

Common-Unique Decomposition Driven Diffusion Model for Contrast-Enhanced Liver MR Images Multi-Phase Interconversion

Chenchu Xu, *Member, IEEE*, Shijie Tian, Boyan Wang, Jie Zhang, Kemal POLAT, *Senior Member, IEEE*, Adi Alhudhaif, and Shuo Li, *Senior Member, IEEE*

Abstract—All three contrast-enhanced (CE) phases (e.g., Arterial, Portal Venous, and Delay) are crucial for diagnosing liver tumors. However, acquiring all three phases is constrained due to contrast agents (CAs) risks, long imaging time, and strict imaging criteria. In this paper, we propose a novel Common-Unique Decomposition Driven Diffusion Model (CUDD-DM), capable of converting any two input phases in three phases into the remaining one, thereby reducing patient wait time, conserving medical resources, and reducing the use of CAs. 1) The Common-Unique Feature Decomposition Module, by utilizing spectral decomposition to capture both common and unique features among different inputs, not only learns correlations in highly similar areas between two input phases but also learns differences in different areas, thereby laying a foundation for the synthesis of remaining phase. 2) The Multi-scale Temporal Reset Gates Module, by bidirectional comparing lesions in current and multiple historical slices, maximizes reliance on previous slices when no lesions and minimizes this reliance when lesions are present, thereby preventing interference between consecutive slices. 3) The Diffusion Model-Driven Lesion Detail Synthesis Module, by employing a continuous and progressive generation process, accurately captures detailed features between data distributions, thereby avoiding the loss of detail caused by traditional methods (e.g., GAN) that overfocus on global distributions. Extensive experiments on a generalized CE liver tumor dataset have demonstrated that our CUDD-DM achieves state-of-the-art performance (improved the SSIM by at least 2.2% (lesions area 5.3%) comparing the seven leading methods). These results demonstrate that CUDD-

This work was supported in part by The National Natural Science Foundation of China (62106001, U1908211), The University Synergy Innovation Program of Anhui Province (GXXT-2021-007), and The Anhui Provincial Natural Science Foundation (2208085Y19). (Corresponding authors: Boyan Wang and Jie Zhang.)

C. Xu, and S. Tian are with the School of Computer Science and Technology, Anhui University, Hefei, China (e-mail: cxu332@gmail.com).

B. Wang is with the Department of Computer Science and Technology, Tsinghua University, Beijing, China (e-mail: wby00000@mail.tsinghua.edu.cn).

J. Zhang is with the Information Center, the First Affiliated Hospital of Anhui Medical University, Hefei, China (e-mail: jiezhang.ayfy@163.com).

Kemal POLAT is with the Department of Electrical and Electronics Engineering, Bolu Abant Izzet Baysal University, Bolu, Turkey (e-mail: kpolat@ibu.edu.tr).

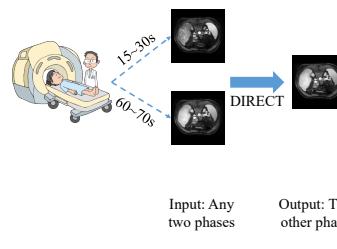
Adi Alhudhaif is with the College of Computer Engineering and Sciences in Al-kharj, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia (e-mail: a.alhudhaif@psau.edu.sa).

S. Li is with the School of Medicine, Case Western Reserve University, Cleveland, United States (e-mail: slishuo@gmail.com).

Our method

Advantages:

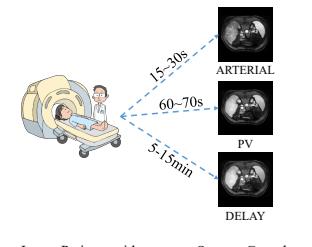
- 1) Reduce the harm to the body caused by excessive use of contrast agents.
- 2) Reduce patient waiting time.
- 3) Complete diagnostic information.



Input: Any two phases

Output: The other phase

Traditional



Input: Patients with contrast undergo MRI

Output: Complete three-phase image

Fig. 1. A potential clinical alternative is proposed to reduce the use of contrast agents (CAs) in CE liver MR imaging. This method not only reduces patient waiting time and conserves the utilization of clinical resources, it is also enable a reduction in the use of CAs.

DM advances CE liver tumor imaging technology.

Index Terms—Contrast-enhanced Liver Magnetic Resonance Imaging, Multi-phase interconversion, Feature Decomposition.

I. INTRODUCTION

ALL three phases of contrast-enhanced (CE) liver MR images (e.g., Arterial, PORTAL VEIN (PV), and Delay) are paramount for accurately diagnosing and formulating treatment strategies for liver tumors, each offering unique insights into hepatic conditions [1]–[5]. The Arterial phase appears 15–30 seconds after contrast agents (CAs) injection, prominently displaying areas supplied by the Arterial system of liver. Subsequently, the PV phase emerges 60–70 seconds after the Arterial phase, illuminating potential lesions like normal hepatocytes and metastatic liver cancer. Lastly, the Delay phase unfolds 5–15 minutes post-injection, highlighting specific lesions that contrast sharply with normal hepatocytes. The simultaneous employment of the three phases is clinically recognized to enable an enhancement in diagnostic efficiency, due to: the combined data from all three phases enables the differentiation of lesions based on their enhancement patterns at different times, which might be indiscernible when viewed in isolation. Furthermore, this three phases approach enables the pivotal detection of subtle lesions or changes in vascular

flow that become apparent only by observing the response of the liver across these successive phases.

However, for some patients, obtaining all three phases presents clear limitations. 1) The extensive use of CAs during the Delay phase poses significant health risks, particularly for those with compromised kidney functionality [6]. Administering a high volume of these agents enables discomfort and exacerbates the physical strain on the patient. For those with kidney issues, this is especially problematic. Reducing the volume of CAs used is thus essential to mitigate these risks and alleviate the associated physical burden on patients. 2) Another issue is the significant time consumption for obtaining all three phases [7]. The need to spend upwards of 15 minutes to acquire comprehensive data through all phases not only puts a strain on medical resources but also extends the waiting time for patients. 3) The stringent imaging requirements of Arterial phase pose challenges for certain patients, hindering the accurate acquisition of data [8]. Stillness is necessary during this phase to prevent motion artifacts, a condition that might be particularly difficult for certain people, such as children. Therefore, this introduces a substantial hurdle in obtaining reliable and precise imaging data.

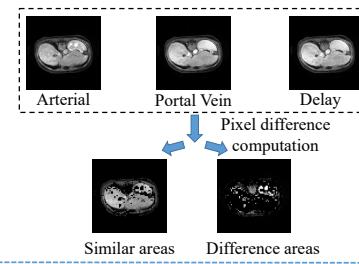
Therefore, a strong clinical desire exists for a method that allows interconversion among the three phases. Particularly, a 2-to-1 phase interconversion (Figure 1) enable leverage comprehensive information from the available phases to synthesize the most accurate images.

However, as shown in Figure 2, accurately achieving interconversion among the three phases of CE liver MR images poses several challenges: 1) The three phases are highly similar, and the characteristic differences between them are small, making it difficult to detect unique characteristics of the lesions. Typically, in the Arterial, PV, and Delay phases, over 95% of the areas in the images are highly similar, with less than 5% displaying unique changes of lesions. 2) Slices with inconsistent dependencies cause mutual interference between adjacent frames. Utilizing the dependencies from adjacent frames can facilitate better interconversion for subsequent slices. However, in CE liver MR images, lesions might appear unexpectedly between consecutive slices, and leveraging the dependencies from these lesions in the interconversion process can result in interference during multi-phase interconversion. 3) Liver tumor require precise diagnosis, and global distribution methods (e.g., GAN) are not always effective in achieving optimal interconversion. Particularly, GANs [9] tend to focus on broader image patterns, often neglecting the subtle details of liver tumor. This oversight can compromise the accuracy of tumor representation, making diagnosis more challenging.

In this paper, we for the first time proposed a Common-Unique Decomposition Driven Diffusion Model (CUDD-DM) for CE liver MR images multi-phase interconversion, capable of converting any two input phases in three phases into the remaining one phase. Firstly, our CUDD-DM method innovatively utilizes spectral decomposition for extracting both common and unique features across input phases. CUDD-DM carries out a spectral decomposition on the two inputs, segregating them into low-frequency and high-frequency features. These are subsequently merged, enabling us to discern the

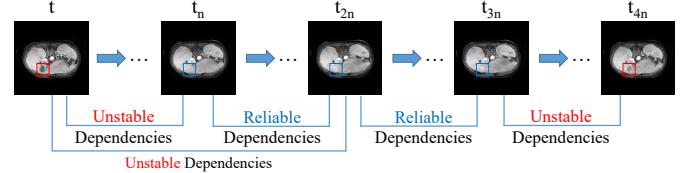
Challenge 1

The three phases are highly similar, and the characteristic differences between them are small, making it difficult to detect unique characteristics of the lesions.



Challenge 2

The randomness of lesions causes inconsistent dependencies between slices, leading to mutual interference between slices.



Challenge 3

Detail missing in synthesized image caused by overfocus global distribution (e.g., GAN).

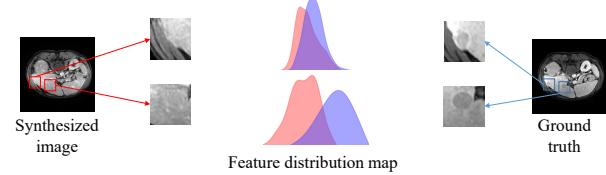


Fig. 2. Three challenges for CE MR Images Multi-Phase Interconversion: 1) The three phases are highly similar, and the characteristic differences between them are small, making it difficult to detect unique characteristics of the lesions. 2) Slices with inconsistent dependencies cause mutual interference between adjacent frames. 3) Liver tumor require precise diagnosis, and global distribution methods (e.g., GAN) are not always effective in achieving optimal interconversion.

common and unique features of both inputs. Secondly, utilizing multiple reset gates, CUDD-DM conducts a comparative analysis between adjacent slices to identify any pathological variations. Finally, CUDD-DM endeavors to integrate the diffusion model into the multi-phase interconversion task of CE liver MR images. Specifically, CUDD-DM includes three parts: 1) Common-Unique Feature Decomposition Module (CUFD). CUFD divides the features into low-frequency and high-frequency through a spectral decomposition approach. By employing a decomposition loss function, it makes the low-frequency features correlated and the high-frequency features uncorrelated. This process distinguishes between the over 95% similarity across the Arterial, PV, and Delay phases and the unique changes that account for less than 5% of the images. Consequently, it results in obtaining low-frequency, minimally varying common features and high-frequency, lesion-unique features, addressing the challenge of detecting unique characteristics in the lesions. 2) Multi-scale Temporal Reset Gates Module (MTRG). MTRG takes input from CUFD and processes slices of CE liver MR images through multiple consecutive temporal reset gates. The key advantage is its ability to utilize dependencies that don't undergo lesion areas and avoid depending on areas that experience lesions. This

design effectively handles the issue of slices with inconsistent dependencies, ensuring that unexpected lesion appearances between consecutive slices don't lead to interference in the interconversion process. 3) Diffusion Model-Driven Lesion Detail Synthesis Module (DM-DLDS). DM-DLDS takes inputs from MTRG and implements a added noise diffusion and denosing process. The DM-DLDS employs a two-process diffusion approach. In the forward process, randomness is intricately introduced using a noise-adding mechanism, mimicking the diffusion of data. Following this, the backward process utilizes a refined denoising process, acting as a corrective mechanism to extract the essential features from the noisy data. This diffusion-centric approach ensures the synthesized images capture nuanced liver tumor details, thereby addressing the challenge where methods like GANs overlook these fine-grained characteristics.

Our contributions can be summarized in four aspects:

- For the first time, a new approach is proposed for CE liver MR images multi-phase interconversion. This unique approach not only reduces patient waiting time and conserves the utilization of clinical resources, it is also enable a reduction in the use of CAs.
- A novel spectral approach is employed to decompose both common and unique features, derived from low and high-frequency information. This process is further refined by an innovative forced decomposition loss function, ensuring precise feature decomposition.
- We propose the Multi-scale Temporal Reset Gates Module (MTRG), which enables our CUDD-DM to process multiple time scales of lesion areas appearing in consecutive slices of CE liver MR images.
- A new multi-phase interconversion framework based on a diffusion model is proposed to ensure accurate preservation of specific lesion details in liver tumor images during different phases of interconversion in MR imaging.

II. RELATED WORK

A. Existing image translation methods for MR images synthesis

MR images translation is critical for improving the accuracy of disease detection and accurate diagnosis. MR images translation involves changing the phase of the MR images, this solves the problem of unavailability of some phase images caused by uncooperative patients and excessive scan times [10]. Currently, the existing MR images translation methods [11]–[14] mainly uses the one-to-one method, that is, directly translating a MR images phase into another phase. However, many organs (e.g., liver, brain, etc.) usually has T1, T2 and DWI and many other phases. At this time, the traditional one-to-one translation method will no longer apply. [15] implements multi-phase translation by extracting shape information of multiple input images, [10] uses multiple Residual transformers to implement multi-phase translation, [16] based on disentanglement representations of anatomical and modality specific features implements multi-phase translation. Compared with existing techniques, our method specifically extracts the common features of multiple input images

as well as their respective unique features through spectral decomposition, providing more comprehensive information for synthesizing the remaining output images, thus providing a richer and more accurate image translation approach for diagnosis and treatment planning.

B. Existing multi-phase interconversion methods for CE liver MR images synthesis

Although several liver tumor analysis methods have recently been published, there is no work on multi-phase interconversion of CE liver MR images. Most of these methods, such as liver tumor detection [1], [6], tumor segmentation [17], [18] and tumor classification [19], [20], are performed on CE liver images. These methods significantly promote the progress of liver tumor analysis, especially liver cancer diagnosis. However, they often struggle with inconsistencies in lesion appearance, potential introduction of artifacts [21], [22]. Recently, harnessing the power of deep learning development, numerous methods related to tumor synthesis have been proposed [23]–[25]. Despite the significant progress in tumor synthesis, no articles on the multi-phase interconversion of CE liver MR images have yet been published. Therefore, in this study, we attempt to facilitate direct interconversion among the Arterial, PV, and Delay phases of CE liver MR images. Our proposed method has the potential to significantly reduce patient waiting times and conserves the utilization of clinical resources.

C. Common and Unique feature for CE liver MR images synthesis

CE liver MR images, by leveraging both common and unique features, enable a more precise depiction of liver tumor and their subtle pathological changes. Common features establish the fundamental structure and morphology of the liver, ensuring stable and consistent representation across all enhancement phases. This provides a robust foundation for the synthesis, ensuring the generated images maintain structural consistency [26]. Unique features, on the other hand, capture the specific nuances, especially details related to pathological changes. These nuances represent the subtle differences between enhancement phases. This not only guarantees the authenticity of the images in specific enhancement phases but also enhances the diagnostic accuracy for liver tumor [27]. Traditional methods employed for common and unique features extraction have demonstrated potential in specific scenarios, yet they typically necessitate intricate network architectures [28]. In contrast, our spectral decomposition method intuitively separates and extracts different frequency components of the image. This approach aims to more precisely capture the desired common and unique features, streamlining the feature extraction process.

D. Diffusion model in medical image synthesis

Recently, diffusion model has shown impressive performance in a variety of image synthesis tasks. Diffusion model, noted for its generative prowess in computer vision, employs

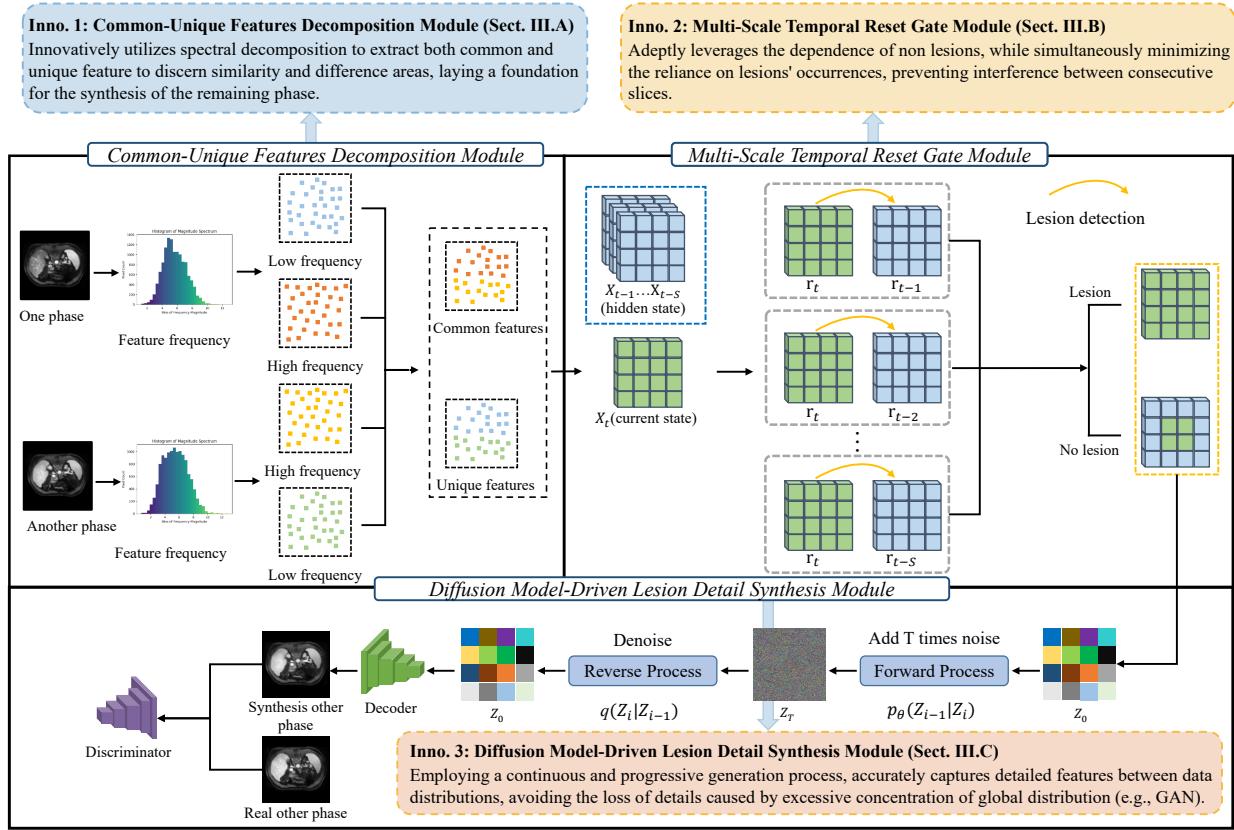


Fig. 3. Our CUDD-DM integrates a CUFD Module, a MTRG Module, and a DM-DLDS Module. The CUFD discerns common features across various input phases and identifies their unique features, while the MTRG addresses potential issues with lesions arising between consecutive slices. Furthermore, the DM-DLDS is engineered to ensure the complex multi-layer detail and high diagnostic needs of liver tumor images.

a dual-phase diffusion process to synthesize high-quality and diverse samples [29]. Thus, it can through its distinctive noise-adding and denoising process, ensuring that synthesized images are not only of elevated fidelity but also hold nuanced and realistic details [30]–[33], crucial for advancing diagnostic imaging applications. [30], [31] utilizes latent diffusion models to achieve High-resolution image synthesis, [32] performs diffusion-based resampling of fMRI samples to make fMRI interpretation rational, [33] utilizes latent diffusion models producing photorealistic synthetic brain images. However, existing diffusion model approaches predominantly focus on computer vision applications [34]–[36], with scarcely explored implementations within the medical domain [37]–[40], and notably, no applications have been ventured upon within the context of CE liver imaging. We innovatively attempt to integrate diffusion models into the interconversion task of CE liver MR images.

III. METHODOLOGY

Our CUDD-DM (Figure 3) is composed of a CUFD Module (Sect. III-A), a MTRG Module (Sect. III-B), and a DM-DLDS Module (Sect. III-C). Initially, the CUFD leverages spectral technology to simultaneously decompose the input from two different phases, effectively decomposing both the common features and the unique features to each phase. Subsequently, the MTRG performs lesion comparisons between the current slice and several adjacent frames using multiple reset gates.

These gates control the influence of adjacent frames on the current slice by adjusting the weighting between them. When a lesion is detected, the influence of adjacent frames is minimized to prevent erroneous correlations; otherwise, the adjacent frames have a greater impact on the current slice to enhance continuity and detail accuracy. Finally, the feature vector generated by MTRG is used to refine the multi-layer detail and high diagnostic needs of liver tumor through the DM-DLDS. DM-DLDS includes a forward process and a reverse process. The forward process turns the feature vector into a pure noise distribution by gradually adding gaussian noise. The reverse process is to denoise the pure noise to obtain a noise distribution that is the same as the input, and then in order to ensure the consistency between the synthetic image and the real image, the discriminator is used to conduct continuous adversarial training between the generated image and the real image.

A. Common-Unique Feature Decomposition Module for common and unique features extraction

The CUFD (as shown in Figure 4) not only creatively extracts the common features between two inputs through spectral decomposition, but also inventively isolates their respective unique features using the same method. The common features establish the basic structure and morphology of the liver, representing more than 95% highly similar regions

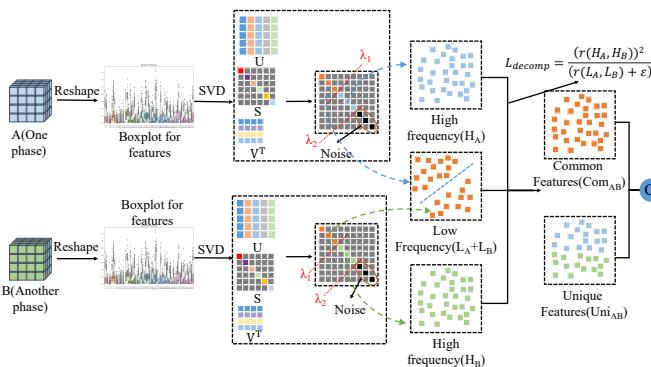


Fig. 4. The CUDF first applies spectral decomposition to the reshaped feature vectors, yielding matrices U , S , and V^T . Using the resultant singular value matrix along with the left and right singular vectors, both high and low-frequency features, as well as noise, are extracted. After noise removal, the low-frequency features L_A and L_B , and the high-frequency features H_A and H_B are merged. With the introduction of L_{decomp} , Com_{AB} and Uni_{AB} are derived. Ultimately, Com_{AB} and Uni_{AB} are integrated to complete the process.

between the two inputs. Unique features capture specific nuances, especially details related to pathological changes, representing less than 5% of regions that differ significantly between the two inputs. This corresponds to low-frequency information and high-frequency information in image processing. In image processing, low-frequency information represents the overall structure and smooth areas of the image while high-frequency information usually represents the details and edges of the image. Therefore, the CUDF regard the overall information of low frequency as common features and the detailed information of high frequency as unique features. Further, a new decomposition loss is introduced to better merge common and unique features.

Specifically, the CUDD-DM initially encodes the two input images to produce a feature vector of dimensions $(1, 64, 256, 256)$, dimensionality reduction to $(1, 64, 32, 32)$, followed by reshaping the resulting features to $A_{(64 \times 1024)}$, and then applying singular value decomposition (1) :

$$A_{64 \times 1024} = U_{64 \times 64} S_{64 \times 1024} V_{1024 \times 1024}^T \quad (1)$$

where the left singular vectors $U_{(64 \times 64)}$ and the right singular vectors $V_{1024 \times 1024}^T$ represent the features of the original data, whereas the singular value matrix $S_{(64 \times 1024)}$ is a diagonal matrix, with the values on the diagonal indicating the degree of data transformation. Among them, larger singular values usually correspond to the main features and structures in the image, while smaller singular values are more related to subtle changes and details of the image. This corresponds to low-frequency information and high-frequency information in image processing.

The CUDD-DM operates by separating the singular value matrix, leveraging the thresholds λ_1 and λ_2 to distinguish features into high-frequency, low-frequency, and noise segments. Following this segregation, the attributes derived from distinct frequency bands are integrated with their respective left (2) and right (3) singular vectors to form the low and high-frequency information pairs. These are denoted as follows:

$$L_A = U_A \cdot S_{L_A} \cdot V_A^T, L_B = U_B \cdot S_{L_B} \cdot V_B^T \quad (2)$$

$$H_A = U_A \cdot S_{H_A} \cdot V_A^T, H_B = U_B \cdot S_{H_B} \cdot V_B^T \quad (3)$$

where L_A, L_B represent the low-frequency information derived from the spectral of inputs A and B respectively, and H_A, H_B represent the high-frequency information similarly derived from inputs A and B. U_A, U_B and V_A^T, V_B^T are the left and right singular vectors from the spectral, and $S_{L_A}, S_{H_A}, S_{L_B}, S_{H_B}$ are segments of the singular value matrix, representing the low and high-frequency details, respectively, determined using thresholds λ_1 and λ_2 . Acquiring high and low-frequency information from the two input phases, a crucial decomposition loss, termed L_{decomp} , is introduced. Subsequently, L_A, L_B, H_A, H_B are merged to obtain Com_{AB} and Uni_{AB} , where Com_{AB} and Uni_{AB} represent the common and unique features derived from inputs A and B, respectively. The decomposition loss (4), L_{decomp} , is formulated as:

$$L_{decomp} = \frac{(r(H_A, H_B))^2}{(r(L_A, L_B) + \varepsilon)} \quad (4)$$

where $r(\cdot, \cdot)$ is the Pearson correlation coefficient operator, H_A, H_B represent the high-frequency details of inputs A and B, L_A, L_B represent the low-frequency details of inputs A and B, and ε is a small constant (set to 1.001 here) ensuring the denominator remains positive. Further, the Pearson correlation coefficient operator (5) is defined as:

$$r = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum (x_i - \mu_x)^2 \sum (y_i - \mu_y)^2}} \quad (5)$$

where x_i and y_i are individual samples, while μ_x and μ_y are the means of samples x and y, respectively.

The introduction of L_{decomp} fosters a higher correlation between low-frequency attributes while preserving the distinctness of high-frequency attributes. This innovative loss function not only accentuates the distinction between common and unique features across phases but also heightens the correlation of low-frequency features. Consequently, it surpasses conventional loss functions by efficiently eliminating redundant information and thereby enhancing the quality of feature extraction.

B. Multi-scale Temporal Reset Gates Module for inconsistent dependencies between slices

The MTRG (Figure 5) is a novel module within the CUDD-DM, designed to manage variable dependencies effectively and avoid mutual interference between slices during the interconversion process. MTRG utilizes multiple temporal reset gates to compare the current state with multiple historical states, incorporating both forward and backward dependencies. The reset gates output a result of 0 when a lesion is detected and 1 otherwise. Subsequent processing through a reset gate calculates the combined results. If any change between a historical state and the current state is detected—that is, the reset gate output is 0—the dependence on that historical state is minimized. Otherwise, the use of historical state dependencies is maximized.

More specifically, for the current slice processed by CUDF, denoted as X_t , it simultaneously enters multiple temporal reset gates ($Reset_1^f, Reset_2^f, \dots, Reset_S^f$) for forward processing

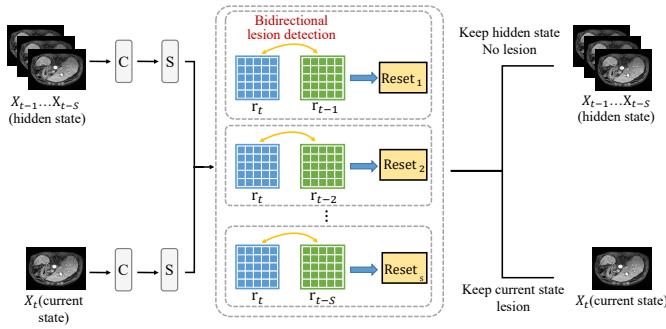


Fig. 5. The MTRG integrates multiple temporal reset gates, allowing for simultaneous bidirectional lesion comparison of the current state (r_t) with its multiple historical states (r_{t-1} to r_{t-S}) and adjusting for potential future states through bidirectional analysis. This innovative approach helps avoid interference when lesions occur between consecutive slices, enhancing the reliability of the interconversion process.

and ($Reset_1^b$, $Reset_2^b$, ..., $Reset_S^b$) for backward processing. The results $r_{t_i}^f$ and $r_{t_i}^b$ from comparing the current slice with each set of reset gates are obtained as follows:

$$r_{t_i}^f = Reset_i^f(X_t) \quad (6)$$

$$r_{t_i}^b = Reset_i^b(X_t) \quad (7)$$

where $Reset_i^f$ and $Reset_i^b$ represent the forward and backward reset gates at different time scales, each consisting of a sigmoid function and 2D convolution. The output $r_{t_i}^f$ and $r_{t_i}^b$ represent the degree of dependence of the current slice on the slices at the i -th scale in both forward and backward directions.

The current slice information is then fused with the slice information on S scales, for both forward and backward states, as expressed by:

$$C_{t_i}^f = (1 - r_{t_i}^f) \cdot X_t + r_{t_i}^f \cdot X_{t-i} \quad (8)$$

$$C_{t_i}^b = (1 - r_{t_i}^b) \cdot X_t + r_{t_i}^b \cdot X_{t+i} \quad (9)$$

where X_{t-i} and X_{t+i} represent the slice information at the i -th scale in the forward and backward directions, respectively. $C_{t_i}^f$ and $C_{t_i}^b$ represent the combined output results.

Upon obtaining the outputs at all scales in both directions, the information from these scales is concatenated to produce the final output Z_t . Simultaneously, the current slice information is updated to the hidden state, defined by:

$$X_{t-i} = (1 - r_{t_i}^f) \cdot C_{t_i}^f + r_{t_i}^f \cdot X_{t-i+1} \quad (10)$$

$$X_{t+i} = (1 - r_{t_i}^b) \cdot C_{t_i}^b + r_{t_i}^b \cdot X_{t+i-1} \quad (11)$$

C. Diffusion Model-Driven Lesion Detail Synthesis Module for high diagnostic needs of liver tumor images.

The DM-DLDS, introduced by CUDD-DM during the feature decoding from MTRG, ensures the complex multi-layer detail and high diagnostic needs of liver tumor images. The diffusion model (Figure 6) plays an irreplaceable role in this module. The forward process continuously adding noise to the feature vector output by the MTRG, and the reverse denoising process composed of a group of U-nets are used to continuously learn the probability distribution of the target

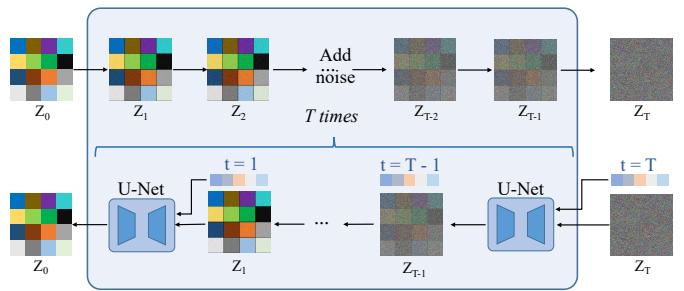


Fig. 6. The detail of main components in the diffusion model: forward process and reverse process. The forward process starting with a true feature representation of a data point, gradually adding noise to these features until they become almost random. The reverse process starting with the random noise features, the noise is gradually removed from these features using the U-net model to try to recover its original or meaningful representation.

image. Subsequently, the DM-DLDS decodes the output of the diffusion model for accurate reconstruction of the target image. In the final phase, a discriminator is employed to conduct adversarial training, aimed at optimizing the differentiation between real and synthesized target images.

The underlying principle of diffusion model design is the gradual approach of the input feature to the target distribution via a diffusion denoising process. Initially, noise is introduced into the original input features, creating a perturbed version of these features. This action is repeated t times, with fresh random noise introduced in each iteration, culminating in a sequence of perturbed features, a sequence termed the diffusion process. These features are then processed in the reverse sequence, with each feature undergoing denoising via the U-Net. This denoising procedure also recurs t times, with the U-Net incrementally converting the noisy feature closer to the target feature during each iteration. The diffusion (12) and denoising (13) processes of DDPM can be expressed as:

$$Z'_i = Z_{i-1} + N_i \quad (12)$$

$$Z''_i = F_i(Z'_i) \quad (13)$$

where Z'_i is the feature after the i -th diffusion, Z''_i is the feature after the i -th denoising, N_i is the noise sampled in the i -th iteration, and F_i is the i -th U-Net model.

Once the output Z from the diffusion is obtained, it can be utilized to synthesize the target phase images. Similar to the pixel-to-pixel synthesis method [41], our generator G tries to generate an image synthesis other phase image $G(Z)$ from the input Z . Meanwhile, the discriminator D tries to distinguish the generated image $G(Z)$ from the real other phase image y . Accordingly, the objective function of the generator loss L_G (14) is defined as:

$$\begin{aligned} L_G = & E_{p_1, p_2} [\log (1 - D(p_1, p_2, G(p_1, p_2)))] \\ & + \lambda E_{p_1, p_2, y} [\|y - G(p_1, p_2)\|_1] \end{aligned} \quad (14)$$

where λ is a non-negative trade-off parameter. The generator G aims to create a realistic image that can deceive the discriminator D , utilizing the L1-norm to quantify the discrepancy between the generated image and the corresponding authentic

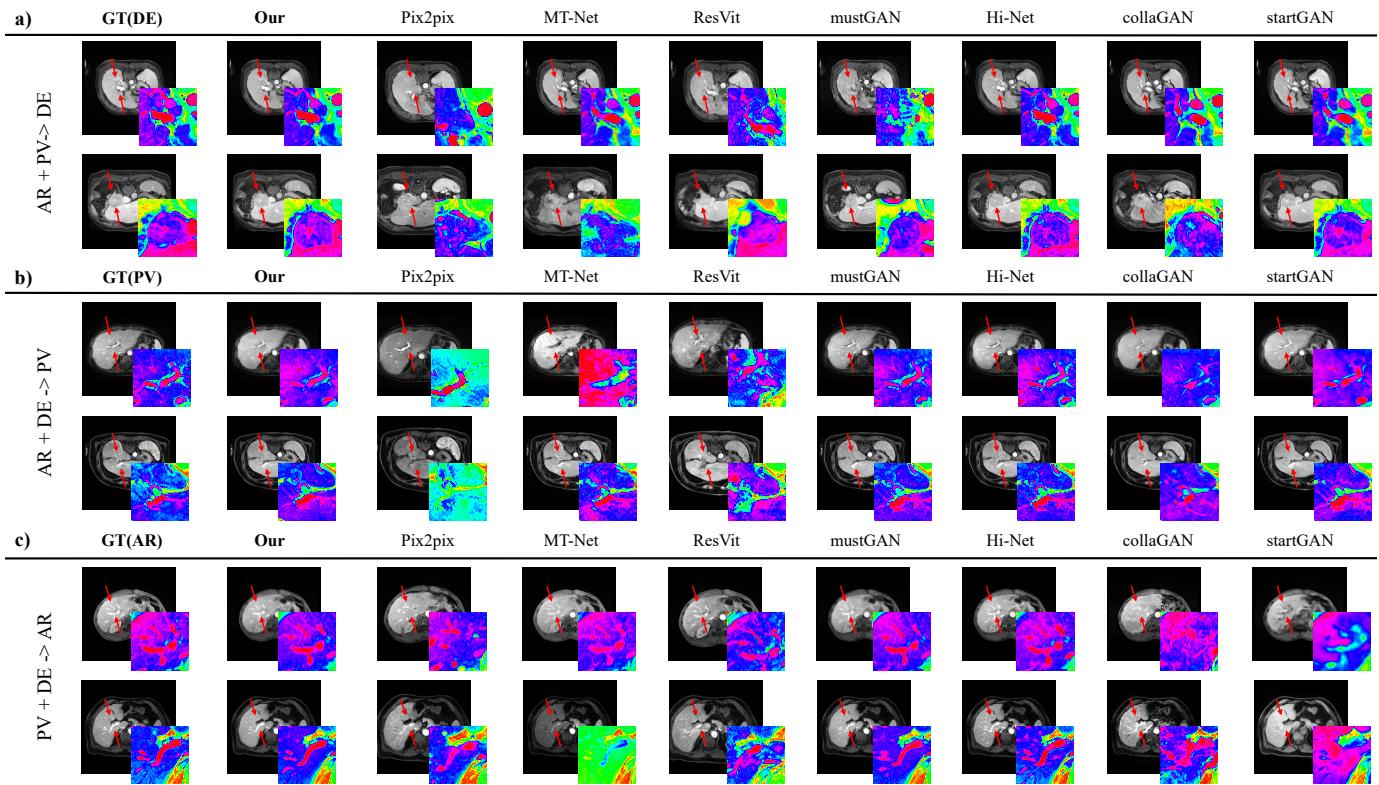


Fig. 7. Qualitative comparison between the proposed synthesis method and other methods for three synthesis tasks on the contrast-enhanced liver MR images dataset: a) Delay phase synthesis from Arterial and Portal Venous phases, b) Portal Venous phase synthesis from Arterial and Delay phases, c) Arterial phase synthesis from Portal Venous and Delay phases. The image in the bottom right corner depicts the heat-map details at key positions.

image in the second term of the equation. The objective function of discriminator, L_D (15), is formulated as:

$$L_D = -E_y [\log D(y)] \\ - E_{p_1, p_2} [\log (1 - D(G(p_1, p_2)))] \quad (15)$$

Consequently, an end-to-end multi-phase interconversion framework can be structured with the subsequent objective function (16):

$$L = L_G + L_D + L_{\text{decomp}} \quad (16)$$

IV. EXPERIMENTAL

A. Datasets

CUDD-DM is trained and tested on a generalized dataset and a public dataset LLD-MMRI [46].

The generalized dataset consists of 10,840 CE liver MR pairwise sequences including hemangiomas, hepatocellular carcinoma, and normal controls. The patient scans contain three phases of co-registered MRI: Arterial, PV, and Delay. Contrast enhanced liver MR images were obtained using 1.5 T MR images scanners (two types: Signa Artist, GE, and Aera, Siemens) using T1-weighted imaging. The CE liver MR images (T1-weighted fat-saturation imaging) were performed in the same orientations and with the same thickness, matrix, and pixel dimensions using a fat saturation sequence after intravenous injection of a gadolinium-based CAs (0.1 mmol/kg). This retrospective study was approved by our institutional

review board (Approval No. F11HRR - 43699) in accordance with local ethics procedures. Further consent was waived with approval.

The LLD-MMRI dataset comprises MRI data from 498 subjects, with each subject undergoing imaging across 8 different phases. Within this dataset, there are a total of 7 distinct lesion types, including 4 benign types (Hepatic hemangioma, Hepatic abscess, Hepatic cysts, and Focal nodular hyperplasia) and 3 malignant types (Intrahepatic cholangiocarcinoma, Liver metastases, and Hepatocellular carcinoma).

B. Implementation

CUDD-DM randomly selected about 4/5 for training and the remaining 1/5 for independent testing. CE three phases liver MR images underwent nonrigid registration as a preprocessing step to prevent potential misalignment caused by slight slice errors. All code are based on the PyTorch, and all high-parameters are tuned to obtain the best results. The ADAM optimizer [47] is employed with a starting learning rate of 1e-4 (0.95 decay, tuned from [1e-3, 1e-4, 1e-5]). The model is trained 40 epochs, a decision derived from evaluations with epochs in the range [20, 40, 60, 80, 100]. The original learning rate is set to 0.0001 for the first 20 epochs and then linearly decays to 0 over the remaining epochs. It required 63 hours for training, and 0.42 sec on average for a test image whose size is 256 × 256 on 4 × Tesla p100 GPUs.

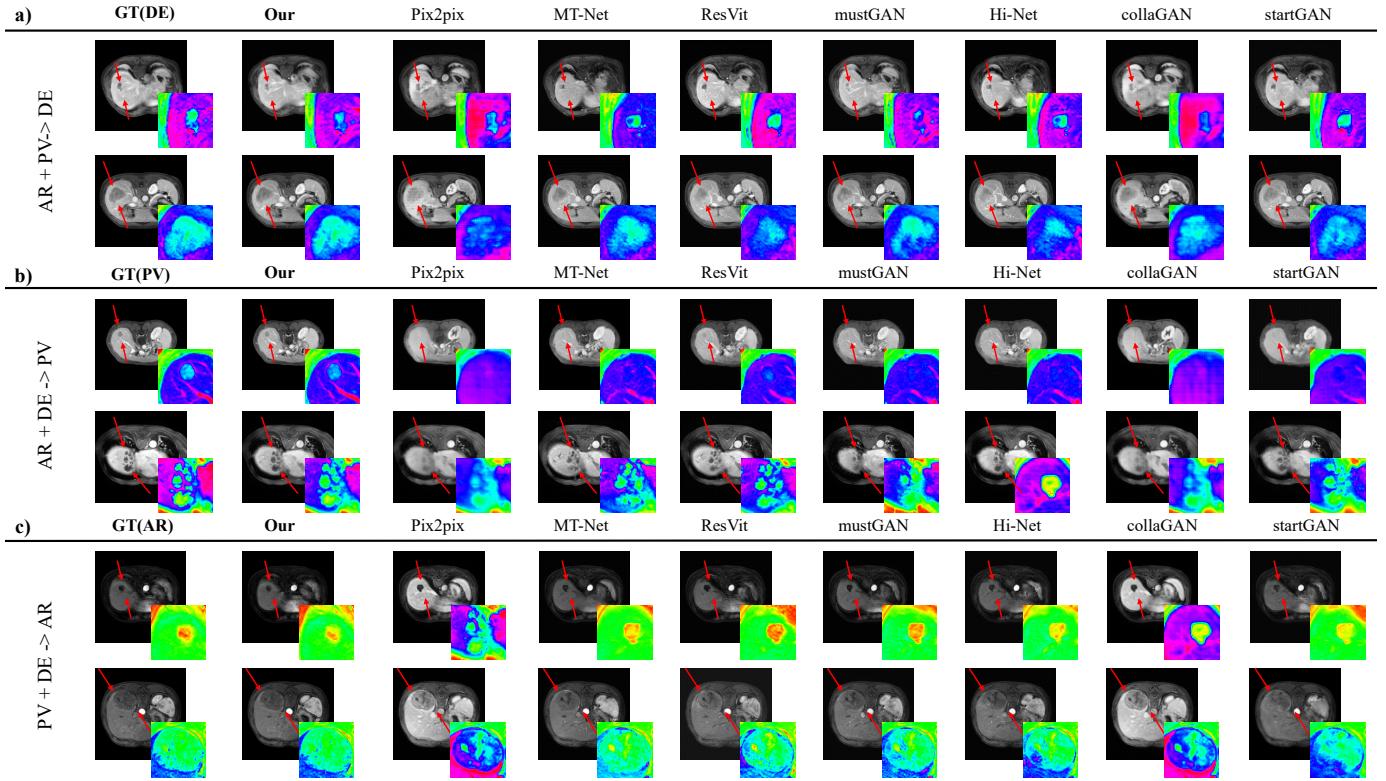


Fig. 8. Qualitative comparison between the proposed synthesis method and other methods for three synthesis tasks on the LLD-MMRI dataset: a) Delay phase synthesis from Arterial and Portal Venous phases, b) Portal Venous phase synthesis from Arterial and Delay phases, c) Arterial phase synthesis from Portal Venous and Delay phases. The image in the bottom right corner depicts the heat-map details at key positions.

C. Experimental setup and evaluation metrics

To comprehensively demonstrate the superiority of CUDD-DM, first we conducted experiments on three tasks: 1) using Arterial and PV phases as inputs to synthesize the Delay phase (i.e., AR + PV \rightarrow DE). 2) using Arterial and Delay phases as inputs to synthesize the PV phase (i.e., AR + DE \rightarrow PV). 3) using PV and Delay phases as inputs to synthesize the Arterial phase (i.e., PV + DE \rightarrow AR). Experimental results on three tasks demonstrate that CUDD-DM achieves state-of-the-art performance in CE liver MR images multi-phase interconversion. We also extended our evaluation to LLD-MMRI dataset, where we performed similar interconversion on AR, PV, and DE phases, as well as T1WI, T2WI, and DWI phases. Secondly, CUDD-DM is compared to seven state-of-the-art methods. They are divided into two types: 1) Four recent approaches for multi-phase image interconversion, compared in this study, are detailed in [28], [43], [45] and [10]. All methods were published in the journal Transactions on medical imaging and Medical image analysis represent the state-of-the-art in this task. 2) Three existing image-to-image synthesis frameworks (pix2pix [41], collaGAN [42], startGAN [44]) in the computer vision community. Finally, we conducted ablation experiments to evaluate the performance of each main novel component of CUDD-DM, including Common-Unique Feature Decomposition Module, Multi-scale Temporal Reset Gates Module, and Diffusion Model-Driven Lesion Detail Synthesis Module. This allowed us to demonstrate the individual superiority of each component. We used four

well-known evaluation metrics in our experiments, including structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), learned perceptual image patch similarity (LPIPS), and normalized mean squared error (NMSE). These metrics allowed us to comprehensively evaluate our method from multiple perspectives.

D. Experimental Results

Comprehensive experiments indicated that CUDD-DM produces high-quality CE liver tumor images. Firstly, CUDD-DM achieved average a SSIM of 0.882, a PSNR of 27.45, an LPIPS of 0.0891 and an NMSE 0.0515 on three multi-phase interconversion tasks for the generalized dataset. Secondly, CUDD-DM achieved average a SSIM of 0.858, a PSNR of 26.76, an LPIPS of 0.0678, and an NMSE 0.0937 on LLD-MMRI dataset. Finally, CUDD-DM achieved average a SSIM of 0.830, a PSNR of 24.89, an LPIPS of 0.116, and an NMSE 0.108 on LLD-MMRI dataset for DWI, T1WI and T2WI three modalities interconversion. These results demonstrate that CUDD-DM has great potential to serve as an accurate clinical alternative for liver tumor diagnosis, reducing patient waiting time and conserving the utilization of clinical resources.

1) The high image quality of the multi-phase CE liver tumor images: Overall excellence in CE liver MR images multi-phase interconversion. Figure 7, 8 demonstrate, compared to the ground truth, CUDD-DM demonstrates high similarity in the multi-phase interconversion of **overall** CE liver tumor images. Furthermore, Table I and II reveals that CUDD-DM

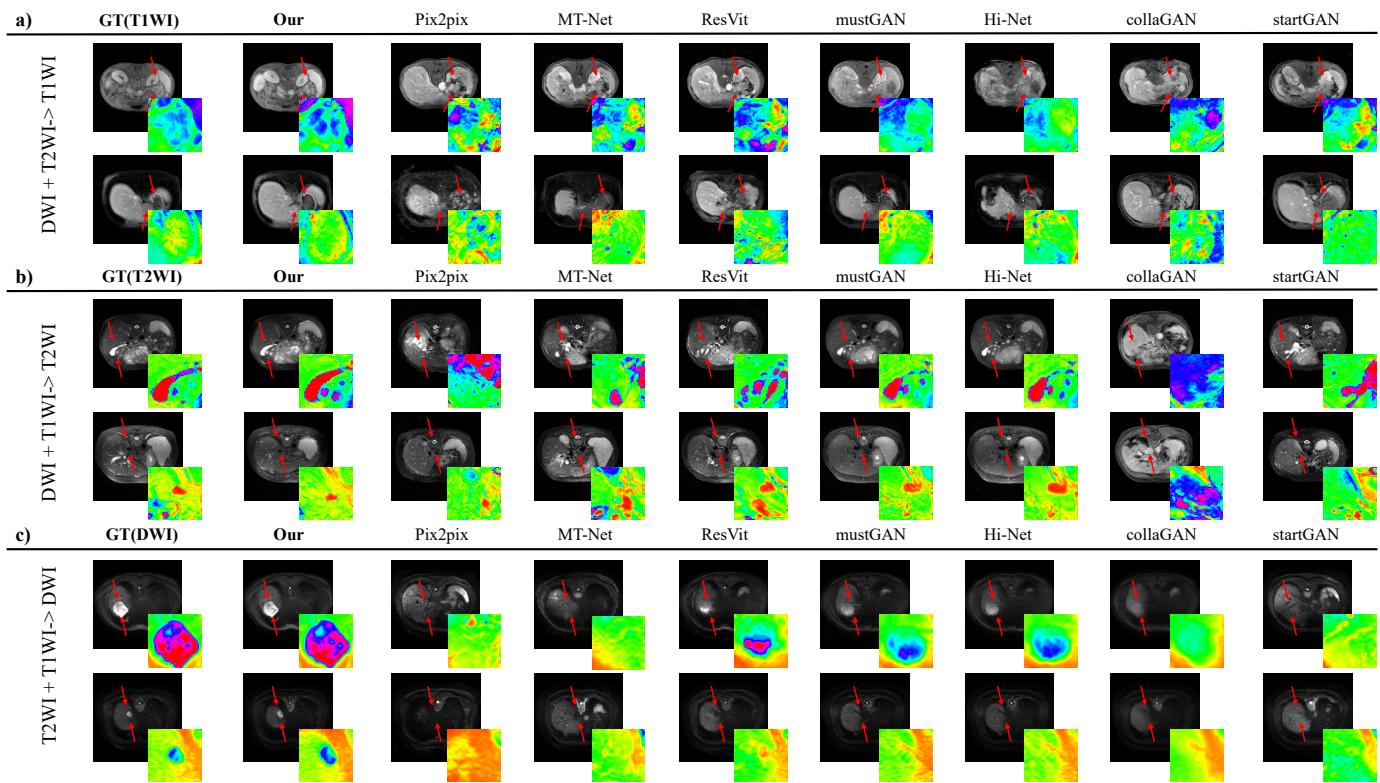


Fig. 9. Qualitative comparison between the proposed synthesis method and other methods for DWI, T2WI, T1WI three modalities interconversion tasks on the LLD-MMRI dataset: a) T1WI image synthesis from DWI and T2WI images, b) T2WI image synthesis from DWI and T1WI images, c) DWI image synthesis from T2WI and T1WI images. The image in the bottom right corner depicts the heat-map details at key positions.

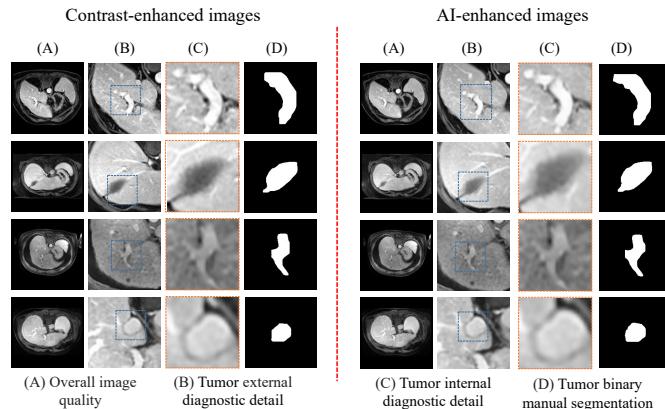


Fig. 10. Compared with CE liver MR images, CUDD-DM produces high quality AI-enhanced liver tumor images in the overall image, external diagnostic details of the tumor, internal diagnostic details of the tumor, and manual tumor segmentation.

achieved impressive scores across three multi-phase interconversion tasks for CE liver tumor images. For AR + PV → DE, the CUDD-DM achieves average a SSIM of 0.876, a PSNR of 27.42, an LPIPS of 0.0821 and an NMSE of 0.0641. For AR + DE → PV, the CUDD-DM achieves average a SSIM of 0.888, a PSNR of 27.94, an LPIPS of 0.0663 and an NMSE of 0.0678. For PV + DE → AR, the CUDD-DM average a SSIM of 0.845, a PSNR of 25.97, an LPIPS of 0.0870 and an NMSE of 0.0862. These results not only validate the efficacy of CUDD-DM in accurate image interconversion for each specific task but also

highlight its potential in clinical applications for liver tumor diagnosis. **Detail excellence in CE liver MR images multi-phase interconversion.** For the **detailed** region containing the liver tumor, the heatmap in the lower right corner of Figure 7, 8 exhibits high similarity using CUDD-DM compared to the ground truth. We further performed a quantitative comparison of regions containing lesions on three tasks, where the CE liver MR of AR + PV → DE, AR + DE → PV and PV + DE → AR are 407, 422 and 265 paired sequences respectively, showing excellent results. Table III (a) demonstrated that the CUDD-DM achieves weighted average a SSIM of 0.767, a PSNR of 23.98, a LPIPS of 0.01 and a NMSE of 0.157 on three tasks detail synthesis of liver tumor.

Accurate manual tumor segmentation in the AI-enhanced liver tumor images. Figure 10 and Table III (b) show the excellent ability of CUDD-DM in manually segmenting liver tumors compared with tumors in real CE liver tumor images. After manual segmentation of two tumor types(i.e., CE images and AI-enhanced images), we generated tumor binary images by specifying tumor areas with a pixel value of 0 and non-tumor areas with a pixel value of 255. Furthermore, using CE liver tumor images as the reference standard, Table III (b) demonstrate that CUDD-DM achieves weighted average a Dice of 0.824, an IoU of 0.738, a Precision of 0.837, and a Recall of 0.834 for the further segmentation of tumor details on three tasks.

2) Outperformance of our CUDD-DM than all comparative methods: AR, PV and DE three phases interconversion. Figure 7, 8 and table I, II demonstrated that the evaluation

TABLE I

OUR CUDD-DM ATTAINS A NEW STATE-OF-THE-ART PERFORMANCE IN THE FIELD OF CONTRAST-ENHANCED LIVER MRI IMAGE INTERCONVERSION THAN ALL SEVEN COMPARATIVE METHODS. NOTE THAT * AND ** REPRESENT THE P-VALUE BETWEEN EXISTING METHODS AND CUDD-DM IS LESS THAN 0.05 AND 0.01, INDICATING THE IMPROVEMENT BY CUDD-DM IS STATISTICALLY SIGNIFICANT.

Method(AR + PV → DE)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.761**	20.99**	0.187**	0.469**
Pix2pix [41]	0.845**	23.83**	0.111**	0.336**
MT-Net [43]	0.796**	21.42**	0.178**	0.168**
startGAN [44]	0.826**	23.24**	0.137**	0.669**
mustGAN [45]	0.826**	25.26**	0.123**	0.134**
ResViT [10]	0.819**	23.57**	0.129**	0.108**
Hi-Net [28]	0.866**	26.2**	0.108*	0.0664**
Ours	0.891	27.80	0.0975	0.0507

(a)				
Method(AR + PV → DE)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.752**	21.52**	0.180**	0.376**
Pix2pix [41]	0.848**	23.21**	0.115**	0.338**
MT-Net [43]	0.823**	23.87**	0.172**	0.179**
startGAN [44]	0.839**	23.84**	0.143**	0.345**
mustGAN [45]	0.875**	26.23**	0.0939*	0.0571**
ResViT [10]	0.840**	24.59**	0.126**	0.119**
Hi-Net [28]	0.881**	26.87**	0.0935*	0.0571*
Ours	0.903	28.41	0.0753	0.0425

(b)				
Method(AR + DE → PV)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.752**	21.52**	0.180**	0.376**
Pix2pix [41]	0.848**	23.21**	0.115**	0.338**
MT-Net [43]	0.823**	23.87**	0.172**	0.179**
startGAN [44]	0.839**	23.84**	0.143**	0.345**
mustGAN [45]	0.875**	26.23**	0.0939*	0.0571**
ResViT [10]	0.840**	24.59**	0.126**	0.119**
Hi-Net [28]	0.881**	26.87**	0.0935*	0.0571*
Ours	0.903	28.41	0.0753	0.0425

(c)				
Method(PV + DE → AR)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.736**	18.44**	0.253**	0.493**
Pix2pix [41]	0.799**	22.38**	0.113**	0.872**
MT-Net [43]	0.775**	21.09**	0.199**	0.165**
startGAN [44]	0.807**	21.75**	0.120**	0.958**
mustGAN [45]	0.817**	23.93**	0.119**	0.143**
ResViT [10]	0.814**	23.34**	0.131**	0.177**
Hi-Net [28]	0.828**	24.09**	0.112**	0.0778*
Ours	0.851	26.14	0.0945	0.0613

TABLE II

OUR CUDD-DM ATTAINS A NEW STATE-OF-THE-ART PERFORMANCE IN THE LLD-MMRI DATASET THAN ALL SEVEN COMPARATIVE METHODS. NOTE THAT * AND ** REPRESENT THE P-VALUE BETWEEN EXISTING METHODS AND CUDD-DM IS LESS THAN 0.05 AND 0.01, INDICATING THE IMPROVEMENT BY CUDD-DM IS STATISTICALLY SIGNIFICANT.

Method(AR + PV → DE)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.755**	23.05**	0.153**	0.228**
Pix2pix [41]	0.774**	21.90**	0.204**	0.907**
MT-Net [43]	0.804**	23.54**	0.0937**	0.157**
startGAN [44]	0.827**	25.07**	0.0823**	0.130**
mustGAN [45]	0.834**	25.22**	0.0767*	0.107**
ResViT [10]	0.823**	24.59**	0.0838**	0.178**
Hi-Net [28]	0.829**	25.06**	0.0785**	0.112**
Ours	0.863	27.03	0.0667	0.0775

(a)				
Method(AR + DE → PV)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.768**	22.57**	0.168**	0.392**
Pix2pix [41]	0.724**	20.81**	0.253**	0.737**
MT-Net [43]	0.814**	23.84**	0.0932**	0.145**
startGAN [44]	0.844**	25.71**	0.08**	0.108**
mustGAN [45]	0.844**	25.70**	0.0747**	0.0969*
ResViT [10]	0.833**	25.06**	0.0783**	0.136**
Hi-Net [28]	0.841**	25.58**	0.0753**	0.0961*
Ours	0.872	27.47	0.0573	0.0931

(b)				
Method(PV + DE → AR)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.769**	23.07**	0.114**	0.158**
Pix2pix [41]	0.742**	20.93**	0.204**	0.915**
MT-Net [43]	0.778**	22.72**	0.113**	0.190**
startGAN [44]	0.807**	24.73**	0.0983**	0.183**
mustGAN [45]	0.809**	24.85*	0.0938*	0.172**
ResViT [10]	0.792**	24.15**	0.103**	0.207**
Hi-Net [28]	0.803**	24.61**	0.0959**	0.181**
Ours	0.839	25.79	0.0795	0.111

(c)				
Tumor Detail(GT and CUDD-DM)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
AR + PV → DE (407 cases)	0.783	24.35	0.071	0.149
AR + DE → PV (422 cases)	0.783	24.42	0.075	0.14
PV + DE → AR (265 cases)	0.734	22.16	0.176	0.187
weighted average	0.767	23.98	0.1	0.157

Tumor Detail(GT and Hi-Net)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
AR + PV → DE (407 cases)	0.724	23.22	0.197	0.254
AR + DE → PV (422 cases)	0.73	23.39	0.105	0.168
PV + DE → AR (265 cases)	0.605	20.47	0.242	0.293
weighted average	0.687	22.62	0.172	0.23

TABLE III

CUDD-DM ACHIEVED NOT ONLY HIGH IMAGE QUALITY BUT ALSO ACCURATE DIAGNOSTIC DETAIL IN THE AI-ENHANCED MR IMAGES.

High diagnostic detail of the tumor				
Tumor Detail(GT and CUDD-DM)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
AR + PV → DE (407 cases)	0.783	24.35	0.071	0.149
AR + DE → PV (422 cases)	0.783	24.42	0.075	0.14
PV + DE → AR (265 cases)	0.734	22.16	0.176	0.187
weighted average	0.767	23.98	0.1	0.157

Accurate manual tumor segmentation				
Tumor Segmentation(GT and CUDD-DM)	Dice↑	IoU↑	Precision↑	Recall↑
AR + PV → DE (407 cases)	82.3%	73.0%	83%	82.5%
AR + DE → PV (422 cases)	83.35%	74.2%	83.7%	83.7%
PV + DE → AR (265 cases)	80.7%	72.0%	82.6%	80.7%
weighted average	82.4%	73.8%	83.7%	83.4%

(b)

TABLE IV

OUR CUDD-DM ATTAINS A NEW STATE-OF-THE-ART PERFORMANCE IN THE T1WI, T2WI AND DWI INTERCONVERSION THAN ALL SEVEN COMPARATIVE METHODS. NOTE THAT * AND ** REPRESENT THE P-VALUE BETWEEN EXISTING METHODS AND CUDD-DM IS LESS THAN 0.05 AND 0.01, INDICATING THE IMPROVEMENT BY CUDD-DM IS STATISTICALLY SIGNIFICANT.

Method(DWI + T2WI → T1WI)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.715**	18.52**	0.257**	0.928**
BicycleGAN [48]	0.722**	19.12**	0.253**	0.165**
MT-Net [43]	0.784**	20.54**	0.177**	0.165**
Pix2pix [41]	0.774**	22.01**	0.192**	0.505**
mustGAN [45]	0.788**	21.41**	0.196**	0.205**
ResViT [10]	0.806**	22.16**	0.185**	0.268**
Hi-Net [28]	0.796**	22.05**	0.156**	0.174**
Ours	0.827	24.03	0.138	0.110

(a)				
Method(DWI + T1WI → T2WI)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.738**	21.75**	0.238**	0.529**
Pix2pix [41]	0.782**	21.18**	0.185**	0.326**
MT-Net [43]	0.774**	21.24**	0.162**	0.185**
startGAN [44]	0.765**	20.71**	0.178**	0.297**
mustGAN [45]	0.778**	21.14**	0.193**	0.262**
ResViT [10]	0.807**	22.60**	0.132*	0.182**
Hi-Net [28]	0.794**	22.47**	0.158**	0.173**
Ours	0.824	24.54	0.122	0.102

(b)				
Method(T1WI + T2WI → DWI)	SSIM↑	PSNR↑	LPIPS↓	NMSE↓
collaGAN [42]	0.758**	20.72**	0.214**	0.667**
Pix2pix [41]	0.792**	21.85**	0.126**	0.208**
MT-Net [43]	0.776**	22.83**	0.111**	0.193**
startGAN [44]	0.806**	24.83**	0.103**	0.133**
mustGAN [45]	0.805**	22.85**	0.138**	0.192**
ResViT [10]	0.821**	23.57**	0.124**	0.175**
Hi-Net [28]	0.824**	24.91**	0.101**	0.146**
Ours	0.840	26.11	0.0883	0.112

(c)				
-----	--	--	--	--

64.39%-0.3%. In Table II (c), PV + DE → AR, comparing seven comparative methods by various synthesis metrics, our CUDD-DM improved the SSIM 9.7%-3.0%, the PSNR 4.86dB-0.94dB, the LPIPS 12.45%-1.47%, the NMSE 80.4%-3.0%. These results demonstrate the efficacy of our proposed Common-Unique Feature Decomposition Module, Multi-scale Temporal Reset Gates Module, and Diffusion Model-Driven Lesion Detail Synthesis Module during the CE overall liver tumor images interconversion.

DWI, T2WI, T1WI three modalities interconversion.

Figure 9 and table IV demonstrated that the evaluation of visualization and all four synthesis metrics demonstrate that our CUDD-DM outperforms all seven compared methods on the three synthesis tasks of multi-modal interconversion. CUDD-DM achieves the highest synthesis performance. In table IV (a), DWI + T2WI → T1WI, comparing seven comparative methods by various synthesis metrics, our CUDD-DM improved the SSIM 11.2%-2.0%, the PSNR 5.51dB-1.87dB, the LPIPS 11.9%-1.8%, the NMSE 81.8%-5.4%. In Table IV (b), DWI + T1WI → T2WI, comparing seven comparative methods by various synthesis metrics, our CUDD-DM improved the SSIM 8.6%-1.7%, the PSNR 3.83dB-1.94dB, the LPIPS 11.6%-1.0%, the NMSE 42.7%-7.1%. In Table IV (c), T1WI + T2WI → DWI, comparing seven comparative methods by various synthesis metrics, our CUDD-DM improved the SSIM

8.2%-1.6%, the PSNR 5.39dB-1.2dB, the LPIPS 12.57%-1.27%, the NMSE 55.5%-2.1%. These results underscore the robust generalization capabilities of our CUDD-DM.

3) Superiority of the Common-Unique Feature Decomposition Module: The quantitative results of CUDD-DM and no-CUFD in Figure 11, 12, and 13 (a) indicate that the use of Common-Unique Feature Decomposition Module can better comprehensively represent the excellent performance of spectral-based CUFD in the multi-phase interconversion task. The no-CUFD means that CUFD has been removed from the original CUDD-DM, while other components of CUDD-DM remain unchanged. CUDD-DM achieves the best performance in SSIM, PSNR, LPIPS, NMSE. SSIM has been improved 2.8%-2.3%, PSNR has been improved 2.79dB-1.17dB, LPIPS has been improved 2.39%-1.3%, and NMSE has been improved 1.43%-1.35%. These improvements can be attributed to the integration of CUFD, which creatively employs spectral technology in the extraction process of common and unique features. The introduction of the CUFD enables the CE liver MR images interconversion to learn the overall appearance of the liver tumor through common features, and to capture the details of the liver tumor through unique features.

4) Superiority of the Multi-scale Temporal Reset Gates Module: Figure 11, 12, and 13 (b) shows that the MTRG used in our CUDD-DM can effectively deal with the mutual interference problem caused by inconsistent dependencies between consecutive slices of CE liver MR images. We compared two ablation versions of MTRG on the multi-phase interconversion task: 1) no MTRG, the no MTRG means that MTRG has been removed from the original CUDD-DM, while other components of CUDD-DM remain unchanged; 2) Gate Recurrent Unit (GRU) [49] replaces MTRG, which means MTRG has been replaced with GRU from the original CUDD-DM, while other components of CUDD-DM remain unchanged. Figure 11, 12, and 13 (b) also indicate that CUDD-DM achieves the best performance in SSIM, PSNR, LPIPS, NMSE. SSIM has been improved 3.2%-1.3%, PSNR has been improved 1.98dB-0.8dB, LPIPS has been improved 2.4%-0.96%, and NMSE has been improved 1.78%-0.95%. These improvements occur because MTRG enables our CUDD-DM model to effectively handle inconsistent dependencies. For consecutive slices where no lesions have occurred, MTRG will strengthen the front slice more. Otherwise, the influence of adjacent frames will be minimized.

5) Superiority of the Diffusion Model-Driven Lesion Detail Synthesis Module: Figure 11, 12, and 13 (c) shows that the DM-DLDS enables our CUDD-DM to better synthesize the details of liver tumor images through the forward and reverse processes of the diffusion model. We compared three ablation versions of DM-DLDS on the multi-phase interconversion task: 1) no DM; 2) Auto-encoder [50] replaces DM-DLDS; 3) pix2pix [41] replaces DM-DLDS. Figure 11, 12, and 13 (c) also indicate that CUDD-DM achieves the best performance in SSIM, PSNR, LPIPS, NMSE. SSIM has been improved 1.8%-0.7%, PSNR has been improved 2.05dB-0.57dB, LPIPS has been improved 1.79%-0.68%, and NMSE has been improved 1.96%-0.78%. The improvement is because DM-DLDS enables our CUDD-DM to better synthesize the details of liver

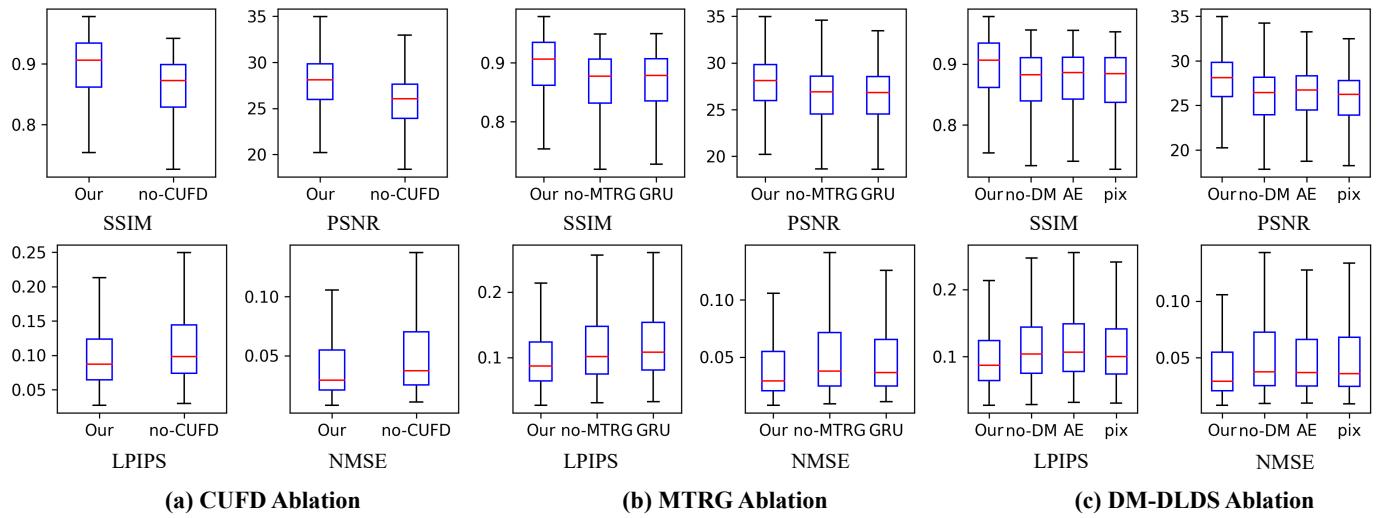


Fig. 11. (a), (b) and (c) respectively represent the ablation of the three innovative modules proposed by our CUDD-DM on AR + PV → DE task. The effectiveness of these components (i.e. CUFD, MTRG, and DM-DLDS) in CE liver MR images interconversion is demonstrated by key metrics such as SSIM, PSNR, LPIPS, and NMSE.

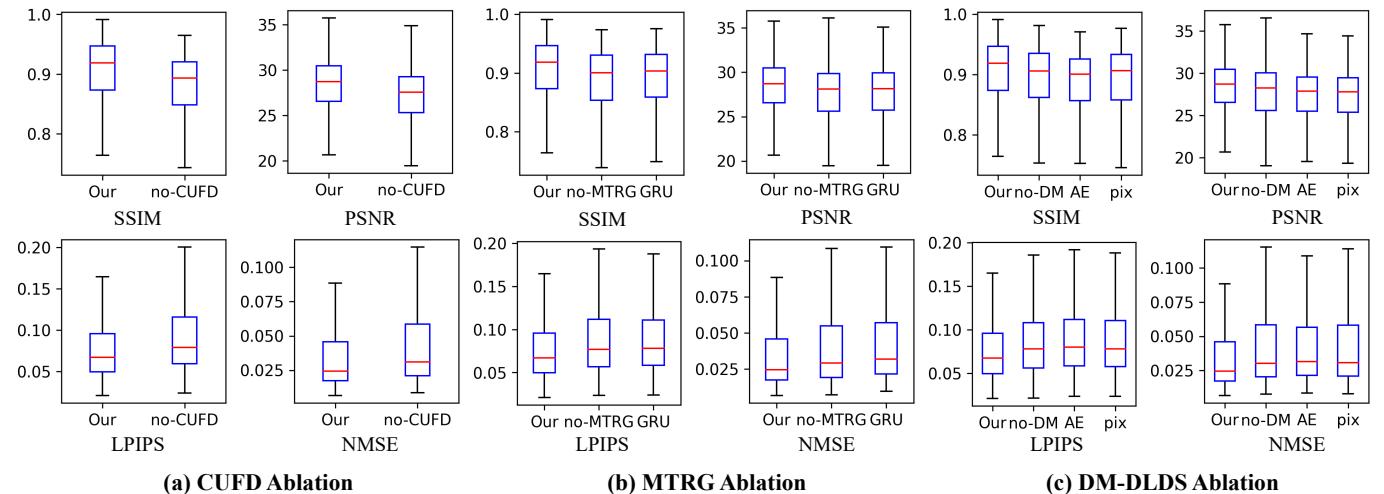


Fig. 12. (a), (b) and (c) respectively represent the ablation of the three innovative modules proposed by our CUDD-DM on AR + DE → PV task. The effectiveness of these components (i.e. CUFD, MTRG, and DM-DLDS) in CE liver MR images interconversion is demonstrated by key metrics such as SSIM, PSNR, LPIPS, and NMSE.

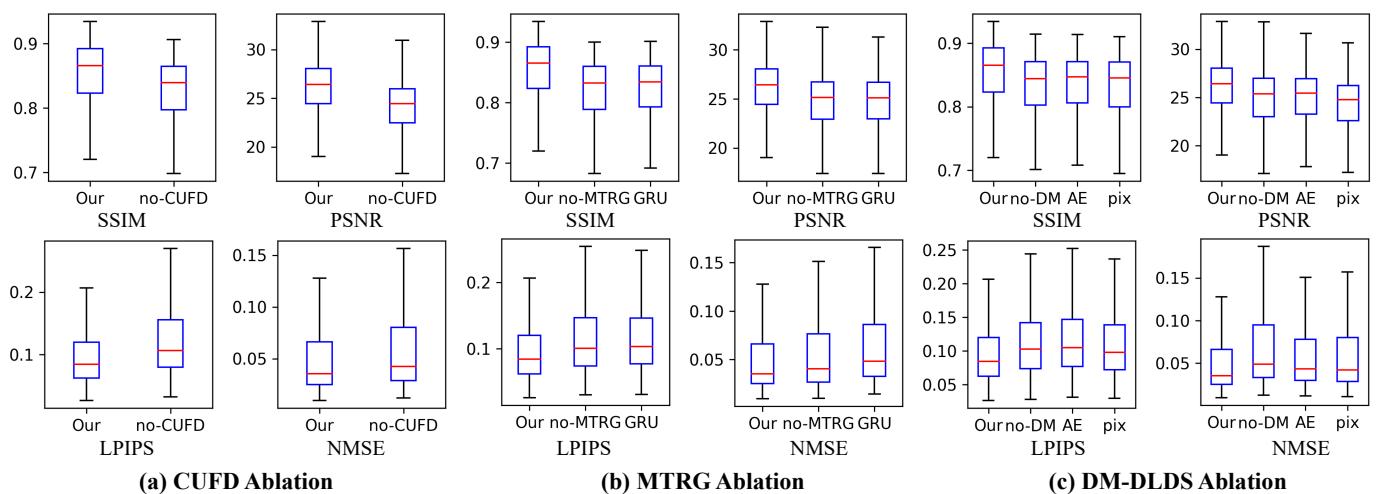


Fig. 13. (a), (b) and (c) respectively represent the ablation of the three innovative modules proposed by our CUDD-DM on PV + DE → AR task. The effectiveness of these components (i.e. CUFD, MTRG, and DM-DLDS) in CE liver MR images interconversion is demonstrated by key metrics such as SSIM, PSNR, LPIPS, and NMSE.

tumor images to meet the needs of accurate diagnosis.

V. DISCUSSION

While our CUDD-DM framework has demonstrated significant advancements in contrast-enhanced liver tumor multi-phases image interconversion, we acknowledge certain limitations and areas for future exploration. Firstly, poor image quality and significant patient movement affect the performance of our model. However, our model mitigates this risk through MTRG and DM-DLDS two innovative modules. The MTRG Module compensating for poor image quality in any current slice by leveraging information from adjacent ones. The DM-DLDS adapts to the impact of poor image quality by forward noise addition and backward denoising. Secondly, compared with other baselines, introducing spectral decomposition increases the overall computational load during training and inference. This increased computational requirement may limit the feasibility of our proposed algorithm in application scenarios with severe computational resource constraints. Future research could explore ways to optimize the computational efficiency of our framework.

VI. CONCLUSION

In this paper, we propose a novel Common-Unique Decomposition Driven Diffusion Model (CUDD-DM) for CE liver MR images multi-phase interconversion. By enabling the interconversion of any two input phases in three into the remaining one, this model holds the potential to reduce patient wait times, conserve medical resources, and decrease the use of CAs. Our CUDD-DM introduces a powerful Common-Unique Feature Decomposition Module, adept at extracting both common and unique features from inputs phases. Through spectral decomposition, it precisely differentiates over 95% of similar regions and less than 5% of lesion-specific unique features, significantly enhancing the synthesis precision of lesions. Additionally, our CUDD-DM incorporates a novel Multi-scale Temporal Reset Gates Module. This module focuses on dependencies on regions without lesions in consecutive slices, considerably reducing reliance on slices with lesions and effectively preventing mutual interference between slices. Lastly, our CUDD-DM utilizes a Diffusion Model-Driven Lesion Detail Synthesis Module. Executing a continuous and progressive generation process, this module accurately captures detailed features within the data distribution, ensuring that the generated liver images meet the stringent diagnostic standards required for clinical applications. Our CUDD-DM achieved a new state-of-the-art performance on multi-phase interconversion of three tasks, with a SSIM of 0.877, a PSNR of 27.32, an LPIPS of 0.0927, and an NMSE 0.0526. Such results demonstrate that the proposed model is a secure and potential auxiliary tool in liver tumor imaging. This method not only enables a reduction in patient waiting times but also has the potential to lessen the harm caused to patients' bodies by the use of CAs.

REFERENCES

- [1] J. Zhao, D. Li, Z. Kassam, J. Howey, J. Chong, B. Chen, and S. Li, "Tripartite-gan: Synthesizing liver contrast-enhanced mri to improve tumor detection," *Medical image analysis*, vol. 63, p. 101667, 2020.
- [2] S. Li, "Contrast-agent-free medical diagnostic imaging," Jun. 30 2022, uS Patent App. 17/138,353.
- [3] C. Xu, D. Zhang, Y. Song, L. Kayat Bittencourt, S. H. Tirumani, and S. Li, "Contrast-free liver tumor detection using ternary knowledge transferred teacher-student deep reinforcement learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 266–275.
- [4] C. Xu, D. Zhang, J. Chong, B. Chen, and S. Li, "Synthesis of gadolinium-enhanced liver tumors on nonenhanced liver mr images using pixel-level graph reinforcement learning," *Medical image analysis*, vol. 69, p. 101976, 2021.
- [5] C. Xu, Y. Song, D. Zhang, L. K. Bittencourt, S. H. Tirumani, and S. Li, "Spatiotemporal knowledge teacher-student reinforcement learning to detect liver tumors without contrast agents," *Medical Image Analysis*, vol. 90, p. 102980, 2023.
- [6] C. A. Hamm, C. J. Wang, L. J. Savic, M. Ferrante, I. Schobert, T. Schlachter, M. Lin, J. S. Duncan, J. C. Weinreb, J. Chapiro *et al.*, "Deep learning for liver tumor diagnosis part i: development of a convolutional neural network classifier for multi-phasic mri," *European radiology*, vol. 29, pp. 3338–3347, 2019.
- [7] A. H. Behzadi, Y. Zhao, Z. Farooq, and M. R. Prince, "Immediate allergic reactions to gadolinium-based contrast agents: a systematic review and meta-analysis," *Radiology*, vol. 286, no. 2, pp. 471–482, 2018.
- [8] Y. Huang, L. Shao, and A. F. Frangi, "Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 815–827, 2017.
- [9] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [10] O. Dalmaz, M. Yurt, and T. Çukur, "Resvit: Residual vision transformers for multimodal medical image synthesis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2598–2614, 2022.
- [11] R. Osuala, G. Skorupko, N. Lazrak, L. Garrucho, E. García, S. Joshi, S. Jouide, M. Rutherford, F. Prior, K. Kushibar *et al.*, "medigan: a python library of pretrained generative models for medical image synthesis," *Journal of Medical Imaging*, vol. 10, no. 6, pp. 061403–061403, 2023.
- [12] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III* 20. Springer, 2017, pp. 417–425.
- [13] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2720–2730, 2018.
- [14] C. Xu, L. Xu, G. Brahm, H. Zhang, and S. Li, "Mutgan: simultaneous segmentation and quantification of myocardial infarction without contrast agents via joint adversarial learning," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II* 11. Springer, 2018, pp. 525–534.
- [15] J. Liu, Y. Li, N. Shi, Y. Zhou, S. Gao, Y. Shi, X.-Y. Zhang, and X. Zhuang, "Multi-phase liver-specific dce-mri translation via a registration-guided gan," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2023, pp. 21–31.
- [16] S. Reuangamornrat, H. Sari, C. Catana, and A. Kamen, "Multimodal image synthesis based on disentanglement representations of anatomical and modality specific features, learned using uncooperative relativistic gan," *Medical image analysis*, vol. 80, p. 102514, 2022.
- [17] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic *et al.*, "Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks," *arXiv preprint arXiv:1702.05970*, 2017.
- [18] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, S. Ma, L. Jin, X. Liu, X. Li, and X. Qian, "Automatic segmentation of liver tumors from multiphase contrast-enhanced ct images based on fcns," *Artificial intelligence in medicine*, vol. 83, pp. 58–66, 2017.

- [19] W. Zhou, L. Zhang, K. Wang, S. Chen, G. Wang, Z. Liu, and C. Liang, "Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced mr images," *Journal of Magnetic Resonance Imaging*, vol. 45, no. 5, pp. 1476–1484, 2017.
- [20] S.-h. Zhen, M. Cheng, Y.-b. Tao, Y.-f. Wang, S. Juengpanich, Z.-y. Jiang, Y.-k. Jiang, Y.-y. Yan, W. Lu, J.-m. Lue *et al.*, "Deep learning for accurate diagnosis of liver tumor based on magnetic resonance imaging and clinical data," *Frontiers in oncology*, vol. 10, p. 680, 2020.
- [21] S. Gul, M. S. Khan, A. Bibi, A. Khandakar, M. A. Ayari, and M. E. Chowdhury, "Deep learning techniques for liver and liver tumor segmentation: A review," *Computers in Biology and Medicine*, vol. 147, p. 105620, 2022.
- [22] K. L. C. Association *et al.*, "2022 klca-ncc korea practice guidelines for the management of hepatocellular carcinoma," *Korean Journal of Radiology*, vol. 23, no. 12, p. 1126, 2022.
- [23] L. Sun, J. Wang, Y. Huang, X. Ding, H. Greenspan, and J. Paisley, "An adversarial learning approach to medical image synthesis for lesion detection," *IEEE journal of biomedical and health informatics*, vol. 24, no. 8, pp. 2303–2314, 2020.
- [24] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "Ea-gans: edge-aware generative adversarial networks for cross-modality mr image synthesis," *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1750–1762, 2019.
- [25] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.
- [26] Y. Chen, X.-H. Yang, Z. Wei, A. A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, and Q. Guan, "Generative adversarial networks in medical image augmentation: A review," *Computers in Biology and Medicine*, vol. 144, p. 105382, 2022.
- [27] X. Yu, J. Wang, Q.-Q. Hong, R. Teku, S.-H. Wang, and Y.-D. Zhang, "Transfer learning for medical images analyses: A survey," *Neurocomputing*, vol. 489, pp. 230–254, 2022.
- [28] T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao, "Hi-net: hybrid-fusion network for multi-modal mr image synthesis," *IEEE transactions on medical imaging*, vol. 39, no. 9, pp. 2772–2781, 2020.
- [29] F.-A. Croitoru, V. Hondu, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [30] Z. Kong, W. Ping, J. Huang, K. Zhao, and B. Catanzaro, "Diffwave: A versatile diffusion model for audio synthesis," *arXiv preprint arXiv:2009.09761*, 2020.
- [31] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," *arXiv preprint arXiv:2307.01952*, 2023.
- [32] H. A. Bedel and T. Çukur, "Dreamr: Diffusion-driven counterfactual explanation for functional mri," *arXiv preprint arXiv:2307.09547*, 2023.
- [33] W. H. Pinaya, P.-D. Tudosi, J. Dafflon, P. F. Da Costa, V. Fernandez, P. Nachev, S. Ourselin, and M. J. Cardoso, "Brain imaging generation with latent diffusion models," in *MICCAI Workshop on Deep Generative Models*. Springer, 2022, pp. 117–126.
- [34] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.
- [35] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [36] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "Repaint: Inpainting using denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11461–11471.
- [37] M. Özbuy, O. Dalmaç, S. U. Dar, H. A. Bedel, Ş. Öztürk, A. Güngör, and T. Çukur, "Unsupervised medical image translation with adversarial diffusion models," *IEEE Transactions on Medical Imaging*, 2023.
- [38] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof, "Diffusion models for medical image analysis: A comprehensive survey," *arXiv preprint arXiv:2211.07804*, 2022.
- [39] H. Ali, S. Murad, and Z. Shah, "Spot the fake lungs: Generating synthetic medical images using neural diffusion models," in *Irish Conference on Artificial Intelligence and Cognitive Science*. Springer, 2022, pp. 32–39.
- [40] S. Pan, T. Wang, R. L. Qiu, M. Axente, C.-W. Chang, J. Peng, A. B. Patel, J. Shelton, S. A. Patel, J. Roper *et al.*, "2d medical image synthesis using transformer-based denoising diffusion probabilistic model," *Physics in Medicine & Biology*, vol. 68, no. 10, p. 105004, 2023.
- [41] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
- [42] D. Lee, J. Kim, W.-J. Moon, and J. C. Ye, "Collagan: Collaborative gan for missing image data imputation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2487–2496.
- [43] Y. Li, T. Zhou, K. He, Y. Zhou, and D. Shen, "Multi-scale transformer network with edge-aware pre-training for cross-modality mr image synthesis," *IEEE Transactions on Medical Imaging*, 2023.
- [44] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8188–8197.
- [45] M. Yurt, S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur, "mustgan: multi-stream generative adversarial networks for mr image synthesis," *Medical image analysis*, vol. 70, p. 101944, 2021.
- [46] M. Lou, H. Ying, X. Liu, H.-Y. Zhou, Y. Zhang, and Y. Yu, "Sdr-former: A siamese dual-resolution transformer for liver lesion classification using 3d multi-phase imaging," *arXiv preprint arXiv:2402.17246*, 2024.
- [47] X. Yang, Y. Lin, Z. Wang, X. Li, and K.-T. Cheng, "Bi-modality medical image synthesis using semi-supervised sequential generative adversarial networks," *IEEE journal of biomedical and health informatics*, vol. 24, no. 3, pp. 855–865, 2019.
- [48] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," *Advances in neural information processing systems*, vol. 30, 2017.
- [49] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [50] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.