Term Project: Abdominal Organ Segmentation from CT Images

Shaghayegh Afshari[a]

[a]*Department of Biomechatronics Engineering, National Taiwan University, Taipei, Taiwan*

Abstract

Accurate segmentation of abdominal organs from CT images is crucial for medical diagnostics and treatment planning. Traditional manual delineation, while common, is laborious, time-consuming, and prone to errors. Deep learning-based medical image segmentation significantly enhances this process by improving accuracy and efficiency. This work utilizes 150 abdominal CT volumes from the Whole abdominal ORgan Dataset (WORD) dataset, focusing on the segmentation of 16 different organs. A state-of-the-art deep learning method employing the U-Net architecture was implemented. It was found that organ segmentation is effectively performed, achieving an average accuracy of 68% across all organs, with particularly high performance noted for larger organs. However, the high-resolution 3D images are computationally expensive, needing networks with more trainable parameters that increase GPU memory usage and need long inference times in the testing stage. Therefore, it was observed that for smaller, more complex structures, the utilization of a more powerful GPU would be beneficial, as it could significantly enhance the processing time and efficiency of the deep learning model.

Keywords: Abdominal organ segmentation, CT imaging, deep learning, U-Net architecture, GPU

1. Introduction

Segmentation of abdominal organs from CT images is essential for accurate medical diagnosis and effective treatment planning. Currently, this process often relies on manual segmentation by radiologists, which is time-intensive and prone to human error. This process demands careful

Email address: `d12631007@ntu.edu.tw` (Shaghayegh Afshari)

attention as annotators need to check each image slice one by one, often taking several hours for a single case. These limitations highlight the need for improved segmentation techniques. Automating this process with advanced computational methods can significantly enhance the accuracy and efficiency of medical diagnoses. This not only supports more precise treatment decisions but also alleviates the workload on medical professionals, potentially leading to better overall patient outcomes. Recently many deep learning-based models applied to various automatic medical image segmentation tasks with great success [1]. However, most of them focused on specific single organs, such as the liver [2], pancreas [3], and kidney [4]. Multi-organ segmentation remains difficult because Abdominal organs in different slices often differ greatly in size, shape, and position, especially elongated organs such as the pancreas. Individual differences in organs from different patients are significant, and the boundaries between different organs are often blurred. Addressing these challenges, Ronneberger et al. [5] designed an U-net, which symmetries the encoding and decoding processes and fuse shallow and deep features by adding skip connections from the encoder to the decoder. The U-net has achieved excellent performance and has become the most widely used benchmark framework. Numerous variants of the U-Net model have been proposed for various medical segmentation tasks.

1.1. Motivation and Relevant Works

Gibson et al. [6] proposed a DenseVNet to segment 8 organs from CT volumes, which enables high-resolution activation maps through memory-efficient dropout and feature reuse. Wang et al. [7] presented a novel framework for abdominal multi-organ segmentation using organ-attention networks with reverse connections and evaluated it on an in-house dataset. Liang et al. [8] combined the inter-and intra-patient deformation data augmentation with multiscale Attention-UNet [9] for accurate abdominal multi-organ segmentation. Tang et al. [10] introduced a batch-based approach combined with a random shifting strategy to enhance the performance of multi-organ segmentation from high-resolution abdominal CT volumes. Although these methods have been effective, they depend on having large datasets with detailed annotations. To cut down on annotation costs, Zhou et al. [11] proposed a co-training-based semi-supervised method for abdominal multi-organ segmentation, which reduces almost half of the annotation cost. Furthermore, Zhou et al. [12] introduced a prior-aware neural network that utilizes anatomical priors about the sizes of abdominal organs, enabling the training of models using several partially labeled datasets.

1.2. Contribution and Scope

In this study, two distinct UNet-based models were employed to explore the segmentation of 16 different organs in abdominal CT images, building upon previous research that focused on fewer organs. The research compares different frameworks to determine which one offers superior performance in terms of accuracy and efficiency. The contributions of this research are as follows: Firstly, the scope of organ segmentation was expanded to address more organs than typically studied, enhancing the clinical utility of the models and establishing a more comprehensive approach for such analyses. Secondly, by evaluating various model configurations and comparing different frameworks, more effective methods were identified that improve performance over others' work. These methods demonstrated potential for adaptation to other imaging modalities. This work focuses on developing segmentation models that are not only more precise but also efficient, making them suitable for integration into practical medical applications and aimed at enhancing diagnostic and treatment processes.

2. Datasets

This project employed the Whole abdominal ORgan Dataset (WORD) [13], which includes 150 high-resolution CT scans collected from 150 patients before the radiation therapy in a single center. All of them are scanned by a SIEMENS CT scanner without appearance enhancement. The clinical characteristics of the WORD dataset are listed in Table 1 Each CT volume consists of 159 to 330 slices of 512×512 pixels, with an in-plane resolution of 0.976 mm $\times$ 0.976 mm and slice spacing of 2.5 mm to 3.0 mm, indicating that the WORD dataset is a very high-resolution dataset. All scans of the WORD dataset are exhaustively annotated with 16 anatomical organs, including the liver, spleen, kidney(L), kidney (R), stomach, gallbladder, esophagus, duodenum, colon, intestine, adrenal, rectum, bladder, head of the femur (L), and head of the femur (R). An example of image and annotation from the WORD dataset is shown in Figure 1. WORD dataset Was randomly split into three parts: 100 scans (20 115 slices) for training, 20 scans (4103 slices) for validation, and 30 scans (6277 slices) for testing. Figure 2 shows the volume distributions of all annotated organs. It shows that the extremely unbalanced distribution among large and small organs may bring some challenges to the segmentation task.

Table 1: Clinical characteristics of WORD. Others include some metastatic tumors, such as bone metastasis and soft tissue metastasis.

Characteristics	Train (n = 100)	Validation (n = 20)	Test (n = 30)
Age (median)	47 (28–75)	52 (32–78)	49 (26–72)
Male	63	12	13
Female	37	8	17
Prostatic cancer	28	7	10
Cervical cancer	29	6	5
Rectal cancer	26	3	8
Others	17	4	7

3. Methodology

In this study, semantic segmentation of 16 abdominal organs was explored using two different approaches to leverage the U-Net architecture, comparing the effectiveness of the MONAI, an open-source PyTorch-based medical imaging framework, against a custom implementation in PyTorch. MONAI (Medical Open Network for AI) [14] is designed specifically for deep learning in healthcare imaging, offering tools for comprehensive image analysis tasks from preprocessing to model training, evaluation, and deployment. It excels in simplifying complex workflows in medical image segmentation, by providing pre-built components and algorithms for tasks including image transformation and segmentation. High-resolution abdominal CT images from the WORD dataset, annotated with distinct organ regions, formed the dataset utilized in this study. the image

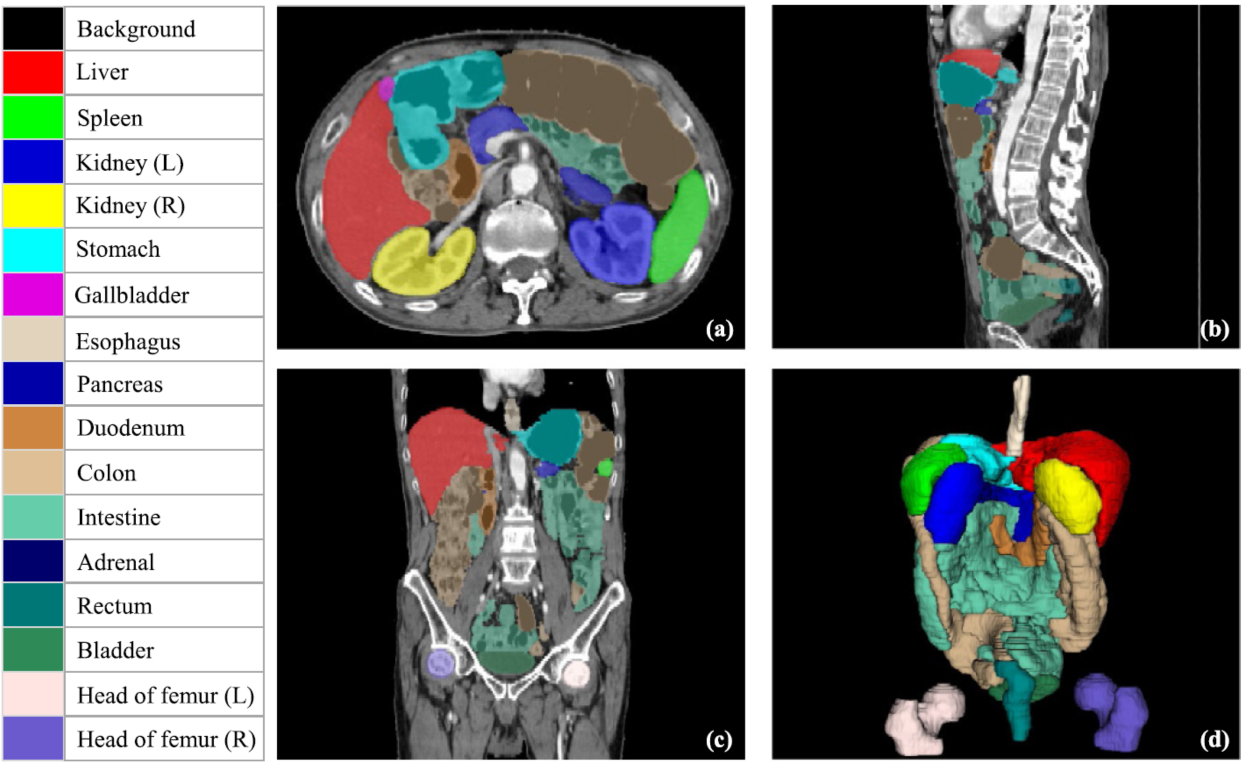

Figure 1: An example of 16 annotated abdominal organs in a CT scan. The left table lists the annotated organs' categories. (a), (b) and (c) denote the visualization in axial, coronal, and sagittal views, respectively. (d) represents the 3D rendering results of annotated abdomen organs.

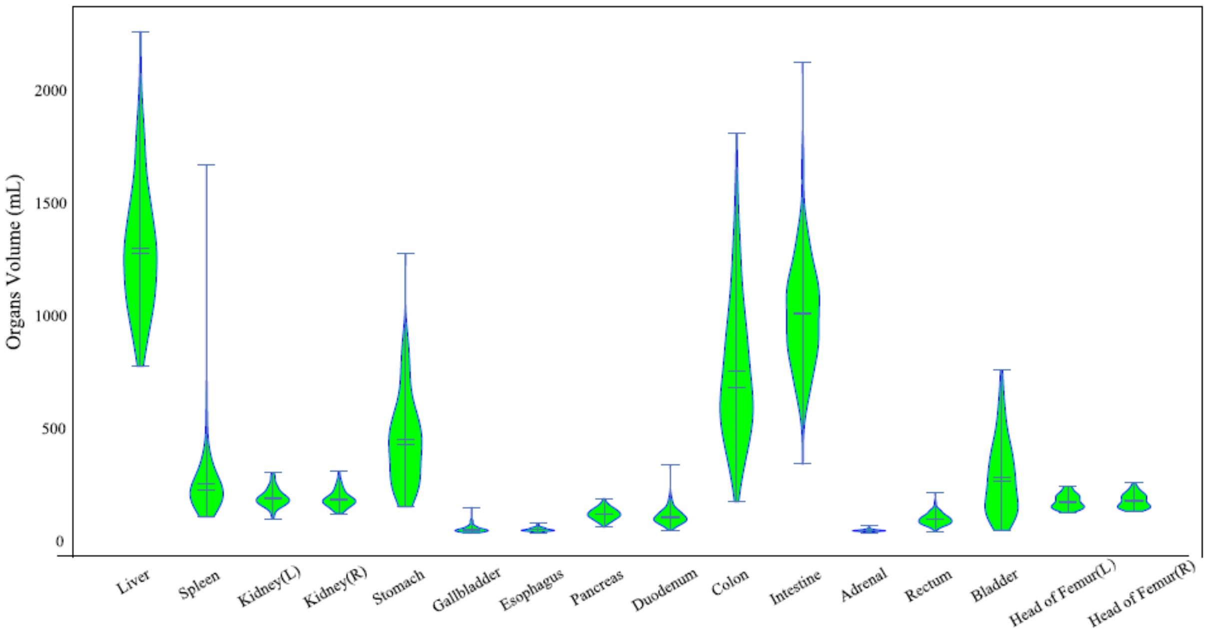

Figure 2: Volume distribution of 16 organs in WORD

data were randomly divided into training, validation, and test sets with counts of 100, 20, and 30, respectively, for both methodologies. In the MONAI-based approach, the preprocessing workflow utilized MONAI's transformation tools to prepare high-resolution abdominal CT images for segmentation. The images underwent normalization to zero mean and unit variance, accompanied by augmentations such as rotations, translations, and elastic deformations to improve model robustness against variations in organ positioning and morphology. Following preprocessing, the images were input into a sophisticated 3D U-Net model configured within the MONAI framework. This model was tailored for the complex task of segmenting multiple abdominal organs from CT scans. It featured three spatial dimensions to align with the three-dimensional nature of the CT images, and used a single input channel, typical of medical imaging data. The model was designed to output seventeen channels, one for each of the sixteen organs plus an additional channel for background segmentation. To effectively capture and synthesize features across multiple scales—a critical factor in accurately delineating complex organ structures—the U-Net's architecture was meticulously designed with progressive channels and strides. This setup allowed the model to handle the intricate spatial hierarchies and the close proximity of the abdominal organs. Training was conducted on an NVIDIA GeForce RTX 3090 GPU, selected for its ability to efficiently handle the large volumes of data typical in 3D medical imaging. The model was trained over 600 epochs using an Adam optimizer with a learning rate of 1×10^{-3} and weight decay of 1×10^{-5}, settings optimized to enhance learning efficiency and model performance. The Dice Loss function was employed for its effectiveness in multi-class segmentation, configured with one-hot encoding, softmax activation, and squared predictions to improve training stability and segmentation accuracy. These adjustments were essential for accurately delineating complex organ structures within the abdominal region.

In the custom PyTorch-based approach, preprocessing was executed using a Python script that resized, normalized, and padded high-resolution CT images and labels for uniformity and efficiency. The processed images and labels were then stored in HDF5 files to facilitate quick access during model training. This preparation ensured that the data was optimally conditioned for effective training of the neural network. Following the initial data preparation, the custom PyTorch-based approach utilized a more hands-on method for training the U-Net model. This approach involved loading the preprocessed data using a custom dataset handler, which ensured that the images and masks were appropriately formatted for network input. The neural network architecture, a variant of the 3D U-Net, was manually defined in PyTorch, providing flexibility to adjust architectural details such as the number of convolutional layers, filter sizes, and activation functions to better suit the specifics of the abdominal CT segmentation task. The training pipeline was explicitly controlled via a Python script, which managed the batching, epoch iterations, and backpropagation. In addition to implementing the Dice Loss function, the model also employed the DiceCE Loss function, which combines Dice Loss with cross-entropy. which was particularly effective at handling class imbalance—a common issue in medical image segmentation where some organs are significantly smaller or less frequently represented than others. This loss function facilitated the network's ability to learn from both the pixel-wise errors and the global distribution of classes, enhancing the segmentation performance across all organ types. The network was trained on a dual-GPU setup to enhance training efficiency through parallel processing. This configuration distributed the computational load across two GPUs—one for performing forward passes and computing loss, and the other for gradient calculations and backpropagation. This strategy not only accelerated the training process but also alleviated memory constraints associated with handling large 3D medical datasets. Model performance was continuously monitored through a custom logging system that recorded losses and metric improvements at each epoch.

4. Results and Discussions

The performance of the U-Net models, both MONAI-based and custom PyTorch-based, was assessed using the WORD dataset for semantic segmentation of 16 abdominal organs. The results underscore the models' segmentation capabilities and highlight differences in effectiveness

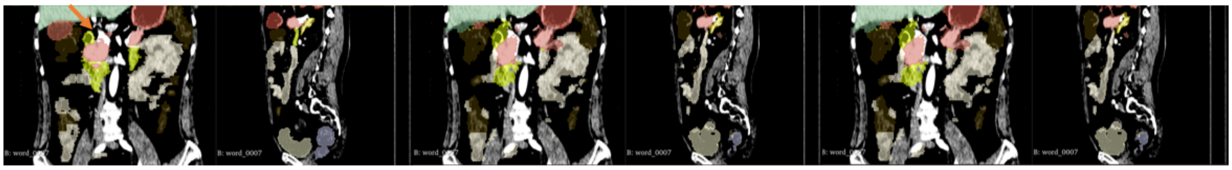

Figure 3: Comparative analysis of abdominal CT scan segmentation by the custom U-Net model. From left to right: Ground Truth, DiceCE Model Output, and Dice Loss Model Output, depicted in both coronal and sagittal views.

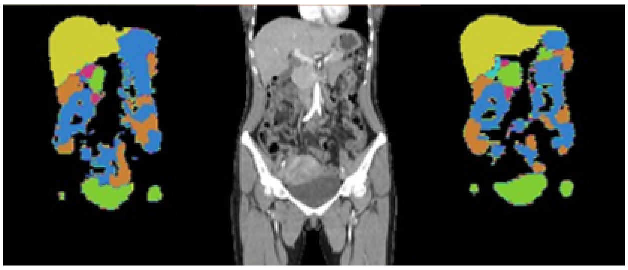

Figure 4: Segmentation of abdominal organs displayed in coronal view. From left to right: Ground Truth segmentation, CT Volume, and Predicted segmentation using the MONAI framework.

across various organ types. In assessing the performance of both U-Net models, Table 2 reveals a detailed comparison across 16 abdominal organs. Both models showed strong performance in segmenting larger organs like the liver, with the Custom U-Net achieving superior Dice scores under both Dice Loss and DiceCE metrics. Notably, the Custom U-Net demonstrated a significant improvement in segmenting complex organs such as the gallbladder and esophagus, which are traditionally challenging due to their smaller size and irregular shapes. This improvement underscores the efficacy of the Custom U-Net in handling varied organ geometries and the effectiveness of using a composite loss function, which seems to enhance sensitivity and accuracy in distinguishing organ boundaries. Figures3 and 4 visually demonstrate the segmentation accuracy of both models. These images, displayed using 3D Slicer [15], offer a direct comparison between the automated segmentations produced by the models and the actual ground truth labels. Notably, the visualizations underscore the enhanced edge detection in complex organs like the pancreas and liver, especially evident in the custom U-Net model.

Table 2: Comparison of segmentation performance using Dice scores for different U-Net configurations on the WORD dataset.

Organ	**MONAI U-Net Dice Loss**	**Custom U-Net Dice Loss**	**Custom U-Net DiceCE**
Liver	88.24	90.16	91.32
Spleen	81.33	84.55	86.01
L Kidney	80.99	82.08	84.56
R Kidney	80.05	81.06	84.62
Stomach	67.76	70.18	73.76
Gallbladder	19.54	27.97	29.07
Esophagus	36.16	53.72	55.23
Pancreas	48.14	53.85	53.09
Duodenum	38.54	42.59	39.08
Colon	56.17	56.86	61.35
Intestine	62.24	63.26	65.70
Adrenal	32.43	38.50	42.31
Rectum	53.64	59.14	58.83
Bladder	70.48	72.45	76.62
Head of L Femur	82.71	86.92	86.16
Head of R Femur	85.27	85.79	85.20
Average	61.48	65.57	67.07

5. Conclusion

In this study, the performance of U-Net-based models for segmenting abdominal organs from CT images was evaluated, highlighting their potential to replace labor- and time-intensive manual segmentation techniques. The research tested two approaches: one using the MONAI framework and another custom-built with PyTorch. The results showed that the custom PyTorch U-Net performed better with an average of 67% accuracy among 16 different organs. However, both methods struggled with segmenting smaller, complex organs. This issue is partly due to differences in CT scan quality and manual segmentations, which can vary widely. Also, the lack of powerful hardware affected the study's outcomes. With limited GPU memory, I had to use a batch size of one, which isn't ideal and can't match the performance seen in other studies with better resources. Furthermore, this study considered nnUNet [16], an innovative framework known for its automatic configuration adjustments tailored to specific segmentation tasks. Unfortunately, due to the same hardware limitations, it was impossible to train the nnUNet model effectively, which requires a minimum batch size of two for optimal performance. In conclusion, while U-Net-based models are promising for automating organ segmentation, their success is heavily influenced by the quality of input data and available computational resources.

6. Acknowledgements

I would like to express my gratitude to Professor Rih-Teng Wu for his guidance in the Deep Learning in Computer Vision course, which provided a comprehensive understanding of both the concepts and mathematics behind deep learning methods. This course deepened my understanding of the theory behind deep learning, beyond just its practical use. Additionally, I am thankful to our CIBI Lab for providing the GPU resources essential for conducting this research.

References

[1] Yunze Man, Yangsibo Huang, Junyi Feng, Xi Li, and Fei Wu. Deep q learning driven ct pancreas segmentation with geometry-aware u-net. IEEE Transactions on Medical Imaging, 38(8):1971–1980, 2019.

[2] Devidas T Kushnure and Sanjay N Talbar. Ms-unet: A multi-scale unet with feature recalibration approach for automatic liver and tumor segmentation in ct images. Computerized Medical Imaging and Graphics, 89:101885, 2021.

[3] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999, 2018.

[4] Geeta Rani, Priyam Thakkar, Akshat Verma, Vanshika Mehta, Rugved Chavan, Vijaypal Singh Dhaka, Ravindra Kumar Sharma, Eugenio Vocaturo, and Ester Zumpano. Kub-unet: segmentation of organs of urinary system from a kub x-ray image. Computer Methods and Programs in Biomedicine, 224:107031, 2022.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pages 234–241. Springer, 2015.

[6] Eli Gibson, Francesco Giganti, Yipeng Hu, Ester Bonmati, Steve Bandula, Kurinchi Gurusamy, Brian Davidson, Stephen P. Pereira, Matthew J. Clarkson, and Dean C. Barratt. Automatic multi-organ segmentation on abdominal ct with dense v-networks. IEEE Transactions on Medical Imaging, 37(8):1822–1834, 2018.

[7] Yan Wang, Yuyin Zhou, Wei Shen, Seyoun Park, Elliot K Fishman, and Alan L Yuille. Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Medical image analysis, 55:88–102, 2019.

[8] Xiaokun Liang, Na Li, Zhicheng Zhang, Jing Xiong, Shoujun Zhou, and Yaoqin Xie. Incorporating the hybrid deformable model for improving the performance of abdominal ct segmentation via multi-scale feature fusion network. Medical Image Analysis, 73:102156, 2021.

[9] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis, 53:197–207, 2019.

[10] Yucheng Tang, Riqiang Gao, Ho Hin Lee, Shizhong Han, Yunqiang Chen, Dashan Gao, Vishwesh Nath, Camilo Bermudez, Michael R Savona, Richard G Abramson, et al. High-resolution 3d abdominal segmentation with random patch network fusion. Medical image analysis, 69:101894, 2021.

[11] Yuyin Zhou, Yan Wang, Peng Tang, Song Bai, Wei Shen, Elliot Fishman, and Alan Yuille. Semi-supervised 3d abdominal multi-organ segmentation via deep multi-planar co-training. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 121–140. IEEE, 2019.

[12] Yuyin Zhou, Zhe Li, Song Bai, Chong Wang, Xinlei Chen, Mei Han, Elliot Fishman, and Alan L Yuille. Prior-aware neural network for partially-supervised multi-organ segmentation. In Proceedings of the IEEE/CVF international conference on computer vision, pages 10672–10681, 2019.

[13] Guotai Wanga and Shaoting Zhanga. Word: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from ct image. 2022.

[14] M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701, 2022.

[15] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. Magnetic resonance imaging, 30(9):1323–1341, 2012.

[16] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods, 18(2):203–211, 2021.