

Fourier Feature Network for 3D Vessel Reconstruction from Biplane Angiograms

Sean Wu

Pepperdine University

Naoki Kaneko

University of California, Los Angeles

David Liebeskind

University of California, Los Angeles

Fabien Scalzo

`fabien.scalzo@pepperdine.edu`

Pepperdine University

Research Article

Keywords: Fourier Feature Mapping, Computed Tomography, Deep Learning, Reconstruction

Posted Date: March 19th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4103573/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Fourier Feature Network for 3D Vessel Reconstruction from Biplane Angiograms

Sean Wu¹, Naoki Kaneko², David S. Liebeskind³, Fabien Scalzo^{1,3*}

¹*Keck Data Science Institute, Pepperdine University, Malibu, 90265 ,
CA, USA.

²Department of Interventional Neuroradiology, UCLA, Los Angeles,
90095, CA, USA.

³Department of Neurology, UCLA, Los Angeles, 90095, CA, USA.

*Corresponding author(s). E-mail(s): fabien.scalzo@pepperdine.edu;
Contributing authors: sean.wu@pepperdine.edu;
nkaneko@mednet.ucla.edu; davidliebeskind@yahoo.com;

Abstract

3D reconstruction of biplane cerebral angiograms remains a challenging, unsolved research problem due to the loss of depth information and the unknown pixelwise correlation between input images. The occlusions arising from only two views complicate the reconstruction of fine vessel details and the simultaneous addressing of inherent missing information. In this paper, we take an incremental step toward solving this problem by reconstructing the corresponding 2D slice of the cerebral angiogram using biplane 1D image data. We developed a coordinate-based neural network that encodes the 1D image data along with a deterministic Fourier feature mapping from a given input point, resulting in a slice reconstruction that is more spatially accurate. Using only one 1D row of biplane image data, our Fourier feature network reconstructed the corresponding volume slices with a peak signal-to-noise ratio (PSNR) of 26.32 ± 0.36 , a structural similarity index measure (SSIM) of 61.38 ± 1.79 , a mean squared error (MSE) of 0.0023 ± 0.0002 , and a mean absolute error (MAE) of 0.0364 ± 0.0029 . Our research has implications for future work aimed at improving backprojection-based reconstruction by first examining individual slices from 1D information as a prerequisite.

Keywords: Fourier Feature Mapping, Computed Tomography, Deep Learning, Reconstruction

1 Introduction

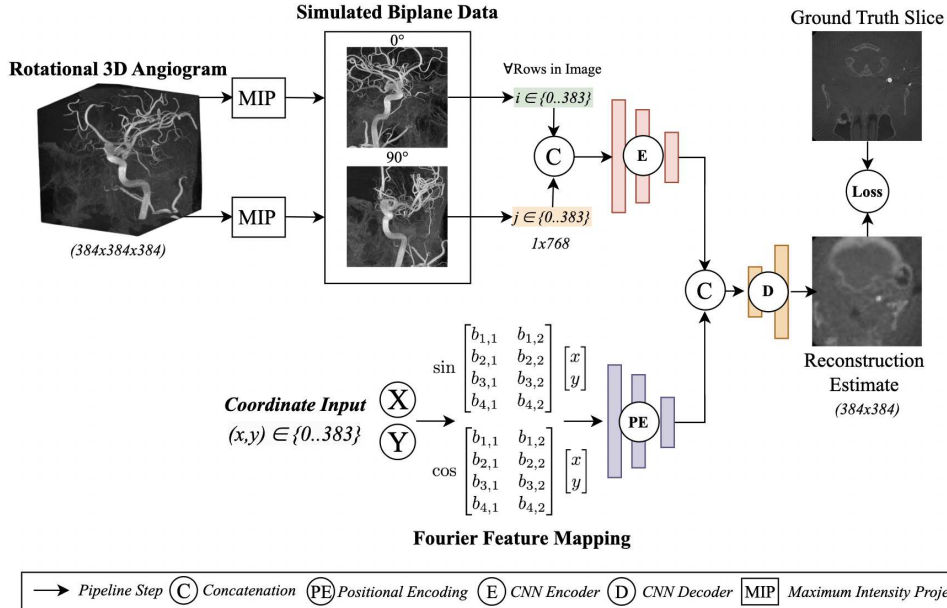


Fig. 1: General end-end overview of our research. (1) We first take a maximum intensity projection to obtain the simulated biplane data. (2) We then extract rows from both images, and by leveraging deep learning, we reconstruct the corresponding 3D slices.

Digital subtraction angiography (DSA) remains the gold standard for visualizing blood vessels in the brain due to its ability to provide high spatial resolution and fine vessel structure[1]. DSA of the brain is typically employed in biplane scenario, which involves visualizing 0th- and 90th-degree views of the brain. These views correspond to the coronal and sagittal planes that are typically used in magnetic resonance imaging (MRI) and computed tomography (CT). Cerebral angiograms are obtained by first injecting the patient with an iodine-based contrast agent[2], which then travels into the brain’s blood vessels. Once in the brain, blood flow in two dimensions is captured by temporal biplane X-ray acquisitions to visualize how the contrast agent flows through the vessels over time. Cerebral angiography provides an extremely precise and accurate representation of the brain’s blood vessels. This modality is important because it can help neurovascular surgeons identify various cerebral ailments, such as strokes, aneurysms, tumors, and malformations[3]. Computed tomography remains the state of the art methodology for transforming hundreds of acquisitions from cerebral angiography into a 3D representation of the vessels 3D[1]. Some classical CT reconstruction algorithms have such as the projection[4], shaded surface display[5], volume rendering[6], and curved planar reformation techniques[7], have been used to

reconstruct 3D geometry with a dense number of input images. While various CT reconstruction algorithms have proven to be extremely effective in successfully reconstructing the fine details of blood vessels[8], hundreds of X-ray acquisitions emit a noticeable amount of ionizing radiation that can in severe cases lead to cancer and other radiation-caused diseases[9]. Currently, blood flow is visualized, and flow parameters can be computed from biplane cerebral angiograms.

In this research, we present a new way to think about the biplane 3D reconstruction problem by breaking it down into a simpler research problem. Our method involves initially predicting individual 2D slices within the 3D volume from single rows of the image (1D Vector), utilizing Fourier features[10] as a positional encoding step to better localize the blood vessels on the 2D plane. For this step, we utilize a coordinate-based neural network with (x, y, z) coordinates as an additional input to our image row vector. The 2D angiographic data we use in our experiments are derived from utilizing the same maximum intensity projection as proposed in our previous paper [11], to simulate 2D data without the need for real-world rotational angiography. By doing so, we demonstrate improved slice reconstruction by leveraging Fourier feature mapping alongside a coordinate-based neural network, taking a step towards solving this difficult task. As a brief summary our contributions are as follows:

- Utilize maximum intensity projections to simulate 2D CT data from 3D volumes, as real-world rotational angiography is hard to obtain.
- This paper demonstrates the utility of leveraging a deterministic Fourier feature mapping through a coordinate-based neural network when predicting individual slices of cerebral angiograms.
- We show the need to reconstruct individual 2D individual slices before attempting 3D reconstruction.

2 Related Works

2D to 3D reconstruction is a well-defined problem in the field of computer vision, in which researchers are typically interested in reconstructing objects or scenes for single or multiview scenarios. Some popular baseline models include nondeep learning photogrammetry techniques such as multiview stereo[12] or structure from motion[13], but such methodologies are inadequate for reconstructing biplane cerebral angiograms due to severe occlusions from only two views. In addition, feature detection methods such as scale-invariant feature transform (SIFT), and ORB (Oriented FAST and Rotated BRIEF), may find it difficult to match features across the two images. State-of-the-art 3D reconstruction includes neural radiance fields[14] that predict a color and density value for each (x, y, z, θ, ϕ) input or even pixelNeRF[15], which can predict a neural representation from only one or multiple views without explicit 3D supervision and has exceeded benchmarks on ShapeNet[16]. Finally, the current state-of-the-art 2D to 3D reconstruction method is Gaussian splatting[17], a rasterization method contrary to NeRF ray tracing[[18], which utilizes structures from motion to predict a sparse point cloud and then “splats” and optimizes Gaussian distributions to reconstruct the scene.

However, such methods are not as useful for 3D biplane angiograms and most medical image reconstruction tasks because they involve many images and overfit a neural network and retrain it to each particular scene or object, resulting in less optimal generalizability to unseen examples. One study demonstrated a simple 3D reconstruction methodology for cerebral angiograms utilizing backprojection and a 3D to 3D denoising autoencoder, but it showed to fail in the biplane scenario[11]. Another study utilized conditional generative adversarial networks (pix2pix)[19] to generate binary segmentation maps and showed that generative models may not be suitable for this problem, as they produce realistic vessels that do not match the geometry of the input images. Because most medical volumes are naturally slicewise, we constructed a method to reconstruct individual slices, as doing so will allow more accurate biplane 3D reconstruction in future works. In addition, deep learning in the \mathbb{R}^2 space trains and scales better than in \mathbb{R}^3 , which can take weeks to train and has long inference times.

3 Methodology

In this research, we construct a Fourier feature network to reconstruct individual 2D slices of the volume. From the given biplane images, we extract row i from each image and pass it in tandem with our coordinate-based network to condition the network based on those row features. We compare these methods both quantitatively and qualitatively to the backprojection method proposed in our previous paper[11] and a standard 2D convolutional neural network[20]. A general overview of our methods is shown in Figure 1.

3.1 Dataset Acquisition and Preprocessing

In this research, we obtained 49 high-quality, precomputed CT images (384,384,384) from a clinical database at a university medical center as ground truth volumes. For the purpose of this study, we utilized the voxel 3D representation instead of the point cloud or surface representation, as it best fits the CT scans taken from the clinical database. To simulate 2D images, we take a maximum intensity projection (MIP) of the CT volumes to obtain an image for the front and side views to obtain biplane data. This methodology allows us to use real-world rotational angiography data because they are rather difficult to obtain.

We set our data for this research as follows. For each biplane pair (two images) of size 384x384, we take 1D rows of the image (1x384) and concatenate them together to obtain a vector of (1x768). For each row i in the images, we match it with the corresponding slice index in the ground truth volume for supervised training. We split our data into 45 training patients and 4 patients for testing. Each patient has two images and 384 training examples, resulting in a total of 17280 training examples and 1536 examples for testing.

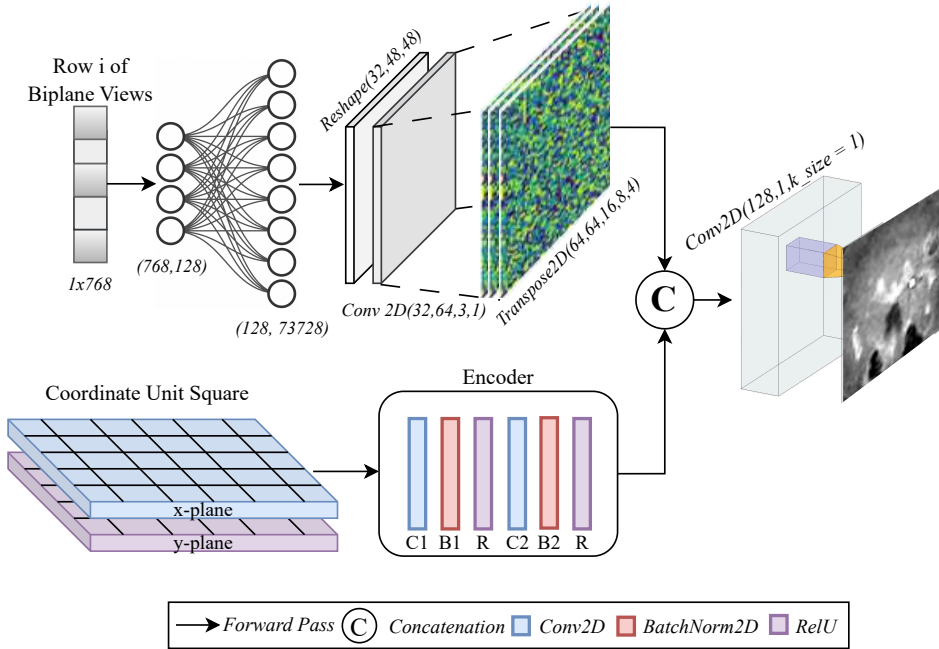


Fig. 2: Detailed visualization of our Fourier feature network, where we show the input channels, output channels, kernel size, padding, and stride for each convolutional layer.

3.2 Baseline Reconstruction

Our baseline 1D to 2D network consists of four transposed convolutional layers. The input to the network is a (1x768) vector, which is first passed through a linear layer, with the output shape being [-1, 147456]. These extracted features are then passed into 2D transposed convolutional layers with 256 input channels and 128 output channels with a kernel size of 4, stride of 2, and padding of 1. This layer is then passed to another transposed convolutional layer from 128 to 64 features with the same filter parameters and to another layer with 32 output channels. Between each convolutional layer, we utilized a 2D batch normalization layer to speed up convergence and a ReLU activation function. We then added a sigmoid activation function at the end of the network to scale the predicted slice between 0 and 1. The forward/backward pass of this network requires approximately 50 MB of memory, and the total size of the network is 435 MB.

3.3 Fourier Feature Network

Our Fourier feature network instead takes two inputs. A (1x768 vector), and a (x,y) coordinate. The output of that corresponding point is the pixel value corresponding to that point. The first input, which consists of concatenated image rows, is passed through two feed-forward layers, which output 737,728 features. We then reshape and

pass the network through a convolution layer with 32 channels and 64 output channels, with a kernel size of 3 and a padding of 1. These features are then upsampled through another convolutional layer and concatenated with coordinate-based positional encoding. This positional encoding will be explained in a later section. Finally, a convolutional layer is utilized to output the predicted slice.

3.3.1 Positional Encoding

In this research, we apply positional encoding for each (x,y) input in our network with Fourier features[10]. Instead of using the raw, low-dimensional (x,y) coordinates as input, we pass them through the Fourier feature mapping, which at a high level, is a set of sin and cos functions that are randomly rotated and scaled. Our positional encoding is deterministic and involves simple mapping from $\gamma(\mathbf{v}) = [\cos(2\pi B\mathbf{v}), \sin(2\pi B\mathbf{v})]^T$, where \mathbf{v} is the input point (x,y) . The B variable corresponds to a generated random Gaussian matrix, in which every sample is taken from a normal distribution $N(0, \sigma^2)$. Because the (x,y) are projected into a higher dimensional space, it allows for the network to learn higher frequency functions[10].

In practice, passing in each coordinate (x,y) into a network is too time and resource computing. For just one image, 147456 forward passes are required for the model. Instead, we created a 2-channel pixel grid, one representing the x-values and the other representing the y-values. In this way, we create a pixel grid in the unit square, which is the input to the model. This positioning encoding has additional parameters, namely, the mapping size and scale. We chose a mapping size of 128, which is the dimensionality of the transformed features, and a scale of 10, which is related to the frequency of sinusoids. The detailed implementation of the Fourier feature network is illustrated in Figure 2.

3.4 Training Details

The backprojection method[11] requires no learning or parameter updates. We trained both the baseline 2D decoder and Fourier feature network for 60 epochs. For the baseline 2D decoder, we set the batch size to 64 to expedite training, while the batch size of the Fourier feature network was set to 16 due to the high dimensionality of the positional encodings. We trained both networks with a learning rate of 1×10^{-4} to ensure consistency. All of the models in this research were trained with a mean square error cost function. This research was performed on a single NVIDIA RTX A6000 GPU with 49 GB of memory. We utilized the Adam optimizer[21] to train both networks, and we also leveraged mixed precision training[22] to train both single and half-bit models to speed up convergence while maintaining the same performance. This research was done with PyTorch, and all inputs were scaled between 0 and 1 for normalization.

4 Results

We compare various methods to reconstruct 3D slices given rows of biplane images. We have three models in this study. First, we utilize the backprojection method proposed

Table 1: Comparison of 1D to 2D Algorithms

Model	PSNR (dB)	SSIM (%)	MSE	MAE
Backprojection [11]	19.31 ± 0.45	33.86 ± 1.27	0.0118 ± 0.0012	0.0885 ± 0.0087
Decoder 2D	25.92 ± 0.69	59.12 ± 1.33	0.0026 ± 0.0004	0.039 ± 0.005
Fourier Feature	26.32 ± 0.36	61.38 ± 1.79	0.0023 ± 0.0002	0.0364 ± 0.0029

in our previous paper[11]. Secondly, we constructed a baseline 2D decoder with only transposed convolutions, and finally we create a Fourier feature network, taking into account positional encoding to reconstruct the slice.

4.1 Evaluation Methods

To evaluate the predictions of all three models, we leverage the peak signal-to-noise ratio (PSNR)[23], the structure similarity index measure (SSIM), the mean square error (MSE), and the mean absolute error (MAE). The PSNR is denoted by the equation below:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right)$$

where MAX is the maximum pixel value in the image, and we divide the square of that value by the mean squared error. This gives us a result in decibels. The PSNR is inversely related to the MSE. The SSIM is another method for determining the similarity between two input images, as shown in the equation below.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

As denoted in the equation above, the SSIM score between two images is determined by their structural similarity, contrast, and luminance. In the equation above, x and y are the two input images, μ_x and μ_y represent the luminance of the two images, and σ_x^2 and σ_y^2 represent the variance of the images. An additional metrics we use is the MSE, which is also the cost function used in training denoted by the equation below.

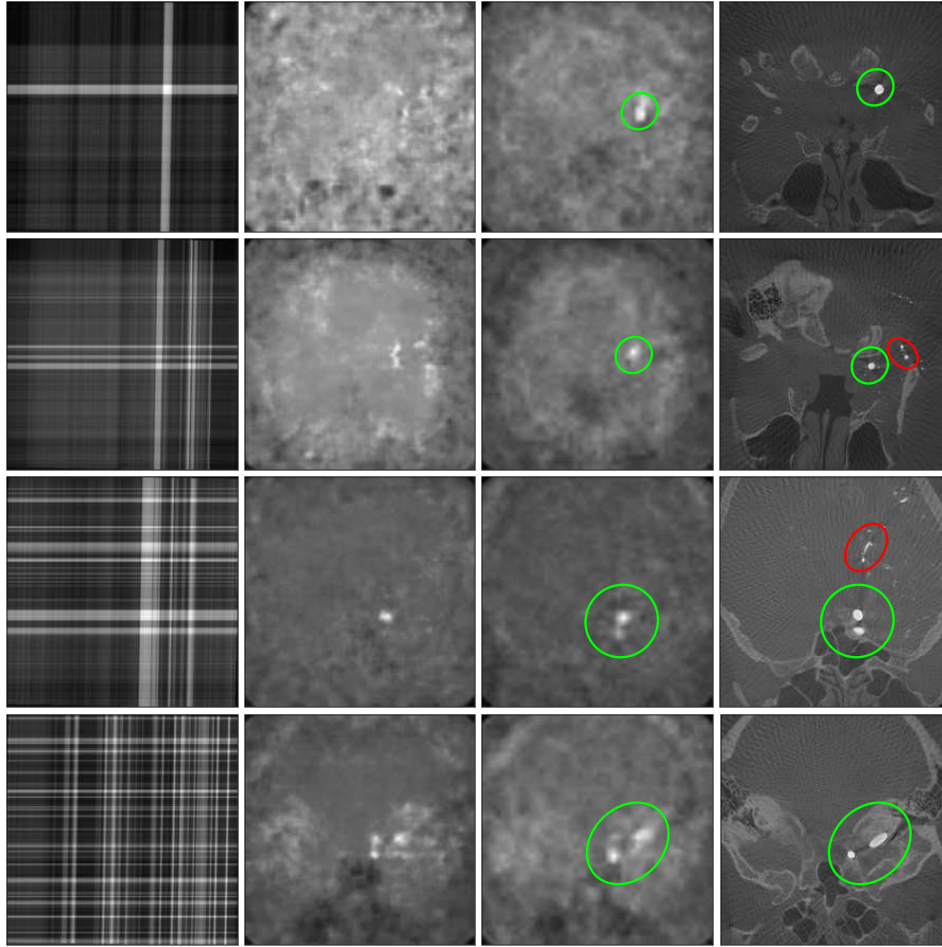
$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Finally, we report the MAE, which is measures the average magnitude of error.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

4.2 Analysis of Methods

Of of the three methods, the Fourier feature network surpassed both the baseline backprojection, and also the 2D decoder. In terms of the PSNR, the Fourier feature network scored the highest with a value of 26.32±0.36, a SSIM score of 61.38±1.79, a MSE of 0.0023±0.0002, and finally a MAE score of 0.0364±0.0029. Following that, the CNN decoder model scored a PSNR score of 25.92±0.69, a SSIM score of 59.12±1.33, a MSE of 0.0026±0.0004, and finally a MAE score of 0.039±0.005. The worst performing model was the backprojection algorithm proposed in our previous paper[11], where the PSNR score of 19.31±0.45, a SSIM score of 33.86±1.27, a MSE of 0.0118±0.0012, and finally a MAE score of 0.0885±0.0087. These results can be better visualized in Table 1, and the error bound was computed by taking the standard deviation of all of the testing patients. We computed all the metrics across the stacked predicted slices vs. the ground truth volumes.



(a) Backprojection (b) 2D Decoder (c) Fourier Feature (d) Ground Truth

Fig. 3: Comparison of Fourier feature 2D reconstruction with a vanilla 2D decoder and the backprojection method from our previous paper. The green regions represent vessels that were properly reconstructed. The red regions are fine vessels that were not reconstructed by the Fourier feature network.

5 Discussion

In this research, we showed that Fourier feature networks can improve the slicewise reconstruction of cerebral angiograms. By doing so, we take a step toward solving the biplane 3D reconstruction problem by first reconstructing 2D slices given a 1D input vector. This paper truly demonstrates the difficulty of the biplane 3D reconstruction problem, as we show that even predicting 2D slices from a given 1D input is not a trivial task. We show that with Fourier feature networks, we can reconstruct 2D slices on our testing set with SSIM scores as high as 0.61. Figure 3 shows the Fourier

feature network outperformed the others in localizing and producing clearer vessels. For instance, in the first row, the Fourier feature network identified the stem, whereas the baseline decoder failed to do so. In the second row, the Fourier feature network more accurately captured the shape of the vessel compared to both baseline models. Additionally, in the third row of Figure 3, it is evident that the Fourier feature network detected two large stems in the cross-section, a detail missed by the 2D decoder.

However, there is still considerable room for improvement in this challenging research problem. As observed in Figure 3, the predicted cross sections are not as sharp as the ground truth. Although the Fourier feature network can localize large vessels more accurately than baseline methods, it is evident that further research is required to generate sharper images. In some examples, we even observe that fine vessels are not included in all models. This performance can be attributed to the severe occurrences present in the biplane data. The simulated 2D data that we use in this paper may not represent real-world rotational angiographic data, as the maximum intensity projection inherently has more visual correspondence to the ground truth reconstructions. We are interested in pursuing several promising research directions in the future. The first involves the use of physics-informed neural networks (PINNs)[24] to accurately model blood vessels by incorporating physical properties. For instance, we could ensure that the modeled blood vessels conform to a cylindrical or vessel-like shape by adding an additional term to the loss function specifically for this purpose. We are also interested in leveraging diffusion models[25], a new generative model, to reconstruct more slices that better match the representation of real-world scenarios. In future work, we aim to expand the research concepts presented in this research to 3D scenes. Our ultimate goal is to accurately predict a 3D representation of blood vessels and estimate flow parameters using fluid simulations.

6 Conclusion

In this paper, we developed a Fourier feature neural network that leverages deep learning and coordinate-based position encoding to reconstruct slices of cerebral angiograms. We utilize the maximum intensity projection to simulate 2D data from the ground truth CT volumes; then, we take rows of biplane image data to reconstruct the corresponding 2D slices. We also show that with a deterministic Fourier feature mapping $\gamma(\mathbf{v}) = [\cos(2\pi B\mathbf{v}), \sin(2\pi B\mathbf{v})]^\top$ from the given (x,y) input coordinates, we can see improved results in reconstruction in terms of PSNR, SSIM, MSE, and MAE. This paper serves as an incremental step to solving the incredibly difficult biplane 2D to 3D reconstruction task by first attempting to properly reconstruct 2D data from 1D input. We hope that this work will provide a foundation for future research in solving the biplane 2D to 3D reconstruction problem, in which it can not only decrease radiation exposure to the patient but also allow neurosurgeons to visualize and simulate diseases and blood flow in real time, which can assist their procedures.

Acknowledgements

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

Author Contributions

Sean Wu: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Naoki Kaneko:** Conceptualization, Data curation, Methodology, Project administration, Resources. **David S. Liebeskind:** Conceptualization, Data curation, Methodology, Project administration, Resources. **Fabien Scalzo:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declarations

- **Conflict of interest/Competing interests:** The authors have no conflict of interest to declare.
- **Code availability:** The code used during the current study are available from the corresponding author on reasonable request.

References

- [1] Scalzo, F., Liebeskind, D.S., et al.: Perfusion angiography in acute ischemic stroke. *Computational and Mathematical Methods in Medicine* **2016** (2016)
- [2] Cho, E.-S., Chung, T.-S., Oh, D.K., Choi, H.S., Suh, S.H., Lee, H.-K., Lee, K.H.: Cerebral computed tomography angiography using a low tube voltage (80 kvp) and a moderate concentration of iodine contrast material: a quantitative and qualitative comparison with conventional computed tomography angiography. *Investigative radiology* **47**(2), 142–147 (2012)
- [3] Medicine, J.H.: Diagnostic Cerebral Angiography. https://www.hopkinsmedicine.org/interventional_neuroradiology/downloads/Cerebral-Angiography-Johns-Hopkins-web.pdf. Accessed: 2023-12-26
- [4] Sakas, G., Grimm, M., Savopoulos, A.: Optimized maximum intensity projection (mip). In: *Rendering Techniques' 95: Proceedings of the Eurographics Workshop in Dublin, Ireland, June 12–14, 1995* 6, pp. 51–63 (1995). Springer
- [5] Magnusson, M., Lenz, R., Danielsson, P.-E.: Evaluation of methods for shaded surface display of ct volumes. *Computerized medical imaging and graphics* **15**(4), 247–256 (1991)
- [6] Perandini, S., Faccioli, N., Zaccarella, A., Re, T., Mucelli, R.P.: The diagnostic contribution of ct volumetric rendering techniques in routine practice. *Indian Journal of Radiology and Imaging* **20**(02), 92–97 (2010)
- [7] Kanitsar, A., Fleischmann, D., Wegenkittl, R., Felkel, P., Groller, E.: CPR-curved Planar Reformation. *IEEE*, ??? (2002)

- [8] Jiménez Sosa, M.S., Cantú González, J.R., Morales Avalos, R., Garza Castro, O., Quiroga Garza, A., Pinales Razo, R., Elizondo Riojas, G., Elizondo Omaña, R.E., Guzmán López, S.: Anatomical variants of anterior cerebral arterial circle: a study by multidetector computerized 3d tomographic angiography. *International Journal of Morphology* **35**(3), 1121–1128 (2017)
- [9] De González, A.B., Mahesh, M., Kim, K.-P., Bhargavan, M., Lewis, R., Mettler, F., Land, C.: Projected cancer risks from computed tomographic scans performed in the united states in 2007. *Archives of internal medicine* **169**(22), 2071–2077 (2009)
- [10] Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems* **33**, 7537–7547 (2020)
- [11] Wu, S., Kaneko, N., Mendoza, S., Liebeskind, D.S., Scalzo, F.: 3d reconstruction from 2d cerebral angiograms as a volumetric denoising problem. In: *International Symposium on Visual Computing*, pp. 382–393 (2023). Springer
- [12] Goesele, M., Curless, B., Seitz, S.M.: Multi-view stereo revisited. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 2402–2409 (2006). IEEE
- [13] Özyeşil, O., Voroninski, V., Basri, R., Singer, A.: A survey of structure from motion*. *Acta Numerica* **26**, 305–364 (2017)
- [14] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021)
- [15] Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelNeRF: Neural radiance fields from one or few images. In: *CVPR* (2021)
- [16] Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago (2015)
- [17] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42**(4), 1–14 (2023)
- [18] Glassner, A.S.: *An Introduction to Ray Tracing*. Morgan Kaufmann, ??? (1989)

- [19] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
- [20] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25** (2012)
- [21] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
- [22] Micikevicius, P., Narang, S., Alben, J., Diamos, G., Elsen, E., Garcia, D., Ginsburg, B., Houston, M., Kuchaiev, O., Venkatesh, G., et al.: Mixed precision training. arXiv preprint arXiv:1710.03740 (2017)
- [23] Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th International Conference on Pattern Recognition, pp. 2366–2369 (2010). IEEE
- [24] Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F.: Scientific machine learning through physics-informed neural networks: Where we are and what’s next. *Journal of Scientific Computing* **92**(3), 88 (2022)
- [25] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. *Advances in neural information processing systems* **33**, 6840–6851 (2020)