MeshVPR: Citywide Visual Place Recognition Using 3D Meshes - Supplementary material

Gabriele Berton[1], Lorenz Junglas[2], Riccardo Zaccone[1], Thomas Pollok[3], Barbara Caputo[1], and Carlo Masone[1]

[1] Politecnico di Torino
[2] Karlsruhe Institute of Technology
[3] Fraunhofer IOSB

1 Examples of training and test images

In this Supplementary we present a large number of randomly chosen images from each of the training and test sets, shown in Fig. 1 (for the San Francisco datasets) and Fig. 2 (for the datasets from Berlin, Paris and Melbourne). These images allow to visually assess the large distribution gap between the various datasets, and, most importantly, between the synthetic and real images.

2 Further Qualitative Results and Failure Cases

Figure 3 shows examples of randomly chosen predictions on synthetic datasets for queries from Berlin, Paris and Melbourne, computed with the best model (SALAD + MeshVPR). In many cases it is visible how the two models (synthetic and real), given two images that are semantically similar but visually different, learn to map them nearby in the features space. For example the trees from the synthetic and real domains look very different, but when a query contains a tree, the predictions often contain a "synthetic" tree. A similar pattern is visible for statues (see examples from Paris), which are hardly recognizable in the synthetic domain, but are mapped in the same features space by the two models.

Fig. 1: Examples of images from the datasets from San Francisco, namely the real database, the High Quality (HQ) synthetic database, the Low Quality (LQ) synthetic database and the queries.

Fig. 2: Examples of synthetic database and real queries from the datasets of Berlin, Paris and Melbourne.

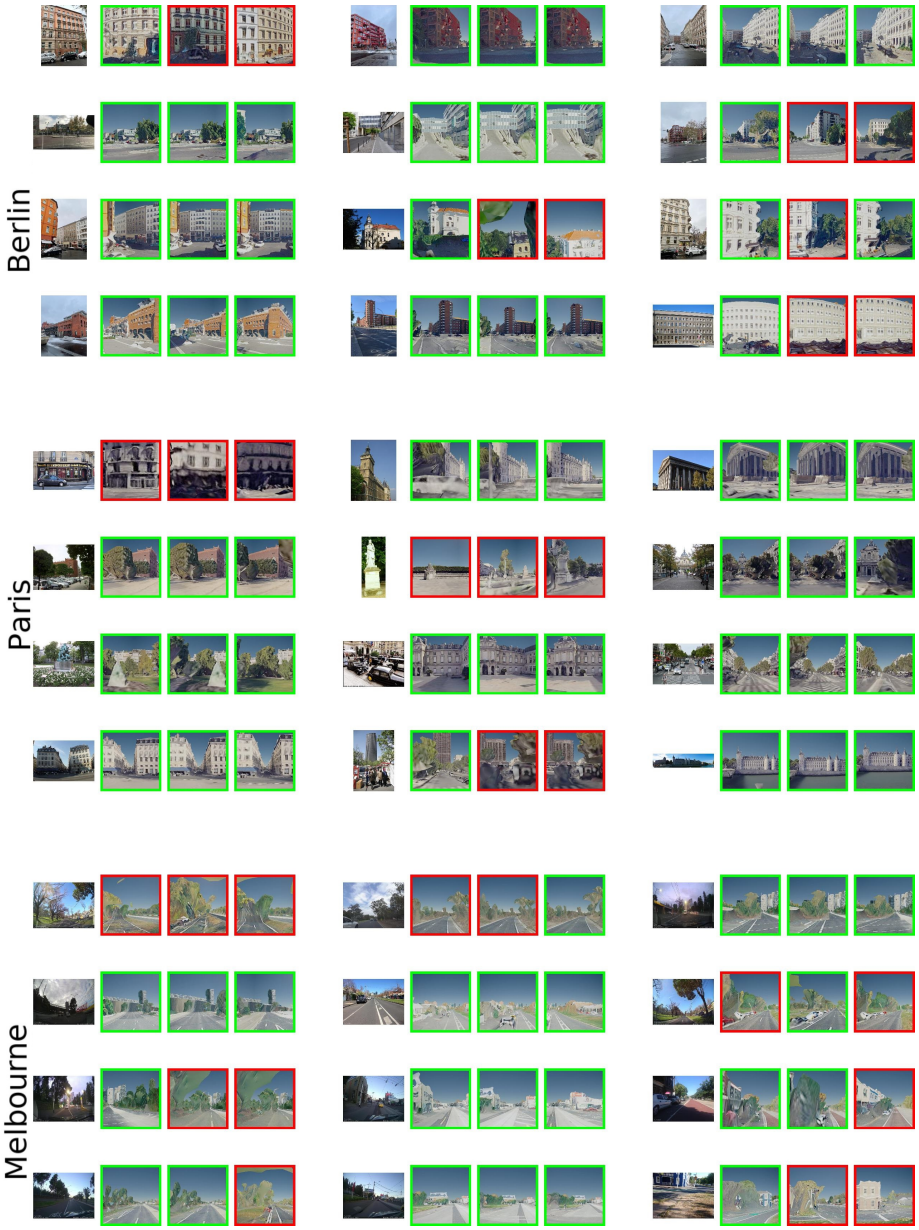

Fig. 3: Qualitative results from the three test sets, randomly picked, computed with the best model (SALAD + MeshVPR).