ChestX-Seg40: A Chest X-ray Dataset for Thoracic Multi-Organ Segmentation

Xiongfeng Zhu, Qianjin Feng

School of Biomedical Engineering, Southern Medical University, Guangzhou, 510515, China

1. Introduction

Thoracic multi-organ segmentation on Chest X-rays (CXR) has many important clinical applications, such as organ quantification, surgical planning, and disease diagnosis. Most existing public datasets for chest X-ray segmentation focus on lungs and hearts, while ignoring other organs. Hence, we introduce a new benchmark dataset called ChestX-Seg40 for thoracic multi-organ segmentation.

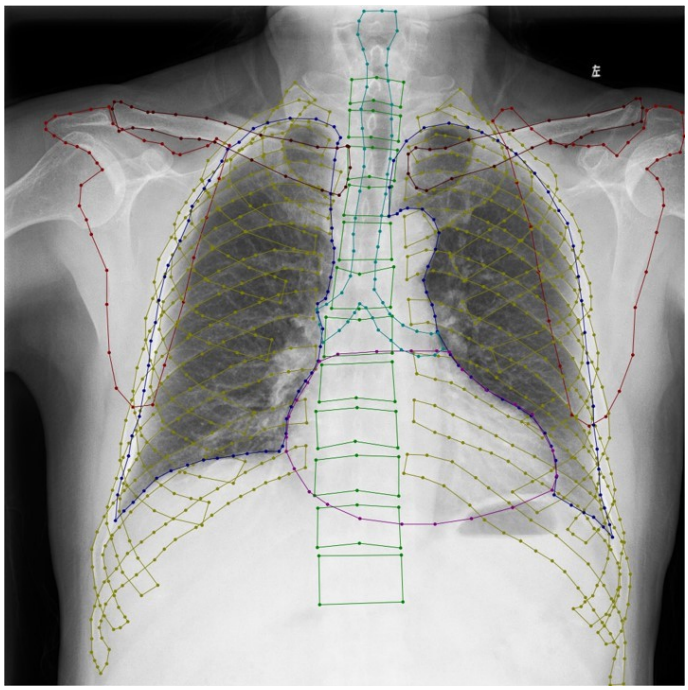

Figure 1. An example of the annotated masks in ChestX-Seg40.

2. Dataset

ChestX-Seg40 was retrospectively collected from Guangdong Provincial People's Hospital. It contains 600 CXR images with corresponding ground truth masks annotated by human experts. In ChestX-Seg40, images are directly extracted from the DICOM files and de-identified to protect patient privacy. Images are resized as 512×512 pixels and saved as npy files. The targets include 40 organs: 20 ribs, 12 thoracic vertebrae, 2 scapula, 2 clavicles, 2 lungs, 1 heart, and 1 trachea. The masks of organs were stored in a CSV file. Figure 1 shows an example of the annotated masks.

3. Evaluation

To develop and evaluate segmentation algorithms, we divided the whole dataset into a training set of 500 images and a validation set of 100 images. Intersection over union (IoU) score was used as the quantitative metric to evaluate the segmentation

performance. IoU score was calculated for an individual organ, and then averaged across all organs.

4. Download

ChestX-Seg40 was released at: https://github.com/fzfs/ChestX-Seg40.