

Article

Spider U-Net: Incorporating Inter-Slice Connectivity Using LSTM for 3D Blood Vessel Segmentation

Kyeorye Lee ¹, Leonard Sunwoo ², Tackeun Kim ² and Kyong Joon Lee ^{2,*}

¹ Interdisciplinary Program in Bioengineering, Seoul National University, Seoul 08826, Korea; siniphia@snu.ac.kr

² Department of Radiology, Seoul National University Bundang Hospital, Seongnam 13620, Korea; leonard.sunwoo@snu.ac.kr (L.S.); tackeun.kim@snu.ac.kr (T.K.)

* Correspondence: kjoon@snubh.org

Abstract: Blood vessel segmentation (BVS) of 3D medical imaging such as computed tomography and magnetic resonance angiography (MRA) is an essential task in the clinical field. Automation of 3D BVS using deep supervised learning is being researched, and U-Net-based approaches, which are considered as standard for medical image segmentation, are proposed a lot. However, the inherent characteristics of blood vessels, e.g., they are complex and narrow, as well as the resolution and sensitivity of the imaging modalities increases the difficulty of 3D BVS. We propose a novel U-Net-based model named Spider U-Net for 3D BVS that considers the connectivity of the blood vessels between the axial slices. To achieve this, long short-term memory (LSTM), which can capture the context of the consecutive data, is inserted into the baseline model. We also propose a data feeding strategy that augments data and makes Spider U-Net stable. Spider U-Net outperformed 2D U-Net, 3D U-Net, and the fully convolutional network-recurrent neural network (FCN-RNN) in dice coefficient score (DSC) by 0.048, 0.077, and 0.041, respectively, for our in-house brain MRA dataset and also achieved the highest DSC for two public datasets. The results imply that considering inter-slice connectivity with LSTM improves model performance in the 3D BVS task.

Keywords: 3D blood vessel segmentation; deep supervised learning; inter-slice connectivity; long short-term memory; medical imaging; semantic segmentation; U-Net



Citation: Lee, K.; Sunwoo, L.; Kim, T.; Lee, K.J. Spider U-Net: Incorporating Inter-Slice Connectivity Using LSTM for 3D Blood Vessel Segmentation. *Appl. Sci.* **2021**, *11*, 2014. <https://doi.org/10.3390/app11052014>

Academic Editor: Syoji Kobashi

Received: 31 December 2020

Accepted: 19 February 2021

Published: 25 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Blood vessel segmentation (BVS) is a prerequisite step that aids in a wide range of clinical procedures such as surgery planning and disease diagnosis. In the radiology field, it is done on images from various imaging modalities such as computed tomography (CT), magnetic resonance imaging (MRI), and magnetic resonance angiography (MRA). For example, there is a study in which coronary arteries are segmented with CT angiography to assess and diagnose coronary heart diseases [1]. Intracranial artery segmentation with MRA images can help radiologists diagnose serious cerebrovascular diseases such as ischemic stroke by displaying areas with stenosis or occlusion [2].

The automation of BVS is an important subject because the manual segmentation of blood vessels is challenging and laborious. The anatomical complexity of the blood vessel structure is generally higher than other organs of the human body, thus it is time consuming for trained technicians to perform. There are various traditional methods for automatic BVS, such as the vessel enhancing model, the deformable model, and the tracking model. The tracking model [3,4] increases vessel contrast compared with the non-vessel area by applying adequate filters to the image. Binarization with a pre-defined threshold value is done to produce the final segmentation mask. The deformable model [5,6] decides the boundary of the vessel contour by considering internal and external forces using mathematical formulae such as B-snake. The tracking model [7,8] repeatably finds the nearby vessel area from the initial seed point with pre-defined rules and constraints.

However, such traditional methods require manual input features or parameters, the optimal value of which is different according to the target dataset. They also have difficulty when dealing with exceptional cases, other than machine learning methods [9].

Recently, with the advent of convolutional neural networks (CNN), segmentation models using deep supervised learning are being actively proposed in the medical imaging field. In particular, U-Net [10] has been recognized as the de facto segmentation model for medical imaging as it enables deliberate prediction due to skip connections that bridge the encoder and decoder of the same level. Furthermore, in 3D BVS, there are many kinds of research using U-Net as the baseline. First, there are approaches using the 2D U-Net family which processes a single plane image at a time [11,12]. Although it is generally applied to 2D BVS regarding optical imaging modalities (e.g., color fundus photography, fluorescein angiography), there is a report that 2D U-Net outperforms the graph cut method—a representative traditional image segmentation method—when segmenting intracranial artery [11]. The limitation of 2D U-Net is that it ignores the z-axis of 3D data. It understands 3D volume as a stack of 2D axial slices and does not consider the relationships with the other slices. Therefore, the information in the z-axis, which can have a positive effect in model training, is not utilized. Next, the 3D U-Net family, which extracts volumetric features instantly using a 3D convolutional operation, is commonly adopted [13–16]. However, 3D U-Net consumes a considerable amount of memory because it requires sufficient 3D data as input. Although volumetric cropping can be processed to bypass the memory issue, it narrows the sight of the model destroying global anatomical information which can be an important clue when dealing with a medical image. Moreover, 3D U-Net has a difficulty when trained with 3D data such that each volume is comprised of various numbers of axial slices caused by the diversity of the thicknesses and spacings of the images [17,18]. This characteristic is common in medical image datasets, especially for CT, and to bypass this problem, the additional pre-processing stage which reconstructs the whole dataset with the same voxel size is necessary.

There are attempts to incorporate a recurrent neural network (RNN) into the U-Net baseline to make the model learn additional information through the z-axis [19,20]. Generally, for 3D medical imaging, the region-of-interest, such as the tumor, lesion, and organ, is spread through several image slices and the nearby images are visually similar to each other. When the 3D volume is interpreted as the image sequence, they form connectivity through the z-axis and it can be extracted by adding an RNN, which is designed to learn features of the consecutive data [21]. A fully convolutional network (FCN)-RNN [20] is a representative architecture for 3D medical image segmentation that arranges arbitrary RNN components after the arbitrary fully convolutional network (FCN) baseline, which is a type of CNN for semantic segmentation with an encoder–decoder structure [22]. FCN-RNN is implemented by adopting convolutional long short-term memory (LSTM) after the customized U-Net. Using an RNN has been demonstrated to be suitable when dealing with 3D medical images achieving a better performance than 2D U-Net and 3D CNN.

For 3D BVS, we believe that emphasizing inter-slice connectivity with the RNN could be more effective for the following two reasons. First, the blood vessel in the 3D modality is represented with small cross-sectional areas of less than 1% of an image. This severe lack of information about the blood vessel causes a class imbalance and causes the training stage to be dominated by the non-vessel area, resulting in a tendency to produce false negative pixels [23]. Augmenting information about the blood vessel with adjacent slices would be helpful to alleviate this problem. Second, the blood vessel has the intrinsic characteristic of strong inter-slice connectivity because the prime role of the blood vessel is to convey fluid to the whole body without leakage. An RNN can be effectively trained when dealing with datasets with this characteristic.

In this context, we propose Spider U-Net—a novel 2D-based deep learning framework for 3D BVS. It segments blood vessels by emphasizing the inter-slice connectivity using an RNN. Spider U-Net is composed of two paths: the warp path and the weft path. The

warp path is formed by stacking several 2D U-Net baselines in parallel to be able to process consecutive axial image slices simultaneously. It extracts spatial features of 2D images one by one. The weft path is implemented by incorporating convolutional LSTM between the encoder and the decoder of the warp path. It additionally learns the connectivity of extracted spatial features through the z-axis. In consequence, the prediction masks are inferred not only by spatial features but also the inter-slice connectivity of adjacent slices. We also suggest a training strategy named striding stencil (SS) to augment data and to train the stable model. Finally, we evaluate Spider U-Net on three 3D BVS datasets that vary in terms of modality and target vessel. Previous representative models re-implemented with an equal baseline are compared with our framework to assess the effectiveness of Spider U-Net.

2. Methods

2.1. Network Architecture

Spider U-Net aims to secure rich information regarding a blood vessel by combining spatial features and inter-slice connectivity with each warp path and weft path, in a similar way to weaving fabric. It does so by following three steps. First, the encoder part of the warp path prepares a set of spatial features from the image sequence of a certain length. Second, the weft path is spread through the warp path creating information flow that crosses over the consecutive images. It then extracts the inter-slice connectivity of the blood vessel using LSTM. Third, the decoder part of the warp path predicts the accurate segmentation mask using rich information augmented by the inter-slice connectivity. Here, we introduce the network architecture of Spider U-Net.

2.1.1. Common Baseline

We start from the plain 2D U-Net, which is composed of an encoder, a decoder, and skip connections. The encoder extracts image features by using convolution layers. Max-pooling layers are located at the end of each level and shrink the input resolution in half. They enable the convolutional operations to extract abundant image features under the various input resolutions. The decoder predicts the blood vessel area and restores the image resolution to the original input size. For prediction accuracy, the up-sampling paths, which double the input size, are composed of deconvolution layers. As the deconvolution layers are trainable, they are dynamically optimized to predict a better segmentation mask during training. Skip connections are designed to improve the performance of the decoder by passing additional image features from each level of the encoder. The softmax function located at the end of the baseline model produces the probability map which indicates whether each pixel is a blood vessel or not. To make the final segmentation mask, pixels with a probability larger than 0.5 are marked as blood vessels.

The overall architecture of the baseline U-Net is shown in Figure 1. The total network depth is set to 5, and each level of the encoder and decoder is linked with skip connections. When the serial arrangement of a convolutional layer, a rectified linear unit, and a batch normalization is denoted as a basic layer set (BLS), each level of the encoder is implemented with two BLS and a max-pooling layer, and the decoder is composed of two BLS and a deconvolutional layer per level. A single BLS is placed in the middle of the encoder and decoder, and the last convolutional layer is activated by the softmax function. This whole architecture is denoted as *B* and is adopted for the common baseline of Spider U-Net and the other comparative models.

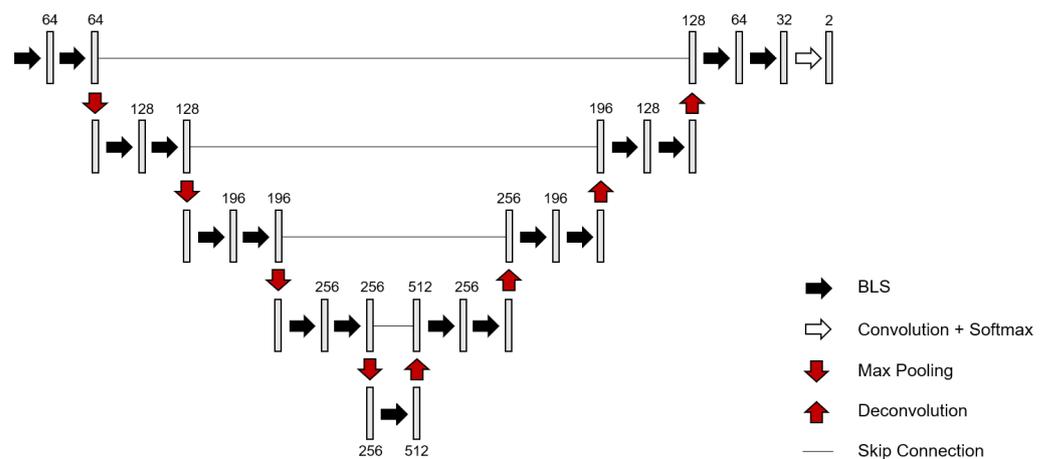


Figure 1. Structure of the U-Net baseline B . The downstream path is an encoder part which extracts convolutional features and the upstream path is a decoder part which restores segmentation masks. The network depth is fixed to 5 and skip connections connect the encoder and decoder on each level. Each level of the encoder is implemented by two basic layer sets (BLS) composed of a 3×3 convolutional layer, a rectified linear unit, and a batch normalization layer, and a single 2×2 max-pooling layer follows. The decoder has two BLS and a deconvolutional layer per level. The last layer, which is activated by the softmax function, produces probabilities regarding whether each pixel is a blood vessel or not.

2.1.2. Warp Path

The warp path is formed by stacking the baseline B several times in parallel, while maintaining parameter sharing, as shown in Figure 2. This structure enables the model to receive and process each slice of the image sequence simultaneously. The encoder part extracts a set of spatial features and serves them side by side like warp threads. The decoder part restores binary vessel segmentation masks from connectivity-aware spatial features augmented by the weft path. These are explained in the next section.

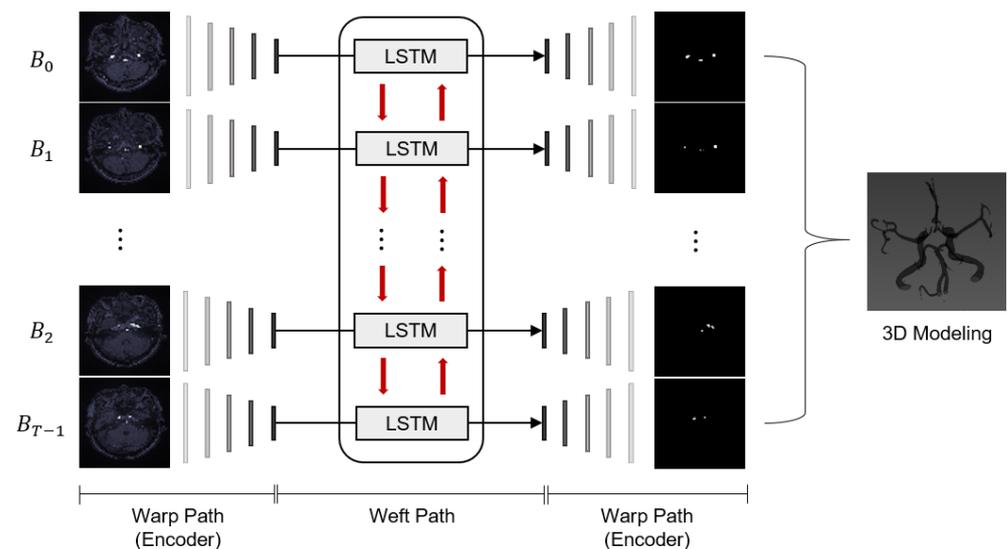


Figure 2. Structure of Spider U-Net. The warp path, which is formed by stacking the 2D U-Net baseline (B) T times in parallel, extracts spatial features of each 2D image and infers each segmentation mask. The weft path is placed between the encoder and decoder of the warp path and bridges information path between consecutive images (red arrows inside the weft path). Therefore, the model can learn additional inter-slice connectivity from the weft path and produces deliberate 2D segmentation masks from the spatial features augmented by the weft path. The final 3D modeling is constructed by simple trilinear interpolation.

The number of baselines B composing the warp path is denoted as T . This determines the input capacity, i.e., the size of the image sequence the model can process at one time. T can be defined freely by the network designer. However, setting T too high can be inefficient, because the context of two slices becomes weaker as the distance between two slices increases. We conducted pilot studies under various settings and found that a T larger than five did not give a notable improvement in the model performance.

2.1.3. Weft Path

The weft path is built by replacing the bottommost convolutional layers between the encoder and decoder of the warp path with a single convolutional LSTM layer [21] in a bidirectional way to capture the inter-slice connectivity through the z-axis, as depicted in Figure 2. We selected convolutional LSTM because the LSTM family is a specialized type of neural network for extracting rules and features of sequential data and can stably maintain contextual information due to the inner memory components. In addition, convolutional LSTM can process features from a CNN without losing spatial information, which is essential to semantic segmentation in which the inner operations are replaced by the convolutional operations.

We placed the weft path between the encoder and decoder of the warp path. As far as we know, previous approaches using RNN to exploit inter-slice connectivity placed the RNN component after the baseline model. We conducted experiments on these two model types and the results show that our RNN location achieves a better performances than the others. More related discussions can be found in Section 5.1.

2.2. Data Feeding Strategy

We propose a customized data feeding strategy for Spider U-Net named striding stencil (SS) to achieve memory-efficient and stable training, as shown in Figure 3. Rather than training the given image sequence at once, SS causes Spider U-Net to process data in several steps. First, Spider U-Net updates the model parameters with a portion of the volume as an image sequence with length T , i.e., the pre-defined input capacity of the warp path. Next, the model moves through the inter-slice level with the given natural number S . These two steps are repeatedly continued until the model reaches the end of the volume.

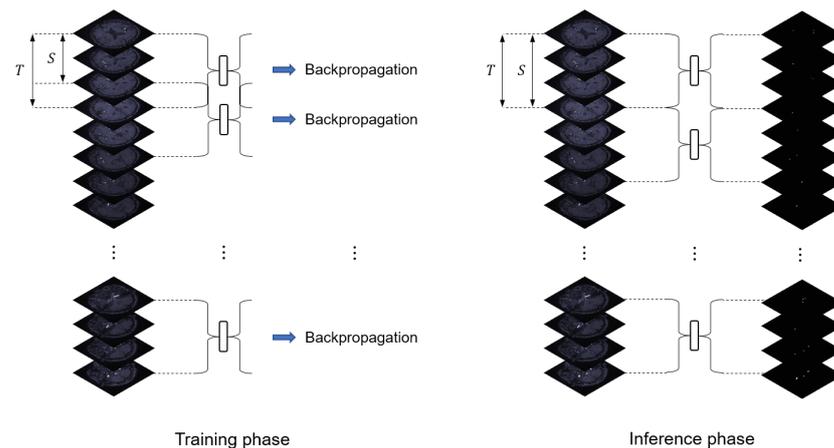


Figure 3. The striding stencil strategy of the training and inference phases. The spider-like figure represents Spider U-Net. When processing a 3D image, Spider U-Net holds the image sequence of length T at a time and updates the model weights with backpropagation. Thereafter, it strides through the z-axis with a hyperparameter S . These two steps are performed iteratively until the model reaches the end of the 3D image. For the training phase, S is set to less than T to ensure that the image sequences are partially overlapped. This has the effect of data augmentation and ensures that the model is robust to the composition of the image sequence. For the inference phase, S is set to be equal to T to reduce the time cost.

For the training phase, S is set to less than T to ensure the current image sequence is partially overlapped with the previous one. This *dense* training both augments the data to prevent overfitting and ensures the model is robust to the composition of the image sequence.

For the inference phase, S is set to be equal to T so that each image sequence is not overlapped. This *sparse* inference saves time because it is the fastest way to process the whole volume. At the same time, as a result of dense training, the model can predict consistent segmentation masks regardless of the composition of the image sequence.

3. Experiments

3.1. Dataset and Image Processing

We selected three different 3D BVS datasets that have varieties in the target organ, vessel size, and modality. Detailed information is described in Table 1.

Table 1. Details of 3D blood vessel segmentation datasets used in experiments.

Dataset	Target Vessel	Organizer	#Volumes			#Slices Per Volume
			Train	Validation	Test	
Brain MRA	Intracranial Artery	SNUBH	18	4	4	140–148
Abdomen CT	Hepatic Vessel	Decathlon [24]	253	25	25	24–181
Cardiac MRI	Left Ventricle Wall	York Univ. [25]	27	3	3	8–15

The first dataset is a brain time-of-flight (TOF) MRA for intracranial artery segmentation provided by Seoul National University Bundang Hospital (SNUBH). Twenty-six TOF MRA studies of normal cases were randomly selected, and only axial slices were downloaded from the picture archiving and communication system of SNUBH. The dataset was split into 18, 4, and 4 cases for the training, validation, and test set, respectively. Each case varied in the machine vendor, slice number per volume, and image size. The target vessel to segment was defined to all visible intracranial arteries except for the peripheral middle cerebral arteries after the M1 branch. The annotation of the groundtruth mask was done by two invited radiologists using the medical imaging interaction toolkit (version 2018.04.2). Because annotation from scratch takes a lot of time and human resources, we worked on the initial segmentation masks produced by a simple thresholding method that marks bright pixels with the intensity of the top 30% as blood vessel and the others as background. All images were normalized from 0 to 1, cropped centrally leaving 75% of the original image, and resized to 256×256 pixels.

The second dataset is an abdomen CT for hepatic vessel segmentation. We used a part of the publicly available Decathlon Challenge datasets [24] that contain 303 abdomen CT cases with annotation masks of liver tumor and hepatic vessel per axial slice. We split the dataset into 253, 25, and 25 cases for the training, validation, and test set, respectively. The image size was 512×512 pixels, and the slice number was diverse. The tumor masks were ignored because our region-of-interest was only the hepatic vessel. For pre-processing, we clipped the Hounsfield Unit (HU) between 0 to 300 to eliminate a vast range of unnecessary HU. All images were normalized from 0 to 1, and 256×256 pixels, centered on the liver area, were cropped and used.

The last dataset was the public cardiac MRI dataset for left ventricle (LV) segmentation provided by York university [25]. Although LV is technically a type of organ rather than vessel, it conveys blood, which forms the definite inter-slice connectivity. The experiment on this dataset was used to validate the applicability of Spider U-Net to general organ segmentation with the inter-slice connectivity. For the given 33 MRI cases, each case had a single heartbeat composed of 3D cardiac volume with 20 frames. We split the dataset into 27, 3, and 3 cases for the training, validation, and test set, respectively. The given labels were initially two oval contours representing the endocardium and epicardium of the LV with 32 coordinates per slice. Slices not containing both contours were eliminated

because we extracted the donut-shaped LV wall area between two labels and used it as a groundtruth mask. All images were normalized from 0 to 1.

3.2. Comparative Models

We selected three models to compare: 2D U-Net, 3D U-Net, and FCN-RNN. The first two models are representative approaches using 2D and 3D convolutional operation, each in the medical image segmentation field. FCN-RNN is an abstract architecture that adds any RNN component after any FCN baseline. It has a similar philosophy to Spider U-Net, which utilizes the inter-slice connectivity of the image sequence.

3.3. Implementation Details

The implementation and execution of all experiments were performed on the Keras platform (version 2.2.4-tf) on the Tensorflow backend (version 2.1.0) using Python programming language (version 3.7.4). The open-source Python package Pydicom (version 1.4.2) and NiBabel (version 3.0.2) were used for loading and processing medical images. To use graphics processing unit acceleration, a single TITAN RTX was exploited by the CUDA toolkit (version 10.1) and cuDNN library (version 7.6).

For the common baseline B , The model depth was set to 5, and the filter number of convolutional layers per level was 64, 128, 196, 256, and 512. All convolutional operations used a filter size of 3, and max-pooling operations halve each axis. This setting was applied equally to Spider U-Net and the other comparative models for a fair comparison. For Spider U-Net, T , i.e., the number of B replicas forming the warp path, was set to 5. The weft path was implemented with a bidirectional convolutional LSTM with 512 convolution filters and replaced the deepest convolution layers of the warp path. S , i.e., the stride parameter of SS, was set to 3 and 5, respectively, for the training and inference phase. The SS strategy was also applied to 3D U-Net and FCN-RNN.

The implementation details of comparative models are as follows. 2D U-Net was implemented with a single B using 2D operations for all layers (e.g., convolution, deconvolution, and max-pooling), and the input was 2D images. 3D U-Net was also based on a single B , and the differences with 2D U-Net was that 3D operations were applied and the input was volumes with a stack of 16 images. As B has four downsampling paths, at least 16 images were required simultaneously for 3D U-Net. For the cardiac MRI dataset in particular, the image stack size was set to 8 because the length of the available image sequence was usually less than 16. In this case, the last downsampling step of 3D U-Net was changed into the 2D operation which halved the x- and y-axis only. For FCN-RNN, the warp and weft paths of the Spider U-Net implementation were applied to FCN and RNN component individually. In other words, the only difference between our customized FCN-RNN and Spider U-Net was the placement of the weft path.

The focal dice loss [26] with Laplace smooth factor 1 was used as a loss function to deal with the harsh class imbalance of the 3D BVS dataset. Adam optimizer was adopted with an initial learning rate of 0.001. The training was performed in an end-to-end manner without pre-training. The early stopping strategy was adapted when the validation loss was not improved within 10 epochs. The best model with the lowest validation loss was selected as the final model for evaluation.

4. Results

We report the performances of the Spider U-Net and the other comparative models on test set of three datasets. Average dice coefficient score (DSC) and Intersection-over-union (IoU) were the evaluation metrics for all experiments.

4.1. Main Experiment

Spider U-Net and three comparative models were trained with three datasets separately, creating 12 trained models. We evaluated every model with the corresponding test sets and report each performance with an average of dice coefficient score (DSC) and

intersection-over-union (IoU). Detailed results are shown in Table 2. Spider U-Net achieved the best performance among all models except for IoU on the brain MRA dataset. Qualitative results are shown in Figure 4. Yellow, red, and blue colors denote true-positive, false-positive, and false-negative each.

Table 2. Average of the dice score (DSC) and intersection-over-union (IoU) of Spider U-Net and the other comparative models on three datasets.

Metric	Dataset	2D U-Net	3D U-Net	FCN-RNN	Spider U-Net
DSC	Brain MRA	0.745	0.716	0.752	0.793
	Abdomen CT	0.327	0.145	0.351	0.456
	Cardiac MRI	0.772	0.726	0.812	0.833
IoU	Brain MRA	0.673	0.617	0.758	0.743
	Abdomen CT	0.277	0.152	0.271	0.375
	Cardiac MRI	0.681	0.624	0.707	0.728

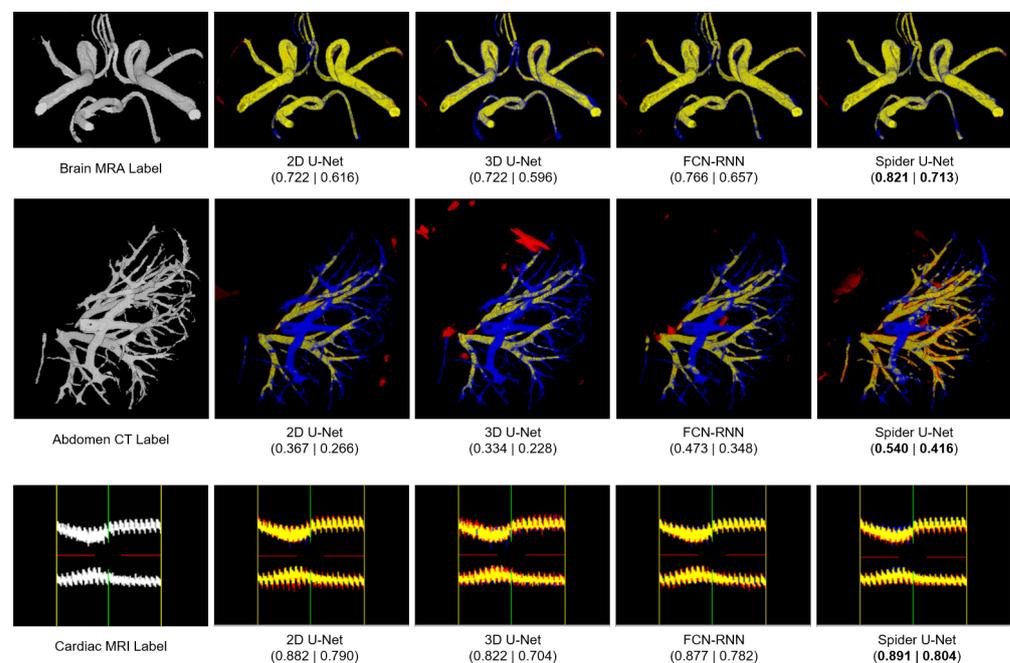


Figure 4. Qualitative results of Spider U-Net and the other comparative models on three datasets. The first row shows the brain MRA, the second row shows the abdomen CT, and the last row shows the cardiac MRI. The last row shows a single pulse of the left ventricle wall over 20 frames. As the cardiac MRI is 4D data, i.e., volume with the time axis, we visualized the cross-section of the left ventricle wall through central coronal plane. Each average value of dice coefficient score (DSC) and intersection-over-union (IoU) according to an example case is reported with DSC | IoU. True-positive, false-positive, and false-negative are denoted as yellow, red, and blue, respectively.

For the brain MRA dataset, 2D U-Net, 3D U-Net, FCN-RNN, and Spider U-Net scored 0.745, 0.716, 0.752, and 0.793 for DSC and 0.673, 0.617, 0.758, and 0.743 for IoU, respectively. As an example, in the first row in Figure 4, entire models failed to successfully predict the posterior cerebral arteries located on the bottom side but could predict relatively thick vessels such as the internal carotid artery and basilar artery. Only Spider U-Net predicted all four visible anterior cerebral arteries which are the thinnest vessel type.

For the abdomen CT dataset, 2D U-Net, 3D U-Net, FCN-RNN, and Spider U-Net achieved 0.327, 0.145, 0.351, and 0.456 for DSC and 0.277, 0.152, 0.271, and 0.375 for IoU, respectively. An example of hepatic vessel segmentation in the second row in Figure 4 shows

clear qualitative differences between Spider U-Net and the others. Spider U-Net produces less false-negative pixels, thus maintaining overall performance. On the other hand, Spider U-Net predicted a part of the vertebra as a blood vessel, producing a lot of false-positive pixels at the left side of the figure. To be utilized in real clinical scenarios, additional image processing techniques such as removing bone areas in the pre-processing stage and deleting the out-of-context blobs in the post-processing stage would be required. The performances of 3D U-Net are very poor, and we believe this is because of the anisotropy of the abdomen CT dataset discussed in Section 5.2.

For the cardiac MRI dataset, 2D U-Net, 3D U-Net, FCN-RNN, and Spider U-Net scored 0.882, 0.822, 0.877, and 0.891 for DSC, and 0.790, 0.704, 0.782, and 0.804 for IoU, respectively. The third row in Figure 4 represents qualitative examples of cardiac MRI predictions. As this dataset is given in 4D, i.e., 20 frames of 3D cardiac volume representing a single cardiac cycle, we consecutively arranged a cross-section view of 20 frames of volumes through the central coronal plane.

4.2. Ablation Experiment

4.2.1. The Impact of the Weft Path

We implemented Spider U-Net without the weft path to verify the influence of the inter-slice connectivity on performance. In this model, the weft path was changed to a set of convolutional layers with 512 filters, disconnecting the bridges between image slices. When the original Spider U-Net scored 0.796 and 0.743 for DSC and IoU, respectively, the adaptation scored 0.700 for DSC and 0.633 for IoU, as shown in Table 3. This implies that the weft path extracting inter-slice connectivity plays a vital role in the performance of Spider U-Net.

Table 3. Dice coefficient score (DSC) and intersection-over-union (IoU) of Spider U-Net adaptations on the brain MRA dataset.

Metric	Spider U-Net (w/o. Weft Path)	Spider U-Net (w. Attention Module)	Spider U-Net
DSC	0.700	0.807	0.796
IoU	0.633	0.740	0.743

4.2.2. The Expandability of Warp Path

The warp path of Spider U-Net can be augmented with various 2D-based modules because it is composed of several *B* baselines which are basically a 2D U-Net. We implemented the expanded Spider U-Net by attaching an attention module [27] to the warp path for assessment. As shown in Table 3, it scored 0.807 for DSC and 0.740 for IoU, which is not significantly different from the original Spider U-Net. We will explore the performances under various options by adopting other modules, such as a residual block and squeeze-and-excitation block, to improve the model performance of Spider U-Net.

5. Discussion

Spider U-Net was designed to improve the performance of the 3D BVS task by reinforcing the insufficient blood vessel features with inter-slice connectivity. The experimental results show the effectiveness of Spider U-Net on the 3D BVS dataset, for which it achieved the highest score among all the tested models. Spider U-Net can also be adopted to segment vessel-like organs such as the lung airway and lymphatic vessels which have strong inter-slice connectivity. In the following subsections, we discuss factors that affect model performance.

5.1. Network Composition

In this study, the position in which the weft path was built affected the segmentation performance. The only difference between our FCN-RNN re-implementation and Spider

U-Net is that the weft path—was placed after the warp path for FCN-RNN, whereas it was placed in the middle of the warp path for Spider U-Net. Considering that Spider U-Net outperformed FCN-RNN in most cases, incorporating the weft path inside the warp path is the experimentally better placement.

For FCN-RNN, the weft path receives a segmentation mask with a single channel and post-processes it by extracting the spatial feature again. This workflow would not be optimal because there are two separate feature extractors in the architecture. In Spider U-Net, on the other hand, the weft path directly augments global features without re-extracting because it is considered as a part of the encoder. We believe this architecture makes the training process stable and has a positive effect on performance.

5.2. Anisotropy of the Dataset

3D medical images vary in slice thickness and inter-slice spacing according to the manufacturer, machine, and reconstruction method. In the volumetric view, this characteristic means that each 3D volume has a diverse voxel size along the z-axis and means the entire dataset is anisotropic. There is some evidence to suggest that 3D U-Net performance tends to deteriorate when trained with an anisotropic dataset [17,18]. We also conducted an ablation study with our in-house isotropic dataset by collecting MRA volumes with the same manufacturer and image acquisition settings. The average DSC of 2D U-Net, 3D U-Net, FCN-RNN, and Spider U-Net was 0.696, 0.697, 0.725, and 0.737, respectively. In this experiment, the performance of 3D U-Net was not inferior to 2D U-Net. This shows that 3D U-Net has difficulty when trained with an anisotropic dataset. On the other hand, Spider U-Net, which relies on LSTM when extracting z-axis information, demonstrates robustness in terms of anisotropy by showing consistent results on every dataset. As a result of this characteristic, an additional reconstruction stage to match up voxel size was not required.

6. Conclusions

We propose a novel deep learning framework, Spider U-Net, for 3D BVS that incorporates an RNN component into the 2D U-Net baseline. The core element is the weft path that unites the broken context between consecutive images by applying convolutional LSTM in a bidirectional way. This helps the entire model infer accurate segmentation masks while taking into consideration adjacent image features. Moreover, a training strategy named striding stencil is proposed to train Spider U-Net. Our proposed model was evaluated on three different datasets and outperformed all comparative models on DSC. This shows that utilizing the inter-slice connectivity with LSTM improves the performance for 3D BVS tasks. In addition, we experimentally validated that our network component placement is more effective than previous models.

A future study to validate Spider U-Net in a real clinical scenario is planned. In order to segment the intracranial artery and diagnose stenosis and occlusion simultaneously, we will implement a multi-task model by attaching an auxiliary classifier to Spider U-Net. We will then compare the diagnostic results with invited radiologists.

Author Contributions: Conceptualization, K.L., L.S. and K.J.L.; methodology, K.L.; software, K.L.; validation, K.L.; formal analysis, L.S. and T.K.; investigation, K.L. and L.S.; resources, L.S., T.K. and K.J.L.; data curation, L.S. and T.K.; writing—original draft preparation, K.L.; writing—review and editing, K.L. and K.J.L.; visualization, K.L.; supervision, T.K. and K.J.L.; project administration, L.S. and K.J.L.; funding acquisition, L.S. and K.J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. 2017R1D1A1B03033610) and by grant (no. 18-2018-016) from the SNUBH Research Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Minkyung Kang and Green Lee for providing technical assistance to data acquisition and annotation.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BVS	Blood Vessel Segmentation
CNN	Convolutional Neural Network
CT	Computed Tomography
DSC	Dice Coefficient Score
FCN	Fully convolutional Network
HU	Hounsfield Unit
IoU	Intersection-over-Union
LSTM	Long Short-Term Memory
LV	Left Ventricle
MRA	Magnetic Resonance Angiography
MRI	Magnetic Resonance Imaging
RNN	Recurrent Neural Network
SNUBH	Seoul National University Bundang Hospital
SS	Striding Stencil
TOF	Time-of-flight

References

1. Tian, Y.; Pan, Y.; Duan, F.; Zhao, S.; Wang, Q.; Wang, W. Automated Segmentation of Coronary Arteries Based on Statistical Region Growing and Heuristic Decision Method. *Biomed. Res. Int.* **2016**, *2016*, 3530251. [[CrossRef](#)] [[PubMed](#)]
2. Chen, L.; Mossa-Basha, M.; Sun, J.; Hippe, D.S.; Balu, N.; Yuan, Q.; Pimentel, K.; Hatsukami, T.S.; Hwang, J.N.; Yuan, C. Quantification of morphometry and intensity features of intracranial arteries from 3D TOF MRA using the intracranial artery feature extraction (iCafe): A reproducibility study. *Magn. Reson. Imaging* **2019**, *57*, 293–302. [[CrossRef](#)] [[PubMed](#)]
3. Frangi, A.F.; Niessen, W.J.; Vincken, K.L.; Viergever, M.A. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 130–137. [[CrossRef](#)]
4. Truc, P.T.; Khan, M.A.; Lee, Y.K.; Lee, S.; Kim, T.S. Vessel enhancement filter using directional filter bank. *Comput. Vis. Image Underst.* **2009**, *113*, 101–112. [[CrossRef](#)]
5. Cheng, Y.; Hu, X.; Wang, J.; Wang, Y.; Tamura, S. Accurate Vessel Segmentation With Constrained B-Snake. *IEEE Trans. Image Process.* **2015**, *24*, 2440–2455. [[CrossRef](#)] [[PubMed](#)]
6. Espona, L.; Carreira, M.; Penedo, M.; Ortega, M. Retinal vessel tree segmentation using a deformable contour model. In Proceedings of the 2008 19th International Conference on Pattern Recognition, Tampa, FL, USA, 8–11 December 2008. [[CrossRef](#)]
7. Yin, Y.; Adel, M.; Bourennane, S. Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recognit.* **2012**, *45*, 1235–1244. [[CrossRef](#)]
8. Vlachos, M.; Dermatas, E. Multi-scale retinal vessel segmentation using line tracking. *Comput. Med Imaging Graph.* **2010**, *34*, 213–227. [[CrossRef](#)]
9. Zhao, F.; Chen, Y.; Hou, Y.; He, X. Segmentation of blood vessels using rule-based and machine-learning-based methods: A review. *Multimed. Syst.* **2017**, *25*, 109–118. [[CrossRef](#)]
10. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science*; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [[CrossRef](#)]
11. Livne, M.; Rieger, J.; Aydin, O.U.; Taha, A.A.; Akay, E.M.; Kossen, T.; Sobesky, J.; Kelleher, J.D.; Hildebrand, K.; Frey, D.; et al. A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease. *Front. Neurosci.* **2019**, *13*, 97. [[CrossRef](#)] [[PubMed](#)]
12. Wang, L.; Liang, D.; Yin, X.; Qiu, J.; Yang, Z.; Xing, J.; Dong, J.; Ma, Z. Coronary artery segmentation in angiographic videos utilizing spatial-temporal information. *BMC Med. Imaging* **2020**, *20*, 110. [[CrossRef](#)] [[PubMed](#)]
13. Sanches, P.; Meyer, C.; Vigon, V.; Naegel, B. Cerebrovascular Network Segmentation of MRA Images With Deep Learning. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019. [[CrossRef](#)]

14. Hilbert, A.; Madai, V.I.; Akay, E.M.; Aydin, O.U.; Behland, J.; Sobesky, J.; Galinovic, I.; Khalil, A.A.; Taha, A.A.; Wuerfel, J.; et al. BRAVE-NET: Fully Automated Arterial Brain Vessel Segmentation in Patients With Cerebrovascular Disease. *Front. Artif. Intell.* **2020**, *3*, 78. [[CrossRef](#)]
15. Huang, Q.; Sun, J.; Ding, H.; Wang, X.; Wang, G. Robust liver vessel extraction using 3D U-Net with variant dice loss function. *Comput. Biol. Med.* **2018**, *101*, 153–162. [[CrossRef](#)] [[PubMed](#)]
16. Zhu, W.; Huang, Y.; Zeng, L.; Chen, X.; Liu, Y.; Qian, Z.; Du, N.; Fan, W.; Xie, X. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med. Phys.* **2018**, *46*, 576–589. [[CrossRef](#)] [[PubMed](#)]
17. Isensee, F.; Petersen, J.; Klein, A.; Zimmerer, D.; Jaeger, P.F.; Kohl, S.; Wasserthal, J.; Koehler, G.; Norajitra, T.; Wirkert, S.; et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv* **2018**, arXiv:1809.10486.
18. Isensee, F.; Jaeger, P.F.; Full, P.M.; Wolf, I.; Engelhardt, S.; Maier-Hein, K.H. Automatic Cardiac Disease Assessment on cine-MRI via Time-Series Segmentation and Domain Specific Features. In *Lecture Notes in Computer Science*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 120–129. [[CrossRef](#)]
19. Cai, J.; Lu, L.; Xie, Y.; Xing, F.; Yang, L. Improving Deep Pancreas Segmentation in CT and MRI Images via Recurrent Neural Contextual Learning and Direct Loss Function. *arXiv* **2017**, arXiv:1707.04912.
20. Chen, J.; Yang, L.; Zhang, Y.; Alber, M.; Chen, D.Z. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3036–3044.
21. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems*; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Hong Kong, China, 2015; Volume 28, pp. 802–810.
22. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.
23. Chan, R.; Rottmann, M.; Huger, F.; Schlicht, P.; Gottschalk, H. Controlled False Negative Reduction of Minority Classes in Semantic Segmentation. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020. [[CrossRef](#)]
24. Simpson, A.L.; Antonelli, M.; Bakas, S.; Bilello, M.; Farahani, K.; van Ginneken, B.; Kopp-Schneider, A.; Landman, B.A.; Litjens, G.; Menze, B.; et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv* **2019**, arXiv:1902.09063.
25. Andreopoulos, A.; Tsotsos, J.K. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Med. Image Anal.* **2008**, *12*, 335–357. [[CrossRef](#)] [[PubMed](#)]
26. Wang, P.; Chung, A.C.S. Focal Dice Loss and Image Dilation for Brain Tumor Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 119–127. [[CrossRef](#)]
27. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.C.H.; Heinrich, M.P.; Misawa, K.; Mori, K.; McDonagh, S.G.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv* **2018**, arXiv:1804.03999.