Hamburg University of Technology
Institute for Medical Technology and Intelligent Systems
Computer Science

Exposé for Bachelor Thesis on the topic:

Impact of different Data Augmentation Techniques for Deep Learning with Optical Coherence Tomography

Bachelor Thesis

for the attainment of the academic degree

BACHELOR OF SCIENCE (B.SC.)

20.11.2021

Submitted by: Lennard Korte
Roggenstieg 9
22927 Großhansdorf
lennard.korte@tuhh.de

Field of Studies: B.Sc. Computer Science
Majoring in Computer- and
Software-Engineering
Semester: 5
Matr.No.: 53673

Reviewers: Prof. Dr. Alexander Schlaefer
Robin Mieling M.Sc.

Table of contents

Definition of the Problem

Ischaemic heart disease is the most common cause of death in the world, which illustrates the importance of medical treatment in particular for cardiology.[1] Many types of coronary disease can be treated or prevented, but often coronary heart diseases lead to death, which is why appropriate treatment is critical. Intravascular Optical Coherence Tomography is one way to threat diseases e.g. stenoses caused by potentially calcified plaque.[2]

For each patient hundreds or thousands of images are created as a result and must be evaluated to decide for the right treatment. Decision support systems are required to decide if plaque is detected. From a medical point of view it is also relevant to find out if plaque is calcified or not.

Deep Learning (DL) algorithms are applied to give decision support for the treating person. Such algorithms need to be trained with data sets, which have influence on the overall performance of the out-coming model. Data Augmentation Techniques can improve such performances by modifying the training data sets in certain ways, which often lead to an increased multiplicity. It has been shown that Data Augmentation can have a positive effect on the performance of DL models with Optical Coherence Tomography. However there are many questions left on how to improve the overall performance further by optimizing Data Augmentation.

Objectives and Research Interest

The objective of this bachelor thesis is to investigate the influence of different Data Augmentation techniques on the DL models in Optical Coherence Tomography (OCT). The foundation for this is mainly laid by the research for OCT by Prof. Dr. Alexander Schlaefer from the Institute for Medical Technology and Intelligent Systems at Hamburg University of Technology (TUHH) on "Automatic Plaque Detection in IVOCT Pullbacks Using Convolutional Neural Networks".[3,4]

Different Data Augmentation techniques for DL models in OCT are investigated and compared. It will be determined which Deep Learning models applied to OCT work well with which Data Augmentation techniques and how they affect the overall performance.

The following are the objectives of the work:

- Illustrate what different, documented and new/proprietary augmentation techniques for DL in OCT can be applied to (given) Intravascular(IV)-OCT data sets.
- Implement in (partially provided) code various Data Augmentation techniques for IV-OCT.
- Apply Data Augmentation techniques for IV-OCT where possible to other OCT data sets that have yet to be collected.
- Evaluate different Data Augmentation techniques for polar as well as cartesian images and test them with the implemented applications.
- Evaluation: benchmark, contrast and relate the impact of the collected Data Augmentation techniques using various metrics.
- Based on the results, answer the following research question: Under what circumstances are which Data Augmentation techniques best suited for DL with OCT?

Data Augmentation techniques are expected to fundamentally enhance the performance of Deep Learning models. Moreover, they can mutually influence their impact on the performance of DL models. Depending on the DL model and data sets used, different consequences and results can be expected.

State of Research and Theoretical Concepts

Today, Optical Coherence Tomography is used in many different areas of medicine. These include not only cardiology, but also, for example, ophthalmology and oncology. It often contributes significantly to the decision making of treating specialists during diagnosis and treatment. Different methods in hardware as well as software lead to a high amount and variety of data, which makes it a challenge to interpret them correctly.

The Deep Learning algorithms, which will be used to assist in the interpretation of the diagnostic results, have so far shown an F1 score of approximately 0.888 under certain real-world conditions.

From this, a relatively safe classification of the data can be derived by DL models. However, in critical areas of medicine, especially in IV-OCT, more reliable algorithms are needed. One way to improve the performance of such models would be to use existing data sets and apply Data Augmentation techniques to them.

In recent publications, simple Data Augmentation techniques have already been applied. However, these have not yet been studied in detail regarding their impact in IV-OCT. There are also other Data Augmentation techniques that have not yet been considered in IV-OCT.

Research Concept

The following questions will be answered in this work:

- What types of Data Augmentation Techniques are suitable for (IV-)OCT?
- To what further data sets can the techniques be applied to?
- How do the techniques suit and perform with the different data sets?
- Optional: How do the Techniques perform with other DL Models?

In order to investigate the effects of Data Augmentation in Deep Learning with Optical Coherence Tomography, the first focus will be on IV-OCT. Materials as well as code modules are already available for this purpose, which can be investigated and further developed. In addition, other data sets will be included if possible. The performance can afterwards also be tested on these.

Within the bachelor thesis the following methodological steps are implemented:

1. Improve code quality of given ML Code
2. Research on which Data Augmentation techniques should be considered
3. Research on other data sets with which the DL models can be trained
4. Analyze, examine the behavior on the different techniques

Common or custom metrics will be used to analyze the performance of the Deep Learning models with the different Data Augmentation techniques. As a first training set, the IV-OCT images described by Prof. Schlaefer will be used. The scans are divided into six different classes and successful training on them has already been demonstrated. For training Robin Mieling provided a code basis on which can be build on.

Preliminary Outline

1. Introduction
2. Theoretical Concepts
    - OCT: Spectral Domain, Swept Source
    - IV-OCT
    - Deep Learning
    - DL Metrics
3. State of the Art Research
    - Problem
4. Objective / Question
5. Methodology
6. Data Augmentation
    - Common Techniques
    - New Techniques
7. Implementation
8. Results
    - Testing
    - Systematic Evaluation
    - Assessment of Performance
9. Behavior on other Data Sets
    - Testing
    - Systematic Evaluation
    - Assessment of performance
10. Discussion
    - Transfer Learning
    - Conclusion / Proposition

Schedule

Duration: Nine Weeks (22.11.2021–30.01.2022)

Every Friday there will be a summary about the work completed in the past week. This report will be submitted to the supervisor Robin Mieling M.Sc. and Prof. Dr. Alexander Schlaefer.

1. Until 30.11.2021: Literature research
2. Until 06.12.2021: Thematic introduction + hypotheses
3. Until 16.12.2021: Coding and Experimenting
4. Until 30.12.2021: Rough draft of main part
5. Until 01.01.2022: Introduction and Conclusion
6. Until 03.01.2021: Oberseminar
7. Until 05.01.2022: Raw version: introduction + conclusion
8. Until 18.01.2022: Revision + correction
9. Until 24.01.2022: Layout + cover page
10. Until 27.01.2022: Print
11. Until 29.01.2022: Submission

Table: Overview

Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
1	28.11.21	29.11.21	**30.11.21**	01.12.21	02.12.21	03.12.21	04.12.21
2	05.12.21	**06.12.21**	07.12.21	08.12.21	09.12.21	10.12.21	11.12.21
3	12.12.21	13.12.21	14.12.21	15.12.21	**16.12.21**	17.12.21	18.12.21
4	19.12.21	20.12.21	21.12.21	**22.12.21**	23.12.21	24.12.21	25.12.21
5	26.12.21	27.12.21	28.12.21	29.12.21	**30.12.21**	31.12.21	**01.01.22**
6	02.01.22	**03.01.22**	04.01.22	**05.01.22**	06.01.22	07.01.22	08.01.22
7	09.01.22	10.01.22	11.01.22	12.01.22	13.01.22	14.01.22	15.01.22
8	16.01.22	17.01.22	**18.01.22**	19.01.22	20.01.22	21.01.22	22.01.22
9	23.01.22	**24.01.22**	25.01.22	26.01.22	**27.01.22**	28.01.22	**29.01.22**

Schedule - Modified

This schedule version includes the replanning, since original plannings for a semester abroad got canceled. This open opportunity to write additional exams. This doesn’t mean it is not possible to complete the work earlier.

Duration: Nine Weeks (30.11.2021–31.03.2022)

Every Friday there will be a summary about the work completed in the past week. This report will be submitted to the supervisor Robin Mieling M.Sc. and Prof. Dr. Alexander Schlaefer.

1. Until 30.11.2021: Literature research
2. Until 12.12.2021: Thematic introduction + hypotheses
3. Until 06.01.2022: Coding and Experimenting
4. Until 28.01.2022: Rough draft of main part
5. Until 01.02.2022: Introduction and Conclusion
6. Until 06.02.2022: Oberseminar
7. Until 23.02.2022: Raw version: introduction + conclusion
8. Until 16.03.2022: Revision + correction
9. Until 21.03.2022: Layout + cover page
10. Until 25.03.2022: Print
11. Until 31.03.2022: Submission

Table: Overview-modified

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Dez	1											2							official Vacation												
Jan						3																						4			
Feb	5					6																	7								
Mar																8					9				10						11

References

[1] https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[2] https://arxiv.org/pdf/1804.03904.pdf

[3] https://ieeexplore.ieee.org/abstract/document/8438495