Topology-Aware Loss Function for Aorta and Great Vessel Segmentation in Computed Tomography Images

Seher Ozcelik, Sinan Unver, Ilke Ali Gurses, Rustu Turkay, and Cigdem Gunduz-Demir

***Abstract*—Segmentation networks are not explicitly imposed to learn global invariants of an image, such as the shape of an object and the geometry between multiple objects, when they are trained with a standard loss function. On the other hand, incorporating such invariants into network training may help improve performance for various segmentation tasks when they are the intrinsic characteristics of the objects to be segmented. One example is segmentation of aorta and great vessels in computed tomography (CT) images where vessels are found in a particular geometry in the body due to the human anatomy and they mostly seem as round objects on a 2D CT image. This paper addresses this issue by introducing a new topology-aware loss function that penalizes topology dissimilarities between the ground truth and prediction through persistent homology. Different from the previously suggested segmentation network designs, which apply the threshold filtration on a likelihood function of the prediction map and the Betti numbers of the ground truth, this paper proposes to apply the Vietoris-Rips filtration to obtain persistence diagrams of both ground truth and prediction maps and calculate the dissimilarity with the Wasserstein distance between the corresponding persistence diagrams. The use of this filtration has advantage of modeling shape and geometry at the same time, which may not happen when the threshold filtration is applied. Our experiments on 4327 CT images of 24 subjects reveal that the proposed topology-aware loss function leads to better results than its counterparts, indicating the effectiveness of this use.**

***Index Terms*—Topology, persistent homology, Vietoris-Rips filtration, encoder-decoder networks, aorta and great vessel segmentation, computed tomography.**

I. Introduction

ENCODER-decoder networks have achieved state-of-the-art results for various segmentation problems on medical images. The training of these networks relies on minimizing a loss function, e.g., mean squared error and cross entropy, which typically defines the loss of each pixel separately and aggregates these pixel-wise losses. This aggregation might be unweighted, assigning the unit weight to each pixel's loss, or weighted, giving higher loss weights to hard-to-learn pixels. In the latter case, the pixels' weights can be assigned beforehand and remains the same during training, e.g., giving higher weights for pixels close to instance boundaries [1] or belonging to the minority foreground classes [2]. Alternatively, these weights can be adaptively changed during the training by modulating them based on the network performance, e.g., reducing the weights of easy-to-learn pixels for which the network gives high posteriors in a given epoch [3], [4].

These typical loss functions define the loss of each pixel only on its true and predicted values, but not considering those of other pixels, and aggregate them by weighted averaging or summing without considering the spatial relations between the predictions. Since this type of definition is of local nature, these loss functions might not sufficiently impose a network to learn the shape of an instance or the geometry between multiple instances. On the other hand, the ability of the network to learn the shape may be important for better segmenting the instances in medical images since these instances typically have an expected shape or a geometry due to their intrinsic characteristics. One example is the formation of the aortic arch and great vessels in a human body. The aorta and the large arteries and veins (also known as great vessels) are not randomly distributed over the human body. Instead, they are found in a particular geometry due to the human anatomy (Fig. 1). Besides, they mostly seem as round objects on a 2D axial image since blood vessels are tubular in 3D. This anatomic information is indeed utilized by human annotators to locate these vessels and delineate their boundaries.

In response to this issue, this paper introduces a new topology-aware loss function to train an encoder-decoder network for segmenting the aortic arch and the great vessels in computed tomography (CT) images. This loss function is defined as a weighted cross entropy, in which the weight for a training sample (and thus, for its pixels) is calculated inversely proportional to topological similarity between the maps of its ground truth and predicted vessels. This paper proposes to quantify topological features of these maps through persistent homology. To this end, it proposes to calculate the persistence diagram of each map by applying the Vietoris-Rips filtration on the point cloud of its vessel contours and enforces the network to minimize the Wasserstein distance between the corresponding persistence diagrams by defining the loss weight as a function of this distance.

This work was partly supported by the Scientific and Technological Research Council of Turkey, project no: TÜBİTAK 120E497.

S. Ozcelik is with the Computational Sciences and Engineering Program and KUIS AI Center, Koc University, 34450 Istanbul, Turkey (e-mail: sozcelik19@ku.edu.tr).

S. Unver is with the Department of Mathematics, Koc University, 34450 Istanbul, Turkey (e-mail: sunver@ku.edu.tr).

I. A. Gurses is with the Department of Anatomy, School of Medicine, Koc University, 34450 Istanbul, Turkey (e-mail: igurses@ku.edu.tr; iagurses@gmail.com).

R. Turkay is with the Department of Radiology, School of Medicine, Haseki SUAM, Medical Sciences University, 34265 Istanbul, Turkey (e-mail: rustu.turkay@sbu.gov.tr; rustuturkay@hotmail.com).

C. Gunduz-Demir is with the Department of Computer Engineering, School of Medicine, and KUIS AI Center, Koc University, 34450 Istanbul, Turkey (e-mail: cgunduz@ku.edu.tr).

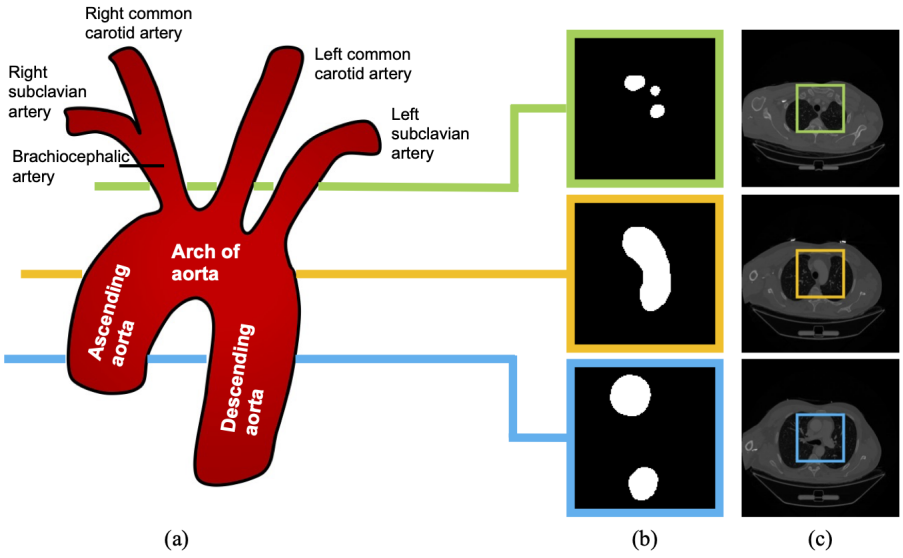


Fig. 1. (a) Anatomic formation of the aortic arch and the large arteries. (b) Manual annotations of the aortic arch and the large arteries on three exemplary axial slices. (c) CT scans for these annotations. Note that the annotations only for the green, orange, and blue squares are illustrated for better visualization. All pixels outside the rectangles are annotated as background.

The proposed approach differs from the existing studies in the construction of the topological loss function and its integration to the network. Although there exist recent studies, such as [5], [6], [7] and [8], that satisfactorily use persistent homology to train neural networks using the *topology* of the ground truth, our approach is different from these studies. We use persistent homology to learn *not only the topology but also the geometry* of the ground truth. Here, by geometry we mean the differential geometric nature of the objects in question, namely the shape of the aorta and the distribution of the great vessels with respect to each other.

We will describe the details of this contribution in more mathematical terms below in Sec. II. In summary, the difference of our contribution from the other studies stems from the type of filtration that we use in persistent homology, from the employment of the full strength of the persistent homology on the ground truth, as well as the use of the special metric to compare the persistent diagrams of the ground truth and the prediction maps. The previous studies use the persistent homology of the prediction based on the threshold filtration associated to a likelihood function predicted by the network and the Betti numbers of the ground truth. Additionally, the loss functions they use only ensure that the *topologies* of the ground truth and the prediction approached each other. On the other hand, different from this previous approach, in this paper, we propose to use the Vietoris-Rips filtration on the persistent homology of both the ground truth and the prediction maps and the Wasserstein distance between the corresponding persistence diagrams. The use of the Vietoris-Rips filtration takes into account the *geometry* of the ground truth and the prediction maps, and the loss function based on the Wasserstein distance ensures that these *two geometries* approach each other.

Besides, this is the first proposal of using a topological-aware loss function in a neural network design for the purpose of segmenting the aortic arch and great vessels, which are indeed found in a particular geometry in the human body, and thus, provides an exemplary showcase to demonstrate the usefulness of the geometry-preserving property of a neural network. Although there exist previous traditional models and network designs to segment the aorta and coronary arteries [9], [10], [11], [12], none of them use persistent homology in their models or in the definition of their loss functions.

II. Related Work and Contribution

A. Persistent Homology

For visual data of medical nature, there are prior restrictions on their shape coming from the human anatomy. At first, it might be thought of as natural to use their topological invariants as part of their features since the topological invariants would comprise a summary of the global properties of the images. On the other hand, one immediately faces the problem that topological invariants are very rigid and real life data are very noisy so that using such rigid invariants to summarize noisy data will lead to complications. A solution to this problem in the context of homology of topological spaces is to use an invariant which is more stable under small perturbations. This invariant we will use, is persistent homology, which we will briefly review in our context.

An important invariant of a topological space X is the n-th homology group $\mathrm{H}_n(X)$, for each $n \geq 0$ [13]. Here and elsewhere in this paper, we always consider homology groups with coefficients in the field of rational numbers $\mathbb{Q}$. This makes the homology groups vector spaces over $\mathbb{Q}$. The dimension $b_n(X)$ of the vector space $\mathrm{H}_n(X)$ is the n-th Betti number of X and is a measure of the number of n dimensional spheres which do not bound an $n+1$ dimensional ball. Thus, $b_0(X)$ is the number of connected components of X and $b_1(X)$ is the number of circles which do not bound a disk. Even though the

Betti invariants are extremely useful in the abstract study of topological spaces, they are too rigid to be of use in machine learning applications. More precisely, all real-world data come with noise and error. This is indeed a very typical case for medical images, in which noise and artifacts commonly exist in an image due to non-ideal conditions in image acquisition and/or technical limitations of the image scanner.

A variant of homology that is more convenient for real-world applications is the persistent homology [14], [15]. Here the input is a topological space $\mathbb{X}$, which is endowed with a filtration $\{X_t\}_{t\in\mathbb{R}}$, indexed by the real numbers $\mathbb{R}$, with $\mathbb{X} = \cup_{t\in\mathbb{R}} X_t$. The condition for being a filtration is that for every $s \leq t$, $X_s \subseteq X_t$. Taking homology of the spaces in the filtration for a fixed integer n, we obtain a persistence module $\{\mathrm{H}_n(X_t)\}_{t\in\mathbb{R}}$, which associates a $\mathbb{Q}$-vector space to each $t \in \mathbb{R}$ and a linear map between these vector spaces for each pair (t, s) with $t \leq s$, coming from the functoriality of homology. Associated to a persistence module, there is a barcode and a persistence diagram that summarizes at which filtration index the holes are born and at which filtration index they die. The flavor of the persistent homology and what it measures depends very much on which filtration one chooses to consider on $\mathbb{X}$.

B. Persistent Homology for Segmentation Networks

There exist only a few studies that employ persistent homology in the design of a segmentation network. Similar to ours, this is achieved through the definition of a topological loss function. On the other hand, the main difference between these previous studies and ours is the type of filtration, the choice of which affects the phenomena persistent homology quantifies, and hence, the phenomena that a network is enforced to learn during its training. In the context of segmentation networks, there are essentially two different ways one obtains a topological space with a filtration.

One of these filtrations, which is the one that we employ in this work, is through the use of a distance function. Here, one starts with a point cloud X in a metric space M with a distance function d. For each $t \in \mathbb{R}$, X_t is the set of $m \in M$ such that there exists an $x \in X$ with $d(m, x) \leq t$. Then $\{X_t\}_{t\in\mathbb{R}}$ gives a filtration of M. The persistence homology of this filtration encodes information about the shape of X. This filtration is called the distance filtration or the Vietoris-Rips filtration below. The other type of filtration is constructed by using a real valued function f on space. If one lets $X_t := f^{-1}((-\infty, t])$ then $\{X_t\}_{t\in\mathbb{R}}$ forms a filtration of the underlying space. In most of the works below, f is chosen to be $1 - p$ where p is a likelihood function on $\mathbb{R}^n$, which aims to predict a shape X in $\mathbb{R}^n$. More precisely, we wish p to have the property that $x \in X$ if and only if $p(x) = 1$.

1) The method of [5]: In this work, if Ω denotes the image, which is viewed as a rectangular domain, there are two filtrations obtained on Ω, which correspond to two different functions on Ω. The first one is a binary function f which assumes the value 0 on the foreground and 1 on the background. The other function is $g := 1 - p$, where p is the likelihood function predicted by the neural network. The functions f and g give two different filtrations on Ω and these result in two different persistent homology data. The topological part of the loss function used in [5] is the square of the 2-Wasserstein distance between the persistence diagrams for these filtrations for both dimensions 0 and 1. The effect of using this topological loss function, in addition to the per-pixel cross-entropy loss, is that the network will emphasize learning the 0-th and 1-st Betti numbers of the ground truth, in addition to learning the pixels.

2) The method of [6]: In this work, the authors use a training set in which they know the ground-truth segmentation for only some of the items, but they know the *topology* of the ground-truth segmentation for all of the items. The topology is known a priori without the use of the network. The method is then to train the network so that the predicted images have the *desired Betti numbers* as well as the pixelwise Dice loss function is minimized on the labeled images. These desired Betti numbers are determined by the correct prior topology. The topological loss function is then constructed in terms of these Betti numbers. The loss function is based on increasing the barcode length of the k-th largest barcode lengths if k is the desired Betti number. More precisely, denoting the birth and death coordinates of a bar by (b, d), if this bar is to be a prominent feature of the image, its contribution to the loss function is $1 - (d - b)^2$, otherwise it is $(d - b)^2$. The same authors extended this method to multi-class image segmentation, including the Betti numbers corresponding to the triplets of the objects of different classes into the prior topology [8].

The method of [16] is somewhat similar. In this paper, the authors define the filtration by using the voxel intensity function on the data. The intensity function is normalized to have values between 0 and 1. The joint loss function is defined in terms of the Dice and cross-entropy losses together with the topological loss, which is the 1-Wasserstein distance between the persistence diagram of the prediction and the persistence diagram of the expected topological space. Denoting the birth and death coordinates of a bar with (b, d), the contribution to the topological loss function is $1 - (d - b)$ if the feature is expected to be a prominent feature, and is $(d - b)$, otherwise. Using such a loss function has the effect of killing non-prominent features and emphasizing prominent ones through the learning process.

3) The method of [17]: In this work, the authors use persistent homology in two different ways to improve the 3D segmentation of objects: First, they use a topological loss function in a similar vein as those in [6] and [16]. This topological loss function is defined as the distance between the persistence diagrams of the likelihood function predicted by the network and those of the ground truth labels. The distance function between the persistence diagrams is defined using the L_∞-norm on $\mathbb{R}^2$, and first finding a matching between the diagrams that realizes the 1-Wasserstein distance between these two diagrams, and then computing the sum of the squares of the distances between the matched points with respect to the ordinary metric on $\mathbb{R}^2$. Besides, the authors integrate persistent homology with a graph convolution network to capture multi-scale structural information. To do so, they form a point cloud

in three dimensions and calculate persistence diagrams in each dimension using the distance filtration. From the persistence diagrams, the persistence image is constructed and put into a vector form, and added as a local feature map to augment the feature map obtained by the graph convolutional network. Even though persistence diagrams are defined on point clouds using the distance filtration, this second use, which involves adding the vectorization of the persistence image as a feature, differs considerably from our method, which is based on defining a loss function using the Wasserstein distance between the persistence diagrams of the Vietoris-Rips filtrations of the prediction and the ground truth. Both of these uses in [17] do not define such kind of loss function to enforce the network to learn the topology and shape of the objects as well as the geometry in between.

4) Other uses: In [18], the authors use persistent homology for a generative adversarial network to synthesize more realistic images in its generator. They map synthetic and real images into a topological feature space and define their topological dissimilarity as an additional loss term. However, different from our proposal, the filtration function is defined on the distance transform, the distance from each background pixel to the closest foreground instance, which would not reflect the shape of an instance or the geometry between multiple instances. In [19], the authors define a topological loss term for a segmentation network, but not using persistent homology. Instead, they measure the difference between the pretrained VGG19 responses of the predicted and ground truth maps and use it as an additional loss term to correct the topology of linear structures in the ground truth.

C. Persistent Homology for Other Network Tasks

Another common use of persistent homology is to design convolutional neural networks where classification is the upstream task. These networks use persistent homology as a tool to obtain better latent representations, which reflect the topological characteristics in the data. Such representations are obtained from persistence landscape [7] and using the filtration associated to the height function [20], and integrated as a topological layer of the classification network. There are works which also use the persistent homology associated to the Vietoris-Rips complex. In [21], a loss function is defined based on the death times of the barcodes for 0-dimensional persistent homology of the latent representation. The loss function measures the difference between these death times and a fixed distance η. Our work is different from [21] in several aspects. First, since we are interested in shape as well as connectivity, our loss function uses both the 0 and 1-dimensional persistent homology groups unlike [21], which only use the 0-dimensional homology. Additionally, our loss function is based on the Wasserstein distance between the persistence diagrams of both the prediction and the ground truth, and hence, the nature of the loss function changes as the ground truth varies, whereas in [21] the loss function is defined with respect to a fixed distance η as described above. Such flexibility is essential in our setting since the shape and arrangement of the great vessels and aorta change through the axial scans due to the inherent nature of the human anatomy. Likewise, in [22], connectivity properties of the latent representation in an autoencoder are improved through the use of a loss function based on the 0-dimensional persistent homology. Moreover, all these studies apply this learning process, involving persistent homology, to classification tasks rather than a segmentation task as we do in our study.

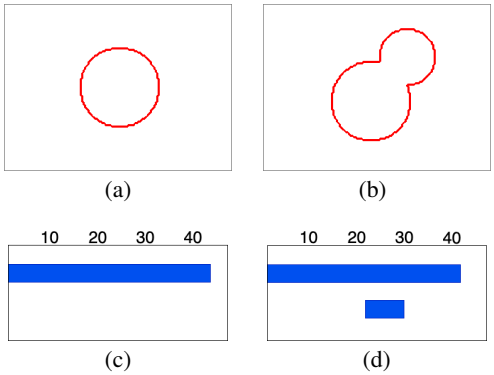


Fig. 2. (a), (b) Boundaries of two homotopy equivalent objects, and (c), (d) 1-dimensional persistent homologies of (a) and (b), respectively. Even though the standard homology groups of (a) and (b) are the same, since they are homotopy equivalent, their persistent homologies are not. The latter takes into account the *shape* of the object in this case. Namely, the smaller circular bump in the upper right part of (b) is responsible for the smaller bar in (d).

Other principal uses of persistent homology in machine learning, which will not have relevance for this work, includes regularizing the weights of a network [23], interpreting the weights of layers in a convolutional neural network [24], and extracting topological features for a classifier [25]. Nevertheless, none of these studies use persistent homology to define a loss function for a segmentation network.

III. Methodology

Our method relies on 1) quantifying the topological features of the ground truth and the predicted segmentation maps by their persistence diagrams, 2) defining a loss function using the Wasserstein distance between the persistence diagrams of the ground truth and the prediction, and 3) training an encoder-decoder network by minimizing the proposed loss function. The following subsections give the details.

A. Persistence Diagram Calculation for the Aortic Arch and Great Vessels

Suppose that we start with a point cloud X in $\mathbb{R}^n$. In our case $n = 2$, and X will be the contours of the aorta and the great vessels for either the ground truth or the prediction of the network at a given epoch. For $t \in \mathbb{R}$, if we let X_t to be the set of points in $\mathbb{R}^n$, whose distance to X is less than or equal to $\max(t, 0)$, then this gives us a filtration $\{X_t\}_{t\in\mathbb{R}}$ of $\mathbb{R}^n = \mathbb{X}$, with $X_0 = X$. The associated persistence diagrams can be thought of as more dynamic and stable versions of the ordinary homology groups of X.

The persistent homology with the Vietoris-Rips (distance) filtration above has an added, somewhat surprising, benefit. It tells us about the *geometry* of X, and not only about its topology. By geometry, we mean the shape of an object

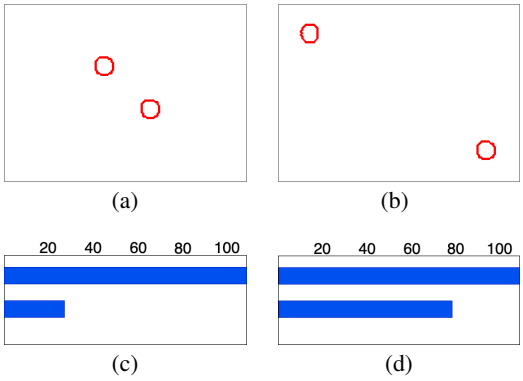


Fig. 3. (a), (b) Boundaries of two pairs of homotopy equivalent objects, and (c), (d) 0-dimensional persistent homologies of (a) and (b), represented by two bars. The long bar in each case represents the connected component which persists forever. The smaller bar in each case corresponds to the distance between the two connected components. This bar is smaller in (c) than in (d), consistent with the fact that the components are closer to each other in (a).

and also the distribution of multiple objects with respect to each other. Fig. 2 depicts the boundaries of two homotopy equivalent objects of different shapes, exhibiting the same topology but different 1-dimensional persistent homologies. Likewise, Fig. 3 sketches the boundaries of two pairs of homotopy equivalent objects with different distributions. These object pairs have the same topology, but this time, different 0-dimensional persistent homologies. Such geometry differences cannot be captured by a filtration associated to the likelihood function, as suggested by the previous segmentation networks [5], [6], [17]. On the other hand, as illustrated in these figures, the Vietoris-Rips filtration that we use in our design produces different barcodes, which allows us to model differences in the object geometries.

This idea will be very important for our segmentation model since the shape of the aorta and the distribution and the distances of the great vessels with respect to each other give us essential global invariants, which will help us improve the network using this prior geometric information. In our model, we define a loss function based on the 0-dimensional persistent homology if the ground truth includes any great vessels, which are indeed smaller in size compared to the aorta. The reason is that when we are dealing with the great vessels, the geometry of the associated point cloud is essentially determined by the distribution and the distances of the connected components in the data. The connected components in the images we consider correspond to the individual great vessels themselves. Even though the number of the connected components, hence the 0-th Betti number, in two point clouds might be the same, the corresponding barcodes associated to their 0-dimensional persistent homology might be quite different (see Fig. 3). If the ground truth includes only the aortic arches, we use the 1-dimensional persistent homology in the loss function since this time the shape becomes more distinctive for these relatively larger veins. The fact that the holes are born and they die at different indices of the Vietoris-Rips filtration (see Fig. 2) gives essential information about the shape of the aortic arch, and this trait can be successfully used to train the network.

B. Topology-Aware Loss Function

Let I be a training image, $i \in I$ be a pixel, p_i be the ground truth for the pixel i, and $\hat{p}_i$ be its posterior probability estimated by a network. Here $p_i = 1$ if i is an aortic arch or a great vessel pixel, and $p_i = 0$ otherwise. The cross-entropy loss CE_I for the image I is defined as:

$$\text{CE}_I = \sum_{i \in I} p_i \log \hat{p}_i + (1 - p_i) \log(1 - \hat{p}_i) \tag{1}$$

In this work, we define our topology-aware loss function $\mathcal{L}_T$ as a weighted sum of cross-entropy losses CE_I

$$\mathcal{L}_T = -\sum_{I} \omega_I \ \text{CE}_I \tag{2}$$

where ω_I is the topological weight term for the training image I calculated based on the difference between the persistence diagrams of its ground truth map $\mathcal{S}_I$ and the prediction map $\widehat{\mathcal{S}}_I$ estimated by the network at the end of each forward pass.

We define the topological weight term ω_I as a linear combination of the Wasserstein distances of the homology group 0 and the homology group 1, $d_0(\mathcal{S}_I, \widehat{\mathcal{S}}_I)$ and $d_1(\mathcal{S}_I, \widehat{\mathcal{S}}_I)$, respectively.

$$\omega_I = 1 + \alpha_I \cdot d_0(\Pi_{\mathcal{S}_I}, \Pi_{\widehat{\mathcal{S}}_I}) + \beta_I \cdot d_1(\Pi_{\mathcal{S}_I}, \Pi_{\widehat{\mathcal{S}}_I}) \tag{3}$$

where α_I and β_I are the constants that determine the importance of a homology group. Based on our discussions given at the end of Sec. III-A, if the ground truth $\mathcal{S}_I$ of the image I includes any great vessel, we consider only the homology group 0 and empirically set $(\alpha_I,\ \beta_I)$ = (5.0e–6, 0.0). Otherwise, if $\mathcal{S}_I$ contains only the aortic arches (without any great vessel), we consider the homology group 1 and set $(\alpha_I,\ \beta_I)$ = (0.0, 1.0e–4).

There are different choices to calculate the distance between two persistence diagrams, namely $\Pi_{\mathcal{S}_I}$ and $\Pi_{\widehat{\mathcal{S}}_I}$. These calculations rely on finding matches between the points in these two persistence diagrams that minimize the cost over all matchings, where the points are allowed to be matched with any point on a diagonal (Fig. 4). The bottleneck distance uses the maximum of the distances between the matched points whereas the Wasserstein distance uses the sum of the powers of the distances between the matched points. The main difference between these two distance functions is that several matched points have a contribution to the Wasserstein distance whereas they have no contribution to the bottleneck distance. In preliminary tests with our data, we noticed that using the Wasserstein distance led to better results. This is consistent with our intuition that the loss function defined in terms of the Wasserstein distance will continue improving the network when there are several predicted components which need minor corrections. In contrast, the loss function defined in terms of the bottleneck distance will not improve the network when all the components would need only minor corrections.

C. Network Architecture and Training

An encoder-decoder network is used to segment the aortic arch and great vessels in CT images. This network is trained

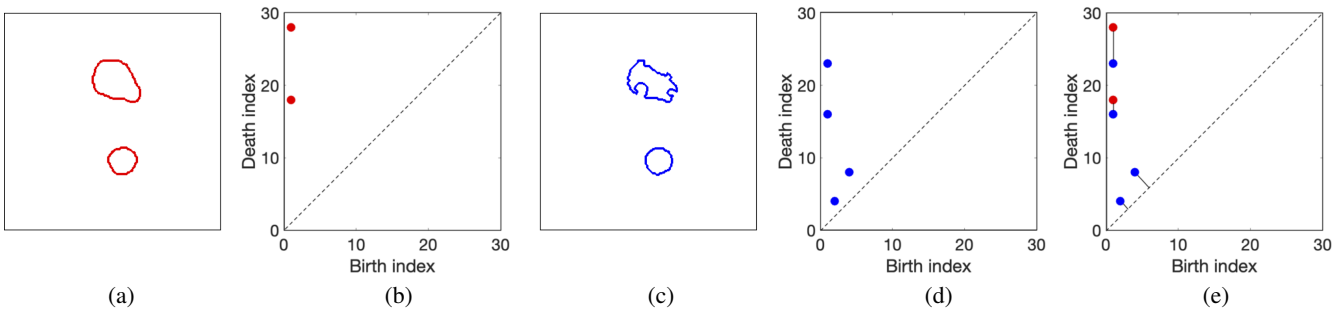


Fig. 4. (a) Boundaries of the objects found in the ground truth map $\mathcal{S}_I$ and (b) its persistence diagram $\Pi_{\mathcal{S}_I}$ for the homology group 1. (c) Boundaries of the objects found in the prediction map $\widehat{\mathcal{S}}_I$ and (d) its persistence diagram $\Pi_{\widehat{\mathcal{S}}_I}$ for the homology group 1. (e) Illustration of the best matching of these two diagrams. The Wasserstein distance is the sum of the powers of the distances between the matched points.

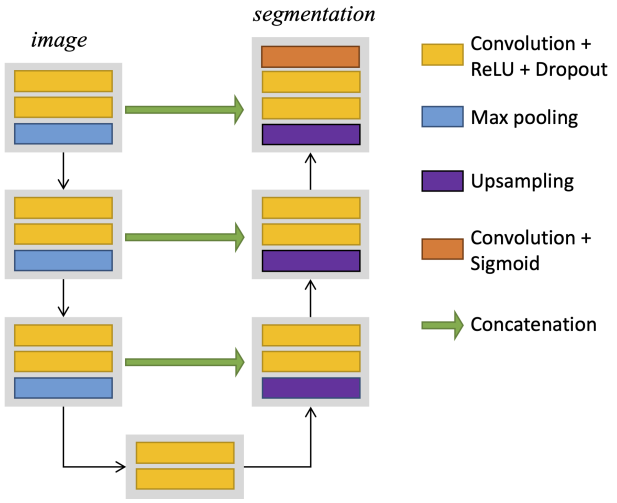


Fig. 5. UNet architecture used as the base model.

to minimize the proposed topology-aware loss function by backpropagation. At each epoch, the forward pass estimates segmentation maps for every training image and updates the topology-aware loss $\mathcal{L}_T$ by calculating cross-entropy losses CE_I as well as topological weight terms ω_I with respect to the difference between the ground truths and the predictions. Then, the backward pass updates the network weights by differentiating the updated loss $\mathcal{L}_T$. This training has a warm-up period for 25 epochs where only the cross-entropy loss is used (i.e., $\omega_I = 1$). Afterwards, it continues with minimizing the proposed topology-aware loss function $\mathcal{L}_T$. It is worth noting that this strategy is also used by the previous studies [5].

In this work, we use a UNet architecture [1], which is illustrated in Fig. 5. The encoder path comprises three blocks of two convolutions, with 3×3 filters, and one max pooling, with a 2×2 filter. The dropout layer with a dropout factor of 0.3 is added to prevent overfitting. The encoder paths starts with 32 feature maps in its first block and doubles the number of feature maps at the end of each block. The bottleneck block has the same two convolution layers without max pooling. The decoder path includes three blocks, each of which consecutively applies upsampling, concatenation, and two convolution operations. Likewise, the convolution and upsampling layers use 3×3 and 2×2 filters, respectively. The number of feature maps are halved at the end of each decoder block. All convolution layers except the last one use the ReLu activation function. The last layer uses the sigmoid function.

This network was implemented in Python using the PyTorch framework. It was end-to-end trained from scratch with an early stopping approach; training was stopped if there was no improvement on the validation set loss in the last 40 epochs. AdaDelta was used as an optimizer to adaptively adjust the learning rate and the momentum. The batch size was selected as 1. The training was conducted on a Tesla T4 GPU. The implementation is available at https://github.com/seherozcelik/TopologyAware.

IV. Experiments

A. Dataset

The proposed topology-aware loss function was tested on a dataset that contains CT scans of 24 subjects with prediagnosis of pulmonary embolism. The CT scans were acquired using a 128 slice Philips Ingenuity CT scanner with 1.5 mm slice thickness. A 60 ml of non-ionic contrast material (iohexol; generic name Opaxol) was introduced with a 100 ml saline chaser at 5 ml/s. The data collection was conducted in accordance with the tenets of the Declaration of Helsinki and was approved by Koc University Institutional Review Board (Protocol number: 2022.161.IRB1.064).

We randomly split the 24 subjects into the training and test sets. The training set contains 2896 images of 16 subjects; 2234 images of 12 subjects were used to learn the network weights by backpropagation and 662 images of 4 subjects were used as validation images for early stopping. The test set comprises 1431 images of 8 subjects; note that the images of none of these subjects were used neither in the training nor for early stopping.

B. Evaluation

Predictions were quantitatively evaluated by calculating the performance metrics both at the pixel- and vessel-level. For the pixel-level evaluation, true positive pixels were found, and the precision, recall, and f-score were calculated for each image,

TABLE I
PERFORMANCE METRICS CALCULATED OVER THE TEST SET IMAGES. THESE ARE THE AVERAGES AND STANDARD DEVIATIONS ACROSS FIVE RUNS.

	Pixel-level metrics			Vessel-level metrics			Hausdorff
	Precision	Recall	F-score	Precision	Recall	F-score	distance
Proposed algorithm	**88.2 ± 1.4**	**87.2 ± 2.1**	**86.6 ± 0.7**	74.6 ± 2.0	84.0 ± 1.2	**79.0 ± 0.9**	**5.3 ± 0.4**
Baseline [1]	87.6 ± 2.0	86.0 ± 2.5	85.6 ± 2.2	71.5 ± 6.3	82.1 ± 4.8	76.4 ± 5.6	5.5 ± 0.9
LikelihoodFiltration [5]	87.2 ± 0.7	86.9 ± 0.8	85.8 ± 0.6	73.3 ± 1.7	**84.1 ± 2.2**	78.3 ± 1.7	5.7 ± 0.6
FourierNet [26]	85.7 ± 2.8	81.2 ± 4.3	81.6 ± 1.6	**76.9 ± 3.1**	73.1 ± 2.9	74.9 ± 1.5	7.6 ± 1.1

TABLE II
PERFORMANCE METRICS CALCULATED OVER (A) THE TEST SET IMAGES CONTAINING ANY GREAT VESSELS AND (B) THE TEST SET IMAGES CONTAINING ONLY AORTIC ARCHES. THESE ARE THE AVERAGES AND STANDARD DEVIATIONS ACROSS FIVE RUNS.

	Pixel-level metrics			Vessel-level metrics			Hausdorff
	Precision	Recall	F-score	Precision	Recall	F-score	distance
Proposed algorithm	**83.2 ± 2.4**	**80.5 ± 3.3**	**79.9 ± 1.0**	**72.1 ± 1.0**	78.6 ± 1.9	**75.2 ± 1.0**	**5.7 ± 0.6**
Baseline [1]	82.5 ± 3.1	78.4 ± 4.7	78.5 ± 3.8	68.9 ± 6.8	76.3 ± 6.3	72.4 ± 6.5	5.9 ± 1.4
LikelihoodFiltration [5]	81.8 ± 1.2	80.0 ± 1.5	78.8 ± 0.9	71.4 ± 1.9	**79.2 ± 3.1**	75.1 ± 2.1	6.0 ± 0.6
FourierNet [26]	80.3 ± 3.8	71.0 ± 6.9	72.7 ± 2.9	71.66 ± 3.2	64.2 ± 4.1	67.6 ± 1.9	9.9 ± 1.9

(a)

	Pixel-level metrics			Vessel-level metrics			Hausdorff
	Precision	Recall	F-score	Precision	Recall	F-score	distance
Proposed algorithm	**93.5 ± 0.7**	**94.5 ± 1.1**	**93.7 ± 0.8**	80.3 ± 5.2	**96.8 ± 1.3**	87.7 ± 3.2	**4.9 ± 0.4**
Baseline [1]	93.1 ± 1.2	94.1 ± 0.2	93.3 ± 0.6	77.2 ± 6.1	95.8 ± 1.5	85.5 ± 4.2	5.0 ± 0.6
LikelihoodFiltration [5]	93.2 ± 0.9	94.2 ± 0.9	93.3 ± 0.7	77.5 ± 3.2	95.9 ± 1.6	85.7 ± 2.4	5.4 ± 0.8
FourierNet [26]	91.5 ± 1.8	92.3 ± 1.8	91.3 ± 1.0	**87.3 ± 2.6**	94.7 ± 2.0	**90.8 ± 2.1**	5.0 ± 0.4

(b)

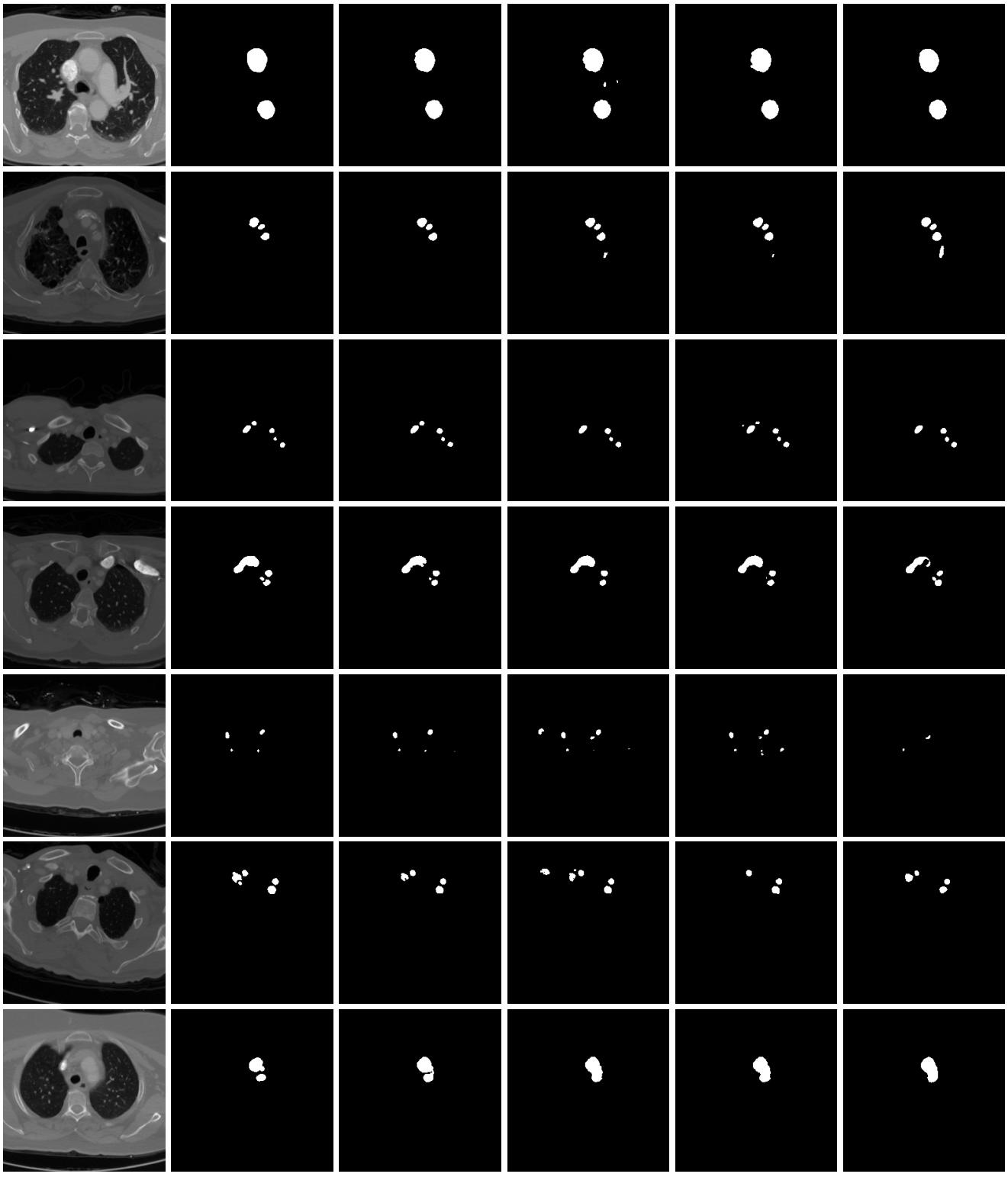
separately. These metrics were then averaged over the test set images. The vessel-level evaluation was conducted as follows: Let s_i be a vessel (a great vessel or an aortic arch) in the ground truth $\mathcal{S}_I$ of a test set image I, and $\hat{s}_j$ be a segmented object in its prediction map $\widehat{\mathcal{S}}_I$. Each vessel $s_i \in \mathcal{S}_I$ was matched with its maximally overlapping object $\hat{s}_j \in \widehat{\mathcal{S}}_I$, and considered as true positive if the intersection-over-union (IoU) for this match was greater than 50 percent. Afterwards, true positive vessels were accumulated over all test set images and the vessel-level metrics were calculated. With TP_I being the number of true positive vessels in the test set image I, precision $= \sum_I TP_I / \sum_I |\widehat{\mathcal{S}}_I|$, recall $= \sum_I TP_I / \sum_I |\mathcal{S}_I|$, and f-score was calculated. Additionally, the Hausdorff distance was found between each ground truth vessel $s_i \in \mathcal{S}_I$ and its maximally overlapping object in the prediction map, and vice versa. If there is no overlap for a vessel, the Hausdorff distance was calculated between this vessel and the closest segmented object. Then, for the test image I, the overall Hausdorff distance was the weighted average of all Hausdorff distances where the weight of a vessel was selected as the ratio of the vessel's area to the area of all vessels in $\mathcal{S}_I$. Note that better segmentations yield higher precision, recall, and f-score metrics, and lower Hausdorff distances.

C. Comparisons

We used three algorithms for comparison and ablation studies. The first one was the *Baseline* algorithm that had the UNet architecture given in Fig. 5. This network and its training were exactly the same with ours except that it used the standard cross-entropy as its loss function. We used this algorithm in our comparisons to understand the importance of using a topology-aware loss function in the network training. Here it is worth noting that we used the same set of initial network weights for this baseline as well as our model. Thus, it is possible to directly observe the effects of a loss function in a controlled setting.

The second algorithm had also the same network design, with the same set of initial network weights, but used another topology-preserving loss function suggested by [5]. As mentioned in the introduction and the related work, this suggested loss relied on calculating the persistent homology based on the threshold filtration associated to a likelihood function predicted by the network and the Betti numbers of the ground truth. We included this *LikelihoodFiltration* algorithm in our comparisons to investigate the benefits of using Vietoris-Rips filtration on the persistent homology, which is effective in modeling the topology but also the geometry of the ground truth vessels (i.e., both the shape of the vessels and the distribution of the vessels with respect to each other). Note that in our experiments, we run this algorithm using the codes provided by its authors.

The last comparison was with the *FourierNet* algorithm that proposed a shape-preserving network design [26]. This algorithm represented the shape prior by extracting Fourier descriptors on the objects' contours and concurrently learned these descriptors with the main task of segmentation. We included this algorithm in our comparisons to observe the effects of modeling the vessels' geometry instead of modeling only the vessels' shape. We also run this algorithm using the codes provided by its authors.



(a) (b) (c) (d) (e) (f)

Fig. 6. (a) Example test set images. (b) Ground truths. (c) Results of our model that used the proposed topology-aware loss function. (d) Results of the *Baseline* algorithm [1], which uses the standard cross entropy loss. (e) Results of the topology-preserving network, the *LikelihoodFiltration* algorithm, proposed by [5]. (f) Results of the shape-preserving *FourierNet* algorithm proposed by [26].

D. Results and Discussion

The quantitative results obtained on all test set images are given in Table I. We run our model as well as the comparison algorithms five times, and these are the quantitative results averaged across these five runs. This table reveals that our model, which uses the proposed topology-aware loss function, leads to better segmentations, giving higher f-scores and lower Hausdorff distances. The visual results obtained on exemplary test set images are also consistent with this observation (Fig. 6).

Comparing with the *Baseline* algorithm, which uses the standard cross-entropy as its loss function, our model with the proposed topology-aware loss is effective to eliminate false positives (the first two rows of Fig. 6) as well as correct false negatives (the third and fourth rows of Fig. 6). It can also fix incorrect segmentations when false positives and false negatives are found in the same segmentation map (the fifth and sixth rows of Fig. 6). Although the *LikelihoodFiltration* and *FourierNet* algorithms can correct them to some extent, the proposed model is more effective than these algorithms, as also reflected in the quantitative results. The last row of Fig. 6 shows an example where all models failed to fix an undersegmentation. However, even on this example, the proposed loss function improved incorrectly segmented pixels better than the other algorithms, partially predicting the boundary pixels between the two vessels.

The main contribution of this work is to use the Vietoris-Rips filtration for calculating the persistent homology of the ground truth and the prediction. This has the benefit of modeling the shape of the objects and their *geometry*, which the persistent homology associated to a likelihood function fails to detect. This concurrent modeling is essential to capture the global invariants in our application. Since CT images contain both aorta and great vessels, the shape of the aorta and the distribution and the distances of the great vessels with respect to each other contain important prior geometric information that could be exploited. To investigate this further, we also calculated the performance metrics separately, for images containing any great vessels and for those containing only the aorta (or the aortic arches). These metrics are reported in Tables II(a) and II(b), respectively. These tables demonstrate that the proposed topology-aware loss function is able to improve the metrics both for the great vessels and the aortic arches. Here one can observe that the *FourierNet* algorithm, with the shape-preserving property, gives the best vessel-level f-score for the aortic arches, which is consistent with our observation that the shape is important for the aorta. On the other hand, it is not successful to model the distribution of the great vessels, and in turn, it yields the worst results for them.

V. Conclusion

This paper presented a topology-aware loss function for automated segmentation of the aorta and great vessels in CT images. This loss function was defined as a weighted cross entropy, in which the weight for an image was the topological dissimilarity between the ground truth map and the segmented map predicted by the network at the end of each epoch. Different from the previously suggested segmentation network designs, this paper proposed to apply the Vietoris-Rips filtration to obtain the persistence diagrams of these maps and calculate their (dis)similarity using the Wasserstein distance between the corresponding persistence diagrams. Experiments on 4327 CT images of 24 subjects revealed that this proposal is more effective than its counterparts in simultaneously modeling the shape of the aorta and the geometry between the great vessels.

In this work, we used the topological dissimilarity between the ground truth and the prediction to define the weight of an image in the loss function. In other words, we used the same dissimilarity metric to penalize every pixel in the same image, regardless of whether they were correctly and incorrectly predicted. One future research direction is to reflect the dissimilarity only to false negative and false positive pixels, and possibly with different extents. The aorta and great vessel segmentation provides an exemplary showcase for the necessity of modeling the shape and geometry at the same time, and hence, the effectiveness of applying the Vietoris-Rips filtration to obtain the persistence diagrams for defining the proposed topology-aware loss function. Using it for other segmentation problems can be considered as another future research direction.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2015, pp. 234–241.

[2] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 240–248.

[3] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comp. Vis.*, 2017, pp. 2980–2988.

[4] G.N. Gunesli, C. Sokmensuer, and C. Gunduz-Demir, "AttentionBoost: Learning what to attend for gland segmentation in histopathological images by boosting fully convolutional networks," *IEEE Trans. Med. Imaging*, vol. 39, no. 12, pp.4262–4273, 2020.

[5] X. Hu, F. Li, D. Samaras, and C. Chen, "Topology-preserving deep image segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019.

[6] J. Clough, N. Byrne, I. Oksuz, V.A. Zimmer, J.A. Schnabel, and A. King, "A topological loss function for deep-learning based image segmentation using persistent homology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, pp. 8766–8778, 2022.

[7] E. Khramtsova, G. Zuccon, X. Wang, and M. Baktashmotlagh, "Rethinking persistent homology for visual recognition," in *Topological, Algebraic and Geometric Learning Workshops*, 2022, pp. 206–215.

[8] N. Byrne, J. Clough, I. Valverde, G. Montana, and A. King, "A persistent homology-based topological loss for CNN-based multiclass segmentation of CMR," *IEEE Trans. Med. Imaging,* vol. 42, no. 1, 2023.

[9] S. Bonechi *et al.,* "Segmentation of aorta 3D CT images based on 2D convolutional neural networks," *Electronics*, vol. 10, no. 20, pp. 2259, 2021.

[10] W.K. Cheung *et al.,* "A computationally efficient approach to segmentation of the aorta and coronary arteries using deep learning," *IEEE Access*, vol. 9, 2021.

[11] J. Zhong, Z. Bian, C.R. Hatt, and N.S. Burris, "Segmentation of the thoracic aorta using an attention-gated U-Net," *Medical Imaging 2021: Computer-Aided Diagnosis*, vol. 11597, 2021.

[12] L. Gu and X.-C. Cai, "Fusing 2D and 3D convolutional neural networks for the segmentation of aorta and coronary arteries from CT images," *Artif. Intell. Med.*, vol. 121, pp. 102189, 2021.

[13] A. Hatcher, "Algebraic Topology," *Cambridge University Press, Cambridge*, 2002.

[14] G. Carlsson, "Topology and data," *Bull. Am. Math. Soc.*, vol. 46(2), pp. 255–308, 2009.

[15] F. Chazal, and B. Michel, "An introduction to topological data analysis: fundamental and practical aspects for data scientists," 2021, arXiv:1710.04019.

[16] M. Haft-Javaherian, M. Villiger, C.B. Schaffer, N. Nishimura, P. Golland, and B.E. Bouma, "A topological encoding convolutional neural network for segmentation of 3D multiphoton images of brain vasculature using persistent homology," in *Proc. IEEE Int. Conf. Comp. Vis. Pattern Recognit. Workshops*, 2020, pp. 990–991.

[17] C.-C. Wong and C.-M. Vong, "Persistent homology based graph convolution network for fine-grained 3D shape segmentation," in *IEEE Int. Conf. Comp. Vis.*, 2021, pp. 7078–7087.

[18] F. Wang, H. Liu, D. Samaras, and C. Chen, "TopoGAN: A topology-aware generative adversarial network," in *Computer Vision – ECCV 2020*, 2020, pp. 118–136.

[19] A. Mosinska, P. Marquez-Neila, M. Kozinski, and P. Fua, "Beyond the pixel-wise loss for topology-aware delineation," in *Proc. IEEE Conf. Comp. Vis. Pattern Recognit.*, 2018, pp. 3136–3145.

[20] C. Hofer, R. Kwitt, M. Niethammer, and A. Uhl, "Deep learning with topological signatures," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017.

[21] C. Hofer, R. Kwitt, M. Niethammer, and M. Dixit, "Connectivity-optimized representation learning via persistent homology," in *Proc. Int. Conf. Mach. Learning*, 2019, pp. 2751–2760.

[22] M. Moor, M. Horn, B. Rieck, and K. Borgwardt, "Topological autoencoders," in *Proc. Int. Conf. Mach. Learning*, 2020, pp. 7045–7054.

[23] R. Gabrielsson, B. Nelson, A. Dwaraknath, P. Skraba, L. Guibas, and G. Carlsson, "A Topology layer for machine learning," in *Proc. Int. Conf. Artif. Intell. Statistics*, 2020, pp. 1553–1563.

[24] R. Gabrielsson, and G. Carlsson, "A look at the topology of convolutional neural networks," 2018, arXiv:1810.03234.

[25] T. Qaiser *et al.,* "Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features," *Med. Image Anal.*, vol.55, pp. 1–14, 2019.

[26] S. Cansiz, C. Kesim, S.N. Bektas, Z. Kulali, M. Hasanreisoglu, and C. Gunduz-Demir, "FourierNet: Shape-preserving network for Henle's fiber layer segmentation in optical coherence tomography images," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 2, pp.1036–1047, 2023.