ETH zürich

Aerial Image Labeling

Quantitative Big Imaging Course 2016

Javier Montoya, PhD
montoya@ethz.ch
PostDoc @ ScopeM

ETH zürich

What do I mean by **Aerial Image Labeling?**

- **Task:** divide a given input image into a set of semantic coherent regions: road networks and buildings.

Input Image

Semantic Segmented Image

ETH zürich

Why is it **Important?**

Earth Observation & Environmental Modeling

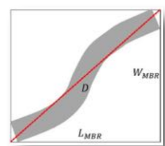
Urban Planning

Virtual Representations / 3D city models

Location-Aware Applications / Navigation Maps

ETH zürich

Which are the **existing techniques?**



Deterministic Representations: *detected pieces of objects are stitched together using low-level image processing, e.g. Miao et.al. (GRSL'13), Poullis et.al. (JPRS'10).*

- ✓ Successful when objects appear more clearly (no occlusions, etc).
- ✗ Many parameters must be tuned empirically.
- ✗ Errors from each step are propagated.

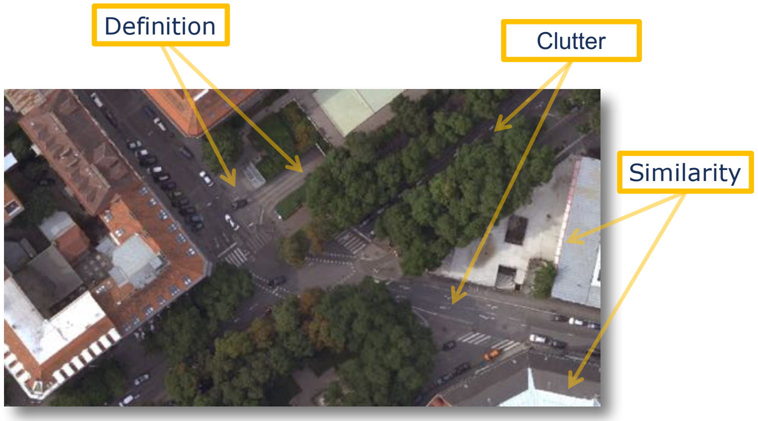
Local Statistical Approaches: *set of features are locally extracted to train class-specific models, e.g. Dollar et.al. (CVPR'06), Mnih et.al. (ECCV'10).*

- ✓ Robust models can be obtained.
- ✗ Availability of training datasets, hand-labeling process.
- ✗ Predictions are focused on local information only.

Probabilistic Representations of Image Context: *high-level semantic knowledge is incorporated, e.g. Lacoste (PAMI'05), Türetken et.al. (CVPR'13).*

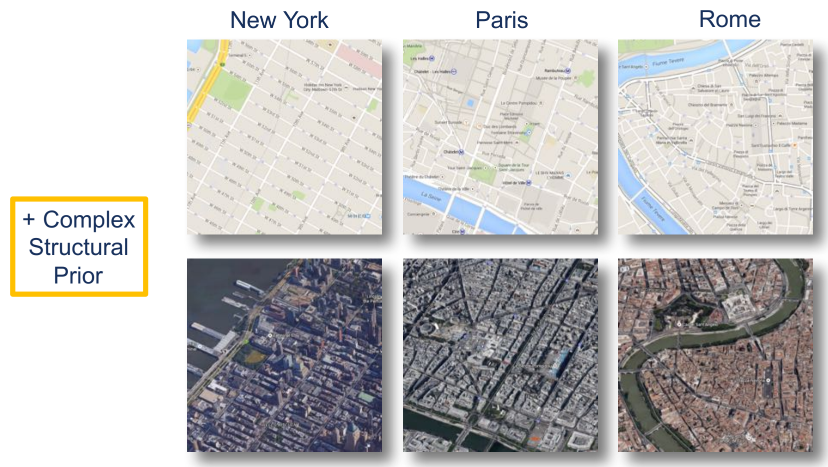
- ✓ Encode rich semantic level information.
- ✓ Smooth and precise segmentations.
- ✗ Inference can be time consuming, e.g. Markov Point Processes.

What are the **Challenges?**

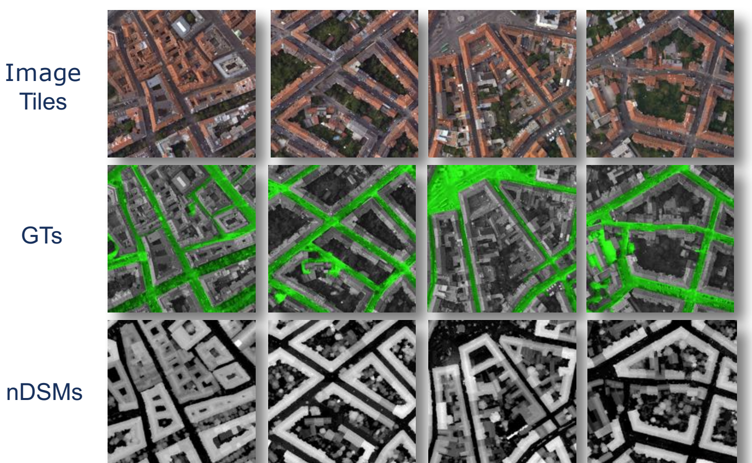


… unsolved problem since almost 40 years! *Bajcsy et al. (TSMC'76)*

What are the **Challenges?**

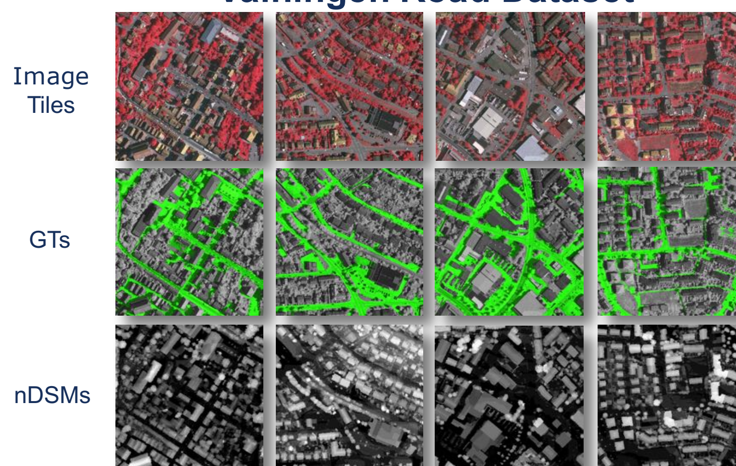


Graz Road Dataset



(i) Urban region, two classes (road vs. background), 67 tiles of 1000x1000 pixels, manual annotations.
(ii) Road networks: major avenues + secondary streets, change slow in width/curvature.
(iii) Presence of occlusions, e.g. trees and cars.

Vaihingen Road Dataset



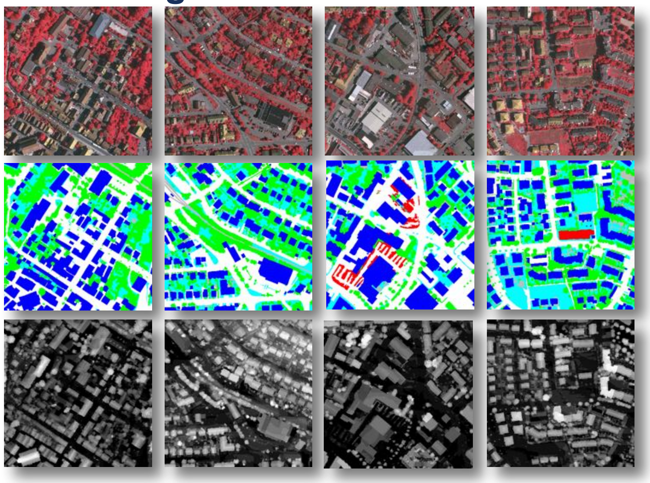
(i) Countryside region, two classes (road vs. background), 16 tiles of 1000x1000 pixels, manual annotations.
(ii) Road networks: very irregular, mainly narrow, partially occluded trees, shadows.

Vaihingen Multi-class Dataset

Image Tiles

GTs

nDSMs



(i) Six classes: natural ground, background, roads, trees, grass, and buildings.
(ii) Buildings: vary strongly in shape and are often densely clustered.

How are the results **measured?**

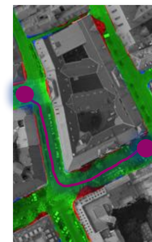
Pixel-wise classification accuracy:

- F1-score, Precision, Recall.

Road Network Topology:

- What fraction of connecting paths between road seeds have the correct length within 5% tolerance, are respectively:

GT



Result

Ideal Case:

How are the results **measured?**

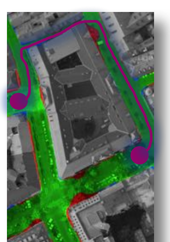
Pixel-wise classification accuracy:

- F1-score, Precision, Recall.

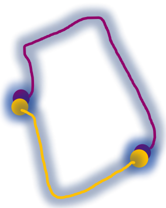
Road Network Topology:

- What fraction of connecting paths between road seeds have the correct length within 5% tolerance, are respectively:

GT



Result



Too Long (2long)

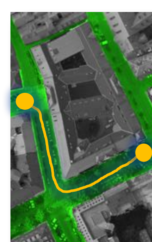
How are the results **measured?**

Pixel-wise classification accuracy:

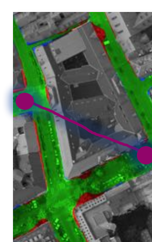
- F1-score, Precision, Recall.

Road Network Topology:

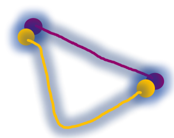
- What fraction of connecting paths between road seeds have the correct length within 5% tolerance, are respectively:



GT



Result



Too Long (2long)

Too Short (2short)

ETH zürich

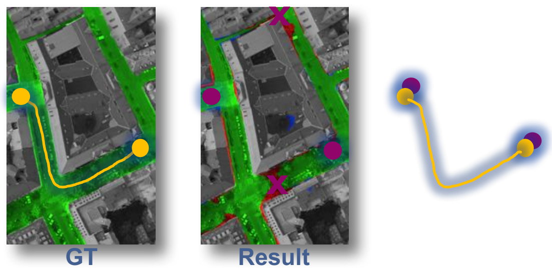
How are the results **measured?**

Pixel-wise classification accuracy:

- F1-score, Precision, Recall.

Road Network Topology:

- What fraction of connecting paths between road seeds have the correct length within 5% tolerance, are respectively:



Too Long (2long)

Too Short (2short)

No Connectivity (NoConn)

ETH zürich

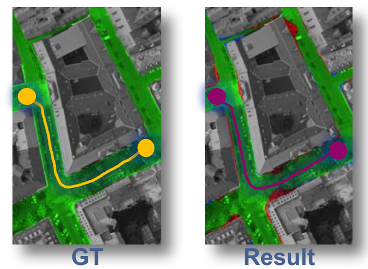
How are the results **measured?**

Pixel-wise classification accuracy:

- F1-score, Precision, Recall.

Road Network Topology:

- What fraction of connecting paths between road seeds have the correct length within 5% tolerance, are respectively:

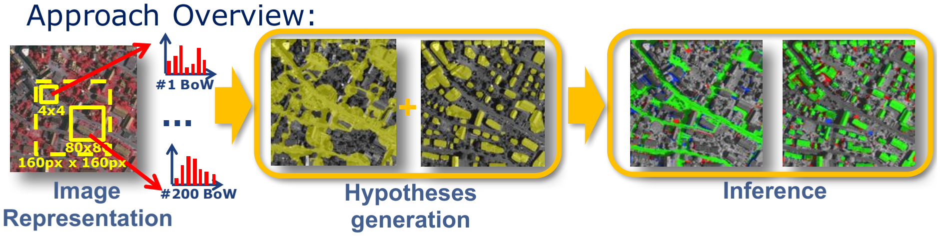


TopoCorrectness = 100% - 2Short – 2Long - noConn

ETH zürich

Class-Specific higher-order cliques

Approach Overview:



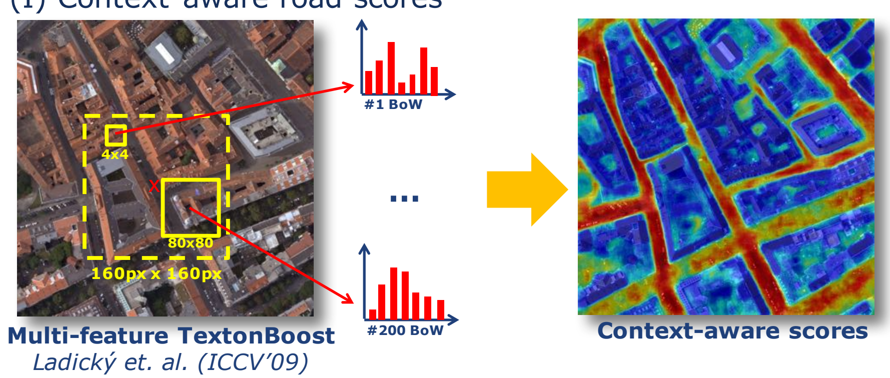
Model:

(i) Multilabel pixelwise classification using powerful neighbor features.
(ii) Overcomplete representation of *building* and *road* candidates.
(iii) Candidates are prunned to optimal subset through CRF.

ETH zürich

Class-Specific higher-order cliques

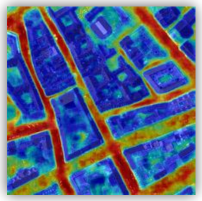
(I) Context-aware road scores



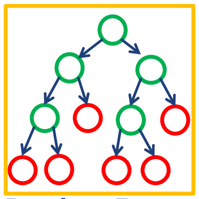
- Multi-label pixelwise classification.
- Self context/Local layout, appearance information is encoded over large spatial neighborhoods.

Class-Specific higher-order cliques

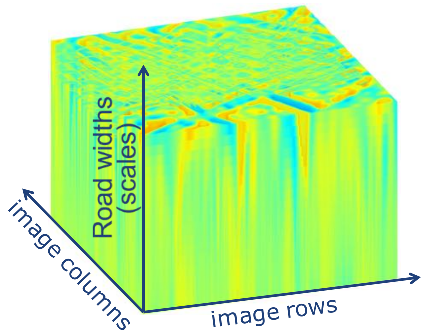
(II) Hyphoteses Generation: Roads



Context-aware road scores



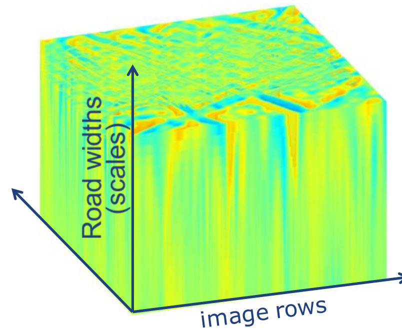
Random Forest estimates road widths



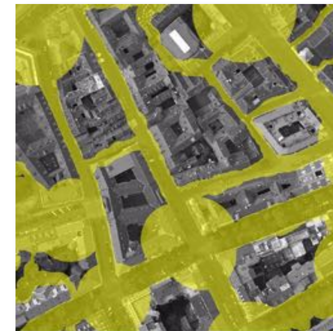
Likelihoods per road width

Class-Specific higher-order cliques

(II) Hyphoteses Generation: Roads



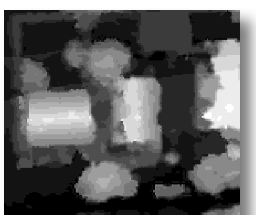
Scale-spatial Road likelihoods

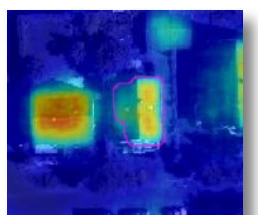


Sampled Road candidates

Class-Specific higher-order cliques

(II) Hyphoteses Generation: Buildings

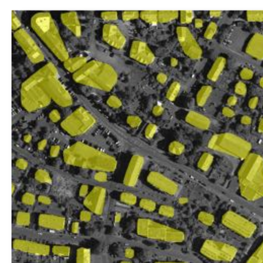




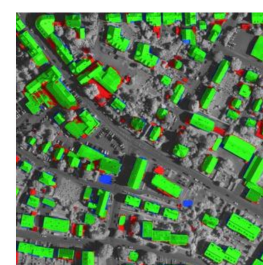
- Building candidates based on classifier scores.
- Connected components with high building likelihood.
- Building segments are approximated through alpha-shapes.

Class-Specific higher-order cliques

(III) Hypotheses Selection: Inference



Sampled candidates (high recall)



Candidate selection (high precision)

$$E = \sum_{pixels} E_u(x_{pix}) + \sum_{pix}\sum_{neigh} E_p(x_{pix}, x_{neigh}) + \sum_{roads} E_R(Q_m) + \sum_{build.} E_B(Q_n)$$

class evidence

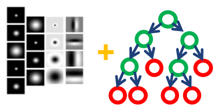
soft-clique membership

CRF-Model for road superpixel segmentation

Baselines:

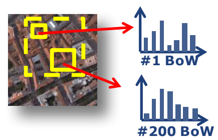
$$E = \sum_{pixels} E_u(x_{pix}) + \sum_{pix}\sum_{neigh} E_p(x_{pix}, x_{neigh}) + \sum_{roads} E_R(Q_m)$$

Winn



Road/Background Pixel-likelihoods

Context

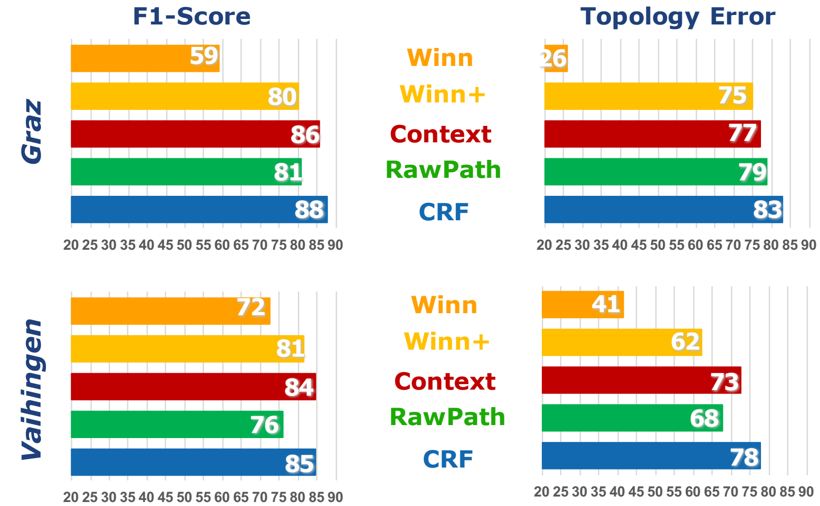


Multi-feature Textonboost

RawPath

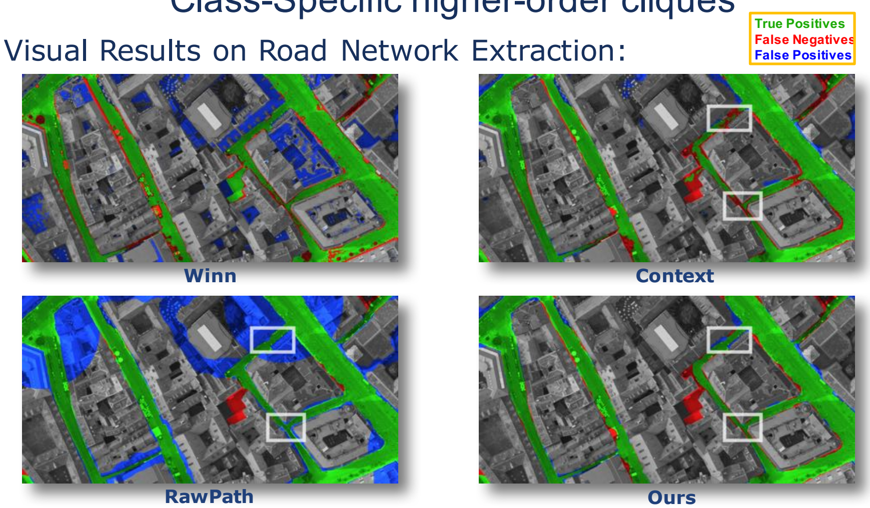
Sampled Road Candidates

Class-Specific higher-order cliques



Class-Specific higher-order cliques

Visual Results on Road Network Extraction:

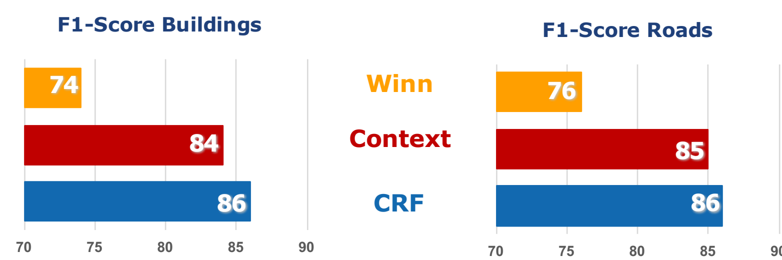


Road network on GRAZ, img6

Class-Specific higher-order cliques

Experimental Results on Joint Road Networks + Buildings:

$$E = \sum_{pixels} E_u(x_{pix}) + \sum_{pix}\sum_{neigh} E_p(x_{pix}, x_{neigh}) + \sum_{roads} E_R(Q_m) + \sum_{build.} E_B(Q_n)$$



Labeling Performance on the **Vaihingen Multi-class Dataset** (all numbers percentages).

ETH zürich

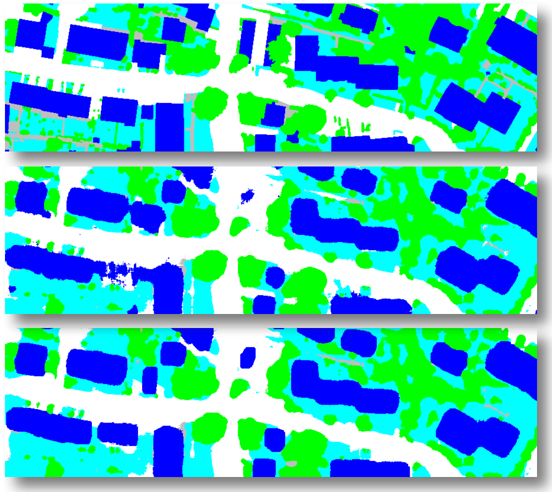
Class-Specific higher-order cliques

Visual Results on Joint Road Networks + Buildings:

GTs

Context

Ours

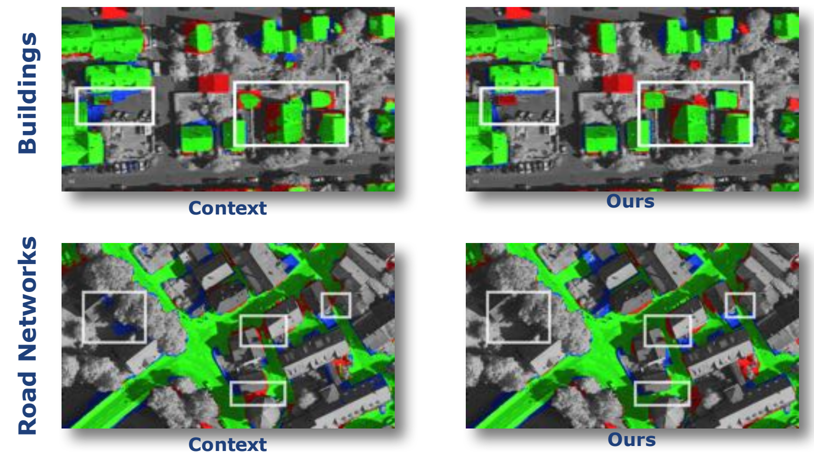


ETH zürich

Class-Specific higher-order cliques

Visual Results on Joint Roads + Buildings:

True Positives
False Negatives
False Positives



ETH zürich

Conclusions

- **Class-specific Priors:**
  - Higher-level representations for buildings and roads are useful multi-class segmentation.
  - Buildings are represented as a set of compact-like polygons.
  - Roads are modeled as a collection of long, narrow segments.

ETH zürich

Outlook & Future Work

- **Generation of object candidates:**
  - Probabilistic Sampling Scheme?
  - Hypotheses parameters as regression task?
- **Training Dataset:**
  - Manual labeling of roads is time-consuming and costly
    => use publicly available data as ground truth, e.g. Open Street Map.
    => deep learning.
- **Applicability in other domains:**
  - Generic model potentially applicable to other networks such as neurons and vessel in medical imaging.


Thank You!