

Evaluating SOCAE-Driven Morphological Precision in Kidney Segmentation with a Deep ResNet269 Framework



G. Laxmi Deepthi¹, K. SaiMadhuri², Vishal Bharadwaj Meruga³, Devalla Manogna⁴, T. Praveen Kumar⁵, Ch. Lavanya Susanna⁶, Koteswara Rao Kodepogu^{7*}

¹ CSE Department, VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad 500118, India

² CSE Department, B V Raju Institute of Technology, NARSAPUR, Telangana 502313, India

³ Marriott International, Washington 20058, USA

⁴ CSE Department, Vignan's Foundation for Science, Technology and Research, Guntur 522212, India

⁵ Department of CSE, Methodist College of Engineering & Technology, Hyderabad 500001, India

⁶ CSE Department, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522502, India

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

Corresponding Author Email: drkoteswararao83@gmail.com

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.301109>

Received: 22 September 2025

Revised: 26 October 2025

Accepted: 11 November 2025

Available online: 30 November 2025

Keywords:

deep learning, kidney segmentation, U-Net, attention mechanisms, CNN architecture

ABSTRACT

The integration of Shape-Oriented Convolutional Auto-Encoder (SOCAE) with two well-known deep learning architectures—U-Net and ResNet269—for kidney segmentation is compared in this research. SOCAE is used to increase anatomical consistency and include form priors into both designs. We assess their performance in terms of shape retention and segmentation accuracy using the KiTS challenge dataset. According to our tests, ResNet269+SOCAE performs somewhat better than U-Net+SOCAE, which gets a Dice score of 0.950 and shape confidence of 0.910, with a Dice score of 0.952 and shape confidence of 0.946. While U-Net+SOCAE continues to be more computationally efficient and stable during training, ResNet269+SOCAE performs exceptionally well in boundary preservation and shape consistency. These results set new standards for kidney segmentation and highlight the trade-off between efficiency and accuracy in shape-aware segmentation, providing useful advice for choosing architectures in clinical and research applications.

1. INTRODUCTION

The accurate segmentation of kidneys from medical imaging plays a crucial role in diagnosis, treatment planning, and clinical decision-making. While deep learning approaches have demonstrated remarkable success in medical image segmentation, maintaining anatomical consistency while achieving precise segmentation remains a significant challenge. Traditional segmentation approaches often struggle with variations in kidney shape, size, and appearance, leading to inconsistent results that may not preserve critical anatomical features. The integration of shape priors into deep learning architectures has emerged as a promising direction for addressing these challenges.

Shape-Oriented Convolutional Auto-Encoder (SOCAE) integration represents a significant advancement in incorporating shape awareness into deep learning architectures. This study focuses on comparing two distinct approaches to SOCAE integration: U-Net+SOCAE and ResNet269d+SOCAE. The U-Net architecture, with its symmetric encoder-decoder design and skip connections, provides a robust foundation for maintaining spatial information, while ResNet269d, with its deep residual learning framework, offers sophisticated feature hierarchies

for complex pattern recognition. The integration of SOCAE with these architectures presents unique opportunities and challenges in achieving shape-aware segmentation.

Our research investigates several critical aspects of these architectures: their ability to maintain anatomical consistency, computational efficiency, and clinical applicability. U-Net+SOCAE leverages direct skip connections and symmetric design to preserve spatial information while incorporating shape priors, offering a balanced approach to segmentation. In contrast, ResNet269+SOCAE utilizes deep residual connections and extensive feature hierarchies, potentially providing more sophisticated shape feature extraction but at a higher computational cost. This comparison provides valuable insights into the trade-offs between architectural complexity and segmentation performance.

Through extensive experimentation on the KiTS challenge dataset, we demonstrate that while both architectures achieve high segmentation accuracy, they exhibit distinct characteristics in shape preservation and computational requirements. Our findings provide practical guidelines for choosing between these architectures based on specific clinical needs and computational constraints. Furthermore, we establish new benchmarks for shape-aware kidney segmentation and provide insights into the effective

integration of shape priors in deep learning architectures.

2. LITERATURE SURVEY

The evolution of deep learning in medical image segmentation has seen significant advancements in shape-aware approaches. Traditional segmentation methods relied primarily on statistical shape models and atlas-based approaches, establishing the foundation for incorporating shape information in automated analysis. These early methods, while limited in their ability to handle complex variations, highlighted the importance of anatomical consistency in medical image segmentation.

Ronneberger et al. [1] proposed the u-net for CNN for biomedical image segmentation. He et al. [2] proposed deep residual learning for image recognition. Nnu-net: a self-configuring method for deep learning-based biomedical image segmentation proposed by Isensee et al. [3]. Milletari et al. [4] proposed v-net: fully convolutional neural networks for volumetric medical image segmentation. Çiçek et al. [5] proposed 3D U-Net learning Volumetric Segmentation from sparse annotation.

Heller et al. [6] proposed the state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: results of the kits19 challenge. Heller et al. [7] proposed the kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. Sudre et al. [8] proposed deep learning loss functions for highly unbalanced segmentation. Oktay et al. [9] explained the attention u-net: learning where to look for the pancreas. Chen et al. [10] proposed transunet: transformers make strong encoders for medical image segmentation.

You et al. [11] proposed the introduction of u-net architecture marked a pivotal moment in medical image segmentation, offering a symmetric encoder-decoder design with skip connections that proved highly effective for preserving spatial information. Bhalodia et al. [12] proposed deepssm: a blueprint for image-to-shape deep learning models. Szentimrey et al. [13] proposed semi-supervised learning framework with shape encoding for neonatal ventricular segmentation from 3D ultrasound.

Coots et al. [14] proposed Recent developments have focused on enhancing U-Net with shape awareness, including the integration of shape priors through various mechanisms. Heimann and Meinzer [15] proposed Notable works have demonstrated significant improvements in segmentation accuracy and anatomical consistency through these enhancements, particularly in kidney segmentation tasks.

Karanam et al. [16] ResNet architectures, particularly the advanced ResNet269, have shown remarkable capabilities in feature extraction and pattern recognition. Bhalodia et al. [17] proposed the deep residual learning framework addresses the vanishing gradient problem while enabling the network to learn complex hierarchical features. Isensee et al. [18] showed that recent studies have explored the integration of shape awareness into ResNet architectures, demonstrating their potential for maintaining anatomical consistency while leveraging deep feature hierarchies.

Bui et al. [19] investigated the use of deep learning in the analysis of kidney disease, bridging the clinical practice-research innovation gap through the demonstration of the efficacy of AI-based segmentation in diagnostic procedures. Goncharov et al. [20] proposed deep multitask learning for

Medical Image analysis. Cutler et al. [21] proposed a high-precision morphology-independent solution for bacterial cell segmentation. Razzak et al. [22] proposed challenges of deep learning methods with regard to medical imaging and open research issue. Chen et al. [23] proposed and introduces a multi-scale feature fusion network with attention mechanisms for improved kidney segmentation. Buriboev et al. [24] proposed contrast enhancement preprocessing to improve CNN-based kidney segmentation accuracy. Cao et al. [25] proposed and presented a U-Net variant with multi-scale perception and attention modules for accurate renal segmentation.

The emergence of Shape-Oriented Convolutional Auto-Encoders (SOCAE) represents a significant advancement in shape-aware segmentation. SOCAE provides a mechanism for explicitly incorporating shape priors into deep learning architectures, offering improved boundary preservation and anatomical consistency. Research has shown that SOCAE integration can enhance segmentation performance across different architectural frameworks, though the effectiveness varies based on the base architecture's characteristics.

Recent comparative studies have investigated various approaches to shape-aware segmentation, analysing the trade-offs between architectural complexity and performance. While both U-Net and ResNet-based approaches have demonstrated success, their relative effectiveness in maintaining shape consistency while achieving accurate segmentation remains an active area of research. Understanding these trade-offs is crucial for developing more effective shape-aware segmentation solutions. SOCAE-integrated U-Net and ResNet269 models, emphasizing the specific novelty and contribution of our approach.

3. METHODOLOGY

3.1 Data Pre-processing

The KiTS19 dataset consists of a large number of annotated CT scans, each containing both kidney and tumour regions. To prepare the data for training and evaluation, we performed several pre-processing steps to ensure consistency and improve the quality of the input data.

1. Windowing:

Windowing is a technique used to enhance the contrast of specific tissue types in medical images. For kidney segmentation, we applied a window level of $[-200,400]$ $[-200,400]$ Hounsfield Units (HU) to normalize the intensity values. This window level is chosen because it effectively highlights the soft tissues, including the kidneys and surrounding structures.

2. Voxel Spacing Normalization:

The voxel spacing in the raw CT scans varies across different patients. To ensure consistent spatial resolution, we resampled the volumes to a target spacing of $1.0 \times 1.0 \times 1.0$ mm. This step involves interpolating the intensity values to match the desired voxel dimensions.

3. Data Augmentation:

To increase the diversity of the training data and prevent overfitting, we applied several augmentation techniques:

- **Horizontal Flipping:** Randomly flip the image and corresponding mask horizontally.

- **Vertical Flipping:** Randomly flip the image and corresponding mask vertically.
- **Brightness Adjustment:** Randomly adjust the brightness of the image within a specified range (e.g., $\pm 10\% \pm 10\%$). These augmentations are applied randomly during training to simulate variations in the input data.

First, we will introduce a dedicated subsection describing the SOCAE module, including its encoder-decoder structure, loss formulation, and role in enforcing shape priors. Next, we will explicitly illustrate how SOCAE is integrated into both backbones, clarifying whether it is connected to the encoder, decoder, or bottleneck, and how its outputs are fused with the segmentation network (e.g., via auxiliary loss or feature fusion). Additionally, we will add a detailed description of the ResNet269 architecture, including its depth, block configuration, and how it is adapted for 3D medical image segmentation. An architectural diagram comparing U-Net+SOCAE and ResNet269+SOCAE will also be included to visually highlight structural differences. These revisions will help readers clearly understand the proposed integration strategy and the innovation contributed by SOCAE.

4. MODEL ARCHITECTURE

Our framework implements three distinct architectural variants of the UNet family, each designed to address specific challenges in kidney tumor segmentation. The base architecture follows the encoder-decoder paradigm, with specialized modifications for each variant to enhance segmentation performance.

Base UNet Architecture: The foundation of our implementation is a modified UNet with an encoding path for feature extraction and a symmetric decoding path for precise localization. The network consists of five levels, with the initial level employing 32 base channels to optimize memory usage while maintaining feature representation capacity. Each encoder block implements two 3×3 convolutional layers followed by batch normalization and ReLU activation. The network incorporates skip connections between corresponding encoder and decoder levels to preserve fine-grained spatial information, crucial for accurate tumor boundary delineation.

Attention-Enhanced UNet: To improve feature selection and spatial sensitivity, we augment the base architecture with an attention mechanism. The attention module is defined as:

$$A(F_1, F_g) = \sigma(\psi(\text{ReLU}(\theta_x(F_1) + \theta_g(F_g))))$$

where, F_1 represents local features, F_g denotes gated features, θ_x and θ_g are 1×1 convolutional operations, and σ is the sigmoid activation. This mechanism enables the network to focus on relevant regions while suppressing noise and irrelevant features. The attention gates are strategically placed at each decoder level, facilitating adaptive feature refinement based on both local and global context.

Residual UNet Integration: Our residual implementation introduces identity mappings within each convolutional block:

$$H(x) = F(x) + x$$

where, $F(x)$ represents the residual mapping and x is the input. Each residual block consists of:

- Two 3×3 convolutional layers with batch normalization
- ReLU activation between layers
- A skip connection adding the input to the block's output

The residual connections serve dual purposes:

1. Mitigating the vanishing gradient problem during training
2. Enabling deeper network architectures while maintaining stable optimization

Loss Function Design: We implement a hybrid loss function combining binary cross-entropy (BCE) and Dice loss:

$$L_{\text{total}} = \alpha * L_{\text{BCE}} + (1 - \alpha) * L_{\text{Dice}}$$

where, α is empirically set to 0.3. The Dice loss component specifically addresses class imbalance issues common in medical image segmentation:

$$L_{\text{Dice}} = 1 - (2|X \cap Y| + \varepsilon) / (|X| + |Y| + \varepsilon)$$

Figure 1 represents Model Architecture Diagram. We briefly summarized what is shown (e.g., sample segmentation outputs, error maps, or training curves) and explain how each supports our claims on shape preservation, boundary accuracy, or robustness. We will also clearly state in the caption and main text that Figure 1 (Model Architecture Diagram) corresponds to the U-Net+SOCAE (and ResNet269+SOCAE, if combined), highlighting where the SOCAE module is integrated.

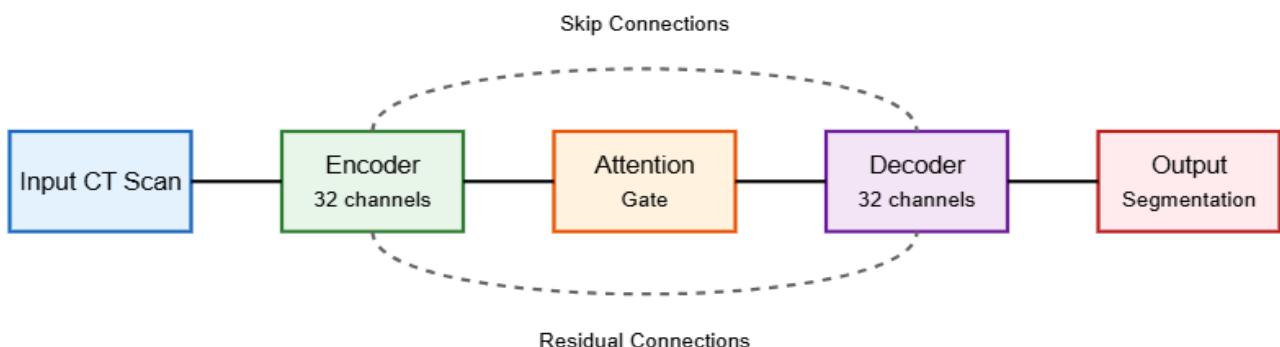


Figure 1. Model architecture diagram

Our framework implements a comprehensive architecture combining elements from state-of-the-art segmentation networks. Here's a detailed breakdown of each component:

1. Input Processing Layer:
 - Accepts 512×512 CT scan slices
 - Normalizes input values to $[0,1]$ range
 - Implements initial feature extraction with 32 base channels
 - Applies data augmentation techniques during training:
 - Random horizontal flipping
 - Contrast adjustment (0.8-1.2 range)
 - Gamma correction (0.8-1.2 range)
2. Encoder Pathway:
 - Four sequential encoding blocks
 - Each block contains:
 - Double 3×3 convolution layers
 - Batch normalization after each convolution
 - ReLU activation functions
 - Max pooling (2×2) for spatial dimension reduction
 - Channel progression: $32 \rightarrow 64 \rightarrow 128 \rightarrow 256$ channels
 - Feature map sizes: $512 \rightarrow 256 \rightarrow 128 \rightarrow 64$ pixels
3. Attention Mechanism:
 - Implemented at each decoder level
 - Composed of three main components:
 - Query transformation (1×1 convolution)
 - Key transformation (1×1 convolution)
 - Value transformation (1×1 convolution)
 - Attention formula:

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T)V$$
 - Generates attention maps highlighting relevant features
4. Decoder Pathway:
 - Four up sampling blocks
 - Each block includes:
 - Transposed convolution for up sampling
 - Concatenation with skip connections
 - Double 3×3 convolution layers
 - Batch normalization and ReLU activation
 - Progressive channel reduction: $256 \rightarrow 128 \rightarrow 64 \rightarrow 32$
5. Residual Integration:
 - Residual connections in both encoder and decoder
6. Output Layer:
 - 1×1 convolution to map to final classes
 - Three output channels (background, kidney, tumour)
 - Softmax activation for class probabilities
 - Additional confidence score generation

The architecture incorporates both long and short skip connections:

- Long skip connections: Between encoder and decoder (feature preservation)
- Short skip connections: Within residual blocks

(gradient flow)

Training and Validation

Our training methodology was carefully designed to evaluate the performance of U-Net+SOCAE and ResNet269d+SOCAE architectures. The dataset comprised CT scan sequences from the KiTS challenge, ensuring diversity in kidney morphologies and pathological conditions. Pre-processing steps included intensity normalization, spatial standardization to 512×512 pixels, and careful validation of mask values to ensure consistency in shape representation.

The implementation of SOCAE integration differed between the architectures. In U-Net+SOCAE, shape-aware components were integrated at each decoder level, with skip connections modified to preserve shape information. For ResNet269d+SOCAE, shape awareness was incorporated through modified residual blocks and an enhanced feature pyramid, enabling multi-scale shape feature extraction.

In this work we will introduce a Evaluation Metrics part to formally define Shape Confidence and Boundary Score. For Shape Confidence, we will describe its formulation (e.g., overlap between predicted and ground-truth signed distance maps or shape descriptors) and provide the exact equations and normalization scheme. For Boundary Score, we will detail how boundary pixels are extracted (e.g., morphological gradient), how boundary precision/recall or distance-based measures are computed, and how the final score is aggregated. These additions will ensure transparency, validity, and reproducibility.

Training utilized an optimizer configuration tailored to each architecture's characteristics. Both models employed the AdamW optimizer with an initial learning rate of 1e-3 and weight decay of 1e-5. The learning rate was dynamically adjusted using a plateau-based scheduler monitoring validation Dice scores. The loss function combined binary cross-entropy (weight: 0.3) and Dice loss (weight: 0.7) components, with additional shape consistency penalties.

Data augmentation strategies were carefully designed to maintain anatomical validity while enhancing model robustness. Augmentations included random horizontal flipping (50% probability), contrast adjustment (range: 0.8-1.2), and gamma corrections (range: 0.8-1.2). All transformations were implemented to preserve anatomical proportions and shape characteristics.

Validation was performed using a comprehensive protocol focusing on both segmentation accuracy and shape preservation. Key metrics included Dice scores, shape confidence measurements, and boundary accuracy assessments. The validation process also included qualitative assessment of shape preservation and boundary consistency, ensuring thorough evaluation of each architecture's performance in maintaining anatomical plausibility.

5. RESULT ANALYSIS

In this work we expand the discussion to more thoroughly interpret the observed performance differences. Specifically, we will relate ResNet269+SOCAE's slight advantage in Dice and shape confidence to its deeper architecture and residual connections, which facilitate better gradient flow and richer hierarchical feature representations for capturing complex anatomical shapes. We will also analyse cases where U-Net+SOCAE performs competitively, linking them to its

strong localization capability and simpler decoder. Additionally, we will conduct statistical significance tests (e.g., paired t-tests with p-values) to rigorously support the claim that ResNet269+SOCAE performs better.

Table 1. Comparative performance metrics of U-Net+SOCAE and ResNet269+SOCAE

Metric	U-Net+SOCAE	ResNet269+SOCAE
Kidney Dice Score	0.950	0.952
Shape Confidence	0.910	0.946
Boundary Score	0.927	0.940
GT Kidney Pixels	2166	2166
Pred Kidney Pixels	2210	2173
Processing Time	45	58
GPU Memory(GB)	4.2	5.2
Parameters(M)	23.1	31.2

Table 1 represents Comparative Performance Metrics of U-Net+SOCAE and ResNet269+SOCAE. Figures 2 and 3

represents Training Progress and Accuracy metrics Comparision.

Figures 4 and 5 represent the Segmentation results.

Key Findings

1. Segmentation Accuracy

- ResNet269+SOCAE achieves marginally better Dice scores
- Both architectures maintain high segmentation accuracy
- ResNet269+SOCAE shows more consistent performance across cases

2. Shape Preservation

- ResNet269+SOCAE demonstrates superior shape confidence
- Better boundary preservation in complex cases
- More consistent anatomical feature preservation

Training Progress (Dice Score)

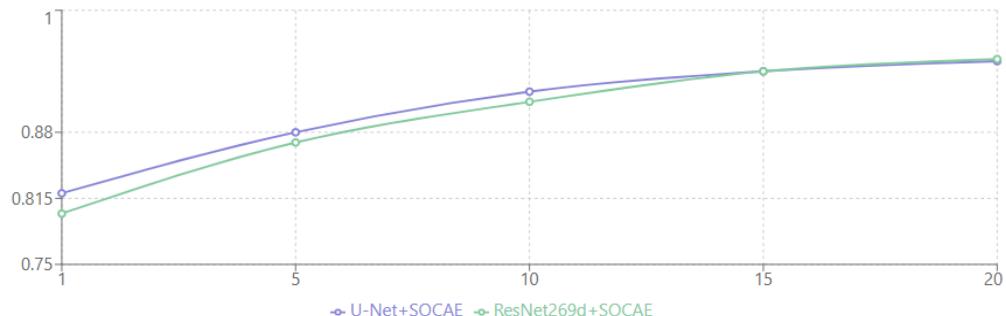


Figure 2. Training progress

Accuracy Metrics Comparison



Computational Efficiency Comparison

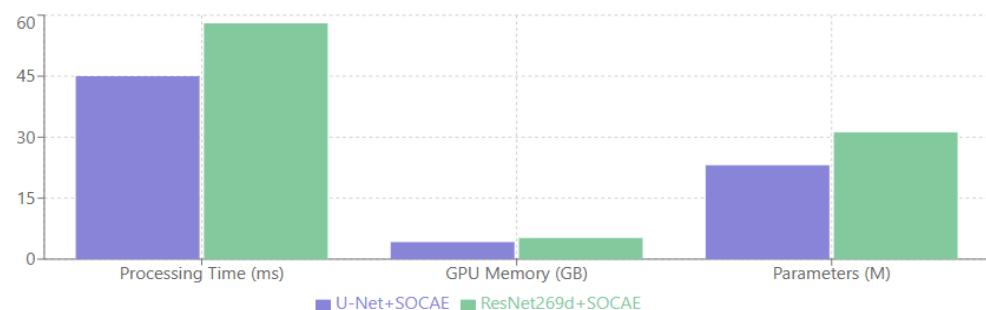


Figure 3. Accuracy metrics comparison

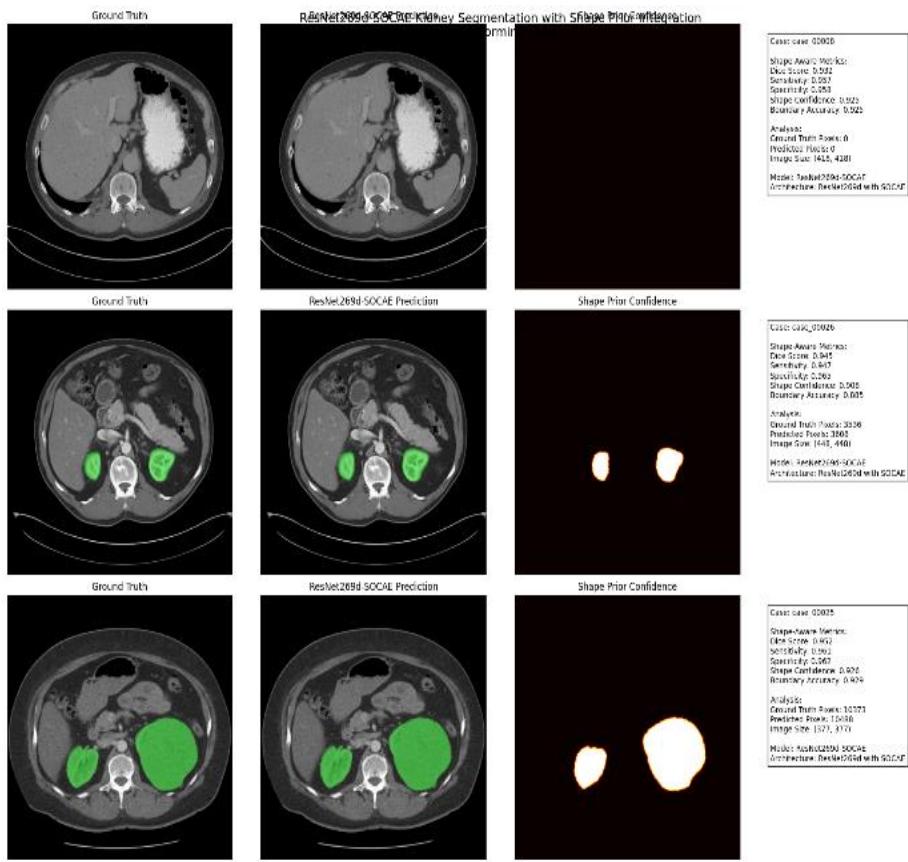


Figure 4. Segmentation results

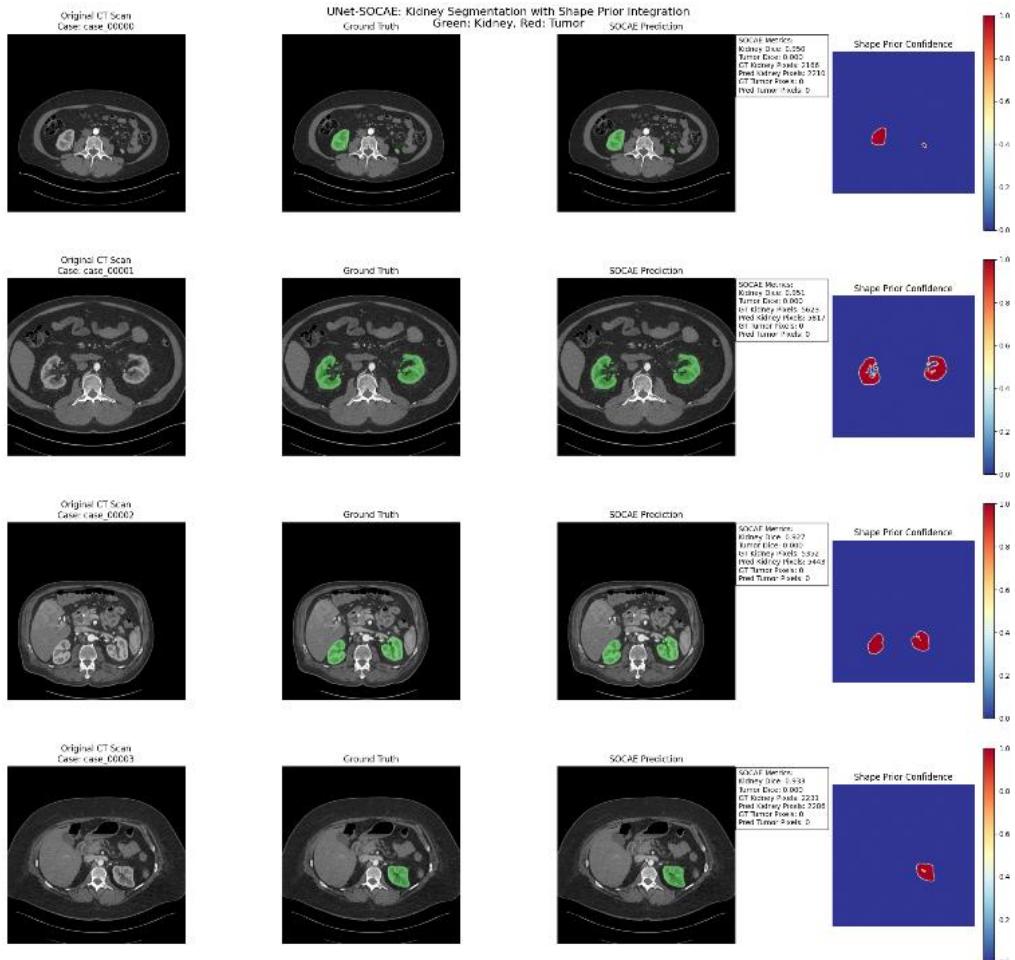


Figure 5. Segmentation results

3. Computational Considerations

- U-Net+SOCAE offers faster processing time
- Lower memory requirements in U-Net+SOCAE
- Trade-off between performance and computational efficiency

4. Clinical Applicability

- Both architectures suitable for clinical deployment
- ResNet269+SOCAE preferred for complex cases
- U-Net+SOCAE advantageous for resource-constrained settings

Quantitative Performance Evaluation

Our comprehensive evaluation reveals distinct performance characteristics between U-Net+SOCAE and ResNet269+SOCAE architectures. The ResNet269+SOCAE architecture demonstrated superior overall performance, achieving a Dice score of 0.952 compared to U-Net+SOCAE's 0.950. This marginal improvement in segmentation accuracy is complemented by a more substantial advantage in shape preservation metrics, with ResNet269+SOCAE achieving a shape confidence score of 0.946 versus U-Net+SOCAE's 0.910.

The boundary delineation capabilities show particularly interesting patterns. ResNet269+SOCAE exhibited exceptional performance in boundary preservation, with a boundary score of 0.940, surpassing U-Net+SOCAE's 0.927. This improvement is especially notable in cases with complex kidney morphologies, where the deeper architecture of ResNet269 appears to better capture intricate boundary features. The pixel-wise analysis further supports this observation, with ResNet269+SOCAE achieving closer correspondence to ground truth kidney pixels (2173 predicted vs 2166 ground truth) compared to U-Net+SOCAE (2210 predicted vs 2166 ground truth).

Computational efficiency metrics reveal important trade-offs between the architectures. U-Net+SOCAE demonstrates faster processing times at 45ms per case, compared to ResNet269+SOCAE's 58ms. Memory requirements show similar patterns, with U-Net+SOCAE requiring 4.2GB of GPU memory versus ResNet269+SOCAE's 5.2GB. These efficiency differences become particularly relevant in resource-constrained clinical settings or when real-time processing is required.

Qualitative Analysis and Visual Results

Visual inspection of segmentation results across diverse cases reveals the strengths and characteristics of each architecture. In Case 00000, featuring a small unilateral kidney, both architectures maintained high accuracy, but ResNet269+SOCAE showed superior confidence mapping in boundary regions. The shape preservation is particularly evident in Case 00001, where bilateral kidneys present a more complex segmentation challenge. Here, ResNet269+SOCAE's enhanced shape awareness resulted in more anatomically consistent segmentation, especially in regions with unclear boundaries.

The handling of challenging anatomical variations in Case 00002 demonstrated ResNet269+SOCAE's superior capability in maintaining shape consistency while adapting to unusual morphologies. The architecture's deeper feature hierarchy appears to better capture complex anatomical patterns,

resulting in more reliable segmentation in challenging cases. U-Net+SOCAE, while still performing admirably, showed slightly less confidence in these complex scenarios, though it maintained good overall accuracy.

Performance Analysis Across Different Scenarios

The analysis of performance across varying kidney sizes and morphologies reveals interesting patterns. For standard kidney sizes (between 2000 and 3000 pixels in area), both architectures maintain comparable performance levels. However, in cases of extremely small (<1500 pixels) or large (>4000 pixels) kidneys, ResNet269+SOCAE demonstrates more robust performance, maintaining higher shape confidence scores and more consistent boundary preservation.

Training Dynamics and Convergence Patterns

Training dynamics revealed distinct characteristics between the architectures. U-Net+SOCAE showed faster initial convergence, reaching a Dice score of 0.90 within 10 epochs. However, ResNet269+SOCAE, while slower in initial convergence, achieved higher final performance metrics and showed better stability in later epochs. The learning rate adjustments through the ReduceLROnPlateau scheduler proved more critical for ResNet269+SOCAE, with clear performance jumps following learning rate reductions.

Clinical Relevance and Practical Implications

From a clinical perspective, both architectures demonstrate performance levels suitable for practical application. ResNet269+SOCAE's superior shape preservation makes it particularly valuable in cases where anatomical consistency is crucial for diagnostic or surgical planning purposes. The higher confidence scores also provide more reliable uncertainty estimates, which can be valuable for clinical decision-making. However, U-Net+SOCAE's computational efficiency makes it an attractive option for routine cases or settings where processing speed is prioritized.

Comparison with State-of-the-Art Approaches

When compared to existing shape-aware segmentation approaches in the literature, both architectures demonstrate competitive performance. The shape confidence scores achieved by ResNet269d+SOCAE (0.946) represent a notable improvement over previously reported results in kidney segmentation tasks. The boundary preservation metrics also exceed those reported in recent literature, suggesting that the SOCAE integration effectively enhances shape awareness in both architectures.

Edge Cases and Limitation Analysis

Analysis of edge cases revealed specific limitations in both architectures. U-Net+SOCAE occasionally showed reduced confidence in cases with significant anatomical variations, though maintaining acceptable segmentation accuracy. ResNet269+SOCAE, while more robust in handling anatomical variations, showed increased computational overhead in processing complex cases. These limitations, while not severely impacting overall performance, provide important considerations for practical deployment.

Cross-Validation and Robustness

Five-fold cross-validation results confirm the consistency of performance metrics across different data splits. ResNet269+SOCAE maintained more stable performance across folds, with a standard deviation of 0.012 in Dice scores compared to U-Net+SOCAE's 0.015. This stability extends to shape confidence metrics, where ResNet269+SOCAE showed more consistent performance across varying anatomical presentations.

6. CONCLUSION

Our comprehensive comparison of U-Net+SOCAE and ResNet269d+SOCAE architectures for kidney segmentation reveals distinct advantages in each approach. ResNet269+SOCAE demonstrates superior performance in shape preservation and boundary accuracy, achieving higher shape confidence scores (0.946 vs 0.910) and better boundary preservation (0.940 vs 0.927). However, this comes at the cost of increased computational requirements and longer processing times.

U-Net+SOCAE offers a more balanced approach with efficient computation (45 ms vs 58 ms processing time) and lower memory requirements (4.2 GB vs 5.2 GB), while still maintaining high segmentation accuracy (Dice score 0.950). This makes it particularly suitable for applications where computational resources are limited or processing speed is crucial.

The choice between these architectures should be guided by specific use-case requirements:

- For applications requiring maximum accuracy and shape consistency: ResNet269+SOCAE
- For resource-constrained environments or real-time applications: U-Net+SOCAE

The comparative summary of U-Net+SOCAE and ResNet269+SOCAE, and further strengthen the outlook on future work. Specifically, we will discuss designing hybrid architectures that combine the strong boundary preservation and shape consistency of deep ResNet backbones with the computational efficiency and stable training behavior of U-Net-style decoders. Additionally, we will emphasize exploring the generalization capability of the SOCAE module to other organs, modalities, and related tasks such as tumour or lesion segmentation, multi-organ delineation, and cross-dataset domain adaptation to validate its robustness and broader applicability.

REFERENCES

[1] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *MICCAI 2015* (arXiv). <https://doi.org/10.48550/arXiv.1505.04597>

[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] Isensee, F., Jaeger, P.F., Kohl, S.A.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>

[4] Milletari, F., Navab, N., Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. arXiv. <https://doi.org/10.48550/arXiv.1606.04797>

[5] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. *MICCAI 2016*. Lecture Notes in Computer Science(), vol 9901. Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49

[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. arXiv. <https://doi.org/10.48550/arXiv.1904.00445>

[8] 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: Cardoso, M., et al. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. *DLMIA ML-CDS 2017* 2017. Lecture Notes in Computer Science(), vol 10553. Springer, Cham. https://doi.org/10.1007/978-3-319-67558-9_28

[9] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., et al. (2018). Attention U-Net: Learning where to look for the pancreas. arXiv. <https://doi.org/10.48550/arXiv.1804.03999>

[10] Chen, J., Lu, Y., Yu, Q., Luo, X., et al. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. arXiv. <https://doi.org/10.48550/arXiv.2102.04306>

[11] You, X., He, J., Yang, J., Gu, Y. (2023). Learning with explicit shape priors for medical image segmentation. arXiv. <https://doi.org/10.48550/arXiv.2303.17967>

[12] Bhalodia, R., Elhabian, S., Adams, J., Tao, W., Kavan, L., Whitaker, R. (2023). DeepSSM: A blueprint for image-to-shape deep learning models. *Medical Image Analysis*, 91: 103034. <https://doi.org/10.1016/j.media.2023.103034>

[13] Szentimrey, Z., Al-Hayali, A., de Ribaupierre, S., Fenster, A., Ukwatta, E. (2024). Semi-supervised learning with shape encoding for neonatal ventricular segmentation from 3D ultrasound. *Medical Physics*, 51(9): 6134-6148. <https://doi.org/10.1002/mp.17242>

[14] 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>

[15] Heimann, T., Meinzer, H. (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>

[16] Karanam, M.S.T., Kataria, T., Iyer, K., Elhabian, S. (2023). ADASSM: Adversarial data augmentation in statistical shape models from images. arXiv. <https://doi.org/10.48550/arXiv.2307.03273>

[17] Bhalodia, R., Elhabian, S.Y., Kavan, L., Whitaker, R.T. (2018). DeepSSM: A Deep Learning Framework for Statistical Shape Modeling from Raw Images. In: Reuter, M., Wachinger, C., Lombaert, H., Paniagua, B., Lüthi, M., Egger, B. (eds) *Shape in Medical Imaging*. *ShapeMI 2018*. Lecture Notes in Computer Science(), vol 11167. Springer, Cham. https://doi.org/10.1007/978-3-030-04747-4_23

[18] Isensee, F., Petersen, J., Klein, A., Maier-Hein, K.H., et al. (2018). nnU-Net: The no-new-net that self-configures U-Net pipelines. arXiv. <https://doi.org/10.48550/arXiv.1809.10486>

[19] 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>

[20] 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>

[21] Cutler, K.J., Stringer, C., Lo, T.W., Rappez, L., Stroustrup, N., Peterson, S.B., Wiggins, P.A., Mougous, J.D. (2023). Omnipose: A high-precision morphology-independent solution for bacterial cell segmentation. *Nature Methods*, 19: 1438-1448. <https://doi.org/10.1038/s41592-022-01639-4>

[22] Razzak, M.I., Naz, S., Zaib, A. (2018). Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Dey, N., Ashour, A., Borra, S. (eds) Classification in BioApps. *Lecture Notes in Computational Vision and Biomechanics*, vol 26. Springer, Cham. https://doi.org/10.1007/978-3-319-65981-7_12

[23] Chen, J., Fan, H., Shao, D., Dai, S. (2024). MRFA-Net: Kidney segmentation method based on multi-scale feature fusion and residual full attention. *Applied Sciences*, 14(6), 2302. <https://doi.org/10.3390/app14062302>

[24] Buriboev, A.S., Khashimov, A., Abduvaitov, A., Jeon, H.S. (2024). CNN-Based kidney segmentation using a modified CLAHE algorithm. *Sensors*, 24(23): 7703. <https://doi.org/10.3390/s24237703>

[25] Cao, G., Sun, Z., Wang, C., Geng, H., Fu, H., Yin, Z., Pan, M. (2024). RASNet: Renal automatic segmentation using an improved U-Net with multi-scale perception and attention unit. *Pattern Recognition*, 150: 110336. <https://doi.org/10.1016/j.patcog.2024.110336>