Topological distance-constrained feature descriptor learning model for vessel matching in coronary angiographies

Xiaojiao SONG, Jianjun ZHU, Jingfan FAN*, Danni AI, Jian YANG

Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China

* Corresponding author, fjf@bit.edu.cn

Received: 20 February 2021 Accepted: 19 April 2021

Supported by the Key Technologies R&D Program of the National Ministry of Science (2018ZX10734–401); the National Science Foundation Program of China (61901031, 62071048, 62025104, 61971040); Beijing Nova Program from Beijing Municipal Science & Technology Commission (Z201100006820004).

Citation: Xiaojiao SONG, Jianjun ZHU, Jingfan FAN, Danni AI, Jian YANG. Topological distance-constrained feature descriptor learning model for vessel matching in coronary angiographies. Virtual Reality & Intelligent Hardware, 2021, 3(4): 287—301
DOI: 10.1016/j.vrih.2021.08.003

Abstract Background Feature matching technology is vital to establish the association between virtual and real objects in virtual reality and augmented reality systems. Specifically, it provides them with the ability to match a dynamic scene. Many image matching methods, of which most are deep learning-based, have been proposed over the past few decades. However, vessel fracture, stenosis, artifacts, high background noise, and uneven vessel gray-scale make vessel matching in coronary angiography extremely difficult. Traditional matching methods perform poorly in this regard. Methods In this study, a topological distance-constrained feature descriptor learning model is proposed. This model regards the topology of the vasculature as the connection relationship of the centerline. The topological distance combines the geodesic distance between the input patches and constrains the descriptor network by maximizing the feature difference between connected and unconnected patches to obtain more useful potential feature relationships. Results Matching patches of different sequences of angiographic images are generated for the experiments. The matching accuracy and stability of the proposed method is superior to those of the existing models. Conclusions The proposed method solves the problem of matching coronary angiographies by generating a topological distance-constrained feature descriptor.

Keywords Vessel matching; Deep learning; Feature descriptor; Coronary angiographies; Geodesic distance; Topological distance-constrained

1 Introduction

Coronary artery disease is a common cardiovascular illness which can be fatal. Globally, it is the "number one killer"[1]. Vessel matching in coronary angiography has been rapidly developing in recent years. This technology is vital in X-ray angiography (XRA)-guided surgery, coronary intervention surgery, endovascular
diseases, and other clinical treatments. However, feature descriptors based on handcrafted methods require professional and complex design procedures, due to which potentially useful information may be lost. Therefore, vessel matching methods with greater accuracy and stability are required urgently.

The concentration of the contrast agent varies with time and space; therefore, the gray level of the angiogram during the imaging process of coronary angiography depends on both the factors. In addition, the X-ray gray level of vessels and that of organs, such as bone and diaphragm, are similar; consequently, the images contain artifacts, which makes it difficult to recognize vessels. Furthermore, medical angiographic images contain few feature points and vessel fracture and stenosis compound the difficulty of vessel matching.

Many matching methods have been proposed over the past few decades to address the aforementioned issues. Feature-based matching methods are widely used in different fields of matching tasks; they are usually more efficient and robust to geometrical deformations. They are divided into two main categories.

1.1 Feature descriptors based on handcrafted design

Gradient statistics-based descriptors. Lowe et al. proposed the scale-invariant feature transform (SIFT) descriptor that uses the constructed scale space to find key points, assigns direction to each key point, and generates a very stable and invariant 128-dimensional descriptor\(^2\). The descriptor can adapt to factors such as rotation, scale scaling, and luminance changes. Bay et al. proposed the speeded up robust features (SURF) descriptor, wherein the gradient calculation in SIFT was replaced using Haar wavelets and integral map approximation, which improved computational efficiency\(^3\). Cao et al. presented a general framework of multi-scale geometric analysis and particle swarm optimization virtual reality (VR) image matching algorithm; they performed experiments on several feature description algorithms\(^4\). Morel et al. proposed the Affine-SIFT (ASIFT), which improves the affine invariance by varying the coordinate orientation parameters of two cameras to simulate all the image views\(^5\). Tola et al. proposed a dense DAISY descriptor that uses a log-polar grid and a Gaussian pooling strategy to approximate the gradient direction histogram\(^6\).

Intensity-based descriptors. Calonder et al. proposed the binary robust independent elementary features (BRIEF) descriptor, a combination of the intensity binary test results of several random pairs of points in an image block\(^7\). Rublee et al. proposed the oriented features from accelerated segment test (FAST) and rotated BRIEF (ORB) algorithm, which combines the rotational BRIEF and FAST operators and used a machine learning strategy to select a robust binary test to reduce sensitivity to rotation and scale changes\(^8\). Inspired by the structure of the retina, Alahi et al. proposed the fast retina keypoint (FREAK) descriptor by comparing the image intensity of retinal sampling patterns\(^9\); the FREAK descriptor has low memory consumption and fast running speed and is robust to rotation and noise. Wang et al. encoded the local order information for each pixel and divided the local patches into sub-regions, using all the sequence information to speed up the calculation\(^10\).

1.2 Feature descriptors based on learning

Classical learning-based descriptors. Ke et al. proposed principal component analysis (PCA)-SIFT, which uses PCA to obtain a highly robust and compact descriptor by reducing the vector dimension obtained from the local image gradient\(^11\). Brown et al. introduced a learning framework with a series of sub-blocks to construct descriptors and used the Powell minimization and linear discriminant analysis to determine the optimal parameters\(^12\). Gionis et al. proposed the locally sensitive hashing (LSH) algorithm, an
unsupervised hashing method that generates an embedding expression by random projection\cite{13}. Subsequently, unsupervised algorithms, such as kernelized locally sensitive hashing (kernelized LSH)\cite{14}, spectral hashing\cite{15}, and semantic hashing\cite{16}, have been proposed.

Deep-learning-based descriptors. Bhowmik et al. used the principle of reinforcement learning to optimize keypoint selection and matching under high-level vision tasks\cite{17}. Simo-Serra et al. proposed DeepDesc, which uses a convolutional neural network (CNN) to learn patch representation and measure the L2 distance\cite{18}. Specifically, they trained twin networks using positive and negative patches by minimizing the hinge loss. The proposed hard-negative matching sampling strategy alleviated the imbalance between positive and negative samples, thereby improving the performance of the descriptor significantly. Balntas et al. proposed TFeat, a CNN-based method implemented by a shallow convolutional network and fast hard-negative mining strategy that used three sets of training samples to describe and match patches\cite{19}. Tian et al. proposed L2-Net, which used an advanced sampling strategy to optimize the distance-based correlation loss function in the Euclidean space\cite{20}. The author focused on intermediate feature mapping and the compactness of the descriptors in the training process. Mishchuk et al. proposed HardNet, whose network structure was similar to that of the L2-Net. Owing to its use of a hinge triplet loss and a "hardest-within-batch" sampling strategy, it achieved better training results than the L2-Net\cite{21}.

The above feature descriptors are used for feature-based image matching to establish the correct image pixel or point correspondence between two images. Smeets et al. proposed a robust matching method for 3D lung vessel trees\cite{22}. For the first time, a three-dimensional tree was expressed as an intrinsic matrix, which included the geodesic distance or Euclidean distance between each pair of detected branches. A global correspondence model was combined with a local bifurcation similarity model based on the shape and local gray-scale distribution of neighboring vessels. Serradell et al. described vessel matching as a process to find the maximum possible correspondence, a process they accelerated using a priori search\cite{23}. Ma et al. proposed a Bayesian formulation for rigid and non-rigid feature matching and image registration. Local linear transformation constraints were introduced to further exploit the geometric information\cite{24}. Recently, they also proposed a guided local preserving matching method\cite{25}. Benseghir et al. proposed the iterative nearest neighbor curve (ICC) method, which uses vessel branches as pairing elements based on the nearest neighbor relationships and estimated a transformation that minimized the sum of distances between the pairing branches\cite{26}. Moriconi et al. defined the closeness function of graph matching by setting node and edge attributes, following which they maximized the function of the quadratic assignment problem to match nodes\cite{27}. Fischler et al. proposed the classical random sample consensus (RANSAC) algorithm, which assumed that two images are coupled by a specific parametric geometric relationship such as a projection transformation or polar geometric transformation\cite{28}. This algorithm was based on a hypothesis and validation strategy, which entailed repeatedly sampling a minimal subset from the data, estimating the model hypothesis, and verifying its quality using a continuous number of internal points.

Finally, correspondences consistent with the optimized model were identified as the interior points. Subsequently, several different approaches have been proposed to improve the performance of RANSAC. Torr et al. proposed maximum likelihood estimation sample consensus (MLESAC), which validated the quality of the model using a maximum likelihood process\cite{29}. Under some assumptions, MLESAC can optimize the results and is less sensitive to a predetermined threshold. Myatt et al. proposed the N adjacent points sample consensus, which assumes that the interior points are continuous\cite{30}. Ni et al. proposed GroupSAC based on the assumption that interior points exist in groups\cite{31}. Chum et al. proposed the locally optimized (LO)\textendash RANSAC\cite{32}. Thus far, it has been observed that the minimum subset can amplify the potential noise and generate a hypothesis far from the basic facts. When the optimal model is reached, the
problem can be solved through an iterative least-squares fitting process and local optimization. Chen et al. proposed a partial intensity invariant feature descriptor for matching retinal images\(^{33}\).

The abovementioned feature-based matching methods rely on obvious feature detection. To make them applicable to coronary angiography images with less information, their matching accuracy and stability need to be enhanced. In this study, a topological distance-constrained feature descriptor learning model for vessel matching in coronary angiography is proposed to address the aforementioned problems. The geodesic distance between patches was used to reflect the connection relationship. Regarding this, the feature differences between the connected and unconnected patches were maximized. Accordingly, additional topological constraints were applied to the training network. Vascular topology is an invariant property of morphology and dimensionality. Thus, topology, as important information, becomes an effective vessel matching strategy. A topology-supervised network was introduced to learn more useful features based on angiography images to solve the problem of insufficient information and other characteristics in medical angiography images. A more accurate matching result can be achieved by the model through topological distance constraints. The proposed model was completely driven by data samples and exhibited superior performance in the field of vessel matching.

2 Methods

We propose a joint loss function \(L\) based on the potential features between matching patches and the topological structure of the coronary arteries. This function propagates information between matching patches of two images and between all the patches with topology in the same image to jointly supervise network training. The joint loss function \(L\) consists of two parts, the feature constraint function part \(L_D\) and the topological distance constraint function part \(L_T\).

Figure 1 shows the flowchart of the proposed method. The input data, training network, output features, and joint loss function were exploited in the training process. First, a CNN was first trained using the matching patches extracted from the vessel images in coronary angiographies to learn the potential feature relationships. The network training was supervised by the topological distance function \(L_T\) and feature constraint function \(L_D\). These functions were combined to update the network parameters. This made it possible to feed the vessel images into the trained network to obtain the corresponding feature descriptors; then, the matching similarity between the descriptors was calculated. We found the topological neighborhood value corresponding to the best model by validating the experimental results using different topological neighborhood values, which improved the accuracy and stability of the descriptors.

![Figure 1](image)

Figure 1  Schematic diagram of proposed method. \(L_D\) denotes the feature constraint function, \(L_T\) denotes the topological distance constraint function, both of which form the joint loss function \(L\).

2.1 Network architecture

Figure 2 shows the network structure used in this study, which was similar to that of the L2-Net\(^{29}\). The network parameters are listed in Table 1. The input to the network was 32×32 patches of coronary...
The goal function for this part was as follows:

Thus the final negative descriptor was selected using.

Accordingly.

where

Each matching pair

and its size was

had two

where

respectively

2

A

was the same as that of the

128.

The dimension of the output of the network

P

and positive

A

were generated

anchor

×

32.

batch of

matching

32

32

The feature loss function

L21

2

had the same dimensions as the SIFT descriptors.

respectively

1

128

1

The feature descriptors output by the network

which were

and

0

they were set to constant;

The

No pooling layer was included in

(A dropout

layer and a rectified linear unit

ReLU).

Each convolutional layer

except the last one

it was normalized by subtracting the mean value of each patch and dividing them by

maximized

the following methods

set of four descriptors

second-nearest neighbor descriptors

where

pairwise distance matrix

SIFT

the potential features

patches

could strongly recognize features of very similar non-matched patches

maximize the feature differences between the matched and non-matched patches to ensure that the network

The feature loss function

L

dimensions

values of

the network because experiments show that pooling reduced the performance of deep descriptors

layer was added before the last convolution layer to prevent overfitting.

angiography images; it was normalized by subtracting the mean value of each patch and dividing them by

the standard deviation of each patch. Each convolutional layer, except the last one, was followed by a

batch normalization (BN) layer and a rectified linear unit (ReLU) nonlinear activation layer. A dropout

layer was added before the last convolution layer to prevent overfitting. No pooling layer was included in

the network because experiments show that pooling reduced the performance of deep descriptors. The

weighting and bias parameters of the BN were not continually updated; rather, they were set to constant

values of 1 and 0, respectively. The feature descriptors output by the network, which were

1×128

dimensions, had the same dimensions as the SIFT descriptors.

2.2 Feature loss function \( L_\theta \)

The feature loss function \( L_\theta \) was the same as that of HardNet(21). The feature loss function \( L_\theta \) was learned to

maximize the feature differences between the matched and non-matched patches to ensure that the network

could strongly recognize features of very similar non-matched patches. First, a batch of \( 32 \times 32 \) matching

patches, anchor (A) and positive (P), were generated. These patches were then fed into the network to learn

the potential features. The dimension of the output of the network, \( 1 \times 128 \), was the same as that of the

SIFT(2) descriptor. \( a_i \) and \( p_i \) denote the anchor and positive descriptors, respectively, where \( i = 1 \cdots n \). A

pairwise distance matrix

\( D = \text{pdist}(a_i, p_i) \)

was established between the anchor and positive descriptors,

where

\( d(a_i, p_j) = \sqrt{2 - 2a_i p_j} \), \( i = 1 \cdots n \), \( j = 1 \cdots n \), and its size was \( n \times n \). Each matching pair\((a_i, p_j)\) had two

second-nearest neighbor descriptors \( a_{i_2} \) and \( p_{i_2} \), respectively, where \( j, k \in (1, n), j \neq i, k \neq i \). Accordingly, a

set of four descriptors \((a_i, p_j, a_{i_2}, p_{i_2})\) were formed. Thus, the final negative descriptor was selected using

the following methods:

\[
\begin{align*}
    a_{i_2}, & \quad \text{if } d(a_{i_2}, p_i) < d(a_i, p_i) \\
    p_{i_2}, & \quad \text{if } d(a_i, p_{i_2}) < d(a_i, p_i)
\end{align*}
\]

(1)

Finally, the distance between the matching descriptor and next-nearest non-matching descriptor was

maximized. The goal function for this part was as follows:

\[
L_\theta = \frac{1}{n} \sum_{i=1}^{n} \max \left( 0, 1 + d(a_i, p_i) - \min \left( d(a_i, p_{i_2}), d(a_{i_2}, p_i) \right) \right)
\]

(2)
2.3 Topological distance loss function $L_T$

The $L_T$ constrained the training process of the network based on the topological distance of the vascular centerline. The topology of vessels was considered the connection relationship of the centerlines; the patches with connection relationships were in proximity to each other. A graph could be constituted from the topological connections of the vessels. For the graph in space, the geodesic distance could reflect the distance relationship between patches more realistically. Accordingly, the real topology of vessels could be maintained. Therefore, the topological distance was considered the geodesic distance between patches, and network training was supervised by maximizing the differences in topological distance between the connected and unconnected patches. The $L_T$ consisted of two parts: the $L_T^1$ and $L_T^2$. Both parts have similar computational processes. They are also the anchor and positive topological distance constraints of the centerline to the network.

The adjacency matrix $\text{Adj} = \{ \text{adj}_{i,j} = 0, 1 \}_{i,j}, i = 1 \cdots n, k = 1 \cdots n, \text{adj}_{i,i} = 0$ can be generated, where 0 is unconnected and 1 is connected based on the connectivity of the vascular centerline. A batch of matching patches $(A_i, P_i), i = 1 \cdots n$ was input into the network to learn the descriptors $(a_i, p_i), i = 1 \cdots n$. The distance matrix between the anchor descriptors $D_A = \text{pdist}(a_i, a_j)$ and positive descriptors $D_P = \text{pdist}(p_i, p_j)$ were established, where $d(a_i, a_j) = \sqrt{2 - 2a_i a_j}, d(p_i, p_j) = \sqrt{2 - 2p_i p_j}, i = 1 \cdots n, j = 1 \cdots n$. The pairwise distance matrix between the anchor patches $D_A = \text{pdist}(A_i, A_j)$ and the positive patches $D_P = \text{pdist}(P_i, P_j)$ were also established, where $d(A_i, A_j) = \sqrt{2 - 2A_i A_j}, d(P_i, P_j) = \sqrt{2 - 2P_i P_j}, i = 1 \cdots n, j = 1 \cdots n$. Then, $D_A$ and $D_P$ were combined with the generated adjacency matrix $\text{Adj}$ as the input of the Floyd algorithm. The outputs were the geodesic distance matrices $G_A = \{ e_{a_i} \}$ and $G_P = \{ e_{p_i} \}$, which contained the shortest path length. Figure 3 illustrates the entire process. Based on the concept of dynamic time programming, the Floyd algorithm performed $n$ recursions, starting from the adjacency matrix $\text{Adj}$ of the graph, with the state transfer equation as

$$\text{dist}(i, j) = \min \{ \text{dist}(i, k) + \text{dist}(k, j), \text{dist}(i, j) \}$$  \hspace{1cm} (3)

where $\text{dist}(i, j)$ represents the shortest distance from $i$ to $j$, and $k$ is an intermediate node.
Therefore, the goal loss function of topological distance could be formulated as follows:

\[
L_T = \frac{1}{2} \times (L_A^t + L_P^t),
\]

where:

\[
L_A^t = \frac{2 \| G_A - D_a \|_F}{\| G_A \|_F - \| D_a \|_F},
\]

\[
L_P^t = \frac{2 \| G_P - D_p \|_F}{\| G_P \|_F - \| D_p \|_F},
\]

Combined with the feature loss function \( L_D \), the final joint loss function \( L \) can be expressed as follows:

\[
L = \alpha L_D + \beta L_T,
\]

where \( \alpha \) and \( \beta \) are the weight coefficients.

We considered the geodesic distance between training patches as the strength of the connection relationship. Thus, there was no connection relationship between the patches after the topological distance reached a certain value, at which point the constraint effect on the network should be zero. Therefore, the best matching model of the topological neighborhood was obtained by verifying different topological neighborhood values, that is, the models that best represent the number of topological distances in the connected and unconnected relationships. This eliminated the influence of unconnected patches on the network.

3 Experiments

3.1 Dataset

To train deep learning models, several samples are required to learn the underlying features, given that they are data-driven. The original data for this study were collected from a clinical database of real patients at a collaborative hospital. The corresponding computed tomography angiography and XRA data for each patient were also collected. The matching patches for training and testing the feature description and adjacency matrix of the vascular centerline were the data required for the experiments. The network was pre-trained using a natural dataset owing to the impossibility of constituting a large number of data samples using only the matching point pairs of vessels. Meanwhile, the generated patches were subjected to data augmentation before being fed into the network. These operations yielded a model with a better generalization capacity.

Generation of matching patches. The complete shapes of the vessel and its branches are presented more clearly when the contrast is full. Accordingly, several image sequences were manually selected in 2D coronary angiography imaging in the overflowing contrast state. All the images had a resolution of 512×512. Accordingly, the coronary artery vessels could be segmented, as shown in Figure 4a. The binary images could also be refined, as shown in Figure 4b. The vascular centerline could be regarded as being composed of endpoints, branch points, and bifurcation points. Thus, among them, one connection point exists for the endpoints, two connection points exist for the branch point, and three connection points exist for the bifurcation point. Therefore, the endpoints and bifurcation points of the vessel were identified based on the number of connection points with the current point. Then, the corresponding serial numbers were labeled, as shown in Figure 4c. Figure 4d shows the constructed simple vessel graph model. Only the bifurcation points and endpoints were extracted as key points in this process for the sparse representation of the vessels to reduce the amount of computation and time redundancy. Thereafter, two different
sequences of labeled coronary images were manually matched. Matching point pairs on the vascular centerline were constructed based on the vessels between the key points to establish matching relationships. Next, 65×65 square patches were extracted, with the matching points as the center, as shown in Figure 5. The number of patches collected could be adjusted based on the sampling intervals. In this experiment, a total of 11435 patch pairs were extracted for the training process.

Extraction of the adjacency matrix $Adj$ of the vascular centerline. The adjacency matrix represents the relationship between the pixels in an image. We used the eight-neighborhood as the definition range for measuring the connection relationship. Therefore, the pixel that satisfied the eight-neighborhood connection relationship was retained in the connection matrix. The adjacency matrix $Adj = \{ Adj_{ij} = 0, 1 \}_{n \times n}, i = 1 \cdots n, j = 1 \cdots n, Adj_{i,j} = 0$ could be constructed by analyzing each pixel, where 0 means unconnected and 1 means connected.

### 3.2 Evaluation metrics

We quantitatively evaluated the experimental results based on two metrics. The first one is FPR95, which means the value of the false positive rate ($FPR$) at the true positive rate ($TPR$) equals 95. The second is the number of mismatches $C_{mis}$. $FPR$ indicates the probability of the model making an incorrect prediction. Therefore, the smaller the value, the lower the error rate and the better the network performance. $FPR$ and $TPR$ are defined by the following equations:

$$FPR = \frac{FN}{TP + FN}$$

$$TPR = \frac{TP}{TP + FN}$$

where the $TP$ indicates that the prediction of the correct sample is a positive example and $FN$ indicates that the prediction of the false sample is a negative example. The number of mismatches, $C_{mis}$, represents the difference between the prediction of the model, $C_{pre}$, and the total matching pair, $C_{total}$. Therefore, the smaller the value of $C_{mis}$, the better the network performance. This can be expressed using the following formula:
3.3 Experimental details

In this study, the network model was trained using Pytorch. In the pre-training process, 500k matching pairs in HPatches\textsuperscript{(34)} natural datasets were extracted to train the network to initialize the relevant parameters. The stochastic gradient descent was used to optimize the network training process. The learning rate, momentum, and number of training were set to 0.1, 0.9, and 20, respectively. The number of samples input into the network for each experiment was set to 1024. The parameters of the pre-trained model were saved, to initialize the network of the topological distance-constrained feature descriptors.

We aimed to improve the topological distance constraint of the network. Thus, the batch size of each input network during the training process was not a fixed value. However, it varied according to the number of the matching patch pairs collected from each patient's sample data, to ensure the authenticity of the topological connection; Adam was used as the optimizer. The weight decay and the learning rate were set to $10^{-6}$ and $10^{-3}$, respectively. A dropout layer with a dropout rate of 0.3 was used before the last convolutional layer to prevent overfitting. ReLU was utilized as the activation function, and the model parameters were initialized using the saved parameters of the pre-trained model. In addition, it was necessary to pre-calculate the geodesic distance matrix between patches before the training process to reduce the computational cost and save runtime.

During the test, a batch of patches, including anchor, positive, and negative patches, were fed into the network. The negative was randomly selected from among the remaining patches, excluding the corresponding anchor and positive. The topological distance-constrained feature descriptors corresponding to each patch were obtained to calculate the FPR\textsubscript{95} value.

3.4 Results and analysis

Experiments on different topological neighborhood values were conducted to eliminate the influence of unconnected patches on the network and identify the best matching model of the topological neighborhood. The corresponding FPR\textsubscript{95} value of the network model based on different topological neighborhood constraints was also obtained, as shown in Figure 6. A topological neighborhood value of zero indicated that there was no topological distance constraint in the network. Consequently, there is only $L_0$ left in the joint loss function $L$, which is the same with HardNet\textsuperscript{(21)} model. We show only the case in which the topological neighborhood value was less than 10. Through experiments, we established that the FPR\textsubscript{95} was least, 1.37, when the topological neighborhood value was 2. The model exhibited the best performance and the strongest matching ability during this experiment.

The value of the FPR\textsubscript{95} was often directly proportional to the topological neighborhood. This trend proved that the patches with no connection relations interfered with the network training and

\[
C_{\text{mis}} = C_{\text{total}} - C_{\text{pre}}
\]

\[\text{(10)}\]

![Figure 6: FPR\textsubscript{95} of different topological neighborhood constraint models. The horizontal coordinates indicate neighborhood value and the vertical coordinates indicate FPR\textsubscript{95}.](image)
affected the performance of the feature descriptors. In a certain neighborhood range, more incorrect connections were made when the neighborhood value was larger, as a result of which the error information learned by the network increased, which degraded the training results. The neighborhood was continuously expanded to the entire threshold range, at which point the model became a topological distance-constrained feature descriptor learning model, as shown in Table 2, and the FPR95 value was 2.58. The FPR95 had a certain correct vascular topological constraint. Thus, the FPR95 of the network model was better than that of the HardNet\(^{(21)}\) model with only feature loss constraints. The FPR95 value of HardNet\(^{(21)}\) was 2.92. However, the topological constraint portion of the network contained several incorrect constraints that did not actually have connection relationships. Therefore, the FPR95 value was higher than that of the optimal topological constraint model, whose topological neighborhood was 2.

We compared the proposed method and the best topological neighborhood value model to other matching algorithms with better performance, and the FPR95 results are shown in Table 2.

Table 2 shows that the FPR95 of the topological distance-constrained feature descriptor learning model was better than that of the other methods when it was applied to vessel matching in coronary angiographies. The FPR95 values of HardNet\(^{(21)}\) and L2-Net\(^{(20)}\) were 2.92 and 7.42, respectively, in Table 2. The results indicate that the proposed model could learn more useful potential features, which made the network more capable of recognition. The FPR95 of the HardNet\(^{(21)}\) model was better than that of the L2-Net\(^{(20)}\) owing to the feature loss constraint and a certain sampling strategy. When the topological neighborhood was 2, the proposed Ours-2 model had the best FPR95 value, 1.37. At that point, the network was trained using a combination of the feature loss constraint and topological loss constraints with a fully correct connection relationship. Similarly, the FPR95 value of the proposed Ours-2 model was still the best result when no augmentation was applied in the network.

The heat map of the L2 distance between the feature descriptors of two images to be matched after the learning of different models was further verified, as shown in Figure 7. The distance on the diagonal is the distance between two corresponding matching descriptors; the matching relationship was stronger when the distance was smaller. Therefore, the Ours-2 model exhibited a better matching performance than the HardNet\(^{(21)}\) model. This was because the diagonal line is clear and the boundary is well-defined, and the color difference with other positions is obvious. Otherwise, according to the color value on the right side, the distance between the two descriptors with matching relationship on the diagonal is small, which indicates that the generated descriptors have good discrimination and high matching accuracy. The diagonal distances of the SIFT\(^{(2)}\) and L2-Net\(^{(20)}\) models were smaller; however, they were indistinguishable from some distances in other positions, which could easily cause mismatching.

The results of the SIFT\(^{(2)}\), HardNet\(^{(21)}\), and Ours and Ours-2 models, when applied to vessel matching in coronary angiographies for the same sample, are shown in Figure 8. The figure indicates that the Ours-2 model could obtain more correct matching pairs using the same gold standard matching pairs. Furthermore, the number of correct matching pairs was close to that of the gold standard matching pairs. Although the matching results of the Ours model were only the second-best, they were still better than those of the SIFT\(^{(2)}\) descriptor and HardNet\(^{(21)}\) model.

Five different groups of sample data were used for the experiments to obtain more matching results. Table 3 shows the number of mismatches, \(C_{mm}\), for the different methods. Figure 8 and Table 3 show that
the number of mismatches of the proposed method was significantly lower, especially for the optimal topological distance constraint model, which was the Ours-2 model in Table 3. The model obtained the fewest mismatches in many different samples. Furthermore, the model could obtain more correct matching relations and had stronger stability than the SIFT\(^{(2)}\) and HardNet\(^{(21)}\) models. The SIFT\(^{(2)}\) descriptors had the most mismatches. This finding indicates that traditional descriptors, which rely heavily on feature detection, have poor matching ability for coronary angiography images with less information.
We evaluated different methods on multiple samples for comparison experiments, as shown in Figure 9, owing to the presence of vessel stenosis, fracture, and small vessel segments in the medical coronary angiography images. The results revealed that the proposed Ours-2 model improved the matching and recognition of the vessel stenosis, fracture, and small vessel segments. Even in the first sample, our model and the HardNet\(^{[21]}\) model had the same matching ability for stenosis. However, the proposed model obtained fewer mismatches. Therefore, the Ours-2 model not only improved the matching ability of vessel stenosis and fracture in angiographic images but also reduced the number of mismatches. At the same time, the proposed method was very stable and could obtain fewer mismatching results than other models on different experimental samples. In the experiment, only the feature points on the centerline were extracted; the background features were ignored. Thus, the influence of uneven gray distribution and high background noise in coronary angiography images was reduced to a certain extent to further improve the accuracy of vessel matching.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Different models</th>
<th>SIFT(^{[2]})</th>
<th>HardNet(^{[21]})</th>
<th>Ours</th>
<th>Ours-2</th>
<th>C(_{total})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>8</td>
<td>12</td>
<td>3</td>
<td></td>
<td>74</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>13</td>
<td>14</td>
<td>12</td>
<td>1</td>
<td>87</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>89</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>87</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td></td>
<td>77</td>
</tr>
</tbody>
</table>

Figure 9  Results of vessel matching in coronary angiographies. From top to bottom, each row is a different sample data. From left to right in the same row is original coronary angiography image, SIFT\(^{[2]}\) feature descriptor matching, and the matching results of the L2-Net\(^{[20]}\), HardNet\(^{[21]}\) and Ours-2. The points with different colors on the vessel are the points that matched the target image accurately. The red box and arrow indicate that the model could get the correct matching result, whereas the yellow box and arrow indicate that the model could not be matched.
4 Conclusion

An image matching task is a fundamental and key technology in the field of computer vision. This technology is widely used in various fields, such as high-dimensional structure reconstruction, VR/augmented reality, image stitching, image fusion, and target recognition and tracking. Vessel matching in coronary angiography, which is an application of image matching in the medical field, is important for the clinical treatment of cardiovascular and cerebrovascular diseases and the development of modern medicine. We propose a topological distance-constrained feature descriptor learning model for vessel matching in coronary angiography. The topological distance was considered as the geodesic distance between patches, which should be small for the connected patches. Thus, the magnitude of the geodesic distance was used to reflect the connected relations between the patches. Additional topological distance constraints were applied to the network training by maximizing the feature differences between the connected and unconnected patches. Moreover, we verified, by varying the topological neighborhood values, that the proposed model achieved its best performance when the neighborhood value was 2, when the network could learn more useful potential features and exhibit a better generalization ability.

The topological distance-constrained feature descriptor learning model and its best neighborhood constraint model were tested on the 2D coronary angiographies dataset. The FPR95 values for the two models were 2.58, and 1.39, respectively. These values were lower than those of the other algorithms. The FPR95 of the best neighborhood constraint model was at its smallest when the neighborhood value was 2. Both algorithms had in the best matching accuracy results of all the algorithms. Specifically, the two-neighborhood topological constraint model could obtain more accurate matching results; it had the best matching performance. The abovementioned experimental results show that the proposed method were very effective for vessel matching tasks. This research provides a basis for future vessel registration, 3D reconstruction, and motion analysis. Better topological distance-constrained feature descriptors could be learned to improve the matching performance. Furthermore, the proposed model could be applied to other natural images with a similar topological structure to that of the vessel images.

Declaration of competing interest
We declare that we have no conflict of interest.

References


4 Cao H J, Shang X H, Qi H M. Virtual reality image technology assists training optimization of motion micro-time scale. IEEE Access, 2020, 8: 123215–123227 DOI:10.1109/access.2020.3007389


21 Mishchuk A, Mishkin D, Radenovic F, Matas J. Working hard to know your neighbor's margins: Local descriptor learning loss. 2017

DOI:10.1109/tpami.2014.2343235

DOI:10.1109/tgrs.2015.2441954

DOI:10.1109/tgrs.2018.2820040

DOI:10.1007/978-3-642-40811-3_23

DOI:10.1007/978-3-030-09284-5_91

DOI:10.1145/358669.358692

DOI:10.1006/cviu.1999.0832

DOI:10.5244/c.16.44

DOI:10.1109/iccv.2009.5459241

DOI:10.1007/978-3-540-45243-0_31

DOI:10.1109/tbme.2010.2042169

DOI:10.1109/cvpr.2017.410