Abstract:

We consider the geo-localization task - finding the pose (position & orientation) of a camera in a large 3D scene from a single image. We aim to experimentally explore the role of geometry in geo-localization in a convolutional neural network (CNN) solution. We do so by ignoring the often available texture of the scene. We therefore deliberately avoid using texture or rich geometric details and use images projected from a simple 3D model of a city, which we term lean images. Lean images contain solely information that relates to the geometry of the area viewed (edges, faces, or relative depth). We find that the network is capable of estimating the camera pose from the lean images, and it does so not by memorization but by some measure of geometric learning of the geographical area. The main contributions of this work are: (i) providing insight into the role of geometry in the CNN learning process; and (ii) demonstrating the power of CNNs for recovering camera pose using lean images.