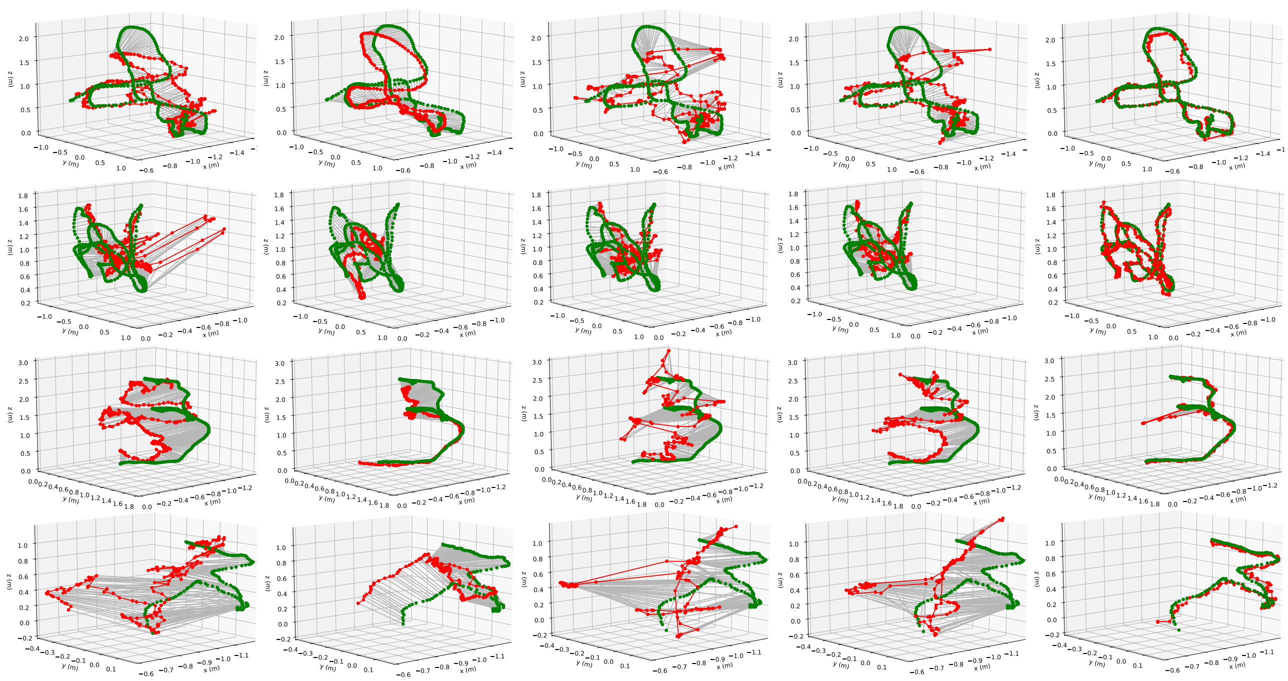
CamNet: Coarse-to-Fine Retrieval for Camera Re-Localization
– Supplementary Material –

This supplementary material provides more comparisons that show our model (CamNet) outperforms the state-of-the-art methods. Figure 1 and Figure 2 show the results for all the testing sequences on the 7-Scenes dataset.

Moreover, the activation of the feature map are visualized as Figure 3, which demonstrate that our model is more robust and mitigates the over-fitting problem. We compute the activated value of the last feature map (7 × 7) for all methods. The results show: (1) The absolute-based pose regression method has a obvious over-fitting problem compared to the relative-based pose regression method. (2) Compared to NN-Net [4] and RelocNet [1], our model focuses on a wider and more accurate area, and mitigates the over-fitting problem. (3) The high activation area of our model is concentrated on static and highly discriminating objects, which means more accurate attention and higher localization accuracy.

BFMatcher (Brute-force matcher): For each descriptor of ORB key-point in the first image, Brute-force matcher is used to find the closest matching from the descriptors of the second image. The L2 distance is used as the distance metric. Figure 4 shows a schematic diagram of BFmatcher.



PoseNet [3] | MapNet++ [2] | NN-Net [4] | RelocNet [1] | Ours

Figure 1: Camera localization results on the 7-Scenes dataset. For each 3D plot, the green lines show the ground truth camera trajectories and the red lines show the camera pose predictions. Sequences (from top to bottom): office-07-fig, office-09-fig, redkitchen-04-fig, stairs-04-fig.

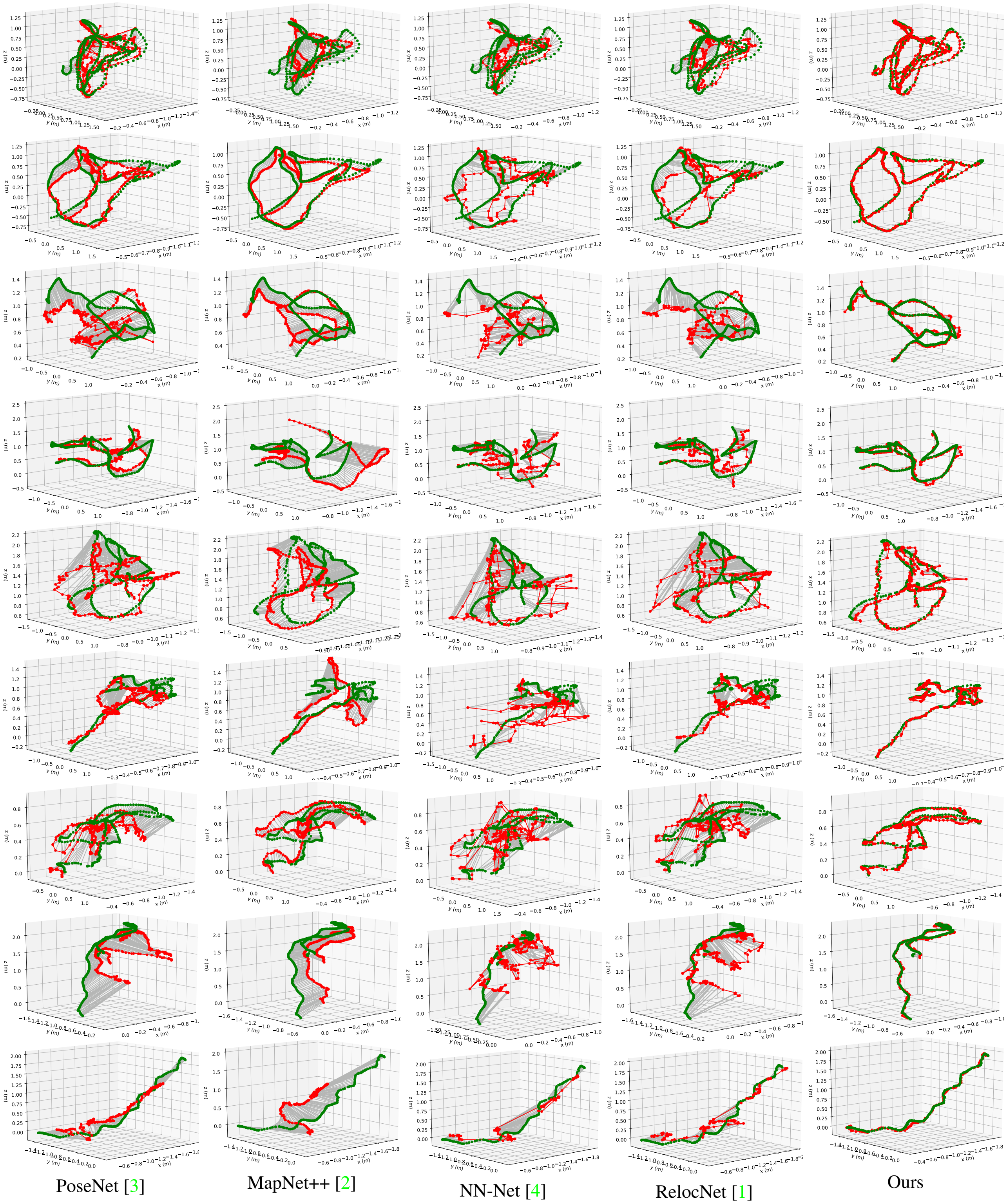


Figure 2: Camera localization results on the 7-Scenes dataset. For each 3D plot, the green lines show the ground truth camera trajectories and the red lines show the camera pose predictions. Sequences (from top to bottom): chess-03-fig, chess-05-fig, fire-03-fig, office-02-fig, office-06-fig, pumpkin-01-fig, pumpkin-07-fig, redkitchen-03-fig, stairs-01-fig.

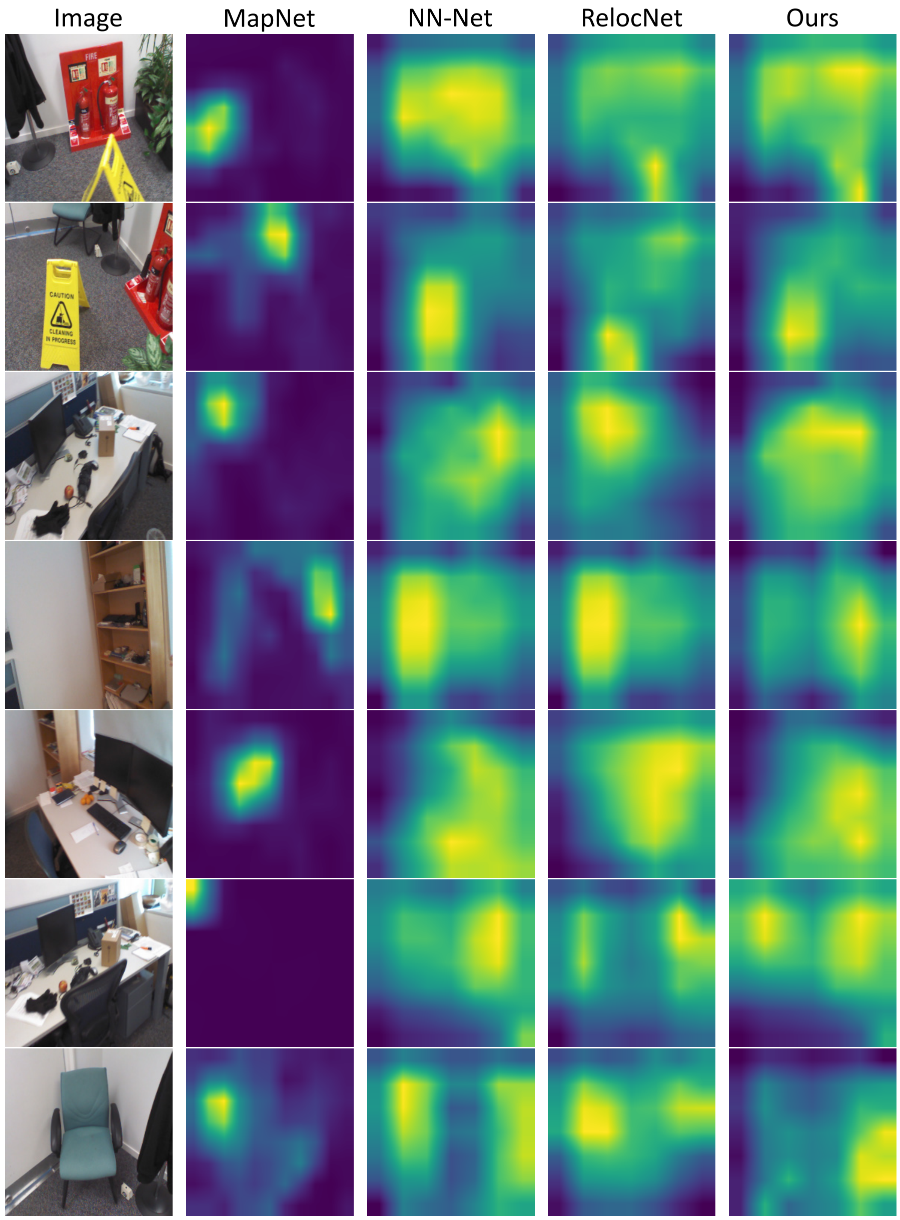


Figure 3: Visualization of the feature map activation of four methods: MapNet [2], NN-Net [4], RelocNet [1], Our CamNet. The activated areas of MapNet [2] are not concentrated on the correct objects for localization (such as the fire extinguisher and warning signs in the first two figures, the computer and chair in the last two figures). Compared to NNnet [4] and RelocNet [1], our model focuses on a wider and more accurate area, means the over-fitting problem is mitigated by our model.

References

[1] V. Balntas, S. Li, and V. Prisacariu. Relocnet: Continuous metric learning relocalisation using neural nets. In *The European Conference on Computer Vision (ECCV)*, September 2018. 1, 2, 3

[2] S. Brahmbhatt, J. Gu, K. Kim, J. Hays, and J. Kautz. Geometry-aware learning of maps for camera localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2616–2625, 2018. 1, 2, 3

[3] A. Kendall, M. Grimes, and R. Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015. 1, 2

Figure 4: First 10 matches of BFMatcher.

[4] Z. Laskar, I. Melekhov, S. Kalia, and J. Kannala. Camera relocalization by computing pairwise relative poses using convolutional neural network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 929–938, 2017. 1, 2, 3