Conditional Diffusion Model for Versatile Temporal Inpainting in 4D Cerebral CT Perfusion Imaging

Juyoung Bae[1], Elizabeth Tong[4], and Hao Chen[1,2,3](✉)

[1] Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Kowloon, Hong Kong
{jbaeaa,jhc}@cse.ust.hk
[2] Department of Chemical and Biological Engineering,
The Hong Kong University of Science and Technology, Kowloon, Hong Kong
[3] Division of Life Science,
The Hong Kong University of Science and Technology, Kowloon, Hong Kong
[4] Department of Pediatric Radiology,
Stanford University School of Medicine, Stanford, CA, USA
etong@stanford.edu

Abstract. Cerebral CT Perfusion (CTP) sequence imaging is a widely used modality for stroke assessment. While high temporal resolution of CT scans is crucial for accurate diagnosis, it correlates to increased radiation exposure. A promising solution is to generate synthetic CT scans to artificially enhance the temporal resolution of the sequence. We present a versatile CTP sequence inpainting model based on a conditional diffusion model, which can inpaint temporal gaps with synthetic scan to a fine 1-second interval, agnostic to both the duration of the gap and the sequence length. We achieve this by incorporating a carefully engineered conditioning scheme that exploits the intrinsic patterns of time-concentration dynamics. Our approach is much more flexible and clinically relevant compared to existing interpolation methods that either (1) lack such perfusion-specific guidances or (2) require all the known scans in the sequence, thereby imposing constraints on the length and acquisition interval. Such flexibility allows our model to be effectively applied to other tasks, such as repairing sequences with significant motion artifacts. Our model can generate accurate and realistic CT scans to inpaint gaps as wide as 8 seconds while achieving both perceptual quality and diagnostic information comparable to the ground-truth 1-second resolution sequence. Extensive experiments demonstrate the superiority of our model over prior arts in numerous metrics and clinical applicability. Our code is available at https://github.com/baejustin/CTP_Inpainting_Diffusion.

Keywords: Diffusion model · CT Perfusion · Radiation reduction · Sequence inpainting.

1 Introduction

Cerebral Computed Tomography Perfusion (CTP) is a widely used modality for fast and cost-effective stroke diagnosis [3]. CTP involves a sequence of 3D CT scans, where an acquisition interval exceeding 1 second could produce inaccurate diagnostic results [11]. However, increasing the temporal resolution with shorter intervals may result in prolonged radiation exposure. Conventional approaches to mitigate this risk are low-dose CT protocols that reduce the radiation dosage of each scan while maintaining sufficient temporal resolution [1, 16]. Yet, they inevitably degrade the quality of individual scans, requiring follow-up post-processing efforts often impractical due to the limited access to raw sinogram data on commercial scanners, high computational demands [15], or scarcity of paired datasets for learnable reconstruction. A more pragmatic solution for sequential imaging like CTP is to acquire the scans with adequate dosage in a reduced temporal resolution and interpolate the sequence with synthetic scans. This strategy also addresses the limitation of certain CT scanning protocols, such as Cone-beam CT, that inherently cannot achieve a sufficiently fast acquisition rate [14].

Several existing works have employed deep learning to address this challenge in CTP [22, 24]. They require an entire sequence with a fixed length and acquisition interval as input. This limits clinical applicability because (1) duration and intervals may vary due to patient or hardware considerations, and (2) corruptions in the known scans could compromise the interpolated scans. Meanwhile, recent advancements in diffusion models demonstrated exceptional capabilities in natural and medical sequence interpolation tasks [2, 6, 12, 20], surpassing Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) in many applications and benchmarks. However, they primarily aim to enhance the visual fluidity only between the adjacent conditioning frames, neglecting the global, sequence-level dynamics crucial for accurate estimation of time-specific concentration levels in CTP. This can be addressed by simply augmenting the condition with even more known scans in the sequence, but this remedy is (1) not readily scalable to longer sequences and (2) ineffective when other known scans are corrupted.

To address the aforementioned issues, we present a novel and flexible CTP sequence inpainting framework using a 3D diffusion model. Our model only needs two known scans before and after the target timepoint to generate a scan at that timepoint. We indicate the temporal position of the target only within the local window between the two reference scans, making our model agnostic to both the length and the interval of the sequence. To include a more holistic context beyond the local window, we avoid using any other known scans in the sequence as conditions. Instead, we indicate the state of concentration changes of the reference scans, which can be easily inferred from the pattern of time-concentration curves inherent in all perfusion imaging. Our simple and effective conditioning scheme allows our model to inpaint gaps of up to eight seconds with accurate and coherent synthetic scans at a one-second interval. To further demonstrate the clinical applicability of our model, we compute the perfusion

parametric maps from the synthetic sequences interpolated using our model. Results suggest that our model can be used to yield perfusion parameter maps with comparable quality to those derived from the ground-truth high temporal resolution sequence, while reducing the radiation exposure to as little as one-eighth. We also showcase the flexibility of our model by successfully applying it to an application of restoring a section of a CTP sequence corrupted by severe motion artifacts with clean, synthetic scans. Extensive experiments show the superior performance and clinical utility of our model in various temporal resolution settings and applications.

2 Methodology

2.1 Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPMs) [10, 19] are powerful generative models that learn to approximate the target data distribution through two main phases: the forward diffusion phase (FD) and the reverse diffusion phase (RD). In FD, the model formulates a Markov chain q with T iterations that progressively corrupt real data $\mathbf{x}_0$ with Gaussian noise. Each iteration, defined as $q(\mathbf{x}_t|\mathbf{x}_{t-1})$, applies noise to the data based on a variance schedule $\{\beta_t\}_{t=1}^{T}$:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}). \tag{1}$$

This variance schedule ensures that, after T steps, $\mathbf{x}_T$ is approximately a sample from a standard Gaussian distribution, $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$.

Meanwhile, RD aims to learn a generative process that can synthesize data by reversing the forward diffusion chain. Since the true posterior $p(\mathbf{x}_{t-1}|\mathbf{x}_t)$ is intractable, RD approximates it with a parameterized Gaussian distribution $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_t)$, where variance Σ_t is a function of t predetermined based on β_t, and mean $\mu_\theta(\mathbf{x}_t, t)$ is a function of t that can be learned through a neural network parameterized by θ. Typically, timestep t is embedded into a higher-dimensional vector via sinusoidal positional encoding and added to the layers of the network [10].

Instead of directly predicting the denoised data $\mu_\theta(\mathbf{x}_t, t)$, the network is trained to predict the noise $\epsilon_\theta(\mathbf{x}_t, t)$ added during the forward process [10]. This prediction is then used to iteratively refine the estimate of the original data point $\mathbf{x}_0$. The objective function for training the neural network θ is given by:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0, \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}), t\sim\mathcal{U}(1,T)}\left[\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|^2\right], \tag{2}$$

where $t \sim \mathcal{U}(1, T)$ indicates that t is uniformly sampled from all possible timesteps.

2.2 Condition Mining Attuned to Time-Concentration Dynamics

To generate the target scan $\mathbf{x}^n$, we condition the RD on two known scans, $\mathbf{x}^A$ and $\mathbf{x}^B$, from the same sequence that precedes and follows the timestamp of $\mathbf{x}^n$. We

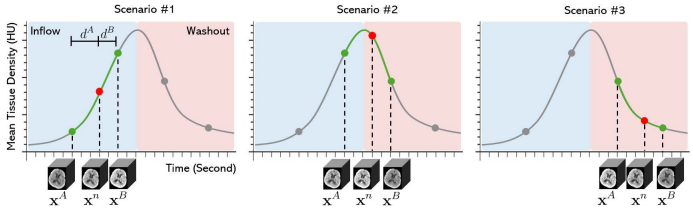

Fig. 1. We can always assign $(\mathbf{x}^A, \mathbf{x}^B)$ pair to one of the above three scenarios based on the mean tissue intensities. This only requires minimal knowledge of the other known scans in the sequence.

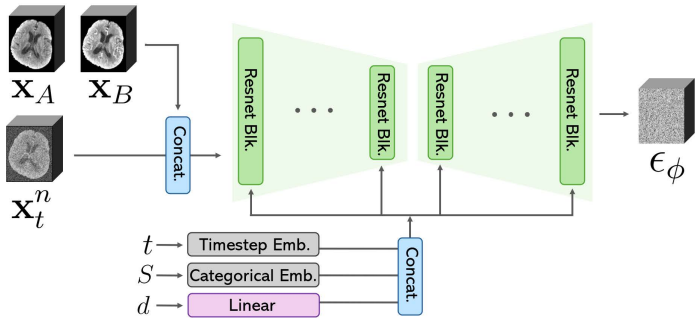

Fig. 2. Conditioning scheme of our model during RD at timestep t.

also employ the temporal distances from $\mathbf{x}^n$ to $\mathbf{x}^A$ and $\mathbf{x}^B$, denoted $d^A \in \mathbb{Z}^+$ and $d^B \in \mathbb{Z}^+$, to specify the temporal positioning of $\mathbf{x}^n$. Yet, CTP sequences exhibit specific patterns in the change of contrast levels across the sequence [13], where d^A and d^B alone cannot provide the global context beyond the local window. To address this, we leverage the innate nature of the time-concentration mapping in a perfusion sequence, characterized by an inflow (rising phase) and a washout (falling phase) of the contrast agent [7, 13]. We categorize the selected $\mathbf{x}^A$ and $\mathbf{x}^B$ pair into one of the three scenarios described in Figure 1 and incorporate it into the condition. This can provide insight into the expected change in motion within the specified local window, with respect to the global context. Combined with the local distance information d^A and d^B, the amount and the direction of the contrast change at a specific timepoint can be inferred.

Our versatile conditioning scheme permits any selection of $\mathbf{x}^A$ and $\mathbf{x}^B$, as long as the combined distance $d^A + d^B$ does not exceed 8 seconds - the maximum interval our model is designed to handle. At the same time, our scheme produces a large number of $\{\mathbf{x}^A, \mathbf{x}^B, d^A, d^B\}$ combinations for a given target $\mathbf{x}^n$ during training, fostering the model to thoroughly learn the global patterns of hemodynamics in a sequence. Overall, we minimize the following loss at each t to train our denoising network ϕ:

$$\mathcal{L}(\phi) = \mathbb{E}_{\mathbf{x}_0^n, \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), t \sim \mathcal{U}(1,T), 2 \leq d^A + d^B \leq 8} \left[\left\| \epsilon - \epsilon_\phi(\mathbf{x}_t^n, \mathbf{x}^A, \mathbf{x}^B, d, S, t) \right\|^2 \right], \quad (3)$$

where d denotes distance vector $\left[d^A, d^B\right]$ and S is a one-hot vector of size 3 that indicates the contrast change combinations of $\mathbf{x}^A$ and $\mathbf{x}^B$.

At each iteration of RD, $\mathbf{x}^A$ and $\mathbf{x}^B$ are concatenated with $\mathbf{x}_t^n$ to form an input to ϕ. For d and S, we encode them via linear and class embedding layers, respectively, and then append them to the timestep embedding. This is a more effective and economical conditioning strategy over the widely-used cross-attention method [9, 18], given the low dimensionality of d and S.

3 Experiments

3.1 Datasets and Implementation Detail

Our dataset comprises 4D CTP sequence data from 551 patients: 103 from ISLES2018 Challenge [4, 8], 115 from UniToBrain Dataset [17], and 333 from our in-house cohort. All sequences are resampled to a consistent 1-second interval, and individual CT scans are skull-stripped and windowed to 40 Hounsfield Unit (HU) level and 80 HU width. We standardize the scans to a spatial size of 192×192 by brain-centered cropping and an axial height of 8 via slab separation or axial zero-padding, as required. Sequences from 50 patients - 25 from the in-house cohort and 25 from the two public datasets - are randomly selected for hold-out testing.

During training, we randomly sample a condition set $\{\mathbf{x}^A, \mathbf{x}^B, d^A, d^B, S\}$ for every non-edge scan in the training sequences. In every iteration of the training dataset, the same non-edge scan $\mathbf{x}^n$ is paired with a different condition set. The denoising neural network of our model follows a 3D residual UNet [5] based on the original implementation of DDPM [10], with channel dimensions of 64, 64, 128, 128, 256, and 256 for each stage. Noise level β follows a linear schedule from 1e-6 to 1e-2, with T=1500. We train the model for 600,000 iterations using Adam optimizer. The batch size is set to 20 and the starting learning rate is set to 1e-4, which is decayed by a factor of 10 after every 200,000 iterations.

3.2 Evaluation Criteria

In our test set, we generate three low-resolution sequences with acquisition intervals of 2, 4, and 8 seconds by omitting the intermediate scans. Missing scans are reconstructed to a 1-second interval and thoroughly compared to the ground truth scans using the Structural Similarity Metric (SSIM), Peak Signal-to-Noise Ratio (PSNR), Fréchet Inception Distance (FID), and Learned Perceptual Image Patch Similarity (LPIPS), with the latter two metrics assessed on axial slices. To evaluate our model in a more clinically relevant context, we implement the standard deconvolution algorithm [7, 23] commonly used in commercial software packages for computing perfusion parameter maps (PPMs), which provide key indicators crucial for stroke assessment. The algorithm derives the hemodynamic parameters by deconvolving the time-concentration curves (TCC) of each voxel with the TCC of the main feeding artery [7]. We compute three types of PPMs

— Time-to-max (Tmax), Cerebral Blood Flow (CBF), and Cerebral Blood Volume (CBV) — from both the synthetic and the corresponding ground-truth sequences. Mean absolute error (MAE), root-mean-squared error (RMSE), and Lin's concordance correlation coefficient (CCC) are used to evaluate the similarity of the PPMs. Only the 25 in-house test sequences with manually annotated artery locations available are used for this follow-up experiment.

We compare our model with three recent state-of-the-art (SOTA) methods: (1) CNN-based CTP temporal upsampling model [24]; (2) MPVF [21], 4D medical sequence inpainting model using pyramid voxel flow; and (3) MCVD [20], diffusion-based 2D natural video interpolation model. Note that for MCVD, scans are generated slice-by-slice.

3.3 Result Analysis

Table 1 reports the reconstruction quality of the scans inpainted from low temporal resolution sequences. Our model consistently outperforms all non-diffusion-based methods across evaluated metrics, particularly in FID scores. While our advantage over MCVD is less pronounced in the simpler 2-seconds interval, it becomes more significant with longer intervals, which we attribute to our model's adept use of both local and global hemodynamic contexts to bridge the gap.

Table 1. Quantitative performance on CT scan synthesis. * and † indicate significant differences from our method with $p < 0.001$ and $p < 0.05$, respectively, by paired t-test.

Acquisition Interval	Methods	SSIM↑	PSNR↑	LPIPS↓	FID↓
2s	MCVD [20]	**99.321**	43.028*	0.0021	0.0004
	MPVF [21]	98.919*	41.138*	0.0116*	0.1446
	CNN [24]	99.105*	41.003*	0.0030*	0.0033
	Ours	99.309	**43.323**	**0.0021**	**0.0002**
4s	MCVD [20]	97.582	36.991*	0.0071*	0.0018
	MPVF [21]	96.537*	35.878*	0.0173*	0.1285
	CNN [24]	97.324*	37.038	0.0083*	0.0125
	Ours	**97.601**	**37.204**	**0.0067**	**0.0016**
8s	MCVD [20]	91.434†	31.843*	0.0188*	0.0012
	MPVF [21]	91.317†	31.782*	0.0292*	0.3133
	CNN [24]	90.261*	31.099*	0.0388*	0.2975
	Ours	**91.527**	**32.002**	**0.0184**	**0.0009**

Meanwhile, the utility of interpolated CTP sequences hinges on their ability to yield accurate PPMs that align with those derived from high temporal resolution counterparts. This requires not only high-quality synthesis of individual scans but also precise replication of hemodynamics throughout the sequence, where all synthetic scans must accurately reflect the expected concentration level at every timepoint. The challenge intensifies with larger temporal gaps due to reduced contextual data for estimating the concentration. As reported in Table 2, our model outperforms competing methods in simulating the reference PPMs,

Table 2. Quantitative evaluation of PPMs from interpolated synthetic sequences."Low Res." indicates baseline PPMs derived without interpolation. Non-parametric signed-rank test is applied to account for small sample size. * and † indicate significant differences from our method with $p < 0.001$ and $p < 0.05$, respectively.

Acquisition Interval	Methods	Tmax (second)			CBF (mL/100g/min)			CBV (mL/100g)		
		RMSE↓	MAE↓	CCC↑	RMSE↓	MAE↓	CCC↑	RMSE↓	MAE↓	CCC↑
2s	Low Res.	1.576*	0.382*	0.9365*	13.783*	6.476*	0.9963*	4.370*	2.114*	0.9934*
	MCVD [20]	1.101	0.127	0.9720	5.602	2.396	0.9995	1.498	0.647	0.9991
	MPVF [21]	1.259*	0.166*	0.9616*	10.622*	5.194*	0.9982*	2.286*	1.069*	0.9982*
	CNN [24]	1.158*	0.176*	0.9691*	**4.981**	2.073	0.9995	1.482†	0.645*	0.9992†
	Ours	**1.008**	**0.101**	**0.9762**	5.044	**2.072**	**0.9996**	**1.330**	**0.552**	**0.9993**
4s	Low Res.	2.689*	0.883*	0.8338*	38.567*	18.937*	0.9794*	13.484*	6.896*	0.9547*
	MCVD [20]	1.755†	0.311†	0.9299*	15.082†	6.948†	0.9968	4.115*	1.886*	0.9949†
	MPVF [21]	1.894*	0.375*	0.9145*	19.915*	9.685*	0.9945*	4.571*	2.150*	0.9938†
	CNN [24]	1.743†	0.367*	**0.9333**	15.046	7.125†	0.9967	3.853	1.748	0.9953
	Ours	**1.724**	**0.297**	0.9322	**14.436**	**6.531**	**0.9968**	**3.838**	**1.747**	**0.9954**
8s	Low Res.	4.464*	1.927*	0.6134*	70.998*	35.024*	0.9328*	26.901*	14.161*	0.8379*
	MCVD [20]	2.641*	0.710*	0.8369*	37.346*	17.539*	0.9817*	10.761*	5.083*	0.9669*
	MPVF [21]	2.692*	0.743*	0.8328*	42.220*	20.394*	0.9760*	12.178*	5.880*	0.9567*
	CNN [24]	3.279*	1.077*	0.7880*	56.573*	28.441*	0.9526*	14.070*	6.834*	0.9421*
	Ours	**2.547**	**0.663**	**0.8503**	**34.703**	**16.138**	**0.9844**	**9.921**	**4.639**	**0.9722**

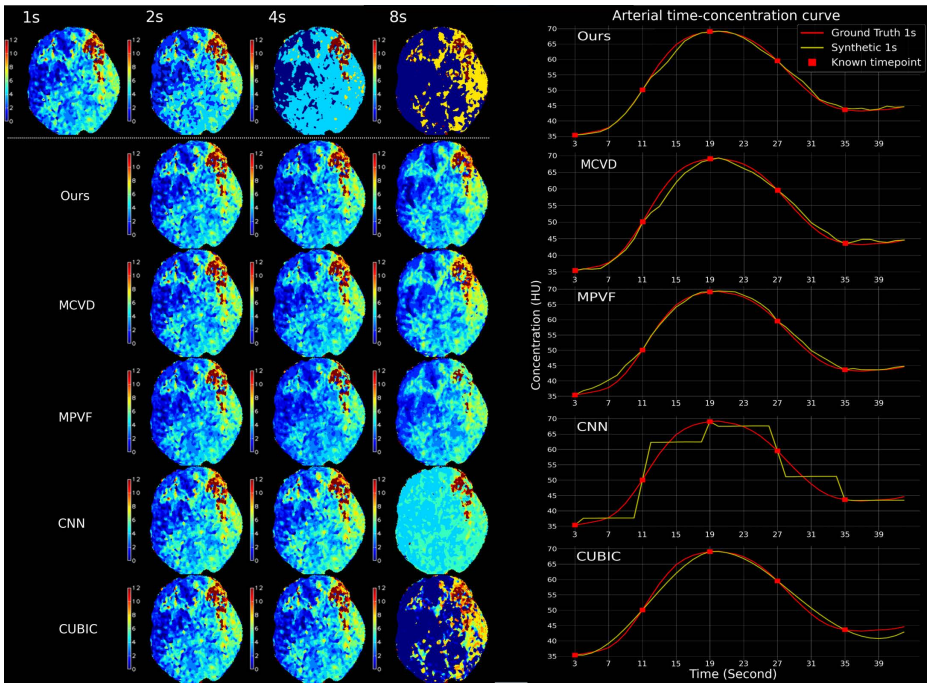

Fig. 3. Left: Tmax maps derived from different temporal resolution settings, with the top row presenting non-interpolated baselines. Right: Arterial TCCs of the interpolated 8-seconds interval sequences and the ground-truth 1-second interval sequence. See supplementary material for other types of PPMs. (Color figure online)

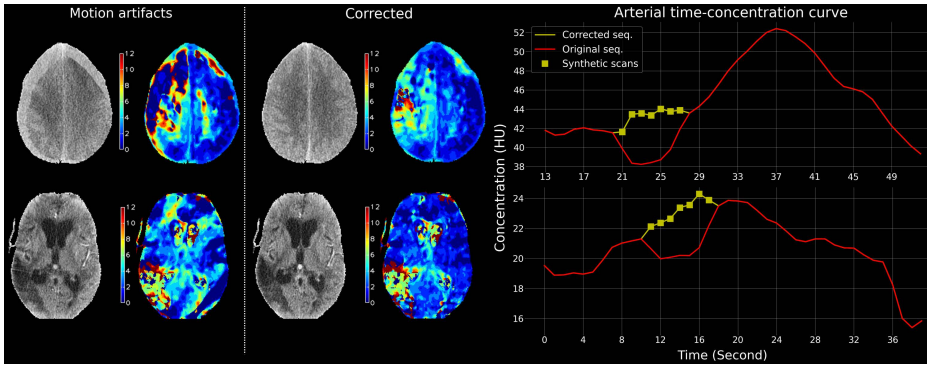

Fig. 4. Left: CT scans and corresponding Tmax maps, before and after the motion correction using our model. Right: Arterial TCCs used to compute the Tmax maps, before and after the correction. (Color figure online)

especially by a significant margin under the most demanding 8-seconds temporal resolution setting. This underscores our model's superior ability to harness contextual information, translating into improved clinical applicability.

Figure 3 presents qualitative visuals of the presented methods under different temporal resolution settings. Under the most challenging 8-seconds resolution setting, Tmax from our method's interpolated sequence demonstrates the highest visual alignment with the Tmax from the ground truth 1-second setting. Notably, the CNN method fails to accurately replicate the arterial TCC, a critical prerequisite for achieving reliable PPMs. Other comparing methods reproduce the arterial TCCs with reasonable accuracy but struggle to capture nuanced contrast changes throughout the tissue. This is particularly noticeable in the Tmax from the most naïve cubic interpolation method, which barely improves upon the non-interpolated baseline. Our model stands out by accurately reflecting both arterial TCCs and tissue contrast changes, ensuring precise PPM derivation.

Application to Motion Correction Our model's key strength is its versatility. It can inpaint temporal gaps of arbitrary duration, agnostic to the total length of the sequence, and does not require comprehensive knowledge of the entire sequence. This enables our model to adaptively perform motion correction by replacing one or more corrupted sections in the sequence with clean, synthetic counterparts. Unlike the traditional alternatives such as registration or rotation, our model is entirely agnostic to the magnitude of the artifact, yielding valuable clinical utility. In Figure 4, the first row example clearly shows artifacts in both the CT scan and the arterial TCC. After correction, the sequence can simulate a plausible arterial TCC and a PPM with clearer diagnostic information. Even when the artifacts in the CT scan are subtle (2nd row), our model can effectively remove the abnormalities and enhance the diagnostic information.

4 Conclusion

In this work, we present a versatile CTP sequence inpainting model based on conditional diffusion. Our model grants exceptional flexibility and reconstruction accuracy through an intuitive conditioning scheme based on time-concentration patterns in CTP. Thorough evaluation demonstrates the competence of our model in both generation quality and clinical applicability compared to existing SOTA methods.

Acknowledgments. This work was supported by the Hong Kong Innovation and Technology Fund (Project No. MHP/002/22) and HKUST (Project No. FS111).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arjah, H., Hjouj, M., Hjouj, F.: Low dose brain ct, comparative study with brain post processing algorithm. In: Proceedings of the 2019 2nd International Conference on Digital Medicine and Image Processing. pp. 1–7 (2019)
2. Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S.W., Fidler, S., Kreis, K.: Align your latents: high-resolution video synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22563–22575 (2023)
3. Campbell, B.C., Christensen, S., Levi, C.R., Desmond, P.M., Donnan, G.A., Davis, S.M., Parsons, M.W.: Cerebral blood flow is the optimal ct perfusion parameter for assessing infarct core. Stroke **42**(12), 3435–3440 (2011)
4. Cereda, C.W., Christensen, S., Campbell, B.C., Mishra, N.K., Mlynash, M., Levi, C., Straka, M., Wintermark, M., Bammer, R., Albers, G.W., Parsons, M.W., Lansberg, M.G.: A benchmarking tool to evaluate computer tomography perfusion infarct core predictions against a dwi standard. Journal of Cerebral Blood Flow & Metabolism **36**(10), 1780–1789 (2016)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 424–432. Springer (2016)
6. Danier, D., Zhang, F., Bull, D.: Ldmvfi: Video frame interpolation with latent diffusion models. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 1472–1480 (2024)
7. Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., Fahrig, R.: Deconvolution-based ct and mr brain perfusion measurement: theoretical model revisited and practical implementation details. Journal of Biomedical Imaging **2011**, 1–20 (2011)
8. Hakim, A., Christensen, S., Winzeck, S., Lansberg, M., Parsons, M., Lucas, C., Robben, D., Wiest, R., Reyes, M., Zaharchuk, G.: Predicting infarct core from computed tomography perfusion in acute ischemia with machine learning: lessons from the isles challenge. Stroke **52** (2021)

9. Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D.P., Poole, B., Norouzi, M., Fleet, D.J., et al.: Imagen video: high definition video generation with diffusion models. arXiv preprint arXiv:2210.02303 (2022)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
11. Kämena, A., Streitparth, F., Grieser, C., Lehmkuhl, L., Jamil, B., Wojtal, K., Ricke, J., Pech, M.: Dynamic perfusion ct: optimizing the temporal resolution for the calculation of perfusion ct parameters in stroke patients. European Journal of Radiology **64**(1), 111–118 (2007)
12. Kim, B., Ye, J.C.: Diffusion deformable model for 4d temporal medical image generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 539–548. Springer (2022)
13. Konstas, A., Goldmakher, G., Lee, T.Y., Lev, M.: Theoretic basis and technical implementations of ct perfusion in acute ischemic stroke, part 1: theoretic basis. American Journal of Neuroradiology **30**(4), 662–668 (2009)
14. Liu, J.S., Zhang, Y.K., Tang, H., Zhang, L.B., Yang, B.Q., Yan, Y., Luo, L.M., Chen, Y.: Rollback reconstruction for tdc enhanced perfusion imaging. Nuclear Science and Techniques **32**(8), 80 (2021)
15. Moghari, M.D., Sanaat, A., Young, N., Moore, K., Zaidi, H., Evans, A., Fulton, R.R., Kyme, A.Z.: Reduction of scan duration and radiation dose in cerebral ct perfusion imaging of acute stroke using a recurrent neural network. Physics in Medicine & Biology **68**, 165005 (2023)
16. Mullins, M.E., Lev, M.H., Bove, P., O'Reilly, C.E., Saini, S., Rhea, J.T., Thrall, J.H., Hunter, G.J., Hamberg, L.M., Gonzalez, R.G.: Comparison of image quality between conventional and low-dose nonenhanced head ct. American Journal of Neuroradiology **25**(4), 533–538 (2004)
17. Perlo, D., Tartaglione, E., Gava, U., D'Agata, F., Benninck, E., Bergui, M.: Unitobrain dataset: a brain perfusion dataset. In: International Conference on Image Analysis and Processing. pp. 498–509. Springer (2022)
18. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
19. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. pp. 2256–2265. PMLR (2015)
20. Voleti, V., Jolicoeur-Martineau, A., Pal, C.: Mcvd-masked conditional video diffusion for prediction, generation, and interpolation. Advances in Neural Information Processing Systems **35**, 23371–23385 (2022)
21. Wei, T.T., Kuo, C., Tseng, Y.C., Chen, J.J.: Mpvf: 4d medical image inpainting by multi-pyramid voxel flows. IEEE Journal of Biomedical and Health Informatics (2023)
22. Xiao, Y., Liu, P., Liang, Y., Stolte, S., Sanelli, P., Gupta, A., Ivanidze, J., Fang, R.: Stir-net: deep spatial-temporal image restoration net for radiation reduction in ct perfusion. Frontiers in Neurology **10**, 647 (2019)
23. Zanderigo, F., Bertoldo, A., Pillonetto, G., Cobelli, C., et al.: Nonlinear stochastic regularization to characterize tissue residue function in bolus-tracking mri: assessment and comparison with svd, block-circulant svd, and tikhonov. IEEE Transactions on Biomedical Engineering **56**(5), 1287–1297 (2009)
24. Zhu, H., Tong, D., Zhang, L., Wang, S., Wu, W., Tang, H., Chen, Y., Luo, L., Zhu, J., Li, B.: Temporally downsampled cerebral ct perfusion image restoration using

deep residual learning. International Journal of Computer Assisted Radiology and Surgery **15**, 193–201 (2020)