

## Abstract

---

### Remote Sensing Image Retrieval using English and Arabic Textual Descriptions

by

Norah Alsharif

Master of Science in the Department of Computer Engineering at the College of Computer and  
Information Sciences.

King Saud University, Riyadh, June 2022

Prof. Yakoub Bazi

As a result of the rapid development of earth observation technologies, the amount of Remote Sensing (RS) imagery has been increasing dramatically. Consequently, effectively managing and retrieving this massive amount of data has become necessary. In the RS community, image retrieval is considered one of the main important and exciting fields of research. Usually, content-based Remote Sensing Image Retrieval (RSIR) approaches rely on a query image to retrieve the relevant images. To increase the flexibility of the retrieval experience, cross-modal representations based on text-image pairs are gaining popularity. Indeed, combining text and image domains is regarded as one of the next frontiers in RSIR. Yet, aligning text to the content of remote sensing images is particularly a challenging task due to the visual-semantic discrepancy between language and vision worlds. In this thesis, we are particularly interested in text-to-image retrieval. In addition, we will explore the Arabic text-to-image retrieval to bridge the gap between the Arabic language and artificial intelligence. In particular, we propose an embedding model based on two transformer encoders a language transformer encoder and a vision transformer encoder, for English text-to-image retrieval. Then in order to increase the flexibility of the retrieval process for non-English speakers, we further fine-tune the model on descriptions from the Arabic language using single language model and bilingual model. The experimental results demonstrate that the proposed model significantly improves the retrieval performances in terms of recall compared to the existing state-of-the-art RSIR methods based on recurrent networks.

**Keywords:** Remote sensing, cross-modal retrieval, transformers, contrastive loss.