LOCALIZED PERSISTENT HOMOLOGIES FOR MORE EFFECTIVE DEEP LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Persistent Homologies have been successfully used to increase the performance of deep networks trained to detect curvilinear structures and to improve the topological quality of the results. However, existing methods are very global and ignore the location of topological features. In this paper, we introduce an approach that relies on a new filtration function to account for location during network training. We demonstrate experimentally on 2D images of roads and 3D image stacks of neuronal processes that networks trained in this manner are better at recovering the topology of the curvilinear structures they extract.

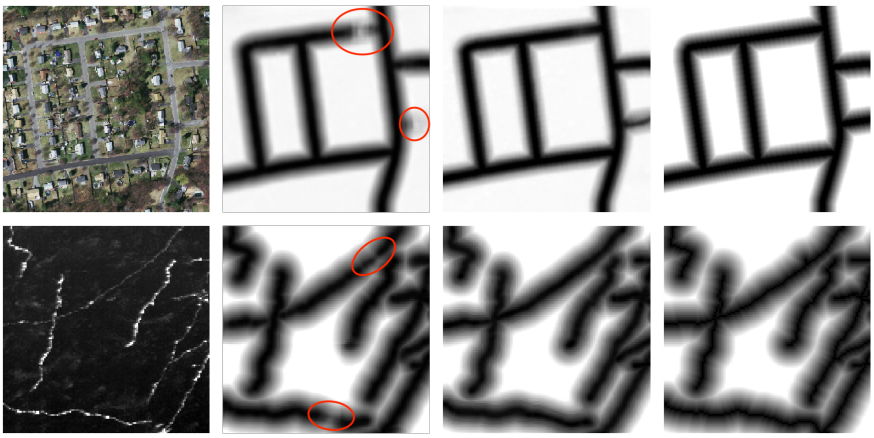
1 INTRODUCTION

In many image segmentation tasks, the topology of the resulting mask is as important as, if not more than, its pixel-wise accuracy. For example, a model of an aortic valve that does not form a ring is biologically implausible. Similarly, networks of curvilinear structures—-be they roads in aerial images, blood vessels in Computer Tomography (CT) scans, or dendrites and axons in Light Microscopy (LM) image stacks—should not feature breaks that disrupt connectivity of false connections between disjoint structures. Unfortunately, deep networks trained by minimizing pixel-wise loss functions, such as the cross-entropy or the mean square error, are subject to such mistakes. This is in part because it often takes very few mislabeled pixels or voxels to alter the topology significantly without significant impact on the overall pixel-wise accuracy of the labeling. In other words, it is possible for a network trained in this manner to deliver both a good point-wise classification accuracy and an incorrect topology.

Specialized solutions to this problem have been proposed in the form of loss functions that compare the topology of the prediction to that of the annotation. They are effective for specific applications but do not naturally generalize. For example, the perceptual loss of Mosińska et al. (2018) penalizes topological differences between the prediction and the ground truth, but cannot be guaranteed to detect them all. Similarly, minimizing the MALIS loss for segmenting electron microscopy scans (Briggman et al., 2009; Funke et al., 2018) yields better region boundaries but does not penalize interruptions in loopy linear structures. This has been addressed by Oner et al. (2021) for delineation of 2D road networks but the proposed solution is not applicable to 3D image stacks.

Persistent Homology (Aktas et al., 2019) is an elegant approach to describing and comparing topological structure of data, which is well-established in the field of topological data analysis. It offers the promise to address the connectivity problem in a generic way, both in 2D and 3D. Homologies are connected components within a structure, such as loops in 2D and closed surfaces in 3D, which appear at a particular scale and disappear at a larger one. The scale interval in which they exist is known as their *persistence* and the set of persistences for all the homologies characterizes the overall topology of the structure. It can be represented by a *persistence diagram*. The similarity of these diagrams across two different structures can then be used to quantify their topological similarity. This has been successfully exploited to train deep networks for delineation (Hu et al., 2019), image segmentation (Clough et al., 2019; Hu et al., 2019) and crowd counting (Abousamra et al., 2021).

However, existing techniques do not unleash the full power of persistent homology because the persistent diagram is a global image feature that ignores the localization of the topological features, which reduces their descriptive power. As shown in Fig. 1, this can result in networks that still fail to enforce the proper topology. This is because, when training a deep network, a persistence-based loss can be low even if the network predicts a structure that is quite different from the ground-truth. To



(a) (b) (c) (d)

Figure 1: **2D and 3D delineation.** (a) Aerial image and slice of a microscopy stack. (b) A network trained using a standard homology-based loss yields road and neurite interruptions. (c) One trained using our localized loss is more topologically accurate and produces predictions that closely resemble the ground truth (d).

remedy this, we introduce a new approach to computing persistence diagrams that takes location into account and increases their descriptive power, as shown in Fig. 3. Our main contribution is a novel filtration technique that introduces an element of localization into the persistence diagram and better distributes the ground truth persistence diagrams in two dimensions. It is completely generic and applies both in 2D images and 3D image stacks and significantly improves performance compared to state-of-the-art topological methods.

2 Related work

Training a deep network that produce topologically correct segmentations has typically been done by designing loss functions that, when minimized, favor plausible topologies. In this section, we briefly review first those that do not rely on Persistent Homology, and then those that do.

Losses designed to enforce topological correctness Several such losses have been proposed already to go beyond pixel-wise classification accuracy by encoding more global properties. In Li et al. (2020), the connectivity between neighboring pixel pairs is used as an additional source of supervision. This approach has been shown to improve connectivity, but since disconnections or false connections are not penalized explicitly, there is no guarantee it captures all such errors. The perceptual loss of Mosińska et al. (2018) is based on the assumption that a pre-trained neural network can capture differences of connectivity between the prediction and the ground-truth. However, even though it has been shown experimentally to improve the topology of masks produced by a deep net, there is no guarantee that this assumption holds in general. Making the Rand index of segmentations produced by the network similar to that of ground truth ones (Briggman et al., 2009; Funke et al., 2018) helps when modeling tree-like structures, both in 2D and in 3D, but cannot prevent disconnections in loopy structures. This shortcoming has been addressed by Oner et al. (2021) by detecting disconnections of 2D loopy structures as interconnections of background regions, but the proposed solution does not generalize to 3D. It has also been proposed to detect disconnections in predicted 2D and 3D structures using the Discrete Morse Theory (Hu et al., 2021). Topological features that are inconsistent with the ground truth are then penalized in the loss function. However, when the annotations lack spatial precision, which is often the case for neurite and road centerline annotation like the ones studied here, ground-truth inaccuracies may confuse the network. By contrast, our technique allows for considerable misalignment between the prediction and the ground truth.

Losses that Rely on Persistent Homologies Persistent Homology is an established topological data descriptor. But, until recently, its applications in computer vision were scarce. For example, it has been shown that it can be used to enforce the correct Betti number on binary masks resulting from inference in Markov Random Fields Chen et al. (2011). However, this has changed recently with the demonstration that persistence diagrams can be differentiated with respect to the input image Clough et al. (2019); Hu et al. (2019). Hence, they can be incorporated in a loss func-

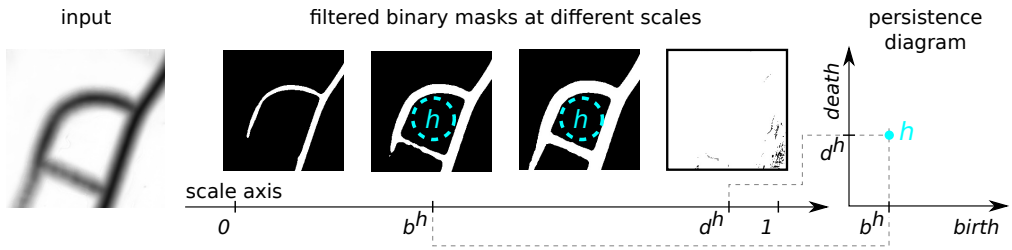


Figure 2: **Filtration.** When the distance map shown on the left is filtered by thresholding, the loop h emerges at scale b^h and is filled at scale d^h. This gives rise to the point (b^h, d^h) in the persistence diagram shown on the right. Here, thresholding means retaining all pixels whose value is lower than the threshold.

tion that penalize topological differences between the predicted and ground truth segmentations of medical images. Since then, PH has been used to train deep networks to produce topologically accurate crowd density maps (Abousamra et al., 2021) and to generate topologically credible synthetic data (Wang et al., 2020). This work also improved upon the original technique by applying it to predicted and ground truth distance maps instead of binary annotations and class affinity maps. As discussed in more detail in section 3, this makes the loss function more effective at detecting and penalizing topological errors. Unfortunately, even the improved technique remains susceptible to errors provoked by incorrectly matching the persistence diagrams of the prediction and the ground truth. By making the diagrams depend on location of the topological features in the image, our method makes them more diverse and minimizes potential for erroneous matches.

3 METHOD

We first introduce Persistent Homologies (PHs) and their application to characterizing two-dimensional images and three-dimensional image stacks. As PHs provide global descriptors that ignore location of homologies, we then introduce our approach to accounting for it.

3.1 PERSISTENT HOMOLOGIES

In the interest of simplicity, we introduce PHs for binary images and image stacks, where homologies are limited to connected components, loops, and closed 3D surfaces. We refer the interested reader to the review by Aktas et al. (2019) for a more general treatment, applicable to non-image and higher-dimensional data.

At the heart of the approach is detecting homologies at many different scales. The ones that exist over a wide scale range are called *persistent* and deemed more likely to represent true features, as opposed to artifacts either of sampling, noise or specific parameter choices. Here *scale* has a very specific meaning. It refers to the parameter of a *filtration function* f that is applied to an image $\mathbf{X}$ to produce a binary mask. In the topology literature, this binary mask is called a *cubical complex* (Kaczynski et al., 2004) to emphasize its equivalence to *simplicial complexes* resulting from filtering point cloud data. Binary masks obtained for different scales form a sequence, such that, for a pair of scale parameters $s_1 < s_2$, the mask $f(\mathbf{X}, s_1)$ is entirely contained within the mask $f(\mathbf{X}, s_2)$. The simplest example of a function for filtering grayscale images is thresholding, where the threshold acts as the scale, as shown in Fig. 2.

As the scale changes, homologies in the filtered binary mask emerge and disappear. To capture this, the scale range is sampled from small to large, the image is filtered at the selected scale values, and homologies in the resulting binary masks are detected algebraically, using matrix reduction (Tauzin et al., 2021), and correspondence is established between homologies found at consecutive scales. For each homology, this yields a pair (b, d), where b is the scale at which the homology appears and d the scale at which it disappears. We will refer to them as *birth* and *death* times and to the interval $[b, d]$ as the *persistence interval* of the homology. The set $P_{\mathbf{X}} = \{(b^h, d^h)\}_{h \in H_{\mathbf{X}}}$, where $H_{\mathbf{X}}$ is the set of all homologies found in image $\mathbf{X}$ at all scales, is called the *persistence diagram* of $\mathbf{X}$. Fig. 2 depicts the birth and death of a specific homology.

To compare images $\mathbf{X}_1$ and $\mathbf{X}_2$, one-to-one matching is performed between their persistence diagrams, $P_{\mathbf{X}_1}$ and $P_{\mathbf{X}_2}$, with the cost of matching a homology $g \in H_{\mathbf{X}_1}$ to a homology $h \in H_{\mathbf{X}_2}$ set

to $c_{g,h} = (b^h - b^g)^2 + (d^h - d^g)^2$ and the cost of leaving a homology unmatched set to a constant hyper-parameter. The optimal matching can be found using the Hungarian algorithm. Its cost that we denote as $C(\mathbf{X}_1, \mathbf{X}_2)$ quantifies the topological discrepancy between $\mathbf{X}_1$ and $\mathbf{X}_2$ by penalizing differences between corresponding homologies and ones that only appear in either $\mathbf{X}_1$ or $\mathbf{X}_2$.

3.2 Training Deep Networks using PH

Let f be a network that associates to an image $\mathbf{X}$ a segmentation mask $\mathbf{Y} = f(\mathbf{X})$ such that for all pixels or voxels $p \in \mathbf{Y}$, $0 \leq \mathbf{Y}[p] \leq 1$ and let $\hat{\mathbf{Y}}$ be the corresponding ground-truth mask. A natural idea then is to train f by minimizing

$$L_{\text{tot}}(\mathbf{Y}, \hat{\mathbf{Y}}) = L(\mathbf{Y}, \hat{\mathbf{Y}}) + \alpha C(\mathbf{Y}, \hat{\mathbf{Y}}) \ , \tag{1}$$

where L is the standard loss function, either the Mean Square Error, or the Cross Entropy, and α is a hyper-parameter, which we set to 0.01 in practice. This is possible because C is sub-differentiable with respect to its inputs when filtration is achieved by thresholding, as shown before (Hu et al., 2019; Clough et al., 2019). However, when the ground truth $\hat{\mathbf{Y}}$ is binary, as it often is, all structures emerge at scale zero and disappear at scale one. Hence, as shown in Fig. 3(a) the persistence intervals all are $[0, 1]$ and filtering it by thresholding is uninformative. An approach to handling this difficulty is to replace the binary ground truth by its distance transform that can be thresholded over a wide range of threshold values to create different binary masks Wang et al. (2020). Unfortunately, computing the persistence diagram of a distance transform still yields persistence diagrams that are spread along the 'death' axis but not along the 'birth' one: The distance value at the structures themselves is zero and, as a result, all the loops appear as soon at the scale value becomes positive. As shown in Fig. 3(b), this may lead to erroneous matches between the persistence diagrams, which in consequence encourages the deep network to produce wrong segmentations. Moreover, this approach ignores the location of homologies within the image. This is sub-optimal, because the predicted homologies should not be too far from the ground truth ones.

3.3 Localized Filtration

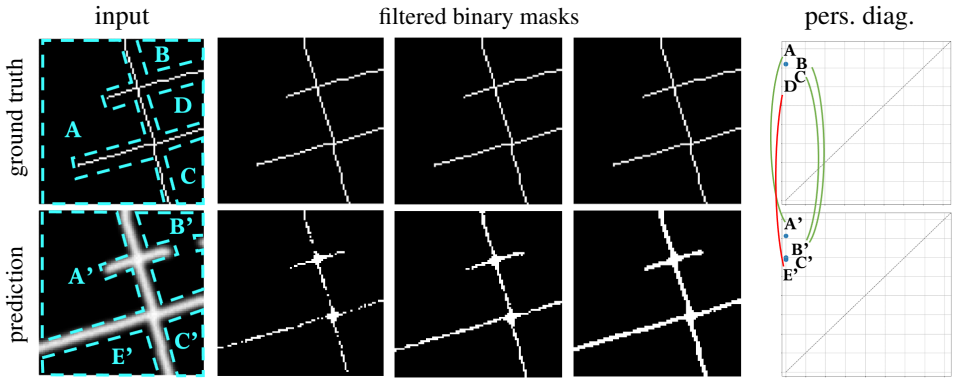
To remedy the above-mentioned drawbacks of traditional PHs, our goal is therefore to spread the persistence diagrams along *both* dimensions while also accounting for where in the image the homologies are. To this end, we draw our inspiration from another filtration technique called the height function (Turner et al., 2014; Garin & Tauzin, 2019). It was originally designed for three-dimensional meshes and can be applied to binary images by assigning to each pixel a *height* value that is the coordinate of its projection along a selected straight line. Filtration is carried out by forming binary masks made of pixels whose height is smaller than the scale parameter. As the scale is increased, the binary image is revealed in scan-lines perpendicular to the height axis, one scan-line at a time. The birth and death times are the heights of pixels responsible for the emergence and disappearing of homologies. As a result, the persistence diagram contains partial information about the location of homologies. Moreover, both birth and death times of different homologies are distributed across scales.

Unfortunately, in this simple form, height filtration can only be applied to binary images and is therefore unsuitable for training deep networks. To achieve a similar effect in a way that can be used to do so, we introduce a novel filtration method that unifies the thresholding- and height-function-based techniques. Given a scale s, the value of the filtered binary mask at coordinates $\mathbf{p}$ is taken to be

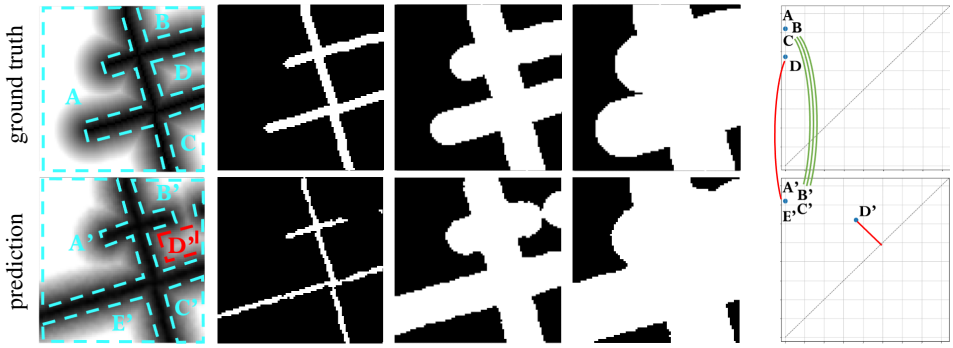
$$F(\mathbf{Y}, s)[\mathbf{p}] = \mathbb{1}(\mathbf{Y}[\mathbf{p}] + g(\mathbf{p}) < s) \ , \tag{2}$$

where $\mathbb{1}(\cdot)$ evaluates to one if the condition in the bracket is satisfied and to zero otherwise. In essence, this amounts to thresholding the sum of the height function g and the pixel values. In its simplest form, g is a linear function of pixel coordinates, and the region highlighted for any s extends along a line perpendicular to the height axis, as shown in Fig. 3(c). But other forms of g are also possible. We tested

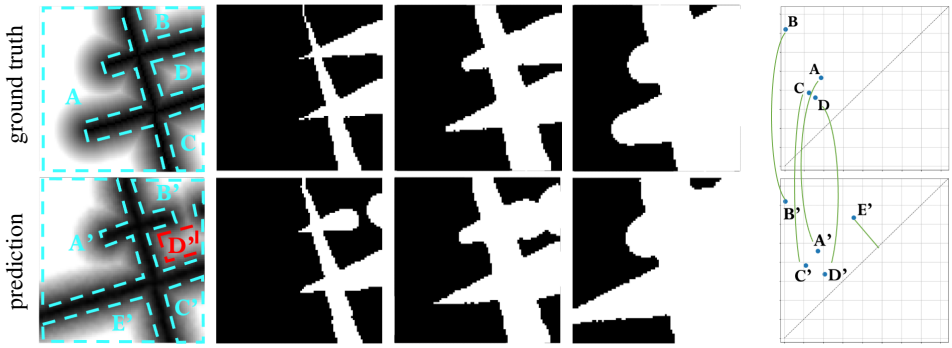
- linear functions $g(\mathbf{p}) = \mathbf{w}^\intercal \mathbf{p}$, where $\mathbf{w}$ is a two-vector hyper-parameter encoding the orientation of the height axis and the slope of the height function;
- a scaled distance to a point $\mathbf{q}$ in the image, $g(\mathbf{p}) = a\|\mathbf{p} - \mathbf{q}\|_2$, where $\mathbf{q}$ and a are hyper-parameters;



(a) **Filtration by thresholding** binary ground truth and predicted class affinity maps. Here, filtration involves decreasing the threshold from 1 to 0, and retaining the pixels greater than the threshold. Note, that the the binary masks resulting from filtering the ground truth at different scales are all the same and that all points in the ground truth persistence diagram (*top-right*) coincide. This results in erroneous matches between the predicted and ground truth homologies. Minimizing a loss function based on such a filtration can magnify the errors.



(b) **Filtration by thresholding distance maps** distributes the persistence diagram of the ground truth along the vertical but not the horizontal axis. This still results in erroneous matching between the predicted and ground truth homologies: Loop D' in the prediction emerges when the threshold is high enough to make the road brake disappear. Hence, it remains unmatched and the E' loop created by the false positive road is matched to the ground truth loop D.



(c) **Our localized filtration of distance maps** distributes the persistence diagram of the ground truth across the plane, promoting correct matches between predicted and ground truth homologies.

Figure 3: **Comparing filtration functions on synthetic data.** The binary ground truth road annotation (*top-left* in each table part) contains four loops, marked with cyan dashed lines. We synthesized a predicted class affinity map (*bottom-left* in each part) by extending one road to the left and interrupting another. In consequence, loop B and D from the ground truth are joined into B' in the prediction, and A is split into A' and E'. For each filtration method, we show binary masks resulting from filtration at different scales, pairs of persistence diagrams, and their optimal matches.

- the square of the height function $g(\mathbf{p}) = \mathbf{p}^\intercal \mathbf{W} \mathbf{p}$, where $\mathbf{W} = \mathbf{w}^\intercal \mathbf{w}$, and $\mathbf{w}$ is the hyper parameter encoding the slope of the function and the orientation of the height axis;

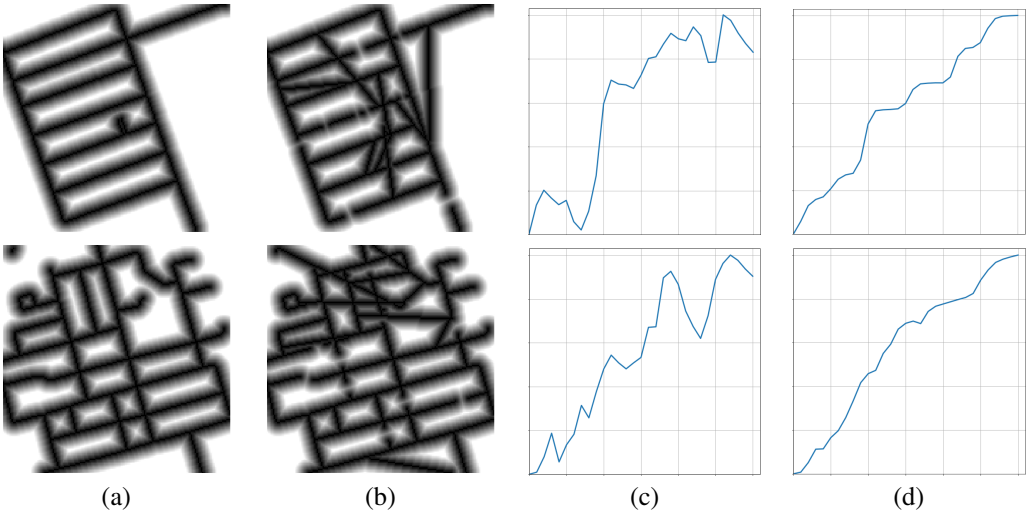


Figure 4: **Sensitivity of the topological discrepancy C to noise** (a) Ground truth distance maps. (b) Distance maps corrupted by introducing additional branches and gaps in existing branches. (c,d) Average value of C over 5 trials as a function of noise. In (c), C is evaluated using the filtration by thresholding distance maps, whereas in (d) we use our localized filtration. Note that in the first case, the curve is chaotic whereas in the second it is much closer to being a monotonic function of the error rate.

The height function g introduces localization into the persistence homology framework. This is illustrated by Fig. 3 where different values of the scale parameter make homologies appear in different parts of the image. But, because the scale parameter must be a scalar, it can only localize homologies in 2D or 3D images along one direction. This could be addressed by evaluating the loss function many times for many different orientations of the height axis, or more generally, for many different hyper-parameters of g. Unfortunately, this would considerably slow down training. Instead, we randomly draw the hyper-parameters of the height function at each training iteration. We show in the supplementary material that, in practice, the simple linear function performs best.

3.4 Validation on Synthetic Data

We have motivated our new filtration technique by the fact that it introduces partial localization of homologies into the persistence diagrams and better spreads the diagrams across the plane. For validation purposes, we use synthetic data to show the resulting loss correlates better with the number of errors injected into ground truth distance maps than the loss based on filtration by simple thresholding.

To that end, we took two crops of ground truth road graphs of the *RTracer* dataset (Bastani et al., 2018) and generated many faulty synthetic distance maps by injecting one error at a time, randomly selected between a road disconnection and a false interconnection with equal probability. We then evaluated the topological discrepancy C of Eq. 1 using either filtration by thresholding distance maps, or our localized filtration. We plotted the results in Fig. 4. When using the standard approach, the average value of C as a function of the noise rate is non-monotonic, meaning that very noisy images can be seen as being very similar to error-free ones. By contrast, when using our approach, C increases almost linearly with the error rate.

4 Experiments

We now describe the dataset we have tested our approach on, the baselines to which we compare our results, and the metrics we used to assess the topological correctness of the segmentations. We then demonstrate that our new loss improves the topological correctness of segmentation masks. We provide additional qualitative results and an ablation study in the supplementary material.

4.1 DATASETS

We experimented on three datasets.

- *RTracer*. A recently published dataset of high-resolution satellite images covering urban areas of forty cities in six different countries (Bastani et al., 2018). The ground truth was obtained from OpenStreetMap. Like Bastani et al. (2018); Li et al. (2019); Yang et al. (2019); Mosińska et al. (2020), we used twenty five cities as the training set and the remaining fifteen as the test set.
- *Massachusetts*. The Massachusetts dataset (Mnih, 2013) features both urban and rural neighborhoods, with many different kinds of roads ranging from small paths to highways. For a fair comparison to (Hu et al., 2019), we split the data into three equal folds and performed a three-way cross validation.
- *Neurons*. The dataset is a part of a proprietary 3D, 2-photon microscopy scan of a whole mouse brain. It contains 14 stacks of size $250 \times 250 \times 200$ voxels and a spatial resolution of $1.0 \times 0.3 \times 0.3\ \mu m$. We used ten stacks for training and the remaining four for testing.

4.2 METHODS TESTED

To test the impact of our proposed filtration function, we used the standard U-Net architecture (Ronneberger et al., 2015), with four blocks, each with two sequences of convolution-ReLU-batch normalization. Max-pooling in 2×2 windows followed each of the blocks. The initial feature size was set to 32 and grew to 512 in the smallest feature map in the network. We augmented the training data with vertical and horizontal flips and random rotations and used the ADAM algorithm (Kingma & Ba, 2015) with the learning rate set to $1e-4$. We then used different version of the L_{tot} of Eq. 1 we minimized to train the network. We tested the following as baselines:

- *UNet-CE*. L is the cross Entropy loss for pixel classification and there is no topological discrepancy loss, that is, $alpha = 0.0$.
- *UNet-MSE*. L is the mean squared error of the truncated distance to the closest foreground pixel, with no topological discrepancy loss.
- *Homo-Pre*. L is the cross Entropy loss and we compute C by thresholding pixel classification maps., as in Hu et al. (2019); Clough et al. (2019).
- *Homo-Reg*. L is the mean squared error and we compute C by thresholding the truncated distance maps, as in Wang et al. (2020).
- *Homo-Ours*. L is the mean squared error and we compute C using our proposed filtration function.

In the last three cases, we set α to 0.01 for all our experiments except the ablation study below. Like Hu et al. (2019), we compute the loss in windows sized 64×64 pixels, and limit the method to homologies order one, that is, loops. This has two advantages. First, by convention, loops are created by the borders of the window, making disconnections in dead-ending roads or neurites detected as broken loops. Second, detection of homologies is computationally expensive, and the time grows cubically with the number of pixels. In our current setup, computing the loss for a single window takes 0.5 seconds. Similarly to Hu et al. (2019), we did not observe any gain in performance from computing the loss for homologies order zero—connected components—in addition to loops.

For completeness, we also compared our approach to recent techniques *not* relying on persistent homologies *Segmentation* (Bastani et al., 2018), *RoadTracer* (Bastani et al., 2018), *Seg-Path* (Mosińska et al., 2020), *RCNNU-Net* (Yang et al., 2019), *DeepRoad* (Máttyus et al., 2017), *PolyMapper* (Li et al., 2019), *DMT* (Hu et al., 2021), and *ConnLoss* (Oner et al., 2021). *Segmentation*, *RoadTracer*, *RCNNU-Net*, and *PolyMapper* do no explicitly enforce topology constraints, while the others do and are discussed in the related work section.

4.3 PERFORMANCE METRICS

Comparing connectivity of segmentation masks is difficult, because the reconstructions rarely overlap with the ground truth, and often deviate from it significantly. There seems to be no consensus

concerning the single best evaluation technique; we found five different connectivity-oriented metrics in concurrently published recent work. To provide an exhaustive evaluation, we computed all of them in our experiments.

- *APLS* for Average Path Length Similarity. It is defined as an aggregation of relative length difference of shortest paths between pairs of corresponding points in the ground truth and predicted maps Etten et al. (2018).
- *TLTS*. It is a statistics of lengths of shortest paths between corresponding pairs of end points randomly selected in the predicted and ground-truth networks Wegner et al. (2013). We report the fraction of paths for which the relative length difference is within 5%.
- *JCT*. It is a junction score that considers the number of roads intersecting at each junction Bastani et al. (2018). It consists of road recall, averaged over the intersections of the ground-truth and road precision, averaged over the intersections of the prediction. We report the corresponding F1 score.
- *Betti*. The Betti error is a function of the number of topological structures seen in the ground-truth and prediction persistence diagrams. We take random patches from predictions and compare them to corresponding ground-truth patches. The mean absolute error between number of topological structures, is reported.
- *CCQ* To complement the connectivity-oriented evaluation, we also computed the most popular metric that measures spatial co-occurrence of annotated and predicted road pixels, rather than connectivity. The Correctness, Completeness and Quality are equivalent to precision, recall and intersection-over-union, where the definition of a true positive has been relaxed from spatial coincidence of prediction and annotation to co-occurrence within a distance of 5 pixels Wiedemann et al. (1998). We report the Quality as our single-number metric.

Table 1: Comparative results on the *Massachusetts* dataset. Our loss function outperforms all PH-based loss functions. The results for our method are means and standard deviations over three independent training runs.

	Connectivity-oriented				pixel-based
Method	*APLS*	*TLTS*	*JCT*	*Betti*	*CCQ*
UNet-CE	60.9	41.6	72.0	3.12	66.9
UNet-MSE	61.3	41.9	71.9	3.09	67.3
DMT	64.7	45.8	80.6	**0.99**	74.9
ConnLoss	**73.4**	**53.2**	**81.4**	1.29	**75.8**
Homo-Pre	62.5	42.1	74.2	1.28	69.3
Homo-Reg	65.0	45.6	76.9	1.09	71.8
Homo-Ours	**68.7** ± 1.2	**50.6** ± 2.3	**79.2** ± 2.6	**0.90** ± 0.3	**74.9** ± 1.8

4.4 Comparative Results

As shown in Tabs 1 and 2, on the *Massachusetts* and *RTracer* data sets, our method outperforms the other methods based on Persistent Homology, which demonstrates that our approach to filtering is truly effective. It also outperforms the other 2D tracing algorithms targeted at handling aerial images, *RoadTracer*, *Seg-Path*, *DeepRoad*, and *PolyMapper*, at the exception of *ConnLoss* that does marginally better. This is presumably because *ConnLoss* explicitly penalizes each disconnection of the prediction, whereas a persistence diagram is a lossy topological descriptor that may fail to penalize some errors. However, *ConnLoss* does not naturally extend to 3D data, whereas our method does. On the 3D *Neurons* data set, it outperforms the competing algorithms, as evidenced by the results shown in Tab. 3. We provide qualitative results in the supplementary material.

5 Conclusion

We proposed an improved approach to using Persistent Homologies to train deep networks to delineate curvilinear structures. It outperforms current such approaches by introducing an element of

Table 2: Comparative results on the *RTracer* dataset. Our loss function outperforms all PH-based loss functions. The results for our method are means and standard deviations over three independent training runs.

	Connectivity-oriented				pixel-based
Method	*APLS*	*TLTS*	*JCT*	*Betti*	*CCQ*
UNet-CE	63.4	37.5	78.0	3.08	59.7
UNet-MSE	66.3	40.0	77.5	2.99	59.5
Segmentation	62.5	33.0	78.2	3.04	54.4
RoadTracer	59.1	40.6	81.2	2.85	47.8
Seg-Path	68.1	46.5	75.4	2.31	54.0
RCNNU-Net	48.2	18.4	75.9	3.25	62.8
DeepRoad	24.6	6.4	51.4	4.95	43.6
PolyMapper	61.3	31.5	80.0	2.90	35.7
ConnLoss	**75.4**	**49.6**	**82.6**	**1.30**	**68.4**
Homo-Pre	67.3	42.3	78.7	1.32	61.9
Homo-Reg	69.9	45.1	79.6	1.07	63.2
Homo-Ours	**73.8** ± 1.8	**47.8** ± 0.9	**81.3** ± 1.6	**0.89** ± 0.2	**66.3** ± 1.7

Table 3: Comparative results on the *Neurons* dataset. Our loss function outperforms all PH-based loss functions. The results for our method are means and standard deviations over three independent training runs.

	Connectivity-oriented			pixel-based
Method	*APLS*	*TLTS*	*Betti*	*CCQ*
UNet-CE	79.9	80.8	2.33	90.6
UNet-MSE	80.2	80.9	2.31	90.4
Homo-Pre	83.5	82.1	1.06	91.2
Homo-Reg	85.4	83.4	0.91	92.5
Homo-Ours	**86.9** ± 1.1	**85.2** ± 1.9	**0.80** ± 0.2	**93.3** ± 1.9

localization in the filtration process. Unlike other powerful approaches Oner et al. (2021) to enforcing topological constraints on the output of deep networks, it generalizes naturally to 3D, which opens the door to future research in a space that is critical for biomedical applications.

To further increase performance, we will address the fact that our proposed loss function has sparse gradients that only depend on values at pixels that are critical for emergence and disappearance of homologies. This limits robustness and our future work will focus on developing more sophisticated topological descriptors with more smooth gradients.

REFERENCES

Shahira Abousamra, Minh Hoai, Dimitris Samaras, and Chao Chen. Localization in the crowd with topological constraints. In *AAAI Conference on Artificial Intelligence*, 2021.

M. Aktas, E. Akbas, and A. El Fatmaoui. Persistence homology of networks: methods and applications. *Appl. Netw. Sci.*, 4(1):61:1–61:28, 2019. doi: 10.1007/s41109-019-0179-3.

F. Bastani, S. He, M. Alizadeh, H. Balakrishnan, S. Madden, S. Chawla, S. Abbar, and D. Dewitt. Roadtracer: Automatic Extraction of Road Networks from Aerial Images. In *Conference on Computer Vision and Pattern Recognition*, 2018.

K. Briggman, W. Denk, S. Seung, M. Helmstaedter, and S. Turaga. Maximin Affinity Learning of Image Segmentation. In *Advances in Neural Information Processing Systems*, pp. 1865–1873, 2009.

C. Chen, D. Freedman, and C. Lampert. Enforcing topological constraints in random field image segmentation. In *Conference on Computer Vision and Pattern Recognition*, pp. 2089–2096, 2011. doi: 10.1109/CVPR.2011.5995503.

J. Clough, I. Oksuz, N. Byrne, J. Schnabel, and A. King. Explicit Topological Priors for Deep-Learning Based Image Segmentation Using Persistent Homology. In *Information Processing in Medical Imaging*, 2019.

A. Van Etten, D. Lindenbaum, and T. Bacastow. Spacenet: A Remote Sensing Dataset and Challenge Series. *CoRR*, abs/1807.01232, 2018.

J. Funke, F. D. Tschopp, W. Grisaitis, A. Sheridan, C. Singh, S. Saalfeld, and S. C. Turaga. Large Scale Image Segmentation with Structured Loss Based Deep Learning for Connectome Reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1669–1680, 2018.

A. Garin and G. Tauzin. A topological "reading" lesson: Classification of MNIST using TDA. In *ICMLA*, pp. 1551–1556, 2019. doi: 10.1109/ICMLA.2019.00256.

X. Hu, F. Li, D. Samaras, and C. Chen. Topology-Preserving Deep Image Segmentation. In *Advances in Neural Information Processing Systems*, pp. 5658–5669, 2019.

X. Hu, Y. Wang, L. Fuxin, D. Samaras, and C. Chen. Topology-aware segmentation using discrete morse theory. In *International Conference on Learning Representations*, 2021.

T. Kaczynski, K. Mischaikow, and M. Mrozek. *Computational homology*. Applied mathematical sciences (Springer-Verlag New York Inc.); v. 157. Springer, New York, 2004.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimisation. In *International Conference on Learning Representations*, 2015.

X. Li, Y. Wang, L. Zhang, S. Liu, J. Mei, and Y. Li. Topology-Enhanced Urban Road Extraction via a Geographic Feature-Enhanced Network. *IEEE Trans. Geosci. Remote. Sens.*, 58(12):8819–8830, 2020.

Z. Li, J. Wegner, and A. Lucchi. Topological Map Extraction from Overhead Images. In *International Conference on Computer Vision*, 2019.

G. Máttyus, W. Luo, and R. Urtasun. Deeproadmapper: Extracting Road Topology from Aerial Images. In *International Conference on Computer Vision*, pp. 3458–3466, 2017.

V. Mnih. *Machine Learning for Aerial Image Labeling*. PhD thesis, University of Toronto, 2013.

A. Mosińska, P. Marquez-Neila, M. Kozinski, and P. Fua. Beyond the Pixel-Wise Loss for Topology-Aware Delineation. In *Conference on Computer Vision and Pattern Recognition*, pp. 3136–3145, 2018.

A. Mosińska, M. Kozinski, and P. Fua. Joint Segmentation and Path Classification of Curvilinear Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1515–1521, 2020.

D. Oner, M. Koziński, L. Citraro, N. C. Dadap, A. G. Konings, and P. Fua. Promoting Connectivity of Network-Like Structures by Enforcing Region Separation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 234–241, 2015.

G. Tauzin, U. Lupo, L. Tunstall, J. Burella Pérez, M. Caorsi, A. Medina-Mardones, A. Dassatti, and K. Hess. giotto-tda: : A topological data analysis toolkit for machine learning and data exploration. *J. Mach. Learn. Res.*, 22:39:1–39:6, 2021.

K. Turner, S. Mukherjee, and D. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3(4):310–344, 12 2014. doi: 10.1093/imaiai/iau011.

Fan Wang, Huidong Liu, Dimitris Samaras, and Chao Chen. TopoGan: A Topology-Aware Generative Adversarial Network. In *European Conference on Computer Vision*, pp. 118–136, 2020.

J.D. Wegner, J.A. Montoya-Zegarra, and K. Schindler. A Higher-Order CRF Model for Road Network Extraction. In *Conference on Computer Vision and Pattern Recognition*, pp. 1698–1705, 2013.

C. Wiedemann, C. Heipke, H. Mayer, and O. Jamet. Empirical Evaluation of Automatically Extracted Road Axes. In *Empirical Evaluation Techniques in Computer Vision*, pp. 172–187, 1998.

X. Yang, X. Li, Y. Ye, R. Y. K. Lau, X. Zhang, and X. Huang. Road Detection and Centerline Extraction via Deep Recurrent Convolutional Neural Network U-Net. *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–12, 2019.

A SUPPLEMENTARY MATERIAL

A.1 QUALITATIVE RESULTS

In this section, we provide qualitative results on our three test datasets. For each method, we display the thresholded predictions with their skeletons overlaid in red. In the case of the 3D dataset, the images we show are maximum intensity projections.

A.2 ABLATION STUDY

To investigate the impact of hyper-parameter choices on performance, we ran three ablation studies.

Weighting the PH Loss We have varied the coefficient α in equation 1, while keeping the other parameters fixed. We report the results in Tab. 4. The best results are achieved for $\alpha = 0.01$, and the performance decreases when α is set ten times higher or lower. This suggests that the standard Mean Square Loss is still important for overall performance, which is not a surprise as the gradient of our persistent-homology-based loss is sparse and concentrated at pixels critical for topological correctness.

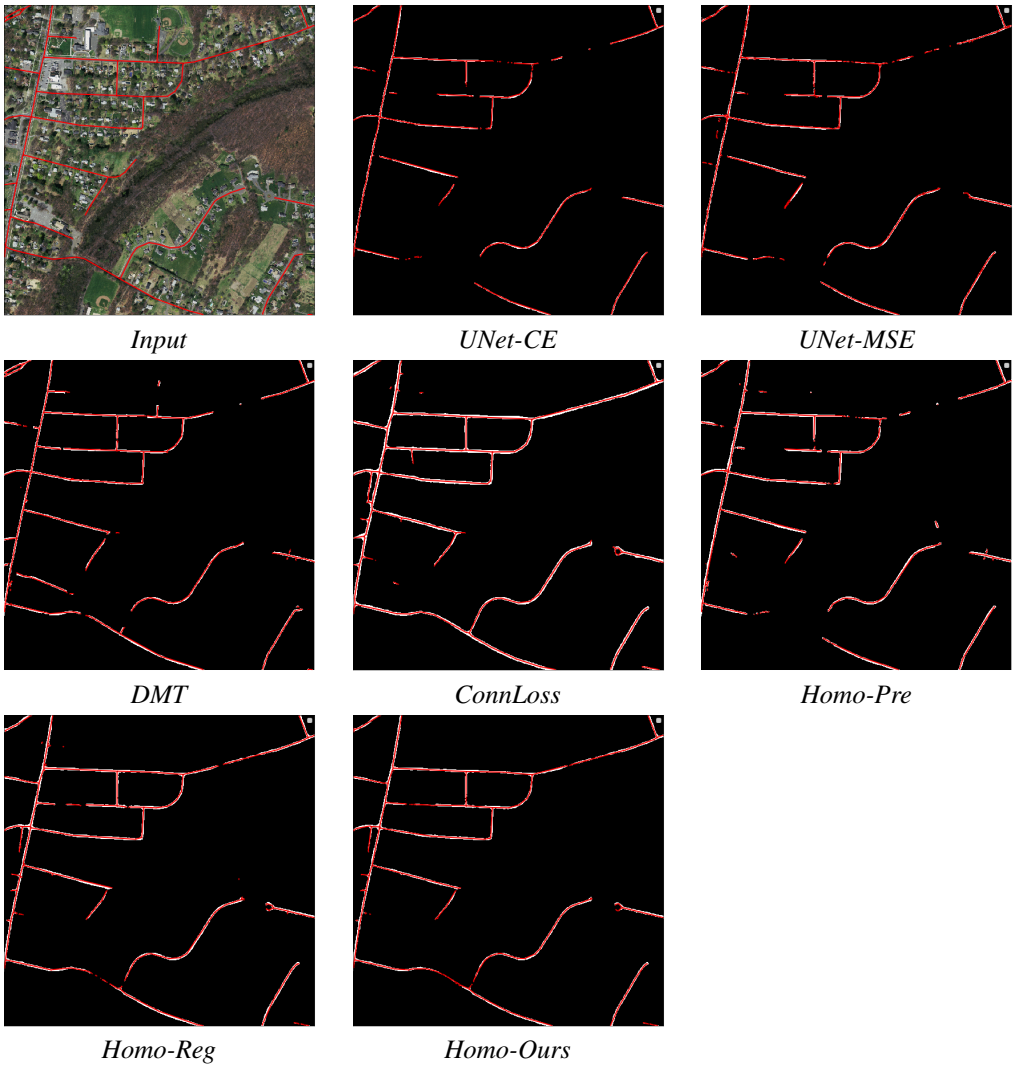


Figure 5: Comparative results on the *Massachusetts* dataset.

Window size We changed the size of the window in which the persistent homology is computed. We report the results in Tab. 5. Our method performs best when using large windows that contain significant portions of the structures of interest. We could not try even larger ones because it would have increased the time needed to detect the homologies and slowed down the training too much.

Height Functions We also evaluated the effect on performance of using different forms of function g in equation 2, that ties homology birth and death times to image coordinates, distributing the points

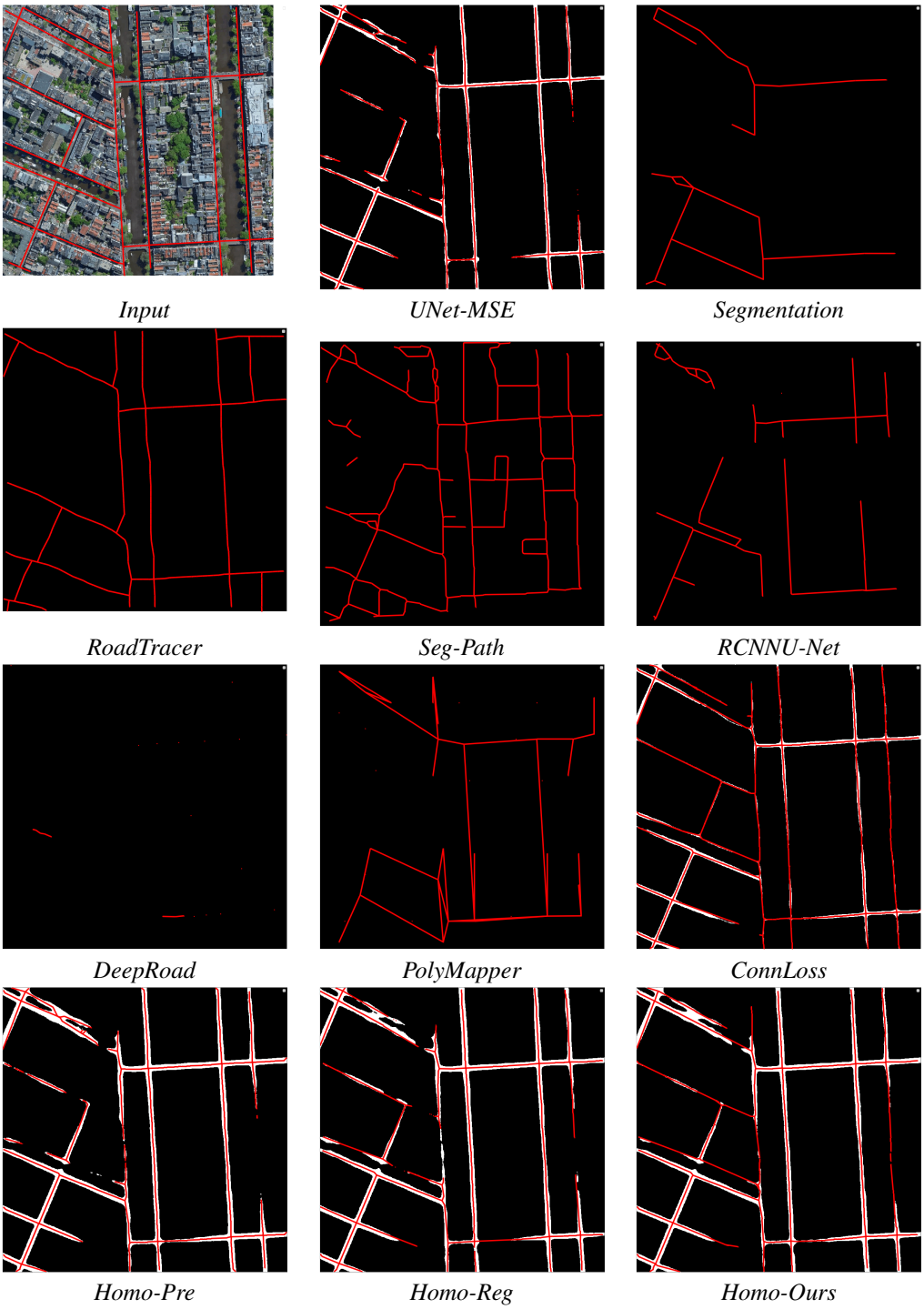


Figure 6: Comparative results on the *RTracer* dataset.

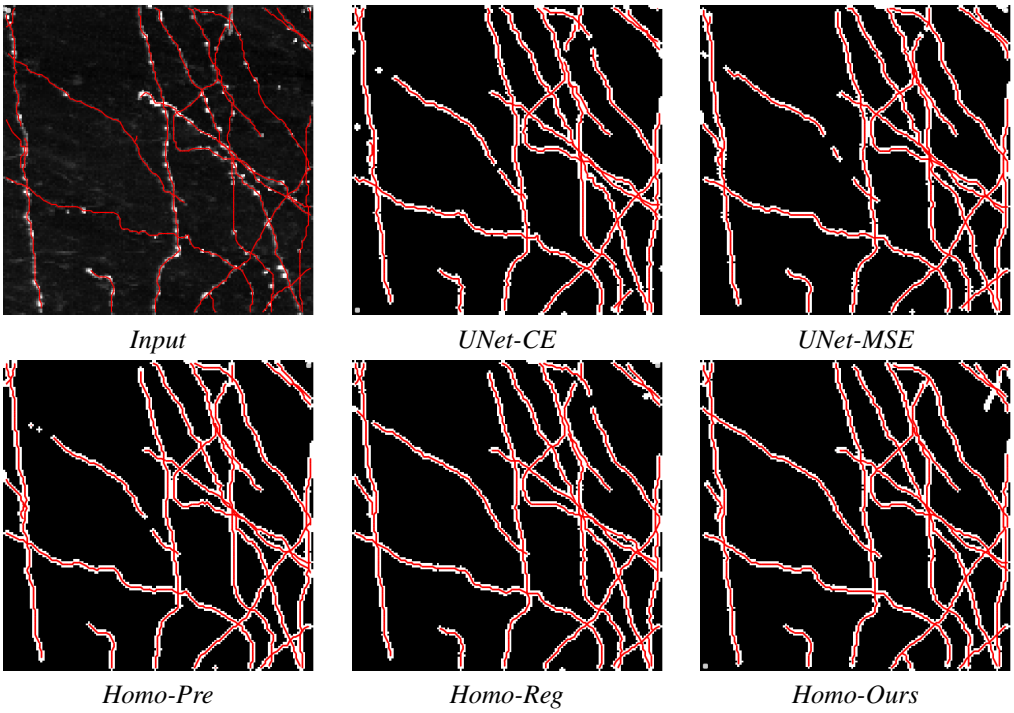


Figure 7: Comparative results on the 3D *Neurons* dataset.

Table 4: Impact of changing the learning coefficient of localized PH loss on the *Massachusetts* dataset. The window size is fixed to 64x64.

	Connectivity-oriented				pixel-based
α	*APLS*	*TLTS*	*JCT*	*Betti*	*CCQ*
1e-3	64.9	46.0	77.1	1.21	72.3
1e-2	**68.7**	**50.6**	**79.2**	**0.90**	**74.9**
1e-1	67.1	48.9	77.8	0.94	74.6
1e-0	64.8	45.8	76.2	1.10	72.0

Table 5: Impact of changing the window size when computing our localized loss on the *Massachusetts* dataset. The learning coefficient is fixed to 1e-2.

	Connectivity-oriented				pixel-based
Window Size	*APLS*	*TLTS*	*JCT*	*Betti*	*CCQ*
8x8	62.1	41.9	73.0	2.84	67.2
16x16	62.7	42.4	74.5	2.09	68.8
32x32	65.4	45.7	77.1	1.17	72.5
64x64	**68.7**	**50.6**	**79.2**	**0.90**	**74.9**

in persistence diagram. We present the results in Tab. 6. The distance to a random image point, or the use of a quadratic instead of linear function of image coordinates do not result in higher performance than the plain linear function.

Table 6: Performances of different height functions used for localized PH loss on the *Massachusetts* dataset. The learning coefficient is fixed to 1e-2 and window size to 64x64

	Connectivity-oriented				pixel-based
Height Function	*APLS*	*TLTS*	*JCT*	*Betti*	*CCQ*
Distance to a point	67.8	49.4	77.9	1.01	73.6
Random Linear	**68.7**	**50.6**	**79.2**	**0.90**	**74.9**
Fixed Linear	67.5	48.7	76.5	1.15	73.0
Square	64.2	45.1	76.3	1.32	70.3