CT Image Harmonization for Enhancing Radiomics Studies

Md Selim
Dept. of Computer Science
University of Kentucky
Lexington, KY
md.selim@uky.edu

Jie Zhang
Dept. of Radiology
University of Kentucky
Lexington, KY
jie.zhang1@uky.edu

Baowei Fei
Dept. of Bioengineering
University of Texas at Dallas
Dallas, TX
BFei@utdallas.edu

Guo-Qiang Zhang
Dept. of Neurology
University of Texas Health Science Center at Houston
Houston, TX
Guo-Qiang.Zhang@uth.tmc.edu

Jin Chen
Inst. for Biomedical Informatics
University of Kentucky
Lexington, KY
chen.jin@uky.edu

2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 978-1-6654-0126-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/BIBM52615.2021.9669448

***Abstract*—While remarkable advances have been made in Computed Tomography (CT), most of the existing efforts focus on imaging enhancement while reducing radiation dose. How to normalize CT images acquired using non-standard protocols is vital for decision-making in cross-center large-scale radiomics studies but remains the boundary to explore. We develop a novel GAN-based image standardization algorithm called *Radiomic-GAN* to mitigate the discrepancy caused by using non-standard acquisition protocols. In RadiomicGAN, a pre-trained U-Net has been adopted as part of the generator to learn radiomic feature distributions efficiently, and a novel training approach, called *Window Training*, has been developed to smoothly transform the pre-trained model to the medical imaging domain. In the experiments, we compared RadiomicGAN with four state-of-the-art CT image standardization approaches on both patient and phantom CT images acquired using three different reconstruction kernels. We objectively evaluated model performance based on more than 1,000 radiomic features. The results show that RadiomicGAN clearly outperforms the compared models. The source code, manual, and sample data are available at https://github.com/selim-iitdu/radiomicGAN.**

***Index Terms*—Computed Tomography, Generative Adversarial Network, Image Synthesis, Standardization, Radiomics.**

I. INTRODUCTION

As one of the most popular diagnostic image modalities routinely used for assessing anatomical tissue characteristics for disease management (1), computed tomography (CT) provides the flexibility of customizing acquisition and image reconstruction protocols to meet an individual's clinical needs (2). However, capturing CT images with non-standardized protocols could result in inconsistent radiomic features in both intra-CT (by changing CT acquisition parameters) and inter-CT (by comparing different scanners with the same acquisition parameters) tests. The low reproducibility regarding radiomic features, such as intensity, shape, and texture, for CT imaging, may forms a barrier to analyzing CT images in a large scale, a.k.a. radiomics (3; 4).

The radiomic feature discrepancy problem can be addressed by either normalizing the radiomic features of all the non-standard images or standardizing CT images and then extracting the radiomic features from the standardized images. The former solution, however, is difficult, since the distributions of the radiomic features are not well defined (5). For the latter one, image synthesis algorithms have been recently developed aiming to synthesize images with similar feature-based distributions compared to that of the target images while preserving anatomic details (6). Mathematically, let x be a CT image acquired using a non-standard reconstruction kernel, y be its corresponding standard image, the model aims to compose a synthetic image y' from x, such that y' follows the feature distributions of y rather than x.

Most of the recent progresses on CT image standardization and normalization are based on deep learning models (6; 7; 8). All these models need to be trained from scratch using a relatively large data which are difficult to obtain. To relax the demand of large data, transfer learning may be adopted. One of the computational challenges is to adopt models pre-trained on natural image domain due to different dynamic ranges. In particular, the pixel intensity of a natural image ranges from 0 to 255 (8-bit), while for the state-of-the-art CT scanners, standard 12-bit depth images are commonly utilized, resulting in a much wider Hounsfield Unit (HU) range that scales from -1,024 to 3,071 (9).

In this paper, we present **RadiomicGAN**, a novel GAN-based deep learning model, for CT image standardization and normalization focused on harmonizing CT images acquired with non-standard reconstruction kernels as it is one of the most critical factors causing feature inconsistency (10). RadiomicGAN employs a hybrid architecture for image texture feature extraction and embedding. Its encoder consists of multiple consecutive neural blocks including both pre-trained and trainable convolutional layers. To address the dynamic pixel range-related problem in transfer learning, Radiomic-

GAN uses a new training strategy named *Dynamic Window-based Training* (DWT), which allows us to train a model using pixels within a selected range called "window". The range of a window can be automatically broaden or shrank based on the pixels where the model suffers most in the previous training iteration, allowing us to fine-tune the trainable layers in RadiomicGAN using the frequently appeared pixels in the window.

In summary, given its hybrid network structure, RadiomicGAN can effectively learn the radiomic feature distributions from the standard CT images and then harmonize non-standard CT images. A dynamic window-based training approach is developed to effectively address the pixel range difference problem and thus enable transfer learning in the medical image domain. Overall, RadiomicGAN has the following advantages:

1) RadiomicGAN effectively learns the radiomic feature distributions of standard CT images from limited number of patients using a novel transfer learning structure.
2) RadiomicGAN adopts a novel window training method to gradually map RGB to HU range for CT images.
3) A systematic evaluation metric for CT image standardization, which extensively measures 1,401 radiomic features, has been developed in RadiomicGAN.
4) Experimental results show RadiomicGAN is clearly better than the state-of-the-art CT image standardization methods.

II. Method

RadiomicGAN is a novel GAN-based model optimized for CT image standardization and normalization. Let x be a non-standard CT image and y be its corresponding standard CT image. Given x, the generator of RadiomicGAN G aims to synthesize a new image y' that has the same data distribution as y rather than x. Meanwhile, the discriminator of RadiomicGAN D determines whether y and y' are from the same distribution. Leveraging a pre-trained VGG-19, the generator of RadiomicGAN can effectively learn the feature distribution of the standard CT images. In addition, a new window training strategy enables the application of the VGG-19 pre-trained with natural images on medical image analysis tasks.

A. RadiomicGAN Architecture

The network architecture of RadiomicGAN shown in Figure 1 consists of a generator and a discriminator. The discriminator of RadiomicGAN D is a typical fully convolutional neural network adopted from the pix2pix network (11). The generator G is the U-Net-like structure, where each trainable hidden layer in the encoder is connected to its corresponding hidden layer in the decoder with a skip connection to preserve the lost features during down-sampling (12).

Inspired by the application of pre-trained VGG for style transfer (13; 14), we construct the encoder of RadiomicGAN as a series of consecutive neural blocks. All the layers in the same neural block have the same dimension. A neural block includes multiple pre-trained VGG layers (Figure 1, light blue), which is frozen during network training, and a trainable layer (Figure 1, dark blue) that works as filter to extract and forward the fine-to-coarse texture features and forward them to the corresponding decoding layers.

For example, the first block includes two 512-by-512 pre-trained VGG layers and a 512-by-512 trainable layer, both the first and the second frozen layer are constricted by using convolution operation with stride=1 on the previous layer. The trainable layers in a neural block, which is constructed from the previous pre-trained VGG layer using convolution, batch-normalization, and ReLU activation, is designed to propagate the domain-specific information onto the corresponding decoding layers.

B. RadiomicGAN Training

Leveraging pre-trained networks, transfer learning has been widely adopted in applications in the medical domain (15; 16). The CT image standardization and normalization problem can be addressed by fine tuning a pre-trained CNN with limited medical data. In RadiomicGAN training, by passing a non-standard CT image x through a neural block, which consists of multiple pre-trained VGG layers and a trainable convolutional layer, the domain-specific texture features of x can be effectively extracted. In the following text, we introduce the feature extraction and embedding, loss function, and window training, which are the three key components in RadiomicGAN model training.

1) Feature Extraction and Embedding.: In a CNN, the first several convolutional layers are used to extract texture-related features and the last a few layers are used to extract shape-related features (13). Since RadiomicGAN is expected to standardize the texture features while keeping the shape features unchanged, we adopt the first four groups of convolutional layers of a pre-trained VGG-19 network aiming to effectively extract fine-to-coarse features from the input CT images.

2) Window Training: Natural images usually use the 8-bit encoding, so the pixel domain covers numbers ranging from 0 to 255. Medical images, since they follow the 16-bit encoding, have a much wider dynamic range. The extended pixel range poses a fundamental challenge for the use of transfer learning in the medical image domain.

We introduce a new training strategy, called *Window Training* to gradually vary the effective pixel range and to continuously train RadiomicGAN, making it possible to map RGB to HU numbers in transfer learning. During the window training, RadiomicGAN is exposed to an effective pixel range called "window". The lower bound and the upper bound of a window can be specified using the fixed growing and dynamic selection approaches consecutively.

Using window training, RadiomicGAN can be continuously trained with the updated training data with specified effective pixel ranges. Algorithm 1 describes the window training strategy, which consists of the fixed growing approach followed by the dynamic selection approach. The former is a bottom-up training approach, where the window starts from a small range and gradually extends to the whole HU range, allowing the model to gradually learns the HU number distribution in the

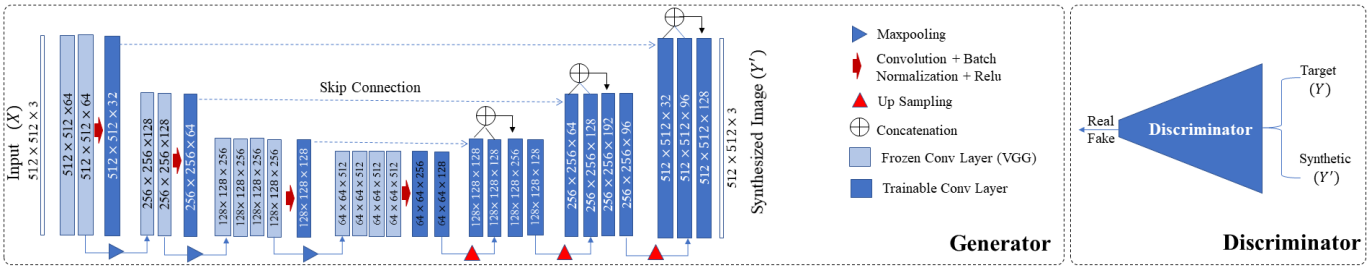

Fig. 1: **Architecture of RadiomicGAN.** The generator G is a combination of a U-Net and a pre-trained VGG convolutional layers. The pre-trained frozen layers with trainable layer(s) encode radiomic features effectively, and layers in the decoder reconstruct synthesized images using the encoded features.

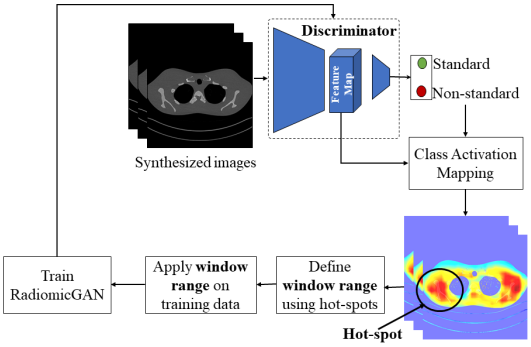

Fig. 2: **Dynamic selection framework.** Given a set of CT images synthesized by the generator G, the non-trainable discriminator D is used to generate the corresponding class activation mappings (CAMs). A window range is defined using all the hot-spots in the CAMs. RadiomicGAN is then continuously trained using the updated training data.

training data. The latter is a top-down training approach, where the training starts with the whole HU range; and the effective pixel range reduces dynamically according to the performance of RadiomicGAN so that RadiomicGAN is always focused on the pixel distributions where it suffers most significantly in the previous training iteration.

Also, the window training strategy can be considered as a novel data augmentation function that generates training data with diverse texture distributions while preserving spatial details. Note that the traditional min-max normalization is a special case of the window training where the window size is fixed to be the full range.

Window Training Step 1. Fixed Growing: Algorithm 1 line 1-7 describes the fixed growing approach. In the fixed growing approach, we gradually select all the pixels by specifying pixel ranges with a fixed pace. The first window includes the HU numbers within a narrow effective pixel range (e.g. $[-1024, -769]$). Then, the window size is gradually increased by 256 throughout the training process (e.g., from $[-1024, -769]$ to $[-1024, -511]$), allowing for more pixels being considered in the training process. We repeat the process until the whole HU range is covered (Algorithm 1 line 3-7). Note the window size is increased only if the model accuracy

Algorithm 1 *Window Training*

Input: model M, training data DS, HU min hu_{min}, HU max hu_{max}, step w, training accuracy threshold th_{acc}, training epoch threshold th_{η}, heat-map threshold th_{fail}, and window width threshold th_{win}.

Output: Trained model M_{θ}

Initialisation: M_{θ}

1: $currentStart \leftarrow hu_{min}$
2: $currentEnd \leftarrow hu_{mmin} + w$
3: **while** $hu_{max} < currentEndPoint$ **do**
4: $\quad DS' \leftarrow clip(DS, currentStart, currentEnd)$
5: $\quad M_{\theta} \leftarrow train(M_{\theta}, DS', th_{acc}, th_{\eta})$
6: $\quad currentEnd \leftarrow currentEnd + w$
7: **end while**
8: $winRange \leftarrow DynamicSelection(DS, M_{\theta}, th_{fail})$
9: $currentStart \leftarrow MIN(winRange)$
10: $currentEnd \leftarrow MAX(winRange)$
11: **while** $|currentEnd - currentStart| > th_{win}$ **do**
12: $\quad DS' \leftarrow clip(DS, currentStart, currentEnd)$
13: $\quad M_{\theta} \leftarrow train(M_{\theta}, DS', th_{acc}, 1)$
14: $\quad winRange \leftarrow DynamicSelection(DS, M_{\theta}, th_{fail})$
15: $\quad currentStart \leftarrow MIN(winRange)$
16: $\quad currentEnd \leftarrow MAX(winRange)$
17: **end while**
18: **return** M_{θ} =0

exceeds threshold th_{acc} or reaches maximum training epoch th_{η} (Algorithm 1 line 5). Function $clip(.)$ selects all pixels whose intensity values are within the current window range (Algorithm 1 line 4). Function $train(.)$ continuously trains model M_{θ} using training data DS' until the model accuracy exceeds threshold th_{acc} or reaches maximum training epoch th_{η} (Algorithm 1 line 5).

Window Training Step 2. Dynamic Selection: Unlike the fixed growing approach, the window range in the dynamic selection approach is determined by using the Class Activation Map (CAM), which identifies the subareas of the input images contributed most for a specific classification task (17). The direction of window expansion or shrinking in this approach is not fixed as well. The motivation is be focused on the pixels where the image synthesis failed, meaning that the discrimi-

nator of RadiomicGAN is able to determine the images as a non-standard images.

The dynamic selection approach is illustrated in Figure 2 and Algorithm 1 line 8-17. Let $\mathcal{P}$ be a randomly selected subset of the training data DS and $\mathcal{P}'$ be its corresponding synthesized images. After each training epoch, we feed P' to the current discriminator D and calculate a set of CAMs C using Grad-CAM (18). During the calculation, D remains freeze. For each image $p_i \in P'$, the layer before the soft-max layer in D indicates the probability of being a standard and non-standard image, which is used to calculate the CAM c_i of p_i. CAM can be visualized as a heat-map where values close to 1 are critical for determining the image being non-standard. With a user given threshold th_{fail}, we can determine the subareas of p_i contributed most for predicting p_i to be a non-standard image ($th_{fail} \in [0, 1]$). For image p_i, if a pixel's value in the CAM is greater than th_{fail}, the corresponding CT image pixel value is added to a pixel intensity list named W_i. All the pixels with their frequency equal to 1 are discarded from W_i considering them as image noise. This process is repeated by randomly selecting subsets of images from the training data. The window range is defined by merging all the pixel intensity lists. The function $DynamicSelection(.)$ in Algorithm 1 line 8 calculate the window range based on a training dataset DS, trained model M, and the heat-map threshold th_{fail}. The window selection and training process continues based on the Algorithm 1 line 11-17. Note the window range is calculated in every epoch and the model get trained on the current window range.

III. Experimental Results

A. Dataset and Model Implementation

In total 14,372 CT image slices from five lung cancer patient scans and ten phantom scans were obtained using three different reconstruction kernels (Bl57, Bl64, and Br40) and four different slice thicknesses (0.5, 1, 1.5, 3mm) using the Siemens CT Somatom Force scanner at the University of Kentucky Medical Center. We adopted Bl64 kernel and 1mm slice thickness as the standard CT imaging protocol, since it has been widely used in clinical practice for lung cancer diagnosis (7). Two testing datasets were prepared for RadiomicGAN performance evaluation. The first testing data were captured using the reconstruction kernel Bl57 and 1mm slice thickness. The second testing data were captured using the reconstruction kernel Br40 and 1mm slice thickness. Each test dataset contains 387 image slices and have paired target standard images (Bl64). The first testing data had relatively similar radiomic features compared with the standard images, while the second one are dramatically different to the standard images. HU number range was set to between -1024 and 1000 in the standardization process, since most pixel values belong to this range.

RadiomicGAN consists of a VGG-based U-Net with 26 hidden layers as the generator and a fully convolutional neural network with six hidden layers as the discriminator. Both the input and output dimension of RadiomicGAN were set to $512 \times 512 \times 3$. The convolutional layers used 4×4 filters. LeakyRelu (19) was adopted as the activation function in all the hidden layers. The discriminator of RadiomicGAN was a fully convolutional neural network with six hidden layers adopted from the pix2pix network (11). Batch-normalization was used in each layer, LeakyRelu (19) was used as an activation function, and soft-max was used in the last layer of the discriminator. The RadiomicGAN model is train using the adversarial loss introduced by Goodfellow et al (20). Random weights were used during the network initialization phase. Maximum training epochs were set to 30 with the learning rate being 0.0001 with momentum 0.5. RadiomicGAN was implemented in TensorFlow on a Linux computer server with eight Nvidia GTX 1080 GPU cards. The model took about 8 hours to train from scratch. Once the model was trained, it took about two seconds to synthesize and normalize a CT image slice. (Source code: https://github.com/selim-iitdu/radiomicGAN)

B. Evaluation Metric

Model performance was evaluated systematically at the whole image (DICOM) level and with randomly selected regions of interest (ROIs) in four HU ranges, including $[-800, -300]$, $[-100, 250]$, $[10, 250]$, and $[300, 800]$. Since RadiomicGAN compares standard and non-standard images in the deep feature space during training, we evaluated the model performance in the radiomic feature domain. For each CT image or ROI, a total 1,401 radiomic features were extracted using IBEX (21). These features belong to seven feature classes: Gray Level Co-occurrence Matrix 2.5D, Gray Level Co-occurrence Matrix 3D, Neighbor Intensity Difference 2.5D, Intensity Direct, Intensity Histogram, Neighbor Intensity Difference 2.5D, and Neighbor Intensity Difference 3D.

Four state-of-the-art CT image standardization models, i.e. Histogram matching (22), Choe et al. (10), GANai (7), and STAN-CT (8), were selected for performance comparison. Here, the first model is based on traditional method and the other three methods are deep-learning based models. All the models, including RadiomicGAN, were developed based on TensorFlow (23) and trained and evaluated using the same training and testing data.

We examined the radiomic features reproducibility performance using Concordance Correlation Coefficient(CCC) (24). CCC represents the correlation between the standard and the synthesized image features in a given features class. CCC ranges from -1 to 1 and is the higher the better. We conclude that a radiomic feature is reproducible if the synthesized image is more than 85% similar to the corresponding standard image (i.e., $CCC > 0.85$) (10; 25).

Mathematically, CCC represents the correlation between the standard and the non-standard image features in the seven features classes:

$$CCC = \frac{2\rho_{s,t}\sigma_s\sigma_t}{{\sigma_s}^2{\sigma_t}^2 + (\mu_s - \mu_t)^2} \quad (1)$$

where μ_s and σ_s (or μ_t and σ_t) are the mean and standard deviation of the radiomic features belong to the same feature

TABLE I: **CT image standardization model performance comparison for images acquired with Bl57 kernel.** The values represent the averaged (±standard deviation) number of reproducible radiomic features of the synthesized images generated using different models. The numbers are rounded to the nearest integer.

HU range	Non-standard	Hist. matching	Coe et al.	GANai	STAN-CT	Radiomic-GAN[1]	Radiomic-GAN[2]	Radiomic-GAN
[-800, -300]	854 ± 14	885 ± 70	905 ± 47	979 ± 47	1053 ± 10	**1226 ± 47**	1153 ± 33	1168 ± 36
[-100, 250]	589 ± 20	592 ± 37	605 ± 47	782 ± 1	824 ± 1	**1257 ± 5**	1183 ± 55	1204 ± 27
[10, 250]	409 ± 20	425 ± 27	483 ± 34	566 ± 15	582 ± 38	**722 ± 41**	690 ± 34	706 ± 66
[300, 800]	457 ± 16	494 ± 100	511 ± 40	817 ± 20	594 ± 9	**902 ± 20**	853 ± 54	853 ± 54
Average	577 ± 18	599 ± 58	626 ± 42	786 ± 21	763 ± 14	**1027 ± 28**	970 ± 44	983 ± 46

TABLE II: **CT image standardization model performance comparison for images acquired with Br40 kernel.** The values represent the averaged (±standard deviation) number of reproducible radiomic features of the synthesized images generated using different models. The numbers are rounded to the nearest integer.

HU range	Non-standard	Hist. matching	Coe et al.	GANai	STAN-CT	Radiomic-GAN[1]	Radiomic-GAN[2]	Radiomic-GAN
[-800, -300]	448 ± 14	530 ± 64	796 ± 14	850 ± 15	933 ± 42	1036 ± 25	1087 ± 33	**1126 ± 43**
[-100, 250]	303 ± 20	355 ± 43	437 ± 38	572 ± 10	588 ± 10	**1131 ± 29**	1098 ± 44	1108 ± 25
[10, 250]	211 ± 53	73 ± 65	167 ± 21	205 ± 37	300 ± 192	348 ± 171	512 ± 49	**611 ± 50**
[300, 800]	246 ± 16	430 ± 17	357 ± 47	520 ± 45	487 ± 8	1030 ± 15	1047 ± 36	**1047 ± 36**
Average	302 ± 26	347 ± 47	439 ± 30	537 ± 27	577 ± 63	886 ± 60	936 ± 40	**973 ± 38**

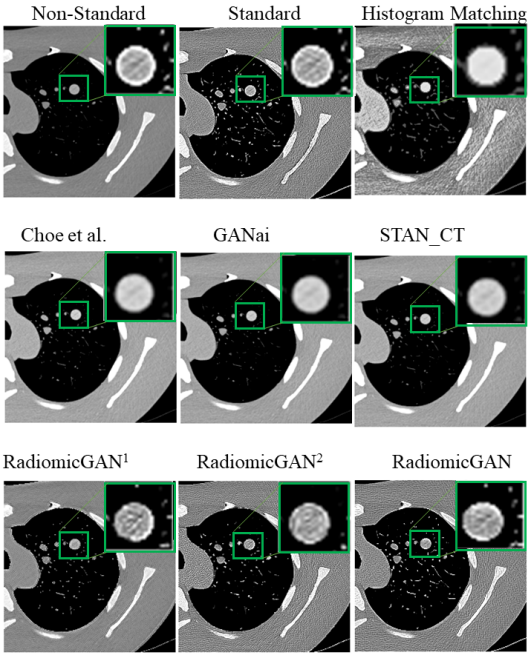

Fig. 3: **CT image standardization using different models.** An image with a tumor is used as a case study to show the visual quality of all the compared models. The green rectangle highlights a tumor in the ROI. The display window is [-800, 800] HU.

class in a synthesized (or standard) image respectively, and $\rho_{s,t}$ is the Pearson correlation coefficient between s and t. CCC ranges from -1 to 1 and is the higher the better.

C. Performance Evaluation

We compared RadiomicGAN with Histogram Matching, Choe et al, GANai, and STAN-CT. We considered three versions of RadiomicGAN. The RadiomicGAN was trained using the proposed window training strategy where both the fixed growing approach and the dynamic selection approach were used consecutively (section II-3). RadiomicGAN[1] was a variation trained using the fixed growing approach only, and RadiomicGAN[2] was another variation trained using the dynamic selection approach only.

Table I and Table II indicate the effectiveness of CT image standardization in different ROIs while standardizing CT images captured using Bl57 and Br40 kernels respectively. RadiomicGAN[1], RadiomicGAN[2], and RadiomicGAN outperformed the models-to-compare in almost all the evaluation results for all ROIs. Here, the column named "non-standard" shows the performance of the images before standardization.

Figure 3 illustrates the performance of all the models compared using a case study. The nonstandard and the standard were phantom CT images acquired with the Br40 and Bl64 kernels respectively. The tumor in the image was highlighted in a green box and was magnified in the left upper corner. The results of all the models were visualized as well. A visual inspection indicates that the tumor image synthesized with RadiomicGAN was the most similar to the standard image. The total number of reproducible features (computed using CCC) was 612 for RadiomicGAN, higher than all the compared methods (Histogram Matching: 82, Choe et al: 183, GANai: 209, and STAN-CT: 307). The RadiomicGAN[1] and RadiomicGAN[2] have 487 and 512 total number of reproducible features respectively which are also better than the compared models. The PSNR score of RadiomicGAN was 33.23, clearly higher than the compared methods (Hist. Matching 24.15, Choe et al: 26.08, GANai: 26.14, and STAN-CT: 26.07). The PSNR score of RadiomicGAN[1] and RadiomicGAN[2] are 27.85 and 30.09 respectively.

D. Evaluation of Window Training

We have seen from Table I and Table II that RadiomicGAN with window training outperforms the existing models

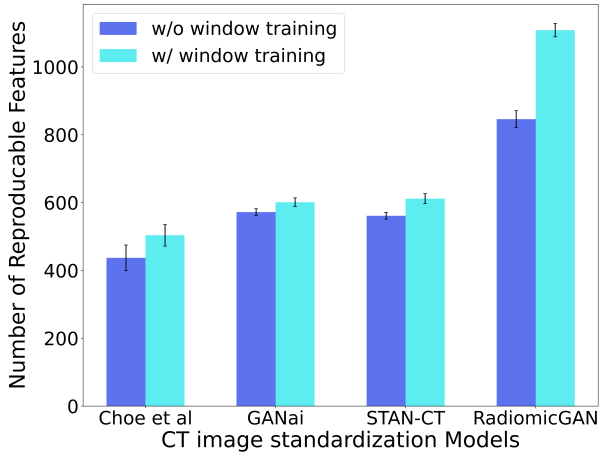

Fig. 4: **The effect of window training in CT image standardization.** Model performance were evaluated with and without window training.

regarding the number of the reproducible radiomic features. To identify the optimal window training strategy, an ablation study was conducted to compare the fixed growing approach, the dynamic selection approach, and the combined approach.

To further determine the effectiveness of the window training, we applied it on all the compared CT image standardization models. Figure 4 shows the model performance of using or not using window training for all the compared CT image standardization models applied on soft tissue ROIs. The x-axis represents the CT image standardization models and the y-axis is the number of the reproducible features measured using CCC. The result shows that RadiomicGAN improved significantly with window training (t-test, p-value$<$0.01), whereas the other models only obtained limited benefit from window training. It implies window training can be best utilized together with transfer learning to fine tune a pre-trained deep learning model.

IV. Conclusion

CT image radiomic feature discrepancy 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. The much wider dynamic range of CT images has been hindering the adaptation of transfer learning onto the CT image synthesis task. RadiomicGAN addresses these challenges by efficiently standardizing and normalizing clinically usable synthetic CT images. A novel window training strategy is proposed in RadiomicGAN allowing the model to be gradually exposes to the data points with more local intensity details, thus significantly improving model performance. In the experiments, we systematically extracted 1,401 radiomic features frequently used in radiomic models and the results show that RadiomicGAN has significantly increased the number of reproducible radiomic features. In the future, we will extend RadiomicGAN to standardize CT images acquired by different CT scanners, further evaluate RadiomicGAN with patient data collected from multiple institutes, and investigate why certain radiomic image features are more difficult to standardize than the others.

Acknowledgment

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

References

[1] J. T. Bushberg and J. M. Boone, *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.
[2] A. Midya, J. Chakraborty, M. Gönen, R. K. Do, and A. L. Simpson, "Influence of ct acquisition and reconstruction parameters on radiomic feature reproducibility," *Journal of Medical Imaging*, vol. 5, no. 1, p. 011020, 2018.
[3] R. Berenguer and M. d. R. e. a. Pastor-Juan, "Radiomics of ct features may be nonreproducible and redundant: Influence of ct acquisition parameters," *Radiology*, p. 172361, 2018.
[4] L. A. Hunter, S. Krafft, and F. e. a. Stingo, "High quality machine-robust image features: Identification in nonsmall cell lung cancer computed tomography images," *Medical physics*, vol. 40, no. 12, 2013.
[5] J. J. Foy, K. R. Robinson, and H. e. a. Li, "Variation in algorithm implementation across radiomics software," *Journal of Medical Imaging*, vol. 5, no. 4, p. 044505, 2018.
[6] M. F. Cohen and J. R. Wallace, *Radiosity and realistic image synthesis*. Elsevier, 2012.
[7] G. Liang, S. Fouladvand, and J. e. a. Zhang, "Ganai: Standardizing ct images using generative adversarial network with alternative improvement," in *2019 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2019, pp. 1–11.
[8] M. Selim, J. Zhang, and B. e. a. Fei, "Stan-ct: Standardizing ct image using generative adversarial network," in *AMIA Annual Symposium Proceedings*, vol. 2020. American Medical Informatics Association, 2020.
[9] L. Gao, H. Sun, X. Ni, M. Fang, and T. Lin, "Effects of 16-bit ct imaging scanning conditions for metal implants on radiotherapy dose distribution," *Oncology letters*, vol. 15, no. 2, pp. 2373–2379, 2018.
[10] J. Choe, S. M. Lee, and K.-H. e. a. Do, "Deep learning–based image conversion of ct reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses," *Radiology*, vol. 292, no. 2, pp. 365–373, 2019.
[11] P. Isola, J.-Y. Zhu, and T. Z. et al, "Image-to-image translation with conditional adversarial networks," in *Computer Vision and Pattern Recognition (CVPR)*, 2017.
[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.
[13] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Advances in neural information processing systems*, 2015, pp. 262–270.
[14] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2414–2423.
[15] H. C. Shin, H. R. Roth, and M. G. et al, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[16] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Advances in neural information processing systems*, 2019, pp. 3347–3357.
[17] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.
[18] R. R. Selvaraju, M. Cogswell, and A. e. a. Das, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
[19] B. Xu, N. Wang, T. Chen *et al.*, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.
[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, 2014, pp. 2672–2680.
[21] L. Zhang, D. V. Fried, and X. J. e. a. Fave, "IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics," *Medical physics*, vol. 42, no. 3, pp. 1341–1353, 2015.
[22] A. K. Jain, *Fundamentals of digital image processing*. Englewood Cliffs, NJ: Prentice Hall,, 1989.
[23] M. Abadi, A. Agarwal, P. Barham *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/
[24] I. Lawrence and K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, pp. 255–268, 1989.
[25] B. Zhao, Y. Tan, and W.-Y. e. a. Tsai, "Reproducibility of radiomics for deciphering tumor phenotype with imaging," *Scientific reports*, vol. 6, no. 1, pp. 1–7, 2016.