segmentation are further demonstrated.

## VIII. CONCLUSION

This review systematically investigates the latest advancements and challenges in pancreatic CT image segmentation using U-Net and its variants. Detailed quantitative assessments reveal that ensemble learning and transformer-based U-Net variants exhibit significant advantages in handling noise and label imbalances, especially when validated on the NIH and MSD datasets. These findings highlight the potential of these technologies to improve the accuracy and efficiency of medical image processing.

Future research should focus on developing models with enhanced interpretability and visual clarity to strengthen their applicability in clinical settings. Additionally, there is an urgent need to devise effective methods for integrating multimodal data from various sources, with an emphasis on optimizing computational efficiency. Addressing the limitations of current datasets by implementing innovative semi-supervised or unsupervised learning methods is a critical step towards enhancing the practical usability of models.

The unique contribution of this paper lies not only in the systematic evaluation of existing U-Net architectures and their improvement strategies but also in thoroughly assessing their performance across multiple publicly available datasets, thereby confirming their effectiveness. These achievements not only provide a solid foundation for future research but also clearly demonstrate how technological innovations can further advance precision medicine and imaging techniques, particularly in the field of pancreatic CT image segmentation.

## REFERENCES

[1] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Jan. 2018.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Jan. 2017.

[3] B. Kenner, S. T. Chari, D. Kelsen, D. S. Klimstra, S. J. Pandol, M. Rosenthal, A. K. Rustgi, J. A. Taylor, A. Yala, N. Abul-Husn, and D. K. Andersen, "Artificial intelligence and early detection of pancreatic cancer: 2020 summative review," *Pancreas*, vol. 50, no. 3, pp. 251–279, 2021.

[4] C. Xia, X. Dong, H. Li, M. Cao, D. Sun, S. He, F. Yang, X. Yan, S. Zhang, N. Li, and W. Chen, "Cancer statistics in China and United States, 2022: Profiles, trends, and determinants," *Chin. Med. J.*, vol. 135, no. 5, pp. 584–590, Mar. 2022.

[5] L. C. Chu, S. Park, S. Kawamoto, Y. Wang, Y. Zhou, W. Shen, Z. Zhu, Y. Xia, L. Xie, and F. Liu, "Application of deep learning to pancreatic cancer detection: Lessons learned from our initial experience," *J. Amer. College Radiol.*, vol. 16, no. 9, pp. 1338–1342, Sep. 2019.

[6] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, Oct. 2015, pp. 234–241.

[7] B. Huang, H. Huang, S. Zhang, D. Zhang, Q. Shi, J. Liu, and J. Guo, "Artificial intelligence in pancreatic cancer," *Theranostics*, vol. 12, no. 16, pp. 6931–6954, 2022.

[8] T. G. W. Boers, Y. Hu, E. Gibson, D. C. Barratt, E. Bonmati, J. Krdzalic, F. van der Heijden, J. J. Hermans, and H. J. Huisman, "Interactive 3D U-Net for the segmentation of the pancreas in computed tomography scans," *Phys. Med. Biol.*, vol. 65, no. 6, Mar. 2020, Art. no. 065002.

[9] H. Kumar, S. V. DeSouza, and M. S. Petrov, "Automated pancreas segmentation from computed tomography and magnetic resonance images: A systematic review," *Comput. Methods Programs Biomed.*, vol. 178, pp. 319–328, Sep. 2019.

[10] H. Xiao, L. Li, Q. Liu, X. Zhu, and Q. Zhang, "Transformers in medical image segmentation: A review," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104791.

[11] I. Poce, J. Arsenjeva, A. Kielaite-Gulla, A. Samuilis, K. Strupas, and G. Dzemyda, "Pancreas segmentation in CT images: State of the art in clinical practice," *Baltic J. Modern Comput.*, vol. 9, no. 1, pp. 25–34, 2021.

[12] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The prisma statement," *Int. J. Surg.*, vol. 8, no. 5, pp. 336–341, 2010.

[13] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quart.*, vol. 26, no. 2, pp. 13–23, 2002.

[14] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decision Support, 4th Int. Workshop, DLMIA, 8th Int. Workshop, ML-CDS, Conjunction MICCAI*, Granada, Spain. Cham, Switzerland: Springer, Sep. 2018, pp. 3–11.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[16] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. 19th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Athens, Greece. Cham, Switzerland: Springer, Oct. 2016, pp. 424–432.

[17] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[18] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal. Int. Workshop Large-Scale Annotation Biomed. Data Expert Label Synth.*, 2016, pp. 179–187.

[19] 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, Feb. 2021.

[20] T. Zhou, Y. Dong, B. Huo, S. Liu, and Z. Ma, "U-Net and its applications in medical image segmentation: A review," *J. Image Graph.*, vol. 26, no. 9, pp. 2058–2077, 2021.

[21] J. Wang, L. Han, and D. Ran, "Architectures and applications of U-Net in medical image segmentation: A review," in *Proc. 9th Int. Symp. Syst. Secur., Saf., Rel. (ISSSR)*, Jun. 2023, pp. 84–94.

[22] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[24] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[25] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[26] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.

[27] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6450–6458.

[28] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149.

[29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[30] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, and B. Kainz, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[31] W. Li, S. Qin, F. Li, and L. Wang, "MAD-UNet: A deep U-shaped network combined with an attention mechanism for pancreas segmentation in CT images," *Med. Phys.*, vol. 48, no. 1, pp. 329–341, Jan. 2021.

[32] Y. Yan and D. Zhang, "Multi-scale U-like network with attention mechanism for automatic pancreas segmentation," *PLoS ONE*, vol. 16, no. 5, May 2021, Art. no. e0252287.

[33] Y. Wang, G. Gong, D. Kong, Q. Li, J. Dai, H. Zhang, J. Qu, X. Liu, and J. Xue, "Pancreas segmentation using a dual-input v-mesh network," *Med. Image Anal.*, vol. 69, Apr. 2021, Art. no. 101958.

[34] L. Chen and L. Wan, "CTUNet: Automatic pancreas segmentation using a channel-wise transformer and 3D U-Net," *Vis. Comput.*, vol. 39, no. 11, pp. 5229–5243, Nov. 2023.

[35] J. Li, G. Liao, W. Sun, J. Sun, T. Sheng, K. Zhu, K. M. von Deneen, and Y. Zhang, "A 2.5D semantic segmentation of the pancreas using attention guided dual context embedded U-Net," *Neurocomputing*, vol. 480, pp. 14–26, Apr. 2022.

[36] M. Nishio, S. Noguchi, and K. Fujimoto, "Automatic pancreas segmentation using coarse-scaled 2D model of deep learning: Usefulness of data augmentation and deep U-Net," *Appl. Sci.*, vol. 10, no. 10, p. 3360, May 2020.

[37] H. Zheng, Y. Chen, X. Yue, C. Ma, X. Liu, P. Yang, and J. Lu, "Deep pancreas segmentation with uncertain regions of shadowed sets," *Magn. Reson. Imag.*, vol. 68, pp. 45–52, May 2020.

[38] Y. Deng, L. Lan, L. You, K. Chen, L. Peng, W. Zhao, B. Song, Y. Wang, Z. Ji, and X. Zhou, "Automated CT pancreas segmentation for acute pancreatitis patients by combining a novel object detection approach and U-Net," *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104430.

[39] M. Li, F. Lian, C. Wang, and S. Guo, "Accurate pancreas segmentation using multi-level pyramidal pooling residual U-Net with adversarial mechanism," *BMC Med. Imag.*, vol. 21, no. 1, pp. 1–8, Dec. 2021.

[40] M. Huang, C. Huang, J. Yuan, and D. Kong, "A semiautomated deep learning approach for pancreas segmentation," *J. Healthcare Eng.*, vol. 2021, pp. 1–10, Jul. 2021.

[41] M. Huang, C. Huang, J. Yuan, and D. Kong, "Fixed-point deformable U-Net for pancreas CT segmentation," in *Proc. 3rd Int. Symp. Image Comput. Digit. Med.* New York, NY, USA: Association for Computing Machinery, Aug. 2019, pp. 283–287, doi: 10.1145/3364836.3364894.

[42] S.-H. Lim, Y. J. Kim, Y.-H. Park, D. Kim, K. G. Kim, and D.-H. Lee, "Automated pancreas segmentation and volumetry using deep neural network on computed tomography," *Sci. Rep.*, vol. 12, no. 1, p. 4075, Mar. 2022.

[43] B. Giddwani, H. Tekchandani, and S. Verma, "Deep dilated V-Net for 3D volume segmentation of pancreas in CT images," in *Proc. 7th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2020, pp. 591–596.

[44] Y. Wang, J. Zhang, H. Cui, Y. Zhang, and Y. Xia, "View adaptive learning for pancreas segmentation," *Biomed. Signal Process. Control*, vol. 66, Apr. 2021, Art. no. 102347.

[45] C. Qiu, Z. Liu, Y. Song, J. Yin, K. Han, Y. Zhu, Y. Liu, and V. S. Sheng, "RTUNet: Residual transformer UNet specifically for pancreas segmentation," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104173.

[46] Y. Chen, C. Xu, W. Ding, S. Sun, X. Yue, and H. Fujita, "Target-aware U-Net with fuzzy skip connections for refined pancreas segmentation," *Appl. Soft Comput.*, vol. 131, Dec. 2022, Art. no. 109818.

[47] K. Fang, B. He, L. Liu, H. Hu, C. Fang, X. Huang, and F. Jia, "UMRFormer-net: A three-dimensional U-shaped pancreas segmentation method based on a double-layer bridged transformer network," *Quant. Imag. Med. Surg.*, vol. 13, no. 3, pp. 1619–1630, Mar. 2023.

[48] L. Cao, J. Li, and S. Chen, "Multi-target segmentation of pancreas and pancreatic tumor based on fusion of attention mechanism," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104170.

[49] M. Li, F. Lian, and S. Guo, "Automatic pancreas segmentation using double adversarial networks with pyramidal pooling module," *IEEE Access*, vol. 9, pp. 140965–140974, 2021.

[50] C. Qiu, Y. Song, Z. Liu, J. Yin, K. Han, and Y. Liu, "CMFCUNet: Cascaded multi-scale feature calibration UNet for pancreas segmentation," *Multimedia Syst.*, vol. 29, no. 2, pp. 871–886, Apr. 2023.

[51] M. Li, F. Lian, and S. Guo, "Multi-scale selection and multi-channel fusion model for pancreas segmentation using adversarial deep convolutional nets," *J. Digit. Imag.*, vol. 35, no. 1, pp. 47–55, Feb. 2022.

[52] Y. Tang et al., "Pancreas CT segmentation by predictive phenotyping," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*, Strasbourg, France. Cham, Switzerland: Springer, Sep./Oct. 2021, pp. 25–35.

[53] J. Li, X. Lin, H. Che, H. Li, and X. Qian, "Pancreas segmentation with probabilistic map guided bi-directional recurrent UNet," *Phys. Med. Biol.*, vol. 66, no. 11, Jun. 2021, Art. no. 115010.

[54] H. Li, J. Li, X. Lin, and X. Qian, "A model-driven stack-based fully convolutional network for pancreas segmentation," in *Proc. 5th Int. Conf. Commun., Image Signal Process. (CCISP)*, 2020, pp. 288–293.

[55] S. Liu, X. Yuan, R. Hu, S. Liang, S. Feng, Y. Ai, and Y. Zhang, "Automatic pancreas segmentation via coarse location and ensemble learning," *IEEE Access*, vol. 8, pp. 2906–2914, 2020.

[56] S. A. Sriram, A. Paul, Y. Zhu, V. Sandfort, P. J. Pickhardt, and R. M. Summers, "Multilevel UNet for pancreas segmentation from non-contrast CT scans through domain adaptation," *Proc. SPIE*, vol. 11314, pp. 116–121, Mar. 2020.

[57] H. Chen, X. Wang, Y. Huang, X. Wu, Y. Yu, and L. Wang, "Harnessing 2D networks and 3D features for automated pancreas segmentation from volumetric CT images," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Shenzhen, China. Cham, Switzerland: Springer, Oct. 2019, pp. 339–347.

[58] N. Zhao, N. Tong, D. Ruan, and K. Sheng, "Fully automated pancreas segmentation with two-stage 3D convolutional neural networks," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Shenzhen, China. Cham, Switzerland: Springer, Oct. 2019, pp. 201–209.

[59] H. D. Yoon, H. Kim, and H. Hong, "Deep pancreas segmentation through quantification of pancreatic uncertainty on abdominal CT images," *Proc. SPIE*, vol. 12033, pp. 953–958, Apr. 2022.

[60] S. Dai, Y. Zhu, X. Jiang, F. Yu, J. Lin, and D. Yang, "TD-net: Trans-deformer network for automatic pancreas segmentation," *Neurocomputing*, vol. 517, pp. 279–293, Jan. 2023.

[61] R. O. Dogan, H. Dogan, C. Bayrak, and T. Kayikcioglu, "A two-phase approach using mask R-CNN and 3D U-Net for high-accuracy automatic segmentation of pancreas in CT imaging," *Comput. Methods Programs Biomed.*, vol. 207, Aug. 2021, Art. no. 106141.

[62] A. Davradou, "Detection and segmentation of pancreas using morphological snakes and deep convolutional neural networks," 2023, *arXiv:2302.06356*.

[63] S. Liu, S. Liang, X. Huang, X. Yuan, T. Zhong, and Y. Zhang, "Graph-enhanced U-Net for semi-supervised segmentation of pancreas from abdomen CT scan," *Phys. Med. Biol.*, vol. 67, no. 15, Aug. 2022, Art. no. 155017.

[64] Y. Man, Y. Huang, J. Feng, X. Li, and F. Wu, "Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net," *IEEE Trans. Med. Imag.*, vol. 38, no. 8, pp. 1971–1980, Aug. 2019.

[65] Y. Tang, K. Zhan, Z. Tian, M. Zhang, S. Wang, and X. Wen, "Curriculum knowledge switching for pancreas segmentation," 2023, *arXiv:2306.12651*.

[66] Z. Gong, Z. Zhu, G. Zhang, D. Zhao, and W. Guo, "Convolutional neural networks based level set framework for pancreas segmentation from CT images," in *Proc. 3rd Int. Symp. Image Comput. Digit. Med.*, Aug. 2019, pp. 27–30.

[67] T. A. Qureshi, C. Lynch, L. Azab, Y. Xie, S. Gaddam, S. J. Pandol, and D. Li, "Morphology-guided deep learning framework for segmentation of pancreas in computed tomography images," *J. Med. Imag.*, vol. 9, no. 2, Apr. 2022, Art. no. 024002.

[68] H. R. Roth, A. Farag, E. Turkbey, L. Lu, J. Liu, and R. M. Summers, "Data from pancreas-CT. The cancer imaging archive," *IEEE Trans. Image Process*, vol. 10, p. K9, 2016.

[69] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, and B. Menze, "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," 2019, *arXiv:1902.09063*.

[70] B. Landman, Z. Xu, J. Igelsias, M. Styner, T. Langerak, and A. Klein, "MICCAI multi-atlas labeling beyond the cranial vault–workshop and challenge," in *Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault-Workshop Challenge*, vol. 5, 2015, p. 12.

[71] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). (2018). *The Clinical Proteomic Tumor Analysis Consortium Pancreatic Ductal Adenocarcinoma Collection (CPTAC-PDA)*. [Online]. Available: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=33948258

**CHAOHUI ZHANG** (Student Member, IEEE) received the M.S. degree from Asia Pacific University of Technology and Innovation, Malaysia, in 2020. He is currently pursuing the Ph.D. degree with the School of Computer Sciences, Universiti Sains Malaysia. In 2021, he joined as a Lecturer with the Center for Technology Enhanced Learning, Guangdong Maoming Health Vocational College, China. His research interests include deep learning, computer vision, and medical image analysis.

**ANUSHA ACHUTHAN** (Member, IEEE) received the Ph.D. degree in computer vision and medical image analysis from Universiti Sains Malaysia, in 2016. She is currently a Senior Lecturer with the School of Computer Sciences, Universiti Sains Malaysia. Her research interests include the areas of computational neuroscience, knowledge guided image analysis, biomarkers analysis for neurological disorders and aging, machine learning, and computer vision. She is a member of ISMRM and IACSIT. Besides research work, she is also very passionate about online teaching and learning resources. She has developed a MOOC course on scientific writing and delivered workshops on e-learning modules.

**GALIB MUHAMMAD SHAHRIAR HIMEL** received the B.Sc. degree in computer science and engineering from the Ahsanullah University of Science and Technology (AUST), in 2016, the M.Sc. degree in computer science and engineering from United International University (UIU), in 2018, the B.Sc. degree in computing from the University of Greenwich (UoG), U.K., in 2021, the M.Sc. degree in computer science in intelligent systems from American International University-Bangladesh (AIUB), in 2022, and the M.Sc. degree in applied physics and electronics from Jahangirnagar University (JU), in 2023. He was a Researcher with Bangladesh University of Business and Technology (BUBT) and Independent University, Bangladesh (IUB). He is currently involved in several researches related to bio-medical image processing using machine learning and a Ph.D. Researcher with Universiti Sains Malaysia (USM). His research interests include artificial intelligence, machine learning, bioinformatics, bio-medical image analysis, and computer vision.