

# UDA-CT: A General Framework for CT Image Standardization

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

Matthew Lewis

Dept. of Radiology  
UT Southwestern Medical Center  
Dallas, TX  
matthew.lewis@utsouthwestern.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

**Abstract**—Large-scale CT image studies often suffer from a lack of homogeneity regarding radiomic characteristics due to the images acquired with scanners from different vendors or with different reconstruction algorithms. We propose a deep learning-based framework called UDA-CT to tackle the homogeneity issue by leveraging both paired and unpaired images. Using UDA-CT, the CT images can be standardized both from different acquisition protocols of the same scanner and CT images acquired using a similar protocol but scanners from different vendors. UDA-CT incorporates recent advances in deep learning including domain adaptation and adversarial augmentation. It includes a unique design for model training batch which integrates non-standard images and their adversarial variations to enhance model generalizability. The experimental results show that UDA-CT significantly improves the performance of the cross-scanner image standardization by utilizing both paired and unpaired data.

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

## I. INTRODUCTION

Computed Tomography (CT) image acquisition using non-standardized protocols is a common practice due to the clinical requirements and patients' diagnostics need [1]–[5]. This flexibility unavoidably encounter inconsistent radiomic features posing significant challenges in large-scale cross-center radiomic data analysis and clinical decision support [6]. Thus, intra-scanner (different acquisition protocols of the same scanner) and cross-scanner (similar acquisition protocols of different scanners) radiomic feature variation remains a critical challenge.

The CT image radiomic feature discrepancy problem could be addressed by defining a universal standard for CT image acquisition. However, this approach may significantly reduce the scope of the CT imaging modality [7], [8]. Recent advancement has been made to address the CT image inconsistency problem by developing deep-learning based post-processing framework to standardize and normalize existing CT images while preserving most of the anatomic details [9]–[13].

There are in general two kinds of CT image standardization models for difference purposes and also depending on data availability. The first category is called intra-scanner image standardization which requires paired image data. In this setting, two images constructed from the same scan but different reconstruction kernels form an image pair, where the image constructed with the non-standard kernel (e.g. Siemens Br40) is called the source image and the image constructed with the standard kernel (e.g. Siemens B164) is called the target image. Given the paired image training data, a model learns how to convert source images to target images. The second category of models were designed for cross-scanner image standardization which does not requiring paired image data. In this setting, paired images are not required. Instead, images acquired with the standard protocol and images acquired with the non-standard protocols are saved separately. Getting paired training data is relatively easy but most of the time is limited within one scanner. Generally the demand of standardization is higher for cross vendor in large-scale radiomic study which cannot be done with models of first category.

We propose a general framework named UDA-CT for both intra- and cross-scanner image standardization. Its network architecture is shown in Fig 1. UDA-CT utilizes both paired and unpaired CT images to learn a mapping from multiple non-standard image distributions to the same standard image distribution. UDA-CT consists of two generators and two discriminators. The two input components are responsible for intra-scanner and cross-scanner image data synthesis respectively. In summary, UDA-CT is a general framework for both intra- and cross-scanner image standardization. Experimental results show that UDA-CT is significantly better than the current image standardization models.

## II. METHOD

UDA-CT is a generalized framework that supports both intra- and cross-scanner CT image standardization, leveraging

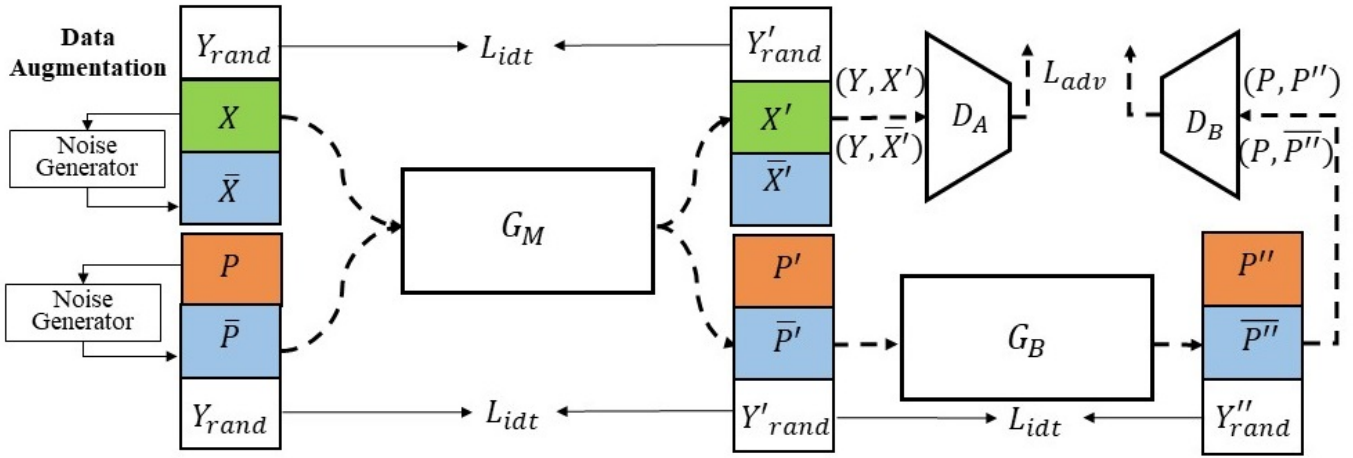


Fig. 1. **Overview of UDA-CT.** UDA-CT has two input components: the first component for domain A image synthesis where paired training data  $(X, Y)$  are provided, the second component for domain B data synthesis where unpaired images  $(P, Y_{rand})$  are provided.  $\bar{X}$  and  $\bar{P}$  are augmented version of perturbation  $X$  and  $P$  with added Gaussian noise.

unsupervised domain adaptation [14] and CycleGAN [15]. UDA-CT has two major components, where the first component is responsible for domain A image synthesis where paired training data are provided, the second component is responsible for domain B data synthesis where unpaired images are provided.

#### A. Problem Definition

Domain  $A$  is defined as a collection of paired CT image  $(x_i, y_i) \in (X, Y)$  where images in  $X$  is captured using non-standard acquisition protocols and  $Y$  is captured using the pre-defined standard acquisition protocol.

Domain  $B$  is defined as a collection of unpaired CT image  $p_i \in P$  captured using non-standard acquisition protocols. For any  $p_i \in P$ , the corresponding standard image does not exist.

The problem is to identify a mapping function from both  $X$  and  $P$  to  $Y$  such that the mapping function can effectively synthesize image  $x'$  from image  $x$  in domain  $A$  and  $p'$  from image  $p$  in domain  $B$  to the standard image distribution  $Y$  such that both  $x'$  and  $p'$  follow the distribution of  $Y$  rather than  $X$  or  $P$ .

#### B. Neural network architecture

Figure 1 shows the architecture of UDA-CT. UDA-CT consists of two generators and two discriminators. The first generator named “master generator”  $G_M$  is responsible for standardizing images acquired using all kinds of protocols to a pre-defined standard domain. Specifically, given paired CT images collected from domain A and unpaired images collected from domain B,  $G_M$  is trained to synthesize the corresponding standard image  $Y$ . The second generator named “ancillary generator”  $G_B$  is a domain-specific generator for converting images from standard image distribution  $Y$  to domain B. Discriminators  $D_A$  classifies target images and synthesized images in domain A, and  $D_B$  detects whether a synthesized image fall in the distribution of domain B.

Both generators  $G_M$  and  $G_B$  consist of seven convolutional layers and each layer has 32 feature maps with stride 1. Discriminators  $D_A$  and  $D_B$  consist with six fully connected convolutional layers [16].

#### C. Data Augmentation

To improve model generalizability, a comprehensive data augmentation approach that perturbs the inputs with adversarial noise is developed in UDA-CT. Data augmentation has long been used for enhancing model generalizability and preventing model over-fitting [14]. The proposed data augmentation approach improves model generalizability by preparing three input sets. Given a batch of images from the same domain, a training batch is constructed with three components: 1) randomly selected standard images  $Y_{rand}$ , 2) images captured with non-standard protocols  $X$  or  $P$ , and 3) adversarial samples  $\bar{N}$ , which is the perturbed version of the second component with systematic noise. The noise is generated from a Gaussian distribution described in [14].

### III. EXPERIMENT SETTING

Five deep learning-based CT image standardization models trained on paired training dataset, i.e. GANai [17], STAN-CT [11], RadiomicGAN [13], CycleGAN [15], and CVH-CT [12], were selected for performance comparison. Here, the first three models were based on paired training data, and the last two models were designed for unpaired training data. A variations of UDA-CT named UDA-CT<sub>BASIC</sub> was developed for ablation study by removing the data augmentation component. All the models, including UDA-CT, were implemented on PyTorch [18]. All the models were trained and evaluated using the same training and testing dataset.

#### A. Data

Both public and in-house image data were used for model training and testing. The public CT image data include 4255

image slices from The Cancer Imaging Archive (TCIA) downloaded in December 2021 [19]. The in-house CT images were collected with two scanners made by Siemens, and GE Medical Systems respectively. It includes images from patients and phantom captured using Siemens Somatom Force in University of Kentucky (UKY), GE Revolution EVO in VA hospital, Lexington, KY. In total, 9900 image slices were collected from Siemens scanner using B164 and Br40 reconstruction kernels. 4255 image slices were collected from GE scanner using LUNG reconstruction kernel.

To build high-quality testing data, we scanned the same Lungman chest phantom using the two aforementioned scanners with three different synthetic tumors. Slice thickness was set to 5 mm as it was the only common slice thickness within the three scanners. Note that phantom images were only used for model testing.

### B. Evaluation Metric

Model performance was evaluated based on the lung tumors. For each tumors, a total 1,401 radiomic features were extracted using IBEX [20]. These features belong to nine feature classes: Gradient Oriented Histogram, Gray Level Co-occurrence Matrix (GLCM) 2.5D, Gray Level Co-occurrence Matrix (GLCM) 3D, Gray Level Run Length Matrix (GLRLM) 2.5D, Neighbor Intensity Difference (NID) 2.5D, Intensity Direct (ID), Intensity Histogram (IH), Neighbor Intensity Difference (NID) 2.5D, and Neighbor Intensity Difference (NID) 3D.

Based on these radiomic features, we evaluated UDA-CT and the models-to-compare using absolute relative error. The absolute relative error (RE, see Eq. 1), defined as the summarized relative difference between a synthesized image and its corresponding standard image regarding radiomic features, was utilized to calculate the linear distance between the standard and the synthesized images for each radiomic feature after standardization:

$$RE = \frac{|f_t - f_s|}{|f_t|} \quad (1)$$

where  $f_t$  and  $f_s$  are the radiomic features of the standard and synthesized image respectively.

A radiomic feature is reproducible if the synthesized image is more than 85% similar to the corresponding standard image [21], [22]. Mathematically, a reproducible feature means that a tumor's  $RE < 0.15$ .

## IV. EXPERIMENTAL RESULTS

UDA-CT was compared with GANai, STAN-CT, and RadiomicGAN in domain A and with CycleGAN and CVH-CT in domain B due to the different requirements of paired/unpaired images for model training.

In this experiment, as a demonstration, Siemens B164 was used as image standard, while the other protocols (Siemens Br40, GE Lung) were considered as non-standard. Specifically, Siemens Br40 and B164 images were defined as domain A, and, GE lung images as domain B. The experiment can be easily extended to any clinically used imaging protocols.

### A. Paired Image Standardization

For domain A applications where paired standard and non-standard images are provided for model training, the testing task is to convert Siemens Br40 to Siemens B164. Table I shows the number of reproducible radiomic features and the relative errors of three tumors respectively. The performance of the baseline was measured using the original input images. For tumor 1, UDA-CT had 44% more reproducible radiomic features than the baseline (GANai: 21%, STAN-CT: 24%, and RadiomicGAN: 35%). For tumor 2, UDA-CT had 33% more reproducible radiomic features than the baseline (GANai: 4%, STAN-CT: 14%, and RadiomicGAN: 19%). For tumor 3, UDA-CT and the compared models had approximately 4% more reproducible radiomic features than the baseline. UDA-CT<sub>BASIC</sub>, which is the vanilla domain adaptation version of the UDA-CT had improved the performance compared with the baseline and other models.

### B. Unpaired Image Standardization

For domain B applications where only non-standard image pairs were provided for model training, the testing task is to convert images from GE Lung to Siemens B164. Table II shows the number of reproducible radiomic features and the relative errors of three tumors respectively respectively. For tumors 1 and 2, all the models had a higher total number of reproducible features compared to the baseline. For tumor 1, UDA-CT had 13% more reproducible radiomic features than the baseline (CycleGAN: 9% and CVH-CT: 12%). For tumor 2, UDA-CT had 12% more reproducible radiomic features than the baseline (CycleGAN: 5% and CVH-CT: 7.5%). For tumor 3, UDA-CT and the compared models did not change the number of reproducible features.

## V. DISCUSSION AND CONCLUSION

Recently, deep learning-based image standardization algorithms have been developed to standardize CT images acquired with different parameters, aiming to improve the prediction power of clinical outcomes using image-based features. We introduce an unsupervised domain adaptation-based deep learning model called UDA-CT. Experimental results show that UDA-CT significantly enhanced the performance of CT image standardization by leveraging both paired and unpaired image data. This work also provided a systematic evaluation workflow for unified systematic intra-scanner and cross-scanner image standardization assessment.

In the future, we will optimize image standardization strategies for different radiomic feature groups. The results show that almost every model achieved better performance on tumor 1 than on the rest tumors. The performance variation is largely located on GLRLM, which uniquely requires an exact match of neighboring pixels. To further improve model performance, strategies to better harmonize GLRLM features are required.

### ACKNOWLEDGMENT

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

TABLE I

THE RELATIVE ABSOLUTE ERROR OF DOMAIN A IMAGE SYNTHESIS. EACH ROW REPRESENTS THE TOTAL NUMBER OF REPRODUCIBLE RADIOMIC FEATURES WITH AN ERROR LESS THAN 0.15 AND THE MEAN  $\pm$  STANDARD DEVIATION OF THE ERROR FOR RESPECTIVE TUMORS.

Tumor	Tumor 1		Tumor 2		Tumor 3	
	# of reproducible features	RE	# of reproducible features	RE	# of reproducible features	RE
Baseline	557	1.59 $\pm$ 4.68	699	1.32 $\pm$ 7.89	651	5.12 $\pm$ 76.99
GANai	851	0.46 $\pm$ 1.29	760	0.74 $\pm$ 4.96	714	0.72 $\pm$ 4.9
STAN-CT	903	0.23 $\pm$ 0.66	896	0.45 $\pm$ 4.36	734	0.74 $\pm$ 7.96
RadiomicGAN	1050	0.26 $\pm$ 0.99	971	0.38 $\pm$ 2.10	867	0.65 $\pm$ 4.81
UDA-CT <sub>BASIC</sub>	1036	0.24 $\pm$ 0.64	902	0.18 $\pm$ 0.28	651	0.24 $\pm$ 1.61
UDA-CT	1174	0.20 $\pm$ 0.56	1162	0.08 $\pm$ 0.45	714	0.21 $\pm$ 2.54

TABLE II

THE RELATIVE ABSOLUTE ERROR OF DOMAIN B IMAGE SYNTHESIS. EACH ROW REPRESENTS THE TOTAL NUMBER OF REPRODUCIBLE RADIOMIC FEATURES WITH AN ERROR LESS THAN 0.15 AND THE MEAN  $\pm$  STANDARD DEVIATION OF THE ERROR FOR RESPECTIVE TUMORS.

Tumor	Tumor 1		Tumor 2		Tumor 3	
	# of reproducible features	RE	# of reproducible features	RE	# of reproducible features	RE
Baseline	908	0.49 $\pm$ 1.46	758	0.27 $\pm$ 0.95	686	0.59 $\pm$ 10.76
CycleGAN	1036	0.25 $\pm$ 0.82	829	0.20 $\pm$ 0.64	646	5.22 $\pm$ 5.14
CVH-CT	1083	0.18 $\pm$ 0.48	863	0.49 $\pm$ 2.13	651	0.55 $\pm$ 5.08
UDA-CT <sub>BASIC</sub>	1049	0.20 $\pm$ 0.56	971	0.38 $\pm$ 2.1	867	0.65 $\pm$ 4.81
UDA-CT	1093	0.17 $\pm$ 0.50	939	0.35 $\pm$ 0.88	673	0.54 $\pm$ 10.63

## REFERENCES

- [1] 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.
- [2] J. L. Prince and J. M. Links, *Medical imaging signals and systems*. Pearson Prentice Hall Upper Saddle River, 2006.
- [3] J. Beutel, H. L. Kundel, and R. L. Van Metter, *Handbook of medical imaging: Physics and psychophysics*. Spie Press, 2000, vol. 1.
- [4] Y. Cui and F.-F. Yin, "Impact of image quality on radiomics applications," *Physics in Medicine & Biology*, vol. 67, no. 15, p. 15TR03, 2022.
- [5] P. Chirra, P. Leo, M. Yim, B. N. Bloch, and et. al, "Empirical evaluation of cross-site reproducibility in radiomic features for characterizing prostate mri," in *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575. International Society for Optics and Photonics, 2018, p. 105750B.
- [6] A. J. Buckler, L. Bresolin, N. R. Dunnick, D. C. Sullivan, and Group, "A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging," *Radiology*, vol. 258, no. 3, pp. 906–914, 2011.
- [7] J. 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*, vol. 81, no. 5, pp. e699–e703, 2012.
- [8] D. S. Gierada, A. J. Bierhals, C. K. Choong, S. T. Bartel, J. H. Ritter, N. A. Das, C. Hong, T. K. Pilgram, K. T. Bae, B. R. Whiting *et al.*, "Effects of ct section thickness and reconstruction kernel on emphysema quantification: relationship to the magnitude of the ct emphysema index," *Academic radiology*, vol. 17, no. 2, pp. 146–156, 2010.
- [9] G. Liang, J. Zhang, M. Brooks, J. Howard, and J. Chen, "radiomic features of lung cancer and their dependency on ct image acquisition parameters," *Medical Physics*, vol. 44, no. 6, p. 3024, 2017.
- [10] M. F. Cohen and J. R. Wallace, *Radiosity and realistic image synthesis*. Elsevier, 2012.
- [11] M. Selim, J. Zhang, B. Fei, G.-Q. Zhang, and J. Chen, "Stan-ct: Standardizing ct image using generative adversarial network," in *AMIA Annual Symposium Proceedings*, vol. 2020. American Medical Informatics Association, 2020.
- [12] M. Selim, J. Zhang, B. Fei, and et. al., "Cross-vendor ct image data harmonization using cvh-ct," in *AMIA Annual Symposium Proceedings*, vol. 2021. American Medical Informatics Association, 2021, p. 1099.
- [13] M. Selim, J. Zhang, and et. al., "CT image harmonization for enhancing radiomics studies," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 1057–1062.
- [14] J. Wang, C. Lan, C. Liu, and et. al., "Generalizing to unseen domains: A survey on domain generalization," *arXiv preprint arXiv:2103.03097*, 2021.
- [15] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2242–2251.
- [16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [17] G. Liang, S. Fouladvand, J. Zhang, M. A. Brooks, N. Jacobs, and J. Chen, "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.
- [18] A. Paszke, S. Gross, and e. a. Massa, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [19] W. S. L. T. L. J. H. Y. Li, P. and D. Wang, "A large-scale ct and pet/ct dataset for lung cancer diagnosis [data set]," in *The Cancer Imaging Archive*. <https://doi.org/10.7937/TCIA.2020.NNC2-0461>, 2020.
- [20] L. Zhang, D. V. Fried, X. J. Fave, and et. al., "Ibex: an open infrastructure software platform to facilitate collaborative work in radiomics," *Medical physics*, vol. 42, no. 3, pp. 1341–1353, 2015.
- [21] B. Zhao, Y. Tan, W.-Y. Tsai, J. Qi, C. Xie, L. Lu, and L. H. Schwartz, "Reproducibility of radiomics for deciphering tumor phenotype with imaging," *Scientific reports*, vol. 6, no. 1, pp. 1–7, 2016.
- [22] J. Choe, S. M. Lee, K.-H. Do, G. Lee, J.-G. Lee, S. M. Lee, and J. B. Seo, "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.