

# Aerial Image Based Localisation via Learned Embeddings

Visual Place Recognition (VPR) has drawn significant attention in the past decade thanks to the advantages it can offer for localisation and navigation. It has been proven in the past that deep learning techniques can be applied to navigate robots based on their optical sensors. However, there are still challenges present when aliasing is present in both rural and urban environments. UAV image-based localisation has been demonstrated to be effective in urban environments where geometric features such as buildings and road shapes act as distinct features to aid in localisation. However, the application of such techniques in rural environments presents additional challenges due to the absence of such distinct features. We apply a cross view shared transformer (CVST) network to train a robust model capable of handling cross time data in urban environments and challenging rural scenes for daylight navigation. The model is tested with sequential images successfully localising a UAV operating over a challenging rural setting regardless of light variations and dynamic scenes. A comparative study of the transformer network's performance on both rural and urban cross time dataset is presented against state-of-the-art architectures, demonstrating its superiority in the localisation task. A short video demonstrating the localisation system can be found at: <https://youtu.be/-tGkSW8GuOI>