Response to Reviewer 2 Comments

Thank you a lot for your careful reading and helpful comments on our manuscript. In this revision, we have addressed all of these comments/suggestions. We highlighted all the revisions in red/blue color. Below are our point-to-point responses to your questions. We tried our best to improve the manuscript and made changes in the manuscript. We appreciate for Editors/Reviewers' careful review and hope for your favorable consideration regarding this manuscript.

Point 1: Figure 2 doesn't contain (Ik and street-view images Ip as inputs) please write I_k and I_p on the figure.

Response 1: We have corrected this error.

Point 2: How to share the weights, why, and what is the time complexity to achieve this task?

Response 2: Shared-weights is a common technique used in contrastive learning, aimed at enabling the network to adapt to the effects of data augmentation. During training, we include both strongly augmented and weakly augmented images in the same batch and feed them into the same network to achieve shared-weights. In this field, we focus on the computational cost of the model, which is mentioned in Section 4.5.3.

Point 3: What is the mean of project head

Response 3: A project head consists of two fully connected layers. It is a common technique used in contrastive learning and we introduced it in Section 2.3.

Point 4: Figure 2 is not logic how the output is loss function?.

Response 4: Its intention is to express which features are used to compute the loss function. We have already made modifications in the figure to avoid ambiguity.

Point 5: I think the results are good but authors can add Statistical Analysis such Friedman test.

Response 5: The Friedman test, a statistical analysis method, may better showcase the performance of our method. However, in the task of cross-view geo-localization, it is often more expressive to evaluate the model's performance through comparative experiments and ablation studies.