Causality-driven Spatio-temporal Generator for Multi-phase Contrast-enhanced CT Synthesis

Qikui Zhu[1], Hao Wu[1], Yanyan Zhang[2], and Shuo Li[2]

[1] Department of Biomedical Engineering, Case Western Reserve University, OH, USA
[2] Department of Computer Science, Case Western Reserve University, OH, USA
shuo.li11@case.edu

Abstract. Synthesizing multi-phase contrast-enhanced CT (CE-CT) images is clinically significant, as it can mitigate clinical risks such as radiation exposure and allergic reactions to contrast agents. However, existing methods treat multi-phase synthesis as separate tasks, failing to maintain the inter-phase dependencies and consistency between synthesized multi-phase CE-CT images. Moreover, the limited variability in CT intensity distributions makes it challenging to capture subtle variations in multi-phase imaging. For the first time, we propose a novel Causality-driven Spatio-temporal Generator (CSGen) for synthesizing multi-phase CE-CT imaging through three key novelties: 1) Using a novel phase-causality to creatively exploit the multi-phase variation content for driving the multi-phase CE-CT synthesizing, addressing the challenge of capturing multi-phase discriminative features through one model. 2) Introducing a new Spatio-temporal Transformer to establish the spatio-temporal correlation between multi-phase CE-CT images for leveraging multi-phase inter- and intra-dependencies and improving synthesis quality. 3) Multi-phase adversarial learning is designed for enhancing multi-phase discriminative feature learning. Experimental results (mean PSNR: 31.15, mean SSIM: 0.9066, mean NMAE: 3.17) demonstrate that CSGen outperforms state-of-the-art synthesis methods, and, for the first time, successfully synthesizes multi-phase CE-CT images.

Keywords: Multi-phase CE-CT synthesis · Causality-driven · Spatio-Temporal Transformer · Multi-phase adversarial learning.

1 Introduction

Multi-phase contrast-enhanced CT (CE-CT) imaging is essential for diagnosing liver tumors [1,2]. However, in clinical practice, multi-phase CE-CT imaging is performed by injecting chemical contrast agents (CAs), which introduces several inefficiencies and risks, such as prolonged examination time, radiation exposure [7,16], and the potential for allergic reactions [9]. If multi-phase CE-CT imaging can be synthesized using non-contrast CT (NCCT) imaging without the administration of chemical CAs, it has the potential to reduce costs and eliminate associated dangers. This would leverage the full capabilities of CT imaging,

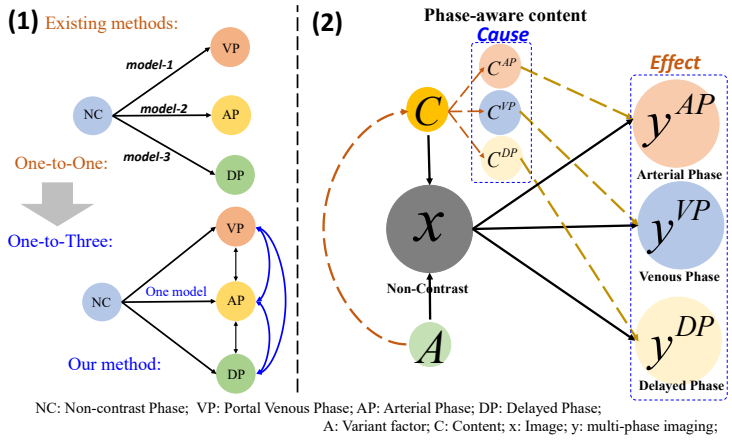

Fig. 1. 1) Existing methods treat multi-phase synthesis as separate tasks, failing to ensure consistency across the synthesized multi-phase images. 2) We use a phase-causality to exploit the multi-phase variation content and formulate multi-phase CE-CT image synthesis as a sequence synthesis task.

leading to a significant impact on clinical applications and providing greater value for patients [17].

Although existing state-of-the-art methods [10,12,15,18,19] achieved remarkable performance in image synthesis, these methods pay all attention to single modality and can not exploit the correlation between multi-phase images to maintain the inter-phase dependencies and consistency (Fig. 1.1). Furthermore, these methods struggle to address the unique challenges associated with synthesizing multi-phase images. Specifically, 1) simultaneously learning multi-phase variation information from one input for these methods is challenging. The contents of CT imaging enhanced by CAs in different phases are different, coronary arteries, aorta in portal venous phase; liver, pancreas, kidney in the arterial phase; and kidney tumor in the delayed phase, simultaneously learning and classifying these contents is challenging. 2) Different from dense sequences (eg.video), while multi-phase CE-CT imaging retains a consistent overall structure across phases, the content differs considerably, posing a significant challenge to capturing inter-phase dependencies [3]. What's more, the inherent intricacies of CT images also amplify the challenge [8]. Specifically, 1) CT imaging exhibits low sensitivity to CAs, making it challenging to obtain an accurate feature representation of CAs [5]. 2) CT imaging exhibits low contrast in soft-tissue structures, which makes extracting discriminative tissue representations for different phases challenging. These challenges have resulted in the synthesis of multi-phase contrast-enhanced CE-CT images remaining largely unexplored.

We propose a novel Causality-driven Spatio-temporal Generator (CSGen) with three key novelties for synthesizing multi-phase CE-CT images. Specifically, *1):* using a novel phase-causality to exploit the multi-phase variation con-

tent for driving the multi-phase CE-CT synthesizing (Fig. 1.2), addressing the challenge of capturing multi-phase discriminative features through one model. Unlike previous approaches that directly intervene from inputs to outputs, our phase-causality builds a causal graph to represent the dependency among multi-phase images, variant factors (eg. CAs), and semantic contents and conducts counterfactual inference over the causal graph to exploit the intrinsic causality of the discrepancy between the multi-phase images. More specifically, it can drive generator in exploiting the multi-phase variation content for learning multi-phase CA-responding discriminative content from each phase. *2):* A new cross-phase Spatio-temporal Transformer (StTransformer) is proposed, establishing the spatio-temporal correlation between multi-phase CE-CT images. Specifically, StTransformer leverages the dependencies within multi-phase CE-CT images from both spatial and temporal perspectives, which empowers the model to harness content and structural correlations across multi-phase CE-CT imaging. Additionally, *3):* a multi-phase adversarial learning (MpAL) is proposed. Rather than focusing solely on the differences in feature distributions between synthetic and real images, MpAL leverages multi-phase information to enhance the discriminator's ability to distinguish among multiple phases, further encouraging the generator to learn discriminative features across all phases. Experimental results on three-phase CE-CT imaging synthesis show that CSGen significantly improves the quality of synthetic multi-phase CE-CT images and outperforms the state-of-the-art methods.

Our main contributions include:

- For the first time, multi-phase CE-CT imaging synthesis through a single model has been achieved. A single, phase-interaction methodology that eliminates the need for chemical CAs in synthesizing multi-phase CE-CT images.
- CSGen, for the first time, converts multi-phase CE-CT imaging synthesis to sequence synthesis, establishing the spatio-temporal correlation between multi-phase CE-CT images, addressing the limitation in exploiting correlation of multi-phase images.
- Phase-causality, for the first time, is used in synthesizing multi-phase CE-CT imaging. Advanced than previous approaches that directly intervene from inputs to outputs, it focuses on the causal inference between cross-phase feature representation space, allowing for more flexible and adaptive interventions.
- StTransformer has the advantage of establishing cross-phase spatio-temporal correlation, which addresses the challenge of Transformer in exploiting multi-phase spatial and temporal dependencies.

2 Method

Our Causality-driven Spatio-temporal Generator (CSGen) integrates causal inference with multi-phase adversarial learning, formulating multi-phase CE-CT image synthesis as a sequence synthesis task (Fig. 2.2). Specifically, given

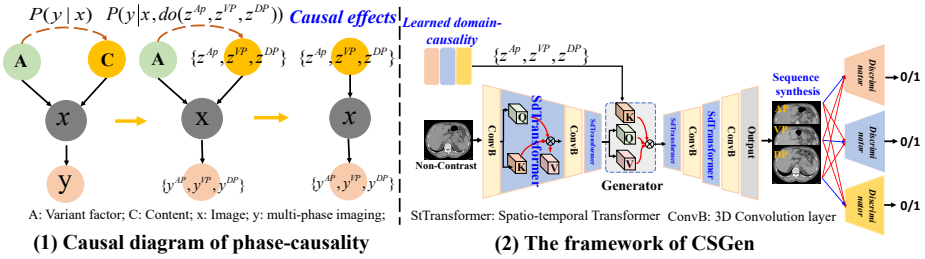

Fig. 2. 1) The causal diagram of our phase-causality. 2)The framework of CSGen.

a NCCT image, CSGen employs a 3D generator—composed of stacked convolutional blocks and a Spatio-temporal Transformer—guided by learned phase causality to generate the multi-phase CE-CT sequence, including Portal Venous Phase (VP), Arterial Phase (AP), and Delayed Phase (DP).

2.1 Phase-causality for learning phase-specific content

According to [11], the causal relationship between NCCT and CE-CT images can be defined as follows: the multi-phase CE-CT images $\{y^{VP}, y^{AP}, y^{DP}\} \in R^{C\times H\times W}$ are generated based on two independent factors: contrast agent:$\boldsymbol{A}$ and content: $\boldsymbol{C}$ (Fig. 2.1). Here, $\boldsymbol{A}$ represents the injected CAs, while $\boldsymbol{C}$ represents the tissue-specific response observable in NCCT. Factor $\boldsymbol{A}$ modulates the pixel intensities of various tissues over different time intervals, thereby yielding multi-phase CE-CT images. Building on this causal inference, we implement our phase-causality learning by extracting the phase-specific content latent code $\{z^{AP}, z^{VP}, z^{DP}\}$. Inspired by latent autoencoders [4, 14], we employ an adversarial latent autoencoder to encode the phase-specific content latent. The adversarial latent autoencoder consists an encoder (E), a decoder (G), a discriminator (D), they learn to represent images with codes from a learned, discrete codebook $\mathcal{Z} = \{z^{AP}, z^{VP}, z^{DP}\}$. More precisely, we approximate a given image y^i by $\hat{x}^i = G(z^i)$. We obtain z^i using the encoder $\hat{z}^i = E(y^i)$ and a subsequent element-wise quantization $q(.)$ or each spatial code $\hat{z}^i$ onto its closet codebook z^i:

$$z^i = q(\hat{z}^i) := (\arg\min \left\|\hat{z}^i - z^i\right\|) \in R^{n\times h\times w} \tag{1}$$

To enable the codebook to learn discriminative feature representations from each phase, a discriminator oversees the learning process. This end-to-end codebook training is guided by the following loss function:

$$\mathcal{L}_{VQ}(E, G, \mathcal{Z}) = \|y^i - \hat{y^i}\|^2 + \|\text{sg}[E(y^i)] - z^i\|_2^2 \tag{2}$$

$$Loss = \arg\min_{E,G,\mathcal{Z}} \max_{D} \mathbb{E}_{x\sim q(x)} \left[\mathcal{L}_{VQ}(E, G, \mathcal{Z}) + \mathcal{L}_{GAN}(\{E, G, \mathcal{Z}\}, D)\right], \tag{3}$$

where, L_{GAN} is adversarial loss, $\|\text{sg}[E(y^i)] - z^i\|_2^2$ denotes commitment loss. To improve the discrimination ability, three independent discriminators are used for

three-phase CE-CT imaging evaluation. After model training, we can obtain the phase-specific content latent.

2.2 Spatio-temporal Transformer establishes cross-phase correlation

To exploit spatio-temporal correlations among multi-phase images, the Spatio-temporal Transformer (StTransformer) employs two key modules: (1) Cross-phase Attention Module (CAM) and (2) Inter-phase Attention Module (IAM). CAM explores both content and structural correlations across multi-phase CE-CT images. Meanwhile, IAM establishes spatial dependencies within each phase by utilizing inter-phase variance information, which captures the distribution of CAs and extracts distinctive feature representations from multi-phase CE-CT images. Given a sequence feature map $\mathrm{X} \in \mathbb{R}^{C\times D\times H\times W}$, where C is the number of channels, D, H, and W represent the spatial dimension. The CAM projects X into sequences $[Q_D, K_D, V_D] \in R^{(CHW)\times D}$, and performs attention calculation from the temporal aspect. Afterward, IAM projects Y_D into new sequences $[X_Q; X_K; X_V] \in R^{\mathrm{D}\times(CHW)}$, and computing the spatial dependencies within each phase:

$$Y_D = \mathrm{softmax}\left(\frac{Q_D K_D^T}{d_D}\right) V_D, \quad X = \mathrm{softmax}\left(\frac{X_Q X_K^T}{d}\right) X_V. \tag{4}$$

where d is a learnable scaling parameter. The two modules address long-range dependencies in both spatial and temporal dimensions, facilitating the exploration of inter- and intra-phase relationships within multi-phase images.

2.3 Multi-phase adversarial learning for discriminative training

To endow the generator with greater discriminative power, CSGen uses multi-phase information for adversarial learning. Specifically, all of target multi-phase CE-CT images that consist of cross-phase feature distribution are used as negative samples for training. The training loss for multi-phase adversarial learning (MpAL) is:

$$L_{MpAL} = \sum_{\tilde{y}^i, y^j} L_{ce}(y^j, 0)\mathbb{1}[\tilde{y}^i \neq y^j] + L_{ce}(\tilde{y}^i, 0) + L_{ce}(y^i, 1) \tag{5}$$

where $\mathbb{1}$ is the binary indicator denoting whether the pixel belongs to same class, L_{ce} is cross entropy loss function, $\tilde{y}_i$, y_i represents the pixel in synthetic multi-phase CE-CT images $\{\tilde{y}^{VP}, \tilde{y}^{AP}, \tilde{y}^{DP}\}$ and CE-CT images $\{y^{VP}, y^{AP}, y^{DP}\}$, respectively.

2.4 Model training

In summary, the pre-learned phase-specific content latent from phase-causality is initially injected into the generator through self-attention (Fig. 2.2). Under the supervision of MpAL, the generator is supervised by

$$Loss = \alpha L_{pixel-wise} + \beta L_{perceptual} + \lambda L_{MpAL} \tag{6}$$

Table 1. Quantitative results of multi-phase CE-CT imaging synthesis evaluated from holistic (top) and tumor (bottom) perspectives.

Method	Mapping	Holistic evaluation											
		PSNR ↑				SSIM ↑				NMAE ↓			
		AP	VP	DP	Mean	AP	VP	DP	Mean	AP	VP	DP	Mean
CycleGAN [15]	One-to-One	29.39	28.93	28.63	28.98	0.8833	0.8802	0.8788	0.8808	3.68	3.92	3.81	3.80
AGAN [6]	One-to-One	29.05	29.16	29.11	29.11	0.8731	0.8666	0.8731	0.8709	4.01	3.92	3.91	3.95
CyTran [13]	One-to-One	29.25	29.36	28.92	29.18	0.8345	0.8718	0.8624	0.8562	4.39	3.97	4.17	4.18
CSGen (our)	One-to-Three	**31.24**	**31.34**	**30.87**	**31.15**	**0.9074**	**0.9081**	**0.9044**	**0.9066**	**3.14**	**3.13**	**3.25**	**3.17**
Method	Mapping	Tumor evaluation											
		AP	VP	DP	Mean	AP	VP	DP	Mean	AP	VP	DP	Mean
CycleGAN [15]	One-to-One	29.46	29.24	28.65	29.12	0.8884	0.8867	0.8844	0.8865	3.58	3.74	3.71	3.68
AGAN [6]	One-to-One	29.14	29.27	29.22	29.21	0.8787	0.8726	0.8792	0.8768	3.95	3.82	3.80	3.86
CyTran [13]	One-to-One	29.33	29.43	29.01	29.26	0.8377	0.8772	0.8682	0.8610	4.32	3.86	4.06	4.08
CSGen (our)	One-to-Three	**31.27**	**31.43**	**30.91**	**31.20**	**0.9125**	**0.9133**	**0.9101**	**0.9120**	**3.02**	**3.00**	**3.12**	**3.05**

$$L_{perceptual}(\tilde{y}^i, y^i) = \left\| VGG(\tilde{y}^i) - VGG(y^i) \right\|_1 \tag{7}$$

where, α, β and λ are weight coefficients. The $L_{pixel-wise}$ computes the difference between synthetic multi-phase CE-CT images $\{\tilde{y}^{VP}, \tilde{y}^{AP}, \tilde{y}^{DP}\}$ and target multi-phase CE-CT images $\{y^{VP}, y^{AP}, y^{DP}\}$ at the pixel level, where L_1 loss is used. The perceptual loss $L_{perceptual}$ is utilized to compute the difference between synthetic and true multi-phase CE-CT images in higher feature representations extracted from pre-trained VGG network.

$$L_{perceptual}(\tilde{y}^i, y^i) = \left\| VGG(\tilde{y}^i) - VGG(y^i) \right\|_1 \tag{8}$$

3 Experiments

3.1 Dataset and implementation

A total of 92 well-registered multi-phase CT volumes are used for evaluation. Each volume has four phases including a non-contrast phase and three CA-enhanced phases, VP, AP, and DP. During training, the voxel intensity was normalized within each subject to a scale of $[-1, 1]$ via division by the maximum intensity. The training, validation, and test sets were randomly split in a 7:1:2 ratio. Our model is implemented using PyTorch and trained end-to-end with Adam optimization. During training, the learning rate is initially set to 0.0001 and decreased by a weight decay of 1.0×10^{-6}. The weight coefficients α, β, λ are set to 1.0, 1.0, 0.5 respectively. The performance is evaluated from **1) Holistic aspect**: evaluating the synthesis quality from a global view; and **2) Local aspect**: evaluating the sensitivity to tumor by using tumor labels. The peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) metrics, and normalized mean absolute error (NMAE) between the synthesized and true CE-CT images are measured.

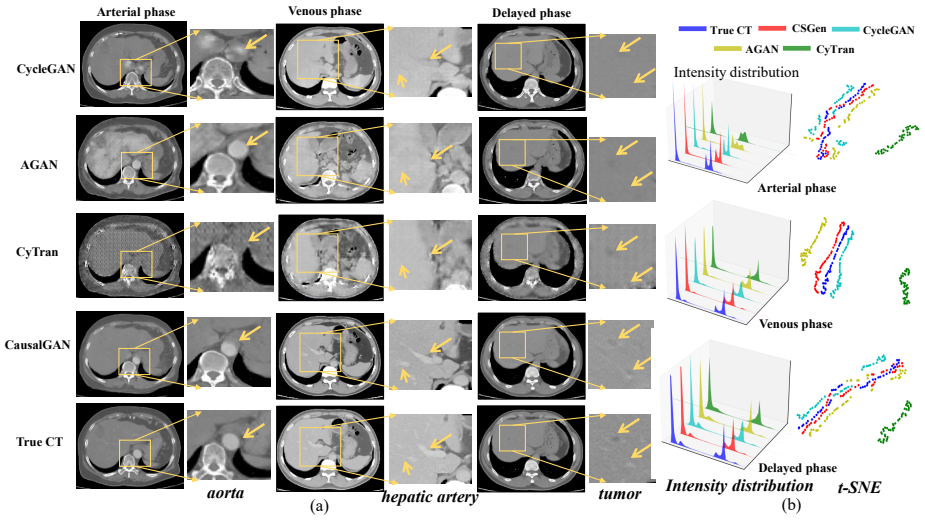

Fig. 3. Experimental results demonstrate that our model can more accurately enhance hepatic artery, aorta, and liver tumors than existing methods. b) The intensity distribution and t-SNE visualization of synthesis CE-CT images.

3.2 Comparison with state-of-the-art methods

Table 1 shows the quantitative result of CSGen and state-of-the-art methods from both holistic and liver tumor perspectives. In all three phases of CE-CT imaging, our CSGen exhibited superior performance in terms of PSNR, SSIM, and NMAE compared to other methods. Additionally, from the quantitative result of synthetic tumors, CSGen also achieved the best quantitative result. Notably, all compared methods employ a one-to-one mapping approach. These experimental results demonstrate that our CSGen has higher sensitivity to CA-aware feature distribution and are able to successfully synthesize multi-phase CE-CT images. Fig. 3 (a) clearly shows that our CSGen consistently enhances these CA-related organs including hepatic artery, aorta, and tumors across all phases, which provides compelling evidence for the effectiveness of our model in the synthesis of multi-phase CE-CT images. The feature distribution (Fig. 3 (b)) of synthesized CE-CT images reveals that the intensity and feature distribution (features extracted by VGG) of CE-CT images synthesized by our CSGen is notably closer to true CE-CT images compared to other methods, which highlights the effectiveness of our approach in leveraging cross-domain discriminative features and proves that existing methods have low sensitivity in the CAs responding areas.

3.3 Effect analysis of CSGen

Table 2 lists the quantitative results of various modules. StTransformer assists the baseline to obtain 1.96, 2.18, and 1.08 holistic improvements and 1.91,

Table 2. Quantitative results of synthesized CE-CT imaging.

Method				Holistic evaluation											
				PSNR↑				SSIM (%)↑				NMAE↓			
Baseline	StTransformer	PC	MpAL	AP	VP	DP	Mean	AP	VP	DP	Mean	AP	VP	DP	Mean
✓				28.39±3.00	28.36±2.88	27.88±2.67	28.21	87.98±5.86	88.04±5.74	86.87±5.54	87.63	4.24±1.43	4.25±1.44	4.43±1.46	4.31
✓	✓			30.31±4.37	30.27±4.41	29.94±4.50	30.17	89.08±5.62	90.32±5.81	90.02±5.84	89.81	3.25±1.56	3.18±1.70	3.27±1.72	3.23
✓		✓		30.44±4.26	30.44±4.28	29.97±4.29	30.28	87.43±5.80	90.35±5.84	89.83±5.78	89.20	3.30±1.55	3.11±1.67	3.27±1.67	3.23
✓			✓	28.76±2.99	28.74±3.12	28.34±2.94	28.61	86.27±2.99	89.34±5.12	88.58±4.91	88.06	3.72±1.33	3.62±1.37	3.76±1.35	3.70
✓	✓	✓	✓	31.24±5.23	31.34±5.37	30.87±5.32	31.15	90.74±6.06	90.81±6.22	90.44±6.29	90.66	3.14±1.69	3.13±1.83	3.25±1.88	3.17
				Tumor evaluation											
Baseline	StTransformer	PC	MpAL	AP	VP	DP	Mean	AP	VP	DP	Mean	AP	VP	DP	Mean
✓				28.44 ± 2.90	28.45±2.77	27.97±2.60	28.29	88.54 ± 5.39	88.62±5.30	87.51±5.02	88.23	4.18±1.36	4.18±1.31	4.33±1.35	4.23
✓	✓			30.29±4.17	30.29±4.27	30.01±4.38	30.20	89.56±4.88	90.78±5.14	90.59±5.07	89.31	3.15±1.41	3.09±1.53	3.15±1.56	3.13
✓		✓		30.41±3.99	30.50±4.08	30.04±4.18	30.32	87.97±4.96	90.89±5.12	90.42±4.99	89.76	3.19±1.38	2.98±1.48	3.15±1.49	3.11
✓			✓	28.97±2.81	28.95±2.90	28.51±2.78	28.81	87.05±4.23	90.11±4.31	89.37±4.06	88.84	3.54±1.17	3.44±1.18	3.59±1.20	3.52
✓	✓	✓	✓	31.27±5.04	31.43±5.22	30.91±5.15	31.20	91.25±5.39	91.33±5.59	91.01±5.50	91.20	3.02±1.55	3.00±1.64	3.12±1.69	3.05

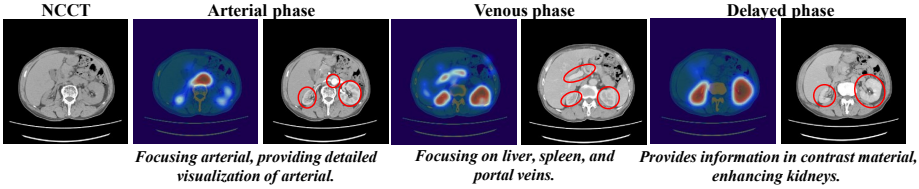

Fig. 4. The heatmap generated by phase-causality on three phases.

1.08, and 1.10 local improvements on mean PSNR, mean SSIM, and mean NMAE, respectively. These improvements reveal the advantage of StTransformer in utilizing the spatio-temporal correlation of multi-phase CE-CT images. The phase-causality (PC) assists the baseline to obtain a 2.07, 1.57, and 1.08 holistic improvement and obtain a 2.03, 1.53, and 1.12 local improvement on mean PSNR, mean SSIM, and mean NMAE PSNR, respectively. Those improvements demonstrate that phase-causality enhanced the generator's ability to learn CA-responding content information. Furthermore, the learned phase-causality contents are illustrated in Fig. 4, which reveals its aligns with discriminative features specific to each phase. In AP phase, the arterial vasculature is captured and enhanced, revealing intricate details of arterial structures and early enhancement within organs. In VP phase, features of the portal venous system are emphasized. In DP phase, the excretory characteristics of the contrast material become more apparent, particularly enhancing kidney visualization. These experimental results underscore the excellent ability of our phase-causality in learning CA-responding content from each CE-CT image. Additionally, MpAL assists the model to obtain a 0.40, 0.43, and 0.61 holistic improvement, and 0.52, 0.61 and 0.71 local tumor improvement on mean PSNR, mean SSIM, and mean NMAE, respectively. These improvements also prove the effectiveness of MpAL in capturing global and local discriminative contextual information.

4 Conclusion

Our CSGen is the first to successfully synthesize three-phase CE-CT imaging from NCCT using a single model. CSGen uses three novel mechanisms: 1) phase-

causality mechanism; 2) spatio-temporal Transformer; 3) multi-phase adversarial learning to address the challenge of three-phase CE-CT synthesis. Experiments prove that the synthesized three-phase CE-CT imaging by CSGen is equivalent to real CE-CT imaging and outperforms state-of-the-art synthesis methods.

Acknowledgments. The National Institutes of Health supported this work under grants R01-HL167199, R01-HL165218, R01-HL165218-01A1W1, and R44-HL156811. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

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

References

1. Abdelhalim, I., Abou El-Ghar, M., Dwyer, A., Ouseph, R., Contractor, S., El-Baz, A.: A New Non-Invasive AI-Based Diagnostic System for Automated Diagnosis of Acute Renal Rejection in Kidney Transplantation: Analysis of ADC Maps Extracted from Matched 3D Iso-Regions of the Transplanted Kidney . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15012. Springer Nature Switzerland (October 2024)
2. Bae, H., Lee, H., Kim, S., Han, K., Rhee, H., Kim, D.k., Kwon, H., Hong, H., Lim, J.S.: Radiomics analysis of contrast-enhanced ct for classification of hepatic focal lesions in colorectal cancer patients: its limitations compared to radiologists. European Radiology **31**(11), 8786–8796 (2021)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
4. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12873–12883 (2021)
5. Holalkere, N.S., Sahani, D.V., Blake, M.A., Halpern, E.F., Hahn, P.F., Mueller, P.R.: Characterization of small liver lesions: added role of mr after mdct. Journal of computer assisted tomography **30**(4), 591–596 (2006)
6. Hu, T., Oda, M., Hayashi, Y., Lu, Z., Kumamaru, K.K., Akashi, T., Aoki, S., Mori, K.: Aorta-aware gan for non-contrast to artery contrasted ct translation and its application to abdominal aortic aneurysm detection. International Journal of Computer Assisted Radiology and Surgery pp. 1–9 (2022)
7. Huang, W., Liu, W., Zhang, X., Yin, X., Han, X., Li, C., Gao, Y., Shi, Y., Lu, L., Zhang, L., et al.: Lidia: Precise liver tumor diagnosis on multi-phase contrast-enhanced ct via iterative fusion and asymmetric contrastive learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 394–404. Springer (2024)
8. Kim, A., Saharkhiz, N., Sizikova, E., Lago, M., Sahiner, B., Delfino, J., Badano, A.: S-SYNTH: Knowledge-Based, Synthetic Generation of Skin Images . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15003. Springer Nature Switzerland (October 2024)

9. Marckmann, P., Skov, L., Rossen, K., Dupont, A., Damholt, M.B., Heaf, J.G., Thomsen, H.S.: Nephrogenic systemic fibrosis: suspected causative role of gadodiamide used for contrast-enhanced magnetic resonance imaging. Journal of the American Society of Nephrology **17**(9), 2359–2362 (2006)
10. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
11. Mitrovic, J., McWilliams, B., Walker, J., Buesing, L., Blundell, C.: Representation learning via invariant causal mechanisms. arXiv preprint arXiv:2010.07922 (2020)
12. Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering **65**(12), 2720–2730 (2018)
13. Ristea, N.C., Miron, A.I., Savencu, O., Georgescu, M.I., Verga, N., Khan, F.S., Ionescu, R.T.: Cytran: A cycle-consistent transformer with multi-level consistency for non-contrast to contrast ct translation. Neurocomputing **538**, 126211 (2023)
14. Tschannen, M., Bachem, O., Lucic, M.: Recent advances in autoencoder-based representation learning. arXiv preprint arXiv:1812.05069 (2018)
15. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
16. Zhu, Q., Wang, Y., Zhu, S., Du, B.: Partial consistent adversarial unified framework for unsupervised non-contrast ct cross-domain adaptation and segmentation. Pattern Recognition **165**, 111638 (2025)
17. Zhu, Q., Wentland, A.L., Li, S.: Contrast-aware network with aggregated-interacted transformer and multi-granularity aligned contrastive learning for synthesizing contrast-enhanced abdomen ct imaging. IEEE Transactions on Computational Imaging (2025)
18. Zhu, Q., Zhu, S., Du, B., Wang, Y.: Cross-domain distribution adversarial diffusion model for synthesizing contrast-enhanced abdomen ct imaging. Pattern Recognition p. 111695 (2025)
19. Zhu, X., Zhang, W., Li, Y., O'Donnell, L.J., Zhang, F.: When Diffusion MRI Meets Diffusion Model: A Novel Deep Generative Model for Diffusion MRI Generation . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15002. Springer Nature Switzerland (October 2024)