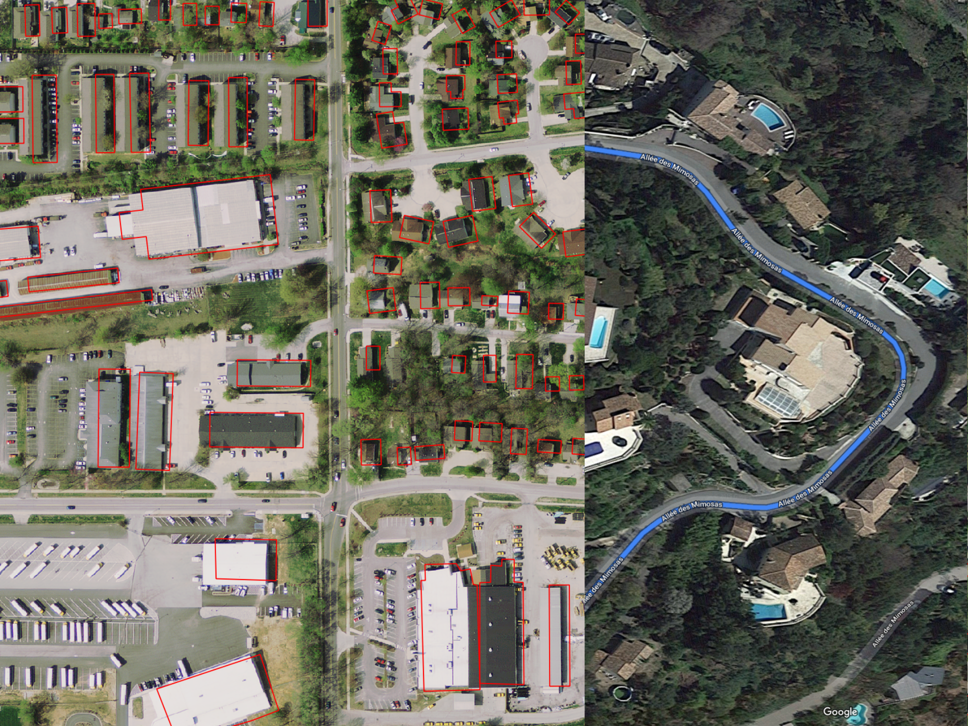
Noisy Supervision for Correcting Misaligned Maps Without Perfect Groundtruth

Nicolas Girard[1], Guillaume Charpiat[2] and Yuliya Tarabalka[1,3]

[1]TITANE team, INRIA, Université Côte d'Azur, France
[2]TAU team, INRIA, LRI, Université Paris-Sud, France
[3]LuxCarta Technology, France
email: nicolas.girard@inria.fr

30th of July, 2019


Allée des Mimosas
Allée des Mimosas
Allée des Mimosas
Allée des Mimosas
Allée des Mimosas
Allée des Mimosas
Allée des Mimosas
Allée des Mimosas
Allée des Mimosas
Google

- Misalignment between image and map (OpenStreetMap and Google Maps)
- Up to 8m (27px here)
- Not constant across the image, with possible building rotations

Causes of misalignments

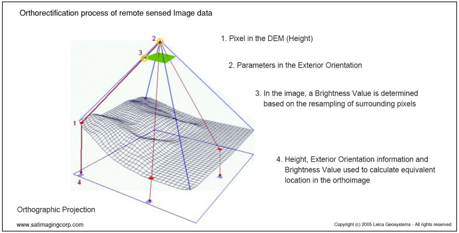


Image orthorectification errors:

- Induced by errors in the Digital Elevation Model[1]
- Elevation mean error: 1m

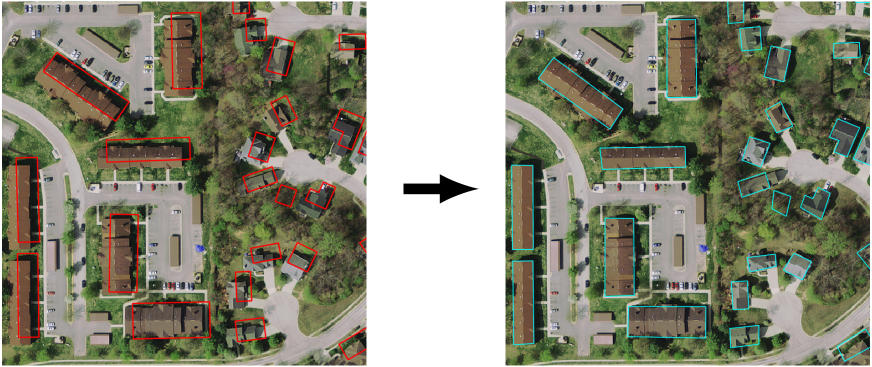
Map errors:

- Human error when drawing maps[2]
- Unprecise source material (scanned map from local authorities)

[1] J. A. Thompson et al., Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling (2001)

[2] J. K. Wright et al., Map Makers Are Human: Comments on the Subjective in Maps (1942)

Objective



- Correct existing maps so that they become useful
- Other machine learning tasks (ex: image segmentation) will be able to use that new perfect groundtruth

Image alignment (or registration) previous works

Monomodal (ex: RGB image to RGB image):

- Keypoint matching (SIFT, HOG)
- FlowNet uses a CNN to predict optical flow [3]
- Quicksilver learns an image similarity measure directly from image appearance[4]

Multimodal (ex: polygon map to RGB image):

- Structural similarity measure [5]
- Double input U-Net-like model to predict large displacements in a multi-resolution approach[6]

None of those machine learning methods learn from misaligned data

[3] P. Fischer et al., FlowNet: Learning Optical Flow with Convolutional Networks (2015)

[4] X. Yang et al., Quicksilver: Fast Predictive Image Registration - a Deep Learning Approach (2017)

[5] Y. Ye et al., Robust Registration of Multimodal Remote Sensing Images Based on Structural Similarity (2017)

[6] A. Zampieri, G. Charpiat, N. Girard and Y. Tarabalka, Coarse to fine non-rigid registration: a chain of scale-specific neural networks for multimodal image alignment with application to remote sensing (2018)

Motivation for noisy-supervision

Context

- If enough good groundtruth data available: apply fully supervised machine learning methods
- Problem: remote sensing has few data with perfect manually-curated annotations
- OpenStreetMap: huge public resource of annotations
- However those annotations can be noisy (misalignments)

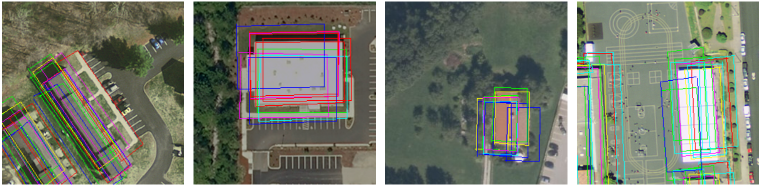
Objective

- Get perfectly aligned annotations

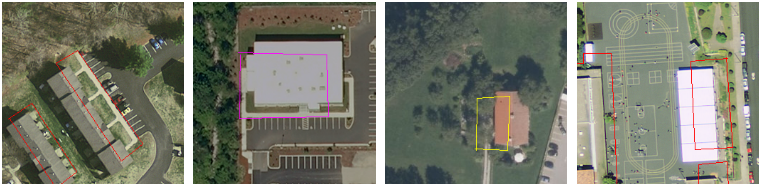
Solution

- Learn from available annotations, even if they are noisy (misaligned)
- Use a multiple-rounds training scheme to correct the ground truth annotations at each round to better train the model at the next round

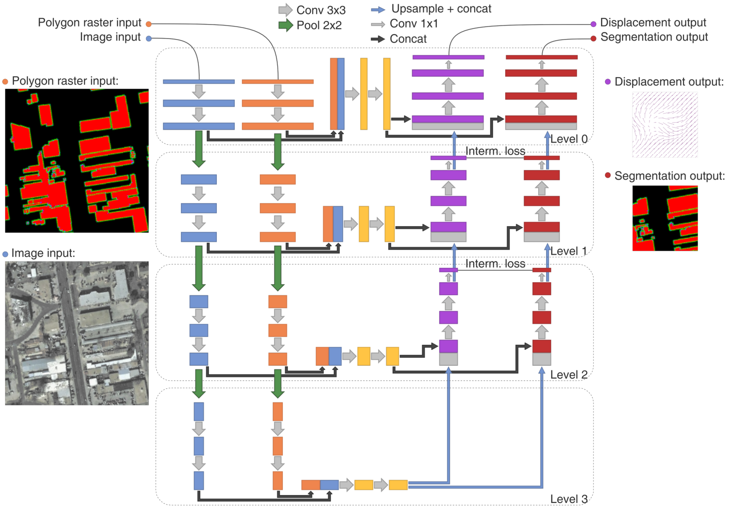
How can noisy annotations be useful?



Ground truth with i.i.d. noise (above) $\rightarrow$ model optimized with L2 loss $\rightarrow$ average predictions $\rightarrow$ corresponds to perfect underlying GT.
However in our case, only one sample of the noise is available:

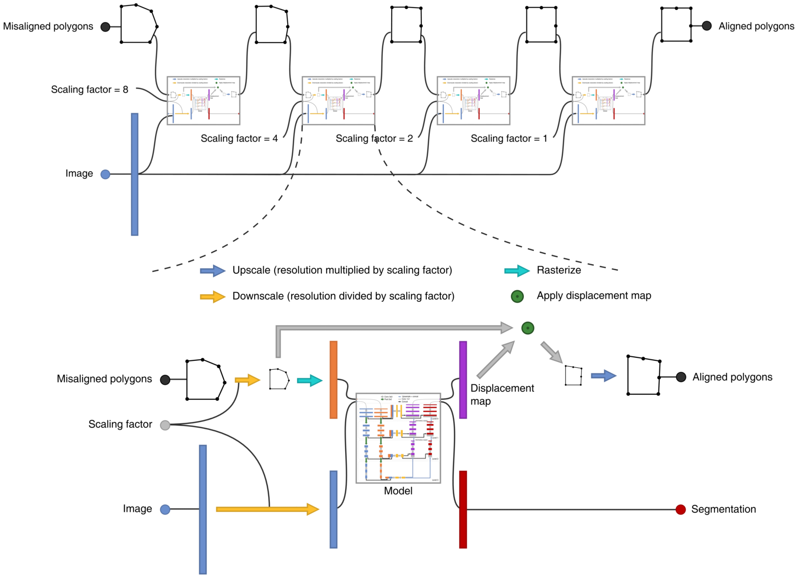


Neural network with double inputs and outputs



[6] N. Girard et al., Aligning and Updating Cadaster Maps with Aerial Images by Multi-Task, Multi-Resolution Deep Learning (2018)

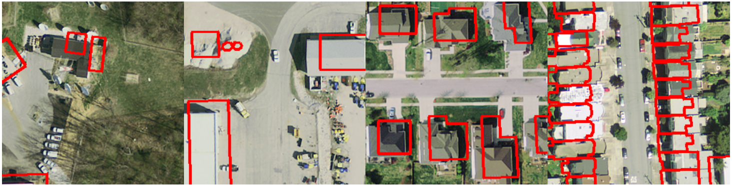
Multi-resolution



Pipeline animation

Animation of the alignment pipeline

Dataset:



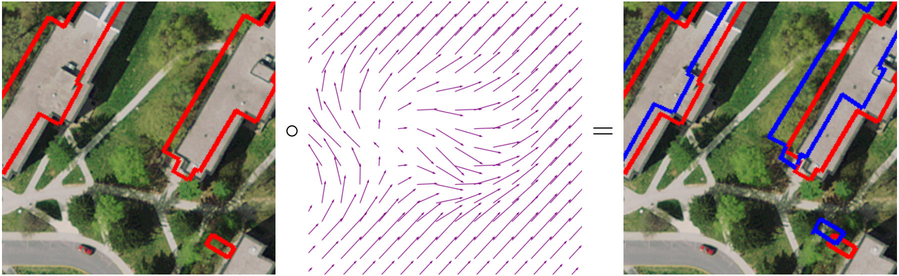
Images from the “Inria” dataset[7] and “Bradbury” dataset [8]:

- 386 images of 5000×5000 px
- 16 cities from Europe and the U.S.
- Each image has in average a few thousand buildings
- Groundtruth building polygons from OpenStreetMap

[8] E. Maggiori et al., Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark (2017)

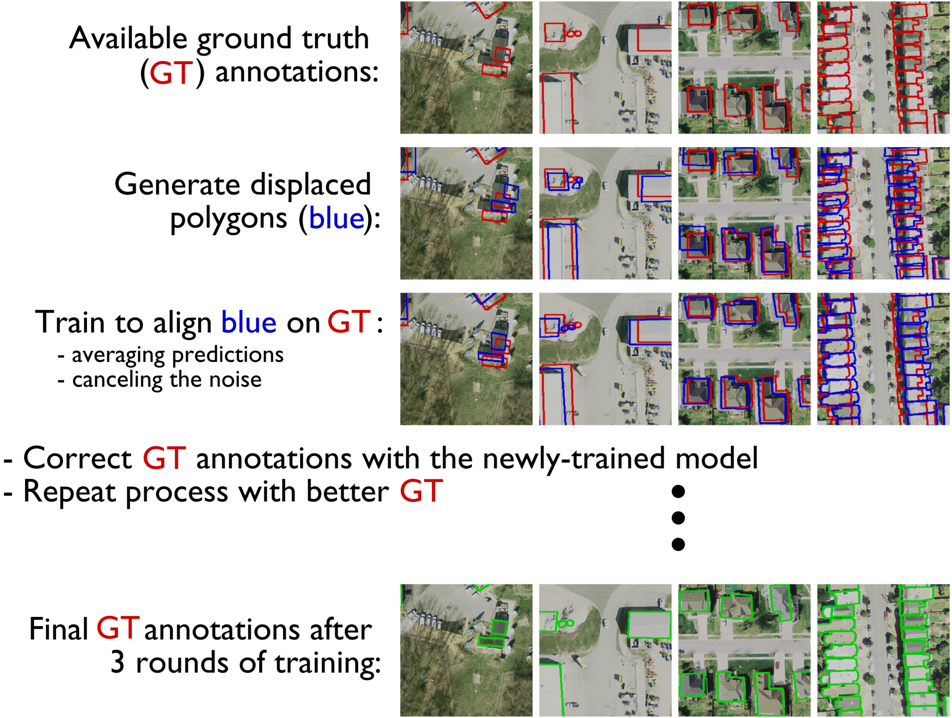
[8] K. Bradbury et al., Aerial imagery object identification dataset for building and road detection, and building height estimation (2016)

Groundtruth generation

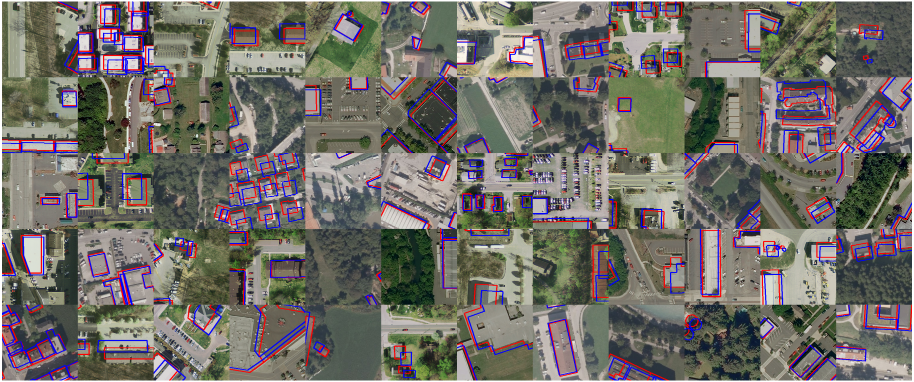


Displacement map generation:

- Random 2D Gaussian fields
- Max displacement of 32 px
- Applied to ground truth polygons to generate misaligned polygons

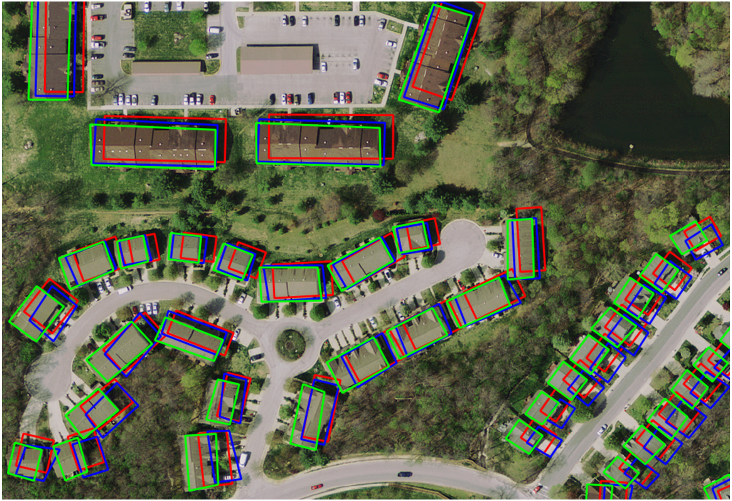

Available ground truth (GT) annotations:
Generate displaced polygons (blue):
Train to align blue on GT :
- averaging predictions
- canceling the noise
- Correct GT annotations with the newly-trained model
- Repeat process with better GT
Final GT annotations after 3 rounds of training:

Training



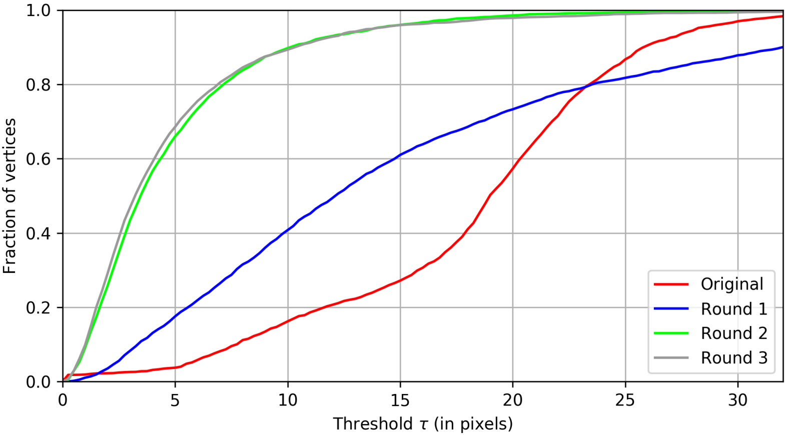
Video of the training process

Qualitative results



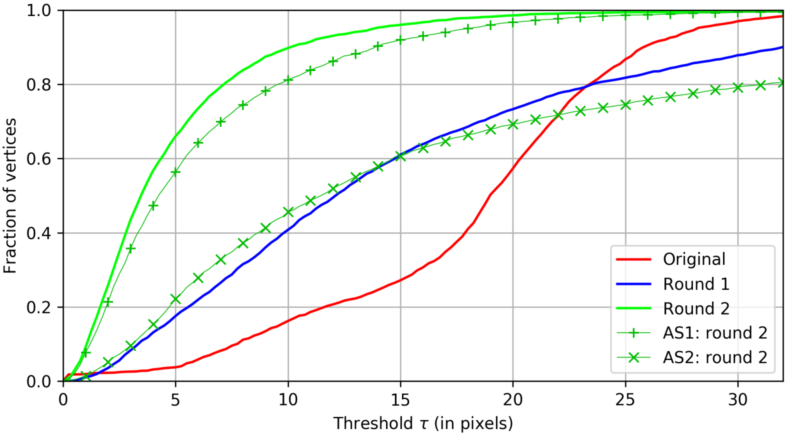
(Red: initial dataset annotations; blue: aligned annotations round 1; green: aligned annotations round 2.)

Quantitative results



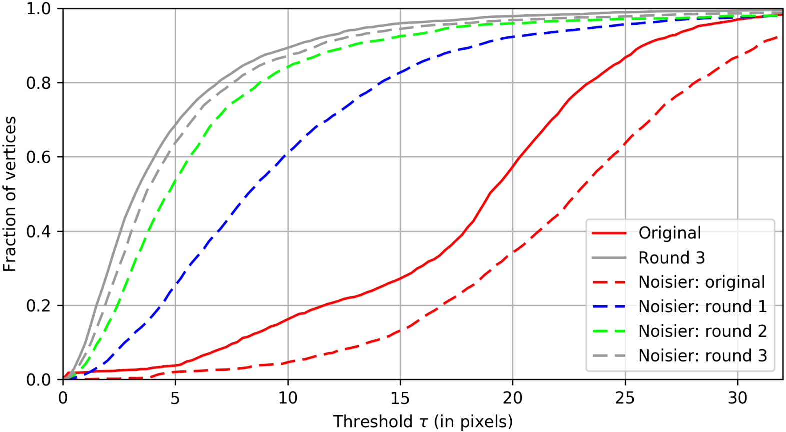
(Accuracy cumulative distributions measured with the manually-aligned annotations of bloomington22 from the Inria dataset.)

Ablation studies



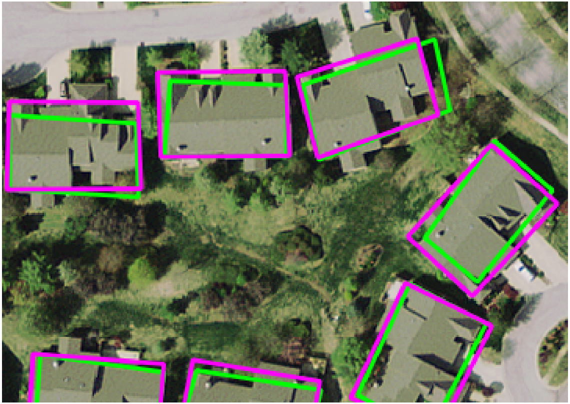
AS1: align annotations from previous round instead of the original annotations
AS2: train the model once on the original annotations and apply it multiple times to iteratively align annotations

Noisier original annotations

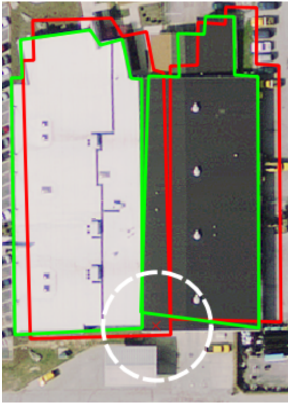


Noisier: added more noise (16 px amplitude) to the original annotations

Sources of errors



Ambiguity of the perfect ground truth annotations (up to 20px difference)



Failure case

Magenta: manually aligned annotations; red: original dataset annotations; green: aligned annotations round 2.

Concluding remarks

Contributions

- Multi-task, multi-resolution map alignment model
- Multiple-rounds training scheme to iteratively train a better model
- We also established theoretical tools for understanding neural networks better[9]

---

[9] G. Charpiat, N. Girard, L. Felardos and Y. Tarabalka: Input similarity from the neural network perspective, pre-print on personal web page, 2019

Future works

- Piece-wise smooth displacements generation method
- Use corrected groundtruth to train a better image segmentation model
- Use the theoretical tools we developed to study the case of noisy-supervision on the alignment task

Thank you for your attention !

Code available on GitHub (github.com/Lydorn/mapalignment):

Any questions ?