

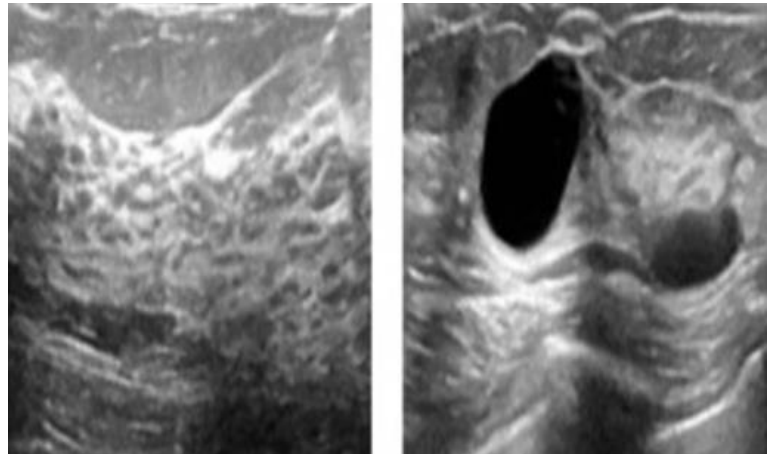
Medical Ultrasound Images for Breast Cancer Classification

Christoph Balada (christoph.balada@dfki.de)

Description

This project is about classifying different types of breast cancer based on ultrasound images. To achieve this a deep neural network should be trained in two phases. First phase concerns an unsupervised pretraining using a modern approach like BYOL, SimCLR, SWAV, etc. and different publicly available datasets. In the second phase, the model will be finetuned on the target dataset. To evaluate the results Captum library should be used to apply different approaches from Explainable AI (ExAI) domain.

To make it all perfect, a minimalist web-based frontend is to be provided for the model. This frontend shall provide the options to upload a new image to be classified and set any hyperparameters for inference. Furthermore, the input image, the result and the findings of the ExAI shall be displayed. Dash is a good choice for implementation, although other libraries are also possible.



Requirements

- Very good Python skills
- Machine Learning / PyTorch
- Experience with Plotly/Dash/Bokeh is a plus
- First hands-on experience with ExAI approaches preferable

Test Task

- Download a small image dataset; which one doesn't matter, but it should not contain more than a few thousand images
 - e.g. Ultrasound breast image dataset:
<https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset>
- Use PyTorchs on ImageNet preinitialized ResNet50 to process the whole dataset
 - Note: ResNet50 expects the input to be of size 224x224 px – take care of this
- Extract the output from the **last feature layer** representing the encoded feature vector (→ output right before the fully connected layers)
- Use t-SNE (e.g., scikit-learn implementation) to compute a low-dimensional representation of all feature vectors in **3 dimensions**
- Use Dash / Plotly or Bokeh to generate a HTML file showing an interactive view of the t-SNE encoded data representations

Submit your code along with the generated HTML file that contains your results.