Learning-based automatic segmentation of arteriovenous malformations on contrast CT images in brain stereotactic radiosurgery

Tonghe Wang*, Yang Lei*, Sibao Tian, Xiaojun Jiang, Jun Zhou, Tian Liu, Sean Dresser, Walter J. Curran, Hui-Kuo Shu, and Xiaofeng Yang^{a)}

Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA

(Received 20 December 2018; revised 23 April 2019; accepted for publication 23 April 2019; published 21 May 2019)

Purpose: Stereotactic radiosurgery (SRS) is widely used to obliterate arteriovenous malformations (AVMs). Its performance relies on the accuracy of delineating the target AVM. Manual segmentation during a framed SRS procedure is time consuming and subject to inter- and intraobserver variation. To address these drawbacks, we proposed a deep learning-based method to automatically segment AVMs on CT simulation image sets.

Methods: We developed a deep learning-based method using a deeply supervised three-dimensional (3D) V-Net with a compound loss function. A 3D supervision mechanism was integrated into a residual network, V-Net, to deal with the optimization difficulties when training deep networks with limited training data. The proposed compound loss function including logistic and Dice losses encouraged similarity and penalized discrepancy simultaneously between prediction and training dataset; this was utilized to supervise the 3D V-Net at different stages. To evaluate the accuracy of segmentation, we retrospectively investigated 80 AVM patients who had CT simulation and digital subtraction angiography (DSA) acquired prior to treatment. The AVM target volume was segmented by our proposed method. They were compared with clinical contours approved by physicians with regard to Dice overlapping, difference in volume and centroid, and dose coverage changes on original plan.

Results: Contours created by the proposed method demonstrated very good visual agreement to the ground truth contours. The mean Dice similarity coefficient (DSC), sensitivity and specificity of the contours delineated by our method were >0.85 among all patients. The mean centroid distance between our results and ground truth was 0.675 ± 0.401 mm, and was not significantly different in any of the three orthogonal directions. The correlation coefficient between ground truth and AVM volume resulting from the proposed method was 0.992 with statistical significance. The mean volume difference among all patients was 0.076 ± 0.728 cc; there was no statistically significant difference. The average differences in dose metrics were all less than 0.2 Gy, with standard deviation less than 1 Gy. No statistically significant differences were observed in any of the dose metrics.

Conclusion: We developed a novel, deeply supervised, deep learning-based approach to automatically segment the AVM volume on CT images. We demonstrated its clinical feasibility by validating the shape and positional accuracy, and dose coverage of the automatic volume. These results demonstrate the potential of a learning-based segmentation method for delineating AVMs in the clinical setting. © 2019 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.13560>]

Key words: AVM, CT, deep learning, segmentation

1. INTRODUCTION

Stereotactic radiosurgery (SRS) has been shown to be effective in obliterating arteriovenous malformations (AVMs) since the 1970s.¹ During SRS treatment, the patient is immobilized with a frame, and precisely targeted radiation is delivered using either LINAC or Gamma Knife systems. Typically, 17.5–20 Gy in a single fraction is prescribed to the margin of the AVM nidus. Progressive stenosis is then induced by radiation effects on the vascular endothelium, which reduces the risk of hemorrhage. Clinically, SRS serves as a noninvasive alternative to microsurgical and endovascular treatments, with minimal acute complications. SRS is favored for AVM

with small and compact nidus (<3 cm in diameter), or in deep and eloquent areas that would engender great neurologic risk with attempted resection.²

Accurate target definition of the nidus is of greatest concern for radiosurgery in order to avoid obliteration failure that may result from partial volume irradiation.^{3,4} During SRS treatment planning, in addition to CT simulation that is standard to radiotherapy workflow, digital subtraction angiography (DSA) is acquired on patients as a primary imaging modality for AVM target delineation. In DSA, two series of two-dimensional (2D) sequential x-ray images are taken frame by frame from orthogonal directions (anteroposterior and lateral) after the injection of a contrast agent. They are

then subtracted from the images taken prior to contrast administration. Background structures such as bones are subtracted from rendered images, leaving contrast-enhancing vessels in real time. Since it provides unique real-time vessel feeding and drainage information, DSA is currently considered as the gold standard for AVM target identification.⁵

However, DSA has its inherent limitations as a treatment planning procedure. First, DSA is not able to provide accurate three-dimensional (3D) target information because of its 2D nature. The 3D target volume defined by DSA is the intersection volume of a pair of 2D projections, based on contours delineated on the orthogonal pair of radiographs. Note that given two projection contours, such defined 3D target volume is just one of the possible volumes corresponding to them. Thus, it usually does not represent the true AVM shape, especially for concave and irregular targets. Resultant errors, with both significant overestimation and underestimation of AVMs have been reported,^{6,7} which may lead neurologic deficits or treatment failures, respectively. Additional adjustments by physicians are usually required based on CT simulation images. Secondly, DSA is an invasive procedure, which has been shown to be associated with nontrivial procedural risks, such as stroke or death.^{8–11} Moreover, DSA involves a much higher radiation dose to the patient than a CT scan; as well as exposure to the medical team due to its fluoroscopy nature.¹² Phantom studies show that the dose on a patient from cerebral vessel DSA is around four to five times that of CT angiography.^{13,14} Other minor issues include registration uncertainty between DSA and CT simulation images, patient discomfort, additional time, labor, and cost.

Given the limitations mentioned above, an alternative image modality capable of accurately localizing AVMs is desirable. Directly using CT simulation images with contrast enhancement to segment AVMs is attractive, as it not only provides volumetric information but also simplifies the workflow by eliminating an additional imaging procedure. However, the loss of dynamic information of contrast flow limits the ability to distinguish the nidus from surrounding normal vasculature. Manual AVM contouring on CT images is time consuming, observer dependent, and may be unacceptable with respect to accuracy in many cases compared with DSA as ground truth,¹⁵ while studies on automatic methods are sparse. Thus, an automatic and accurate AVM segmentation method using CT images is worth investigating.¹⁶

The aim of this study is to develop an automatic segmentation method to delineate the AVM target from CT simulation images. In recent years, machine learning methods are being integrated into segmentation process.^{17–22} In this study, we proposed a novel deep learning-based method using a deeply supervised 3D V-Net with a compound loss function, which is an end-to-end fully convolution neural network. To evaluate our proposed method, we retrospectively investigated 80 AVM patients who were treated with SRS. AVM contours that were automatically segmented by the proposed method solely on CT images were compared with those manually delineated by physicians based on DSA as ground truth.

2. MATERIALS AND METHODS

The proposed AVM segmentation method consists of a training stage and a segmentation stage. Each brain CT dataset contains the clinically implemented physician-drawn AVM contours. These clinical contours were converted into binary maps and used as the learning-based target of the CT images. The training CT images were cropped into a $64 \times 64 \times 64$ voxel volume of interest (VOI) enclosing the AVM target to reduce computational cost. The 3D deep learning networks were trained based on the entire extracted VOI. The 3D V-Net architecture was introduced to enable end-to-end learning. A deep supervision strategy combined with a compound loss was used to supervise the network.^{20,23,24} During the segmentation stage, the physician first selected a VOI enclosing the AVM on a newly acquired brain CT image; the general location and shape of the AVM target is usually apparent on CT images. A 3D volume within this VOI was fed into the well-trained networks which performed an end-to-end AVM segmentation. The segmented AVM volume was then located back to the image to get the final contour. Figure III outlines the workflow schematic of our segmentation method.

Our proposed network architecture (as shown in yellow and green part of Fig. III) was inspired by a widely known end-to-end V-Net.²⁵ Figure III shows the schematic flowchart of the proposed algorithm for AVM segmentation. The upper (yellow and green) part of this figure shows the training stage of our proposed method, which consists of V-Net (yellow) and deep supervision architecture (green). The input patch of the training stage is a cropped VOI of a brain CT image in the size of $64 \times 64 \times 64$ voxels. The lower part (orange) shows the segmentation stage. In segmenting stage, a new CT volume within the VOI is fed into the well-trained model to get an AVM segmentation. As shown in Fig. III yellow part, the network consists of a compression path (on left side of yellow part), decompression path (on right side of yellow part), and a bridge path (on middle side of yellow part) connecting these two paths. The compression path is composed of two and three convolutional layers for final/high-resolution stage and modest/low-resolution stage, respectively. Such convolutional component is followed by a “down” convolutional layer to reduce the resolution. In each convolution layer, feature representations can be extracted via 3D convolutions followed by the parametric rectified linear unit (PReLU). The PReLU is proposed as an effective activation function to prevent overfitting. Instead of a max-pooling operation for down sampling, a strided convolution (convolution with stride size of $2 \times 2 \times 2$ voxels), called “down” convolution, is adopted to produce the input volumes for the next convolution layer where the size of feature volumes is reduced while the number is increased. The decompression path is constructed by two and three convolutional layers and is followed by an “up” convolutional layer to enhance the resolution. The decompression path has a similar structure to the compression path above without strided convolution. Both compression path and decompression path use an element-wise sum, which is a

which is followed by a softmax and threshold operator to obtain an equal-size segmentation.

Recent reports have used either logistic or Dice loss functions in their networks.^{25,27} We proposed a compound loss function to supervise our network at four stages. It combines logistic loss and Dice loss, which penalizes dissimilarity and encourages similarity between the prediction and training dataset, respectively. Voxel-wise binary cross entropy (BCE) loss is widely used as one of the logistic losses. Since the segmentation task can be regarded as a binary regression, we used voxel-wise BCE loss as the logistic loss function. The BCE loss is defined as follows:

$$L_{BCE}(C, \hat{C}) = - \sum_j [C_j \log \hat{C}_j + (1 - C_j) \log(1 - \hat{C}_j)] \quad (1)$$

where C_j and $\hat{C}_j$ denote the j th voxel in clinical contour C and prediction $\hat{C}$, respectively. A Dice similarity coefficient (DSC) loss was also introduced, defined as:

$$L_{DSC}(C, \hat{C}) = 1 - \frac{2 \times V(C \cap \hat{C})}{V(C) + V(\hat{C})} \quad (2)$$

where V indicates the volume of the region enclosed in the contours. Combining the above two loss functions, the compound loss function for deep supervision at the different stages is defined as follows:

$$L_{final}(C, \hat{C}) = \sum_{l=1}^4 \lambda_l (L_{BCE}(C, \hat{C}^l) + \mu L_{DSC}(C, \hat{C}^l)) \quad (3)$$

where l denotes the stage of our network, λ_l is the regularization weights of the l th stage's loss, and μ is a balancing parameter, which balance the BCE and DSC losses. They were all set empirically to achieve optimal performance. Since the segmented contour in different stages has different resolution, the DSC loss weighting parameter λ_l should be provided differently. We set $\lambda_l = \rho^{l-1}$ ($0 < \rho < 1$) to enable lower weight for lower resolution and higher weight for higher resolution. We employed fourfold cross validation to evaluate the parameter μ and ρ setting, that is, the random selected 75% samples were used to train the model, and the remaining 25% samples were used for validation. By trying different values of parameters, it is shown that the performance is not sensitive when μ is between 1.7 and 2.3, thus we set $\mu = 2$. It also shown that when $\rho = 0.8$, the segmentation achieves the best performance. Thus, the λ_l is set as $\lambda_l = 0.8^{l-1}$.

To evaluate the performance of our proposed method for AVM segmentation, we compared the difference between contours generated by our method to those generated by clinicians. Our algorithm was implemented in python 3.6 and Tensorflow with Adam gradient descent optimizer and was trained and tested on a NVIDIA TITAN XP GPU with 12 GB of memory. Each epoch took 23 s. We also used several libraries and toolboxes such as numpy, scikit-image, pydicom, h5py, and scipy. Retrospectively we reviewed data from 80 patients (35 were male) who were treated with SRS.

Patient age ranged from 18 to 75 yr; median age was 45 yr. Each patient underwent standard treatment planning workflow, that is, CT simulation with contrast followed by DSA acquisition. The CT images were acquired by a Siemens SOMATOM Definition AS CT scanner at 120 kVp, patients received 200 cc of Isovue-300 contrast injected intravenously. Each 0.6-mm-thick slice had a resolution of 512×512 pixels, with pixel spacing of 0.79 mm. A stereotactic frame was mounted on the patient's head during the entire imaging process. Image registration between CT and DSA was done with the aid of fiducial markers. AVM 2D projection contours were delineated on DSA images by a radiation oncologist and then reconstructed as a 3D target volume on CT. A second radiation oncologist would adjust the contour based on CT landmarks until consensus was achieved, and a neurosurgeon would further adjust the contour with the two radiation oncologists until consensus was achieved among the three. Note that it is not always the same two radiation oncologists and one neurosurgeon that provided the contour for all cases, but the neurosurgeon and at least one of the radiation oncologists were attending physicians with more than 10 yr experience in treating AVM. All patients in this study had this clinical AVM volume, which was considered as ground truth. Corresponding target volumes were also generated by our proposed method. Our method was evaluated by leave-five-out experiment strategy, that is, for each experiment, we excluded five patients from the dataset for training our deep learning-based segmentation model. After training, the excluded patients' brain CT images were used for segmentation test.

We quantitatively characterized the accuracy of the proposed method using three common metrics: DSC, sensitivity and specificity. The DSC describes the overlap of the segmented volumes between ground truth and our proposed method, which can be calculated as,

$$DSC = \frac{2 \times V(C_{ground\ truth} \cap C_{proposed\ method})}{V(C_{ground\ truth}) + V(C_{proposed\ method})}, \quad (4)$$

where V indicates the volume of the region enclosed in the ground truth contour ($C_{ground\ truth}$), proposed method contour ($C_{proposed\ method}$), or their shared area ($C_{ground\ truth} \cap C_{proposed\ method}$). Sensitivity measures the portion of the ground truth contour that was correctly occupied by the proposed method. Specificity measures the portion of the ground truth volume that is not contoured by the proposed method. They are defined as,

$$Sensitivity = \frac{V(C_{ground\ truth} \cap C_{proposed\ method})}{V(C_{ground\ truth})}, \quad (5)$$

$$Specificity = \frac{V(\bar{C}_{ground\ truth} \cap \bar{C}_{proposed\ method})}{V(\bar{C}_{ground\ truth})}, \quad (6)$$

where C and $\bar{C}$ indicate the region inside and outside of each contour, respectively. A magnitude DSC/sensitivity/specificity closer to 1 indicates greater overlap with ground truth,

and thus greater accuracy of the proposed method. The difference between the target volume centroid of the ground truth and our results were measured in three orthogonal directions [left–right (L–R), anterior–posterior (A–P), and superior–inferior (S–I)]. In order to quantify the error in treatment isocenter setup, a 3D centroid distance was further calculated from centroid differences in all directions. The volume accuracy was evaluated by Pearson linear regression analysis between ground truth and the proposed method among all patients, and its correlation coefficient was calculated. A correlation coefficient closer to 1 indicates higher accuracy of the proposed method. A student's *t*-test was further applied on centroid and volume difference to determine whether our results were significantly different from ground truth.

In addition, the dosimetric consequence of the proposed method was evaluated by comparing the dose distribution of the original treatment plan (based on AVM contours of ground truth) to that derived from our results. Clinical relevant dose-volume histogram (DVH) metrics were extracted for comparison.

3. RESULTS

Representative segmentation results of the proposed method are compared side-by-side with the clinical ground truth at different axial slices for 2 patients (Fig. III). The contours from our proposed method closely resemble those of the ground truth. Our method successfully differentiated the AVM from surrounding normal vasculature in patient #1 (upper panel). Minimal difference can be seen with respect to

the small details of curvatures. DSC/sensitivity/specificity are 0.892/0.953/0.996 and 0.878/0.921/0.996, 3D centroid distances were 0.401 mm and 0.250 mm, and volume errors were 0.216 and 0.150 cc for patients #1 and #2, respectively.

The DSC, sensitivity and specificity for all 80 patients is plotted in Fig. III and summarized in Table III. All metrics are greater than 0.75 for all patients, with mean values greater than 0.85. These results quantitatively demonstrate the accuracy of contours delineated by our proposed method.

The differences between ground truth and our results for target volume centroid in three orthogonal directions as well as for 3D centroid distance for all 80 patients are shown in Fig. III. No obvious bias can be seen in any direction, and the ranges of centroid differences are similar among the three directions, with maximum around 1 mm. As listed in Table III, there were no statistically significant differences found for all three directions. The error in treatment isocenter setup can be estimated by the 3D centroid distance, which was 0.675 ± 0.401 mm on average.

The mean contoured volume of our method was 5.459 ± 5.451 cc, corresponding well with the ground truth mean volume of 5.383 ± 5.586 cc. The linear correlation of target volumes between the two groups (Fig. III) is 0.992, statistically significant at $P < 0.001$. As seen in Table III, the difference in size of the contoured target volumes between the two groups was 0.076 ± 0.728 cc, which was not statistically significant ($P = 0.387$).

Figure III shows the dosimetric comparison on patient #1. The corresponding DVH curves for AVM target coverage for ground truth and our method were essentially identical. The

FIG. 2. The axial CT images showing the ground truth AVM contours and proposed method contours of patient #1 (upper) and #2 (lower) at different slices. [Color figure can be viewed at wileyonlinelibrary.com]

differences in DVH metrics among all patients are shown by the box plot in Fig. III. The statistics of DVH metric differences among patients are summarized in Table III. The average dose difference was less than 0.2 Gy, with standard deviation less than 1 Gy. There were no statistically significant differences in any of the selected metrics (P -value > 0.05).

4. CONCLUSIONS AND DISCUSSION

In this study, we proposed a novel deep learning-based method to segment an AVM target volume based on CT simulation images. The mean DSC, sensitivity and specificity of the AVM contours delineated by our method were all greater than 0.85 for the entire cohort of 80 patients. The mean 3D centroid distance of AVM contours between the two groups was 0.675 ± 0.401 mm, with no statistically significant differences in any single direction. The correlation coefficient of the AVM volume between ground truth and the proposed method's results was 0.992 with $P < 0.001$, which indicates statistically significant correlation. The mean volume difference among all patients was 0.076 ± 0.728 cc ($P > 0.05$). The mean differences in DVH metrics were all less than

0.2 Gy ($P > 0.05$), with standard deviation less than 1 Gy. These results indicate quantitatively that the proposed method accurately reproduces the AVM volume on CT images. It also demonstrates the great potential of a learning-based AVM segmentation method for clinical use.

This paper is the first study of automatic AVM segmentation on CT images using a deep learning-based method. Compared to manual AVM contouring solely based on CT images, this automatic method achieves higher accuracy. Zhang *et al.* reported the average error of AVM centroid for manual contours only based on CT was around 0.5 mm with STD of 2.5 mm in any direction among 18 patients.¹⁵ Such a variance is too large to determine the position of the target. The average overlapping volume, that is, sensitivity values reported here, was $63.7 \pm 19.3\%$. The mean difference in our results was less than 0.1 mm with STD no more than 0.5 mm in any direction. This is at least five times as accurate in target positioning as manual results. The sensitivity was also 40% higher by our method.

The results of the proposed method can be further examined by comparing with interobserver variation in current manual AVM delineation based on DSA and CT. Buis *et al.* reported an average agreement ratio of 0.45 ± 0.18

FIG. 3. Box plot of DSC, sensitivity and specificity of AVM contours between ground truth and the results of the proposed method among all 80 patients. The central mark of each box indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the “+” symbol. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 4. Box plot of centroid differences in left-right (L-R), anterior-posterior (A-P) and superior-inferior (S-I) directions, and 3D centroid distance. The central mark of each box indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the “+” symbol. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE I. Quantitative metrics of AVM target difference among all 80 patients.

	DSC	Sensitivity	Specificity	Centroid Error (mm)				Volume Error (cc)
				L-R	A-P	S-I	3D Distance	
Mean	0.852	0.880	0.990	0.034	0.005	0.022	0.675	0.076
SD	0.041	0.056	0.011	0.414	0.482	0.430	0.401	0.728
P -value ^a	N/A	N/A	N/A	0.734	0.800	0.752	N/A	0.387

L-R, left-right; A-P, anterior-posterior; S-I, superior-inferior; AVM, arteriovenous malformations; DSC, Dice similarity coefficient.

^aThe P values were calculated from the t-test of the corresponding metrics between ground truth and our results.

(equivalent to DSC of 0.62 ± 0.12) among all possible pairs with 6 observers and 31 patients. Average 3D distance of AVM target centroid was 2.8 ± 2.6 mm, and the mean dose coverage of these contours in the study was around 10% lower than original plan.²⁸ Forkert *et al.* showed an average DSC of 0.830 ± 0.079 between 2 observers on 15 patients.²⁹ Al-Shahi *et al.* found that the standard deviation of the difference of transverse nidus size between paired observers among 5 observers on 40 patients was around 10 mm, which is equivalent to a standard deviation of 3.5 cc in volume difference.³⁰ The corresponding results by our method are close or better when compared with the above findings. It indicates that manual contouring is prone to the random error from different operators. The method proposed here can provide an observer-independent segmentation method to improve reproducibility and efficiency with comparable accuracy.

In our proposed method, the deep supervision is combined with a compound loss and integrated into the network. We justified the necessity of the addition of deep supervision by investigating its resulted improvement in segmentation accuracy. We applied our network without deep supervision

FIG. 5. Linear regression analysis of target volume between the ground truth and proposed method. Blue circles indicate individual patient measurement; the dashed red line is the line of identity. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 6. Left: dose distribution from original plan with target contours of ground truth (red) and our results (blue). Right: DVH comparison of dose in target contours of ground truth (red) and our results (blue). [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 7. DVH metrics comparison for the AVM target volume for all patients. The central mark of each box indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the “+” symbol. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II. DVH metrics of AVM target difference among all 80 patients.

	D99	D95	Dmean	Dmax
Mean (Gy)	−0.153	−0.063	−0.026	−0.010
±SD (Gy)	0.929	0.343	0.128	0.037
P-value	0.188	0.139	0.094	0.052

DVH, dose-volume histogram; AVM, arteriovenous malformations.

TABLE III. DSC of segmentation results among all 80 patients using current VOI and VOI shifted up to 10 voxels in all directions.

DSC	Current VOI	Shifted voxels in all directions				
		2	4	6	8	10
Average	0.852	0.851	0.847	0.847	0.847	0.848
SD	0.041	0.041	0.043	0.043	0.042	0.041
P-value ^a		0.883	0.113	0.096	0.078	0.099

VOI, volume of interest; DSC, dice similarity coefficient.

^aThe P values were calculated from the t-test of the DSC results between current VOI and the VOI with corresponding shifts.

among all the 80 patients, and the average DSC of segmentation results is 0.808 ± 0.071 , which is lower than the results of our proposed “deep supervised V-net” (0.852 ± 0.041 as presented above), and such a decrease in DSC is statistically significant (P -value <0.001). Thus, we can conclude that the deep supervision step improves the segmentation results.

In our method, the physicians estimated the general position of AVM target for each patient on CT images, and then set a VOI of $64 \times 64 \times 64$ voxels to enclose and center the target for segmentation. We found that estimating the general position of AVM target involves minimal variation for physicians since they usually have the information about target from diagnosis. The average 3D deviation of VOI center from target centroid in our study was 7.2 ± 5.6 voxels, which is equivalent to 5.1 ± 4.1 mm. To evaluate the sensitivity of our method to VOI location selection, we applied a similar test study as above but we shifted the current VOI selections of all patients in the three orthogonal directions by 2, 4, 6, 8 and 10 voxels to simulate different VOI selections. Table III shows the DSC of the segmentation results of shifted VOI. Compared with the current results, the results of voxel shift do not show a significant decrease in DSC (P -value >0.05). Thus, we can conclude that the performance of our method is not sensitive to the VOI location selection.

The size of AVMs in this study ranged widely from $8 \times 8 \times 8$ voxels to $35 \times 35 \times 35$ voxels, all of which fit in the $64 \times 64 \times 64$ voxels VOI size. Note that we used the $64 \times 64 \times 64$ voxels VOI size to test the feasibility of our method. In clinical application, such a VOI size can be changed appropriately to fit the size of AVM by physicians. As long as the portion between AVM and VOI is reasonable, we expect the performance of our method is not sensitive to the VOI size selection.

The proposed method has critical clinical utility in simplifying the AVM treatment planning workflow since it provides accurate AVM target volumes based solely on the CT simulation image. For physicians and physicists, treatment plan parameters (isocenter, treatment modality, etc.) can be determined rapidly after CT simulation without the need to await DSA acquisition. A separate procedure can entail logistical issues such as patient transportation, interdepartmental communication, etc. If the physician does not feel confident with the segmentation result in certain circumstances, a DSA can still be obtained after CT. In this situation, our method provides a reference contour based on the experience of previously treated patients. For patients, it is highly possible to avoid an invasive procedure and shorten the time staying in clinic.

In this study, we proposed a novel method for automatic AVM segmentation, and demonstrated its feasibility using data from 80 clinical patient cases. Future study would involve a comprehensive evaluation with a larger cohort of patients with diverse disease characteristic, including variables such as hemorrhage status, embolization status, diffuse or compact nidus. This study validated the proposed method by quantifying the shape similarity of contours. Small differences from ground truth were observed; however, its potential

clinical impact due to such differences in dose coverage is not yet understood. Thus, a further investigation in clinical outcomes of the proposed method in SRS treatment planning would be of great interest and needed for eventual adoption in general clinical use.

ACKNOWLEDGMENTS

This research was supported in part by the National Cancer Institute of the National Institutes of Health Award Number R01CA215718 and the Emory Winship Cancer Institute pilot grant. We are also grateful for the GPU support from NVIDIA Corporation.

*Equal contribution, co-first author.

^{a)}Author to whom correspondence should be addressed. Electronic mail: xiaofeng.yang@emory.edu; Telephone: 404-778-8622.

REFERENCES

1. Friedman WA, Bova FJ. Radiosurgery for arteriovenous malformations. *Neurol Res.* 2011;33:803–819.
2. Schulder M. *Handbook of Stereotactic and Functional Neurosurgery*. Boca Raton, FL: CRC Press; 2003.
3. Flickinger JC, Kondziolka D, Maitz AH, Lunsford LD. An analysis of the dose-response for arteriovenous malformation radiosurgery and other factors affecting obliteration. *Radiother Oncol.* 2002;63:347–354.
4. Gallina P, Gallina P, Merienne L, et al. Failure in radiosurgery treatment of cerebral arteriovenous malformations. *Neurosurgery.* 1998;42:996–1002.
5. Khajuria R, Gross BA, Du R. Chapter 12 - Image-Guided Open Cerebrovascular Surgery. In: Golby AJ. ed. *Image-Guided Neurosurgery*. Boston, MA: Academic Press; 2015:277–296.
6. Spiegelmann R, Friedman WA, Bova FJ. Limitations of angiographic target localization in planning radiosurgical treatment. *Neurosurgery.* 1992;30:619–623.
7. Bova FJ, Friedman WA. Stereotactic angiography: an inadequate database for radiosurgery? *Int J Radiat Oncol Biol Phys.* 1991;20:891–895.
8. Tomycz L, Bansal NK, Hawley CR, Goddard TL, Ayad MJ, Mericle RA. “Real-world” comparison of non-invasive imaging to conventional catheter angiography in the diagnosis of cerebral aneurysms. *Surgical neurology international.* 2011;2:134.
9. Cloft HJ, Joseph GJ, Dion JE. Risk of cerebral angiography in patients with subarachnoid hemorrhage, cerebral aneurysm, and arteriovenous malformation: a meta-analysis. *Stroke.* 1999;30:317–320.
10. Leffers AM, Wagner A. Neurologic complications of cerebral angiography. A retrospective study of complication rate and patient risk factors. *Acta Radiol (Stockholm, Sweden: 1987).* 2000;41:204–210.
11. Ringer AJ, Lanzino G, Veznedaroglu E, et al. Does angiographic surveillance pose a risk in the management of coiled intracranial aneurysms? A multicenter study of 2243 patients. *Neurosurgery.* 2008;63:845–849.
12. Yi HJ, Sung JH, Lee DH, Kim SW, Lee SW. Analysis of radiation doses and dose reduction strategies during cerebral digital subtraction angiography. *World Neurosurg.* 2017;100:216–223.
13. Manninen AL, Isokangas JM, Karttunen A, Siniluoto T, Nieminen MT. A comparison of radiation exposure between diagnostic CTA and DSA examinations of cerebral and cervicocerebral vessels. *AJNR Am J Neuroradiol.* 2012;33:2038–2042.
14. Schueler BA, Kallmes DF, Cloft HJ. 3D cerebral angiography: radiation dose comparison with digital subtraction angiography. *AJNR Am J Neuroradiol.* 2005;26:1898–1901.
15. Zhang XQ, Shirato H, Aoyama H, et al. Clinical significance of 3D reconstruction of arteriovenous malformation using digital subtraction angiography and its modification with CT information in stereotactic radiosurgery. *Int J Radiat Oncol Biol Phys.* 2003;57:1392–1399.

16. Sharma N, Aggarwal LM. Automated medical image segmentation techniques. *J Med Phys.* 2010;35:3–14.
17. Dong X, Lei Y, Wang T, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. *Med Phys.* 2019;46:2157–2168.
18. Wang T, Lei Y, Tang H, et al. A learning-based automatic segmentation and quantification method on left ventricle in gated myocardial perfusion SPECT imaging: A feasibility study. *Journal of Nuclear Cardiology.* 2019. <https://doi.org/10.1007/s12350-019-01594-2>
19. Wang B, Lei Y, Tian S, et al. Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation. *Med Phys.* 2019;46:1707–1718.
20. Lei Y, Wang T, Wang B, et al. Ultrasound prostate segmentation based on 3D V-Net with deep supervision. *SPIE Medical Imaging.* 2019;10955.
21. Wang B, Lei Y, Wang T, et al. Automated prostate segmentation of volumetric CT images using 3D deeply supervised dilated FCN. *SPIE Medical Imaging.* 2019;10949.
22. Zabihollahy F, White JA, Ukwatta E. Convolutional neural network-based approach for segmentation of left ventricle myocardial scar from 3D late gadolinium enhancement MR images. *Med Phys.* 2019;46:1740–1751.
23. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. International Conference on 3D Vision. 2016.
24. Dou Q, Yu L, Chen H, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal.* 2017;41:40–54.
25. Milletari F, Navab N, Ahmadi S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. Paper presented at: 2016 Fourth International Conference on 3D Vision (3DV); 25–28 Oct. 2016, 2016.
26. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27–30 June 2016, 2016.
27. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention*. Berlin, Germany: Springer; 2015:234–241.
28. Buis DR, Lagerwaard FJ, Barkhof F. Stereotactic radiosurgery for brain AVMs: role of interobserver variation in target definition on digital subtraction angiography. *Int J Radiat Oncol Biol Phys.* 2005;62:246–252.
29. Forkert ND, Illies T, Goebell E, Fiehler J, Säring D, Handels H. Computer-aided nidus segmentation and angiographic characterization of arteriovenous malformations. *Int J Comput Assist Radiol Surg.* 2013;8:775–786.
30. Al-Shahi R, Pal N, Lewis SC, Bhattacharya JJ, Sellar RJ, Warlow CP. Observer agreement in the angiographic assessment of arteriovenous malformations of the brain [published online ahead of print 2002/06/08]. *Stroke.* 2002;33:1501–1508.