Supplemental Material : A Unified Model for Near and Remote Sensing

Scott Workman[1] scott@cs.uky.edu
Menghua Zhai[1] ted@cs.uky.edu
David J. Crandall[2] djcran@indiana.edu
Nathan Jacobs[1] jacobs@cs.uky.edu
[1]University of Kentucky [2]Indiana University Bloomington

This document contains additional details and experiments related to our methods.

1. Brooklyn and Queens Dataset

Figure 1 shows the spatial coverage of the Brooklyn and Queens regions in our dataset. Figure 2 visualizes the label distributions for the Brooklyn and Queens test sets. Compared to Brooklyn, Queens has significantly different label occurrence. For example, for land use classification, Brooklyn has more "Public Buildings", while Queens has more "Open Space/Recreation".

2. Adaptive Bandwidth Visualization

In Figure 3 we visualize the estimated kernel bandwidth parameters, computed using our *unified (adaptive)* method for the task of land use classification, as a map for the Brooklyn and Queens regions. For each location, we display the mean of the diagonal entries of the kernel bandwidth matrix, Σ. These results show that the adaptive method is adjusting based on the underlying terrain.

3. Semantic Segmentation Results

Figure 4 shows confusion matrices for all three labeling tasks we consider (land use, age, function), each computed using the *unified (adaptive)* approach, for the Brooklyn test set. For building function estimation, we aggregate the 206 building classes into 30 higher-level classes. Classes are merged according to a hierarchy outlined by the New York City Department of City Planning in the PLUTO dataset. Despite the challenging nature of these tasks, our method seems to make sensible mistakes. For example, for the task of estimating building age, nearby decades are most often confused.

We report performance, top-1 accuracy and mean region intersection over union (mIOU), for building function estimation after aggregating the classes. For *unified (adaptive)*, on the Brooklyn test set, top-1 accuracy increases to 61.08% and mIOU increases to 30.40%. Similarly for Queens, top-1 accuracy increases to 52.01% and mIOU increases to 14.99%.

In our experiments, we considered the $N = 20$ closest ground-level images, chosen empirically based on available computational resources. Theoretically, there is no downside to including as many ground-level images as possible. However, we explored at what point performance might saturate. We performed this experiment for land use classification, using our *unified (adaptive)* approach, varying N in increments of 5 up to 25, and found that performance saturated at $N = 15$, but this was just one dataset/task. Figure 5 visualizes the results of this experiment using top-1 accuracy.

For each labeling task, we show additional semantic labeling results in Figure 6.

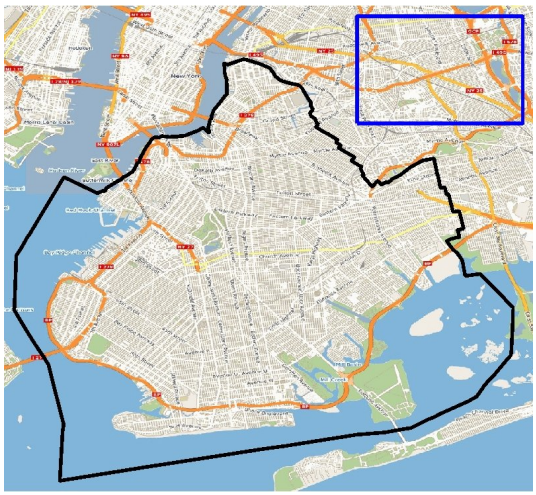
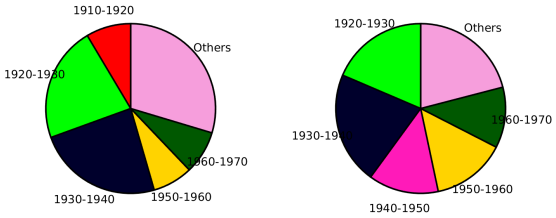
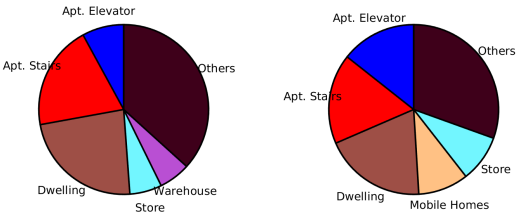


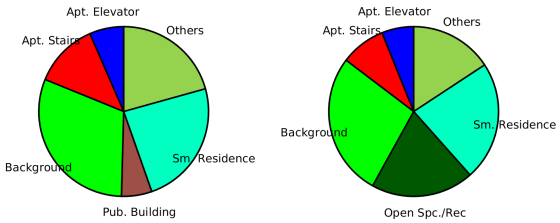
Figure 1: A coverage map for the Brooklyn (black) and Queens (blue) regions in our dataset.



(a) Age

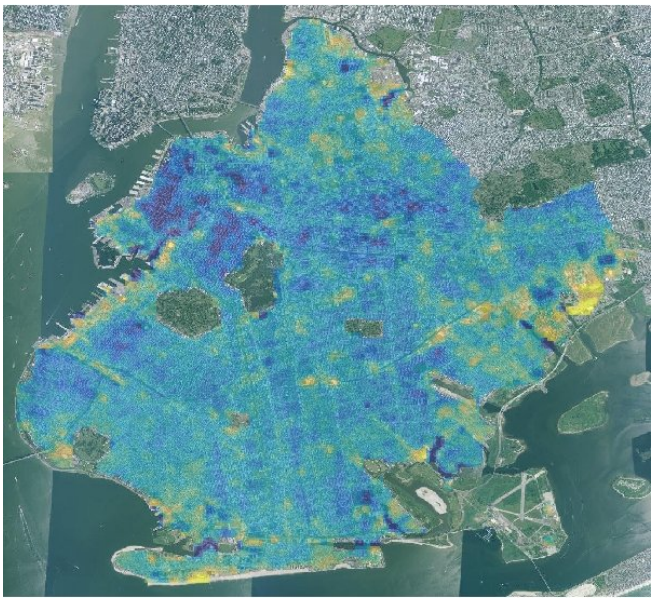


(b) Function

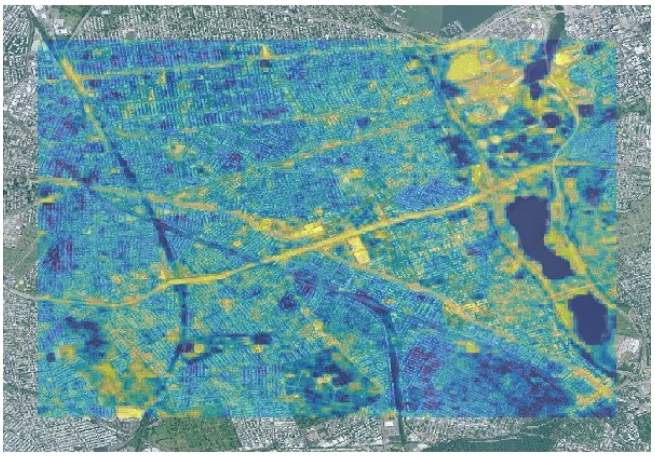


(c) Land Use

Figure 2: Distribution of labels for the Brooklyn (left) and Queens (right) test sets.



(a) Brooklyn



(b) Queens

Figure 3: Adaptive kernel bandwidth estimation. For each location we show the mean of the estimated optimal kernel bandwidth parameters, for the task of land use classification, computed using the *unified (adaptive)* method.

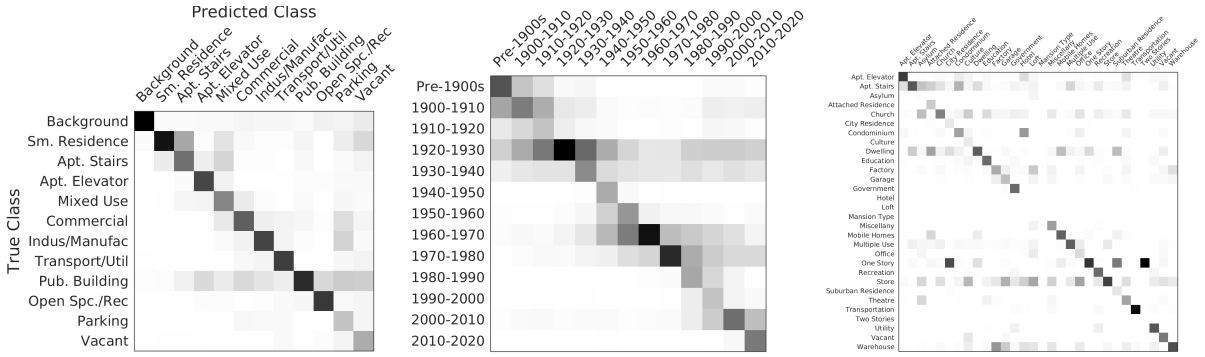


Figure 4: Confusion matrices for classifying land use (left), estimating building age (middle), and identifying building function (right). These results were computed using our *unified (adaptive)* approach for the Brooklyn test region.

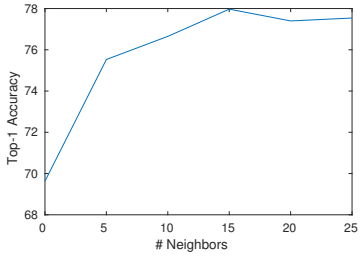


Figure 5: Varying the number of nearby ground-level images (land use classification). Each point corresponds to an instance of our *unified (adaptive)* method, except $N = 0$ which reflects the performance of the *remote* baseline.

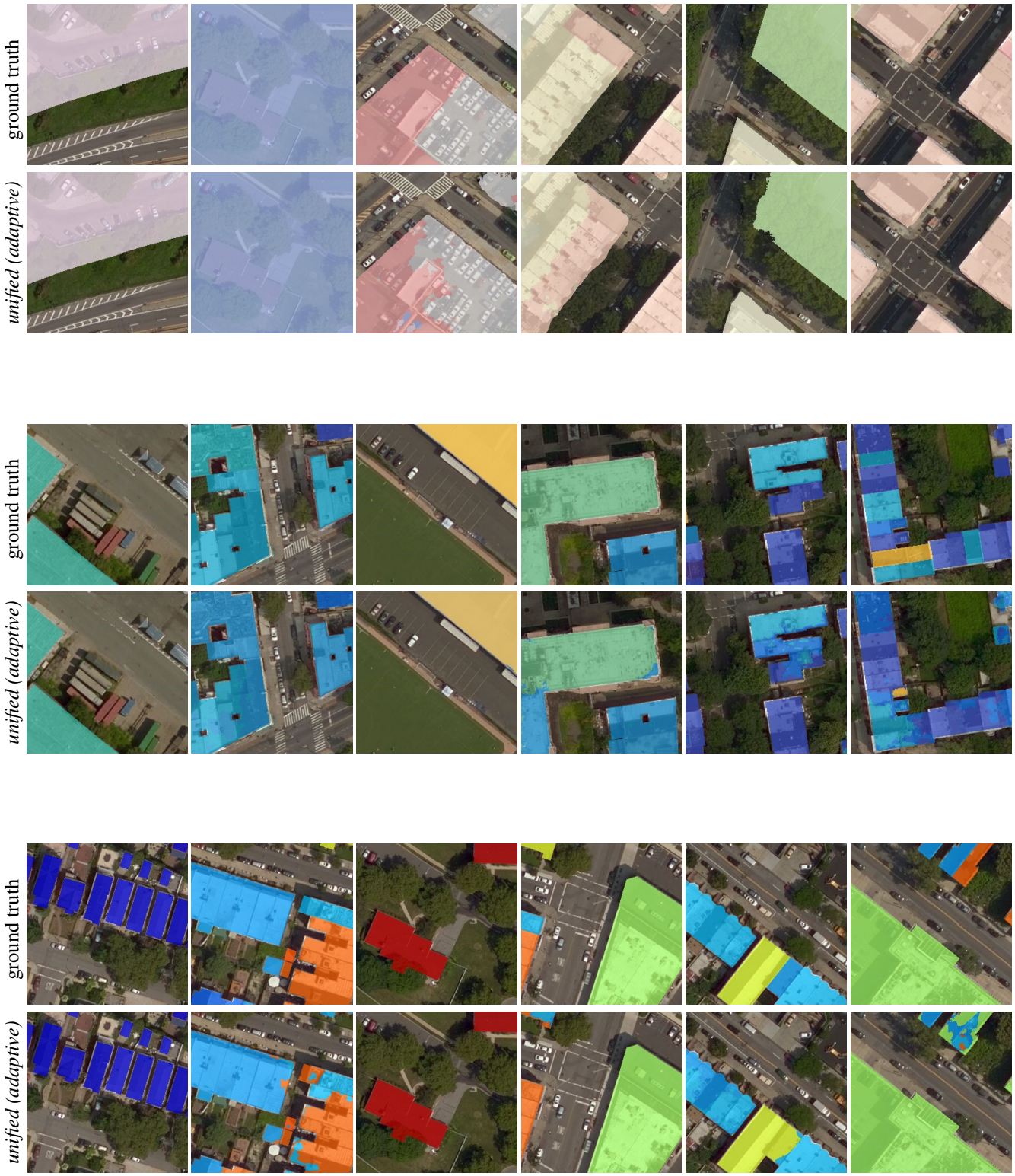


Figure 6: Additional semantic labeling results for classifying land use (top), estimating build age (middle) and identifying building function (bottom).