# ShapeAtlas-UNet: Enhancing U-Net Performance for Liver Segmentation in CT Images Using Shape Atlas Information

Yubraj Bhandari

December 2024

**Abstract**

This study presents a novel approach to enhance the performance of U-Net models for liver segmentation in CT images, particularly focusing on improving efficiency for smaller networks and resource-constrained environments. We introduce a method that incorporates shape atlas information into the U-Net architecture, aiming to boost segmentation accuracy without significantly increasing computational demands during inference.

Our approach involves generating a probabilistic shape atlas from the training data and integrating it into the U-Net model through an additional encoder. This encoder transforms the shape atlas into a latent representation that is concatenated with the main U-Net's latent space. We evaluate this method on the Abdomen Atlas mini dataset, using various U-Net configurations with different latent space dimensions and training data sizes.

Results demonstrate substantial improvements in segmentation performance, especially for smaller models and limited training data scenarios. For instance, a 32-dimensional latent space model showed an 8% increase in Dice score (from 0.78 to 0.86) with only a modest increase in parameters. The method's efficacy diminished for larger models, suggesting that they may already implicitly capture much of the shape information.

This research contributes to the field of medical image segmentation by offering a computationally efficient way to enhance performance, particularly beneficial for deployment on low-end devices or in settings with limited resources. The findings pave the way for more accessible and accurate medical image analysis tools, potentially impacting clinical practice in diverse healthcare settings.

Code: https://github.com/yubrajbhandari923/shapeatlasUnet

## 1 Introduction

Medical image segmentation is a critical task in computer vision, playing a pivotal role in healthcare diagnostics, treatment planning, and medical research [1]. This process involves precisely delineating anatomical structures or regions of interest within medical images,

enabling healthcare professionals to analyze specific organs, tissues, or abnormalities with greater accuracy and efficiency [2].

In recent years, deep learning approaches, particularly convolutional neural networks (CNNs), have revolutionized the field of medical image segmentation. Among these, the U-Net architecture, introduced by Ronneberger et al. in 2015, has emerged as a cornerstone in this domain [3]. U-Net's distinctive encoder-decoder structure, coupled with skip connections, allows it to capture both fine-grained details and broader contextual information, making it particularly well-suited for medical image segmentation tasks [4].

While U-Net and its variants have demonstrated remarkable performance across various medical imaging modalities, including X-rays, MRI, and ultrasound [5], this report focuses specifically on Computed Tomography (CT) imaging. CT scans provide detailed three-dimensional representations of internal body structures, offering invaluable insights for diagnosing and monitoring a wide range of conditions. However, the inherent complexity of CT data presents unique challenges for segmentation tasks.

The three-dimensional nature of CT scans significantly increases the computational demands of segmentation models. Traditional U-Net architectures, when adapted for 3D data, often require a substantial increase in the number of parameters to effectively capture the spatial relationships across multiple slices. This increase in model complexity leads to longer training times, increased memory requirements, and slower inference speeds. Consequently, the deployment of these sophisticated models may be limited in resource-constrained environments or on less powerful hardware platforms. This limitation poses a significant barrier to the widespread adoption of advanced medical image segmentation techniques, particularly in less technologically advanced regions of the world.

Addressing this challenge requires innovative approaches that can enhance the performance of smaller, more efficient models without significantly increasing their computational footprint. By improving the accuracy and efficiency of compact U-Net variants, we can potentially democratize access to advanced medical image analysis tools like 3D Slicer [8], making them viable for deployment on a broader range of devices and in diverse healthcare settings.

This research project aims to contribute to this crucial area by exploring novel techniques to boost the performance of small U-Net models, specifically targeting liver segmentation in CT images. By incorporating additional contextual information through a shape atlas encoder, we seek to demonstrate that significant improvements in segmentation accuracy can be achieved for models with limited latent space dimensions, without substantially increasing the computational requirements during inference.

The findings of this study have the potential to not only advance the field of medical image segmentation but also to make a meaningful impact on healthcare accessibility. By developing more efficient and accurate segmentation models, we can help bridge the technological gap and ensure that the benefits of AI-assisted medical imaging are more widely available, regardless of geographical or economic constraints.

# 2 Related Work

Recent advancements in medical image segmentation have led to remarkable improvements in performance and versatility. Two notable examples are the nnU-Net and the CLIP-Driven Universal Model, which have achieved state-of-the-art results on various segmentation tasks.

The nnU-Net, introduced by Isensee et al. [12], has demonstrated exceptional performance across a wide range of medical image segmentation challenges. It achieved the highest mean Dice scores in the Medical Segmentation Decathlon [9], showcasing its adaptability to different anatomical structures and imaging modalities. Similarly, the CLIP-Driven Universal Model, proposed by Liu et al. [11], has shown impressive results in organ segmentation and tumor detection tasks, ranking first on the Medical Segmentation Decathlon public leaderboard. However, these high-performing models come with significant computational costs. The nnU-Net, for instance, requires substantial GPU memory (up to 32 GB for some configurations) and can take several days to train on large datasets [6]. The CLIP-Driven Universal Model, while more efficient than dataset-specific models, still demands considerable computational resources due to its use of large-scale pre-trained language models.

## 2.1 Hypothesis

Motivated by the success of the CLIP-Driven Universal Model in leveraging additional contextual information, this research explores a more lightweight approach to enhancing segmentation performance. By encoding and integrating shape atlas information into the U-Net architecture, the segmentation performance of smaller models can be improved without significantly increasing their computational requirements during inference. This approach could be particularly beneficial for models with limited representational capacity, providing them with valuable prior knowledge about the expected shape and location of the liver. By doing so, we aim to bridge the gap between the high performance of large, resource-intensive models and the practical needs of resource-constrained environments in medical imaging applications.

# 3 Methodology

Our methodology focuses on enhancing the performance of small U-Net models for liver segmentation in 3D CT images. We first describe the baseline U-Net architecture and then present our proposed enhancement using shape atlas information.

## 3.1 Problem Formulation

Let

$$X \in \mathbb{R}^{H \times W \times D \times 1}$$

be an input 3D CT image with height $H$, width $W$, depth $D$, and a single channel. The segmentation task aims to find a function $f$ that maps the input image to a segmentation mask

$$Y \in \{0, 1\}^{H \times W \times D \times 2}$$

for liver segmentation, where the two channels represent background and liver masks respectively:

$$Y = f(X; \theta)$$

where $\theta$ represents the parameters of the segmentation model. The final binary segmentation is obtained by applying argmax along the channel dimension.

# 4 Methodology

Our methodology focuses on enhancing the performance of small U-Net models for liver segmentation in 3D CT images. We first describe the baseline U-Net architecture and then present our proposed enhancement using shape atlas information.

## 4.1 Baseline U-Net Architecture

The standard U-Net consists of an encoder-decoder architecture with skip connections. Let

$$X \in \mathbb{R}^{H \times W \times D \times 1}$$

be an input 3D CT image with height $H$, width $W$, depth $D$, and a single channel. The segmentation task aims to find a function $f$ that maps the input image to a segmentation mask

$$Y \in \{0, 1\}^{H \times W \times D \times 2}$$

for liver segmentation:

$$Y = f(X; \theta)$$

where $\theta$ represents the model parameters.

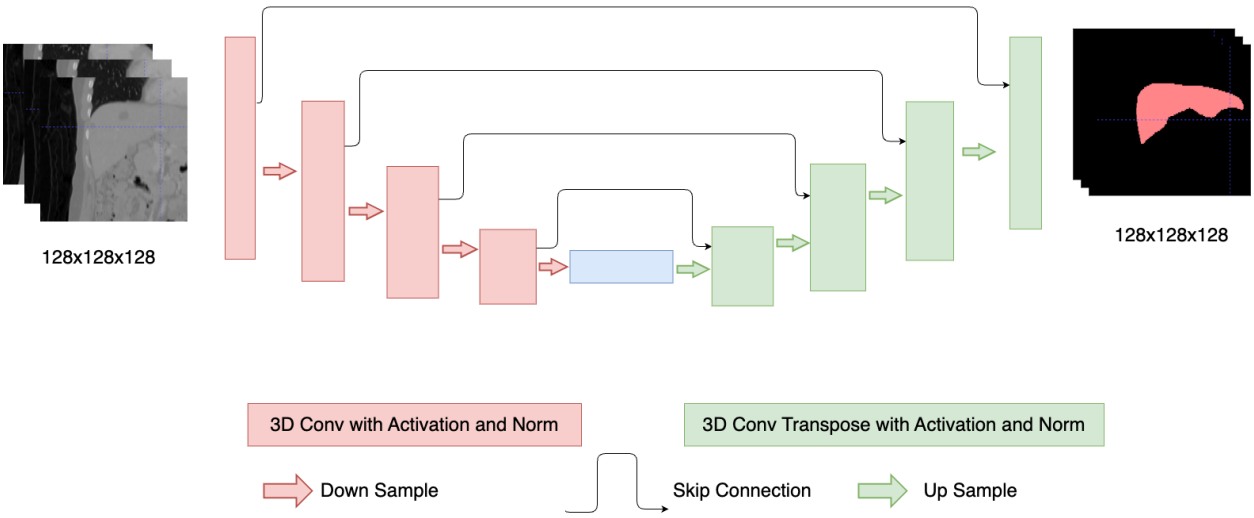


Figure 1: Baseline 3D U-Net architecture for medical image segmentation. The network consists of an encoder path (left) with consecutive 3D convolutions and downsampling operations, and a decoder path (right) with upsampling operations. Skip connections (curved arrows) allow the network to preserve fine spatial details by combining high-resolution features from the encoder with upsampled features in the decoder. The input is a $128 \times 128 \times 128$ CT volume patch, and the output is a binary segmentation mask of the same dimensions.

The encoder progressively reduces spatial dimensions while increasing the number of channels through consecutive convolutional and downsampling operations. At each level $l$, the encoder produces feature maps $E^l(X)$. The decoder then upsamples these features and combines them with corresponding encoder features through skip connections:

$$Y_l = D_l([U(Y_{l-1}); E^l(X)])$$

where $Y_l$ is the output of the $l$-th decoder layer, $U$ is an upsampling operation, and $[\cdot;\cdot]$ denotes concatenation.

## 4.2 Enhanced Architecture with Shape Atlas

To improve segmentation performance, particularly for smaller models, we propose incorporating shape prior information through an additional encoder pathway.

### 4.2.1 Shape Atlas Generation

We create a shape atlas

$$S \in [0,1]^{128 \times 128 \times 128}$$

representing the probability of liver presence at each spatial location. This atlas is generated by centering and cropping all liver label masks to 128x128x128 dimensions, then averaging across the training dataset:

$$S = \frac{1}{N} \sum_{i=1}^{N} Y_i^{cropped}$$

where $N$ is the number of training samples and $Y_i^{cropped}$ is the centered and cropped binary segmentation mask for the $i^{\text{th}}$ sample. This approach allows us to create a consistent shape atlas despite variations in input CT image dimensions.

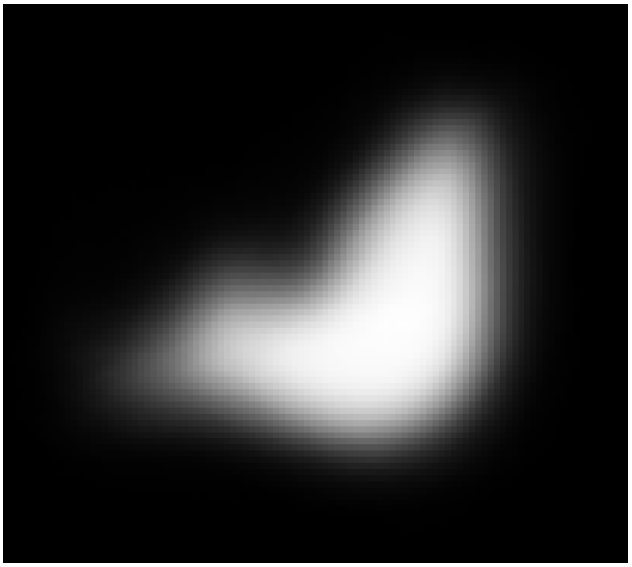

Figure 2: Mid-slice visualization of the probabilistic shape atlas generated by averaging aligned liver segmentation masks. Brighter regions indicate higher probability of liver presence across the training dataset, showing the characteristic liver shape and its spatial distribution. The diffused appearance reflects the natural anatomical variation in liver size and position across different patients.

#### 4.2.2 Dual Encoder Architecture

Our enhanced U-Net architecture incorporates two parallel encoders with skip connections:

1. The main encoder $E_X : \mathbb{R}^{H \times W \times D \times 1} \rightarrow \mathbb{R}^{h \times w \times d \times c_x}$ for the input CT image.
2. A shape atlas encoder $E_S : \mathbb{R}^{128 \times 128 \times 128} \rightarrow \mathbb{R}^{h \times w \times d \times c_s}$ for the pre-computed shape atlas.

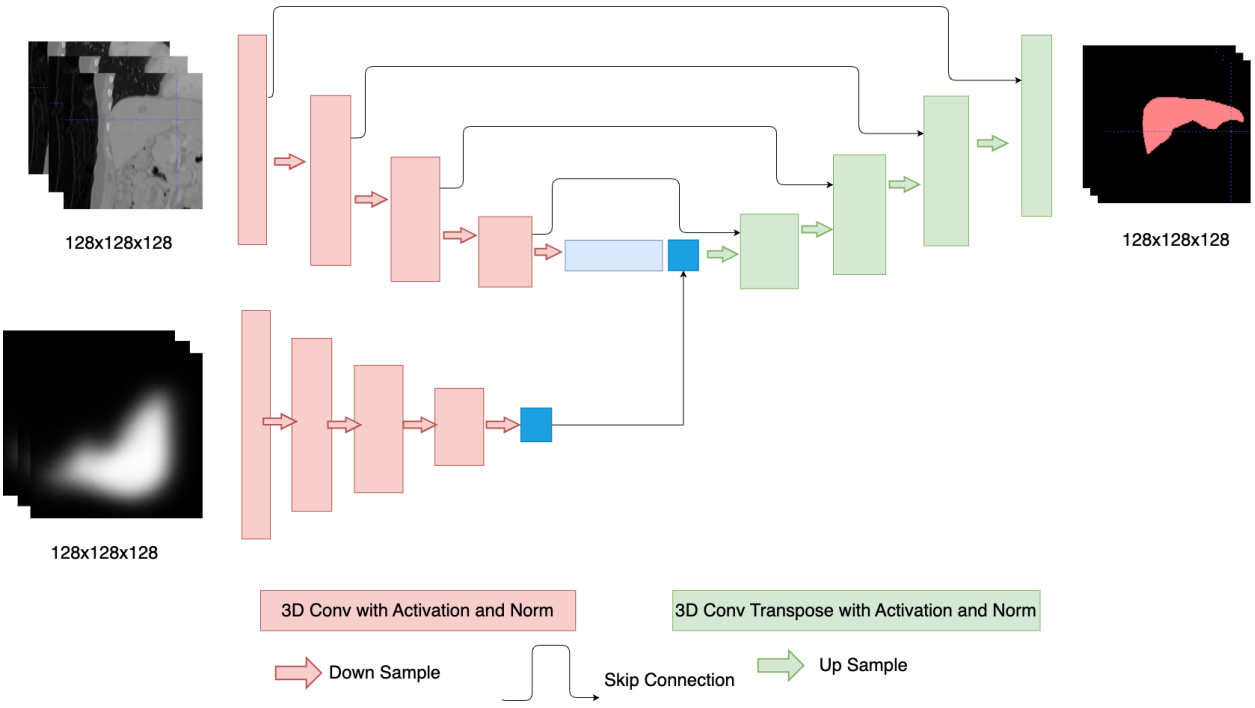


Figure 3: Proposed dual-encoder U-Net architecture for liver segmentation. The upper pathway processes 3D CT image patches (128×128×128) through a standard U-Net encoder with skip connections. The lower pathway encodes the shape atlas (128×128×128) through a parallel encoder. The shape atlas latent representation is concatenated with the CT image latent representation at the bottleneck layer before decoding. Red blocks represent 3D convolutions with downsampling, green blocks represent 3D transpose convolutions with upsampling, and gray arrows indicate skip connections. During inference, only the pre-computed shape atlas latent representation is needed, reducing computational overhead.

Both encoders transform their inputs to the same spatial dimensions ($h \times w \times d$) but with different numbers of channels ($c_x$ and $c_s$ respectively). This allows for concatenation along the channel dimension in the latent space:

$$Z = [E_X(X); E_S(S)]$$

where $[\cdot;\cdot]$ denotes channel-wise concatenation.

The decoder $D : \mathbb{R}^{h \times w \times d \times (c_x + c_s)} \to \mathbb{R}^{H \times W \times D \times 2}$ then processes this enhanced latent representation to produce the final segmentation:

$$Y = \text{softmax}(D(Z))$$

where softmax is applied along the channel dimension.

At each decoder level, we combine features from only the main image encoder pathway through skip connections:

$$Y_l = D_l([U(Y_{l-1}); E_X^l(X)])$$

where $Y_l$ is the output of the $l$-th decoder layer, $U$ is an upsampling operation, and $E_X^l(X)$ represents the skip connection features from the $l$-th level of the main image encoder. The shape atlas encoder only contributes its final latent representation which is concatenated at the bottleneck layer, and does not participate in the skip connections.

This approach solves the problem of incorporating shape atlas information by allowing the model to learn the relationship between the CT image features and the shape prior in the latent space. Despite the shape atlas and CT images having different shapes and spatial characteristics, their latent representations can be easily concatenated, enabling the decoder to leverage both sources of information effectively.

### 4.2.3 Loss Function and Optimization

We employ the Dice loss function, which is well-suited for 3D segmentation tasks and addresses class imbalance issues common in medical imaging. The Dice coefficient between the predicted segmentation $\hat{Y}$ and ground truth $Y$ is defined as:

$$DSC = \frac{2|\hat{Y} \cap Y|}{|\hat{Y}| + |Y|}$$

The Dice loss is then formulated as:

$$L_{Dice} = 1 - DSC = 1 - \frac{2|\hat{Y} \cap Y|}{|\hat{Y}| + |Y|}$$

We train the model using the Adam optimizer to minimize the Dice loss.

### 4.2.4 Optimization for Inference

During inference, we optimize the model by removing the deterministic part of the shape atlas encoder. Instead, we pre-compute the latent representation of the shape atlas

$$Z_S = E_S(S)$$

and directly concatenate it with the output of the main encoder:

$$Z = [E_X(X); Z_S]$$

This approach reduces the computational overhead during inference while still leveraging the benefits of the shape atlas information. This trick aims to enhance the performance of small U-Net models for liver segmentation in 3D CT images, particularly for models with limited latent space dimensions, without significantly increasing their computational requirements during inference.

# 5 Experiments and Results

To evaluate the effectiveness of our proposed method, we conducted experiments using the Abdomen Atlas mini 1.0 dataset [10]. The data preprocessing and experiments were implemented using monai framework [7]

## 5.1 Dataset and Preprocessing

The dataset consists of approximately 5,000 CT images with varying input dimensions. For our experiments, we randomly selected different training set sizes (1000, 512, 256, 128, 64 images) while maintaining a fixed validation set of 100 images and a test set of 265 images. All images were preprocessed using the following steps:

1. Resampling to 1.5×1.5×1.5mm spacing in LPI orientation
2. Intensity normalization
3. Patching into 128×128×128 segments with 80% overlap during inference

## 5.2 Experimental Setup

We implemented UNet models with the following channel configurations:

- (16, 32, 64, 128, 256)
- (8, 16, 32, 64, 128)
- (4, 8, 16, 32, 64)
- (2, 4, 8, 16, 32)

The models were constructed using the `monai.networks.blocks.Convolution` function with a stride of 2 to halve the input dimensions at each layer. This resulted in a latent representation of size `latent channel × 8 × 8 × 8` for all configurations.

For the shape atlas integration, we used latent embeddings with channel dimensions (64, 32, 16, 8), which were transformed to `8 × 8 × 8 × 8` and concatenated with the UNet's latent representation.

## 5.3 Results

Table 1 presents the experimental results, showcasing the performance of different model configurations with and without the shape atlas information.

Table 1: Experimental Results

| Latent Layer Dimension | Number of Parameters (at inference) | Number of Training Images | Atlas Map Provided | Test Set Mean Dice Score |
|---|---|---|---|---|
| 32 | 41,813 | 128 | False | 0.78 |
| 32 + 8 | 45,269 | 128 | True | 0.86 |
| 64 | 166,127 | 128 | False | 0.85 |
| 64 + 8 | 173,039 | 128 | True | 0.88 |
| 64 | 166,127 | 256 | False | 0.88 |
| 64 + 8 | 173,039 | 256 | True | 0.90 |
| 128 | 662,291 | 128 | False | 0.911 |
| 128 + 8 | 676,115 | 128 | True | 0.913 |
| 256 | 2,644,763 | 1024 | False | 0.959 |
| 256 + 8 | 2,672,411 | 1024 | True | 0.960 |

### 5.4 Analysis

The results demonstrate several key findings:

1. **Impact of Shape Atlas**: For smaller models (32 and 64 latent dimensions), the inclusion of the shape atlas information led to significant improvements in segmentation performance. For instance, the 32-dimensional model saw an increase in Dice score from 0.78 to 0.86 with the addition of the shape atlas.

2. **Diminishing Returns for Larger Models**: As the model size increased, the impact of the shape atlas became less pronounced. For the largest model (256 latent dimensions), the improvement was minimal (0.959 to 0.960).

3. **Data Efficiency**: The shape atlas appeared to be particularly beneficial when training data was limited. For example, the 64-dimensional model with 128 training images and shape atlas (0.88) outperformed the same model without shape atlas and 256 training images (0.85).

4. **Parameter Efficiency**: The addition of the shape atlas information resulted in only a modest increase in the number of parameters, making it a cost-effective approach for improving performance, especially for smaller models.

These results suggest that our proposed method of incorporating shape atlas information is particularly effective for enhancing the performance of smaller UNet models and in scenarios where training data is limited. This approach offers a promising solution for improving medical image segmentation in resource-constrained environments or on low-end devices.

## 6 Conclusion

This study demonstrates the effectiveness of incorporating shape atlas information to enhance the performance of U-Net models for liver segmentation in CT images, particularly

for smaller models and limited training data scenarios. Our proposed method significantly improved segmentation accuracy while maintaining computational efficiency, addressing the crucial need for high-performance medical image segmentation on resource-constrained devices.

Key findings from our experiments include:
1. Substantial performance improvements for smaller models (32 and 64 latent dimensions) with the integration of shape atlas information.
2. Enhanced data efficiency, allowing models to achieve better performance with fewer training examples.
3. Minimal increase in computational overhead during inference, making the approach suitable for deployment on low-end devices.

These results highlight the potential of our method to democratize access to advanced medical image segmentation techniques, particularly in settings with limited computational resources or data availability.

## 6.1 Future Work

While our current study focused on liver segmentation, there are several promising directions for future research:

1. **Multi-organ Segmentation**: Extending the approach to multiple organs by incorporating separate shape atlas encoders for each organ of interest. This could involve developing a modular architecture where organ-specific shape atlases can be selectively integrated based on the segmentation task.

2. **Dynamic Atlas Selection**: Implementing a mechanism to dynamically select or weight different shape atlases based on the input image characteristics, potentially improving adaptability to diverse anatomical variations.

3. **Transfer Learning**: Investigating the transferability of learned shape atlas encoding across different datasets or even different imaging modalities, potentially reducing the need for large annotated datasets for new segmentation tasks.

4. **Integration with Other Architectures**: Exploring the applicability of the shape atlas approach to other segmentation architectures beyond U-Net, such as attention-based models or transformer architectures.

5. **Uncertainty Quantification**: Incorporating uncertainty estimation techniques to provide confidence measures alongside segmentation predictions, which could be particularly valuable in clinical decision-making processes.

6. **Real-time Segmentation**: Optimizing the model for real-time performance on mobile or edge devices, enabling applications in intra-operative guidance or point-of-care diagnostics.

In conclusion, our work presents a promising approach for improving medical image segmentation performance, particularly in resource-constrained environments. By bridging the gap between model complexity and computational efficiency, this method has the potential to significantly impact clinical practice, enabling more widespread adoption of advanced segmentation techniques in diverse healthcare settings. Future research in this direction could lead to more robust, adaptable, and accessible medical image analysis tools, ultimately contributing to improved patient care and outcomes.

# References


[1] Hanan, Sabbar. et al. (2024). Overview of Medical Image Segmentation Techniques through Artificial Intelligence and Computer Vision. International Journal of Computing and Digital Systems. 15. 10.12785/ijcds/160183.

[2] Pinto-Coelho L. (2023). How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. Bioengineering (Basel, Switzerland), 10(12), 1435. https://doi.org/10.3390/bioengineering10121435

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[4] Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285.

[5] Fnu Neha et. al. (2024) U-Net in Medical Image Segmentation: A Review of Its Applications Across Modalities. arXiv preprint arXiv:2412.02242.

[6] Wang, T., Wen, Y. & Wang, Z (2024). nnU-Net based segmentation and 3D reconstruction of uterine fibroids with MRI images for HIFU surgery planning. BMC Med Imaging 24, 233 . https://doi.org/10.1186/s12880-024-01385-3

[7] Cardoso, M. J. et al. (2022). MONAI: An open-source framework for deep learning in healthcare. https://doi.org/https://doi.org/10.48550/arXiv.2211.02701

[8] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., ... & Kikinis, R. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magnetic resonance imaging, 30(9), 1323-1341.

[9] Antonelli, M., Reinke, A., et al. (2021). The Medical Segmentation Decathlon. Nature Communications, 13.

[10] Wenxuan Li, Chongyu Qu, et. al.(2024) AbdomenAtlas: A large-scale, detailed-annotated, & multi-center dataset for efficient transfer learning and open algorithmic benchmarking, Medical Image Analysis, Volume 97, 103285, ISSN 1361-8415, https://doi.org/10.1016/j.media.2024.103285.

[11] Liu, J., Zhang, Y., Chen, J., Xiao, J., Lu, Y., Landman, B.A., Yuan, Y., Yuille, A.L., Tang, Y., & Zhou, Z. (2023). CLIP-Driven Universal Model for Organ Segmentation and Tumor Detection. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 21095-21107.

[12] 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.