Globally Optimal Closed-surface Segmentation for Connectomics

Bjoern Andres[1⋆], Thorben Kroeger[1⋆], Kevin L. Briggman[2], Winfried Denk[3], Natalya Korogod[4], Graham Knott[4], Ullrich Koethe[1], and Fred A. Hamprecht[1]

[1] HCI, University of Heidelberg
[2] NIH, Bethesda
[3] MPI for Medical Research, Heidelberg
[4] EPFL, Lausanne

Abstract. We address the problem of partitioning a volume image into a previously unknown number of segments, based on a likelihood of merging adjacent supervoxels. Towards this goal, we adapt a higher-order probabilistic graphical model that makes the duality between supervoxels and their joint faces explicit and ensures that merging decisions are consistent and surfaces of final segments are closed. First, we propose a practical cutting-plane approach to solve the MAP inference problem to global optimality despite its NP-hardness. Second, we apply this approach to challenging large-scale 3D segmentation problems for neural circuit reconstruction (Connectomics), demonstrating the advantage of this higher-order model over independent decisions and finite-order approximations.

1 Introduction

This paper studies the problem of partitioning a volume image into a previously unknown number of segments, based on a likelihood of merging adjacent supervoxels.

We choose a graphical model approach in which binary variables are associated with the joint faces of supervoxels, indicating for each face whether the two adjacent supervoxels should belong to the same segment (0) or not (1). Models of low order can lead to inconsistencies where a face is labeled as 1 even though there exists a path from one of the adjacent segments to the other along which all faces are labeled as 0. As a result, the union of all faces labeled as 1 need not form closed surfaces. Such inconsistencies can be excluded by a higher-order conditional random field (CRF) [1] that constrains the binary labelings to the multicut polytope [2], thus ensuring closed surfaces. While the number of multicut constraints can be exponential [3], constraints that are violated by a given labeling can be found in quadratic time [4]. The MAP inference problem can therefore be addressed by the cutting-plane method, i.e. by solving a sequence of relaxed problems to global optimality until no more constraints are violated [4].

Here, we show that the optimization scheme described in [1] is unsuitable for large 3D segmentations where the supervoxel adjacency graph is denser and non-planar. We therefore extend the cutting-plane approach by adding only constraints which are facet-defining by a property of the multicut polytope (Section 4.3), a double-ended, parallel search to find violated constraints (Section 4.2) and a problem-specific warm-start

⋆ Contributed equally.

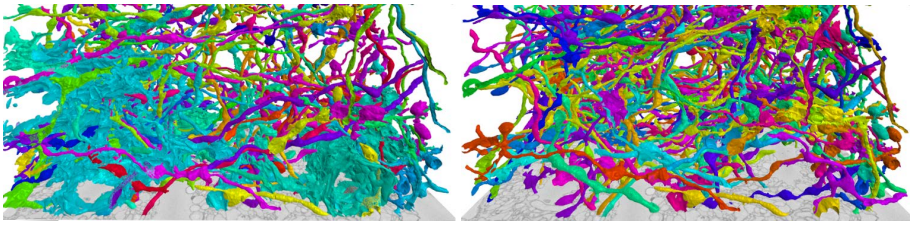


Fig. 1: Segmenting volume images based on a likelihood of merging adjacent supervoxels is difficult if merging decisions are made independently. *Left:* Segmentation errors remain a problem even if the model is biased optimally with $\beta = 0.8$ in Eq. 5 (note the under-segmentation at the bottom and missing segments at the top). *Right:* Multicut constraints alleviate this problem and allow for an unbiased, parameter-free model.

heuristic (Section 4.4). This approach scales gracefully to volume images that consist of 10^9 voxels and are initially segmented into 10^6 supervoxels (Section 5).

The fact that exact MAP inference remains tractable at this scale is important for the reconstruction of neural circuits form electron microscopic volume images (Fig. 1) where multicut constraints substantially improve the quality of segmentations across different imaging techniques (Section 5).

For this quantitative analysis, we manually segmented 10^9 voxels of two different datasets acquired at different laboratories using different imaging techniques and assess the performance of three models: the simplest, *local model* uses learned unary conditional probabilities that state, independently, if each face should be *on* (establish part of an object boundary) or *off* (merge adjacent supervoxels). Optimization of this model almost always results in inconsistent labelings (Figs. 3b, 3e). The intermediate, *finite-order model* guarantees consistency across a small local horizon, leading to better results. The best results are obtained with the *fully constrained model* which admits only labelings that are globally consistent. With the cutting-plane approach proposed here, the full model is often faster to optimize than the finite-order approximation. Figs. 3, 4, and 5 summarize why the fully constrained method is the one we recommend.

2 Related Work

The problem we address is known as the multicut problem [5] in combinatorial optimization and as correlation clustering [6,7] in statistics, and it is a special case of the partition problem [2]. Both problems are NP-hard [6,8]. Instances of the multicut problem have been solved by tightening an outer approximation of the multicut polytope [9] via cutting planes [10,4]. In computer vision, this technique has been applied in [11,12,13] where the relaxed LP is solved first, as well as in [1] where the integrality constraints are kept throughout the cutting-plane loop. Cutting planes have also been used to enforce connectivity in foreground vs. background image segmentation [14,15,16]. Here, we build on the probabilistic formulation in [1] but without the likelihood terms w.r.t. geometry that were shown to have a negligible effect on segmentations of photographs.

We concentrate on exact MAP inference for which we propose a cutting-plane approach that is efficient for large-scale 3D segmentation. In particular, we discuss the efficient search for violated constraints that are facet-defining and the parallelization of this search. We measure how the optimization runtime scales with the size of the image and the number of variables, respectively, with and without improvements.

Results were shown in [1] for photographs in the Berkeley Segmentation Dataset [17] which consist of 10^5 pixels that were initially partitioned into 10^4 superpixels, leading to optimization problems with at most 10^4 variables. Here, we segment 3D images of up to 10^9 voxels which are initially partitioned into 10^5 supervoxels, leading to 100 times larger optimization problems with 10^6 variables.

3 Probabilistic Model

In order to make the duality between supervoxels and their joint faces explicit, we build a cell-complex $(C, \prec, \dim)$ representation [18] of the supervoxel segmentation by connected component labeling of the finest possible grid-cell topology. This yields disjoint sets of supervoxels C_3, joint-faces between supervoxels C_2, lines C_1 and points C_0. The function $\dim : C \to \mathbb{N}$ maps each cell to its dimension. For any cells $c, c' \in C = \cup_{j=0}^{3} C_j$, $c \prec c'$ indicates that c bounds c', which implies $\dim(c) < \dim(c')$. This representation has the advantage that multiple disconnected joint-faces seperating the same pair of supervoxels can be treated independently for feature extraction and classification (see below). We model the posterior probability of a joint labeling $y \in \{0,1\}^{|C_2|}$ of all faces, given

- $m \in \mathbb{N}$ features $f_c \in \mathbb{R}^m$ of each face $c \in C_2$, summarized in a vector $F \in \mathbb{R}^{m|C_2|}$
- the bounding relation between faces and supervoxels, encoded in a topology matrix $T \in \{0,1\}^{|C_2| \times |C_3|}$ in which $T_{cc'} = 1$ if and only if $c \prec c'$.

We assume that features are independent of the topology, a) $F \perp\!\!\!\perp T$, b) $F \perp\!\!\!\perp T \mid y$, that features of any faces $c \neq c'$ are independent, c) $f_c \perp\!\!\!\perp f_{c'}$, d) $f_c \perp\!\!\!\perp f_{c'} \mid y$ and that the label of any face c is independent of the label of any face $c' \neq c$ given the features f_c, e) $y_c \perp\!\!\!\perp y_{c'} \mid f_c$. (Note, however, that $y_c \not\perp\!\!\!\perp y_{c'} \mid T$). From these conditional independence assumptions follows

$$
\begin{aligned}
p(y|F,T) &= \frac{p(F,T|y)p(y)}{p(F,T)} \overset{(a,b)}{=} \frac{p(F|y)p(T|y)p(y)}{p(F)p(T)} \\
&\overset{(c,d)}{=} \frac{p(T|y)}{p(T)} \prod_{c \in C_2} \frac{p(f_c|y)p(y)}{p(f_c)} = \frac{p(T|y)}{p(T)} \prod_{c \in C_2} p(y|f_c) \\
&\overset{(e)}{=} \frac{p(T|y)}{p(T)} \prod_{c \in C_2} p(y_c|f_c) = \frac{p(T|y)}{p(T)} \prod_{c \in C_2} \frac{p(f_c|y_c)p(y_c)}{p(f_c)} \\
&\propto \; p(T|y) \prod_{c \in C_2} p(f_c|y_c)p(y_c) \;\; . \qquad (1)
\end{aligned}
$$

The prior $p(y_c)$ is assumed to be identical for all faces and is specified with a single design parameter $\beta \in (0,1)$ as $p(y_c = 1) = \beta$ and $p(y_c = 0) = 1 - \beta$.

The likelihood $p(T|y)$ is the instrument that is used to enforce consistency: While uninformative for all consistent labelings, it assigns zero probability to all inconsistent labelings. Consistent labelings $y \in \mathrm{MC}$, i.e. labelings inside the multicut polytope[5], are defined in (3) w.r.t. the set $\mathrm{SC}(n)$ of all simple cycles $(s_1, t_1, \dots s_n, t_n, s_1)$ over $n \in \mathbb{N}$ pairwise distinct supervoxels $(s_1, \dots, s_n)$ via pairwise distinct faces $(t_1, \dots t_n)$. The inequalities in (3) ensure that along each cycle either none or more than one face is labeled as 1:

$$\mathrm{SC}(n) = \left\{ \begin{array}{l|l} (c_j)_{j \in \mathbb{N}_{2n+1}} \\ \text{in } C \end{array} \begin{array}{l} c_1 = c_{2n+1} \wedge \forall j \in \mathbb{N}_n : \ c_{2j-1} \in C_3 \wedge c_{2j} \in C_2 \\ \wedge \ \forall j \in \mathbb{N}_n : c_{2j} \prec c_{2j-1} \wedge c_{2j} \prec c_{2j+1} \\ \wedge \ \forall j,k \in \mathbb{N}_n : j = k \vee (c_{2j} \neq c_{2k} \wedge c_{2j-1} \neq c_{2k-1}) \end{array} \right\}$$

$$\mathrm{MC}(n) = \left\{ y \in \{0,1\}^{|C_2|} \,\middle|\, \forall (s_1, t_1, \dots s_n, t_n, s_1) \in \mathrm{SC}(n) : \ y_{t_1} \leq \textstyle\sum_{j=2}^{n} y_{t_j} \right\} \quad (2)$$

$$\mathrm{MC} = \bigcap_{j=1}^{|C_2|} \mathrm{MC}(j) \quad (3)$$

The likelihood $p(f_c|y_c)$ is learned by means of a random forest. More precisely, we learn $\hat{p}(y_c|f_c)$ from class-balanced training data, i.e. with $\hat{p}(y_c) = 0.5$, and assume $p(f_c|y_c) = \hat{p}(f_c|y_c)$. Therefore, $p(f_c|y_c) \propto \hat{p}(y_c|f_c)\hat{p}(f_c)$ and thus,

$$p(y|F,T) \propto p(T|y) \prod_{c \in C_2} \hat{p}(y_c|f_c) p(y_c) \ . \quad (4)$$

For comparison, we consider two simpler models, a *local model* in which $p(T|y)$ is uniform and thus, faces are labeled independently,

$$p'(y|F,T) \ \propto \prod_{c \in C_2} \hat{p}(y_c|f_c) p(y_c) \ , \quad (5)$$

and a *finite-order approximation* of (4) in which not all multicut constraints need to be fulfilled but only those that correspond to cycles up to length 4. With $p''(T|y) = \text{const.}$ if $y \in \cap_{j=1}^{4} \mathrm{MC}(j)$ and $p''(T|y) = 0$, otherwise,

$$p''(y|F,T) \ \propto \ p''(T|y) \prod_{c \in C_2} \hat{p}(y_c|f_c) p(y_c) \ . \quad (6)$$

4 MAP Inference

4.1 Integer Linear Programming Problem (ILP)

Instead of maximizing (4), we minimize its negative log likelihood by solving the ILP

$$\begin{array}{l} \min\limits_{y \in \{0,1\}^{|C_2|}} w^{\mathrm{T}} y \\ \text{subject to } y \in \mathrm{MC} \end{array} \quad (7)$$

[5] The multicut polytope is the convex hull of MC, by Lemma 2.2 in [2].

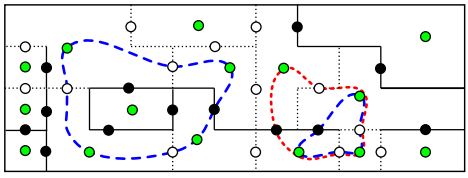


Fig. 2: Depicted is one slice of a supervoxel segmentation. Supervoxels (green dots) are bounded by joint faces which are labeled as 0 (white dots) or 1 (black dots). The simple circles drawn in long-dashed blue (short-dashed red) are chordless (chordal).

with $w \in \mathbb{R}^{|C_2|}$ such that $\forall j \in \{1, \ldots, |C_2|\}$:

$$w_j = \log \frac{p(y_j = 0|f_j)}{p(y_j = 1|f_j)} + \log \frac{1-\beta}{\beta} \tag{8}$$

We start by solving the (trivial) ILP without multicut constraints. We then search for constraints that are violated by the solution, add these to the constraint pool and re-solve the constrained ILP using the branch-and-cut algorithm of a state-of-the-art solver. This procedure is repeated until no more multicut constraints are violated and thus, the original problem (7) has been solved to optimality.

4.2 Search for Violated Constraints

If any multicut constraints are violated by a given labeling y, then at least one can be found by considering all faces $c \in C_2$ with $y_c = 1$ and looking for a shortest cycle $(r_1, t_1 = c, r_2, t_2, \ldots, r_n, t_n, r_1) \in \mathrm{SC}(n)$ with $n \in \mathbb{N}$ along which all other faces $t_2, \ldots, t_n$ are labeled as 0 [1,4]. Shortest cycles correspond to constraints with minimal numbers of variables.

Double-ended search. Although a breadth-first search for such a cycle can be carried out in time $O(|C_2| + |C_3|)$, starting from either r_1 or r_2, the absolute runtime to perform this task for all relevant faces can be comparable to that of solving the ILP (Section 5). We therefore propose to grow two search trees, rooted at r_1 and r_2, simultaneously. This saves runtime because both trees are only half as deep as a single one would be, at the point when a shortest cycle is found.

Parallelization. Shortest paths need to be found for many faces, from one adjacent supervoxel to the other, and yet not so many that it would be profitable to solve the All Pairs Shortest Path problem. We therefore propose to solve Single Pair Shortest Path problems in parallel with a space complexity that is linear in the number of threads. In practice, we use OpenMP for this embarrassingly parallel task.

4.3 Chordality Check

Not all inequalities in (3) define a facet of the multicut polytope due to the following

Theorem [2]: If $n \in \mathbb{N}$ and $(c_j) = (s_1, t_1, \ldots s_n, t_n, s_1) \in \mathrm{SC}(n)$, the inequality $y_{t_1} \leq \sum_{j=2}^{n} y_{t_j}$ is facet-defining if and only if (c_j) is chordless[6] (Fig. 2).

[6] A path is chordless if each node is connected only to its successor and predecessor. Here, the path of segments via joint faces is chordless if each segment is connected by a face only to its successor and predecessor.

We exploit this algorithmically by adding violated inequalities only if these correspond to chordless cycles (Fig. 2).

4.4 Warm Start Heuristic

When violated constraints are added, the solution of the relaxed problem becomes infeasible and thus, the upper bound on the global minimum is lost. However, the structure of the problem allows us to find a new feasible solution efficiently: A given labeling $y \in \{0,1\}^{|C_2|}$ is mapped to a labeling on the multicut polytope by labeling all variables of all violated inequalities as 0. Since violated inequalities have already been found, this heuristic does not change the runtime complexity of the optimization scheme overall.

5 Applications

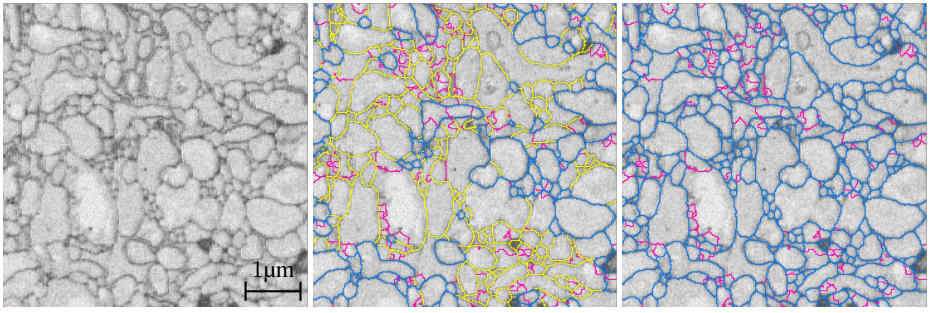
5.1 SBEM Volume Image

A volume image referred to as E1088 in [19] was acquired using serial block-face electron microscopy (SBEM) [20] and shows a section of rabbit retina at the almost isotropic resolution of $22 \times 22 \times 30$ nm^3. A small subset is shown in Fig. 3a. The bright intra-cellular space that makes up more than 90% of the volume contrasts the stained extra-cellular space that forms thin membranous faces. This staining [21] simplifies the automated segmentation because no intra-cellular structures such as mitochondria or vesicles are visible (Fig. 3d shows a different staining). A supervoxel segmentation is obtained as described in Appendix C.

Features f_c of each face $c \in C_2$ described in Appendix B include statistics of voxel features over c as well as characteristics of the two supervoxels that are bounded by c. In order to learn $p(y_c|f_c)$ by means of a random forest, 437 faces per class have been labeled interactively in a subset of 250^3 voxels in about one hour, starting with obvious cases and continuing where the predictions needed improvement. A custom extension of ILASTIK [22] implements this online learning workflow. The user can mark faces as "on", "off" or "unlabeled" via a single mouse click and inspect intermediate predictions by viewing faces colored according to $\hat{p}(y_c|f_c)$.

Qualitative results on independent test data are shown in Fig. 3a-c. The global maximum of the local model (5) is inconsistent, i.e. not all surfaces are closed. A consistent labeling with closed surfaces and thus a segmentation is obtained by merging supervoxels transitively, i.e. whenever there exists a path from one supervoxel to the other along which all faces are labeled as 0 (cf. Section 4.4), regardless of how many faces between these supervoxels are labeled as 1. This mapping to the multicut polytope is biased towards under-segmentation (Fig. 3b). In contrast, the global maximum of the fully constrained model (4) is consistent and directly yields a segmentation with closed surfaces (Fig. 3c). Note also the quality of the initial supervoxel segmentation from which only 19.3% of all faces are removed in the found global optimum.

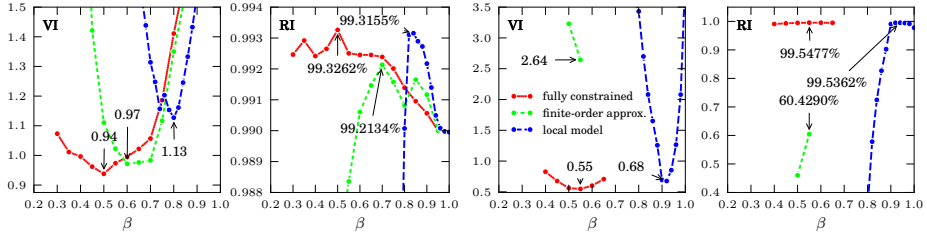
Quantitatively, the effect of introducing multicut constraints is shown in Fig. 4a where maxima of (4), the local model (5) and the finite-order approximation (6) are compared to a man-made segmentation of $400 \times 200 \times 200$ voxels (Appendix A) in



(a) SBEM raw data (b) w/o constraints, $\beta = 0.5$ (c) with constraints, $\beta = 0.5$

(d) FIBSEM raw data (e) w/o constraints, $\beta = 0.5$ (f) with constraints, $\beta = 0.5$

Fig. 3: Segmentations of the SBEM dataset (a-c, 242^2 voxels) and FIBSEM dataset (d-f, 512^2 voxels). Faces are colored to show if the algorithm decides that these are part of a segment boundary (blue) or not (magenta). Importantly, yellow faces are decided to be part of a segment boundary, but are ignored because the bounded supervoxels are merged elsewhere.



(a) SBEM ($400 \times 200 \times 200$ voxels) (b) FIBSEM ($900 \times 900 \times 900$ voxels)

Fig. 4: Variation of information (VI) and Rand index (RI) from a comparison of man-made segmentations (Appendix A) with segmentations obtained as maxima of the fully constrained model (Eq. 4), the finite-order approximation (Eq. 6), and the local model (Eq. 5), for various priors β. Exact optimization of the finite-order approximation becomes intractable for most β on the FIBSEM data. In contrast, optima of the full model are found in less than 13 minutes. (Fig. 5b).

Dataset	Problem size			Runtime [min.] for optimizing (4)			
				[1]		Our work	
	Voxels	$\lvert C_3\rvert$	$\lvert C_2\rvert$	Cplex	Gurobi	Cplex	Gurobi
SBEM	800^3	194 809	1 603 683	249.0	299.0	**11.4**	**12.9**
FIBSEM	800^3	73 099	624 747	386.2	291.7	**9.2**	**7.2**

Table 1: Globally optimal closed-surface segmentations of the SBEM and and FIBSEM volume images of 800^3 voxels are found in less than 13 minutes using the proposed cutting-plane method that is about 22 times as fast as the optimization scheme in [1].

terms of the Variation of Information (VI) [23] and Rand Index (RI) [24]. The overall best segmentation is obtained from (4), i.e. with all multicut constraints, and without an artificial bias $\beta \neq 0.5$; note that the bias term in (8) vanishes for $\beta = 0.5$. Without any multicut constraints, the best segmentation, obtained for $\beta = 0.8$, is worse in terms of both VI and RI. Segmentations of intermediate quality are obtained from the finite-order approximation (6).

Runtimes for optimizing (4) are shown in Fig. 5a and Tab. 1 for our C++ implementation. The ILP (7) is solved with IBM ILOG Cplex and Gurobi alternatively, with the duality gap set to 0 in order to obtain globally optimal solutions. Global optima of (4), for blocks of 800^3 voxels with $1.6 \cdot 10^6$ and $6 \cdot 10^5$ faces (and variables), respectively, are found in less than 13 minutes (Tab. 1), about 22 times as fast as with the optimization scheme in [1].

Maximizing (4) can be faster than maximizing the finite-order approximation (6) that does not guarantee closed surfaces and yields worse segmentations empirically (runtimes not shown). We therefore recommend to use all multicut constraints.

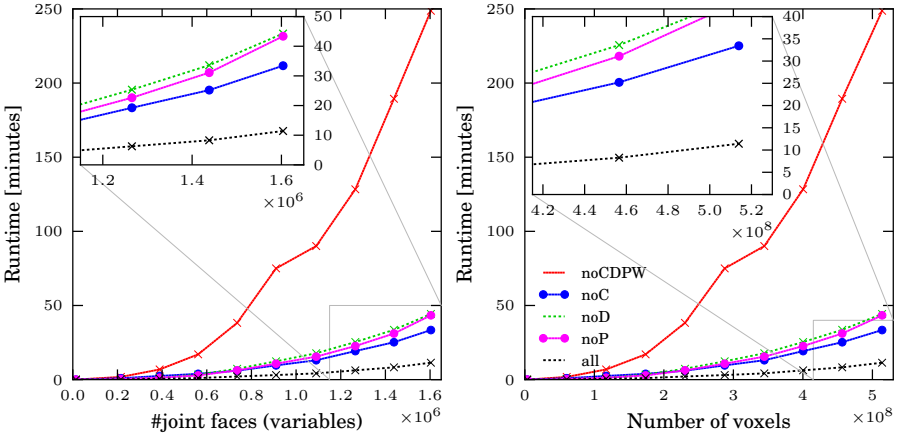
5.2 FIBSEM Volume Image

A volume image acquired with a focused ion beam serial-section electron microscope (FIBSEM) [25] shows a section of adult mouse somatosensory cortex at the almost isotropic resolution of $5 \times 5 \times 6$ nm^3. A small subset is shown in Fig. 3d. Intra- and extracellular space are indistinguishable by brightness and texture. Not only cell membranes but also intra-cellular structures such as mitochondria and vesicles are visible due to a different staining. The resolution is four times as high as that of the SBEM image. However, intra-cellular structures have membranous surfaces themselves and thus make the problem of segmenting entire cells more difficult. A supervoxel segmentation is obtained as described in Appendix C.

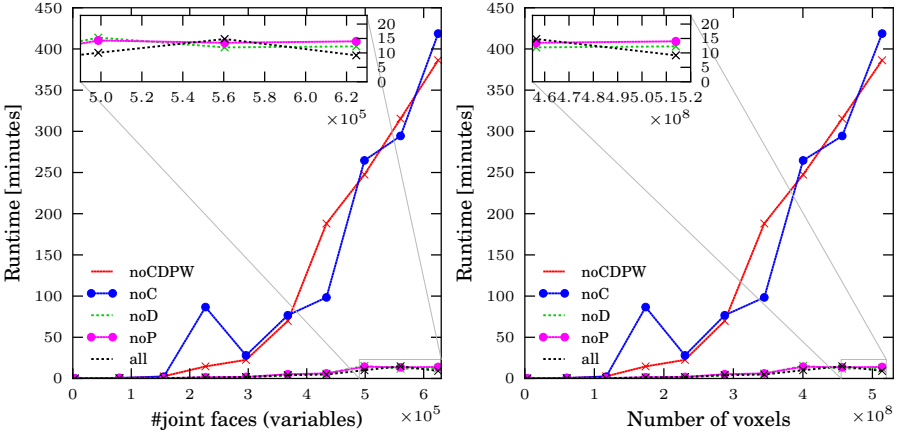
We use the same features as for the SBEM dataset but adjusted in scale. Our labeling strategy was to annotate membranes of both cells and mitochondria[7] as $y_c = 1$.

Qualitative results are shown in Fig. 3d-f for a slice of 512^3 voxels. The MAP labeling of the local model (5) is inconsistent and almost all faces are removed by

[7] It is still possible to obtain the geometry of cells because mitochondria can be detected reliably [26], even via their mean gray-value once a segmentation is available. Segments which are classified as mitochondria are disregarded when computing VI [23] and RI [24].

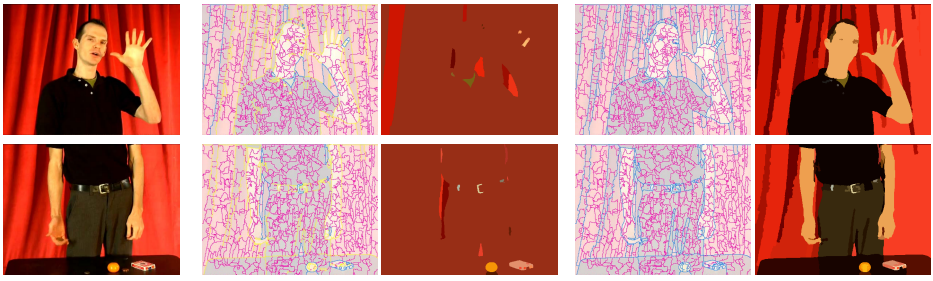


(a) SBEM dataset



(b) FIBSEM dataset

Fig. 5: Wall clock runtimes for optimizing (4) with $\beta = 0.5$ (on an 8-core Intel i7 at 2.8 GHz), with and without improvements to the optimization scheme in [1], for increasing problem size (number of joint faces of supervoxels, left, and number of voxels, right), for datasets ranging in size from 150^3 through 800^3 voxels. *all*: proposed optimization scheme, *noC*: no chordality check, *noD*: no double-ended search, *noP*: no parallel search, *noCDPW*: no improvements. Analogous plots for Gurobi instead of Cplex are provided as supplementary material.



(a) frames 1285, 1950 (b) without multicut constraints, $\beta = 0.5$ (c) with multicut constraints, $\beta = 0.5$

Fig. 6: Color video segmentation. Supervoxel faces are colored blue if $y_c = 1$, magenta if $y_c = 0$ and yellow if the decision is inconsistent, i.e. $y_c = 1$ but the adjacent supervoxels are merged. Segments are filled with their mean color.

transitivity (note the lack of blue faces in Fig. 3e). In contrast, the global maximum of the fully constrained model (4) is consistent (Fig. 3f). In contrast to the SBEM dataset in which only 19.3% of the faces are removed, 80.7% of all faces are removed here, due to the inferior supervoxel segmentation.

Quantitative results are shown in Fig. 4b. VI and RI are w.r.t. a complete segmentation of 900^3 voxels carried out by a neurobiologist (Appendix A). Similarly as for the SBEM volume segmentation, the best segmentations are obtained from the full model (4). Exact MAP inference for the finite-order approximation (6) becomes intractable for most β.

Runtimes are shown in Fig. 5b for problem instances from blocks between 150^3 and 800^3 voxels. Unlike for the SBEM dataset (Fig. 5a), the overall speedup is dominated by the chordality check.

5.3 Video Segmentation

As a proof-of-concept, we applied the same model to bottom-up video segmentation, treating the video[8] as a three-dimensional (x,y,t)-volume. Obtaining a supervoxel segmentation that strikes a balance between negligible under-segmentation and a small number of excessive supervoxels has proven difficult. We settled for a marker-based watershed transformation of the color gradient magnitude in the Lab color space. A 486×360 video with 1000 frames (Fig. 6a) is thus partitioned into $|C_3| = 22\,056$ supervoxels with $|C_2| = 256\,734$ joint faces.

As features f_c, we use (i) the mean (over the face c) of the 2D patch features in [1] which are computed per frame, and (ii) the absolute distance of color histograms of the bounded supervoxels. A training set of 153 labels per class was acquired using the

[8] Frames 1000–2000 of the video at youtube.com/watch?v=YN0I-TZFn58

same tool and protocol as for the SBEM dataset. Only faces intersecting frame 1015 were labeled.

Qualitative results are shown in Fig. 6. A video of the segmentation is provided as supplementary material. While the finite-order model is intractable for this problem, global optimization of the fully constrained model (4) with $\beta = 0.5$ takes 131 seconds.

6 Discussion

In a fully connected graph with $n \in \mathbb{N}$ nodes, the $\binom{n}{3}$ cycle constraints of order 3 imply all $\sum_{j=0}^{n-4} \frac{n!}{j!}$ higher-order cycle constraints. However, in the sparse graphs that we consider, the higher-order constraints need to be dealt with explicitly. An example is depicted in Fig. 2 where the constraint that corresponds to the blue cycle on the left has order 4, excluding from the feasible set a locally closed loop that is globally inconsistent. Our experiments have shown that including these higher-order constraints is essential to achieve the best performance w.r.t. ground truth.

When inconsistent labelings are permitted, unlike in the fully constrained model (4), and mapped to the multicut polytope as described in Section 4.4, the risk of false mergers is higher in 3D than in 2D because there are on average more and shorter paths from one supervoxel to another along which faces can be incorrectly labeled as 0. We therefore expect multicut constraints to be more important in 3D than in 2D.

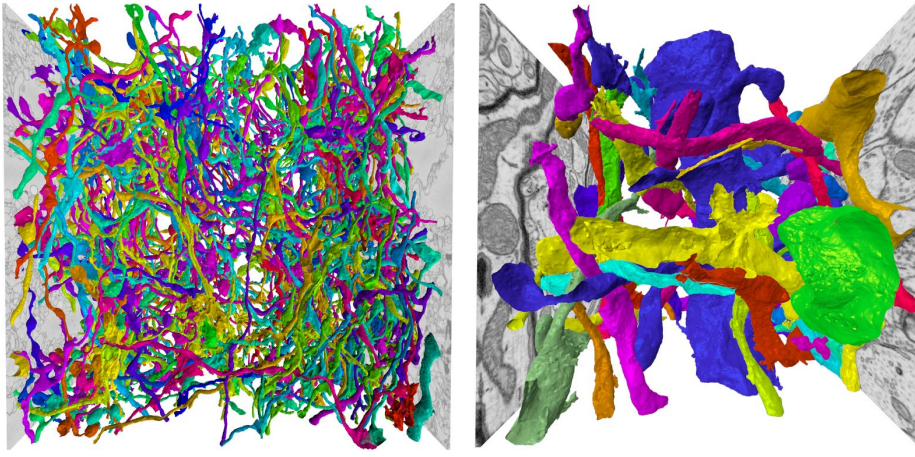
Solving (7) as proposed here requires the solution of problems of an NP-hard class. Whether or not this is tractable in practice depends on the quality of the predictions $\hat{p}(y_c|f_c)$. The learning of this function is therefore important.

The warm start heuristic described in Section 4.4 is biased maximally towards under-segmentation. Smarter heuristics that explore local neighborhoods in the cell complex are subject of future work. The model (4) can be grafted on any method, 2D or 3D, that finds superpixels or supervoxels and probabilities that joint boundaries of these should be preserved or removed. Whether it can improve segmentations of images acquired by transmission electron microscopy is subject of future research.

The recontruction of neural circuits such as a neocortical column or the central nervous system of *Drosophila melanogaster* will eventually require the segmentation of volume images of 10^{12} voxels. The result that 10^9 voxels can be segmented by optimizing a non-submodular higher-order multicut objective exactly on a single computer in 13 minutes is encouraging.

7 Conclusion

We have addressed the problem of segmenting volume images based on a learned likelihood of merging adjacent supervoxels. To solve this problem, we have adapted a probabilistic model that enforces consistent decisions via multicut constraints to 3D cell topologies and suggested a fast scheme for exact MAP inference. The resulting 22-fold speedup has allowed us to systematically study the positive effect of multicut constraints in large-scale 3D segmentation problems for neural circuit reconstruction. The best segmentations have been obtained for an unbiased parameter-free model with multicut constraints.

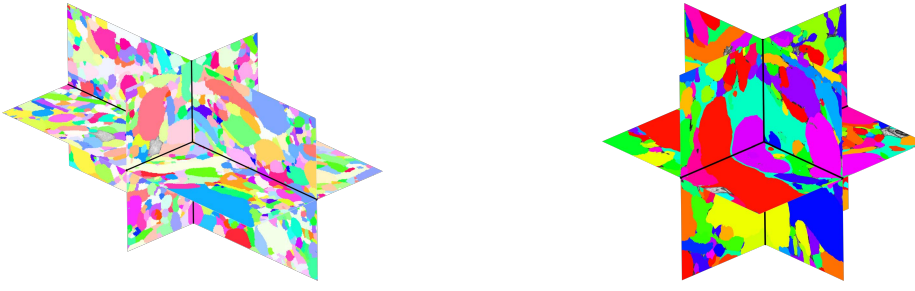


(a) SBEM, $1400 \times 1400 \times 600$ voxels (b) FIBSEM, $900 \times 900 \times 900$ voxels

Fig. 7: The largest volume images we have segmented with multicut constraints. Shown are 600 segments in the SBEM dataset and 50 objects in the FIBSEM dataset.

A Acquisition of Ground Truth

We manually segmented subsets of the SBEM and the FIBSEM dataset (Fig. 8) using the interactive method [27]. Each object was segmented independently. Some independently segmented objects needed correction because there was overlap. This asserts a consistent ground truth and shows that the segmentation problem is non-trivial, even for a human. For the SBEM dataset, 528 objects (90.8% of $400 \times 200 \times 200$ voxels) were segmented by one expert in two weeks. For the FIBSEM dataset, 514 objects (96.8% of a cubic block of 900^3 voxels) were segmented by a neurobiologist in three weeks.



(a) SBEM $(400 \times 200 \times 200)$ (b) FIBSEM (900^3)

Fig. 8: Man-made segmentations (ground truth). Shown are orthogonal slices in which a random color is assigned to each segment.

Index	Feature	Details
1	Size of the face	
2–3	Sizes v_1 and v_2 of	$(v_1 + v_2)^{1/3}$
	adjacent supervoxels	$\lvert v_1 - v_2\rvert^{1/3}$
4–10	Bilateral filter b	*
11–17	Gradient magnitude	*
18–24	Hessian matrix of	max. eigenvalue,*
25–31	Voxel classifier	*

(a) Features of supervoxel faces

Index	Feature
1	Volume image
2	Bilateral filter
	$w_{\sigma_s}(r) = \frac{1}{\sigma_s(2\pi)^{3/2}} \exp\left(-\frac{r^2}{2\sigma_s^2}\right)$
	$w_{\sigma_v}(v) = \frac{1}{1+\frac{v^2}{\sigma_v^2}}$
3–4	Gradient magnitude
5–16	Structure Tensor eigenvalues
17–28	Hessian matrix eigenvalues

(b) Features of voxel neighborhoods

Table 2: Features of supervoxel faces. The statistics (*) include the min, max, mean, median, standard deviation, 0.25- and 0.75-quantile over all voxels adjacent to the face.

B Features

From every joint face of adjacent supervoxels, 31 features are extracted (Tab. 2a). One of these features is the response of a Random Forest that discriminates between membranes on the one hand and intra-/extra-cellular tissue on the other hand, based on 28 rotation-invariant non-linear features of local neighborhoods of 11^3 voxels (Tab. 2b).

C Supervoxel Segmentation

We computed supervoxels by marker-based watersheds. To obtain an elevation map and markers for the SBEM dataset, we train a random forest classifier to distinguish between two classes of voxels, *extra-cellular space* and *intra-cellular space*, based on rotation invariant features of local neighborhoods (Tab. 2b and Appendix B). As training data, 1600 voxels per class were labeled interactively in two subsets of 150^3 voxels which has taken three hours using ILASTIK [22]. Predicted probabilities are used as elevation levels. Connected components of at least three voxels classified as intra-cellular space are used as markers. For the FIBSEM dataset, the elevation level is defined as the largest eigenvalue of the Hessian matrix at scale 1.6. Markers are taken to be maximal plateaus of the raw data that consist of at least two voxels.

References

1. Andres, B., Kappes, J.H., Beier, T., Köthe, U., Hamprecht, F.A.: Probabilistic image segmentation with closedness constraints. In: ICCV. (2011) 1, 2, 3, 5, 8, 9, 10
2. Chopra, S., Rao, M.R.: The partition problem. Math. Program. **59** (1993) 87–115 1, 2, 4, 5
3. Alt, H., Fuchs, U., Kriegel, K.: On the number of simple cycles in planar graphs. In Möhring, R., ed.: Graph-Theoretic Concepts in Computer Science. Volume 1335 of LNCS. Springer (1997) 15–24 1

4. Grötschel, M., Wakabayashi, Y.: A cutting plane algorithm for a clustering problem. Math. Program. **45** (1989) 59–96 1, 2, 5
5. Costa, M.C., Letocart, L., Roupin, F.: Minimal multicut and maximal integer multiflow: A survey. European J. of Oper. Res. **162** (2005) 55–69 2
6. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Machine Learning **56** (2004) 89–113 2
7. Demaine, E.D., Emanuel, D., Fiat, A., Immorlica, N.: Correlation clustering in general weighted graphs. Theoretical Computer Science **361** (2006) 172–187 2
8. Garey, M., Johnson, D.: Computers and Intractability: A Guide to the Theory of NP-Completeness. Freeman, New York (1979) 2
9. Sontag, D., Jaakkola, T.: New outer bounds on the marginal polytope. In: NIPS. (2008) 2
10. Barahona, F., Grötschel, M., Jünger, M., Reinelt, G.: An application of combinatorial optimization to statistical physics and circuit layout design. Oper. Res. **36** (1988) 493–513 2
11. Kappes, J.H., Speth, M., Andres, B., Reinelt, G., Schnörr, C.: Globally optimal image partitioning by multicuts. In: EMMCVPR. (2011) 2
12. Kim, S., Nowozin, S., Kohli, P., Yoo, C.D.D.: Higher-order correlation clustering for image segmentation. In: NIPS. (2011) 2
13. Nowozin, S., Jegelka, S.: Solution stability in linear programming relaxations: graph partitioning and unsupervised learning. In: ICML. (2009) 2
14. Vicente, S., Kolmogorov, V., Rother, C.: Graph cut based image segmentation with connectivity priors. In: CVPR. (2008) 2
15. Nowozin, S., Lampert, C.H.: Global connectivity potentials for random field models. In: CVPR. (2009) 2
16. Lempitsky, V., Kohli, P., Rother, C., Sharp, T.: Image segmentation with a bounding box prior. In: ICCV. (2009) 2
17. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV. (2001) 3
18. Hatcher, A.: Algebraic Topology. Cambridge Univ. Press (2002) 3
19. Helmstaedter, M., Briggman, K.L., Denk, W.: High-accuracy neurite reconstruction for high-throughput neuro-anatomy. Nature Neuroscience **14** (2011) 1081–1088 6
20. Denk, W., Horstmann, H.: Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS Biology **2** (2004) e329 6
21. Briggman, K.L., Helmstaedter, M., Denk, W.: Wiring specificity in the direction-selectivity circuit of the retina. Nature **471** (2011) 183–188 6
22. Sommer, C., Straehle, C., Koethe, U., Hamprecht, F.A.: Ilastik: Interactive Learning and Segmentation Toolkit. In: ISBI. (2011) 6, 13
23. Meilă, M.: Comparing clusterings – an information based distance. J. of Multivariate Anal. **98** (2007) 873–895 8
24. Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association **66** (1971) pp. 846–850 8
25. Knott, G., Marchman, H., Wall, D., Lich, B.: Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. J. Neurosci. **28** (2008) 2959–2964 8
26. Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P.: Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. IEEE Transactions on Medical Imaging **31** (2012) 474–486 8
27. Straehle, C., Köthe, U., Knott, G., Hamprecht, F.: Carving: Scalable interactive segmentation of neural volume electron microscopy images. In: MICCAI. (2011) 12