



Figure 1: Localization examples on CVACT_val dataset. The aerial image bordered by green square is the groundtruth.

1 A Additional Qualitative Analysis

2 To better understand the superiority of the EgoTR, in Figure 1, we present some representative
3 cross-view geo-localization results by our EgoTR on CVACT_val dataset. From the localization
4 results, one can find that the EgoTR can properly distinguish the correct match, even among aerial
5 images that are highly similar to each other (*e.g.*, cases 2, 5 and 8). These results demonstrate that
6 the EgoTR can learn discriminative representations for cross-view geo-localization. In the first case,
7 by comparing the top 1 and top 2 returned results, we could also find that the EgoTR can correctly
8 match the groundtruth by corresponding geometric configurations between ground and aerial images.
9 This result shows that the EgoTR is capable of learning position-aware representations for cross-view
10 geo-localization. Furthermore, our EgoTR is versatile for a wide range of test scenarios, such as
11 traffic intersections (*e.g.*, case 7) and suburban areas (*e.g.*, case 6), which manifests the broad practical
12 applicability of our model.