Cyprus University of Technology

杭州電子科技大學

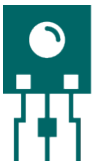
HANGZHOU DIANZI UNIVERSITY

**Master's thesis**

# Multimodal Medical Imaging with application in Histotripsy and Image Fusion

**Yuhan Lyu**

**Limassol, May 2025**



MSc in Electronics Science and Technology

CYPRUS UNIVERSITY OF TECHNOLOGY

Faculty of Engineering and Technology

Department of Electrical Engineering, Computer Engineering, and Informatics

Master's thesis

# Multimodal Medical Imaging with application in Histotripsy and Image Fusion

Yuhan Lyu

Supervisor

Christakis Damianou

Limassol, May 2025

**Approval Form**

Master's thesis

# Multimodal Medical Imaging with application in Histotripsy and Image Fusion

Presented by

Yuhan Lyu

**Supervisor**: Christakis Damianou

**Member of the committee**: Committee member 1

**Member of the committee**: Committee member 2

Cyprus University of Technology
Limassol, May 2025

# Copyrights



The approval of the dissertation by the Department of Electrical Engineering, Computer Engineering, and Informaticsdoes not necessarily imply the approval by the Department of the views of the writer.

## Acknowledgements

This work would not have been possible without the guidance of my supervisors, Dr. Christakis Damianou from the Faculty of Engineering and Technology at the Cyprus University of Technology, and Dr. Jufeng Zhao from Hangzhou Dianzi University. I would also like to extend my heartfelt gratitude to my senior colleagues, PhD candidates Antria Filippou and Nikolas Evripidou from the Cyprus University of Technology, for their invaluable support. Their mentorship and constructive criticism have significantly contributed to the progress of this research.

# ABSTRACT

The rapid development of medical imaging technology has led to the increasing application of various imaging modalities, providing physicians and researchers with diverse modal information. The utilization of multimodal medical imaging techniques for disease diagnosis and research analysis has also grown significantly. Such approaches can enhance the accuracy of disease diagnosis and research outcomes. This thesis addresses two key aspects of medical imaging research: conducting histotripsy experiments based on multimodal medical image analysis; in addition, we propose an innovative fusion method to effectively integrate images from different modalities into a single, information-rich composite image.

In the histotripsy experiments, MRI was used for the first time to accurately visualize tissue damage caused by tissue fragmentation. Through extensive experimentation, we established precise relationships between various treatment protocols and their respective effects, including the extent of tissue damage induced by different acoustic power levels and pulse counts, as well as their relationship with erosion diameter and length. These findings contribute to the advancement of histotripsy research and may aid in refining ultrasound parameters for future clinical translation of histotripsy applications.

Furthermore, we propose a novel multimodal medical image fusion method, which was compared to state-of-the-art fusion techniques. Our method demonstrated superior performance by effectively addressing the limitations of current fusion methods, such as inadequate feature extraction, poor inheritance of complementary information between modalities, and insufficient evaluation of color information in color and grayscale fusion tasks. This method provides a comprehensive and efficient solution for multimodal image fusion in medical imaging research. In the future, we plan to use fusion technique in histotripsy by combining CT and MRI images for detecting cavition in histotripsy lession.

**Keywords**: Medical image, Multimodal, histotripsy, image fusion

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| MRI | Magnetic resonance imaging |
| CT | Computed tomography |
| PET | Positron emission tomography |
| B-mode | Brightness-mode |
| US | ultrasound |
| FUS | focused ultrasound |
| FDA | Food and Drug Administration |
| PRF | pulse repetition frequency |
| PRP | pulse repetition perio |
| NSST | nonsubsampled shearlet transform |
| CNN | convolutional neural network |
| NSCT-SR | Non-Subsampled Contourlet Transform with sparse representation |
| GAN | generative adversarial networ |
| DDPM | denoising diffusion probabilistic model |
| EEDB | Edge Enhancement Dense Block |
| ECA | Edge Detail Coordinate Attention mechanism |
| FER | Feature-enhancement Reconstruction Network |
| ESA | Edge Spatial Attention mechanism |
| CIEDE2000 | color difference Formula |
| ACD | Average Color Difference |
| PLA | Polylactic acid |
| %w/v | weight/volume percentage concentrations |
| MSCA | Multi-scale Cross-axis Attention |
| GB | Gigabytes |
| EI | Edge Intensity |
| SF | Spatial Frequency |
| DF | Definition |
| AG | Average Gradient |
| CIE | International Commission on Illumination |
| SPECT | Single-Photon Emission Computed Tomography |
| ECT | Emission Computed Tomography |

# 1 Introduction

As medical imaging technology rapidly advances, different medical imaging modalities have been used for disease detection, clinical decision-making, and scientific research. For example, Magnetic Resonance Imaging (MRI) is a non-invasive diagnostic imaging technique that reconstructs high-resolution images by utilizing the signals generated from the resonance of atomic nuclei in a magnetic field. Its fundamental principle involves exciting the hydrogen nuclei in body tissues with radiofrequency pulses in a static magnetic field, causing them to undergo magnetic resonance. After the cessation of the radiofrequency pulse, these nuclei emit MR signals during their relaxation process. By receiving these signals, applying spatial encoding, and performing image reconstruction, high-quality MR images are produced. MRI offers the advantages of being non-invasive, free from ionizing radiation, and providing high resolution with detailed anatomical and structural information, although it may be less effective in detecting calcifications or metabolic changes.

Computed Tomography (CT) is another imaging modality that employs precisely controlled X-ray beams and highly sensitive detectors to perform layer-by-layer scans of the human body. The data acquired during these scans are processed by a computer to generate high-resolution cross-sectional, coronal, or sagittal images. CT images represent varying degrees of X-ray absorption by different tissues and organs in grayscale, offering high density resolution that allows for the clear visualization of both soft tissues and bone structures. However, CT may have limitations in detecting soft tissue tumors or small lesions in regions such as the cerebellum.

Positron Emission Tomography (PET) is a nuclear medicine imaging technique used to observe biological processes and functional activity within the body. Unlike conventional X-ray and CT scans, which primarily provide anatomical information, PET focuses on the metabolic and biochemical activities of tissues. The principle behind PET is based on the decay of radioactive isotopes. During a PET scan, a radiotracer containing a radioactive isotope is injected into the patient, targeting specific metabolic processes. As the radiotracer decays, it emits positrons, which, upon colliding with electrons, result in annihilation events that produce two photons traveling in opposite directions. These photons are captured by a PET scanner, which comprises one or more rings of detectors arranged around the patient. By measuring the arrival times and positions of these photons, the computer reconstructs the distribution of radioactivity within the body. By accumulating a large number of such positron annihilation events, the system generates a series of slice images that reflect the metabolic activity of tissues, such as the uptake of glucose.

An illustration of MRI, CT, and PET modalities is shown in Figure 1.1. These imaging modalities provide diverse sources of information, allowing researchers and clinicians to select the most appropriate imaging technique for disease diagnosis and medical research analysis based on the distinct characteristics of each medical image. Recently, the demand for the integrated utilization of multimodal medical images has been steadily increasing. For example, as shown in Figure 1.2, in the diagnosis of Alzheimer's disease, MRI is the preferred imaging modality, primarily focusing on hippocampal atrophy as a significant indicator. However, hippocampal atrophy can result from various causes. When combined with PET images, a significant reduction in blood flow signals and decreased metabolism in the hippocampal region can be

MRI CT PET

Figure 1.1: Examples of multimodal medical images. From left to right: MRI, CT and PET.

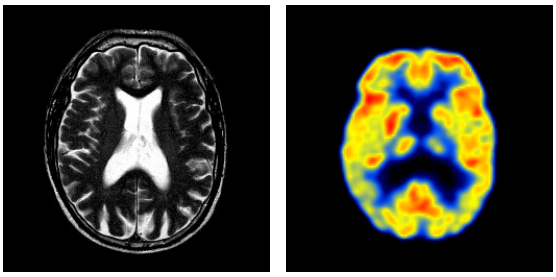



Figure 1.2: Multimodal medical imaging of Alzheimer's disease

clearly observed, enabling the definitive diagnosis of Alzheimer's disease. In addition, in the emerging non-invasive treatment technology of histotripsy, medical imaging technologies also play an indispensable role. Brightness-mode (B-mode) ultrasound (US) imaging is typically used for real-time monitoring of cavitation bubble formation during the histotripsy process (indicated by high echogenic areas) and for assessing tissue liquefaction by capturing hypoechoic regions post-treatment [1–4]. Meanwhile, previous studies have indicated that MRI may be a promising modality for providing intraprocedural monitoring during histotripsy [5, 6] as well as for assessing histotripsy-induced liquefaction post-treatment [5, 7]. These imaging techniques provide critical support for histotripsy research, enabling researchers to visually observe the effects of different treatment protocols on tissues.

However, the application of a single imaging modality in the aforementioned scenarios is often limited by its own constraints. Doctors and researchers often need to conduct comprehensive analyses based on medical images from different modalities. Currently, due to the wide variety of medical imaging equipment, the images obtained by doctors vary significantly, which increases the workload for doctors and researchers, thereby reducing their work efficiency and affecting diagnostic accuracy. Furthermore, in histotripsy research, existing methods lack a comprehensive evaluation of the impact of treatment parameters and fail to precisely quantify the effects of different treatment parameters (such as acoustic power, applied pulses, etc.) on human tissues.

To address these issues, this study focuses on two key areas of Multimodal medical imaging. First, we uti-

lize multimodal medical imaging techniques to investigate and evaluate the impact of various parameters on tissue liquefaction outcomes during the histotripsy treatment process, providing scientific evidence for its development. Second, in response to the growing demand for multimodal medical image analysis, we propose an efficient fusion method that generates a high-quality fused image containing rich information from different imaging modalities. This method offers multimodal information to doctors and researchers without visual bias, significantly assisting medical diagnosis and scientific research, thereby reducing the workload of doctors and researchers and enhancing their work efficiency.

The thesis structure is organized as follows: Section 2 reviews state-of-the-art developments in histotripsy and multimodal medical image fusion. Section 3 provides a detailed description of the methodologies for the two studies. Section 4 comprehensively demonstrates the experimental results through rigorous systematic analysis. The concluding section (Section 5) systematically synthesizes the principal research contributions and methodological advancements achieved in this investigation.

# 2 Literature Review

This chapter presents the related work of two research areas, highlighting the limitations of existing methods and how our approach addresses these shortcomings. Section 2.1 provides an overview of histotripsy technology, while Section 2.2 focuses on multimodal medical image fusion techniques.

## 2.1 Histotripsy technology

Histotripsy is a transformative technology introduced approximately two decades ago that uses focused ultrasound (FUS) to non-invasively destroy tissue mechanically, through controlled acoustic cavitation [8, 9]. The technique employs repetitive high-pressure FUS pulses of short duration, delivered at low duty cycles, that initiate the fast expansion and violent collapse of intrinsic bubble clouds in the tissue [10, 11] Destruction occurs in regions where the applied FUS pressure exceeds the intrinsic cavitation threshold of the targeted tissue, resulting in well-defined treatment zones with sharp boundaries, with the produced tissue debris eventually resorbed by the body within 2 months [12, 13].

Currently, Preclinical animal studies have explored the feasibility of histotripsy for treating various tumours such as pancreatic [14], renal [15], bone [16], breast [17] and prostate [18]. Additionally, histotripsy has been found promising for managing neurological [19] and cardiac [1], pathologies as well as an alternative thrombolytic therapy [20]. Notably, due to its precise tissue-destruction abilities, histotripsy has been recently approved by the Food and Drug Administration (FDA) for clinical treatment of liver tumours [12], while its preliminary therapeutic safety and efficacy in treating calcified valve stenosis [21] and benign prostatic hyperplasia [22] has also been realised successfully through human trials.

Therefore, as histotripsy advances toward broader clinical applications, it is crucial for researchers to thoroughly understand the interaction between US waves and human tissues. However, conducting experiments on human subjects poses numerous challenges; therefore, some researchers have turned to using animal tissues to simulate the behavior of human tissues. For example, previous studies executed on excised porcine atrial wall tissue have demonstrated that the volume of tissue liquefaction is dependent on the applied ultrasonic parameters such as the pulse intensity and duration as well as the pulse repetition frequency (PRF) [23, 24]. Specifically, for exposures applied with a 788-kHz single element transducer at a fixed pulse duration, increased histotripsy-induced tissue perforation volume is observed with higher intensity [24] and increasing PRF [23], while for pulses delivered at a constant duty factor, higher erosion rate is observed at shorter pulse durations [23].

Although the histotripsy-induced erosion in excised animal tissues can qualitatively mimic in-vivo results [23], the repetitive use of such models can be quite costly. As such, tissue mimicking phantoms that simulate the acoustic and mechanical properties of biological tissues are often used as evaluation tools, providing a cost-effective and controlled experimental environment for establishing essential foundational work for histotripsy studies before in-vivo translation [25–27].

Agar-based phantoms are prominently used because they can replicate the mechanical properties of tissues, within a physiologically relevant range, just by adjusting the agar concentration [28]. Almost a

decade ago, Xu and colleagues executed experiments on agar-based phantoms to examine the impact of the mechanical stiffness of the phantoms, the sonication time, and treatment direction of the FUS transducer on the dimensions of the histotripsy-induced fractionation [27]. Sonications were executed in phantoms that were developed with varying agar concentrations of 1.5-3% to attain different stiffnesses and were delivered with varying exposure times, and thus varied pulses, whilst also mechanically moving the transducer bilaterally along the axis of the ultrasonic beam. Phantom dissection post-exposures demonstrated that larger-sized liquefied areas were achieved with longer exposure times and in phantoms with lower agar concentration, indicating the need of choosing treatment parameters based on the tissue's mechanical properties.

In fact, numerous follow-up phantom studies have been executed to systematically ascertain the relationship of the cavitation dynamics and the effectiveness of histotripsy with the mechanical properties of the host medium. Vlaisavljevich et al. [28] performed multi-cycle histotripsy experiments on phantoms and excised pork tissue having varying Young's moduli to examine the required threshold for histotripsy. Findings revealed that mediums with increased stiffness exhibited increased thresholds for cavitation initiation, suggesting the requirement for applying high intensities at higher tissue stiffnesses to initiate bubble growth. However, in a follow-up in-vitro study by Vlaisavljevich et al. [25] employing single-cycle histotripsy on agar phantoms with tuneable mechanical stiffness, it was revealed that the intrinsic threshold remains unaffected by the medium stiffness for Young's moduli less than 1 MPa.

Later in-vitro studies of multi-cycle histotripsy demonstrated that the mechanical properties of the tissue not only affect the cavitation threshold, but also the extent of mechanical fractionation of the treated medium [29–31]. Specifically, the application of multi-cycle histotripsy bursts on agar-based phantoms having varied mechanical strength in a study by Vlaisavljevich et al. [29], yielded in smaller-sized lesions on phantoms having greater mechanical strength, indicating a correlation between mechanical strength and histotripsy fractionation. Exposures that were executed on various ex-vivo porcine tissues also in the same study [29], demonstrated that low mechanical stress and density as well as high water content and fractional strain enhance the sensitivity of tissues to histotripsy-induced damages. Consequently, the authors emphasized the necessity of adjusting the histotripsy parameters based on the mechanical properties of the treated tissue to ensure effective histotripsy and maintain precision [29]. Similarly, Xu et al. [30] examined the impact of various mechanical properties of agar-based phantoms on the histotripsy-induced erosion and confirmed that increased lesion volumes were achieved in phantoms with lower mechanical strength, as quantified by the Young's modulus. However, the authors concluded that other mechanical properties such as varied fracture toughness and bending strength do not significantly affect the extent of phantom fractionation [30]. Notably, in a more recent study characterising the effect of the stiffness of agar-based phantoms on bubble activity and resulting liquefaction, no notable differences in the eroded areas were observed for phantoms with Young's moduli between 12 kPa to 142 kPa [31].

Although significant progress has been made in histotripsy research, there is still a lack of studies assessing the impact of specific ultrasound parameters on human tissue segmentation, which is essential for optimizing histotripsy treatment parameters. Therefore, based on previous studies and the effectiveness of phantoms as human tissue analogs for such experiments, we aim to address this issue through the following approaches:

a) To avoid ethical concerns, minimize economic losses, and align with sustainability goals, we use

phantoms as human tissue analogs for this experiment.

b) We investigate the effects of ultrasound treatment by varying acoustic power, the number of applied pulses, duty factor and pulse repetition period (PRP) to analyze the resulting erosion dimensions of the phantoms.

c) We employ agar-based phantoms with relatively low agar concentrations, as these phantoms closely mimic the properties of human soft tissues and allow for clear observation of erosion effects.

d) We utilize multimodal medical imaging techniques, including MRI, US imaging, and visible light imaging, to evaluate the results comprehensively.

## 2.2 Multimodal medical image fusion

Multimodal medical image fusion aims to combine the advantages of various modalities into a single image. This method can provide clinicians with more comprehensive information, improving disease diagnosis. It also offers richer data for subsequent computer vision tasks, thereby improving the effectiveness of image-guided diagnosis and assessment of medical problems [32].

Multimodal medical image fusion techniques can be broadly classified into traditional approaches and deep learning-based methods. Azam *et al.* [33] present a thorough review of multimodal medical image fusion techniques, emphasizing their significance in disease diagnosis and treatment. In the initial stages of this technology's development, traditional image processing methods were primarily used for fusion. This method primarily depends on designing specific image decomposition strategies and fusion rules based on these decomposition strategies for image fusion. Various decomposition strategies are applied, which primarily include pyramid decomposition, filter decomposition, and frequency transform decomposition. Pyramid decomposition divides the source image into multiple sub-images arranged in a pyramid structure through successive filtering and downsampling operations. Each layer's sub-images are then fused separately, and finally, the fused image is reconstructed. For example, Burt *et al.* [34] were among the first to apply Laplacian pyramid decomposition techniques to image fusion. However, the Laplacian pyramid cannot effectively represent the contours and contrast of images. To overcome this limitation, Du *et al.* [35] proposed a joint Laplacian pyramid with multiple features for medical image fusion. This method first converts the input images into their multiscale representations using the Laplacian pyramid. Then, both contrast feature maps and contour feature maps are extracted from images at each scale. Afterward, a fusion strategy is applied to combine the pyramid coefficients, and the fused image is then reconstructed through inverse pyramid transformation. Although pyramid decomposition methods have achieved certain fusion results, they also face challenges such as detail loss and high computational cost. In contrast, filter-based decomposition methods can address these issues. This method typically uses filters to decompose the image into structure and base layers. Each decomposed layer is then fused separately according to specific fusion rules, and finally, the fused results of each layer are combined to obtain the final fused image. It features simplicity in computation and relatively good detail preservation. A representative example is the work of Li *et al.* [36], who used joint bilateral filtering to decompose the image into an energy layer and a structure layer. The structure layer is fused using a local gradient energy operation, while the energy layer is fused using a maximum absolute value operation. Finally, the fused image is reconstructed by combining the fusion results of the energy and structure layers. Du *et al.* [37]

decomposed the image into smooth, texture, and detail layers using local extrema and low-pass filters in the spatial domain. They then applied three different fusion rules to fuse each layer and obtained the final fused image by combining the fused layers. Notably, this method enhances contrast in the texture and edge layers, preserving illumination closely related to tumor regions. Although filter-based decomposition methods offer advantages such as computational simplicity, relatively good detail preservation, and robustness to noise, they still have certain limitations, such as spatial distortions in the fused image and reduced image clarity. Frequency transform decomposition methods primarily involve converting input images into the frequency domain, applying fusion algorithms to the transformed images, and finally performing an inverse transformation to obtain the fused image. For example, Kumar *et al.* [38] proposed a method for medical image fusion using an improved fast discrete curvelet transform and Type-2 fuzzy entropy. They first employed a weighted fast discrete curvelet transform to separate the high-frequency and low-frequency components of the images to be fused. Then, they fused the low-frequency information using Type-2 fuzzy entropy and averaged the high-frequency information. Finally, the image was reconstructed using the inverse fast discrete curvelet transform. Gupta *et al.* [39] employed the nonsubsampled shearlet transform (NSST) for medical image fusion. Notably, to enhance the edge information of medical images and increase the contrast of the fused image, they utilized a fusion method based on anisotropic diffusion and incorporated its results into the NSST-based fusion process.

In summary, traditional methods decompose an image into multiple layers using various strategies, with each layer representing different types of information. Fusion rules are then applied to combine each layer individually, with common approaches including weighted averaging, the $l_1$-max rule, and local energy functions [40]. Finally, the fused image is reconstructed using inverse transformations or similar techniques.

Although traditional approaches for medical image fusion have achieved some success, they require complex rule design, demand substantial human resources, and often fail to produce optimal fusion results across different scenarios, inevitably reducing fusion efficiency. To address these issues, deep learning-based medical image fusion methods have emerged, leveraging the strong feature extraction and representation capabilities of deep learning to avoid the complexity of manual rule design. It is mainly divided into methods based on convolutional neural network (CNN) [41–44] and those based on generative network [45–48].

CNN, with its powerful feature extraction capabilities, is extensively used in computer vision tasks. Similarly, in Multimodal medical image fusion tasks, many researchers use CNN-based methods to address various challenges. As a representative, Liu *et al.* [41] proposed utilizing CNN for medical image fusion. A Siamese convolutional network is adopted to generate a weight map that integrates the pixel activity information from two source images and employs an image pyramid in a multiscale manner for fusion. Liang *et al.* [42] proposed an end-to-end deep learning network that uses two separate branch networks with non-shared weights to extract complementary information from medical images of two different modalities. The network then reconstructs the final fused image from the extracted information. Shibu *et al.* [43] proposed a medical image fusion method based on multi-scale decomposition with convolutional neural networks and sparse representation. They fused high-frequency information using a CNN fusion rule and low-frequency information using a Non-Subsampled Contourlet Transform with sparse representation (NSCT-SR) fusion rule. Xu and Ma [44] utilized surface-level and deep-level constraints

to enhance information preservation. Notably, they enhanced the color information of the fused image by using the MRI images to improve the CB and CR channels of the PET images.

Generative model-based medical image fusion techniques harness the powerful generative capabilities of these models to integrate information from different medical imaging modalities, resulting in a single, information-rich fused image. As a representative, Ma *et al.* [45] propose the dual-discriminator conditional generative adversarial network, referred to as DDcGAN, which employs a single-generator and dual-discriminator structure for generating fused images. Jun *et al.* [46] used a multi-generator multi-discriminator conditional generative adversarial network for medical image fusion. Specifically, they employed two generators: one to produce realistic fused images and the other to enhance dense structural information in the final fused image, ensuring that functional information is preserved. Although GAN-based methods have achieved some results, issues such as unstable training processes and difficulty in model convergence affect the quality of the generated images. Recent advancements in diffusion models [49] have enriched the methods for addressing challenges in computer vision tasks. Their ease of training and high-quality image generation capabilities make them increasingly prominent in medical image fusion tasks. As a representative, Zhao *et al.* [47] were the first to apply the diffusion model to accomplish multimodal medical image fusion. They formulated the fusion task as a conditional generation problem within the denoising diffusion probabilistic model (DDPM) sampling framework, further dividing it into an unconditional generation subproblem and a maximum likelihood subproblem. Yi *et al.* [48] employed a diffusion model combined with prior fusion knowledge for medical image fusion tasks, addressing the issue of the diffusion model lacking ground truth in image fusion tasks. In conclusion, deep learning-based methods have achieved impressive results in producing information-rich fused images by designing specific models and loss functions to optimize fusion outcomes.

Based on our observations, although current medical image fusion algorithms have achieved promising results, they still face the following limitations:

**Limitation** 1: Some key detailed features of multimodal images are not fully extracted, and certain detail features may be lost during the fusion stage, resulting in a loss of critical details in the fused image and impacting the effectiveness of information fusion.

**Limitation** 2: Most medical image fusion methods, when handling tasks that involve the fusion of color and grayscale images, typically convert the color images to grayscale before performing the fusion using the grayscale images, and then map the fused results back to color images. This approach often results in information loss, leading to poor inheritance of complementary information between modalities and ultimately resulting in a low richness of information in the fused image. For example, in PET-MRI image fusion tasks, this approach can lead to some detailed information from the MRI being obscured by the color information from the PET, resulting in a lack of clear texture information from the MRI in the fused image. Additionally, some edge information from the PET may also be lost.

**Limitation** 3:Many multimodal medical image fusion methods, particularly those involving the fusion of color and grayscale images (e.g., PET-MRI fusion), often overlook the evaluation of color information. However, color information in medical images is vital for diagnosis and research; for instance, in PET images, color reflects physiological functional data such as tissue blood flow and metabolism.

**Limitation** 4: In some deep learning-based medical image fusion methods, the design of the loss function

is often based on model and information theory rather than tailored to specific characteristics and advantages of each of the modalities. While such designs offer general applicability, they can result in poor inheritance of complementary information between modalities, causing the loss of valuable information from some modalities and ultimately impacting diagnostic accuracy.

To address the aforementioned limitations in medical image fusion, inspired by the successful applications of DDPM [50, 51] and previous image fusion tasks, in this work we develop a novel multimodal medical image fusion strategy based on DDPM. Specifically, The main solutions are as follows:

**Solution** 1: To address **Limitation 1**, we design a novel Edge Enhancement Dense Block (EEDB), in which includes the newly designed Edge Detail Coordinate Attention mechanism (ECA). We use DDPM and EEDB together as feature extractor to capture multimodal medical image features, ensuring the thorough extraction of critical detail features and overcoming the limitations of inadequate detail feature extraction in medical image fusion tasks.

**Solution** 2: To address **Limitation 1**, we design a Feature-enhancement Reconstruction Network (FER) to enhance image features while obtaining the fusion result. By embedding an Edge Spatial Attention mechanism (ESA) [52], the network heightens attention to detail information, ensuring minimal loss of detail during reconstruction.

**Solution** 3: To address **Limitation 2**, we introduce a multi-channel joint learning method that simultaneously learns the R, G, and B channels of color medical images and grayscale medical images to address the issue of information loss in the color and grayscale medical image fusion task.

**Solution** 4: To address **Limitation 3**, we introduce the CIEDE2000 color difference formula [53] and design an Average Color Difference (ACD) metric based on it for assessing color information in fused images. This metric enables the evaluation of color fidelity in fused images from both color and grayscale medical image fusion tasks, thereby addressing the common lack of color information assessment in most existing fusion methods.

**Solution** 5: To address **Limitation 4**, We design different joint loss functions for fusion based on the advantages and characteristics of different modalities, ensuring the comprehensive inheritance of advantageous information from each modality to guarantee the richness of information in the final fused image.

## 2.3 summary

In summary, this chapter presents a comprehensive and systematic review of histotripsy technology and multimodal medical image fusion. It analyzes the current research status in both fields while highlighting key advancements and existing limitations. Through a literature survey, the chapter identifies prevailing challenges in these domains. Finally, it provides detailed discussions on how the proposed approaches address these challenges by offering potential solutions to enhance the effectiveness and reliability of both histotripsy technology and multimodal medical image fusion methods.

# 3 Research Methodology

In this section, we provide a detailed description of the experimental details for the two studies. The first subsection elaborates on the experimental investigations of histotripsy technology, while the second subsection presents a comprehensive introduction to our proposed multimodal fusion method.

## 3.1 Materials and Methods of histotripsy

### 3.1.1 Agarose tissue-mimicking phantoms

Agar-based phantoms were prepared and used as a controlled sonication environment for the histotripsy experiments. The phantoms were fabricated with specific agar concentrations after previous studies demonstrating that this can tune their mechanical (new insight into agarose gel mechanical properties), acoustic (ultrasonic attenuation), thermal (acoustic and thermal characterization) and magnetic (mr relaxation times) properties to human tissue levels. Phantoms were prepared by slowly mixing the appropriate amount of agar (101614, Merck KGgA, Darmstadt, Germany) into deionized/degassed water heated to $50\,^{\circ}$C and stirring the solution with a magnetic hot plate until its temperature exceeded $80\,^{\circ}$C. The solution was then cooled under continuous stirring, and at approximately $60\,^{\circ}$C it was poured into dedicated 3D-printed (Raise3D E2, Raise3D, California, USA) Polylactic acid (PLA) molds and was left overnight at ambient temperature to solidify. The molds had dimensions of 10 cm (width) × 20 cm (length) × 10 cm (height) and allowed removal of their bottom surface. The bottom surface was fitted for fabrication and was removed upon phantom jellification. Before the experiments, the jellified phantoms (as integrated in the molds) were independently submerged in water and were degassed for 3 hours in a vacuum chamber (VC2523AG, Vacuum Chambers, Jodlowa, Poland) with a dual-stage vacuum pump (VP260, Vacuum Chambers) at a pressure of -1 bar.

### 3.1.2 Histotripsy experimental set-up

A custom-built single element spherically focused ultrasonic transducer developed in-house (piezoceramic element from Hubei Hannas Tech Co., Hubei, Wuhan, China) with a central frequency of 2.6 MHz was employed for all experiments in this study. The FUS transducer had a 50 mm diameter, a geometric focal length of 65 mm and an effective F-number of 1.3. The transducer was immersed in an acrylic tank filled with deionized/degassed water having an oxygen content less than 4 mg/L at a temperature of 25.3 oC as measured with an oxygen meter (HI5421, Hanna Instruments, Limena, Italy). The transducer was submerged through a 3D-printed (Raise 3D E2, Raise 3D) PLA mechanical positioning system, a ttached to the edges of the acrylic tank, that allowed transducer fixation centrally at the bottom of the tank and its manual motion in the lateral and vertical planes. The phantom mold was secured on top of the tank via dedicated couplings, resulting in immersion of the enclosed agar-based phantom in the deionized/degassed water, above the transducer as shown in Figure 3.1. The transducer was connected to an RF power amplifier (AG1012, T&C Power Conversion Inc., Rochester, NY, USA) which was driven in the pulsed mode to produce the ultrasonic bursts for the histotripsy exposures. For controlling the hardware and ultrasonic exposures an in-house MRgFUS software [54] on an ordinary laptop was employed. Pa-

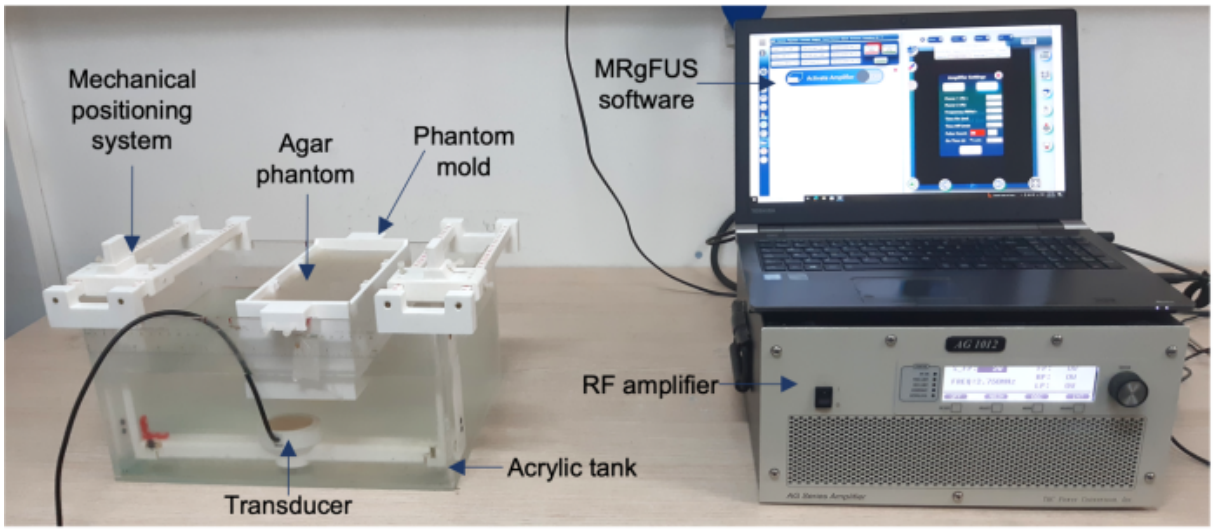



Figure 3.1: Picture of the experimental setup.

rameters for the pulsed histotripsy sonications were selected by appropriately commanding the required power, pulse length, pause between consecutive pulses and number of applied pulses through the user interface as explicitly described previously [54]. Ultrasonic protocols were applied at different single-spot treatment locations by appropriately aligning the transducer at varied sites below the bottom surface of the agar-based phantom. This was achieved by laterally moving the transducer using the mechanical positioning system and sliding the mold couplings along the tank's edge, thus moving the longitudinal axis of the phantom sideways. For all tested protocols, the transducer focus was set 2 cm deep within the phantom.

### 3.1.3 Evaluation Methodology

In this study, we evaluated the experimental results using multimodal medical imaging techniques, including US imaging, MRI, and visible light imaging. For US imaging, after histotripsy exposures, the agar-based phantoms were removed from the acrylic tank, and their sonicated surfaces were initially scanned using a diagnostic US system (DP-50, Shenzhen Mindray Bio-Medical Electronics Co., Shenzhen, China) equipped with a 7.5 MHz linear probe (75L38EA, Shenzhen Mindray Bio-Medical Electronics Co.). B-mode US images were captured at each treatment site to examine the formation of histotripsy-induced phantom erosion along the axial direction (parallel to the beam direction). For visible light imaging, we used direct visual assessment to evaluate the results. The agar-based phantoms were dissected to examine the induced fractionated lesions based on gross morphology. Initially, the phantoms were cut 2 cm from the sonicated side to observe the damage caused by each protocol on the transverse plane relative to the direction of the ultrasound beam's propagation. Successfully formed lesions were then sliced along the axial direction to reveal the extent of erosion along the beam axis. In cases where excess moisture was present inside the lesions, paper tissue was used to absorb the liquid slurry. Finally, the dimensions (diameter and length) of the clean resultant lesions were measured using a digital caliper. For MRI imaging, the phantom samples were removed from the acrylic tank and positioned on the table of a clinical 3T MRI scanner (Magnetom Vida, Siemens Healthineers, Erlangen, Germany). Imaging was performed using a 64-channel head-and-neck coil (Siemens Healthineers) with a high-resolution T2-Weighted Turbo Spin

Echo (T2-W TSE) sequence. The experimental setup is illustrated in Figure 3.2. The acquired T2-W

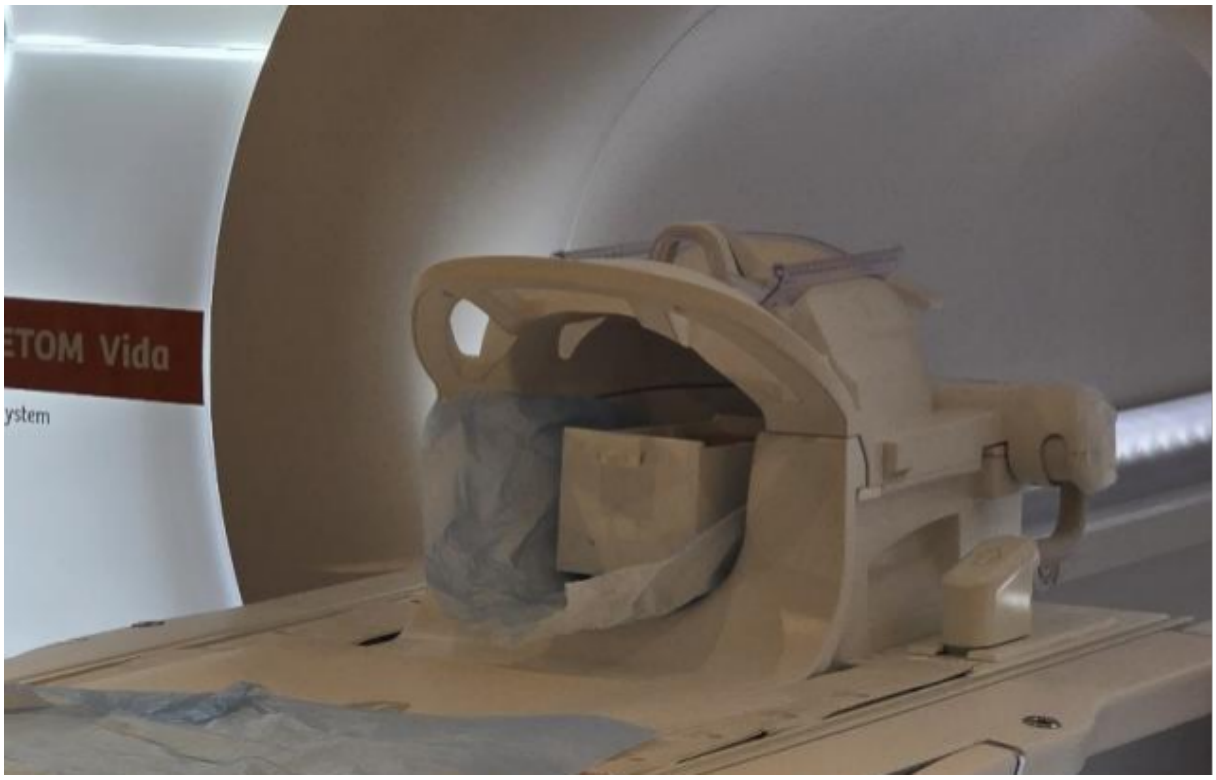



Figure 3.2: Experimental setup arranged on the table of the 3 T MRI scanner, with the imaging coil fixed above.

TSE images were analyzed post hoc using Digital Imaging and Communication in Medicine (DICOM) software (MicroDicom, MicroDicom Ltd., Sofia, Bulgaria). Distance measurement tools within the software were used to quantify the dimensions of the liquefied lesions. The lengths of the histotripsy-induced lesions were determined by visually outlining the boundaries of each liquefied region on the T2-W TSE images and measuring their extent along the vertical axis. Similarly, the diameters of the lesions were measured within these boundaries along a horizontal axis located 20 mm from the inferior side of the phantom, corresponding to the focal depth within the phantom.

### 3.1.4 Benchtop assessment of experimental parameters on phantom erosion

#### 3.1.4.1 Effect of agar concentration on phantom erosion and lesion dimensions

Pulsed ultrasonic exposures were performed on three agar-based phantoms, each fabricated with different weight/volume percentage concentrations (% w/v) of agar. The effect of agar concentration on the extent of histotripsy-induced phantom erosion was evaluated using visible light imaging for detailed assessment. Specifically, phantoms with 2%, 2.5%, and 3% w/v agar concentrations were sonicated using identical histotripsy protocols that were delivered at a low duty cycle of 2 %. Ultrasonic bursts were generated at an acoustic power of 129 W and were applied on each of the three phantoms at varied PRP of 250, 500 and 750 ms. Noteworthy, the pulse duration varied accordingly for each PRP (250-750 ms) so as to maintain a constant duty cycle, leading to pulse lengths of 5 ms, 10 ms, and 15 ms being employed, respectively. A total of 1000 pulses were delivered to each phantom at the PRP of 250 ms and 500 ms.

A higher dose of 5000 pulses was also explored for the PRP of 500 ms, while at the highest PRP of 750 ms, 1500 pulses were applied on each agar-based phantom.

#### 3.1.4.2 Effect of acoustic power on phantom erosion and lesion dimensions

Histotripsy pulses were also delivered at varied acoustic power to different treatment points on an agar-based phantom consisting of 2 % w/v agar to examine any differences arising in the extent of phantom fractionation. Specifically, exposures were performed at varied acoustic power of 200, 210 and 215 W. The effect of varied acoustic power was examined at two different PRP of 500 and 750 ms. As the duty cycle remained at 2%, burst periods of 10 ms (PRP of 500 ms) and 15 ms (PRP of 750 ms) were used for sonications executed at each acoustic power. Bursts at each acoustic power consisted of 1000 pulses at the PRP of 500 ms and 1500 at the PRP of 750 ms corresponding to active exposure times of 10 s and 22.5 s, respectively. Finally, we analyzed the experimental results using both US imaging and visible light imaging.

#### 3.1.4.3 Effect of applied pulses on phantom erosion and lesion dimensions

Sonications were also performed on a 2% w/v agar-based phantom to assess the impact of the histotripsy pulses on the dimensions of the induced fractionated lesions. Exposures were executed at multiple treatment locations on the phantom, with a total of either 200, 500 or 1000 pulses delivered at each spot. Notably, the effect of varying the ultrasonic pulses was evaluated at three different PRPs of 250, 500, and 750 ms. It is worth stating that at the PRP of 750 ms, an additional sonication with increased pulses of 1500 was also examined. For all different exposures the applied acoustic power was maintained constant at 215 W, while the duty factor remained at 2%. Therefore, across the 10 different protocols bursts of 5 ms, 10 ms and 15 ms duration were delivered at the PRP of 250, 500 and 750 ms, respectively. These burst periods corresponded to active FUS times of 1, 2.5, and 5 s and total treatment durations of 50, 125, and 250 s for the varying pulses delivered at the PRP of 250 ms. Correspondingly, the active exposure times were 2, 5 and 10 s for the PRP of 500 ms and 3, 7.5, 15, and 22.5 s for the PRP of 750 ms. These corresponded to total treatment durations of 100, 250, and 500 s for the PRP of 500 ms, and 150, 375, 750, and 1125 s for the PRP of 750 ms.

#### 3.1.4.4 Effect of PRP on phantom erosion and lesion dimensions

The effect of the PRP on the dimensions of the histotripsy lesions was investigated through pulsed exposures that were executed on a 2% w/v agar phantom using varied PRP values of 250, 500 and 750 ms. Exposures at the three different PRPs were consistently delivered to the agar-based phantom at a constant duty factor of 2% and an acoustic power of 215 W. Due to the constant duty factor of the sonications, the pulse durations varied accordingly for each PRP, lasting 5, 10, and 15 ms for the PRP of 250, 500, and 750 ms, respectively. Pulsed sonications at each different PRP were executed by applying three different number of pulses to examine the effect of varying the PRP at different histotripsy doses. In this sense, exposures at each PRP of 250-750 ms consisted of either 200, 500, or 1000 pulses delivered at identical ultrasonic power.

#### 3.1.4.5 MRI-based assessment of Effect of acoustic power on phantom erosion and lesion dimensions

In addition, we employed both MRI imaging and visible light imaging to investigate the effects of histotripsy on phantoms under different acoustic power levels. Histotripsy pulses were also delivered at varied acoustic power to different treatment points on an agar-based phantom consisting of 2 % w/v agar to examine any differences arising in the extent of phantom fractionation. Specifically, exposures were performed at varied acoustic power of 110, 140, 170, 200 and 220 W. The effect of varied acoustic power was examined at 1000 ms. As the duty cycle remained at 2%, burst periods of 20 ms were used for sonications executed at each acoustic power. For each acoustic power level, 1000 pulses were delivered in bursts, corresponding to an active exposure time of 20 seconds.

#### 3.1.4.6 MRI-based assessment of Effect of applied pulses on phantom erosion and lesion dimensions

Sonications were also performed on a 2 % w/v agar-based phantom to assess the impact of the histotripsy pulses on the dimensions of the induced fractionated lesions. Exposures were executed at multiple treatment locations on the phantom, with a total of either 100, 200 ,300 ,500 or 1000 pulses delivered at each spot. Notably, the effect of varying the ultrasonic pulses was evaluated at PRP of 1000 ms. For all exposure conditions, the applied acoustic power was kept constant at 215 W, with a duty factor of 2%. This corresponds to a burst period of 20 ms. Finally, we analyzed the results using both MRI and visible light imaging.

#### 3.1.4.7 MRI-based assessment of Effect of duty factor on phantom erosion and lesion dimensions

Histotripsy sonication was also performed on a 2% w/v agarose-based phantom to evaluate the effects of different duty factors on phantom erosion using MRI and visible light imaging. Five different locations were treated on the phantom by applying 1000 histotripsy pulses at an acoustic power of 216 W. Different duty factors of 1%, 2%, 3%, 4%, and 5% were applied to five distinct treatment locations. Thus, pulse durations of 10 ms, 20 ms, 30 ms, 40 ms, and 50 ms were used for the corresponding duty factors.

#### 3.1.4.8 MRI-based assessment of Effect of PRP on phantom erosion and lesion dimensions

Histotripsy sonications were also performed on a 2% w/v agar-based phantom to assess the effect of varied PRP on mechanical phantom erosion using both MRI and visible light imaging. Ten different locations were treated on the phantom by applying 1000 histotripsy pulses at an acoustic power of 215 W. Exposures at half of the treatment locations were performed at a PRP of 750 ms, while the other five locations were sonicated with a higher PRP of 1000 ms. A constant duty factor of 2 % was used for executing the varying PRP protocols. Consequently, varying pulse durations of 15 ms and 20 ms were employed for the PRPs of 750 and 1000 ms, respectively, resulting in active FUS exposure times of 15 s and 20 s at each treatment point.

## 3.2 Methods of multimodal medical image fusion

### 3.2.1 Overall Framework

In this work, we perform two typical multimodal medical image fusion tasks:

1. grayscale image fusion (CT and MRI);
2. color and grayscale image fusion (PET and MRI);

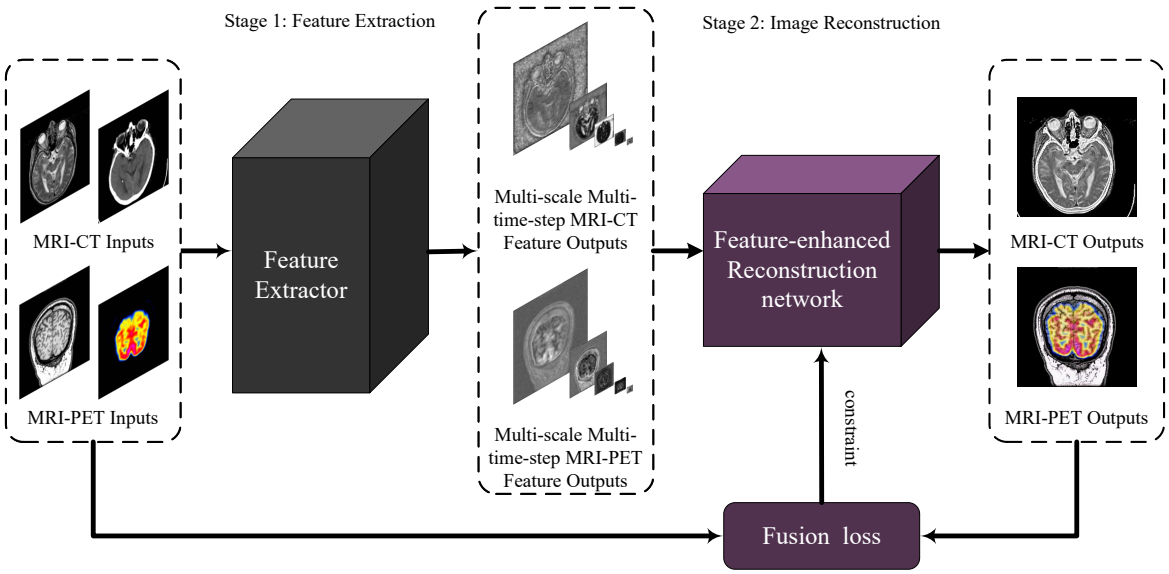



Figure 3.3: Overall framework of the proposed method. Source images are first input into a feature extractor to extract features, which are then fed into the Feature-enhancement Reconstruction Network(FER) for feature enhancement and reconstruction to obtain well-performed fusion results.

Figure 3.3 illustrates the overall procedure for image fusion with the proposed DDPM-EMF, which includes two training stages. The first stage trains the feature extractor to learn and extract joint features from multimodal medical images. The second stage trains the designed reconstruction network to reconstruct the final fused image based on the learned joint features. Multimodal medical images are combined to form a multi-channel image, which serves as the input to the network. For the CT images ($\mathbf{I}_{\text{CT}} \in \mathbb{R}^{HW1}$) and MRI images ($\mathbf{I}_{\text{MRI}} \in \mathbb{R}^{HW1}$), where $H$ and $W$ represent the height and width, respectively, which are grayscale images. These images are concatenated along the channel dimension and then input into the feature extractor. For PET images ($\mathbf{I}_{\text{PET}} \in \mathbb{R}^{HW3}$) and MRI images ($\mathbf{I}_{\text{MRI}} \in \mathbb{R}^{HW1}$), we apply a multi-channel joint learning method to address the information loss issue in this fusion task, where PET images are in color and MRI images are grayscale. These images are concatenated along the channel dimension to form a four-channel image. The combined multi-channel image is then input into the feature extractor for feature extraction, and finally, FER uses the extracted features to reconstruct the final result.

### 3.2.2 Principles of the Denoising Diffusion Probabilistic Model (DDPM)

The DDPM belongs to the class of generative models and has been applied in various computer vision tasks due to its powerful image generation capabilities, including tasks such as image generation [55], [56], [57], image super-resolution [58] [59] [60], and image deblurring [61], [62]. Recently, DDPM has also demonstrated effectiveness in medical image fusion research [47, 48].

DDPM consists of two processes: the forward process and the reverse process. The forward process refers to the gradual addition of Gaussian noise to the data until it becomes random noise, resulting in an isotropic Gaussian distribution $N(0, I)$. The noise operation at time step $t$ is shown as follows:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t}\, x_{t-1}, (1-\alpha_t)\mathbf{I}) \tag{3.1}$$

Here, $x_t$ and $x_{t-1}$ represent the noisy data generated by adding Gaussian noise at time steps $t$ and $t-1$, respectively. $\alpha_t$ represents the variance of the noise added at step $t$. The reverse process is a denoising process that starts from random noise and, using a neural network, performs a series of small denoising operations step-by-step to eventually recover the original image. At each step in the reverse process, we perform a denoising operation on the image $x_t$ to recover $x_{t-1}$. The denoise operation at time step $t$ is shown as follows:

$$p(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 \mathbf{I}) \tag{3.2}$$

where $\sigma_t^2$ is the variance of the conditional distribution $p(x_{t-1} \mid x_t)$.

### 3.2.3 Feature extractor

Currently, multimodal medical image fusion methods often face issues such as insufficient feature extraction and the loss of key details. To address this, we use DDPM as the feature extractor. DDPM [49] generates high-quality images by modeling the diffusion process to restore noise-corrupted images, demonstrating advantages over other generative models. Its ability to efficiently model complex training distributions enables DDPM to extract highly informative and compressed feature representations, which is especially beneficial for multimodal medical image fusion tasks, as it helps retain crucial information from different modalities to achieve more accurate fusion results. Notably, DDPM has also been successfully applied as an effective feature extractor in remote sensing image detection [50] and infrared-visible light image fusion tasks [51].

As shown in Figure 3.4, the main function of DDPM is to work in conjunction with EEDB as a feature extractor, learning relevant information from multimodal medical images and extracting Multi-scale and Multi-time step features from them.

The multi-scale features include feature maps with sizes $(W/16)(H/16)$, $(W/8)(H/8)$, $(W/4)(H/4)$, $(W/2)(H/2)$, and $WH$. The multi-time-step features correspond to the features at time steps $T_1 = 5$, $T_2 = 50$, and $T_3 = 100$. Here, we use the DDPM structure from SR3 [58]. In the backbone network of SR3, we incorporated EEDB to enhance the network's focus on high-frequency information during training. This enables the feature extractor to capture more detailed information, thereby further enhancing its feature extraction capability. The feature ($F$) extracted from the feature extractor can be obtained as

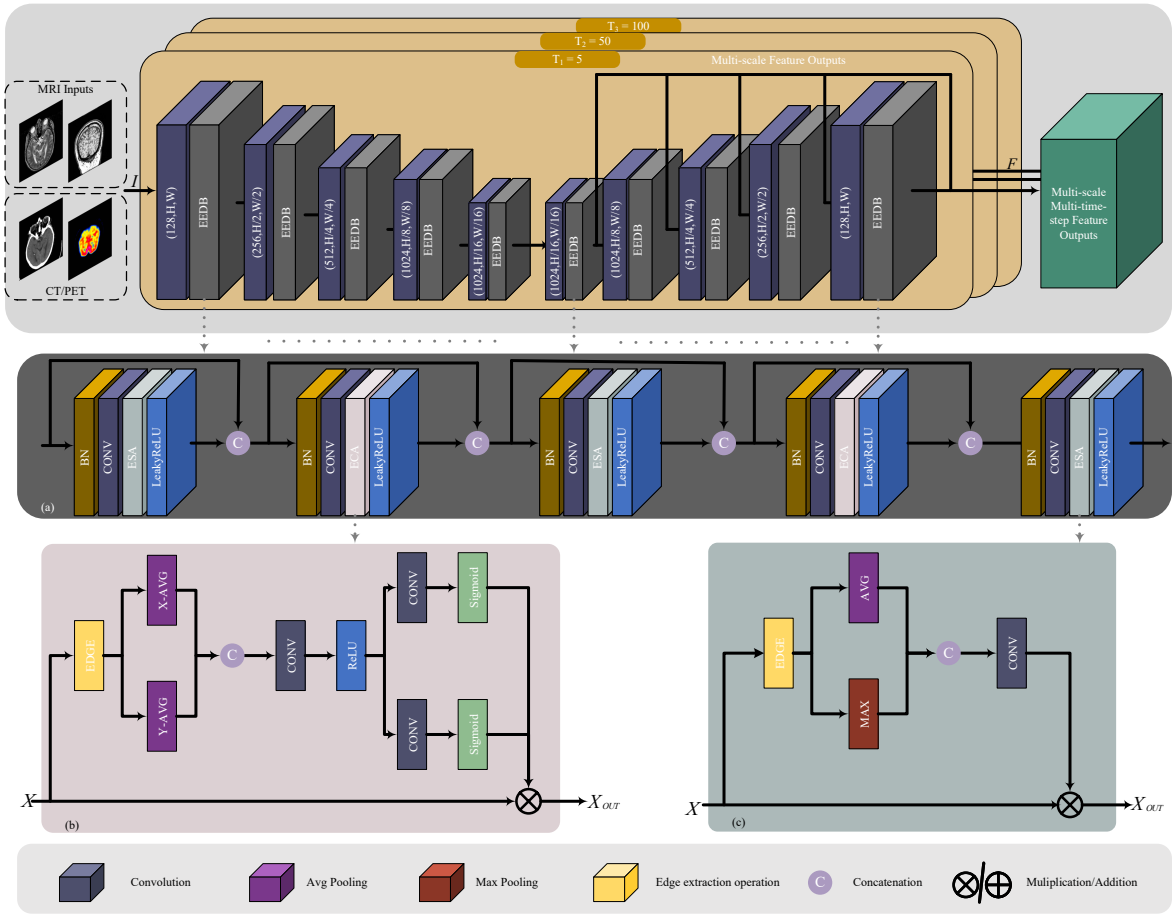



Figure 3.4: Architecture of the feature extractor in the proposed method. "'(n, H, W)' indicates that the output of this convolutional layer contains n channels, with feature maps of size $H \times W$." (a) Architecture of the EEDB. (b) and (c) show the detailed structures of the ECA and ESA, respectively. The bottom position shows the legend of the proposed fusion network.

follows:

$$F = f_{\mathrm{f}}(I) + \theta \cdot \Lambda(I), \tag{3.3}$$

where $f_{\mathrm{f}}$ denotes the feature extraction operation, and $\Lambda$ represents the feature extraction enhancement operation. $\theta$ is a hyper-parameter. For the CT-MRI task, $\theta = 1$ was used. For the PET-MRI task, $\theta = 0$ was used.

### 3.2.4 Edge Enhancement Dense Block

Currently, some other multimodal medical image fusion methods often do not sufficiently extract image detail information, leading to the loss of critical details in the fused images. However, the details in the medical images are crucial, as they often influence the doctor's diagnosis. To ensure that detailed information in medical images is preserved during the feature extraction process, we design an EEDB (as illustrated in Figure 3.4a), which consists of ECA (as illustrated in Figure 3.4b), Edge Spatial Attention mechanism (ESA) (as illustrated in Figure 3.4c) [52] and dense connections. Specifically, we first design ECA based on coordinate attention [63], and then combine ECA with ESA to create an EEDB. Finally, we embedded this block into the DDPM, enabling the feature extraction process to capture more details, thereby enhancing the feature extractor's capacity to provide more comprehensive and detailed information for the subsequent reconstruction process.

Edge attention mechanisms have been proven to be more effective in various computer vision tasks, including image segmentation and image restoration. This enhances the network's focus on detailed information, thereby improving the feature extraction capability of the feature extractor and enabling the final fused image to retain more detailed information. However, most current edge attention mechanisms use convolution to exploit spatial information. Convolution can only capture local relationships and cannot model the long-term dependencies necessary for visual tasks [63]. To capture long-term dependency details in multimodal medical images, we design the ECA using coordinate attention to address this issue. Coordinate Attention embeds positional information into the channel attention mechanism, enabling the network to focus on a larger area while avoiding excessive computational overhead. Specifically, it involves two operations. The first operation, called Coordinate Information Embedding ($f_{\mathrm{CIE}}$), uses two one-dimensional global pooling operations to aggregate the input features in the vertical and horizontal directions into two separate direction-aware feature maps. The second operation, called Coordinate Attention Generation ($f_{\mathrm{CAG}}$), uses convolutions and sigmoid functions to generate an attention weight map with positional information.

The designed ECA is described in Figure 3.4b. Specifically, for a given input feature $X$, we extract edge detail information from $X$, obtaining the multi-channel edge detail feature map $X_{\mathrm{EDGE}}$ as follow:

$$X_{\mathrm{EDGE}} = f_{\mathrm{edge}}(X). \tag{3.4}$$

Here, $f_{\mathrm{edge}}$ denotes the edge extraction operation, which employs a bidirectional gradient-based algorithm to capture high-frequency edge details. The operation is defined by Equation (3.5):

$$E(i,j) = \max\left(E_x(i,j), E_y(i,j)\right), \tag{3.5}$$

where $E_x(i,j)$ and $E_y(i,j)$ represent the maximum gradient magnitudes in the horizontal and vertical directions, respectively. These gradients are computed as:

$$E_x(i,j) = \max\left(\left|I(i,j) - I(i-1,j)\right|, \left|I(i,j) - I(i+1,j)\right|\right), \tag{3.6}$$

$$E_y(i,j) = \max\left(\left|I(i,j) - I(i,j-1)\right|, \left|I(i,j) - I(i,j+1)\right|\right). \tag{3.7}$$

In these equations, $I(i,j)$ denotes the pixel value at position $(i,j)$ in the input image. This operation generates an edge detail feature map for each channel, which is then combined to form $X_{\text{EDGE}}$.

Subsequently, we performed the $f_{\text{CIE}}$ on the edge feature map. This step integrates spatial positional information into the edge features, enabling the attention block to capture long-range dependencies with precise spatial localization, as formulated below:

$$X_{\text{EDGES}} = f_{\text{CIE}}(X_{\text{EDGE}}). \tag{3.8}$$

Here, $X_{\text{EDGES}}$ represents edge feature map with embedded positional information. Through directional feature aggregation along the $x$- and $y$-axes in the edge detail feature map, we generate a pair of axis-aware feature maps. Subsequently, $X_{\text{EDGES}}$ is obtained via concatenation and convolution operations. This design enables the attention mechanism to capture long-range dependencies along one spatial axis while preserving precise positional cues along the orthogonal axis. This modeling of edge detail feature maps in medical images is particularly beneficial for medical image processing tasks, as it allows the attention mechanism to focus more effectively on the anatomical structures within the images.

Afterward, the $f_{\text{CAG}}$ was used to fully utilize the captured positional information, accurately highlighting the regions of interest in the edge detail feature map. This results in an attention weight map with positional information based on the edge detail feature map. The formula is as follows:

$$\{G_H, G_W\} = f_{\text{CAG}}(X_{\text{EDGES}}), \tag{3.9}$$

where $G_H$ and $G_W$ are the attention weights.

Then, by multiplying $G_H$, $G_W$ and the input feature $X$. We get the features denoted as $X_{\text{OUT}}$:

$$X_{\text{OUT}} = X G_H G_W. \tag{3.10}$$

To fully utilize edge detail features, we design a dense connection structure to connect ECA and ESA, forming EEDB. As shown in Figure 3.4a, the EEDB consists of five alternating 'BN layer - Convolution layer - Edge Attention layer - LeakyReLU layer' structures. In this EEDB, the edge attention layers in the second and fourth layers utilize ECA, while the remaining edge attention layers use ESA. They are connected through dense connections to ensure feature propagation. This EEDB will assist the DDPM in extracting more detailed features, enhancing its feature extraction capability.

### 3.2.5 Feature-enhanced Reconstruction network

To improve the utilization of features extracted by the feature extractor and further enhance detailed features, we design a FER. It enhances detail features while reconstructing the fused image. Figure 3.5

shows the overall architecture of FER. As depicted, the FER consists of two parts: five acquisition heads

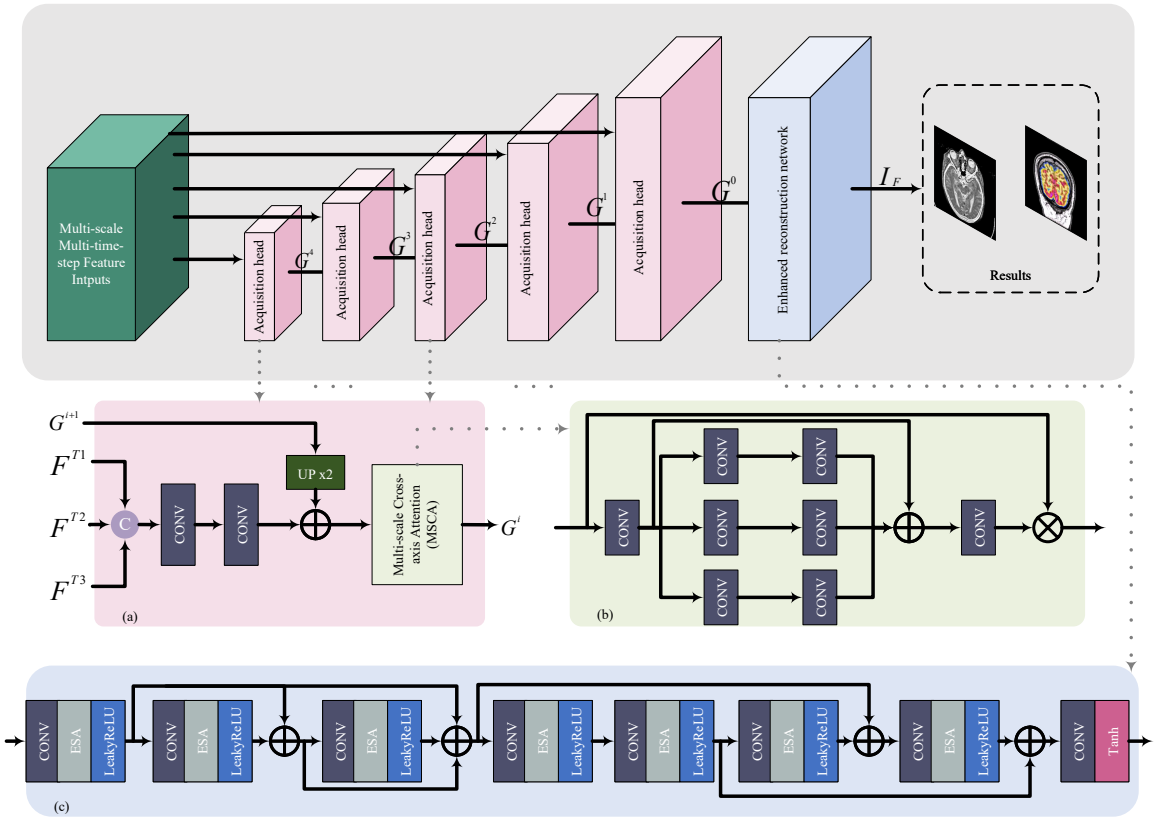



Figure 3.5: Architecture of the FER in the proposed method.(a) Architecture of the acquisition head. (b) Architecture of the Multi-scale Cross-axis Attention (MSCA). (c) Architecture of Enhanced reconstruction network.

(as illustrated in Figure 3.5a) and an Enhanced reconstruction network (as illustrated in Figure 3.5c). With the multi-time step features $F^{T_1}$, $F^{T_2}$ and $F^{T_3}$ as inputs, the role of the acquisition head is to acquire and enhance features of different scales and multiple time steps from the feature extractor. Here, we improved the hierarchical change decoder from [50] and named it the acquisition head. Unlike [50], which uses Channel and Spatial Squeeze and Excitation to enhance channel and spatial re-calibration of the difference feature representations at each scale, we introduce a Multi-scale Cross-axis Attention (MSCA) mechanism [64] in the acquisition head to efficiently capture multi-scale information and build long-range dependencies among pixels. This ensures that the acquisition head can capture long-range dependencies in the feature maps, thereby increasing feature utilization to achieve enhanced feature representation. As shown in Figure 3.5a, it mainly includes convolution, attention blocks, and upsampling. The features captured by the acquisition head can be represented as follows:

$$G = f_{\text{ah}}\left(F^{T_1}, F^{T_2}, F^{T_3}\right), \tag{3.11}$$

where $f_{\text{ah}}$ denotes the operation of the acquisition head, $G$ represents the enhanced features captured from the feature extractor, and $F^{T_1}$, $F^{T_2}$ and $F^{T_3}$ represent the features at different time steps, respectively. To further enhance the detail features extracted and obtain final fusion results, we designed the Enhanced reconstruction network. Figure 3.5c shows the detailed architecture of the designed Enhanced

reconstruction network. As shown, it consists of eight convolutional layers and seven ESA blocks. Skip connections are used to improve training efficiency and facilitate the propagation of information to subsequent layers, thereby avoiding information loss and enhances the information flow. The embedded ESA block in the network enables it to focus on edge details within the feature map, assigning greater weight to edge information and thereby achieving feature enhancement. Finally, the number of channels in the feature map is gradually reduced to reconstruct the final fused image. The following formulation shows how we obtained the final image $I_F$:

$$I_F = f_{\text{ef}}(G). \tag{3.12}$$

Here, $f_{\text{ef}}$ denotes the operation of the enhancement network.

### 3.2.6 Loss fusion

Unlike other fusion tasks, such as multi-exposure image fusion [65] and multi-focus image fusion [66], multi-exposure image fusion addresses the limited dynamic range of a single image by combining multiple images captured at different exposures (e.g., underexposed, normal, and overexposed). Its primary goal is to generate an image that is rich in detail and evenly exposed, thereby avoiding regions that are either too bright or too dark. In contrast, multi-focus image fusion synthesizes images focused on different regions of the same scene into a single, uniformly sharp image. Due to the depth-of-field limitations inherent in optical lenses, a single image often cannot clearly capture both the foreground and background details. Multi-focus fusion overcomes this by integrating several partially focused images to produce a globally clear composite.

Both of these fusion tasks benefit from well-defined ground truths—such as uniformly exposed or sharply focused images—whereas multimodal medical image fusion lacks such fundamental references. This absence poses a significant challenge in the design of effective loss functions.

#### 3.2.6.1 Loss Function for CT-MRI Fusion Task

For the fusion of CT and MRI images, the fused image should simultaneously display the tissue texture details from the CT and MRI images as much as possible, while fully preserving the bone information reflected in the CT images. The loss function for CT and MRI image fusion is illustrated in (3.13):

$$L_{\text{CM}} = L_{\text{G}}^1 + L_{\text{I}}^1. \tag{3.13}$$

Here, $L_{\text{G}}^1$ denotes a single-channel gradient loss, and $L_{\text{I}}^1$ denotes an intensity loss. $L_{\text{G}}^1$ is used to constrain the network, ensuring that the fused image better inherits the texture information from both CT and MRI images. On the other hand, $L_{\text{I}}^1$ ensures that the fused image more effectively inherits the brightness information from these images. $L_{\text{G}}^1$ is given as follows:

$$L_{\text{G}}^1 = \frac{1}{HW}\|\nabla I_F - \max(\nabla|I_{\text{MRI}}|, \nabla|I_{\text{CT}}|)\|_1, \tag{3.14}$$

where $\nabla$ represents the gradient operator. $L_{\text{I}}^1$ is can be estimated as follows:

$$L_{\text{I}}^1 = \frac{1}{HW}\left(L_{\text{IMAX}} + 0.5 \times L_{\text{CT}} + 0.5 \times L_{\text{IMRI}}\right), \tag{3.15}$$

$$L_{\text{IMAX}} = \|I_F - \max(I_{MRI}, I_{CT})\|_1, \tag{3.16}$$

$$L_{\text{IMRI}} = \|I_F - I_{\text{MRI}}\|_1, \tag{3.17}$$

$$L_{\text{ICT}} = \|I_F - I_{\text{CT}}\|_1, \tag{3.18}$$

where $L_{\text{IMAX}}$ denotes the maximum intensity loss to constrain the maximum brightness of the overall image, ensuring the inheritance of bone information from the CT image. $L_{\text{IMRI}}$ and $L_{\text{ICT}}$ represent the intensity losses between the fused image and the MRI image, and between the fused image and the CT image, respectively, which constrain the overall intensity of the fused image.

#### 3.2.6.2 Loss Function for PET-MRI Fusion Task

For the fusion of PET and MRI images, the fused image should contain detailed tissue structure information from MRI and body function information information represented by PET colors. Therefore, the loss function for the PET-MRI fusion task, $L_{\text{PM}}$, consists of both gradient loss and color loss, as shown below:

$$L_{\text{PM}} = L_{\text{G}}^3 + L_{\text{C}}^3. \tag{3.19}$$

Here, $L_{\text{G}}^3$ denotes the multi-channel gradient loss [51], which constrains the inheritance of detailed textures from both MRI and PET images as follows:

$$L_{\text{G}}^3 = \frac{1}{HW} \sum_{i=1}^{3} \|\nabla I_F^i - \max(\nabla|I_{\text{PET}}^i|, \nabla|I_{\text{CT}}|)\|_1. \tag{3.20}$$

$L_{\text{C}}^3$ denotes the color loss, primarily constraining the color information in the fused image as follows:

$$L_{\text{C}}^3 = \frac{1}{HW}\left(w1 \times \left(\sum_{i=1}^{3} \|I_F - I_{\text{MRI}}\|_1\right) + w2 \times \left(\sum_{i=1}^{3} \|I_F - I_{\text{PET}}\|_1\right)\right), \tag{3.21}$$

where $i$ denotes the channel, while $w1$ and $w2$ represent hyperparameters. In this paper, we set both $w1$ and $w2$ to 0.5.

# 4 Results of Histotripsy Experiments and Multi-modal Medical Image Fusion

## 4.1 Histotripsy Experimental Results

### 4.1.1 Effect of agar concentration on phantom erosion and lesion dimensions

Histotripsy sonications that were executed on the three agar-based phantoms of varying agar concentration generated different levels of erosion in each phantom. As shown in the measured lesion diameters in Figure 4.1, no fractionation was created in the tissue mimicking phantoms developed with 2.5 % or 3

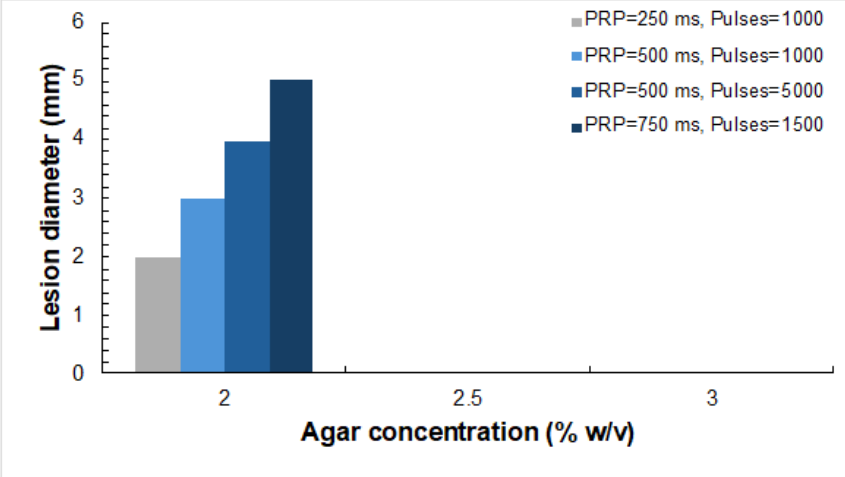



Figure 4.1: Effect of Agar Concentration on Lesion Diameter

% w/v agar concentrations for none of the applied histotripsy protocols despite the use of numerous PRP and number of applied pulses. Contrary, for the phantom with the lowest concentration of agar (2 % w/v), fractionated lesions were successfully created at all PRPs. As expected, differences in the diameters of the formed lesions were observed between the four different protocols applied on the 2 % w/v agar phantom. For example, increasing the PRP from 250 to 500 ms while keeping the applied pulses constant at 1000, resulted in an increase in the lesion diameter from 2 to 3 mm. Similarly, increasing the number of applied pulses to 5000 at the PRP of 500 ms further increased the lesion diameter (4 mm), while an even increased 5 mm diameter was observed at the highest PRP of 750 ms. Notably, these protocols resulted only in partial erosion of the 2 % w/v phantom characterised by the formation of small lesions on the transverse plane having no observed length along the beam axis.

### 4.1.2 Effect of acoustic power on phantom erosion and lesion dimensions

Histotripsy exposures executed on the 2 % w/v agar-based phantom using varied applied acoustic power successfully generated lesions at all PRPs. The mechanically fractionated liquefied regions were visible on B-mode US images acquired post-exposures as indicatively shown in Figure 4.2 for the lesion created after application of an acoustic power of 210 W at the PRP of 750 ms. Lesions appeared on US images


Figure 4.2: B-mode ultrasound image of mechanical fractionation lesions in a 2% w/v agar phantom under PRP = 750 ms.

as tadpole-shaped hypoechogenic areas that could be easily delineated from the surrounding undamaged echogenic phantom, thus indicating the axial extent of the cavitation-induced damage within the phantom. Figure 4.3 shows photos of the dissected phantom after sonications delivered at the PRP of 750 ms, indicating the erosion on a plane perpendicular to the ultrasonic beam transmission. Applying varied

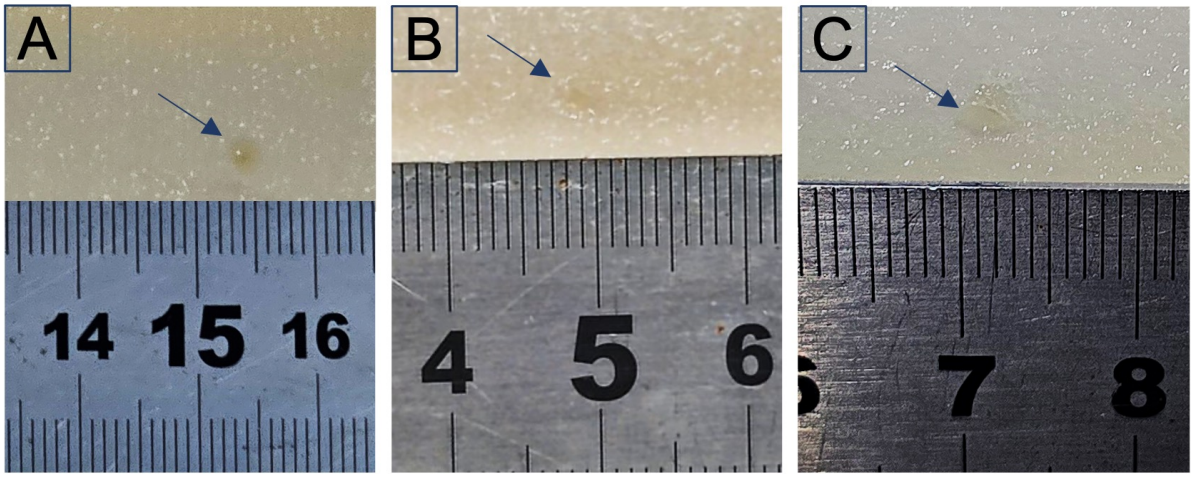



Figure 4.3: Visible light image of mechanical fractionation lesions in a 2% w/v agar phantom under PRP = 750 ms.

acoustic power of 200, 210, and 215 W resulted in the formation of sharply delineated circular void lesions at the focal point, hat were filled with liquid phantom debris, with no mechanical fractionation observed outside the liquefied regions as shown in Figure 4.3A, Figure 4.3B, and Figure 4.3C, respectively. Phantom dissection along the axial direction revealed the formation of characteristic tadpole-shaped lesions with clear erosion boundaries at all applied acoustic power as shown in Figure 4.4A, Figure 4.4B and

Figure 4.4C. The diameter and length of the histotripsy-induced lesions differed with the varied applied

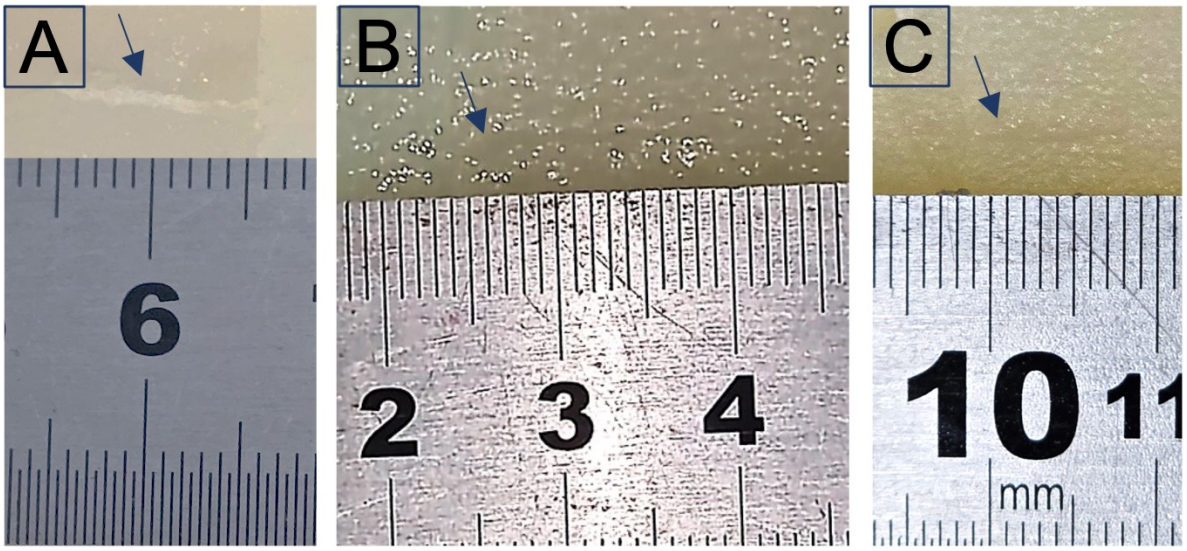



Figure 4.4: Visible light image of mechanically fractionated lesions in a 2% w/v agar phantom under PRP = 750 ms, shown in the axial plane.

acoustic power at all PRPs as shown in Figure 4.5A and Figure 4.5B, respectively, with different erosion trends observed with increasing power at the two PRPs. Increasing the acoustic power from 200 W to 210 W for the 1000-burst sonications delivered at the PRP of 500 ms resulted in a decrease in the lesion diameter from 3 mm to 2 mm as shown in Figure 4.5A, but an identical lesion length of 4 mm as shown in Figure 4.5B. Further increasing the acoustic power to 215 W resulted in the same lesion diameter of 2 mm (Figure 4.5A), and an increased lesion length of 7.5 mm (Figure 4.5B). Regarding exposures that were performed at the increased PRP of 750 ms using a higher number of pulses, an increase in both the lesion diameter from 2 mm to 4 mm, and lesion length from 10 mm to 13 mm was observed for an increase in the acoustic power from 200 W to 210 W as shown in Figure 4.5A and Figure 4.5B, respectively. However, further increasing the acoustic power to 215 W, resulted in an approximately similar erosion diameter (3.5 mm) but a decreased lesion length of 9 mm.

### 4.1.3 Effect of applied pulses on phantom erosion and lesion dimensions

The effect of the number of applied histotripsy pulses on the diameter and length of the fractionated lesions formed in the 2% w/v agar phantom is shown in Figure 4.6A and Figure Figure 4.6B, respectively. At the PRP of 250 ms, no erosion was found for sonications delivered with 200 or 500 pulses. A fractionation was formed for a higher number of applied pulses of 1000, however this 2 mm-wide lesion had no significant length (Figure 4.6B). At the PRP of 500 ms, lesions were successfully formed at all varied applied pulses. Notably, applying 200, 500 or 1000 pulses resulted in an identical lesion diameter of 2 mm (Figure 4.6A). However, when increasing the pulses from 200 to 500, the lesion length decreased from 4 mm to 2 mm, and then increased to 7.5 mm with a further increase of the applied pulses to 1000 (Figure 4.6B). Sonications that were delivered at the highest PRP of 750 ms also created phantom fractionation at all applied pulses. The lesion diameter remained constant at 2 mm when pulses of 200 and 500 were applied and increased with the application of a greater number of pulses. Lesions with observed

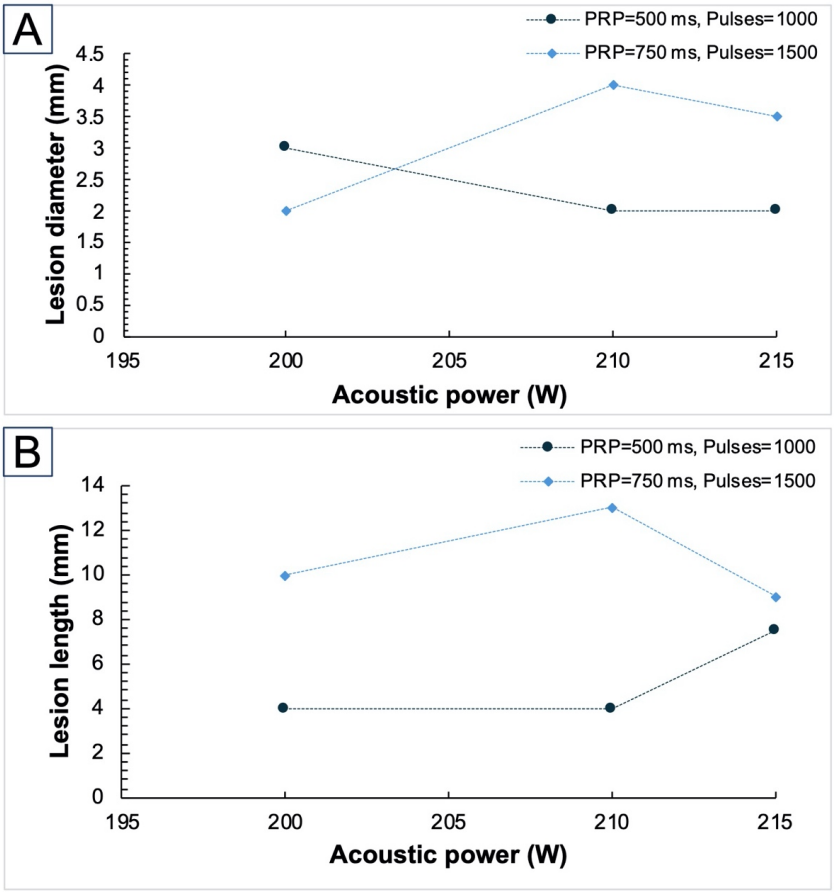



Figure 4.5: Tissue erosion under different acoustic power levels.(a) Relationship between acoustic power and erosion diameter.(b) Relationship between acoustic power and erosion length.

lengths ranging from 2 to 9 mm were formed when 500 or more histotripsy pulses were applied, with the lesion length increasing with an increased number of pulses of 500 to 1500.

### 4.1.4 Effect of PRP on phantom erosion and lesion dimensions

Figure 4.7A and Figure 4.7B show the effects of the varied PRP of 250 to 750 ms on the diameter and length of the histotripsy-induced lesions in the 2% w/v agar-based phantom, respectively, for the three different histotripsy conditions of 200 to 1000 pulses. Notably, at the smallest PRP of 250 ms, an erosion with a 2 mm diameter and no length was formed only when the highest pulse count of 1000 was applied. Increasing the PRP to 500 ms whilst increasing the pulse duration from 5 ms to 10 ms resulted in fractionations that had both diameter and length for all three pulse protocols. Similar erosion observations were made at the increased PRP of 750 ms, however, when 200 pulses were applied an erosion that had a 2 mm diameter and no significant extent was formed (Figure 4.7B). Noteworthy, for the PRPs of 500 and 750 ms, lesions with identical diameters of 2 mm were formed for all applied pulses, with the only exception of an erosion having a slightly increased diameter of 3 mm which was formed when 1000 pulses were delivered at the 750 ms PRP (Figure 4.7A). As seen in Figure 4.7B, when increasing the PRP from 500 ms to 750 ms, a consistent lesion length of 2 mm was produced for 500 pulses. However, for a higher pulse count of 1000, a decreased erosion length from 7.5 mm to 4 mm was observed for an increase in

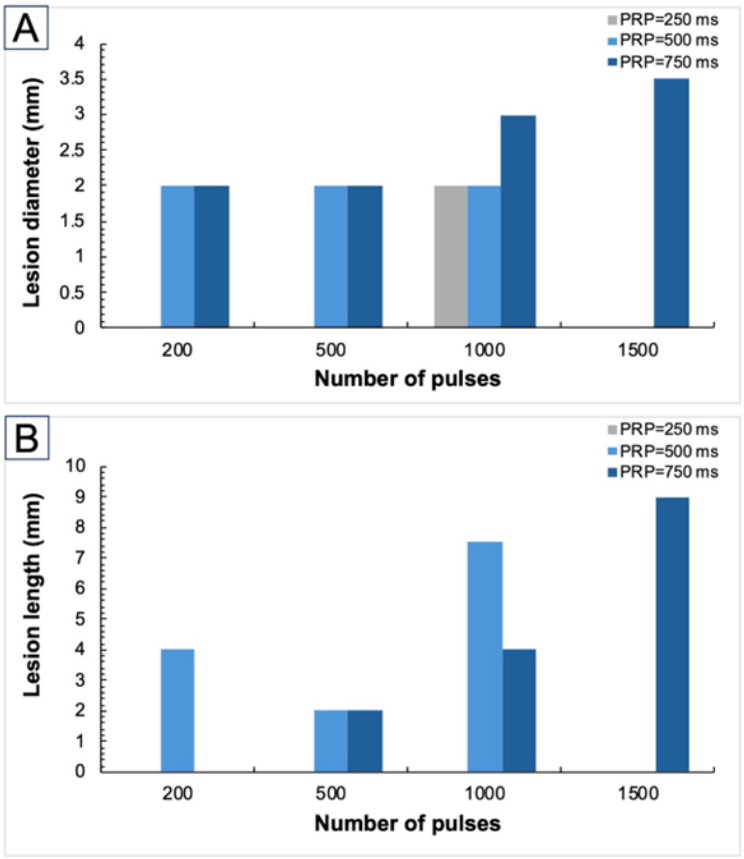



Figure 4.6: Tissue erosion under different number of pulses.(a) Relationship between number of pulses and erosion diameter.(b) Relationship between number of pulses and erosion length.

the PRP from 500 ms to 750 ms.

### 4.1.5 MRI-based assessment of Effect of acoustic power on phantom erosion and lesion dimensions

The MRI cross-sectional and sagittal views of fractionated lesions formed in a 2% w/v agarose phantom under different applied acoustic power levels are shown in Figure 4.8A and Figure 4.8B. These images clearly illustrate the erosion of the phantom under different applied acoustic power levels. Notably, no erosion was observed in the phantom following sonication at acoustic power levels of 110 W or 140 W. However, when an acoustic power of 170 W was applied, erosion became apparent, and the erosion length increased with higher acoustic power levels. Measurements obtained from the MRI images were used to generate plots depicting the relationship between different applied acoustic power levels and the erosion diameter and length, as shown in Figure 4.9A and Figure 4.9B. These plots further confirm that the erosion length increases with higher applied power levels, while the erosion diameter also increases but gradually stabilizes as the applied power increases. The effects of different applied acoustic power levels on the phantom, as observed under visible light imaging, are shown in Figure 4.10. From the visible light images, it is evident that erosion in the phantom is not detectable under an applied power of

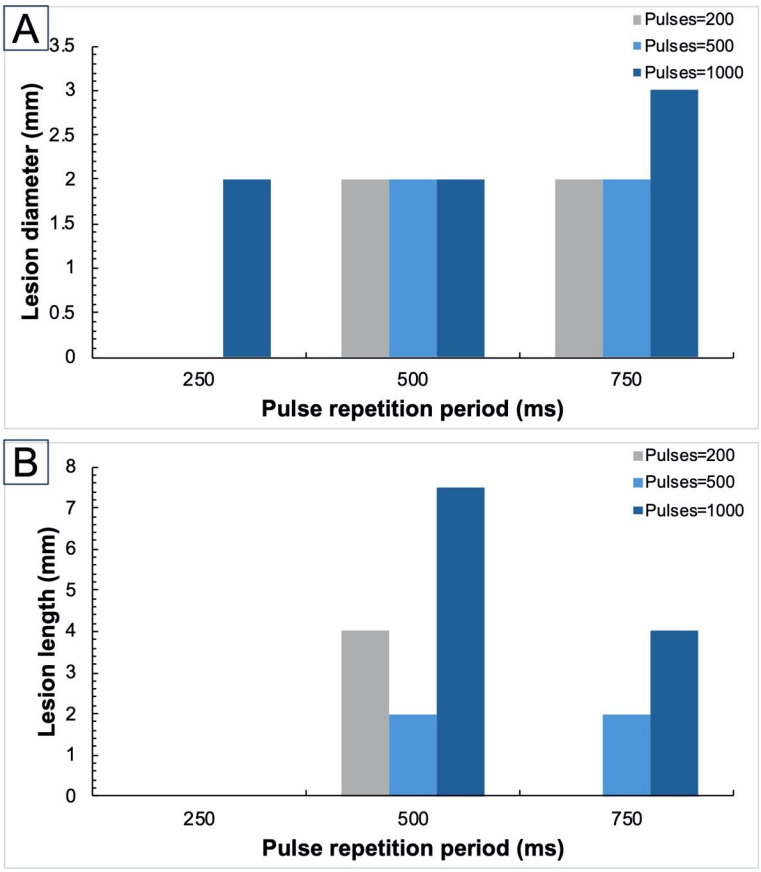



Figure 4.7: Tissue erosion under different PRP.(a) Relationship between PRP and erosion diameter.(b) Relationship between PRP and erosion length.

170 W. However, erosion becomes clearly visible at applied power levels of 200 W and 220 W.

### 4.1.6 MRI-based assessment of effect of applied pulses on phantom erosion and lesion dimensions

The MRI cross-sectional and sagittal views of fractionated lesions formed in a 2% w/v agarose phantom under varying numbers of histotripsy pulses are shown in Figure 4.11A and Figure 4.11B. These images clearly illustrate the relationship between the number of applied pulses and the erosion diameter and length. Notably, no erosion was observed after sonication with 100 or 200 pulses. However, erosion became evident when 300 pulses were applied. It is noteworthy that the lesion diameter and length increased progressively with the number of pulses. Measurements derived from the MRI images were used to generate the plots of pulse number versus erosion diameter and length, as shown in Figure 4.12A and Figure 4.12B. These plots further confirm that the erosion diameter and length increase with the number of pulses. As shown in Figure 4.13, an example of a photomicrograph obtained under visible light imaging demonstrates the occurrence of erosion. However, compared to MRI images, it is challenging to observe clear contours in visible light imaging. This indirectly highlights the advantage of MRI imaging in providing a clear view of erosion formation, making it particularly suitable for assessing tissue

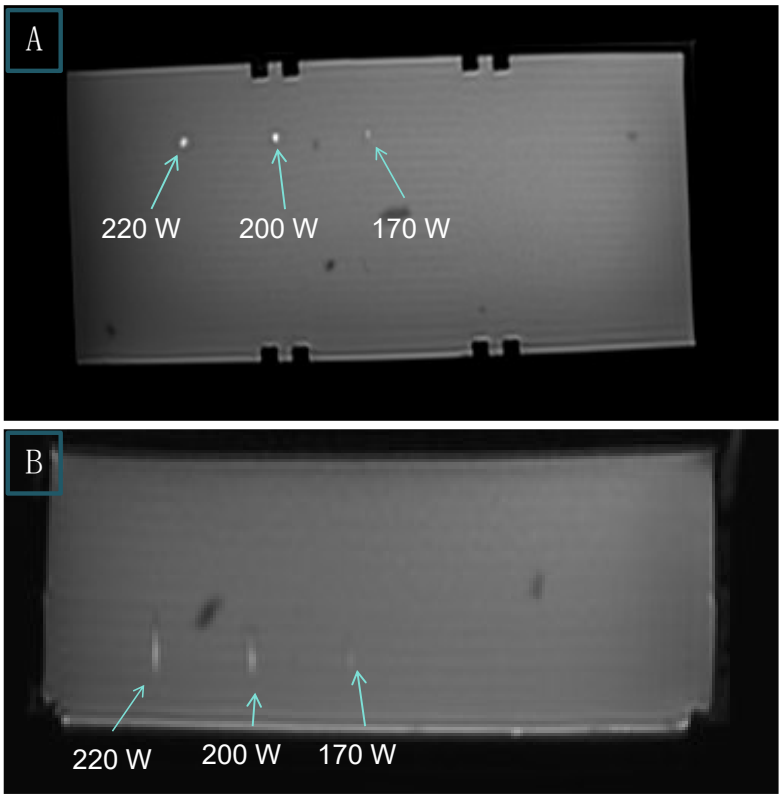



Figure 4.8: MRI cross-sectional and sagittal images of fragmentation lesions formed in a 2% w/v agarose model under different acoustic power levels. The numbers in the images represent the acoustic power.

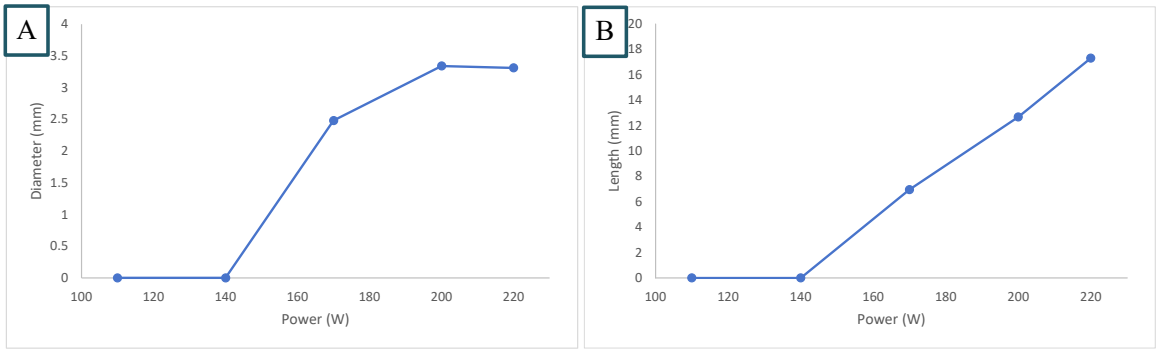



Figure 4.9: Tissue erosion under different acoustic power levels.(a) Relationship between acoustic power and erosion diameter.(b) Relationship between acoustic power and erosion length.

liquefaction caused by fractionation post-treatment. Although visible light imaging makes it difficult to discern erosion, it has the advantage of revealing internal details of the erosion that are not readily visible in MRI imaging.

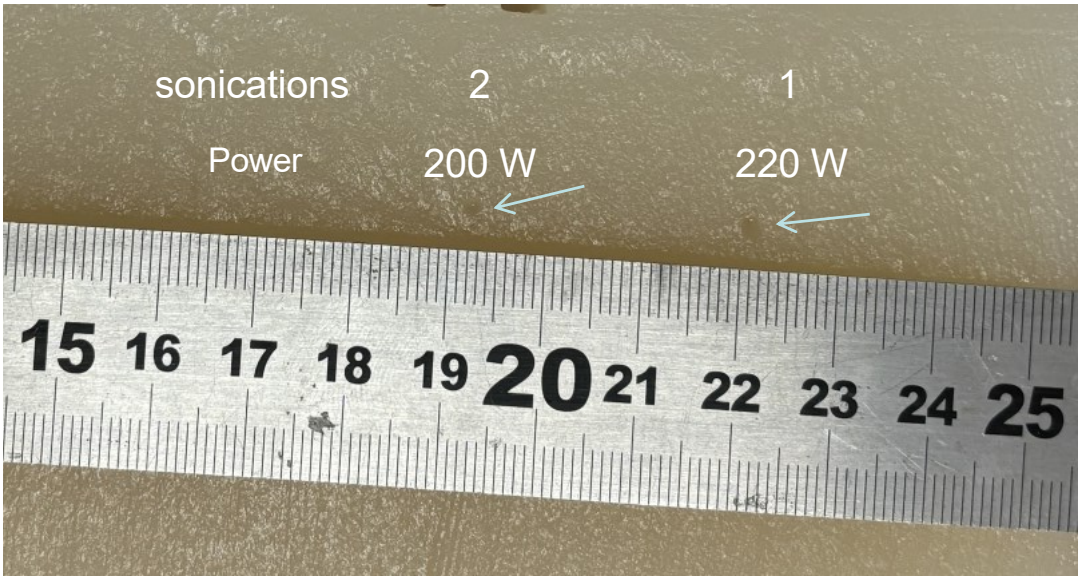



Figure 4.10: Visible light imaging of cavitation lesions formed in 2% w/v agarose models under different acoustic power levels.

### 4.1.7 MRI-based assessment of effect of duty factor on phantom erosion and lesion dimensions

The MRI cross-sectional and sagittal images of fractionated lesions in a 2% w/v agarose phantom under various duty factor conditions are presented in Figure 4.14A and Figure 4.14B. The MRI images clearly demonstrate the erosion of the phantom under different duty factor sonication conditions. It is worth noting that no erosion was observed in the phantom after sonication with a 1% duty factor. However, when a duty factor of 2% was applied, erosion became evident, and the erosion length increased progressively with higher duty factors up to the maximum applied value of 5%. The relationship between various duty factors and the erosion diameter and length was established using measurements from the MRI images, as illustrated in Figure 4.15A and Figure 4.15B. As shown in Figure 4.15A, it is further confirmed that the erosion length increases with higher duty factors. Additionally, Figure 4.15B demonstrates that the erosion diameter gradually stabilizes at 3.5 mm as the duty factor increases. Figure 4.16 presents the erosion conditions of the phantom under visible light imaging. It is evident from the images that erosion occurs in the phantom under different duty factors, further validating the effectiveness of visible light imaging for analyzing and studying histotripsy.

### 4.1.8 MRI-based assessment of effect of PRP on phantom erosion and lesion dimensions

The sagittal T2-W TSE image of the mechanical fractionation induced within the 2 % w/v agar-based phantom resulting histotripsy exposures delivered at the PRP of 750 ms is shown in Figure 4.17A. Similarly, Figure 4.17B shows the acquired sagittal T2-W TSE image of the five histotripsy-induced lesions formed after pulsed sonications executed at the 1000 ms PRP. Resulting lesions formed by both varying

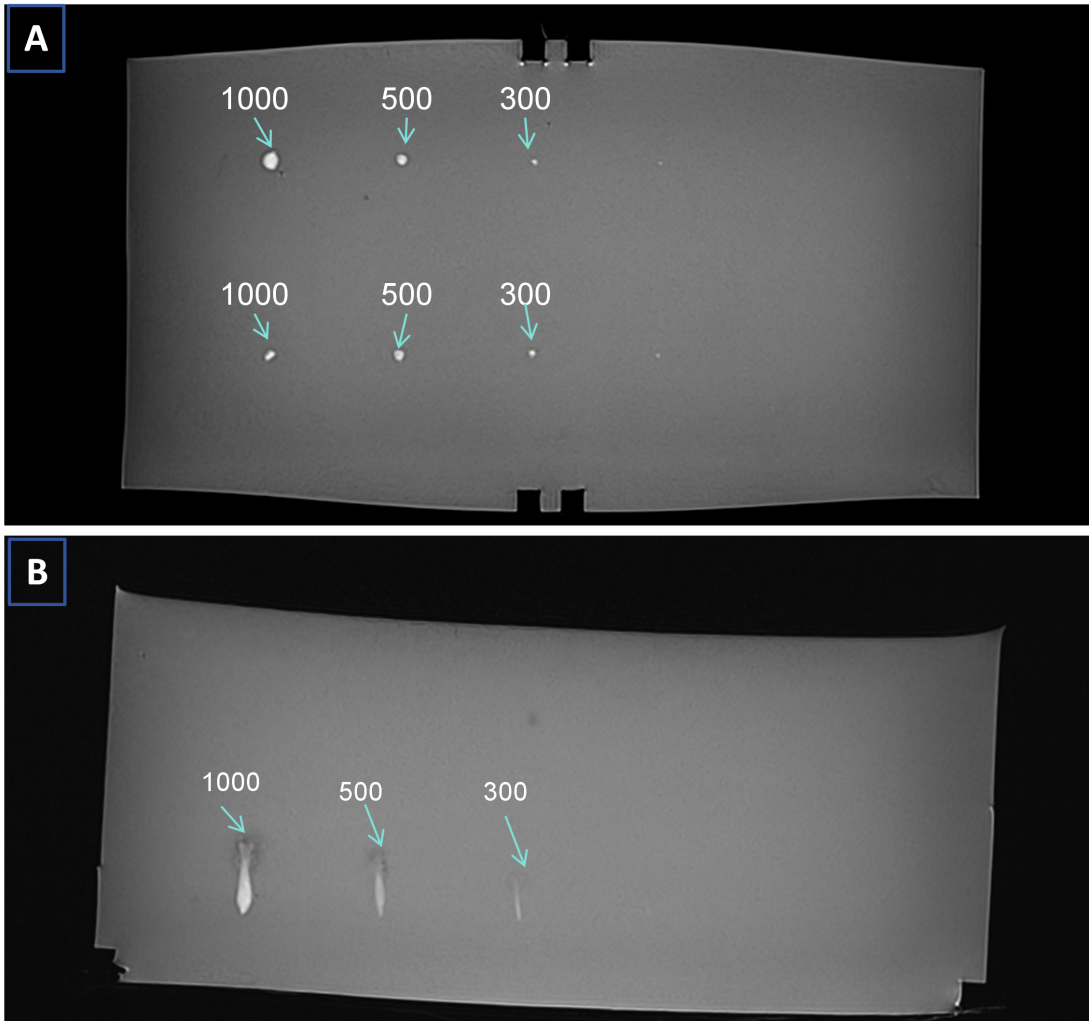



Figure 4.11: MRI cross-sectional and sagittal images of fragmentation lesions formed in a 2% w/v agarose model under different numbers of tissue fragmentation pulses. The numbers in the images represent the number of pulses.

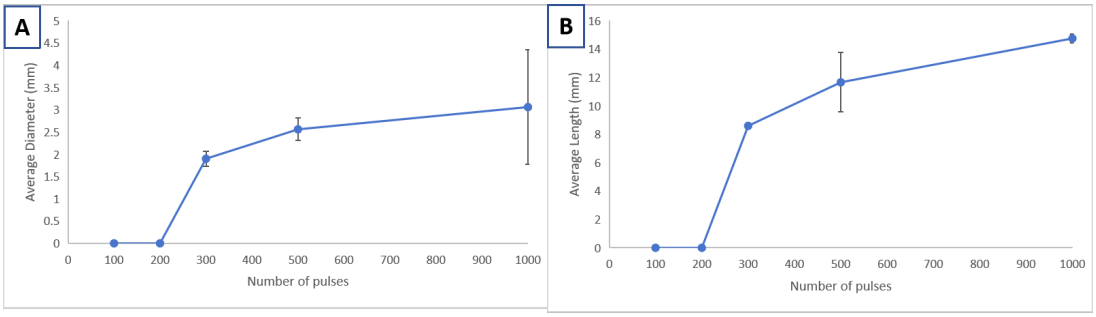



Figure 4.12: Relationship between the number of pulses and the average erosion diameter and average erosion length.

PRP protocols, appeared as tadpole-shaped hyperintense areas surrounded by a hypointense rim at the end of the lesion tail. The diameters and lengths of the histotripsy lesions inflicted by sonications executed at the PRP of 750 ms as measured from the T2-W TSE image, were between 2.25-2.36 mm and 14.25-15.45 mm, respectively, resulting in an average diameter of 2.31 ± 0.05 mm and an average erosion

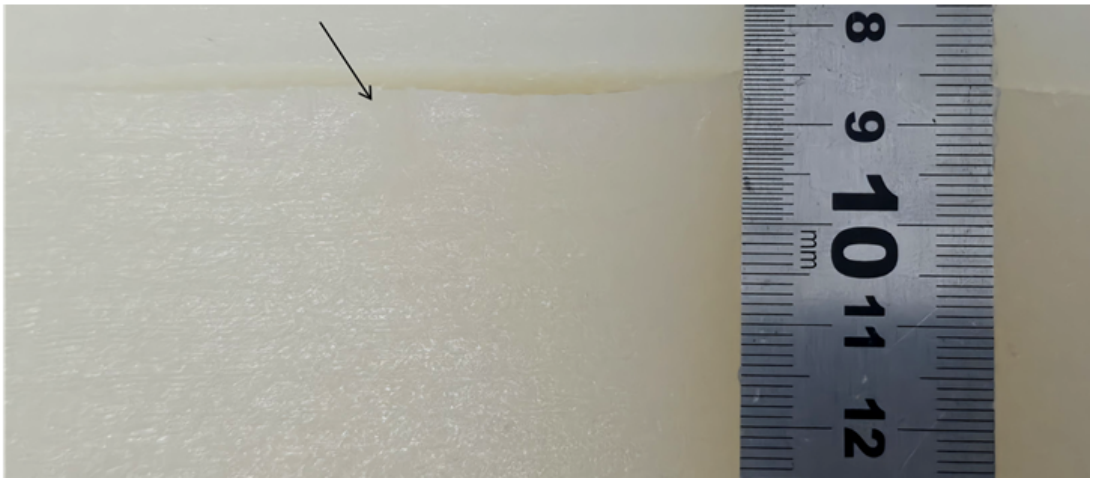



Figure 4.13: Examples of visible light imaging.

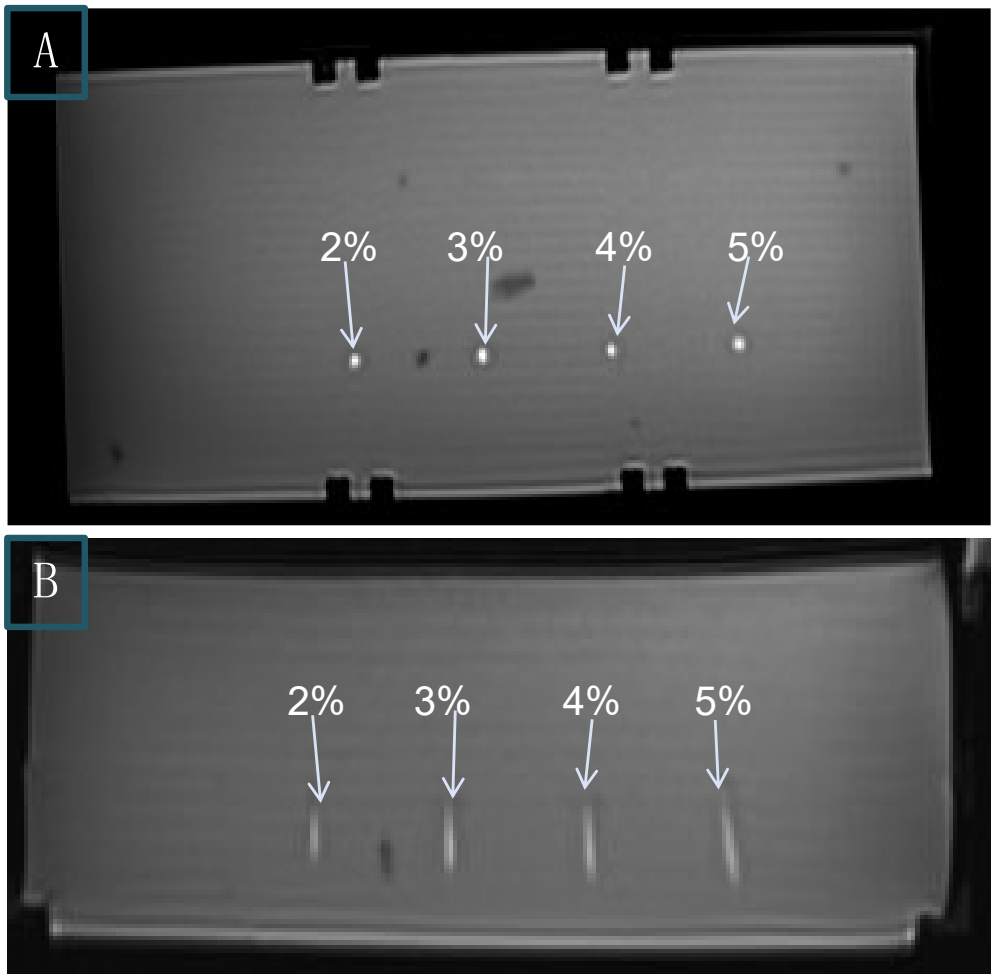



Figure 4.14: MRI cross-sectional and sagittal images of fragmentation lesions formed in a 2% w/v agarose model under different duty factor. The annotations in the figure represent different duty factors.

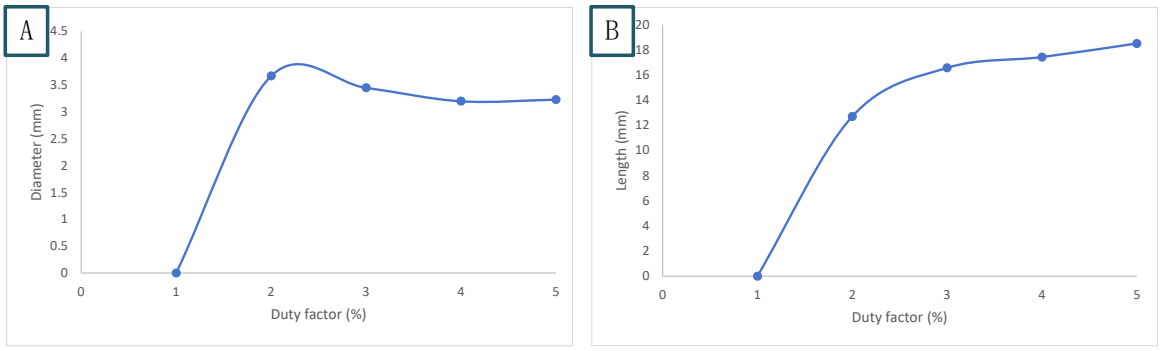



Figure 4.15: Relationship between the duty factor and the erosion diameter and erosion length.

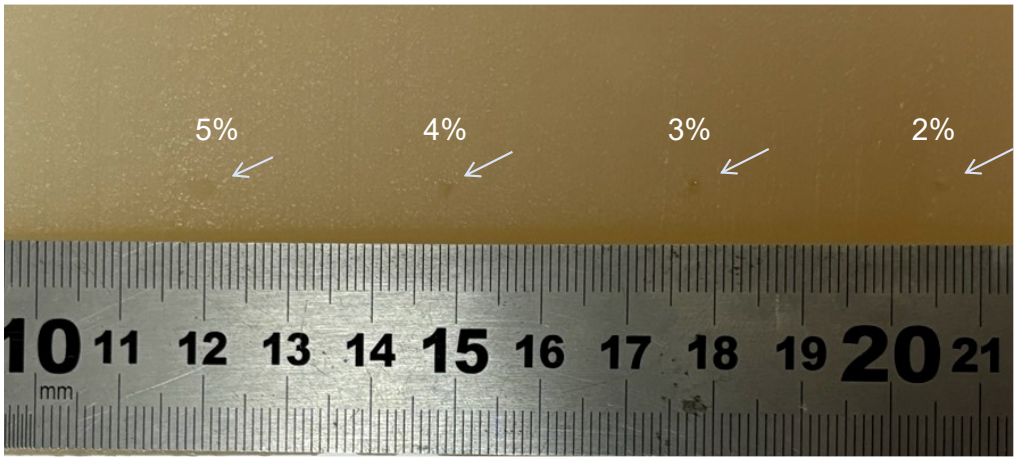



Figure 4.16: Visible light imaging of cavitation lesions formed in 2% w/v agarose models under different duty factors.

length of 15.04 ± 0.49 mm. T2-W TSE measurements of the erosion diameters created by the PRP of 1000 ms resulted in individual values between 2.36-3.17 mm, indicating an average diameter of 2.89 ± 0.32 mm. Accordingly, the five individual lesions had a length in the range of 9.98-16.22 mm, resulting in an average length of 13.94 ± 2.96 mm.

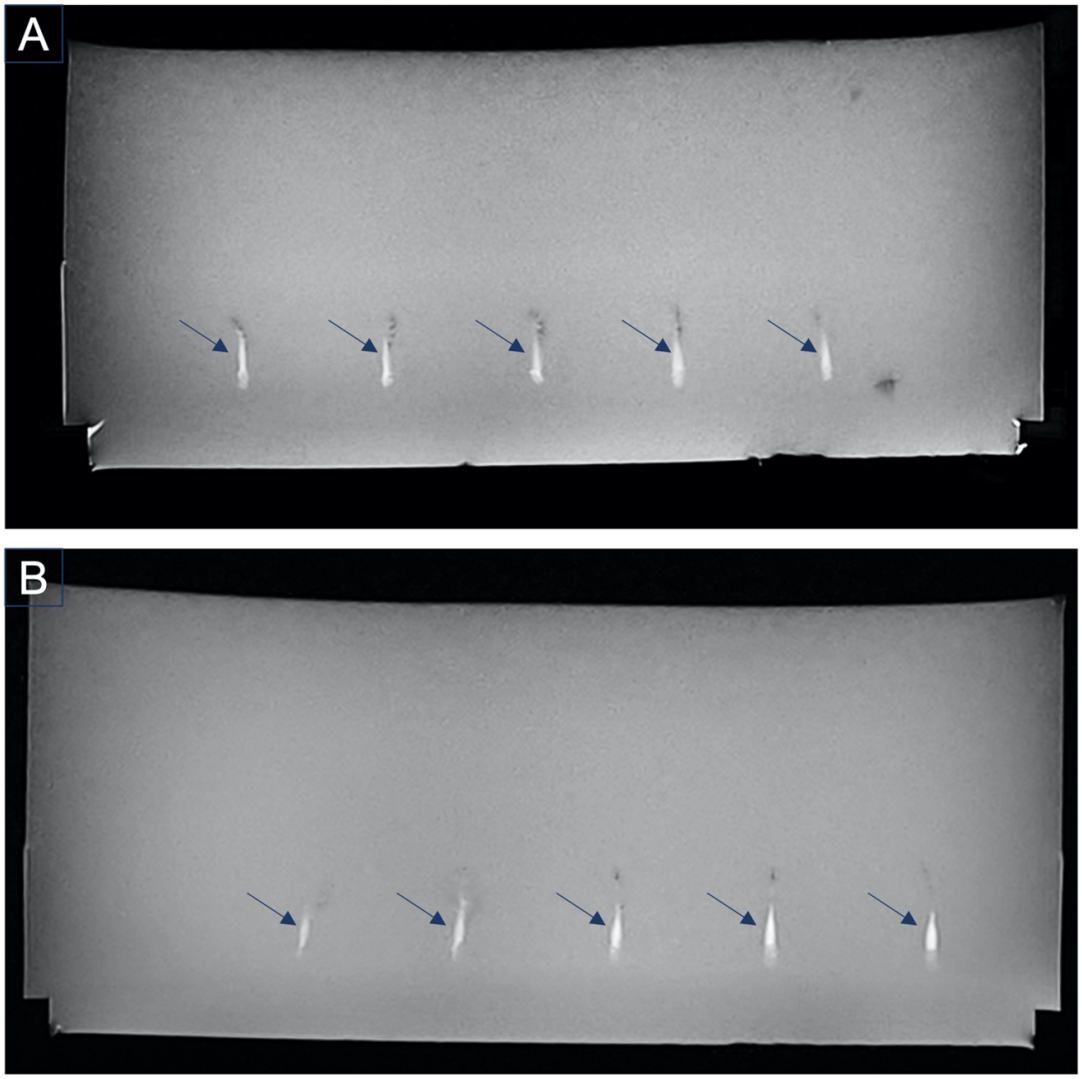



Figure 4.17: (a) Sagittal MRI image of tissue fragmentation exposure under PRP = 750 ms. (b) Sagittal MRI image of tissue fragmentation exposure under PRP = 1000 ms.

## 4.2 Multimodal Medical Image Fusion Results

### 4.2.1 Experimental Setup of Our Method

#### 4.2.1.1 Dataset

The training and test datasets is from the publicly available Harvard medical dataset [67] which provides normal and pathological brain images in multiple modalities such as MRI, CT, and PET. Since our work focuses on addressing the problem of multimodal medical image fusion, we utilized 184 pairs of CT and MRI images and 264 pairs of PET and MRI images from the Harvard medical dataset. To obtain sufficient training data, a random cropping strategy is employed for data augmentation, a commonly used approach in image fusion [44]. Therefore, all the images in the training set are cropped into patch pairs of size $64 \times 64$. As the cropping strategy is adopted for data augmentation, it is not employed for validation or testing, so that the entire image can be directly fed into the trained model to generate fusion results [68].

#### 4.2.1.2 Implementation Details

In the proposed DDPM-EMF, there are two networks which need to be trained in turn: (i) feature extractor and (ii) FER. We describe the details of training each network in turn. During the training process of the feature extractor, the AdamW optimizer was used with a learning rate of 0.00001, a batch size of 2, and 1,000,000 iterations. The GPU memory consumption was 8 Gigabytes (GB) for the PET-MRI task and 12 GB for the CT-MRI task. For the FER network, the Adam optimizer was used with a learning rate of 0.0001, a batch size of 8, and 300 epochs. The GPU memory consumption was 10 GB for the PET-MRI task and 5 GB for the CT-MRI task. All the experiments involved were carried out on a workstation containing the NVIDIA RTX A4000 GPU and 3.80 GHz Intel(R) Xeon(R) Gold 5320 CPU.

### 4.2.2 Comparison approaches and evaluation metrics

#### 4.2.2.1 Comparison approaches

We compared our proposed methods with seven state-of-the-art methods, including EMFusion [44], SwinFuse [69], GeSeNet [70], CDDFuse [71], DDFM [47], Diff-IF [48] and EMMA [72]. These methods are based on different foundational deep learning networks: EMFusion and GeSeNet are CNN-based image fusion networks; SwinFuse is a Transformer-based fusion network; CDDFuse combines CNN and Transformer; DDFM and Diff-IF are based on diffusion models; and EMMA is an end-to-end self-supervised learning paradigm for equivariant multimodal image fusion. It is worth noting that for PET and MRI image fusion, the SwinFuse, CDDFuse, DDFM and EMMA methods output grayscale images. Therefore, like most methods [73, 74], we used a mapping technique to convert these images into the RGB space, obtaining color images.

#### 4.2.2.2 Evaluation Metrics

To quantify the merits of our fusion results, we selected four quantitative evaluation metrics Edge Intensity (EI) [75], Spatial Frequency (SF) [76], Definition (DF) [77] and Average Gradient (AG) [78] to compare with other seven state-of-the-art methods. EI primarily represents the contrast intensity between edge

information and neighboring pixels. The higher the EI, the more prominent the edge detail information, resulting in clearer tissue structure. SF reflects the rate of grayscale variation in an image. A higher spatial frequency indicates a sharper image and better fusion quality. The calculation formula is as follows:

$$SF = \sqrt{RF^2 + CF^2}, \tag{4.1}$$

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|I(i,j) - I(i,j-1)|^2}, \tag{4.2}$$

$$CF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}|I(i,j) - I(i-1,j)|^2}, \tag{4.3}$$

Here, $I$ represents the fused image, while $M$ and $N$ denote the height and width of the image, respectively. DF represents the sharpness of an image, with higher DF values indicating a clearer image. AG can also be used to assess the sharpness of the fused image. A higher AG generally indicates better image clarity and higher fusion quality. The calculation formula is as follows:

$$\text{AG} = \frac{1}{(M-1)(N-1)}\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\sqrt{\frac{(H(i+1,j) - H(i,j))^2 + (H(i,j+1) - H(i,j))^2}{2}}. \tag{4.4}$$

The four evaluation metrics mentioned above are that a higher value indicates a higher quality of the fused image. For the PET-MRI task, we specifically used CIEDE2000 [53] to evaluate the color information of the fused images. CIEDE2000 is a color difference formula proposed by the International Commission on Illumination (CIE) in 2000, designed to more accurately quantify human visual perception of color differences. Specifically, the color discrepancies between each fused image and its corresponding source images were quantified using CIEDE2000, and their average was computed to yield the ACD. The formula for ACD is as follows:

$$\text{ACD} = \frac{\text{CIE}(I_1, I_F) + \text{CIE}(I_2, I_F)}{2}, \tag{4.5}$$

where $I_1$ and $I_2$ represent the original images, $I_F$ denotes the fused image, and $CIE(I_1, I_2)$ represents the color difference between two images calculated using the CIEDE2000 formula. A lower ACD signifies a closer alignment of the fused image's colors with those of the original images, thereby addressing the prevalent shortcoming of inadequate color evaluation in current medical image fusion methodologies.

### 4.2.3 CT-MRI Comparison Results

#### 4.2.3.1 Qualitative Analysis

The qualitative fusion results in three typical CT and MRI image pairs which are shown in Figure 4.18. In the results of EMFusion and SwinFusion, skeletal information in the CT images are weakened when making a comparison to other methods. In comparison, our results effectively preserve the skeletal information. Furthermore, compared with other methods, DDPM-EMF preserves more information, especially more texture details. Two examples are given in the first and second rows in Figure 4.18. As shown

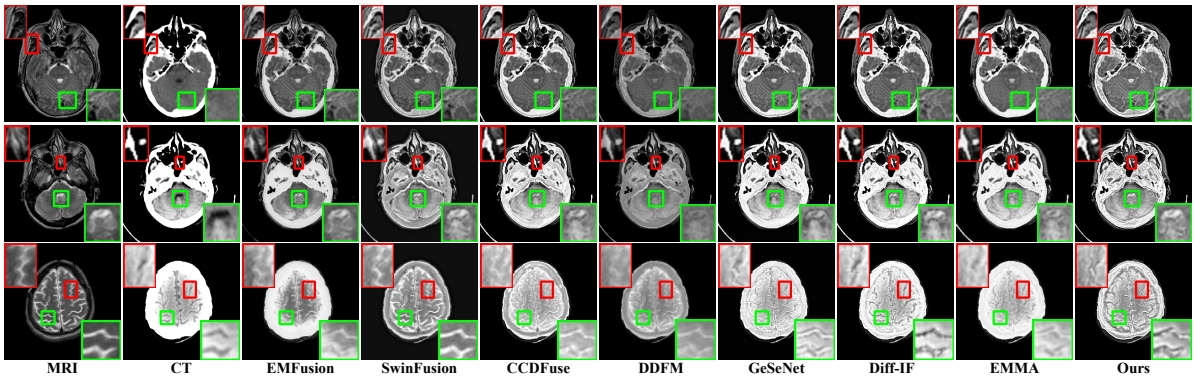



Figure 4.18: Qualitative results on fusion of three typical CT and MRI images. They are MRI images, CT images, and fused results of EMFusion, SwinFusion, CCDFuse, DDFM, GeSeNet, Diff-IF, EMMA and ours.

in the first row, the red boxes clearly indicate that the results from the EMFusion and the SwinFusion lead to slightly more skeletal loss, and the results from the DDFM method are overall darker. In contrast, our results effectively preserve the skeletal information. Additionally, the green boxes show that the detail information in our results is more prominent. Moreover, as in the second row, our method effectively preserves critical information from different modalities. Specifically, as seen in the green boxes, all methods except for EMFusion and our method exhibit varying degrees of tissue detail loss. Only EMFusion and our method display complete tissue detail information. While both methods successfully extract full tissue details, our method achieves better clarity. The same phenomenon is evident in the third row, where some competitors fail to achieve effective information retention. Especially in the red boxes, some methods did not effectively incorporate tissue information from MRI and CT images. Furthermore, in the green boxes and throughout the entire image, our method excellent clarity, enabling enhanced visualization of tissue details. Overall, our fusion results not only emphasize the distinct features of CT and MRI source images while effectively preserving complementary information from different modalities but also ensure comprehensive feature extraction, producing high-quality fused images with clearly visible details.

#### 4.2.3.2 Quantitative Analysis

Table 4.1: Quantitative comparison results of the proposed method with seven competitors on CT and MRI image fusion. On 25 test image pairs, the quantitative results of fusion results obtained by different fusion methods on four metrics are shown below (mean and standard deviation are shown, red: optimal, blue: suboptimal)

| Metrics | EMFusion | SwinFusion | CCDFuse | DDFM | GeSeNet | Diff-IF | EMMA | OURS |
|---|---|---|---|---|---|---|---|---|
| **EI** | 57.508±15.498 | 73.835±18.792 | 80.621±22.856 | 47.308±12.243 | 80.910±23.007 | 83.58±23.831 | 67.651±20.455 | 90.310±24.985 |
| **SF** | 21.432±3.783 | 24.931±4.915 | 34.868±7.215 | 19.684±3.711 | 34.342±6.807 | 36.614±7.445 | 25.808±5.228 | 35.681±6.910 |
| **DF** | 6.406±1.871 | 8.637±2.369 | 9.581±2.948 | 5.484±1.542 | 9.716±2.801 | 10.075±2.976 | 7.549±2.381 | 10.770±3.078 |
| **AG** | 5.521±1.551 | 7.225±1.933 | 7.989±2.367 | 4.631±1.253 | 8.030±2.324 | 8.324±2.438 | 6.508±2.017 | 9.021±2.556 |

For a more comprehensive comparison, we give the four aforementioned evaluation indicators, as shown in Table 4.1. Our fusion results achieved the best marks on EI, DF and AG, and the second best on SF. From the perspective of quantitative metrics, a higher EI value indicates that our generated results contain more edge detail information. This is because our EEDB enables the network to focus more on

edge details, enhancing its feature extraction capabilities and ensuring thorough extraction of detailed information. Additionally, by leveraging the powerful feature extraction and representation capabilities of DDPM, along with our design of FER to further enhance features, AG and DF achieved optimal results. This indicates that our images have higher quality and less distortion. Finally, the proposed method maintains excellent quantitative indicators while ensuring the visual quality of the fused images.

### 4.2.4 PET-MRI Comparison Results

#### 4.2.4.1 Qualitative Analysis

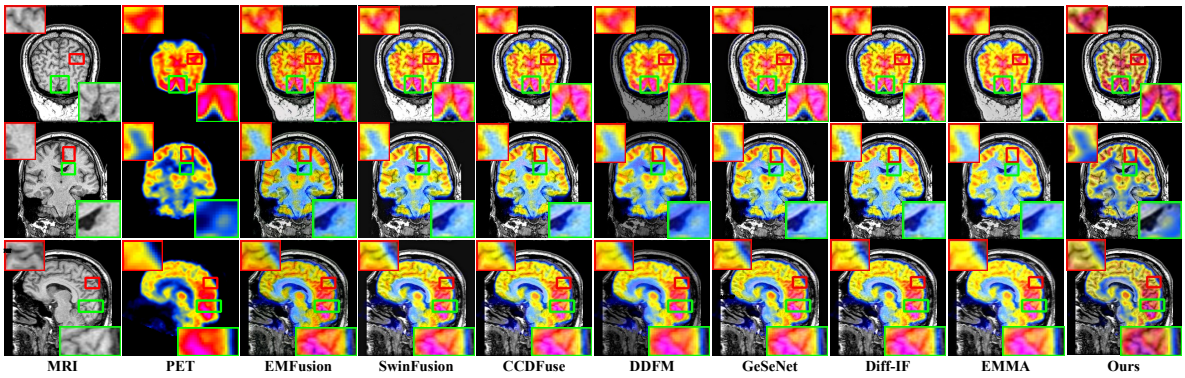



Figure 4.19: Qualitative results on fusion of three typical PET and MRI images. They are MRI images, PET images, and fused results of EMFusion, SwinFusion, CCDFuse, DDFM, GeSeNet, Diff-IF, EMMA and ours.

Three typical qualitative comparison results are shown in Figure 4.19. Our results have three typical advantages. The first point is that our method effectively inherits complementary information between different modalities. This is especially evident in the first and third rows, where our method effectively preserves the color information from the PET images and the tissue detail information from the MRI images, making them clearly visible. In contrast, due to the influence of color information from the PET images, it is difficult to clearly see the tissue detail information from the MRI images in the fusion results of other methods. For example, within the red and green boxes in these rows, our method effectively displays the texture details from the MRI images and the color information from the PET images. In contrast, in the results of other methods, although the color information from the PET images is visible, the texture details from the MRI images cannot be clearly distinguished. Second, our method effectively preserves the edge detail information of the PET images. As shown in the green box in the second row, the yellow spot within the green box in the PET image are not visible in the results of other methods, but are clearly displayed in the result of our method. Moreover, as shown within the red boxes, our method preserves the edge contour information of the PET image more clearly compared to other methods. Third, in our method, the feature extractor captures joint features from MRI and PET images, so the mosaic effect in the PET images is mitigated by the high-quality edge information from the MRI images, as shown in the green boxes in the first row. Overall, our fusion results not only effectively inherit information from multiple modalities but also exhibit excellent image quality.

Table 4.2: Quantitative comparison results of the proposed DDPM-EMF with seven competitors on PET and MRI image fusion. On 25 test image pairs, the quantitative results of fusion results obtained by different fusion methods on five metrics are shown below (mean and standard deviation are shown, red: optimal, blue: suboptimal)

| Metrics | EMFusion | SwinFusion | CCDFuse | DDFM | GeSeNet | Diff-IF | EMMA | OURS |
|---|---|---|---|---|---|---|---|---|
| **EI** | 103.340±11.589 | 114.830±13.253 | 115.275±13.231 | 72.820±9.203 | 113.551±13.045 | 116.968±13.761 | 97.962±10.541 | 123.295±14.503 |
| **SF** | 33.462±4.384 | 36.465±4.952 | 36.889±5.047 | 22.747±3.482 | 35.548±4.402 | 37.211±5.003 | 28.377±2.791 | 39.138±5.212 |
| **DF** | 12.735±1.877 | 14.102±2.112 | 14.173±2.135 | 8.892±1.405 | 13.870±1.971 | 14.490±2.191 | 11.216±1.365 | 14.797±2.132 |
| **AG** | 10.247±1.279 | 11.362±1.444 | 11.398±1.450 | 7.144±0.976 | 11.170±1.385 | 11.599±1.498 | 9.443±1.069 | 12.165±1.540 |
| **ACD** | 7.422±1.170 | 10.294±1.369 | 9.012±1.535 | 6.805±0.974 | 9.001±1.288 | 9.175±1.586 | 9.097±1.608 | 6.938±0.932 |

#### 4.2.4.2 Quantitative Analysis

Quantitative comparison results about the PET-MRI fusion task are shown in Table 4.2. The statistical results showed that our method achieved the highest average values in EI, SF, DF, and AG. This indicates that our approach transfers more useful information from the source images, providing researchers with richer texture details and greater clarity. For the ACD metric, the average value of our fusion results achieved the second-best score. This is because, in the fusion results of our method, the MRI color information affects the comparison with PET colors. In contrast, in the results obtained by the DDFM method, the MRI colors are darker, which has a smaller impact on the color comparison. This also demonstrated that our method performs well in preserving the color information of the original images.

### 4.2.5 Ablation study

To validate the effectiveness of our model design, we conducted ablation experiments focused on three key components: the EEDB in the feature extractor, the FER module, and the multi-channel joint learning strategies specifically for color and grayscale medical image fusion tasks such as PET-MRI image fusion. First, we will prove the module we designed.We employ the same metrics as CT-MRI and PET-MRI tasks to evaluate the performance of ablation experiments. As shown in Tables 4.3 and 4.4, no configuration

Table 4.3: Quantitative Comparison Results of the Ablation Study on Whether to Use the Designed Module in CT-MRI Image Fusion Tasks (Showing Mean Values; Red Indicates Optimal)

| EEDB | FER | DDPM | EI | SF | DF | AG |
|---|---|---|---|---|---|---|
| | | √ | 84.816 | 35.445 | 9.661 | 8.302 |
| √ | | √ | 89.211 | 33.871 | 10.227 | 8.762 |
| √ | √ | √ | 90.310 | 35.681 | 10.770 | 9.021 |

Table 4.4: Quantitative Comparison Results of the Ablation Study on Whether to Use the Designed Module in PET-MRI Image Fusion Tasks (Showing Mean Values; Red Indicates Optimal)

| FER | DDPM | EI | SF | DF | AG | ACD |
|---|---|---|---|---|---|---|
| | √ | 122.513 | 38.958 | 14.406 | 12.025 | 7.177 |
| √ | √ | 123.295 | 39.138 | 14.797 | 12.165 | 6.938 |

achieved the same level of performance as our full model with all modules in place. Specifically, the increase in EI values confirms that our modules better retain fine details within the fused images, while improvements across other parameters indicate overall enhanced image quality, underscoring the effectiveness of EEDB and FER. Furthermore, Figure 4.20 illustrates that the fusion results with EEDB and

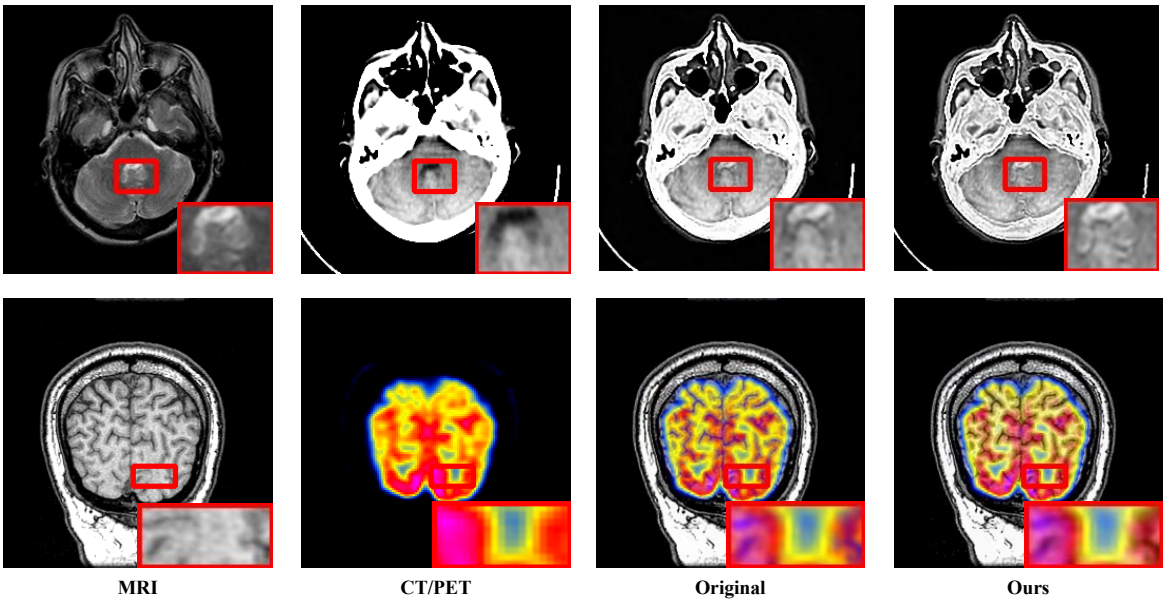



Figure 4.20: Qualitative results of the DDPM-EMF under different ablation experiments. The first row represents the CT-MRI fusion task, while the second row represents the PET-MRI fusion task.

FER preserve more details and yield higher-quality images compared to those without these modules. Additionally, we assessed the effectiveness of the multi-channel joint learning strategy. As illustrated in

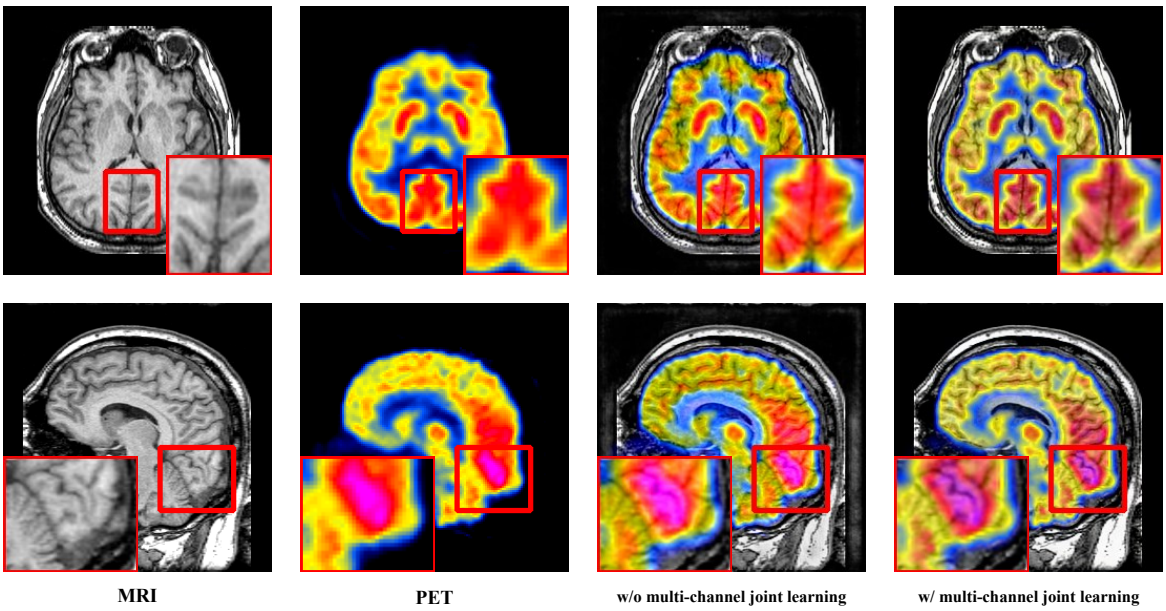



Figure 4.21: Qualitative Results of DDPM-EMF: Impact of Multi-Channel Joint Learning Method.

Figure 4.21, this approach improves complementary information inheritance across modalities. For example, in the red boxes in the first and second rows, the fusion results with multi-channel joint learning display more MRI texture details and clearer structural information than the results without it. Moreover, as shown in Table 4.5, the results obtained using the multi-channel joint learning method exhibit

Table 4.5: Quantitative Comparison Results of the Ablation Study on Whether to Use the Multi-Channel Joint Learning Method (Showing Mean Values; Red Indicates Optimal)

| multi-channel joint learning method | ACD |
|---|---|
| | 7.57 |
| √ | 6.94 |

lower ACD values, indicating that the color information of the fused images is closer to that of the source images. This further confirms that the strategy enhances the retention of complementary information in color and grayscale image fusion tasks.

Additionally, we further validated the generalizability of the model by selecting 17 pairs of Single-Photon Emission Computed Tomography (SPECT) and MRI images from the Harvard Medical Dataset [67] and directly testing them using the network trained on the PET-MRI task without any fine-tuning. Some of the obtained results are shown in Figure 4.22. SPECT and PET are two nuclear medicine-based CT techniques. Since both methods image the gamma rays emitted from within the patient's body, they are collectively referred to as Emission Computed Tomography (ECT). The fundamental imaging principle of

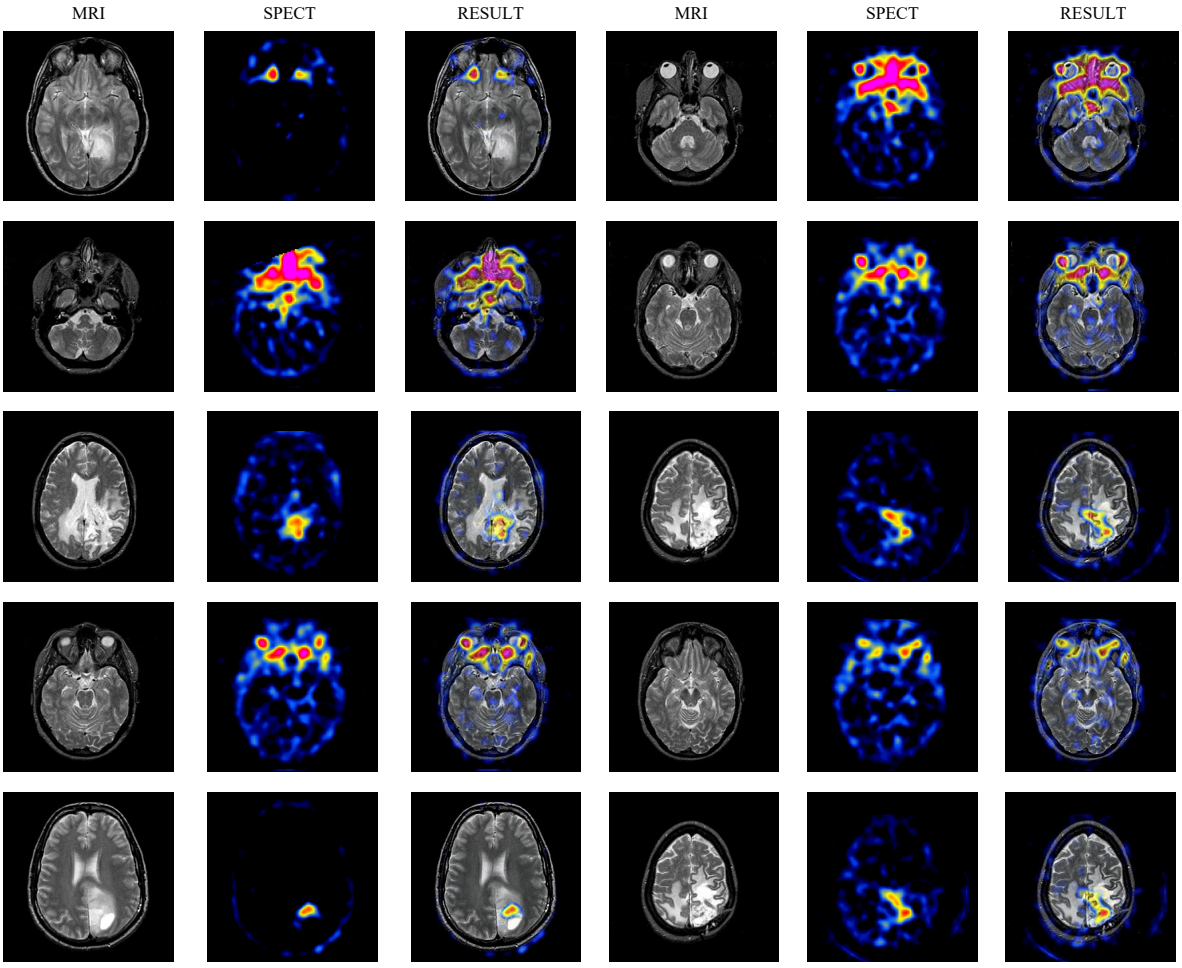



Figure 4.22: Partial Fusion Results for the SPECT-MRI Image Fusion Task.

SPECT is as follows: First, the patient must ingest or be injected with a radiopharmaceutical containing a

radioactive isotope with an appropriate half-life. Once the radiopharmaceutical reaches the target imaging slice, gamma photons are emitted due to radioactive decay. The gamma camera probe detects these photons along specific projection lines (rays) using its sensitive points. A scintillator converts the detected high-energy gamma rays into a large number of lower-energy light signals, which are then transformed into electrical signals and amplified by a photomultiplier tube. The measured values represent the total radioactivity along the projection line in the human body.

The sensitive points along the same projection line detect the distribution of the radiopharmaceutical in a particular slice of the body, producing a one-dimensional projection of that slice. In the imaging setup, all projection lines are perpendicular to the detector and parallel to each other, forming what is known as a parallel beam. The angle between the detector's normal and the X-axis is referred to as the viewing angle. Since the gamma camera is a two-dimensional detector, when equipped with a parallel-hole collimator, it can simultaneously acquire parallel-beam projections of multiple slices, forming a planar image. However, planar images do not provide depth information along the projection lines. To determine the internal structure of the body in the depth direction, multiple projections from different angles are required. It can be mathematically proven that if the one-dimensional projections of a given slice are obtained at all viewing angles, the original slice image can be reconstructed. The process of deriving the slice image from its projections is known as reconstruction. Since this imaging technique relies heavily on computational processing, it is referred to as CT.

The primary advantage of SPECT imaging is its ability to provide dynamic functional information about organs and tissues within the body, making it highly valuable for disease diagnosis and treatment planning. For example, in cardiology, SPECT can be used to assess myocardial blood flow, helping to determine whether the myocardium is ischemic or to delineate the extent of an acute myocardial infarction. In oncology, SPECT can be employed to evaluate tumor metabolism, allowing for the identification of tumor location, size, and activity level. Another key advantage of SPECT is its capability for quantitative analysis. By measuring pixel intensity values in the images, it is possible to calculate the distribution of the radiopharmaceutical within the body. This quantitative assessment is crucial for evaluating disease severity and guiding treatment decisions. However, SPECT imaging also has certain limitations. First, since it involves the use of radiopharmaceuticals, there is some level of radiation exposure to the patient. Second, SPECT has relatively low spatial resolution, which may limit its accuracy in detecting small organs or lesions. Lastly, as a functional imaging technique, SPECT does not provide anatomical structural details of the body. Therefore, in the SPECT-MRI fusion task, the fused image should simultaneously preserve the functional information reflected in the SPECT image and the anatomical structural details provided by the MRI image. Our fusion results effectively achieve this goal. For example, in the two sets of images in the third row, the SPECT image reveals a distinct area of high metabolism in the brain (indicated by yellow-to-red tones). While the MRI image displays local anatomical structures, it is insufficient on its own to determine the nature of the abnormality. However, the high metabolic activity observed in the SPECT image—such as significantly increased metabolism associated with tumor cell proliferation—suggests a high likelihood of a neoplastic brain lesion (e.g., a malignant tumor, as tumor tissue metabolism is often higher than that of normal brain tissue). In our fusion results, the high-metabolism regions are precisely aligned with the brain structures, allowing both functional metabolic information from SPECT and anatomical details from MRI to be clearly visualized. This provides clinicians and researchers with

a more comprehensive view, aiding in accurate diagnosis and medical decision-making. Additionally, other results also demonstrate that our fusion framework effectively preserves both the functional information represented by the SPECT colors and the anatomical structural details provided by MRI. This validates the effectiveness of our model and its ability to achieve high-quality fusion.

# 5 Conclusion

This article is summarized as follows.

In the histotripsy experiments, a single-element 2.6 MHz FUS transducer was used to investigate the effects of specific treatment parameters on the mechanical erosion of agar-based phantoms. Through a series of benchtop experiments and multimodal medical imaging, we precisely analyzed the impact of agar concentration, acoustic power, pulse number, duty factor, and PRP on lesion size. This study aims to fill the gap in histotripsy research regarding the influence of ultrasound treatment parameters on lesion formation and size in agar-based phantoms.

First, ultrasound treatment was applied to agar-based phantoms with different agar concentrations to investigate its effect on phantom fractionation. No histotripsy-induced fractionation was observed in tissue-mimicking phantoms developed with 2.5% or 3% w/v agar concentration. Although the mechanical strength of the fabricated phantoms was not measured in this study, previous research has shown that higher agar concentrations result in greater mechanical strength [79]. Therefore, our results confirm the dependence of fractionation efficacy on agar concentration and, ultimately, on the mechanical strength of the agar-based phantoms. The absence of erosion in the two stiffer phantoms further highlights the role of mechanical properties in initiating cavitation events and the resulting mechanical fractionation. This is consistent with previous findings indicating that erosion size is inversely related to the mechanical strength of the phantom [24, 27, 30].

Additionally, according to existing literature, media with higher mechanical stiffness require high-intensity ultrasound exposure to induce cavitation events [28]. Thus, the insufficient fractionation observed in our study suggests that the applied acoustic power (129 W) may not have been sufficient to facilitate bubble nucleation, indicating that higher acoustic power should be applied to phantoms with higher agar concentrations. Future studies could investigate the degree of fractionation induced by histotripsy protocols using higher acoustic power in stiffer phantoms.

In contrast, successful lesion formation was achieved in the softer phantom (2% w/v agar), indicating that models developed with lower agar concentrations are more susceptible to histotripsy-induced damage. Moreover, the lesion diameter ranged between 2–5 mm and increased gradually with PRP. However, the observed partial phantom erosion (without axial fractionation) suggests that the cavitation-induced mechanical strain was limited for the adopted protocol, possibly due to the relatively low applied acoustic power. Nevertheless, since the 2% w/v agar phantom exhibited lower resistance to histotripsy-induced fractionation, it was prioritized as the ultrasound treatment target in subsequent experiments investigating the effects of histotripsy treatment parameters on phantom erosion.

Interestingly, the dimensions of the histotripsy-induced lesions differed with the applied acoustic power, with notable differences observed across the two PRPs (500, and 750 ms). At the lower PRP of 500 ms, increasing the acoustic power from 200 W to 210 W reduced the lesion diameter by approximately 0.7-fold, without affecting lesion length (4 mm), while further increasing the acoustic power to 215 W maintained a consistent erosion diameter (2 mm) but increased the lesion length (7.5 mm) by approximately 1.9-fold. Following linear regression analysis, a strong negative correlation ($R_2$=0.8929) was observed between lesion diameter and acoustic power, while a moderate positive correlation ($R_2$=0.5724) was noticed for

lesion length and acoustic power. In contrast, at the higher PRP of 750 ms, both lesion diameter and length initially increased by 2-fold and 1.3-fold, respectively, with increased acoustic power from 200 W to 210 W, indicating enhanced cavitation activity. Although a 0.7-fold reduction in lesion length was observed for further increases in the acoustic power, no correlation was found between lesion length and acoustic power ($R_2$=0.0027) at the increased PRP. Contrary, the relation between erosion diameter and power was moderate and positive ($R_2$=0.7033). Unlike previous studies that showed that the acoustic power does not impact the total volume of the eroded phantom lesions [7], herein results indicate that the acoustic power differently influences cavitation dynamics in the transverse and axial focal directions. For example, when increasing acoustic power at low PRPs, the axial extent of lesions can be controlled, while applying higher acoustic powers at higher PRPs favours a broader lateral erosion. Moreover, the fact that at the highest acoustic power (215 W) similar lesion lengths were recorded for both PRPs (7.5 mm for 500 ms and 9 mm for 750 ms) potentially implies that a limit in lateral erosion efficiency might exist at higher acoustic powers. Therefore, future studies are needed to examine any potential saturation points that may limit the axial extension of fractionation. Nevertheless, since the highest acoustic power (215 W) generated damages with sufficient diameter and length, it was maintained for sonications performed in the following experiments to ensure that any observed differences in phantom erosion would be directly attributed to the time-dependent treatment parameters (e.g., number of pulses, and PRP), and not to acoustic power differences.

Differences in phantom erosion were also observed after the exposures that were executed using a varied number of pulses (200-1500), with distinct effects observed across the varying PRPs (250, 500, and 750 ms). At the PRP of 250 ms, a minimum threshold of 1000 pulses was required to induce a partial erosion with no axial extent within the phantom, suggesting that low pulse counts are insufficient for significant tissue erosion at short PRPs, probably as a result of inadequate cavitation activity. Therefore, in future studies, pulse counts greater than 1000 should be employed for efficient phantom fractionation at low PRPs, corroborating previous studies that revealed that high pulses are needed to enable erosion at high PRFs (low PRPs) [80, 81]. In contrast, higher PRPs (500 ms and 750 ms) facilitated phantom erosion across all pulse numbers, suggesting improved cavitation effects. A consistent lesion diameter (2 mm) was noticed for 200 and 500 pulses, irrespective of the PRP, while it increased for a greater pulse count at the highest PRP (750 ms). Specifically, at the PRP of 750 ms, lesion diameter was found to linearly increase ($R_2$=0.9448) with a higher pulse count, demonstrating that cumulative mechanical energy can increase the efficiency of lateral erosion. Although a 2-fold decrease in lesion length was observed at the 500 ms PRP when increasing the pulses from 200 to 500 (length of 4 mm reduced to 2 mm), further increases observed in the axial lesion extent for a greater number of pulses suggest that the lesion length linearly increases with the pulse count ($R_2$=0.5377). This was further substantiated from the lesions formed resulting exposures at the 750 ms PRP. Specifically, a strong positive linear relationship was observed ($R_2$=0.9638) between lesion length and number of pulses, suggesting that enhanced axial erosion can be achieved for a higher pulse count at prolonged sonications probably due to cumulative mechanical stress. These results are in accordance with previous studies executed on various phantoms that demonstrated that higher histotripsy pulses enable a greater spatial distribution of the fractionation effect [7, 82].

Above, we analyzed the results of histotripsy using visible light imaging and ultrasound imaging. Sub-

sequently, we combined MRI imaging with visible light imaging for further analysis of histotripsy outcomes. Notably, this is the first time MRI has been used to accurately visualize tissue damage caused by histotripsy.

Sonications were performed at varied acoustic power to investigate its effect on the extent of phantom fractionation. At low acoustic powers (110-140 W) no histotripsy-induced fractionation was observed on acquired T2-W images, indicating that these power levels were insufficient to initiate cavitation events. MR images revealed a threshold acoustic power of 170 W for effective formation of lesions with both lateral (2.48 mm) and axial (6.96 mm) extents, with subsequent acoustic power increases, increasing both lesion dimensions. Increasing the acoustic power from 200 to 220 W resulted in lesions with approximately similar diameter ( 3.3 mm) and about 1.35-fold increased length. The positive correlations observed between acoustic power, and lesion diameter, and length align with prior phantom studies wherein greater erosion volumes were observed with increased peak negative pressures, and thus electronic driving voltages of the histotripsy pulses [83, 84].

The critical role of pulse count in the formation, and dimensions of the histotripsy-induced lesions was also examined for sonications delivered at pulse counts ranging from 100 to 1000. Complete phantom fractionation with both axial, and lateral extents was effectively achieved with 300 pulses, providing insights on a pulse count threshold for effective mechanical fractionation. Beyond this threshold, lesion morphology along the beam axis shifted to a tadpole-like structure with wider heads at higher pulse counts, suggesting enhanced cavitation dynamics, and pronounced erosions in the proximal regions of the focal zone. These findings are consistent with a previous study which reported wider lateral sizes of the heads of histotripsy-induced lesions with higher pulse counts due to the formation of broader cavitation clouds [85]. Quantitative T2-W TSE lesion dimension measurements provided additional insights into the effect of pulse count on phantom fractionation, revealing an approximately 1.6-fold increase in average lesion diameter, and a 1.7-fold increase in lesion length for a pulse count increase from 300 to 1000. The significant linear correlations found between pulse count, and both lesion dimensions demonstrate that higher pulse counts can enhance the efficiency of erosion in both focal directions. Similar patterns of lesion growth have been previously reported in phantoms with an increasing number of histotripsy pulses [7, 85]corroborating herein findings.

Notable differences in lesion size were also observed with the duty factor of the sonications. T2-W TSE imaging, and phantom sectioning after exposures demonstrated that phantom fractionation was not achieved at the lowest duty factor (1%), likely due to limited cavitation effects generated by the low treatment dose. Contrary, all tested duty factors above 1% successfully induced phantom erosion, as evidenced by the hyperintense areas surrounded by a hypointense border on T2- W TSE images, and the structural changes observed in the phantom with macroscopic inspection. Moreover, the damages observed in the sectioned phantom correlated with the location of the hypointense-bordered hyperintense areas on T2-W images, demonstrating the capability of these MR sequences in accurately detecting histotripsy-induced fractionations.

Dimension measurements showed that while lesion diameter remained relatively consistent across all duty factors, lesion length increased with increasing duty factor, resulting in a 1.46- fold difference between duty factors of 2%, and 5%. Linear correlations were observed between duty factor, and lesion diameter, and length, suggesting a potential relationship between duty factor, and erosion size as indicated

in previous studies [86, 87].

Finally, the effects of different PRPs of 750 and 1000 ms were examined under the 1000 pulse count condition. T2-W TSE lesion dimension measurements revealed that fractionations generated at the lowest PRP (750 ms) had approximately a 0.8-fold decreased average diameter and about a 1.08-fold increased average length relative to the corresponding dimensions of the lesions created by the higher PRP (1000 ms). These findings suggesting that higher PRPs (lower PRFs) improve the spatial distribution of phantom erosion in the lateral direction but lead to a decrease in lesion length are consistent with the results of previous histotripsy studies wherein decreases in lesion volume were noticed with lower PRFs [7, 23, 26]. Furthermore, they provide insights on the distinct impact of PRP on the erosion dimensions along both focal directions, enhancing existing literature that described the influence of PRF on the size of the entire focal region [7, 23, 26].

Overall, this study focuses on analyzing the effects of key ultrasound parameters in histotripsy on lesion formation within agarose models using multimodal medical imaging. The findings demonstrate that acoustic power, pulse number, duty factor, and PRF can independently influence lesion size and morphology. By employing various imaging modalities to meticulously observe the impact of histotripsy on agarose models, this study provides a precise evaluation of the relationship between key ultrasound parameters and the degree of erosion in the agarose model. To the best of the authors' knowledge, this is the first study to investigate these parameter effects using T2-weighted MRI. Lesions were effectively detected in post-treatment images, confirming that the T2-weighted sequence can accurately characterize histotripsy outcomes. These findings validate the practicality of MRI in assessing histotripsy-induced lesions and provide critical insights for optimizing ultrasound treatment protocols to achieve effective mechanical fractionation in vitro, making a valuable contribution to the field of preclinical histotripsy research.

Medical image fusion plays a crucial role in clinical diagnosis and treatment by integrating complementary information from different imaging modalities to generate more informative and visually comprehensive fused images. However, existing fusion algorithms suffer from several limitations, including insufficient feature extraction, poor inheritance of complementary information, and inadequate evaluation of color information in color and grayscale image fusion tasks. To address these challenges, we propose a novel fusion method that significantly enhances the quality and effectiveness of multimodal medical image fusion.

Our proposed method aims to overcome these limitations by employing advanced feature extraction mechanisms to ensure that key details in the fused image are preserved. Conventional fusion methods often fail to extract sufficient features from multimodal medical images, resulting in loss of details. However, our method demonstrates superior feature extraction capabilities, effectively capturing complex details from the source images while ensuring a high level of texture preservation in the final fused image. This is particularly evident in CT-MRI fusion, where our method successfully preserves high-resolution structural details from both CT and MRI images.

Furthermore, we introduce a multi-channel joint learning strategy to improve the inheritance of complementary information between different imaging modalities. This strategy is particularly effective in fusing color and grayscale medical images, where traditional methods often struggle to maintain a balance between the two modalities. By leveraging this strategy, our approach minimizes information loss and

enhances the quality of the final fusion results. This is particularly evident in PET-MRI fusion, where our method successfully retains the vibrant color information from PET images while maintaining the high-resolution structural details of MRI images.

To further enhance the effectiveness of our fusion approach, we introduce the CIEDE2000 color difference formula and design the Average Color Difference (ACD) metric. This metric comprehensively evaluates the retention and accuracy of color information in fused images, addressing the longstanding issue of inadequate color evaluation in multimodal medical image fusion. Since color plays a pivotal role in the diagnosis of diseases in color medical imaging, the integration of this metric provides a more accurate and holistic assessment of the fusion quality.

Our fusion framework is built upon the diffusion probabilistic model (DDPM) in combination with an enhanced edge detail block (EEDB) for feature extraction. This design enables the model to extract a broader spectrum of relevant features, ensuring improved edge preservation and overall image clarity. The extracted features are then processed through a feature enhancement and reconstruction (FER) module, which reconstructs the final fused image while amplifying essential feature representations. Additionally, to ensure that the fused images fully inherit the complementary information from multiple modalities, we design a set of joint loss functions specifically tailored to the distinct characteristics of each imaging modality. This ensures optimal integration of information while mitigating the loss commonly observed in multimodal fusion tasks.

Extensive experiments have been conducted to evaluate the performance of our proposed method, including an ablation study on the SPECT-MRI image fusion task using our trained parameters directly. The results demonstrate that our approach significantly outperforms existing state-of-the-art fusion methods, achieving superior feature extraction, enhanced detail preservation, and improved color information retention. The fused images generated by our method exhibit higher clarity, richer textures, and greater vividness, making them more useful for clinical applications and diagnostic purposes.

Looking ahead, we plan to expand the application of our proposed fusion method to real-world clinical scenarios. For example, in the histotripsy experiment, we combined CT and MRI images to detect cavitation in tissue fragmentation lesions to provide a better basis for this experiment. By integrating our method into the medical imaging workflow, we aim to further verify its practicality and effectiveness to assist medical professionals in obtaining more accurate and reliable diagnostic information. This advancement in multimodal medical image fusion has great potential for improving medical imaging analysis and ultimately improving patient care.

# BIBLIOGRAPHY


[1] Z. Xu, G. Owens, D. Gordon, C. Cain, and A. Ludomirsky, "Noninvasive creation of an atrial septal defect by histotripsy in a canine model," *Circulation*, vol. 121, no. 6, pp. 742–749, 2010.

[2] T. D. Khokhlova, Y.-N. Wang, J. C. Simon, B. W. Cunitz, F. Starr, M. Paun, L. A. Crum, M. R. Bailey, and V. A. Khokhlova, "Ultrasound-guided tissue fractionation by high intensity focused ultrasound in an in vivo porcine liver model," *Proceedings of the National Academy of Sciences*, vol. 111, no. 22, pp. 8161–8166, 2014.

[3] T. D. Khokhlova, G. R. Schade, Y.-N. Wang, S. V. Buravkov, V. P. Chernikov, J. C. Simon, F. Starr, A. D. Maxwell, M. R. Bailey, W. Kreider *et al.*, "Pilot in vivo studies on transcutaneous boiling histotripsy in porcine liver and kidney," *Scientific reports*, vol. 9, no. 1, p. 20176, 2019.

[4] T. L. Hall, J. B. Fowlkes, and C. A. Cain, "A real-time measure of cavitation induced tissue disruption by ultrasound imaging backscatter reduction," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 3, pp. 569–575, 2007.

[5] M. Hoogenboom, D. Eikelenboom, M. H. den Brok, A. Veltien, M. Wassink, P. Wesseling, E. Dumont, J. J. Fütterer, G. J. Adema, and A. Heerschap, "In vivo mr guided boiling histotripsy in a mouse tumor model evaluated by mri and histopathology," *NMR in Biomedicine*, vol. 29, no. 6, pp. 721–731, 2016.

[6] S. P. Allen, T. L. Hall, C. A. Cain, and L. Hernandez-Garcia, "Controlling cavitation-based image contrast in focused ultrasound histotripsy surgery," *Magnetic resonance in medicine*, vol. 73, no. 1, pp. 204–213, 2015.

[7] A. Eranki, N. Farr, A. Partanen, K. V. Sharma, H. Chen, C. T. Rossi, S. V. Kothapalli, M. Oetgen, A. Kim, A. H. Negussie *et al.*, "Boiling histotripsy lesion characterization on a clinical magnetic resonance imaging-guided high intensity focused ultrasound system," *PloS one*, vol. 12, no. 3, p. e0173867, 2017.

[8] Z. Xu, T. L. Hall, E. Vlaisavljevich, and F. T. Lee Jr, "Histotripsy: the first noninvasive, non-ionizing, non-thermal ablation technique based on ultrasound," *International Journal of Hyperthermia*, vol. 38, no. 1, pp. 561–575, 2021.

[9] M. A. Kisting, M. S. Jentink, M. G. Wagner, Z. Xu, J. L. Hinshaw, and P. F. Laeseke, "Imaging for targeting, monitoring, and assessment after histotripsy: a non-invasive, non-thermal therapy for cancer," *blood vessels*, vol. 10, pp. 15–21, 2023.

[10] V. A. Khokhlova, J. B. Fowlkes, W. W. Roberts, G. R. Schade, Z. Xu, T. D. Khokhlova, T. L. Hall, A. D. Maxwell, Y.-N. Wang, and C. A. Cain, "Histotripsy methods in mechanical disintegration of tissue: Towards clinical applications," *International journal of hyperthermia*, vol. 31, no. 2, pp. 145–162, 2015.

[11] J. E. Parsons, C. A. Cain, G. D. Abrams, and J. B. Fowlkes, "Pulsed cavitational ultrasound therapy for controlled tissue homogenization," *Ultrasound in medicine & biology*, vol. 32, no. 1, pp. 115–129, 2006.

[12] Z. Xu, T. D. Khokhlova, C. S. Cho, and V. A. Khokhlova, "Histotripsy: a method for mechanical tissue ablation with ultrasound," *Annual Review of Biomedical Engineering*, vol. 26, 2024.

[13] M. F. Iqbal, M. A. Shafique, M. A. Raqib, T. K. F. Ahmad, A. Haseeb, A. M. Mhjoob, and A. Raja, "Histotripsy: an innovative approach for minimally invasive tumour and disease treatment," *Annals of Medicine and Surgery*, vol. 86, no. 4, pp. 2081–2087, 2024.

[14] A. Hendricks-Wenger, J. Sereno, J. Gannon, A. Zeher, R. M. Brock, N. Beitel-White, A. Simon, R. V. Davalos, S. Coutermarsh-Ott, E. Vlaisavljevich *et al.*, "Histotripsy ablation alters the tumor microenvironment and promotes immune system activation in a subcutaneous model of pancreatic cancer," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 68, no. 9, pp. 2987–3000, 2021.

[15] E. A. Knott, J. F. Swietlik, K. C. Longo, R. F. Watson, C. M. Green, E. J. Abel, M. G. Lubner, J. L. Hinshaw, A. R. Smolock, Z. Xu *et al.*, "Robotically-assisted sonic therapy for renal ablation in a live porcine model: initial preclinical results," *Journal of Vascular and Interventional Radiology*, vol. 30, no. 8, pp. 1293–1302, 2019.

[16] L. Arnold, A. Hendricks-Wenger, S. Coutermarsh-Ott, J. Gannon, A. N. Hay, N. Dervisis, S. Klahn, I. C. Allen, J. Tuohy, and E. Vlaisavljevich, "Histotripsy ablation of bone tumors: Feasibility study in excised canine osteosarcoma tumors," *Ultrasound in medicine & biology*, vol. 47, no. 12, pp. 3435–3446, 2021.

[17] A. D. Hendricks, J. Howell, R. Schmieley, S. Kozlov, A. Simon, S. L. Coutermarsh-Ott, E. Vlaisavljevich, and I. C. Allen, "Histotripsy initiates local and systemic immunological response and reduces tumor burden in breast cancer," *The Journal of Immunology*, vol. 202, no. 1_Supplement, pp. 194–30, 2019.

[18] G. R. Schade, J. Keller, K. Ives, X. Cheng, T. J. Rosol, E. Keller, and W. W. Roberts, "Histotripsy focal ablation of implanted prostate tumor in an ace-1 canine cancer model," *The Journal of urology*, vol. 188, no. 5, pp. 1957–1964, 2012.

[19] J. R. Sukovich, C. A. Cain, A. S. Pandey, N. Chaudhary, S. Camelo-Piragua, S. P. Allen, T. L. Hall, J. Snell, Z. Xu, J. M. Cannata *et al.*, "In vivo histotripsy brain treatment," *Journal of neurosurgery*, vol. 131, no. 4, pp. 1331–1338, 2018.

[20] A. D. Maxwell, G. Owens, H. S. Gurm, K. Ives, D. D. Myers Jr, and Z. Xu, "Noninvasive treatment of deep venous thrombosis using pulsed ultrasound cavitation therapy (histotripsy) in a porcine model," *Journal of vascular and interventional radiology*, vol. 22, no. 3, pp. 369–377, 2011.

[21] E. Messas, A. IJsselmuiden, G. Goudot, S. Vlieger, S. Zarka, E. Puymirat, B. Cholley, C. Spaulding, A. A. Hagège, E. Marijon *et al.*, "Feasibility and performance of noninvasive ultrasound therapy in patients with severe symptomatic aortic valve stenosis: a first-in-human study," *Circulation*, vol. 143, no. 9, pp. 968–970, 2021.

[22] T. G. Schuster, J. T. Wei, K. Hendlin, R. Jahnke, and W. W. Roberts, "Histotripsy treatment of benign prostatic enlargement using the vortx rx system: initial human safety and efficacy outcomes," *Urology*, vol. 114, pp. 184–187, 2018.

[23] Z. Xu, A. Ludomirsky, L. Y. Eun, T. L. Hall, B. C. Tran, J. B. Fowlkes, and C. A. Cain, "Controlled ultrasound tissue erosion," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 51, no. 6, pp. 726–736, 2004.

[24] Z. Xu, J. B. Fowlkes, A. Ludomirsky, and C. A. Cain, "Investigation of intensity thresholds for ultrasound tissue erosion," *Ultrasound in medicine & biology*, vol. 31, no. 12, pp. 1673–1682, 2005.

[25] E. Vlaisavljevich, K.-W. Lin, A. Maxwell, M. T. Warnez, L. Mancia, R. Singh, A. J. Putnam, B. Fowlkes, E. Johnsen, C. Cain *et al.*, "Effects of ultrasound frequency and tissue stiffness on the histotripsy intrinsic threshold for cavitation," *Ultrasound in medicine & biology*, vol. 41, no. 6, pp. 1651–1667, 2015.

[26] J. Xu, T. A. Bigelow, and H. Lee, "Effect of pulse repetition frequency and scan step size on the dimensions of the lesions formed in agar by hifu histotripsy," *Ultrasonics*, vol. 53, no. 4, pp. 889–896, 2013.

[27] J. Xu and T. A. Bigelow, "Experimental investigation of the effect of stiffness, exposure time and scan direction on the dimension of ultrasound histotripsy lesions," *Ultrasound in medicine & biology*, vol. 37, no. 11, pp. 1865–1873, 2011.

[28] E. Vlaisavljevich, A. Maxwell, M. Warnez, E. Johnsen, C. A. Cain, and Z. Xu, "Histotripsy-induced cavitation cloud initiation thresholds in tissues of different mechanical properties," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 61, no. 2, pp. 341–352, 2014.

[29] E. Vlaisavljevich, Y. Kim, G. Owens, W. Roberts, C. Cain, and Z. Xu, "Effects of tissue mechanical properties on susceptibility to histotripsy-induced tissue damage," *Physics in Medicine & Biology*, vol. 59, no. 2, p. 253, 2013.

[30] J. Xu, T. A. Bigelow, G. Davis, A. Avendano, P. Shrotriya, K. Bergler, and Z. Hu, "Dependence of ablative ability of high-intensity focused ultrasound cavitation-based histotripsy on mechanical properties of agar," *The Journal of the Acoustical Society of America*, vol. 136, no. 6, pp. 3018–3027, 2014.

[31] S. A. Hendley, V. Bollen, G. J. Anthony, J. D. Paul, and K. B. Bader, "In vitro assessment of stiffness-dependent histotripsy bubble cloud activity in gel phantoms and blood clots," *Physics in Medicine & Biology*, vol. 64, no. 14, p. 145019, 2019.

[32] Y. Yang, Y. Que, S. Huang, and P. Lin, "Multimodal sensor medical image fusion based on type-2 fuzzy logic in nsct domain," *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3735–3745, 2016.

[33] M. A. Azam, K. B. Khan, S. Salahuddin, E. Rehman, S. A. Khan, M. A. Khan, S. Kadry, and A. H. Gandomi, "A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics," *Computers in biology and medicine*, vol. 144, p. 105253, 2022.

[34] P. J. Burt and E. H. Adelson, "Merging images through pattern decomposition," in *Applications of digital image processing VIII*, vol. 575. SPIE, 1985, pp. 173–181.

[35] J. Du, W. Li, B. Xiao, and Q. Nawaz, "Union laplacian pyramid with multiple features for medical image fusion," *Neurocomputing*, vol. 194, pp. 326–339, 2016.

[36] X. Li, F. Zhou, H. Tan, W. Zhang, and C. Zhao, "Multimodal medical image fusion based on joint bilateral filter and local gradient energy," *Information Sciences*, vol. 569, pp. 302–325, 2021.

[37] J. Du, W. Li, and H. Tan, "Three-layer image representation by an enhanced illumination-based image fusion method," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1169–1179, 2019.

[38] N. Nagaraja Kumar, T. Jayachandra Prasad, and K. S. Prasad, "An intelligent multimodal medical image fusion model based on improved fast discrete curvelet transform and type-2 fuzzy entropy," *International Journal of Fuzzy Systems*, vol. 25, no. 1, pp. 96–117, 2023.

[39] P. Gupta and N. Jain, "Anisotropic diffusion filter based fusion of nsst transformed medical images," *Biomedical Signal Processing and Control*, vol. 90, p. 105819, 2024.

[40] P.-H. Dinh, "Multi-modal medical image fusion based on equilibrium optimizer algorithm and local energy functions," *Applied Intelligence*, vol. 51, no. 11, pp. 8416–8431, 2021.

[41] Y. Liu, X. Chen, J. Cheng, and H. Peng, "A medical image fusion method based on convolutional neural networks," in *2017 20th international conference on information fusion (Fusion)*. IEEE, 2017, pp. 1–7.

[42] X. Liang, P. Hu, L. Zhang, J. Sun, and G. Yin, "Mcfnet: Multi-layer concatenation fusion network for medical images fusion," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7107–7119, 2019.

[43] D. S. Shibu and S. S. Priyadharsini, "Multi scale decomposition based medical image fusion using convolutional neural network and sparse representation," *Biomedical Signal Processing and Control*, vol. 69, p. 102789, 2021.

[44] H. Xu and J. Ma, "Emfusion: An unsupervised enhanced medical image fusion network," *Information Fusion*, vol. 76, pp. 177–186, 2021.

[45] J. Ma, H. Xu, J. Jiang, X. Mei, and X.-P. Zhang, "Ddcgan: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4980–4995, 2020.

[46] J. Huang, Z. Le, Y. Ma, F. Fan, H. Zhang, and L. Yang, "Mgmdcgan: medical image fusion using multi-generator multi-discriminator conditional generative adversarial network," *IEEE Access*, vol. 8, pp. 55 145–55 157, 2020.

[47] Z. Zhao, H. Bai, Y. Zhu, J. Zhang, S. Xu, Y. Zhang, K. Zhang, D. Meng, R. Timofte, and L. Van Gool, "Ddfm: denoising diffusion model for multi-modality image fusion," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 8082–8093.

[48] X. Yi, L. Tang, H. Zhang, H. Xu, and J. Ma, "Diff-if: Multi-modality image fusion via diffusion model with fusion knowledge prior," *Information Fusion*, vol. 110, p. 102450, 2024.

[49] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[50] W. Gedara Chaminda Bandara, N. Gopalakrishnan Nair, and V. M. Patel, "Ddpm-cd: Denoising diffusion probabilistic models as feature extractors for change detection," *arXiv e-prints*, pp. arXiv–2206, 2022.

[51] J. Yue, L. Fang, S. Xia, Y. Deng, and J. Ma, "Dif-fusion: Toward high color fidelity in infrared and visible image fusion with diffusion models," *IEEE Transactions on Image Processing*, vol. 32, pp. 5705–5720, 2023.

[52] J. Zhang, G. Cui, J. Zhao, and Y. Chen, "High-frequency attention residual gan network for blind motion deblurring," *IEEE Access*, vol. 10, pp. 81 390–81 405, 2022.

[53] G. Sharma, W. Wu, and E. N. Dalal, "The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, vol. 30, no. 1, pp. 21–30, 2005.

[54] A. Filippou, A. Georgiou, A. Nikolaou, N. Evripidou, and C. Damianou, "Advanced software for mrgfus treatment planning," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107726, 2023.

[55] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[56] Y. Lu, M. Zhang, A. J. Ma, X. Xie, and J. Lai, "Coarse-to-fine latent diffusion for pose-guided person image synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 6420–6429.

[57] L. Höllein, A. Božič, N. Müller, D. Novotny, H.-Y. Tseng, C. Richardt, M. Zollhöfer, and M. Nießner, "Viewdiff: 3d-consistent image generation with text-to-image models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5043–5052.

[58] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 4, pp. 4713–4726, 2022.

[59] Y. Zhang, J. Zhang, H. Li, Z. Wang, L. Hou, D. Zou, and L. Bian, "Diffusion-based blind text image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 25 827–25 836.

[60] K. V. Gandikota and P. Chandramouli, "Text-guided explorable image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 25 900–25 911.

[61] Z. Chen, Y. Zhang, D. Liu, J. Gu, L. Kong, X. Yuan *et al.*, "Hierarchical integration diffusion model for realistic image deblurring," *Advances in neural information processing systems*, vol. 36, 2024.

[62] M. Ren, M. Delbracio, H. Talebi, G. Gerig, and P. Milanfar, "Multiscale structure guided diffusion for image deblurring," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 721–10 733.

[63] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 713–13 722.

[64] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "Segnext: Rethinking convolutional attention design for semantic segmentation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 1140–1156, 2022.

[65] X. Deng, Y. Zhang, M. Xu, S. Gu, and Y. Duan, "Deep coupled feedback network for joint exposure fusion and image super-resolution," *IEEE Transactions on Image Processing*, vol. 30, pp. 3098–3112, 2021.

[66] X. Zhang, "Deep learning-based multi-focus image fusion: A survey and a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 4819–4838, 2022.

[67] K. A. Johnson and J. A. Becker, "Atlas: The whole brain," http://www.med.harvard.edu/AANLIB/.

[68] W. Tang, F. He, Y. Liu, and Y. Duan, "Matr: Multimodal medical image fusion via multiscale adaptive transformer," *IEEE Transactions on Image Processing*, vol. 31, pp. 5134–5149, 2022.

[69] J. Ma, L. Tang, F. Fan, J. Huang, X. Mei, and Y. Ma, "Swinfusion: Cross-domain long-range learning for general image fusion via swin transformer," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1200–1217, 2022.

[70] J. Li, J. Liu, S. Zhou, Q. Zhang, and N. K. Kasabov, "Gesenet: A general semantic-guided network with couple mask ensemble for medical image fusion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 11, pp. 16 248–16 261, 2024.

[71] Z. Zhao, H. Bai, J. Zhang, Y. Zhang, S. Xu, Z. Lin, R. Timofte, and L. Van Gool, "Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 5906–5916.

[72] Z. Zhao, H. Bai, J. Zhang, Y. Zhang, K. Zhang, S. Xu, D. Chen, R. Timofte, and L. Van Gool, "Equivariant multi-modality image fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 25 912–25 921.

[73] P.-H. Dinh, V.-H. Vu, N. L. Giang *et al.*, "A new approach to medical image fusion based on the improved extended difference-of-gaussians combined with the coati optimization algorithm," *Biomedical Signal Processing and Control*, vol. 93, p. 106175, 2024.

[74] X. Xie, X. Zhang, S. Ye, D. Xiong, L. Ouyang, B. Yang, H. Zhou, and Y. Wan, "Mrscfusion: Joint residual swin transformer and multiscale cnn for unsupervised multimodal medical image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–17, 2023.

[75] C. S. Xydeas, V. Petrovic *et al.*, "Objective image fusion performance measure," *Electronics letters*, vol. 36, no. 4, pp. 308–309, 2000.

[76] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[77] H.-N. Wang, W. Zhong, J. Wang, and D. Xia, "Research of measurement for digital image definition," *Journal of Image and Graphics*, vol. 9, no. 7, pp. 828–831, 2004.

[78] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," *Optics Communications*, vol. 341, pp. 199–209, 2015.

[79] V. Normand, D. L. Lootens, E. Amici, K. P. Plucknett, and P. Aymard, "New insight into agarose gel mechanical properties," *Biomacromolecules*, vol. 1, no. 4, pp. 730–738, 2000.

[80] T.-Y. Wang, Z. Xu, T. L. Hall, J. B. Fowlkes, and C. A. Cain, "An efficient treatment strategy for histotripsy by removing cavitation memory," *Ultrasound in medicine & biology*, vol. 38, no. 5, pp. 753–766, 2012.

[81] A. P. Duryea, C. A. Cain, W. W. Roberts, and T. L. Hall, "Removal of residual cavitation nuclei to enhance histotripsy fractionation of soft tissue," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 62, no. 12, pp. 2068–2078, 2015.

[82] T.-Y. Wang, T. L. Hall, Z. Xu, J. B. Fowlkes, and C. A. Cain, "Imaging feedback of histotripsy treatments using ultrasound shear wave elastography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 59, no. 6, pp. 1167–1181, 2012.

[83] Y. Zhou and X. Wang, "Effect of pulse duration and pulse repetition frequency of cavitation histotripsy on erosion at the surface of soft material," *Ultrasonics*, vol. 84, pp. 296–309, 2018.

[84] K.-W. Lin, Y. Kim, A. D. Maxwell, T.-Y. Wang, T. L. Hall, Z. Xu, J. B. Fowlkes, and C. A. Cain, "Histotripsy beyond the intrinsic cavitation threshold using very short ultrasound pulses: microtripsy," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 61, no. 2, pp. 251–265, 2014.

[85] K. J. Pahk, P. Gélat, D. Sinden, D. K. Dhar, and N. Saffari, "Numerical and experimental study of mechanisms involved in boiling histotripsy," *Ultrasound in Medicine & Biology*, vol. 43, no. 12, pp. 2848–2861, 2017.

[86] T. D. Khokhlova, M. S. Canney, V. A. Khokhlova, O. A. Sapozhnikov, L. A. Crum, and M. R. Bailey, "Controlled tissue emulsification produced by high intensity focused ultrasound shock waves and millisecond boiling," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 3498–3510, 2011.

[87] M. S. Canney, T. D. Khokhlova, V. A. Khokhlova, M. R. Bailey, J. Ha Hwang, and L. A. Crum, "Tissue erosion using shock wave heating and millisecond boiling in hifu fields," in *AIP Conference Proceedings*, vol. 1215, no. 1. American Institute of Physics, 2010, pp. 36–39.