Region-Specific Masked Loss for Preserving Bone HU in CECT-to-VNC CT Translation

Juhyeong Ki[1] JUHYEONGKI@UNIST.AC.KR
Hyeongmin Jin[2] HMJIN@SNU.AC.KR
Seongmoon Jung[3] SMJUNG@KRISS.RE.KR
Jimin Lee[1] JIMINLEE@UNIST.AC.KR
[1] *Department of Nuclear Engineering, Ulsan National Institute of Science and Technology, Ulsan Republic of Korea*
[2] *Department of Radiation Oncology, Seoul National University Hospital, Seoul, Republic of Korea*
[3] *Biomedical Metrology, Ionizing Radiation Group, Korea Research Institute of Standards and Science, Daejeon, Republic of Korea*

Editors: Under Review for MIDL 2026

Abstract

Deep learning-based virtual non-contrast computed tomography (VNC CT) can enable accurate dose calculation in radiotherapy planning without additional radiation exposure. However, when dual-energy CT (DECT) derived VNC images are used as training targets, limitations of the DECT-based VNC algorithm can lead to reduced HU values in bone regions. To address this issue, a Region-Specific Masked Loss is proposed using bone masks extracted through a SAM-assisted semi-automated pipeline. A bVNC-Net model was proposed and trained on DECT data from 24 liver cancer patients (5,146 slices). The proposed method effectively removed soft-tissue contrast enhancement while preserving bone HU values close to those of the original images.

Keywords: Dual-energy computed tomography, virtual non-contrast CT, radiotherapy planning.

1. Introduction

Using contrast-enhanced CT (CECT) directly for radiotherapy planning can introduce dose calculation errors because contrast agents alter CT attenuation values (Yamada et al., 2014). Although true non-contrast CT (TNC) is preferred, acquiring an additional scan increases radiation exposure and may introduce registration mismatch. Deep learning-based VNC CT translation offers a practical alternative, and DECT-derived VNC images are attractive training targets because they provide sufficient paired data (Kim et al., 2024). However, DECT-based VNC algorithms can undesirably reduce bone HU values even though bone itself shows little contrast enhancement (Li et al., 2020; Parakh et al., 2021). As a result, models trained with such targets inherit the same limitation, which is problematic because accurate bone HU values are important for dose attenuation calculation. To address this issue, a Region-Specific Masked Loss is proposed, in which the original image is used as the target in bone regions and the DECT-derived VNC image is used as the target in soft-tissue regions.

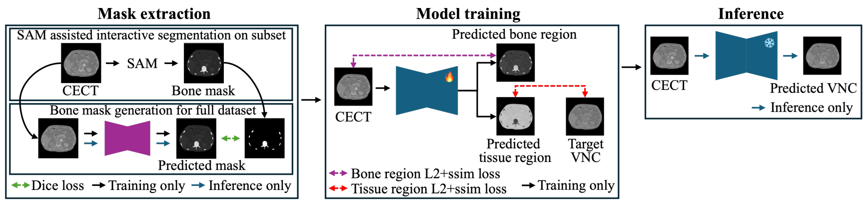

Figure 1: Overview of the proposed pipeline: SAM-assisted bone mask extraction, region-specific masked training for CECT-to-VNC CT translation, and mask-free inference.

2. Methods

Figure 1 summarizes the proposed pipeline, which consists of three stages: semi-automated bone mask extraction, region-specific masked training, and mask-free inference.

2.1. Dataset

CT data from 24 liver-cancer patients at Seoul National University Hospital were used. All scans were acquired with a Philips IQon DECT scanner. The cohort was split by patient into training, validation, and test sets of 17, 4, and 3 patients (3,851, 803, and 492 slices; 5,146 total), where each slice had a matrix size of 512 $\times$ 512.

2.2. SAM-Assisted Bone Mask Extraction

Bone masks were obtained in two steps: interactive labeling of a subset of CECT slices with SAM (Kirillov et al., 2023) in CVAT (Corporation, 2026), then training a bone segmentation model on those labels and applying it across the dataset. Masks were used only during model training.

2.3. Region-Specific Masked Loss

As shown in Figure 1, a U-Net-based (Ronneberger et al., 2015) bVNC-Net model was proposed and trained to translate CECT to VNC CT. The bone mask was used to preserve

Table 1: Quantitative comparison between the baseline and proposed models in terms of SSIM and PSNR.

Model	Region	SSIM ↑	PSNR (dB) ↑
bVNC-Net (Baseline)	Tissue	0.997586 ± 0.000826	48.629502 ± 2.642235
	Bone	0.997049 ± 0.001606	27.998624 ± 2.107025
bVNC-Net (Region-Specific)	Tissue	**0.997807 ± 0.000752**	**49.968323 ± 2.721779**
	Bone	**0.999996 ± 0.000003**	**63.982071 ± 3.315680**

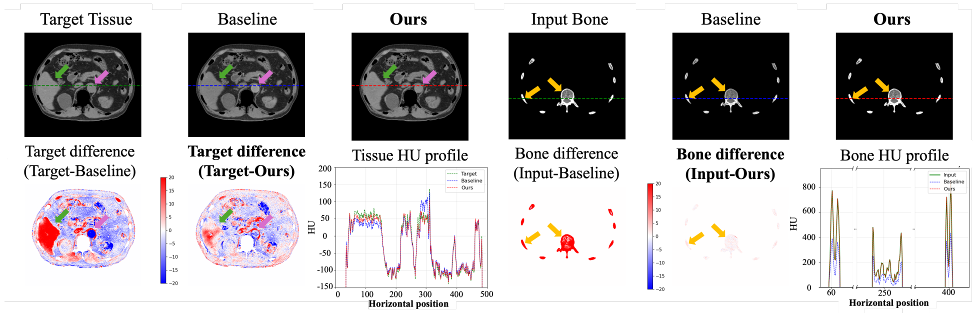

Figure 2: Qualitative comparison of tissue and bone regions between the baseline and proposed methods. The first row shows reconstructed tissue and bone images with the locations of the line profiles, and the second row shows difference maps and HU profiles.

HU values, while the inverted bone mask was used to match soft tissue to the DECT VNC target for contrast removal. In both regions, the loss combined masked L_2 and SSIM in a 0.9:0.1 ratio, with region weights of 0.8 for tissue and 0.2 for bone. Inference required only the input CECT image, and no bone mask was needed after training.

3. Results

Table 1 compares the proposed Region-Specific loss with a baseline model trained using only DECT-derived VNC images as targets. In the tissue region, the proposed model improved SSIM by 0.000221 and PSNR by 1.338821 dB over the baseline, indicating better contrast removal. In the bone region, the proposed model showed a PSNR gain of 35.98 dB over the baseline, with an SSIM of 0.999996, indicating superior preservation of bone HU values. As shown in Figure 2, the proposed model reduced the HU difference in the liver (green arrow) and aorta (pink arrow) compared with the baseline, while better preserving the high-HU outer cortical bone indicated by the orange arrow. These observations are consistent with the corresponding HU profiles, where the proposed model more closely matched the reference trend in both the tissue and bone regions.

4. Conclusion

This study demonstrates that the proposed Region-Specific Masked Loss for CECT-to-VNC CT mitigates tissue and bone HU loss caused by DECT-derived training targets. The proposed approach may also be useful for radiotherapy planning and can be further extended to multi-organ weighting for broader clinical applications.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00344958).

References

CVAT.ai Corporation. Computer vision annotation tool (cvat), April 2026. URL https://doi.org/10.5281/zenodo.19387363.

Jungye Kim, Jimin Lee, Bitbyeol Kim, Sangwook Kim, Hyeongmin Jin, and Seongmoon Jung. Generation of deep learning based virtual non-contrast ct using dual-layer dual-energy ct and its application to planning ct for radiotherapy. *PloS one*, 19(12):e0316099, 2024.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.

Baojun Li, Mark Pomerleau, Avneesh Gupta, Jorge A Soto, and Stephan W Anderson. Accuracy of dual-energy ct virtual unenhanced and material-specific images: a phantom study. *American Journal of Roentgenology*, 215(5):1146–1154, 2020.

Anushri Parakh, Simon Lennartz, Chansik An, Prabhakar Rajiah, Benjamin M Yeh, Frank J Simeone, Dushyant V Sahani, and Avinash R Kambadakone. Dual-energy ct images: pearls and pitfalls. *Radiographics*, 41(1):98–119, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Sachiko Yamada, Takashi Ueguchi, Toshiyuki Ogata, Hirokazu Mizuno, Ryota Ogihara, Masahiko Koizumi, Takeshi Shimazu, Kenya Murase, and Kazuhiko Ogawa. Radiotherapy treatment planning with contrast-enhanced computed tomography: feasibility of dual-energy virtual unenhanced imaging for improved dose calculations. *Radiation oncology*, 9(1):168, 2014.