On TotalSegmentator's Performance on Low-Dose CT Images

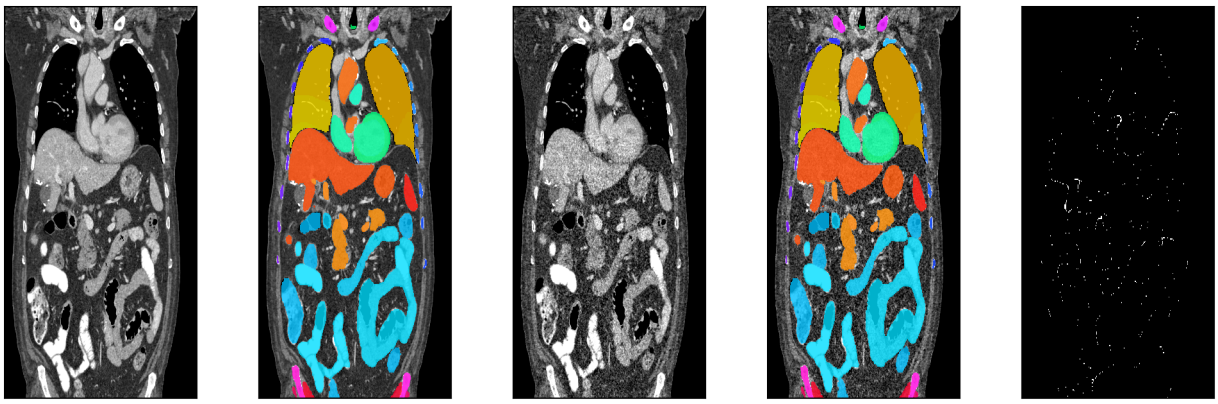
A. Tsanda[a, b], H. Nickisch[a], T. Wissel[a], T. Klinder[a], T. Knopp[b, c], and M. Grass[a]

[a]Philips Research, Hamburg, Germany
[b]Institute for Biomedical Imaging, Hamburg University of Technology, Hamburg, Germany
[c]Section for Biomedical Imaging, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

ABSTRACT

Deep neural networks have emerged as the preferred method for semantic segmentation of CT images in recent years. However, understanding their limitations and generalization properties remains an active area of research and a relevant topic for clinical applications. One crucial factor among many is the X-ray radiation dose, which is always kept as low as reasonably possible during CT acquisition. Therefore, potential dose reductions may pose a challenge for existing segmentation models. In this paper, we investigate robustness of the recently proposed TotalSegmentator model for anatomical segmentation with respect to dose reduction. TotalSegmentator combines a large CT dataset and the well-established nnU-Net framework to train deep learning models, resulting in state-of-the-art performance for anatomical segmentation. Our method relies on accurate low-dose simulations derived from acquired full-dose projections. For a set of registered low- and full-dose CT images, we measure the Dice score between the corresponding segmentations. Our results reveal a high level of robustness in the segmentation outcomes. Comprehensive quantitative comparisons demonstrate that at a 20% dose level, the Dice score declines by at most 3%. Visual comparisons reveal only minor differences at the boundaries of the segmented organs. These findings may have a large potential for dose reduction when CT data are acquired predominantly for segmentation purposes, such as for the planning of interventional or surgical procedures.

Keywords: Low-dose CT, semantic segmentation, deep learning



(a) A CT image at 100% dose level. (b) Segmentation results at 100% dose level (c) A CT image at 20% dose level. (d) Segmentation results at 20% dose level (e) Differences between the segmentation results

Figure 1. An example of a low-dose CT image (c) simulated from a full-dose image (a) along with the corresponding segmentation results (d), (b). The Dice score between (d) and (b) equals 0.994. The difference image (e) shows that the segmentation results are robust against dose reduction down to 20%.

1. INTRODUCTION

Semantic segmentation plays a key role in the computer-aided diagnosis based on CT images, helping to automate the routine annotation process required for the quantitative analysis, which is an essential measure for a

reliable diagnosis. In recent years, methods based on deep neural networks have surpassed the state-of-the-art in various visual applications, including segmentation. Among different neural network architectures, U-Net[1] has emerged as one of the most widely used for semantic segmentation in all medical image modalities. However, the neural architecture alone is not the sole contributor to a successful solution; proper data pre-processing also plays a significant role. The nnU-Net framework[2] combines these two factors, facilitating better segmentation performance. Another integral part is a large amount of annotated medical data which often remains a limiting factor in developing high-quality segmentation models. Addressing this limitation, the TotalSegmentator approach[3] introduces a large dataset for human organ segmentation in CT images along with trained nnU-Net models.

As CT acquisition tends to use lower doses of X-ray radiation whenever possible, it becomes crucial to understand the robustness of existing segmentation models with respect to dose reduction. Studies in this direction face a trade-off between using real and simulated low-dose CT data. Real data typically leads to unpaired datasets of low- and full-dose images, while simulated data raises concerns about the applicability of the results to real-world scenarios due to introduced simplifications. Hooper *et al.*[4] investigated robustness of a 3D classification network trained to triage head CT data. The authors reprojected existing CT volumes to simulate lower tube currents by adding Gaussian white noise in projections. Although the simulation of acquired projections may introduce additional errors, and the physics of CT acquisition implies the Poisson noise model, the authors reported a high level of generalization to lower tube currents. Aiello *et al.*[5] examined the applicability of existing deep learning models for lungs and COVID-19 lesions segmentation, trained on full-dose CT scans, to low-dose scenarios. In the absence of registered image pairs, the authors compared the results of radiomics and reported their stability.

In this paper, we propose a method to accurately assess the dose robustness of the TotalSegmentator approach. To achieve this, we extensively utilize raw CT data, such as projections and tube currents, to simulate realistic low-dose acquisitions. Our method generates pairs of co-registered low- and normal-dose CT data, enabling an analysis of robustness at the level of segmentation masks.

2. MATERIALS AND METHODS

Our method relies on accurate simulations of low dose acquisitions. To achieve this, we conduct simulations using raw projection data, where the noise model is known. Due to the nature of the radiation process, the number of photons acquired by the detector follows the Poisson distribution. We employ a method proposed by Žabić *et al.*[6] to sample new projection values with lower dose levels. Specifically, given the known number of acquired photons n^{α} corresponding to the current α, a new number of photons $n^{\alpha\to\beta}$ for a lower current β follows a Poisson distribution:

$$n^{\alpha\to\beta} \sim \frac{\alpha-\beta}{\alpha}\mathcal{P}\left(\frac{\beta}{\alpha-\beta}n^{\alpha}\right). \tag{1}$$

After sampling photon counts for lower dose levels, we reconstruct 3D images using the Aperture Weighted Wedge Filtered Backprojection algorithm[7] from both the initial raw and the simulated projections resulting in a pair of low- and full- dose images. We employ 5 dose levels $\beta/\alpha = 20\%, 40\%, 60\%, 80\%, 100\%$. Due to the considered range of tube currents, we neglect electronic noise in CT detectors.

In our experiments, we used data acquired from the Philips Spectral CT 7500 System (Philips Medical Systems, Cleveland, USA). The data represents raw projections of the abdominal area in helical geometry with 8 cm collimation and pitch factor 1.38. The data is reconstructed using 420 mm field of view with pixel numbers along X and Y set to 512. The slice thickness is 1 mm with a slice increment of 0.5 mm. We utilized data from 42 patients, comprising a total of 99 scans. Approximately half of the CT studies include contrast in different phases of the injection as well as both healthy patients and patients with pathologies.

The data is then passed to the TotalSegmentator model. Since it consists of five models targeting different organ groups (*totsegm-organs*, *totsegm-vertebrae*, *totsegm-cardiac*, *totsegm-muscles*, and *totsegm-ribs*), we consider them both as a whole and individually. The model's evaluation code was used as provided by the authors' repository*. For inference, we disable test time augmentations and ensembling, taking only the model trained on

*https://github.com/wasserth/TotalSegmentator

the first cross-validation split. We compare each predicted label from low-dose segmentations against full-dose predictions using the Dice similarity coefficient.

3. RESULTS

Table 1 shows the Dice scores calculated between segmentations of low- and full-dose CT images at different dose levels. The median and the median absolute deviation aggregate the results across multiple CT images as some scores are falling into the extreme of the value range. While the segmentation results for lower dose levels gradually deviate from those of the full-dose images as the dose level decreases, the absolute decline does not exceed 3%, even at the 20% dose level.

Figure 1 illustrates an example of segmentation results at 20% and 100% dose levels. The difference image (Figure 1(e)) provides additional support for the quantitative results. It is noteworthy that even for smaller body parts (e.g., the gallbladder) the segmentation results demonstrate a high degree of robustness.

Table 1. Dice scores calculated between segmentations of full- and low-dose CT images for different dose levels. Since the full-dose segmentation was used as a reference, the scores for the 100% dose level equal 1.0.

Model Name	Dose Level				
	20%	40%	60%	80%	100%
totsegm-all	0.983 ± 0.008	0.990 ± 0.005	0.993 ± 0.003	0.996 ± 0.002	1.000 ± 0.000
totsegm-organs	0.987 ± 0.011	0.992 ± 0.007	0.995 ± 0.004	0.997 ± 0.003	1.000 ± 0.000
totsegm-vertebrae	0.988 ± 0.002	0.993 ± 0.002	0.995 ± 0.001	0.997 ± 0.001	1.000 ± 0.000
totsegm-cardiac	0.982 ± 0.010	0.989 ± 0.006	0.993 ± 0.004	0.996 ± 0.002	1.000 ± 0.000
totsegm-muscles	0.986 ± 0.004	0.992 ± 0.003	0.995 ± 0.002	0.997 ± 0.001	1.000 ± 0.000
totsegm-ribs	0.974 ± 0.006	0.984 ± 0.004	0.990 ± 0.002	0.994 ± 0.002	1.000 ± 0.000

4. CONCLUSION

With the increasing integration of deep neural networks into clinical practice, it becomes crucial to understand their limitations. In this paper, we analyzed the robustness of the state-of-the-art TotalSegmentator model for anatomical segmentation with respect to lower doses in CT imaging. By working with CT projections, we could accurately simulate low-dose CT data that were intrinsically registered with the original CT images. Our approach enabled clinically relevant analysis of the model at the segmentation level. Although the nnU-Net framework incorporates the Gaussian noise augmentation and the data used to train the TotalSegmentator model contains a small portion of low-dose images, the results of our study reveal surprising robustness of the TotalSegmentator approach. Even when reducing the dose level down to 20%, outputs of the model showed minimal differences at the pixel level. The remarkable robustness of the model opens up possibilities for further dose reduction and expands the application of existing models to new low-dose data.

5. ACKNOWLEDGMENT

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1615 – 503850735.

REFERENCES

[1] Ronneberger, O., Fischer, P., and Brox, T., "U-Net: Convolutional networks for biomedical image segmentation," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*], 234–241, Springer (2015).

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

[3] Wasserthal, J., Meyer, M., Breit, H.-C., Cyriac, J., Yang, S., and Segeroth, M., "TotalSegmentator: robust segmentation of 104 anatomical structures in CT images," *arXiv preprint arXiv:2208.05868* (2022).

[4] Hooper, S. M., Dunnmon, J. A., Lungren, M. P., Gambhir, S. S., Ré, C., Wang, A. S., and Patel, B. N., "Assessing robustness to noise: Low-cost head CT triage," *CoRR* **abs/2003.07977** (2020).

[5] Aiello, M., Baldi, D., Esposito, G., Valentino, M., Randon, M., Salvatore, M., and Cavaliere, C., "Evaluation of AI-Based Segmentation Tools for COVID-19 Lung Lesions on Conventional and Ultra-low Dose CT Scans," *Dose-Response* **20**, 155932582210828 (Apr. 2022).

[6] Žabić, S., Wang, Q., Morton, T., and Brown, K. M., "A low dose simulation tool for CT systems with energy integrating detectors," *Medical Physics* **40**, 031102 (Feb. 2013).

[7] Brown, K. M., Heuscher, D. J., and Kling, P., "Conebeam computed tomography imaging," (2004). US Patent 6,775,346.