Efficient Automatic Detection of Scanned Body Regions in CT Scans

Bertram Sabrowsky-Hirsch, Ahmed Alshenoudy, Stefan Thumfart
RISC Software GmbH, Hagenberg
{first.lastname}@risc-software.at

Thomas Potrusil, Johanna Gottwald, Mario Moser
CADS GmbH, Perg
{first.lastname}@cads.at

Abstract

Accurate identification of body regions in CT scans is critical for clinical and research workflows, yet DICOM metadata is often incomplete or unreliable. We propose a fast and lightweight method for automatic body region detection using 2D projections of CT volumes and a 2D U-Net segmentation model. Our method, trained on both public and proprietary anatomically labeled datasets, achieves high accuracy (median DSC of 0.97) and reduces inference time by more than 600x compared to full 3D segmentation. This enables efficient, scalable image screening and improves anatomical metadata consistency in large imaging archives.

1 Introduction

An essential step in many clinical workflows involving medical imaging databases is the accurate retrieval of relevant imaging data corresponding to indication-specific target anatomical regions. The *Digital Imaging and Communications in Medicine* (DICOM) standard defines a series of optional metadata tags, such as *Body Part Examined* and *Anatomic Region Sequence*, which can support the retrieval process. Additionally, datasets often include tabular metadata that may provide complementary anatomical information. However, this information is often inconsistent, ambiguous, incorrect, or removed entirely during de-identification [5, 7, 3]. Such issues significantly inhibit automated retrieval of relevant medical imaging data in both clinical and research settings. As a result, developing robust automated methods for extracting reliable labels has become increasingly important for effective image organization and data analysis.

Convolutional Neural Networks (CNNs) have demonstrated strong performance in classifying anatomical structures and image properties in CT scans [7, 8, 9], though they are less effective when multiple targets must be handled or precise localization is required. Grid-based classification, such as the method proposed by Hammami et al. [4], enables simultaneous localization and classification but often relies on slice-wise analysis and complex logic to make decisions at the volume level. Semantic segmentation methods, particularly those based on U-Net, have significantly advanced multi-organ segmentation [2, 6] and build the foundation for frameworks like nnDetection [1] and TotalSegmentator [11]. Nonetheless, the substantial computational time and resources required for full-volume segmentation limit their practicality for large-scale image screening.

To address these limitations, we propose a lightweight method for fast and accurate detection of body regions in CT scans. Our approach uses 2D projection images instead of full volumetric data, which significantly reduces model complexity, computation time and resource usage, while still achieving high segmentation accuracy suitable for reliably inferring and identifying imaged body regions.

The Second Austrian Symposium on AI and Vision (AIROV25).

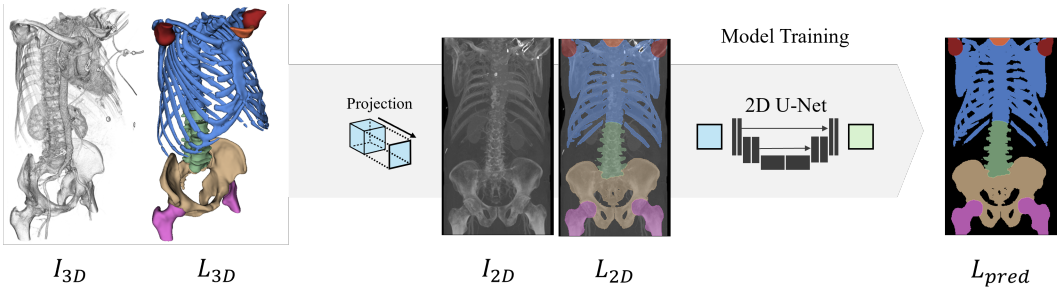

Figure 1: Our proposed approach, highlighting the generation of training data for body region detection. The CT volume I_{3D} and corresponding annotations L_{3D} are projected to 2D, which are then used to train a 2D U-Net model. During inference, $I_{3D} \longrightarrow I_{2D}$ is computed and fed into the trained model to obtain L_{pred}, which delineates the anatomical regions of interest.

2 Method and Preliminary Results

Our proposed method is illustrated in Figure 1. Starting with an input CT volume I_{3D} and the target annotation of body regions L_{3D}, we reduce the dimensionality of $I_{3D} \longrightarrow I_{2D}$ by computing the Maximum Intensity Projection (MIP) on the coronal plane. This provides a frontal view that clearly reveals structures extending along the body's length, making it the preferred choice for our detection task. Similarly, we project the ground-truth annotations $L_{3D} \longrightarrow L_{2D}$ on the coronal plane, resulting in a paired 2D dataset consisting of MIP images and ground-truth labels for different anatomical regions.This dataset is then used to train a 2D-UNet model using the nnUNet framework [6]. To enhance model generalization and extend our training dataset, we apply additional data augmentations prior to the 2D projection. First, we rotate I_{3D} around the longitudinal axis to make the model robust to varying patient orientations. These rotations are performed within the range of $[-40°, +40°]$ in $20°$ increments. We also crop the volumes along the Anterior-Posterior (AP) axis to simulate partial acquisitions, performed at 66% of both the AP and PA axes. In addition to the augmentations applied prior to the 2D projection of I_{3D}, the *DA5* trainer within the nnU-Net framework is configured to apply advanced augmentations to I_{2D} and L_{2D} during model training.

Our experiments were conducted using a combination of 1,221 subjects from the TotalSegmentator dataset [11] and a smaller in-house proprietary dataset of 66 subjects. Additional data was required due to the limited number of forearm and lower leg images in the TotalSegmentator dataset, which also lacked corresponding ground-truth annotations. To address this, we used the method MOOSE [10] to generate pseudo-labels for the forearms and lower legs in the TotalSegmentator dataset, as well as for 31 lower leg scans from our in-house dataset. The remaining 35 forearm CT scans in our in-house dataset were manually annotated. We categorized the body into six regions — head, thorax, abdomen, pelvis, arms, and legs — by grouping the relevant anatomical structures. Importantly, we limited our approach exclusively to bone structures, which are clearly visible in MIP images. The abdomen was identified by the lumbar vertebrae; the thorax included all cervical and thoracic vertebrae along with chest and shoulder bones; and the pelvis comprised the sacrum and hip bones. The arms and legs were defined as starting from the humerus and femur, respectively. In overlap areas, labels were assigned based on the following priority order, with later regions overriding earlier ones: thorax, head, pelvis, abdomen, arms, and legs.

Training and evaluation were performed on an NVIDIA RTX 4090 GPU using five-fold cross-validation. The median Dice Similarity Coefficient (DSC) across all labels was 0.97. A detailed breakdown of the performance for each individual body region is presented in Table 1 and Figure 2. Figure 3 presents examples of outliers with low Dice scores. Inference averaged 0.12 seconds per image, over 600 times faster than the 75-second average required by the 3D TotalSegmentator model on the same system.

We further evaluated the accuracy of the detected body regions from segmentation against the available metadata. The TotalSegmentator dataset provides a metadata table, where the columns *study-type* and *pathology-location* can be used to infer a set of body regions, denoted as R_P. Similarly, for our method, we derive R_P from the predicted segmentation L_{pred}. To filter out small fragments, we include only regions that cover an area of at least $\tau_A = 9\,\text{cm}^2 (3 \times 3\,\text{cm})$.

Region	Overall	TotalSegmentator	In-House
Head	0.98 [0.95–0.99]	0.98 [0.95–0.99]	–
Thorax	0.96 [0.95–0.97]	0.96 [0.95–0.97]	–
Abdomen	0.97 [0.95–0.98]	0.97 [0.95–0.98]	–
Pelvis	0.99 [0.98–0.99]	0.99 [0.98–0.99]	0.96 [0.95–0.96]
Leg	0.99 [0.99–0.99]	0.99 [0.99–0.99]	0.98 [0.98–0.98]
Arm	0.97 [0.93–0.98]	0.97 [0.93–0.98]	0.98 [0.96–0.98]
All	0.97 [0.95–0.99]	0.97 [0.95–0.99]	0.97 [0.96–0.98]

Table 1: Median and IQR of the DSC metric across the full dataset and the two subdatasets.

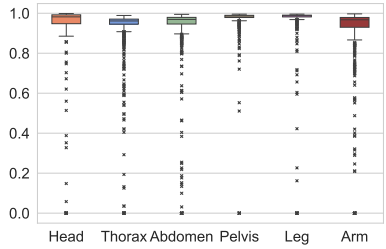

Figure 2: DSC value distribution by body region. Outliers mainly reflect fragmented structures (see Figure 3).

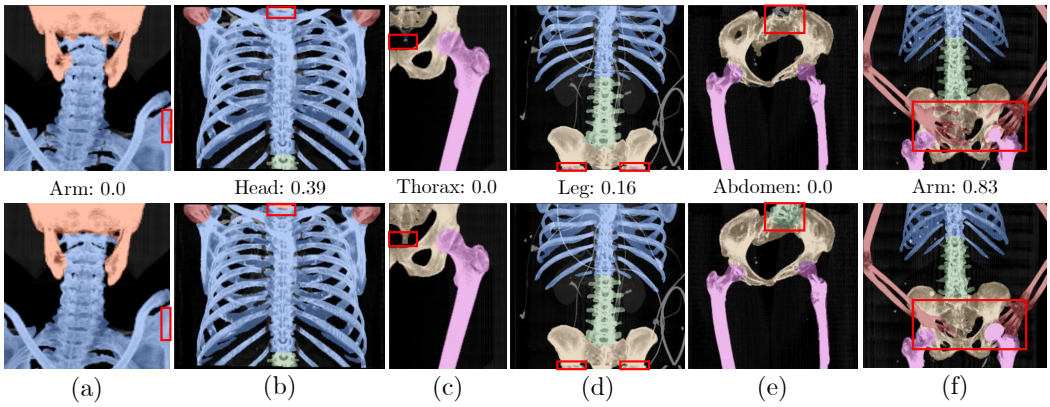

Figure 3: Examples of discrepancies between ground-truth (top) and predicted (bottom) segmentations that lead to low DSC scores (middle). Most errors are due to small ground-truth fragments, often near image boundaries, that the model fails to detect. These fragments disproportionately lower the DSC score, as shown in (a) to (d). Many fragments originate from minor annotation errors in TotalSegmentator, such as a sternum fragment mislabeled as thorax in (c). In some cases, ground-truth labels are missing entirely, like the lumbar vertebrae in (e). Only few cases reflect true segmentation challenges, such as the overlapping hands and pelvis in (f), which are underrepresented in the dataset.

Source	Available	Inferred	Contained	Equal
study-type	99.6%	91.5%	82.1%	11.2%
pathology-location	65.0%	51.4%	48.2%	4.8%
Ours	100.0%	100.0%	99.0%	98.0%

Table 2: Evaluation of the accuracy of inferred body regions R_P using two metadata sources compared to our method. We report the percentage of cases where metadata is present (Available) and successfully linked to body regions (Inferred). Additionally, we evaluate whether the inferred regions are contained within (Contained) or exactly match (Equal) the ground-truth body regions R_T.

Ground-truth body regions R_T are obtained in the same manner from the projected annotations L_{2D}. Table 2 summarizes the percentage of cases in which metadata is available and can be linked to body regions ($R_P \neq \emptyset$), and where the inferred regions are either contained within ($R_P \subseteq R_T$) or exactly match ($R_P = R_T$) the ground-truth regions. The study-type was available in nearly all cases, however, 8.1% of entries contained ambiguous values, such as *intervention*. Pathology-location, which was already available in only 65% of cases, included the ambiguous values *unclear* or *bones* in 21% of those. Among the available data, the inferred labels R_P were incorrect in 10.3% and 6.1% of cases for study-type and pathology-location, respectively, as they did not match any values in R_T. A complete match was achieved in only 11.2% and 4.8% of cases, respectively. Figure 4 shows cases where the body region has been incorrectly inferred ($R_P \not\subseteq R_T$) based on the study-type

(top), pathology-location (middle), or our method (bottom). We found that the study-type attribute was often labeled too broadly, including multiple regions not present in the image. In contrast, the pathology-location attribute was frequently outright incorrect. For our model, only one genuine error was observed: a rib was misclassified as an arm (bottom right image). The remaining cases did contain the target label, although only at levels slightly below the minimum area threshold τ_A.

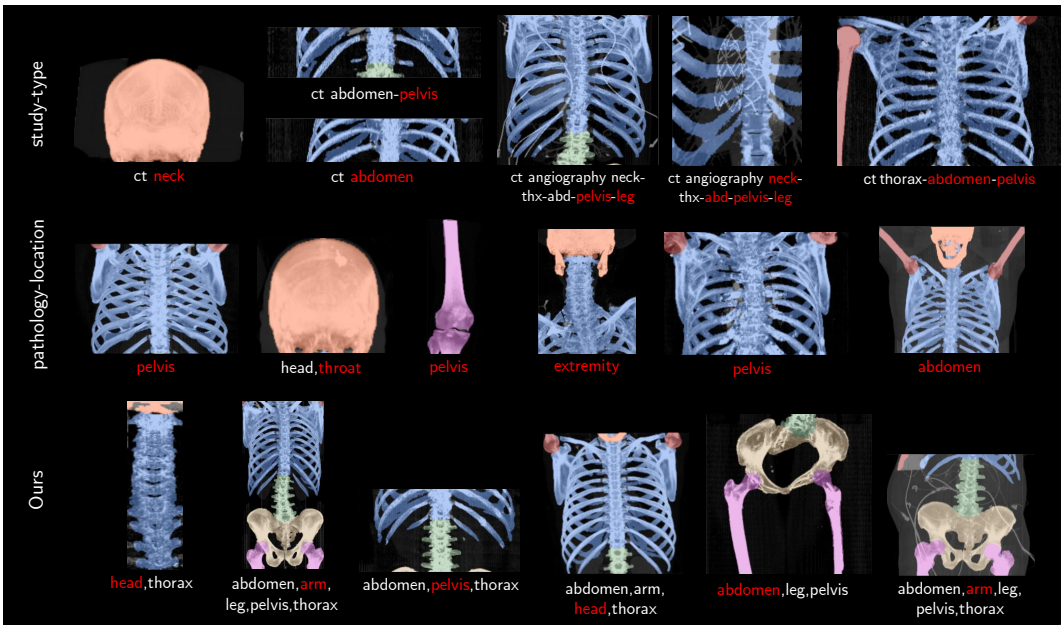

Figure 4: Instances of incorrectly inferred regions ($R_P \not\subseteq R_T$) are shown for the study-type (top), pathology-location (middle) and our method (bottom). Ground-truth segmentations are labeled with the original metadata or segmented body regions, with mismatched regions highlighted in red.

3 Discussion and Conclusion

To enable fast and accurate body region detection in CT scans, our method addresses the high computational demands of 3D segmentation. By using coronal projections to reduce image dimensionality, we achieve a substantial reduction in inference time. With an average inference time of just 0.1 seconds per image, our approach is well suited for large-scale screening of medical databases. For the TotalSegmentator dataset, we demonstrated that our method predicted body regions far more reliably than the provided metadata. Our approach offers the added benefit of producing segmented regions, which facilitates result visualization and quantification of region sizes, thereby enhancing explainability and enabling more advanced tasks beyond region detection.

The selection and grouping of anatomical regions were tailored to our specific use case. However, segmenting finer-grained structures may further improve model accuracy and generalizability by enabling the model to distinguish a broader range of anatomical features. In this case, grouping into broader body regions could be deferred to a postprocessing step. Expanding segmentation to include non-skeletal structures, such as muscles and organs, may enhance robustness, particularly in the abdominal region where the model is currently limited to the lumbar vertebrae and could benefit from incorporating larger structures like the colon. Soft-tissue structures may also benefit from alternative projection techniques, such as Average Intensity Projection (AIP) or Digitally Reconstructed Radiographs (DRR), which could serve as additional input channels. Future improvements should consider multi-class training to accommodate overlapping structures, allowing the model to assign multiple anatomical labels within the same spatial region. The impact of these extensions should be carefully evaluated, as they also increase model complexity.

We are currently evaluating these proposed extensions to develop a 2D counterpart to TotalSegmentator, with body region detection integrated as a secondary processing step. As part of this effort, we also plan to release models trained exclusively on public datasets.

References

[1] M. Baumgartner, P. F. Jäger, F. Isensee, and K. H. Maier-Hein. nndetection: a self-configuring method for medical object detection. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pages 530–539. Springer, 2021.

[2] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang. A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica*, 85:107–122, 2021.

[3] M. O. Gueld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann. Quality of DICOM header information for image categorization. In E. L. Siegel and H. K. Huang, editors, *Medical Imaging 2002: PACS and Integrated Medical Information Systems: Design and Evaluation*, volume 4685, pages 280 – 287. International Society for Optics and Photonics, SPIE, 2002.

[4] M. Hammami, D. Friboulet, and R. Kechichian. Cycle gan-based data augmentation for multi-organ detection in ct images via yolo. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 390–393, 2020.

[5] A. Iancu, J. Bauer, M. S. May, H.-U. Prokosch, A. Dörfler, M. Uder, and L. A. Kapsner. Large-scale integration of dicom metadata into hl7-fhir for medical research. *Methods of Information in Medicine*, 63(03/04):077–084, 2024.

[6] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[7] F. Jonske, M. Dederichs, M.-S. Kim, J. Keyl, J. Egger, L. Umutlu, M. Forsting, F. Nensa, and J. Kleesiek. Deep learning–driven classification of external dicom studies for pacs archiving. *European radiology*, 32(12):8769–8776, 2022.

[8] W. Li, H. M. Lin, A. Lin, M. Napoleone, R. Moreland, A. Murari, M. Stepanov, E. Ivanov, A. S. Prasad, G. Shih, et al. Machine learning classification of body part, imaging axis, and intravenous contrast enhancement on ct imaging. *Canadian Association of Radiologists Journal*, 75(1):82–91, 2024.

[9] P. Raffy, J.-F. Pambrun, A. Kumar, D. Dubois, J. W. Patti, R. A. Cairns, and R. Young. Deep learning body region classification of mri and ct examinations. *Journal of Digital Imaging*, 36(4):1291–1301, 2023.

[10] L. K. S. Sundar, J. Yu, O. Muzik, O. C. Kulterer, B. Fueger, D. Kifjak, T. Nakuz, H. M. Shin, A. K. Sima, D. Kitzmantl, et al. Fully automated, semantic segmentation of whole-body 18f-fdg pet/ct images based on data-centric artificial intelligence. *Journal of Nuclear Medicine*, 63(12):1941–1948, 2022.

[11] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, et al. Totalsegmentator: robust segmentation of 104 anatomic structures in ct images. *Radiology: Artificial Intelligence*, 5(5), 2023.