Photoacoustic imaging which combines high contrast of optical imaging and high resolution of ultrasound imaging, can provide functional information, potentially playing a crucial role in the study of breast cancer diagnostics. However, open source dataset for PA imaging research is insufficient on account of lacking clinical data.

To tackle this problem, we propose a method to automatically generate breast numerical model for photoacoustic imaging. The different type of tissues is automatically extracted first by employing deep learning and other methods from mammography. And then the tissues are combined by mathematical set operation to generate a new breast image after being assigned optical and acoustic parameters.

Segmenting the breast is the most important step when generating dataset. The breast can be mainly divided into four types of tissues: skin, fat, fibroglandular, and tumor. Among them, extracting fibroglandular is the most crucial and difficult task, which is the key to the implementation of automatic generation of photoacoustic numerical model dataset. The feature extraction of fibroglandular is complicated, because its shape is irregular and has no clear boundary, so that even manual labeling is very troublesome.

The CBIS-DDSM (Curated Breast Imaging Subset of DDSM) dataset will be used, which is an upgraded version of DDSM (Digital Database for Screening Mammography), containing 10,239 processed mammography. In particular, manual cropping will be performed to process the fibroglandular tissue, which is the input of our method based on the guidance of clinicians.

Deep learning provides an effective automatic extraction method, which has been widely used in the field of medical image segmentation. Convolutional neural network will be utilized in our work. In order to implement the algorithm better, we preprocess the dataset by data augmentation.

Finally, Dice loss and Focal loss function will be utilized to evaluate fibroglandular segmentation performance.