Using Object Detection Models to Detect Lesions in Breast Ultrasound Videos

Project Management and Software Development
for Medical Applications

General Info

Contact Person: Zhikai Yang

Contact Email: zhikai@kth.se

Project Abstract

Digital Breast Tomosynthesis (DBT) is an advanced breast imaging technique that provides three-dimensional (3D) information of the breast tissue, offering improved detection and characterization of breast lesions compared to conventional two-dimensional (2D) mammography. However, the iterative reconstruction algorithms employed in DBT are computationally expensive, leading to long reconstruction times. This research proposal aims to develop a deep learning-based approach for efficient and accurate DBT reconstruction, leveraging the power of artificial neural networks to improve the clinical value of DBT.

Background and Motivation

Breast ultrasound (BUS) imaging is widely used for detecting breast lesions due to its safety, accessibility, and effectiveness, particularly for dense breast tissue where mammography may have limitations. While traditional breast ultrasound diagnosis relies on still images, ultrasound videos can provide richer spatiotemporal information, capturing the movement and variations of the tissue. However, manual interpretation of these videos is time-consuming, prone to human error, and highly dependent on the radiologist's experience.

Automated lesion detection in breast ultrasound videos can significantly reduce the diagnostic workload and improve early diagnosis accuracy. Deep learning-based object detection models have shown impressive results in detecting objects across various domains, including medical imaging. Applying such models to breast ultrasound videos holds great potential for improving lesion detection accuracy by leveraging temporal information across frames. This project aims to develop and evaluate object detection models for automatically identifying and localizing breast lesions in ultrasound videos.

To develop a deep learning-based object detection model for identifying and localizing breast lesions in ultrasound videos.

To explore temporal coherence between video frames to improve detection accuracy.

To evaluate the model's performance on publicly available datasets, comparing it to baseline methods.

Student's Tasks Description

Data Collection and Preprocessing

Data Collection: Publicly available breast ultrasound video datasets will be used. If unavailable, synthetic video data could be generated from high-resolution breast ultrasound images by simulating probe motion.

Preprocessing: The ultrasound video data will be preprocessed to reduce noise and enhance contrast. Standard techniques, such as normalization and filtering, will be applied. Data augmentation techniques like frame jittering, video rotation, and temporal cropping will be used to make the model more robust to different conditions.

Model Architecture

Object Detection Model: The core detection system will be based on state-of-the-art object detection frameworks such as YOLOv8, Faster R-CNN, or RetinaNet, adapted for video inputs. These models will be modified to incorporate temporal information from consecutive frames in ultrasound videos.

Evaluation

Metrics: Standard object detection metrics such as mean Average Precision (mAP), Recall, and Precision will be used to evaluate model performance. Additionally, video-specific metrics such as frame-level accuracy and temporal consistency (evaluating lesion tracking across frames) will be considered.

Baseline Comparison: The proposed model's performance will be compared with other state-of-the-art object detection models applied to individual frames of ultrasound videos without exploiting temporal information.

Technical Prerequisites

The student is required to possess a good knowledge of Python coding. In addition, a basic understanding of the machine learning knowledge is required. Furthermore, have some experience of image processing and deep learning experiences

References

Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. IEEE International Conference on Computer Vision (ICCV).

Feichtenhofer, C., Fan, H., Malik, J., & He, K. (2019). SlowFast Networks for Video Recognition. IEEE/CVF International Conference on Computer Vision (ICCV).