Synthesizing Delayed-Phase Contrast-Enhanced Breast MR Images from Early-Phase Images Using an Iterative Deep Network

Woojin Chung[1], Junghwa Kang[1], Ga Eun Park[2], Sung Hun Kim[2], and Yoonho Nam[1]

[1] Department of Biomedical Engineering, Hankuk University of Foreign Studies, Yongin, Korea
{goglxych97,kangjung9592,yoonhonam}@hufs.ac.kr
[2] Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea
hoonhoony@naver.com
rad-ksh@catholic.ac.kr

Abstract. Acquisition of dynamic contrast-enhanced MR imaging with gadolinium-based contrast agents at multiple time points provides valuable diagnostic information. In breast MRI, dynamics of enhancement serve as key indicators for differentiating malignant from benign tumors. However, acquiring delayed-phase images requires extended scan times and could lead to patient discomfort and increased costs. Furthermore, some protocols acquire only early-phase images, limiting the ability to capture dynamics of enhancement over time. In this study, we propose an iterative deep neural network that sequentially generates post-contrast images using prior outputs. By synthesizing delayed-phase images at multiple time points from early acquisitions, the proposed network enables the temporal prediction of enhancement. We evaluate our approach using a breast MRI dataset consisting of images acquired at six time points, including the pre-contrast phase. The results indicate that the proposed method can approximate delayed-phase images from early-phase images, suggesting its potential to support abbreviated scan protocols in dynamic contrast-enhanced MRI. Our code is available at: https://github.com/goglxych97/iterU-Net.git

Keywords: Breast MRI · Contrast-Enhanced · Image Synthesis · Iterative Deep Network

1 Introduction

Dynamic contrast-enhanced (DCE) MRI with gadolinium-based contrast agents acquired at multiple time points provides insight into the temporal dynamics of tissue enhancement, reflecting physiological information [8]. However, prolonged scan times for delayed-phase imaging result in patient discomfort and increased examination costs. Furthermore, some protocols such as the abbreviated protocol

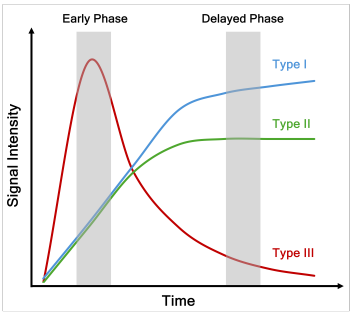

Fig. 1. Representative kinetic curve types in DCE-MRI, illustrating signal intensity changes over time. The three primary enhancement patterns are shown: Type I (persistent), Type II (plateau), and Type III (washout).

include only early-phase imaging. These limitations could restrict the evaluation of enhancement patterns.

In breast imaging, DCE-MRI is particularly valuable for distinguishing malignant from benign lesions. The temporal variation in enhancement contains critical information on tumor vascularity and tissue characteristics [5, 12, 19]. These enhancement-based signal intensity patterns allow tumors to be categorized into three primary kinetic curve types: persistent, plateau, and washout. The three kinetic curve types are illustrated in Figure 1. Each curve type is characterized as follows: the persistent curve gradually increases, the plateau curve rises and stabilizes, and the washout curve rises and then declines. These dynamics of enhancement are helpful for lesion characterization and diagnostic decisions. Malignant lesions are more frequently associated with washout or plateau curves, whereas benign lesions often exhibit a persistent pattern[12, 19].

Recently, deep learning models have been applied to synthesize breast MR imaging [2], due to their capability of generating high-quality images and capturing complex patterns from data. Previous studies [6, 13] have demonstrated the ability of deep learning models in contrast-enhanced breast MRI, particularly in synthesizing delayed-phase images. However, conventional approaches are limited to synthesizing images at a single time point.

In this study, we propose iterU-Net, an iterative deep neural network that generates post-contrast images at delayed time points from a pre-contrast and early-phase image. At each iterative step, the network progressively predicts the contrast-enhanced image for the next time point using prior outputs. This approach enables synthesizing time-series contrast-enhanced images and predicting delayed enhancement patterns without requiring extended scans.

2 Related Work

2.1 Iterative Deep Networks in Medical Imaging

Several studies have explored the use of iterative deep networks in medical imaging. In MRI reconstruction, these models improve image quality through progressive steps [10, 16, 22]. In medical image segmentation and object detection, they refine predictions by leveraging prior outputs and iteratively correcting errors [14, 15]. While these studies have focused on iterative refinement, a recent study introduced an iterative approach that sequentially synthesizes low-dose contrast-enhanced images from high-dose contrast-enhanced images [25]. This study highlights that an iterative and progressive network could incorporate information at different levels, thereby enabling the flexible generation of outputs at various stages.

2.2 Synthesizing Contrast-Enhanced Imaging

Studies on synthesizing delayed-phase image from early-phase acquisition in DCE-MRI have explored approaches [1, 6, 9] such as GANs [7] and physics-informed models. Some models incorporate multiple time points to improve the generation of delayed-phase images [13]. Recently, studies have applied deep learning models to predict delayed-phase images and kinetic curves in breast MRI, demonstrating the feasibility of kinetic curve estimation [6]. However, existing approaches have been limited to predicting a single time point, restricting their ability to capture the temporal progression of enhancement. Our proposed method effectively models the sequential dynamics of enhancement using an iterative deep-learning framework.

3 Method

3.1 Image Normalization

Normalizing input images is recommended for stable model training, and using dataset-appropriate normalization helps the network learn meaningful representations more effectively. Because DCE-MRI has different imaging distributions over time, it requires consistent normalization across all time frames. In this study, we adopted the Time-Intensity pattern-based normalization (TI-norm), which has been shown to be effective for dynamic images [6]. The normalized image y_k at each time point image x_k is defined as follows:

$$y_k = (x_k - \mu_p)/\sigma_p \tag{1}$$

Here, μ_p and σ_p denote the mean and standard deviation at a specific time point. In this study, we used the first acquired image after contrast administration as the reference time point.

3.2 ROIs and Lesion Segmentation

For model training, regions of interest (ROIs) were defined using a previously developed segmentation model [18]. This model, trained on pre-contrast images, was used to predict fibroglandular tissue and background parenchymal enhancement, both of which were included in the breast ROIs.

Since the cancerous region had a relatively limited spatial extent within the breast ROIs, there was a potential risk of bias during model training. To address this issue, an additional segmentation method trained on both pre-contrast and post-contrast enhanced images was performed [20]. The resulting segmented regions were used for foreground-aware random cropping, as described in a later section.

3.3 Proposed Method

Unlike previous models that synthesize contrast-enhanced images at a single time point, our approach iteratively predicts sequential time-series images, effectively capturing the dynamics of enhancement. The iterU-Net, a 3D U-Net [4] based network with an iterative architecture, is designed to predict contrast-enhanced images in DCE-MRI. The proposed network has two main contributions. First, the network iteratively generates sequential outputs and, at each iteration step, uses time embeddings and a ConvLSTM [23] block to effectively integrate temporal dependencies. Second, a warm-up stage was employed to configure the initial model parameters, facilitating model convergence. The overall training process is shown in Figure 2.

Architecture of iterU-Net Figure 2(b) illustrates the architecture of the proposed iterative network: iterU-Net. Based on a U-Net [21] design, this network takes a pre-contrast image and a contrast-enhanced image at time step k as inputs, and generates the contrast-enhanced image for time step $k + 1$.

To incorporate temporal information into the model, sinusoidal encoding [24] was applied, followed by projection layers that embed time information into all blocks of the U-Net based architecture. The projection layers consist of two linear layers with a GELU [11] activation in between. Additionally, a ConvLSTM block was employed to integrate information from previous iterations. As the network iterates, it updates its hidden state and cell state at each step, which are then used in subsequent iterations to maintain dynamic information. The ConvLSTM block was applied prior to the downsampling block.

Since structural details are preserved between time points in contrast-enhanced images, a residual connection was added at the output-level. This enables an iterative process where the output from the previous time point is refined to generate the next, ensuring that the network focuses on temporal changes.

Foreground-Aware Random Cropping To address the imbalanced distribution of lesion areas during 3D patch-based training, we applied a foreground-aware random cropping strategy that selectively extracts regions containing both

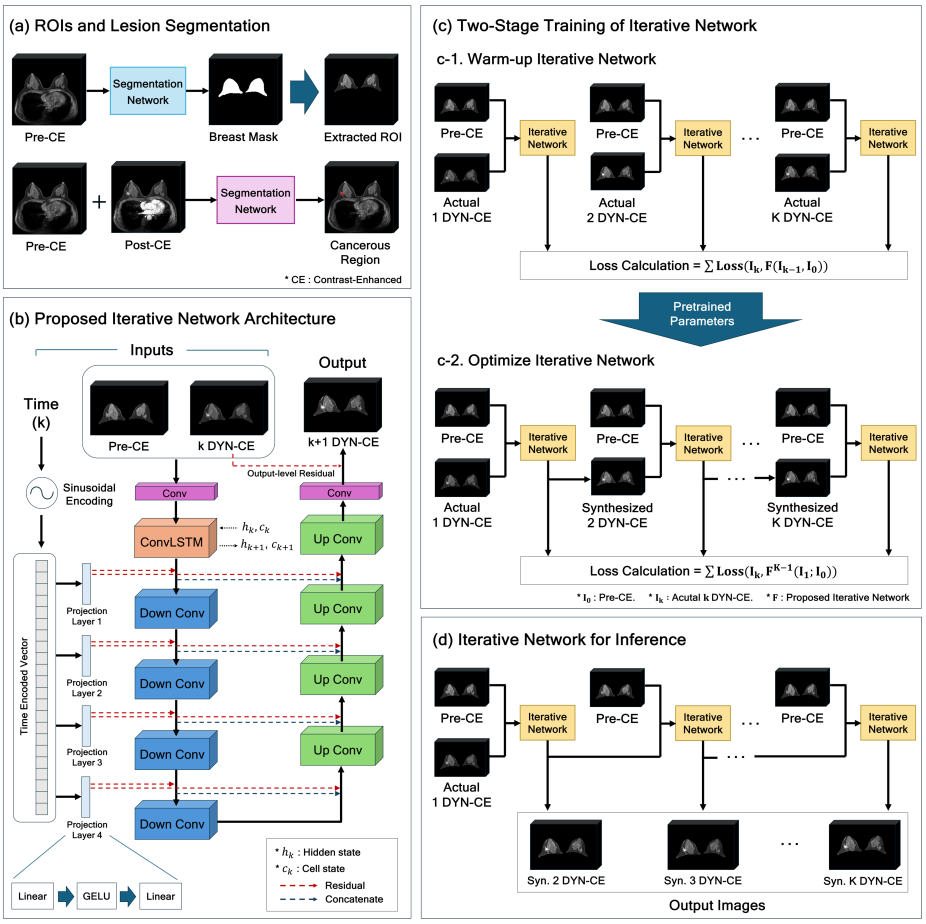

Fig. 2. Overview of the training process and model architecture. (a) The process of extracting ROIs and cancerous regions using pre-contrast and post-contrast images was presented. (b) Architecture of the proposed iterative network: The network generates the contrast-enhanced image for the next time point based on the prior contrast-enhanced image and pre-contrast image. (c) Two-stage training of the proposed network: c-1. Warm-up stage, where actual images are used to achieve initial convergence, and c-2. Optimization stage, where the model's output is recursively used as input. (d) Application of the iterative network for inference: The model takes a pre-contrast image and one dynamic contrast-enhanced image as inputs, and iteratively predicts sequential outputs.

the foreground (lesion) and the background. One patch was randomly sampled from the segmented cancerous region, and another from the entire image to form a paired input. This approach helped alleviate bias caused by data imbalance and ensured that the model captured critical lesion details.

Warm-up Training for Pretraining To prevent unstable outputs from being recursively fed into the network during training, we divided the training process into two stages, as illustrated in Figure 2(c): warm-up stage and optimization stage. During the warm-up stage, actual images were used as inputs to achieve initial convergence in the iterative network. Then, in the optimization stage, leveraging the parameters pretrained during the warm-up stage, the prior output was used as input for each iteration.

Loss Calculation In training, we used a combination of loss functions to ensure both structural fidelity and contrast preservation. L1 loss was used for pixel-wise accuracy, SSIM loss preserved structural similarity, perceptual loss (Med3D, Chen et al. [3]) captured high-level contextual features, and gradient difference loss [17] emphasized fine contrast variations. Loss weights were empirically set to 0.1, 1, 1, and 0.1, respectively.

4 Result

4.1 Dataset

The Breast DCE-MRI dataset was acquired at our institution and consists of images taken at six time points: before and at 85, 155, 225, 295, and 365 seconds after the injection of the contrast agent gadolinium DTPA (0.1 mmol/kg). Scans were performed with patients in the prone position using a 3-T Ingenia scanner (Philips, Best, The Netherlands). T1-weighted axial images were obtained with THRIVE sequences (TR/TE = 4.0/1.8 ms, flip angle 12°, slice thickness 1 mm). From an initial cohort of 200 subjects, two cases were excluded due to inconsistent image dimensions, resulting in a final dataset of 198 individuals with 160 (42 type I, 85 type II, 33 type III) for model training and 38 (16 type I, 14 type II, 8 type III), for testing.

4.2 Evaluation of Proposed Method

The proposed network was compared with 3D U-Net baseline models (one for each time point) that were trained to predict a single time point contrast-enhanced image. All models were trained under the same conditions. We evaluated overall image quality using PSNR and SSIM, and quantified the enhancement error in the segmented cancerous region by measuring squared error in mean intensity between the actual and synthesized images. The results for each time point are presented in Table 1.

Table 1. Results between baseline model and proposed iterative network

	Time Point	PSNR(↑)	SSIM(↑)	Enhancement Error(↓)
Basline Model (Single Time point)	2DYN (155 s)	41.0328	0.9800	0.1176
	3DYN (225 s)	40.5878	0.9795	0.1200
	4DYN (295 s)	39.5314	0.9795	0.2397
	5DYN (365 s)	39.1023	0.9792	0.2331
Proposed Iterative Network	2DYN (155 s)	41.7476	0.9808	**0.0346**
	3DYN (225 s)	39.9487	0.9747	**0.0738**
	4DYN (295 s)	39.0312	0.9718	**0.0884**
	5DYN (365 s)	38.1176	0.9688	**0.0921**

While the overall image similarity metrics (as measured by PSNR and SSIM) remains comparable between the proposed network and the baseline models, the proposed network exhibits a substantial reduction in enhancement error. For instance, at the 5DYN time point, the enhancement error in the baseline model rises to 0.2331, whereas the proposed network maintains a lower error of 0.0921. This demonstrates the effectiveness of our iterative framework in preserving the temporal dynamics of lesion enhancement.

Figure 3 presents representative examples of the actual images in the test set and the synthesized images generated by the proposed network for three types, along with the enhancement kinetic curves of signal intensity in the cancerous region. Each kinetic curve includes the actual image, multiple outputs from the proposed iterative network, and a single output from a baseline model trained to predict the 5DYN time point. As shown in the kinetic curves, the baseline model struggles to accurately capture fine-grained dynamic enhancement. In contrast, the iterU-Net more accurately approximates the changes in curve slope, leading to improved predictions. The model refines predictions step-by-step, ensuring consistency across multiple time points.

4.3 Ablation Study

The ablation study evaluated the effect of warm-up training and temporal information integration, which includes time embedding and ConvLSTM. For each experimental setting, kinetic curves generated for the segmented cancerous region were analyzed using Spearman's rank correlation coefficient (SCC) and dynamic time warping (DTW), while enhancement error at the 5DYN time point was also measured. The results can be found in Table 2

Table 2. Ablation study on warm-up training and temporal information integration.

	SCC(↑)	DTW(↓)	Enhancement Error on 5DYN(↓)
w/o Warm-Up Training	0.8390	0.8031	0.1233
w/o Temporal Information	0.8390	0.8046	0.1146
Proposed Method	0.9203	0.6300	0.0921

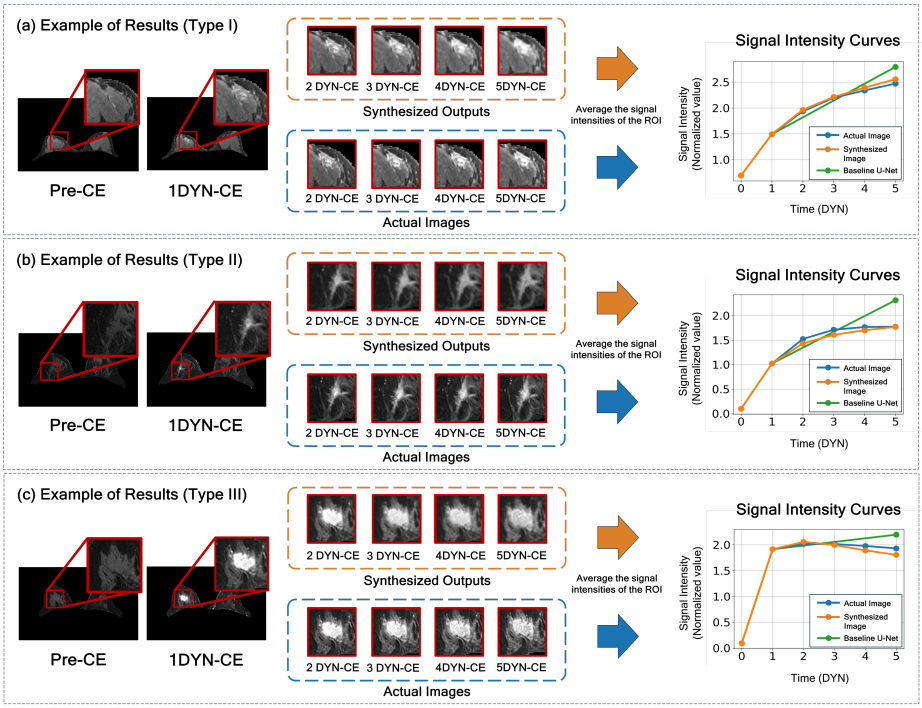

Fig. 3. Examples of actual images and synthesized outputs from the proposed network for Type I (persistent), Type II (plateau), and Type III (washout) enhancement patterns. On the right, the signal intensity curves represent the mean intensity value of the segmented cancerous region at each time point.

As shown in Table 2, the proposed method achieved the highest SCC (0.9203) and the lowest DTW (0.6300) and enhancement error (0.0921), demonstrating the effectiveness of integrating temporal information and warm-up training. In contrast, removing these components significantly reduced the network's ability to dynamics of enhancement.

5 Discussion and Conclusion

In this study, we proposed iterU-Net, an iterative network for synthesizing delayed contrast-enhanced images in breast MRI. Our findings demonstrate that the synthesized kinetic curves closely approximate those from actual delayed-phase images, outperforming those from single time-point prediction methods. Unlike conventional trained approaches that require a specific time point as input, our network flexibly handles inputs from various time points, making it more adaptable to diverse scenarios. By predicting dynamic contrast enhancement using only early-phase post-contrast images, our approach enables characterization of enhancement patterns while potentially reducing scan times in DCE-MRI.

However, our study has some limitations. The iterative network requires a warm-up stage for optimal performance, adding complexity to the training process. Additionally, its iterative prediction approach results in higher computational costs compared to single time point prediction models. Our study was also conducted on a limited dataset, highlighting the need for further validation with larger and more diverse datasets. While our evaluation primarily focused on the correlation between signal intensity kinetic curves, future studies should employ larger, multi-vendor, multi-center cohorts using diverse datasets and include clinical assessments.

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

References

1. Bharti, D., Ramanarayanan, S., Ram, K., Agarwal, H., Venkatesan, R., Sivaprakasam, M., et al.: Aad-dce: An aggregated multimodal attention mechanism for early and late dynamic contrast enhanced prostate mri synthesis. arXiv preprint arXiv:2502.02555 (2025)
2. Carriero, A., Groenhoff, L., Vologina, E., Basile, P., Albera, M.: Deep learning in breast cancer imaging: State of the art and recent advancements in early 2024. Diagnostics **14**(8), 848 (2024)
3. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. pp. 424–432. Springer (2016)
5. El Khouli, R.H., Macura, K.J., Jacobs, M.A., Khalil, T.H., Kamel, I.R., Dwyer, A., Bluemke, D.A.: Dynamic contrast-enhanced mri of the breast: quantitative method for kinetic curve type assessment. American Journal of Roentgenology **193**(4), W295–W300 (2009)
6. Fonnegra, R.D., Hernández, M.L., Caicedo, J.C., Díaz, G.M.: Synthesizing late-stage contrast enhancement in breast mri: A comprehensive pipeline leveraging temporal contrast enhancement dynamics. arXiv preprint arXiv:2409.01596 (2024)
7. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
8. Gordon, Y., Partovi, S., Müller-Eschner, M., Amarteifio, E., Bäuerle, T., Weber, M.A., Kauczor, H.U., Rengier, F.: Dynamic contrast-enhanced magnetic resonance imaging: fundamentals and application to the evaluation of the peripheral perfusion. Cardiovascular diagnosis and therapy **4**(2), 147 (2014)
9. Gravina, M., Maddaluno, M., Marrone, S., Sansone, M., Fusco, R., Granata, V., Petrillo, A., Sansone, C.: A physiological-informed generative model for improving breast lesion classification in small dce-mri datasets. IEEE Journal of Biomedical and Health Informatics (2024)

10. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated mri data. Magnetic resonance in medicine **79**(6), 3055–3071 (2018)
11. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
12. Kuhl, C.K., Mielcareck, P., Klaschik, S., Leutner, C., Wardelmann, E., Gieseke, J., Schild, H.H.: Dynamic breast mr imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? Radiology **211**(1), 101–110 (1999)
13. Li, W., Liu, J., Wang, S., Feng, C.: Mtfn: multi-temporal feature fusing network with co-attention for dce-mri synthesis. BMC Medical Imaging **24**(1), 47 (2024)
14. Liang, Z., Guo, X., Noble, J.A., Kamnitsas, K.: Itermask 2: Iterative unsupervised anomaly segmentation via spatial and frequency masking for brain lesions in mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 339–348. Springer (2024)
15. Liao, X., Li, W., Xu, Q., Wang, X., Jin, B., Zhang, X., Wang, Y., Zhang, Y.: Iteratively-refined interactive 3d medical image segmentation with multi-agent reinforcement learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9394–9402 (2020)
16. Lonning, K., Putzky, P., Caan, M., Welling, M.: Recurrent inference machines for accelerated mri reconstruction. (2018)
17. Mathieu, M., Couprie, C., LeCun, Y.: Deep multi-scale video prediction beyond mean square error. arXiv preprint arXiv:1511.05440 (2015)
18. Nam, Y., Park, G.E., Kang, J., Kim, S.H.: Fully automatic assessment of background parenchymal enhancement on breast mri using machine-learning models. Journal of Magnetic Resonance Imaging **53**(3), 818–826 (2021)
19. O'Flynn, E.A., DeSouza, N.M.: Functional magnetic resonance: biomarkers of response in breast cancer. Breast Cancer Research **13**, 1–10 (2011)
20. Park, G.E., Kim, S.H., Nam, Y., Kang, J., Park, M., Kang, B.J.: 3d breast cancer segmentation in dce-mri using deep learning with weak annotation. Journal of Magnetic Resonance Imaging **59**(6), 2252–2262 (2024)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
22. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D.: A deep cascade of convolutional neural networks for dynamic mr image reconstruction. IEEE transactions on Medical Imaging **37**(2), 491–503 (2017)
23. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems **28** (2015)
24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
25. Wang, D., Pasumarthi, S., Zaharchuk, G., Chamberlain, R.: Simulation of arbitrary level contrast dose in mri using an iterative global transformer model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 88–98. Springer (2023)