15th Conference on

DATA ANALYSIS METHODS for Software Systems

November 28–30, 2024

Druskininkai, Lithuania, Hotel “Europa Royale”
https://www.mii.lt/DAMSS

Co-Chairmen:

Prof. **Gintautas Dzemyda** (Vilnius University, Lithuanian Academy of Sciences)
Dr. **Saulius Maskeliūnas** (Lithuanian Computer Society)

Programme Committee:

Dr. **Jolita Bernatavičienė** (Lithuania)
Prof. **Juris Borzovs** (Latvia)
Prof. **Janis Grundspenkis** (Latvia)
Prof. **Janusz Kacprzyk** (Poland)
Prof. **Ignacy Kaliszewski** (Poland)
Prof. **Bożena Kostek** (Poland)
Prof. **Tomas Krilavičius** (Lithuania)
Prof. **Olga Kurasova** (Lithuania)
Assoc. Prof. **Tatiana Tchemisova** (Portugal)
Assoc. Prof. **Gintautas Tamulevičius** (Lithuania)
Prof. **Julius Žilinskas** (Lithuania)

Organizing Committee:

Dr. **Jolita Bernatavičienė**
Prof. **Olga Kurasova**
Assoc. Prof. **Viktor Medvedev**
Laima Paliulionienė
Assoc. Prof. **Martynas Sabaliauskas**
Prof. **Povilas Treigys**

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Tel. (+370 5) 2109 315
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt

https://doi.org/10.15388/DAMSS.15.2024
ISBN 978-609-07-1112-5 (digital PDF)

Change Detection in Satellite Imagery Using Transformer Models and Machine Learning Techniques: A Comprehensive Captioning Dataset

Kürşat Kömürcü, Linas Petkevičius

Institute of Computer Science
Vilnius University

kursatkomurcu@gmail.com

This paper addresses the growing need for automating the caption generation of satellite image pairs, focusing on change detection tasks. The study leverages four major datasets—CLCD, LEVIR-CD, DSIFN, and S2Looking—to create a satellite image caption change detection dataset containing descriptions of 16,753 image pairs. The primary aim is to generate descriptive and accurate captions using the Llama model and evaluate the performance of various machine learning and transformer models in detecting changes between pre-event and post-event images based on these captions. The introduction highlights the importance of automated change detection in remote sensing for applications such as urban planning, environmental monitoring, and disaster management. Traditional manual interpretation of satellite images is time-consuming and requires expertise, underscoring the value of machine learning models in automating this process. The study uses a combination of deep learning techniques, particularly transformer models like BERT, DistilBERT, RoBERTa, and XLNET, and classical machine learning models, including Logistic Regression, Naive Bayes, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), to evaluate the generated captions. Four datasets are utilised for this task. The CLCD dataset focuses on multi-temporal and multi-sensor analysis for land cover changes, while the LEVIR-CD dataset is centered on building change detection in urban areas. DSIFN enhances high-resolution image change detection with a dual-stream network approach, and S2Looking provides Sentinel-2 images for large-scale change detection studies. Data augmentation

techniques, including random rotations and scaling, are employed to address dataset imbalances and improve the model's ability to generalise across different scenarios. For caption generation, the Llama model is used due to its robust transformer-based architecture, which excels at processing complex visual data and converting it into coherent and contextually relevant natural language descriptions. The generated captions are further refined for clarity and accuracy. The model is trained to maximise the likelihood of the correct word sequences, using a self-attention mechanism to capture dependencies and relationships within the data. The results section presents an in-depth evaluation of various machine learning and transformer models. SVM consistently outperforms traditional machine learning models, achieving the highest accuracy across all datasets. Transformer-based models also demonstrate strong performance, with BERT and RoBERTa leading the pack. BERT excels in training accuracy due to its bidirectional training on masked language modeling, while RoBERTa shows better generalisation on validation datasets, particularly those that benefit from optimisation techniques. DistilBERT offers a faster alternative with slightly reduced accuracy, and XLNET, while competitive, does not outperform BERT and RoBERTa. The study concludes that the Llama model, combined with transformer models, effectively generates accurate and descriptive captions for satellite image pairs, facilitating automated change detection tasks. This research contributes to a new satellite image caption change detection dataset and provides valuable insights into the performance of various models in this domain. The findings underscore the potential of using advanced deep learning techniques, such as transformer models, to translate complex visual data into descriptive language, enabling more informed decision-making across a range of applications.

Acknowledgements: This research was funded by the European Union (project No S-MIP-23-44) under the agreement with the Research Council of Lithuania (LMTLT).