JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Tutorial: Large-Scale Medical Segmentation Benchmarking (1st Edition)

Dear Colleagues,

Thanks for your contribution to our Touchstone benchmark.

Our annotated dataset is attached below.

- Data & annotations (328.35G): Dropbox (option #1, option #2), Baidu Wangpan, Huggingface (option #1, option #2)
- Script (*feel free to modify or completely rewrite*): GitHub

Although there is no hard deadline, we hope you can start training at your earliest convenience. Considering the great size of the dataset, it might take **a couple of weeks** for training. Once you finish training, please send us your training & testing scripts and checkpoint(s). Please refer to the **Appendix** for detailed instructions of download and submission.

My colleagues, Pedro Bassi <pedro.salvadorbassi2@unibo.it> and Wenxuan Li <wli131@jh.edu>, are also here to assist you during the benchmark. Please let us know if you have any questions, we are very happy to help, and Zoom with you if needed.

Lastly, you have my permission to use this dataset if it is helpful for your upcoming submissions (e.g., ICLR, AAAI, CVPR, etc.). The related papers are

> *Qu, Chongyu, Tiezheng Zhang, Hualin Qiao, Yucheng Tang, Alan L. Yuille, and Zongwei Zhou. "Abdomenatlas-8k: Annotating 8,000 CT volumes for multi-organ segmentation in three weeks." Advances in Neural Information Processing Systems 36 (2023). https://www.cs.jhu.edu/~alanlab/Pubs23/qu2023abdomenatlas.pdf*

> *Li, Wenxuan, Alan Yuille, and Zongwei Zhou. "How well do supervised models transfer to 3d image segmentation?" In The Twelfth International Conference on Learning Representations. 2024. https://www.cs.jhu.edu/~alanlab/Pubs23/li2023suprem.pdf*

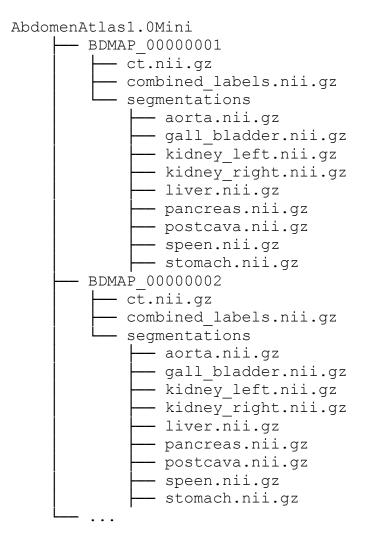For any questions, please contact Zongwei Zhou at zzhou82@jh.edu

Thank you.

Appendix

A.1 Training Data Structure

For superior stability, use wget to download from Dropbox:

```
wget
https://www.dropbox.com/scl/fi/bqk780hn794gep6upfrge/AbdomenAtla
s1.0Mini.tar.gz?rlkey=gjiiz731us18p9u7iouhin3gw&dl=0
```

We provide the following statistics to help you verify that you have successfully downloaded all the data (~328GB). AbdomenAtlas1.0Mini has **5,195** subjects (BDMAP_00000001 – BDMAP_00005195), totaling **57,145** nii.gz files (5,195 x 11); each subject is structured as:

```
AbdomenAtlas1.0Mini
        ├── BDMAP_00000001
        │       ├── ct.nii.gz
        │       ├── combined_labels.nii.gz
        │       └── segmentations
        │               ├── aorta.nii.gz
        │               ├── gall_bladder.nii.gz
        │               ├── kidney_left.nii.gz
        │               ├── kidney_right.nii.gz
        │               ├── liver.nii.gz
        │               ├── pancreas.nii.gz
        │               ├── postcava.nii.gz
        │               ├── speen.nii.gz
        │               ├── stomach.nii.gz
        ├── BDMAP_00000002
        │       ├── ct.nii.gz
        │       ├── combined_labels.nii.gz
        │       └── segmentations
        │               ├── aorta.nii.gz
        │               ├── gall_bladder.nii.gz
        │               ├── kidney_left.nii.gz
        │               ├── kidney_right.nii.gz
        │               ├── liver.nii.gz
        │               ├── pancreas.nii.gz
        │               ├── postcava.nii.gz
        │               ├── speen.nii.gz
        │               ├── stomach.nii.gz
        └── ...
```
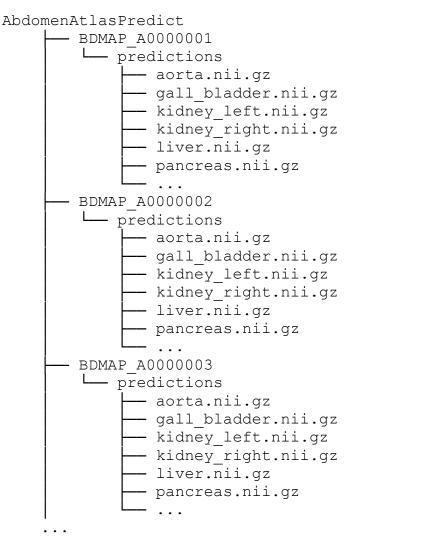
A.2 Test Script Preparation

The JHU team will evaluate your checkpoint(s) using a dataset of over 5,000 fully annotated CT volumes. This dataset will NOT be available to the public and has a very different demographic profile from that of the AbdomenAtlas1.1Mini dataset.

Our test dataset is structured as follows:

```
AbdomenAtlasTest
    ├── BDMAP_A0000001
    │   └── ct.nii.gz
    ├── BDMAP_A0000002
    │   └── ct.nii.gz
    ├── BDMAP_A0000003
    │   └── ct.nii.gz
    ...
```

Please make sure that your test script can (1) read the test dataset structure and (2) generate predictions structured as follows:

```
AbdomenAtlasPredict
    ├── BDMAP_A0000001
    │   └── predictions
    │       ├── aorta.nii.gz
    │       ├── gall_bladder.nii.gz
    │       ├── kidney_left.nii.gz
    │       ├── kidney_right.nii.gz
    │       ├── liver.nii.gz
    │       ├── pancreas.nii.gz
    │       └── ...
    ├── BDMAP_A0000002
    │   └── predictions
    │       ├── aorta.nii.gz
    │       ├── gall_bladder.nii.gz
    │       ├── kidney_left.nii.gz
    │       ├── kidney_right.nii.gz
    │       ├── liver.nii.gz
    │       ├── pancreas.nii.gz
    │       └── ...
    ├── BDMAP_A0000003
    │   └── predictions
    │       ├── aorta.nii.gz
    │       ├── gall_bladder.nii.gz
    │       ├── kidney_left.nii.gz
    │       ├── kidney_right.nii.gz
    │       ├── liver.nii.gz
    │       ├── pancreas.nii.gz
    │       └── ...
    ...
```

Before sending us the training, test scripts and checkpoint(s), please make sure that your code can successfully generate 9-class predictions using the test examples [here](#). We kindly ask for a README file with installation and running instructions, including a complete list of all your code's dependencies and their versions (e.g., Python 3.8, PyTorch 1.11.0, ...).