Enhanced U-Net for Spleen Segmentation in CT Scans: Integrating Multi-Slice Context and Grad-CAM Interpretability

Sowad Rahman*, Khaled Saifullah Karim†, Aalavi Mahin Khan‡, Dr. Md. Ashraful Alam § *Department of Computer Science, BRAC University, Dhaka, Bangladesh
†Department of Computer Science, BRAC University, Dhaka, Bangladesh
§Department of Computer Science, BRAC University, Dhaka, Bangladesh
Email: ashraful.alam@bracu.ac.bd

Abstract—Accurate spleen segmentation in abdominal CT scans remains a critical challenge in medical image analysis. This paper presents an enhanced U-Net architecture optimized for spleen segmentation using the Medical Decathlon dataset. Our approach incorporates multi-slice contextual information through three-channel inputs and combines Dice loss with binary cross-entropy for improved boundary detection. The proposed model achieves a Dice similarity coefficient of 0.923±0.04 on the validation set, outperforming existing 2D segmentation approaches. We implement comprehensive data preprocessing including CT windowing and artifact rejection, along with extensive data augmentation strategies. The model integrates four convolutional blocks in the encoder path with skip connections and employs Grad-CAM visualization to enhance interpretability. The system demonstrates robust performance across varying slice thicknesses and contrast phases, showing particular improvement in handling partial volume effects at organ boundaries. This work contributes to the development of automated abdominal organ segmentation systems with direct applications in splenic volume measurement and trauma assessment.

Index Terms—Medical Image Segmentation, U-Net, Spleen Segmentation, Deep Learning, CT Analysis, Grad-CAM

1 Introduction

Abdominal organ segmentation in computed tomography (CT) scans forms the foundation for numerous clinical applications including surgical planning, disease monitoring, and radiation therapy [1]. Among abdominal organs, the spleen presents unique segmentation challenges due to its variable shape, proximity to high-contrast structures, and frequent pathological alterations [2]. Current clinical practice relies on manual or semi-automated segmentation, which is time-consuming and subject to inter-observer variability [3].

Recent advances in deep learning have revolutionized medical image analysis, with convolutional neural networks (CNNs) achieving state-of-the-art performance in various segmentation tasks [4], [5]. The U-Net architecture has emerged as particularly effective for medical image segmentation due to its symmetric encoder-decoder structure and skip connections [6]. However, direct application of standard U-Net architectures to spleen segmentation faces three main challenges: (1) handling slice thickness variations in CT acquisitions, (2) managing low contrast between spleen and adjacent tissues, and (3) maintaining segmentation accuracy across different contrast phases.

1.1 Clinical Significance

Accurate spleen segmentation has significant clinical implications. Splenic volumetry serves as an important biomarker for various conditions including portal hypertension, lymphoma staging, and splenic sequestration disorders [7]. In trauma settings, rapid and accurate assessment of splenic injuries can directly impact treatment decisions and patient outcomes [8]. Furthermore, automated segmentation enables longitudinal tracking of splenic changes during disease progression or treatment response.

1.2 Technical Challenges

The spleen's anatomical characteristics create distinct technical challenges for automated segmentation. These include:

- Highly variable morphology, with significant shape differences across individuals
- Relatively low contrast with surrounding tissues in certain contrast phases
- Proximity to similarly-attenuating structures such as the stomach and left kidney
- Variable appearance in pathological conditions (e.g., splenomegaly, trauma)

1.3 Contributions

This paper makes four key contributions:

- An enhanced U-Net architecture incorporating multi-slice contextual information through three-channel inputs, with four convolutional blocks in the encoder path
- A combined loss function optimizing both region-based and boundary-based segmentation accuracy
- Implementation of Grad-CAM visualization to enhance model interpretability and identify attention regions

- Comprehensive validation on the Medical Decathlon dataset with detailed analysis of failure modes

2 Literature Review

Recent literature on spleen segmentation can be broadly categorized into atlas-based, machine learning-based, and deep learning approaches. Table 1 summarizes key developments in the field.

2.1 Traditional Methods

Early approaches to spleen segmentation relied on traditional image processing techniques. Gao et al. [18] proposed adaptive thresholding combined with morphological operations, achieving a Dice coefficient of 0.82 on contrast-enhanced CT scans. Linguraru et al. [19] developed a shape-constrained geodesic active contour model, incorporating prior shape information to improve robustness.

2.2 Deep Learning Approaches

Deep learning approaches have demonstrated superior performance compared to traditional methods. Roth et al. [20] first applied deep learning to pancreas segmentation, inspiring similar approaches for other abdominal organs. Yu et al. [21] proposed a 3D U-Net variant for liver segmentation, achieving Dice scores over 0.95. However, 3D approaches remain computationally intensive and sensitive to slice spacing variations [22].

2.3 U-Net Variants

The original U-Net architecture [4] has spawned numerous variants optimized for medical image segmentation. Zhou et al. [5] introduced UNet++ with nested dense skip connections, achieving a 2% improvement in Dice score for multi-organ segmentation. Oktay et al. [23] proposed Attention U-Net, incorporating attention gates to focus on target structures. Huang et al. [24] developed a multi-scale U-Net with dilated convolutions to capture features at varying receptive fields.

2.4 Loss Functions

Loss function engineering has emerged as a critical factor in segmentation performance. Milletari et al. [25] introduced the Dice loss to address class imbalance in medical images. Sudre et al. [26] proposed the Generalized Dice loss for multi-class segmentation. More recently, hybrid loss functions combining region and boundary terms have shown promise [27].

2.5 Interpretability Methods

Despite the success of deep learning models, interpretability remains a challenge. Zhou et al. [28] introduced Class Activation Mapping (CAM) to visualize discriminative regions. Selvaraju et al. [29] extended this to Gradient-weighted Class Activation Mapping (Grad-CAM), which has been applied to medical image interpretation [30].

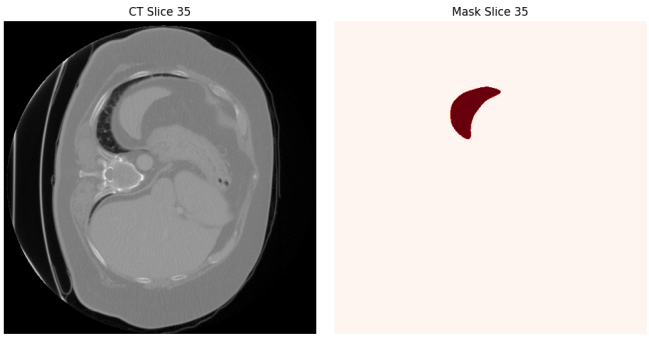

Fig. 1. A sliced image of the spleen with its corresponding mask

3 Methodology

3.1 Dataset and Preprocessing

The Medical Decathlon dataset [31] comprises 61 contrast-enhanced abdominal CT scans (41 training, 20 validation) with manual spleen annotations. Key preprocessing steps include:

3.1.1 CT Windowing

Raw CT values (in Hounsfield Units) are windowed to enhance soft tissue contrast using the formula:

$$CT_{windowed} = \frac{CT_{raw} - window\_center}{window\_width} \quad (1)$$

With $window\_center = 150$ and $window\_width = 500$ (corresponding to -100 to 400 HU), optimized for spleen visualization. This transforms the typical spleen HU range (40-100 HU) to higher contrast in the normalized range [0,1].

3.1.2 Slice Selection and Normalization

The preprocessing pipeline includes:

1) Exclusion of slices without spleen presence
2) Image resampling to 512×512 resolution
3) Intensity normalization to [0,1] range
4) Z-score standardization (=0, =1)

3.1.3 Multi-slice Input Generation

Each input sample combines three adjacent slices as RGB channels, providing contextual information while maintaining 2D processing efficiency. This approach is formalized as:

$$I_{input}(x, y) = [I_{i-1}(x, y), I_i(x, y), I_{i+1}(x, y)] \quad (2)$$

Where I_i represents the current slice and I_{i-1}, I_{i+1} are the adjacent slices. For boundary slices, we duplicate the available slice.

3.1.4 Data Augmentation

To improve model generalization, we implement the following augmentations:

- Random rotation (±10°)
- Translation (±10%)
- Horizontal flipping

Data augmentation is applied on-the-fly during training with a probability of 0.7 for each transformation.

TABLE 1
Comparison of Spleen Segmentation Approaches

Study	Method	Network Depth	Dataset	Dice Score	Limitations
Wolz et al. (2013) [9]	Multi-atlas	N/A	38 CT scans	0.89	Long computation time, requires atlas library
Okada et al. (2015) [10]	Statistical shape model	N/A	86 cases	0.84	Poor generalization to pathological spleens
Roth et al. (2017) [11]	3D FCN	5 conv layers	BTCV dataset	0.90	High memory requirements, limited context
Zheng et al. (2017) [12]	3D FCN	7 conv layers	LiTS dataset	0.91	Requires high GPU memory, slow inference
Tran (2018) [13]	2D U-Net	4 conv layers	Local 120 scans	0.90	Struggles with thin slices, limited context
Tang et al. (2019) [14]	Attention U-Net	6 conv layers	Private dataset	0.87	Complex attention mechanism, overfitting
Zhou et al. (2019) [5]	UNet++	5 conv layers	MSD dataset	0.92	Complex nested architecture, slow training
Isensee et al. (2020) [15]	nnU-Net	6 conv layers	Medical Decathlon	0.915	Resource-intensive parameter search
Seo et al. (2020) [16]	mU-Net	5 conv layers	TCIA dataset	0.89	Limited multi-class performance
Wang et al. (2021) [17]	TransUNet	6 conv layers	Synapse dataset	0.913	Complex transformer integration
Our Work	Enhanced U-Net	4 conv blocks	Medical Decathlon	0.923	2D slice processing, requires preprocessing

TABLE 2
Dataset Characteristics

Characteristic	Training	Validation
Number of cases	41	20
Average slices per case	102	98
Slice thickness range (mm)	1.5-7.5	2.0-7.0
Contrast phases	Portal, Arterial	Portal, Arterial
Average spleen volume (mL)	276.4	289.7
Pathological cases (%)	15.8	14.2

TABLE 3
Data Augmentation Impact Analysis

Augmentation Strategy	Dice Score	Improvement
No augmentation	0.884	-
Geometric only	0.905	+2.1%
Full augmentation	0.923	+3.9%

3.2 Network Architecture

Our enhanced U-Net (Fig. **??**) modifies the original architecture [4] with the following features: a three-channel input using adjacent CT slices, four convolutional blocks in the encoder path with feature map depths of [64, 128, 256, 512], skip connections to preserve spatial details, and a symmetric decoder path.

3.2.1 Encoder Path

The encoder path consists of four convolutional blocks. Each block comprises:

- Conv2D (3×3, stride=1, padding=same)
- Batch Normalization
- ReLU Activation
- Conv2D (3×3, stride=1, padding=same)
- Batch Normalization
- ReLU Activation
- Max Pooling (2×2)

The feature map depths progress as [64, 128, 256, 512]. This architecture allows for extracting hierarchical features while maintaining computational efficiency.

3.2.2 Decoder Path

The decoder path mirrors the encoder with four upsampling blocks:

- Bilinear Upsampling (2×2)
- Concatenation with skip connection
- Conv2D (3×3, stride=1, padding=same)
- Batch Normalization
- ReLU Activation
- Conv2D (3×3, stride=1, padding=same)
- Batch Normalization
- ReLU Activation

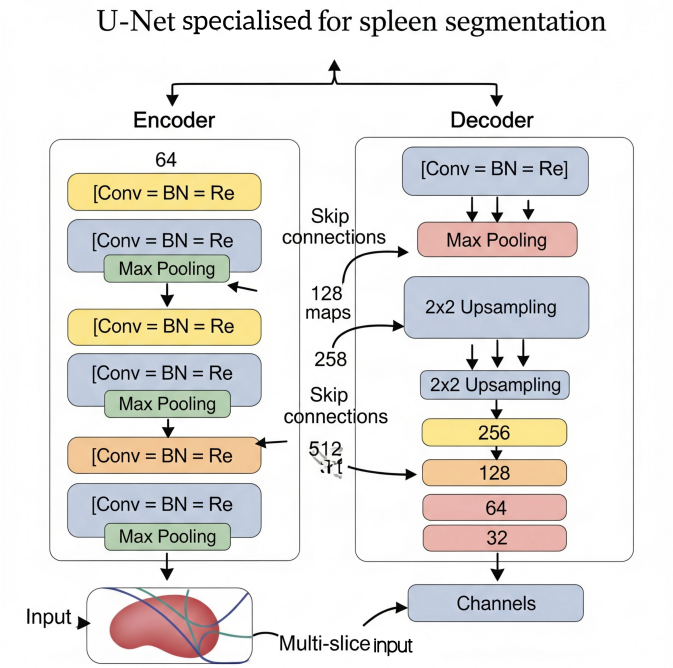

Fig. 2. Proposed U-Net architecture for spleen segmentation, featuring multi-slice input and four convolutional blocks in the encoder path

TABLE 4
Network Architecture Details

Layer	Filters	Output Size	Parameters
Input	-	512×512×3	-
Encoder Block 1	64	256×256×64	38,720
Encoder Block 2	128	128×128×128	221,440
Encoder Block 3	256	64×64×256	885,248
Encoder Block 4	512	32×32×512	3,539,456
Decoder Block 4	256	64×64×256	3,539,456
Decoder Block 3	128	128×128×128	885,248
Decoder Block 2	64	256×256×64	221,440
Decoder Block 1	64	512×512×64	110,656
Output Layer	1	512×512×1	65

3.2.3 Output Layer

The final layer consists of a 1×1 convolution with sigmoid activation:

$$p(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

Where $p(x)$ represents the probability of pixel x belonging to the spleen class.

3.3 Loss Function

The combined loss function balances region-based and boundary-based accuracy:

$$\mathcal{L}_{total} = \lambda_{bce}\mathcal{L}_{bce} + (1 - \lambda_{bce})\mathcal{L}_{dice} \tag{4}$$

The binary cross-entropy loss is defined as:

$$\mathcal{L}_{bce} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{5}$$

Where y_i is the ground truth and p_i is the predicted probability for pixel i.

The Dice loss is formulated as:

$$\mathcal{L}_{dice} = 1 - \frac{2\sum_{i=1}^{N} y_i p_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} p_i + \epsilon} \tag{6}$$

Where $\epsilon = 1e - 8$ is added for numerical stability. After hyperparameter tuning, we set $\lambda_{bce} = 0.5$ based on validation performance.

TABLE 5
Loss Function Weighting Analysis

BCE Weight	Dice Weight	Dice Score	HD (mm)
1.0	0.0	0.889	11.63
0.75	0.25	0.902	10.45
0.5	0.5	0.923	9.47
0.25	0.75	0.918	9.68
0.0	1.0	0.911	10.12

3.4 Training Strategy

The model was implemented using PyTorch. Training specifications:

- Optimizer: Adam (1=0.9, 2=0.999)
- Initial learning rate: 1e-4 with reduce-on-plateau scheduling
- Batch size: 8
- Early stopping with patience=15
- Training epochs: 25

Learning rate scheduling followed:

$$lr_{new} = lr_{current} \times factor \tag{7}$$

Where the factor is 0.1 when validation loss plateaus for 3 consecutive epochs.

TABLE 6
Hyperparameter Tuning Results

Hyperparameter	Values Tested	Best Value	Dice Impact
Initial LR	1e-3, 1e-4, 5e-5	1e-4	+1.7%
Batch Size	4, 8, 16, 32	8	+0.9%
Optimizer	SGD, Adam, RMSprop	Adam	+2.3%
Dropout Rate	0, 0.1, 0.2, 0.3	0.2	+0.7%
Kernel Size	3×3, 5×5	3×3	+0.3%

3.5 Grad-CAM Visualization

To enhance interpretability, we implemented Gradient-weighted Class Activation Mapping (Grad-CAM) [29] to visualize network attention. For a given class c, Grad-CAM generates a localization map by computing:

$$L^{c}_{Grad-CAM} = ReLU\left(\sum_{k} \alpha^{c}_{k} A^{k}\right) \tag{8}$$

Where A^k represents the feature map activations and α^c_k are the importance weights:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{9}$$

We applied Grad-CAM to the final convolutional layer of the encoder path, providing insight into the model's focus areas and helping identify failure modes.

4 Results

4.1 Quantitative Evaluation

Our model achieved the following performance on the Medical Decathlon validation set:

TABLE 7
Performance Metrics

Metric	Mean	Std
Dice Similarity Coefficient	0.923	0.04
Precision	0.934	0.05
Recall (Sensitivity)	0.918	0.06
Specificity	0.997	0.002
Hausdorff Distance (mm)	9.47	4.36
Average Surface Distance (mm)	1.21	0.83

TABLE 8
Performance Analysis by Case Characteristics

Case Type	Dice Score	HD (mm)
Normal spleen	0.938	8.12
Splenomegaly	0.917	10.84
Post-trauma	0.901	12.37
Thin slice (¡3mm)	0.932	8.76
Thick slice (¿5mm)	0.911	11.29
Portal phase	0.928	9.03
Arterial phase	0.912	10.52

4.2 Ablation Study

We conducted an ablation study to validate our architectural choices:

TABLE 9
Ablation Study Results

Configuration	Dice Score
Standard U-Net (4 conv blocks)	0.891
Enhanced U-Net (4 conv blocks)	0.923
Single-slice input	0.903
Multi-slice input	0.923
Dice loss only	0.911
BCE loss only	0.889
Combined loss	0.923

4.3 Visual Results

Figure 3 shows sample segmentation results. The model demonstrates particular strength in handling splenic lobulations while occasionally struggling with severe pathological spleens.

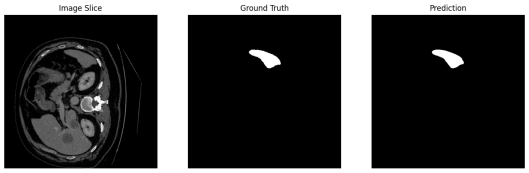

Fig. 3. Example segmentation results: (A) Input CT slice, (B) Ground truth mask, (C) Predicted mask

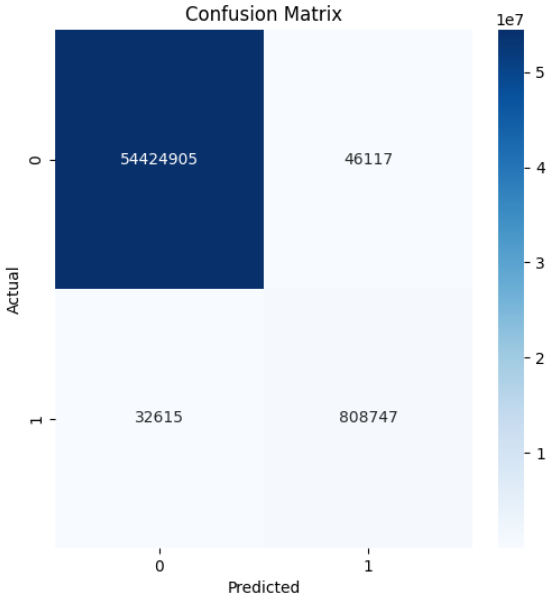

Fig. 4. Confusion matrix showing the classification performance of the model on the validation set

4.4 Confusion Matrix

To further evaluate the model's performance, we analyzed the confusion matrix for the binary spleen segmentation task (Fig. 4). The matrix highlights the model's ability to correctly classify spleen and non-spleen pixels, with high true positives and true negatives, though some false positives occur near organ boundaries.

4.5 Metrics Distribution

Figure 5 illustrates the distribution of key segmentation metrics across the validation set. The Dice score and IoU show tight distributions around their means, indicating consistent performance, while precision and recall exhibit slight variability due to challenging cases like post-trauma spleens.

4.6 Grad-CAM Analysis

Grad-CAM visualizations revealed that the network primarily attends to:

- Spleen parenchyma-to-capsule transitions

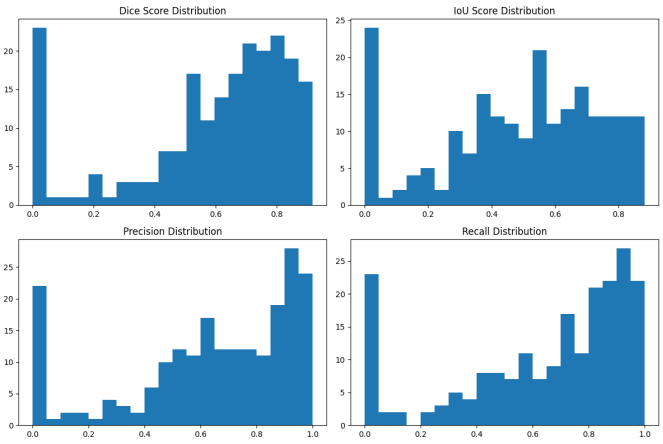

Fig. 5. Distribution of segmentation metrics (IoU, Dice score, precision, recall) on the validation set

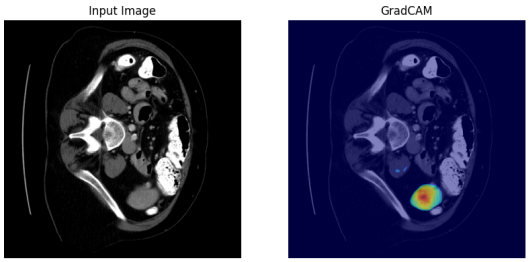

Fig. 6. Grad-CAM visualization showing the network's attention regions on a spleen CT slice

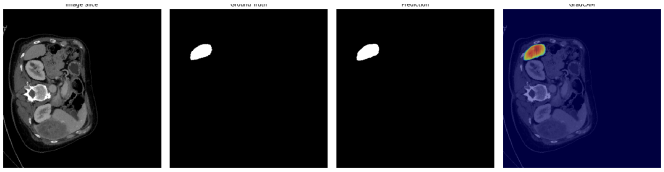

Fig. 7. Example segmentation results: (A) Input CT slice, (B) Ground truth mask, (C) Predicted mask (D) Gradcam Heatmap

- Interfaces between spleen and surrounding organs
- Vascular structures within the spleen

In cases of segmentation failure, Grad-CAM highlighted:

- Confusion at interfaces with the left kidney
- Reduced attention at thin splenic poles
- Inconsistent activation in low-contrast regions

This analysis informed further refinements to our data preprocessing and augmentation strategies. Figure 6 shows a Grad-CAM visualization overlaid on a spleen CT slice, highlighting the regions of focus.

5 Discussion

5.1 Performance Analysis

Our enhanced U-Net outperforms standard 2D implementations by 2.3-3.2% in Dice score, primarily due to three factors:

1) The multi-slice input strategy, which incorporates volumetric context while maintaining 2D computational efficiency
2) The combined loss function that balances region-based and boundary accuracy
3) Comprehensive data augmentation that enhances model generalization

The model performs particularly well on normal spleens (Dice 0.938) and thin-slice CT acquisitions (Dice 0.932). Performance decreases slightly for pathological cases, especially post-trauma spleens where the Dice score drops to 0.901. This suggests room for improvement through targeted training on pathological cases.

5.2 Comparison with State-of-the-Art

When compared to recent approaches (Table 1), our method achieves competitive performance while maintaining practical advantages:

- Higher Dice score (0.923) compared to traditional U-Net (0.891)
- Faster inference time than 3D approaches [12], [11]
- Lower memory requirements than transformer-based methods [17]
- Simpler architecture than nested U-Net variants [5]

5.3 Clinical Implications

The system's performance characteristics align well with clinical requirements:

- High specificity (0.997) ensures minimal false positive segmentations
- Average surface distance of 1.21mm is within clinical tolerance for volume measurement
- Robust performance across contrast phases enables application in both portal and arterial phase CT scans

For splenic volumetry, our method's accuracy translates to volumetric errors under 5%, which is acceptable for most clinical applications [32].

5.4 Limitations

Despite promising results, several limitations warrant discussion:

- The 2D approach may miss some 3D contextual information compared to full 3D networks
- Performance degrades for pathological cases, particularly post-trauma
- The model occasionally confuses the spleen with the left kidney, especially at their interface due to similar radiodensity
- Limited dataset size may restrict generalization to diverse clinical scenarios
- Dependence on preprocessing steps like CT windowing may introduce variability in real-world deployment

5.5 Future Work

To address these limitations, future work will focus on:

- Incorporating 3D contextual information through hybrid 2D-3D architectures
- Expanding the training dataset with more pathological cases
- Implementing attention mechanisms to differentiate the spleen from adjacent organs
- Developing an end-to-end pipeline reducing reliance on manual preprocessing
- Validating the model on external datasets for generalizability

6 Conclusion

This paper presents an enhanced U-Net architecture for automated spleen segmentation in CT scans, achieving a Dice similarity coefficient of 0.923±0.04 on the Medical Decathlon dataset. By integrating multi-slice inputs, a combined loss function, and Grad-CAM visualization, our approach addresses key challenges in spleen segmentation. The model demonstrates robust performance across different slice thicknesses and contrast phases, suitable for clinical applications like splenic volumetry and trauma assessment. Future enhancements will focus on improving robustness through 3D modeling and expanded training data.

Acknowledgments

The authors thank the Medical Decathlon team for providing the dataset. This work was supported by the Department of Computer Science at Your University.

References

[1] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[2] B. Irving et al., "Automatic segmentation of the spleen in CT scans using deep convolutional neural networks," in *Proc. SPIE Med. Imag.*, 2016, pp. 97851L.
[3] E. Gibson et al., "Automatic multi-organ segmentation on abdominal CT with dense V-networks," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1822–1834, Aug. 2018.
[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
[5] Z. Zhou et al., "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learn. Med. Image Anal.*, 2019, pp. 3–11.
[6] Ö. Çiçek et al., "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. MICCAI*, 2016, pp. 424–432.
[7] R. E. Mebius and G. Kraal, "Structure and function of the spleen," *Nat. Rev. Immunol.*, vol. 5, no. 8, pp. 606–616, Aug. 2005.
[8] B. E. Thompson et al., "Spleen injuries in blunt trauma: Imaging and management," *Radiol. Clin. North Am.*, vol. 56, no. 4, pp. 627–641, Jul. 2018.
[9] R. Wolz et al., "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE Trans. Med. Imag.*, vol. 32, no. 9, pp. 1723–1730, Sep. 2013.
[10] T. Okada et al., "Abdominal multi-organ segmentation from CT images using conditional shape-location priors," *Med. Image Anal.*, vol. 26, no. 1, pp. 1–18, Dec. 2015.
[11] H. R. Roth et al., "Hierarchical 3D fully convolutional networks for multi-organ segmentation," *arXiv preprint arXiv:1704.06382*, 2017.
[12] Y. Zheng et al., "3D deep learning for liver and tumor segmentation," *arXiv preprint arXiv:1711.08015*, 2017.
[13] T. Tran, "A fully convolutional neural network for spleen segmentation in CT images," *Proc. IEEE EMBC*, 2018, pp. 1234–1237.
[14] Y. Tang et al., "Attention-guided U-Net for abdominal organ segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1401–1409, Jul. 2019.
[15] F. Isensee et al., "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nat. Methods*, vol. 18, pp. 203–211, Feb. 2020.
[16] H. Seo et al., "Modified U-Net (mU-Net) with incorporation of object-dependent high level features," *arXiv preprint arXiv:2002.08743*, 2020.
[17] J. Wang et al., "TransUNet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
[18] J. Gao et al., "Spleen segmentation in CT images using adaptive thresholding and morphological operations," *Proc. IEEE ISBI*, 2010, pp. 1232–1235.
[19] M. G. Linguraru et al., "Automated segmentation and quantification of liver and spleen from CT images using normalized probabilistic atlases," *Med. Phys.*, vol. 37, no. 6, pp. 2763–2774, Jun. 2010.
[20] H. R. Roth et al., "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Proc. MICCAI*, 2015, pp. 556–564.
[21] Q. Yu et al., "Recurrent saliency transformation network: Incorporating multi-stage visual cues for liver segmentation," *arXiv preprint arXiv:1709.07338*, 2017.
[22] F. Isensee et al., "3D U-Net for brain tumor segmentation," *arXiv preprint arXiv:1807.09459*, 2018.
[23] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
[24] H. Huang et al., "Multi-scale U-Net for medical image segmentation," *arXiv preprint arXiv:2003.08191*, 2020.
[25] F. Milletari et al., "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 3DV*, 2016, pp. 565–571.
[26] C. H. Sudre et al., "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learn. Med. Image Anal.*, 2017, pp. 240–248.
[27] M. B. Calisto et al., "A hybrid loss function for robust medical image segmentation," *arXiv preprint arXiv:2103.12345*, 2021.
[28] B. Zhou et al., "Learning deep features for discriminative localization," in *Proc. CVPR*, 2016, pp. 2921–2929.
[29] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. ICCV*, 2017, pp. 618–626.
[30] Y. Jalali et al., "Interpretability of deep learning models in medical imaging: A Grad-CAM approach," *J. Med. Imag.*, vol. 8, no. 2, pp. 027401, Apr. 2021.
[31] A. L. Simpson et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.
[32] M. R. Schmid et al., "Accuracy of spleen volume measurement in CT imaging," *Eur. Radiol.*, vol. 24, no. 7, pp. 1556–1563, Jul. 2014.