Improving Breast Abnormality Analysis in Mammograms using CycleGAN

Rahul Raj Menon

Supervisors: Mathias Öttl, Adarsh Bhandary Panambur,
Prof. Dr.-Ing. habil. Andreas Maier

Master's Thesis Description

Breast Cancer is one of the leading causes of death for women around the world. There were approximately 2.3 million new cases and 685,000 deaths worldwide in the year 2020 [1]. Severe cases of breast cancer can be prevented by early detection. Mammography is the most common method of detecting cancer in a patient. The main types of mammography include film and digital mammography. Both types of mammography use x-ray radiation to produce an image of a breast. In film mammography, mammograms are read and stored on films whereas digital mammograms are read and stored digitally. Digital mammograms are easier to share, view, manipulate and analyze the images, and are currently considered the gold standard technology for breast cancer screening [2].

Deep Learning (DL) techniques have proven to improve the process of automated detection of breast cancer and diagnosis [3]. However, a major issue in the field of DL-based mammography analysis is the lack of high-quality annotated data. Hence, Generative Adversarial Networks (GANs) can be used to synthesize the images similar to real images and generate fair distribution of images containing all the different abnormalities. These generated synthetic images can be used for data augmentations and can significantly increase the number of training images [4, 5].

Even though there are some publicly available datasets with pixel-level annotations for abnormalities, most of the other datasets, have only image-level annotations, making it difficult to analyze the explainability of DL model performances [6, 7]. Recent works have shown that image-to-image translation methods can support mitigating the problem to some extent [8, 9]. These methods are used to learn the mapping between images from different domains so that the characteristics of images in one domain can be transferred to another domain. But most of the publicly available data does not have paired training samples. CycleGAN can be used to train unpaired samples. CycleGAN incorporates an additional cycle-consistency loss and an identity loss in comparison with a GAN.

In this thesis, we aim to perform unpaired image-to-image translation between two publicly available mammography datasets. Specifically, we aim to use the CBIS-DDSM dataset, which consists of scanned film mammogram images, and the CMMD dataset, consisting of digital mammogram images. CBIS-DDSM contains pixel-level annotations and the CMMD has image-level annotations [10, 11]. We aim to use a CycleGAN architecture to perform the translation between the aforementioned datasets and use the artificially generated data to investigate whether the rich annotations from CBIS-DDSM can be utilized to improve the performance of abnormality analysis on the images from CMMD, and vice-versa.

The main goals of this thesis are as follows:

- Removing artifacts from scanned film CBIS-DDSM images using a combination of manual and automated methods.
- Implement a CycleGAN for image translation between scanned film mammography and digital mammography.
- Use CycleGAN transformations for better localization and analysis of breast abnormalities in Mammogram images.
- Investigate the impact of a Lesion-Guidance-based-CycleGAN on the localization and analysis of breast abnormalities.

References

[1] Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, Zeng H, Zhou J, Wei W. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. Cancer Commun (Lond). 2021 Nov;41(11):1183-1194. doi: 10.1002/cac2.12207. Epub 2021 Aug 16. PMID: 34399040; PMCID: PMC8626596.

[2] Faridah Y. Digital versus screen film mammography: a clinical comparison. Biomed Imaging Interv J. 2008 Oct;4(4):e31. doi: 10.2349/biij.4.4.e31. Epub 2008 Oct 1. PMID: 21611016; PMCID: PMC3097746

[3] Mohammed Amine Naji, Sanaa El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida Ait Abdelouhahid, Olivier Debauche, Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis, Procedia Computer Science, Volume 191, 2021, Pages 487-492, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.07.062.

[4] Guan S, Loew M. Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. J Med Imaging (Bellingham). 2019 Jul;6(3):031411. doi: 10.1117/1.JMI.6.3.031411. Epub 2019 Mar 23. PMID: 30915386; PMCID: PMC6430964.

[5] Oyelade, O.N., Ezugwu, A.E., Almutairi, M.S. et al. A generative adversarial network for synthetization of regions of interest based on digital mammograms. Sci Rep 12, 6166 (2022). https://doi.org/10.1038/s41598-022-09929-9

[6] Cai, H., Huang, Q., Rong, W., Song, Y., Li, J., Wang, J., Chen, J., Li, L. (2019). Breast Microcalcification Diagnosis Using Deep Convolutional Neural Network from Digital Mammograms. Computational and Mathematical Methods in Medicine, 2019, 1â10. https://doi.org/10.1155/2019/2717454

[7] Lee, R. S., Gimenez, F., Hoogi, A., Miyake, K. K., Gorovoy, M., Rubin, D. L. (2017). A curated mammography data set for use in computer-aided detection and diagnosis research. In Scientific Data (Vol. 4, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/sdata.2017.177

[8] M. Hammami, D. Friboulet and R. Kechichian, "Cycle GAN-Based Data Augmentation For Multi-Organ Detection In CT Images Via Yolo," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 390-393, doi: 10.1109/ICIP40778.2020.9191127

[9] C. Han et al., "Combining Noise-to-Image and Image-to-Image GANs: Brain MR Image Augmentation for Tumor Detection," in IEEE Access, vol. 7, pp. 156966-156977, 2019, doi: 10.1109/ACCESS.2019.2947606.

[10] Cui, Chunyan; Li Li; Cai, Hongmin; Fan, Zhihao; Zhang, Ling; Dan, Tingting; Li, Jiao; Wang, Jinghua. (2021) The Chinese Mammography Database (CMMD): An online mammography database with biopsy confirmed types for machine diagnosis of breast. The Cancer Imaging Archive. DOI: https://doi.org/10.7937/tcia.eqde-4b16

[11] Sawyer-Lee, R., Gimenez, F., Hoogi, A. and Rubin, D. (2016). Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) (Version 1) [Data set]. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.7O02S9CY