STRUCTURED LEARNING OF ASSIGNMENT MODELS FOR NEURON RECONSTRUCTION TO MINIMIZE TOPOLOGICAL ERRORS

Jan Funke[⋆‡]   *Jonas Klein*[‡]   *Francesc Moreno-Noguer*[⋆]   *Albert Cardona*[†]   *Matthew Cook*[‡]

[⋆] Institut de Robòtica i
Informàtica Industrial
UPC/CSIC Barcelona

[†] Janelia Research Campus
VA, Ashburn

[‡] Institute of Neuroinformatics
UZH/ETH Zurich

ABSTRACT

Structured learning provides a powerful framework for empirical risk minimization on the predictions of structured models. It allows end-to-end learning of model parameters to minimize an application specific loss function. This framework is particularly well suited for discrete optimization models that are used for neuron reconstruction from anisotropic electron microscopy (EM) volumes. However, current methods are still learning unary potentials by training a classifier that is agnostic about the model it is used in. We believe the reason for that lies in the difficulties of (1) finding a representative training sample, and (2) designing an application specific loss function that captures the quality of a proposed solution. In this paper, we show how to find a representative training sample from human generated ground truth, and propose a loss function that is suitable to minimize topological errors in the reconstruction. We compare different training methods on two challenging EM-datasets. Our structured learning approach shows consistently higher reconstruction accuracy than other current learning methods.

1. INTRODUCTION

The reconstruction of neurons from volumes of electron microscopy (EM) images is an important step in gaining new insights into the function of nervous systems. However, at present these findings are the result of slow and tedious manual reconstruction work, and further insights are hindered by this bottleneck. Consequently, the (semi-)automatization of the reconstruction is of great importance and became an active field of research over the last decade.

Given the challenging nature of EM images, current models need sensible cost functions to be robust to noise and missing data. Using machine learning methods, these costs can be learned from human annotated training data. Given that the provision of ground truth is tedious, and consequently the available datasets are small, it is necessary to find machine learning methods that generalize well without overfitting.

In this paper, we present a structured learning framework to train *assignment models* [1, 2] for anisotropic neuron reconstruction. Our contributions are: (1) We show how to gen-erate a *training sample* suitable for structured learning from human annotated ground truth. (2) We introduce a loss for structured learning, which minimizes topological errors during learning. (3) We show how our learning framework can be used to train on skeleton (*i.e.*, non-volumetric) annotations, which are in practice much faster to obtain.

Related Work. Assignment models constitute the current state of the art for the reconstruction of neurons from anisotropic volumes, as obtained by serial section EM [1, 2]. These models enumerate and price possible assignments of candidate segments across sections of EM stacks (see Figure 1 for an overview and Section 2 for details). A final segmentation is found by selecting a cost minimal and consistent subset of all assignments.

Learning these models consists of finding suitable assignment costs. Currently, these costs are set by hand [3, 2, 4], learned from a random forest classifier based on positive and negative assignment examples [1, 5], or found via grid-search on linear weights for a small number of features [6]. Except for grid-search, which does not scale to larger sets of parameters, none of the currently used training methods implements real end-to-end learning. In this paper, we show how to overcome these limitations by performing structured learning on a sensible loss function.

2. LEARNING OF ASSIGNMENT MODELS

Assignment models for anisotropic neuron reconstruction introduce n binary indicator variables $\mathbf{z} \in \{0,1\}^n$ to represent all possible assignments of 2D neuron candidates across all pairs of sections of a volume [1, 2]. Linear constraints are formulated on the binary assignment indicators to ensure that a solution is consistent, *i.e.*, no pair of overlapping candidates is selected.

Consequently, the set of consistent solutions $\mathcal{Z}$ to an assignment problem is characterized by linear constraints $(A, \mathbf{b})$:

$$\mathcal{Z} = \{\mathbf{z} \in \{0,1\}^n | A\mathbf{z} \preceq \mathbf{b}\}, \tag{1}$$

where we write $\mathbf{a} \preceq \mathbf{b}$ to say that $\mathbf{a}$ is element-wise less than or equal to $\mathbf{b}$. Given a cost vector $\mathbf{c}$ for the assignment variables, the optimal assignment is the solution to the integer
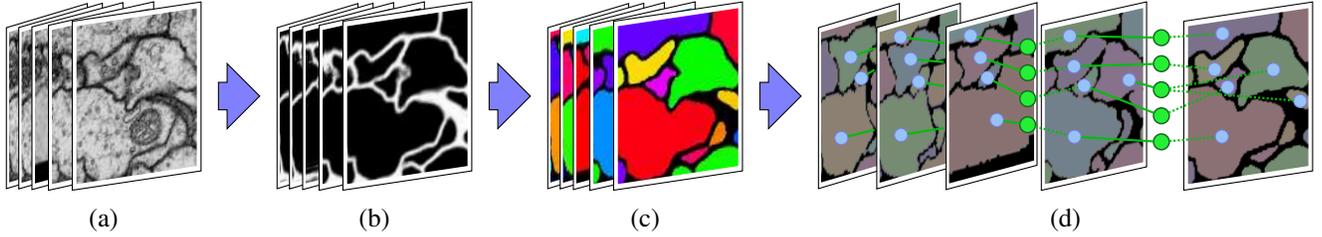
Fig. 1: Assignment model for anisotropic neuron reconstruction. From a stack of raw images (a), a pixel classifier is used to predict membrane locations (b). 2D neuron candidates are extracted for each section (c), and possible assignments are enumerated between candidates of adjacent sections (d). In the model, each assignment (green nodes) is represented by a binary variable z_i and has an associated cost c_i for selecting it. Learning of an assignment model consists of learning these costs from annotated ground truth. Note that for the purpose of illustration, only a few candidates and possible assignments are shown.

linear program $\min_{\mathbf{z} \in \mathcal{Z}} \langle \mathbf{c}, \mathbf{z} \rangle$. Without loss of generality, we assume that the cost c_i for selecting an assignment z_i is a weighted sum of features $\boldsymbol{\phi}_i$ extracted for this assignment:

$$\mathbf{c} = \Phi \cdot \boldsymbol{w} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_n]^\mathsf{T} \cdot \boldsymbol{w}. \qquad (2)$$

Using the structured learning framework [7], we find the optimal $\boldsymbol{w}$ given annotated training data $(\boldsymbol{\phi}, \mathbf{z}')$. More specifically, we use the margin rescaling variant to find the weights $\boldsymbol{w}^*$ as the minimizer of

$$L(\boldsymbol{w}) = \lambda |\boldsymbol{w}|^2 + \max_{z \in \mathcal{Z}} \; [\langle \Phi \boldsymbol{w}, \mathbf{z}' \rangle - \langle \Phi \boldsymbol{w}, \mathbf{z} \rangle] + \Delta(\mathbf{z}', \mathbf{z}), \; (3)$$

where λ is the regularizer weight and $\Delta(\mathbf{z}', \mathbf{z})$ is an application specific loss function. In order for this method to be successful, two problems need to be solved: (1) a representative training sample $\mathbf{z}'$ has to be found, and (2) a sensible loss function $\Delta(\mathbf{z}', \mathbf{z})$ has to be designed.

2.1. Training Sample $\mathbf{z}'$

Even apart from the difficulties in obtaining unambiguous human generated ground truth for the neuron reconstruction problem in the first place, the provision of $\mathbf{z}'$ is not trivial: We have to find a member of $\mathcal{Z}$, *i.e.*, the set of all possible reconstructions *using the found 2D neuron candidates*, that is as close as possible to the human annotated ground truth. We have to note that the extracted 2D neuron candidates can be imperfect and thus there might not be a $\mathbf{z} \in \mathcal{Z}$ that corresponds to the human annotated ground truth. Consequently, we have to accept that the training sample $\mathbf{z}'$ will only represent a best-effort reconstruction and not the ground truth.

In order to find this best-effort reconstruction in a principled way, we assign a local ground truth matching score g_i to each assignment. Let $\Omega = [1, W] \times [1, H] \times [1, D]$ be the set of all discrete pixel locations in a stack of size $W \times H \times D$. We assume a ground truth labeling $x : \Omega \mapsto K$ that assigns a unique label $k \in K$ to each ground truth segment in the

volume. Let $u(i)$ and $v(i)$ denote the section indices that are linked by assignment z_i. We denote by $A_i \subset \Omega$ the set of pixels of section $u(i)$ and $v(i)$ that are merged by the assignment z_i. Similarly, let $G_i^k \subset \Omega$ denote the set of pixels that are labelled to belong to the same region k in the ground truth, limited to the sections $u(i)$ and $v(i)$. For each pair of assignment i and ground truth label k, we compute a similarity g_i^k that rewards overlap between the sets A_i and G_i^k and punishes set differences:

$$g_i^k = \underbrace{|G_i^k \cap A_i|}_{\text{overlap}} - \underbrace{|G_i^k \setminus A_i| + |A_i \setminus G_i^k|}_{\text{set difference}}. \qquad (4)$$

The final matching score g_i of an assignment z_i is the maximal similarity with any ground truth label:

$$g_i = \max_{k \in K} g_i^k. \qquad (5)$$

To find the training sample $\mathbf{z}'$, we solve an inference problem where we replace the costs in (2) with the negative matching score.

$$\mathbf{z}' = \arg\min_{\mathbf{z} \in \mathcal{Z}} \; <-\mathbf{g}, \mathbf{z}> . \qquad (6)$$

This way, assignments are chosen linking candidates with the largest overlap with the same ground truth region.

2.2. Loss $\Delta(\mathbf{z}', \mathbf{z})$

Ideally, we would use the error measure that we use to evaluate the results of our automatic reconstruction as $\Delta(\mathbf{z}', \mathbf{z})$. However, we have to make sure that the maximization in (3) is still tractable.

We propose to minimize the Tolerant Edit Distance (TED) [8] during training. The TED measures the minimal weighted sum of split and merge errors between two labelings x and y, but ignores small tolerable errors explainable by boundary shifts up to a threshold distance θ:

$$\text{TED}(x, y) = \min_{y' \in \mathcal{Y}^\theta(y)} \alpha \, \text{s}(x, y') + \beta \, \text{m}(x, y'), \qquad (7)$$

where $\mathcal{Y}^\theta(y)$ is the set of all relabelings of y obtained by shifting boundaries by at most θ units, and $s(x, y)$ and $m(x, y)$ measure the number of splits and merges between x and y, respectively. By providing different weights α and β, we can prioritize errors during training. In fact, if the weights correspond to the average time needed to fix a split or merge error, we are minimizing the time-to-fix (TTF) during training.

For the computation of the TED of a solution $\mathbf{z}$, we have to transform it into a labeling $y_{\mathbf{z}} : \Omega \mapsto K_{\mathbf{z}}$, which assigns a unique label to each found neuron in the volume Ω. However, since the relation between a solution $\mathbf{z}$ and its labeling $y_{\mathbf{z}}$ is not trivial, there is no straightforward way to incorporate the TED into our learning objective (3). Therefore, we propose to use a first-order approximation, that we refer to as *structured learning with TED* (SL-TED), where we set

$$\Delta(\mathbf{z}', \mathbf{z}) = \langle \mathbf{l}, \mathbf{z} \rangle + c \approx \text{TED}(y_{\mathbf{z}'}, y_{\mathbf{z}}). \quad (8)$$

Here, the assignment losses l_i in $\mathbf{l}$ represent a reward (if negative) or punishment (if positive) of using assignment i. Since this approximation is linear in $\mathbf{z}$, the structure of the maximization in (3) is very similar to the assignment inference problem, making it tractable in practice. To determine the loss l_i, we measure the TED between the best-effort labeling $y_{\mathbf{z}'}$ and the labeling $y_{\bar{\mathbf{z}}(i)}$ obtained by inverting decision z_i in $\mathbf{z}'$:

$$l_i = (1 - 2z_i')\,\text{TED}(y_{\mathbf{z}'}, y_{\bar{\mathbf{z}}(i)}) \quad \text{and} \quad c = \sum_{i:z_i'=1} -l_i, \quad (9)$$

where the factor $(1 - 2z_i')$ ensures that we get a reward for selecting an assignment that is part of the best-effort. This way, we punish not selecting the corresponding z_i. Since the constraints (1) might not allow inverting single variables in isolation, we identify a minimal group of variables that have to be inverted as well to obtain a consistent solution: for each assignment i, we find a reconstruction $\bar{\mathbf{z}}(i) \in \mathcal{Z}$ that has z_i inverted and minimizes the Hamming distance to $\mathbf{z}'$:

$$\bar{\mathbf{z}}(i) = \underset{\substack{z \in \mathcal{Z}: \\ z_i = 1 - z_i'}}{\arg\min} \ \sum_j |z_j - z_j'|. \quad (10)$$

3. RESULTS

We use two publicly available datasets for our experiments, which we refer to as DROSOPHILA [9], which consists of two stacks of 20 EM sections with $4 \times 4 \times 40nm$ resolution ($1024 \times 1024 \times 20$ pixels), and MOUSE CORTEX [10], which consists of two stacks of 100 EM sections with $6 \times 6 \times 30nm$ resolution ($1024 \times 1024 \times 100$ pixels).

We split the parts for which ground truth was available into two stacks of equal size (2×10 sections for DROSOPHILA and 2×50 sections for MOUSE CORTEX). For each dataset, we trained all methods on a sample $\mathbf{z}'$ (see Section 2.1) extracted from the first stack and report the results on the second stack.

We trained and evaluated the assignment model implemented in SOPNET [1], using membrane predictions from [11], and 2D neuron candidates extracted from component trees [1]. We used the default features implemented in SOPNET for Φ.

Comparison of Learning Methods. We compare the structured learning method proposed in Section 2 to random forests (RF) as proposed in [1, 12], support vector machines (SVM), and *overlap*. RF and SVM learn to score each assignment, based on positive and negative examples provided by $\mathbf{z}'$ (see Section 2.1). As a baseline, *overlap* uses the number of overlapping pixels of an assignment across sections as score. Since these methods need a prior for the selection of assignments, we trained RF and SVM on a subset of the training data (5 sections for DROSOPHILA, 40 sections for MOUSE CORTEX) and used the rest to validate a prior for RF, SVM, and *overlap* with a grid-search on the Hamming distance to $\mathbf{z}'$.

To study the performance of the structured learning method, we compare our loss SL-TED (see Section 2.2) against a baseline, SL-Ham, which uses the Hamming distance of $\mathbf{z}$ to $\mathbf{z}'$ for $\Delta(\mathbf{z}', \mathbf{z})$. For the computation of SL-TED, we evaluated the TED allowing boundary shifts up to $\theta = 100nm$, with weights $\alpha = 1$ and $\beta = 2$ to account for the fact that merges lose geometric information and thus usually take more time to repair than splits.

Results are shown in Table 1. We report errors for several commonly used measures for neuron reconstruction: Rand Index (RI), Variation of Information (VOI), Anisotropic Edit Distance [1] (AED, note that we refer to the inter FP/FN as FS/FM), and TED. The TED counts topological errors that are not considered boundary shifts as false splits (FS) and false merges (FM). Splits of the ground truth background label are false positives (FP) and merges involving the reconstruction background label false negatives (FN). For the time-to-fix (TTF) estimate, we again set the time needed for fixing a split to $\alpha = 1$ and for fixing a merge to $\beta = 2$. The structured learning methods are in general superior to *overlap*, RF, and SVM, with the best results being obtained by training on SL-TED. Training on the TED-approximation SL-TED does indeed minimize the TTF. Furthermore, RI, VOI, and AED are minimized. Our results also reveal interesting differences between error measures: Although the best solutions in terms of TED have also best RI, VOI, and AED, we see a discrepancy in the mid-field: on DROSOPHILA, SVM scores much better than RF in terms of VOI and slightly better in terms of RI. However, TED on a clearly defined criterion shows that the numbers are misleading and in fact RF has less errors in total and shorter TTF.

Learning from Skeletons. We show on MOUSE CORTEX that our method to find a training sample $\mathbf{z}'$ allows us to train on skeleton annotations as well. Skeleton annotations are not volumetric, *i.e.*, instead of labeling every pixel, only the cen-

method	Rand	VOI			AED					TED				
		split	merge	total	FP	FN	FS	FM	total	FP	FN	FS	FM	TTF
overlap	0.9939	0.668	0.192	0.860	1,553	2,404	3,114	1,666	8,737	155	179	678	57	1,305
RF	0.9936	0.375	0.291	0.666	1,048	2,546	3,014	1,451	8,059	23	151	273	68	734
SVM	0.9572	0.507	1.434	1.940	2,998	3,761	5,155	4,587	16,501	4	147	129	167	761
SL-Ham	0.9933	0.348	0.309	0.657	895	2,258	2,735	1,333	7,221	23	138	243	82	706
SL-TED	**0.9948**	0.331	0.275	**0.606**	838	2,297	2,752	1,268	**7,155**	18	135	229	82	**681**

method	Rand	VOI			AED					TED				
		split	merge	total	FP	FN	FS	FM	total	FP	FN	FS	FM	TTF
overlap	0.9906	0.309	0.340	0.648	179	517	648	254	1,598	13	58	201	99	528
RF	0.9864	0.934	0.518	1.452	181	585	556	252	1,574	1	175	108	35	529
SVM	0.9890	0.804	0.230	1.034	366	357	593	537	1,853	10	86	224	84	574
SL-Ham	0.9959	0.309	0.080	0.389	241	234	375	250	1,100	14	63	227	47	461
SL-TED	**0.9960**	0.299	0.087	**0.386**	224	249	382	239	**1,094**	15	63	215	50	**456**

Table 1: Comparison of reconstruction results of different learning methods on two anisotropic EM datasets.

MOUSE CORTEX DATASET

method	TED				
	FP	FN	FS	FM	TTF
volumetric ground truth	18	135	229	82	**681**
skeleton ground truth	17	114	188	152	737

Table 2: Reconstruction results on MOUSE CORTEX after training on different ground truth types: *volumetric* uses the original ground truth, *skeleton* a skeletonized version. We show false splits and false merges (FS and FM), false positives and false negatives (FP and FN), and an estimated time-to-fix (TTF), as reported by the TED measure.

terline of the neuron is provided as training data. In practice, this saves a lot of manual labeling effort such that larger volumes can be annotated. To simulate skeleton annotations and compare them to the learning outcome of complete ground truth, we skeletonized each ground truth label of the training stack. For that, we shrunk each 2D connected component of one label in each EM section to a single pixel at its center of mass. Consequently, we adjusted the search for the training sample $\mathbf{z}'$ to not consider the set difference term in (4). The results of training with SL-TED on the $\mathbf{z}'$ obtained this way are shown in Table 2. There is almost no loss in accuracy compared to training from volumetric annotations.

Runtimes. The bottleneck of our method is the computation of the coefficients l_i needed for the TED approximations SL-TED, since for every binary variable in the $\mathbf{z}'$ the TED has to be evaluated. For MOUSE CORTEX and DROSOPHILA, $\mathbf{z}'$

contained 277,874 and 20,890 variables, respectively. Computing the coefficients took 64.3h for MOUSE CORTEX and 4.8h for DROSOPHILA on a 12 core Intel Xeon CPU with 3.47 GHz. By noting that the influence of a single variable flip is usually local, the computation of the TED could be limited to constant size subvolumes around the variable of interest, such that the effort of computing the coefficients scales linearly with the best-effort size. Structured learning with SL-TED took 30m for DROSOPHILA and 1h45m for MOUSE CORTEX on 10 cores of a Intel Xeon CPU with 2.6 GHz[1].

4. DISCUSSION

We could show that structured learning improves the accuracy of assignment models for anisotropic neuron reconstruction. We believe that the key to this improvements is the consideration of topological errors during training. Previous attempts tried to minimize the classification error to a best-effort solution and did not take into account the severity of the deviation from the best-effort in terms of split and merge errors in the result. Training on a TED approximation overcomes this problem.

We used a simple approximation of the TED measure, which tolerates boundary shifts of up to $100nm$. It is worth noting that this is just an example of how to use a measure like TED for training. Depending on the biological question, more or less deviations from the ground truth can be permitted. For example, boundary shifts could be tolerated to an extent that locally depends on the diameter of the ground truth neuron.

---

[1]Using our own implementation of a bundle-method solver, available at http://github.com/funkey/sbmrm.

5. REFERENCES

[1] Jan Funke, Bjoern Andres, Fred A. Hamprecht, Albert Cardona, and Matthew Cook, "Efficient Automatic 3D-Reconstruction of Branching Neurons from EM Data," in *CVPR*, 2012, pp. 1004–1011. 1, 3

[2] Verena Kaynig, Amelio Vazquez-Reina, Seymour Knowles-Barley, Mike Roberts, Thouis R. Jones, Narayanan Kasthuri, Eric Miller, Jeff Lichtman, and Hanspeter Pfister, "Large-scale automatic reconstruction of neuronal processes from electron microscopy images," *IEEE Transactions on Medical Imaging*, vol. 22, no. 1, pp. 77–88, 2015. 1

[3] Hao Jiang Hao Jiang, S. Fels, and J.J. Little, "A Linear Programming Approach for Multiple Object Tracking," in *CVPR*, 2007. 1

[4] Florian Jug, Tobias Pietzsch, Dagmar Kainmüller, Jan Funke, Matthias Kaiser, Erik van Nimwegen, Carsten Rother, and Eugene W. Myers, "Optimal Joint Segmentation and Tracking of Escherichia Coli in the Mother Machine," in *BAMBI*, 2014. 1

[5] Amelio Vazquez-Reina, Daniel Huang, Michael Gelbart, Jeff W. Lichtman, Eric Miller, and Hanspeter Pfister, "Segmentation Fusion for Connectomics," in *ICCV*, 2011. 1

[6] Bernhard X. Kausler, Martin Schiegg, Bjoern Andres, Martin Lindner, Ullrich Köthe, Heike Leitte, Jochen Wittbrodt, Lars Hufnagel, and Fred A. Hamprecht, "A discrete chain graph model for 3d+t cell tracking with high misdetection robustness," in *ECCV*, 2012. 1

[7] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, Yasemin Altun, and Yoram Singer, "Large margin methods for structured and interdependent output variables," *Journal of Machine Learning Research*, vol. 6, pp. 1453–1484, 2005. 2

[8] Jan Funke, Francesc Moreno-Noguer, Albert Cardona, and Matthew Cook, "TED: A Tolerant Edit Distance for Segmentation Evaluation," *arXiv:1503.02291*, 2016. 2

[9] Stephan Gerhard, Jan Funke, Julien Martel, Albert Cardona, and Richard D. Fetter, "Segmented anisotropic ssTEM dataset of neural tissue," 2013. 3

[10] Ignacio Arganda-Carreras, Sebastian H. Seung, Ashwin Vishwanathan, and Daniel R Berger, "SNEMI 3D: 3D Segmentation of Neurites in EM Images," 2013. 3

[11] Dan C. Ciresan, Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber, "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images," in *NIPS*, 2012, vol. 25, pp. 2843–2851. 3

[12] Jan Funke, Julien Martel, Stephan Gerhard, Bjoern Andres, Dan C. Ciresan, Alessandro Giusti, Luca M. Gambardella, Jürgen Schmidhuber, Hanspeter Pfister, Albert Cardona, and Matthew Cook, "Candidate Sampling for Neuron Reconstruction from Anisotropic Electron Microscopy Volumes," in *MICCAI*, 2014, pp. 17–24. 3