A review of TotalSegmentator

Anton Zhitomirsky

January 2, 2024

A review of the paper [2] with github link [1]

1 Review of Paper

- Model used: nnU-Net segmentation algorithm

- Performance measure: Dice similarity coefficient

- Trained on 1204 CT examinations

- Trained once more on 4004 whole-body CT examinations to investigate age dependent volume and attenuation changes.

- segments most anatomically relevant structures throughout the body

1.1 Datasets

1.1.1 Trianing dataset

1368 CT images randomly sampled with $(37 + 87 + 40)$ samples excluded due to missing sections or human annotation missing were exluded leading to $(1368 - 164)$ 1204.

- Training set: 90% lead to 1082 patients

- Validation dataset: 5% lead to 57 patients

- Test dataset: 5% lead to 65 patients

1.1.2 Aging-study dataset

This dataset includes patients with polytrauma[1] who received whole-body CT scans. 4102 CT images randomly sampled with $(30 + 33 + 35)$ patients were excluded leaving $(4102 - 98)$ 4004.
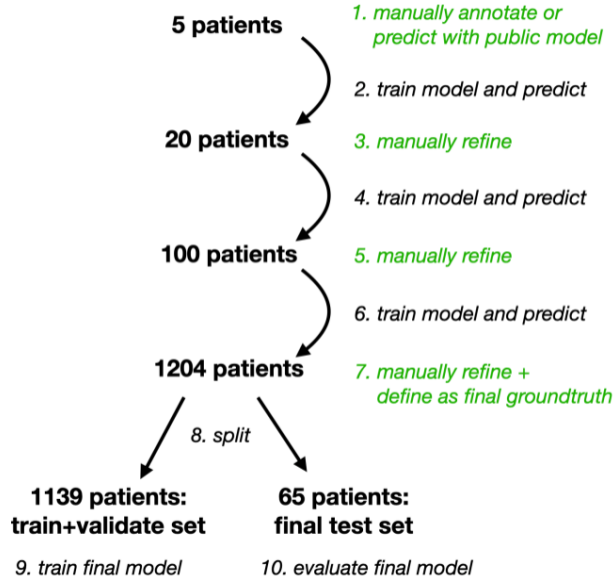
1.2 Annotation workflow

Iterative learning approach was used.

---

[1]Polytrauma and multiple trauma are medical terms describing the condition of a person who has been subjected to multiple traumatic injuries, such as a serious head injury in addition to a serious burn.

Annotation workflow

5 patients
1. manually annotate or predict with public model
2. train model and predict

20 patients
3. manually refine
4. train model and predict

100 patients
5. manually refine
6. train model and predict

1204 patients
7. manually refine + define as final groundtruth
8. split

1139 patients: train+validate set
9. train final model

65 patients: final test set
10. evaluate final model

(1) After manual segmentation of the first 5 patients was completed, (2) a preliminary nnU-Net was trained, (3) and its predictions were manually refined, if necessary. (4) Retraining of the nnU-Net was performed after (5) reviewing and refining 5 patients, 20 patients, and (6) 100 patients.

In the end, all 1204 CT examinations had annotations that were manually reviewed and corrected whenever necessary. These final annotations served as the ground truth for training and testing. The model was trained on the dataset of 1082 patients, validated on the dataset of 57 patients and tested on the dataset of 65 patients. This final model was independent of the intermediate models trained during the annotation workflow, which reduced bias in the test set to a minimum. Using completely manual annotations in the test set would have introduced a distribution shift and thus greater bias.

1.3 Model

The nnU-Net was used because of its ability to automatically configure hyperparameters based on the dataset characteristics.

1.4 Problems they had

1.4.1 Ribs

Some patients had ribs missing, which a clinitian would typically count from top/bottom to identify the individual ribs - However, on some scans only a subset is visible. This can extend to our case as some have missing kidneys or large abnormalities that aren't common to other cases.

1.4.2 Varying organ properties

Texture, position and size relative to the body was found to cause problems, specifically 'The colon's shape, size, texture and position in the abdomen can vary greatly and is sometimes difficult to distinguish from the small bowel'

2 GitHub

References

[1] wasserth et al. URL: https://github.com/wasserth/TotalSegmentator.

[2] Jakob Wasserthal et al. "TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images". In: (2023). URL: https://arxiv.org/pdf/2208.05868.pdf.