Cycle-CTFlow: A CycleGAN–Normalizing Flow Harmonization Framework for Improved Nodule Detection

Anil Yadav[*1,2] ayadav01@ucla.edu
[1] *Department of Bioengineering, Samueli School of Engineering, University of California, Los Angeles, 90095, CA, USA*
[2] *Medical & Imaging Informatics Group, Department of Radiological Sciences, David Geffen School of Medicine, UCLA, Los Angeles, 90095, CA, USA*
Nathan Tran[2] nathantran3819@gmail.com
Spencer Welland[3] SWelland@mednet.ucla.edu
[3] *Center for Computer Vision and Imaging Biomarkers, Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA*
John M. Hoffman[3] JMHoffman@mednet.ucla.edu
Grace Hyun J. Kim[3] GraceKim@mednet.ucla.edu
Ashley E. Prosper[2] AProsper@mednet.ucla.edu
Denise R. Aberle[1,2] DAberle@mednet.ucla.edu
Michael F. McNitt-Gray[3] MMcNittGray@mednet.ucla.edu
William Hsu[1,2] WHsu@mednet.ucla.edu

Editors: Under Review for MIDL 2025

Abstract

Although many deep learning models exist for nodule detection, characterization, and other related tasks, their widespread clinical adoption is hindered by substantial variability in computed tomography (CT) acquisitions. Differences in scanner hardware, reconstruction methods, dose levels, and patient demographics cause domain shifts, often leading models trained on one dataset to underperform on another. This highlights the need for harmonization and adaptation strategies to ensure consistent performance across diverse clinical settings. To address this challenge, we propose a training scheme in which knowledge of the downstream model's training distribution is incorporated into the training process of the harmonization model. The goal is to guide the harmonization model to transform input images so that their distribution closely aligns with that of the downstream model. We demonstrate that the proposed approach, Cycle-CTFlow, leads to improvements in nodule detection performance: a 5.2% increase in sensitivity and a 2.6% increase in CPM compared to no harmonization on the MiniDeepLesion dataset, and a 1.9% increase in sensitivity and an 8.5% increase in CPM on the UCLA in-house dataset.

Keywords: Computed Tomography, Harmonization, Normalizing Flow

[*] Corresponding author

1. Introduction

Computed tomography (CT) is widely recognized for its pivotal role in the detection and management of various diseases such as lung cancer, where early and accurate identification of pulmonary nodules can significantly improve patient outcomes (Aberle et al., 2011). The rapid growth of CT datasets has fueled a surge in AI-driven diagnostic tools (Hosny et al., 2018); however, deploying these models in real-world clinical settings remains challenging due to limited generalizability. Substantial variability in CT images—due to differences in scanner hardware, acquisition settings, and patient populations—can lead to domain shifts that degrade the performance of deep learning models when applied to unseen data (Emaminejad et al., 2018).

We propose a framework that incorporates knowledge of the downstream model's target distribution into the harmonization process. We select an open-source nodule detection model from MONAI as our downstream model, which is trained on the publicly available Lung Nodule Analysis 2016 (LUNA16) dataset, and aim to improve its generalizability on external datasets. We train a normalizing flow–based harmonization model using our in-house low-dose CT (LDCT) dataset, which includes paired high-dose reconstructions, and introduce a cycleGAN that transforms the texture of the reference high-dose images to resemble that of the LUNA16 dataset. We hypothesize that the flow model will learn to map low-dose images to their texture-transformed high-dose counterparts, resulting in improved performance on the downstream task (i.e., nodule detection).

2. Methods

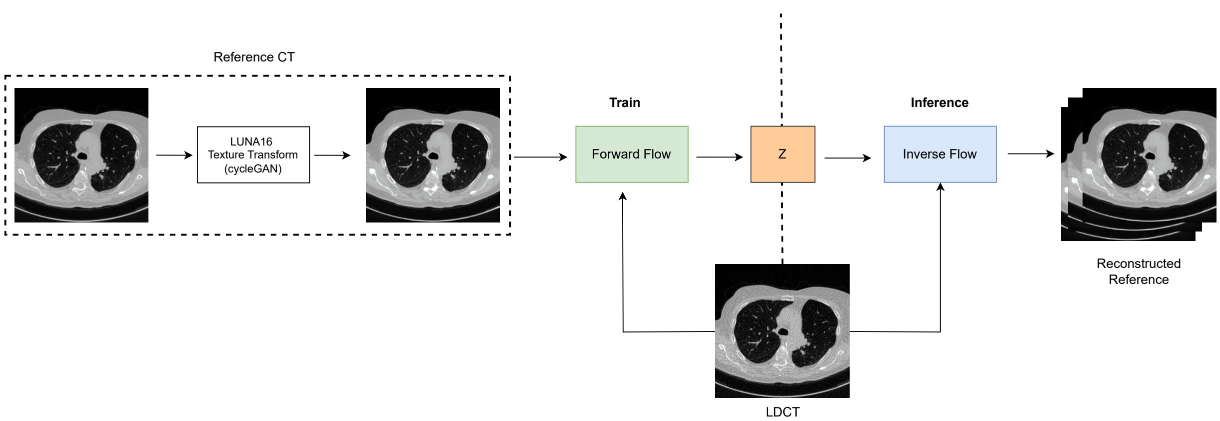

Figure 1: Training scheme of the Cycle-CTFlow model.

We used two datasets for training. The first was an in-house dataset of 100 LDCT exams (31,308 slices), acquired on a Siemens Definition AS64 scanner with 1mm slices using the B45 kernel (i.e., reference condition). To simulate a lower-dose setting, Poisson noise equivalent to 25% of the original dose was added to the raw projection data—referred to as the Medium/25% condition. The second dataset was LUNA16, from which we extracted an equal number of slices to match the in-house LDCT dataset.

Table 1: Performance comparison of harmonization methods across two datasets. CPM represents the average sensitivity at seven predefined false positive rates, as defined by the LUNA16 evaluation criteria.

Datasets	Model	TP	FP	FN	Sensitivity	CPM
MiniDeepLesion	No Harmonization	76	442	12	0.8636	0.5633
	CTFlow-Base	77	1447	11	0.8750	0.5649
	Cycle-CTFlow	80	454	8	0.9090	0.5779
UCLA in-house	No Harmonization	104	445	8	0.9285	0.6011
	CTFlow-Base	103	509	9	0.9196	0.6172
	Cycle-CTFlow	106	534	6	0.9464	0.6525

Using 30% of the 31,308 slices, we first trained a previously published CycleGAN model with a CBAM module (You et al., 2019) to transform the texture of reference images to resemble those from the LUNA16 dataset. The authors had previously shown that integrating CBAM into CycleGAN improves both PSNR and SSIM compared to the standard CycleGAN. We evaluated this model on 10% of the data, obtaining a mean PSNR of 30.9083 and SSIM of 0.8757. Next, using 70% of the data, we trained a normalizing flow model called CTFlow (Wei et al., 2023), which we previously demonstrated to produce more consistent predictions across varying CT reconstructions compared to a baseline generative network. In this model, the latent variable z serves as an invertible representation of the input CT, enabling the reconstruction of a distribution of plausible harmonized images rather than a single deterministic output.

We applied two training schemes: in the first, CTFlow was trained to map Medium/25% images to the reference images (i.e., CTFlow-Base); in the second, it was trained to map Medium/25% images to the texture-transformed reference images generated by CycleGAN (i.e., Cycle-CTFlow). We evaluated the methods on two datasets. The first is the publicly available MiniDeepLesion dataset from Kaggle, which contains 88 lung nodule cases derived from the NIH DeepLesion dataset (Yan et al., 2018). The second dataset comprises our in-house collection of CT scans from never-smokers ($<$ 100 lifetime cigarettes) and ever-smokers, referred to as the UCLA in-house dataset, and includes 112 lung nodule cases.

3. Results and Discussion

Table 1 shows that Cycle-CTFlow consistently outperforms the baseline model across both datasets, achieving the highest sensitivity and CPM scores. It yields fewer false negatives, which is crucial in medical imaging where missing a true nodule can lead to serious clinical consequences. Although Cycle-CTFlow results in more false positives than No Harmonization, its reduction in false negatives is a critical advantage. This trade-off is often preferred in diagnostic settings, where maximizing true positive detections outweighs the cost of additional false alarms. We demonstrate that the texture-transformed training scheme is a more effective approach to harmonization, and our future efforts will focus on improving this module, extending beyond the current reliance on CycleGAN.

Acknowledgments

The authors acknowledge funding for this work from the NIH/National Institute for Biomedical Imaging and Bioengineering under award number R01 EB031993. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsoring agency.

References

Denise R Aberle, Amanda M Adams, Christine D Berg, William C Black, Jonathan D Clapp, Richard M Fagerstrom, Ilana F Gareen, Constantine Gatsonis, Sicks Jorean D Marcus, Pamela M, and The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.*, 365(5):395–409, August 2011.

Nastaran Emaminejad, Muhammad Wahi-Anwar, John Hoffman, Grace H. Kim, Matthew S. Brown, and Michael McNitt-Gray. The effects of variations in parameters and algorithm choices on calculated radiomics feature values: initial investigations and comparisons to feature variability across CT image acquisition conditions. In Nicholas Petrick and Kensaku Mori, editors, *Medical Imaging 2018: Computer-Aided Diagnosis*, volume 10575, pages 886 – 895. International Society for Optics and Photonics, SPIE, 2018. doi: 10.1117/12.2293864. URL https://doi.org/10.1117/12.2293864.

Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo J. W. L. Aerts. Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8):500–510, Aug 2018. ISSN 1474-1768. doi: 10.1038/s41568-018-0016-5. URL https://doi.org/10.1038/s41568-018-0016-5.

Leihao Wei, Anil Yadav, and William Hsu. Ctflow: Mitigating effects of computed tomography acquisition and reconstruction with normalizing flows. In Hayit Greenspan, Anant Madabhushi, Parvin Mousavi, Septimiu Salcudean, James Duncan, Tanveer Syeda-Mahmood, and Russell Taylor, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 413–422, Cham, 2023. Springer Nature Switzerland.

Ke Yan, Xiaosong Wang, Le Lu, and Ronald M. Summers. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of Medical Imaging*, 5(03):1, July 2018. ISSN 2329-4302. doi: 10.1117/1.jmi.5.3.036501. URL http://dx.doi.org/10.1117/1.JMI.5.3.036501. https://www.kaggle.com/datasets/kmader/nih-deeplesion-subset?select=minideeplesion.

Qijing You, Cheng Wan, Jing Sun, Jianxin Shen, Hui Ye, and Qiuli Yu. Fundus image enhancement method based on cyclegan. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4500–4503, 2019. doi: 10.1109/EMBC.2019.8856950.