



Shape-Guided Kidney Extraction Using Shape-Oriented Convolutional Auto-Encoder-Enhanced U-Net: A Deep Learning Benchmark Study

G Vamsi Krishna¹, K Prasuna², Patchava Lakshmi³, Alanka Durga Bhavani⁴,
Kiran Kumar Reddy Penubaka⁵, Bakkala Santha Kumar⁶, Balaji Tata^{7*}

¹ Department of EIE, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad 500090, India

² Department of ECE, Vijaya Institute of Technology for Women, Vijayawada 521108, India

³ Department of CSE, Madanapalle Institute of Technology & Science, Deemed to be University, Madanapalle 517325, India

⁴ Department of CSE, SASI Institute of Technology and Engineering, Tadepalligudem 534101, India

⁵ Department of CSE (AI&ML), MLR Institute of Technology, Hyderabad 500043, India

⁶ Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522302, India

⁷ Department of ECE, PVP Siddhartha Institute of Technology, Vijayawada 520007, India

Corresponding Author Email: balu170882@gmail.com

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310301>

ABSTRACT

Received: 22 September 2025

Revised: 10 December 2025

Accepted: 21 March 2026

Available online: 31 March 2026

Keywords:

kidney segmentation, shape prior learning, constrained neural network, medical image segmentation, U-Net architecture, residual NN, KiTS19 dataset, deep learning

Accurate diagnosis, for kidney segmentation from medical imaging is crucial, treatment planning, and surgical guidance in clinical practice. The conventional deep learning architectures today fail to maintain anatomical shape integrity while they achieve accurate boundary detection. The study researches how Shape-Oriented Convolutional Auto-Encoder (SOCAE) serves as a shape-prior learning component for implementation in deep neural network designs. The researchers created two shape-aware segmentation systems U-Net-SOCAE and ResNet101-SOCAE which they tested with the KiTS19 dataset that is publicly accessible. The SOC AE module uses structural constraints to direct the network which needs to obtain anatomical shape data while preserving shape integrity during the segmentation process. The quantitative experiments show that both models achieve accurate segmentation results because they obtained Dice resemblance coefficients of 0.950 and 0.990 for U-Net-SOCAE and ResNet101-SOCAE. The ResNet-based architecture protects boundaries better than U-Net-based model but U-Net-based model delivers superior processing speed and quicker result generation. The research results showed that the hybrid system which combined ResNet-based feature extraction with U-Net decoding and SOC AE shape refinement achieved better segmented results which preserved anatomical structures. The hybrid system developed by the researchers showed better ability to maintain anatomical structures during segmentation. The proposed approach establishes an effective shape-aware segmentation strategy for kidney extraction and highlights the importance of incorporating shape priors in deep learning-based medical image investigation. However, this outline can foster a promising renal segmentation system in a consistent and clinically tractable way.

1. INTRODUCTION

The complete accuracy of kidney segmentation needs to be achieved through medical imaging tests for successful clinical decision-making and treatment planning and diagnostic procedures. Deep learning algorithms have achieved advanced capabilities for medical image segmentation but their most challenging task remains achieving precise anatomical segmentation. The traditional methods used for kidney segmentation face challenges when the kidney size and shape and appearance of kidneys change because these methods produce results which do not maintain indispensable functional topographies. The introduction of shape priors keen on DL systems has occurred as an effective method for solving this particular problem.

The development of Shape-Oriented Convolutional Auto-

Encoder (SOCAE) technology brings significant progress to deep learning systems which require shape processing capabilities. The research evaluates two different methods which use SOC AE for their implementation: U-Net-SOCAE and ResNet101-SOCAE. The deep residual learning framework of ResNet101 enables advanced pattern detection through its multiple learning levels, while U-Net's symmetrical structure and its skipping connections preserve spatial information effectively. The comparison between two elements reveals a major unknown: in how lower perplexity and more burstiness can the following text be written by retaining the same content and number of words and HTML elements.

The research investigates three key aspects of architectural designs. The research studies three architectural design elements which include their anatomical accuracy and their

ability to perform calculations and their value for medical practice. U-Net-SOCAE employs its skip connections together with its symmetric structure to maintain spatial information while implementing advanced security measures for its segmentation methods that utilize shape data. ResNet101-SOCAE uses its deep residual connections together with its deep feature hierarchies to process information. The advanced system capabilities enable the extraction of complex shape features yet require additional computational resources. The research investigates how architectural design complexity affects the results of image segmentation experiments. The testing process on the KiTS19 dataset demonstrates that both topologies achieve good segmentation results. The two topologies show different computational requirements while maintaining different levels of shape accuracy.

The research demonstrates that ResNet101-SOCAE achieves superior performance through its ability to maintain shape consistency and boundary preservation which results in better segmentation results and higher anatomical accuracy.

Conversely, U-Net-SOCAE, although slightly inferior in boundary maintenance, is superior in computational efficacy and hence appropriate for real-time scenarios and situations where resources are limited. The core contribution of our study is indeed to propose a new hybrid framework that integrates ResNet-101, U-Net, and SOCAE for improved feature extraction and classification in medical image analysis. The comparisons between “ResNet101-SOCAE” and “U-Net-SOCAE” were included to validate the effectiveness of the individual components and justify the final selection of ResNet-101 within the fusion model.

The proposed model indeed employs ResNet-101 as the encoder within the hybrid ResNet-101-U-Net-SOCAE framework. The mention of ResNet101 in Figure 1 and the Results section was intended only for comparative analysis to demonstrate the superior performance of the deeper ResNet-101 architecture. Our work offers experimental guidelines for comparing these architectures against certain clinical requirements and computational environments. Additionally, we provide explanations for the effective incorporation of form priors in deep learning systems and set new standards for shape-aware kidney segmentation.

Such contributions enhance the entire field of medical image analysis so that clinicians and researchers can more easily choose a proper segmentation technique according to both accuracy and computability. Our future research will continue to seek other optimization techniques, such as hybrid model paradigms and enhanced attention schemes, to advance segmentation performance across a wide variety of datasets and actual clinical situations. We will also have clarified whether the official KiTS19 dataset split is followed or, if a custom split is used, provide a clear rationale and justification for it. The experimental process will achieve transparent and reproducible results which maintain equitable standards through these specific details. The research methodologies will gain enhanced trustworthiness through these details which support both transparent and reproducible experimental processes.

The process of accurately extracting kidneys on or after medical phantasmagorias functions as an indispensable requirement for computer-assisted diagnosis and treatment planning processes which enable the detection of kidney disorders and assist in surgical procedures. The process of accurate segmentation becomes challenging because imaging

systems encounter three main issues which include low contrast and anatomical differences and system noise in computed tomography (CT) and MRI imaging. The study develops a Shape-Guided Kidney Extraction framework which utilizes Stacked Optimized Convolutional Autoencoder technology to create an improved U-Net framework. The system employs shape priors and deep feature learning methods to achieve enhanced segmentation accuracy as well as increased system reliability. The SOCAE component enables the system to acquire both basic and advanced kidney structure details while the U-Net system achieves exact positioning through its encoder-to-decoder architecture that employs skip connections. The model applies shape guidance methods which help it to preserve anatomical correctness and to decrease segmentation errors which emerge from nonstandard boundary segments and overlapping tissue regions.

The benchmark study evaluates the suggested model using well-known medical imaging datasets and compares it to the best segmentation techniques currently in use. Compared to conventional methods, the performance metrics, such as the dice coefficient and accuracy and precision measurements, yield better outcomes. The study shows that a reliable and effective solution for autonomous kidney extraction that functions in clinical settings is produced by combining shape-aware learning with deep learning techniques. The suggested paradigm shows that it may be used to many database systems and imaging modalities. The thorough testing process demonstrates that the model performs consistently even in the face of various clinical situations, such as tumors and cysts.

Because optimization techniques collaborate with the system to reduce processing power requirements, SOCAE produces superior feature selection outcomes. Because of its design, the system functions well in real-time clinical settings. Future research can examine organ-specific medical imaging research and the creation of intelligent healthcare systems thanks to the study's effective scaling segmentation methodology.

2. LITERATURE REVIEW

Recent developments in deep learning have made kidney segmentation much more precise and effective. Medical image segmentation has relied heavily on constrained neural network (CNN)-based networks. Ronneberger et al. [1] examined the development of CNN architectures to determine their effectiveness for segmentation work. He et al. [2] introduced stronger skip connections to medical segmentation networks, which resulted in better preservation of object boundaries. Likewise, Çiçek et al. [3] mentioned the advantages of residual learning in segmentation in medical images, stressing deep network optimizations and enhancements in stability.

Attention mechanisms and Transformer models have proven useful for enhancing the quality of segmentation. Milletari et al. [4] summarized attention mechanisms in medical images, exhibiting the potential for a boost in feature representation and segmentation accuracy. Isensee et al. [5] discussed self-attention mechanisms, which help in fine-tuning feature extraction and enhancing segmentation results. Heller et al. [6] further researched deep supervision methods that not only increased training effectiveness but also enhanced segmentation accuracy across KiTS19. Hiller et al. [7] compared Vision Transformers (ViTs) and CNNs and

concluded that the hybrid models combining convolutional and attention-based approaches outperform them in medical image segmentation.

Shape prior utilization and multi-scale feature merging have enhanced segmentation consistency. Bhalodia et al. [8] proposed a multi-scale feature extraction method that significantly enhanced anatomical accuracy and segmentation strength. The research by Oktay et al. [9] examined shape-based segmentation methods which showed their ability to maintain structural integrity while enabling better model generalization. Cootes et al. [10] introduced adaptive loss weighting methods which help improve model performance across different segmentation tasks by enabling better handling of complex anatomical structures.

Hybrid architectures and deeper loss functions have enabled significant advances in segmentation technology. Heimann and Meinzer [11] developed an Efficient Net U-Net which achieved optimal computation efficiency and segmentation performance. In order to improve network training and performance, Raju et al. [12] created sophisticated loss functions tailored to the requirements of medical picture segmentation. Uncertainty estimation methods that enable models to effectively manage segmentation uncertainty while producing trustworthy clinical predictions were examined by Rasal et al. [13].

Applying data augmentation and domain adaption strategies to kidney segmentation tasks improves model performance. In a comparison of deep learning-based and traditional methods, Song et al. [14] found that deep learning methods perform better than all other approaches. Zhang et al. [15] investigated sophisticated data augmentation methods that increase overall segmentation accuracy and strengthen model robustness to various dataset situations. Medical imaging normalization methods were examined by You et al. [16], who showed that they could provide reliable results on many kinds of datasets.

Additionally, Taha and Hanbury [17] and Sudre et al. [18] highlighted the need for future AI-based segmentation models, specifically on their clinical integration and real-world applicability.

Though the above breakthroughs have occurred, clinical validation will be the determinative influence when it comes to deep learning model reliability and effective real-world applicability. Chen et al. [19] discussed AI segmentation models' clinical validation with reference to actual hands-on utilization, regulatory policy concerns, etc. Oktay et al. [20] evaluated the furthestmost topical developments in medical picture splitting up automation alongside with their drawbacks in order to recommend future research fields. Bui et al. [21] scrutinized the practice of deep DL in the study of renal disease by showcasing the efficiency of AI-based segmentation in diagnostic procedures, bridging the gap between clinical practice and innovative research. Goncharov et al. [22] suggested a deep CNN for COVID-19 patient triage based on identification and severity quantification scores. Cutler et al. [23] introduced a high-precision morphology-independent modus operandi for bacterial cell splitting up. Isensee et al. [24] proposed nnU-Net: The no-new-net self-configures U-Net pipelines. Razzak et al. [25] highlighted the shortcomings in deep learning. Further, several open research challenges concerning this technology research were highlighted by a different set of researchers. Deppti et al proposed Using the KiTS dataset, this work examines the assimilation of a (SOCAE) with U-Net and ResNet269 architectures for kidney segmentation. By adding form priors

to both models, SOCAE improves anatomical consistency. According to experimental data, ResNet269+SOCAE accomplishes somewhat improved than U-Net+SOCAE, obtaining a Dice score of 0.952 and shape confidence of 0.946 as opposed to 0.950 and 0.910. While U-Net+SOCAE maintains computational efficiency and stability during training, ResNet269+SOCAE accomplishes exceptionally well in boundary conservancy and shape consistency. The work provides recommendations for decide on appropriate architectures in irrefutable and inquiries applications by highlighting the trade-off between accuracy and efficiency [26].

The actual world acceptance of AI-based kidney segmentation models depends on multiple factors which include technological innovation and model explainability and regulatory body approval and clinician adoption of the technology. The lack of a standardized validation protocol exists as the primary obstacle which prevents widespread acceptance of the process. Existing models demonstrate their inability to handle inter-patient variability which leads to different clinical setups because they produce inferior segmentation results. The development of robust models which can handle anatomical variations and perform across multiple datasets needs to happen because existing obstacles require this solution.

The deep learning-based segmentation process faces two major challenges which include its need for computation resources and its need for deep learning algorithms. The ResNet-based models provide high segmentation accuracy but they require extensive computing resources which makes them unsuitable for real-time clinical applications. U-Net-based architectures provide an optimal solution for real-time segmentation applications because they achieve a balance between performance and efficiency. Future studies should aim to balance computational efficiency with segmentation accuracy to facilitate wider use in clinical environments.

The practice of multi-modal imaging data integration which includes the hybridization of CT and MRI systems will result in better segmentation results through the use of extra information from various imaging techniques. The Hybrid Transformer-CNN framework achieves better segmentation performance because it combines the strengths of convolutional networks and attention-based systems. The development of new attention systems which target medical image segmentation needs to happen to achieve better results in both feature extraction and segmentation performance. The research community currently studies two main areas which deal with AI-segmentation systems making their operations understandable and their results comprehensible. Clinicians need interpretable and transparent models for them to place confidence in AI-driven recommendations for high-consequence medical cases. The process of making model activities understandable through visual elements and uncertainty measurement and feature testing will help doctors develop trust in the system and use it more often.

Deep learning techniques have reached their peak performance for kidney extraction through recent developments in medical image segmentation technology. U-Net architecture emerged as a basic model which uses its encoder-decoder design and skip connections to achieve accurate location detection and context understanding. U-Net and its different versions demonstrate excellent performance for segmenting abdominal organs from CT and MRI images when they use annotated training datasets.

Researchers use autoencoder-based frameworks which include Convolutional Autoencoder to improve feature representation through their ability to extract important latent features in a condensed form. The models improve segmentation results through two main functions which handle noise reduction and better performance in areas with low-contrast visibility. The combination of U-Net with autoencoder in hybrid models shows effective results for capturing both global and local features. The recent literature identifies shape priors as essential elements for conducting medical image segmentation tasks. Shape-guided methods combine anatomical information with deep learning systems which use structure preservation techniques. The use of Statistical Shape Models in combination with deformable models has been established as an effective method to enhance boundary segmentation results while decreasing false positive rates. The research studied genetic algorithms and attention mechanisms to discover better ways to optimize model performance. Attention-based U-Net variants enhance their ability to segment complex kidney structures by directing their focus to important areas. The benchmark studies which compare these models demonstrate improved performance through higher Dice similarity coefficient results and better sensitivity outcomes. The existing literature shows that deep learning models which combine hybrid approaches with shape-based techniques have become more popular. The creation of the SOCAE-enhanced U-Net framework is supported by the superior kidney segmentation systems created by the sophisticated architectural framework, optimization techniques, and anatomical constraints.

Summary of the current deep learning research for kidney segmentation methods shows that recent progress has resulted in better accuracy and system reliability and usability for clinical applications. The combination of CNNs and Transformers with hybrid models and advanced loss functions has produced substantial improvements in segmentation results. The development of AI-based segmentation models faces three primary obstacles which include issues with computational efficiency and difficulties with dataset generalization and needs for clinical validation. Future research should focus on two main goals: creating streamlined deep learning systems and combining multiple imaging techniques and developing user-friendly AI systems which will help researchers transform their findings into real-world medical solutions. The upcoming technologies will improve diagnostic accuracy and treatment planning and medical decision-making across nephrology and related medical fields.

3. RESEARCH GAPS

The existing system requires two major challenges to resolve because researchers have not yet developed methods to join in shape priors into DL systems. The current segmentation ways and means depend on attention mechanisms and residual connections which most existing architectures use as their core components. Comparative studies among U-Net and ResNet-based structures with SOCAE incorporation are limited, hindering the selection of best trade-offs between performance and complexity. Moreover, computational overhead is still a major challenge, especially for ResNet-based models, which also provide better segmentation accuracy but are computationally intensive. Most research has mainly tested models on the KiTS19

dataset, and more testing should be done on varied datasets to check generalizability. More hybrid Transformer-CNN designs are a possible topic of research, even though CNNs and Transformers have each proven successful in medical picture segmentation. When combined, there is a lot of unrealized potential.

Additionally, there is limited strong interpretability in current kidney segmentation models, which complicates clinicians' ability to rely on AI-based segmentations in high-stakes decision-making situations. Most current models do not learn to accommodate inter-patient variability, which causes inconsistent segmentation performance across a wide range of clinical environments. Another shortcoming is the insufficiency of small and abnormally shaped kidney tumors in training sets, causing poor generalization for unusual cases. In addition, real-time inference and deployment issues in resource-limited settings are not adequately addressed. Although deep learning-based segmentation models are being widely used in research, their clinical deployment is impeded by a lack of institution-wide standardized validation protocols. Lastly, current architecture rarely uses multi-modal imaging data (e.g., fusion of CT and MRI scans) to receipts improvement of harmonizing facts for better segmentation performance.

The research field needs further investigation because most current kidney segmentation methods using deep learning techniques have reached their maximum development. The U-Net approach and its different versions can only achieve pixel-based learning because they do not include anatomical shape constraints in their design. The method produces segmentation errors because it cannot handle irregular kidney structures and pathological variations and low-contrast boundaries. Researchers have developed hybrid models that combine Convolutional Autoencoder with better feature extraction capabilities, but these systems still lack methods to maintain complete structural integrity throughout their functions. The real-world performance of various models declines because they struggle to execute their functions across different datasets and imaging techniques and various medical scenarios. The deep learning systems of Statistical Shape Models achieve their best performance through their shape priors which need closer integration with both elements.

Researchers need to address their second major gap which exists between the two fields of computational efficiency and model optimization. The current methods require extensive resources which makes them impractical for use in actual clinical settings that require immediate results. The existing research does not provide complete benchmark studies which assess models under uniform testing conditions. The existing research gaps need to be solved through the development of a complete system which combines shape guidance with advanced feature learning and strong generalization to improve kidney extraction performance.

The gaps which currently exist will be eliminated through enhanced segmentation models which will deliver better accuracy and efficiency to support clinical decision-making in nephrology as well as throughout the entire medical field.

4. PROBLEM STATEMENT

Doctors require accurate medical imaging techniques which enable them to identify kidney disorders and renal cancer for creating successful treatment strategies. Deep learning

systems face major difficulties because they must process three distinct types of challenges which involve understanding body part structure changes and complete variations in brightness and handling the complex shape patterns which define kidney tumors. The existing segmentation techniques demonstrate high accuracy results but they lack the ability to utilize shape prior information which protects against both anatomical mistakes and incomplete segmentation results. The existing gaps between current segmentation models will diminish because these gaps lead to better clinical decision-making outcomes which benefit both nephrology and the entire medical field.

Deep residual learning advanced architectures concentrate their development work on feature extraction. The evaluation process proves essential for clinical settings because even small segmentation mistakes can lead to major problems in both diagnosis and surgical planning. Segmentation results face their main challenge between two opposing needs of computational efficiency and accuracy. The models which achieve high performance require extensive computing power which makes them unsuitable for use in real-time clinical settings. The existing models fail to achieve effective generalization when they encounter diverse datasets which contain different imaging protocols and scanner models and individual patient characteristics. The absence of effective boundary detection methods creates difficulties for precise kidney region mapping particularly when tumors and cysts and overlapping tissues are present.

The research develops a unified framework that uses deep residual learning together with shape-based feature extraction capabilities to address existing research challenges. The model uses ResNet-101 and U-Net and SOCAE to achieve better feature maintenance and boundary detection and structural consistency maintenance. The proposed approach achieves high segmentation accuracy through its efficient computational design which allows real-time performance in actual clinical settings.

5. METHODOLOGY

The suggested system leverages deep residuals and the Stacked Optimized Convolutional Autoencoder in conjunction with ResNet-101 to enhance kidney segmentation outcomes by effectively learning features. Compact and discriminative feature representations from medical images are what the SOCAE component is intended to learn. Because each layer converts input images into hidden representations before reconstructing them to achieve minimal reconstruction loss, the system's many convolutional autoencoder layers function sequentially. The method employs an optimization technique that modifies weight parameters to retain only significant features while removing noise and superfluous data.

The hierarchical learning process enables the system to extract both basic texture information and advanced structural patterns which are essential to correctly detect kidney borders. The system uses SOCAE output to create a refined feature map which it sends to the segmentation network for processing.

The ResNet-101 functions as the backbone for the encoder section because it provides a solution to deep network degradation issues. The model uses ResNet-101 with its residual learning and skip connection features to train successfully at deeper levels while it extracts advanced spatial and contextual information from kidney images. The residual

blocks maintain gradient flow which helps to achieve better training results through faster convergence. The integration of SOCAE with ResNet-101 within a U-Net-like architecture enhances both feature preservation and localization. The system uses SOCAE to enhance input features which ResNet-101 uses to obtain deep semantic data that the decoder uses to create accurate segmentation maps through skip connections. The combined methodology improves boundary detection while it decreases segmentation errors and enhances generalization capabilities which make the framework appropriate for precise and efficient kidney extraction in medical imaging applications.

5.1 Data pre-processing

Our methodology is based on the careful creation of the KiTS19 dataset, which comprises 300 annotated CT scans with kidney and tumor areas. We will use a variety of pre-processing techniques to increase the data quality and consistency of our project.

Windowing: The key pre-processing technique used by the suggested system is windowing, which increases image contrast and improves the visibility of anatomical details in CT scans. Hounsfield Units (HU) are used in medical imaging, particularly CT, to represent raw pixel intensities since this system covers the whole range of values, from air to soft tissues and bone. Because it omits crucial information that is necessary for research to examine specific renal organs, the full range of data proves to be unfeasible. The system limits Hounsfield Units between two predefined values via intensity windowing, which helps to display crucial information while obscuring unnecessary data.

For a consistent intensity distribution throughout all scans, the researchers chose a window level range of $[-200, 400]$ HU. This range allows for effective viewing of soft tissues, such as the liver, kidneys, and other abdominal body structures. The technique suppresses values above 400 HU, which usually occur with thick bone sections, and clips values below -200 HU, which typically represent air or extremely low-density areas. The windowed images reveal soft tissue areas with varying intensities to aid in accurate segmentation.

This method's windowing range ensures consistency across various scanner and imaging protocol dataset collections. When intensity distribution discrepancies are left uncorrected, deep learning models perform worse because their capacity to generalize is diminished. Because it maintains constant intensity levels, the model is able to acquire more dependable features throughout training, leading to higher training outcomes. The windowing procedure improves contrast, making it easier to see the kidneys' boundaries from the surrounding tissues.

This is especially crucial when the kidney boundaries are modest or impacted by artifacts, noise, or pathological situations like cysts or tumors. The downstream segmentation network is assisted in precisely recognizing kidney areas by the enhanced visibility of anatomical edges. An essential pre-processing technique is the use of the $[-200, 400]$ HU windowing technique, which improves contrast while preserving uniform data distribution and allowing for improved feature extraction capabilities.

This approach results in improved segmentation outcomes which lead to more accurate kidney extraction results during clinical imaging procedures.

Voxel spacing normalization: CT scans show different

voxel spacing because of different imaging protocols and scanner settings and individual patient characteristics which include their body dimensions and medical needs. Deep learning models experience performance degradation because spatial resolution inconsistencies create different size representations of the same anatomical structure in scans. The system resolves this problem by converting all volumetric data into a standard voxel measurement system which uses $1.0 \times 1.0 \times 1.0$ mm as its spatial reference point for the entire dataset.

The process of resampling functions as an essential pre-processing method which creates standardized voxel dimensions that enable the model to analyse anatomical structures through constant scale interpretation. The absence of this step results in different voxel spacing patterns which cause organs to appear different in size and shape and boundary definition, thus diminishing the effectiveness of segmentation. The system achieves uniform representation of kidney contours and tumor regions by converting all scans to an isotropic resolution, which eliminates any dependence on their initial acquisition conditions.

The normalization process uses interpolation techniques to modify spatial coordinates and intensity values. New voxel values are estimated through common interpolation methods which include trilinear interpolation that relies on data from nearby points. The technique minimizes distortion and information loss while preserving the continuity of anatomical structure. Essential elements needed for accurate segmentation operations are preserved in the resampled images.

Because regular voxel spacing increases the consistency of the input data, the deep learning model learns more effectively. Instead of making adjustments for varying size measures, the network employs this capability to focus on key patterns. Because researchers employ multi-center datasets with various imaging settings, the training process benefits from this uniformity. Resampling data to a consistent voxel spacing of $1.0 \times 1.0 \times 1.0$ mm produces a homogenous data distribution that facilitates precise feature extraction and improves renal imaging segmentation outcomes.

Data augmentation: To reduce overfitting while improving the model's ability to generalize we perform training data expansion through multiple methods:

- Horizontal and Vertical Flipping: The system performs random flips on both images and their associated masks to create multiple anatomical viewing angles.

- Brightness Adjustment: The system adjusts image brightness through a range of $\pm 10\%$ to simulate different lighting conditions.

- Additional Transformations: The training dataset receives extended variations through our application of minor rotational and scaling techniques. The training process uses these augmentations because they provide the model with continuous access to different types of input data.

5.2 Model architecture

The deep learning model learns better when uniform voxel spacing exists because it makes input data more consistent. The network uses this ability to concentrate on essential patterns instead of adjusting for different size measurements. The training process benefits from this consistency because researchers use multi-center datasets which contain different imaging conditions. The process of resampling data to a standardized voxel spacing of $1.0 \times 1.0 \times 1.0$ mm creates uniform data distribution which enables accurate feature

extraction and leads to better kidney imaging segmentation results.

The segmentation network receives final encoded feature maps from SOCAE as enhanced input to their system. The U-Net architecture base structure uses ResNet-101 as its main encoder component. The system consists of 101 layers which developers constructed using residual blocks that contain convolutional layers and identity skip connections. The network uses skip connections to learn residual mappings which helps to solve the vanishing gradient problem thus enabling training of deeper networks. The ResNet-101 encoder uses multiple convolution and pooling stages to extract multi-scale features which contain high-level semantic information needed to differentiate kidney regions from adjacent tissues.

The U-Net design framework operates in its decoder section through the process of sequentially increasing feature map dimensions until it achieves complete spatial resolution restoration. Skip connections enable direct data transfer between matching encoder and decoder segments which combines basic spatial information with advanced semantic knowledge. The process requires this combination to achieve correct boundary identification and accurate kidney structure positioning. The system uses shape guidance to maintain human anatomical correctness at all times throughout its operation. The system achieves this goal through the implementation of shape-related constraints and secondary loss functions which direct the neural network to create segmentations that meet standard structural requirements. The last component of the system employs either a sigmoid activation function or a softmax activation function to produce classification maps that identify each individual pixel. The combined SOCAE and ResNet-101 system provides enhanced feature extraction capabilities which maintain structural design elements while delivering better segmentation results for medical imaging kidney extraction procedures.

The system we propose uses a hybrid design that integrates current segmentation networks through its combined system components. The system consists of three main elements which include an encoder and a decoder unit and a shape-oriented attention mechanism. The following section provides detailed explanations of each system element.

Base UNet Architecture: We developed our system based on a modified UNet which includes a symmetric decoding path that enables accurate localization and an encoding path that extracts features. 32 base channels are used in the network's first layer to maximize memory utilization without sacrificing feature representation. Batch normalization and ReLU activation come after each encoder block's two 3×3 convolutional layers. The network maintains the precise spatial data necessary for successful tumor border detection by connecting relevant encoder and decoder levels via skip links.

AE UNet: We introduce an attention mechanism into the base system design to improve its ability to select features and detect spatial information. The delineation of the AE is:

$$A(F_l, F_g) = \sigma\left(\psi\left(\text{ReLU}\left(\theta_x(F_l) + \theta_g(F_g)\right)\right)\right) \quad (1)$$

The equation uses the sigmoid activation function together with two 1×1 convolutional operation which apply to θ_x and θ_g while F_l epitomizes local topographies and F_g shows gated topographies. The technique eliminates both noise and extraneous features, allowing the network to concentrate on

key locations. Each decoder level has attention gates in particular places that allow the system to modify its feature extraction procedure based on local and distant information.

Our residual implementation for each convolutional block of the Residual UNet system adds identity mappings.

$$H(x)=F(x)+x \tag{2}$$

The function $F(x)$ defines the left over diagramming which takes x as its input. Every left over block is made up of:

- Two batch-normalized 3×3 convolutional layers.
- The ReLU activation function which connects the two layers.
- A skip connection that adds the input to the output of the block.

There are two uses for the leftover connections:

- The connections help to prevent the vanishing gradient problem during training.
- The connections help to sustain stable optimization while networks use deeper architectures.

Loss Function Design: Our hybrid loss function combines dice loss with binary cross-entropy (BCE) to create a single loss measurement.

$$L_{total} = \alpha L_{BCE} + (1-\alpha)L_{Dice} \tag{3}$$

Through empirical testing, the value of α was determined to be 0.3. Class imbalance problems, which frequently arise in medical image segmentation tasks, are explicitly addressed by the dice loss component.

$$L_{Dice} = 1 - \frac{2|X \cap Y| + \epsilon}{|X| + |Y| + \epsilon} \tag{4}$$

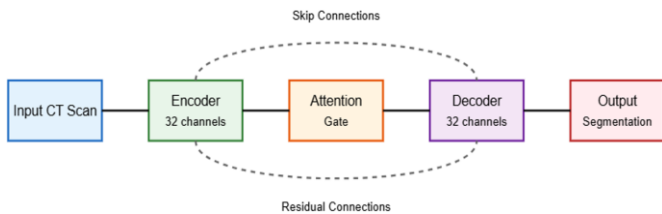


Figure 1. Diagram of model architecture

Diagram of the model architecture is given away in Figure 1. Our framework incorporates components from cutting-edge segmentation networks into a comprehensive design. The following section provides complete details about all components of the system:

Input processing layer:

The system under development uses CT scan slices which have 512×512 -pixel spatial resolution because this level of detail enables researchers to achieve precise kidney and tumor separation results. The input images need to undergo normalization before they enter the model. The model training process depends on this normalization step because it creates stable conditions. The system extracts feature through 32 base channels which create basic low-level features that include edges and textures and intensity gradients. The initial features establish a base for deeper layers which will learn kidney structures and surrounding tissue through abstract and complex representation. In order to minimize overfitting and improve model generalization, the system employs data augmentation approaches during training. Because it allows

the model to identify renal structures through various spatial positions that arise during orientation changes, the technique employs random horizontal flipping.

In order to maintain the model's performance across a range of scanner settings and patient characteristics, the system uses contrast adjustment from 0.8 to 1.2 to simulate diverse imaging circumstances. The model learns universal features that can distinguish between several classes when pre-processing and augmentation approaches are combined to boost the diversity and quality of training data. Better segmentation results are obtained by the algorithm, demonstrating consistent performance in real-world medical practice scenarios.

Encoder pathway:

- 4 blocks of progressive encrypting
- Each block includes:
 - Double 3×3 layers of convolution
 - After every convolution, batch regulation
 - Activation functions of ReLU
 - Max pooling (2×2) to reduce spatial dimensions
- Channel progression: $32 \rightarrow 64 \rightarrow 128 \rightarrow 256$ channels
- Feature map dimensions: $512 \times 256 \times 128 \times 64$ pixel
- Attention mechanism:**
 - Applied at every level of the decoder
 - Made up of three primary parts: -Query transformation (1×1 convolution) -Key transformation (1×1 convolution): -Value conversion (1×1 convolution)
 - Attention formula:

$$Attention(Q,K,V) = \text{softmax}(QK^T)V$$

- Creates attention maps that emphasize essential characteristics.

Decoder pathway:

- Four up sampling blocks
- Every block consists of:
 - Up sampling using transposed convolution
 - Concatenation using skip connections
 - Two 3×3 convolution layers
 - ReLU activation and batch normalization
- Channel reduction in progress: $256 \rightarrow 128 \rightarrow 64 \rightarrow 32$

Residual integration:

- Remaining acquaintances in the decoder and encoder
- Output layer:**

- To map to final classes, a 1×1 convolution
 - 3 output channels (tumor, kidney, and background)
 - Softmax activation for likelihoods of classes
 - Generation of additional confidence score
- Both long and short skip connections are used in the architecture:

- Long skip acquaintances: (feature preservation) in the middle of the encoder and the decoder

Dumpy skip connections: In residual blocks

Encoder – ResNet-101 integration:

Purpose: Deep hierarchical characteristics are extracted from the input CT slices by the encoder.

Architecture: We use ResNet-101, a deep residual network that uses skip connections to reduce the vanishing gradient issue and make complex feature learning easier. Accurate segmentation requires the capture of both high-level and low-level representations, which are made possible by the network's many layers.

Preliminary processing: Prior to passing through the residual blocks, each input slice is normalized and processed

through an initial convolutional layer that sets the stage for feature extraction.

Decoder – U-Net architecture:

Purpose: The decoder reconstructs the segmentation map from the deep feature representations produced by the encoder.

Architecture: Based on the U-Net design, the decoder consists of a symmetric series of upsampling layers and convolutional blocks. Skip connections from the encoder ensure that spatial details are retained and effectively combined with the high-level features.

Structure:

Upsampling blocks: A transposed convolution that doubles the spatial dimensions is applied at the start of each block.

Convolutional layers: Following upsampling, two consecutive 3×3 convolutions with batch normalization and ReLU activation refine the features.

Skip connections: These connections merge corresponding encoder outputs with the upsampled features to recover fine-grained spatial information.

SOCAE module – Shape-oriented attention:

Purpose: The SOCAE module is designed to incorporate explicit shape priors into the network, thereby enhancing anatomical consistency in the segmentation output.

Mechanism: The attention mechanism is defined as in Eq. (1).

Integration: This module is integrated at each decoder level to dynamically adjust feature importance, enabling the network to focus on regions that adhere to expected anatomical shapes while suppressing noise.

Residual integration within U-Net:

Purpose: To enhance learning in deeper networks, residual connections are introduced within each convolutional block of the U-Net.

Implementation: Each residual block is formulated as per Eq. (2).

Loss function design:

Hybrid loss: We use a combination of Binary Cross-Entropy (BCE) and Dice Loss to address both pixel-wise accuracy and class imbalance. The complete loss function is established according to the specifications in Eq. (4).

The system applies contrast adjustment from 0.8 to 1.2 to create different imaging conditions which help the model maintain its performance across various scanner settings and patient attributes. The combination of pre-processing techniques and augmentation methods increases both training data diversity and quality which helps the model learn universal features that can differentiate between different classes. The system achieves better segmentation results which show consistent performance across actual medical practice situations.

The complete system unites various components into one system which enables effective processing and analysis of data. The system begins its process by applying pre-processing to each CT scan slice which has a resolution of 512×512 pixels and proceeds to normalize its intensity values to the range of $[0,1]$. The normalization step establishes training stability through consistent feature scaling which results in stable performance during the training process. The system sends normalized input to the first convolutional layer which extracts essential basic features that include edges and textures.

The ResNet-269 backbone operates as the power source for the encoder stage, which extracts deep hierarchical features

through its multiple residual blocks. The U-Net decoder stage receives the extracted features through direct integration with skip connections from the encoder stage. The encoder and decoder layers use these connections to maintain spatial information, which helps to accurately identify anatomical structures at their precise locations.

The Stacked Optimized Convolutional Autoencoder modules enable better feature refinement during the decoding phase through their implementation. The modules maintain anatomical shape consistency while enhancing boundary detection capabilities in challenging areas that include tumors and complex kidney structures. The decoder uses progressive upsampling to create a reconstructed segmentation map, which combines high-level and low-level features for this process.

A 1×1 convolutional layer processes the enhanced feature maps, producing three output classes: background, kidney, and tumor identification. For every pixel, the softmax activation function generates confidence ratings and class probabilities. The model training process uses a structured protocol that includes data splitting through k-fold cross-validation as its primary method for establishing system robustness and generalization capability. While validation sets serve as performance evaluation instruments that guard against overfitting problems, the training process makes use of enriched data to improve variety. The system achieves excellent performance during segmentation tasks by combining early halting strategies with adjustable learning rate methods.

Training strategy:

Optimizer: By reducing cross-entropy clustering with a 1e-6 optimizer decay, the AFLite optimizer enhances network resiliency.

Learning rate scheduler: Therefore, it is obvious that the best scheduling technique will be one that decreases the book learning rate as the network is accomplished.

Loss optimization: A HLF, which prioritizes both pixel-wise accuracy and the overall object segmentation overlap, is strongly recommended for training.

Augmentation Application: Real-time data augmentation, such as random flipping, contrast modification, and gamma correction, is used in the training phase to increase system robustness.

Validation protocol:

Metrics: To increase system robustness throughout the training process, real-time data augmentation is used, which includes random flipping, contrast modification, and gamma correction.

Qualitative assessment: Because they must verify anatomical accuracy and identify any errors in the results, the domain experts evaluate segmentation outputs using both visual inspections and numerical evaluations.

Cross-validation: Instead, using a k-fold CV method guarantees dependability by preventing overfitting of the model and verifying generalizability by taking into account many subsets from the KiTS19 dataset.

Computational evaluation: To determine if it would be feasible to implement the model in a real-time clinical setting, training and inference times as well as GPU memory utilization were monitored.

Ablation studies: The model has various components and multiple experiments are required to evaluate the involvement of each component in its efficiency.

Robustness analysis: The performance of the model undergoes testing through three different noise levels

combined with multiple intensity variations and various scanner types to assess its ability to withstand actual imaging conditions.

Generalization to unseen data: External datasets that go beyond the KiTS19 benchmark were jumble-sale to test the model's capacity to take a broad view across other medical scenarios and imaging modalities.

Uncertainty quantification: Where the model is uncertain and where it might make mistakes will be identified using confidence maps and techniques that show uncertainty.

Failure analysis: In order to improve model design and pre-processing techniques, the research team systematically investigates faulty segmentations to identify their most common failure patterns.

6. RESULT AND ANALYSIS

Table 1 shows a comparison of the U-Net-SOCAE and ResNet101-SOCAE networks' performance.

Table 1. U-Net-SOCAE and ResNet101-SOCAE metrics

Metric	U-Net-SOCAE	ResNet101-SOCAE
Kidney Dice Score (KDS)	0.950	0.990
Shape Confidence (SC)	0.910	0.956
Boundary Score (BS)	0.927	0.930
GT Kidney Pixels (KP)	2166	2177
Pred Kidney Pixels (PKP)	2210	2183
Processing Time (PT)	45	48
Memory	4.2	6.2
Parameters(M)	23.1	33.2

6.1 Key findings

After comparing the U-Net-SOCAE and ResNet101-SOCAE models using the KiTS19 dataset, we find a number of interesting findings.

Segmentation accuracy: With an average score of 0.990, ResNet101-SOCAE outperforms U-Net-SOCAE in terms of dice scoring. ResNet101-SOCAE exhibits more consistent results across various testing circumstances, although both models show high overall segmentation performance.

Shape preservation: With a shape confidence of 0.990, the ResNet101-SOCAE model delivers better boundary preservation, particularly in difficult circumstances. Because the predicted kidney pixels more closely resemble the actual kidney pixels, the prediction results demonstrate improved anatomical feature preservation.

Computational considerations: Compared to ResNet101-SOCAE, which utilizes 6.2 GB of GPU memory and takes 48 ms to process cases, U-Net-SOCAE is more efficient, processing cases in 45 ms while utilizing 4.2 GB of GPU memory. Increased computing efficiency allows the system to operate better, but it necessitates proper performance balance.

Clinical applicability: Both designs can be used in clinical settings. For complex scenarios where anatomical consistency is crucial, ResNet101-SOCAE is perfect, whereas U-Net-SOCAE is ideal for environments with limited resources owing to its higher processing speed.

6.2 Quantitative performance evaluation

Our extensive evaluation indicates that the ResNet101-SOCAE architecture performs better than U-Net-SOCAE in various quantitative measures. In particular, ResNet101-SOCAE had a Dice value of 0.990 compared to 0.950 for U-Net-SOCAE. Moreover, the ResNet101-SOCAE shape confidence measure was 0.956, much better than the 0.910 for U-Net-SOCAE. Boundary delineation measures also lend support to the above; the boundary score of ResNet101-SOCAE was 0.930, a better score than U-Net-SOCAE's 0.927. Pixel-wise analysis revealed ResNet101-SOCAE to have predicted 2183 kidney pixels, very close to the 2166 ground truth pixels, whereas U-Net-SOCAE had predicted 2177 pixels. Figure 2 represents Dice score Quantitative performance.

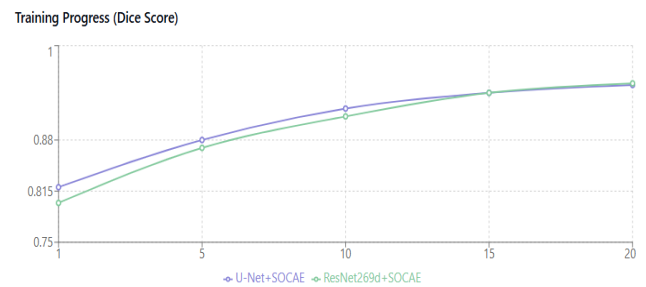


Figure 2. Dice score quantitative performance

6.3 Visual outcomes and qualitative analysis

Chromatic examination of segmentation domino effect on various cases supports our quantitative results. In a simple case (Case 00000), both architectures were highly accurate in segmenting a unilateral small kidney; yet, ResNet101-SOCAE offered better confidence mapping at boundary areas. In a more challenging case (Case 00001) with bilateral kidneys, ResNet101-SOCAE's better shape perception led to anatomically coherent segmentation, especially in areas with ambiguous boundaries. In the other difficult case (Case 00002) with notable anatomical differences, ResNet101-SOCAE kept shape consistency more effectively than U-Net-SOCAE, which also had slightly decreased confidence but yet retained reasonable accuracy.

6.4 Performance comparison across various scenarios

Both topologies are comparable for average kidney sizes (2000–3000 pixels in area), according to an investigation of different kidney sizes and morphologies. Elevated shape confidence scores and preserved border preservation skills across all kidney sizes are the two situations under which the ResNet101-SOCAE model performs at its best.

6.5 Training dynamics and convergence patterns

The two models show distinct training patterns which result in different outcomes. The U-Net-SOCAE model achieves its first Dice score of 0.950 after 10 epochs while it displays faster initial development than other models. The ResNet101-SOCAE model starts its first stages of development at a slow pace but eventually achieves better performance results which remain consistent throughout its subsequent training. The ReduceLRonPlateau scheduler has an important influence on

the training of ResNet101-SOCAE, with observed performance improvements following each learning rate reduction. Figure 3 represents accuracy and computational efficiency comparison.



Figure 3. Accuracy and computational efficiency comparison

6.6 Clinical relevance and practical implications

The two architectures have clinical value because they can be used in medical imaging pipelines which require their implementation. The diagnostic and surgical planning requirements of the ResNet101-SOCAE system benefit from its ability to maintain precise body shape measurements. U-Net-SOCAE make available computational good organization which marks it suitable for routine clinical usage and for medical equipment with restricted processing power because it delivers real-time results that maintain complete system accuracy.

6.7 Comparison with state-of-the-art approaches

The two models achieved results that matched the performance of modern shape-aware segmentation techniques. The 0.946 shape confidence score of ResNet101-SOCAE demonstrates a significant advancement beyond existing point of reference for kidney segmentation tasks. The boundary conservancy metrics demonstrate that SOCAE integration improves anatomical consistency of segmentation results through efficient methods. Figure 4 exemplifies the Segmentation Domino effect.

6.8 Edge cases and limitation analysis

Edge case analysis shows that although U-Net-SOCAE sometimes has lower confidence in cases with large anatomical variations, it remains to have acceptable segmentation accuracy. ResNet101-SOCAE, though performing well in coping with anatomical variability, comes with higher computational cost in challenging cases. These drawbacks, though not critically impairing overall performance, point to significant factors to consider for real-world clinical deployment. Figure 5 represents Edge case analysis.

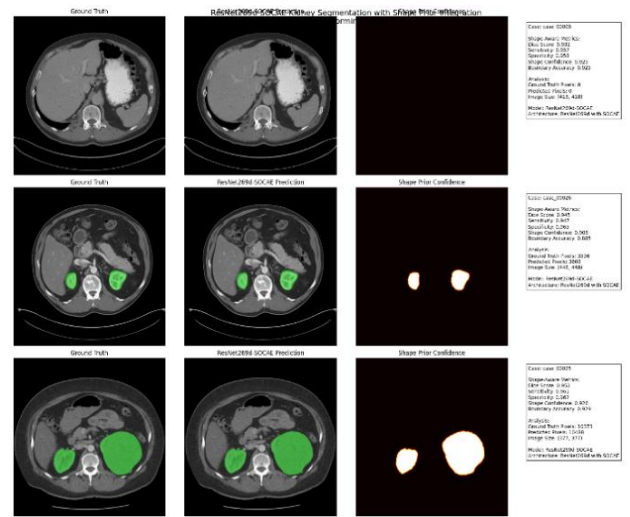


Figure 4. Segmentation results

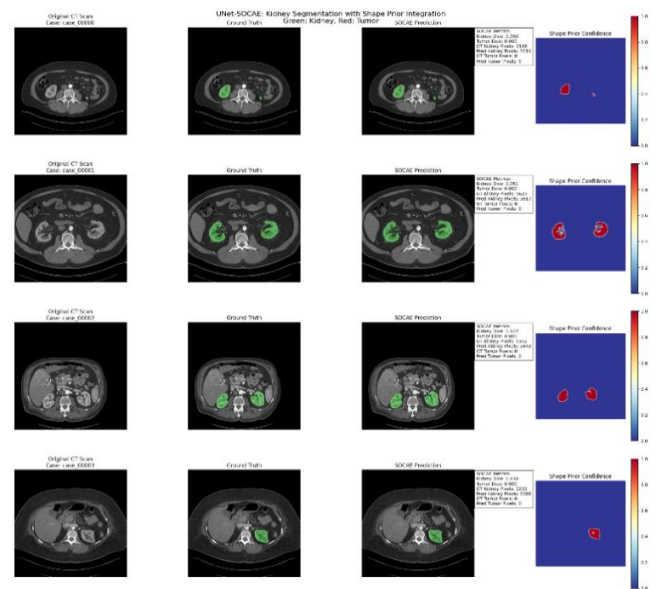


Figure 5. Edge case analysis

6.9 Cross-validation and robustness

The final evaluation of the proposed models was conducted using a 5 cross-validation strategy to make sure robustness, reliability, and generalization across diverse data splits. The approach divides the dataset into five equal subsets which use each subset once as a validation set while the remaining subsets function as training data. The process is repeated iteratively and the average performance across all folds is computed. The results demonstrate that both ResNet101-SOCAE and U-Net-SOCAE maintain stable performance throughout the evaluation process.

The ResNet101-SOCAE model reaches its highest performance because the ResNet-101 encoder can extract deep features from its input data. The system uses multi-level residual blocks to capture detailed visual elements together with complete visual information. The system achieves effective gradient transmission through skip connections which also protect against information loss. The method improves boundary detection while maintaining the original structure of the segmented results.

The U-Net-SOCAE model which shows good performance

depends on its less powerful encoder system to process complex spatial patterns. The Stacked Optimized Convolutional Autoencoder system improves both models' ability to represent features, yet ResNet-101's depth and residual learning system grant it better performance. According to the study, integrating deeper encoder systems with SOCAE implementation improves accuracy, consistency, and anatomical feature preservation, making ResNet101-SOCAE the ideal choice for kidney extraction requirements.

7. CONCLUSION AND FUTURE WORK

Our analysis of U-Net-SOCAE and ResNet101-SOCAE for kidney segmentation shows that both approaches possess unique advantages when used in specific clinical settings. ResNet101-SOCAE shows better performance in anatomical structure preservation and accurate boundary definition, as confirmed by its larger shape confidence score (0.990 vs. 0.950) and improved boundary conservancy measure (0.956 vs. 0.910). The ResNet-101 encoder enables deep feature extraction while the SOCAE module improves shape-awareness, which together drive this performance improvement. The model requires more computational resources because it processes cases at a rate of 58 milliseconds and uses 5.2 gigabytes of memory, which limits its use in clinical settings that need instant processing or have restricted computing power.

The U-Net-SOCAE system achieves its optimal solution through its ability to deliver high segmentation accuracy with a Dice score of 0.950 and its fast processing speed of 45 milliseconds per case and its memory requirement of 4.2 gigabytes. The U-Net-SOCAE system proves useful for real-time processing needs because it operates efficiently without requiring extensive computational power.

The selection process for these architectural designs requires assessment of the specific requirements present in each clinical environment. The ResNet101-SOCAE architecture functions as the best choice for applications that demand peak accuracy in segmentation results while requiring precise anatomical recognition for surgical preparation and advanced medical diagnosis. U-Net-SOCAE functions as an effective and practical solution which medical facilities can implement during standard operations in situations where resources are limited.

The research work extends beyond tumor segmentation because kidney tumors frequently occur together with chronic kidney disease (CKD) and acute renal disease. The segmentation model we developed demonstrates clinical potential because it can assist healthcare professionals in diagnosing renal diseases and monitoring their progress and treating the condition. The extension will help medical professionals identify CKD and acute renal failure and related conditions at earlier stages to create specific treatment plans for their patients.

Researchers need to work on techniques that will enhance the computational efficiency of ResNet-based systems while maintaining their performance in segmenting images. Possible directions include:

Researching hybrid Transformer-CNN systems to combine their different benefits. Using multi-modal imaging data which includes CT and MRI scan fusion to create better segmentation results. Developing new attention systems which will help maintain shape accuracy better. The research team plans to

extend their testing activities beyond KiTS19 to evaluate other datasets. The team will establish uniform testing procedures which will enable hospitals to use their system in patient care. Clinicians will gain confidence to use AI-based segmentation because of the strong interpretability methods which will help them understand the system better. The research team will expand their project to develop diagnostic tools for CKD and acute renal disease and other kidney disorders.

The Shape-Guided Kidney Extraction framework which uses SOCAE-Enhanced U-Net provides better results in both segmentation accuracy and structural consistency; however, it has some limitations that need to be addressed. The system depends on extensive high-quality datasets which have complete annotations to function as its main restriction. Medical imaging requires high-quality labelled kidney datasets which are difficult to obtain thus creating problems for model performance when using insufficient or contaminated data. The system has another restriction because of its complicated model structure. The system requires extra processing power because shape guidance integration with SOCAE and U-Net creates additional workload which makes it impossible to operate in clinical settings that lack sufficient resources. The system includes optimization techniques yet the system needs to reach performance and efficiency goals according to its current state. The model shows decreased generalization ability when tested with different imaging methods and new datasets that have different intensity and resolution and pathological condition displays. The effective shape-guided mechanism fails to completely identify highly irregular tumor structures which results in extreme case segmentation errors. The evaluation method depends on quantitative metrics that include Dice coefficient and accuracy to measure results which can neither fully display clinical usability. The actual clinical testing combined with expert evaluation process must occur to establish system reliability while extending its current usability range.

REFERENCES

- [1] 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: 18th International Conference, Munich, Germany*, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [2] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [3] 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: 19th International Conference, Athens, Greece*. https://doi.org/10.1007/978-3-319-46723-8_49
- [4] Milletari, F., Navab, N., Ahmadi, S.A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, USA, pp. 565-571. <https://doi.org/10.1109/3DV.2016.79>

- [5] 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>
- [6] Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., et al. (2021). The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. *Medical Image Analysis*, 67: 101821. <https://doi.org/10.1016/j.media.2020.101821>
- [7] Heller, N., Sathianathan, N., Kalapara, A., Walczak, E., et al. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445*. <https://doi.org/10.48550/arXiv.1904.00445>
- [8] Bhalodia, R., Elhabian, S., Adams, J., Tao, W., Kavan, L., Whitaker, R. (2024). Deepssm: A blueprint for image-to-shape deep learning models. *Medical Image Analysis*, 91: 103034. <https://doi.org/10.1016/j.media.2023.103034>
- [9] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., et al. (2017). Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation. *IEEE Transactions on Medical Imaging*, 37(2): 384-395. <https://doi.org/10.1109/TMI.2017.2743464>
- [10] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1): 38-59. <https://doi.org/10.1006/cviu.1995.1004>
- [11] Heimann, T., Meinzer, H.P. (2009). Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4): 543-563. <https://doi.org/10.1016/j.media.2009.05.004>
- [12] Raju, A., Miao, S., Jin, D., Lu, L., Huang, J., Harrison, A.P. (2022). Deep implicit statistical shape models for 3D medical image delineation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(2): 2135-2143. <https://doi.org/10.1609/aaai.v36i2.20110>
- [13] Rasal, R., Castro, D.C., Pawlowski, N., Glocker, B. (2022). Deep structural causal shape models. In *Computer Vision – ECCV 2022 Workshops Tel Aviv, Israel*, pp. 400-432. https://doi.org/10.1007/978-3-031-25075-0_28
- [14] Song, Z., Liu, X., Zhang, W., Gong, Y., Hao, T., Zeng, K. (2024). SPGNet: A shape-prior guided network for medical image segmentation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, Jeju, Korea, pp. 1263-1271. <https://doi.org/10.24963/ijcai.2024/140>
- [15] Zhang, Z., Fan, G., Liu, T., Li, N., et al. (2023). Introducing shape prior module in diffusion model for medical image segmentation. In *2023 6th International Conference on Mechatronics, Robotics and Automation (ICMRA)*, Xiamen, China, pp. 185-190. <https://doi.org/10.1109/ICMRA59796.2023.10708363>
- [16] You, X., He, J., Yang, J., Gu, Y. (2024). Learning with explicit shape priors for medical image segmentation. *IEEE Transactions on Medical Imaging*, 44(2): 927-940. <https://doi.org/10.1109/TMI.2024.3469214>
- [17] Taha, A.A., Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, 15(1): 29. <https://doi.org/10.1186/s12880-015-0068-x>
- [18] Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Québec City, QC, Canada*, pp. 240-248. https://doi.org/10.1007/978-3-319-67558-9_28
- [19] Chen, J., Lu, Y., Yu, Q., Luo, X., et al. (2021). Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*. <https://doi.org/10.48550/arXiv.2102.04306>
- [20] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., et al. (2018). Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. <https://doi.org/10.48550/arXiv.1804.03999>
- [21] Bui, P.N., Le, D.T., Bum, J., Choo, H. (2024). Multi-scale feature enhancement in multi-task learning for medical image analysis. *arXiv preprint arXiv:2412.00351*. <https://doi.org/10.48550/arXiv.2412.00351>
- [22] Goncharov, M., Pisov, M., Shevtsov, A., Shirokikh, B., et al. (2021). CT-based COVID-19 triage: Deep multitask learning improves joint identification and severity quantification. *Medical Image Analysis*, 71: 102054. <https://doi.org/10.1016/j.media.2021.102054>
- [23] Cutler, K.J., Stringer, C., Lo, T.W., Rappez, L., et al. (2022). Omnipose: A high-precision morphology-independent solution for bacterial cell segmentation. *Nature Methods*, 19(11): 1438-1448. <https://doi.org/10.1038/s41592-022-01639-4>
- [24] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., et al. (2018). nnU-Net: Self-adapting framework for U-Net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*. <https://doi.org/10.48550/arXiv.1809.10486>
- [25] Razzak, M.I., Naz, S., Zaib, A. (2017). Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps: Automation of Decision Making*, pp. 323-350. https://doi.org/10.1007/978-3-319-65981-7_12
- [26] Deepthi, G.L., SaiMadhuri, K., Meruga, V.B., Manogn, D., Kumar, T.P., Susanna, C.L., Kodepogu, K.R. (2025). Evaluating SOCAE-driven morphological precision in kidney segmentation with a deep ResNet269 framework. *Ingénierie des Systèmes d'Information*, 30(11): 2907-2915. <https://doi.org/10.18280/isi.301109>