# UCLA
## UCLA Previously Published Works

**Title**

Synthesizing MR Image Contrast Enhancement Using 3D High-Resolution ConvNets

**Permalink**

https://escholarship.org/uc/item/7hm533pz

**Journal**

IEEE Transactions on Biomedical Engineering, 70(2)

**ISSN**

0018-9294

**Authors**

Chen, Chao

Raymond, Catalina

Speier, William

et al.

**Publication Date**

2023-02-01

**DOI**

10.1109/tbme.2022.3192309

Peer reviewed

# Synthesizing MR Image Contrast Enhancement Using 3D High-resolution ConvNets

**Chao Chen**,

Information Science and Electrical Engineering Department, Zhejiang Univeristy, 310027 Hangzhou, China.

**Catalina Raymond**,

UCLA Brain Tumor Imaging Laboratory, Center for Computer Vision and Imaging Biomarkers, Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles.

**Bill Speier**,

Computational Diagnostics Lab, the Department of Radiological Sciences, the Department of Pathology and Laboratory Medicine, and the Department of Bioengineering at the University of California Los Angeles, 924 Westwood Blvd, Suite 600, CA 90024 USA

**Xinyu Jin**,

Information Science and Electrical Engineering Department, Zhejiang Univeristy, 310027 Hangzhou, China.

**Timothy F. Cloughesy**,

Department of Neurology, David Geffen School of Medicine at the University of California Los Angeles.

**Dieter Enzmann**,

Department of Radiological Sciences, David Geffen School of Medicine at the University of California Los Angeles.

**Benjamin M. Ellingson**,

UCLA Brain Tumor Imaging Laboratory, Center for Computer Vision and Imaging Biomarkers, Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles.

**Corey W. Arnold**

Computational Diagnostics Lab, the Department of Radiological Sciences, the Department of Pathology and Laboratory Medicine, and the Department of Bioengineering at the University of California Los Angeles, 924 Westwood Blvd, Suite 600, CA 90024 USA

## Abstract

**Objective:** Gadolinium-based contrast agents (GBCAs) have been widely used to better visualize disease in brain magnetic resonance imaging (MRI). However, gadolinium deposition within the brain and body has raised safety concerns about the use of GBCAs. Therefore, the development

(chench@zju.edu.cn).

of novel approaches that can decrease or even eliminate GBCA exposure while providing similar contrast information would be of significant use clinically.

**Methods:** In this work, we present a deep learning based approach for contrast-enhanced T1 synthesis on brain tumor patients. A 3D high-resolution fully convolutional network (FCN), which maintains high resolution information through processing and aggregates multi-scale information in parallel, is designed to map pre-contrast MRI sequences to contrast-enhanced MRI sequences. Specifically, three pre-contrast MRI sequences, T1, T2 and apparent diffusion coefficient map (ADC), are utilized as inputs and the post-contrast T1 sequences are utilized as target output. To alleviate the data imbalance problem between normal tissues and the tumor regions, we introduce a local loss to improve the contribution of the tumor regions, which leads to better enhancement results on tumors.

**Results:** Extensive quantitative and visual assessments are performed, with our proposed model achieving a PSNR of 28.24dB in the brain and 21.2dB in tumor regions.

**Conclusion and Significance:** Our results suggest the potential of substituting GBCAs with synthetic contrast images generated via deep learning. Code is available at https://github.com/chenchao666/Contrast-enhanced-MRI-Synthesis

### Keywords

Medical Image Synthesis; GBCAs; Brain MRI; Contrast Enhancement; Fully Convolutional Networks

## I. Introduction

Magnetic resonance imaging (MRI) is one of the most important techniques to distinguish different tissue properties and lesions in brain. To better visualize different kinds of disease, gadolinium-based contrast agents (GBCAs) have been widely used for brain MRI image enhancement [1]. Initially, the use of GBCAs was felt to carry minimal risk, with GBCAs administered in up to 35% of all MRI examinations [2]. However, some recent studies have demonstrated the deposition of gadolinium contrast agents in body tissues, including the brain [3], [4], which has raised broad safety concerns about the use of GBCAs in medical imaging. Previous studies have also suggested that GBCA dose should be as low as required, since advanced renal disease and the development of nephrogenic systemic fibrosis (NFS) are linked to high exposure to GBCAs [2], [5]. Even though deposition can be minimized by reducing the dose of gadolinium used, using low-dose contrast-enhanced MRI images may ignore some important information provided by contrast [6]. It is of importance to minimize or even eliminate the use of GBCAs, while preserving high-contrast information.

Recent development of deep learning methods have demonstrated success for medical image analysis [7], especially in the fields of segmentation [8]-[11], detection [12], [13], reconstruction, [14]-[16] and synthesis [6], [17]-[20]. In this study, we focus on developing a deep learning based approach to synthesize contrast-enhanced brain MRI sequences from non-contrast brain MRI sequences. Specifically, synthetic contrast-enhanced MRI images would especially useful for certain patients, such as: (1) pediatric patients, (2) patients with benign or low grade (slow growing) brain tumors who are undergoing routine clinical exams

over time to look at tumor growth, and (3) patients with impaired renal function or who can otherwise not get GBCAs.

Most recently, Gong et al. proposed to learn the reconstruction of full-dose T1 scans from pre-contrast T1 scans and 10% low-dose T1 scans [6]. In order to completely eliminate the dependence on GBCAs, Kleesiek et al. proposed to predict contrast enhancement sequences directly from non-contrast brain MRI sequences [21]. To introduce additional information from other modalities, they utilized 10 multi-parametric scans as inputs. There are several limitations of existing studies. First, datasets used for training and evaluation are small, containing no more than 100 subjects. Second, off-the-shelf network architectures and loss functions are used, which likely limits performance. Finally, existing work has insufficient performance on tumors and small vessels. For these reasons, in this work, we introduce a larger scale dataset containing more than 400 MRI sequences, and design a 3D high resolution fully convolutional network (FCN) to synthesize contrast-enhanced T1 (CE-T1) images. Fig. 1 illustrates the four MRI modalities used in this work. Specifically, T1, T2 and ADC, are used as inputs to synthesize the post-contrast T1 with the proposed 3D FCN model. The main contributions of this paper are:

- A dataset of over 400 MRI sequences are analyzed, the largest explored thus far for the task of MRI virtual contrast enhancement.

- A 3D high-resolution FCN model is designed to generate the CE-T1 from the precontrast MRI scans. The presented model outperforms existing virtual contrast enhancement methods in two ways: (1) it maintains high-resolution information throughout processing, and (2) it repeats multi-scale fusion and aggregates multi-scale information in parallel.

- Since the voxels that compose tumor regions are limited relative to the entire MRI volume, deep learning methods with global loss functions struggle to accurately synthesize contrast in these regions. Therefore, a local loss is introduced to re-balance the contribution of the tumor regions, which leads to improved performance on tumors.

- Extensive experiments, visual assessments, and ablation studies are conducted. As a result, we achieved a peak-signal-to-noise ratio (PSNR) of 28.24dB in brain regions and 21.2dB in tumor regions. Numerical and visual assessments demonstrate that the presented method significantly outperforms existing work.

## II. Related Work

In this section, we review the deep network architectures and loss functions that are widely used for image-to-image translation. We then discuss recent applications in medical image synthesis that are related to our study.

### A. Image-to-Image Translation

Image-to-image (I2I) translation has been explored in recent years with the aim of translating an input image in a source domain to an image in a target domain. The basic idea of I2I methods is to learn a non-linear feature mapping given the input and output

image pairs as training data. A large number of network architectures has been proposed to act as the non-linear mapping. For example, Long et al. [22] proposed to utilize a FCN model for image-to-segmentation. In Ronneberger et al. [8], a U-Net architecture was proposed for biomedical image segmentation, which is currently widely used in medical image translation tasks. In Chen et al. [23], the authors introduced the dilated convolution to enlarge the receptive field of neural networks. In Zhao et al. [24], a pyramid pooling module was proposed to fuse features under four different pyramid scales, which enable the model to utilize local and global context information for pixel-wise prediction. In Newell et al. [25], the stacked hourglass module was proposed to capture and consolidate information across all scales of the image for human pose estimation. Different from the traditional high-to-low and low-to-high FCN architectures, Sun et al. [26] proposed a high-resolution net, which maintains high-resolution feature maps throughout processing.

A large number of training objectives have been introduced to measure the difference between the generated image and the ground truth image in I2I translation tasks. The typical choices are $\ell_1$ and $\ell_2$ loss. In [27], a differentiable variant of $\ell_1$ loss, named the Charbonnier penalty function, was proposed to handle outliers. Compared with $\ell_1$ or $\ell_2$ loss, which may lead to blurry images [28], adversarial loss [18], [20], [28], [29] has become a popular choice for I2I tasks. This process trains a discriminator to distinguish generated images from ground truth images. Additionally, perceptual loss, which measures the difference in feature space, has also been widely used in I2I translation [30], [31].

### B. Medical Image Synthesis

Recently, an increasing number of machine learning and deep learning methods have shown potential in medical image synthesis, which estimates a desired imaging modality from other modalities or scans. For example, Li et al. applied the 3D-CNN model to predict missing PET patterns from MRI data [32]. To improve the quality of 3T MR images, Bahrami et al. collected a dataset with paired 3T and 7T images scanned from the same subjects, and proposed to reconstruct 7T-like images from 3T images [17], [33]. In Xiang et. al [34], the authors proposed to estimate standard-dose positron emission tomography (PET) images from low-dose PET and MRI images. In Huang et al. [35], the authors proposed a weakly-supervised convolutional sparse coding method to simultaneously solve the problems of super-resolution and cross-modality image synthesis. In Dar et al. [20], the authors proposed a new approach for multi-contrast MRI synthesis based on conditional generative adversarial networks, employing adversarial loss to preserve intermediate-to-high frequency details. In Han et al. [29], a GAN model is employed to synthesize rich and diverse brain MR images from existing MR images. In Nie et al. [18], [36], a 3D FCN model was trained to transform MRI to CT images using an adversarial strategy to train the FCN, which enforces the generated images to be more realistic. Finally, additional works have investigated methods for MRI to CT image synthesis [37]-[39], multimodal MRI synthesis [40], [41], and high-quality PET synthesis [42].

The studies that most relevant to ours can be seen in [6], [21], [43]. Specifically, in [6], the author utilized a U-Net-like model to synthesize full-dose CE-MRI from zero-dose pre-contrast MRI and the 10% low-dose postcontrast MRI. In [21], a 3D U-Net model was

developed to generate CE-MRI, which utilizes 10 multiparametric MRI sequences acquired prior to GBCA application as inputs. As a result, their model demonstrates a peak signal to noise ratio (PSNR) of 22.967dB and a structural similarity index (SSIM) of 0.872 dB for the whole brain region. In [43], the author utilized the residual attention U-Net architecture to estimate CE-MRI from non-contrast T2 MRI for cerebral blood volume (CBV) mapping in mice brain.

## III.  Materials and Methods

### A.  Data

Our dataset was acquired at UCLA on Siemens 3 Tesla MRI systems as part of standard-of-care for brain tumor patients. The protocol used was consistent with the International Standard Brain Tumor Protocol [44] and includes 3D MPRAGE T1-weighted pre- and post-contrast imaging, axial 2D T2-weighted imaging, and axial 2D diffusion-weighted imaging used in the calculation of the ADC map. A total of 426 scans from 300 brain tumor patients were included in this study. The data includes two parts, A and B. Set A consisted of 411 scans. It was used for training purposes and therefore further subdivided in 369 scans for training, and 42 scans randomly selected for validation. Set B contained 15 test samples with precise tumor masks and were used to evaluate the quantitative performance on tumor regions. Note that scans in Set B are patients from the UCLA brain tumor trial (IRB# 14-001261), and were selected randomly from the available data making sure both enhancing and non-enhancing tumor were part of the cohort. The experts with more than 10 years of experience created the tumor ROIs as part of the clinical trial reads.

Pre-contrast T1, T2 and ADC map, were utilized as input images and the contrast-enhanced T1 is utilized as the target image. Note that apparent diffusion coefficient (ADC) maps was chosen to augment T1 and T2 because it is independent of these image contrasts and may provide additional information for CE-T1 image synthesis. ADC maps were derived from standard, isotropic diffusion weighted images (DWIs) with and without diffusion weighting according to the standardized brain tumor imaging protocol (BTIP) [45]. Simply, ADC was calculated from b=0, 500, and 1000 s/mm2 by fitting the equation $ADC = -1/b*ln(S(b)/S0)$, where b = 500 and 1,000 s/mm2, ln is the natural log, S(b) is the signal intensity for an MR image at the given b-value, and S0 is the signal intensity of the MR image without any diffusion weighting (b=0).

All the sequences of the MRI data were co-registered to match the targeted 3D contrast-enhanced T1. Bilinear interpolation was utilized to resize all the MRI data to the volume size of 192×256×192 voxels. To remove the skull, brain masks were created for the 3D contrast-enhanced T1 sequences utilizing FSL's brain extraction tool [46]. Besides, in order to remove the side effects of the background slices, in the training and evaluation stage, we only selected the foreground slices and remove the top and bottom background slices which are less informative. Finally, all MRI scans were pre-processed by image equalization and the intensity values of the voxels within the brain region were normalized to [0,1].

## B.  Model Architecture

Let $\mathbf{X} = [\mathbf{X}_{T1}; \mathbf{X}_{T2}; \mathbf{X}_{ADC}] \in \mathbb{R}^{h \times w \times d \times 3}$ denote the MRI sequences with three MRI scans, and $\mathbf{Y} \in \mathbb{R}^{h \times w \times d \times 1}$ denote the contrast-enhanced T1 sequence. To synthesize the CE-T1 sequences from the non-enhanced MRI scans, a 3D FCN model is designed to work as a non-linear mapping function $f_{\boldsymbol{\theta}}$ such that $f_{\boldsymbol{\theta}}: \mathbf{X} \to \mathbf{Y}$, where $\boldsymbol{\theta}$ is the model parameter to be learned. As shown in Fig. 2, the introduced 3D FCN generator is composed of stacked 3D convolutional layers, batch normalization layers, and non-linear activation layers. The first stage is a stem network [26] which is composed of three individual FCN branches that used to handle different input modalities. The resulting feature maps corresponding to different modalities are then fused by concatenation. Following the stem net are repeated multi-resolution subnetworks with generation multi-scale fusion stages. We start from a high-resolution subnetwork in the second stage and add high-to-low resolution subnetworks one-by-one gradually. Specifically, in the second stage, a $3 \times 3 \times 3$ convolution with stride 2 is used to obtain the 2× downsampling feature maps. In the third stage, 4× downsampling feature maps are obtained from the higher resolution feature maps. As a result, we have three different resolution feature maps for different subnetworks, which corresponds to multiple-scale information. To fuse the multi-resolution information comprehensively, multi-scale fusion stages are introduced, which ensure multi-resolution information exchange across different parallel subnetworks. Specifically, in the 3rd and 4th stages, different subnetworks aggregate the feature maps from the other parallel subnetworks. During the multi-scale fusion stages, upsample and downsample operators are utilized to match the size of the feature maps in different subnetworks. In the last stage, feature maps from different branches are fused by concatenation, and a ResBlock is utilized to obtain the final CE-T1.

As illustrated in Fig. 2, Let $F_{mn}$ denotes the feature maps generated by the $m$-th stage and the $n$-th subnetwork, then the feature maps can be calculated as: $F_{11} = \phi(T1) \oplus \phi(T2) \oplus \phi(ADC)$, $F_{12} = D_2(F_{11})$, $F_{21} = \phi(F_{11})$, $F_{22} = \phi(F_{12})$, $F_{23} = D_2(F_{22})$, $F_{31} = \phi(F_{21} \oplus U_2(F_{22}))$, $F_{32} = \phi(F_{22} \oplus D_2(F_{21}))$, $F_{33} = \phi(F_{23} \oplus D_4(F_{21}))$, $F_{41} = \phi(F_{31} \oplus U_2(F_{32}) \oplus U_4 F_{33}))$, $F_{42} = \phi(D_2(F_{31}) \oplus F_{32} \oplus U_2(F_{33}))$, $F_{43} = \phi(D_4(F_{31}) \oplus D_2(F_{32}) \oplus F_{33})$ and $F_5 = F_{41} \oplus U_2(F_{42}) \oplus U_4(F_{43})$. Here, $\phi(x)$ denotes the feature mapping function determined by the FCN module parameters, $\oplus$ denotes feature map concatenation along channel dimension, $D_2$ denotes 2x downsampling, $D_4$ denotes 4x downsampling, $U_2$ denotes 2x upsampling and $U_4$ denotes 4x upsampling. The FCN module consists of four ResBlocks and each ResBlock is composed of two "Conv-BN-Relu" layers. The widths (number of feature maps) of the three parallel subnetworks are 64, 128, and 256. Detailed information regarding the model architecture can be seen in our source code, which is available at https://github.com/chenchao666/Contrast-enhanced-MRI-Synthesis.

The advantages of the presented model are four fold. First, we utilize three MRI scans as inputs and employ three individual stem nets for different modalities, which is able to preserve the modality-specific information. Second, the presented model maintains a high resolution representation throughout the processing pipeline, and contains three parallel subnetworks that can generate and process multi-scale information in parallel. Third, the repeated multi-scale fusion stages ensure better feature fusion across different scale. Fourth, 3D convolution was utilized to exploit additional information from neighboring slices.

### C. Loss Function

Let $\mathbf{X}$ denote the input non-enhanced MRI sequences and $\mathbf{Y}$ denote the contrast-enhanced T1. Our goal is to learn a mapping function $f$ which can generate the CE-T1 sequences $\widehat{\mathbf{Y}} = f_\theta(\mathbf{X})$, such that the synthetic CE-T1 is close to the ground truth $\mathbf{Y}$. The loss function utilized to train the model consists of three terms: pixel-wise MAE loss $\mathscr{L}_{MAE}$, Structural Similarity loss (SSIM) $\mathscr{L}_{SSIM}$ and a local loss $\mathscr{L}_{local}$ to focus performance on tumor regions.

• **Pixel-wise Loss:** The MAE loss and MSE loss are the most widely used pixel-wise losses for image synthesis. We found that using MSE loss resulted in blurrier images compared to MAE loss in this task. Therefore, the pixel-wise MAE loss was utilized in our model, which is given as

$$\mathscr{L}_{MAE} = \|f_\theta(\mathbf{X}) - \mathbf{Y}\|_1$$

• **SSIM Loss:** Using pixel-wise loss alone may ignore image structures. Therefore, we also utilize SSIM loss [47], which is perceptually motivated and leads to more realistic images. The SSIM loss is defined as

$$\mathscr{L}_{SSIM} = \frac{1}{n} \sum_{i=1}^{n} \|1 - SSIM(f_\theta(\mathbf{X})_i, \mathbf{Y}_i)\|_1$$

where $n$ denotes the number of slices of the output 3D MRI sequences, and $\mathbf{Y}_i$ denotes the $i$-th slice of the ground truth CE-T1. $SSIM(\mathbf{x}, \mathbf{y})$ outputs a scalar between 0 and 1, which indicates the structural similarity between images $\mathbf{x}$ and $\mathbf{y}$. The definition of the SSIM metric can be seen in IV-A.

• **Local Loss:** Tumor regions are of particular interest, but account for a very small proportion of voxels in the entire MRI seqeuences. This data imbalance problem leads to under-fitting and poor performance for the tumor regions. Therefore, we introduce a local loss to increase the contribution of the tumor regions. The local loss is defined as

$$\mathscr{L}_{local} = \|(f_\theta(\mathbf{X}) - \mathbf{Y}) \odot \mathbf{M}\|_1$$

where $\mathbf{M}$ is a binary mask of the tumor regions, and $\odot$ denotes voxel-wise multiplication. Since it is very expensive to assign voxel-level labels for each slice, we do not have precise tumor masks for the training samples. Fortunately, compared to the non-enhanced T1 images, the tumors and vessels are significantly enhanced in the CE-T1 images due to the utilization of GBCAs. Therefore, we can calculate a rough tumor mask by thresholding the difference of the T1 and CE-T1 images as follows,

$$\mathbf{M} = \begin{cases} 1 & \mathbf{Y} - \mathbf{X}_{T1} > \delta \\ 0 & \text{else} \end{cases}$$

where $\delta$ is the threshold to control the size of the mask, $\mathbf{X}_{T1}$ is the non-enhanced T1w image, and $\mathbf{Y}$ is the contrast-enhanced T1 image. Fig. 3 shows several examples of the generated binary masks. The generated masks are able to identify regions that are highly enhanced, such as tumors and vessels. These regions are used with our local loss to correct for data imbalance.

• **Overall Loss:** The overall loss function is defined as the weighted sum of all the three terms,

$$\mathcal{L}_{local} = \lambda_1 \mathcal{L}_{MAE} + \lambda_2 \mathcal{L}_{SSIM} + \lambda_3 \mathcal{L}_{local} \tag{1}$$

where $\lambda_1$, $\lambda_2$ and $\lambda_3$ are trade-off parameters to balance the contribution of each loss term.

### D. Implementation Details

The proposed network was implemented in Python with the Keras library and trained on an NVIDIA DGX system with eight NVIDIA V100 CPUs and 512G memory. The Adam optimizer was utilized for training. Since feeding the whole 3D MRI sequences into the model leads to out-of-memory (OOM) problem, we follow [18], [39] to adopt a patch-based strategy for training, and set the batch size to three. In each iteration step, three-slices of MRI sequence with size $3 \times 256 \times 192$ are randomly sampled from each sequence volume. Therefore, we have three input channels with size $3 \times 3 \times 256 \times 192$ and one output with the size $3 \times 256 \times 192$. Model training is divided into two stages. In the first stage, we set $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, and $\lambda_3 = 1.0$, and train the model for the first 40 epochs with a learning rate $lr = 0.0001$. In the second stage, we alter the trade-off parameter of the local loss by setting $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $\lambda_3 = 10$ and fine-tune the model for another 10 epochs with the learning rate $lr = 0.00001$. We empirically the threshold utilized to obtain the brain mask to $\delta = 0.1$ throughout the experiments.

## IV. Experiments

In this section, we first introduce the baseline models and evaluation metrics that we utilized in the experiments. Then, the qualitative and quantitative performance of different models are presented. Finally, we provide ablation experiments that show the impact of different input MRI scans and the impact of the introduced local loss.

### A. Baseline Model and Evaluation Metric

• **Baseline Model:** To evaluate the effectiveness of our proposed model, we implemented a 2D U-Net, a 3D U-Net, as well as the 2D version of our proposed network. All the baseline models are evaluated on the same training and test sets. For the 2D U-Net, we expanded the model depth to 2x of the original U-Net network. For the 3D U-Net, we expanded the model depth to 1.5x of the original model depth.

• **Evaluation Metric:** The quantitative performance of our model and the baseline models was measured using mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Given two images $\mathbf{y} \in \mathbf{Y}$ and $\hat{\mathbf{y}} \in \widehat{\mathbf{Y}}$, where $\mathbf{y}$ is the ground truth image and $\hat{\mathbf{y}}$ is the predicted image. Then,

$$MAE(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{1}{\Omega_{\mathbf{Y}}} \|\mathbf{Y} - \widehat{\mathbf{Y}}\|_1$$

$$MSE(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{1}{\Omega_{\mathbf{Y}}} \|\mathbf{Y} - \widehat{\mathbf{Y}}\|_2^2$$

$$PSNR(\mathbf{Y}, \widehat{\mathbf{Y}}) = 10 \log_{10}(\frac{MAX_I^2}{MSE})$$

$$SSIM(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{1}{\Phi_{\mathbf{Y}}} \sum_{\mathbf{y}, \widehat{\mathbf{y}}} \frac{(2\mu_y \mu_{\widehat{y}} + C1)(2\sigma_{y\widehat{y}} + C2)}{(\mu_y^2 + \mu_{\widehat{y}}^2 + C1)(\sigma_y^2 + \sigma_{\widehat{y}}^2 + C2)}$$

where $\Omega_{\mathbf{Y}}$ is the number of voxels in $\mathbf{Y}$, $\Phi_{\mathbf{Y}}$ is the number of slices in $\mathbf{Y}$, $\mu_{\mathbf{y}}$ and $\sigma_{\mathbf{y}}$ are the mean and variance value of image $\mathbf{y}$, and $\sigma_{y\widehat{y}}$ is the covariance between the ground truth image $\widehat{\mathbf{y}}$ and the predicted image $\widehat{\mathbf{y}}$. $MAX_I$ is the maximum value of the image $\mathbf{y}$. Theoretically, lower MAE values and higher PSNR and SSIM values indicate better image generation quality. The statistical significance of experimental results was evaluated using paired t-tests.

## B. Experimental Results

**Qualitative and Visual Assessment—**The results of representative test samples are shown in Fig. 4. The first two rows demonstrate a normal subject without any tumor and the remaining six rows demonstrate subjects with tumors. We compared the visual performance between our model (both 2D and 3D model) and the 2D/3D U-Net, which are widely used techniques for medical image synthesis tasks [6], [21]. As can be seen, the results reveal several interesting observations. First, both U-Net models and our proposed model generate promising visual performance for the normal patients. The vessels and high-frequency details in the generated images are very close to the ground truth images. Second, since the tumor voxels are limited relative to the entire MRI volume, models are more likely to over-fit to normal regions. As a result, the performance of the tumor regions is much worse than the performance on the normal regions. Third, compared with the U-Net model, which misses or under-estimates most of the tumors for the abnormal samples, our model often yields better visual performance for the tumor regions. Fourth, compared with the 2D model, the 3D model achieves better performance, especially for tumor regions. We believe this is because the 3D model takes advantage of the information from the neighboring slices. Fifth, while achieving promising visual performance for the whole brain MRI, our model sometimes misses or under-estimates some tumors, especially for those that are not distinct enough in the non-contrast images. Finally, Fig. 5 shows two representative MRI sequences in test set B. As observed, our model synthesized satisfactory contrast enhanced T1 images for different slices of the given 3D volume.

**Quantitative Evaluation—**As shown in Table I, we show the quantitative performance of our proposal and comparison methods in test set A. All metrics are computed based on the brain region. Our model significantly outperforms the U-Net model in PSNR and SSIM metrics and the 3D FCN model outperforms its 2D counterpart, which is consistent with our visual assessment. Specifically, U-Net(3D) outperforms U-Net(2D) by 0.35 points in PSRN (p-value=0.021), and Ours(3D) outperforms Our(2D) by 0.37 points (p-value=0.018), which demonstrates the effectiveness of utilizing 3D spatial information. Besides, the proposed Ours(3D) significantly outperforms the U-Net(3D) by more than one point (P-value=0.00037), we believe this is because our proposed model maintains a high resolution representation throughout the processing pipeline and contains three parallel subnetworks with multi-scale fusion stages. Compared with an existing method [21] that performs virtual contrast enhancement with deep learning, our proposal has more than five points improvement in PSNR, and has five points improvement in SSIM. Note that in [21], the author utilized 10-channel multiparametric MRI data as input while we only utilize three, which is a subset of their data. Our model outperforms [21] by a large margin even with less input data, which demonstrates the superiority of our proposed framework.

In order to evaluate the quantitative performance on the tumor region, we also collected 15 test patients with precise tumor masks in set B. The test performance in set B is shown in Table II, with performance on both brain region and tumor region presented. The overall performance on the brain region is similar to the results on set A. Our proposed method significantly outperforms the U-Net and existing work [21]. For results on tumors, the proposed model outperforms the U-Net(3D) model by a large margin (p=0.0036), and the best performance on tumors is 21.2 in PSNR. Note that the quantitative performance on tumors is far from perfect and much worse than the performance on the whole brain region, this is because the tumor pixels are out-of-distribution and the model therefore tends to underestimate on tumor regions. It is worth noting that [21] utilizes a U-Net shape model to segment the tumor masks, which they utilize to evaluate the performance on tumors. These masks were not reviewed by a radiologist and therefore their quantitative performance on tumors may be inaccurate due to segmentation errors.

**Comparison with State-of-the-Arts—**In addition to the U-Net structure that are used for MRI virtual contrast enhancement [6], [21], we also compare our proposal with several state-of-the-art medical image synthesis methods, including Pix2Pix [28], DECNN [37], LA-GANs [42] and MedGAN [48]. Quantitative comparison between our proposal and other state-of-the-art methods is presented in Table IV. Pipx2Pix produces the worst result, while DECNN and LA-GANs achieve similar performance and both outperform Pix2Pix. The MedGAN outperforms the previous methods by a large margin, we believe this is because MedGAN utilizes a cascade U-Blocks as generator which is deeper and better designed than the generator in [28], [37]. Compared with the MedGAN, our proposal shows impressive improvement (p-value=0.0064) on brain region, and also significantly outperforms the MedGAN on tumor region by more than 1.3 points (p-value<1e-4), which demonstrate the superiority of the introduced framework. Note that the performance improvement on the tumor region is much significant than the improvement on the whole brain region, this is because the comparison methods does not take into the performance on tumor regions,

while we introduce a local loss to improve the performance on tumors. The significant improvement on tumors also demonstrate the effectiveness of the local loss.

## C. Ablation Study

**Impact of the Model Architecture—**To demonstrate the effectiveness of the proposed framework, we compare the performance of the final model with four degraded model architectures, which are: (1) Model-A: utilize one FCN module to handle different input modalities, i.e., the three modalities (T1, T2 and ADC) are fused in the input layer; (2) Model-B: remove the 2x downsampling subnetwork and the 4x downsampling subnetwork, only utilizing the high-resolution branch as the FCN generator; (3) Model-C: remove the 1x high resolution subnetwork; (4) Model-D: remove the repeated multi-scale fusion in stage 3 and stage 4. Table V shows the qualitative comparison between different model architectures. As can be seen, our proposed model yields notable improvement over the comparison degraded model architectures. In particular, the final model outperforms the Model-A by more than 0.5 points in PSNR, which demonstrates that employ three individual FCN modules for different input modalities is able to preserve the modality-specific information and lead to better performance. The final model outperforms the Model-B by more than 1 point, this is because it is difficult for the model to capture the global information only using the high resolution branch. Besides, The final model also outperforms the Model-C by more than 0.8 points on average, which indicates the effectiveness of maintaining the high resolution representation. Furthermore, the final model also outperforms the degraded Model-D, which proves the effectiveness of using repeated multi-scale fusion stages.

**Impact of Different Input Modalities—**To investigate the influence of different input modalities, we train the model with different sequences. Specifically, in addition to using all three modalities (T1, T2 and ADC) as inputs, we also train the model: (1) using only T1 as input, and (2) using both T1 and T2 modalities as inputs. Experimental results are presented in Table III. We can conclude from the results that: (1) using T1 alone as input obtains a satisfactory performance; (2) the incorporated T2 and ADC modalities introduced additional information, which further improves performance; and (3) compared to ADC, T2 is more informative and improves performance significantly. Besides, we also present the visual performance of training the model with different input modalities in Fig. 6.

## D. Parameter Sensitivity Analysis

**Sensitivity of the Local Loss—**In order to investigate the influence of the introduced local loss, we used five groups of representative trade-off parameters to train the model. The parameters $(\lambda_1, \lambda_3)$ were selected from $\{(1.0, 1.0), (1.0, 10), (1.0, 30), (0.1, 10), (0.01, 10)\}$. Note that we can not only increase the $\lambda_3$ to improve the influence of the tumor regions as it will cause the gradient to be too large and the network will not converge. Therefore, we promote the influence of the local loss by increasing a ratio $r = \frac{\lambda_3}{\lambda_1}$. $\lambda_2$ is set to the same value as $\lambda_1$ throughout the experiments. Results are presented in Fig. 7 and indicate that by increasing the ratio $r = \frac{\lambda_3}{\lambda_1}$, the PSNR performance of the brain region decreased and the PSNR performance of the tumor region increased. We present two representative samples in

Fig. 8 to demonstrate the influence of local loss visually. By increasing the ratio $\frac{\lambda_3}{\lambda_1}$, tumor enhancement becomes more salient while the overall image becomes more blurry, which is consistent with our quantitative results. The experimental results show that one suitable ratio can be set to $r = \frac{\lambda_3}{\lambda_1} = 100$, which leads to better enhancement results for tumors and satisfactory performance for the whole brain.

**Sensitivity of the threshold $\delta$**—To demonstrate the robustness of the thresholded mask, we also perform a parameter sensitivity analysis experiment on the thresholds $\delta$. Fig. 9 shows the variation of PSNR performance on brain and tumor regions when the threshold is changed from $\delta \in \{0.001, 0.01, 0.03, 0.05, 0.1, 0.15, 0.2, 0.3\}$. As can be seen, the PSNR performance on the brain region increases as the threshold increases and gradually stabilize when $\delta > 0.1$. Besides, the PSNR performance on the tumor region increases first and then decrease as $\delta$ increases, and shows a bell-shaped curve. We believe this is because when the threshold is set to $\delta < 0.03$, the generated mask M will include more non-tumor areas, and when $\delta > 0.15$, the generated mask M will underestimate the tumor regions. Therefore, based on the performance on brain and tumor region, the best threshold can be set to $\delta \in$ [0.03, 0.15].

## E. Inference with Missing Modalities

The proposed model works well when all three input modalities are available. However, rather than having complete three modalities, it is common to have missing modalities in clinical scenarios. To understand how the model performs when only a subset of modalities are available, we visualize the inference performance when: (b) T1 is unavailable; (c) T2 is unavailable; (d) ADC is unavailable; (e) both T2 and ADC are unavailable in Fig. 10. Note that we utilize a zero matrix to represent the unavailable input modality during inference. The results shows that the model performance will be severely degraded when one or two modalities are not available. Besides, compared with the model performance when T2 or ADC is unavailable, the model performs much worse when T1 is unavailable. To ensure that the model is able to produce satisfactory results when only a subset of inputs are available, we also train a T1→T2 and a T1→ADC synthesis model utilizing a similar framework. In this way, when T2 or ADC is unavailable, we can first generate the T2 and ADC and then utilize the generated data to synthesize the contrast-enhanced T1 image. Fig. 11 shows the visual performance of our trained T1→T2 and T1→ADC synthesis model, which demonstrates that our proposed framework also generates promising results for cross-modal image synthesis task. Fig. 10(f) illustrates the model prediction of using T1, synthetic T2 and synthetic ADC as input, which shows that using the synthetic input data, our model can generate similar results as using full three inputs.

## V.  Discussion

Several studies have presented methods to generate synthetic contrast. Briefly, Gong et al. [6] proposed a 2D U-Net-like model to synthesize the full-dose postcontrast images from precontrast and low dose images. In Kleesiek at al. [21], a 3D Bayesian U-Net was utilized to predict contrast enhanced images from 10 multiparametric zero-dose MRI sequences, and

Sun et al. [43] proposed a 2D residual attention U-Net to produce contrast in mice brain MR images directly from noncontrast structural images. Compared to these previous studies, our model has the following advantages:

- In [6], [43] only one MRI sequence was used for generating full-dose MRI compared to 10 multiparametric MRI scans in [21]. We believe a single low-dose or noncontrast MRI may miss important information, while the utilization of 10 MRI scans requires a long time to acquire. According to the International Standardized Brain Tumor Imaging Protocol (BTIP) [44], we utilized three informative noncontrast MRI scans (T1, T2 and ADC) for postcontrast MRI synthesis. Our ablation results suggest that using all three of these sequences maximizes performance.

- To investigate the feasibility of predicting contrast-enhanced MRI sequences from non-contrast or low-contrast MRI sequences, 60 patients are used in [6], 82 patients are used in [43], and [43] test their idea in mice as a proof of concept. Our study utilized more than 400 patients, allowing us to train a deeper FCN model and obtain state-of-the-art performance.

- Different from previous methods [6], [21] that utilize the off-the-shelf model architectures (2D/3D Unet) and loss function, we present a 3D high-resolution FCN model that maintains high-resolution information throughout the fully convolution stage and aggregates the multi-scale information in parallel. As a result, our model outperforms the 2D/3D U-Net counterparts by more than 1 point in PSNR, and also outperforms several state-of-the-art medical synthesis methods. In addition, we introduced a local loss to improve performance in tumor regions.

- Previous studies [6], [21] obtained imperfect enhancement results in vessels and tumors due to the difficulty of the problem. For example, in Gong et al. [6], enhancement results appear rough in vessels compared to our model. Furthermore, our method achieves a result of 28dB in PSNR on the whole brain region, which is similar to Gong et al. [6] despite the fact that our method requires no contrast agent compared to the the low-dose MRI sequences used as input in their work. In terms of PSNR, our model also outperformed Kleesiek et al [21]. by a large margin even with less input data.

While our model demonstrates promising results, there are several limitations. First, since the tumor regions account for a very small proportion of the entire MRI sequences, the performance on these regions remains sub-optimal. As we do not have precise tumor masks for training, the introduced local loss that we used to balance the contribution of the tumors can be improved. Providing precise tumor masks for the local loss during training will likely improve performance further and is in our future work. In addition, some advanced methods in data imbalance learning or long-tailed distribution learning, such as BBN [49], can be introduced to balance the performance of the whole brain region and tumor region. Second, while our cohort was larger than any previously published cohort for the task, performance can likely be further improved by including more training patients, especially a large number

of abnormal patients with high diversity. Beyond increasing and diversifying the dataset, future work will also investigate recent advancements in data augmentation.

## VI. Conclusion

In conclusion, the objective of this investigation was to formulate and implement a deep learning model to generate contrast-enhanced MRI sequences from non-contrast MRI sequences, which was expected to eliminate the risk of gadolinium deposition during standard-of-care for brain tumor patients. For this purpose, the largest dataset for the task of MRI virtual contrast enhancement was explored, and a novel high resolution 3D FCN model was designed, which showed superior performance than the counterparts. Besides, we also introduced a local loss to re-balance the contribution of the tumor regions, which leaded to improved performance on tumors. We demonstrate promising visual and numerical results and obtain state-of-the-art performance. The results suggest great potential in substituting the GBCAs with deep learning to obtain the contrast information in brain MRI. Future work will focus on defective performance on abnormal regions.

## Acknowledgments

## References

[1]. Choi JW and Moon W-J, "Gadolinium deposition in the brain: current updates," Korean journal of radiology, vol. 20, no. 1, pp. 134–147, 2019. [PubMed: 30627029]

[2]. Fraum TJ, Ludwig DR, Bashir MR, and Fowler KJ, "Gadolinium-based contrast agents: A comprehensive risk assessment," Journal of Magnetic Resonance Imaging, vol. 46, no. 2, pp. 338–353, 2017. [PubMed: 28083913]

[3]. Kanda T, Nakai Y, Oba H, Toyoda K, Kitajima K, and Furui S, "Gadolinium deposition in the brain," Magnetic resonance imaging, vol. 34, no. 10, pp. 1346–1350, 2016. [PubMed: 27613998]

[4]. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB et al. , "Gadolinium deposition in the brain: summary of evidence and recommendations," The Lancet Neurology, vol. 16, no. 7, pp. 564–570, 2017. [PubMed: 28653648]

[5]. Khawaja AZ, Cassidy DB, Al Shakarchi J, McGrogan DG, Inston NG, and Jones RG, "Revisiting the risks of mri with gadolinium based contrast agents—review of literature and guidelines," Insights into imaging, vol. 6, no. 5, pp. 553–558, 2015. [PubMed: 26253982]

[6]. Gong E, Pauly JM, Wintermark M, and Zaharchuk G, "Deep learning enables reduced gadolinium dose for contrast-enhanced brain mri," Journal of magnetic resonance imaging, vol. 48, no. 2, pp. 330–340, 2018. [PubMed: 29437269]

[7]. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, and Sánchez CI, "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, pp. 60–88, 2017. [PubMed: 28778026]

[8]. Ronneberger O, Fischer P, and Brox T, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[9]. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, and Ronneberger O, "3d u-net: learning dense volumetric segmentation from sparse annotation," in International conference on medical image computing and computer-assisted intervention. Springer, 2016, pp. 424–432.

[10]. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, Rueckert D, and Glocker B, "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," Medical image analysis, vol. 36, pp. 61–78, 2017. [PubMed: 27865153]

[11]. Pereira S, Pinto A, Alves V, and Silva CA, "Brain tumor segmentation using convolutional neural networks in mri images," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1240–1251, 2016. [PubMed: 26960222]

[12]. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, and Summers RM, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1285–1298, 2016. [PubMed: 26886976]

[13]. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K et al. , "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.

[14]. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, and Knoll F, "Learning a variational network for reconstruction of accelerated mri data," Magnetic resonance in medicine, vol. 79, no. 6, pp. 3055–3071, 2018. [PubMed: 29115689]

[15]. Mardani M, Gong E, Cheng JY, Vasanawala SS, Zaharchuk G, Xing L, and Pauly JM, "Deep generative adversarial neural networks for compressive sensing mri," IEEE transactions on medical imaging, vol. 38, no. 1, pp. 167–179, 2018. [PubMed: 30040634]

[16]. He Z, Zhu Y-N, Qiu S, Wang T, Zhang C, Sun B, Zhang X, and Feng Y, "Low-rank and framelet based sparsity decomposition for interventional mri reconstruction," IEEE Transactions on Biomedical Engineering, 2022.

[17]. Bahrami K, Shi F, Rekik I, and Shen D, "Convolutional neural network for reconstruction of 7t-like images from 3t mri using appearance and anatomical features," in Deep Learning and Data Labeling for Medical Applications. Springer, 2016, pp. 39–47.

[18]. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Wang Q, and Shen D, "Medical image synthesis with deep convolutional adversarial networks," IEEE Transactions on Biomedical Engineering, vol. 65, no. 12, pp. 2720–2730, 2018. [PubMed: 29993445]

[19]. Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, and Išgum I, "Deep mr to ct synthesis using unpaired data," in International workshop on simulation and synthesis in medical imaging. Springer, 2017, pp. 14–23.

[20]. Dar SU, Yurt M, Karacan L, Erdem A, Erdem E, and Çukur T, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," IEEE transactions on medical imaging, vol. 38, no. 10, pp. 2375–2388, 2019. [PubMed: 30835216]

[21]. Kleesiek J, Morshuis JN, Isensee F, Deike-Hofmann K, Paech D, Kickingereder P, Köthe U, Rother C, Forsting M, Wick W et al. , "Can virtual contrast enhancement in brain mri replace gadolinium?: a feasibility study," Investigative radiology, vol. 54, no. 10, pp. 653–660, 2019. [PubMed: 31261293]

[22]. Long J, Shelhamer E, and Darrell T, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[23]. Chen L-C, Papandreou G, Kokkinos I, Murphy K, and Yuille AL, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834–848, 2017. [PubMed: 28463186]

[24]. Zhao H, Shi J, Qi X, Wang X, and Jia J, "Pyramid scene parsing network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.

[25]. Newell A, Yang K, and Deng J, "Stacked hourglass networks for human pose estimation," in European conference on computer vision. Springer, 2016, pp. 483–499.

[26]. Sun K, Xiao B, Liu D, and Wang J, "Deep high-resolution representation learning for human pose estimation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2019, pp. 5693–5703.

[27]. Lai W-S, Huang J-B, Ahuja N, and Yang M-H, "Deep laplacian pyramid networks for fast and accurate super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 624–632.

[28]. Isola P, Zhu J-Y, Zhou T, and Efros AA, "Image-to-image translation with conditional adversarial networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134.

[29]. Han C, Hayashi H, Rundo L, Araki R, Shimoda W, Muramatsu S, Furukawa Y, Mauri G, and Nakayama H, "Gan-based synthetic brain mr image generation," in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 734–738.

[30]. Johnson J, Alahi A, and Fei-Fei L, "Perceptual losses for real-time style transfer and super-resolution," in European conference on computer vision. Springer, 2016, pp. 694–711.

[31]. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Kalra MK, Zhang Y, Sun L, and Wang G, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," IEEE transactions on medical imaging, vol. 37, no. 6, pp. 1348–1357, 2018. [PubMed: 29870364]

[32]. Li R, Zhang W, Suk H-I, Wang L, Li J, Shen D, and Ji S, "Deep learning based imaging data completion for improved brain disease diagnosis," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2014, pp. 305–312.

[33]. Bahrami K, Shi F, Zong X, Shin HW, An H, and Shen D, "Reconstruction of 7t-like images from 3t mri," IEEE transactions on medical imaging, vol. 35, no. 9, pp. 2085–2097, 2016. [PubMed: 27046894]

[34]. Xiang L, Qiao Y, Nie D, An L, Lin W, Wang Q, and Shen D, "Deep auto-context convolutional neural networks for standard-dose pet image estimation from low-dose pet/mri," Neurocomputing, vol. 267, pp. 406–416, 2017. [PubMed: 29217875]

[35]. Huang Y, Shao L, and Frangi AF, "Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6070–6079.

[36]. Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, and Shen D, "Medical image synthesis with context-aware generative adversarial networks," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2017, pp. 417–425.

[37]. Xiang L, Wang Q, Nie D, Zhang L, Jin X, Qiao Y, and Shen D, "Deep embedding convolutional neural network for synthesizing ct image from t1-weighted mr image," Medical image analysis, vol. 47, pp. 31–44, 2018. [PubMed: 29674235]

[38]. Emami H, Dong M, Nejad-Davarani SP, and Glide-Hurst CK, "Generating synthetic cts from magnetic resonance images using generative adversarial networks," Medical physics, vol. 45, no. 8, pp. 3627–3636, 2018.

[39]. Lei Y, Wang T, Liu Y, Higgins K, Tian S, Liu T, Mao H, Shim H, Curran WJ, Shu H-K et al., "Mri-based synthetic ct generation using deep convolutional neural network," in Medical Imaging 2019: Image Processing, vol. 10949. International Society for Optics and Photonics, 2019, p. 109492T.

[40]. Chartsias A, Joyce T, Giuffrida MV, and Tsaftaris SA, "Multimodal mr synthesis via modality-invariant latent representation," IEEE transactions on medical imaging, vol. 37, no. 3, pp. 803–814, 2017. [PubMed: 29053447]

[41]. Zhou T, Fu H, Chen G, Shen J, and Shao L, "Hi-net: hybrid-fusion network for multimodal mr image synthesis," IEEE transactions on medical imaging, 2020.

[42]. Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, Lin W, Wu X, Zhou J, and Shen D, "3d auto-context-based locality adaptive multi-modality gans for pet synthesis," IEEE transactions on medical imaging, vol. 38, no. 6, pp. 1328–1339, 2018. [PubMed: 30507527]

[43]. Sun H, Liu X, Feng X, Liu C, Zhu N, Gjerswold-Selleck SJ, Wei H-J, Upadhyayula PS, Mela A, Wu C-C et al., "Substituting gadolinium in brain mri using deepcontrast," in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 908–912.

[44]. Ellingson BM, Bendszus M, Boxerman J, Barboriak D, Erickson BJ, Smits M, Nelson SJ, Gerstner E, Alexander B, Goldmacher G et al. , "Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials," Neuro-oncology, vol. 17, no. 9, pp. 1188–1198, 2015. [PubMed: 26250565]

[45]. Ellingson B, Bendszus M, Boxerman J, Barboriak D, Erickson B, Smits M, Nelson S, Gerstner E, Alexander B, Goldmacher G et al. , "Jumpstarting brain tumor drug development coalition

imaging standardization steering c. consensus recommendations for a standardized brain tumor imaging protocol in clinical trials," Neuro Oncol, vol. 17, no. 9, pp. 1188–1198, 2015. [PubMed: 26250565]

[46]. Jenkinson M, Pechaud M, Smith S et al., "Bet2: Mr-based estimation of brain, skull and scalp surfaces," in Eleventh annual meeting of the organization for human brain mapping, vol. 17. Toronto., 2005, p. 167.

[47]. Zhao H, Gallo O, Frosio I, and Kautz J, "Loss functions for image restoration with neural networks," IEEE Transactions on computational imaging, vol. 3, no. 1, pp. 47–57, 2016.

[48]. Armanious K, Jiang C, Fischer M, Küstner T, Hepp T, Nikolaou K, Gatidis S, and Yang B, "Medgan: Medical image translation using gans," Computerized medical imaging and graphics, vol. 79, p. 101684, 2020. [PubMed: 31812132]

[49]. Zhou B, Cui Q, Wei X-S, and Chen Z-M, "Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9719–9728.

(a) T1    (b) T2    (c) ADC    (d) CE-T1
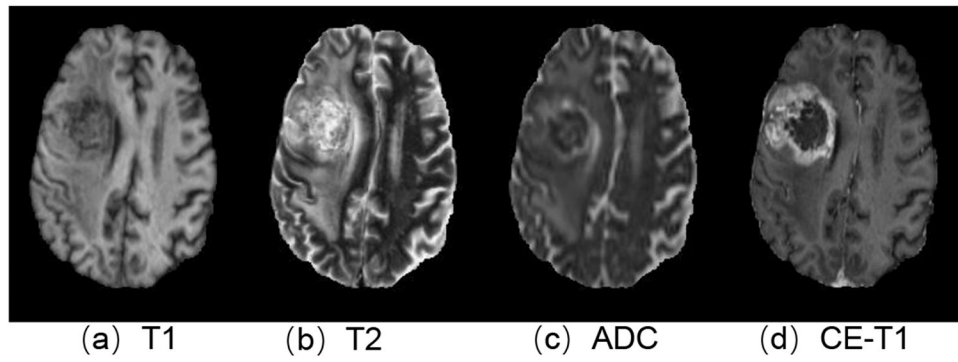
**Fig. 1:**
Contrast and non-contrast MRI sequences used for brain tumor diagnosis and clinical monitoring. Three non-contrast scans, T1, T2, and the ADC were used to estimate the contrast-enhanced T1-weighted image (CE-T1) using a 3D FCN generator.
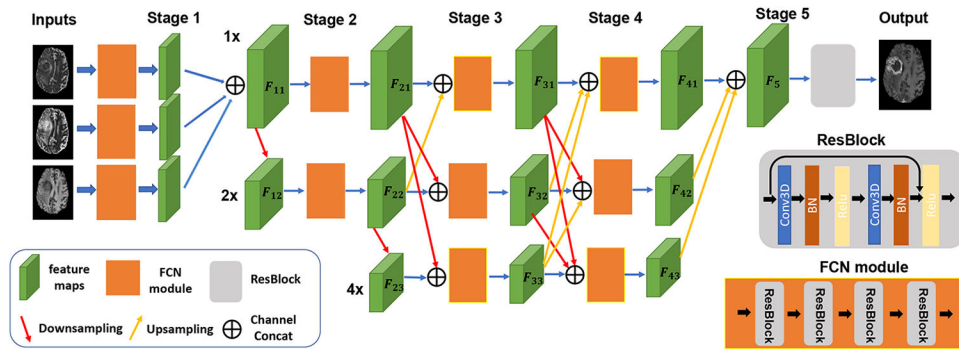
**Fig. 2:**

Overview of our proposed model. A high-resolution FCN model was trained as a generator to synthesize contrast-enhanced T1 images. Three non-contrast scans, including T1, T2, and ADC, were utilized as input images.
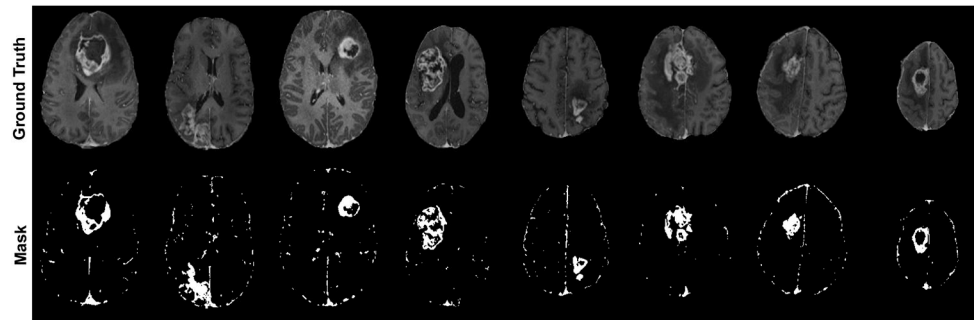
**Fig. 3:**
Illustration of six representative CE-T1 images and corresponding binary masks. The generated masks are able to identify the tumors, vessels and other high frequency details, which are enhanced by GBCAs.
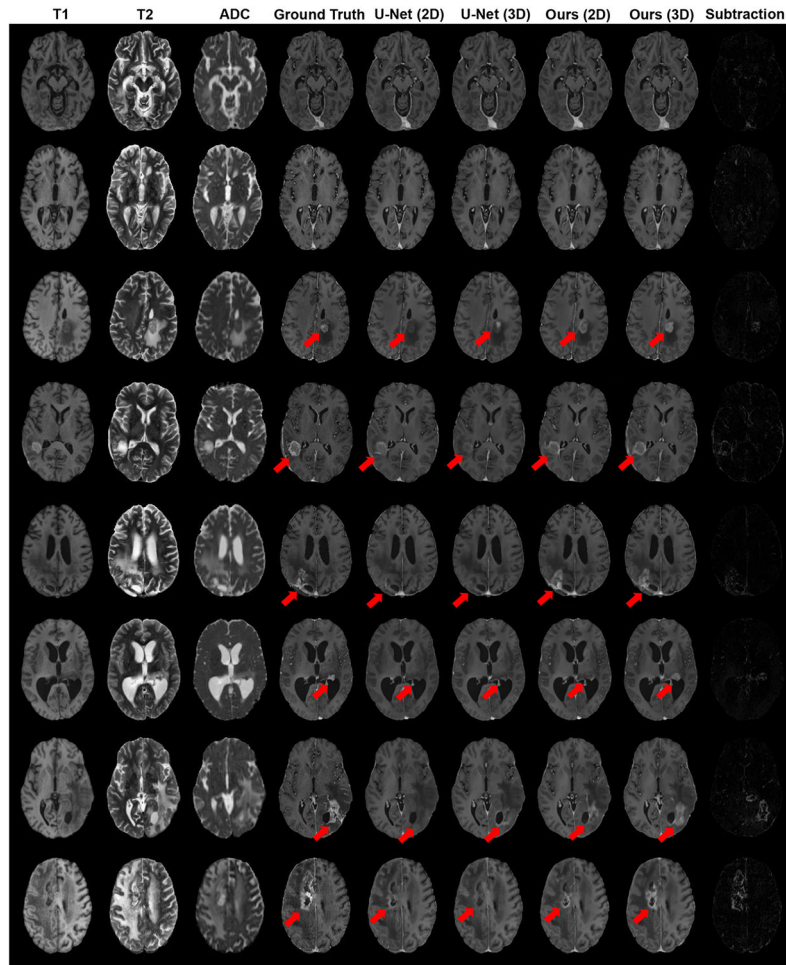
**Fig. 4:**
Qualitative evaluation of our proposal and baseline models. From left to right, input T1, input T2, input ADC, ground truth image CE-T1, synthetic image of a 2D U-Net, synthetic image of a 3D U-Net, synthetic image of the proposed 2D FCN model, synthetic image of the proposed 3D FCN model, and the absolute difference between the results of our 3D FCN and the ground truth. The first two rows are from normal patients and the other rows are from patients with tumors. Different rows are from different subjects in the test set.
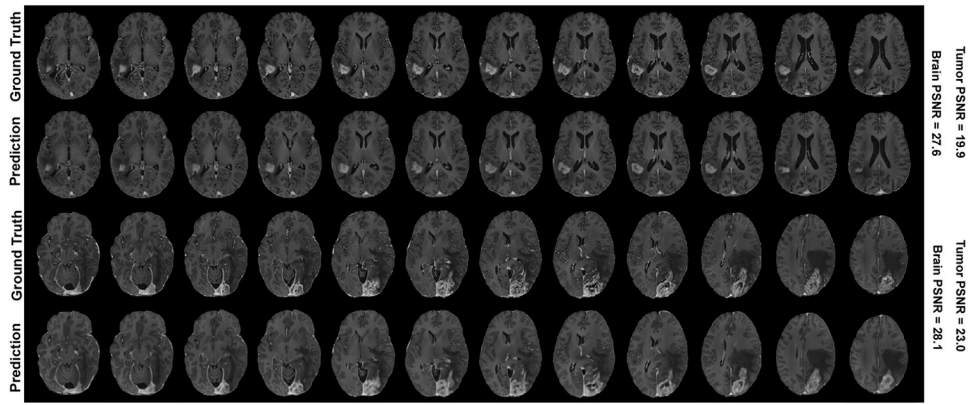
**Fig. 5:**
Visual assessment of our proposed method in two representative test samples in set B. From top to bottom, the ground truth slices of test patient A, the synthetic CE-T1 for test patient A, the ground truth slices of test patient B, the synthetic CE-T1 for test patient B. The synthetic CE-T1 images are generated by the proposed 3D FCN model. The images in the same row represent different slices of the same subjects.
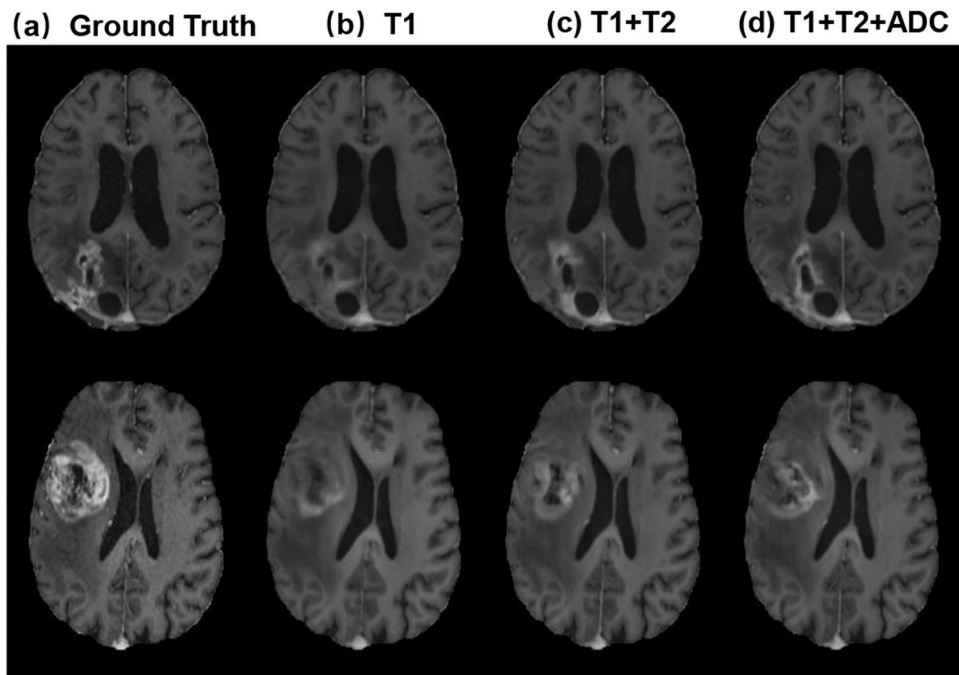
**Fig. 6:**
Visual performance of training the model by different input modalities. (a) Ground truth images, (b) results by training the model with T1 as input, (c) results by training the model with T1 and T2 as inputs, and (d) results by training the model with T1, T2 and ADC as inputs.
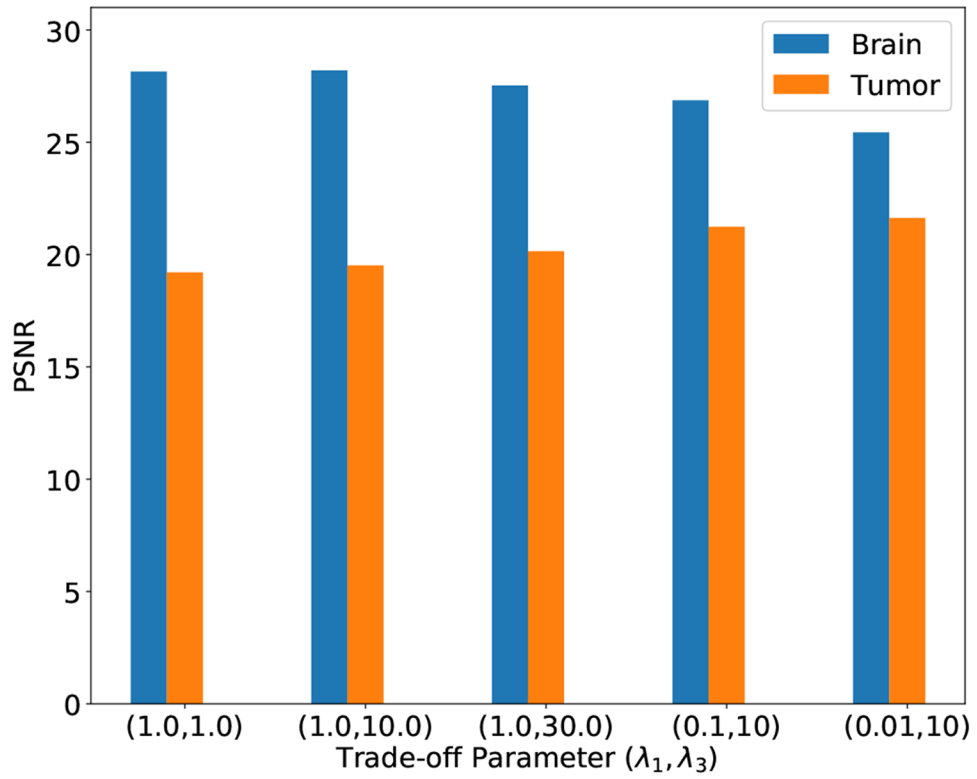
**Fig. 7:**
Quantitative evaluation of the impact of local loss. The blue bar indicates the PSNR performance on the Brain region of Set A and the orange bar indicates the PSNR performance of the Tumor region of Set B. Note that $\lambda_2$ is set to the same value as $\lambda_1$ throughout the experiments.
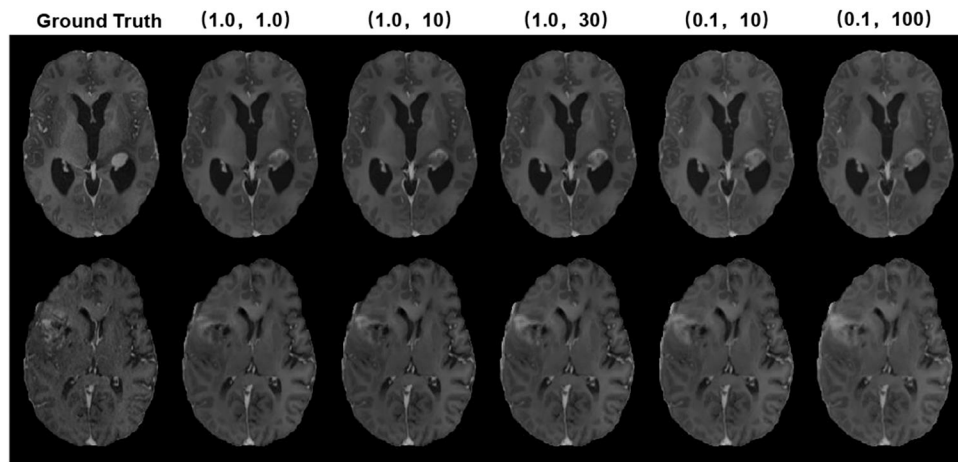
**Fig. 8:**
Visual performance of training the model by using different trade-off parameters $(\lambda_1, \lambda_3)$. From left to right, the trade-off parameters are set to $(\lambda_1, \lambda_3)$ = {(1.0, 1.0), (1.0, 10), (1.0, 30), (0.1, 10), (0.01, 10)}.
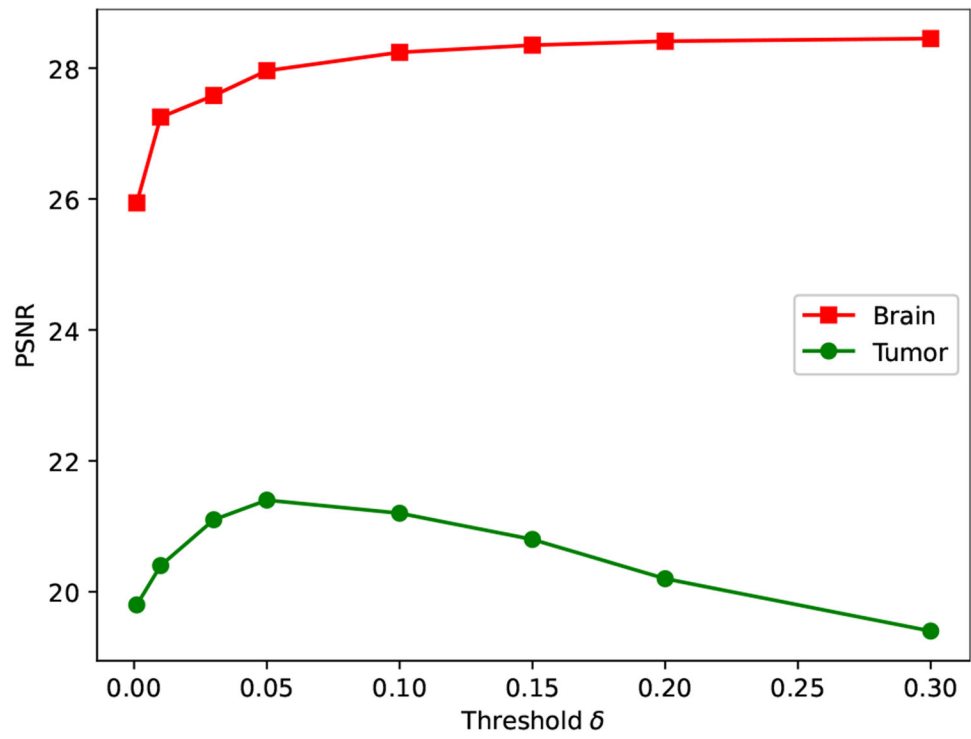
**Fig. 9:**
PSNR performance on brain region tumor region when threshold is changed from $\delta \in$ {0.001, 0.01, 0.03, 0.05, 0.1, 0.15, 0.2, 0.3}.
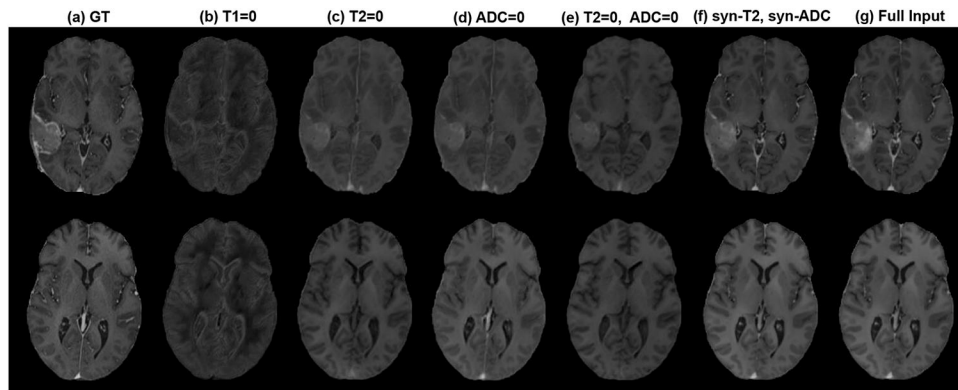
**Fig. 10:**
Illustration of (a) Group Truth (GT) and model performance when: (b) T1 is unavailable; (c) T2 is unavailable; (d) ADC is unavailable; (e) T2 and ADC are unavailable; (f) utilize T1, synthetic T2 and synthetic ADC as input; (g) utilize full three inputs. Note that we utilize a zero matrix to represent the unavailable input modality during inference.
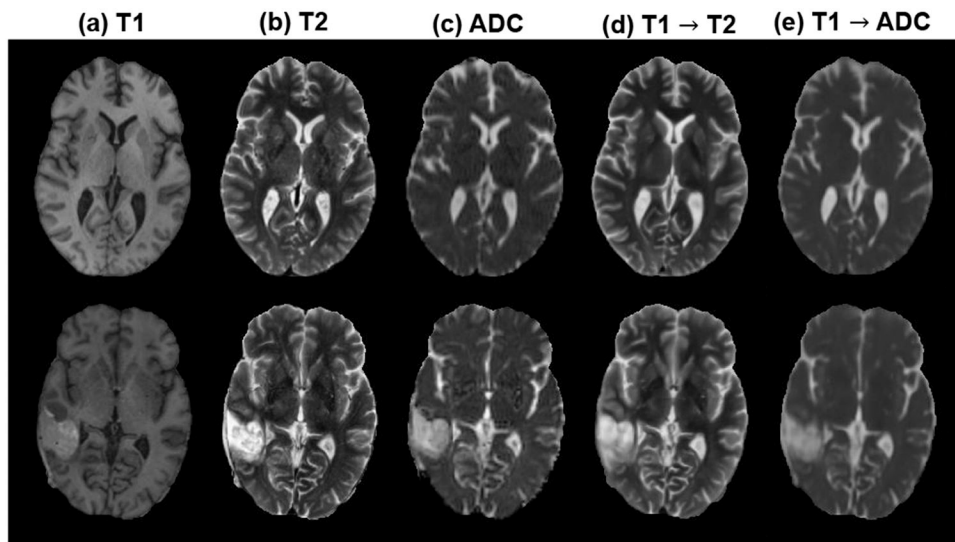
**Fig. 11:**
The visual performance of the T1→T2 and T1→ADC synthesis model. (a) T1; (b) T2; (c) ADC; (d) generated T2 by T1→T2 synthesis model; (e) generated ADC by T1→ADC synthesis model.

**TABLE I:**

Quantitative comparison between our model and baseline models. The models are evaluated on test set A, and the performance on the whole brain region is presented.

| Model | MAE | PSNR | SSIM |
|---|---|---|---|
| **Ref** [21] | N/A | 22.97±1.16 | 0.872±0.031 |
| **U-Net** (2D) | 0.033±0.005 | 26.86±1.05 | 0.905±0.038 |
| **U-Net** (3D) | 0.032±0.004 | 27.21±1.18 | 0.908±0.038 |
| **Ours** (2D) | 0.030±0.005 | 27.87±1.30 | 0.915±0.039 |
| **Ours** (3D) | **0.029±0.005** | **28.24±1.26** | **0.923±0.041** |

**TABLE II:**

Quantitative performance evaluated on test set B. The performance on the brain region and tumor region are presented.

| Model | PSNR | | SSIM |
|---|---|---|---|
| | **Brain** | **Tumor** | **Brain** |
| **Ref** [21] | 22.97±1.16 | 20.15±4.70 | 0.872±0.031 |
| **UNet (2D)** | 26.44±1.40 | 18.45±2.22 | 0.896±0.022 |
| **UNet (3D)** | 26.79±1.26 | 18.89±2.38 | 0.899±0.022 |
| **Ours (2D)** | 27.22±1.21 | 19.64±2.59 | 0.903±0.023 |
| **Ours (3D)** | **27.62±1.34** | **21.2±2.36** | **0.909±0.023** |

**TABLE III:**

Comparison experiments by using different modalities as inputs.

| Input | | T1 | T1+T2 | T1+T2+ADC |
|---|---|---|---|---|
| **A-Brain** | **PSNR** | 27.4±1.28 | 27.9±1.24 | **28.24±1.26** |
| | **SSIM** | 0.914±0.040 | 0.920±0.040 | **0.923±0.041** |
| **B-Brain** | **PSNR** | 26.7±1.23 | 27.3±1.27 | **27.62±1.34** |
| | **SSIM** | 0.898±0.022 | 0.905±0.023 | **0.909±0.023** |
| **B-Tumor** | **PSNR** | 19.8±2.32 | 20.8±2.30 | **21.2±2.36** |

**TABLE IV:**

Quantitative comparison between our proposal and state-of-the-art medical image synthesis methods on test set B

| Model | PSNR | | SSIM |
|---|---|---|---|
| | **Brain** | **Tumor** | **Brain** |
| **Pix2Pix** [28] | 25.90±1.52 | 18.20±2.64 | 0.887±0.025 |
| **DECNN** [37] | 26.48±1.38 | 18.74±2.31 | 0.898±0.023 |
| **LA-GANs** [42] | 26.23±1.30 | 18.69±2.47 | 0.894±0.024 |
| **MedGAN** [48] | 27.04±1.26 | 19.88±2.42 | 0.901±0.023 |
| **Ours (3D)** | **27.62±1.34** | **21.2±2.36** | **0.909±0.023** |

**TABLE V:**

Qualitative comparison between the final model and four degraded model architectures.

| Model | A-Brain | B-Brain | B-Tumor |
|---|---|---|---|
| **Model-A** | 27.68±1.31 | 27.26±1.28 | 20.63±2.46 |
| **Model-B** | 26.42±1.12 | 26.32±1.20 | 19.54±2.38 |
| **Model-C** | 27.20±1.22 | 27.04±1.19 | 20.39±2.44 |
| **Model-D** | 27.36±1.20 | 27.08±1.18 | 20.68±2.32 |
| **Final-Model** | **28.24±1.26** | **27.62±1.34** | **21.20±2.36** |