

In this paper, we address the complex problem of multi-label pixelwise classification. We present our distinct solution based on a convolutional neural network (CNN) for performing multi-label pixelwise classification and its application to large-scale urban reconstruction. A supervised learning approach is followed for training a 13-layer CNN using both LiDAR and satellite images. An empirical study has been conducted to determine the hyperparameters which result in the optimal performance of the CNN. Scale invariancy is introduced by training the network on six different scales of the input and labeled data. This results in six pixelwise classifications for each different scale. Finally, an SVM maps the six pixelwise classifications into a single-label. The pixelwise classification is then used to accurately extract and reconstruct the buildings in large-scale urban areas. The proposed approach has been extensively tested and the results are reported in comparison tables and realistic 3D visual renderings.