Automatic bed removal in small animal CT using deep learning based segmentation

*Jonathan Fisher, *Maitar Asher, Josep F. Oliver, and Craig Levin

***Abstract*—**

I. Introduction

X-ray CT provides a high spatial resolution 3D image of the scanned subject, offering a clear view of internal structures, such as bones, organs, and soft tissues. CT scans are valuable for identifying diseases or injuries in different body regions. For instance, CT scans are employed for screening and detecting potential tumors or lesions in the abdomen. A heart CT scan may be prescribed in cases of suspected heart disease or anomalies. CT head imaging is done to locate injuries, tumors, blood clots leading to stroke, hemorrhages, and other conditions. CT lung imaging is used to uncover tumors, pulmonary embolisms (blood clots), excess fluid, and conditions like emphysema or pneumonia. CT scans are also used for visualizing intricate bone fractures, severely deteriorated joints, or bone tumors. CT is also being used in multimodal imaging, such as PET/CT, to provide anatomical information and for PET attenuation and scatter corrections.

In many medical imaging applications, it is desirable and sometimes important to localize and subsequently remove the subject's bed from CT images:

1. For 3D visualization of CT or PET/CT, since the table can potentially obscure essential data.
2. In radiation treatment planning, including the patient table in images can significantly affect the accuracy of dose modeling. This is because the table used during imaging typically exhibits distinct attenuation properties compared to the table used during treatment [1]. In such instances, it becomes necessary to adjust the image by replacing the imaging table with the treatment table to ensure more precise planning.
3. In image registration, it is advantageous for the two registered images to possess similar image content. However, in modalities like PET, magnetic resonance imaging (MRI), or (single photon emission computed tomography (SPECT), where the patient table is not visible, the presence of a table in a CT image can have detrimental effects on the registration process when aligning CT with these other modalities. Even in the case of CT-CT (CHECK, typo?) registration, variations in table design, shape, and placement can introduce disparities that can negatively influence the registration procedure.
4. In emission-based attenuation correction (EBAC) through machine learning (ML) image-to-image translation [**?**], this approach focuses on estimating only the subject's pseudo-CT (pCT) while excluding the bed. To generate training set labels, the tables must be removed from the CT scans.

There is a limited body of literature addressing table removal. The available methods can be divided into manual, semi-automatic, or automatic.

In manual methods, the user is responsible for drawing a contour or manually marking the desired segment with software tools such as 3D Slicer [2]. Manual delineation is a time-intensive and labor-intensive process and suffers from a lack of reproducibility.

Semi-automatic techniques enable users to incorporate prior knowledge into the segmentation process [3], [4] and expedite it. Nevertheless, when applied to a whole-body CT volume comprising hundreds of slices, this approach may still involve labor-intensive and monotonous tasks.

Automatic algorithms can be based on one or several of the following techniques:

1. The template-based approach [5] acquires an extra bed-only scan and uses it as a template. This method didn't provide perfect results when applied to human data, leaving linen partially present and leading to the unintended removal of parts like fingers or ears. The approach fails to remove the restraint tubes used to secure the rat in place. Additionally, when used on small animals, the method will fail since the assumption that the bed remains in the exact same position doesn't apply (based on our dataset).
2. Thresholding, morphological filtering, and connected components analysis [6], [7]. These methods are founded on assumptions that do not hold in whole-body scans of small animals, such as the assumption that the subject and table do not make contact and that each slice contains only two connected components.
3. Using Hough transform on the vertical-edge-detected sagittal image to detect the bed boundary and remove it [8]. When applied to small animals, this method falls short in eliminating the nose cone, as well as the restraint tubes.

*These authors contributed equally to this work.

Jonathan Fisher is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA.

Maitar Asher is with the Department of Computer Science, Columbia University, New York City, NY 10027 USA.

Josep F. Oliver, is with Bruker BioSpin, Preclinical Imaging Division, Valencia, Spain. (e-mail: Jose.Oliver@bruker.com).

Craig S. Levin is with the Molecular Imaging Instrumentation Laboratory, Department of Radiology, Stanford University, Stanford, CA 94305 USA (e-mail: cslevin@stanford.edu)

In this study, we investigated a novel automatic method for bed removal in small animal CT. The solution applies semantic segmentation using 2D Unet models to accurately segment the rat and remove the bed from the CT scan.

II. Materials and Methods

A. Dataset

A total of twelve rats underwent PET/CT scanning using Bruker BioSpin's Albira Si system. The acquired data was reconstructed using Bruker's reconstruction software. This study exclusively utilized the CT images with a spatial resolution of 0.25x0.25x0.25 $[mm^3]$. The CT images served as the input data for the different models, while the models' targets were segmentation masks of the rats, which were manually created using 3D Slicer [2]. Before the manual segmentation process, the dimensions of the 3D volume were adjusted from 320x320x600 to 256x256x480. This adjustment was made to ensure that the dimensions of the transaxial slices are powers of 2, as required when using Unet models. Additionally, this modification aimed to reduce the time and labor required for the segmentation process.

Among these rats, eleven were scanned at one location using the same scanner, while the twelfth rat [9] was scanned at a different location but with an identical model scanner.

Initially, we focused on scans conducted with the same scanner and bed and used them to perform hyperparameter tuning, training, and testing. Subsequently, our solution was applied to the twelfth rat to evaluate our solution's generalizability across different scanners and beds.

The model inputs are single channel 2D slices in 32 float bit format.

B. Network architecture and cost function

Unet structures with the following backbones were used for the hyperparameter tunning "vgg19" "resnet152" "seresnet152" "resnext101" "seresnext101" "senet154" "densenet201" "inceptionresnetv2" "mobilenetv2" "efficientnetb7"

C. Network output processing

A recurring issue observed in the outputs of most, if not all, networks is that the generated mask, in addition to capturing the rat, tends to include some undesired artifacts in the background, as seen in Fig.1.A. To address this problem and eliminate these small artifacts, a post-processing step was introduced. For each rat, a 3D mask was generated by stacking the 2D slices of the network output. The post-processed 3D mask was defined as the largest connected component within the generated mask. Furthermore, the largest connected component often contained holes as can be seen in Fig.1.B., and to address this, the following steps were taken:

1. The 3D mask was inverted, rendering the inverted mask as representing the air surrounding the rat and the holes within the rat's mask.
2. The largest connected component was then selected from this inverted mask, which corresponds to the air surrounding the rat.
3. The outcome was inverted once more to obtain the rat's mask without any holes, ensuring a more accurate segmentation result.

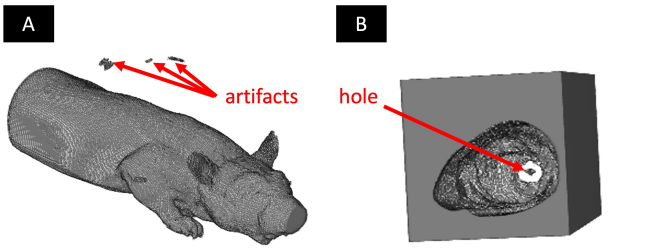

Fig. 1: A. The preprocessed 3D rat mask with small artifacts in the background. B. The inverse of the preprocessed 3D rat mask with a hole in it

D. Network performance analysis

To quantitatively analyze the different networks' performance, the following metrics were used:

1. Intersection over union (IoU): $\frac{|T \cap E|}{|T \cup E|}$
2. Dice similarity coefficient (DSC): $\frac{2 \cdot |T \cap E|}{|T| + |E|}$
3. False positive (FP): $\frac{|E| - |T \cap E|}{|T|}$
4. False negative (FN): $\frac{|T| - |T \cap E|}{|T|}$

Where T represents the target masks of the rats and E represents the estimated masks of the rats (model's output).

The performance was estimated on the raw and post-processed network's predictions.

E. Hyperparameter tuning and network testing

Two rats out of the eleven scanned at the same site were randomly chosen to be the validation set for hyperparameter tuning. The hyperparameters subject to tuning included the Unet structure/backbone and the cost function, as discussed in the network architecture subsection, as well as the batch size and the number of epochs. We employed batch sizes of 2, 4, 8, 16, 32, and 64 where feasible. It's worth noting that certain networks encountered memory issues when using larger batch sizes on our GPU, leading to their inability to run. For each combination of network, cost function, and batch size, we trained for 60 epochs. The optimal number of epochs was determined based on the minimal loss observed on the preprocessed validation set by the chosen cost function. The optimal network structure/backbone, cost function, and batch size appeared most frequently in the best-performing groups according to the metrics mentioned in the network performance analysis section (IoU, DSC, FP, FN measured on the processed networks' outputs). The best-performing group, according to an individual metric, includes those that performed as well as the best-performing hyperparameters set (not significantly worse than it).

To assess the solution's performance, cross-validation was conducted on a subset of eight rats randomly chosen from the remaining nine (excluding the validation set). In each cross-validation fold, a different pair of rats was designated as the testing set, while the remaining nine rats, including the validation set, were used for training.

To assess the generalizability of our solution, we tested the performance of the four trained models on the single rat scan taken at a different site using a different bed.

III. Results

A. Hyperparameter tuning

The hyperparameter tuning results are shown in Fig.2. Only the 40 best-performing networks according to each metric out of 102 combinations of Unet structure/backbone, cost function, and batch size are shown. The red line indicates the best-performing group criteria. 4 hyperparameter combinations were in 3 out of the 4 best-performing groups:

- Efficientnetb7 with BCE + Jaccard loss, batch size 4, trained for 19 epochs.
- Efficientnetb7 with BCE + Jaccard loss, batch size 2, trained for 23 epochs.
- Efficientnetb7 with Jaccard loss, batch size 8, trained for 10 epochs.
- senet154 with Jaccard loss, batch size 4, trained for 38 epochs.

Out of these 4 options, we chose to use the first.

B. Segmentation performance

C. generalization across different sites and beds

IV. Discussion

V. Conclusion and future work

Acknowledgment

Imaging data was acquired through the University of Virginia Molecular Imaging Core Lab, with NIH S10OD021672 funding for the Albira Si trimodal scanner, and through the University of Leeds, with ePIC-funding, British Heart Foundation, UK (SI/14/1/30718). The authors thank Dr. Garry Chinn for helpful discussions, Arianna Fozzato, Lauryn New, and Beatrice M Filippi for providing Leeds dataset. This work was supported by Bruker BioSpin.

References

[1] I. Mihaylov, P. Corry, Y. Yan, V. Ratanatharathorn, and E. Moros, "Modeling of carbon fiber couch attenuation properties with a commercial treatment planning system," *Medical physics*, vol. 35, no. 11, pp. 4982–4988, 2008.

[2] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[3] W. Zhou, Y. Xie *et al.*, "Interactive medical image segmentation using snake and multiscale curve editing," *Computational and mathematical methods in medicine*, vol. 2013, 2013.

[4] W. Zhou and Y. Xie, "Interactive contour delineation and refinement in treatment planning of image-guided radiation therapy," *Journal of Applied Clinical Medical Physics*, vol. 15, no. 1, pp. 141–166, 2014.

[5] J. Kim, Y. Hu, S. Eberl, D. Feng, and M. Fulham, "A fully automatic bed/linen segmentation for fused pet/ct mip rendering," 2008.

[6] L. Chen, S. Wu, Z. Zhang, S. Yu, Y. Xie, and H. Zhang, "Real-time patient table removal in ct images," in *Health Information Science: 5th International Conference, HIS 2016, Shanghai, China, November 5-7, 2016, Proceedings 5*. Springer, 2016, pp. 1–8.

[7] A. A. A. Rahni, M. F. M. Fuzaie, and O. I. Al Irr, "Automated bed detection and removal from abdominal ct images for automatic segmentation applications," in *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 2018, pp. 677–679.

[8] Y.-M. Zhu, S. M. Cochoff, and R. Sukalac, "Automatic patient table removal in ct images," *Journal of digital imaging*, vol. 25, no. 4, pp. 480–485, 2012.

[9] A. Fozzato, L. E. New, J. C. Griffiths, B. Patel, S. A. Deuchars, and B. M. Filippi, "Manipulating mitochondrial dynamics in the nts prevents diet-induced deficits in brown fat morphology and glucose uptake," *Life Sciences*, vol. 328, p. 121922, 2023.

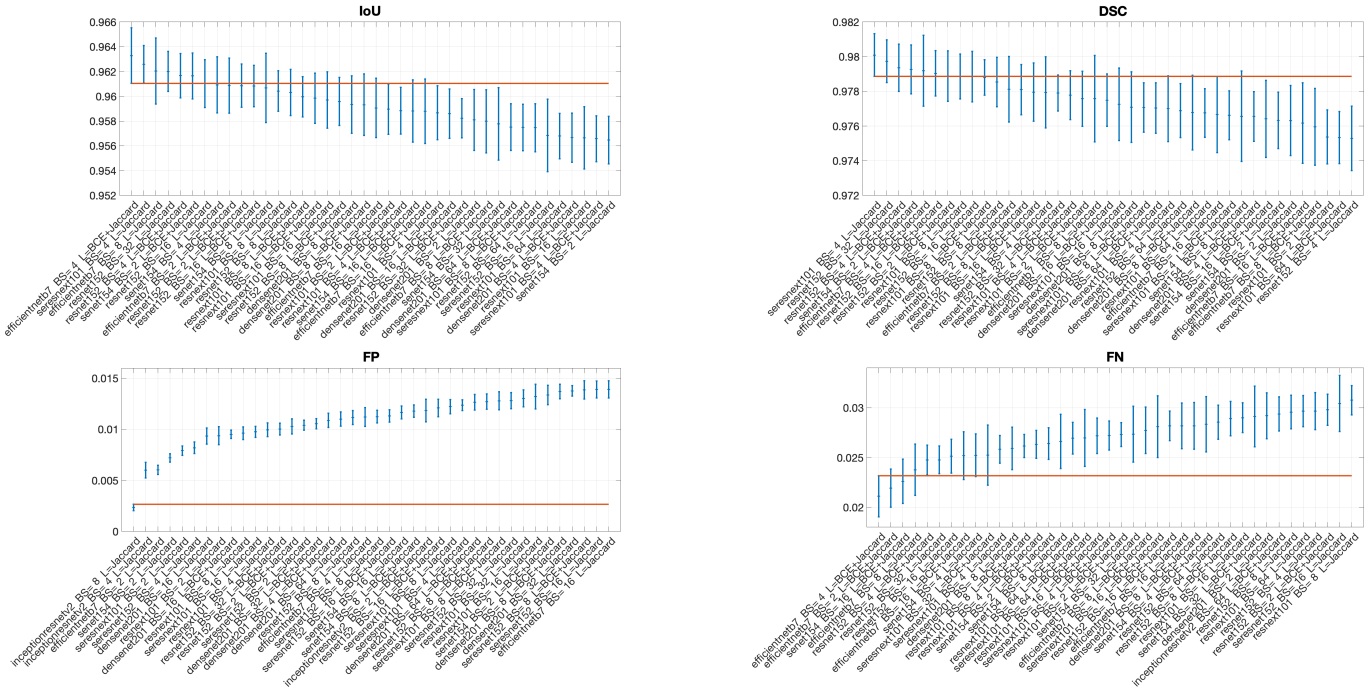

Fig. 2: need to write