Computerized Medical Imaging and Graphics 114 (2024) 102364

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Multi-task global optimization-based method for vascular landmark detection

Zimeng Tan [a], Jianjiang Feng [a,*], Wangsheng Lu [b], Yin Yin [b], Guangming Yang [b], Jie Zhou [a]

[a] Department of Automation, Tsinghua University, Beijing, China
[b] UnionStrong (Beijing) Technology Co.Ltd, Beijing, China

ARTICLE INFO

Keywords:
Deep learning
Global optimization
Multi-task network
Anatomical landmark detection
Vascular structure

ABSTRACT

Vascular landmark detection plays an important role in medical analysis and clinical treatment. However, due to the complex topology and similar local appearance around landmarks, the popular heatmap regression based methods always suffer from the landmark confusion problem. Vascular landmarks are connected by vascular segments and have special spatial correlations, which can be utilized for performance improvement. In this paper, we propose a multi-task global optimization-based framework for accurate and automatic vascular landmark detection. A multi-task deep learning network is exploited to accomplish landmark heatmap regression, vascular semantic segmentation, and orientation field regression simultaneously. The two auxiliary objectives are highly correlated with the heatmap regression task and help the network incorporate the structural prior knowledge. During inference, instead of performing a max-voting strategy, we propose a global optimization-based post-processing method for final landmark decision. The spatial relationships between neighboring landmarks are utilized explicitly to tackle the landmark confusion problem. We evaluated our method on a cerebral MRA dataset with 564 volumes, a cerebral CTA dataset with 510 volumes, and an aorta CTA dataset with 50 volumes. The experiments demonstrate that the proposed method is effective for vascular landmark localization and achieves state-of-the-art performance.

1. Introduction

Anatomical landmarks in vascular structures, which are distributed at the bifurcation positions and hierarchically divide the vessel into multiple morphological and functional units, play an important role in clinical diagnosis and treatment planning (Zheng et al., 2015). Anatomical landmark detection is also a prerequisite for subsequent medical image analysis tasks, such as vessel centerline extraction (Chen et al., 2020), segmentation initialization (Oktay et al., 2017), and image registration (Han et al., 2015; Almasi et al., 2018). Since manual annotation is always tedious and time-consuming (Elattar et al., 2016), robust and automatic landmark detection methods are meaningful for clinical usages.

Recently, deep learning-based methods have demonstrated superior performance in anatomical landmark detection due to their strong ability to learn task-oriented features (Zhang et al., 2017; Payer et al., 2019; Zhong et al., 2019; Qian et al., 2019; Ghesu et al., 2019; Alansary et al., 2019; Noothout et al., 2020; Lang et al., 2020; Al and Yun, 2020; Oh et al., 2020; Zeng et al., 2021; Xu et al., 2021; Chen et al., 2022; Laiz et al., 2023). However, in terms of vascular landmark detection, there still remain many challenges. Firstly, vascular structures around the same landmark may have various shapes (Fig. 1(c)), meanwhile different landmarks may have similar local appearances. The large intra-class variations and small inter-class differences bring landmark confusion problems. Taking the popular heatmap regression based methods (Payer et al., 2019; Zhong et al., 2019; Lang et al., 2020; Oh et al., 2020; Xu et al., 2021) as an example, the heatmap prediction of a certain landmark may have strong responses at several positions (Fig. 1(d)). These false positive responses may lead to mislocalization of the target landmark to other bifurcation positions and a very large detection error. Secondly, it is difficult to model the spatial distribution of landmarks due to the complex vascular topology. Physiological variations, such as loss of one or multiple cerebrovascular segments (Fig. 1(a)), and disease-related effects, such as vessel deformation caused by aneurysm and stenosis (Fig. 1(b)), may change the local appearance around landmarks, bringing further challenges. Moreover, the optimization process of deep learning networks relies on large-scale annotated dataset, while it is especially difficult to annotate vascular structures. The limited size of the training dataset severely limits the development of vascular landmark detection algorithms.

* Corresponding author.
E-mail address: jfeng@tsinghua.edu.cn (J. Feng).

https://doi.org/10.1016/j.compmedimag.2024.102364
Received 16 July 2023; Received in revised form 4 December 2023; Accepted 22 February 2024
Available online 1 March 2024

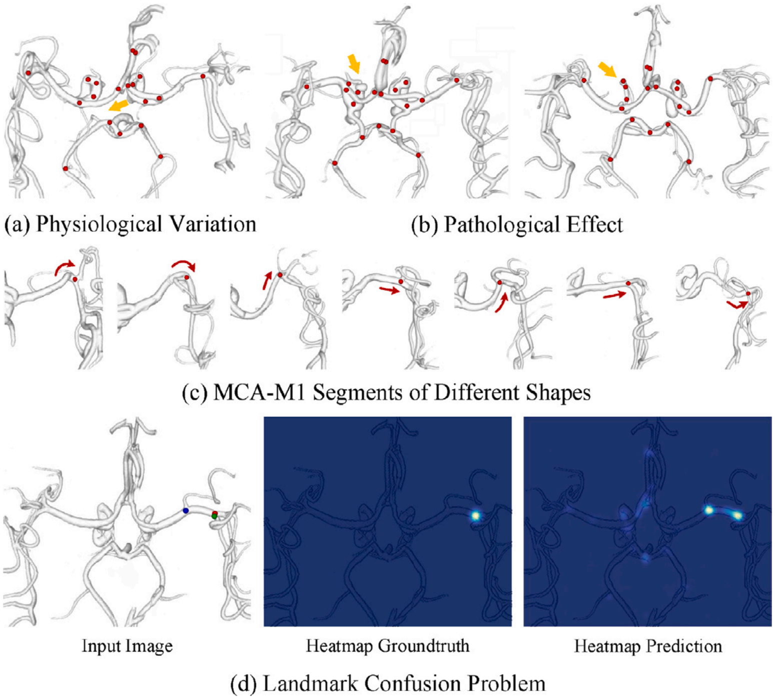

(a) Physiological Variation
(b) Pathological Effect
(c) MCA-M1 Segments of Different Shapes
Input Image
Heatmap Groundtruth
Heatmap Prediction
(d) Landmark Confusion Problem

Fig. 1. Illustration of vascular landmark detection challenges. (a) Physiological variation of missing left PCoA segment. (b) Pathological effects of aneurysm and stenosis on left ICA segments. (c) MCA-M1 segments of different shapes with arrows indicating the directions of vessel extension. (d) Landmark confusion problem exemplified by one landmark. The ground truth and predicted heatmaps are shown in 2D form using maximum intensity projection. The red, blue, and green dots denote the landmark ground truth, predicted positions of the max-voting strategy and the proposed method, respectively.

Different from organ landmarks such as skull (Chen et al., 2022) and prostate (Tuysuzoglu et al., 2018), vascular landmarks are connected by vascular segments, providing potential clues for localization. It is essential to incorporate structural prior information and spatial relationships between landmarks, which have not been fully explored in existing methods. The vascular network can be modeled as a graph model by taking landmarks and vascular segments as vertices and edges, respectively. In this way, the landmark spatial distribution and adjacent relationships are encoded and contribute to the detection performance.

In this paper, we present a novel framework for anatomical landmark detection in vascular structures, which consists of a multi-task learning network and a global optimization-based landmark decision strategy. The multi-task network is exploited for initial landmark heatmap regression, where vascular semantic segmentation and orientation field regression are introduced as auxiliary objectives to guide the backbone network to capture more discriminative features. Given a heatmap prediction, a general landmark inference method is the max-voting strategy (i.e., selecting the voxel with the highest temperature). However, disturbed by false positive responses, this strategy may misdetect the target landmark at other bifurcation locations (blue dot in Fig. 1(d)). To overcome this problem, a global optimization method is proposed for final landmark decision by evaluating the spatial relationships between pairs of landmarks explicitly (green dot in Fig. 1(d)). The proposed method considerably extends our previous work (Tan et al., 2021). In addition to a more detailed literature review, other major extensions in this work include (1) a more reasonable orientation field design along the direction of vessel extension, (2) a global optimization-based post-processing method to tackle the landmark confusion problem, (3) comprehensive evaluations with three different vascular landmark detection tasks, (4) a new metric to evaluate three types of topological detection errors automatically, and (5) further discussion of the detection performance in clinical applications.

Our contributions can be summarized as follows:

- We develop a deep learning framework specific for vascular landmark detection, combining a multi-task U-shape network and a global optimization based landmark decision algorithm, where the multi-task network accomplishes landmark heatmap regression, vascular semantic segmentation, and orientation field regression simultaneously.
- To overcome the landmark confusion problem, we propose a global optimization-based post-processing algorithm for landmark decision, where the structural prior and the spatial relationships between landmarks are incorporated explicitly.
- We present a new metric to evaluate three types of topological errors in vascular landmark detection, which reflects the anatomical rationality of landmark predictions and is relevant to clinical applications.
- We evaluated our method on three datasets with different vascular structures and imaging modalities. The experiments demonstrate that our method is effective for vascular landmark detection and achieves state-of-the-art performance.

2. Related work

2.1. Anatomical landmark detection

There have been numerous efforts for anatomical landmark detection in medical images over the past decades. Atlas-based (Isgum et al., 2009) and statistical shape models (SSMs)-based (Norajitra and Maier-Hein, 2017) methods perform non-rigid alignment between template images with landmark annotations and target images, which are limited by anatomical abnormalities and imaging artifacts. Random forests (RFs)-based methods (Criminisi et al., 2013; Gao and Shen, 2015; Urschler et al., 2018) estimate landmark locations using a regression-voting mechanism. These methods rely heavily on hand-crafted features and calculation processes, which are difficult to be optimized with training data.

Recently, deep learning-based methods have generated overwhelming enthusiasm and achieved remarkable progress in anatomical landmark detection. They can be broadly categorized into coordinate regression, heatmap regression, and deep reinforcement learning (DRL).

Coordinate regression based methods (Zhang et al., 2017; Qian et al., 2019; Noothout et al., 2020; Zeng et al., 2021) analyze local image patterns and predict absolute landmark coordinates or displacement vectors towards the target landmark. For example, Noothout et al. (2020) performed patch classification and displacement regression in parallel. Zeng et al. (2021) treated landmark detection task as a multi-level regression problem, where cascaded convolutional neural networks estimate the coarse positions of all landmarks simultaneously and refine each landmark independently. The limitation of these methods is that the mapping from image to coordinates involves highly nonlinear complexity and may lead to overfitting problem (Payer et al., 2019). Heatmap regression based methods (Payer et al., 2019; Zhong et al., 2019; Lang et al., 2020; Oh et al., 2020; Xu et al., 2021) have yielded state-of-the-art performance and become increasingly popular. They suggest generating a pseudo-saliency map for each landmark, then the position with the maximum response is chosen as the prediction. Compared with coordinate regression based approaches, these methods are intrinsically more suitable for landmark detection by changing the focus from the point of interest to the region of interest. Payer et al. (2019) introduced an end-to-end fully-convolutional network for heatmap regression and further refined the results by introducing a spatial configuration component to model the geometric relationships between landmarks. Xu et al. (2021) designed a dependency mining module and a local voting algorithm for hip landmark detection. More recently, Ao and Wu (2023) employed an encoder–decoder architecture named FARNet with a feature aggregation module for multi-scale feature fusion and a feature refinement module for high-resolution heatmap regression. Some studies combined these two methods. Chen et al. (2022) first performed heatmap regression on down-sampled volumes for coarse predictions, and then progressively refined the landmarks by attentive offset regression on multi-resolution patches. In contrast to the first two paradigms, DRL-based methods (Ghesu et al., 2019; Alansary et al., 2019; Al and Yun, 2020; Browning et al., 2021) detect landmarks by searching for an optimal path from an initial starting position towards the target landmark. Ghesu et al. (2019) adopted a deep RL-agent to navigate in the image space exploiting multi-scale image representations. Alansary et al. (2019) presented an extensive evaluation of several different RL-based models. However, multi-agent system has high computational complexity, making it difficult to locate multiple landmarks simultaneously.

A relevant problem to vascular landmark detection is anatomical labeling, which refers to dividing the centerline into several semantic categories according to landmarks. Most existing methods formulate the tubular structure in a graph representation and realize labeling using rule-based (Mori et al., 2005), geometric feature classification-based (Matsuzaki et al., 2015), or graph matching-based (Bogunović et al., 2013; Robben et al., 2016; Liu et al., 2022) approaches. These methods usually assume an existing centerline-plus-diameter model (Matsuzaki et al., 2015; Liu et al., 2022), or need to extract the tubular structure semi-automatically (Mori et al., 2005; Bogunović et al., 2013). An exception is the work of Robben et al. (2016), where the authors first computed an overcomplete segmentation, and optimized the centerlines and labels using integer programming. This is the biggest difference from the landmark detection task, which expects to localize landmarks directly from the input image without dependence on segmentation performance.

2.2. Structural prior modeling

How to model structural prior information and incorporate the relationships between landmarks is a key issue in anatomical landmark detection, which has drawn increasing attention in previous literature. Yang et al. (2017) interactively evolved the probability maps with the message passing schemes and refined the final predictions with sparsity regularization. Tuysuzoglu et al. (2018) estimated anatomical landmarks and contour of the prostate in parallel, making the algorithm more contextually aware. Aiming for vertebrae identification and localization, Liao et al. (2018) developed a multi-task bidirectional recurrent network to encode the long-range contextual information. Furthermore, Zhang et al. (2020) proposed to learn bone segmentation and landmark digitization jointly. A more similar work is the multi-task network for cerebral landmark detection (Tan et al., 2022), which associates vascular variations with local bifurcation appearance changes. The main difference between Tan et al. (2022) and this study is that the former can only be applied to cerebral landmark detection and does not take full advantage of the spatial relationships between landmarks.

More broadly, there are many tasks in computer vision domain similar to anatomical landmark detection, such as facial landmark localization (Zeng et al., 2017) and human pose estimation (Wang et al., 2020; Cao et al., 2021). For example, Zeng et al. (2017) modeled facial landmarks as a tree model and characterized the relationships in hierarchical order. Wang et al. (2020) proposed a graph pose refinement module for top-down human pose estimation. As a milestone innovation, Cao et al. (2021) utilized part affinity fields (PAFs) to encode the localization and orientation of limbs, which provides a new insight into the problem of anatomical landmark detection. We note that the human keypoints connected by limbs are similar to the anatomical landmarks connected by vascular segments, both of which can be regarded as a graph structure.

Inspired by Cao et al. (2021), we propose to employ orientation fields distributed on vascular segments to model the spatial relationships between neighboring landmarks. Moreover, vascular semantic segmentation is introduced to enhance the contextual information. During inference, different from the multi-task schemes mentioned above (Tuysuzoglu et al., 2018; Zhang et al., 2020; Tan et al., 2022), where the auxiliary tasks are only designed to guide the network to learn more discriminative features, in our method, both semantic segmentation and orientation field predictions participate in the landmark decision process. In this way, the landmark correlations can be considered explicitly, boosting localization performance.

3. Method

In this work, we provide a deep learning-based framework for vascular landmark detection. The overall workflow is presented in Fig. 2. Firstly, we employ a multi-task network for landmark heatmap regression, where vascular semantic segmentation and orientation field regression are introduced as auxiliary objectives. For each landmark, several candidate positions are generated by finding local maxima values in the heatmap prediction. Subsequently, a global optimization-based post-processing algorithm is proposed for final landmark decision. Besides, we introduce a new evaluation metric to calculate the topological errors in vascular landmark detection, which reflects the anatomical rationality of landmark predictions and is relevant to clinical applications.

3.1. Multi-task network

Recently, multi-task learning has shown promising performance by leveraging the synergy among associated tasks (Vandenhende et al., 2022) and is widely applied in many domains, including image classification (Kuang et al., 2017), semantic segmentation (Dai et al., 2016), and medical image analysis (Graham et al., 2019; He et al., 2020; Song et al., 2020; Zhou et al., 2021; Gende et al., 2022; Brenes et al., 2022; Choi et al., 2023). In this section, we propose a multi-task network for initial landmark detection, where heatmap regression, vascular semantic segmentation, and orientation field regression are learned in parallel. The goal is to learn a shared representation among different tasks and guide the network to capture more discriminative features.

The backbone of the proposed multi-task network is based on an improved 3D UNet (Ronneberger et al., 2015), following the encoder–decoder architecture. The detailed structure of the network is shown

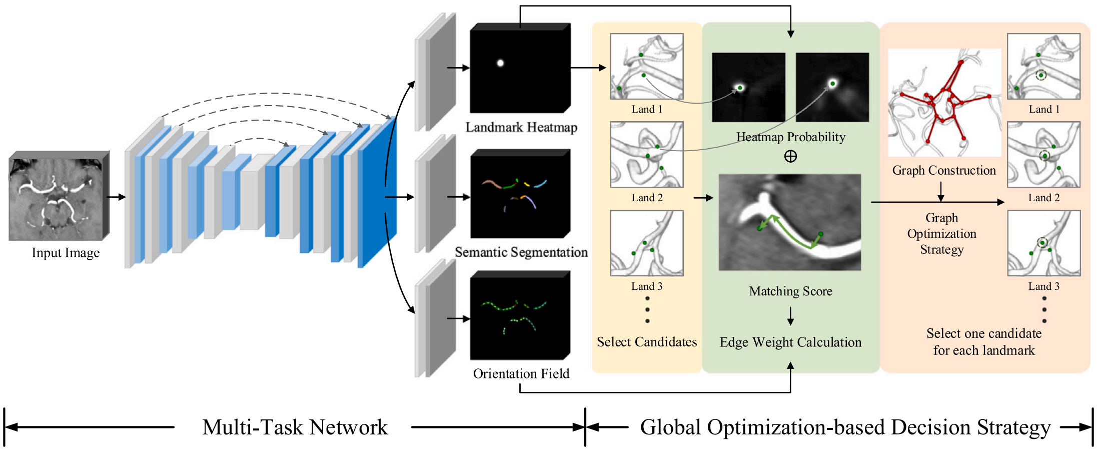


Fig. 2. Illustration of the proposed framework for vascular landmark detection, which contains a multi-task network and a global optimization-based landmark decision strategy.

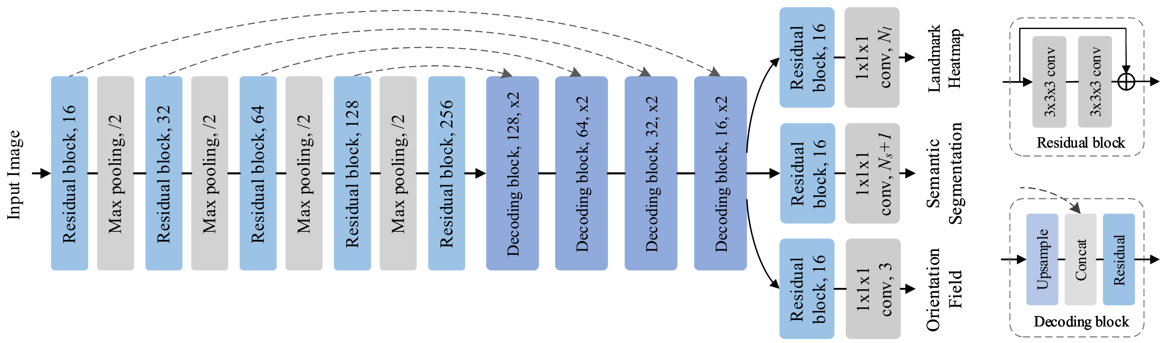


Fig. 3. Architecture of the proposed multi-task network, which accomplishes landmark heatmap regression, vascular semantic segmentation, and orientation field regression simultaneously.

in Fig. 3. The encoder module consists of multiple residual blocks (He et al., 2016) to generate increasingly abstract feature representations, with each block followed by a 2 × 2 × 2 max pooling layer with the stride 2 for downsampling. Each residual block contains two 3 × 3 × 3 convolutional layers with batch-normalization operators (Ioffe and Szegedy, 2015) and ReLU activations (Nair and Hinton, 2010), while a shortcut connection is applied between the input and output to avoid gradient vanishing problem. The decoder module contains several decoding blocks, roughly symmetrical to the residual blocks in the encoder. In each decoding block, the input is first fed to a deconvolutional layer. Then, the upsampled intermediate features are concatenated with the corresponding ones from the encoder module and processed using a residual block. These skip connections between the encoder and decoder modules incorporate low-level fine features with high-level abstract features, preserving more spatial details for better localization. The number of filters starts with 16, which doubles at each residual block and halves at each decoding block. After the decoder module, the network is split into three branches to accomplish tasks simultaneously. Each branch contains a same residual block as the encoder module to capture task-specific features and a 1 × 1 × 1 convolutional layer to generate pixel-wise prediction. A softmax activation is applied only for the semantic segmentation branch.

Following the encoder–decoder architecture, the backbone network is adapted from the widely used 3D U-Net (Ronneberger et al., 2015). We replace the plain convolutional layer with a residual block (He et al., 2016) to avoid the gradient vanishing problem, which contains two convolution operators and a shortcut connection between the input and output. Skip connections between the encoder and decoder modules incorporate low-level fine features with high-level abstract features, preserving more spatial details for better localization. The network is then split into three branches to accomplish tasks simultaneously. The details of each task are described as follows.

3.1.1. Landmark heatmap regression

In the first branch, inspired by Payer et al. (2019), we convert the landmark detection problem into a heatmap regression task. The discrete coordinates of a landmark are modeled as a channel heatmap with a Gaussian distribution centered at the landmark position. The heatmap temperature $H_i^*(x)$ on voxel x ranging in [0,1] represents the probability to be the ith landmark. The distribution is determined according to the distance from voxel x to landmark position x_i, with the standard deviation δ controlling the size. Formally, the multi-channel heatmap for N_l landmarks is defined as follows:

$$H_i^*(x) = e^{-\frac{1}{2\delta^2}(x-x_i)^2}, i = 1, 2, \ldots, N_l. \tag{1}$$

In order to deal with the class imbalance problem, we apply a weighted L2 loss function $\mathcal{L}_{heat}$ between the predicted heatmaps H_i and the ground truth heatmaps H_i^*. The weights are set to be the exponential powers of the ground truth volume.

3.1.2. Vascular semantic segmentation

In the second branch, the network is responsible for predicting the vascular semantic segmentation. According to the anatomical topology, an artery can be hierarchically decomposed into multiple morphological and functional units. These vascular segments are viewed as different semantic classes. To prepare the semantic segmentation ground truth, the entire vascular structure is divided into different vascular segments according to the corresponding bifurcation landmarks. That is, the landmark is located at the interface of two adjacent substructures, which makes the semantic segmentation task highly correlated with the landmark detection objective. In this way, we enhance contextual information and provide richer feature representations for modeling structural prior knowledge.

Specifically, the vascular semantic segmentation task is regarded as a multi-channel voxel-wise classification problem. The output contains $N_s + 1$ channels, where the first N_s channels correspond to the N_s vascular segment classes and the last one belongs to the background. We train this network branch using a weighted Dice loss function $\mathcal{L}_{seg}$ to deal with the class imbalance problem. During inference, we obtain the final semantic segmentation prediction by assigning each voxel the class with the highest probability.

3.1.3. Orientation field regression

In the third branch, taking inspiration from Cao et al. (2021), we employ 3D orientation fields to model the spatial relationships between neighboring landmarks explicitly. Considering that paired landmarks are connected by vascular segments, the orientation field is constructed based on the vascular semantic segmentation. In the preliminary conference publication (Tan et al., 2021), the orientation field was simply defined as a set of unit vectors in the hierarchical structure pointing from the upper to the lower bifurcation landmarks. However, we argue that this may be unreasonable especially for long and curved vessels, since it is difficult for stacked convolutional layers to capture long-range dependency due to the limited receptive field.

More intuitively, we define the orientation field as the tangential direction of the lumen centerline. To prepare the orientation field ground truth, the centerline of each vascular segment is extracted by performing skeletonization algorithm on the semantic segmentation annotation. The lumen slope of each point on the centerline is then set as the 3D unit vector between the voxel position and the adjacent centerline point. In the lumen segmentation, the orientation field values of the voxels on each section perpendicular to the centerline are set to be the same as the corresponding centerline sampling point. For voxels belonging to the background, the vector is zero-valued. In this way, the orientation field changes slowly along the direction of vessel extension, which is easier to learn based on the local image pattern. In order to reduce memory occupation, we combine the vector fields of all vascular segments into a single 3D volume, that is, the output has three channels corresponding to the x, y, and z coordinate axes. Assisted by semantic segmentation and orientation field regression, the multi-task network preserves both location and orientation information across the vascular structure.

Similarly, a weighted L2 loss function $\mathcal{L}_{ori}$ is applied for this branch. The final loss function is formulated by the linear combination of all losses:

$$\mathcal{L} = \mathcal{L}_{heat} + \alpha\mathcal{L}_{seg} + \beta\mathcal{L}_{ori}, \tag{2}$$

where the hyperparameters α and β are dynamically adjusted to make the different components having the same scale after the training stabilization.

3.2. Global optimization-based decision strategy

Given a heatmap prediction, a general landmark inference method is to perform max-voting or weighted-voting strategy. However, this scheme mainly relies on local appearance, and the spatial relationships among landmarks are not taken into consideration. In vascular landmark detection, false positive responses at other landmark positions are frequent and may lead to misdetection of the target landmark. In this paper, we propose a global optimization-based post-processing method, which combines the predictions of vascular semantic segmentation and orientation field regression to select the most anatomically reasonable landmark configuration. Overall, the proposed landmark decision strategy comprises three main steps: (1) graph construction, (2) edge weight calculation, and (3) global optimization. The schematic diagram is illustrated in Fig. 4.

3.2.1. Graph construction

Anatomical landmarks are located at bifurcation positions of the vascular structure, dividing the vessel into multiple segments with unique physiological functions. According to the anatomical topology, we model the vascular structure as an undirected graph $G(V, E)$, where the vertex set $V = \{v_i\}_{i=1}^{N_l}$ and edge set $E = \{e_i\}_{i=1}^{N_s}$ are defined as landmarks and vascular segments, respectively. If a pair of landmarks are located at the ends of a vascular segment, they are connected in the graph. Considering that the correct landmark location is often included in the highlighted response area of the heatmap prediction but may not have the highest temperature, for each landmark, we obtain n candidate points $v_i = \{v_i^1, v_i^2, \dots, v_i^n\}$ by finding local maxima values in the corresponding heatmap prediction. In this way, there are n^2 candidate edges between two types of candidate points in pairs. Particularly, the edges in the graph G is automatically adjusted according to the semantic segmentation result to deal with the situations where some vascular segments may not exist (e.g., physiological variations of cerebral vessels). If the number of voxels belonging to a certain vascular segment in the semantic segmentation result is less than a threshold (set to 5 in the experiments), the corresponding edge is removed from G.

During inference, a graph prediction G (i.e., a predicted landmark configuration) can be formed by selecting one candidate point for each type of landmarks. The key insight of the decision strategy is to find an optimal graph G^* conforming the anatomical prior by calculating the edge weights.

3.2.2. Edge weight calculation

To evaluate the reasonability of a set of candidate points, both local appearance and anatomical prior information need to be taken into consideration. Benefiting from the strong ability of convolutional neural networks (CNNs) to learn feature representations, the heatmap prediction reflects the confidence of the local appearance around each voxel position. To incorporate the structural prior knowledge, we define *matching score* between paired landmarks. Therefore, the edge weight in the constructed graph G consists of vertex confidence and edge matching score. Note that we do not define the vertex weight separately for simplicity.

Specifically, as shown in Fig. 4(b), consider two candidate points A and B belonging to the endpoints of cerebrovascular segment MCA-M1. Let A_{gt} and B_{gt} be the landmark ground truth positions. Given the outputs of the multi-task network, the local orientation field on MCA-M1 segment is obtained from the overall orientation field prediction filtered by the corresponding semantic segmentation result. The pseudo centerline is extracted from the segmentation using skeletonization algorithm. Then we find two points A' and B' on the pseudo centerline closest to A and B respectively, and a path $A' \to B'$ from point A' to B' along the centerline. Connect point A and A', B' and B to obtain a path from point A to B:

$$A \to B = (A \to A') + (A' \to B') + (B' \to B). \tag{3}$$

Subsequently, we sample uniformly on the path $A \to B$ and calculate the matching score point by point. If a sampling point p lies on the centerline (i.e., $p \in (A' \to B')$), the sampling point is defined as a vector product; for other sampling points, the matching score is defined as a penalty term. Formally, we define the matching score S_p at an sampling point p as:

$$S_p = \begin{cases} d_{AB} \cdot O(p), & \text{if } p \in (A' \to B') \\ -\tau, & \text{otherwise} \end{cases} \tag{4}$$

where d_{AB} denotes the unit vector from point A to B and $O(p)$ is the predicted vector at voxel p in the orientation field. $\tau(\tau > 0)$ is the penalty coefficient.

The matching score of point A and B is given by:

$$\begin{aligned} S_{AB} &= S_{AA'} + S_{A'B'} + S_{B'B} \\ &= \sum_{p \in (A' \to B')} d_{AB} \cdot O(p) - \tau(l_{AA'} + l_{B'B}), \end{aligned} \tag{5}$$

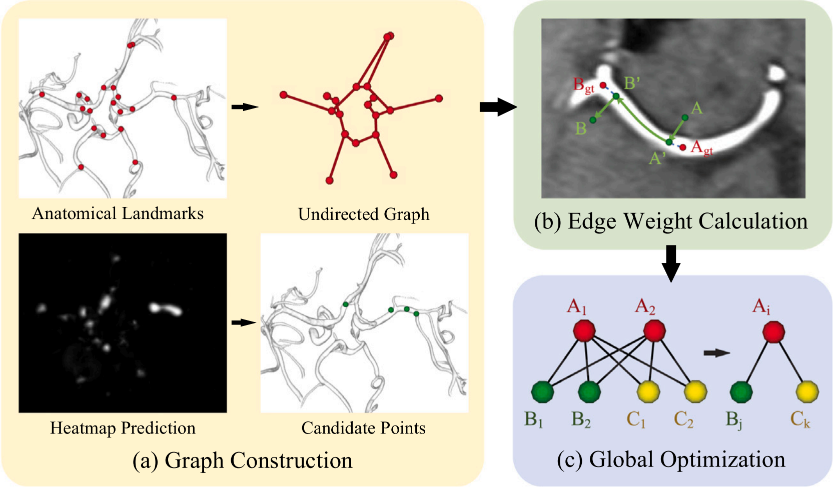


Fig. 4. Schematic diagram of the proposed global optimization-based landmark decision strategy.

where $l_{AA'}$ and $l_{B'B}$ denote the number of sampling points of path $A \rightarrow A'$ and $B' \rightarrow B$ respectively, proportional to the path length.

The vertex confidence is chosen as the predicted heatmap probability $H(A)$ and $H(B)$ on the candidate points. In this way, the weight of the edge connecting point A and B is defined as the weighted sum of the matching score and the vertex confidence, with γ controlling the trade-off:

$$W_{AB} = \frac{H(A) + H(B)}{2} + \gamma \cdot S_{AB}. \tag{6}$$

We emphasize that the essence of matching score is to simultaneously constrain the spatial relationships between adjacent landmarks and the degree of deviation from the vascular segment region. The first term in Eq. (5) is the discrete form of the line integral of the vector product along the path $A' \rightarrow B'$, which constrains the relative orientation and distance between paired candidate points to be consistent with the extension direction and length of the corresponding vascular segment. The second term penalizes candidate points deviating from the vascular segment. The predicted position is considered unacceptable if it is located on the surrounding tissue.

Additionally, perfect semantic segmentation prediction is not necessary in our method, since only a pathway in the vascular segment is required instead of a strict centerline. In other words, the network is supposed to detect the main vascular region. The vascular boundary does not need to be finely segmented, which is more challenging due to poor contrast, irregular shape, and varying noise. This also avoids the misleading of suboptimal semantic segmentation results in the landmark decision strategy.

3.2.3. *Global optimization*

The edge weights measure the correlations between candidate points belonging to paired landmarks. Given the constructed graph and the calculated edge weights, the landmark detection task is transformed into selecting a candidate position for each landmark such that the constructed graph G^* has the maximum sum of edge weights. More formally, consider a graph $G(V, E)$ with N_l vertices $V = \{v_i\}_{i=1}^{N_l}$ and N_s edges $E = \{e_i\}_{i=1}^{N_s}$. Let each landmark has n candidate points $v_i = \{v_i^1, v_i^2, \ldots, v_i^n\}, v_i \in V$. For a pair of candidate point $(v_i^{c_i}, v_j^{c_j}), c_i, c_j \in [1, n]$ belonging to paired landmarks (v_i, v_j), let the corresponding edge weight be $W_{v_i^{c_i} v_j^{c_j}}$. Then, the optimization objective can be written as:

$$\max_{c_i, c_j \in [1,n]} \sum\nolimits_{v_i, v_j \in P} W_{v_i^{c_i} v_j^{c_j}}, \tag{7}$$

where P denotes the set of paired landmarks connected by vascular segments. Fig. 4(c) illustrates a simple case of $N_l = 3, N_s = 2, n = 2$ as an example, which can be easily generalized to complex graph structures. In this way, the spatial relationships between landmarks and global structural prior are incorporated explicitly, while the anatomically implausible landmark configurations due to the false positive responses are suppressed. A basic solution for this optimization problem is to traverse all candidate point combinations, with exponential time complexity (i.e., for N_l landmarks and n candidate points per landmark, the number of iterations required is n^{N_l}). We use the Markov Random Field (MRF) model (Li, 2009) to reduce the time complexity to be linear with the number of edges (Schwarz et al., 2012).

It is noteworthy that in the proposed post-processing method, only the correlations between pairs of landmarks connected by vascular segments are taken into consideration, since there is no obvious spatial dependence between landmarks far away. Our method has no restrictions on the vascular topology and can be applied to different tubular structures. For example, for “isolated” landmark without edge connection in the graph (e.g., aortic landmark 2 and 3 in Fig. 5(c)), only the vertex confidence participates in edge weight calculation. Furthermore, the hyperparameters can be flexibly adjusted in practical applications. For example, the penalty coefficient τ can be increased appropriately to encourage the predicted landmarks located within the vascular region.

4. Experiments and results

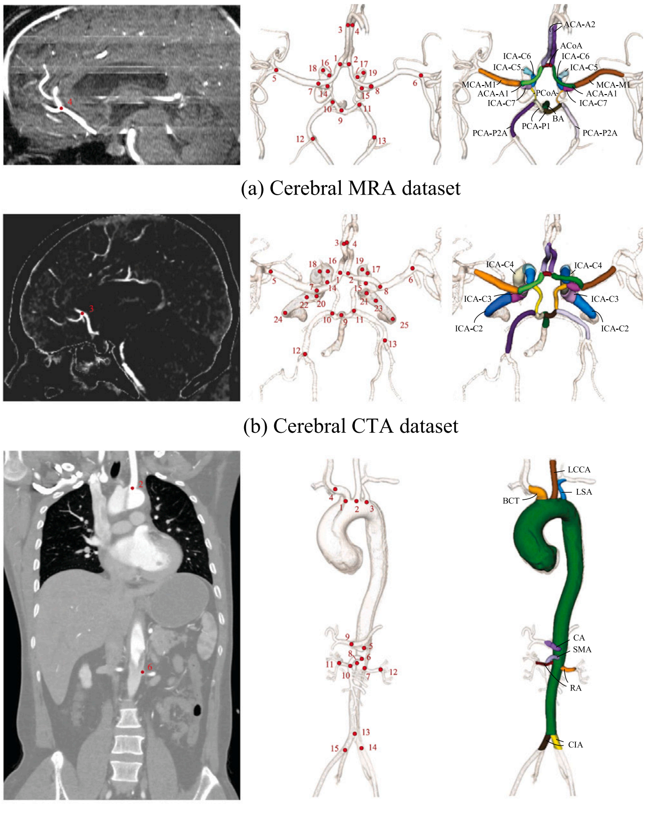
We evaluated our method on three 3D volume datasets with different vascular structures and imaging modalities (Fig. 5). The experimental results show that our approach achieves superior performance compared to state-of-the-art methods. Furthermore, we conducted an ablation study to investigate the contribution of each component in the proposed framework.

4.1. *Datasets*

4.1.1. *Cerebral MRA dataset*

The cerebral magnetic resonance angiography (MRA) datasets consist of a public part and a private part. The public dataset contains 104 scans of healthy volunteers selected from the UNC dataset[1] (for a total of 109 scans), where the samples with incomplete ICA-C5 segments due to limited scanning range (2 scans) and deletion variation of unilateral ACA-A segment (3 scans) were excluded. These two situations will lead to the missing of corresponding landmarks. The private dataset contains 460 scans collected clinically with aneurysms or stenosis. For

[1] https://public.kitware.com/Wiki/TubeTK/Data.



(a) Cerebral MRA dataset

(b) Cerebral CTA dataset

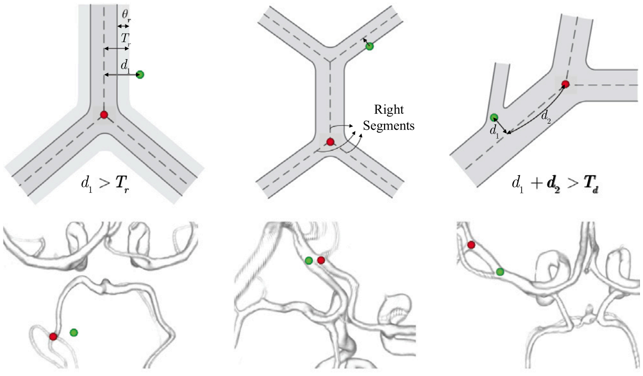
(c) Aorta CTA dataset

Fig. 5. Example images in three datasets with anatomical landmark and semantic segmentation annotations.

each volume, nineteen clinically relevant cerebrovascular landmarks and binary segmentation around the Circle of Willis (CoW) region were annotated manually by one of the authors and verified by an experienced neurosurgeon. The landmark annotations of the public dataset have been released by us.[2] The annotation rules refer to Bradac (2014). Then the semantic segmentation label containing 20 classes was established based on the binary segmentation and landmarks, followed by manual correction. The semantic segmentation generation was performed on 104 public and 40 private scans. The network was trained and tested on the public and private datasets separately to enable fair comparisons on the public part. The public dataset was randomly divided into 70 training scans, 7 validation scans, and 27 test scans. For the private dataset, 10 and 150 scans with only landmark annotations were selected randomly as the validation and the test sets respectively, while the remaining 300 data composed the training set. All scans were spatially normalized to 0.513 mm × 0.513 mm × 0.8 mm firstly, and intensity-based rigid registration was performed by taking a training sample as the template. Then, the scans were automatically cropped to 192 × 160 × 96 according to the mean landmark distribution.

4.1.2. Cerebral CTA dataset

Computed tomography angiography (CTA) is another common imaging modality used in cerebrovascular examination and treatment. The cerebral CTA dataset consists of 510 scans collected clinically with acute ischemic stroke. Twenty-five landmarks were annotated manually for each scan. Note that there are 6 landmarks defined only on the CTA modality (landmark 20–25), since the public MRA images do not cover the complete ICA vascular segments. We annotated 60 scans with binary segmentation around the CoW region, and the cerebrovascular structure was divided into 26 semantic classes. All images were registered rigidly to one of training samples based on intensity and cropped to 208 × 160 × 144 with voxel spacing as 0.488 mm × 0.488 mm × 0.625 mm. We randomly selected 10 and 150 scans with only landmark annotations as the validation and the test sets respectively, while the remaining 350 scans were used for training.



(a) Type I Error

(b) Type II Error

(c) Type III Error

Fig. 6. Illustration of three types of topological detection errors and corresponding test samples, where the red and green dots denote the landmark ground truth and predicted positions, respectively.

4.1.3. Aorta CTA dataset

The aorta CTA dataset contains 50 scans with isotropic voxel spacing of 1 mm. All patients have severe aortic dissection, where the torn aortic intima brings a challenge to landmark detection. Fifteen landmarks and binary segmentation were annotated manually by a radiologist, then the semantic segmentation label containing 10 segments was established accordingly. Two scans were selected randomly as the validation set, and four-fold cross-validation was performed on the remaining 48 images. We did not apply rigid registration due to the large individual variations in vessel size. According to the average distribution of landmarks in the training set, we divided the images into three subregions along the longitudinal axis, and the models were trained for each subregion separately. The data preprocessing operations were repeated for each fold of validation.

4.2. Evaluation metrics

We evaluate the performance of vascular landmark detection with two metrics commonly used in the literature: Mean Radial Error (MRE) and Successful Detection Rate (SDR). The MRE calculates the Euclidean distance (in millimeter) between the predicted and ground truth landmark locations. The associated standard deviation (SD) is also reported. The SDR measures the successful detection percentage, where a landmark with the detection error within the predefined precision threshold is regarded to be detected successfully. In our experiments, we used four precision thresholds (2 mm, 3 mm, 4 mm, and 5 mm) for all datasets.

In addition, we propose a novel evaluation metric to calculate the topological errors in vascular landmark detection, which refers to the predictions that result in significant deformations, deviations, or distortions of the vascular topology. Empirically, the topological errors can be classified into three types (see the second row of Fig. 6): (I) detections outside vessels (e.g., on bones or organ tissues), (II) detections on the wrong vascular segments, and (III) detections at other locations of the correct vascular segment (e.g., on other bifurcations). These three types of topological errors are mutually exclusive and clinically unacceptable,

[2] http://ivg.au.tsinghua.edu.cn/dataset/Cerebral-MRA/Cerebral-MRA.html.

which can be misleading for subsequent applications, such as vascular labeling. We regard the semantic centerline as the correct vascular topology, then the topological errors can be quantitatively evaluated by calculating the deviation of landmark predictions from the centerline. If the prediction is too far from the centerline, it is considered to fall outside the vascular region (type I error). If the prediction falls near the centerline part corresponding to other vascular segments, it is mislocated (type II error). If the prediction falls near the correct centerline part but is off the correct position, it may be located at other bifurcations that cannot accurately describe the distribution of the target segment (type III error).

Specifically, we first generate a pseudo-semantic centerline from the original image using landmark annotations by the minimum cost path algorithm. The centerline of each segment is obtained by finding the optimal path between the two corresponding landmarks. Part of the centerlines needs to be extended manually to consider the farther bifurcation points (required only in metric calculation). Then, we find the nearest point on the centerline for each landmark prediction. If the distance d_1 from the landmark to the centerline is greater than the radius threshold T_r, the prediction is classified as a type I error (Fig. 6(a)). Otherwise, if the closest point is on the wrong vascular segment, the prediction is classified as a type II error (Fig. 6(b)). For trifurcation landmarks, the three adjacent vascular segments are all correct segments. If the closest point is on the right vascular segment, calculate the sum of the distances of the landmark prediction perpendicular to the centerline d_1 and along the centerline d_2. If the sum is greater than the distance threshold T_d, the prediction is classified as a type III error (Fig. 6(c)), otherwise the prediction is topologically correct. The radius threshold T_r for each vascular segment is defined as the average radius of the segmentation annotations with a margin θ_r. Considering the different vessel sizes, θ_r was set to 3 mm for cerebral vessel and 4.5 mm for aorta empirically. T_d was set to 4.5 mm for cerebral vessel and 8.5 mm for aorta. We report the topological error rates (ERs, in %) for all experiments. The ER is the number of each type of errors divided by the total number of landmarks, where the latter is the product of the number of predefined landmarks and test samples.

Furthermore, we report the average time cost of the data preprocessing and landmark detection methods (in second).

4.3. Implementation details

The proposed framework was implemented in PyTorch on an NVIDIA GeForce RTX 3090 GPU. During training, the backbone network was first trained with only vascular semantic segmentation branch, then the heatmap regression and orientation field regression branches were added and trained jointly. The multi-task network was trained for around 500 epochs using an Adam optimizer ($\beta_1 = 0.5, \beta_2 = 0.999$) with a learning rate of 0.0001.

For the number of candidate points n, a large n provides a higher probability of containing the right locations, but introduces computational burden and potential error interference. During inference, we empirically chose n as 4 for all experiments. The hyperparameters in Eqs. (4) and (6) were selected on the validation sets and vary with the datasets. The effect of hyperparameter settings on detection performance is discussed in ablation experiments. The structure of graph G follows the anatomical topological definition. In particular, for the aorta dataset, the trunk was not considered in G, since the spatial correlations are mainly reflected on the nine branches.

4.4. State-of-the-art comparison

To evaluate the proposed method, we compared it with several state-of-the-art methods for anatomical landmark detection: (1) heatmap regression-based methods SCN (Payer et al., 2019) and FARNet (Ao and Wu, 2023), (2) coordinate regression-based methods (Noothout et al., 2020; Zeng et al., 2021), (3) a reinforcement learning-based method DQN (Alansary et al., 2019), and (4) a deep learning method combining heatmap and coordinate regression architectures SA-LSTM (Chen et al., 2022). It is worth noting that these methods can solve the landmark confusion problem to a certain extent. SCN (Payer et al., 2019) introduced a spatial configuration component to model the landmark spatial distribution. FARNet (Ao and Wu, 2023) suggested aggregating multi-scale features and applying coarse-to-fine supervisions. Noothout et al. (2020) and Zeng et al. (2021) and SA-LSTM (Chen et al., 2022) employed multi-stage frameworks and performed global-to-local estimation of landmark localization. DQN (Alansary et al., 2019) proposed a deep Q-network based model with novel hierarchical action steps. For SCN (Payer et al., 2019), DQN (Alansary et al., 2019), SA-LSTM (Chen et al., 2022), and FARNet (Ao and Wu, 2023), we adapted the codes made publicly available by the authors,[3],[4],[5],[6] following the default settings. For DQN (Alansary et al., 2019), we split the predefined landmarks into several groups and detected 2–3 landmarks at a time while the default is 2. We reimplemented the methods of Noothout et al. (2020) and Zeng et al. (2021) since there is no public implementation. The performance of the modified U-Net (Ronneberger et al., 2015) (the backbone network with only heatmap regression branch) is also investigated. We note that the proposed post-processing algorithm relies on predictions of the multi-task network, and thus was not performed on these comparison methods. The same data preprocessing steps were performed for all methods.

4.4.1. Public cerebral MRA dataset

For the public cerebral MRA dataset, the results on the test set evaluated by MRE, associated SD, SDR, and ER are listed in Table 1. Our method achieves an average accuracy of 1.27 ± 0.48 mm, which shows significant improvements by 0.48 mm (27% reduction), 0.82 mm (39% reduction), 0.81 mm (39% reduction), 0.58 mm (31% reduction), 0.4 mm (24% reduction), 0.42 mm (25% reduction) over SCN, DQN, Noothout et al. (2020) and Zeng et al. (2021), SA-LSTM, and FARNet, respectively. The methods in DQN, Noothout et al. (2020) and Zeng et al. (2021), and SA-LSTM detect landmarks in a coarse-to-fine manner, where the coarse stages prevent outlier predictions, resulting in a smaller SD of MRE and higher SDR within large distance thresholds. However, due to the highly nonlinear complexity involved in the mapping from image to landmark locations, it is not trivial to directly regress the coordinates or residuals at the refinement stages. In contrast, the heatmap-based methods (U-Net, SCN, and FARNet) focus on the local features of each voxel, which may result in landmark confusion problem due to similar local appearances, leading to significant localization errors in some cases. Compared with these approaches, our proposed method maintains superior and robust performance in SDR within various precision thresholds.

For topological detection errors, the methods based on heatmap regression (U-Net, SCN, FARNet, and ours) hardly show type I errors benefiting from the high contrast of vessels in the MRA images. Compared with U-Net, previous state-of-the-art methods show fewer type II errors and more type III errors, which indicates that their predictions are affected by redundant vascular branches or curvature mutations without sufficient structural prior knowledge, such as the length and extension direction of the vascular segments. In contrast, our multi-task learning network (MTN) significantly corrects the topological errors of types II and III, while the post-processing algorithm further improves the landmark detection performance.

The MREs for each landmark are shown in Fig. 7. Some landmarks are intrinsically more challenging due to variable vessel shape and

[3] https://github.com/christianpayer.

[4] https://github.com/amiralansary/rl-medical.

[5] https://github.com/runnanchen/SA-LSTM-3D-Landmark-Detection.

[6] https://github.com/JuvenileInWind/Farnet/.

Table 1
Quantitative comparisons with state-of-the-art methods on the public cerebral MRA dataset. The results were evaluated by MRE, associated SD, SDRs in four distance ranges, and ERs of three error types, with the best performance shown in bold. “MTN” stands for the proposed multi-task network.

Method	MRE (SD) (mm, ↓)	SDR (%, ↑)				ER (%, ↓)		
		2 mm	3 mm	4 mm	5 mm	I	II	III
U-Net (Ronneberger et al., 2015)	3.29 (1.82)	81.48	84.60	87.72	89.08	0.00	7.99	9.36
SCN (Payer et al., 2019)	1.75 (0.50)	75.44	84.41	90.25	93.57	0.00	2.14	9.94
DQN (Alansary et al., 2019)	2.09 (0.44)	63.74	80.90	90.25	92.79	1.17	1.75	12.67
Noothout et al. (2020)	2.08 (0.40)	57.89	80.31	92.98	95.13	3.70	2.34	8.77
Zeng et al. (2021)	1.85 (**0.33**)	67.45	86.35	93.37	**95.91**	2.14	1.95	7.99
SA-LSTM (Chen et al., 2022)	1.67 (0.35)	75.05	86.74	92.98	95.52	1.17	1.17	8.19
FARNet (Ao and Wu, 2023)	1.69 (1.11)	85.38	89.67	91.62	93.76	0.19	2.34	7.02
MTN	1.73 (0.90)	86.16	90.06	91.81	93.76	0.00	2.34	6.04
Proposed	**1.27** (0.48)	**87.72**	**92.01**	**93.57**	95.71	**0.00**	**1.17**	**5.26**

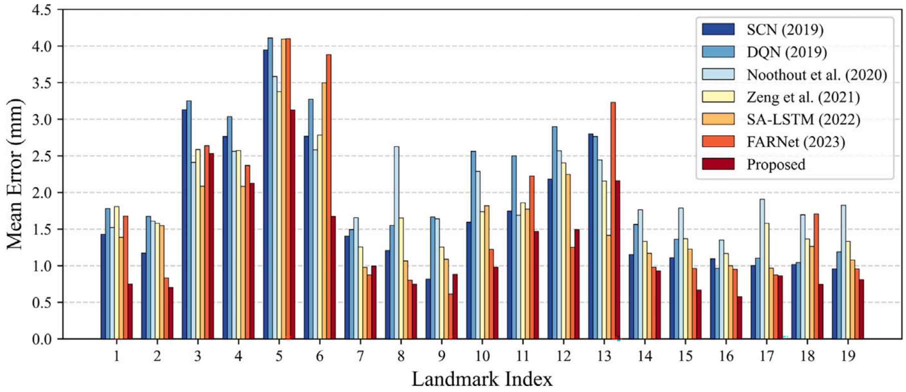


Fig. 7. Mean radial errors (MREs, in mm) for nineteen landmarks in the public cerebral MRA dataset. See Fig. 5(a) for the meaning of landmark index.

interference of additional branches, such as the bifurcation landmarks between MCA-M1, M2 segments (landmark 5 and 6), and between PCA-P2 A, P2P segments (landmark 12 and 13). Symmetric landmarks have similar detection difficulty in theory, but exhibit different detection errors, which may be caused by the limited test set. On larger test sets (e.g., the private cerebral MRA dataset and the cerebral CTA dataset), symmetric landmarks have similar detection performance (see Figs. 8 and 9).

A typical sample is illustrated in Fig. 11(a). It can be observed that our method suppresses the anatomically unreasonable predictions and solves the landmark confusion problem effectively. On average, the data preprocessing times per scan (including resize, registration, and crop operations) for all methods was 9.82 ± 2.80 s, varying with the original data size. The average inference times of each method are listed in Table 5. The inference time of our method is 3.11 ± 0.47 s, where the post-processing stage is the most time-consuming part (2.60 ± 0.22 s) due to the complex edge weight calculation process.

4.4.2. Private cerebral MRA dataset

The private dataset contains 40 scans with semantic segmentation and landmark annotations (called full annotation) and 420 scans with only landmark annotations. 150 and 10 scans were randomly selected from the latter as the test and validation sets respectively, consistent across all experiments. It is noteworthy that the proposed method requires landmark and semantic segmentation annotations simultaneously, while the comparison methods (SCN, DQN, Noothout et al. (2020) and Zeng et al. (2021), SA-LSTM, FARNet) rely only on landmark annotations. Considering that semantic segmentation annotation is more time-consuming and labor-intensive, we trained the proposed method with 200 scans (40 scans with full annotations and 160 scans with only landmark annotations) and the other three methods with all remaining 300 scans for fair comparison. The multi-task network was optimized using $\mathcal{L}$ for the fully annotated scans and $\mathcal{L}_{heat}$ for the data with only landmark annotations. The experimental results on the test set are shown in Table 2 and Fig. 8. It can be observed that our method obtains the lowest detection error of 1.36 ± 0.53 mm (25% reduction than SCN, 38% reduction than DQN, 33% reduction than Noothout et al. (2020), 28% reduction than Zeng et al. (2021), 16% reduction than SA-LSTM, and 21% reduction than FARNet. Our method also achieves higher SDRs especially for small precision thresholds and the lowest ERs for the three types of topological errors. Typical detection results are illustrated in Fig. 11(b).

Moreover, we compared our method trained using the whole training set and only 40 fully annotated scans respectively. As shown in Table 2, larger training dataset reduces the MRE by 0.41 mm (23% reduction) and increases the SDR within 2 mm by 6.63%, which indicates that our method does not rely on a large number of semantic segmentation annotations, and the detection performance can be significantly improved by expanding training dataset with only landmark annotations. The average data preprocessing time and inference times per scan are shown in Table 5.

4.4.3. Cerebral CTA dataset

For the cerebral CTA dataset, we randomly selected 10 and 150 scans with only landmark annotations as the validation and test sets, respectively. Similar to the private MRA dataset, we trained the proposed method with 60 fully annotated scans and 140 scans with only landmark annotations, while 350 scans were used in the comparison methods (SCN, DQN, Noothout et al. (2020) and Zeng et al. (2021), SA-LSTM, FARNet). The results on the test set are presented in Table 3 and Fig. 9. Compared with MRA scans, the cerebral CTA dataset is more challenging due to widespread noise interference, resulting in more serious landmark confusion problems. Therefore, the proposed multi-task configuration and post-processing scheme achieve greater performance

Table 2
Quantitative experiments on the private cerebral MRA dataset, with the best performance shown in bold. "MTN" stands for the multi-task network.

Method	MRE (SD) (mm, ↓)	SDR (%, ↑)				ER (%, ↓)		
		2 mm	3 mm	4 mm	5 mm	I	II	III
U-Net (Ronneberger et al., 2015)	2.86 (1.97)	82.60	87.82	90.74	92.32	0.00	7.68	7.93
SCN (Payer et al., 2019)	1.82 (0.53)	75.79	84.21	89.68	92.95	0.00	3.33	9.96
DQN (Alansary et al., 2019)	2.18 (0.50)	63.12	80.60	89.61	93.61	0.81	2.53	13.23
Noothout et al. (2020)	2.03 (0.49)	61.54	83.23	92.81	94.95	3.33	3.05	8.42
Zeng et al. (2021)	1.89 (0.43)	68.81	86.56	92.60	95.26	1.19	2.28	9.30
SA-LSTM (Chen et al., 2022)	1.62 (**0.40**)	74.60	87.23	93.82	**95.86**	0.81	2.60	6.77
FARNet (Ao and Wu, 2023)	1.73 (1.12)	83.40	89.89	92.46	94.42	0.04	3.79	6.25
MTN	1.68 (0.98)	86.14	90.81	93.16	94.32	0.00	2.67	5.82
Proposed	**1.36** (0.53)	**86.70**	**91.61**	**93.96**	95.12	**0.00**	**1.72**	**5.79**
MTN[a]	2.06 (1.22)	79.72	86.56	90.18	92.32	0.00	3.72	8.32
Proposed[a]	1.77 (0.85)	80.07	87.09	90.77	92.91	0.00	2.77	8.63

[a] Indicates the proposed method trained using only 40 fully annotated data.

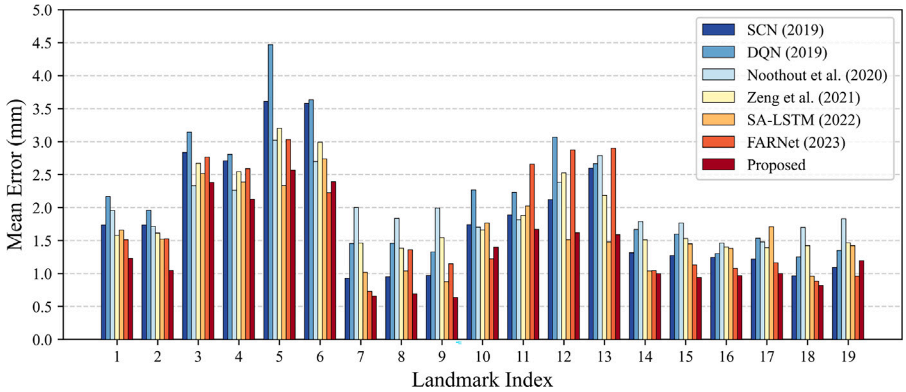


Fig. 8. Mean radial errors (MREs, in mm) in the private cerebral MRA dataset.

Table 3
Quantitative results on the cerebral CTA dataset measured by MRE, associated SD, SDRs in four distance ranges, and ERs of three error types, with the best performance shown in bold. "MTN" stands for the proposed multi-task network.

Method	MRE (SD) (mm, ↓)	SDR (%, ↑)				ER (%, ↓)		
		2 mm	3 mm	4 mm	5 mm	I	II	III
U-Net (Ronneberger et al., 2015)	4.47 (2.68)	62.67	75.97	83.31	86.19	0.08	11.79	**6.51**
SCN (Payer et al., 2019)	2.60 (0.74)	54.77	73.01	83.25	89.07	0.19	4.99	12.27
DQN (Alansary et al., 2019)	2.73 (1.01)	50.11	71.76	84.29	90.16	1.31	4.08	12.88
Noothout et al. (2020)	2.86 (0.96)	42.80	66.56	80.48	87.87	5.52	5.63	8.96
Zeng et al. (2021)	2.69 (0.52)	39.36	68.03	84.19	92.08	1.01	3.92	13.63
SA-LSTM (Chen et al., 2022)	2.54 (**0.41**)	41.01	66.93	84.83	92.85	2.45	4.19	9.55
FARNet (Ao and Wu, 2023)	2.59 (1.53)	63.89	79.95	87.44	91.20	0.67	5.47	9.92
MTN	2.58 (1.27)	65.87	80.91	88.75	92.35	0.03	4.75	7.57
Proposed	**2.18** (0.72)	**66.11**	**81.28**	**89.31**	**93.09**	**0.03**	**3.52**	8.19

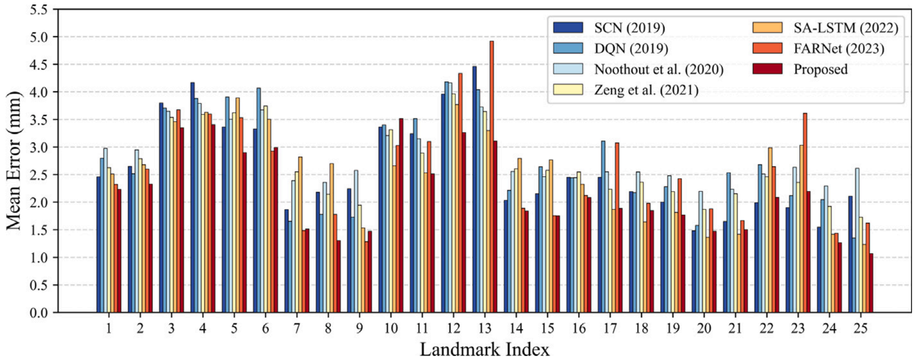


Fig. 9. Mean radial errors (MREs, in mm) for twenty-five landmarks in the cerebral CTA dataset. See Fig. 5(b) for the meaning of landmark index.

Table 4
Quantitative results on the aorta CTA dataset evaluated by MRE, associated SD, SDRs in four distance ranges, and ERs of three error types, with the best performance shown in bold. “MTN” stands for the proposed multi-task network.

Method	MRE (SD) (mm, ↓)	SDR (%, ↑)				ER (%, ↓)		
		2 mm	3 mm	4 mm	5 mm	I	II	III
U-Net (Ronneberger et al., 2015)	4.38 (2.35)	52.64	75.69	84.31	86.11	0.69	3.89	7.78
SCN (Payer et al., 2019)	2.82 (1.71)	52.64	78.06	88.06	91.25	0.83	0.97	4.31
DQN (Alansary et al., 2019)	2.93 (1.02)	49.31	69.72	88.19	91.81	1.25	0.83	5.56
Noothout et al. (2020)	3.03 (0.95)	38.61	58.19	87.22	**92.36**	2.36	0.69	4.58
Zeng et al. (2021)	2.91 (0.85)	42.5	66.81	88.06	91.81	1.81	0.56	4.86
SA-LSTM (Chen et al., 2022)	2.86 (**0.69**)	46.94	67.78	79.17	88.19	1.53	0.69	5.14
FARNet (Ao and Wu, 2023)	2.89 (1.15)	53.06	73.75	86.53	89.03	0.69	1.11	5.28
MTN	2.91 (1.22)	55.83	77.78	86.11	89.44	0.42	1.53	6.11
Proposed	**2.52** (0.99)	**57.22**	**79.44**	**88.47**	91.53	**0.28**	**0.56**	**4.17**

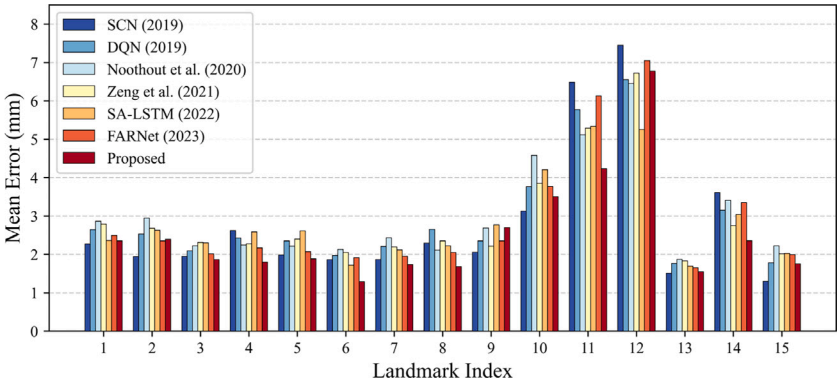


Fig. 10. Mean radial errors (MREs, in mm) for fifteen landmarks in the aorta CTA dataset. See Fig. 5(c) for the meaning of landmark index.

improvements over the traditional heatmap regression network (U-Net) by 1.89 mm (42% reduction) and 2.29 mm (51% reduction) in MRE, respectively. As illustrated in Fig. 9, similar to the MRA dataset, the bifurcation landmarks between MCA-M1, M2 segments (landmark 5 and 6), and between PCA-P2 A, P2P segments (landmark 12 and 13) have larger detection error due to variable vascular structures. The landmarks on the bilateral ICA segments (landmark 20 to 25) exhibit small detection errors, which are more spatially consistent with less variations.

Compared with the state-of-the-art approaches, our method outperforms SCN by 0.42 mm (16% reduction), DQN by 0.55 mm (20% reduction), Noothout et al. (2020) by 0.68 mm (24% reduction), Zeng et al. (2021) by 0.51 mm (19% reduction), SA-LSTM by 0.36 mm (14% reduction), and FARNet by 0.41 mm (16% reduction) in MRE metric. In particular, after the post-processing algorithm, the proposed method exhibits fewer type II errors but more type III errors, since the post-processing algorithm may correct the detection errors of type II to other locations on the right vascular segments. Qualitative results are shown in Fig. 11(c). The average data preprocessing time and inference times are reported in Table 5.

4.4.4. Aorta CTA dataset

In the aorta CTA dataset, the trunk was not included in the constructed graph, and the nine major branches were independently optimized. For the landmarks without edge connection (landmark 2 and 3), the predicted positions of the multi-task network were not changed in the post-processing stage. We report the results on the test set in Table 4 and Fig. 10. Our method obtains lower detection error by 0.3 mm (11% reduction), 0.41 mm (14% reduction), 0.51 mm (17% reduction), 0.39 mm (13% reduction), 0.34 mm (12% reduction), and 0.37 mm (13% reduction) than SCN, DQN, Noothout et al. (2020) and Zeng et al. (2021), SA-LSTM, and FARNet, respectively.

For topological detection errors, compared with cerebral vessels, the aorta is more prone to type I errors due to the interference of the obvious vertebrae, while the ER of type II exhibits significant reductions, since the vascular segments of the aorta are easier to distinguish. The MREs for each landmark are shown in Fig. 10. Landmark confusion problems in the aorta are mainly found at the secondary bifurcation landmarks in the abdominal region (landmark 9–12), which are disturbed by tiny branches and show large detection errors. Qualitative comparisons are presented in Fig. 11(d).

Experiments were conducted on the top (aortic arch), middle (abdominal aorta), and bottom (common iliac artery) parts of the aorta separately. On average, the overall preprocessing time (only crop operation) was 3.15 ± 0.69 s. The total inference times are listed in Table 5.

4.5. Ablation study

To validate the components of the proposed method, we performed an ablation study on the public cerebral MRA dataset using the backbone network with different task configurations. Quantitative results are summarized in Table 6. Qualitative predictions with the multi-task network and overall framework are shown in the last two rows of Fig. 11.

The heatmap regression network (U-Net in Table 1) obtains high SDRs even within small distance thresholds but the highest ER of type II error, which indicates that the network learns representative local features but suffers from severe landmark confusion problems. By adding semantic segmentation and orientation field regression as auxiliary tasks successively, the detection error of the network shows significant reduction by 0.96 mm and 0.6 mm respectively, with the SDR and ER metrics improving accordingly. In particular, utilizing orientation field regression as the single auxiliary task is inferior to semantic segmentation. Although the orientation field contains more structural information (distribution and direction of vascular segments), it is difficult to learn without the assistance of semantic segmentation. By replacing the max-voting strategy with the proposed post-processing method, the MRE metric further reduces by 0.46 mm. Visualization examples show that the post-processing method corrects unreasonable

Table 5
Average time cost per scan (in second) and associated standard deviation for data preprocessing and comparison methods on different datasets.

Method	Public cerebral MRA	Private cerebral MRA	Cerebral CTA	Aorta CTA
Data preprocessing	9.82 (2.80)	10.35 (3.01)	17.04 (4.87)	3.15 (0.69)
U-Net (Ronneberger et al., 2015)	0.45 (0.09)	0.49 (0.10)	1.07 (0.21)	0.73 (0.15)
SCN (Payer et al., 2019)	1.16 (0.05)	1.13 (0.03)	2.47 (0.11)	1.65 (0.03)
DQN (Alansary et al., 2019)	5.25 (2.75)	5.63 (2.79)	8.21 (4.05)	4.15 (1.79)
Noothout et al. (2020)	0.12 (0.01)	0.12 (0.01)	0.83 (0.56)	0.57 (0.08)
Zeng et al. (2021)	1.53 (0.09)	1.79 (0.10)	2.94 (0.33)	3.11 (0.17)
SA-LSTM (Chen et al., 2022)	1.31 (0.16)	1.30 (0.11)	1.87 (0.21)	1.65 (0.19)
FARNet (Ao and Wu, 2023)	1.29 (0.18)	1.27 (0.19)	2.51 (0.37)	2.30 (0.24)
Proposed	3.11 (0.47)	3.23 (0.53)	4.33 (0.74)	5.73 (0.86)

Table 6
Ablation study on the public cerebral MRA dataset of different task configurations and hyperparameter settings in the post-processing algorithm, with the best performance shown in bold. "MTN" stands for the proposed multi-task network.

Method	MRE (SD) (mm, ↓)	SDR (%, ↑)				ER (%, ↓)		
		2 mm	3 mm	4 mm	5 mm	I	II	III
Heat	3.29 (1.82)	81.48	84.60	87.72	89.08	0.00	7.99	9.36
Heat+Seg	2.33 (1.74)	84.21	87.91	90.06	92.01	0.00	5.07	5.65
Heat+Ori	2.63 (1.67)	84.80	87.13	88.89	91.03	0.00	6.04	5.85
MTN	1.73 (0.90)	86.16	90.06	91.81	93.76	0.00	2.34	6.04
$\tau = 0$, $\gamma = 75$	1.35 (0.51)	86.74	91.03	92.79	94.93	0.19	1.36	5.85
$\tau = 3$, $\gamma = 75$	1.67 (0.64)	85.58	89.86	91.42	93.37	0.00	**0.97**	7.21
$\tau = 0.3$, $\gamma = 0$	1.63 (0.56)	83.63	87.72	89.47	91.81	0.58	1.17	9.36
$\tau = 0.3$, $\gamma = 7500$	1.66 (0.88)	86.35	90.25	92.01	93.96	0.00	2.14	6.04
$\tau = 0.75$, $\gamma = 125$	1.31 (0.50)	87.52	91.62	93.18	95.32	0.00	1.17	5.65
$\tau = 0.3$, $\gamma = 75$[a]	**1.27 (0.48)**	**87.72**	**92.01**	**93.57**	**95.71**	**0.00**	1.17	**5.26**

[a] Indicates the optimal hyperparameter setting selected on the validation set.

predictions while keeping the correct predictions unchanged, which also demonstrates that our framework solves the landmark confusion problem effectively.

Furthermore, we compared the performance of different hyperparameter settings in the proposed post-processing algorithm, all of which were performed on the predictions of the multi-task network. The optimal hyperparameter setting selected on the validation set is $\tau = 0.3$, $\gamma = 75$. The penalty coefficient τ constrains the predictions near the corresponding vascular segments, which contributes to the reduction of the ERs of type I and type II. When τ is 0, the ERs of types I and II show a small increase. When τ is very large (e.g., $\tau = 3$), the algorithm may correct the predictions to the wrong positions of the right segments, bringing higher ER of type III. On the other hand, the weight γ controls the proportion of matching score and heatmap confidence in the final score. When γ is 0, the decision strategy only considers the matching score without utilizing the local image features encoded in the heatmaps. When γ is very large (e.g., $\gamma = 7500$), the post-processing algorithm is approximately equivalent to the max-voting strategy. Except for the extreme cases mentioned above, changing the hyperparameter setting will have only a slight impact on the final experimental results (e.g., $\tau = 0.75$, $\gamma = 125$).

5. Discussion

Overall, we propose a deep learning-based framework for vascular landmark detection in medical images. We embed the anatomical structural prior in the multi-task network and utilize the spatial relationships between neighboring landmarks explicitly to optimize the final predictions. The experimental results demonstrate that our method can be applied to vascular landmark detection tasks differing in landmark number, vascular structure, and imaging modality, without the need for complex modifications. Compared with other state-of-the-art approaches (Payer et al., 2019; Noothout et al., 2020; Alansary et al., 2019; Chen et al., 2022), our method achieves superior detection performance, higher location accuracy especially within small distance thresholds, and less topological detection errors, which are more important clinically. Furthermore, considering that our method requires semantic segmentation annotations in addition to landmark annotations, we introduced more scans when training these state-of-the-art approaches for fairer comparisons.

To study the clinical application of vascular landmark detection, we developed a non-learning algorithm to generate vascular semantic segmentation from landmark prediction and manual annotation of binary segmentation. The main flow of the algorithm is to segment the refinement centerline by landmark predictions, and label the remaining voxels in the binary segmentation following the nearest neighbor strategy. The mean Dice coefficient of all vascular segments was evaluated for quantitative comparison. Taking 10 private MRA test data for example, our method, the methods proposed in Payer et al. (2019), Noothout et al. (2020), Alansary et al. (2019), and Chen et al. (2022) yield results of 90.20%, 83.16%, 82.84%, 83.05%, and 85.02%, respectively, which indicates that our method captures the structural prior information more effectively.

There are some limitations in our proposed method. We assume that all predefined landmarks exist in the image and do not discuss the case where some landmarks are missing (e.g., some landmarks are invisible due to vascular obstruction), which may be solved by setting a threshold for the confidence of candidate points. Another limitation is to perform multi-task network and post-processing method in a cascade manner, where the complex edge weight calculation process greatly increases the inference time. In the future, we will investigate how to integrate the optimization process into an end-to-end deep learning framework.

6. Conclusion

In this paper, we propose a multi-task global optimization-based method for automatic vascular landmark detection. A multi-task network is exploited for initial heatmap prediction, where vascular semantic segmentation and orientation field regression are introduced as auxiliary objectives to incorporate anatomical prior information. During inference, instead of performing a max-voting strategy, we present a global optimization-based post-processing algorithm for reliable landmark decision. The spatial relationships between landmarks are utilized

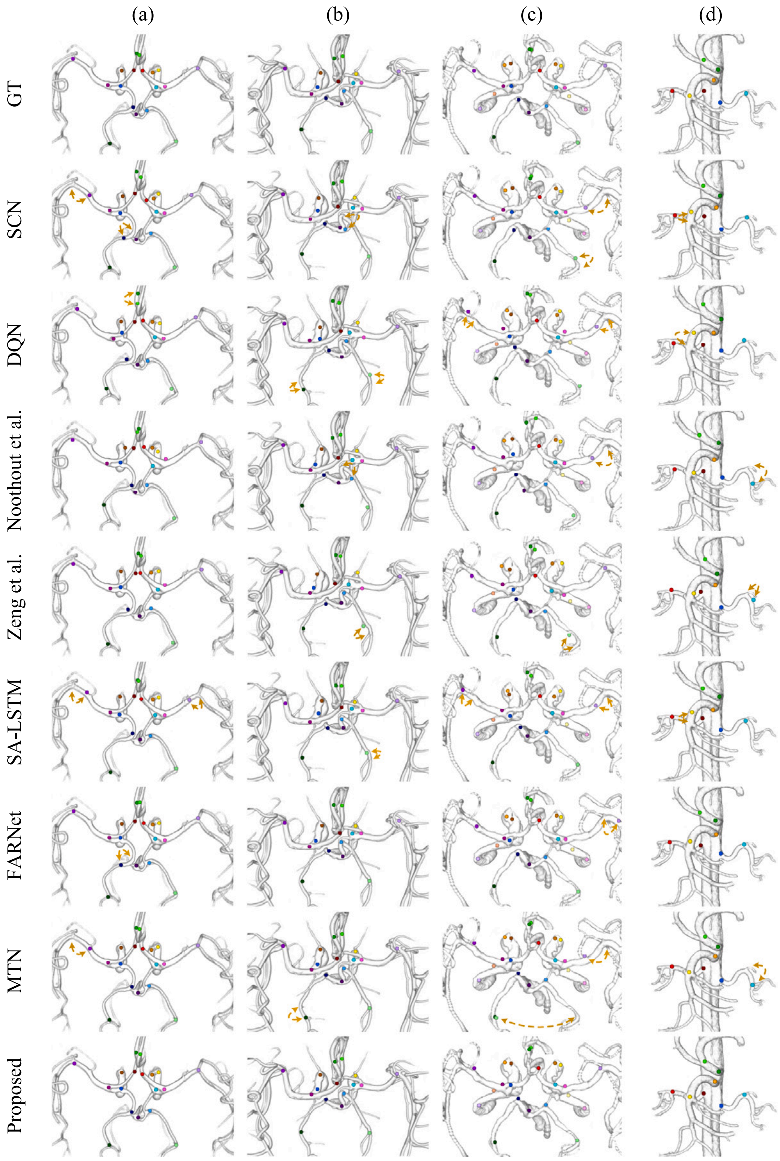


Fig. 11. Qualitative comparisons on (a) public cerebral MRA, (b) private cerebral MRA, (c) cerebral CTA, and (d) aorta CTA datasets. The yellow dashed lines connect the landmark ground truth and predicted positions with large detection errors. Different landmarks are differentiated by color. Some landmarks are not drawn due to occlusion.

explicitly to tackle the landmark confusion problem. The proposed method was evaluated using three datasets with different vascular structures and imaging modalities. Experimental results demonstrate that our framework provides a significant improvement and achieves state-of-the-art performance. Future studies will include extending the proposed method to other anatomical tubular structures.

CRediT authorship contribution statement

Zimeng Tan: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft. **Jianjiang Feng:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. **Wangsheng Lu:** Data curation, Funding acquisition. **Yin Yin:** Data curation, Funding acquisition. **Guangming Yang:** Data curation, Funding acquisition. **Jie Zhou:** Funding acquisition, Resources, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This study got ethical approval of Xuanwu Hospital of Capital Medical University (2020009) for using the clinically collected cerebral MRA dataset, cerebral CTA dataset, and aorta CTA dataset.

References

Al, W.A., Yun, I.D., 2020. Partial policy-based reinforcement learning for anatomical landmark localization in 3D medical images. IEEE Trans. Med. Imaging 39 (4), 1245–1255.

Alansary, A., Oktay, O., Li, Y., Folgoc, L.L., Hou, B., Vaillant, G., et al., 2019. Evaluating reinforcement learning agents for anatomical landmark detection. Med. Image Anal. 53, 156–164.

Almasi, S., Lauric, A., Malek, A., Miller, E.L., 2018. Cerebrovascular network registration via an efficient attributed graph matching technique. Med. Image Anal. 46, 118–129.

Ao, Y., Wu, H., 2023. Feature aggregation and refinement network for 2D anatomical landmark detection. J. Digit. Imaging 36 (2), 547–561.

Bogunović, H., Pozo, J.M., Cárdenes, R., Román, L.S., Frangi, A.F., 2013. Anatomical labeling of the circle of willis using maximum a posteriori probability estimation. IEEE Trans. Med. Imaging 32 (9), 1587–1599.

Bradac, G.B., 2014. Cerebral Angiography: Normal Anatomy and Vascular Pathology. Springer-Verlag Berlin Heidelberg.

Brenes, D., Barberan, C.J., Hunt, B., Parra, S.G., Salcedo, M.P., Possati-Resende, J.C., et al., 2022. Multi-task network for automated analysis of high-resolution endomicroscopy images to detect cervical precancer and cancer. Comput. Med. Imaging Graph 97, 102052.

Browning, J., Kornreich, M., Chow, A., Pawar, J., Zhang, L., Herzog, R., et al., 2021. Uncertainty aware deep reinforcement learning for anatomical landmark detection in medical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 636–644.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y., 2021. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans. Pattern Anal. Mach. Intell. 43, 172–186.

Chen, L., Hatsukami, T., Hwang, J.-N., Yuan, C., 2020. Automated intracranial artery labeling using a graph neural network and hierarchical refinement. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 76–85.

Chen, R., Ma, Y., Chen, N., Liu, L., Cui, Z., Lin, Y., Wang, W., 2022. Structure-aware long short-term memory network for 3D cephalometric landmark detection. IEEE Trans. Med. Imaging 41 (7), 1791–1801.

Choi, D., Kim, T., Jang, J., Sunwoo, L., Lee, K.J., 2023. Intracranial steno-occlusive lesion detection on time-of-flight MR angiography using multi-task learning. Comput. Med. Imaging Graph 107, 102220.

Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K., 2013. Regression forests for efficient anatomy detection and localization in computed tomography scans. Med. Image Anal. 17, 1293–1303.

Dai, J., He, K., Sun, J., 2016. Instance-aware semantic segmentation via multi-task network cascades. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3150–3158.

Elattar, M., Wiegerinck, E., Van Kesteren, F., Dubois, L., Planken, N., Vanbavel, E., et al., 2016. Automatic aortic root landmark detection in CTA images for preprocedural planning of transcatheter aortic valve implantation. Int. J. Cardiovasc. Imaging 32, 501–511.

Gao, Y., Shen, D., 2015. Collaborative regression-based anatomical landmark detection. Phys. Med. Biol. 60 (24), 9377.

Gende, M., de Moura, J., Novo, J., Ortega, M., 2022. End-to-end multi-task learning approaches for the joint epiretinal membrane segmentation and screening in OCT images. Comput. Med. Imaging Graph 98, 102068.

Ghesu, F.-C., Georgescu, B., Zheng, Y., Grbic, S., Maier, A., Hornegger, J., Comaniciu, D., 2019. Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans. IEEE Trans. Pattern Anal. Mach. Intell. 41 (1), 176–189.

Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N., 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med. Image Anal. 58, 101563.

Han, D., Gao, Y., Wu, G., Yap, P.-T., Shen, D., 2015. Robust anatomical landmark detection with application to mr brain image registration. Comput. Med. Imaging Graph. 46, 277–290.

He, T., Hu, J., Song, Y., Guo, J., Yi, Z., 2020. Multi-task learning for the segmentation of organs at risk with label dependence. Med. Image Anal. 61, 101666.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: 32nd International Conference on Machine Learning, ICML 2015. Lille, France, pp. 448–456.

Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B., 2009. Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans. IEEE Trans. Med. Imaging 28 (7), 1000–1010.

Kuang, Z., Li, Z., Zhao, T., Fan, J., 2017. Deep multi-task learning for large-scale image classification. In: Proc. IEEE Conf. Multimedia Big Data. pp. 310–317.

Laiz, P., Vitrià, J., Gilabert, P., Wenzek, H., Malagelada, G., Watson, A.J.M., Seguí, S., 2023. Anatomical landmarks localization for capsule endoscopy studies. Comput. Med. Imaging Graph 108, 102243.

Lang, Y., Lian, C., Xiao, D., Deng, H., Yuan, P., Gateno, J., et al., 2020. Automatic localization of landmarks in craniomaxillofacial CBCT images using a local attention-based graph convolution network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 817–826.

Li, S.Z., 2009. Markov Random Field Modeling in Image Analysis. Springer-Verlag, London, U.K.

Liao, H., Mesfin, A., Luo, J., 2018. Joint vertebrae identification and localization in spinal CT images by combining short- and long-range contextual information. IEEE Trans. Med. Imaging 37 (5), 1266–1275.

Liu, Y., Wang, X., Wu, Z., López-Linares, K., Macía, I., Ru, X., et al., 2022. Automated anatomical labeling of a topologically variant abdominal arterial system via probabilistic hypergraph matching. Med. Image Anal. 75, 102249.

Matsuzaki, T., Oda, M., Kitasaka, T., Hayashi, Y., Misawa, K., Mori, K., 2015. Automated anatomical labeling of abdominal arteries and hepatic portal system extracted from abdominal CT volumes. Med. Image Anal. 20, 152–161.

Mori, K., Ema, S., Kitasaka, T., Mekada, Y., Ide, I., Murase, H., 2005. Automated nomenclature of bronchial branches extracted from CT images and its application to biopsy path planning in virtual bronchoscopy. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 854–861.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML 2010. Madison, WI, USA, pp. 807–814.

Noothout, J.M.H., De Vos, B.D., Wolterink, J.M., Postma, E.M., Smeets, P.A.M., Takx, R.A.P., et al., 2020. Deep learning-based regression and classification for automatic landmark localization in medical images. IEEE Trans. Med. Imaging 39 (12), 4011–4022.

Norajitra, T., Maier-Hein, K.H., 2017. 3D statistical shape models incorporating landmark-wise random regression forests for omni-directional landmark detection. IEEE Trans. Med. Imaging 36 (1), 155–168.

Oh, K., Oh, I.-S., van nhat Le, T., Lee, D.-W., 2020. Deep anatomical context feature learning for cephalometric landmark detection. IEEE J. Biomed. Health Inf. 25 (3), 806–817.

Oktay, O., Bai, W., Guerrero, R., Rajchl, M., de Marvao, A., O'Regan, D.P., et al., 2017. Stratified decision forests for accurate anatomical landmark localization in cardiac images. IEEE Trans. Med. Imaging 36 (1), 332–342.

Payer, C., Štern, D., Bischof, H., Urschler, M., 2019. Integrating spatial configuration into heatmap regression based CNNs for landmark localization. Med. Image Anal. 54, 207–219.

Qian, J., Cheng, M., Tao, Y., Lin, J., Lin, H., 2019. CephaNet: An improved faster R-CNN for cephalometric landmark detection. In: International Symposium on Biomedical Imaging. IEEE, pp. 868–871.

Robben, D., Türetken, E., Sunaert, S., Thijs, V., Wilms, G., Fua, P., Maes, F., Suetens, P., 2016. Simultaneous segmentation and anatomical labeling of the cerebral vasculature. Med. Image Anal. 32, 201–215.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Schwarz, C.G., Fletcher, E., Singh, B., Liu, A., Smith, N., DeCarli, C., et al., 2012. Most edges in markov random fields for white matter hyperintensity segmentation are worthless. In: Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. pp. 2684–2687.

Song, L., Lin, J.P., Wang, Z.J., Wang, H., 2020. An end-to-end multi-task deep learning framework for skin lesion analysis. IEEE J. Biomed. Health Inf. 24 (10), 2912–2921.

Tan, Z., Feng, J., Lu, W., Yin, Y., Yang, G., Zhou, J., 2022. Cerebrovascular landmark detection under anatomical variations. In: International Symposium on Biomedical Imaging. IEEE, pp. 1–5.

Tan, Z., Feng, J., Zhou, J., 2021. Multi-task learning network for landmark detection in anatomical tree structures. In: International Symposium on Biomedical Imaging. IEEE, pp. 1975–1979.

Tuysuzoglu, A., Tan, J., Eissa, K., Kiraly, A.P., Diallo, M., Kamen, A., 2018. Deep adversarial context-aware landmark detection for ultrasound imaging. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 151–158.

Urschler, M., Ebner, T., Štern, D., 2018. Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization. Med. Image Anal. 43, 23–36.

Vandenhende, S., Georgoulis, S., Gansbeke, W.V., Proesmans, M., Dai, D., Gool, L.V., 2022. Multi-task learning for dense prediction tasks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 44 (7), 3614–3633.

Wang, J., Long, X., Gao, Y., Ding, E., Wen, S., 2020. Graph-PCNN: Two stage human pose estimation with graph pose refinement. In: Proceedings of the European Conference on Computer Vision. pp. 492–508.

Xu, J., Xie, H., Liu, C., Yang, F., Zhang, S., Chen, X., Zhang, Y., 2021. Hip landmark detection with dependency mining in ultrasound image. IEEE Trans. Med. Imaging 40 (12), 3762–3774.

Yang, D., Xiong, T., Xu, D., Huang, Q., Liu, D., Zhou, S.K., et al., 2017. Automatic vertebra labeling in large-scale 3D CT using deep image-to-image network with message passing and sparsity regularization. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 633–644.

Zeng, J., Liu, S., Li, X., Mahdi, D.A., Wu, F., Wang, G., 2017. Deep context-sensitive facial landmark detection with tree-structured modeling. IEEE Trans. Image Process. 27 (5), 2096–2107.

Zeng, M., Yan, Z., Liu, S., Zhou, Y., Qiu, L., 2021. Cascaded convolutional networks for automatic cephalometric landmark detection. Med. Image Anal. 68, 101904.

Zhang, J., Liu, M., Shen, D., 2017. Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks. IEEE Trans. Image Process. 26 (10), 4753–4764.

Zhang, J., Liu, M., Wang, L., Chen, S., Yuan, P., Li, J., et al., 2020. Context-guided fully convolutional networks for joint craniomaxillofacial bone segmentation and landmark digitization. Med. Image Anal. 60, 101621.

Zheng, Y., Liu, D., Georgescu, B., Nguyen, H., Comaniciu, D., 2015. 3D deep learning for efficient and robust landmark detection in volumetric data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 565–572.

Zhong, Z., Li, J., Zhang, Z., Jiao, Z., Gao, X., 2019. An attention-guided deep regression model for landmark detection in cephalograms. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 540–548.

Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., et al., 2021. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. Med. Image Anal. 70, 101918.