

Nils Feldhus and
Dr. Philippe Thomas

Joint Theses between TU Berlin and DFKI: Open Topics for Master Theses

DFKI's Speech and Language Technology lab is led by Prof. Dr. Sebastian Möller, which is why the QU lab and DFKI's SLT lab collaborate closely. One of the areas of collaboration concerns joint Master Theses. The thesis will also be accompanied by the Universität Lübeck (Institut für Medizinische Informatik), who will provide expertise on the analysis of medical images.

INTERESTED? If you are interested in exploring the topic further, please get in touch with Nils Feldhus <Nils.Feldhus@dfki.de> and Dr. Philippe Thomas <Philippe.Thomas@dfki.de> to discuss the next steps.

IMPORTANT: Please provide a brief motivational note, your CV and a current transcript of records. Please also indicate your level of experience with regard to software development (e.g., programming languages, tools, approaches), natural language processing and modern machine learning techniques.

DFKI GmbH
Speech & Language Technology Lab
Projektbüro Berlin
Alt-Moabit 91c
10559 Berlin

Telefon: +49 (0)30 23895-1852
Telefax: +49 (0)30 23895-1810
E-Mail: nils.feldhus@dfki.de |
philippe.thomas@dfki.de
Internet: www.dfki.de

Deutsches Forschungszentrum für
Künstliche Intelligenz GmbH (DFKI)

Firmensitz
Kaiserslautern

Weitere Standorte und Betriebsstätten:
Saarbrücken, Bremen, Osnabrück,
Oldenburg, Berlin, St. Wendel

Geschäftsführung
Prof. Dr. Antonio Krüger

Vorsitzender des Aufsichtsrats
Dr. Gabriël Clemens

Amtsgericht Kaiserslautern HRB 2313
USt-ID-Nummer DE 148 646 973
Steuernummer 19/672/50006

Stadtsparkasse Kaiserslautern
IBAN: DE60 5405 0110 0028 0004 79
BIC/SWIFT: MALADE51KLS

Title:

Biomedical Image Caption Generation

Problem statement:

Automatic image captioning is the problem of generating a text description for an image. It has seen ever-growing interest from both computer vision and natural language processing communities [1]. Image captioning models have been applied to the medical domain quite successfully [2]. However, the access to high-quality datasets for training is very restricted [3] and evaluation metrics assessing clinical correctness and coherence remain an unsolved problem [4].

Approach and goal:

In collaboration with Universität Lübeck, we will discuss and select a dataset of x-ray thorax images from real patients. Each image is accompanied by a short transcript with medical findings. We foresee two options to tackle the image captioning problem in the biomedical domain:

- 1.) Applying state-of-the-art neural machine learning models to the problem (full image-to-text pipeline). This includes encoder-decoder and transformer-based neural architectures. The evaluation includes an in-depth analysis on model performance.
- 2.) Trying a novel approach using image tags to automatically generate the caption for the image. Tags will be automatically generated by a state-of-the-art image classification model provided by U-Lübeck. These tags will be used as input in order to generate an image caption. Similar work has been done for text-to-text generation tasks [5].

References:

- [1] Bernardi et al., 2016: “Automatic Description Generation from Images: A Survey of Models, Datasets, and Evaluation Measures”
<http://arxiv.org/abs/1601.03896>
- [2] Pavlopoulos et al., 2021: “Diagnostic Captioning: A Survey”
<http://arxiv.org/abs/2101.07299>
- [3] Johnson et al., 2019: “MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs”
<http://arxiv.org/abs/1901.07042>
- [4] Miura et al., 2020: “Improving Factual Completeness and Consistency of Image-to-Text Radiology Report Generation”
<http://arxiv.org/abs/2010.10042>
- [5] Raffel et al., 2019: "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer"
<http://arxiv.org/abs/1910.10683>