View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by INRIA a CCSD electronic archive server

Supplementary materials for: Aligning and Updating Cadaster Maps with Aerial Images by Multi-Task, Multi-Resolution Deep Learning

Nicolas Girard[1], Guillaume Charpiat[2], and Yuliya Tarabalka[1]

[1] TITANE team, INRIA, Université Côte d'Azur, France
[2] TAU team, INRIA, LRI, Université Paris-Sud, France
firstname.lastname@inria.fr

1 Generalization across dataset

We performed another experiment to check the generalization across datasets of our method. We used a third dataset for this: the Mapping Challenge dataset from CrowdAI. It has 300×300 px images of buildings, 280741 for training and 60317 for validation. We first trained the networks on all training samples from the 3 datasets: Bradbury et al. [1], Mapping Challenge and Inria [8] datasets. We secondly trained on training samples from Bradbury and Mapping Challenge datasets only, excluding the Inria dataset entirely. We finally tested those networks on 44 images (each 5000×5000 px) from the Inria dataset, which constitutes our testing samples. 21 of those images are from Austin and 23 from Chicago as those were the areas for which the OSM ground truth is the most precise. See in the Fig. 1 the accuracy plots for these 2 experiments.

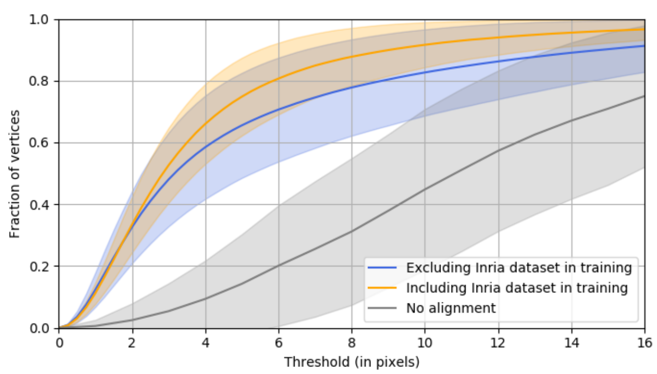

Fig. 1: Alignment accuracies. Solid curves are the average, shaded regions are within the standard deviation.

We lose 8.3% relative mean accuracy in terms of area under the average curve when excluding the Inria dataset from training. That is, we measure the area under the 2 solid curves (representing the average accuracy) of Fig. 1 and compare their relative value. We do not loose mean accuracy until threshold 2 px, we lose 4% mean accuracy for threshold 3 px and 10% for threshold 6 px. There is almost no difference for well-aligned vertices: in both cases, about half of the vertices have an error of less than 3 px. The loss of mean accuracy mainly happens between threshold 3 px and 6 px, where it drops by $10 - 4 = 6\%$. The difference between the 2 curves has to be considered relatively to the grey curve (no alignment): we can see that the blue curve is much closer to the orange one, showing that most of the alignment accuracy is recovered by the networks trained on different datasets.

2 Use of the regularization term from Zampieri et al.

During our experiments we tried to use the regularization term from Zampieri et al., but we did not see any improvement with it. One of the reasons is that real-case displacements are not very smooth: buildings of different heights have different displacements if the image angle is a cause of the displacement. Furthermore, some building footprints come from a different source compared to surrounding buildings and thus have an uncorrelated displacement. Our multi-task learning and intermediary losses guide the optimization at the start of training, so that it works well without regularization. We however observed that the regularization term naturally decreases without explicitly optimizing it.

We might consider in the future to try better-suited regularization criteria. For example a criterion that encourages smoothness within polygons but does not discourage large variations of displacement across polygon boundaries. To design such a criterion, inspiration could be taken from the BV (Bounded Variation) norm or the Mumford-Shah functional, which are piece-wise smooth regularizers.

3 Splitting of the 2 datasets into train, validation and test sets

Here we detail the splitting of the Inria [8] and the Bradbury et al. [1] datasets we used for the experiments in the paper. As relatively few images have perfect ground-truth data, we made sure the validation and test sets are composed only of those good images while also putting some in the train set. Tables 1, 2 and 3 specify the exact dataset split we used.

The city of San Francisco appears in both the train and test sets however the images do not have the same capture conditions, ensuring enough dissimilarity:

- They were taken at different times (some buildings appear in the images from Bradbury compared to Inria)
- The capture angle is different (we can see building facades on Inria images because of the angle, which is not the case for the Bradbury images)

- The sun angle is very different, resulting in different shadow angles and lengths
- Colors are very different as well
- Finally the Bradbury images have a lot of noise, differing them further from the Inria images

Table 1: Train split

Dataset	City	Image numbers
Bradbury	Arlington	3
Bradbury	Atlanta	1, 2, 3
Bradbury	Austin	1, 2, 3
Bradbury	NewYork	2
Inria	Kitsap	1 to 24 and 26 to 36
Inria	Austin	2 to 9, 11 to 19 and 21 to 36
Inria	Chicago	2 to 9, 11 to 19 and 21 to 36
Inria	Tyrol West	2 to 9, 11 to 19 and 21 to 36
Inria	Vienna	2 to 9, 11 to 19 and 21 to 36
Inria	Tyrol East	2 to 9, 11 to 19 and 21 to 36
Inria	San Francisco	2 to 9, 11 to 19 and 21 to 36
Inria	Innsbruck	2 to 9, 11 to 19 and 21 to 36
Inria	Bloomington	2 to 9, 11 to 19 and 21 to 36
Inria	Bellingham	12 to 9, 11 to 19 and 21 to 36

Table 2: Validation split

Dataset	City	Image numbers
Bradbury	Norfolk	1, 2, 3
Inria	Austin	1, 10, 20
Inria	Chicago	1, 20
Inria	Tyrol West	1, 10, 20
Inria	Vienna	1, 10, 20
Inria	Tyrol East	1, 10, 20
Inria	San Francisco	1, 10, 20
Inria	Innsbruck	1, 10, 20
Inria	Bloomington	1, 10, 20
Inria	Bellingham	1, 10, 20

Table 3: Test split

Dataset	City	Image numbers
Bradbury	San Francisco	1, 2, 3