Tumor Synthesis conditioned on Radiomics

Jonghun Kim[1], Inye Na[1], Eun Sook Ko[2], Hyunjin Park[1 †]
[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea
[2] Department of Radiology and Center for Imaging Science, Samsung Medical Center,
Sungkyunkwan University School of Medicine, Suwon, Korea
{iproj2,niy0404,hyunjinp}@skku.edu, mathilda0330@gmail.com

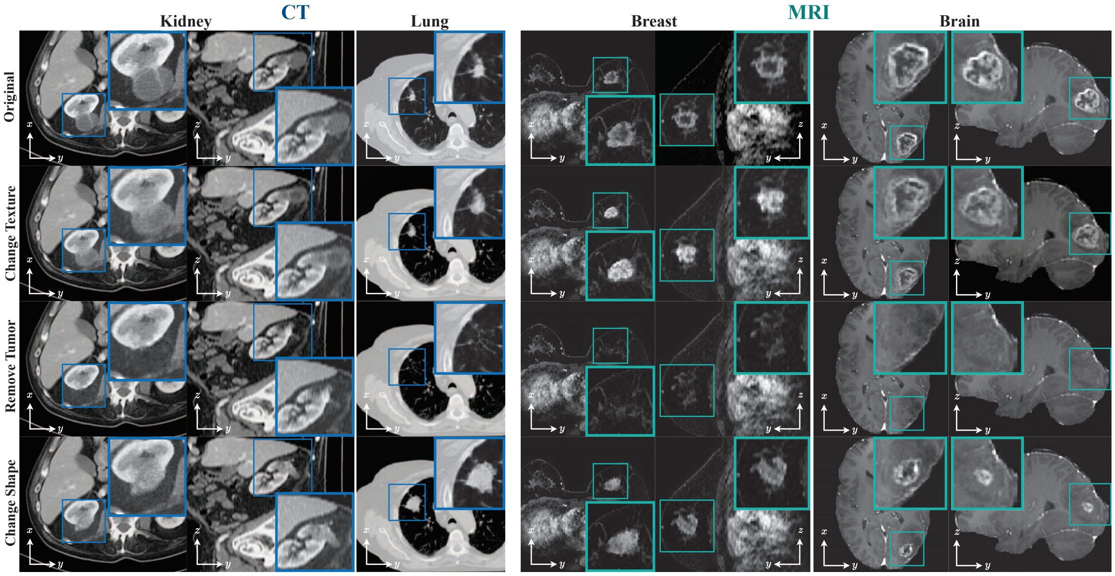


Figure 1. Tumor synthesis results with proposed method. The image displays 2D planes of a volumetric image, with each plane labeled at the bottom left. The x-y plane illustrates the axial view, while the z-y plane depicts the sagittal view. First row: Original images, second row: Images with changed tumor texture, third row: Images with tumors removed, fourth row: Images with changed tumor shape and size.

Abstract

Due to privacy concerns, obtaining large datasets is challenging in medical image analysis, especially with 3D modalities like Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Existing generative models, developed to address this issue, often face limitations in output diversity and thus cannot accurately represent 3D medical images. We propose a tumor-generation model that utilizes radiomics features as generative conditions. Radiomics features are high-dimensional handcrafted semantic features that are biologically well-grounded and thus are good candidates for conditioning. Our model employs a GAN-based model to generate tumor masks and a diffusion-based approach to generate tumor texture conditioned on radiomics features. Our method allows the user to generate tumor images according to user-specified radiomics features such as size, shape, and texture at an arbitrary location. This enables the physicians to easily visualize tumor images to better understand tumors according to changing radiomics features. Our approach allows for the removal, manipulation, and repositioning of tumors, generating various tumor types in different scenarios. The model has been tested on tumors in four different organs (kidney, lung, breast, and brain) across CT and MRI. The synthesized images are shown to effectively aid in training for downstream tasks and their authenticity was also evaluated through expert evaluations. Our method has potential usage in treatment planning with diverse synthesized tumors. Our code is available at github.com/jongdory/TS-Radiomics.

[†] Corresponding Author

This study was supported by National Research Foundation (RS-2024-00408040), Institute for Basic Science (IBS-R015-D1), AI Graduate School Support Program (Sungkyunkwan University) (RS-2019-II190421), ICT Creative Consilience program (RS-2020-II201821), and the Artificial Intelligence Innovation Hub program (RS-2021-II212068).

1. Introduction

In medical image analysis, obtaining large datasets can be challenging due to privacy concerns [49,57]. To mitigate this, extensive research has been conducted on data augmentation based on generative models [10, 45, 55]. Training generative models, particularly those based on Generative Adversarial Networks (GANs) [18], faces significant challenges, including difficulty in model stability and interpretability of results in terms of their medical relevance. A notable issue with GANs is mode collapse, leading to a limited variety of outputs and reduced image diversity, which is particularly problematic in medical settings where accurate representation is vital. These challenges are further heightened with 3D medical images, requiring more extensive training samples for effective model training.

Early detection and prognosis diagnosis of tumors are important. Medical imaging, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), plays a pivotal role in the detection, diagnosis, staging, treatment response monitoring, and recurrence monitoring of tumors [11, 26]. Radiologists can use medical imaging to understand a patient's condition and help select the best treatment method. However, visually evaluating tumors is subjective and can often miss subtle information due to the difficulty in recognizing fine textures or patterns [8, 40].

Radiomics is the method of extracting hundreds to thousands of handcrafted semantic features from routine medical images, enabling quantitative analysis of subtle changes and complex patterns [2, 17, 36]. These features are based on the shape, pixel value distribution, texture, and other patterns of the region of interest. Radiomics extracts meaningful information in medical scenarios such as cancer diagnosis, prognosis, and treatment response prediction for various organs and scanners [36, 39]. With proven medical efficacy, radiomics features are considered biologically well grounded [65] and thus could be rich bases for tumor generation. This provides deep insights that are challenging to obtain through traditional manual analysis and can complement the judgments of clinicians [22, 33, 36]. However, some radiomics features such as complex texture are non-intuitive and challenge medical experts to grasp their significance.

We propose a tumor image generation model conditioned on radiomics features, leveraging the recent diffusion-based generative models [20, 37, 51] and conditioning techniques through the cross-attention mechanism [68]. We demonstrate the ability to generate desired tumor images by adjusting low-dimensional radiomics features. This process involves manipulating intuitive radiomics features such as size to produce 3D tumor images. By converting radiomics features into images, we can provide visual insights. Furthermore, since our approach creates images through adjustable radiomics features, the rationale behind the outcomes for generation is clear. Our model facilitates the simulation of tumor characteristics, including location, size, shape, and texture in 3D medical imaging, enabling the creation of diverse and rare samples as needed. Validated through experiments on tumors in four different organs, it also allows for the generation of tumors with adjustable shapes and textures, as depicted in Figure 1. Our model might have future usage in treatment planning and prognosis prediction with diverse synthesized tumors leading to better personalized treatment options.

Contribution:

- We suggest a tumor shape generator that uses a conditional GAN-based model using shape features and a tumor texture generator based on the Diffusion model to alter the texture of the tumor.
- We enable tumor synthesis through adjustable radiomics features. We propose a diffusion-based model capable of erasing tumors, changing their texture, and manipulating their shape, offering a more comprehensive approach to tumor analysis and simulation.
- We have validated our generative model on tumors across four different organs using CT and MRI images, demonstrating our model's effectiveness through visual results.
- The synthesized images have been useful in aiding downstream tasks and deemed realistic according to expert evaluations.

2. Related Work and Backgrounds

Generative Adversarial Networks are generative models with a structure where a generator and a discriminator learn through competition [18, 29]. With the enhancement in sampling quality and diversity of GANs, they have been deployed across various computer vision applications such as text-to-image synthesis [50, 64, 72, 76], image-to-image translation [12, 23, 25, 35, 75], and image editing [46, 78]. They have rapidly advanced image generation and led to various applications across different domains. They have also been studied in medical image analysis to solve various problems. [12, 35, 75]. For instance, GANs have been utilized for image translation, converting MRI to CT or PET images [7, 13], or facilitating modality transitions within MRIs [12, 35, 75]. Moreover, they have been employed in denoising low-dose CT images to enhance the quality [38]. Given the typical scarcity of data in medical image domains, traditional training can be challenging. However, research using GANs to augment data in medical imaging has shown that training with the generated data can improve performance [10,55]. Additionally, there are studies on synthesizing tumors in 2D images using radiomics features in GANs [45]. Nonetheless, GANs can have unstable learning phases and issues like mode collapse, leading to generating specific data only [56, 62].

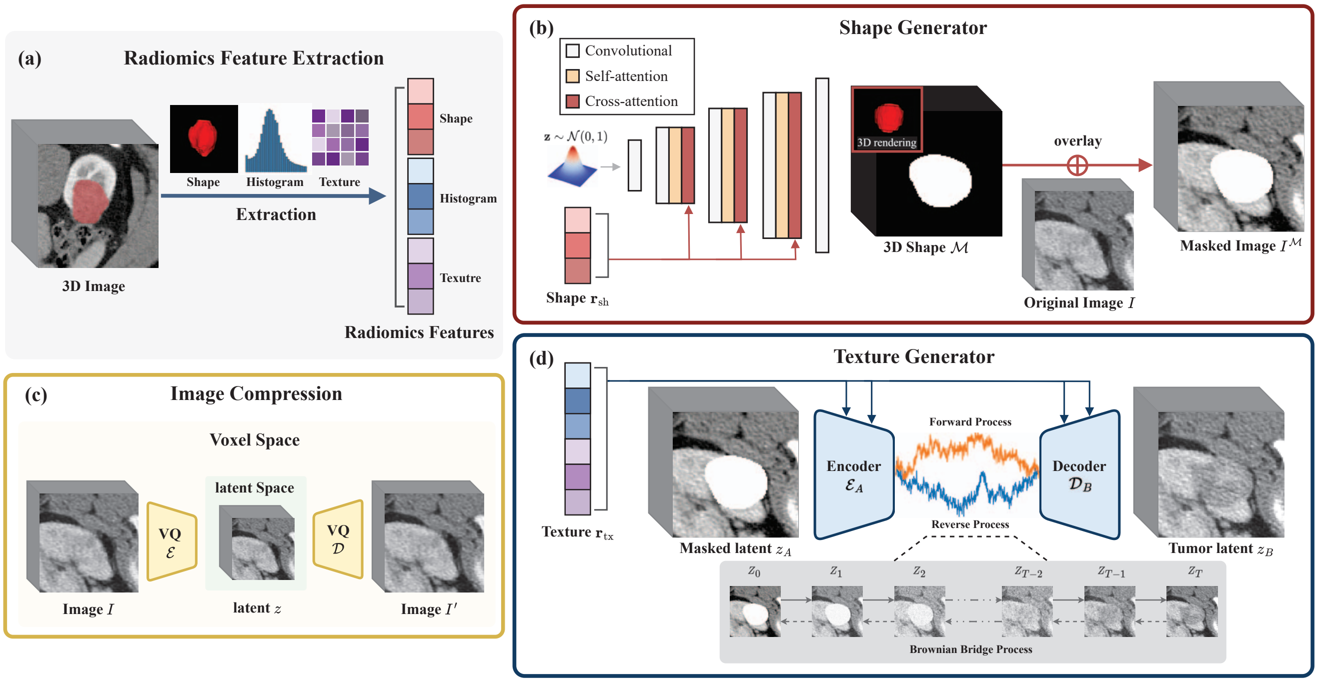


Figure 2. The pipeline of the proposed method. (a) Radiomics feature extraction illustrates the extraction of shape, histogram, and texture features from 3D image and tumor mask. (b) Shape generator depicts training a GAN-based model conditioned on shape features to generate a tumor mask. (c) Image compression into latents using VQ-GAN. (d) Texture generator demonstrates overlaying the generated mask from (b) onto a 3D image to create a masked latent and training a diffusion-based model conditioned on texture features to generate a tumor latent, illustrating the transition from masked domain A to tumor domain B.

Diffusion probabilistic models (DPMs) [20, 59, 61] aim to learn the diffusion process that generates the probability distribution of a given dataset. Unlike GANs, DPMs are known for converging well even with fewer hyperparameters and for producing sharp and detailed images [73]. They have been applied in various domains including text-to-image [48, 51, 77], image-to-image [32, 37, 53, 77], text-to-video [58, 70], data augmentation [16, 66], super-resolution [21, 54, 79], image inpainting and outpainting [41, 51, 53, 77]. Additionally, they have been utilized in the medical domain for various task-specific tasks such as generation, segmentation, translation, anomaly detection, and registration [30–32, 44, 47, 69, 71]. The Latent Diffusion Model (LDM) [51] uses the latent space for high-resolution image generation with efficient computation compared to traditional diffusion models. It utilizes the Vector Quantized GAN (VQ-GAN) [15] to effectively quantize and compress the latent space, aiding in representing and manipulating image features. LDM has shown potential in various image synthesis tasks, demonstrating its ability to handle diverse generation tasks efficiently [9, 27, 47, 51].

Brownian Bridge Diffusion Model (BBDM) assumes the diffusion process as a probabilistic Brownian bridge process in image-to-image translation tasks [37]. By constructing a direct mapping between the source and target domains, it provides a potentially more efficient and generalized model for image-to-image translation tasks, showcasing its applicability across a range of domains. In BBDM, Given the source domain $\boldsymbol{x}_0$ and target domain $\boldsymbol{y}$, the forward process is defined as:

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0,\boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}_t;(1-m_t)\boldsymbol{x}_0+m_t\boldsymbol{y},\delta_t\boldsymbol{I}) \tag{1}$$

, where $m_t = \frac{t}{T}$ with T representing the total steps of the diffusion process and δ_t is a fixed variance. The intermediate state $\boldsymbol{x}_t$ is defined in a discrete form as follows:

$$\boldsymbol{x}_t = (1-m_t)\boldsymbol{x}_0 + m_t\boldsymbol{y} + \sqrt{\delta_t}\boldsymbol{\epsilon}_t \tag{2}$$

, where $\epsilon_t \sim \mathcal{N}(\mathbf{0},\boldsymbol{I})$ is the Gaussian noise. Then, BBDM is trained to approximate the reverse process:

$$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, y) = \mathcal{N}(\boldsymbol{x}_{t-1};\boldsymbol{\mu}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t),\tilde{\delta}_t\boldsymbol{I}) \tag{3}$$

, where $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t,t)$ represents the predicted mean value of the noise and $\tilde{\delta}_t$ denotes the variance of noise at each step. $\boldsymbol{\mu}_t$ is a mean parameterized by the noise predictor $\boldsymbol{\epsilon}_\theta$:

$$\boldsymbol{\mu}_\theta(\boldsymbol{x}_t,\boldsymbol{y},t) = c_{xt}\boldsymbol{x}_t + c_{yt}\boldsymbol{y} + c_{\epsilon t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,t) \tag{4}$$

, where $c_{xt}, c_{yt}, c_{\epsilon t}$ are constants varying with respect to time step t. In the translation process, the sample can be obtained from the Gaussian noise by iterative reverse process: $\boldsymbol{x}_{t-1} = \mu_\theta(\boldsymbol{x}_t,\boldsymbol{y},t) + \sqrt{\tilde{\delta}_t}\boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}(0,1)$ In this study, we utilize radiomics texture features as conditioning in the BBDM to perform the translation task from tumor-masked images to tumor images.

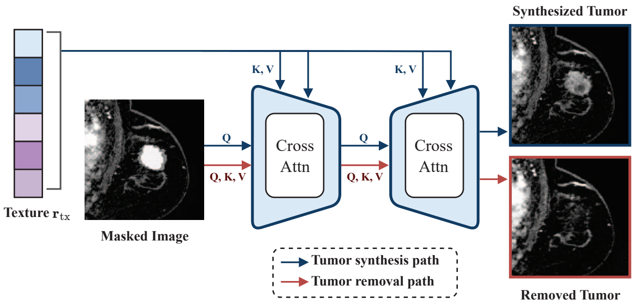


Figure 3. Two generation paths within the texture generator. Blue path: the tumor synthesis path, where texture features are used as key and value for cross-attention, generating the synthesized tumor image. Red path: tumor removal path, where self-attention is performed to generate the removed tumor image.

3. Method

We generate tumor images by utilizing radiomics features extracted using the tumor mask and the underlying 3D medical images. The entire pipeline of our proposed method is depicted in Figure 2. To effectively synthesize tumors, we generate the shape and texture separately. The shape generator and texture generator are trained independently. First, the shape is generated and masked onto the desired image location, then the texture is generated using radiomics features. For normal tissue generation, the tumor area is masked out, and the texture is generated without using radiomics features. Details on the shape generator are in Section 3.1, and the texture generator in Section 3.2.

3.1. Tumor Shape Generator

We employed a GigaGAN [28] with cross-attention [68] for our shape generator, adapting it to a 3D method. We use radiomics features rather than text for feature-to-image generation, focusing on shape-feature conditioning. Since shape masks are relatively easy to generate, we used a simple GAN instead of a complex diffusion model. Traditional convolution filters are limited to their receptive fields [3] and this limitation is significant in the context of tumor shape, where features like volume, surface area, sphericity, and diameter are influenced by long-range relationships. Therefore, integrating these relationships using attention layers is essential. We utilize self-attention to assimilate long-range relationships and cross-attention to enable the generation of tumor shapes with texture features, as depicted in Figure 2 (b). Our shape generator G, generates the shape $\mathcal{M}$, in conjunction with a latent $\mathbf{z} \sim \mathcal{N}(0,1)$, and shape feature $\mathbf{r}_{\text{sh}}$.

$$\mathcal{M} = G(\mathbf{z}, \mathbf{r}_{\text{sh}}) \tag{5}$$

, where shape is $\mathcal{M} \in \mathbb{R}^{H \times W \times D}$ and shape feature is $\mathbf{r}_{\text{sh}} \in \mathbb{R}^{d_{\text{sh}}}$. We can obtain a masked image $I^{\mathcal{M}}$ by applying a mask at the desired location for tumor generation using shape $\mathcal{M}$.

3.2. Tumor Texture Generator

In contrast to the shape generator, to create complex tumor texture patterns in 3D medical images like CT and MRI, we use BBDM [37], which applies a Brownian Bridge process to a diffusion model. This generates texture from the masked image to synthesize the tumor image. Unlike DDPM, which generates the target image from noise, BBDM is specialized for Image-to-Image translation, generating the target domain image from a source domain image. We employ this to translate from the masked domain A to the tumor domain B, as illustrated in Figure 2 (d). Due to the large dimensions of the original 3D volume, we use image compression with VQ-GAN [15, 51] and then conducted the diffusion process in the latent space, as shown in Figure 2 (c). Similar to previous studies [20, 37], we utilize time conditioning UNet [20, 52] ϵ_θ as the backbone. We enable the generation of tumor images corresponding to the given texture condition by utilizing texture features. We follow the standard training object for BBDM and use texture feature $\mathbf{r}_{\text{tx}}$ as conditioning:

$$\mathbb{E}_{\boldsymbol{x}_0,\boldsymbol{y},\boldsymbol{\epsilon}}[c_{\epsilon t}||m_t(\boldsymbol{y}-\boldsymbol{x}_0)+\sqrt{\delta_t}\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,t,\mathbf{r}_{\text{tx}})||^2]. \tag{6}$$

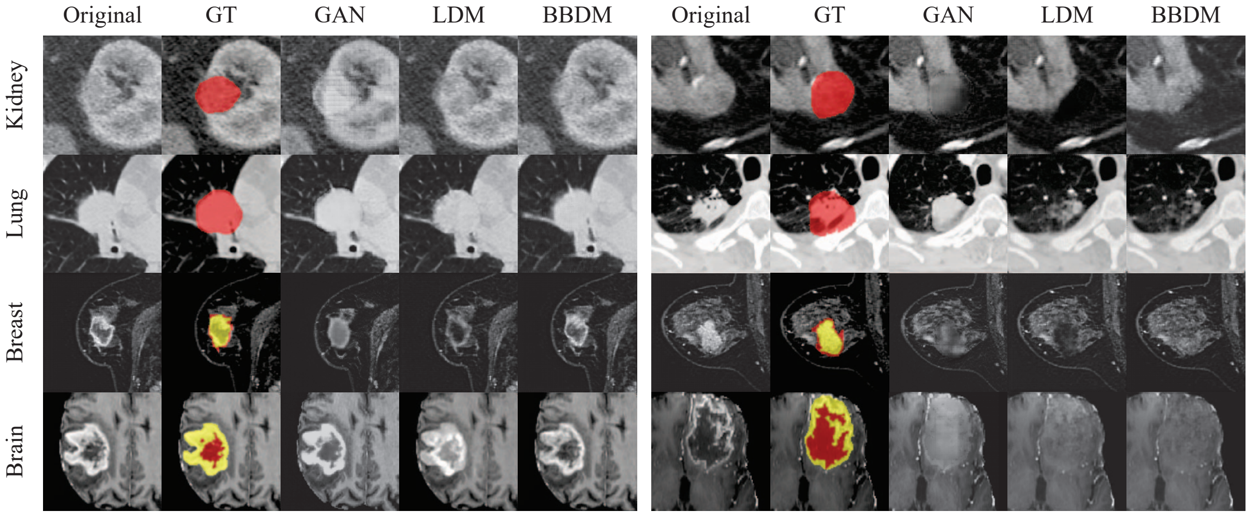
To allow more flexibility in adjusting tumors, we train both tumor synthesis and tumor removal simultaneously. This method uses self-attention when texture features are not provided as conditions and switches to cross-attention when texture conditions are provided. This approach not only helps the model learn the characteristics of surrounding the organ but also eliminates the need to train separate models for tumor generation and removal. This process is illustrated in Figure 3. Similar to previous studies [51], the conditioning mechanisms through cross-attention in the intermediate layers of the UNet are defined as follows: Attention(Q, K, V) = softmax$(\frac{QK^T}{\sqrt{d}} \cdot V)$ with

$$Q = W_Q^{(i)} \cdot q, K = W_K^{(i)} \cdot k, V = W_V^{(i)} \cdot v \tag{7}$$

, where $W_Q^{(i)} \in \mathbb{R}^{d \times d_q^i}, W_K^{(i)} \in \mathbb{R}^{d \times d_k^i}$ and $W_V^{(i)} \in \mathbb{R}^{d \times d_v^i}$ are learnable projection matrices. q, k, v depend on the existence of an input texture feature $\mathbf{r}_{\text{tx}}$:

$$\begin{cases} q = \varphi_i(z_t), \quad k, v = \mathbf{r}_{\text{tx}} & \text{if } \mathbf{r}_{\text{tx}} \text{ is given,} \\ q, k, v = \varphi_i(z_t) & \text{otherwise} \end{cases} \tag{8}$$

, where $\varphi_i(z_t) \in \mathbb{R}^{N \times d_\epsilon^i}$ denotes an intermediate representation of the UNet implementing ϵ_θ. Therefore, when texture features are not provided, we can reconstruct the masked area and by utilizing this, we can also remove the tumor, allowing for flexible alteration of the tumor shape. When training the tumor removal path, it creates a masked image by randomly applying a mask to an area that does not overlap with the existing tumor location. During inference for tumor removal, the tumor area is masked by the user-provided mask and then removed via the removal path. For



(a) Tumor Synthesis

(b) Tumor Removal

Figure 4. Qualitative results of the Baseline on four organs. This displays the outcomes of each model performing two tasks. Synthesis: presents the tumor synthesis result in the given mask when provided with radiomics texture features. Removal: shows the result after removing the tumor corresponding to the mask. In breast, red: peri-tumor, yellow: tumor. In brain, red: necrosis, yellow: enhancing tumor.

tumor synthesis, the normal area is masked by the synthesized mask from the shape generator and texture features are provided as input to synthesize the tumor through the synthesis path. For detailed training and inference processes, please refer to Algorithms 1, 2 and 3 of supplementary.

4. Experiments

Datasets. To substantiate the effectiveness of our method, we validated tumor synthesis performance across two modalities and four organs. The kidneys and lungs were examined using CT data with the kidney data derived from the KiTS23 [19] and the lung data from the NSCLS [1]. For the breast and brain, MRI data were utilized with the breast data coming from a private dataset and the brain data from the BraTS2021 [5, 6, 42]. We allocated 80% of all datasets for training and 20% for testing. Details of the datasets are available in the Sec. A of supplementary.

Models and hyperparameters. Our tumor shape generator is built upon GigaGAN [28], while the tumor texture generator is based on BBDM [37]. Due to the large dimensions of 3D volumes and the consequent high computational cost, we employ VQGAN [15] to compress and reduce the model size. BBDM consists of two components: a pretrained VQ-GAN and a Brownian Bridge diffusion model. The VQ-GAN is later utilized as a comparative model, employing the same model as the Latent Diffusion Model used in subsequent comparisons. During the training stage, the number of time steps for the Brownian Bridge was set to 1000, while in the inference stage, 200 sampling steps were used. The implementation was done using PyTorch[1] and MONAI[2] libraries. We train the network by using the Adam [34] optimizer with a learning rate of 5×10^{-6}. Training was conducted on four A100 80GB GPUs with a batch size of 1 per GPU. For details on the model architecture and hyperparameters, refer to the Sec. B of supplementary.

Baseline Methods and Metrics. To our knowledge, there has not been much previous research that used radiomics features to create tumors for 3D images. Therefore, to substantiate the efficiency of our tumor texture generator, we compare ours with two baselines: one based on GAN and the other on LDM. The GAN-based model is adapted by modifying the 3D image translation model Ea-GAN [75] (based on pix2pix [25]), with the addition of cross-attention for conditioning radiomics. All the baselines we used in the experiment, as well as our method, are 3D methods using radiomics conditioning. For tumor generation, all texture generators used the same shape generator, and in the case of LDM, the same image compression method as BBDM was used. We employed Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM), commonly used quantitative evaluations in medical image generation, to assess the quality of image synthesis [74]. Since the radiomics features can significantly change based on the preprocessing and normalization of the given modality, we assessed whether the features of generated images match the original radiomics features given as conditions by measuring the Pearson and Spearman correlation coefficient.

5. Results

Results of Comparison Method: To demonstrate the effectiveness of our method, we compared its generative performance on a test set against that of the baselines. We con-

[1] https://pytorch.org/
[2] https://monai.io/

Task	(a) Tumor Synthesis						(b) Tumor Removal					
Model	GAN		LDM		**Ours**		GAN		LDM		**Ours**	
Organ (Modality)	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
Kidney (CT)	25.62	0.6962	30.57	0.8829	**33.95***	**0.9346***	25.88	0.6985	31.02	0.8877	**34.16***	**0.9370***
Lung (CT)	23.10	0.6628	29.35	0.9287	**32.68***	**0.9471***	23.61	0.6704	28.42	0.9205	**32.81***	**0.9484***
Breast (MRI)	23.44	0.6420	29.51	0.8726	**31.85***	**0.9177***	24.04	0.6516	29.86	0.8741	**32.07***	**0.9193***
Brain (MRI)	30.17	0.9038	35.83	0.9616	**36.67***	**0.9634***	29.59	0.8943	35.91	0.9609	**37.24***	**0.9681***

$*p$-value < 0.05 comparing two best results

Table 1. Quantitative comparison results for four organs and two modalities were measured using PSNR and SSIM.

Model	GAN		LDM		**Ours**	
Organ	PCC ↑	SCC ↑	PCC ↑	SCC ↑	PCC ↑	SCC ↑
Kidney	0.553	0.587	0.793	0.818	**0.829***	**0.865***
Lung	0.591	0.624	0.759	0.827	**0.801***	**0.859***
Breast	0.610	0.657	0.828	0.810	**0.852***	**0.851***
Brain	0.645	0.672	0.849	0.883	**0.864***	**0.906***

$*p$-value < 0.05 comparing two best results

Table 2. Correlation between conditioned texture features $\mathbf{r}_{tx}$ and the texture features extracted from generated images.

Metric	Model	Kidney	Lung	Breast	Brain	Avg.
Pearson Corr.	GAN	0.937	0.928	0.893	0.902	0.915
	BBDM	0.925	0.933	0.913	0.905	0.919
Spearman Corr.	GAN	0.954	0.941	0.907	0.916	0.930
	BBDM	0.932	0.939	0.924	0.921	0.929

Table 3. Correlation between conditioned shape features $\mathbf{r}_{sh}$ and the shape features extracted from generated tumor masks.

ducted both qualitative and quantitative evaluations. The results of the qualitative assessment are depicted in Figure 4. Each model performed tumor synthesis and tumor removal tasks simultaneously through cross-attention mechanisms. GAN-based model struggled to produce natural depictions for both tasks. This is largely due to the difficulty of obtaining ample data in the medical imaging field and the challenge of dealing with 3D images, which prevented the GAN-based model from generating high-quality images in both tasks across organs. However, models based on diffusion processes could generate good-quality images that closely resembled the real images. Additionally, the BBDM performed the tumor removal task quite naturally, especially in comparison to the LDM. The quantitative evaluation results for each baseline are in Table 1. Consistent with the qualitative assessment, the BBDM exhibits the best performance in generating images for all organs.

We also evaluated how well the generated images reflect the given radiomics features used as conditions. We re-extracted the radiomics features from the generated images and conducted a correlation comparison with the radiomics features provided as conditions. Table 2 shows the correlations between the original texture feature and the extracted texture features from the images generated by each baseline. We observed that the fidelity in terms of correlation of reproducing radiomics features was relatively higher in MRI images, with the highest fidelity occurring in the brain.

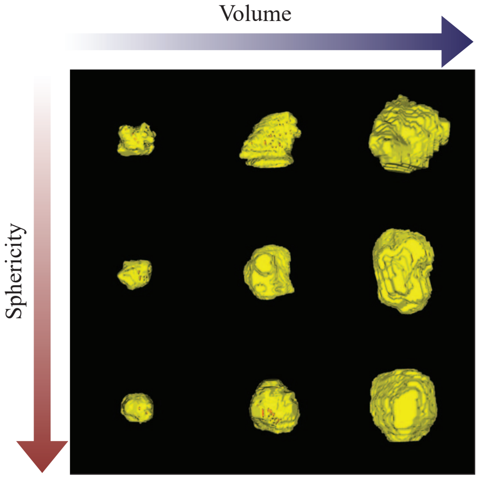


Figure 5. The results of the Tumor Shape Generator for breast. The tumors are conditioned by shape features. Volume denotes the size of the tumor, increasing from left to right. Sphericity refers to the degree to which the shape approaches that of a sphere, assessed from top to bottom.

Moreover, the degree of reproduction was somewhat lower in CT images compared to MRI, possibly because tumor regions in CT are generally more ambiguous.

Manipulation of Tumor Shape: We generated 3D tumor masks with the appropriate conditions using shape features in the tumor shape generator. We quantitatively assessed the generated tumor masks by comparing the shape feature values of the ground truth with those of the generated tumor masks using correlations. Table 3 shows the correlations between the original shape feature and the extracted shape features from the tumor mask generated by the shape generator. This demonstrates that the generation process accurately integrated the shape features showing that there is little difference in performance between GAN and BBDM for shape generation. It also proves that even a simple GAN is sufficient for effective modeling. The shape features used are specified in Sec. C of the supplementary.

We attempted to create tumors by adjusting the intuitive radiomics features of volume and sphericity. Figure 5 shows the results of the tumors generated by manipulating these two features, presented through 3D rendering. The outcomes indicate that our shape generator has successfully reflected the shape features in creating the tumor masks.

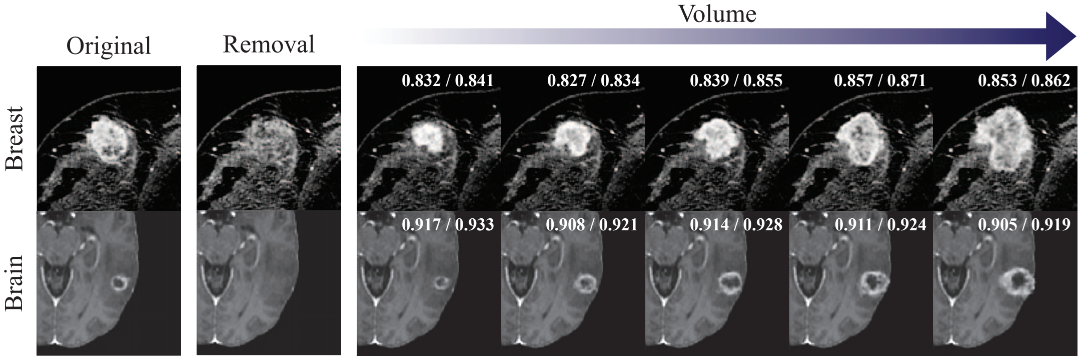


Figure 6. Results of tumor generation for various volume sizes using a tumor mask created by the tumor shape generator according to specified volume size features. The texture generator then removes the tumor and regenerates tumors of different sizes in the same location. The tumor volume increases from left to right. In the upper right corner, the Pearson / Spearman correlation values between the texture features of the original and generated image are displayed. Breast: sagittal view, brain: axial view.

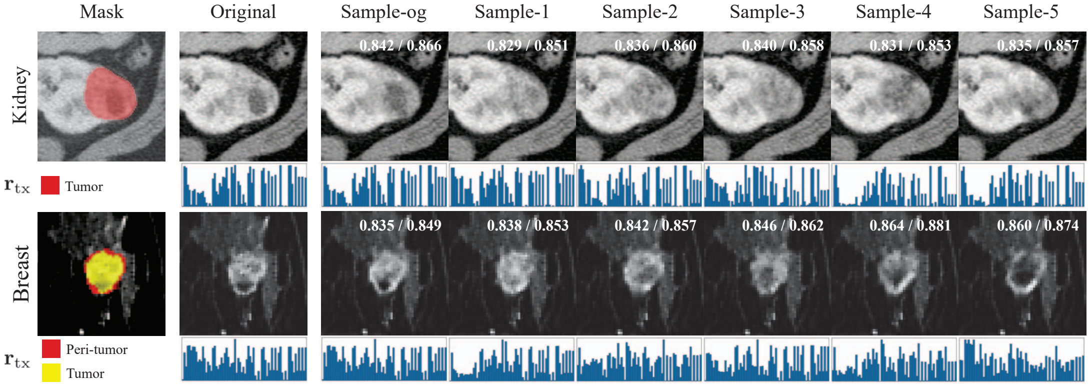


Figure 7. Results of tumor generation based on texture features. In the upper right corner, the Pearson / Spearman correlation values between the radiomics features given as the condition ($\mathbf{r}_{\text{tx}}$) and the radiomics features extracted from the generated image are displayed. Sample-og denotes the generated images conditioned on the original radiomics features. A bar plot representing the various texture values is shown at the bottom of the image. For the kidney and breast, axial views are shown.

Therefore, this suggests that we can generate tumor masks of desired shapes by controlling the intuitive shape features.

In our study, we adjusted the volume size in the shape feature to generate tumor masks of different sizes. We then used these tumor masks to simulate changes based on tumor volume size. This experiment is depicted in Figure 6. Our proposed method not only generates tumors in a remarkably natural manner but also demonstrates the ability to simulate according to the volume size. Additionally, it shows that even as the volume size increases, the original texture is accurately reflected in the tumor generation as shown by correlation values over 0.8.

Change Tumor Texture: To verify if our proposed method accurately reflects texture features in tumor generation, we conducted experiments by generating tumors using texture features from different samples. That is, we replaced the original texture features of one subject with those of another subject. This experiment is depicted in Figure 7. To check how well the given texture is replicated, we extracted texture features from the generated image in the mask and compared their correlation with the features provided as conditions. The quantitative metrics shown in the figure indicate that the texture features were well-replicated in the generation process (correlation > 0.8). This demonstrates that our model is capable of accurately reflecting even subtle textures that are not easily noticeable.

Change Tumor Position: We conducted experiments on tumor generation by altering the shape and position of tumors, depicted in Figure 8. Tumors were generated in the normal brain without tumors and they were successfully created according to the specified shapes and positions. Our results show that our model can generate tumors of various shapes at different locations in an organ, without any constraints on position and shape. This highlights the potential for generating and utilizing a wider variety of samples in medical imaging, where data is often scarce.

Validation in a Downstream Task. We tested our generative model's effectiveness for downstream tasks by training a standard segmentation model with tumor-synthesized images to see if there was an improvement. The results are detailed in Table 4. We trained the baseline segmentation model, nnU-Net [24], and compared it across two tasks. In

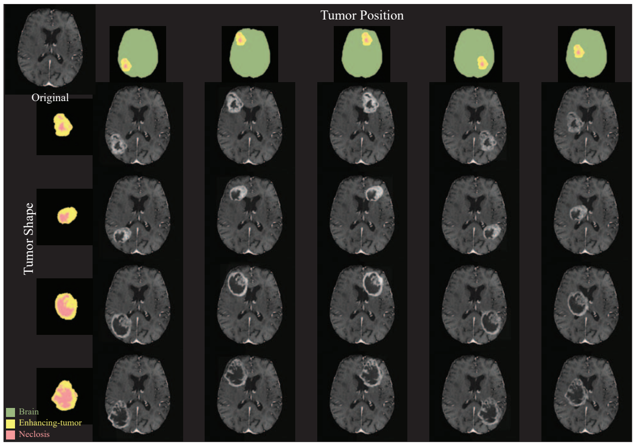


Figure 8. Results of tumor generation in the brain with various shapes and positions. The vertical axis represents different shapes, while the horizontal axis depicts various positions. The upper left indicates the original normal brain.

Region	Brain Tumor		
Model	ET	TC	All
nnU-Net [24]	0.9073$\pm$0.176	0.7641$\pm$0.301	0.8357$\pm$0.256
nnU-Net (+Aug)	**0.9113$\pm$0.166***	**0.7758$\pm$0.294***	**0.8435$\pm$0.256***
Region	Breast Tumor		
Model	Peri-Tumor	Tumor	All
nnU-Net [24]	0.8563$\pm$0.043	0.8633$\pm$0.048	0.8598$\pm$0.046
nnU-Net (+Aug)	**0.8753$\pm$0.031***	**0.8828$\pm$0.032***	**0.8791$\pm$0.031***

$*p$-value < 0.05

Table 4. Performance comparison between the baseline models and the model trained with additional synthesized images for downstream tasks of brain and breast tumor segmentation. Performance metrics include DICE scores.

the brain tumor segmentation task, the baseline was trained with images from 1000 subjects, while another model was trained with an additional 1000 synthesized images for a total of 2000 images. In the breast tumor segmentation task, the baseline learned from 88 subject images, and a comparative model was trained with an additional 88 synthesized images, totaling 176 images. Brain tumor segmentation typically occurs in a multimodal setting. Since training was conducted solely with the T1ce modality, samples where the brain tumor is not clearly visible in T1ce present challenges for tumor detection, leading to a larger standard deviation in performance. The results showed a clear performance improvement when training with additional synthesized images, notably a larger boost in the less data-abundant breast tumor case than in brain tumors with more baseline data. This suggests that our method can significantly enhance model performance through augmentation, especially in situations with scarce or imbalanced datasets, highlighting its utility in augmenting models particularly when normal patient data outnumber abnormal cases.

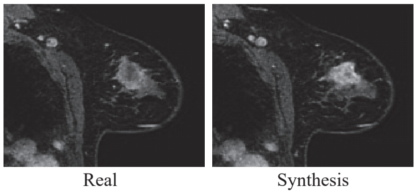


Figure 9. Illustration of real and synthesized images used for expert evaluation.

Experts	#1	#2	Avg
Real vs. Synthesis Acc	55%	60%	57.5%

Table 5. Evaluation by two experts on real vs synthesized images.

Qualitative Evaluation by Experts. To verify the authenticity of synthesized images, we conducted expert evaluations by board-certified radiologists on 20 subjects. As depicted in Figure 9, experts were tasked with identifying real images from a mix of real and synthesized ones. This process assessed how convincingly the generated images mimic real ones. The results are detailed in Table 5. These findings indicate that our model can produce images of such quality that even an expert would find it difficult to distinguish them from the real ones. Additional results of generating rare samples are in the Sec. D of supplementary.

6. Discussion

Limitations: In the case of tumors, there can be related changes in the tissues surrounding the tumor [4, 14]. For instance, edema can occur around a brain tumor [6, 63]. Moreover, as a tumor grows, it also can push and distort the surrounding tissue [43]. In our experiment, we did not include such surrounding areas, which may limit the natural generation of tumors. Additionally, because we use radiomics features extracted from a single dataset for a given organ and modality, the learned distribution could be insufficient leading to degraded generalization on unseen data.

Potential: If it becomes possible to generate tumors of desired shapes and locations with desired textures, this would enable various simulations crucial for providing personalized treatment plans for tumor treatment. For instance, we could simulate scenarios of disease progression using simulated tumor images under an established prognosis model and possibly direct patients to alternative treatments. In short, our technology allows for simulations that assess risks based on the shape, texture, and position of the tumor.

7. Conclusion

In this study, we demonstrate the ability to synthesize 3D tumor images using a diffusion model conditioned on biologically grounded radiomics features. This generation capability not only enables appropriate data augmentation in medical imaging where data are scarce but also serves as a tool for simulating data for personalized treatment plans.

References

[1] HJWL Aerts, E Rios Velazquez, RT Leijenaar, Chintan Parmar, Patrick Grossmann, S Cavalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. Data from nsclc-radiomics. *The cancer imaging archive*, 2015. 5, 1

[2] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*, 5(1):4006, 2014. 2

[3] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74, 2021. 4

[4] Nicole M Anderson and M Celeste Simon. The tumor microenvironment. *Current Biology*, 30(16):R921–R925, 2020. 8

[5] Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021. 5, 1

[6] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017. 5, 8, 1

[7] Avi Ben-Cohen, Eyal Klang, Stephen P Raskin, Shelly Soffer, Simona Ben-Haim, Eli Konen, Michal Marianne Amitai, and Hayit Greenspan. Cross-modality synthesis from ct to pet using fcn and gan networks for improved automated lesion detection. *Engineering Applications of Artificial Intelligence*, 78:186–194, 2019. 2

[8] Wenya Linda Bi, Ahmed Hosny, Matthew B Schabath, Maryellen L Giger, Nicolai J Birkbak, Alireza Mehrtash, Tavis Allison, Omar Arnaout, Christopher Abbosh, Ian F Dunn, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA: a cancer journal for clinicians*, 69(2):127–157, 2019. 2

[9] Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your latents: High-resolution video synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22563–22575, 2023. 3

[10] Yizhou Chen, Xu-Hua Yang, Zihan Wei, Ali Asghar Heidari, Nenggan Zheng, Zhicheng Li, Huiling Chen, Haigen Hu, Qianwei Zhou, and Qiu Guan. Generative adversarial networks in medical image augmentation: A review. *Computers in Biology and Medicine*, 144:105382, 2022. 2

[11] Jin-Young Choi, Jeong-Min Lee, and Claude B Sirlin. Ct and mr imaging diagnosis and staging of hepatocellular carcinoma: part i. development, growth, and spread: key pathologic and imaging aspects. *Radiology*, 272(3):635–654, 2014. 2

[12] Onat Dalmaz, Mahmut Yurt, and Tolga Çukur. Resvit: Residual vision transformers for multimodal medical image synthesis. *IEEE Transactions on Medical Imaging*, 41(10):2598–2614, 2022. 2

[13] Xue Dong, Tonghe Wang, Yang Lei, Kristin Higgins, Tian Liu, Walter J Curran, Hui Mao, Jonathon A Nye, and Xiaofeng Yang. Synthetic ct generation from non-attenuation corrected pet images for whole-body pet imaging. *Physics in Medicine & Biology*, 64(21):215016, 2019. 2

[14] Mikala Egeblad, Elizabeth S Nakasone, and Zena Werb. Tumors as organs: complex tissues that interface with the entire organism. *Developmental cell*, 18(6):884–901, 2010. 8

[15] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12873–12883, 2021. 3, 4, 5, 1

[16] Chun-Mei Feng, Kai Yu, Yong Liu, Salman Khan, and Wangmeng Zuo. Diverse data augmentation with diffusions for effective test-time prompt tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2704–2714, 2023. 3

[17] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016. 2

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2

[19] Nicholas Heller, Fabian Isensee, Dasha Trofimova, Resha Tejpaul, Zhongchen Zhao, Huai Chen, Lisheng Wang, Alex Golts, Daniel Khapun, Daniel Shats, Yoel Shoshan, Flora Gilboa-Solomon, Yasmeen George, Xi Yang, Jianpeng Zhang, Jing Zhang, Yong Xia, Mengran Wu, Zhiyang Liu, Ed Walczak, Sean McSweeney, Ranveer Vasdev, Chris Hornung, Rafat Solaiman, Jamee Schoephoerster, Bailey Abernathy, David Wu, Safa Abdulkadir, Ben Byun, Justice Spriggs, Griffin Struyk, Alexandra Austin, Ben Simpson, Michael Hagstrom, Sierra Virnig, John French, Nitin Venkatesh, Sarah Chan, Keenan Moore, Anna Jacobsen, Susan Austin, Mark Austin, Subodh Regmi, Nikolaos Papanikolopoulos, and Christopher Weight. The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct, 2023. 5, 1

[20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 3, 4, 1

[21] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *The Journal of Machine Learning Research*, 23(1):2249–2281, 2022. 3

[22] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz, and Hugo JWL Aerts. Artificial in-

telligence in radiology. *Nature Reviews Cancer*, 18(8):500–510, 2018. 2

[23] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018. 2

[24] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021. 7, 8

[25] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 2, 5, 1

[26] Veena R Iyer and Susanna I Lee. Mri, ct, and pet/ct for ovarian cancer detection and adnexal lesion characterization. *American Journal of Roentgenology*, 194(2):311–321, 2010. 2

[27] Lan Jiang, Ye Mao, Xiangfeng Wang, Xi Chen, and Chao Li. Cola-diff: Conditional latent diffusion model for multimodal mri synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 398–408. Springer, 2023. 3

[28] Minguk Kang, Jun-Yan Zhu, Richard Zhang, Jaesik Park, Eli Shechtman, Sylvain Paris, and Taesung Park. Scaling up gans for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10124–10134, 2023. 4, 5, 2

[29] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 2

[30] Amirhossein Kazerouni, Ehsan Khodapanah Aghdam, Moein Heidari, Reza Azad, Mohsen Fayyaz, Ilker Hacihaliloglu, and Dorit Merhof. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis*, page 102846, 2023. 3

[31] Boah Kim, Inhwa Han, and Jong Chul Ye. Diffusemorph: unsupervised deformable image registration using diffusion model. In *European Conference on Computer Vision*, pages 347–364. Springer, 2022. 3

[32] Jonghun Kim and Hyunjin Park. Adaptive latent diffusion model for 3d medical image to image translation: Multimodal magnetic resonance imaging study. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 7604–7613, January 2024. 3

[33] Jonghun Kim and Hyunjin Park. Radiomics-guided multimodal self-attention network for predicting pathological complete response in breast mri. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, page 1–5. IEEE, May 2024. 2

[34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5, 1

[35] Lingke Kong, Chenyu Lian, Detian Huang, Yanle Hu, Qichao Zhou, et al. Breaking the dilemma of medical image-to-image translation. *Advances in Neural Information Processing Systems*, 34:1964–1978, 2021. 2

[36] Philippe Lambin, Ralph TH Leijenaar, Timo M Deist, Jurgen Peerlings, Evelyn EC De Jong, Janita Van Timmeren, Sebastian Sanduleanu, Ruben THM Larue, Aniek JG Even, Arthur Jochems, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology*, 14(12):749–762, 2017. 2

[37] Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdm: Image-to-image translation with brownian bridge diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1952–1961, 2023. 2, 3, 4, 5

[38] Zhihua Li, Weili Shi, Qiwei Xing, Yu Miao, Wei He, Huamin Yang, and Zhengang Jiang. Low-dose ct image denoising with improving wgan and hybrid loss function. *Computational and Mathematical Methods in Medicine*, 2021, 2021. 2

[39] Zhenyu Liu, Shuo Wang, Di Dong, Jingwei Wei, Cheng Fang, Xuezhi Zhou, Kai Sun, Longfei Li, Bo Li, Meiyun Wang, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*, 9(5):1303, 2019. 2

[40] Meghan G Lubner, Andrew D Smith, Kumar Sandrasegaran, Dushyant V Sahani, and Perry J Pickhardt. Ct texture analysis: definitions, applications, biologic correlates, and challenges. *Radiographics*, 37(5):1483–1503, 2017. 2

[41] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022. 3

[42] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 5, 1

[43] Hamid Mohammadi and Erik Sahai. Mechanisms and impact of altered tumour mechanics. *Nature cell biology*, 20(7):766–774, 2018. 8

[44] Inye Na, Jonghun Kim, Eun Sook Ko, and Hyunjin Park. Radiomicsfill-mammo: Synthetic mammogram mass manipulation with radiomics features. *arXiv preprint arXiv:2407.05683*, 2024. 3

[45] Inye Na, Jonghun Kim, and Hyunjin Park. Synthetic tumor manipulation: With radiomics features. *arXiv preprint arXiv:2311.02586*, 2023. 2

[46] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. Styleclip: Text-driven manipulation of stylegan imagery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2085–2094, 2021. 2

[47] Walter HL Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F Da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M Jorge Cardoso. Brain imaging generation with latent diffusion models. In *MICCAI Work-*

shop on Deep Generative Models, pages 117–126. Springer, 2022. 3

[48] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022. 3

[49] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps: Automation of Decision Making*, pages 323–350, 2018. 2

[50] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. In *International conference on machine learning*, pages 1060–1069. PMLR, 2016. 2

[51] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2, 3, 4, 1

[52] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 4, 1

[53] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022. 3

[54] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022. 3

[55] Veit Sandfort, Ke Yan, Perry J Pickhardt, and Ronald M Summers. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific reports*, 9(1):16884, 2019. 2

[56] Divya Saxena and Jiannong Cao. Generative adversarial networks (gans) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)*, 54(3):1–42, 2021. 2

[57] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017. 2

[58] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, and Yaniv Taigman. Make-a-video: Text-to-video generation without text-video data. In *The Eleventh International Conference on Learning Representations*, 2023. 3

[59] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015. 3

[60] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 1

[61] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. 3

[62] Akash Srivastava, Lazar Valkov, Chris Russell, Michael U Gutmann, and Charles Sutton. Veegan: Reducing mode collapse in gans using implicit variational learning. *Advances in neural information processing systems*, 30, 2017. 2

[63] Walter Stummer. Mechanisms of tumor-related brain edema. *Neurosurgical focus*, 22(5):1–7, 2007. 8

[64] Ming Tao, Hao Tang, Fei Wu, Xiao-Yuan Jing, Bing-Kun Bao, and Changsheng Xu. Df-gan: A simple and effective baseline for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16515–16525, 2022. 2

[65] Michal R Tomaszewski and Robert J Gillies. The biological meaning of radiomic features. *Radiology*, 298(3):505–516, 2021. 2

[66] Brandon Trabucco, Kyle Doherty, Max Gurinas, and Ruslan Salakhutdinov. Effective data augmentation with diffusion models. *arXiv preprint arXiv:2302.07944*, 2023. 3

[67] Joost JM Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017. 2

[68] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2, 4

[69] Junde Wu, Wei Ji, Huazhu Fu, Min Xu, Yueming Jin, and Yanwu Xu. Medsegdiff-v2: Diffusion-based medical image segmentation with transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6030–6038, 2024. 3

[70] Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Stan Weixian Lei, Yuchao Gu, Yufei Shi, Wynne Hsu, Ying Shan, Xiaohu Qie, and Mike Zheng Shou. Tune-a-video: One-shot tuning of image diffusion models for text-to-video generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7623–7633, 2023. 3

[71] Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 650–656, 2022. 3

[72] Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1316–1324, 2018. 2

[73] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 2022. 3

[74] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552, 2019. 5

[75] Biting Yu, Luping Zhou, Lei Wang, Yinghuan Shi, Jurgen Fripp, and Pierrick Bourgeat. Ea-gans: edge-aware generative adversarial networks for cross-modality mr image synthesis. *IEEE transactions on medical imaging*, 38(7):1750–1762, 2019. 2, 5, 1

[76] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 5907–5915, 2017. 2

[77] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3836–3847, 2023. 3

[78] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A Efros. Generative visual manipulation on the natural image manifold. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*, pages 597–613. Springer, 2016. 2

[79] jianyi Wang Zongsheng Yue and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 3

[80] Alex Zwanenburg, Martin Vallières, Mahmoud A Abdalah, Hugo JWL Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, Spyridon Bakas, Roelof J Beukinga, Ronald Boellaard, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338, 2020. 2