Degree project 30 credits in Biomedical Engineering

Medical Data Augmentation using GANs

Biomedical Engineering R&D (MT-FoU) is a research and development department at the Center for Information Technology and Biomedical Engineering at Norrland University Hospital, Region Västerbotten. The department conducts international research, development and education in the field of biomedical engineering, with expertise in, for example, sensors and measurement systems, image and signal analysis and biomechanical models. MT-FoU is also a part of the competence center AI for Medicine in Northern Sweden, AIM North, which supports clinical research projects with technical method expertise in machine learning and AI.

Background

One of the biggest issues facing the use of Artificial Intelligence (AI) in medical imaging is the need for more availability of large, labelled datasets. The annotation of medical images is not only expensive and time-consuming but also highly dependent on the availability of expert observers. The limited amount of training data can inhibit the performance of supervised machine learning algorithms, which need large quantities of data on which to train to avoid overfitting. Image-based Data Augmentation (DA) techniques, such as cropping, rotating, and flipping input images, alleviates the burden of the limited medical dataset by using existing data more effectively. However, standard data augmentation produces only certain restricted variations in the data. To handle this challenge, a class of generative machine learning models called Generative Adversarial Networks (GANs) offer a potentially valuable addition to the set of augmentation techniques and offer a novel way to unlock additional information from a dataset by generating synthetic samples with the appearance of real images. Hence, GANs can be used to generate new samples that can be used to train ML models.

Aim of the project

Using GAN to generate high-quality synthetic images could allow us to overcome the problem of the limited medical dataset. To achieve this goal, we developed a GAN-based DA procedure that augments each instance in the original dataset increasing the information provided to a classification ML algorithm. We first apply our DA algorithm to a cohort of 190 patients with Non-Small Cell Lung Cancer (NSCLC), finding encouraging results. This project leverages two steps. First, we plan to apply the developed algorithm to publicly available medical datasets of different modalities, e.g. MRI, X-RAY, etc. that suffers from the problem of limited medical data. As for the second step, we aim to further develop the GAN-based DA algorithm to maximise the information provided by the classification network increasing its robustness and generalisation.

Work description

The student will join an ongoing project on this topic recently started, where we have implemented the StyleGAN model and designed the DA algorithm. The student work will consist of the following main steps:

- Literature analysis and personal study of DA techniques in the medical domain;
- Literature analysis and personal study on GAN-based medical data augmentation methods;
- Data collection of publicly available medical datasets of different modalities (MRI, X-RAY, etc.) that suffer from the problem of limited data;
- Understanding the developed GAN-based DA algorithm;
- Understanding the decision support system for classification;
- Test the DA algorithm on found medical datasets and evaluate its performance;
- Eventually improve the DA algorithm;
- Thesis writing

Supervisor at MT-FoU
Paolo Soda

CIMT, Medicinsk teknik – FoU
Region Västerbotten
E-post: paolo.soda@regionvasterbotten.se
Web: https://www.regionvasterbotten.se/medicinsk-teknik-forskning-och-utveckling