

# The Mutex Watershed and its Objective: Efficient, Parameter-Free Graph Partitioning

Steffen Wolf , Alberto Bailoni , Constantin Pape , Nasim Rahaman, Anna Kreshuk , Ullrich Köthe , and Fred A. Hamprecht

**Abstract**—Image partitioning, or segmentation without semantics, is the task of decomposing an image into distinct segments, or equivalently to detect closed contours. Most prior work either requires seeds, one per segment; or a threshold; or formulates the task as multicut / correlation clustering, an NP-hard problem. Here, we propose an efficient algorithm for graph partitioning, the “Mutex Watershed”. Unlike seeded watershed, the algorithm can accommodate not only attractive but also repulsive cues, allowing it to find a previously *unspecified* number of segments without the need for explicit seeds or a tunable threshold. We also prove that this simple algorithm solves to global optimality an objective function that is intimately related to the multicut / correlation clustering integer linear programming formulation. The algorithm is deterministic, very simple to implement, and has empirically linearithmic complexity. When presented with short-range attractive and long-range repulsive cues from a deep neural network, the Mutex Watershed gives the best results currently known for the competitive ISBI 2012 EM segmentation benchmark.

**Index Terms**—Image segmentation, partitioning algorithms, greedy algorithms, optimization, integer linear programming, machine learning, convolutional neural networks



## 1 INTRODUCTION

MOST image partitioning algorithms are defined over a graph encoding purely attractive interactions. No matter whether a segmentation or clustering is then found agglomeratively (as in single linkage clustering / watershed) or divisively (as in spectral clustering or iterated normalized cuts), the user either needs to specify the desired number of segments or a termination criterion. An even stronger form of supervision is in terms of seeds, where one pixel of each segment needs to be designated either by a user or automatically. Unfortunately, clustering with automated seed selection remains a fragile and error-fraught process, because every missed or hallucinated seed causes an under- or oversegmentation error. Although the learning of good edge detectors boosts the quality of classical seed selection strategies (such as finding local minima of the boundary map, or thresholding boundary maps), non-local effects of seed placement along with strong variability in region sizes and shapes make it hard for any learned predictor to place *exactly one* seed in every true region.

In contrast to the above class of algorithms, multicut / correlation clustering partitions vertices with both attractive and repulsive interactions encoded into the edges of a graph. Multicut has the great advantage that a “natural” partitioning of a graph can be found, without needing to specify a desired number of clusters, or a termination criterion, or one seed per region. Its great drawback is that its optimization is NP-hard.

The main insight of this paper is that when both attractive and repulsive interactions between pixels are available, then a generalization of the watershed algorithm can be devised that segments an image *without* the need for seeds or stopping criteria or thresholds. It examines all graph edges, attractive and repulsive, sorted by their weight and adds these to an active set iff they are not in conflict with previous, higher-priority, decisions. The attractive subset of the resulting active set is a forest, with one tree representing each segment. However, the active set can have loops involving more than one repulsive edge. See Fig. 1 for a visual abstract.

In summary, our principal contributions are, first, a fast deterministic algorithm for graph partitioning with both positive and negative edge weights that does not need prior specification of the number of clusters (Section 4); and second, its theoretical characterization, including proof that it globally optimizes an objective related to the multicut correlation clustering objective (4).

Combined with a deep net, the algorithm also happens to define the state-of-the-art in a competitive neuron segmentation challenge (Section 5).

This is an extended version of [1], with the second principal contribution (Section 4) being new.

• S. Wolf, A. Bailoni, N. Rahaman, U. Köthe, and F.A. Hamprecht are with HCI/IWR, Heidelberg University, 69120 Heidelberg, Germany.  
E-mail: {steffen.wolf, alberto.bailoni, nasim.rahaman, ullrich.koethe, fred.hamprecht}@iwr.uni-heidelberg.de.

• C. Pape and A. Kreshuk are with EMBL, 69117 Heidelberg, Germany.  
E-mail: {constantin.pape, anna.kreshuk}@embl.de.

Manuscript received 17 Nov. 2018; revised 26 Feb. 2020; accepted 1 Mar. 2020.

Date of publication 16 Mar. 2020; date of current version 2 Sept. 2021.

(Corresponding author: Fred A. Hamprecht.)

Recommended for acceptance by O. Veksler.

Digital Object Identifier no. 10.1109/TPAMI.2020.2980827

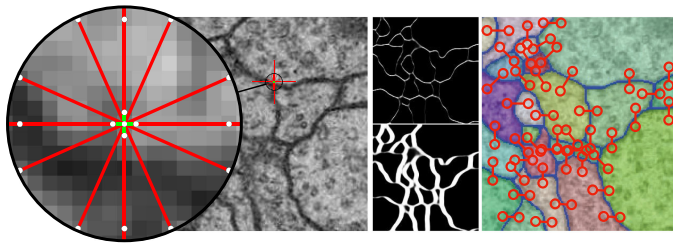


Fig. 1. Left: Overlay of raw data from the ISBI 2012 EM segmentation challenge and the edges for which attractive (green) or repulsive (red) interactions are estimated for each pixel using a CNN. Middle: vertical / horizontal repulsive interactions at intermediate / long range are shown in the top / bottom half. Right: Active mutual exclusion (mutex) constraints that the proposed algorithm invokes during the segmentation process.

## 2 RELATED WORK

In the original watershed algorithm [2], [3], seeds were automatically placed at all local minima of the boundary map. Unfortunately, this leads to severe over-segmentation. Defining better seeds has been a recurring theme of watershed research ever since. The simplest solution is offered by the seeded watershed algorithm [4]: It relies on an oracle (an external algorithm or a human) to provide seeds and assigns each pixel to its nearest seed in terms of minimax path distance.

In the absence of an oracle, many automatic methods for seed selection have been proposed in the last decades with applications in the fields of medicine and biology. Many of these approaches rely on edge feature extraction and edge detection like gradient calculation [5], [6]. Other types of methods generate seeds by first performing feature extraction [7], [8], whereas others first extract region of interests and then place seeds inside these regions by using thresholding [9], binarization [10],  $k$ -means [11] or other strategies [12], [13].

In applications where the number of regions is hard to estimate, simple automatic seed selection methods, e.g., defining seeds by connected regions of low boundary probability, don't work: The segmentation quality is usually insufficient because multiple seeds are in the same region and/or seeds leak through the boundary. Thus, in these cases seed selection may be biased towards over-segmentation (with seeding at all minima being the extreme case). The watershed algorithm then produces superpixels that are merged into final regions by more or less elaborate post-processing. This works better than using watersheds alone because it exploits the larger context afforded by superpixel adjacency graphs. Many criteria have been proposed to identify the regions to be preserved during merging, e.g., region dynamics [14], the waterfall transform [15], extinction values [16], region saliency [17], and  $(\alpha, \omega)$ -connected components [18]. A merging process controlled by criteria like these can be iterated to produce a hierarchy of segmentations where important regions survive to the next level. Variants of such hierarchical watersheds are reviewed and evaluated in [19].

These results highlight the close connection of watersheds to hierarchical clustering and minimum spanning trees/forests [20], [21], which inspired novel merging strategies and termination criteria. For example, [22] simply terminated hierarchical merging by fixing the number of surviving

regions beforehand. [23] incorporate predefined sets of generalized merge constraints into the clustering algorithm. Graph-based segmentation according to [24] defines a measure of quality for the current regions and stops when the merge costs would exceed this measure. Ultrametric contour maps [25] combine the gPb (global probability of boundary) edge detector with an oriented watershed transform. Superpixels are agglomerated until the ultrametric distance between the resulting regions exceeds a learned threshold. An optimization perspective is taken in [26], [27], which introduces  $h$ -increasing energy functions and builds the hierarchy incrementally such that merge decisions greedily minimize the energy. The authors prove that the optimal cut corresponds to a different unique segmentation for every value of a free regularization parameter.

An important line of research is given by partitioning of graphs with both attractive and repulsive edges [28]. Solutions that optimally balance attraction and repulsion do not require external stopping criteria such as predefined number of regions or seeds. This generalization leads to the NP-hard problem of correlation clustering or (synonymous) multicut (MC) partitioning. Fortunately, modern integer linear programming solvers in combination with incremental constraint generation can solve problem instances of considerable size [29], and good approximations exist for even larger problems [30], [31]. Reminiscent of strict minimizers [32] with minimal  $L_\infty$ -norm solution, our work solves the multicut objective optimally when all graph weights are raised to a large power.

Related to the proposed method, the greedy additive edge contraction (GAEC) [33] heuristic for the multicut also sequentially merges regions, but we handle attractive and repulsive interactions separately and define edge strength between clusters by a maximum instead of an additive rule. The greedy fixation algorithm introduced in [34] is closely related to the proposed method; it sorts attractive and repulsive edges by their absolute weight, merges nodes connected by attractive edges and introduces no-merge constraints for repulsive edges. However, similar to GAEC, it defines edge strength by an additive rule, which increases the algorithm's runtime complexity compared to the presented Mutex Watershed. Also, it is not yet known what objective the algorithm optimizes globally, if any.

Another beneficial extension is the introduction of additional long-range edges. The strength of such edges can often be estimated with greater certainty than is achievable for the local edges used by watersheds on standard 4- or 8-connected pixel graphs. Such repulsive long-range edges have been used in [35] to represent object diameter constraints, which is still an MC-type problem. When long-range edges are also allowed to be attractive, the problem turns into the more complicated lifted multicut (LMC) [36]. Realistic problem sizes can only be solved approximately [33], [37], but watershed superpixels followed by LMC post-processing achieve state-of-the-art results on important benchmarks [38]. Long-range edges are also used in [39], as side losses for the boundary detection convolutional neural network (CNN); but they are not used explicitly in any downstream inference.

In general, striking progress in watershed-based segmentation has been achieved by learning boundary maps with

CNNs. This is nicely illustrated by the evolution of neuro-segmentation for connectomics, an important field we also address in the experimental section. CNNs were introduced to this application in [40] and became, in much refined form [41], the winning entry of the ISBI 2012 Neuro-Segmentation Challenge [42]. Boundary maps and superpixels were further improved by progress in CNN architectures and data augmentation methods, using U-Nets [43], FusionNets [44] or inception modules [38]. Subsequent postprocessing with the GALA algorithm [45], [46], conditional random fields [47] or the lifted multicut [38] pushed the envelope of final segmentation quality. MaskExtend [48] applied CNNs to both boundary map prediction and superpixel merging, while flood-filling networks [49] eliminated superpixels altogether by training a recurrent neural network to perform region growing one region at a time.

Most networks mentioned so far learn boundary maps on pixels, but learning works equally well for edge-based watersheds, as was demonstrated in [50], [51] using edge weights generated with a CNN [52], [53]. Tailoring the learning objective to the needs of the watershed algorithm by penalizing critical edges along minimax paths [53] or end-to-end training of edge weights and region growing [54] improved results yet again.

Outside of connectomics, [55] obtained superior boundary maps from CNNs by learning not just boundary strength, but also its gradient direction. Holistically-nested edge detection [56], [57] couples the CNN loss at multiple resolutions using deep supervision and is successfully used as a basis for watershed segmentation of medical images in [58].

We adopt important ideas from this prior work (hierarchical single-linkage clustering, attractive and repulsive interactions, long-range edges, and CNN-based learning). The proposed efficient segmentation framework can be interpreted as a generalization of [23], because we also allow for soft repulsive interactions (which can be overridden by strong attractive edges), and constraints are generated on-the-fly.

### 3 THE MUTEX WATERSHED ALGORITHM AS AN EXTENSION OF SEEDED WATERSHED

In this section we introduce the Mutex Watershed Algorithm, an efficient graph clustering algorithm that can ingest both attractive and repulsive cues. We first reformulate seeded watershed as a graph partitioning with infinitely repulsive edges and then derive the generalized algorithm for finitely repulsive edges, which obviates the need for seeds.

#### 3.1 Definitions and Notation

Let  $\mathcal{G} = (V, E, w)$  be a weighted graph. The scalar attribute  $w : E \rightarrow \mathbb{R}$  associated with each edge is a merge affinity: the higher this number, the higher the inclination of the two incident vertices to be assigned to the same cluster. Conversely, large negative affinity indicates a greater desire of the incident vertices to be in different clusters. In our application, each vertex corresponds to one pixel in the image to be segmented. We call an edge  $e \in E$  repulsive if  $w_e < 0$  and we call it attractive if  $w_e > 0$  and collect them in  $E^- = \{e \in E \mid w_e < 0\}$  and  $E^+ = \{e \in E \mid w_e > 0\}$  respectively.

In our application, each vertex corresponds to one pixel in the image to be segmented. The Mutex Watershed

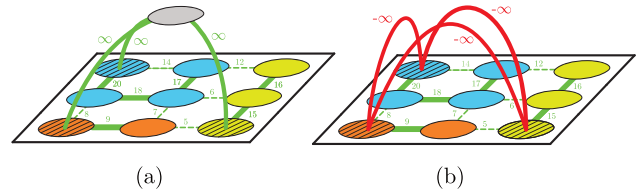


Fig. 2. Two equivalent representations of the seeded watershed clustering obtained using (a) a maximum spanning tree computation or (b) Algorithm 1. Both graphs share the weighted attractive (green) edges and seeds (hatched nodes). The infinitely attractive connections to the auxiliary node (gray) in (a) are replaced by infinitely repulsive (red) edges between each pair of seeds in (b). The two final clusterings are defined by the active sets (bold edges) and are identical. Node colors indicate the clustering result, but are arbitrary.

algorithm, defined in Section 3.3, maintains disjoint active sets  $A^+ \subseteq E^+$ ,  $A^- \subseteq E^-$ ,  $A^+ \cap A^- = \emptyset$  that encode merges and mutual exclusion constraints, respectively. Clusters are defined via the “connected” predicate

$$\begin{aligned} \forall i, j \in V : \\ \Pi_{i \rightarrow j} &= \{\text{paths } \pi \text{ from } i \text{ to } j \text{ with } \pi \subseteq E^+\} \\ \text{connected}(i, j; A^+) &\Leftrightarrow \exists \text{ path } \pi \in \Pi_{i \rightarrow j} \text{ with } \pi \subseteq A^+ \\ \text{cluster}(i; A^+) &= \{i\} \cup \{j : \text{connected}(i, j; A^+)\}. \end{aligned}$$

Conversely, the active subset  $A^- \subseteq E^-$  of repulsive edges defines mutual exclusion relations by using the following predicate:

$$\begin{aligned} \text{mutex}(i, j; A^+, A^-) &\Leftrightarrow \exists e = (k, l) \in A^- \text{ with} \\ &k \in \text{cluster}(i; A^+) \text{ and} \\ &l \in \text{cluster}(j; A^+) \text{ and} \\ &\text{cluster}(i; A^+) \neq \text{cluster}(j; A^+). \end{aligned}$$

Admissible active edge sets  $A^+$  and  $A^-$  must be chosen such that the resulting clustering is consistent, i.e., nodes engaged in a mutual exclusion constraint cannot be in the same cluster:  $\text{mutex}(i, j; A^+, A^-) \Rightarrow \text{not connected}(i, j; A^+)$ . The “connected” and “mutex” predicates can be efficiently evaluated using a union find data structure.

#### 3.2 Seeded Watershed From a Mutex Perspective

One interpretation of the proposed method is in terms of a generalization of the edge-based watershed algorithm [20], [59], [60] or image foresting transform [61]. This algorithm can only ingest a graph with purely attractive interactions,  $E^- = \emptyset$ . Without further constraints, the algorithm would yield only the trivial result of a single cluster comprising all vertices. To obtain more interesting output, an oracle needs to provide seeds (e.g., one node per cluster). These seed vertices are all connected to an auxiliary node (see Fig. 2a) by auxiliary edges with infinite merge affinity. A maximum spanning tree (MST) on this augmented graph can be found in linearithmic time; and the maximum spanning tree (or in the case of degeneracy: at least one of the maximum spanning trees) will include the auxiliary edges. When the auxiliary edges are deleted from the MST, a forest results, with each tree representing one cluster [20], [59], [61].

We now reformulate this well-known algorithm in a way that will later emerge as a special case of the proposed

---

**Algorithm 1.** Mutex Version of Seeded Watershed Algorithm. The Output Clustering is Defined by the Connected Components of the Final Attractive Active Set  $A^+$

---

**Seeded Watershed:**

```

WS ( $\mathcal{G}(V, E)$ , pos. weights  $w : E \rightarrow \mathbb{R}^+$ , seeds  $S \subseteq V$ ):
   $A^+ \leftarrow \emptyset$ ;
   $A^- \leftarrow \{(s, t) \in S \times S \mid s \neq t\}$ ;
  ▷ Equivalent to introducing infinitely repulsive edges
  between seeds
  for  $(i, j) = e \in E$  in descending order of  $w_e$  do
    if not connected  $(i, j; A^+)$  and not mutex  $(i, j; A^+, A^-)$  then
       $A^+ \leftarrow A^+ \cup e$ ;
      ▷ merge  $i$  and  $j$  and inherit the mutex constraints of
      the parent clusters
    end
  end
  return  $A^+ \cup A^-$ 

```

---

Mutex Watershed: we eliminate the auxiliary node and edges, and replace them by a set of infinitely repulsive edges, one for each pair of seeds (Fig. 2b). Algorithm 1 is a variation of Kruskal’s MST algorithm operating on the seed mutex graph just defined, and gives results identical to seeded watershed on the original graph.

This algorithm differs from Kruskal’s only by the check for mutual exclusion in the if-statement. Obviously, the modified algorithm has the same effect as the original algorithm, because the final set  $A^+$  is exactly the maximum spanning forest obtained after removing the auxiliary edges from the original solution.

In the sequel, we generalize this construction by admitting less-than-infinitely repulsive edges. Importantly, these can be dense and are hence much easier to estimate automatically than seeds with their strict requirement of only-one-per-cluster.

### 3.3 Mutex Watersheds

We now introduce our core contribution: an algorithm that is empirically no more expensive than a MST computation; but that can ingest both attractive and repulsive cues and partition a graph into a number of clusters that does not need to be specified beforehand. Neither seeds nor hyperparameters that implicitly determine the number of resulting clusters are required.

The Mutex Watershed, Algorithm 2, proceeds as follows. Given a graph  $\mathcal{G} = (V, E)$  with signed weights  $w : E \rightarrow \mathbb{R}$ , do the following: sort all edges  $E$ , attractive or repulsive, by their absolute weight in descending order into a priority queue. Iteratively pop all edges from the queue and add them to the active set one by one, provided that a set of conditions are satisfied. More specifically, assuming `connect_all` is `False`, if the next edge popped from the priority queue is attractive and its incident vertices are not yet in the same tree, then connect the respective trees provided this is not ruled out by a mutual exclusion constraint. If on the other hand the edge popped is repulsive, and if its incident vertices are not yet in the same tree, then add a mutual exclusion constraint between the two trees. The output clustering is defined by the connected components of the final attractive active set  $A^+$ .

---

**Algorithm 2.** Mutex Watershed Algorithm. The Output Clustering is Defined by the Connected Components of the Final Attractive Active Set  $A^+$ . The `Connect_All` Parameter Changes the Internal Cluster Connectedness From Trees to Fully Connected, but does not Change the Output Clustering. The Connected Predicate can be Efficiently Evaluated Using Union Find Data Structures

---

**Mutex Watershed:**

```

MWS ( $\mathcal{G}(V, E)$ ,  $w : E \rightarrow \mathbb{R}$ , boolean connect_all):
   $A^+ \leftarrow \emptyset$ ;  $A^- \leftarrow \emptyset$ ;
  for  $(i, j) = e \in E$  in descending order of  $|w_e|$  do
    if  $e \in E^+$  then
      if not mutex  $(i, j; A^+, A^-)$  then
        if not connected  $(i, j; A^+)$  or connect_all then
          merge  $(i, j)$ :  $A^+ \leftarrow A^+ \cup e$ ;
          ▷ merge  $i$  and  $j$  and inherit the mutex constraints
          of the parent clusters
        end
      end
    else
      if not connected  $(i, j; A^+)$  then
        addmutex  $(i, j)$ :  $A^- \leftarrow A^- \cup e$ ;
        ▷ add mutex constraint between  $i$  and  $j$ 
      end
    end
  end
  return  $A^+ \cup A^-$ 

```

---

The crucial difference to Algorithm 1 is that mutex constraints are no longer pre-defined, but created dynamically whenever a repulsive edge is found. However, new exclusion constraints can never override earlier, high-priority merge decisions. In this case, the repulsive edge in question is simply ignored. Similarly, an attractive edge must never override earlier and thus higher-priority must-not-link decisions.

The boolean value of the `connect_all` input parameter of the algorithm does not influence the final output clustering, but defines the internal cluster connectedness: when it is set to `True`, the algorithm adds all attractive intra-cluster edges to the active set  $A^+$ . When it is set to `False`, then a maximum spanning tree is built for each cluster similarly to the seeded watershed. This variant of the algorithm will be helpful in the next section 4 to highlight the relation between the Mutex Watershed and the multicut problem.

Fig. 3 illustrates the proposed algorithm: Figs. 3a and 3b show examples of an unconstrained merge and an added mutex constraint, respectively; Figs. 3c and 3d show, respectively, an example of an attractive edge ( $w_e = 14$ ) and repulsive edge ( $w_e = -13$ ) that are not added to the active set because their incident vertices are already “connected” and belong to the same tree of the forest  $A^+$ ; finally, Fig. 3e shows an attractive edge ( $w_e = 12$ ) that is ruled out by a previously introduced mutual exclusion relation.

### 3.4 Time Complexity Analysis

Before analyzing the time complexity of Algorithm 2 we first review the complexity of Kruskal’s algorithm. Using a union-find data structure (with path compression and union by rank) the time complexity of `merge`( $i, j$ ) and

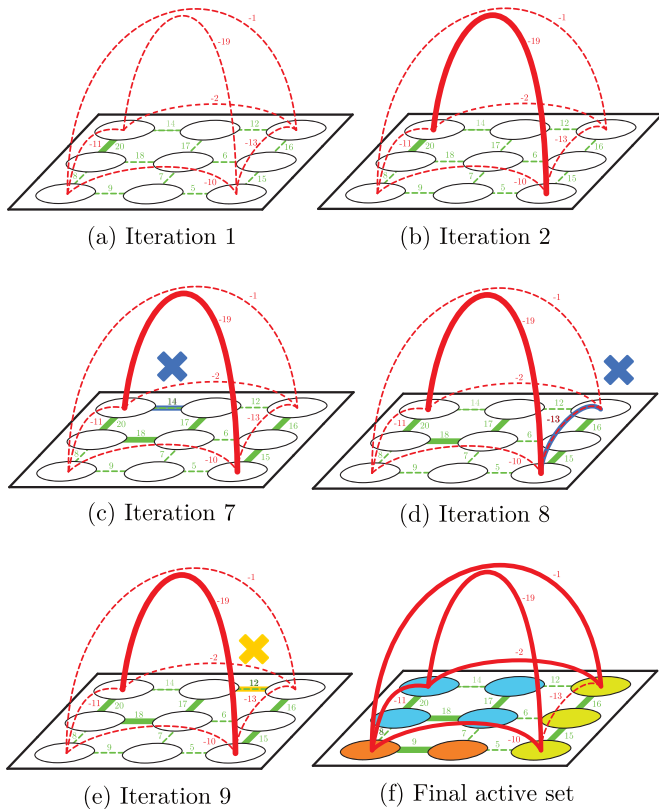


Fig. 3. Some iterations of the Mutex Watershed Algorithm 2 applied to a graph with weighted attractive (green) and repulsive (red) edges. Edges accumulated in the active set  $A$  after a given number of iterations are shown in bold. The `connect_all` parameter of the algorithm is set to False, so that only the positive edges belonging to the maximum spanning tree of each cluster are added to the active set. Once the algorithm terminates, the final active set (f) defines the final clustering (indicated using arbitrary node colors). Some edges are not added to the active set because they are mutex constrained (yellow highlight) or because the associated nodes are already connected and in the same cluster (blue highlight).

connected( $i, j$ ) is  $\mathcal{O}(\alpha(V))$ , where  $\alpha$  is the slowly growing inverse Ackerman function, and the total runtime complexity is dominated by the initial sorting of the edges  $\mathcal{O}(E \log E)$  [62].

To check for mutex constraints efficiently, we maintain a set of all active mutex edges

$$M[C_i] = \{(u, v) \in A^- \mid u \in C_i \vee v \in C_i\},$$

for every  $C_i = \text{cluster}(i)$  using hash tables, where insertion of new mutex edges (i.e., `addmutex`) and search have an average complexity of  $\mathcal{O}(1)$ . Note that every cluster can be efficiently identified by its union-find root node. For mutex ( $i, j$ ) we check if  $M[C_i] \cap M[C_j] = \emptyset$  by searching for all elements of the smaller hash table in the larger hash table. Therefore mutex ( $i, j$ ) has an average complexity of  $\mathcal{O}(\min(|M[C_i]|, |M[C_j]|))$ . Similarly, during merge ( $i, j$ ), mutex constraints are inherited by merging two hash tables, which also has an average complexity  $\mathcal{O}(\min(|M[C_i]|, |M[C_j]|))$ .

In conclusion, the average runtime contribution of attractive edges  $\mathcal{O}(\max(|E^+| \cdot \alpha(V), |E^+| \cdot M))$  (checking mutex constraints and possibly merging) and repulsive edges  $\mathcal{O}(\max(|E^-| \cdot \alpha(V), |E^-|))$  (insertion of one mutex edge) result in a total average runtime complexity of Algorithm 2

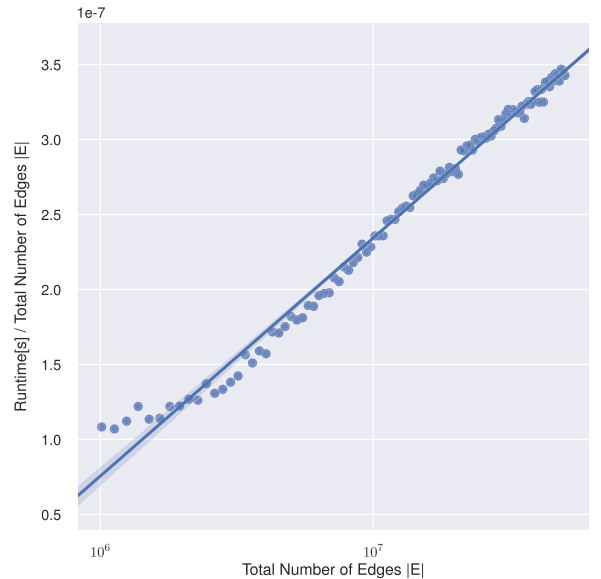


Fig. 4. Runtime  $T$  of Mutex Watershed (without sorting of edges) measured on sub-volumes of the ISBI challenge of different sizes (thereby varying the total number of edges  $E$ ). We plot  $\frac{T}{|E|}$  over  $|E|$  in a logarithmic plot, which makes  $T \sim |E| \log(|E|)$  appear as straight line. A logarithmic function (blue line) is fitted to the measured  $\frac{T}{|E|}$  (blue circles) with ( $R^2 = 0.9896$ ). The good fit suggests that empirically  $T \approx \mathcal{O}(E \log E)$ .

$$\mathcal{O}(\max(E \log E, EM)). \quad (1)$$

where  $M$  is the expected value of  $\min(|M[C_i]|, |M[C_j]|)$  and  $\alpha(V) \in \mathcal{O}(\log V) \in \mathcal{O}(\log E)$ .<sup>1</sup>

In the worst case  $\mathcal{O}(M) \in \mathcal{O}(E)$ , the Mutex Watershed Algorithm has a runtime complexity of  $\mathcal{O}(E^2)$ . Empirically, we find that  $\mathcal{O}(EM) \approx \mathcal{O}(E \log E)$  by measuring the runtime of Mutex Watershed for different sub-volumes of the ISBI challenge (see Fig. 4), leading to a

$$\text{Empirical Mutex Watershed Complexity: } \mathcal{O}(E \log E). \quad (2)$$

## 4 THEORETICAL CHARACTERIZATION

*Towards the Multicut Framework.* In Section 3.3, we have introduced the Mutex Watershed (MWS) algorithm as a generalization of seeded watersheds and the Kruskal algorithm in particular. However, since we are considering graphs with negative edge weights, the MWS is conceptually closer to the multicut problem and related heuristics such as GAEC and GF [34]. Fortunately, due to the structure of the MWS it can be analyzed using dynamic programming. This section summarizes our second contribution, i.e., the proof that the Mutex Watershed Algorithm globally optimizes a precise objective related to the multicut.

### 4.1 Review of the Multicut Problem and its Objective

In the following, we will review the multicut problem not in its standard formulation but in the *Cycle Covering Formulation* introduced in [63], which is similar to the MWS formulation as it also considers the set of *attractive* and *repulsive* edges separately. Previously, in Section 3.1, we defined a clustering by

1. In the worst case  $G$  is a fully connected graph, with  $|E| = |V|^2$ , hence  $\log |V| = \frac{1}{2} \log |E|$ .

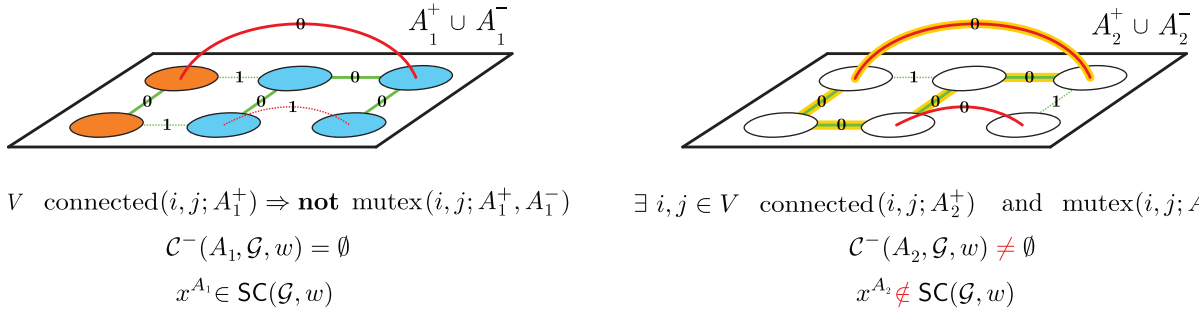


Fig. 5. *Consistent and inconsistent active sets* – Two different active edge sets  $A_1 \subseteq E$  (on the left) and  $A_2 \subseteq E$  (on the right) on identical toy graphs with six nodes, attractive (green) and repulsive (red) edges. The value of the edge indicator  $x^A \in \{0, 1\}^{|E|}$  defined in Eq. (4) is shown for every edge. Members of the active sets are shown as solid lines. *On the left*, the active set  $A_1$  is *consistent*, i.e., does not include any conflicted cycle  $\mathcal{C}^-(\mathcal{G}, w)$  (see Definition 4.1): Therefore, it is associated with a clustering (represented by arbitrary node colors). *On the right*, the active set  $A_2$  is not consistent and includes at least one conflicted cycle (highlighted in yellow), thus it cannot be associated with a node clustering.

introducing the concept of an active set of edges  $A = A^+ \cup A^- \subseteq E$  and the connected/mutex predicates. In particular, an active set describes a valid clustering if it does not include *both* a path of only attractive edges *and* a path with exactly one repulsive edge connecting any two nodes  $i, j \in V$

$$\text{connected}(i, j; A^+) \implies \text{not mutex}(i, j; A^+, A^-). \quad (3)$$

In other words, an active set is *consistent* and describes a clustering if it does not contain any cycle with exactly one repulsive edge (known as conflicted cycles).

**Definition 4.1 (Conflicted cycles).** We call a cycle of  $\mathcal{G}$  *conflicted* w.r.t.  $(\mathcal{G}, w)$  if it contains precisely one repulsive edge  $e \in E^-$ , s.t.  $w_e < 0$ . We denote by  $\mathcal{C}^-(\mathcal{G}, w) \subseteq \mathcal{C}(\mathcal{G}, w)$  the set of all conflicted cycles. Furthermore, given a set of edges  $A \subseteq E$ , we denote by  $\mathcal{C}^-(A, \mathcal{G}, w) \subseteq \mathcal{C}^-(\mathcal{G}, w)$  the set of conflicted cycles involving only edges in  $A$ .

From now on, in order to describe different clustering solutions in the framework of (integer) linear programs, we associate each active set  $A$  with the following edge indicator  $x^A$

$$x^A := \mathbb{1}\{e \notin A\} \in \{0, 1\}^{|E|}. \quad (4)$$

In this way, the cycle-free property  $\mathcal{C}^-(A, \mathcal{G}, w) = \emptyset$  of an active set can be reformulated in terms of linear inequalities

$$\forall C \in \mathcal{C}^-(\mathcal{G}, w) : \sum_{e \in E_C} x_e^A \geq 1 \iff \mathcal{C}^-(A, \mathcal{G}, w) = \emptyset. \quad (5)$$

In words, the active set cannot contain conflicted cycles; or vice versa, every conflicted cycle must contain at least one edge that is not part of the active set. Following [63], via this property we describe the space of all possible clustering solutions by defining the convex hull  $\text{SC}(\mathcal{G}, w)$  of all edge indicators corresponding to valid clusterings of  $(\mathcal{G}, w)$ :

**Definition 4.2.** Let  $\text{SC}(\mathcal{G}, w)$  denote the convex hull of all edge indicators  $x \in \{0, 1\}^{|E|}$  satisfying the following system of inequalities

$$\forall C \in \mathcal{C}^-(\mathcal{G}, w) : \sum_{e \in E_C} x_e \geq 1. \quad (6)$$

That is,  $\text{SC}(\mathcal{G}, w)$  contains all edge labelings for which every conflicted cycle is broken at least once. We call  $\text{SC}(\mathcal{G}, w)$  the

*set covering polyhedron with respect to conflicted cycles, similarly to [63].*

Fig. 5 summarizes these definitions and provides an example of consistent and inconsistent active sets with their associated clusterings and edge indicators.

As shown in [63], the *multicut optimization problem* can be formulated with constraints over conflicted cycles in terms of the following integer linear program (ILP), which is NP-hard

$$\min_{x \in \text{SC}(\mathcal{G}, w)} \sum_{e \in E} |w_e| x_e. \quad (7)$$

The solution of the multicut problem is given by the clustering associated to the connected components of the active set  $\hat{A}^+ = \{e \in E^+ | \hat{x}_e = 0\}$ , where  $\hat{x} \in \{0, 1\}^{|E|}$  is the solution of (7).

## 4.2 Mutex Watershed Objective

We now define the Mutex Watershed objective that is minimized by the Mutex Watershed Algorithm (proof in Section 4.3) and show how it is closely related to the multicut problem defined in Eq. (7). Lange et al. [63] introduce the concept of dominant edges in a graph. For example, an attractive edge  $f \in E^+$  is called dominant if there exists a cut  $B$  with  $f \in E_B$  such that  $|w_f| \geq \sum_{e \in E_B \setminus \{f\}} |w_e|$ . These highlight an aspect of the multicut problem that can be used to search for optimal solutions more efficiently. Not all weighted graphs contain dominant edges; but if, assuming no ties, we raise all graph weights to a large enough power a similar property emerges.

**Definition 4.3 (Dominant Power).** Let  $\mathcal{G} = (V, E, w)$  be an edge-weighted graph, with unique weights  $w : E \rightarrow \mathbb{R}$ . We call  $p \in \mathbb{N}^+$  a dominant power if

$$|w_e|^p > \sum_{t \in E, w_t < w_e} |w_t|^p \quad \forall e \in E, \quad (8)$$

In contrast to dominant edges [63], we do not consider edges on a cut but rather all edges with smaller absolute weight. Note that there exists a dominant power for any finite set of edges, since for any  $e \in E$  we can divide (8) by  $|w_e|^p$  and observe that the normalized weights  $|w_t|^p / |w_e|^p$  (and any finite sum of these weights) converges to 0 when  $p$  tends to infinity.

**Algorithm 3.** Equivalent Formulation of the Mutex Watershed Algorithm 2, With Input Parameter `connect_all = True`. The Set of Conflicted Cycles  $\mathcal{C}^-(A, \mathcal{G}, w)$  is Defined in Definition 4.1. The Output Clustering is Defined by the Connected Components of the Final Attractive Active set  $A^+ = A \cap E^+$

**Conflicted-Cycles Mutex Watershed:**

CCMWS ( $\mathcal{G}(V, E), w : E \rightarrow \mathbb{R}$ ):

```

A ← ∅;
for (i, j) = e ∈ E in descending order of |we| do
  if C-(A ∪ {e}, G, w) = ∅ then
    A ← A ∪ e;
  end
end
return A;

```

**Algorithm 4.** Mutex Watershed Algorithm Starting From Initial Active set  $\tilde{A}$ . An Initial Set  $\tilde{A}$  of Active Edges is Given as Additional Input and the Final Active Set is Such That  $A \subseteq E \setminus \tilde{A}$ . Note That Algorithm 3 is a Special Case of This Algorithm When  $\tilde{A} = \emptyset$ . Differences With Algorithm 3 are Highlighted in Blue.

**Initialized Mutex Watershed:**

IMWS ( $\mathcal{G}(V, E), w : E \rightarrow \mathbb{R}$ , initial active set  $\tilde{A}$ ):

```

A ← ∅;
for e ∈ E \  $\tilde{A}$  in descending order of weight do
  if C-(A ∪  $\tilde{A}$  ∪ {e}, G, w) = ∅ then
    A ← A ∪ e;
  end
end
return A;

```

By considering the multicut problem in Eq. (7) and raising the weights  $|w_e|$  to a dominant power  $p$ , we fundamentally change the problem structure:

**Definition 4.4 (Mutex Watershed Objective).** Let  $\mathcal{G} = (V, E, w)$  be an edge-weighted graph, with unique weights  $w : E \rightarrow \mathbb{R}$  and  $p \in \mathbb{N}^+$  a dominant power. Then the Mutex Watershed Objective is defined as the integer linear program

$$\min_{x \in \text{SC}(\mathcal{G}, w)} \sum_{e \in E} |w_e|^p x_e, \quad (9)$$

where  $\text{SC}(\mathcal{G}, w)$  is the convex hull defined in Definition 4.2.

In the following section, we will prove that this modified version of the multicut objective, which we call Mutex Watershed Objective, is indeed optimized by the Mutex Watershed Algorithm:

**Theorem 4.1.** Objective Let  $\mathcal{G} = (V, E, w)$  be an edge-weighted graph, with unique weights  $w : E \rightarrow \mathbb{R}$  and  $p \in \mathbb{N}^+$  a dominant power. Then the edge indicator given by the Mutex Watershed Algorithm 2

$$x^{\text{MWS}} := \mathbb{1} \left\{ e \notin \text{MWS}(\mathcal{G}, w, \text{connect\_all} = \text{True}) \right\},$$

minimizes the Mutex Watershed Objective in Eq. (9).

### 4.3 Proof of Optimality via Dynamic Programming

In this section we prove Theorem 4.1, i.e., that the Mutex Watershed Objective defined in Definition 4.4 is solved to optimality by the Mutex Watershed Algorithm 3. Particularly, in the following Section 4.3.1 we show that the edge indicator associated to the solution of the MWS algorithm lies in  $\text{SC}(\mathcal{G}, w)$ , whereas in Section 4.3.2 we prove that it solves Eq. (9) to optimality.

#### 4.3.1 Cycle Consistency

The Mutex Watershed algorithm introduced in Section 3 iteratively builds an active set  $A = A^+ \cup A^-$  such that nodes engaged in a mutual exclusion constraint (encoded by edges in  $A^-$ ) are never part of the same cluster. In other words, this means that the active set built by the Mutex Watershed at every iteration does never include a conflicted cycle and is

always consistent. In particular, for any attractive edge  $(i, j) = e^+ \in E^+$  and any consistent set  $A$  that fulfills  $\mathcal{C}^-(A, \mathcal{G}, w) = \emptyset$

$$\text{not mutex}(i, j, A^+, A^-) \Leftrightarrow \mathcal{C}^-(A \cup \{e^+\}, \mathcal{G}, w) = \emptyset.$$

Similarly, for any repulsive edge  $(s, t) = e^- \in E^-$

$$\text{not connected}(s, t, A^+) \Leftrightarrow \mathcal{C}^-(A \cup \{e^-\}, \mathcal{G}, w) = \emptyset.$$

Therefore, we can rewrite Algorithm 2 in the form of Algorithm 3. This new formulation makes it clear that

$$\mathcal{C}^-(\text{MWS}(\mathcal{G}, w, \text{connect\_all} = \text{True})) = \emptyset. \quad (10)$$

Thus, thanks to Eq. (5) and Definition 4.2, it follows that the MWS edge indicator  $x^{\text{MWS}}$  defined in Theorem 4.1 lies in  $\text{SC}(\mathcal{G}, w)$

$$x^{\text{MWS}} \in \text{SC}(\mathcal{G}, w). \quad (11)$$

#### 4.3.2 Optimality

We first note that the Mutex Watershed Objective 4.4 and Theorem 4.1 can easily be reformulated in terms of active sets to minimize

$$\arg \min_{A \subseteq E} - \sum_{e \in A} |w_e|^p \quad \text{s.t.} \quad \mathcal{C}^-(A, \mathcal{G}, w) = \emptyset. \quad (12)$$

We now generalize the Mutex Watershed (see Algorithm 4) and the objective such that an initial consistent set of active edges  $\tilde{A} \subseteq E$  is supplied:

**Definition 4.5 (Energy optimization subproblem).** Let  $\mathcal{G} = (V, E, w)$  be an edge-weighted graph. Define the optimal solution of the subproblem as

$$S(\mathcal{G}, \tilde{A}) := \arg \min_{A \subseteq (E \setminus \tilde{A})} T(A) \quad \text{with} \quad T(A) := - \sum_{e \in A} |w_e|^p, \quad (13)$$

$$\text{s.t.} \quad \mathcal{C}^-(A \cup \tilde{A}, \mathcal{G}, w) = \emptyset, \quad (14)$$

where  $\tilde{A} \subseteq E$  is a set of initially activated edges such that  $\mathcal{C}^-(\tilde{A}, \mathcal{G}, w) = \emptyset$ .

We note that for  $\tilde{A} = \emptyset$ , the optimal solution  $S(\mathcal{G}, \emptyset)$  is equivalent to the solution minimizing the Mutex Watershed Objective and Eq. (12).

**Definition 4.6 (Incomplete, consistent initial set).** For an edge-weighted graph  $\mathcal{G} = (V, E, w)$  a set of edges  $\tilde{A} \subseteq E$  is consistent if

$$\mathcal{C}^-(\tilde{A}, \mathcal{G}, w) = \emptyset. \quad (15)$$

$\tilde{A}$  is incomplete if it is not the final solution and there exists a consistent edge  $\tilde{e}$  that can be added to  $\tilde{A}$  without violating the constraints.

$$\exists \tilde{e} \in E \setminus \tilde{A} \quad \text{s.t.} \quad \mathcal{C}^-(\tilde{A} \cup \{\tilde{e}\}, \mathcal{G}, w) = \emptyset. \quad (16)$$

**Definition 4.7 (First greedy step).** Let us consider an incomplete, consistent initial active set  $\tilde{A} \subseteq E$  on  $\mathcal{G} = (V, E, w)$ . We define

$$g := \arg \max_{e \in (E \setminus \tilde{A})} |w(e)| \quad \text{s.t.} \quad \mathcal{C}^-(\tilde{A} \cup \{e\}, \mathcal{G}, w) = \emptyset. \quad (17)$$

as the feasible edge with the highest weight, which is always the first greedy step of Algorithm 4.

In the following two lemmas, we prove that the Mutex Watershed problem has an *optimal substructure property* and a *greedy choice property* [62], which are sufficient to prove that the Mutex Watershed algorithm finds the optimum of the Mutex Watershed Objective.

**Lemma 4.2 (Greedy-choice property).** For an incomplete, consistent initial active set  $\tilde{A}$  of the Mutex Watershed, the first greedy step  $g$  is always part of the optimal solution

$$g \in S(\mathcal{G}, \tilde{A}).$$

**Proof.** We will prove the theorem by contradiction by assuming that the first greedy choice is not part of the optimal solution, i.e.,  $g \notin S(\mathcal{G}, \tilde{A})$ . Since  $g$  is by definition the feasible edge with highest weight, it follows that:

$$|w(e)| < |w(g)| \quad \forall e \in S(\mathcal{G}, \tilde{A}). \quad (18)$$

We now consider the alternative active set  $A' = \{g\}$ , that is a consistent solution, with

$$T(A') = -|w_g|^p \stackrel{(8)}{<} - \sum_{t \in S(\mathcal{G}, \tilde{A})} |w_t|^p = T(S(\mathcal{G}, \tilde{A})), \quad (19)$$

which contradicts the optimality of  $S(\mathcal{G}, \tilde{A})$ .  $\square$

**Lemma 4.3 (Optimal substructure property).** Let us consider an initial active set  $\tilde{A}$ , the optimization problem defined in Equation (13), and assume to have an incomplete, consistent problem (see Definition 4.6). Then it follows that:

- 1) After making the first greedy choice  $g$ , we are left with a subproblem that can be seen as a new optimization problem of the same structure;
- 2) The optimal solution  $S(\mathcal{G}, \tilde{A})$  is always given by the combination of the first greedy choice and the optimal solution of the remaining subproblem.

**Proof.** After making the first greedy choice and selecting the first feasible edge  $g$  defined in Equation (17), we are clearly left with a new optimization problem of the same structure that has the following optimal solution:  $S(\mathcal{G}, \tilde{A} \cup \{g\})$ .

In order to prove the second point of the theorem, we now show that

$$S(\mathcal{G}, \tilde{A}) = \{g\} \cup S(\mathcal{G}, \tilde{A} \cup \{g\}). \quad (20)$$

Since Algorithm 4 fulfills the greedy-choice property,  $g \in S(\mathcal{G}, \tilde{A})$  and we can add the edge  $g$  as an additional constraint to the optimal solution

$$S(\mathcal{G}, \tilde{A}) = \arg \min_{A \subseteq (E \setminus \tilde{A})} T(A) \quad (21)$$

$$\text{s. t.} \quad \mathcal{C}^-(A \cup \tilde{A}, \mathcal{G}, w) = \emptyset; \quad g \in A.$$

Then it follows that:

$$S(\mathcal{G}, \tilde{A}) = \{g\} \cup \arg \min_{A \subseteq E \setminus (\tilde{A} \cup \{g\})} T(A) \quad (22)$$

$$\text{s. t.} \quad \mathcal{C}^-(A \cup \{g\} \cup \tilde{A}, \mathcal{G}, w) = \emptyset,$$

which is equivalent to Equation (20).  $\square$

**Proof of Theorems 4.1** In Lemmas 4.2 and 4.3 we have proven that the optimization problem defined in (12) has the optimal substructure and a greedy choice property. It follows through induction that the final active set **MWS** ( $\mathcal{G}, w, \text{connect\_all} = \text{True}$ ) found by the Mutex Watershed Algorithm 3 is the optimal solution for the Mutex Watershed objective (12) [62].  $\square$

#### 4.4 Relation to the Extended Power Watershed Framework

The Power Watershed [64] is an important framework for graph-based image segmentation that includes several algorithms like seeded watershed, random walker and graph cuts. Recently, [65] extended the framework to even more general types of hierarchical optimization algorithms thanks to the use of  $\Gamma$ -theory and  $\Gamma$ -convergence [66], [67]. In this section, we show how the Mutex Watershed algorithm can also be included in this extended framework<sup>2</sup> and how the framework suggests an optimization problem that is solved by the Mutex Watershed.

##### 4.4.1 Mutex Watershed as Hierarchical Optimization Algorithm

We first start by introducing the extended Power Watershed framework and restating the main theorem from [65]:

**Theorem 4.4 ([65] Extended Power Watershed Framework).** Consider three strictly positive integers  $p, m, t \in \mathbb{N}^+$  and  $t$  real numbers

$$1 \geq \lambda_0 > \lambda_1 > \dots > \lambda_{t-1} > 0. \quad (23)$$

2. The connection between the Mutex Watershed and the extended Power Watershed framework was kindly pointed out by an anonymous reviewer.



**Algorithm 5.** Generic Hierarchical Optimization Algorithm Introduced in [65]. The Sequence of Continuous Functions  $Q_k : \mathbb{R}^m \rightarrow \mathbb{R}$  is Sorted According to the Associated Scales  $\lambda_k$  (Eq. (23))

**Generic hierarchical optimization:**

**GHO**( $Q_0, \dots, Q_{t-1}$ ):

$M_0 = \arg \min_{x \in \mathbb{R}^m} Q_0(x)$

**for**  $k \in 1, \dots, t-1$  **do**

$M_k = \arg \min_{x \in M_{k-1}} Q_k(x)$

**end**

**Return:** some  $x^* \in M_{t-1}$

Given  $t$  continuous functions  $Q_k : \mathbb{R}^m \rightarrow \mathbb{R}$  with  $0 \leq k < t$ , define the function

$$Q^p(x) := \sum_{0 \leq k < t} \lambda_k^p Q_k(x). \quad (24)$$

Then, if any sequence  $(x_p)_{p>0}$  of minimizers  $x_p$  of  $Q^p(x)$  is bounded (i.e., there exists  $C > 0$  such that for all  $p > 0$ ,  $\|x_p\|_\infty \leq C$ ), the sequence is convergent, up to taking a subsequence, toward a point of  $M_{t-1}$ , which is the set of minimizers recursively defined in Algorithm 5.

**Proof.** See [65] (Theorem 3.3).  $\square$

We now show that the Mutex Watershed algorithm can be seen as a special case of the generic hierarchical Algorithm 5, for a specific choice of scales  $\lambda_k$  and functions  $Q_k(x) : \mathbb{R}^m \rightarrow \mathbb{R}$  (see definitions ((25), (26)) below).

**Scales  $\lambda_k$ :** Let  $\tilde{w}_k$  be the signed edge weights  $w : E \rightarrow \mathbb{R}$  ordered by decreasing absolute value  $|\tilde{w}_1| > |\tilde{w}_2| > \dots > |\tilde{w}_{t-1}|$ . If two edges share the same weight, then the weight is called  $\tilde{w}_k$  for both and  $E_k \subseteq E$  denotes the set of all edges with weight  $\tilde{w}_k$ . We then define the scales  $\lambda_k$  as

$$\lambda_k := \begin{cases} 1 & \text{if } k = 0 \\ \left| \frac{\tilde{w}_k}{2\tilde{w}_1} \right| & \text{otherwise.} \end{cases} \quad (25)$$

The continuous functions  $Q_k(x) : \mathbb{R}^{|E|} \rightarrow \mathbb{R}$  are defined as follows

$$Q_k(x) := \begin{cases} |E| \cdot \min_{x' \in \text{ISC}(\mathcal{G}, w)} \|x' - x\| & \text{if } k = 0 \\ \sum_{e \in E_k} x_e & \text{otherwise,} \end{cases} \quad (26)$$

where  $\text{ISC}(\mathcal{G}, w)$  is defined as

$$\text{ISC}(\mathcal{G}, w) := \text{SC}(\mathcal{G}, w) \cap \{0, 1\}^{|E|}. \quad (27)$$

In words,  $Q_0(x)$  is proportional to the distance between  $x$  and the closest point on the set  $\text{ISC}(\mathcal{G}, w)$ , whereas  $Q_k(x)$  depends only on the indicators  $x_e$  of edges in  $E_k$ , for  $k > 0$ .

Algorithm 6 is obtained by substituting the scales  $\lambda_k$  and functions  $Q_k(x)$  (respectively defined in Eqs. (25) and (26)) into Algorithm 5. The algorithm starts by setting  $M_0$  to  $\text{ISC}(\mathcal{G}, w)$ , i.e., by restricting the space of the solutions only to integer edge labelings  $x$  that do not include any conflicted cycles. Then, in the following iterations  $k \in 1, \dots, t-1$ , the algorithm solves a series of minimization sub-problems that in the most general case are NP-hard, even though they

**Algorithm 6.** Special Case of the General Hierarchical Algorithm 5 Obtained by Substituting Definition (25) and (26). With the Additional Assumption of Unique Signed Weights  $w : E \rightarrow \mathbb{R}$ , This Algorithm is Equivalent to the Mutex Watershed Algorithm 3. The Sequence of Functions  $Q_k : \mathbb{R}^m \rightarrow \mathbb{R}$  Defined in Eq. (26) is Sorted According to the Associated Scales  $\lambda_k$  in Eq. (25).  $\text{ISC}(\mathcal{G}, w)$  is Defined in Eq. (27)

**PWSMWS**( $Q_0, \dots, Q_{t-1}$ ):

$M_0 = \arg \min_{x \in \mathbb{R}^{|E|}} Q_0(x) = \text{ISC}(\mathcal{G}, w)$

**for**  $k \in 1, \dots, t-1$  **do**

$M_k = \arg \min_{x \in M_{k-1}} \sum_{e \in E_k} x_e$

**end**

**Return:** some  $x^* \in M_{t-1}$

involve a smaller set of edges  $E_k \subseteq E$ . Nevertheless, if we assume that all weights are distinct, then  $|E_k| = 1$  for all  $k$  and the solution to the sub-problems amounts to checking if the new edge can be labeled with  $x_e = 0$  without introducing any conflicted cycles. This procedure is identical to Algorithm 2: at every iteration, the Mutex Watershed tries to add an edge to the active set  $A$ , provided that no mutual exclusion constraints are violated.

In summary, the framework in [65] provides a new formulation of the Mutex Watershed Algorithm that is even applicable to graphs with tied edge weights. In practice, when edge weights are estimated by a CNN, we do not expect tied edge weights.

#### 4.4.2 Convergence of the Sequence of Minimizers

In this section, we see how Theorem 4.4 also suggests a minimization problem that is solved by the Mutex Watershed algorithm. A short summary is given in the final paragraph of the section.

First, we make sure that the conditions of Theorem 4.4 are satisfied when we apply it to Algorithm 6:

**Lemma 4.5.** *boundedsequence* Let us consider the scales  $\lambda_k$  and continuous functions  $Q_k(x) : \mathbb{R}^{|E|} \rightarrow \mathbb{R}$  respectively defined in Eqs. (25) and (26). For any value of  $p \in \mathbb{N}^+$ , let  $x_p \in \mathbb{R}^{|E|}$  be a minimizer of the function  $Q^p(x)$  defined in Eq. (24). Then, the minimizer  $x_p$  lies in the set  $\text{ISC}(\mathcal{G}, w)$ . From this, it follows that any sequence of minimizers  $(x_p)_{p>0}$  is bounded and the conditions of Theorem 4.4 are satisfied.

**Proof.** See Appendix A, which can be found on the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2020.2980827>.  $\square$

Then, given any  $p \in \mathbb{N}^+$  and the Definition ((25), (26)), we have that the minimization of the function  $Q^p(x)$  defined in Eq. (24) is given by the following problem:

$$\arg \min_{x \in \mathbb{R}^m} Q^p(x) = \arg \min_{x \in \mathbb{R}^m} \sum_{0 \leq k < t} \lambda_k^p Q_k(x) \quad (28)$$

$$= \arg \min_{x \in \text{ISC}(\mathcal{G}, w)} \sum_{1 \leq k < t} \left| \frac{\tilde{w}_k}{2\tilde{w}_1} \right|^p \sum_{e \in E_k} x_e \quad (29)$$

$$= \arg \min_{x \in \text{ISC}(\mathcal{G}, w)} \frac{1}{|2\tilde{w}_1|^p} \sum_{e \in E} |w_e|^p x_e, \quad (30)$$

where we used Lemma 4.5 and restricted the domain of the  $\arg \min$  operation to  $\text{ISC}(\mathcal{G}, w)$ , so that  $Q_0(x) = 0$  for all  $x \in \text{ISC}(\mathcal{G}, w)$ .

It follows from Lemma 4.5 and Theorem 4.4 that a sequence of minimizers  $(x_p)_{p>0}$  of the problem (30) converge, up to taking a subsequence, to the solution  $x^*$  returned by Algorithm 6. More specifically, we know that any minimizer  $x_p$  of (30) is in the discrete set  $\text{ISC}(\mathcal{G}, w)$ . Hence, the convergent sequence of minimizers  $(x_p)_{p>0}$  eventually becomes constant and there exists a  $p' \in \mathbb{N}^+$  large enough such that  $x_p = x^*$  for all  $p \geq p'$ . In other words, in the case of unique weights and  $p \geq p'$  large enough, the solution  $x^*$  of the Mutex Watershed Algorithm 6 solves the problem (30), which is just a rescaled version of the Mutex Watershed Objective we introduced in Section 4.2.

To summarize, we used the extended Power Watershed framework to show that the Mutex Watershed provides a solution to the minimization problem in Eq. (30) for  $p$  large enough. In particular, this problem suggested by the Power Watershed framework is the same one previously derived in Section 4.2 by linking the Mutex Watershed Algorithm to the multicut optimization problem.

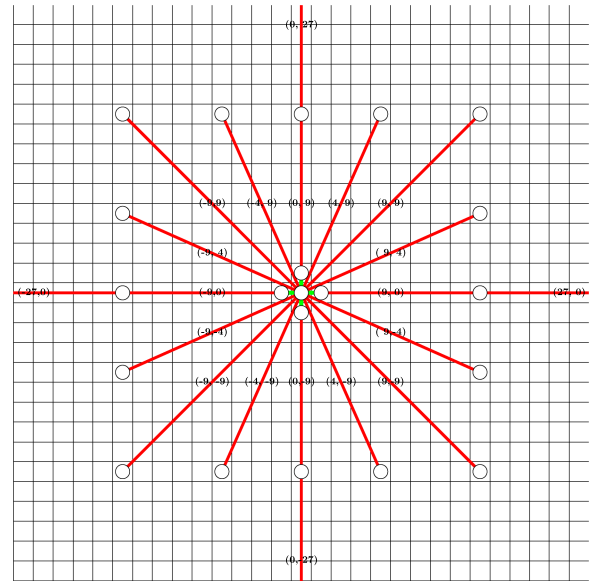
## 5 EXPERIMENTS

We evaluate the Mutex Watershed on the challenging task of neuron segmentation in electron microscopy (EM) image volumes. This application is of key interest in connectomics, a field of neuro-science that strives to reconstruct neural wiring digrams spanning complete central nervous systems. The task requires segmentation of neurons from electron microscopy images of neural tissue—a challenging endeavor, since segmentation has to be based only on boundary information (cell membranes) and some of the boundaries are not very pronounced. Besides, cells contain membrane-bound organelles, which have to be suppressed in the segmentation. Some of the neuron protrusions are very thin, but all of those need to be preserved in the segmentation to arrive at the correct connectivity graph. While a lot of progress is being made, currently only manual tracing or proof-reading yields sufficient accuracy for correct circuit reconstruction [68].

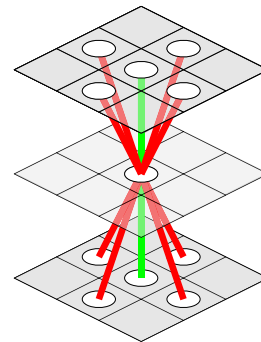
We validate the Mutex Watershed algorithm on the most popular neural segmentation challenge: ISBI2012 [42]. We estimate the edge weights using a CNN as described in Section 5.1 and compare with other entries in the leaderboard as well as with other popular post-processing methods for the same network predictions in Section 5.2.

### 5.1 Estimating Edge Weights With a CNN

The common first step to EM segmentation is to predict which pixels belong to a cell membrane using a CNN. Different post-processing methods are then used to obtain a segmentation, see Section 2 for an overview of such methods. The CNN can either be trained to predict boundary pixels [38], [41] or undirected affinities [39], [69] which express how likely it is for a pixel to belong to a different cell than its neighbors in the 6-neighborhood. In this case, the output of the network contains three channels, corresponding to left, down and next imaging plane neighbors in 3D. The affinities do not have to be limited to immediate



(a) XY-plane neighborhood with local attractive edges (green) and sparse repulsive edges (red) with approximate radius 9 and further long-range connections with distance 27



(b) Due to the great anisotropy of the data we limit the Z-plane edges to a distance of 1. The direct neighbors are attractive, whereas the indirect neighbors are repulsive.

Fig. 6. Local neighborhood structure of attractive (green) and repulsive (red) edges in the Mutex Watershed graph.

neighbors—in fact, [39] have shown that introduction of long-range affinities is beneficial for the final segmentation even if they are only used to train the network. Building on the work of [39], we train a CNN to predict short- and long-range affinities and then use those directly as weights for the Mutex Watershed algorithm.

We estimate the affinities / edge weights for the neighborhood structure shown in Fig. 6. To that end, we define local attractive and long-range repulsive edges. When attractive edges are only short-range, the solution will consist of spatially connected segments that cannot comprise “air bridges”. This holds true for both (lifted) multicut and for Mutex Watershed. We use a different pattern for in-plane and between-plane edges due to the great anisotropy of the data set. In more detail, we pick a sparse ring of in-plane repulsive edges and additional longer-range in-plane edges which are necessary to split regions reliably (see Fig. 6a). We also added connections to the indirect neighbors in the lower adjacent slice to ensure correct 3D connectivity (see Fig. 6b). In our experiments, we pick a subset of

repulsive edges, by using strides of 2 in the XY-plane in order to avoid artifacts caused by occasional very thick membranes. Note that the stride is not applied to local (attractive) edges, but only to long-range (repulsive) edges. The particular pattern used was selected after inspecting the size of typical regions. The specific pattern is the only one we have tried and was *not* optimized over.

In total,  $C^+$  attractive and  $C^-$  repulsive edges are defined for each pixel, resulting in  $C^+ + C^-$  output channels in the network. We partition the set of attractive / repulsive edges into subsets  $H^+$  and  $H^-$  that contain all edges at a specific offset:  $E^+ = \bigcup_{c=1}^{C^+} H_c^+$  for attractive edges, with  $H^-$  defined analogously. Each element of the subsets  $H_c^+$  and  $H_c^-$  corresponds to a specific channel predicted by the network. We further assume that weights take values in  $[0,1]$ .

### Network Architecture and Training

We use the 3D U-Net [43], [70] architecture, as proposed in [69].

Our training targets for attractive / repulsive edges  $w^{\pm}$  can be derived from a groundtruth label image  $L$  according to

$$w_{e=(i,j)}^+ = \begin{cases} 1, & \text{if } L_i = L_j \\ 0, & \text{otherwise} \end{cases} \quad (31)$$

$$w_{e=(i,j)}^- = \begin{cases} 0, & \text{if } L_i = L_j \\ 1, & \text{otherwise} \end{cases} \quad (32)$$

Here,  $i$  and  $j$  are the indices of vertices / image pixels. Next, we define the loss terms

$$\mathcal{J}_c^+ = - \frac{\sum_{e \in H_c^+} (1 - w_e^+) (1 - w_e^+)}{\sum_{e \in H_c^+} ((1 - w_e^+)^2 + (1 - w_e^+)^2)} \quad (33)$$

$$\mathcal{J}_c^- = - \frac{\sum_{e \in H_c^-} w_e^- w_e^-}{\sum_{e \in H_c^-} ((w_e^-)^2 + (w_e^-)^2)}, \quad (34)$$

for attractive edges (i.e., channels) and repulsive edges (i.e., channels).

Equation (33) is the Sørensen-Dice coefficient [71], [72] formulated for fuzzy set membership values. During training we minimize the sum of attractive and repulsive loss terms  $\mathcal{J} = \sum_c \mathcal{J}_c^+ + \sum_c \mathcal{J}_c^-$ . This corresponds to summing up the channel-wise Sørensen-Dice loss. The terms of this loss are robust against prediction and / or target sparsity, a desirable quality for neuron segmentation: since membranes are locally two-dimensional and thin, they occupy very few pixels in three-dimensional the volume. More precisely, if  $w_e^+$  or  $w_e^-$  (or both) are sparse, we can expect the denominator  $\sum_e ((w_e^+)^2 + (w_e^-)^2)$  to be small, which has the effect that the numerator is adaptively weighted higher. In this sense, the Sørensen-Dice loss at every pixel  $i$  is conditioned on the global image statistics, which is not the case for a Hamming-distance based loss like Binary Cross-Entropy or Mean Squared Error.

We optimize this loss using the Adam optimizer [73] and additionally condition learning rate decay on the Adapted Rand Score [42] computed on the training set every 100

TABLE 1  
Results on the ISBI 2012 EM Segmentation Challenge

Method	Rand-Score	VI-Score
UNet + MWS	<b>0.98792</b>	<b>0.99183</b>
ResNet + LMC [76]	0.98788	0.99072
SCN + LMC [77]	0.98680	0.99144
M2FCN-MFA [78]	0.98383	0.98981
FusionNet + LMC [44]	0.98365	0.99130

(a) Top five entries at time of submission. Our Mutex Watershed (MWS) is state-of-the-art without relying on the complex lifted multicut postprocessing used by most other top entries.

Method	Rand-Score	VI-Score	Time [s]
MWS	<b>0.98792</b>	<b>0.99183</b>	43.3
MC-FULL	0.98029	0.99044	9415.8
LMC	0.97990	0.99007	966.0
THRESH	0.91435	0.96961	0.2
WSDT	0.88336	0.96312	4.4
MC-LOCAL	0.70990	0.86874	1410.7
WS	0.63958	0.89237	4.9

(b) Comparison to other segmentation strategies, all of which are based on our CNN. Runtimes were measured on a single thread of a Intel Xeon CPU E5-2650 v3 @ 2.30GHz.

iterations. During training, we augment the data set by performing in-plane rotations by multiples of 90 degrees, flips along the  $X$ - and  $Y$ -axis as well as elastic deformations. At prediction time, we use test time data augmentation, presenting the network with seven different versions of the input obtained by a combination of rotations by a multiple of 90 degrees, axis-aligned flips and transpositions. The network predictions are then inverse-transformed to correspond to the original image, and the results averaged.

## 5.2 ISBI Challenge

The ISBI 2012 EM Segmentation Challenge [42] is the neuron segmentation challenge with the largest number of competing entries. The challenge data contains two volumes of dimensions  $1.5 \times 2 \times 2$  microns and has a resolution of  $50 \times 4 \times 4$  nm per pixel. The groundtruth is provided as binary membrane labels, which can easily be converted to a 2D, but not 3D segmentation. To train a 3D model, we follow the procedure described in [38].

The test volume has private groundtruth; results can be submitted to the leaderboard. They are evaluated based on the Adapted Rand Score (Rand-Score) and the Variation of Information Score (VI-Score) [42].

Our method holds the top entry in the challenge's leaderboard<sup>3</sup> at the time of submission, see Table 1 a. This is especially remarkable insofar as it is simpler than the methods holding the other top entries. Three out of four rely on a CNN to predict boundary locations and postprocess its output with the complex pipeline described in [38]. This post-processing first generates superpixels via distance transform watersheds. Then it computes a merge cost for local and long-range connections between superpixels. Based on this, it defines a lifted multicut partitioning problem that is

3. [http://brainiac2.mit.edu/isbi\\_challenge/leaders-board-new](http://brainiac2.mit.edu/isbi_challenge/leaders-board-new)

solved approximately. In contrast, our method finds an optimal solution of its objective purely on the pixel level.

#### Comparison With Other Segmentation Methods

The weights predicted by the CNN described above can be post-processed directly by the Mutex Watershed algorithm. To ensure a fair comparison, we transform the same CNN predictions into a segmentation using basic and state-of-the-art post-processing methods. We start from simple thresholding (THRESH) and seeded watershed. Since these cannot take long-range repulsions into account, we generate a boundary map by taking the maximum<sup>4</sup> values over the attractive edge channels. Based on this boundary map, we introduce seeds at the local minima (WS) and at the maxima of the smoothed distance transform (WSDT). For both variants, the degree of smoothing was optimized such that each region receives as few seeds as possible, without however causing severe under-segmentation. The performance of these three baseline methods in comparison to Mutex Watershed is summarized in Table 1 b. The methods were applied only in 2D, because the high degree of anisotropy leads to inferior results when applied in 3D. In contrast, the Mutex Watershed can be applied in 3D out of the box and yields significantly better 2D segmentation scores.

Qualitatively, we show patches of results in Fig. 7. The major failure case for WS (Fig. 7e) and WSDT (Fig. 7f) is over-segmentation caused by over-seeding a region. The major failure case for THRESH is under-segmentation due to weak boundary evidence (see Fig. 7d). In contrast, the Mutex Watershed produces a better segmentation, only causing minor over-segmentation (see Fig. 7a and 7b).

Note that, in contrast to most pixel-based postprocessing methods, our algorithm can take long range predictions into account. To compare with methods which share this property, we turn to the multicut and lifted multicut-based partitioning for neuron segmentations as introduced in [29] and [36]. As proposed in [74], we compute costs corresponding to edge cuts from the affinities estimated by the CNN via

$$s_e = \begin{cases} \log \frac{w_e^+}{1-w_e^+}, & \text{if } e \in E^+ \\ \log \frac{1-w_e^-}{w_e^-}, & \text{otherwise,} \end{cases} \quad (35)$$

We set up two multicut problems: the first is induced only by the short-range edges (MC-LOCAL), the other by short- and long-range edges together (MC-FULL). Note that the solution to the full connectivity problem can contain “air bridges”, i.e., pixels that are connected only by long-range edges, without a path along the local edges connecting them. However, we found this not to be a problem in practice. In addition, we set up a lifted multicut problem from the same edge costs.

Both problems are NP-hard, hence it is not feasible to solve them exactly on large grid graphs. For our experiments, we use the approximate Kernighan Lin [33], [75] solver. Even this allows us to only solve individual 2D problems at a time. The results for MC-LOCAL and MC-FULL can be found in Table 1 b. The MC-LOCAL approach scores poorly because it under-segments heavily. This observation

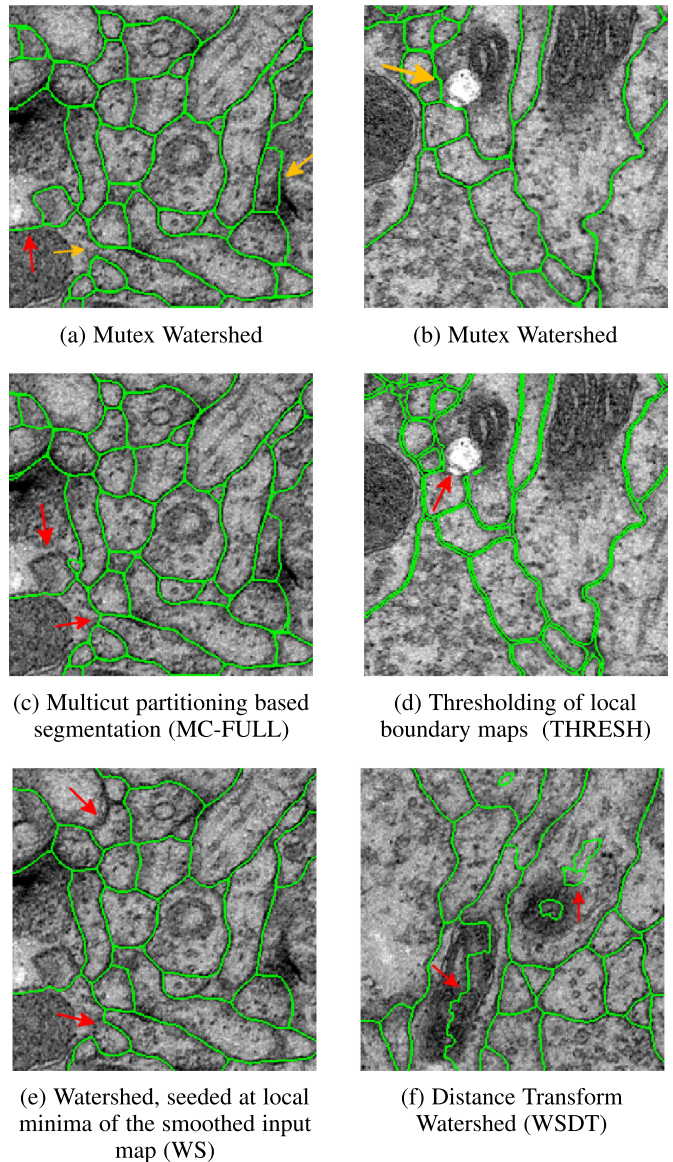


Fig. 7. Mutex Watershed and baseline segmentation algorithms applied on the ISBI Challenge test data. Red arrows point out major errors. Orange arrows point to difficult, but correctly segmented regions. All methods share the same input maps.

emphasizes the importance of incorporating the longer-range edges. The MC-FULL and LMC approaches perform well. Somewhat surprisingly, the Mutex Watershed yields a better segmentation still, despite being much cheaper in inference. We note that both MC-FULL, LMC and the Mutex Watershed are evaluated on the same long-range affinity maps (i.e., generated by the same CNN with the same set of weights).

## 6 CONCLUSION AND DISCUSSION

We have presented a fast algorithm for the clustering of graphs with both attractive and repulsive edges. The ability to consider both gives a valid alternative to other popular graph partitioning algorithms that rely on a stopping criterion or seeds. The proposed method has low computational complexity in imitation of its close relative, Kruskal’s algorithm. We have shown which objective this algorithm

4. The maximum is chosen to preserve boundaries.

optimizes exactly, and that this objective emerges as a specific case of the multicut objective. It is possible that recent interesting work [63] on partial optimal solutions may open an avenue for an alternative proof.

Finally, we have found that the proposed algorithm, when presented with informative edge costs from a good neural network, outperforms all known methods on a competitive bioimage partitioning benchmark, including methods that operate on the very same network predictions.

## ACKNOWLEDGMENTS

This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projekt-nummer 240245660 - SFB 1129 and the Baden-Württemberg Stiftung Elite PostDoc Program. Steffen Wolf and Alberto Bailoni contributed equally to this work.

## REFERENCES

- [1] S. Wolf *et al.*, “The mutex watershed: Efficient, parameter-free image partitioning,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 571–587.
- [2] L. Vincent and P. Soille, “Watersheds in digital spaces: An efficient algorithm based on immersion simulations,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [3] S. Beucher and C. Lantuéjoul, “Use of watersheds in contour detection,” in *Proc. Int. Workshop Image Process.*, 1979, pp. 2.1–2.12.
- [4] S. Beucher and F. Meyer, “The morphological approach to segmentation: The watershed transformation,” *Opt. Eng.*, vol. 34, pp. 433–433, 1992.
- [5] R. Pohle and K. D. Toennies, “Segmentation of medical images using adaptive region growing,” in *Proc. Med. Imag.: Image Process.*, 2001, pp. 1337–1346.
- [6] M. A. Alattar, N. F. Osman, and A. S. Fahmy, “Myocardial segmentation using constrained multi-seeded region growing,” in *Proc. Int. Conf. Image Anal. Recognit.*, 2010, pp. 89–98.
- [7] S. Poonguzhali and G. Ravindran, “A complete automatic region growing method for segmentation of masses on ultrasound images,” in *Proc. Int. Conf. Biomed. Pharm. Eng.*, 2006, pp. 88–92.
- [8] J. Wu, S. Poehlman, M. D. Noseworthy, and M. V. Kamath, “Texture feature based automated seeded region growing in abdominal MRI segmentation,” in *Proc. Int. Conf. BioMed. Eng. Informat.*, 2008, vol. 2, pp. 263–267.
- [9] A. Q. Al-Faris, U. K. Ngah, N. A. M. Isa, and I. L. Shuaib, “Computer-aided segmentation system for breast MRI tumour using modified automatic seeded region growing (BMRI-MASRG),” *J. Digit. Imag.*, vol. 27, no. 1, pp. 133–144, 2014.
- [10] J. Shan, H.-D. Cheng, and Y. Wang, “A novel automatic seed point selection algorithm for breast ultrasound images,” in *Proc. 19th Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.
- [11] D. M. N. Mubarak, M. M. Sathik, S. Z. Beevi, and K. Revathy, “A hybrid region growing algorithm for medical image segmentation,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 4, no. 3, 2012, Art. no. 61.
- [12] M. Abdelsamea, “An enhancement neighborhood connected segmentation for 2D-Cellular Image,” *Int. J. Biosci., Biochemistry Bioinf.*, vol. 1, no. 4, pp. 256–260, 2011.
- [13] A. Q. Al-Faris, U. K. Ngah, N. A. M. Isa, and I. L. Shuaib, “Breast MRI tumour segmentation using modified automatic seeded region growing based on particle swarm optimization image clustering,” in *Proc. Soft Comput. Ind. Appl.*, 2014, pp. 49–60.
- [14] M. Grimaud, “New measure of contrast: The dynamics,” in *Proc. Image Algebra Morphological Process.*, 1992, vol. 1769, pp. 292–305.
- [15] S. Beucher, “Watershed, hierarchical segmentation and waterfall algorithm,” in *Proc. Int. Symp. Memory Manage.*, 1994, vol. 94, pp. 69–76.
- [16] C. Vachier and F. Meyer, “Extinction value: A new measurement of persistence,” in *Proc. Workshop Nonlinear Signal Image Process.*, 1995, vol. 1, pp. 254–257.
- [17] L. Najman and M. Schmitt, “Geodesic saliency of watershed contours and hierarchical segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 12, pp. 1163–1173, Dec. 1996.
- [18] P. Soille, “Constrained connectivity for hierarchical image decomposition and simplification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1132–1145, Jul. 2008.
- [19] B. Perret, J. Cousty, S. J. F. Guimaraes, and D. S. Maia, “Evaluation of hierarchical watersheds,” *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1676–1688, Apr. 2018.
- [20] F. Meyer, “Morphological multiscale and interactive segmentation,” in *Proc. Workshop Image Anal. Multimedia Interactive Services*, 1999, pp. 369–377.
- [21] L. Najman, “On the equivalence between hierarchical segmentations and ultrametric watersheds,” *J. Math. Imag. Vis.*, vol. 40, no. 3, pp. 231–247, 2011.
- [22] P. Salembier and L. Garrido, “Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval,” *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 561–576, Apr. 2000.
- [23] F. Malmberg, R. Strand, and I. Nyström, “Generalized hard constraints for graph segmentation,” in *Proc. Scandinavian Conf. Image Anal.*, 2011, pp. 36–47.
- [24] P. F. Felzenszwalb and D. P. Huttenlocher, “Efficient graph-based image segmentation,” *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.
- [25] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
- [26] B. R. Kiran and J. Serra, “Global–local optimizations by hierarchical cuts and climbing energies,” *Pattern Recognit.*, vol. 47, no. 1, pp. 12–24, 2014.
- [27] L. Guigues, J. P. Cocquerez, and H. Le Men, “Scale-sets image analysis,” *Int. J. Comput. Vis.*, vol. 68, no. 3, pp. 289–317, 2006.
- [28] M. Keuper, S. Tang, Y. Zhongjie, B. Andres, T. Brox, and B. Schiele, “A multi-cut formulation for joint segmentation and tracking of multiple objects,” 2016, *arXiv:1607.06317*.
- [29] B. Andres *et al.*, “Globally optimal closed-surface segmentation for connectomics,” in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 778–791.
- [30] J. Yarkony, A. Ihler, and C. C. Fowlkes, “Fast planar correlation clustering for image segmentation,” in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 568–581.
- [31] C. Pape, T. Beier, P. Li, V. Jain, D. D. Bock, and A. Kreshuk, “Solving large multicut problems for connectomics via domain decomposition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1–10.
- [32] Z. Levi and D. Zorin, “Strict minimizers for geometric optimization,” *ACM Trans. Graph.*, vol. 33, no. 6, pp. 185:1–185:14, Nov. 2014.
- [33] M. Keuper, E. Levinkov, N. Bonneel, G. Lavoué, T. Brox, and B. Andres, “Efficient decomposition of image and mesh graphs by lifted multicuts,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1751–1759.
- [34] E. Levinkov, A. Kirillov, and B. Andres, “A comparative study of local search algorithms for correlation clustering,” in *Proc. German Conf. Pattern Recognit.*, 2017, pp. 103–114.
- [35] C. Zhang, J. Yarkony, and F. A. Hamprecht, “Cell detection and segmentation using correlation clustering,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2014, pp. 9–16.
- [36] A. Hornáková, J.-H. Lange, and B. Andres, “Analysis and optimization of graph decompositions by lifted multicuts,” in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1539–1548.
- [37] T. Beier, B. Andres, U. Köthe, and F. A. Hamprecht, “An efficient fusion move algorithm for the minimum cost lifted multicut problem,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 715–730.
- [38] T. Beier, C. Pape, N. Rahaman, and T. E. A. Prange, “Multicut brings automated neurite segmentation closer to human performance,” *Nature Methods*, vol. 14, no. 2, pp. 101–102, 2017.
- [39] K. Lee, J. Zung, P. Li, V. Jain, and H. S. Seung, “Superhuman accuracy on the SNEMI3D connectomics challenge,” 2017, *arXiv preprint arXiv:1706.00120*.
- [40] V. Jain, J. F. Murray, F. Roth, S. Turaga, V. Zhigulin, K. L. Briggman, M. N. Helmstaedter, W. Denk, and H. S. Seung, “Supervised learning of image restoration with convolutional networks,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
- [41] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 2843–2851.

- [42] I. Arganda-Carreras *et al.*, "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Front. Neuroanatomy*, vol. 9, 2015, Art. no. 142.
- [43] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.
- [44] T. M. Quan, D. G. Hilderbrand, and W.-K. Jeong, "FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics," 2016, *arXiv:1612.05360*.
- [45] J. Nunez-Iglesias, R. Kennedy, T. Parag, J. Shi, and D. Chklovskii, "Machine learning of hierarchical clustering to segment 2D and 3D images," *PLoS One*, vol. 8, 2013, Art. no. e71715.
- [46] S. Knowles-Barley *et al.*, "RhoanaNet pipeline: Dense automatic neural annotation," 2016, *arXiv:1611.06973*.
- [47] M. G. Uzunbas, C. Chen, and D. Metaxas, "Optree: A learning-based adaptive watershed algorithm for neuron segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2014, pp. 97–105.
- [48] Y. Meirovitch *et al.*, "A multi-pass approach to large-scale connectomics," 2016, *arXiv preprint:1612.02120*.
- [49] M. Januszewski *et al.*, "High-precision automated reconstruction of neurons with flood-filling networks," *Nature Methods*, vol. 15, pp. 605–610, 2018.
- [50] A. Zlateski and H. S. Seung, "Image segmentation by size-dependent single linkage clustering of a watershed basin graph," 2015, *arXiv:1505.00249*.
- [51] T. Parag *et al.*, "Anisotropic EM segmentation by 3D affinity learning and agglomeration," 2017, *arXiv preprint 1707.08935*.
- [52] S. C. Turaga *et al.*, "Convolutional networks can learn to generate affinity graphs for image segmentation," *Neural Comput.*, vol. 22, no. 2, pp. 511–538, 2010.
- [53] K. Briggman, W. Denk, S. Seung, M. N. Helmstaedter, and S. C. Turaga, "Maximin affinity learning of image segmentation," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2009, pp. 1865–1873.
- [54] S. Wolf, L. Schott, U. Köthe, and F. Hamprecht, "Learned watershed: End-to-end learning of seeded segmentation," *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2030–2038.
- [55] M. Bai and R. Urtasun, "Deep watershed transform for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2858–2866.
- [56] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1395–1403.
- [57] I. Kokkinos, "Pushing the boundaries of boundary detection using deep learning," 2015, *arXiv:1511.07386*.
- [58] J. Cai, L. Lu, Z. Zhang, F. Xing, L. Yang, and Q. Yin, "Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2016, pp. 442–450.
- [59] F. Meyer, "Topographic distance and watershed lines," *Signal Process.*, vol. 38, no. 1, pp. 113–125, 1994.
- [60] F. Meyer, "Minimum spanning forests for morphological segmentation," in *Mathematical Morphology and its Applications to Image Processing*, Berlin, Germany: Springer, 1994, pp. 77–84.
- [61] A. X. Falcão, J. Stolfi, and R. de Alencar Lotufo, "The image foresting transform: Theory, algorithms, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 19–29, Jan. 2004.
- [62] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Third Edition*, 3rd ed., Cambridge, MA, USA: The MIT Press, 2009.
- [63] J.-H. Lange, A. Karrenbauer, and B. Andres, "Partial optimality and fast lower bounds for weighted correlation clustering," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2898–2907.
- [64] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watershed: A unifying graph-based optimization framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1384–1399, Jul. 2011.
- [65] L. Najman, "Extending the power watershed framework thanks to  $\Gamma$ -convergence," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 2275–2292, 2017.
- [66] G. Dal Maso, *An Introduction to  $\Gamma$ -Convergence*, vol. 8. Berlin, Germany: Springer, 2012.
- [67] A. Braides, "A handbook of  $\gamma$ -convergence," in *Handbook of Differential Equations: Stationary Partial Differential Equations*, vol. 3. Amsterdam, Netherlands: Elsevier, 2006, pp. 101–213.
- [68] P. Schlegel, M. Costa, and G. S. X. E. Jefferis, "Learning from connectomics on the fly," *Current Opinion Insect Sci.*, vol. 24, pp. 96–105, 2017.
- [69] J. Funke *et al.*, "Large scale image segmentation with structured loss based deep learning for connectome reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1669–1680, Jul. 2019.
- [70] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2016, pp. 424–432.
- [71] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [72] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons," *Biol. Skr.*, vol. 5, pp. 1–34, 1948.
- [73] D. P. Kinga and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015. [Online]. Available: <https://iclr.cc/archive/www/doku.php?id=iclr2015:accepted-main.html>
- [74] B. Andres *et al.*, "Globally optimal closed-surface segmentation for connectomics," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 778–791.
- [75] B. W. Kernighan and S. Lin, "An efficient heuristic procedure for partitioning graphs," *The Bell Syst. Tech. J.*, vol. 49, no. 2, pp. 291–307, 1970.
- [76] C. Xiao, J. Liu, X. Chen, H. Han, C. Shu, and Q. Xie, "Deep contextual residual network for electron microscopy image segmentation in connectomics," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 378–381.
- [77] M. Weiler, F. A. Hamprecht, and M. Storath, "Learning steerable filters for rotation equivariant CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 849–858.
- [78] W. Shen, B. Wang, Y. Jiang, Y. Wang, and A. L. Yuille, "Multi-stage multi-recursive-input fully convolutional networks for neuronal boundary detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2410–2419.



**Steffen Wolf** is currently working toward the PhD degree in the Interdisciplinary Center for Scientific Computing, at Heidelberg University. His research interests include computer vision, structured prediction, and deep learning with applications in image segmentation.



**Alberto Bailoni** is currently working toward the PhD degree in the Interdisciplinary Center for Scientific Computing, at Heidelberg University. His research interests include computer vision, image segmentation, deep learning and clustering algorithms, with a focus on their application to automated reconstruction of neural circuit connectivity.



**Constantin Pape** is currently working toward the PhD degree at Heidelberg University, visiting the EMBL Heidelberg. His research interests include biomedical imaging with a focus on deep learning and instance segmentation for large EM datasets.



**Nasim Rahaman** is currently working toward his PhD degree at Max-Planck-Institute for Intelligent Systems Tübingen and Mila, Québec. His research interests lie in the intersection of causal inference, deep representation learning and applications to spatiotemporal modeling of partially observable systems.



**Ullrich Köthe** received the diploma degree in physics from the University of Rostock, Rostock, Germany, and the PhD degree in computer science from the University of Hamburg, Hamburg, Germany. He is currently an associate professor for computer science in the Interdisciplinary Center for Scientific Computing, at Heidelberg University. His research focuses on the connection between image analysis and machine learning, and in particular on the interpretability of machine learning results.



**Anna Kreshuk** received the diploma degree in maths from Lomonosov Moscow State University, Moscow, Russia, and the PhD degree in computer science in Heidelberg. She is currently a group leader in EMBL Heidelberg. Her research focuses on automating analysis of microscopy images with machine learning.



**Fred A. Hamprecht** is professor for image analysis and learning, at Heidelberg University. His primary research interests are algorithms and how they can be used to solve biological problems.

▷ For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/csdl](http://www.computer.org/csdl).