SCR database: Data description

All chest radiographs are taken from the JSRT database. This is a publicly available database with 247 PA chest radiographs collected from 13 institutions in Japan and one in the United States. The images were scanned from films to a size of 2048 by 2048 pixels, a spatial resolution of .175 mm/pixel and 12 bit gray levels. 154 images contain exactly one pulmonary lung nodule each; the other 93 images contain no lung nodules. The JSRT database was established by the Japanese Society of Radiological Technology (JSRT). This data can be obtained on CD from the JSRT website. The JSRT database has been used in many studies on nodule detection on chest radiographs. More details about the database can be found in the following publication:

- J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules", *American Journal of Roentgenology*, vol. 174, p. 71-74, 2000.

Each object has been delineated by clicking points along its boundary using a mouse pointer device. These points are connected by straight line segments. Two observers segmented 5 objects in each image: both lungs, the heart and both clavicles. The segmentations of the first observer are taken as reference standard. The independent segmentations from the second observer can be compared to segmentations by a computer algorithm to allow for comparisons between `human' and `computer' results.

Note that, by convention, a chest radiograph is displayed as if one is facing the patient. This means that the right lung and clavicle are on the left in the image.

The following definition for the lung fields is adopted: any pixel for which radiation passed through the lung, but not through the mediastinum, the heart, structures below the diaphragm, and the aorta. The vena cava superior, when visible, is not considered to be part of the mediastinum.

The heart is defined as those pixels for which radiation passes through the heart. From anatomical knowledge the heart border at the central top and bottom part can be drawn. The hilar vessels can be assumed to lie on top of the heart.

For the clavicles, only those parts superimposed on the lungs and the rib cage have been indicated (as the peripheral parts of the clavicles are not always visible on a chest radiograph).

An example segmentation is shown below.

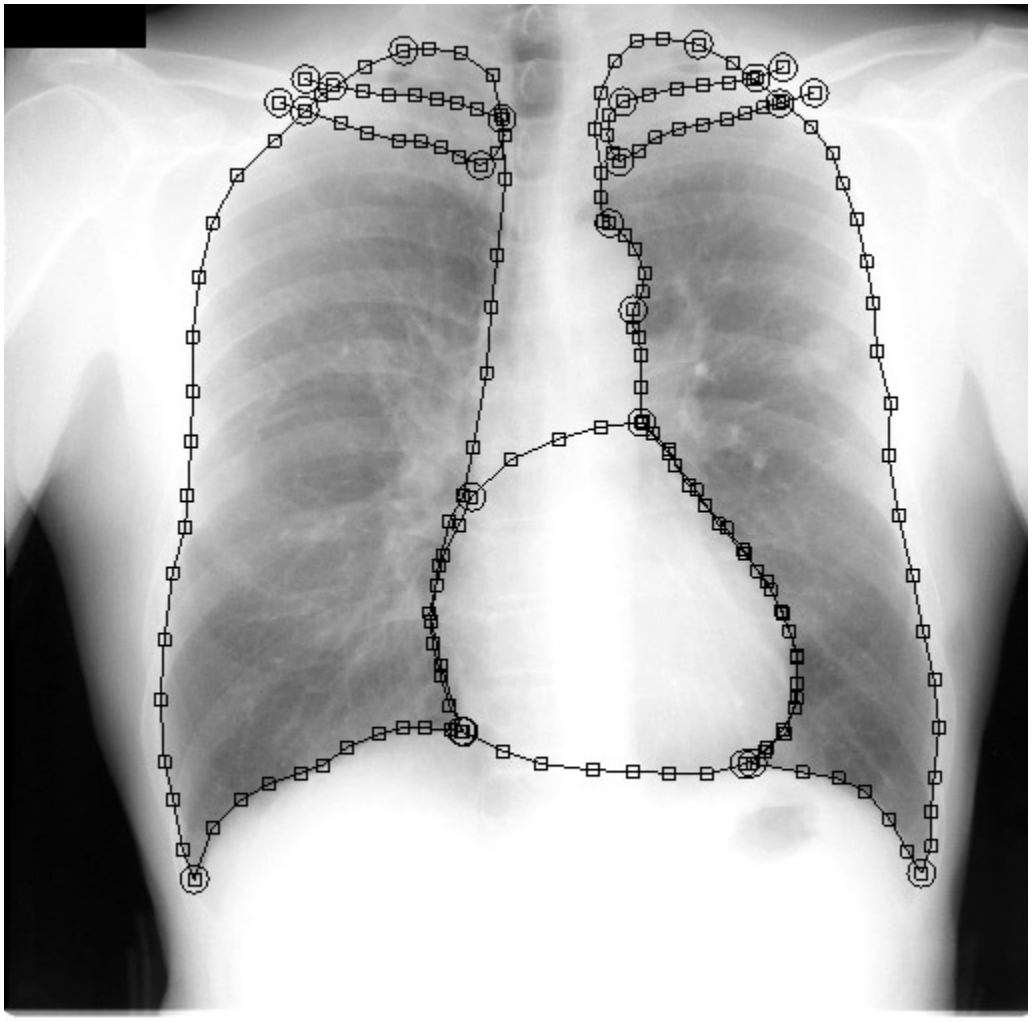

Figure 1. Manual segmentation by the first observer of the first image of the JSRT database. Lung fields, the heart, and the clavicles are delineated. Anatomical or distinctive points (see below) are circled. The right lung contains 3 of these points, the left lung 5, the heart 4 and each clavicle 6.

There are segmentation algorithms that require that in each training image, a set of corresponding points has been indicated. To this end several distinguishable points on each contour have been clicked by the observers, indicating anatomical or other characteristic landmarks. These characteristic points are assumed to correspond. After the complete boundary has been defined, all but the corresponding points are discarded and subsequent points are obtained by equidistantly sampling a certain fixed number of points along the contour between the aforementioned indicated points. This is illustrated in the figure below.

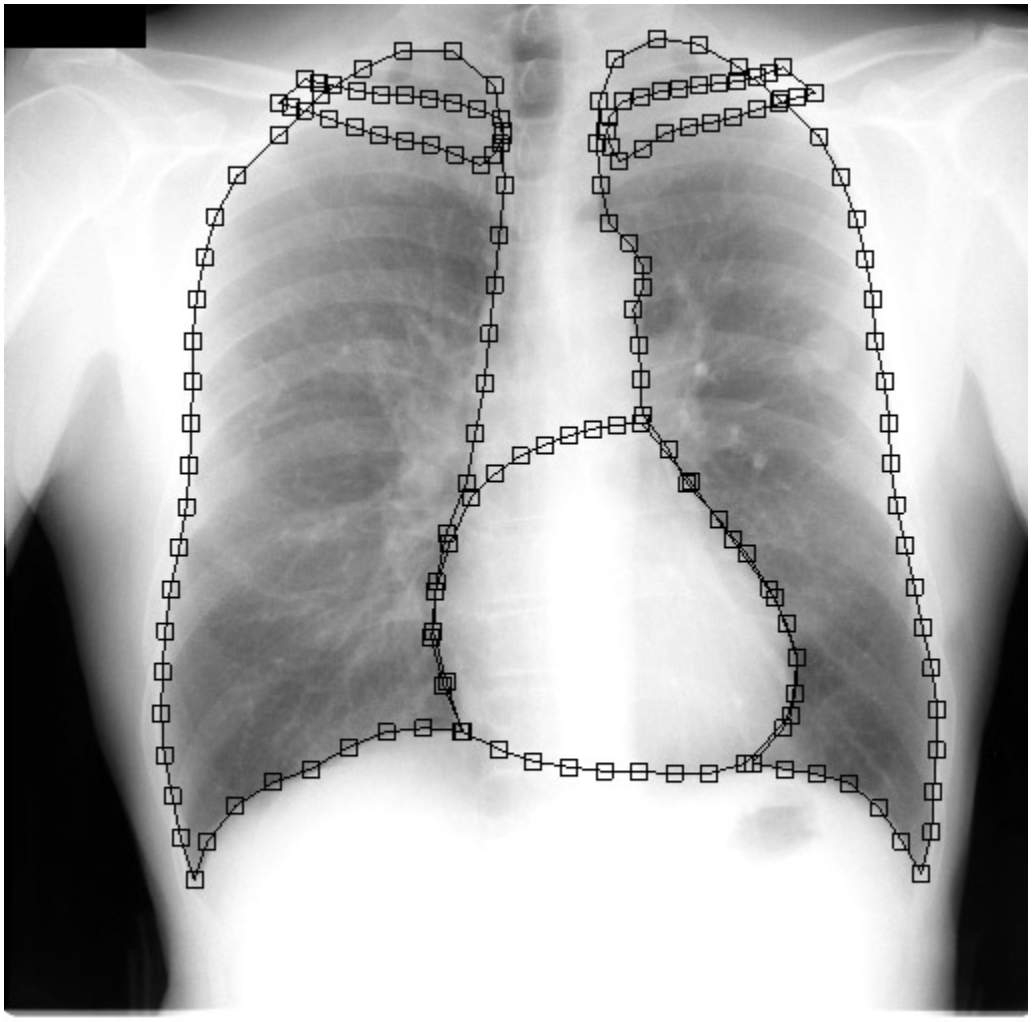

Figure 2. The landmarks interpolated between the anatomical landmarks along the contours indicated in Figure 1. The total number of landmarks is 166, with 44, 50, 26, 23 and 23 points in right lung, left lung, heart, right clavicle, and left clavicle, respectively.

For more information, consult

- B. van Ginneken, M.B. Stegmann, M. Loog, Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database, *Medical Image Analysis*, nr. 1, vol. 10, pp. 19-40, 2006.

-