Mind the gap: modeling local and global context in (road) networks

Javier A. Montoya-Zegarra[a], Jan D. Wegner[a],
Ľubor Ladický[b], Konrad Schindler[a]

[a]Photogrammetry and Remote Sensing, ETH Zürich
[b]Computer Vision Group, ETH Zürich

Abstract. We propose a method to label roads in aerial images and extract a topologically correct road network. Three factors make road extraction difficult: (i) high intra-class variability due to clutter like cars, markings, shadows on the roads; (ii) low inter-class variability, because some non-road structures are made of similar materials; and (iii) most importantly, a complex structural prior: roads form a connected network of thin segments, with slowly changing width and curvature, often bordered by buildings, etc. We model this rich, but complicated contextual information at two levels. Locally, the context and layout of roads is learned implicitly, by including multi-scale appearance information from a large neighborhood in the per-pixel classifier. Globally, the network structure is enforced explicitly: we first detect promising stretches of road via shortest-path search on the per-pixel evidence, and then select pixels on an optimal subset of these paths by energy minimization in a CRF, where each putative path forms a higher-order clique. The model outperforms several baselines on two challenging data sets, both in terms of precision/recall and w.r.t. topological correctness.

1 Introduction

In this paper we deal with automated extraction of the road network from overhead images.[1] The emergence of on-line services like Google Maps, navigation systems, and location-based services has lead to an increased demand for up-to-date maps, particularly in densely populated urban areas. Road extraction is a classical problem which dates back almost 40 years [1] and considerable progress has been achieved, see overviews in [10, 24]. Still, no automatic method is robust enough to be employed in practice, and roads are digitized by hand, which is slow and costly. What makes roads (and other linear structures like waterways) special is that topological completeness of the network is often more important than pixel-accurate segmentation. Consider a routing task where the shortest connection from A to B is sought. While slightly misplaced road boundaries will not harm the routing, even a very narrow gap may cause a lengthy detour. We

[1] As often done in aerial imaging, when it is available we regard the height-field from dense matching as an additional image channel, and do not separately refer to it.

thus put an emphasis on network quality, i.e. our main objective is extracting *topologically* complete and correct road networks.

Road extraction in urban environments is challenged by varying road appearance, occlusions as well as heterogeneous background. Unlike highways in the countryside, city streets are frequently occluded (e.g. by trees) or lie in cast shadow. Shape properties like road width, straightness and network density exhibit greater variation. Moreover, many background objects have road-like appearance when viewed from above, e.g. concrete roofs. Thus, classification based on local appearance is unreliable. On the other hand, roads offer a lot of structure and context: *locally*, road pixels form narrow, elongated strips, often bordered by buildings or lined with trees; *globally*, they form a connected network with (mostly) slowly changing segment width. Importantly, these structural properties are quite universal, whereas geometric properties vary from place to place, e.g. American cities have wider roads laid out in a rectangular grid, whereas central European cities have narrower, and more irregular road networks.

We pose road extraction as a pixel-wise labeling task with two classes "road" and "background", and address local context and long-range structure separately. Context is learned directly from data, by training a classifier that uses rich appearance features extracted from a large window, and in this way implicitly includes the local co-occurrence patterns, similar to [17]. The network is modeled explicitly: from the pixel-wise road score, we predict the likelihood that a road of a certain width is present. Based on the resulting $(x, y, width)$-volume of road likelihoods we apply a *recover-and-select* strategy: in the *recover* step many candidates for larger stretches of road are sampled. The *select* step then picks a subset of these candidates that best explains the image evidence (i.e. optimally covers the roads). The selection is formulated as a higher-order CRF, in which the pixels belonging to each road candidate form a large clique, and the clique potential favors consistent labeling of the member pixels. The CRF thus models two preferences: *(i)* pixels should only be labeled as road if they lie on a well-supported long-range connection (or large square) of the network, thus improving precision; and *(ii)* if in a clique the evidence for road outweighs the one for background, then (almost) all of its pixels should be labeled as road, thus improving recall and preventing gaps in the network.

CRF models are currently the standard way of encoding dependencies between pixels in labeling problems, and the search for maximum-evidence (resp. minimum-cost) paths is a classical approach to reconstruct networks – not only roads, but also blood vessels or neurons in medical imaging. Our work is, to our knowledge, the first road extraction framework that attempts to embed minimum-cost paths in a CRF framework, leading on the one hand to a more global solution (because paths can overlap without "double counting"), and on the other hand to a more accurate segmentation (because pixel-to-path membership is a soft constraint and can change during inference).

In the experiments section, we show that together the proposed measures significantly improve the resulting road network. The local context resolves problems due to ambiguous appearance, and yields a significantly higher labeling

accuracy than a baseline classifier (16-45% gain in $F1$-score), which by itself approximately doubles the topological correctness. The long-range network prior on the other hand only brings moderate additional improvement (up to 2% in $F1$-score) in terms of labeling accuracy, but further increases the topological correctness of the final network by 7%.

2 Related Work

Since the first early works appeared on road extraction from satellite imagery [1], a large number of methods have been proposed that model road networks with comprehensive sets of ad-hoc rules (*e.g.*, [6, 25, 28, 9, 35, 26]). Most often they are bottom-up processes that hierarchically stitch together short road segments detected with low-level image processing. The strategy can be successful in rural and suburban areas, where the roads stand out more clearly, there are fewer shadows and occlusions, and the background is relatively homogeneous. Typically many parameters must be tuned empirically. Also, rule-based "expert systems" rely on hard thresholds rather than probabilistic formulations, so they cannot recover from mistakes made at early stages. Some authors have also tried a rule-based approach for more challenging urban environments [11], leading to even more extensive rule sets.

Marked point processes (MPP) offer a probabilistic framework in which road elements (e.g. short line segments) are the basic variables, and allow one to impose high-level topological constraints [31, 16, 18]. Chai *et al.* [4] recently proposed a comprehensive prior that models both line-segments and junctions of the road network. MPPs are a powerful tool to formulate priors at the level of road elements, but inference is hard and has to rely on all-purpose sampling methods like reversible jump Markov Chain Monte Carlo (RJMCMC), which are computationally expensive and need careful tuning.

A conceptually appealing strategy is to view the road network as a set of minimum cost paths. Such an approach is more flexible in terms of road shape and directly enforces connectivity. Already in early work [7] an A^*-type algorithm is used to iteratively assemble line detector responses to road networks. In medical imaging, minimum-cost paths are widely used to reconstruct vessel trees and neurons (*e.g.*, [20, 34, 3, 2, 39]). Recently [33] also tested an algorithm originally developed for medical data on suburban road networks.

Perhaps the most related methods to ours are [36] and [32]. In [36] the road network is also modeled with the help of a CRF with long-range, higher-order cliques (over super-pixels), using a variant of the P^N-Potts model [14]. The road segments are assumed to be piece-wise straight, and only serve to repair false negatives of the original labeling. Here, we allow for arbitrarily shaped segments, which are found in a data-driven manner. Moreover, our method does not have a foreground bias, but also suppresses false positives in the unaries. Türetken *et al.* [32] (and the earlier [33] for cycle-free networks) compute multi-scale local tubularity scores (reminiscent of our road likelihoods), connect seeds with high foreground scores by minimum cost paths, and prune the over-complete graph

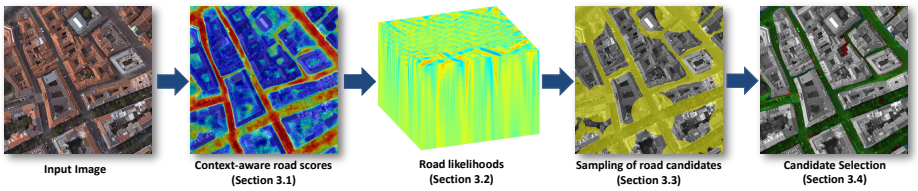

Fig. 1. Given an input image, our method first classifies pixels into road and background (Sec. 3.1). Next, the presence of a road with a specific width is predicted (Sec. 3.2). An over-complete set of road candidates is generated (Sec. 3.3), and pruned to an optimal subset (Sec. 3.4).

to an optimal sub-graph with mixed integer programming. They report good performance on aerial images with unoccluded roads, although the method was developed for neurons. The focus is on the center-lines, whereas road pixels are not individually labeled (and the width is not explicitly recovered).

3 Model

We model the road network as the union of elongated segments (termed *paths*) and large compact regions (termed *blobs*). Both *paths* and *blobs* come with an associated scale, *i.e.* we do not only represent road centerlines, but also the local width of the road, respectively the size/diameter of large undirected parts of the network like squares or parking lots. We start with a conventional pixel-wise classification into road and background pixels (Sec. 3.1). In this stage we already include local context via per-pixel feature vectors that encode appearance information over a large spatial neighborhood. In the raw map of road (foreground) scores the local scale (width) of the roads is contained only implicitly. To make it explicit we add a further classification step (Sec. 3.2), which takes as input statistics about the local distribution of the raw scores, and predicts the likelihood that a road *with a specific width* is present.[2] A set of putative *paths* and *blobs* (together referred to as road *candidates*) are then sampled on the basis of the resulting $(x, y, width)$-volume of road likelihoods (Sec. 3.3), by connecting random seed points with paths of maximal cumulative road likelihood, respectively finding large blobs with maximal cumulative likelihood. In order to achieve high recall we follow a recover-and-select strategy: the set of *candidates* is generated such that it is over-complete, but covers as many of the actual road pixels as possible. Finally, a subset of all candidates is selected by energy minimization, resp. MAP estimation, in a CRF (Sec. 3.4). The original road scores form the pixel-wise unaries, and each *candidate* is a higher-order clique with a robust P^N-Potts potential that encourages clique members to take

[2] In principle the two-stage classification could potentially be replaced by some form of structured prediction. This would require significantly more training data.

on the same label. Our binary labeling problem allows for globally optimal CRF inference with the min-cut algorithm. The approach is summarized in Fig. 1.

3.1 Context-aware road scores

To obtain pixel-wise road/background scores that take into account the context in the vicinity of a pixel we adopt the multi-feature extension [17] of the TextonBoost algorithm [30]. The pipeline works as follows: first multiple types of features – SIFT [22], textons [23], local ternary patterns [13] and self-similarity [29] – are extracted densely for all pixels. Each feature is soft-quantized to the 8 nearest neighbors in a dictionary of 512 words, using distance-based weighting with the exponential kernel [8]. To include the context, the quantized words are then accumulated into bags-of-words over a (fixed) set of 200 random rectangles that cover a large image region around a pixel. Rectangles range from 4×4 to 80×80 pixels in size, and their locations are sampled in a neighborhood of 160×160 neighborhood, from a Gaussian distribution (*i.e.* their density decreases with distance). The final feature set is the concatenation of all 200 bags-of-words, and thus is aware of the context, in the form of the feature distribution in a large 160-pixel neighborhood around a pixel (compared to an average road width of $\approx$ 30 pixels in our data).

A classifier is then learned by 5000 rounds of boosting decision stumps on single feature dimensions (also called "shape filters" [30]). Features are not kept in memory, but extracted on the fly using integral images. The boosting output is transformed to pseudo-probabilities S_{road}, S_{bg} in the standard way, by mapping it to the range [0..1] with a sigmoid.

3.2 Road likelihoods

In order to generate promising candidates in the subsequent step, we need for each pixel not only a single road score, but the likelihood to encounter a piece of road of a particular width w. To estimate that likelihood (for a range of discrete widths), we generate a scale-space representation from the road scores S_{road}, by computing a pyramid of pixel-wise responses to scale-normalized Laplacian-of-Gaussian (LoG) filters of different scales [21]. We found that rather than using the raw LoG responses (or simple transformations of them), it is more robust to resort to a second round of discriminative classification: we feed the mean, median and standard deviation of the pixel-wise LoG responses into a random forest (20 trees, max-depth 15) trained on ground truth road widths, to predict the local likelihood for each possible road width, resulting in a volume $L(x, y, w)$ of likelihoods. In our experience this "learned mapping" from LoG responses to road width likelihoods works better than obvious ad-hoc mappings like rescaling or sigmoid fitting, presumably because it can learn, from the additional information contained in the training labels, to correct typical failures and noise in the raw road scores. At the conceptual level this is in agreement with [27], who also observe an improvement by "cleaning up" raw classifier responses with a second round of classification that looks at their local distribution. In [33] the

order is reversed: paths are constructed after the first round, then segments of those paths are reclassified to get more reliable scores.

3.3 Sampling of road candidates

The goal of this step is to generate a large set of putative road *candidates*. Candidates are either long-range curvilinear *paths* or large isotropic *blobs* that are likely to belong to the road network. Candidate generation aims for high recall: the union of all *candidates* must contain as much as possible of the road network, even at the cost of low precision. Weak candidates are discarded in the subsequent selection step, but missing candidates cannot be recovered later.

Elongated *paths* are generated by picking two random points with reasonably high road likelihood in $L(x, y, w)$ and connecting them with a path through the volume that maximizes the cumulative likelihood. That path is found with the 3D Fast Marching algorithm [5]. By allowing seed points that are far from each other the paths are a means to impose the long-range network prior: a *path* is always an uninterrupted connection between the seeds and bridges gaps with low road likelihood where necessary. Note that the *path candidates* have an explicit *road width* assigned at each pixel, which changes smoothly (because the path through the volume is continuous also in the w-dimension).

Paths alone are not sufficient to represent the road network. In practice the network also contains large regions without a clear direction like parking lots, squares, roundabouts etc. Paths between different seed points will always traverse such regions along the same routes, where the costs are lowest due unavoidable fluctuations of the likelihood. To nevertheless include such *blob* regions we model them separately. We scan only the top scales w of $L(x, y, w)$, above the maximum road width, for local maxima, and perform non-maxima suppression to obtain a set of *blob candidates*.

3.4 Candidate selection

The final step is to select a subset of candidates that best covers the road network. Among the given candidates, some paths will pass (partially) through background; different paths will overlap because the fast marching search tends to use the same high-likelihood regions to connect different seeds; and blobs will also overlap smaller blobs as well as many paths.

On the other hand, even the best paths will not always perfectly correspond to roads, especially along the road boundaries. Therefore it is desirable to include a correction step that slightly modifies the candidates where required, but prefers to change them as little as possible in order to maintain coverage and connectivity of the road network. We cast this "selection with correction" as probabilistic inference in a CRF, *i.e.* we minimize an energy $E = \sum_j E_u(x_j) + \sum_i E_p(Q_i)$ over all pixels x_j of the image.[3] The pixel-wise unaries $E_u(x_j) = -\log(S)$ are negative log-likelihoods of the raw road scores from the original classifier (Sec. 3.1).

[3] If desired the P^N-Potts model would also allow for conventional pairwise potentials. We did not find them necessary, the context-based unaries are already locally smooth.

The candidates enter in the form of higher-order cliques Q_i, which contain all pixels that belong to candidate i. Their potentials $E_p(Q_i) = \min(\alpha, N_k \cdot \frac{\alpha - \beta}{\gamma} + \beta)$ are robust P^N-Potts potentials [15], which encourage all member pixels to have the same label. Here N_k denotes the sum of nodes in the clique that take label k, and $\{\alpha, \beta, \gamma\}$ are the parameters of a truncated linear function that governs how the energy increases as more pixels deviate from the dominant label. The cliques encourage all their member pixels to have the same label. If *sufficient road evidence* is accumulated inside the clique, pixels are pulled to the road class, which helps to correct false negatives in the unaries and maintain connectivity of the network, while still allowing to correct individual pixels that were wrongly assigned to the clique. If, taken together, the pixels in a clique have *too little road evidence*, then it is discouraged to label only small, scattered parts of it as road, which helps to suppress false positives not connected to the road network.

Finally, we still need to encode the model assumption that the candidate set is over-complete, *i.e.* pixels that are not covered by any path should never be labeled as roads. In CRF terms this corresponds to a large higher-order clique Q_{bg} which spans all pixels that are *not member of any candidate* path or blob. That clique has an asymmetric potential which imposes an infinite penalty if any of its pixels is labeled as road, and no penalty otherwise. In practice the same effect can be achieved more efficiently by setting the road likelihoods of the pixels in Q_{bg} to zero, $\forall x_j \in Q_{bg} : S_{road}(x_j) = 0\,,\ E_u(x_j) = \infty$. The binary P^N-Potts model can be solved to global optimality with a graph cut, hence our inference is guaranteed to find a global minimum of the energy with a single run of the min-cut algorithm.

4 Experimental Results

We perform experiments on two data sets of urban scenes, GRAZ (Austria) and VAIHINGEN (Germany).[4] Both data sets are orthophoto mosaics with 3 color channels plus a normalized height channel computed via dense image matching. The pixel size is 0.25 m on the ground. In order to enable parallelization and to reduce the memory footprint, we split up each data set into overlapping tiles of 1500×1500 pixels. Computations are done on the full tiles to avoid boundary artifacts, while the evaluation is done only for the non-overlapping part of 1000×1000 pixels to avoid double counting. The data sets depict rather different road networks. GRAZ covers the city center of a major city with big building blocks, inner court yards and parks. There are 67 tiles overall (30 training, 12 validation, 25 testing). Color channels are standard RGB. VAIHINGEN is a small historic town in hilly countryside, with small buildings, irregular layout, and narrow, winding roads. There are 16 tiles (4 training, 4 validation, 8 testing). Color channels are near infrared, red, and green.

[4] GRAZ was kindly provided by Microsoft Photogrammetry. VAIHINGEN is part of the ISPRS benchmark *http://www.itc.nl/ISPRS_WGIII4/tests_datasets.html*

4.1 Evaluation metrics

As quality measures we report both conventional pixel-based classification scores and topological correctness. Classification accuracy is measured in the standard way with pixel-wise *precision*, *recall*, and $F1$-*score*. In aerial imaging, variants of the measures called *correctness*, *completeness* and *quality* are popular (*e.g.* [19, 24, 12, 27]) which allow for a few pixels of slack orthogonal to the road centerline to account for geometric uncertainty [37]. We found no significant difference to the standard measures, but nevertheless report both sets. Furthermore, we give the κ-value to assess pixel-wise segmentation accuracy. For a confusion matrix C computed from N pixels, $\kappa = \frac{N\sum_i c_{ii} - \sum_i (\sum_j c_{ij} \cdot \sum_j c_{ji})}{N^2 - \sum_i (\sum_j c_{ij} \cdot \sum_j c_{ji})}$. It quantifies how much the predicted labels differ from a random image with the same label counts.[5]

All these measures are based on pixel area and do not capture the topological correctness of the extracted network. A tiny gap in a road can lead to lengthy detours, but has little impact on recall, and vice versa only few false positive pixels are necessary to produce an inexistent shortcut. We thus additionally report the topological metrics of [36]. These measure what fraction of connecting paths between road points have the correct length within 5% tolerance, respectively are too short (*2short*), too long (*2long*), or completely infeasible (*noC*). The metrics are computed by randomly sampling paths and counting the occurrence of the four cases until the numbers converge.

4.2 Results

Tab. 1 shows the results for GRAZ and VAIHINGEN. We report both results of raw classification with the context-aware unaries (*Context*) and results after adding the long-range prior (*CRF*), in order to separate their contributions. Additionally, we add standard baselines for each of the two steps.

As baseline classification we extract per-pixel features with the filter bank of Winn *et al.* [38] and classify them with a random forest (*Winn*). These features consist of multi-scale intensity and derivative responses. They capture the texture properties immediately around the pixel and in our experience work as well as other texture filter banks, but do not capture context in the sense of object-scale shape and co-occurrence patterns. Moreover, as a baseline for a complete system built on top of the features of *Winn* we use our earlier work [36]. That method (*Winn+*) starts from raw road likelihoods obtained by classifying *Winn* features (averaged over superpixels). Straight line segments serve as cliques in a CRF, which is designed to bridge gaps in the road network.

As baseline for the influence of the long-range prior we start from the more powerful *Context* classifier, run the candidate generator in the same way as for *CRF*, and discard all candidates whose average unary score is below 0.7 (the threshold which empirically maximizes the $F1$-score). All pixels of the remaining paths and blobs are labeled as road (*RawPath*).

[5] κ avoids biases due to uneven class distribution. E.g., for an image with 10% *road* pixels a result without a single *road* pixel has 90% overall accuracy, but κ=0%.

	Method	*Qual.*	*Compl.*	*Corr.*	κ	*F1*	*Rec.*	*Prec.*	*Corr.*	*2long*	*2short*	*NoC.*
GRAZ	*Winn*	42.8	55.1	65.7	46.9	58.9	54.9	65.8	26.0	10.3	**0.6**	63.1
	Winn+	67.0	84.7	77.1	72.9	80.1	84.8	77.4	74.8	4.3	12.0	8.9
	Context	74.8	**88.5**	83.1	80.3	85.6	88.5	83.3	77.1	9.1	3.4	10.4
	RawPath	68.2	85.3	78.2	74.0	80.9	85.3	78.4	78.6	**3.2**	16.4	**1.8**
	CRF	**78.0**	**88.5**	**86.9**	**83.1**	**87.6**	**88.6**	**87.1**	**82.8**	5.8	8.4	3.0
VAIH	*Winn*	56.9	68.2	77.2	62.2	72.4	68.2	77.3	41.3	22.6	**3.4**	32.7
	Winn+	68.7	85.6	78.6	73.5	81.4	85.6	78.7	62.1	5.3	22.8	9.8
	Context	73.0	89.6	79.8	77.6	84.4	89.6	79.9	72.6	8.0	12.1	7.3
	RawPath	61.0	**91.2**	64.9	63.8	75.8	**91.3**	64.9	67.7	**1.8**	30.0	**0.5**
	CRF	**73.3**	88.4	**81.1**	**78.0**	**84.6**	88.4	**81.2**	**77.7**	6.2	12.8	3.3

Table 1. Performance of road extraction methods. All numbers are percentages.

For both datasets the local context in the unaries drastically improves the pixel-wise performance – see Tab. 1. The largest contribution comes from increased recall, as the context repairs errors in areas where shadows, trees, unusual surface colour etc. perturb the local appearance – see Fig. 2. Moreover, there is also a significant gain in precision as false positives on concrete roofs, asphalted courtyards etc. are suppressed if they are not supported by the context. Naturally, the greatly improved labeling accuracy is also reflected in much higher topological correctness. *Winn+* does greatly improve the result over the raw labeling of *Winn*, but is still dominated by raw *Context* unaries, which confirms the intuition that one should already include context at the feature level to get stronger unaries.

The proposed *CRF* model further increases per-pixel performance over *Context*, but as expected the effect is comparatively small. Many gaps and false positive patches are cleaned up, but their pixel area is relatively small. Still, these changes significantly increase the topological correctness, mainly by repairing gaps in the network and reducing the number of too long or impossible connections: i.e. the model does what it is designed for, and fills in missing links. The price to pay is that the fraction of too short connections also increases a bit, since some correct gaps are bridged.

On the contrary, heuristically fixing the network (*RawPath*) does not achieve the desired effect. Neither the pixel-wise nor the topological performance of the *Context* unaries is increased, mostly because of false positives. The results suggest that the proposed probabilistic model successfully balances the image evidence against the network prior. It manages to drag concealed roads to the foreground, while at the same time also suppressing false alarms (contrary to *RawPath*, which increases them).

5 Conclusions and Future Work

We have proposed methods to exploit context for the semantic segmentation of roads, both at the local and global level. At the local level, expressive features

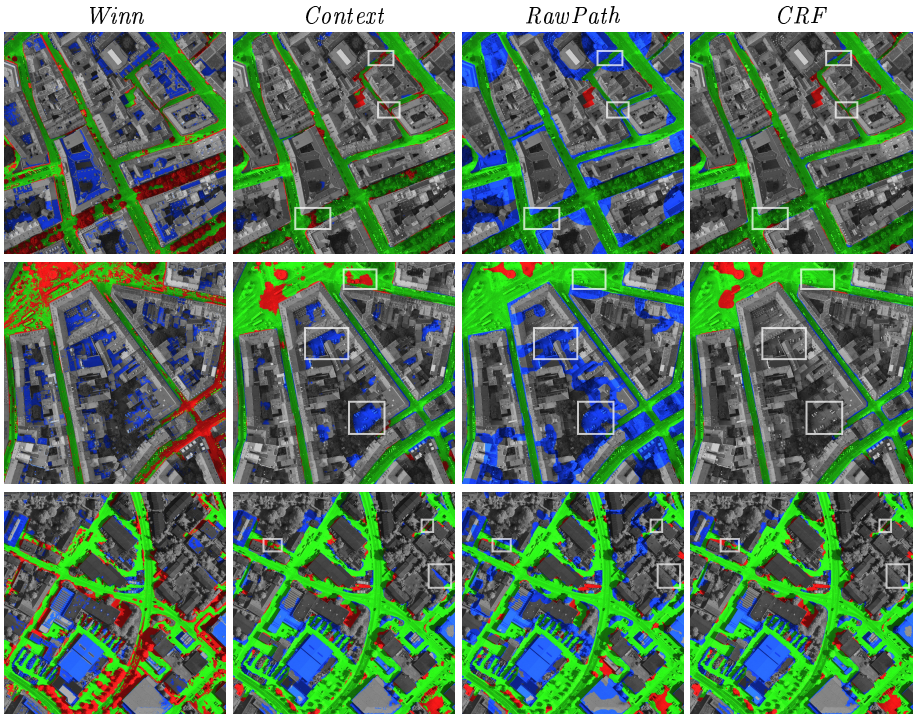

Fig. 2. Road networks extracted in two patches of the orthophoto mosaic of GRAZ (two top rows) and one patch of VAIHINGEN (bottom row). True positives are displayed green, false positives blue, and false negatives red. White boxes highlight improvements by the long-range prior.

extracted over large neighborhoods implicitly capture the shape and layout of roads and surrounding objects, and lead to much improved classification scores. At the global level the combination of optimal path search and a higher-order CRF models makes it possible to construct an an explicit prior about the shape of (pieces of) roads, while still optimizing for pixel-accurate labeling.

Nevertheless, important properties of the road network are still not used. For example, T-junctions and crossings are characteristic network parts that could help to obtain a complete and plausible road network [36, 4], and also aspects like a preference for grid layouts and orthogonal intersections are still missing. Moreover, labeling ground truth for training data is time-consuming and costly, and in our scheme must be repeated not only for different sensors or imaging conditions, but also for different building styles, because of the changing context. Since map data is publicly available for many cities (*e.g.*, Open Street Map) it seems natural to use these as ground truth. This would allow one to directly learn road appearance, shape parameters (*e.g.*, width, straightness), and network topology (*e.g.*, intersection angles at junctions) from big data.

References

1. Bajcsy, R., Tavakoli, M.: Computer recognition of roads from satellite pictures. IEEE T. Systems, Man, and Cybernetics 6(9), 623 – 637 (1976)
2. Bas, E., Erdogmus, D.: Principal Curves as Skeletons of Tubular Objects. Neuroinformatics 9, 181 – 191 (2011)
3. Benmansour, F., Cohen, L.D.: Tubular Structure Segmentation Based on Minimal Path Method and Anisotropic Enhancement. IJCV 92, 192 – 210 (2011)
4. Chai, D., Förstner, W., Lafarge, F.: Recovering Line-networks in Images by Junction-Point processes. In: CVPR (2013)
5. Deschamps, T., Cohen, L.D.: Fast extraction of minimal paths in 3d images and applications to virtual endoscopy. Medical Image Analysis 5(4), 281–299 (2001)
6. Doucette, P., Agouris, P., Stefanidis, A.: Automated Road Extraction from High Resolution Multispectral Imagery. Photogrammetric Engineering & Remote Sensing 70(12), 1405 – 1416 (2004)
7. Fischler, M., Tenenbaum, J., Wolf, H.: Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. Computer Graphics and Image Processing 15, 201 – 223 (1981)
8. Gemert, J.C.V., Geusebroek, J., Veenman, C.J., Smeulders, A.W.M.: Kernel codebooks for scene categorization. In: European Conference on Computer Vision (2008)
9. Grote, A., Heipke, C., Rottensteiner, F.: Road network extraction in suburban areas. Photogrammetric Record 27(137), 8 – 28 (2012)
10. Heipke, C., Mayer, H., Wiedemann, C.: Evaluation of automatic road extraction. In: 3D Reconstruction and Modeling of Topographic Objects (1997)
11. Hinz, S., Baumgartner, A.: Automatic extraction of urban road networks from multi-view aerial imagery. ISPRS J. Photogrammetry and Remote Sensing 58, 83 – 98 (2003)
12. Hu, J., Razdan, A., Femiani, J.C., Cui, M., Wonka, P.: Road network extraction and intersection detection from aerial images by tracking road footprints. IEEE TGRS 45(12), 4144 – 4157 (2007)
13. Hussain, S.u., Triggs, B.: Visual recognition using local quantized patterns. In: European Conference on Computer Vision (2012)
14. Kohli, P., Ladicky, L., Torr, P.H.S.: Robust higher order potentials for enforcing label consistency. In: CVPR (2008)
15. Kohli, P., Ladicky, L., Torr, P.H.S.: Robust higher order potentials for enforcing label consistency. IJCV 82(3), 302–324 (2009)
16. Lacoste, C., Descombes, X., Zerubia, J.: Point Processes for unsupervised line network extraction in remote sensing. PAMI 27(10), 1568 – 1579 (2005)
17. Ladicky, L., Russell, C., Kohli, P., Torr, P.H.S.: Associative hierarchical CRFs for object class image segmentation. In: ICCV (2009)
18. Lafarge, F., Gimel'farb, G., Descombes, X.: Geometric Feature Extraction by a Multimarked Point Process. PAMI 32(9), 1597–1609 (2010)
19. Laptev, I., Mayer, H., Lindeberg, T., Eckstein, W., Steger, C., Baumgartner, A.: Automatic extraction of roads from aerial images based on scale space and snakes. MVA 12, 23 – 31 (2000)
20. Li, H., Yezzi, A.: Vessels as 4-D Curves: Global Minimal 4-D Paths to Extract 3-D Tubular Surfaces and Centerlines. IEEE TMI 26(9), 1213 – 1223 (2007)
21. Lindeberg, T.: Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics pp. 224–270 (1994)

22. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2004)
23. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and texture analysis for image segmentation. International Journal of Computer Vision (2001)
24. Mayer, H., Hinz, S., Bacher, U., Baltsavias, E.: A test of automatic road extraction approaches. In: IAPRS. vol. 36(3), pp. 209 – 214 (2006)
25. Mena, J., Malpica, J.: An automatic method for road extraction in rural and semi-urban areas starting from high resolution satellite imagery. Pattern Recognition Letters 26, 1201 – 1220 (2005)
26. Miao, Z., Shi, W., Zhang, H., Wang, X.: Road Centerline Extraction From High-Resolution Imagery Based on Shape Features and Multivariate Adaptive Regression Splines. IEEE GRSL 10(3), 583 – 587 (2013)
27. Mnih, V., Hinton, G.E.: Learning to detect roads in high-resolution aerial images. In: ECCV (2010)
28. Poullis, C., You, S.: Delineation and geometric modeling of road networks. ISPRS J. Photogrammetry and Remote Sensing 65, 165 – 181 (2010)
29. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: Conference on Computer Vision and Pattern Recognition (2007)
30. Shotton, J., Winn, J., Rother, C., Criminisi, A.: TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: ECCV (2006)
31. Stoica, R., Descombes, X., Zerubia, J.: A Gibbs Point Process for road extraction from remotely sensed images. IJCV 57(2), 121 – 136 (2004)
32. Türetken, E., Benmansour, F., Andres, B., Pfister, H., Fua, P.: Reconstructing Loopy Curvilinear Structures Using Integer Programming. In: CVPR (2013)
33. Türetken, E., Benmansour, F., Fua, P.: Automated Reconstruction of Tree Structures using Path Classifiers and Mixed Integer Programming. In: CVPR (2012)
34. Türetken, E., González, G., Blum, C., Fua, P.: Automated Reconstruction of Dendritic and Axonal Trees by Global Optimization with Geometric Priors. Neuroinformatics 9, 279 – 302 (2011)
35. Ünsalan, C., Sirmacek, B.: Road Network Detection Using Probabilistic and Graph Theoretical Methods. IEEE TGRS 50(11), 4441 – 4453 (2012)
36. Wegner, J.D., Montoya-Zegarra, J.A., Schindler, K.: A higher-order CRF model for road network extraction. In: CVPR (2013)
37. Wiedemann, C., Heipke, C., Mayer, H., Jamet, O.: Empirical evaluation of automatically extracted road axes. In: CVPR Workshops (1998)
38. Winn, J., Criminisi, A., Minka, T.: Object categorization by learned universal visual dictionary. In: CVPR (2005)
39. Zhao, T., Xie, J., Amat, F., Clack, N., Ahammad, P., Peng, H., Long, F., Myers, E.: Automated reconstruction of neuronal morphology based on local geometrical and global structural models. Neuroinformatics 9, 247 – 261 (2011)