electronics

MDPI

Article

Cervical Intervertebral Disc Segmentation Based on Multi-Scale Information Fusion and Its Application

Yi Yang [1,†], Ming Wang [2,†], Litai Ma [1], Xiang Zhang [1], Kerui Zhang [1], Xiaoyao Zhao [2], Qizhi Teng [2,*] and Hao Liu [1,*]

[1] Department of Orthopedics, Orthopedic Research Institute, West China Hospital, Sichuan University, Chengdu 610041, China; hxyangyi@wchscu.edu.cn (Y.Y.); malitai@scu.edu.cn (L.M.); 2021224020102@stu.scu.edu.cn (X.Z.); zhangkerui@stu.scu.edu.cn (K.Z.)

[2] College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China; 2022222055213@stu.scu.edu.cn (M.W.); 2021222050073@stu.scu.edu.cn (X.Z.)

* Correspondence: qzteng@scu.edu.cn (Q.T.); liuhao6304@126.com (H.L.)

† These authors contributed equally to this work.

Abstract: The cervical intervertebral disc, a cushion-like element between the vertebrae, plays a critical role in spinal health. Investigating how to segment these discs is crucial for identifying abnormalities in cervical conditions. This paper introduces a novel approach for segmenting cervical intervertebral discs, utilizing a framework based on multi-scale information fusion. Central to this approach is the integration of multi-level features, both low and high, through an encoding–decoding process, combined with multi-scale semantic fusion, to progressively refine the extraction of segmentation characteristics. The multi-scale semantic fusion aspect of this framework is divided into two phases: one leveraging convolution for scale interaction and the other utilizing pooling. This dual-phase method markedly improves segmentation accuracy. Facing a shortage of datasets for cervical disc segmentation, we have developed a new dataset tailored for this purpose, which includes interpolation between layers to resolve disparities in pixel spacing along the longitudinal and transverse axes in CT image sequences. This dataset is good for advancing cervical disc segmentation studies. Our experimental findings demonstrate that our network model not only achieves good segmentation accuracy on human cervical intervertebral discs but is also highly effective for three-dimensional reconstruction and printing applications. The dataset will be publicly available soon.

Keywords: cervical intervertebral disc segmentation; deep learning; multi-scale information fusion

Citation: Yang, Y.; Wang, M.; Ma, L.; Zhang, X.; Zhang, K.; Zhao, X.; Teng, Q.; Liu, H. Cervical Intervertebral Disc Segmentation Based on Multi-Scale Information Fusion and Its Application. *Electronics* **2024**, *13*, 432. https://doi.org/10.3390/electronics13020432

Academic Editor: Heung-Il Suk

Received: 31 December 2023
Revised: 14 January 2024
Accepted: 15 January 2024
Published: 20 January 2024

1. Introduction

Computer tomography (CT) images, captured through cross-sectional body scans, enable physicians to promptly identify and address patient lesions. Notably, the prevalence of cervical spondylosis in younger individuals, primarily caused by the degeneration of cervical intervertebral discs, has been on the rise [1]. By examining CT images of the cervical spine, doctors can ascertain the presence of cervical intervertebral disc lesions and consider interventions like artificial disc replacement. Segmenting these cervical intervertebral discs and constructing their three-dimensional (3D) models from the images can give doctors a more detailed understanding of the disc structures, thus potentially decreasing surgical risks. However, the segmentation process is complicated due to the dense, fibrous nature of the cervical intervertebral discs, which display grayscale values on CT images similar to those of tissue fluid.

Traditional image segmentation is mainly based on features such as image gray-level and contours to segment the target regions [2]. For example, the commonly used threshold segmentation method [2] segments the target based on the image gray-level value. This method is easy to implement, but it requires manual determination of the threshold and the segmentation accuracy is not high. The fuzzy C-means clustering method [3] obtains several sample points for all class centers by optimizing the objective function, and selects

the one with the smallest distance to the clustering center as the belonging category. These traditional image algorithms generally require some post-processing methods for CT images to refine the segmentation results, such as basic morphological operations: filling holes, erosion, dilation, and noise reduction.

With the development of computer hardware, there has been increasing research on using deep learning to segment the regions of interest in medical images [4–19]. For instance, UNet [4] utilizes an encoder–decoder structure with skip connections to fuse features at different scales, making it widely applicable in medical image segmentation. Meanwhile, medical images are typically composed of a sequence of images. A series of images can be converted into 3D for analysis. Some researchers have used 3D image segmentation to classify targets, such as V-Net [5], which has a similar overall model structure as UNet and utilizes Resnet [6] to address the problem of network degradation. However, these networks are not specifically tailored for cervical intervertebral disc segmentation, highlighting a need for further research. In addition, there is a lack of comprehensive datasets in the cervical intervertebral disc segmentation area.

To address the previous issues, a multi-scale information fusion framework for segmentation is proposed and effective datasets are constructed for cervical intervertebral disc segmentation. A brief illustration of our method is given in Figure 1. The main contributions of this paper are as follows:

1. A multi-scale information fusion framework for segmentation is proposed. This framework consists of multi-scale low–high level feature encoding–decoding fusion module and multi-scale semantic fusion module, with the use of adjacent layer information assisted segmentation strategy. Building upon the conventional hierarchical encoding–decoding framework, it further merges mid-layer semantic information at multiple scales, achieving progressive precision extraction of segmentation features. This framework demonstrates exceptional performance in intervertebral disc segmentation tasks and is effectively applicable to 3D reconstruction and 3D printing.
2. An effective multi-scale semantic fusion module is introduced, which can be further divided into two stages: scale interaction based on convolution and scale interaction based on pooling. This two-stage high-precision fusion method significantly enhances the final segmentation performance.
3. Datasets specifically aimed at cervical intervertebral disc segmentation are developed. The proposed datasets incorporate inter-layer interpolation to address the inconsistency in longitudinal and transverse pixel spacing in CT sequence images. By selecting frames with prominent intervertebral disc regions through data significance selection and then constructing data groups where three consecutive layer images correspond to one label via manual annotation, the datasets provide important support for research in cervical disc segmentation.

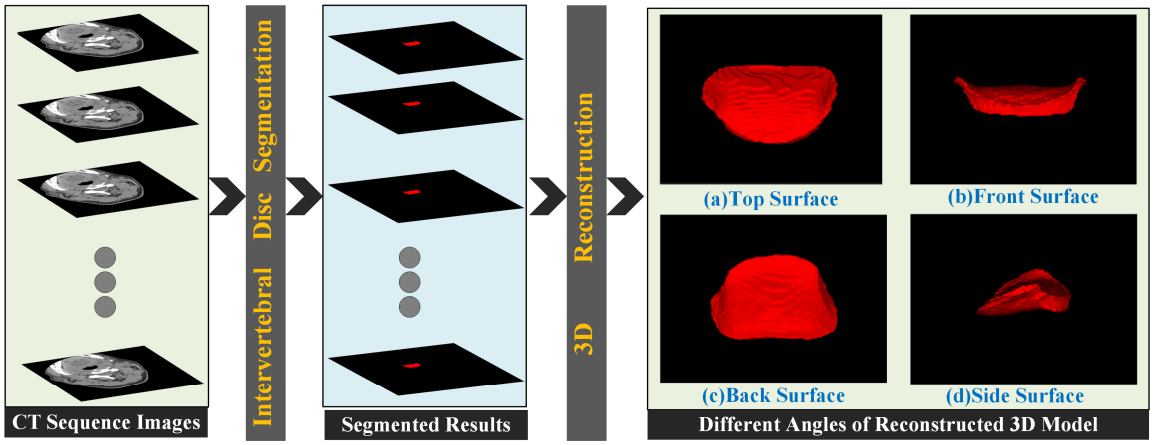


Figure 1. Different angles of the three-dimensional model of cervical intervertebral disc segmented by the network in this article.

2. Related Work

Image segmentation is one of the most important parts in image research. By extracting the regions of interest in an image, it can be used in many fields such as medical image analysis, video surveillance, and autonomous driving. Traditional image segmentation algorithms utilize information such as the grayscale, edges, and shapes of the objects to be segmented in the image to complete the segmentation task. Traditional segmentation algorithms require manual interaction to improve segmentation accuracy, and during execution, they need to continuously iterate to find the optimal solution, which generally requires a long computation time.

In recent years, with the continuous development of computer hardware technology, the development of image segmentation algorithms based on deep learning has been accelerating [7–14]. Long propose the fully convolutional network [20], which applies convolutional neural networks to the field of image segmentation. This network extracts key features of the image using convolutional kernels, and then uses these features to complete the image segmentation task. Ronneberger propose the UNet network [4], which uses a symmetric encoding–decoding structure and connects low-level image features with high-level image features through skip connections. PSPNet [21] uses ResNet [6] as the backbone network and proposed a pyramid pooling module to further expand the receptive field. Different resolution image features are upsampled to the same size for fusion. The pyramid pooling module improves the segmentation ability of the network. BiseNet [22] divides image features into semantic features and contextual features. Semantic features have high image resolution and contain detailed features such as edges and textures, while contextual features are highly abstract features obtained by repeatedly extracting features from the image. Semantic features and contextual features are fused using a feature fusion module to complete the final image segmentation task. Deeplabv3 [23] uses Xception [24] as the backbone and used depth-wise separable convolutions to reduce the number of model parameters. The network also uses the Atrous Spatial Pyramid Pooling (ASPP) module to improve the receptive field and better extract image features. ASPP improves the multi-scale feature extraction ability through dilated convolutions of different kernel sizes. HRNet [25] fuses different scales of image features continuously to complete the image segmentation task, in contrast to the previous networks that always first reduced the resolution and then increased it. The popularity of Transformer in image segmentation has promoted the development of image segmentation [26–32]. Google proposes the vision transformer [33], which decomposes images into several image blocks and uses these image blocks as input to the transformer, greatly improving the performance of semantic segmentation. Chen combines Transform er with Unet network [34], fully utilizing the local feature extraction ability of convolutional kernels and the self-attention mechanism of transform. Segformer [35] uses a layered Transformer encoder to obtain coarse segmentation features with high resolution and high-quality features with low resolution, and then designs an MLP decoder to fuse multi-level features to complete semantic segmentation. UNet-2022 [36] designs an encoder network by parallelizing self-attention and convolution and uses parallel non-isomorphic blocks to enhance the ability to extract image features. Wan propose SeaFormer [37], which is a lightweight semantic segmentation network that balances efficiency while ensuring segmentation accuracy. Yuan propose an effective CNN and Transformer complementary network for medical image segmentation, and achieve good performance [38]. Recently, the release of SAM has pushed the boundaries of segmentation and greatly contributed to the development of basic models for computer vision [39].

The aforementioned deep learning-based image segmentation methods have achieved good experimental results in their respective tasks. However, there is room for improvement in their segmentation capabilities, especially through specific enhancements tailored to the needs of particular applications to further boost performance. Additionally, the scarcity of datasets specifically for cervical intervertebral disc segmentation also poses challenges for practical research in this area. In summary, the aforementioned issues urgently need to be addressed.

3. Proposed Method

3.1. Overview of Proposed Method

To address the previous issues, we propose a new framework as shown in Figure 2. To the scarcity of datasets specifically for cervical intervertebral disc segmentation, we construct datasets via inter-layer interpolation, significant data selection, and manual label construction. More details are provided in Section 3.5.

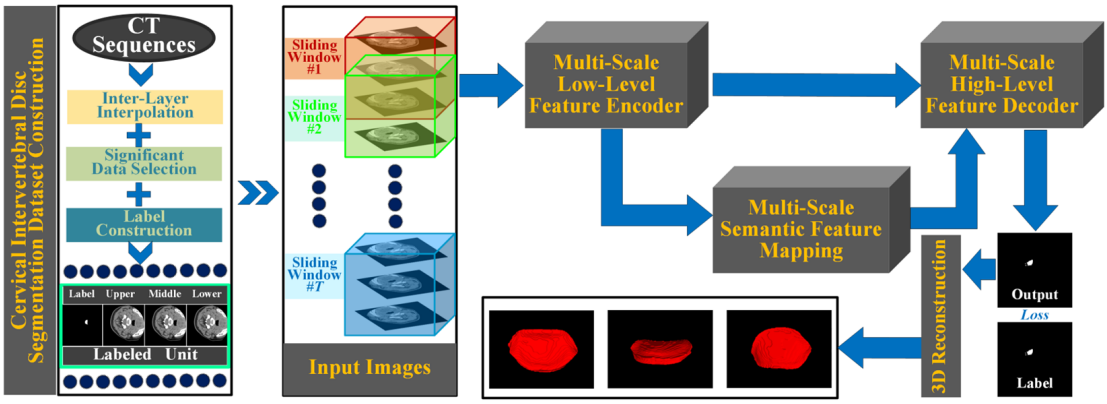


Figure 2. The framework of the proposed method, with cervical intervertebral disc segmentation dataset construction, cervical intervertebral disc segmentation network, 3D reconstruction.

To improve the performance of segmentation for cervical intervertebral disc, we propose an effective network with multi-scale information fusion. We can see from Figure 2 that the input of the network is not a single image, instead, adjacent layer images are used as input with a sliding window of with a width of 3 frams to guide the segmentation results of the middle layer image. Then, the input images are input into the multi-scale low-level feature encoder to extract the low-level feature. After that, the feature is input into the multi-scale semantic feature mapping module for further segmentation semantic feature extraction. Based on the encoding–decoding framework, the multi-scale semantic feature mapping module further merges mid-layer semantic information at multiple scales, achieving progressive precision extraction of the segmentation features. Finally, the semantic feature is projected into pixel space to obtain the predicted segmented result via the multi-scale high-level feature decoder module.

For training the proposed network, the binary cross-entropy (BCE) function was used as the loss function for evaluating the similarity between the label value and the predicted value in the semantic segmentation model. The BCE function is expressed as follows, where p represents the label value and q represents the predicted value:

$$Loss = -p \times \log(q) - (1 - p) \times \log(1 - q) \quad (1)$$

After the model is trained, the test CT sequence can be input into the network with sliding window scheme, and the segmented results of all images can be obtained. Thus, we can further obtain the 3D reconstructed cervical intervertebral disc with the segmented results.

3.2. Adjacent Layer Information Assisted Segmentation

Due to the identical gray information between cervical intervertebral discs and tissue fluids in CT images in Figure 3, it is difficult to distinguish the intervertebral disc region from the surrounding fluid region. For better visualization, we have circled the intervertebral disc area in Figure 3. Using only one image for intervertebral disc segmentation provides limited information and cannot accurately locate the external contour edge points of the intervertebral disc region. However, after CT slice scanning, the human cervical spine CT sequence image is composed of a series of horizontal plane images. Each image, along with its two adjacent images, not only reflects the change trend of the cervical intervertebral

disc but also embodies the overall morphological contour characteristics of the cervical intervertebral disc as a whole. The shape of the intervertebral disc gradually changes in the adjacent three images, and these external contour edge points can be located by using the adjacent images. If the network can learn these gradually changing external contour features, it will be more helpful for accurate intervertebral disc segmentation. Therefore, this study inputs three adjacent images into the network, and the segmentation result of the middle layer image is used as the output for the intervertebral disc segmentation task.

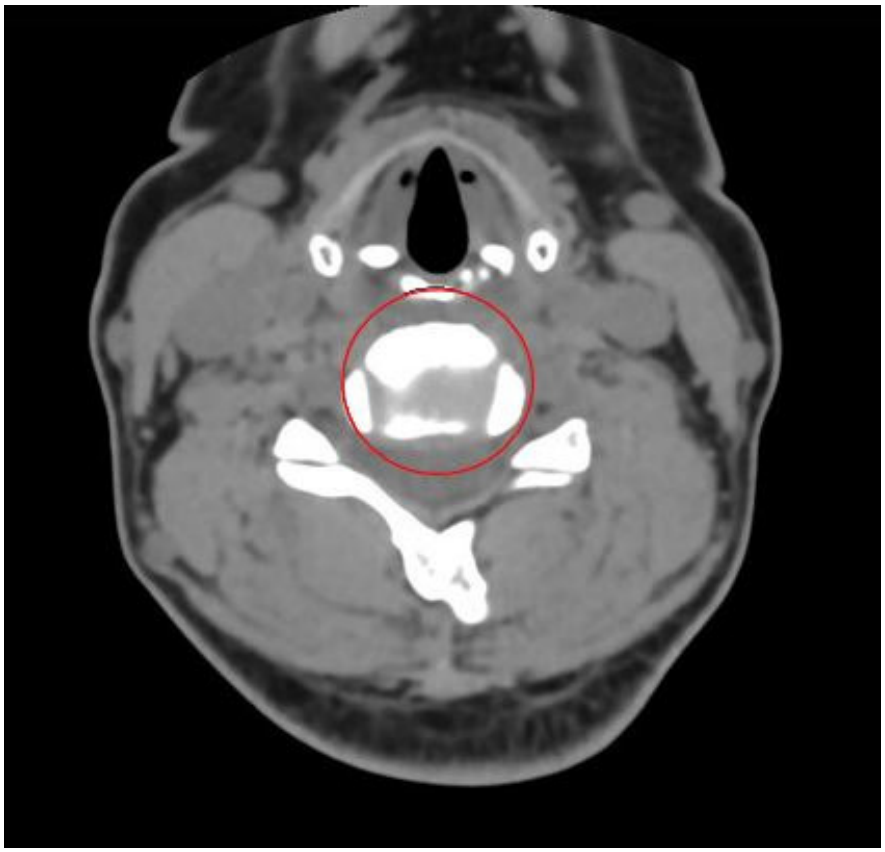


Figure 3. CT image of the cervical spine.

3.3. Multi-Scale Low–High Level Feature Encoding–Decoding Module

To extract the multi-scale feature from input images and project the mapped feature into a segmentation result, the multi-scale low–high level feature encoding–decoding module is adopted. As shown in Figure 4, the multi-scale low-level feature encoder consists of three sequent "CBR + CBR" blocks (CBR means 3 × 3 Conv layer, BatchNorm layer, and ReLU layer) of different scales. The Down layer is used after each CBR block to obtain the coarse-level feature with a larger perceptive field. Specifically, the Down layer is implemented by a 3 × 3 Conv layer with stride 2. Similarly, the multi-scale high-level feature encoder also consists of three sequent "CBR + CBR" blocks of different scales. The UP layer is used before each CBR block to obtain the fine-level feature with a smaller perceptive field. Specifically, the UP layer is implemented by a Bilinear interpolation layer with factor 2. Two long skip connections are used to achieve residual operations between low- and high-level features.

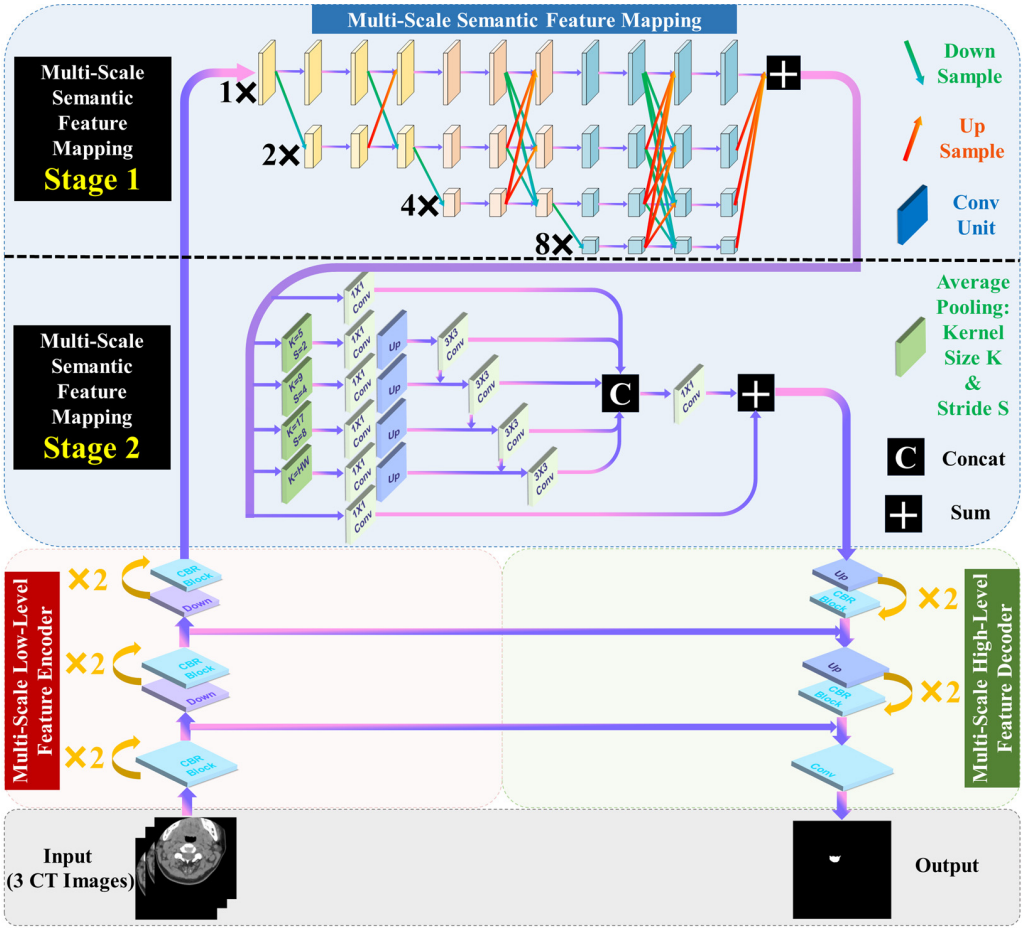


Figure 4. Multi-scale information fusion network for cervical intervertebral disc segmentation.

3.4. Multi-Scale Semantic Feature Mapping Module

Since the feature extracted by the multi-scale low-level feature encoder is relatively coarse, finer features still need to be extracted. Thus, an effective multi-scale semantic fusion module is proposed in this section, which can be further divided into two stages. This two-stage fusion scheme significantly enhances the final segmentation performance.

Multi-scale semantic feature mapping module, Stage 1: The segmentation of the cervical intervertebral disc needs to identify the target area in the original image, and the semantic information can enrich the edge structural features of the image, making the segmentation result more precise. By using the Stage 1 module, we can achieve feature extraction by continuously integrating information from different scales by convolutional operators, while maintaining high-resolution feature maps, which is inspired by [11]. During the feature extraction process, it continuously fuses and exchanges information between feature maps of different scales. High-resolution features enable more precise segmentation of boundaries and contours, while low-resolution features provide more abundant contextual information.

Specifically, as shown in Figure 4, the Stage 1 module is used to obtain various 1×, 2×, 4×, and 8× down-sampled features, and fuse them via at the end of each scale via up-sample and down-sample operators. The down-sample is implemented by a 3 × 3 Conv layer with stride 1 or 2 or 4 or 8, and the up-sample is implemented by a Bilinear interpolation layer with the factor 1 or 2 or 4 or 8. The Conv unit is implemented by "3 × 3 Conv layer + ReLU + 3 × 3 Conv layer" for extracting features. Finally, features of all scales are summed together via up-sample operators.

Multi-scale semantic feature mapping module, Stage 2: In order to further increase the receptive field and extract features of different scales, this paper uses the Stage 2 module,

which can further fuse multi-scale contextual information. Different from Stage 1 which uses convolutional operators for obtaining multi-scales, Stage 2 use average pooling operators (the global average pooling operator is also included) inspired by [14], which can largely increase the receptive field. Specifically, this module extracts features of different scales using different sizes of pooling kernels, adjusts the channel number through 1×1 convolutions, and performs up-sampling via Bilinear interpolation layers. Finally, all scale features are concatenated and convolved to obtain the feature fusion map of different scales, as shown in Figure 4.

3.5. Construction of Cervical Intervertebral Discs Segmentation Datasets

The human cervical spine constitutes a complex integrated structure. In the medical imaging processing of cervical intervertebral discs, using adjacent layer images for auxiliary segmentation not only can significantly enhance the interlayer continuity of the segmentation results but also plays a crucial role in accurately determining the upper and lower boundaries of the cervical discs. Based on this, our study proposes an effective strategy for creating segmentation datasets for cervical intervertebral discs, which involves using three adjacent layer CT images (i.e., upper (($t-1$)-th layer), middle (t-th layer), and lower (($t+1$)-th layer) images) to predict the single label of the middle image. This process is illustrated in Figure 5.

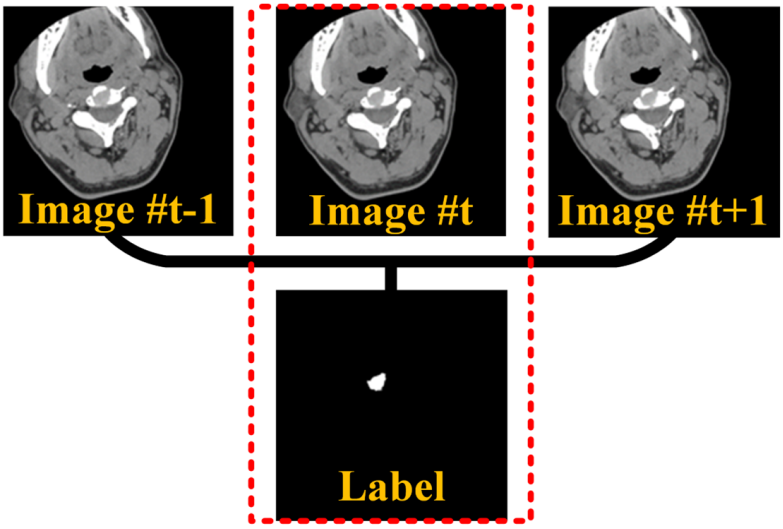
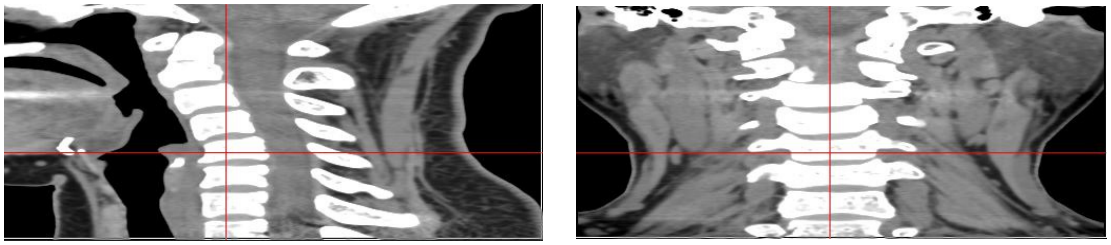


Figure 5. Illustration of the cervical intervertebral disc's segmentation dataset, where every three adjacent layer images (($t-1$)-th layer, t-th layer, ($t+1$)-th layer) group corresponds to one label. The red mark indicates "the single label of the middle image." in the text.

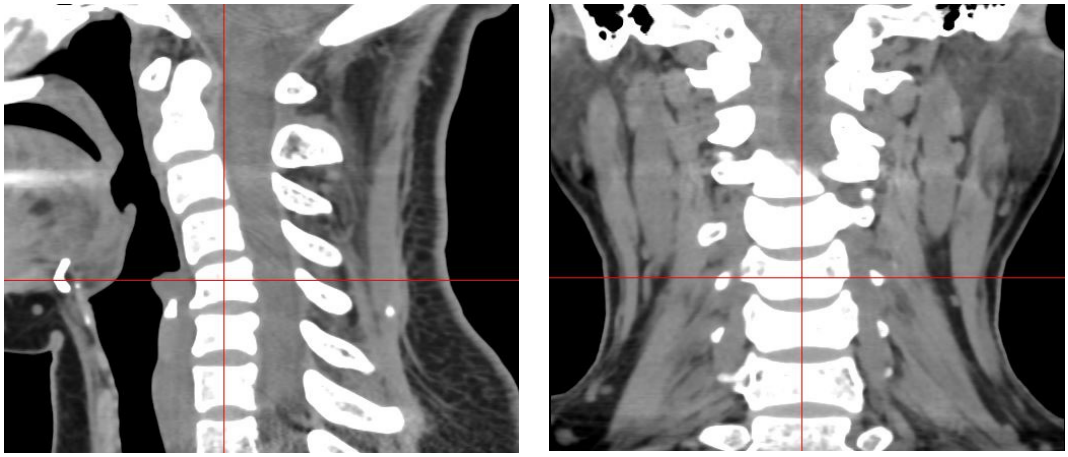
Due to the performance limitations of CT scanners' detectors, when scanning human tissues with the same radiation dose, the thinner the slice thickness of the CT images, the greater the image noise. In addition, the slice thickness of a CT image essentially corresponds to the longitudinal pixel spacing between sequence images. To reduce radiation dosage and image noise, usually, the longitudinal pixel spacing in CT sequence images is significantly larger than the transverse pixel spacing within each slice. If left unaddressed, this disparity in pixel spacing can lead to noticeable deformations of the cervical spine structure in sagittal and coronal plane images, deviating from the spine's actual anatomy. The deformed cervical spine structure appears overly flattened in the images (as shown in Figure 6), adversely affecting the segmentation of cervical vertebrae and intervertebral discs, as well as their subsequent three-dimensional (3D) reconstruction.



(**a**) Sagittal plane image (**b**) Coronal image

Figure 6. Deformed sagittal and coronal images of human cervical spine.

To resolve this issue, we propose using an interpolation method [15] for inter-slice interpolation of human cervical spine CT sequence images, which not only reduces the difference in the length of the horizontal and vertical points but also reduces the drastic changes in the adjacent layers. The interpolated image is shown in Figure 7, and the image information in Figure 7 is more detailed than that of Figure 6 without interpolation. The cervical spine CT sequence images, post-interpolation, more closely resemble the true structure of the human cervical spine in sagittal and coronal views. Subsequently, the vertebrae and intervertebral discs in the processed images are segmented and 3D reconstructed. This approach leads to richer morphological information of the cervical vertebrae and discs, facilitating more detailed studies of the morphology of the cervical endplates.



(**a**) Sagittal plane image (**b**) Coronal image

Figure 7. Interpolated images of the sagittal and coronal planes of the human cervical spine.

More specifically, in the segmentation of cervical intervertebral discs in human cervical spine CT images, considering that the cervical intervertebral discs are located only in the gaps between two cervical vertebrae, we initially selected CT images that prominently feature cervical intervertebral discs. Subsequently, accurate labels were created manually by professionals. Our work produces labels for the cervical intervertebral discs of 100 sets of patient data. Each set consists of CT images of uniform size of 512 $\times$ 512, and every three adjacent layer images are combined with one label, forming a data group. The dataset is divided into training, validation, and test sets, in a 6:2:2 ratio. Specifically, the training dataset comprises 7920 image pairs, the validation dataset 2167 pairs, and the test dataset 2338 pairs (general test dataset). The entire dataset construction process is referenced in Figure 8.

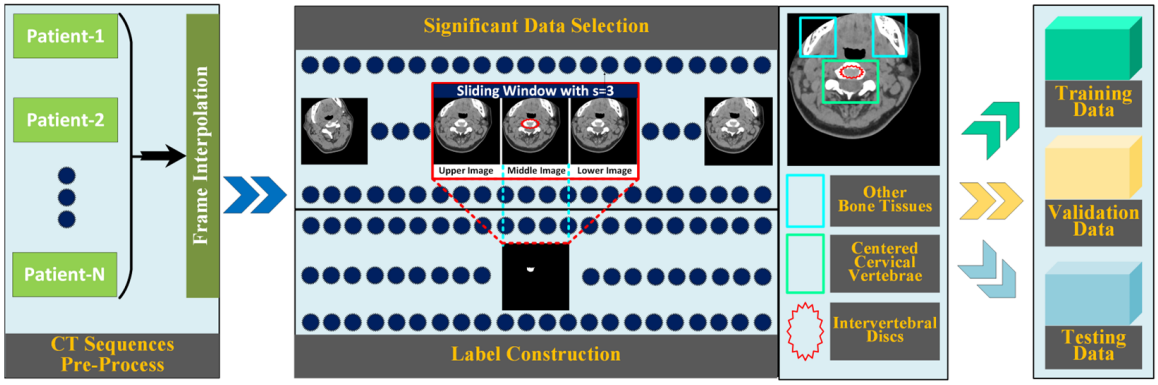


Figure 8. Illustration of cervical intervertebral disc segmentation dataset.

4. Experiments

4.1. Experimental Settings

In the experiments, the operating environment is Ubuntu 20.04, the CPU is Intel(R) Core(TM) i7-9700 with 32 GB of memory, and the GPU is NVIDIA RTX2080Ti with 11 GB of display memory.

The network is built on the PyTorch 1.7.0 platform and trained on the constructed cervical intervertebral disc dataset. The learning rate is set to 0.0001 and the batch size is set to four. The ADAM optimizer is used for gradient backpropagation to optimize the network model parameters, and the network can converge after about 30 epochs during training.

During testing, we compare the proposed method with other typical methods like PSPNet, Deeplabv3, HRNet, and UNet-2022 on the testing dataset. For further validation on the generalization of our method, we compare our method with other baselines on a different testing dataset called T1500.

4.2. Evaluation Indicators

To objectively evaluate the segmentation results of cervical intervertebral discs, the intersection over union (*IOU*) and the Dice similarity coefficient (*DICE*) are used in this study. FN is an area "labeled as true but predicted as false". TP is the area "labeled as true and predicted as true". FP is an area where "the label is false but the prediction is true". TN is an area where "the label is false and the prediction is false". These are illustrated in Figure 9. In order to measure the model size of the proposed network in this paper, the parameter number (Params) is introduced.

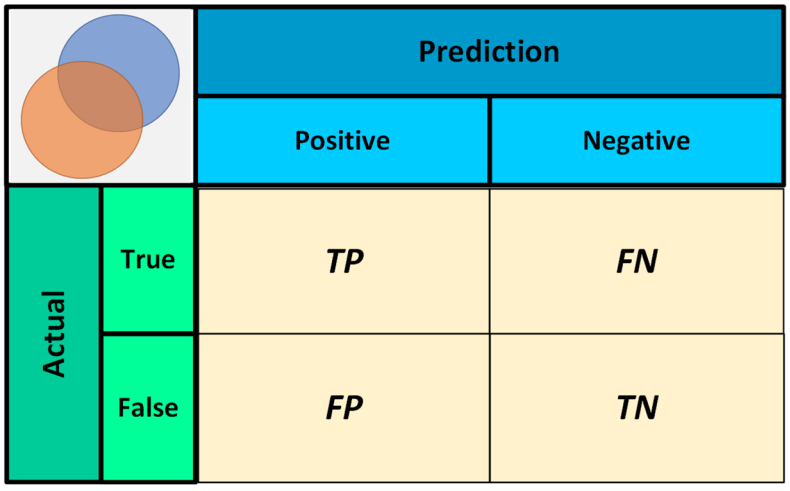


Figure 9. Schematic diagram of FN, TP, FP, and TN.

IOU can be represented by the following equation:

$$IOU = \frac{TP}{TP + FN + FP} \tag{2}$$

The DICE similarity coefficient is represented by the following equation:

$$DICE = \frac{2 \times TP}{(TP + FN) + (TP + FP)} \tag{3}$$

4.3. Comparison with Other Methods

In order to verify the effectiveness of the proposed method, comparative experiments are conducted with current mainstream segmentation algorithms. The network models in the compared methods are trained on the dataset constructed in this study and tested on the constructed test dataset. The comparison methods in this study include PSPNet [15], Deeplabv3 [17], HRNet [19], and UNet-2022 [30]. The segmentation results of cervical intervertebral discs are shown in Table 1, and Figure 10.

Table 1. Experimental results of different networks on the general test dataset.

Network	Index		
	IOU (%)	Dice (%)	Params (M)
PSPNet [16]	59.22	79.61	46.70
Deeplabv3 [18]	70.00	78.99	54.71
HRNet [20]	70.93	79.61	63.59
UNet-2022 [31]	69.06	79.16	41.90
Ours	73.63	82.98	63.83

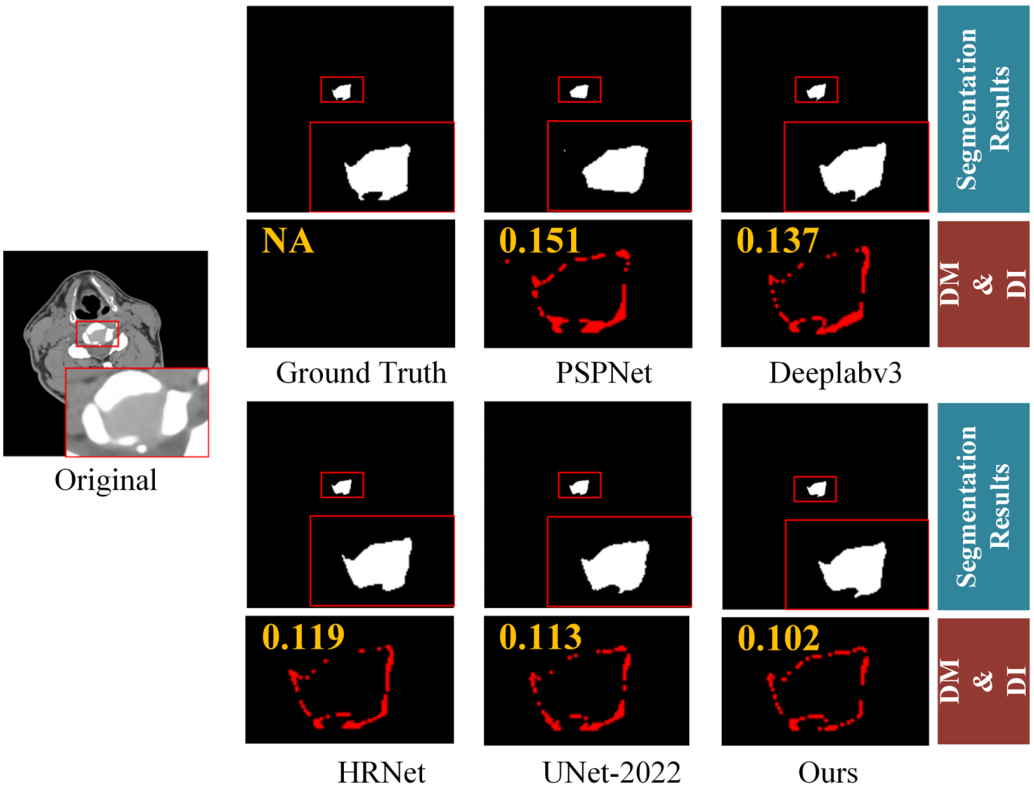


Figure 10. Cervical disc segmentation results of different methods.

As shown in Table 1, the proposed model outperforms the other compared methods in three aspects: IOU, DICE coefficients and Params. For example, compared with HRNet,

the parameter numbers are similar, i.e., 63.83 M vs. 63.59 M, but our method still achieves higher objective results. For the other methods, e.g., Deeplabv3, although it has slightly smaller parameter number, its IOU and Dice indices are much lower. In addition, according to Figure 10, our method achieves better visual segmentation results. For better visualization, we also provide the difference map (*DM*) and difference indicator (DI). *DM* is calculated by:

$$I_{DM} = \left| M_{GT} - M_{Seg} \right| \tag{4}$$

where I_{DM} is the *DM*, M_{GT} is the ground-truth segmentation label, and M_{Seg} is the predicted segmentation label.

DI is calculated by:

$$DI = \frac{SUM(I_{DM})}{SUM(M_{GT})} \tag{5}$$

where SUM is the summation function. This formula calculates the ratio of the difference area between the segmentation result and the ground-truth segmentation label to the total area of the ground-truth segmentation label. It is used to determine the degree of difference between the segmentation result and the ground-truth segmentation label. The larger the DI value, the greater the difference between the two, and the worse the segmentation result, and vice versa.

Both DM and DI also show that the proposed method can obtain better segmentation results with lower difference to the ground-truth label. This is because the adjacent three-layer images increase the interlayer constraint of cervical disc segmentation, and the multi-scale information extraction and fusion module improves the details of cervical discs, and thus enhances the accuracy of cervical disc segmentation.

At the same time, this paper also uses the marching cubes (MC) algorithm to perform iterative reconstruction of cervical intervertebral disc segmentation results from this network, and the 3D reconstruction results of two samples of different surfaces are shown in Figure 11. MC is a representation of 3D objects and is used for volume painting or surface reconstruction. It is to rasterize a 3D object and then use cubes (voxels) to represent it. The details are explained later in Section 4.6. We can observe that the 3D models are reasonable and can well reflect 3D shapes.

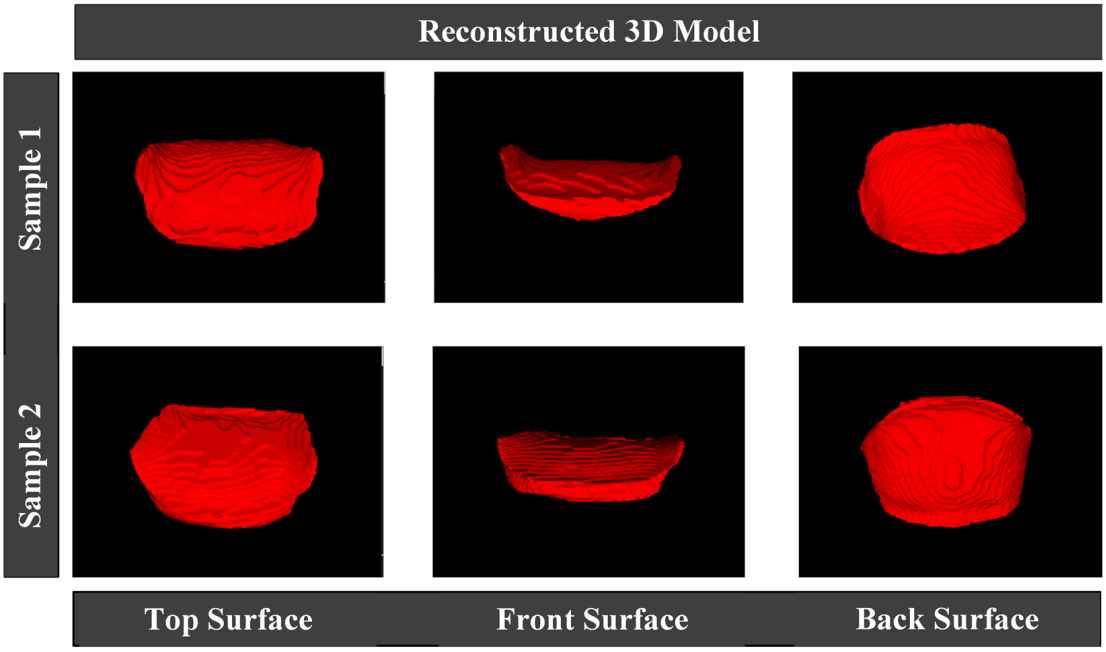


Figure 11. Each row represents a 3D model of a cervical intervertebral disc.

4.4. Generalization to Other Dataset

To further validate the effectiveness of our proposed method, we conduct tests on a new intervertebral disc segmentation test dataset using the previously trained model in Section 4.3. The construction of this test dataset is as follows: CT sequence images of cervical spines from 30 additional patients are selected to create the dataset. Each image size is 512 × 512 pixels, and every three adjacent images are paired with one label, forming a group. There are a total of 1500 sample groups, and we call it T1500 dataset. The test results are shown in Table 2. The performance on this test dataset still surpasses that of comparative methods, demonstrating the strong generalizability of the proposed approach. The segmentation results on the new testing dataset are displayed in Table 2 and Figure 12.

Table 2. Experimental results of different networks on the T1500 dataset.

Network	Index		
	IOU (%)	Dice (%)	Params (M)
PSPNet [16]	58.77	67.30	46.70
Deeplabv3 [18]	69.94	77.53	54.71
HRNet [20]	72.39	80.11	63.59
UNet-2022 [31]	68.93	76.95	41.90
Ours	73.67	81.07	63.83

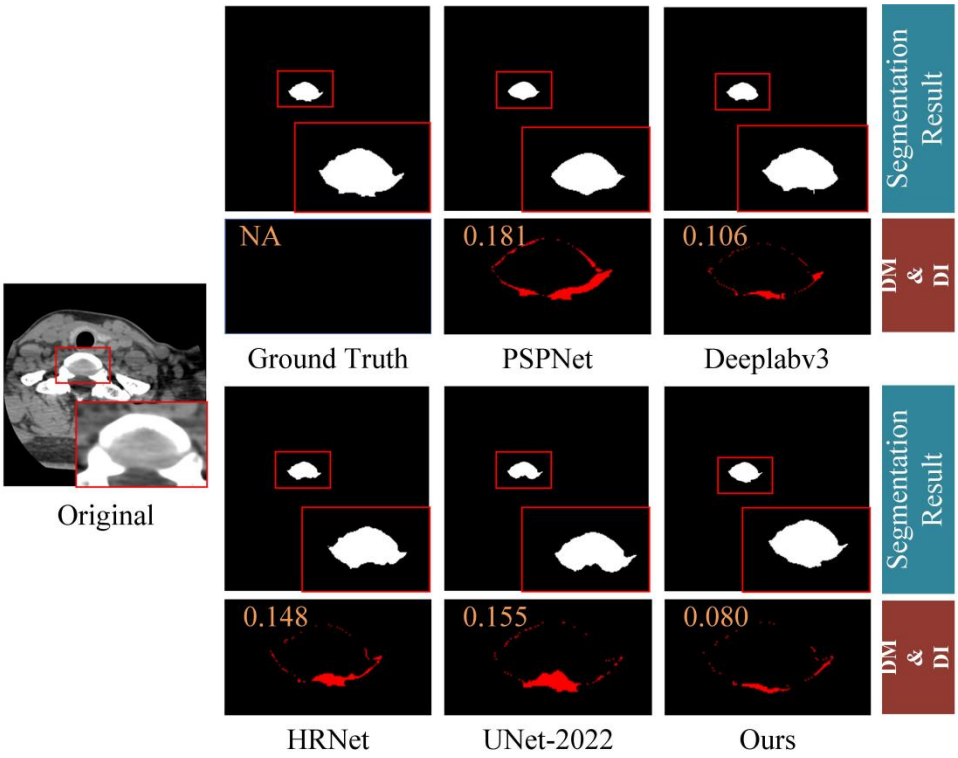


Figure 12. Schematic diagram of cervical disc segmentation results.

4.5. Ablation Experiments

In order to verify the effectiveness of the adjacent layer information aided segmentation, multi-scale low–high level feature encoder–decoder, and multi-scale semantic information mapping module proposed in this paper, this paper compares four variants, including Model0 (without multi-scale low–high level feature encoder–decoder, and without Stage 2), Model1 (without multi-scale low–high level feature encoder–decoder), Model2 (without adjacent layer information assisted segmentation), Model3 (Ours). The results obtained on the test dataset are shown in Table 3. The results show that by removing adjacent layer information assisted segmentation, i.e., only using one image as input, the IOU and Dice will decrease by 0.28 and

0.32, respectively. By removing multi-scale low–high level feature encoder–decoder, the IOU and Dice will decrease by 1.05 and 0.71, respectively. By removing multi-scale low–high level feature encoder–decoder and without Stage 2, the IOU and Dice will significantly decrease by 2.70 and 3.37, respectively. These results verify that the main components of the proposed method are crucial for the segmentation performance.

Table 3. Ablation experiments on adjacent layer information assisted segmentation, multi-scale low–high level feature encoder–decoder, and multi-scale semantic information mapping module (Stage 1 and Stage 2) on the general test dataset.

Module	Index	
	IOU (%)	Dice (%)
Baseline (Model0)	70.93	79.61
Model1	72.58	82.27
Model2	73.35	82.66
Model3	73.63	82.98

4.6. Application to 3D Reconstruction and 3D Printing

This section sequentially introduces the Marching Cubes (MC) algorithm involved in 3D reconstruction, the visualization toolkit (VTK) pipeline, and 3D printing technology, along with a specific example of 3D reconstruction and printing.

MC algorithm based on isosurface extraction. This algorithm generates intermediate geometric primitives by extracting isosurfaces between two-dimensional sequence images, and these primitives are then assembled and displayed using visualization software. Therefore, this method is also known as the indirect rendering method. The intermediate geometric primitives include triangular and square patches. The MC algorithm constructs these primitives only for the object's surface and cannot display the internal structure. The moving cubes algorithm proposed by Lorensen and others, known for its high precision and wide applicability, is one of the most popular methods in 3D display. We have adopted this method for our 3D reconstruction.

Introduction to VTK 8.2.0 and its visualization pipeline. VTK is an open-source software library specifically for 3D image processing and visualization. Its core is written in C++ using object-oriented principles. Thanks to its open-source nature, many researchers have further developed VTK, enriching its functional interfaces and promoting its development. VTK has now been widely used in various fields such as medicine, geological exploration, and fluid dynamics, facilitating the processing and visualization of three-dimensional data. Thus, we have chosen VTK for 3D visualization.

3D printing technology. 3D Printing is a technique that constructs objects layer by layer using powder-like metal or plastic materials that can be bonded together, based on digital model files. 3D printing plays a significant role in intervertebral disc replacement surgeries. We attempt to 3D print the results of our segmentation. A specific example of 3D reconstruction and printing is shown in Figure 13. The left side displays the 3D reconstructed cervical vertebrae and intervertebral discs, while the right side shows the corresponding 3D printed models of the cervical vertebrae and intervertebral discs. The 3D models printed based on the segmentation results demonstrate good effectiveness, providing significant technical support for surgeries like intervertebral disc replacement.

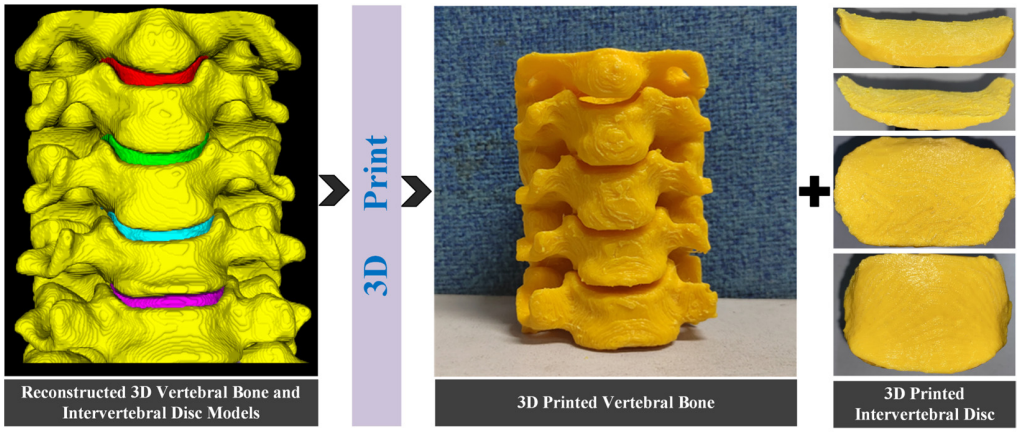


Figure 13. A 3D model of a cervical vertebral bone and intervertebral disc, and the corresponding 3D printed samples.

5. Conclusions

This paper introduces a significant advancement in the field of cervical intervertebral disc segmentation, employing a multi-scale information fusion method. This method uniquely combines multi-scale feature encoding–decoding with semantic fusion modules, facilitating an enhanced precision in the extraction of segmentation features. The two-stage process of the semantic fusion module, which involves convolution and pooling-based scale interactions, has been instrumental in improving the segmentation accuracy. The development of a specialized dataset for cervical intervertebral disc segmentation represents a critical contribution of this study, addressing the previously existing gap in this area. By incorporating inter-layer interpolation, this dataset effectively mitigates issues related to pixel spacing inconsistencies in CT images, thereby bolstering the reliability of segmentation results. Our experimental findings underscore the high segmentation accuracy of the proposed network model in the context of human cervical intervertebral discs. The potential applications of this model in 3D reconstruction and printing further highlight its practical utility. This research not only provides a robust methodological framework for cervical disc segmentation but also lays the groundwork for future explorations in spinal health diagnostics and treatment planning. In the future, we will explore how to further decrease the inference complexity.

Author Contributions: Conceptualization, Y.Y. and M.W.; Data curation, L.M., X.Z. (Xiang Zhang), K.Z. and X.Z. (Xiaoyao Zhao); Funding acquisition, Y.Y. and Q.T.; Investigation, M.W. and X.Z. (Xiaoyao Zhao); Methodology, Y.Y., M.W. and X.Z. (Xiaoyao Zhao); Resources, Q.T. and H.L.; Experiment, M.W.; Supervision, Q.T. and H.L.; Validation, Y.Y. and M.W.; Writing—original draft, Y.Y. and M.W.; Writing—review and editing, Q.T. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Health Commission of Sichuan Province, Project of China (project number: 21PJ037).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the researchers that participated in this study for providing the study environment.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Epstein Nancy, E. A Review of Complication Rates for Anterior Cervical Diskectomy and Fusion (ACDF). *Surg. Neurol. Int.* **2019**, *10*, 100. [CrossRef] [PubMed]
2. Ramesh, K.K.D.; Kumar, G.K.; Swapna, K. A review of medical image segmentation algorithms. *EAI Endorsed Trans. Pervasive Health Technol.* **2021**, *7*, e6. [CrossRef]

3. Abdellahoum, H.; Mokhtari, N.; Brahimi, A. CSFCM: An improved fuzzy C-Means image segmentation algorithm using a cooperative approach. *Expert Syst. Appl.* **2021**, *166*, 114063. [CrossRef]
4. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; pp. 234–241.
5. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016: 4th International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; IEEE: New York, NY, USA, 2016; pp. 565–571.
6. He, K.; Zhang, X.; Ren, S. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June 2016; pp. 770–778.
7. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122.
8. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
9. Chen, L.C.; Papandreou, G.; Kokkinos, I. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
10. Chen, L.C.; Papandreou, G.; Kokkinos, I. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
11. Lin, G.; Milan, A.; Shen, C. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1925–1934.
12. Peng, C.; Zhang, X.; Yu, G. Large kernel matters--improve semantic segmentation by global convolutional network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4353–4361.
13. Zhang, H.; Dana, K.; Shi, J. Context encoding for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7151–7160.
14. Yang, M.; Yu, K.; Zhang, C. Denseaspp for semantic segmentation in street scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3684–3692.
15. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
16. Zhao, H.; Shi, J.; Qi, X. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
17. Yu, C.; Wang, J.; Peng, C. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
18. Chen, L.C.; Zhu, Y.; Papandreou, G. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
19. Chollet, F. Xception: Deep learning with depthwise separable convolutions. *arXiv* **2017**, arXiv:1610.02357.
20. Sun, K.; Xiao, B.; Liu, D. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 5693–5703.
21. Zheng, S.; Lu, J.; Zhao, H. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 6881–6890.
22. Wang, W.; Xie, E.; Li, X. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 568–578.
23. Liu, Z.; Lin, Y.; Cao, Y. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
24. Liu, Z.; Hu, H.; Lin, Y. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12009–12019.
25. Ren, S.; Zhou, D.; He, S. Shunted self-attention via multi-scale token aggregation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10853–10862.
26. Strudel, R.; Garcia, R.; Laptev, I. Segmenter: Transformer for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 7262–7272.
27. Cheng, B.; Schwing, A.; Kirillov, A. Per-pixel classification is not all you need for semantic segmentation. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 17864–17875.
28. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A. An image is worth 16 $\times$ 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
29. Chen, J.; Lu, Y.; Yu, Q. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306.
30. Xie, E.; Wang, W.; Yu, Z. SegFormer: Simple and efficient design for semantic segmentation with transformers. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 6–14 December 2021; Volume 34, pp. 12077–12090.
31. Guo, J.; Zhou, H.Y.; Wang, L. UNet-2022: Exploring Dynamics in Non-isomorphic Architecture. *arXiv* **2022**, arXiv:2210.15566.

32. Wan, Q.; Huang, Z.; Lu, J. Seaformer: Squeeze-enhanced axial transformer for mobile semantic segmentation. *arXiv* **2023**, arXiv:2301.13156. [CrossRef]
33. Jiao, R.; Zhang, Y.; Ding, L. Learning with limited annotations: A survey on deep semi-supervised learning for medical image segmentation. *Comput. Biol. Med.* **2023**, *169*, 107840. [CrossRef]
34. Zhang, Y.; Zhou, T.; Wang, S. Input augmentation with sam: Boosting medical image segmentation with segmentation foundation model. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Vancouver, BC, Canada, 8–12 October 2023; Springer Nature: Cham, Switzerland, 2023; pp. 129–139.
35. Tragakis, A.; Kaul, C.; Murray-Smith, R. The fully convolutional transformer for medical image segmentation. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2023; pp. 3660–3669.
36. Yuan, F.; Zhang, Z.; Fang, Z. An effective CNN and Transformer complementary network for medical image segmentation. *Pattern Recognit.* **2023**, *136*, 109228. [CrossRef]
37. Kirillov, A.; Mintun, E.; Ravi, N. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–3 October 2023; pp. 4015–4026.
38. Ali, M.; Jabreel, M.; Valls, A. LezioSeg: Multi-Scale Attention Affine-Based CNN for Segmenting Diabetic Retinopathy Lesions in Images. *Electronics* **2023**, *12*, 4940. [CrossRef]
39. You, Z.; Yu, H.; Xiao, Z. CAS-UNet: A Retinal Segmentation Method Based on Attention. *Electronics* **2023**, *12*, 3359. [CrossRef]