STAN-CT: Standardizing CT Image using Generative Adversarial Networks

Md Selim[1,3], Jie Zhang, PhD[2], Baowei Fei, PhD[5,6], Guo-Qiang Zhang, PhD[7] , Jin Chen, PhD[1,3,4]
[1]Department of Computer Science, University of Kentucky, Lexington, KY
[2]Department of Radiology, University of Kentucky, Lexington, KY
[3]Institute for Biomedical Informatics, University of Kentucky, Lexington, KY
[4]Department of Internal Medicine, University of Kentucky, Lexington, KY
[5]Department of Bioengineering, University of Texas at Dallas, Richardson, TX
[6]Department of Radiology, UT Southwestern Medical Center, Dallas, TX
[7]The University of Texas Health Science Center at Houston, Houston, TX

Abstract *Computed Tomography (CT) plays an important role in lung malignancy diagnostics, therapy assessment, and facilitating precision medicine delivery. However, the use of personalized imaging protocols poses a challenge in large-scale cross-center CT image radiomic studies. We present an end-to-end solution called STAN-CT for CT image standardization and normalization, which effectively reduces discrepancies in image features caused by using different imaging protocols or using different CT scanners with the same imaging protocol. STAN-CT consists of two components: 1) a Generative Adversarial Networks (GAN) model where a latent-feature-based loss function is adopted to learn the data distribution of standard images within a few rounds of generator training, and 2) an automatic DICOM reconstruction pipeline with systematic image quality control that ensures the generation of high-quality standard DICOM images. Experimental results indicate that the training efficiency and model performance of STAN-CT have been significantly improved compared to the state-of-the-art CT image standardization and normalization algorithms.*

1 Introduction

Computed Tomography (CT), one of the widely used modalities for cancer diagnostics[1,2], provides a flexible image acquisition and reconstruction protocol that allows adjusting kernel function, amount of radiation, slice thickness, etc. to meet clinical requirements[3]. The non-standard protocol setup broadens the scope of CT uses effectively, but at the same time, it creates a data discrepancy problem among the acquired images[4]. For example, the same clinical observation with two different CT acquisition protocols may result in images with significantly different radiomic features, esp. intensity and texture[5,6]. As a result, this hinders the effectiveness of inter-clinic data sharing and the performance of large-scale radiomics studies[5].

The CT data discrepancy problem could be potentially addressed by defining and using a standard image acquisition protocol. However, it is impractical to use the same image acquisition protocol in all the clinical practices, not only because there are already multiple CT scanner manufactures in the market[7], but also the limitations of using a fixed protocol for all patients under all situations in diagnosis, staging, therapy selection, and therapy assessment of tumor malignancies[8]. We propose to develop an image standardization and normalization tool to "translate" any CT images acquired using non-standard protocols into the standard one while preserving most of the anatomic details[4]. Mathematically, let target image x be an image acquired using a standard protocol, given any non-standard source image x', the image standardization and normalization tool aims to compose a synthetic image $\hat{x}$ from x' such that $\hat{x}$ is significantly more similar to x than to x' regarding radiomic features.

In recent years, deep-learning-based algorithms have been developed for image or data synthesis[9,10]. U-Net is a special kind of fully connected U-shaped neural network for image synthesis[11]. Built upon U-Net or a similar neural network structure, Generative Adversarial Network (GAN) is a class of deep learning models, in which two neural networks contest with each other[9]. Being one of the mostly-used deep learning architectures for image synthesis, GAN has been utilized on CT image standardization[10]. In GANai, a customized GAN model is trained using an alternative training strategy to effectively learn the data distribution, thus achieving significantly better performance than the classic GAN model and the traditional image processing algorithm called Histogram matching[10,12,13]. However, GANai focuses on the relatively easier image patch synthesis rather than the whole DICOM image synthesis problem[10].

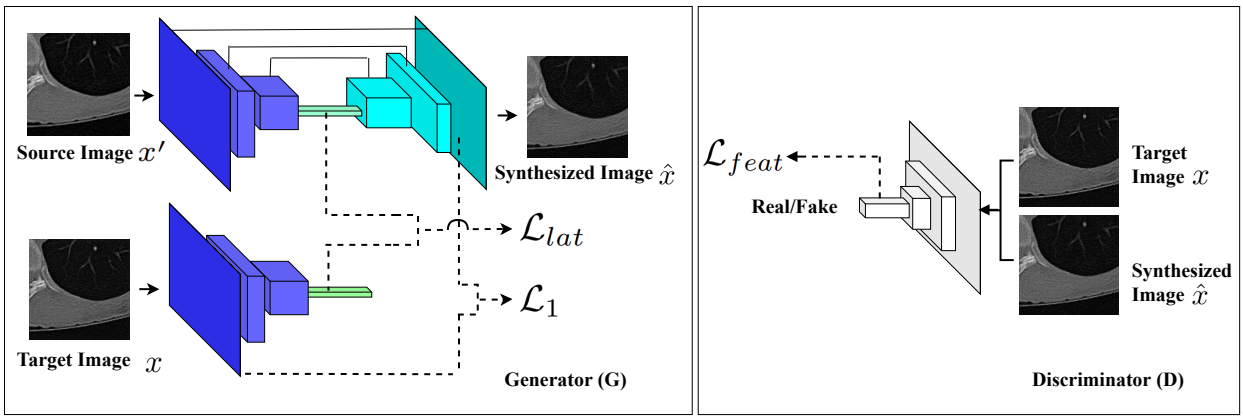


Figure 1: GAN architecture of STAN-CT. The generator G is a U-Net with a new latent loss for synthesizing image patches. The discriminator D is an Fully Convolutional Network classifier for determining whether a synthesized image patch is fake or real.

To address the CT image standardization and normalization problem, tools are needed to reconstruct synthesized data that have the common feature space as the target data[10]. This poses two fundamental computational challenges: 1) to effectively map between target images and synthesized images with great pixel-level details, and 2) to maintain the texture consistency among the synthesized images. In this paper, we present an end-to-end solution called STAN-CT. In STAN-CT, we introduce two new constrains in GAN loss shown in Fig. 1. Specifically, we adopt a latent-space-based loss for the generator to establish a one-to-one mapping from target images to synthesized images. Also, a feature-based loss is adopted for the discriminator to critic the texture features of the standard and the synthesized images. Furthermore, to synthesize CT images in the Digital Imaging and Communications in Medicine (DICOM) format[14], STAN-CT introduces a DICOM reconstruction framework that can integrate all the synthesized image patches to generate a DICOM file for clinical use. The framework ensures the quality of the synthesized DICOM by systematically identifying and pruning low-quality image patches. In our experiment, by comparing the synthesized images with the ground truth, we demonstrate that STAN-CT significantly outperforms the current state-of-the-art models. In summary, STAN-CT has the following advantages:

1. STAN-CT provides an end-to-end solution for CT image standardization and normalization. The outcome of STAN-CT is DICOM image files that are ready to be loaded into clinical systems directly.
2. STAN-CT adopts a novel one-to-one mapping loss function on the latent space. It enforces the generator to draw samples from the same distribution where the standard image belongs to.
3. STAN-CT uses a novel feature-based loss to improve the performance of the discriminator.
4. STAN-CT is effective in model training. It can quickly converge within a few rounds of training processes.

2 Background

CT images are one of the key modalities in tumor malignancy studies[15]. The CT image discrepancy problem due to the common use of non-standard imaging protocols poses a gap between CT imaging and radiomics studies. To fill the gap, clinical image synthesis tools need to be developed to *translate* images acquired using non-standard protocols into standard ones.

Image or data synthesis

Image or data synthesis is an active research area in computer vision, computer graphics, and natural language processing[9]. By definition, image synthesis is a process of generating synthetic images using limited information[16]. The given information includes text description, random noise, or any other types of information. With the recent break-

through in deep learning, image synthesis algorithms have been applied in the areas of text-to-image generation[17], detecting lost frame in a video[18], image-to-image transformation[19], and medical imaging[20] successfully.

U-net

U-Net is a special fully connected neural network originally proposed for medical image segmentation[11]. Precise localization and relatively small training data requirements are the major advantages of using U-Net[11]. A U-Net usually has three parts, down-sampling, bottleneck, and up-sampling, where the up-sampling and down-sampling are symmetric. There are also connections from down-sampling layers to the corresponding up-sampling layers to recover lost information during down-sampling. However, while U-net is effective on generating structural information, it suffers from learning and keeping texture details[21]. This issue can be overcome by adopting U-net in a more sophisticated generative model called Generative Adversarial Networks (GANs)[9].

Generative Adversarial Network

Generative Adversarial Networks (GAN), which are often used for data and image synthesis[9], normally consist of a generator G and a discriminator D. The generator that could be a U-Net is responsible for generating fake data from noise, and the discriminator tries to identify whether its input is drawn from the real or fake data. Among all the GAN models, cGAN is capable of synthesizing new images based on a prior distribution[22]. However, since the image features of the synthesized data and that of the target data may not fall into the same distribution, cGAN may not be directly applicable for the CT image standardization problem. GANai is a customized cGAN model, in which the generator and the discriminator are trained alternatively to learn the data distribution, thus achieving significantly better performance than the vanilla cGAN model. However, GANai focuses on the relatively easier image patch synthesis problem rather than the whole DICOM image synthesis problem.

Disentanglement

In a generative model, the latent space often plays a vital role in target domain mapping. Appropriate latent space learning is crucial for generating high quality data. Disentanglement is an effective metric that provides a deep understanding of a particular layer in a neural network[23]. Network disentanglement can assist to uncover the important factors that contribute to the data generation process[24].

Alternative Training Strategy

Model training is one of the most crucial parts of GAN because of the special network architecture (i.e. the generator needs to fool the discriminator while the discriminator tries to detect true data distribution from the false one). In the alternative training mechanism, when one component is in training, the other one remains unchanged. Also, each component has a fixed number of training iterations. A variant of alternative training was proposed in Liang et al.[10] named *fully-optimized alternative training*, where the model training is divided into two phases called G-phase and D-phase. In the G-phase, D is fixed, and G needs to achieve a certain accuracy θ_G before reaching the maximum training step t_{max}. In the D-phase, G is fixed, and D needs to achieve a pre-defined performance θ_D or it stops when reaching a maximum training step t_{max}. When one training phase is completed, the other phase will begin. The GANai training will continue until an optimal result is achieved or the maximum number of epochs is reached. Furthermore, instead of performance competing between a single copy of D and G, multiple copies of Gs and Ds compete with each other. For example, a G needs to fool multiple Ds before its phase is over. Also, a rollback mechanism is implemented in GANai so that if a component is not able to fool its counterpart within limited steps, it rolls back to the beginning status of the current phase and starts again. This alternative training mechanism has been successfully applied to address the CT image standardization problem.

3 Method

With STAN-CT, we attempt to address the long-standing CT image standardization problem. STAN-CT consists of a novel GAN model and a dedicated DICOM synthesis framework to meet the clinical requirements.

Standardizing CT image patches

Similar to the conventional GAN models, the STAN-CT GAN model has two components, the generator G and the discriminator D. G is a **U**-shaped network[11] consisting of an encoder and a decoder. Both the encoder and the decoder consist of seven hidden layers. There is a skip connection from each layer of the encoder to the corresponding layer of the decoder to address the information loss problem during the down-sampling. D consists of five fully connected convolutional layers. Fig. 1 illustrates the GAN architecture of STAN-CT. The detailed architecture and the hyperparameters of STAN-CT used in our experiments are specified in section 4 sub-section titled *STAN-CT architecture and hyperparameters*. Mathematically, let x be a standard image and x' be its corresponding non-standard image. The aim of the generator is to create a new image $\hat{x}$ that has the same data distribution as x. Meanwhile, the discriminator determines whether x and $\hat{x}$ are from the standard image distribution.

Discriminator Loss. In a GAN model, the performance of D and G increases accordingly. We propose to adopt two losses for the discriminator training, i.e. the WGAN[25] adversarial loss function to critic the standard and non-standard images and the fetcher-based loss. WGAN is a stable GAN training mechanism that provides a learning balance between G and D[26]. STAN-CT adopts the WGAN-based adversarial loss of the discriminator defined as:

$$\mathcal{L}_{adv(D)} = \nabla_w \frac{1}{m} \sum_{i=1}^{m} [f(x^{(i)}) - f(G(x'^{(i)}))] \tag{1}$$

where w is the hyper-parameters of D, m is the batch size, $x'^{(i)}$ is the input (non-standard image), and $x^{(i)}$ is the corresponding standard image. In addition to the WGAN-based adversarial loss, STAN-CT introduces a new feature-based loss function $\mathcal{L}_{feat}$. To improve generator diversity, a similar feature-based loss function has been used in Yang et al.[27]. Here, we use the feature space of D instead of a secondary pre-trained network to maintain a balanced network (i.e. D and G are not too strong or too weak compared with other). The feature-based loss is described in Eq 2:

$$\mathcal{L}_{feat} = \mathbb{E}_{(x)}[\frac{1}{V} \|\phi(D(G(x'))) - \phi(D(x))\|] \tag{2}$$

where ϕ is the feature extractor and V is the volume of the feature space, and $G(x')$ is an image generated by G and x is the target image. Finally, let λ_1 be a wight factor ($\lambda_1 \in [0, 1]$), the total loss of D that combines the WGAN-based loss $\mathcal{L}_{adv(D)}$ and the feature-based loss $\mathcal{L}_{feat}$ is defined as:

$$\mathbb{L}(D) = \mathcal{L}_{adv(D)} + \lambda_1 \mathcal{L}_{D_{feat}} \tag{3}$$

Generator Loss. The generator loss consists of three components, i.e. the WGAN-based loss, the latent loss, and the L1 regularization. The WGAN-based loss, which is used to improve network convergence, is defined as:

$$\mathcal{L}_{adv(G)} = \nabla_\theta \frac{1}{m} \sum_{i=1}^{m} f(G(x'^{(i)})) \tag{4}$$

where θ represents all the hyper-parameters of G, $x'^{(i)}$ is a source image, and f is 1-Lipschitz function, which returns the Earth-Mover (EM) distance from one metric space to another.

It is the latent space that connects the distribution of an input domain and an output domain in the same generative model, which allows a smooth domain translation[28]. Inspired by Mao et al.[28], we propose a new latent-feature-based loss function to enforce one-to-one mapping between the synthesized image and the standard image. Specifically, the latent loss $\mathcal{L}_{lat}$ aims to minimize the distance between the latent distribution of the synthesized images and their corresponding standard images.

$$\mathcal{L}_{lat} = \left\| z_x - z_{G(x')} \right\| \tag{5}$$

where z stands for the latent vector, $G(x')$, and x is its corresponding standard image. Finally, the total loss of G $\mathbb{L}(G)$ is defined as:

$$\mathbb{L}(G) = \mathcal{L}_{adv(G)} + \lambda_2 \mathcal{L}_{G_{lat}} + \lambda_3 \frac{1}{m} \sum_{i=1}^{m} |x - G(x')| \tag{6}$$

where $\lambda_2 \in [0, 1]$ and $\lambda_3 \in [0, 1]$ are wight factors. $\frac{1}{m} \sum_{i=1}^{m} |x - G(x')|$ is the $\mathcal{L}_1$ regularization function.

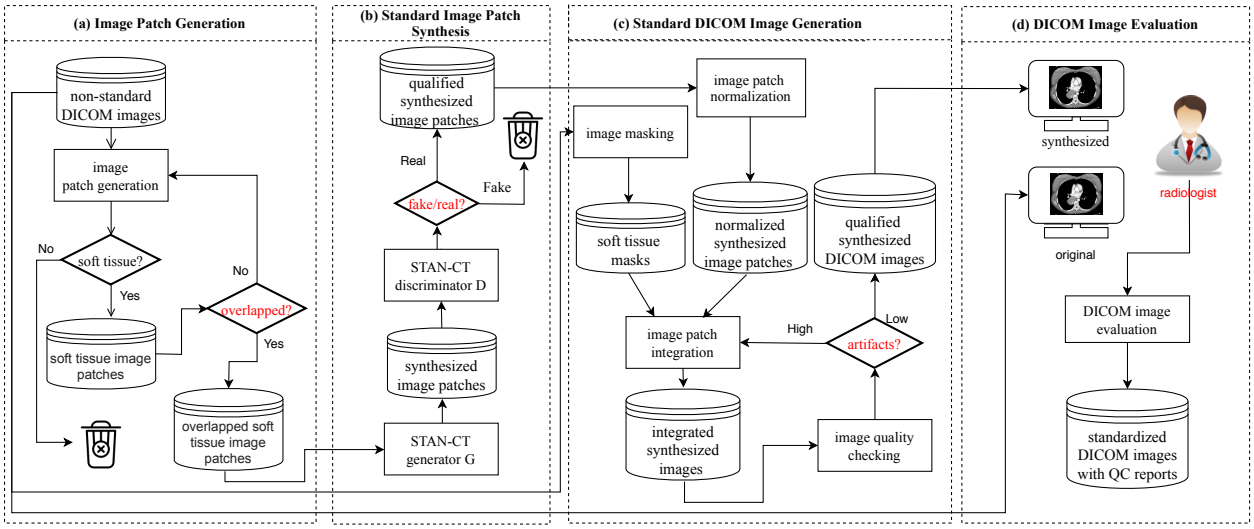


Figure 2: STAN-CT DICOM-to-DICOM image standardization framework. (a) Soft tissue image patches are generated from the input DICOM files, ready for image standardization and normalization. (b) For all the soft tissue image patches, new image patches are synthesized using STAN-CT generator. The quality of the new image patches is checked using STAN-CT discriminator. (c) All the synthesized soft-tissue image patches are integrated and are filtered by a soft tissue image mask generated using the input DICOM image. DICOM image quality is checked by examining box artifacts and empty pixels. (d) The synthesized and the original non-standard DICOM image files will be viewed side-by-side by radiologists using a PACS reading workstation. The radiologists' reports will be used to further evaluate the quality of the standardized CT images. Meanwhile, image texture features will be extracted for automatic performance evaluation.

DICOM Reconstruction Framework

STAN-CT presents a DICOM-to-DICOM reconstruction framework for systematic DICOM reconstruction. The DICOM-to-DICOM reconstruction framework includes four additional components to facilitate processes such as image patch generation and fusion (see Fig. 2). Each component has a unique quality control unit (red diamond box) that ensures the outputs are free from defects.

Step 1. soft tissue image patch generation: The first step of STAN-CT DICOM-to-DICOM image standardization is soft tissue image patch generation. Image patches with size between 100 and 256 are randomly generated using the input DICOM image. An image patch is a soft tissue image patch if at least 70% of the pixels are in the soft tissue range (Hounsfield unit value ranging from -1000 to 900). The process will continue until each soft-tissue image patch contains at least 50% overlapped pixels.

Step 2. standard image patch synthesis: With a trained STAN-CT generator, a soft-tissue image patch obtained in the previous step will be standardized (see Section 3 for details).

Then, the synthesized image patches will be examined by STAN-CT discriminator. If a synthesized image patch can fool the discriminator, it is considered as a *qualified synthesized image patch*. Otherwise, the synthesized image patch will be discarded. This step ensures the quality of the synthesized image patches.

Step 3. standard DICOM image generation: Given all the qualified synthesized image patches, we first normalize the pixel intensity from gray-scale to the Hounsfield unit using:

$$P_{HU} = \frac{P_g - min(\hat{x}_g)}{max(\hat{x}_g) - min(\hat{x}_g)} (MAX - MIN) \tag{7}$$

where P_{HU} and P_g is the pixel value in Hounsfield unit and gray-scale unit respectively, $\hat{x}_g$ is a qualified synthesized image patch, and MAX and MIN are the maximum and minimum CT number of a source DICOM.

Meanwhile, with a soft tissue image mask created from the original DICOM images with Hounsfield unit ranging from

-1000 to 900, the non-soft tissue parts of the synthesized and normalized image patches will be discarded. Finally, we integrate all the valid soft tissue patches to generate the integrated synthesized images.

The quality of the integrated synthesized images will be checked using a quality control unit, which inspects whether there is any box artifacts or missing values. If some artifacts are identified, the corresponding image patches will be re-integrated by cropping boundary pixels.

Step 4. DICOM image evaluation: In the DICOM image evaluate step, both the synthesized and the original non-standard DICOM image files will be viewed side-by-side by radiologists using a PACS reading workstation. Radiologists will be asked to evaluate image quality, estimate the acquisition protocol, and extract tumor properties. The radiologists' reports will be used to evaluate the quality of the standardized CT images. Meanwhile, with all the synthesized DICOM files generated in the previous step, image texture features will be automatically extracted and compared for automatic performance evaluation.

4 Experimental result

Data

For the training data, we used total of 14,688 CT image slices captured using three different kernels (BL57, BL64, and BR40) and four different slice thicknesses (0.5, 1, 1.5, 3mm) using Siemens CT Somatom Force scanner at the University of Kentucky Medical Center. STAN-CT adopted BL64 kernel and 1mm slice thickness as the standard protocol since it has been widely used in clinical practice. Random cropping was used for the image patch extraction and resized into 256×256 pixel patches. Data augmentation was done by rotating and shifting image patches. Finally, a total of 49,000 soft-tissue image patches were generated from the CT slices and were used as the training data of STAN-CT. Two testing data sets were prepared for STAN-CT performance evaluation. Both data sets were captured using Siemens CT Somatom Force scanner at the University of Kentucky Medical Center hospital. The first testing data were captured using the non-standard protocol BR40 and 1mm slice thickness. The second testing data were captured using the non-standard protocol BL57 and 1mm slice thickness. The image patch generation step was the same as that of the training data. Each test data set contains 3,810 image patches.

STAN-CT architecture and hyperparameters

STAN-CT GAN model consists of a U-Net with fifteen hidden layers and an FCN with five hidden layers. The 4×4 kernel is used in the convolutional layer. LeakyRelu[29] is adopted as the activation function in all the hidden layers. Softmax is used in the last layer of FCN. Random weight is used during the network initialization phase. The prediction thresholds for determining fake or real images is 0.01 and 0.99 respectively. Maximum training epochs were set to 100 with a learning rate of 0.0001 with momentum 0.5. A fully optimized alternative training mechanism (the same as GANai) was used for the network training. STAN-CT was implemented in TensorFlow[30] on a Linux computer server with eight Nvidia GTX 1080 GPU cards. The model took about 36 hours to train from scratch. Once the model was trained, it took about 0.1 seconds to synthesize and normalize every image patch.

Evaluation Metric

For performance evaluation, we computed five radiomic texture features (i.e. dissimilarity, contrast, homogeneity, energy, and correlation) using Gray Level Co-occurrence Matrix (GLCM). The absolute error $\mathbb{E}$ of each radiomic texture feature was computed using:

$$\mathbb{E}(I_{syn}, I_{target}, f) = \frac{|\varphi(I_{syn}, f) - \varphi(I_{target}, f)|}{\varphi(I_{target}, f)} \tag{8}$$

where φ is the GLCM feature extractor I_{syn} and I_{target} is the synthesized image from STAN-CT and the target image respectively. f is the corresponding feature space.

Kernel	Features	GANai	STAN-CT	STAN-CT w/o $\mathcal{L}_{Lat}$	STAN-CT w/o $\mathcal{L}_{feat}$
BL57	dissimilarity	0.313	**0.228**	0.234	0.245
	contrast	0.313	**0.228**	0.234	0.245
	homogeneity	0.012	**0.009**	0.009	0.012
	energy	0.032	0.035	0.038	**0.022**
	correlation	0.085	0.058	**0.057**	0.120
BR40	dissimilarity	0.683	**0.407**	0.441	0.545
	contrast	0.683	**0.407**	0.441	0.545
	homogeneity	0.028	**0.018**	0.019	0.024
	energy	**0.040**	0.041	0.057	0.045
	correlation	0.315	**0.203**	0.273	0.303

Table 1: Texture feature comparison between GANai, STAN-CT and its two variants. Five texture features (dissimilarity, contrast, homogeneity, energy and correlation) were extracted from DICOM image patches. The absolute error was reported for each feature.

Kernel	Features	straight-forward	w/ overlapped check	w/ real/fake check	STAN-CT
BL57	dissimilarity	0.727	0.485	0.334	**0.201**
	contrast	0.727	0.485	0.334	**0.201**
	homogeneity	0.031	0.019	0.012	**0.009**
	energy	0.072	0.063	0.046	**0.032**
	correlation	0.319	0.149	0.075	**0.054**
BR40	dissimilarity	0.849	0.653	0.496	**0.405**
	contrast	0.849	0.653	0.496	**0.405**
	homogeneity	0.035	0.027	0.022	**0.016**
	energy	0.048	0.045	0.051	**0.040**
	correlation	0.386	0.345	0.289	**0.201**

Table 2: Texture feature comparison. Five texture features were extracted from DICOM images constructed from the same image patches using four different DICOM reconstruction methods. The averaged absolute error is reported for each feature.

Table 3: Performance on tumor-specific tissues. Five texture features, i.e. dissimilarity, contrast, homogeneity, energy, and correlation, were extracted from all the image patches of a tumor. The mean absolute error of each feature was reported.

Kernel	Model	dissimilarity	contrast	homogeneity	energy	correlation
BL57	GANai	0.124 ± 0.098	0.124 ± 0.098	0.018 ± 0.012	0.098 ± 0.054	0.877 ± 0.368
	STAN-CT	0.114 ± 0.026	0.114 ± 0.026	0.017 ± 0.005	0.041 ± 0.019	0.847 ± 0.408
BR40	GANai	0.616 ± 0.104	0.616 ± 0.104	0.091 ± 0.018	0.222 ± 0.061	1.184 ± 0.646
	STAN-CT	0.602 ± 0.081	0.602 ± 0.081	0.076 ± 0.011	0.182 ± 0.061	1.184 ± 0.646

Performance of image patch synthesis

Table 1 shows the absolute error of five GLCM-based texture features of STAN-CT, GANai (the current state-of-the-art model), and two disentangled representation of STAN-CT on the soft tissues. In the model named "STAN-CT w/o $\mathcal{L}_{lat}$", we discarded from STAN-CT the latent loss function $\mathcal{L}_{lat}$ of G. In the second one named "STAN-CT w/o $\mathcal{L}_{feat}$", we discarded the feature-based loss $\mathcal{L}_{feat}$ of D from STAN-CT. All the models were tested using kernel BL57 and BR40 with the same slice thickness ($1mm$). For kernel BL57, STAN-CT and its variants outperformed GANai in all the texture features. For kernel BR40, STAN-CT was significantly better than GANai in four out of five features. Also, Table 3 shows the feature comparison on the same tumor tissue. On all the five GLCM-based texture features extracted from the images scanned using BL57 and BR40 kernels, STAN-CT clearly outperforms GANai.

The first four generators of each GAN models were selected for further analysis. Fig. 3 illustrates the change of the absolute errors of the five GLCM-based texture features using the generators produced in the first four iterations of alternative training of STAN-CT or GANai. The result indicates that STAN-CT can quickly reduce the errors in the first a few iteration of the alternative training, while no clear trend was observed in the results of GANai.

Performance of DICOM reconstruction

A straightforward patch-based image reconstruction approach has three steps: 1) splitting a DICOM slice into overlapped or non-overlapped image patches; 2) standardizing each image patch; and 3) merging the standardized image patches into one DICOM slice. A common problem in such a patch-based image reconstruction process is image artifacts, such as boundary artifact or inconsistent texture. As shown in Table 2, the straightforward approach has the highest absolute error on all the tested image features.

In STAN-CT, three quality control units were inserted into the framework, each being adopted to address a specific image quality problem. Table 2 shows that STAN-CT achieved significantly better performance than the straightforward method regarding the absolute errors on five selected texture features. Fig. 4 visualized the reconstructed DICOM

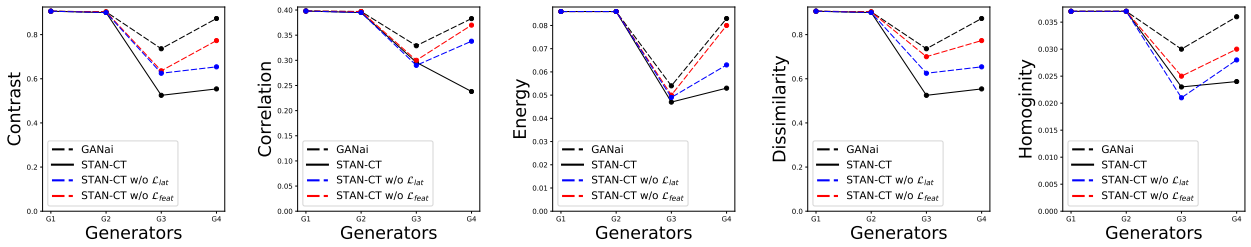
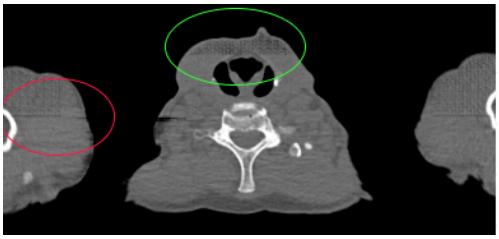
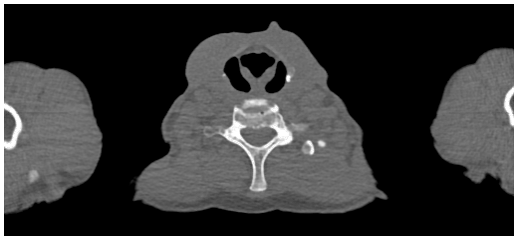


Figure 3: Performance evaluation of STAN-CT generator. The first four generators of STAN-CT and GANai were compared using the GLCM-based features. The x-axis denotes the training phase, and the y-axis denotes the absolute error of each selected texture feature. The result indicates STAN-CT archived overall the best performance.



(a) DICOM reconstructed using a straightforward method. The red circle highlights the boundary effect where two image patches were merged. The green circle shows texture inconsistency.



(b) The same DICOM reconstructed using STAN-CT. No visual artifacts were found according to the radiologist's report.

Figure 4: DICOM Reconstruction Comparison

images using the two methods. The red (green) circle highlights the boundary effect where two image patches were merged (texture inconsistency within a DICOM slice) using the straightforward method. In the same DICOM reconstructed using STAN-CT, no visual artifacts were found according to the radiologist's report.

We further compared STAN-CT with its two variants. The method named "w/ overlapped check" used only the first quality control unit to check whether there were enough overlapped soft-tissue image patches. The method named "w/ real/fake check" used the first two quality control units, which not only checked if there were enough image patches, but also examined whether the image patches were successfully standardized. Table 2 shows that both approaches achieve better results than the straightforward method, but none of is better than STAN-CT, indicating all the three quality control units are critical regarding artifact detection and removal. The standardized DICOM images, along with the corresponding standard images, were reviewed by radiologists at the Department of radiology, University of Kentucky using the picture archiving and communication system (PACS) viewer (Barco, GA, USA). The radiologists, who were blinded to the image reconstruction algorithms, reported that no obvious difference was observed in lung regions between the two kinds of images.

5 Discussion

First, by systematically removing every single component in the GAN model and in the DICOM reconstruction pipeline using the the leave-one-out approach, we analyzed the impact of every component of STAN-CT. In STAN-CT, both the latent loss $\mathcal{L}_{Lat}$ and the feature loss $\mathcal{L}_{feat}$ are key components. To evaluate the impact of the loss functions, two versions of STAN-CT GAN, where the latent loss or the feature loss has been removed respectively, were created. Table 1 shows that none of them can achieve the same performance as that of STAN-CT regarding the GLCM-based texture features. Also, Fig. 5 shows that the latent loss of STAN-CT $\mathcal{L}_{Lat}$ decreases during G-training, indicating that the generator can reduce the gap between the distributions of the target image and the synthetic image effectively, while maintaining flat during the D-training phases.

Second, we analyzed the STAN-CT training phase switches. During the STAN-CT training, the discriminator loss (shown in Fig. 6) bounced between 0 and 2 rapidly in the early phase switches, indicating an efficient discriminator and generator training. In the later phases, however, the training time increased significantly, indicating that both the generator and the discriminator were converging. At the blue colored point, the discriminator failed to distinguish real

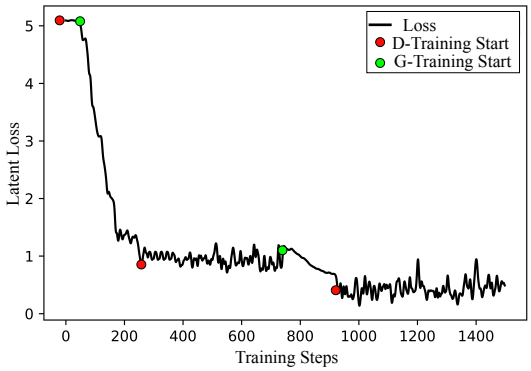


Figure 5: Latent loss during GAN training. Latent loss is decreasing effectively in the G-training phase, while it keeps stable in the D-training phase.

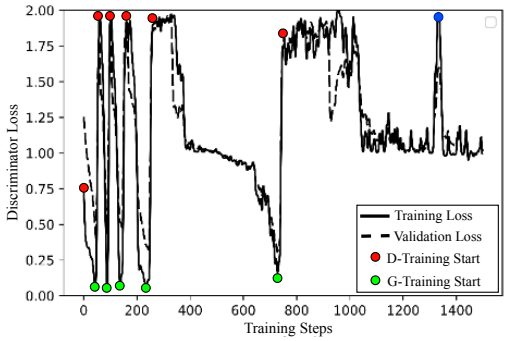


Figure 6: Performance of the discriminator of STAN-CT in different training phases. The discriminator loss decreases in the D-training phase and increases in the G-training phase. The blue-colored point indicates a failed-then-restart D-training.

from fake images after certain iterations of discriminator training. In this situation, the new discriminator was ruled out and the D-training was restarted. If the D-training continued to fail, the STAN-CT alternative training can stop [10].

Finally, the DICOM reconstruction pipeline includes four quality control units, each contributing to the improvements of the quality of the resulting DICOM images. Table 2 shows that the contrast error of the straightforward DICOM reconstruction (without using any of the quality control units) is 0.727, which can be reduced to 0.485 by adding the overlapped soft tissue quality control, which provides consistent texture throughout the DICOM. It can be further reduced to 0.334 (54% improvement) by adding the discriminator checker that ensures the success of image synthesis. Eventually, if all the four quality control units were used, the contract error was reduced to 0.201 (72% improvement).

6 Conclusion

Data discrepancy in CT images due to the use of non-standard image acquisition protocols adds extra burden to radiologists and also creates a gap in large-scale cross-center radiomic studies. We propose STAN-CT, a novel tool for CT DICOM image standardization and normalization. In STAN-CT, new loss functions are introduced for efficient GAN training, and a dedicated DICOM-to-DICOM image reconstruction framework has been developed to automate the DICOM standardization and normalization process. The experimental results show that STAN-CT is significantly better than the existing tools on CT image standardization. Our experiments demonstrate that inconsistency in CT image acquisition can be effectively harmonized using STAN-CT. This work fits well with large-scale radiomic studies in cancer researches where radiomic features can be extracted from standardized images rather than the original ones.

Acknowledgements

This research is supported by NIH NCI (grant no. 1R21CA231911) and Kentucky Lung Cancer Research (grant no. KLCR-3048113817).

References

1. Jerry L Prince and Jonathan M Links. *Medical imaging signals and systems*. Pearson Prentice Hall Upper Saddle River, 2006.
2. Mahadevappa Mahesh. Fundamentals of medical imaging. *Medical Physics*, 38(3):1735–1735, 2011.
3. Abhishek Midya, Jayasree Chakraborty, Mithat Gönen, et al. Influence of ct acquisition and reconstruction parameters on radiomic feature reproducibility. *Journal of Medical Imaging*, 5(1):011020, 2018.
4. G Liang, J Zhang, M Brooks, et al. radiomic features of lung cancer and their dependency on ct image acquisition parameters. *Medical Physics*, 44(6):3024, 2017.
5. Roberto Berenguer, María del Rosario Pastor-Juan, Jesús Canales-Vázquez, et al. Radiomics of ct features may be nonreproducible and redundant: Influence of ct acquisition parameters. *Radiology*, page 172361, 2018.
6. Luke A Hunter, Shane Krafft, Francesco Stingo, et al. High quality machine-robust image features: Identification in nonsmall

cell lung cancer computed tomography images. *Medical physics*, 40(12), 2013.
7. Jijo Paul, B Krauss, R Banckwitz, et al. Relationships of clinical protocols and reconstruction kernels with image quality and radiation dose in a 128-slice ct scanner: study with an anthropomorphic and water phantom. *European journal of radiology*, 81(5):e699–e703, 2012.
8. David S Gierada, Andrew J Bierhals, Cliff K Choong, et al. Effects of ct section thickness and reconstruction kernel on emphysema quantification: relationship to the magnitude of the ct emphysema index. *Academic radiology*, 17(2):146–156, 2010.
9. He Huang, Philip S Yu, and Changhu Wang. An introduction to image synthesis with generative adversarial nets. *arXiv preprint arXiv:1803.04469*, 2018.
10. Gongbo Liang, Sajjad Fouladvand, Jie Zhang, et al. Ganai: Standardizing ct images using generative adversarial network with alternative improvement. *bioRxiv*, page 460188, 2018.
11. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
12. Arthur Robert Weeks, Lloyd J Sartor, and Harley R Myler. Histogram specification of 24-bit color images in the color difference (cy) color space. *Journal of electronic imaging*, 8(3):290–301, 1999.
13. Anil K Jain. *Fundamentals of digital image processing*. Englewood Cliffs, NJ: Prentice Hall,, 1989.
14. Peter Mildenberger, Marco Eichelberg, and Eric Martin. Introduction to the dicom standard. *European radiology*, 12(4):920–927, 2002.
15. Stefania Rizzo, Francesca Botta, Sara Raimondi, et al. Radiomics: the facts and the challenges of image analysis. *European radiology experimental*, 2(1):36, 2018.
16. Roy A Hall and Donald P Greenberg. A testbed for realistic image synthesis. *IEEE Computer Graphics and Applications*, 3(8):10–20, 1983.
17. Scott Reed, Zeynep Akata, Xinchen Yan, et al. Generative adversarial text to image synthesis. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48, pages 1060–1069, 2016.
18. Yong-Hoon Kwon and Min-Gyu Park. Predicting future frames using retrospective cycle gan. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1811–1820, 2019.
19. Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4401–4410, 2019.
20. Zekuan Yu, Qing Xiang, and Jiahao et al. Meng. Retinal image synthesis from multiple-landmarks input with generative adversarial networks. *Biomedical engineering online*, 18(1):62, 2019.
21. Hariharan Ravishankar, Rahul Venkataramani, Sheshadri Thiruvenkadam, et al. Learning and incorporating shape models for semantic segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 203–211. Springer, 2017.
22. Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv:1411.1784v1*, 2014.
23. Quan-shi Zhang and Song-Chun Zhu. Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1):27–39, 2018.
24. Irina Higgins and David et al. Amos. Towards a definition of disentangled representations. *arXiv preprint arXiv:1812.02230*, 2018.
25. Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.
26. Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, et al. Improved training of wasserstein gans. In *Advances in neural information processing systems*, pages 5767–5777, 2017.
27. Qingsong Yang, Pingkun Yan, Yanbo Zhang, et al. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018.
28. Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, et al. Mode seeking generative adversarial networks for diverse image synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1429–1437, 2019.
29. Bing Xu, Naiyan Wang, Tianqi Chen, et al. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.
30. Martín Abadi, Ashish Agarwal, Paul Barham, et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.