




Article

Magnetic Resonance-Based Synthetic Computed Tomography Using Generative Adversarial Networks for Intracranial Tumor Radiotherapy Treatment Planning

Chun-Chieh Wang^{1,2,†}, Pei-Huan Wu^{1,†}, Gigin Lin^{3,4} , Yen-Ling Huang³, Yu-Chun Lin³ ,
Yi-Peng (Eve) Chang⁵ and Jun-Cheng Weng^{1,6,7,8,*} 

- ¹ Department of Medical Imaging and Radiological Sciences, and Graduate Institute of Artificial Intelligence, Chang Gung University, Taoyuan 33302, Taiwan; jjwangucla@gmail.com (C.-C.W.); zxc0914@gmail.com (P.-H.W.)
- ² Department of Radiation Oncology, Chang Gung Memorial Hospital at Linkou, Taoyuan 33302, Taiwan
- ³ Department of Medical Imaging and Intervention and Institute for Radiological Research, Chang Gung Memorial Hospital at Linkou and Chang Gung University, Taoyuan 33302, Taiwan; giginlin@gmail.com (G.L.); b9102091@cgmh.org.tw (Y.-L.H.); jack805@gmail.com (Y.-C.L.)
- ⁴ Clinical Metabolomics Core Lab, Chang Gung Memorial Hospital at Linkou, Taoyuan 33302, Taiwan
- ⁵ Department of Counseling and Clinical Psychology, Columbia University, New York, NY 10027, USA; tiramisueve@gmail.com
- ⁶ Medical Imaging Research Center, Institute for Radiological Research, Chang Gung University and Chang Gung Memorial Hospital at Linkou, Taoyuan 33302, Taiwan
- ⁷ Department of Psychiatry, Chang Gung Memorial Hospital, Chiayi 61363, Taiwan
- ⁸ Department of Medical Imaging and Radiological Sciences, Chang Gung University, No. 259, Wenhua 1st Rd., Guishan Dist, Taoyuan 33302, Taiwan
- * Correspondence: jcweng@mail.cgu.edu.tw; Tel.: +886-3-2118800 (ext. 5394)
- † These authors contributed equally to this work.



Citation: Wang, C.-C.; Wu, P.-H.; Lin, G.; Huang, Y.-L.; Lin, Y.-C.; Chang, Y.-P. (E.); Weng, J.-C. Magnetic Resonance-Based Synthetic Computed Tomography Using Generative Adversarial Networks for Intracranial Tumor Radiotherapy Treatment Planning. *J. Pers. Med.* **2022**, *12*, 361. <https://doi.org/10.3390/jpm12030361>

Academic Editor: Gianluca Ingrassio

Received: 20 January 2022

Accepted: 24 February 2022

Published: 26 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The purpose of this work is to develop a reliable deep-learning-based method that is capable of synthesizing needed CT from MRI for radiotherapy treatment planning. Simultaneously, we try to enhance the resolution of synthetic CT. We adopted pix2pix with a 3D framework, which is a conditional generative adversarial network, to map the MRI data domain into the CT data domain of our dataset. The original dataset contains paired MRI and CT images of 31 subjects; 26 pairs were used for model training and 5 were used for model validation. To identify the correctness of the synthetic CT of models, all of the synthetic CTs were calculated by the quantized image similarity formulas: cosine angle distance, Euclidean distance, mean square error, peak signal-to-noise ratio, and mean structural similarity. Two radiologists independently evaluated the satisfaction score, including spatial, detail, contrast, noise, and artifacts, for each imaging attribute. The mean (\pm standard deviation) of the structural similarity indices (CAD, L2 norm, MSE, PSNR, and MSSIM) between five real CT scans and the synthetic CT scans were 0.96 ± 0.015 , 76.83 ± 12.06 , 0.00118 ± 0.00037 , 29.47 ± 1.35 , and 0.84 ± 0.036 , respectively. For synthetic CT, radiologists rated the results as evincing excellent satisfaction in spatial geometry and noise level, good satisfaction in contrast and artifacts, and fair imaging details. The similarity index and clinical evaluation results between synthetic CT and original CT guarantee the usability of the proposed method.

Keywords: deep learning; generative adversarial net (GAN); attenuation correction; MR-only simulation; radiotherapy planning; brain tumor

1. Introduction

Computed tomography (CT) simulation is a necessary procedure performed after mold customization for every patient who will undergo radiation therapy. It provides information on electron density and geometry. Radiotherapy treatment planning, image reconstruction, and daily treatment guidance are all based on electron density and geometric information

provided by CT. Daily treatment is also based on the CT simulation image for image guidance and positioning error correction. Therefore, the most critical issue of MR-only simulation workflows is retrieving this information only through MRI. It can be done by rigid or deformable registration, but errors are inevitable, so high-quality results cannot be expected. Therefore, a CT-independent method should be developed as a better solution.

Magnetic resonance imaging (MR) and CT are both important medical imaging systems for radiotherapy treatment planning. MR is used to segment the tumor contour and volume in radiotherapy treatment planning, and CT is used to calculate the radiation dose. MR imaging-only radiotherapy planning is a novel application. Electron density information typically acquired from CT images is a prerequisite for attenuation correction and radiotherapy treatment planning. However, MR images represent only longitudinal tissue, transverse relaxation times, and proton-density information. To solve this problem, several methods have been developed to produce synthetic CT images for both MRI-only radiotherapy treatment planning and MRI-based attenuation correction (MRAC) [1–4]. Therefore, the methods that produce accurate attenuation correction on PET–MRI could potentially be applied in MRI-only radiotherapy treatment planning and vice versa. Furthermore, if a single synthetic CT generation method could be used in both applications, there would be no need to use independent sequences or processing pipelines for producing synthetic CT between different systems and modalities. However, a thorough evaluation of the method's robustness in patients with brain tumors should be performed [5].

MR simulation can provide better sensitivity, specificity, and contrast in soft tissue for image-guided radiation therapy (IGRT) treatment planning. However, a remaining challenge lies in obtaining a reliable X-ray attenuation correction map, which is crucial for the calculation of treatment planning. Many novel strategies [6–8] have been introduced to directly estimate bone information for MR imaging-based attenuation correction, including atlas-based methods and image-segmentation-based methods, particularly those using ultrashort echo time and zero echo time (ZTE) approaches [6]. Zero echo time pulse sequences have been used to successfully generate synthetic CT images that can be used for accurate MRAC and MRI-only radiotherapy treatment planning of the brain [9]. Methods based on deep learning have been successfully used to generate synthetic CTs from contrast-enhanced images [10]; however, the effect of contrast agents in synthetic CT generation and their effects on radiation therapy (RT) plan quality have not been studied extensively in an individual study. Although each of these proposed solutions has specific advantages and limitations, the development of rapid and robust MRAC is still currently an unmet need.

Radiotherapy for brain tumors starts at the time of simulation, and the simulation procedure retrieves three-dimensional (3D) images for target delineation and treatment planning. Therefore, magnetic resonance imaging (MRI) has superior soft tissue contrast and improves the target delineation of radiotherapy for brain tumors. Using CT imaging alone, the segmentation of the brain tumor may overestimate the tumor volume. If overestimation can be corrected by MRI, normal tissue damage will be reduced in the future. Unfortunately, MRI cannot be used for radiotherapy planning directly since it lacks electron density information for radiation dose calculation. Although MRI can be registered with CT images and incorporated into radiotherapy planning, this approach is not accurate enough due to misregistration or image distortions [11]. Redundant imaging also increases costs and consumes more time in clinical practice. Developing an MR-only simulation workflow can significantly improve the quality of radiotherapy treatment planning, and time and cost can also be saved.

U-net is proposed by Ronneberger et al. for biomedical image segmentation, named after is the architecture of the model [12]. The structure of u-net is similar to an alphabet "U". The input side is mainly several layers of CNN-base encoder which can reduce dimensions of input data. On the contrary, the output side is a decoder for expanding dimensions. The outstanding design of shortcut connection between encoder and decoder transfer information greatly enhances the performance. Owing to its powerful capacity, the usage of u-net extends further to image translation.

PatchGAN is a classifier for classification tasks [13]. Traditional classifiers present the judgment of input images with a single number, whereas patchGAN does not. The output of patchGAN is a matrix. Each element inside the matrix is the judgment of the corresponding receptive field. That is to say, the mechanism of patchGAN is to separate the original image into different patches and utilize the distinguishing characteristics of all the patches within a matrix.

Generative adversarial networks (GANs) [14], which are made of MLPs, have achieved great success and have spread explosively since being proposed in 2014. The GAN approach in the method of competition constantly modified the parameters of its MLP. There are many GAN extensions and applications currently, and pix2pix [13] is one of them. U-net [12] of pix2pix has shortcut structures that make information pass through from the encoder to the decoder, producing a more precise pattern. At the same time, the patchGAN of pix2pix forces u-net to its limit on generating pseudoimages.

Currently, many researchers try to solve biomedical imaging issues with deep-learning-based approaches. The latest studies have proven that deep-learning methods are able to learn the nonlinear mapping relationship of the MR domain and CT domain in a two-dimensional framework. Now the issue is pushed to a 3D framework. Dong et al. presented a 3D fully convolutional network to estimate pelvic CT images from MRI data. Liu Y et al. took advantage of 3D-based cycleGAN to convert MR to CT [15–17], achieving a great result on 3D-based MRI to CT translation. The advantage of unsupervised-learning cycleGAN is that the dataset where images from two domains are not necessarily paired is easier to prepare. These studies take advantage of a single model to manage their dataset, which is separated into several patches in common. The model of these studies convert a single patch at a time, and all patched predictions are merged at the final stage. The usage of these methodologies is analogous taking a microscope to see a huge object. However, a model has limited capacity to learn all features from a complex 3D-based dataset, and the important features vary from the location of the body. The content of the dataset should be based on the location of the body and each location should have its own corresponding model to deal with. In this paper, we propose a new approach that is expandable to prevent limitations from the model's learning capacity and to enlarge the image size (resolution) of synthetic CT.

2. Methods

2.1. Prepare MR–CT Paired Dataset

This study was approved by the Institutional Review Board of Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan (No. 202002387B0). All methods were carried out in accordance with relevant guidelines and regulations. Both MRI and CT scans of each participant were acquired at Chang Gung Memorial Hospital at LinKou, Taoyuan, Taiwan. Multimodal MRI examinations were performed on a 1.5T MRI scanner (GE, Boston, MA, USA) with a standard head coil. The three-dimensional T1-weighted gradient-echo sequence (MPRAGE) was obtained with repetition time (TR)/echo time (TE)/inversion time (TI)/flip angle (FA) = 7.91 ms/3.27 ms/450 ms/12°; voxel size = 0.39 × 0.39 × 1.0 mm³, and number of average = 2. CT examinations were performed on a GE Light Speed RT16 with a standard brain protocol. Scan parameters were as follows: slice thickness = 1.25 mm of slice thickness, 120 kV of slice thickness, 300 mA of tube current, and 1071 msec of exposure time.

2.2. Data Preprocessing

The original dataset has 31 pairs in total, including 26 pairs for training and 5 pairs for tests. To verify the algorithm and make sure the testing dataset covers symptoms in different degrees simultaneously, the five testing pairs are selected. In comparison with the subjects of training pairs, some subjects of the testing pair had wider tumor volumes and others had previously undergone major surgery; they were the outliers of the full dataset. The details of the dataset are listed in Supplementary Materials Table S1. Each

image pair is made of an MRI and a CT of a subject. Figure 1 shows the data preprocessing workflow diagram, and Supplementary Materials Table S2 shows the modified image parameters after reFOV and resampling for a few more subjects from the training and test data. MRI and CT are different imaging modalities with different voxel sizes (resolutions), FOVs, window widths, and levels of intensity. Dataset preprocess has 3 stages. In the first stage, we define a new FOV for each pair, and any voxels located outside the new FOV are abandoned. In our original dataset, every MR image has a smaller FOV than CT, so the definition of a new FOV is based on MR. After the re-FOV stage, every paired image has the same FOV but a different number of voxels. In the second stage, we use NiBabel, which is a Python library for medical images, for dataset resampling. The resampling method is spline interpolation with an order of 3. Therefore, each paired image has the same FOV, and the number of voxels of the whole dataset is $200 \times 200 \times 128$. In the third stage, every image was normalized and standardized by Formulas (1) and (2), respectively. The intensity of all images are shifted to $-1 \sim 1$.

$$img' = \frac{img - \mu(img)}{\sigma(img)} \tag{1}$$

$$\begin{cases} img'' = 2 \times \left(\frac{img' - m}{d} \right) \\ m = \frac{\max(img') + \min(img')}{2} \\ d = \max(img') - \min(img') \end{cases} \tag{2}$$

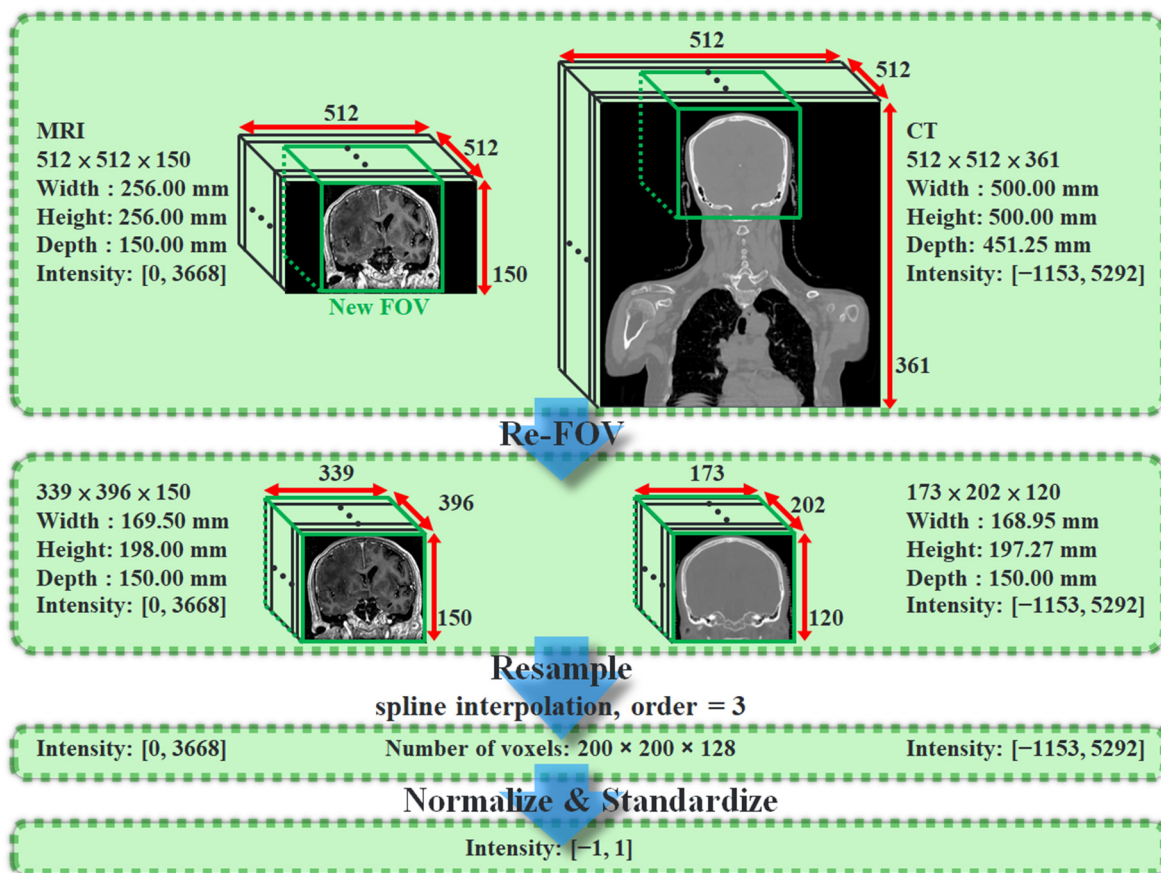


Figure 1. The workflow of data preprocessing. The parameters in Figure 1 come from one of the real cases in the original dataset. The workflow is applied to all pairs of original datasets.

2.3. Patch-Based Datasets

MRI and CT are medical images with high complexity. To increase the similarity of prediction and prevent trainable parameters from exceeding the limitation of hardware, we obtain 5 different patched datasets, denoted as datasets p1, p2, and p5, respectively, from the complete re-FOV dataset. We set the number of voxels to $128 \times 128 \times 128$ for every patch, such that all 3D-pix2pix models can apply with the same architecture. A schematic diagram of patch-based datasets is shown in Figure 2. For example, dataset p1 contains the upper left corner cubes of all subjects in the re-FOV dataset.

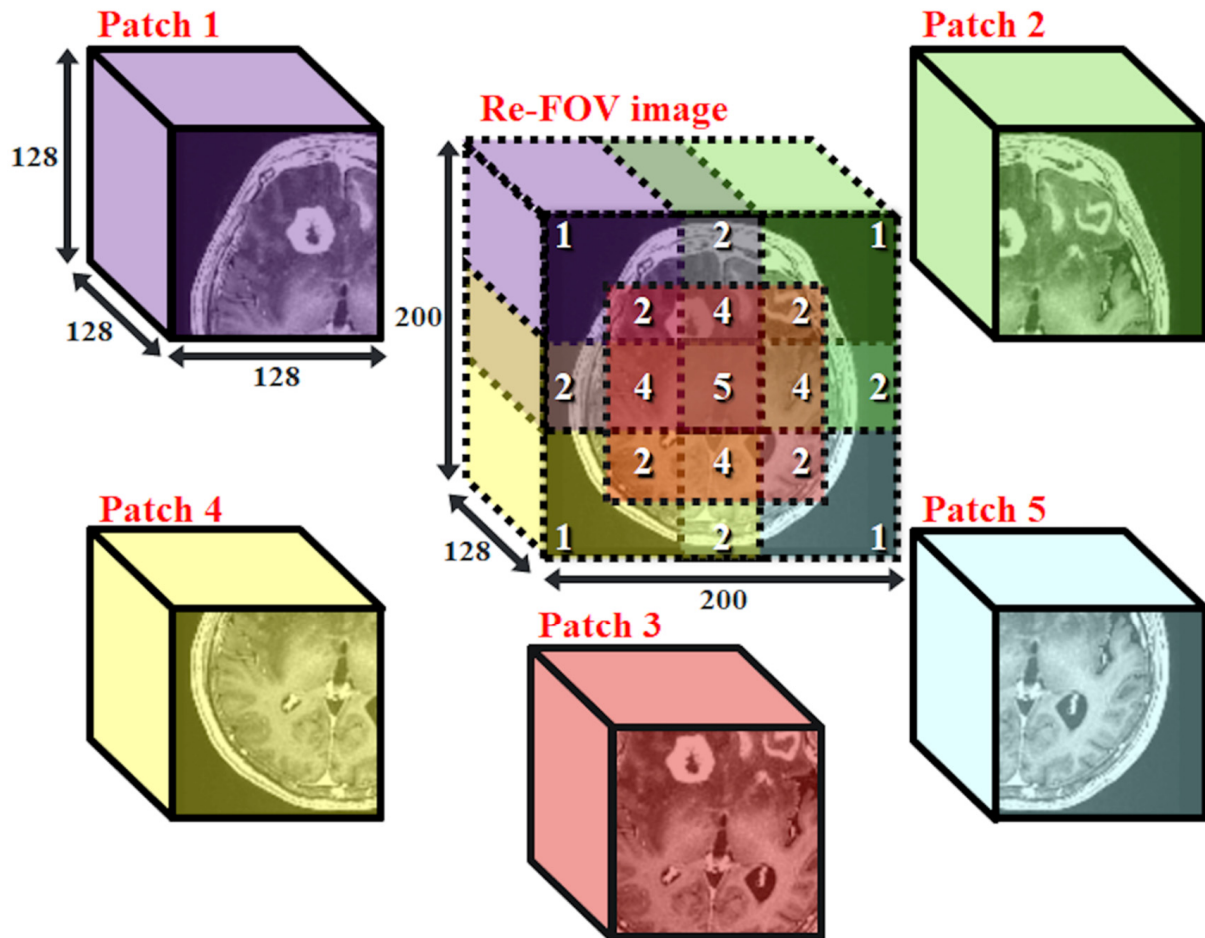


Figure 2. The correlation between five patch and re-FOV image. Patches p1~p5 have their own spatial locations at the coordinates of the re-FOV image. In this paper, we call the collection of the same patch from all subjects the patched dataset.

2.4. Data Augmentation

To overcome the problem of lacking training data, we adopt a data augmentation technique in the training process. Many studies have proven that adopting data augmentation is effective for improving model accuracy for classification tasks [18–20]. We rotate the voxel coordinate of every image, and the rotation angle is different at each epoch. The rotate angle along 3 axes is shown in (3) and denoted as $(\theta_x, \theta_y, \theta_z)$. A rotation angle contains a fixed part and a random part. The fixed part is denoted as θ_f , and the subscript f means fixed. The random part is denoted as θ_r , and subscript r means random. As the iteration increases, the value of θ_f is chosen from 1, -1, 3, -3, 5, -5, 7, and -7, in order. The total 512 combinations are listed as follows:

$$(1 + \theta_r, 1 + \theta_r, 1 + \theta_r), (1 + \theta_r, 1 + \theta_r, -1 + \theta_r), (1 + \theta_r, 1 + \theta_r, 3 + \theta_r),$$

$$\begin{aligned}
 & (1 + \theta_r, 1 + \theta_r, -7 + \theta_r), (1 + \theta_r, -1 + \theta_r, 1 + \theta_r), (1 + \theta_r, -1 + \theta_r, -1 + \theta_r), \\
 & (1 + \theta_r, -1 + \theta_r, -7 + \theta_r), (1 + \theta_r, 3 + \theta_r, 1 + \theta_r), \dots\dots\dots, (1 + \theta_r, -7 + \theta_r, -7 + \theta_r), \\
 & (-1 + \theta_r, 1 + \theta_r, 1 + \theta_r), \dots\dots\dots, (-7 + \theta_r, -7 + \theta_r, -7 + \theta_r), \text{ and } (-7 + \theta_r, -7 + \theta_r, -7 + \theta_r).
 \end{aligned}$$

$$\left\{ \begin{aligned}
 (\theta_x, \theta_y, \theta_z) &= (\theta_f[i] + \theta_r, \theta_f[j] + \theta_r, \theta_f[k] + \theta_r), \text{ where } i, j, k = 0, 1, 2, \dots, 7 \\
 \theta_f &= \{1, -1, 3, -3, 5, -5, 7, -7\} \\
 \theta_r &= \{\theta \mid -1 < \theta < 1\}
 \end{aligned} \right. \tag{3}$$

2.5. Generative Adversarial Nets

Since being proposed by Ian Goodfellow, generative adversarial nets (GANs) have become prosperous in many fields, especially in computer vision (CV). A well-trained GAN is capable of generating pseudoimages that are extremely similar to real images. A GAN is composed of a discriminator and a generator. The task of a discriminator is to judge whether an image is real or not. On the other hand, the generator has entirely opposite goals to the goal of the discriminator. The generator tries to deceive the discriminator such that the discriminator makes the wrong judgment. As the number of iterations increases, the discriminator becomes more capable at judgment, and the generator becomes better at forging. The GANs objective function is shown in (4):

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \tag{4}$$

where the data from the real world are denoted as x ; the random noise is denoted as z ; the discriminator and generator are denoted as D and G , respectively, and the synthetic data are denoted as $G(z)$. The distributions of real-world data and random variables are denoted as $p_{data}(x)$ and $p_z(z)$, respectively. The meaning of subscript $x \sim p_{data}$ is that x belongs to $p_{data}(x)$, and the meaning of subscript $z \sim p_z(z)$ is that z belongs to $p_z(z)$. The symbol \mathbb{E} represents the expected value. The output range of the discriminator is designed to be in the range 0~1. According to the loss function, the parameters of the discriminator are modified to distinguish real data x and synthetic data $G(z)$, and, to deceive the discriminator, the parameters of the generator are modified to forge data $G(z)$ that appear real.

2.6. Pix2pix

Pix2pix is one of the most powerful deep-learning models on image translation for two different styles of images. It is an extended topology of GANs but has a more complex structure. A pix2pix contains a PatchGAN classifier and u-net. The functions of the PatchGAN classifier and u-net are similar to the discriminator and generator, respectively. A normal (or traditional) classifier maps an image onto a single number. In contrast to the normal classifier, the PatchGAN classifier maps an input image onto a $M \times M$ patch, and every element in this patch has its own receptive field from the input image. To do so, the weight number of the classifier can be reduced. It is not necessary to consider the whole input image at a time. The topology of u-net is similar to an autoencoder. The major difference in topology is that u-net has a shortcut between the encoder and decoder but an autoencoder does not. The feature map of the encoder is passed through the shortcut and then concatenated to the feature map of the decoder. Both of them have a bottleneck in the middle of the structure. From the category aspect, pix2pix belongs to a kind of conditional GAN (cGAN) [21] that needs to be provided an extra condition. According to the original paper, the condition can be any type of data, such as a label of class or image data of a

certain modality. The training process restricts the output distribution of the u-net and PatchGAN classifiers. Formula (5) is the loss function of cGAN:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{c,x}[\log D(c, x)] + \mathbb{E}_{c,z}[\log(1 - D(c, G(c, z)))] \tag{5}$$

where we set the MRI as c and CT as x . An additional loss function (6) is applied to the algorithm to enhance sharpness:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{c,x,z}[\|x - G(c, z)\|_1] \tag{6}$$

Therefore, the goal of the algorithm is shown in (7):

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \tag{7}$$

where we set the value of λ to 100. We modified the framework from 2D to 3D to prevent slices from forming discontinuities on pseudoinages. In addition, we also modified hyperparameters for the PatchGAN classifier and u-net, such as the learning rate, to match our dataset.

2.7. Implementation

We use an open source API, TensorFlow 2.4.1, for deep-learning model-building training and testing. The algorithm is deployed on and executed by the server ESC8000 G4 with a GeForce 1080 Ti (Nvidia, Santa Clara, CA, USA). Figure 3 shows the main architecture of a single model used for every patched dataset. The kernel size is denoted as k_y , which means that the kernel size is $y \times y \times y$. The base number of filters (or kernel) is denoted as f_z , which means that the base number of filters is z . We have three combinations of the number of filters and kernel size, which are denoted as $k4_f80$, $k6_f60$, and $k8_f30$. Consequently, we have 15 (5 patches \times 3 filter numbers and kernel size) specific models. For example, the model that has an 80 base filter number and $4 \times 4 \times 4$ kernel size, being trained by the p1 dataset, is denoted as the p1_k4_f80 model. In the algorithm, the stride of 3D convolution and transpose convolution is (2, 2, 2), and the batch size is 1.

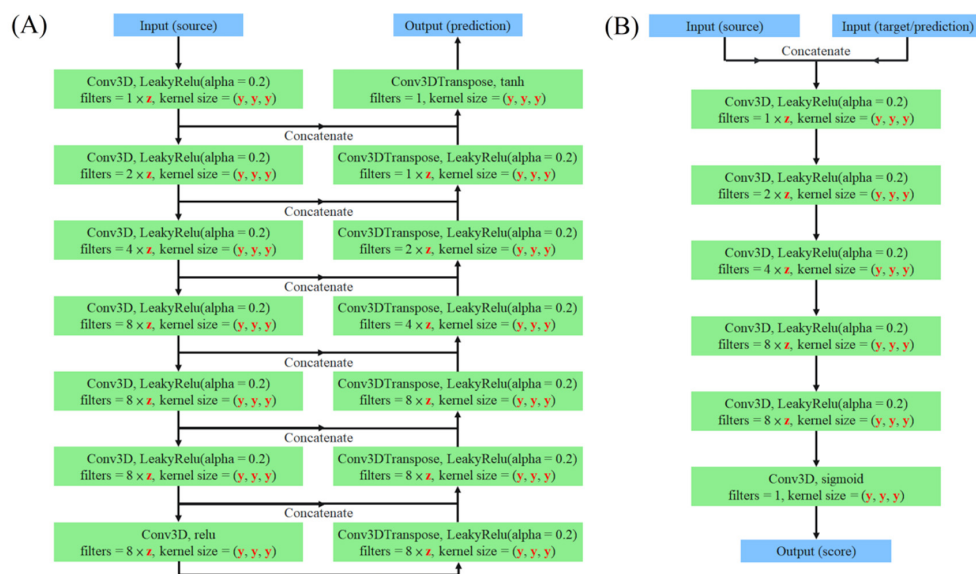


Figure 3. The structure of 3D pix2pix model. A single pix2pix model contains the (A) u-net and a (B) PatchGAN classifier. The task of u-net is to learn the relationship between “source” and “target” and make the distribution of “prediction” approach the distribution of “target” as similarly as possible. In contrast to u-net, the PatchGAN classifier needs to learn how to cause a high score for the real pair of “target” and “source” rather than the pseudopair of “target” and “prediction”.

2.8. Merge Prediction from the k4_f80/k6_f60/k8_f80 Model

The kernel size is an important factor in convolutional neural networks [22–24]. In general, a model with a small kernel is more efficient, and a large kernel is more accurate. The element in the feature map has its own receptive field, which is related to kernel size. That is, if we choose different kernel sizes, it results in obtaining different feature maps. Although a larger kernel considers more information from the former layer to produce feature maps at a time, a larger kernel does not present better performance as long as the kernel size is over a certain number. As shown in Figure 4, we mixed 3 outputs of well-trained models for a single patched dataset. The base kernel sizes of these 3 well-trained models are $4 \times 4 \times 4$, $6 \times 6 \times 6$, and $8 \times 8 \times 8$. The proportion of ingredients of the mixed image is 1:1:1.

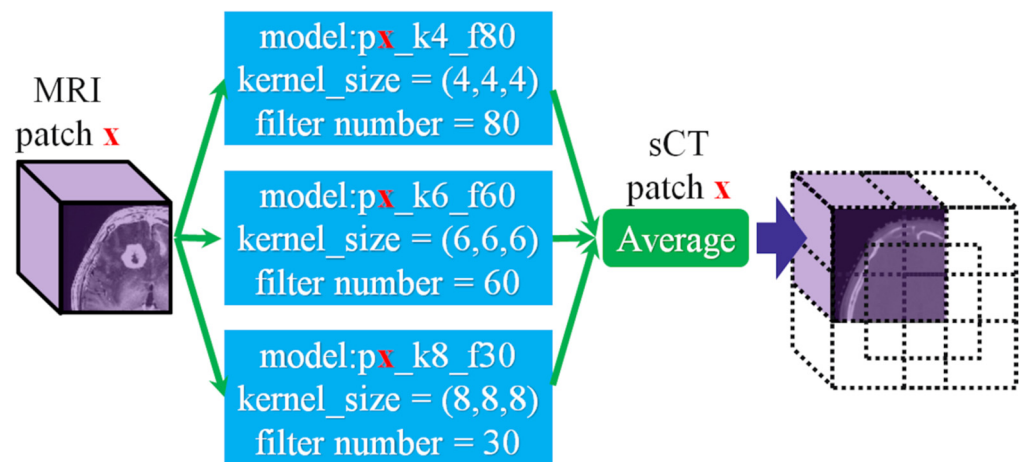


Figure 4. Merging prediction from the k4_f80, k6_f60, and k8_f30 models. A patch of synthetic CT is composed of output from the k4_f80, k6_f60, and k8_f30 models.

2.9. Merge Prediction from the p1~p5 Model

To eliminate discontinuous boundaries that appear at the merged prediction, the center locations of each patch in the re-FOV image are designed such that there is overlap among these patches. We build and train 5 specific models, and each model is specialized for the corresponding patched dataset. After the training process, we create an all-zero image, called the base map, with a size of $200 \times 200 \times 128$. The output value of all patched models is linearly converted to the same output range as the p3 model. Then, based on the location of each specific patch at the re-FOV image, the prediction of each specific model is added back to the base map. Every voxel in the base map is divided by the number of overlapping times. The numbered marks of the re-FOV image in Figure 2 present the amount of overlap in each space.

2.10. Image Similarity Evaluation

The two images *A* and *B* are given and shown in (8):

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix} \tag{8}$$

where a_1, a_2, \dots and a_N are the voxels in image *A*, and b_1, b_2, \dots and b_N are the voxels in image *B*. To objectively determine the difference between images *A* and *B*, we adopt the quantized index for evaluation. Formulas (9)–(11) are used to calculate the cosine angle distance (CAD) [25,26], L2 norm (also known as Euclidean distance) [25,27], and mean structural similarity index (MSSIM) [28], respectively.

CAD:

$$CAD(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \times \|B\|} \tag{9}$$

L2 norm:

$$L2(A, B) = \sqrt{\sum_{i=1}^N (a_i - b_i)^2} \tag{10}$$

MSSIM:

$$\left\{ \begin{aligned} &SSIM(A'_i, B'_i) = [l(A'_i, B'_i)]^\alpha [c(A'_i, B'_i)]^\beta [s(A'_i, B'_i)]^\gamma \\ &l(A'_i, B'_i) = \frac{2^{\mu_{A'_i} \mu_{B'_i} + C_1}}{(\mu_{A'_i})^2 + (\mu_{B'_i})^2 + C_1} \\ &c(A'_i, B'_i) = \frac{2^{\sigma_{A'_i} \sigma_{B'_i} + C_2}}{(\sigma_{A'_i})^2 + (\sigma_{B'_i})^2 + C_2} \\ &s(A'_i, B'_i) = \frac{\sigma_{A'_i B'_i} + C_3}{\sigma_{A'_i} \sigma_{B'_i} + C_3} \\ &w = \{w_i \mid i = 1, 2, \dots, N\} \\ &\mu_{A'_i} = \sum_{j=1}^N w_j a'_{i,j}, \mu_{B'_i} = \sum_{j=1}^N w_j b'_{i,j} \\ &\sigma_{A'_i} = \sqrt{\sum_{j=1}^N w_j (a'_{i,j} - \mu_{A'_i})^2}, \sigma_{B'_i} = \sqrt{\sum_{j=1}^N w_j (b'_{i,j} - \mu_{B'_i})^2} \\ &\sigma_{A'_i B'_i} = \sum_{j=1}^N w_j (a'_{i,j} - \mu_{A'_i})(b'_{i,j} - \mu_{B'_i}) \\ &MSSIM(A, B) = \frac{1}{M} \sum_{i=1}^M SSIM(A'_i, B'_i) \end{aligned} \right. \tag{11}$$

When images A and B are the same, the values of the CAD, L2 norm, and MSSIM are 1, 0, and 1, respectively. These similarity indices provide quantized values and different aspects to evaluate two images. Images A and B are treated as two vectors when the value of CAD is calculated. The angle between these two vectors shows whether image A is similar to image B . The value of the L2 norm value represents the accumulation comparison for voxelwise differences of images A and B . The MSSIM considers three important factors: luminance, contrast, and structure. In Formula (11), M is the number of local patches in full image A . The i -th patch A'_i of image A contains voxels $a'_{i,1}, a'_{i,2}, \dots, a'_{i,N}$; w is a circular-symmetric Gaussian weighting function with a standard deviation of 1.5 samples, normalized to the unit sum $\left(\sum_{i=1}^N w_i = 1\right)$. We take advantage of scikit-image, which is a collection of algorithms for image processing, to calculate MSSIM. It is the function "skimage.metrics.structural_similarity()" that we use, leaving all the parameters as default.

2.11. Clinical Evaluation

Two radiologists independently evaluated the satisfaction score, including spatial, detail, contrast, noise, and artifacts, and categorized them into excellent, good, fair, or bad for each imaging attribute. Reader agreement regarding invasion depth was analyzed using weighted kappa statistics ($0.00 \leq k < 0.40$ indicated poor agreement; $0.40 \leq k \leq 0.70$ indicated fair agreement; $k > 0.70$ indicated excellent agreement). The Mann–Whitney U test was used to compare the clinical satisfaction scores between the bone and soft tissue window images from the synthetic CT.

3. Result

Figure 5 shows the first two testing pairs and their synthetic CT. Each merged prediction is composed of outputs from 15 models. The models of participation are p1_k4_f80, p2_k4_f80, p3_k4_f80, p4_k4_f80, p5_k4_f80, p1_k6_f50, p2_k6_f50, p3_k6_f50, p4_k6_f50,

p5_k6_f50, p1_k8_f30, p2_k8_f30, p3_k8_f30, p4_k8_f30, and p5_k8_f30. Because the algorithm is based on a 3D framework, we plot axial, coronal, and sagittal views for each sample. The similarity indices for all tested CT scans and their corresponding synthetic CT scans are shown in Figure 6. The additional two similarity indices, MSE and PSNR, of CT vs. synthetic CT before and after merging were calculated to evaluate the model performance (Supplementary Materials Table S3).

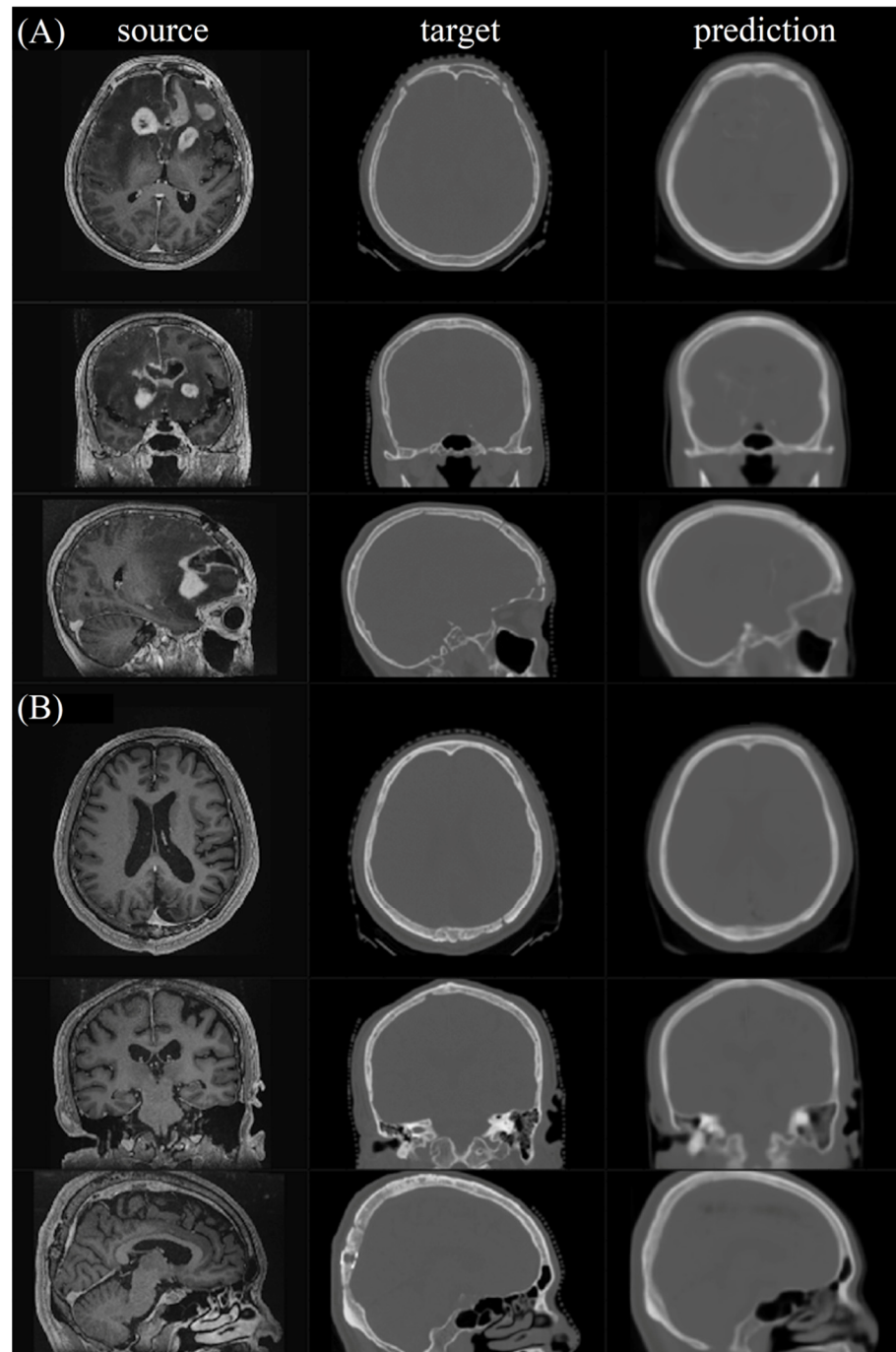


Figure 5. The final merged prediction. (A,B) show the two testing pairs, which are No. 001 and No. 003, and their corresponding merged predictions, respectively.

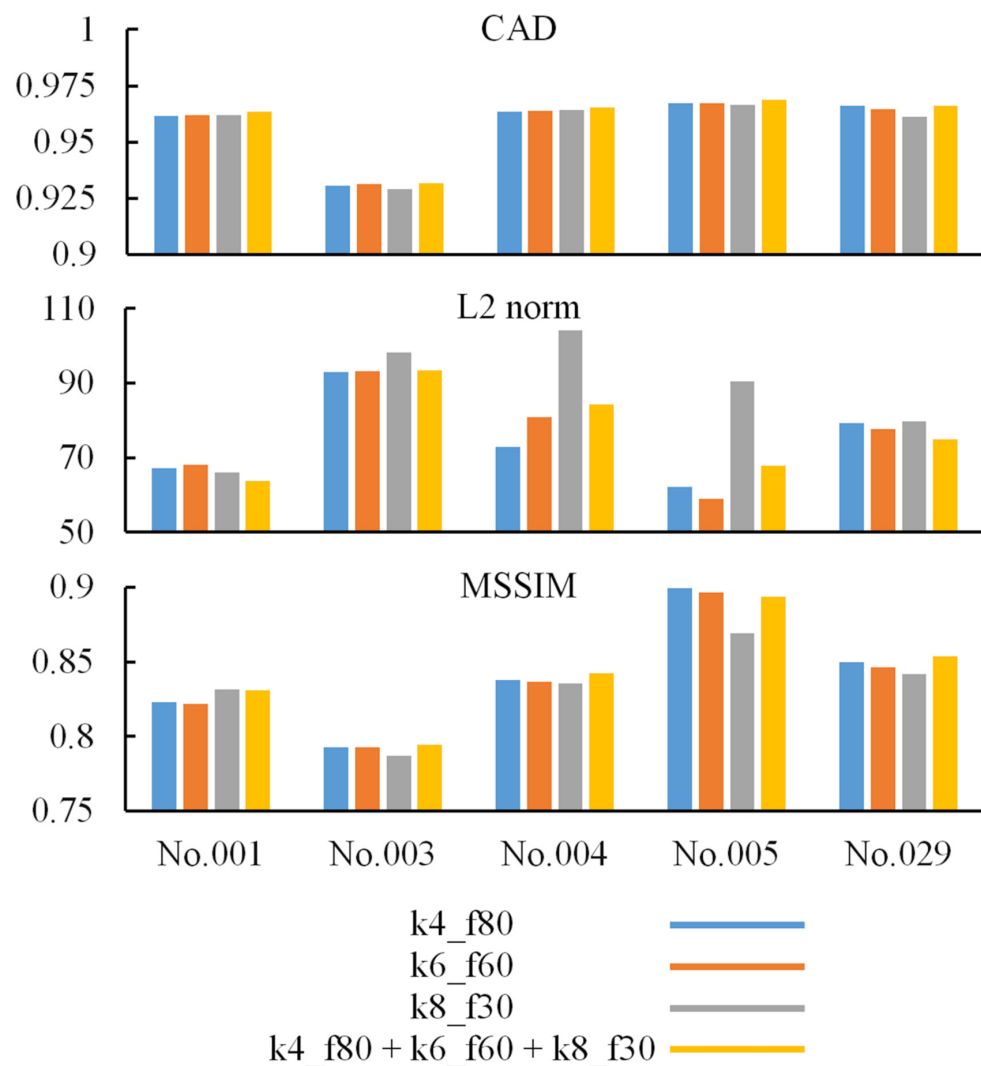


Figure 6. The individual similarity indices before and after merging. To compare the contributions of different kernel sizes, the calculations of similarity indices are performed according to kernel size. The entire synthetic CT evaluated by the indices is made of outputs of p1~p5 models.

For synthetic CT, radiologists valued excellent satisfaction in spatial geometry and noise level, good satisfaction in contrast and artifacts, and fair imaging details for the bone and soft tissue window images (Supplementary Materials Tables S4 and S5). Higher satisfaction scores were observed on the bone window images than on the soft tissue window images for evaluation of the details, contrast, and artifact (Table 1). There was no statistically significant difference in synthetic CT on axial, coronal, or sagittal planes. The reader agreement rate was excellent in terms of spatial, detail, contrast, noise, and artifact in axial, coronal, and sagittal planes for both the bone and soft tissue window images from the synthetic CT. Interestingly, the metallic artifact was reduced, and the air density of the paranasal sinuses and mastoid air cells were well preserved on the bone window images of the synthetic CT (Figure 7a,b). Of note, perifocal hyperdensities on the soft tissue window images of the synthetic CT might lead to a false impression of intracranial hemorrhage, which should have been postoperative encephalomalacia and white matter edema (Figure 7c).

Table 1. Clinical satisfaction score based on the bone and soft tissue window synthetic CT.

	Case	Bone		Soft Tissue		<i>p</i>
		Median	Range	Median	Range	
AXL	spatial	4	3–4	4	3–4	0.71
	detail	2	2–3	1	1–2	<0.001
	contrast	4	2–4	3	3–4	<0.001
	noise	4	3–4	4	3–4	0.88
	artifact	3	1–4	3	2–4	<0.001
COR	spatial	4	3–4	4	3–4	1.00
	detail	2	2–3	2	1–2	<0.001
	contrast	4	2–4	3	3–4	<0.001
	noise	4	4–4	4	3–4	0.32
	artifact	4	3–4	4	2–4	<0.001
SAG	spatial	4	3–4	4	3–4	1.00
	detail	2	2–3	2	1–2	<0.001
	contrast	4	3–4	3	3–4	<0.001
	noise	4	3–4	4	3–4	1.00
	artifact	4	3–4	3	2–4	<0.001

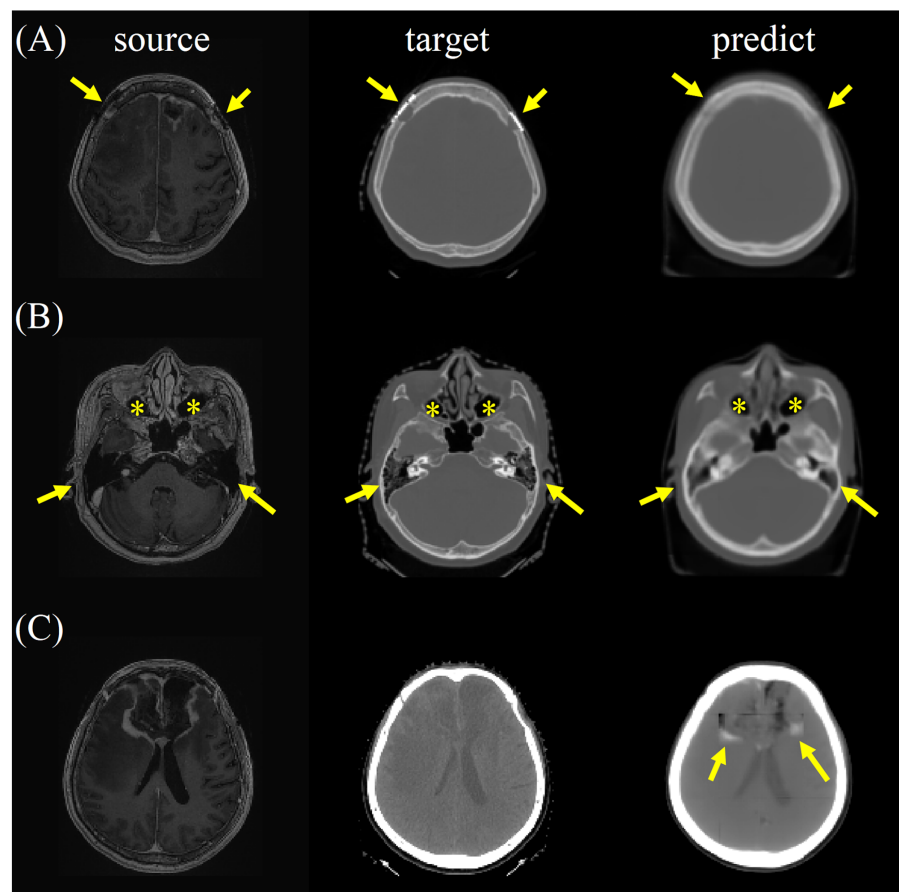


Figure 7. (A) The metallic artifact was reduced (arrows), and (B) the air density of the paranasal sinuses (asterisks) and mastoid air cells (arrows) were well preserved on the bone window images of the synthetic CT. (C) Perifocal hyperdensities (arrows) on the soft tissue window images of the synthetic CT might lead to a false impression of intracranial hemorrhage, which should have been postoperative encephalomalacia and white matter edema.

4. Discussion

4.1. Acquiring Paired Data

Paired data are essential for training models in 3D pix2pix, and reports have shown the feasibility of 3D pix2pix in synthetic images from multiparametric MRI. One novelty of this study is synthetic CT from contrast-enhanced MRI is important for radiotherapy treatment planning. However, the amount of treatment planning data cannot be comparable to diagnostic CT or MRI. Therefore, we developed a preprocessing pipeline to overcome different imaging modalities with different voxel sizes (resolutions), FOVs, window widths, and levels of intensity. Furthermore, data augmentation was established to expand the utility of sparse clinical imaging data. Our process involved little human involvement and is believed to be scalable to a larger dataset.

4.2. Effect of Data Augmentation

Although several studies [18–20] have indicated that data augmentation is a useful solution for limited data, none of them focused on the 3D image translation field. To ensure the effect of data augmentation for our models, we performed two simple experiments. A model was trained by a dataset without augmentation in Experiment 1. In Experiment 2, a model was trained by an augmented dataset wherein the voxel coordinates of all images were rotated by a different angle once at the beginning of each epoch. All the other conditions are the same for Experiments 1 and 2. Figure 8 shows the MSSIM trends of the two experiments.

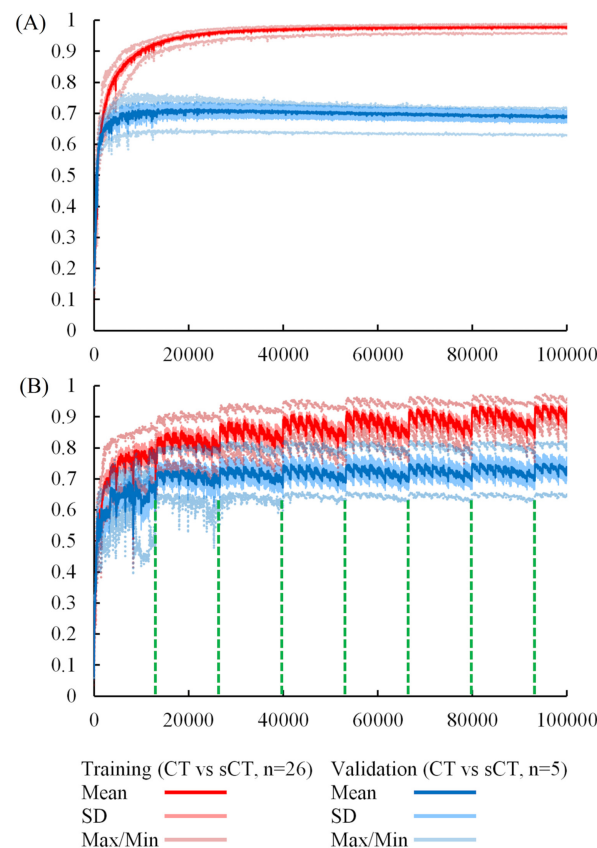


Figure 8. The MSSIM of training process. (A,B) show the results of Experiment 1 and Experiment 2, respectively. The performance of the p3_k4_f80 model is validated with original training data ($n = 26$) and testing data ($n = 5$) after 26 iterations. One epoch equivalent to 26 iterations and the period of data augmentation is 13,312 iterations (26×512). The data augmentation cycles end at iterations 13,312; 26,624; 39,936; 53,248; 66,560; 79,872; and 93,184. “SD” is the abbreviation for “standard deviation”.

The trend of Figure 8 implies that data augmentation indeed prevents a model from overfitting. It is obvious that the model overfits seriously at the first beginning in Experiment 1, and the growth of the mean MSSIM score of the testing data stops at approximately 0.71. Moreover, we inspected every synthetic CT from Experiment 1, and it is obvious that the image was blurry and unclear. On the other hand, the overfitting phenomena of the other experiments improved quite well. The best mean MSSIM scores are breakthroughs over 0.75. Furthermore, according to the slope of the curve, the index is still growing after 100,000 iterations. Therefore, the data augmentation technique we proposed does increase feature variety, which is helpful for model generalization.

4.3. The Approach Is Extendable for Higher-Resolution Datasets

In this paper, we obtain five patched datasets from the original dataset wherein each image has a very high number of voxels. It is impossible to create a model to manage our original dataset directly, and the number of trainable parameters is too numerous to be affordable for hardware. Compared to the original dataset, the benefit of a patched dataset is that it has fewer features and fewer voxels. Even if we take advantage of patched datasets to shrink the model, the number of trainable parameters of the model is approximately 300~400 million in this work for a single model. We think MR of the same body part among different patients has similar features, as does CT. Therefore, it is rational to build and train a specific model for a specific patched dataset that contains the MRI and CT of the same body part from all subjects. The solution can be adopted on any other paired dataset that has higher resolutions as long as the dataset is patched into smaller parts.

4.4. Demand for Pair Data

The 3D pix2pix model can learn the existing relationship between two imaging modalities from a training dataset as long as the relationship is certain. However, there are several uncertain relationships within our paired dataset. The material itself is the main cause of uncertain relationships. Some of the subjects had metal dentures, but some did not. Both metal dentures and normal teeth form a bright area on CT images, whereas they can hardly be seen on MR images. The shape of soft tissue and posture of a patient are also common causes of uncertain relationships. A patient might face a different angle when taking a brain MRI and CT exam or have a different bladder shape when taking a pelvis MRI and CT exam. The resolution of the imaging modality itself is another cause of uncertainty, and a medical image might be affected by the partial volume effect. In addition, MR and CT imaging have different resolutions in our original dataset, and the information amounts (voxel numbers) of the two modalities in space are not equal. Even if the dataset is resampled, some information might be lost during the process or cannot reflect reality. Some of the other uncertain relationships are caused by phantoms, noise, distortion, and inaccurate image registration.

4.5. Effect of Multiple Kernel Sizes

In Figure 6, most indices based on models with multiple kernel sizes are higher than the indices based on models with a single kernel size. We infer that it is helpful to improve prediction by merging output from different models with multiple kernel sizes. The reason why some index values are not the highest on the comparison is affected by the uncertain relationship of testing paired data. Because of the uncertainty mentioned above, some synthetic CT scans are more correct than the corresponding real CT scans (i.e., target or ground truth) on the shape or the position. Therefore, the value of similarity indices not only implies the correctness of prediction but also implicitly presents uncertainty in paired data.

4.6. Performance Variations between Kernels

The manner of data preprocessing affects the performance variances most deeply. For our early stage attempt, the composite image possesses mosaic style. Some predicted

areas are bright and others are dark. The boundary between different predicted patch areas is quite sharp. The reason for performance variation is that we split the dataset into five patched datasets before performing normalization and standardization. The normalization and standardization are based on every patched dataset itself. Therefore, the five patched datasets have no common baseline. We made several tries in order to eliminate the performance variations, such as controlling the input training data originating from the same people for the five patched models that have the same iterations. Compared to the previous version of the process, the final version improves the issue considerably. The performance variations of the proposed version are tiny. The other important factor is the native dataset itself. We found that the p3 model is the most difficult one to train and its performance is slightly poorer than all the other models. The rational explanation is that the image FOV in patched dataset 3 is allocated at the center of the brain. Thus, the patched dataset 3 contains a larger number of complicated features.

4.7. Limitation

The algorithm belongs to supervised learning, and a paired dataset is needed. The p1~p5 models are trained independently such that the phenomenon of grid-like boundary or nonuniform luminance appears in the final compound image. The drawback of expandable methodology is that cost of storage capacity and the quantities of calculation for training and prediction are proportional to the number of patches. The data form in the dataset makes the size of the dataset is difficult to enlarge. In the future, we hope the CT datasets generated from diagnostic MRIs can be used directly for radiation planning. The electron densities are the basis of the work of the irradiation planning programs. Therefore, the comparison of calculated synthetic electron densities and electron densities of the real CT scans will be addressed in the future.

5. Conclusions

In this paper, we propose an extensible solution for 3D image translation with a high-resolution dataset and demonstrate the effectiveness of data augmentation under the circumstance of insufficient training data. With the 3D image patching technique, the model size is no longer a major obstacle for the hardware. A set of well-trained models are capable of converting MRI to CT, which provides helpful information for clinical examination and reduces the radiation damage of CT imaging. For synthetic CT, radiologists reported excellent satisfaction in spatial geometry and noise level, good satisfaction in contrast and artifacts, and fair imaging details. There was no statistically significant difference in synthetic CT on the axial, coronal, or sagittal planes. The present project presents a pivotal role in connecting the MRI and CT datasets, and the application could be expandable from the cranium to other body parts. We will focus on eliminating the clinical testing difference between synthetic CT and real CT in the future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm12030361/s1>, Table S1: The detail of original dataset. The table lists the voxel numbers and voxel size for each pairs in the original dataset; Table S2: The modified image parameters after reFOV and resampling for few more subjects from training and test data; Table S3: The additional similarity indices for CT vs synthetic CT before and after merge. The other two similarity indices, MSE and PSNR, are calculated for evaluation of model performance; Table S4: Clinical satisfaction score based on the bone window synthetic CT.

Author Contributions: J.-C.W. planned and conducted the study. C.-C.W. conducted the study and helped with data collection. P.-H.W. conducted the study and mainly analyzed data. G.L. helped with the plan of study and drafted the manuscript. Y.-L.H. helped with data collection and writing the instructions for the manuscript. Y.-C.L. provided experimental suggestions. Y.-P.C. helped with revising and formatting the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the research programs MOST110-2221-E-182-027, MOST110-2628-B-182A-018, NMRPD1L0311, and CMRPD1H0421, which were sponsored by the Ministry of Science and Technology, Taipei, Taiwan, and Chang Gung University and Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan. The funders had no role in the conduct of the study or the collection, analysis, and interpretation of data.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan (protocol code 202002387B0 of approval).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Due to the ethical approval and requirements of the data protection legislation, the data set will only be made available on a restricted basis according to the data sharing policies at the Chang Gung Memorial Hospital at Linkou, Taoyuan, Taiwan. Applications for access to anonymized data can be obtained by sending an e-mail to jcweng@mail.cgu.edu.tw.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Johnstone, E.; Wyatt, J.; Henry, A.M.; Short, S.C.; Sebag-Montefiore, D.; Murray, L.; Kelly, C.G.; McCallum, H.M.; Speight, R. Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging-Only Radiation Therapy. *Int. J. Radiat. Oncol.* **2018**, *100*, 199–217. [[CrossRef](#)]
2. Mehranian, A.; Arabi, H.; Zaidi, H. Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities. *Med. Phys.* **2016**, *43*, 1130–1155. [[CrossRef](#)] [[PubMed](#)]
3. Edmund, J.M.; Andreasen, D.; Mahmood, F.; Van Leemput, K. Cone beam computed tomography guided treatment delivery and planning verification for magnetic resonance imaging only radiotherapy of the brain. *Acta Oncol.* **2015**, *54*, 1496–1500. [[CrossRef](#)]
4. Korhonen, J.; Kapanen, M.; Tenhunen, M.; Keyriläinen, J.; Seppälä, T. A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. *Med. Phys.* **2013**, *41*, 011704. [[CrossRef](#)] [[PubMed](#)]
5. Ranta, I.; Teuvo, J.; Linden, J.; Klén, R.; Teräs, M.; Kapanen, M.; Keyriläinen, J. Assessment of MRI-Based Attenuation Correction for MRI-Only Radiotherapy Treatment Planning of the Brain. *Diagnostics* **2020**, *10*, 299. [[CrossRef](#)] [[PubMed](#)]
6. Ladefoged, C.N.; Law, I.; Anazodo, U.; Lawrence, K.S.; Izquierdo-Garcia, D.; Catana, C.; Burgos, N.; Cardoso, M.J.; Ourselin, S.; Hutton, B.; et al. A multi-centre evaluation of eleven clinically feasible brain PET/MRI attenuation correction techniques using a large cohort of patients. *NeuroImage* **2016**, *147*, 346–359. [[CrossRef](#)]
7. Beyer, T.; Lassen, M.L.; Boellaard, R.; Delso, G.; Yaqub, M.; Sattler, B.; Quick, H.H. Investigating the state-of-the-art in whole-body MR-based attenuation correction: An intra-individual, inter-system, inventory study on three clinical PET/MR systems. *Magn. Reson. Mater. Phys. Biol. Med.* **2016**, *29*, 75–87. [[CrossRef](#)]
8. Hofmann, M.; Pichler, B.; Schölkopf, B.; Beyer, T. Towards quantitative PET/MRI: A review of MR-based attenuation correction techniques. *Eur. J. Nucl. Med. Mol. Imaging* **2008**, *36*, 93–104. [[CrossRef](#)]
9. Wiesinger, F.; Bylund, M.; Yang, J.; Kaushik, S.; Shanbhag, D.; Ahn, S.; Jonsson, J.H.; Lundman, J.A.; Hope, T.; Nyholm, T.; et al. Zero TE-based pseudo-CT image conversion in the head and its application in PET/MR attenuation correction and MR-guided radiation therapy planning. *Magn. Reson. Med.* **2018**, *80*, 1440–1451. [[CrossRef](#)]
10. Kazemifar, S.; McGuire, S.; Timmerman, R.; Wardak, Z.; Nguyen, D.; Park, Y.; Jiang, S.; Owrangi, A. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. *Radiother. Oncol.* **2019**, *136*, 56–63. [[CrossRef](#)]
11. Ulin, K.; Urie, M.M.; Cherlow, J.M. Results of a Multi-Institutional Benchmark Test for Cranial CT/MR Image Registration. *Int. J. Radiat. Oncol.* **2010**, *77*, 1584–1589. [[CrossRef](#)] [[PubMed](#)]
12. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
13. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
14. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.
15. Liu, Y.; Lei, Y.; Wang, Y.; Shafai-Erfani, G.; Wang, T.; Tian, S.; Patel, P.; Jani, A.B.; McDonald, M.; Curran, W.J.; et al. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. *Phys. Med. Biol.* **2019**, *64*, 205022. [[CrossRef](#)]
16. Liu, Y.; Lei, Y.; Wang, T.; Kayode, O.; Tian, S.; Liu, T.; Patel, P.; Curran, W.J.; Ren, L.; Yang, X. MRI-based treatment planning for liver stereotactic body radiotherapy: Validation of a deep learning-based synthetic CT generation method. *Br. J. Radiol.* **2019**, *92*, 20190067. [[CrossRef](#)] [[PubMed](#)]

17. Lei, Y.; Harms, J.; Wang, T.; Liu, Y.; Shu, H.; Jani, A.B.; Curran, W.J.; Mao, H.; Liu, T.; Yang, X. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med. Phys.* **2019**, *46*, 3565–3581. [[CrossRef](#)] [[PubMed](#)]
18. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the International Interdisciplinary PhD Workshop (IIPhDW), Swinoujscie, Poland, 9–12 May 2018; pp. 117–122. [[CrossRef](#)]
19. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
20. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.
21. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
22. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large Kernel Matters—Improve Semantic Segmentation by Global Convolutional Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1743–1751. [[CrossRef](#)]
23. Agrawal, A.; Mittal, N. Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy. *Vis. Comput.* **2019**, *36*, 405–412. [[CrossRef](#)]
24. Tan, M.; Le, Q.V. Mixconv: Mixed depthwise convolutional kernels. *arXiv* **2019**, arXiv:1907.09595.
25. Qian, G.; Sural, S.; Gu, Y.; Pramanik, S. Similarity between Euclidean and cosine angle distance for nearest neighbor queries. In Proceedings of the 2004 ACM Symposium on Applied Computing, Nicosia, Cyprus, 14–17 March 2004; pp. 1232–1237. [[CrossRef](#)]
26. Jhansi, Y.; Sreenivasa Reddy, E. Sketch Based Image Retrieval with Cosine Similarity. *Int. J. Adv. Res. Comput. Sci.* **2017**, *8*, 691–695.
27. Wang, L.; Zhang, Y.; Feng, J. On the Euclidean distance of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1334–1339. [[CrossRef](#)]
28. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]