Abstract

Aerial Camera-Based Localization Using Deep Learning Methods

Judit Hermán

Camera-based localization in GPS-denied environments is becoming increasingly important in the fields of computer vision and autonomous navigation. The aim of this thesis is to develop a deep learning-based method that enables accurate positioning of Unmanned Aerial Vehicles (UAVs) using only image data. The core idea of the approach is to match images captured by the drone to pre-stored satellite images, using this similarity to estimate the UAV's location.

Throughout this work, I applied various types of autoencoder architectures, including convolutional autoencoders and pretrained convolutional neural networks (VGG16, ResNet50, EfficientNet-B0) combined with decoder blocks. These models are designed to generate compact latent representations of the images, making it possible to compare them in a lower-dimensional space. In addition, a CycleGAN model was used to reduce visual discrepancies between UAV and satellite images, enhancing the robustness of the matching process.

The models are trained and tested on real-world aerial datasets captured over Hungary, with evaluation based on metrics such as RMSE, MAE, NDCG, and top-k accuracy across multiple search radii. Results show that ensemble models combining different autoencoders significantly improve localization accuracy. This work demonstrates the viability of autoencoder-based representations and adversarial image alignment for robust visual localization in real-world UAV applications.