3D Reconstruction from 2D Cerebral Angiograms as a Volumetric Denoising Problem

Sean Wu[1(✉)], Naoki Kaneko[2], Steve Mendoza[1], David S. Liebeskind[3], and Fabien Scalzo[1,3]

[1] Keck Data Science Institute, Pepperdine University, Malibu, CA 90265, USA
sean.wu@pepperdine.edu
[2] Department of Interventional Neuroradiology, UCLA, Los Angeles, CA 90095, USA
[3] Department of Neurology, UCLA, Los Angeles, CA 90095, USA

Abstract. Accurately capturing the 3D geometry of the brain's blood vessels is critical in helping neuro-interventionalists to identify and treat neurovascular disorders, such as stroke and aneurysms. Currently, the gold standard for obtaining a 3D representation of angiograms is through the process of 3D rotational angiography, a timely process requiring expensive machinery, which is also associated with high radiation exposure to the patient. In this research, we propose a new technique for reconstructing 3D volumes from 2D angiographic images, thereby reducing harmful X-ray radiation exposure. Our approach involves parameterizing the input data as a back-projected noisy volume from the images, which is then fed into a 3D denoising autoencoder. Through this method, we have achieved clinically relevant reconstructions with varying amounts of 2D projections from 49 patients. Additionally, our 3D denoising autoencoder outperformed previous generative models in biplane reconstruction by 15.51% for intersection over union (IOU) and 3.5% in pixel accuracy due to keeping a semi-accurate input with back projection. This research highlights the significant role of back-projection in achieving relative visual correspondence in the input space to reconstruct 3D volumes from 2D angiograms. This approach has the potential to be deployed in future neurovascular surgery, where 3D volumes of the patient's brain blood vessels can be visualized with less X-ray radiation exposure time.

Keywords: 3D Reconstruction · Deep Learning · Denoising Autoencoder · Rotational Angiography · Cerebral Angiograms

1 Introduction

Digital subtraction angiography (DSA) is a medical imaging technique that gives clinicians insightful information regarding the blood vessels in a patient's brain. DSA operates by injecting a contrast agent into the patient's bloodstream and acquiring an X-ray scan to obtain a clear image of the blood vessels after digital subtraction [1]. DSA is typically used to monitor the blood flow in the brain and

G. Bebis et al. (Eds.): ISVC 2023, LNCS 14361, pp. 382–393, 2023.
https://doi.org/10.1007/978-3-031-47969-4_30

allows for radiologists to have a stronger understanding of certain cerebral disorders such as ischemic stroke and brain aneurysms [2]. 3D rotational angiography remains the gold standard for cerebral aneurysms and other diseases [3] and extends the concept of DSA to a volumetric visualization of the brain's blood vessels. Rotational angiography generates a 3D mesh of the brain's blood vessels by using a C-arm to capture the brain at different angles in a time series while the contrast agent is flowing through it [4].

While 3D rotational angiography provides crucial information regarding the blood flow in the brain in an interpretable modality, acquiring these volumetric meshes cannot be done during surgical procedures due to large machinery and lengthy procedures. Additionally, there may be safety concerns regarding the long acquisition times that could expose patients to high radiation levels for extended periods of time. While X-rays can provide highly detailed insights into bone and vessel structure, the ionizing radiation emitted from X-Ray beams can be so harmful that when passed through the body, they may cause mutations in cells and molecules, potentially causing cancer and other diseases [5]. This is precisely why we are interested in creating a deep-learning pipeline that can accurately reconstruct the three-dimensional CT angiogram from fewer two-dimensional images to reduce the X-Ray exposure time for patients and also to help radiologists evaluate neurovascular properties given these sparse viewpoints. By performing 3D reconstruction, we also hope to eventually be able to visualize blood flow properties in the patient that cannot usually be interpreted from a 2D DSA acquisition. This paper also discusses the minimum number of viewpoints that provide a clinically relevant 3D reconstruction for future model training.

1.1 Related Works

3D object reconstruction from 2D images has been a popular field in computer vision for many years and exploded in popularity with the release of ShapeNet in 2015, where thousands of everyday objects were being reconstructed with various algorithms [6]. Many works have solved the single and multi-view reconstructive problems with high precision on real-world landscapes and commodities all around us such as airplanes, clothing, and even people [7,8]. Some papers deal with the 3D reconstruction of images through an implicit encoder-decoder-like architecture, where the encoder learns to embed the features of the 2D image to a lower dimensional space, then transforms it to a 3D surface through the decoder layers. An amazing breakthrough in 2020 was the neural radiance field paper (NeRF) [9] that aimed to reconstruct a scene's positional depth, lightning, and geometry from arbitrary camera angles. NeRF, at a high level, is mapping the 3D scene as a 5D continuous function in which both a 3D coordinate and a 2D camera angle are taken as input in tandem to generate a volume and color.

One method for 3D reconstruction has been a spherical deformation technique, where a Convolutional Neural Network (CNN) outputs the vertex offset for each point and edits the original sphere based on its predictions [10]. Although 3D reconstruction is a well-researched field with many new contributions each year, there have been few studies applying deep 3D reconstruction algorithms to specific

medical tasks such as angiographic data because of expensive medical machinery and sensitive patient data. There have been attempts at this task, such as a contour-interpolation [11], and even self-supervised deep learning approaches [12] that have solved the 3D reconstruction task with 4–12 projections. The self-supervised paper [12] created a 2D to 3D reconstruction pipeline of angiographic images by leveraging the input digital subtracted angiogram images as a learning objective, meaning they took a 2D acquisition from the predicted 3D volume and compared it to the original input image. We present a supervised 2D to 3D reconstruction algorithm because we had access to high-quality volumetric 3D data obtained from a university medical center, which gives our model a clear learning objective. There have been few works researched in angiogram reconstruction because of the difficulties to obtain high-quality volumetric datasets from medical centers. Other papers [13] have attempted this challenging task by implementing a generative adversarial network [14] to generate angiographic slices. However, due to the nature of this generative approach, this resulted in a significant amount of noise and inaccurate geometry in the reconstructed volumes. In order to tackle this challenging problem, we first consider the 2D to 3D reconstruction task using different numbers of projections to observe how many projections are needed for a non-generative model to converge. Additionally, we attempt to tackle the biplane problem by creating a deep-learning model capable of performing the reconstruction task with just two projections.

1.2 Problem Definition

We define our research as creating a deep learning algorithm to perform 3D reconstruction from 2D projections of the brain's blood vessels. One reason we chose the angiogram medium for our study is the high-quality images captured by the X-ray machine. Reconstructing the blood vessels in the brain is a problem that is difficult due to the complexity and intricacy of finer blood vessels. However, we show that our proposed method of a deep learning algorithm can solve this complex task due to the nature of the high-quality images, and registration within angiographic images. Researchers are commonly interested in either explicit reconstruction, where convolutional layers are utilized to output a volume or implicit reconstruction is utilized, where every point in space is sampled to be either inside or outside of the mesh. In our case, we use explicit 3D reconstruction because sampling every point in space with an implicit decoder [15] for a $384 \times 384 \times 384$ volume can take roughly 56 million passes, which can make the model slow on inference time.

In this paper, we discover the minimum number of projections needed to reconstruct the 3D geometry of the brain's blood vessels accurately, and also present an interesting 2D to 3D reconstruction methodology that exploits the properties of computed tomography angiogram acquisitions to learn the input-output mapping function between a sets of angiographic images and the corresponding 3D CT volume. This project contains two primary steps. Firstly, the data is pre-processed to ensure exact visual correspondence between the input and output, eliminating the use of separate images, which poses an inherently ambiguous problem. Secondly,

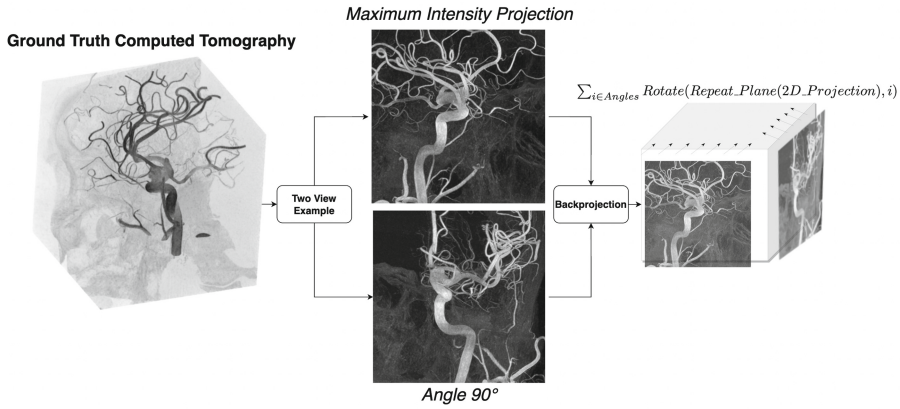

Fig. 1. Illustration of the pre-processing pipeline. We start with the given patient volume (1), then we take a maximum intensity projection (2) for all n viewpoints, then finally, we use back-projection to get a noisy volume (3).

we proceed to train a 3D denoising autoencoder to acquire a semi-photorealistic volume.

2D to a 3D reconstruction poses an inherent challenge due to the limited depth information available from a small number of views and the absence of visual correspondence between the input images and the ground truth volume. Without a perfect visual correspondence where the predicted volume is in spatial alignment with the original CT volume, it is difficult to compute a loss comparing the prediction with the ground truth for training deep learning models. In this paper, we present our methodology for transforming this ill-defined problem [16] into a 3D denoising autoencoder encoder [17]. Our objective is to achieve relative visual correspondence in the model's input space, enabling the network to primarily focus on learning the volume-to-volume translation process.

2 Methodology

To address this intricate problem, we implemented a back-projection technique utilizing angiographic images to obtain a "noisy" three-dimensional volume that matches with relative visual correspondence with the ground truth. This allowed us to obtain a voxel-based representation of the blood vessels, enabling us to apply convolutional operations on the 3D volumes using deep neural networks (Fig. 2).

2.1 Data Pre-processing

For this dataset, we acquired high-quality computed tomographic volumes in the form of DICOM files using [18] from 49 patients who were screened for various neurovascular conditions, including stroke and aneurysms in a clinical database at a university medical center. These CT scans are of size (384,384,384) and have

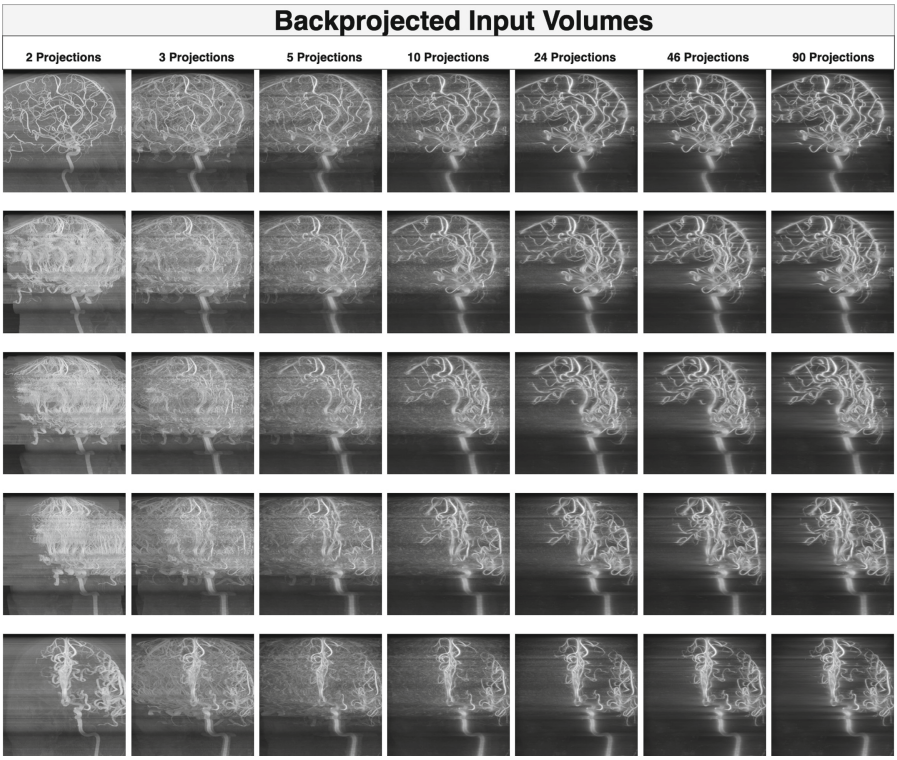

Fig. 2. Depiction of a max-projection at various angles (0,22.5,45,67.5,90) for the back-projected volumes at different viewpoints. The quality of the projection becomes visually worse with fewer than 5 projections.

already been pre-computed with a CT machine and contain information not only about the vessels but also about the skull and other elements in the patient's head. To obtain the 2D acquisitions from this pre-computed ground truth volume, we employed a max-projection technique with [19]. This involved taking the maximum value along the second dimension of an (x, y, z) volume, essentially collapsing the volume and providing us with a clear two-dimensional image of the blood vessels from that angle. We repeated this procedure 90 times for each patient so each patient held a total of 90 projections that we sampled from and utilized for training. In this study, we analyzed 2D to 3D reconstruction from 90, 46, 24, 10, 5, 3, and 2 projections to visualize how many projections are needed to obtain an accurate 3D reconstruction. In all cases, we used the first (0°) and last (90°) projections as input, alongside the other viewpoints obtained by looping through the 90 projections and evenly sampling images depending on how many projections we used. For example, to sample 46 images, we utilized every other angle from 0° to 89° while ensuring the first and last projections are always used.

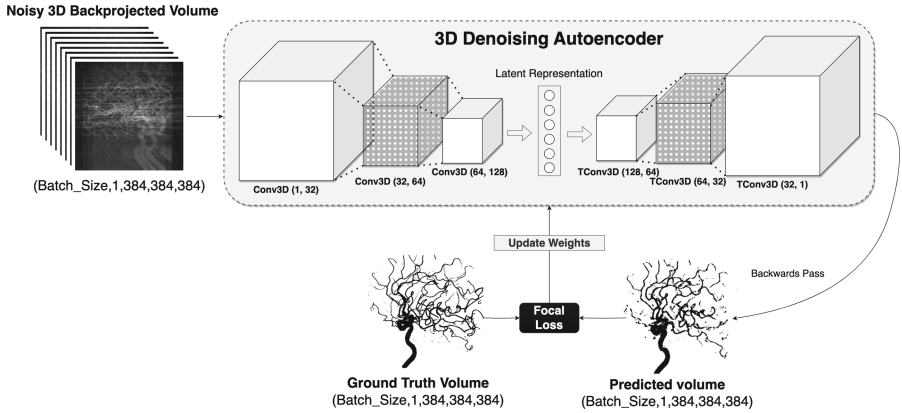

Fig. 3. End to end representation of the forward and backward passes of our 3D denoising autoencoder. (Real example from testing set)

To obtain input volumes, we apply a simple back-projection summation algorithm from n 2D projections by first initializing an empty volume of (384, 384, 384). Then we take each 2D image and repeat that image 384 times to get a volume (384, 384, 384). Finally, we rotate volume based on what projection the 2D image was taken at and sum it to the initialized volume (Fig. 1). By doing this, we ensure that the blood vessels will have larger numbers in the summation, therefore giving us some visual corespondence in the input space. We follow this process where for 90, 46, 24, 10, 5, 3, and 2 projections. Finally, we standardized each volume to hold voxel values between 0 and 1. To obtain ground truth binary data, we appy Otsu's segmentation model [20] from the scikit-learn package [21] to all of the original .dcm files to obtain 3D binary masks as a learning objective. Given the limited availability of data, we adopt an approximate 0.9/0.1 training/testing split, omitting a separate validation set. There are no overlap between patient ids in the training and testing set. The testing set contains 5 unique patients. This decision ensures that we have a larger amount of data for training purposes (Fig. 3).

2.2 Deep Learning Methods

After obtaining the ground truth, we created a standard 3D denoising autoencoder to encode the features into a smaller latent feature tensor. We then used transposed convolution to reconstruct the volume into a denoised representation of it. Our model consists of three 3D convolutional layers. The first layer encoded from one channel to 32 channels, the second layer encoded from 32 channels to 64, and the third layer encoded from 64 channels to 128 output channels. After each convolutional layer, we included a 3D batch normalization layer to standardize the distribution of parameters to have zero mean and standard deviation of one. We also included the ReLU nonlinear activation function after each batch normalization layer to introduce nonlinearity to the network. In addition, we included a

3D max pooling layer after the first two convolutional layers to perform dimensionality reduction on the tensors in the model's forward pass. Our 3D decoder was created in a similar fashion, where we utilized the transposed convolution in PyTorch [22] with kernel size and stride of 2. Our network consists of 360,481 training parameters, each input size is 226.49 (MB), the forward/backwards pass size is 43.034 (GB), and the estimated total size is 43.261 (GB).

2.3 Model Training and Hyper-parameters

To train our 3D denoising autoencoder, we set the batch size to just 1 to stay within the computational limits of our RTX A6000 GPU and set the learning rate to 1e−3 with the Adam optimizer [23]. We chose the focal loss function [24], primarily used for semantic segmentation, as the criterion for our models.

$$\text{Focal Loss(pred)} = -\alpha(1 - pred)^{\gamma} \log(pred)$$

This function assigns weights and gives greater attention to data samples that are misclassified, emphasizing more on challenging tasks. The variable "pred" represents the probability of the binary prediction, assuming that it has already been passed through a sigmoid or equivalent activation function (in this case, BCE with logits). The parameter α correlates to the balance between precision and recall of positive class errors, while γ determines the tradeoff for penalizing hard-to-classify data versus non-hard-to-classify data. In our case, we left the α and γ parameters to their default values of 0.5 and 0.2, respectively. Finally, we utilized mixed precision training [25] to train our 3D denoising autoencoder due to the massive size of the (384,384,384) input tensors to our network. This allows us to keep the quality of the original angiographic data, while still having enough GPU memory to compute the forward and backwards passes of the model. Mixed precision training leverages both float16 (half-float precision) and float32/float64 (full and double precision) during network training to significantly reduce the training time by nearly half.

3 Results

In this paper, five unique patients were in the testing set to evaluate how well our 3D denoising autoencoder can generalize on unseen data. For precise analysis, we test each volume with the mean squared error (MSE), mean absolute error (MAE), and intersection over union (IoU) [26], which is a commonly used metric for semantic segmentation tasks. We can use metrics such as this because we are confident that the predicted volumes have an extremely close spatial orientation to the ground truth volumes as a result of our back-projecting technique. The output of our denoising autoencoder is a volume of shape (384, 384, 384) of probabilities, $>= 0.5$ being in the surface and < 0.5 being part of the background. To test the model, we first thresholded each predicted volume to be binary based on these criteria, then computed the metrics to score the prediction based on IoU, MSE, and MAE.

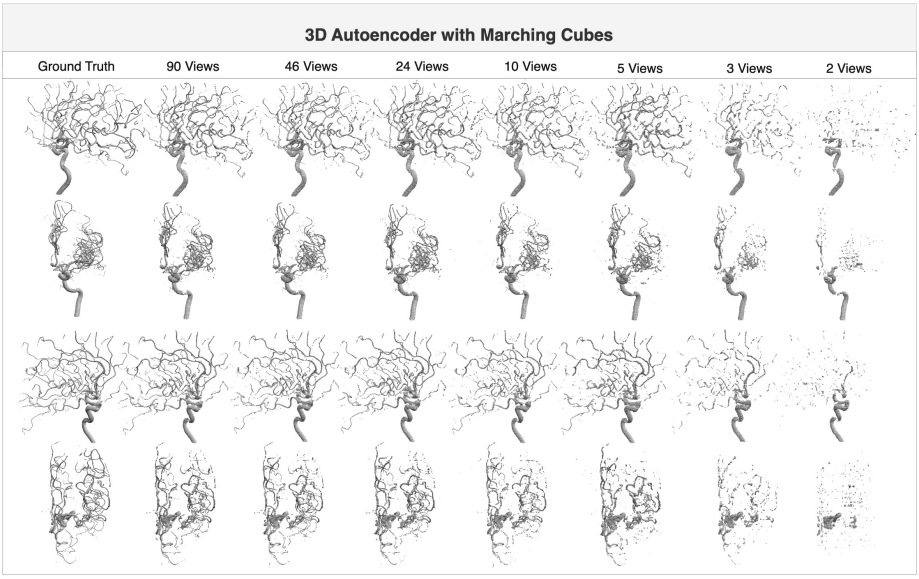

Fig. 4. Visualization of the predicted angiographic volumes from a few different viewpoints of one of the samples in our testing set. The left column represents the ground truth, and the remaining columns are the model outputs at various projections.

The results of our experiments are as expected, with 90, 46, 24, and 10 projections all performing similarly in terms of IoU. While there is a slight decrease in IoU from 90 to 10 views, the fluctuation of IoU in 90, 46, and 24 projections can be accounted for with the random weight initialization (Table 1). To visualize our final outputs, we used the marching cubes [27] surface reconstruction algorithm to extract vertices and faces of the numpy volumes. In this study, we compare our findings with a similar paper by [13], which is one of the only similar papers we could find on the task of biplane 2D to 3D reconstruction. This previous work utilized a 2D conditional generative adversarial network to generate the volume slice-wise. Our 3D denoising algorithm reports a pixel accuracy of 0.9985 and an IoU score of 0.2683 (Table 2), significantly improving both metrics from the previous works.

4 Discussion

In this paper, we have presented a novel technique of supervised 2D to 3D reconstruction given angiographic images by back-projecting the viewpoints into a noisy volume to keep visual correspondence in the input and ground truth, then we train a 3D denoising autoencoder to learn the correct volume translation to the binary ground truth. Our results show that it is possible to train a clinically accurate 3D reconstruction from 90, 46, 24, 10, and even 5 2D projections. However less projections (3, 2) fail to reconstruct fine vessels (Fig. 4), and demonstrates large

rooms for improvement. Additionally, we have shown that this approach has an improved pixel accuracy and IoU score than previous generative methods. While there are a few possible reasons for this improvement, it is possible that a generative model is not ideal for this case due to it "guessing" the unseen vessels based on the information learned during training. However, in this case, that may mean generating realistic vessels that do not actually match the geometry of the input images, meaning scarce vessels will have a higher IoU score than incorrectly predicted vessels.

The results of our paper have some implications for future work on reconstructing 3D geometry from fewer projections, such as two or three views. Rather than combating the ill-defined 2D to 3D reconstruction task, we transformed the problem into a 3D to 3D denoising problem, allowing us to keep relative geometry between the two orthogonal images. This may mean that with cleaner data/more computational power, the results may be convincing enough to be deployed in clinical practice. Despite the successes of outperforming previous generative models in a similar 2D to 3D reconstruction task of angiographic images, some limitations to our research are important to discuss. Firstly having only 49 patients may affect our model's generalizability because each CT volume has a different zooming factor, and the vessel structures are highly variant.

Additionally, we did not apply data augmentations in this study due to the computational time required to rotate or transform a 3D volume of (384, 384, 384) for every batch in our training data. Another limitation to discuss is the batch size of the one used for this research. To prevent a loss of information for the intricate and fine details of the blood vessels, we decided to keep the original quality of the data (384, 384, 384) volumes for training. However, because of this, we were required to set the batch size to one to match the memory requirements of our GPU. Some issue that may have arisen from the small batch size is noisy gradients, the gradient computed for each volume may be noisy and not be explanative of the dataset as a whole. While this paper shows a solid proof of concept for this method of 2D to 3D reconstruction transformed as a denoising problem, some

Table 1. Table representing the error and intersection of union (IoU) scores for each model.

Number of Projections	IoU	MSE	MAE
Ninety	56.81%	0.00124	0.0196
Forty-Six	57.06%	0.00103	0.0153
Twenty-Four	56.43%	0.001396	0.0197
Ten	52.22%	0.00173	0.0241
Five	45.35%	0.00277	0.0275
Three	32.17%	0.00224	0.0256
Two	26.83%	0.00264	0.02516

Table 2. Depiction of our biplane (two-view) reconstruction results where we surpass previous generative methods.

Biplane (Two-Projection) Analysis				
Model Type	Pixel Accuracy	IoU	MSE	MAE
ConditionalGan [13]	0.9631	0.0726	N/A	N/A
L1 + ConditionalGAN [13]	0.9587	0.1132	N/A	N/A
3D Denoising (Ours)	0.9985	0.2683	0.0026	0.0251

significant areas and methods can improve the reconstruction to get a watertight volume almost indistinguishable from the ground truth.

In this study, it is evident that the biplane reconstruction task has not been solved. There is naturally a loss of information when back projecting two orthogonal viewpoints into one volume. This loss of information on the 180° angle from the images is due to the binary thresholding. One possible improvement to our pipeline to mitigate this loss is to include another input channel for the original grayscale angiographic images. This additional channel would retain depth and fine detail for all of the vessels. Previous studies, specifically [12] have shown remarkable results for reconstructing the 3D volume from 4+ viewpoints. One interesting method for future research is generating the 2D straight angle (180°) given a single angiographic image. By doing this successfully, it is possible to have two natural and two artificial viewpoints for a new deep-learning model, and we expect the results to be much improved. For clinically relevant medical, the geometry of the reconstructed vessels must match what is possible in the physical realm. Therefore for future projects, we hope to constrain the predictions of the network with a physical-based loss function to our model to penalize predicted vessels that deviate from the dictionary of valid vessel geometries.

5 Conclusion

This paper presents a new methodology for performing 2D to 3D reconstruction of angiographic images by utilizing the properties of the 2D max-projected images and transforming the 2D to 3D problem into a 3D to 3D denoising task. We also conclude that without a generative model, at least three projections at approximately (0, 45, 90) degrees should be utilized to obtain a reasonable 3D reconstruction. Although preliminary, our network may be utilized for future works where more high-quality data, better computational resources, and additional input channels are obtained to produce higher-quality reconstructions where more depth information is recovered. This research utilized a max-projected angiographic image from the ground truth computed tomography volumes. However, due to the nature of our back-projection algorithm, our pipeline should also be compatible with digital subtraction angiograms or other mediums where vessel information is captured in a rotational fashion. With further improvements to the network in future iterations of the project, the problem of 2D to 3D reconstruction

of angiographic images will have major impacts by decreasing radiation exposure and the need for expensive machinery in neurological procedures. This will help neurosurgeons better visualize many disorders such as aneurysms, blood clots, and strokes.

Acknowledgements. We would like to thank the Keck Foundation for their grant to Pepperdine University to support our Data Science program and this research.

References

1. Scalzo, F., Liebeskind, D.S., et al.: Perfusion angiography in acute ischemic stroke. Comput. Math. Methods Med. **2016** (2016)
2. Cieściński, J., Serafin, Z., Strześniewski, P., Lasek, W., Beuth, W.: DSA volumetric 3D reconstructions of intracranial aneurysms: a pictorial essay. Pol. J. Radiol. **77**, 47 (2012)
3. van Rooij, W.J., Sprengers, M., de Gast, A.N., Peluso, J., Sluzewski, M.: 3D rotational angiography: the new gold standard in the detection of additional intracranial aneurysms. Am. J. Neuroradiol. **29**, 976–979 (2008)
4. Ishihara, S., Ross, I., Piotin, M., Weill, A., Aerts, H., Moret, J.: 3D rotational angiography: recent experience in the evaluation of cerebral aneurysms for treatment. Interv. Neuroradiol. **6**, 85–94 (2000)
5. Frenz, M., Mee, A.: Diagnostic radiation exposure and cancer risk. Gut **54**, 889–890 (2005)
6. Chang, A.X., et al.: Shapenet: An information-rich 3D model repository. arXiv preprint arXiv:1512.03012 (2015)
7. Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H.: PIFu: pixel-aligned implicit function for high-resolution clothed human digitization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2304–2314 (2019)
8. Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G.: Pixel2mesh: generating 3d mesh models from single RGB images. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 52–67 (2018)
9. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: representing scenes as neural radiance fields for view synthesis. Commun. ACM **65**, 99–106 (2021)
10. Kato, H., Ushiku, Y., Harada, T.: Neural 3D mesh renderer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3707–3916 (2018)
11. Galassi, F., et al.: 3D reconstruction of coronary arteries from 2D angiographic projections using non-uniform rational basis splines (NURBS) for accurate modelling of coronary stenoses. PLoS ONE **13**, e0190650 (2018)
12. Zhao, H., et al.: Self-supervised learning enables 3D digital subtraction angiography reconstruction from ultra-sparse 2D projection views: a multicenter study. Cell Rep. Med. **3**, 100775 (2022)
13. Zuo, J.: 2D to 3D neurovascular reconstruction from biplane view via deep learning. In: 2021 2nd International Conference on Computing and Data Science (CDS), pp. 383–387. IEEE (2021)
14. Goodfellow, I.J., et al.: Generative adversarial networks (2014)

15. Chen, Z., Zhang, H.: Learning implicit fields for generative shape modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5939–5948 (2019)
16. Pontes, J.K., Kong, C., Eriksson, A., Fookes, C., Sridharan, S., Lucey, S.: Compact model representation for 3D reconstruction. arXiv preprint arXiv:1707.07360 (2017)
17. Venkataraman, P.: Image denoising using convolutional autoencoder. arXiv preprint arXiv:2207.11771 (2022)
18. Mason, D.: SU-E-T-33: pydicom: an open source DICOM library. Med. Phys. **38**, 3493–3493 (2011)
19. Harris, C.R., et al.: Array programming with NumPy. Nature **585**, 357–362 (2020)
20. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979)
21. Pedregosa, F., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
22. Paszke, A., et al.: Automatic differentiation in pytorch (2017)
23. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
24. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
25. Micikevicius, P., et al.: Mixed precision training. arXiv preprint arXiv:1710.03740 (2017)
26. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union: a metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 658–666 (2019)
27. Lorensen, W.E., Cline, H.E.: Marching cubes: a high resolution 3D surface construction algorithm. In: Seminal Graphics: Pioneering Efforts that Shaped the Field, pp. 347–353 (1998)