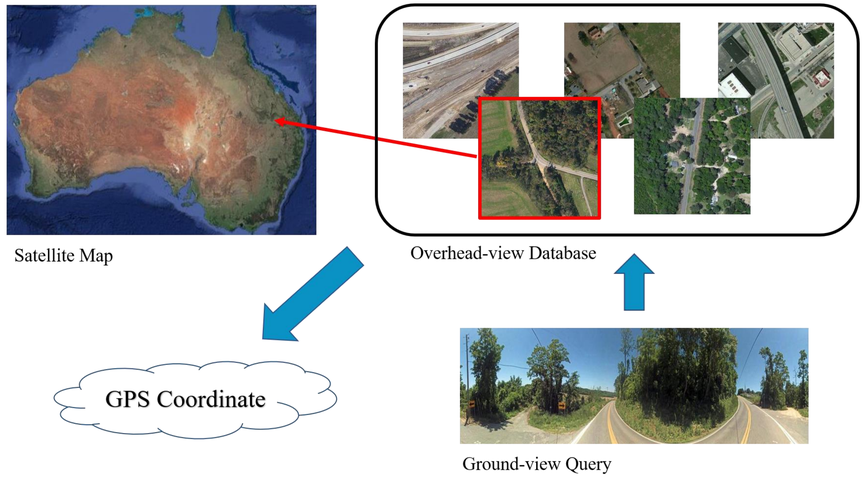
Spatial-Aware Feature Aggregation for Cross-View Image based Geo-Localization

Autralian National University, Australian Centre for Robotic Vision
Yujiao Shi, Liu Liu, Xin Yu, Hongdong Li
{firstname.lastname}@anu.edu.au

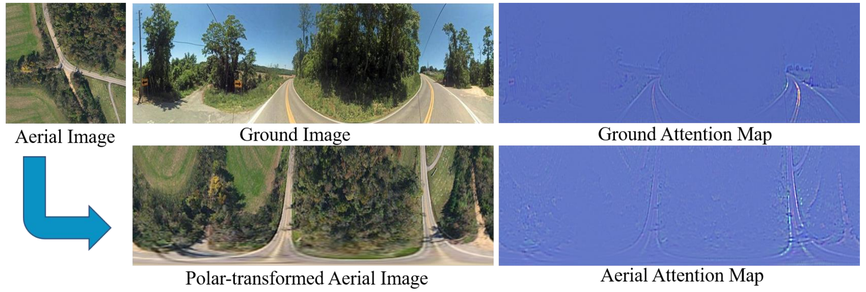
Task Description

- We addresses the problem of cross-view image based localization, where the geographic location of a ground-level street-view query image is estimated by matching it against a large scale aerial map (e.g., a high-resolution satellite image).

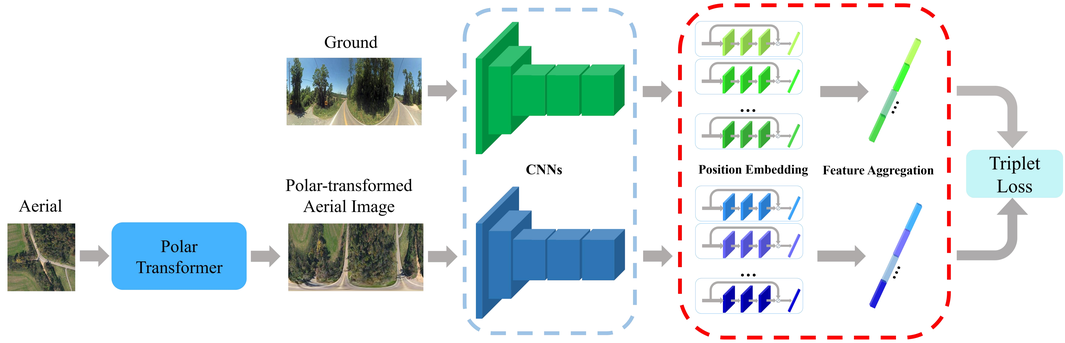


Motivation:

- The first step is to apply a regular polar transform to warp an aerial image such that its domain is closer to that of a ground-view panorama.
- Then, we add a subsequent spatial-attention mechanism which further brings corresponding deep features closer in the embedding space.



Framework:



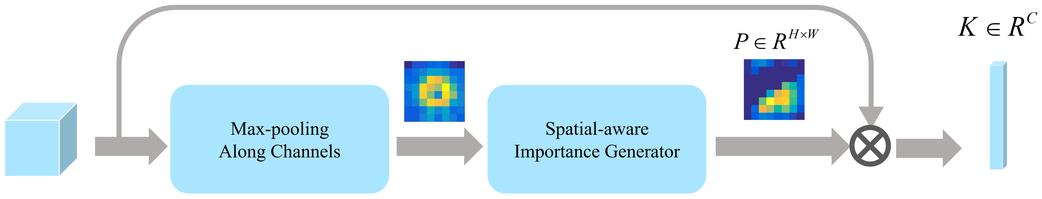
Polar Transform

- We take the center of each aerial image as the polar origin and the north direction as angle 0-degree in the polar transform.

$$x_i^s = \frac{A_a}{2} + \frac{A_a}{2}\frac{y_i^t}{H_g}\sin\left(\frac{2\pi}{W_g}x_i^t\right)$$

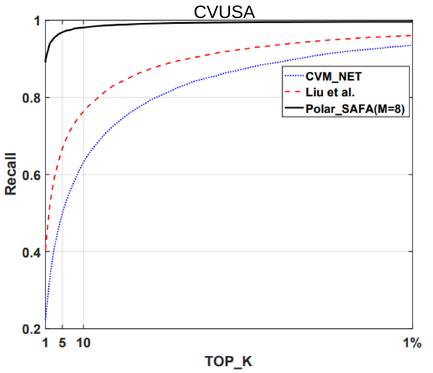
$$y_i^s = \frac{A_a}{2} - \frac{A_a}{2}\frac{y_i^t}{H_g}\cos\left(\frac{2\pi}{W_g}x_i^t\right)$$

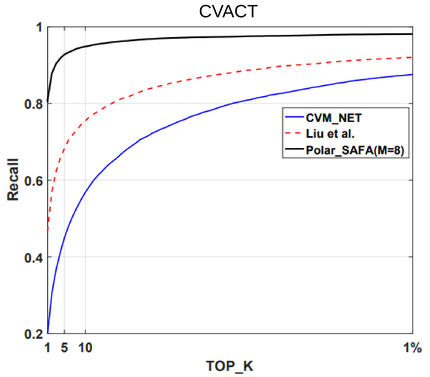
Spatial-aware Position Embedding



Experimental Results:

- Given a ground-level query image, it is regarded as ``successfully localized'' if its ground-truth aerial image is within the nearest top K retrieved images..





Code available ☺