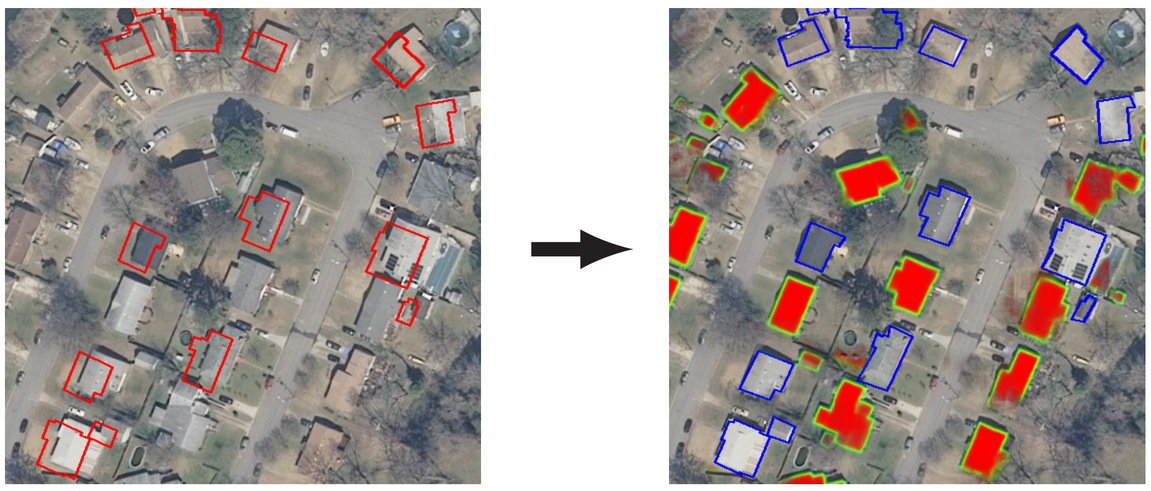
Aligning and Updating Cadaster Maps with Aerial Images by Multi-Task, Multi-Resolution Deep Learning

Nicolas Girard (nicolas.girard@inria.fr)[1] Guillaume Charpiat[2] Yuliya Tarabalka[1]
[1] TITANE team, INRIA, Université Côte d'Azur. [2] TAU team, INRIA Saclay, LRI, Université Paris-Sud.

ACCV 2018
2 – 6 December 2018 Perth Western Australia

I - Introduction

- **Goal**: joint **multi-modal alignment** and **semantic segmentation**



- **Data**:
  - **optical data:** RGB/multi-spectral images from satellite/airplane
  - **binary cadaster maps:** building rooftops as polygons
- **Why?**: existing cadaster maps can be **misaligned** and **outdated** (missing new buildings), which makes **segmentation algorithms** difficult to train. Causes of misalignment:
  - **different angles of capture**, making rooftops move (even on orthorectified images, as Digital Terrain Model is not precise and does not include buildings)
  - **human error** when annotating buildings
  - **lack of precision** of the ground truth data

II - Multi-task learning

- Model optimized for multiple tasks simultaneously [1]
- **Primary task:**
  - Output: **displacement map** $\hat{\mathbf{f}}$ aligning polygons to the optical image
  - Loss:

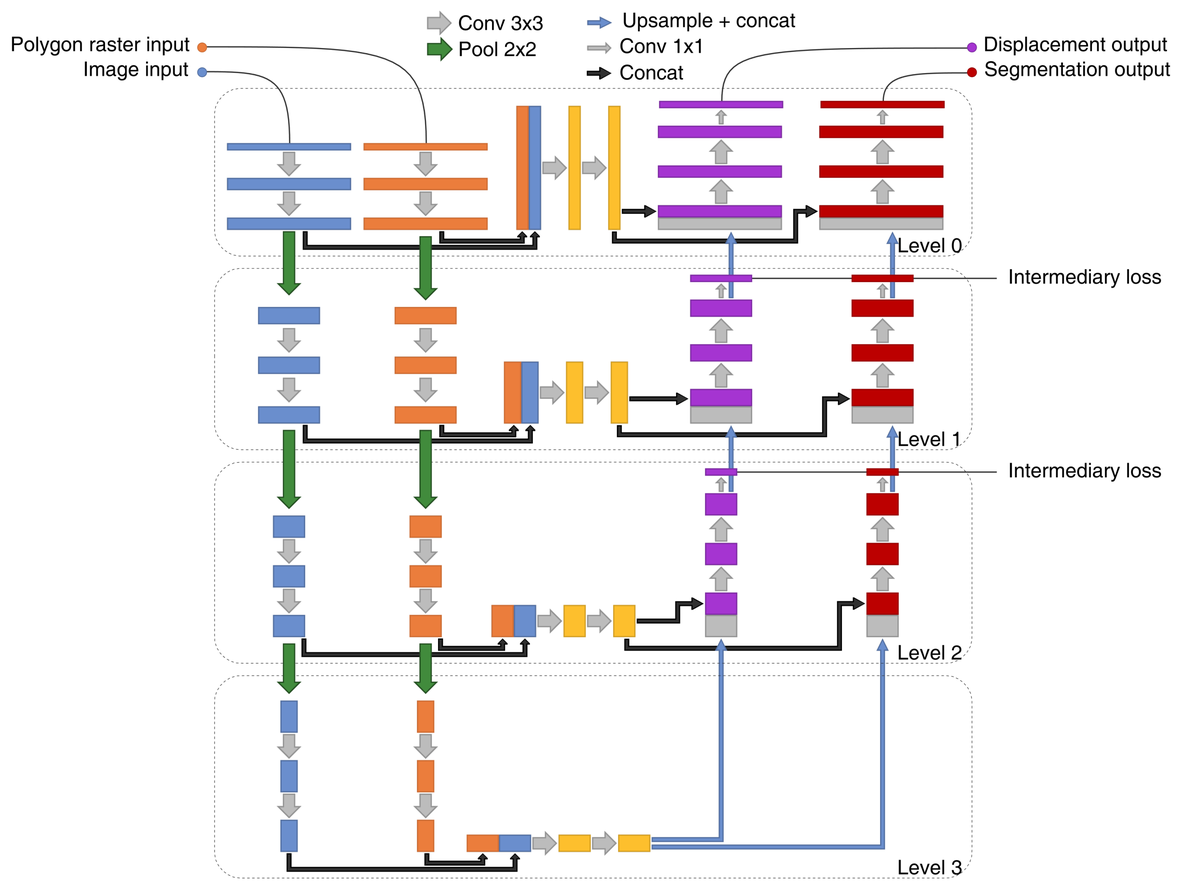
$$L^{\text{disp}}(\hat{\mathbf{f}}) = \sum_{\mathbf{x}\in[1,H]\times[1,W]} \frac{w_{c(\mathbf{x})}}{n_{c(\mathbf{x})}} \left\| \hat{\mathbf{f}}(\mathbf{x}) - \mathbf{f}_{\text{gt}}(\mathbf{x}) \right\|_2^2 \quad (1)$$

- **Secondary task:**
  - Output: **semantic segmentation** $\hat{\mathbf{p}}$ of the optical image
  - Loss:

$$L^{\text{seg}}(\hat{\mathbf{p}}) = \frac{1}{HW} \sum_{\mathbf{x}\in[1,H]\times[1,W]} \sum_{c=1}^{4} w'_c \, KL\left(\mathcal{D}(p^c_{\text{gt}}) \,\|\, \mathcal{D}(\hat{p}^c)\right) \quad (2)$$

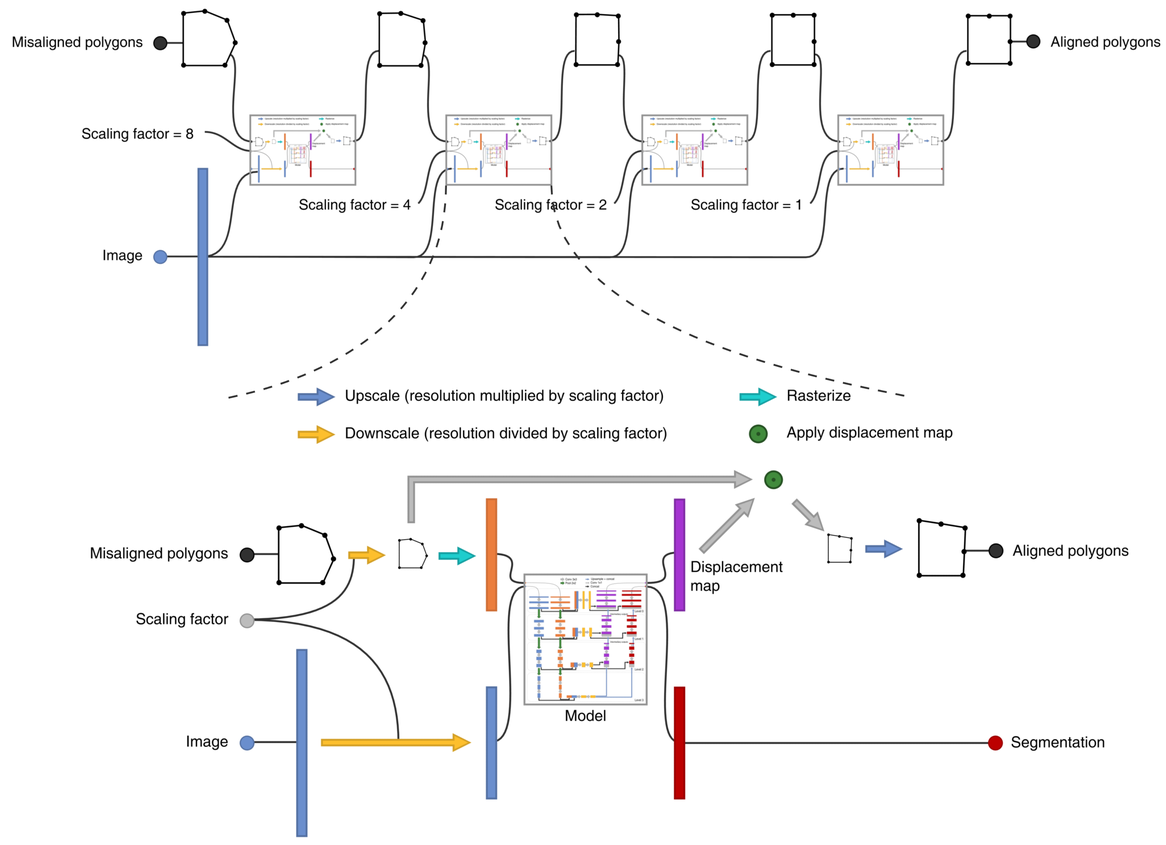
- Multi-task advantage: learning to segment buildings helps in aligning them
- **How:** Deep Learning method building on [2], resulting in a double input U-Net [3] with double outputs

III - Double input U-Net with double outputs



[1] S. Ruder. An overview of multi-task learning in deep neural networks. *CoRR*, 2017.
[2] A. Zampieri et al. Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing. *ECCV*, 2018.
[3] O. Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, 2015.
[4] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org , 2017.

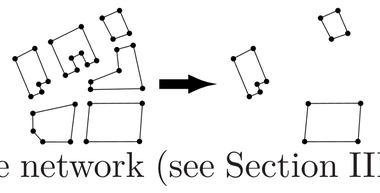
IV - Multi-resolution approach



- Models applied **iteratively at increasing resolutions**, expecting small displacements at each scale (±4 px)

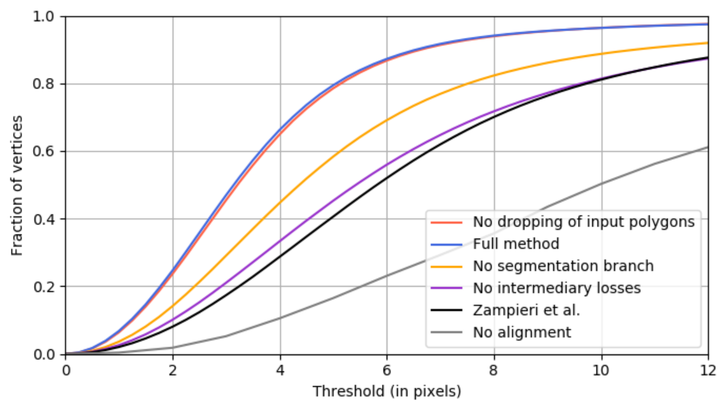
V - Training

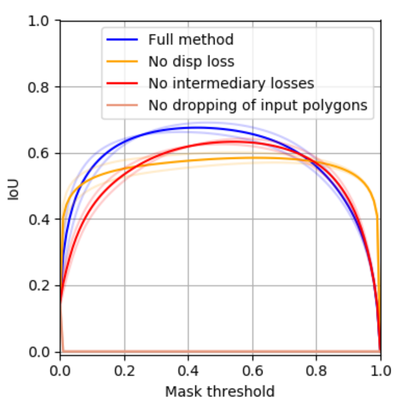
- Supervised approach with ground truth polygons from OpenStreetMap
- **Displacement map ground truth**: generated by applying random deformations to polygons. Maximum displacement: ±32 px
- **Random dropping** of input polygons, forcing the detection of new objects for the segmentation task



- Use of **intermediate losses** at each level of the network (see Section III), helping gradients flow and improving final performance on both tasks
- 4 types of pixels: background, polygon interior, edge and vertex. Different loss coefficient per type: $w_1 < w_2 < w_3 < w_4$ respectively

VI - Results

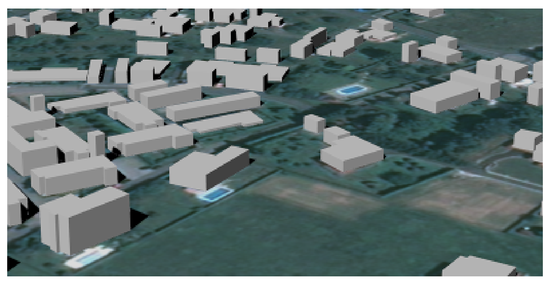




- Test data: 3 images of 5000 × 5000 px with 13614 buildings in total
- **Alignment accuracy measure** (left figure): proportion of vertices v for which $\|\hat{\mathbf{f}}(v) - \mathbf{f}_{\text{gt}}(v)\| \leq$ threshold in abscissa
- **Segmentation accuracy measure** (right figure): Intersection over Union at various thresholds in abscissa

VII - Building height estimation

- Input: 2 **stereo satellite images** with misaligned building polygons
- Process: align building polygons on both images
- Output: **height** of each building



VIII - Conclusion

- Effective at aligning **existing maps** over a **new image**
- Also **detects new buildings** with a segmentation map
- Each task helps training the other
- Results in better performance on both tasks
- **Intermediate losses** inside the network provide better gradient flow
- **Code:** `https://github.com/Lydorn/mapalignment`