Pattern Recognition 146 (2024) 110028

Contents lists available at ScienceDirect

Pattern Recognition

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

Deep orientated distance-transform network for geometric-aware centerline detection

Zheheng Jiang [a,*], Hossein Rahmani [a], Plamen Angelov [a], Ritesh Vyas [b], Huiyu Zhou [c], Sue Black [d], Bryan Williams [a]

[a] School of Computing and Communications, Lancaster University, Lancaster, United Kingdom
[b] Pandit Deendayal Energy University, Gandhinagar, Gujarat, India
[c] School of Computing and Mathematical Sciences, University of Leicester, Leicester, United Kingdom
[d] St John's College of the University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Centerline detection
Geometric properties
Graph representation
Graph refinement

ABSTRACT

The detection of structure centerlines from imaging data plays a crucial role in the understanding, application and further analysis of many diverse problems, such as road mapping, crack detection, medical imaging and biometric identification. In each of these cases, pixel-wise segmentation is not sufficient to understand and quantify overall graph structure and connectivity without further processing that can lead to compound error. We thus require a method for automatic extraction of graph representations of patterning. In this paper, we propose a novel Deep Orientated Distance-transform Network (DODN), which predicts the centerline map and an orientated distance map, comprising orientation and distance in relation to the centerline and allowing exploitation of its geometric properties. This is refined by jointly modeling the relationship between neighboring pixels and connectivity to further enhance the estimated centerline and produce a graph of the structure. The proposed approach is evaluated on a diverse range of problems, including crack detection, road mapping and superficial vein centerline detection from infrared/ color images, improving over the state-of-the-art by 2.1%, 10.9% and 17.3%/ 4.6% respectively in terms of quality, demonstrating its generalizability and performance in a wide range of mapping problems.

1. Introduction

Curvilinear structures are very common occurrences in many systems, including geographical and biological. Examples of such structures include road networks in satellite images, blood vessels in the body, and cracks in structures. The automatic detection of these objects can be useful in numerous domains and applications. For example, accurate road centerline detection is essential for functional and up-to-date navigation systems [1]; crack detection is important in evaluating structural integrity and road conditions to prompt essential maintenance [2]; many artificial intelligence-based aspirations in medical imaging demand understanding of vessel and nerve maps [3]; and in cases of serious crime, comparing superficial vein patterns can help with perpetrator identification [4,5].

Centerline detection is a key step in accurately extracting structure maps from images. Alternative techniques have attempted this using hand-crafted filters [6,7], which are designed to have a strong response when computed on line-like structures. These models are designed for use in controlled conditions and their performance can be sensitive to external imaging conditions such as variation of illumination and contrast. Recent learning based methods [8–10], which aim to train a classification model to predict the label of each pixel, tend to outperform hand-crafted approaches when processing images with complex scenes. Models based on features extracted from deep neural networks, such as U-net [11], also fall into this category and have achieved great success in computer vision tasks. Moreover, Zou et al. [2] modify the U-net and propose to fuse multi-scale feature maps for effectively capturing information of thin objects. However, similar to other U-net-based methods [11–17], the authors classify individual pixels and suffer from several limitations: (1) they require high-quality pixel-level annotations, which are not easy to obtain; (2) it is difficult for the classifier to distinguish centerline pixels from neighboring pixels, since they have similar appearance; (3) due to the absence of connectivity supervision and constraints, isolated erroneous responses, discontinuities, and topological errors are often present in the resulting maps, particularly when close to junctions. While Mosinska et al. [18] proposed a topology loss to reduce topological mistakes, their output cannot preserve global connectivity.

* Corresponding author.
E-mail address: z.jiang11@lancaster.ac.uk (Z. Jiang).

https://doi.org/10.1016/j.patcog.2023.110028
Received 17 January 2023; Received in revised form 8 August 2023; Accepted 2 October 2023
Available online 5 October 2023

To tackle these obstacles, we propose to perform centerline detection by designing a multitask deep network to predict not only a centerline map, but also an orientated distance map, consisting of the distance transform value and direction from each neighboring pixel to the centerline. Such representation encodes geometric properties of the centerline and is used to determine and refine the centerline map from pixel-level classification. In contrast, Zhou et al. [19] encodes geometric information of body joints to associate dense human semantics with sparse keypoints, while our approach encodes geometric information of structure centerlines to construct a graph-cut energy model to improve connectivity. Specifically, the proposed deep learning model, namely, Deep Oriented Distance-transform Network (DODN), takes images, centerline ground truths and oriented distance-transform representations as input for training while only the image is required for testing. This deep architecture is designed to learn three tasks. **First**, it learns to classify each pixel, as with most existing deep fully convolutional neural networks (CNN). However, such per-pixel classification models can only produce diffused centerlines. Existing centerline/edge detection methods [2,9,20] apply non-maximum suppression (NMS) as a post processing step to obtain thin centerlines, resulting in poor localization and poor connectivity. Differing from previous work, we address this issue by learning an Oriented Distance-transform representation. **Second**, the orientation of each pixel to the centerline is predicted and used to localize centerline points. **Third**, the distance of each pixel to the centerline is predicted. The derived representations are used to construct a graph-cut energy model to improve connectivity. Instead of training a separate model for each task, we combine these tasks into a single model, which allows the tasks to be learned simultaneously and improved results to be achieved while reducing computations. Our approach is designed to be more robust to imperfect annotation that does not perfectly coincide with the actual image structures, which is typically the case in the annotation of satellite imagery of roads, blood vessels and cracks. Obtaining high-quality pixel-level annotations of curvilinear structures for such imagery is a challenging and time-consuming process, which can be exceptionally difficult to achieve accurately. Training networks that heavily depend on pixel-level annotations can be prone to errors with curvilinear structures since small imperfections in the annotation can cause significant detail to be missed. While our approach also requires pixel-wise annotation, our DODN+GC model effectively reduces the heavy dependency on pixel-level annotation. As shown in our ablation study (Table 8), if we only use pixel-level loss as in experiment A, the performance is inferior compared to the other experiments. For clarity, the main contributions of our work are summarized as follows:

1. We present a novel multitask learning architecture (DODN), which is designed for pixel-wise classification, orientation prediction and distance regression. Furthermore, we incorporate task uncertainty to simultaneously learn multiple losses.
2. A connectivity refinement approach is proposed, where a connectivity constraint and relationship between neighboring pixels are modeled using a graph-cut approach. To the best of our knowledge, this is the first work to incorporate graph-cut for centerline refinement.
3. The proposed approach is thoroughly evaluated on three datasets, including a large benchmark road mapping dataset (Massachusetts Roads), a crack detection dataset (CrackTree200), and a superficial vein pattern tracing dataset in both color and infrared modalities (SuperID). Our experimental results demonstrate that the proposed approach outperforms current state-of-the-art methods. We present an ablation study to further demonstrate the effectiveness of each component of the proposed model.

2. Related work

Centerline detection methods can be divided into two main categories. The first category relies on hand-designed filters. The computations of Hessian [6,21] and Optimally Oriented Flux [22,23] are two popular techniques, where eigenvalues are extracted to estimate the likelihood that a pixel lies on a centerline. However, since their hyperparameters are set empirically, these approaches are sensitive to background noise and limited to simple scenes and images with high contrast.

A second category of approaches, that rely on machine learning or deep learning techniques, performs better at dealing with poor contrast, noise and complicated background problems. Some works formulate centerline detection as a per-pixel classification problem. Carlos et al. [12] adopted a Gradient Boosting framework to jointly learn the filters for feature extraction and a classifier, and employed them to label the pixels as belonging to a linear structure of interest or to the background. Sironi et al. [9] transformed the per-pixel classification problem to a regression one by adopting the idea of a distance transform. As a state-of-the-art machine learning technique, CNNs have shown great success in many computer vision tasks, such as classification, segmentation and detection. Several CNN architectures, including Inceptionv3 [24], Resnet101 [25], Densenet121 [26], VGG16 and VGG19 [27], have been developed and used to extract deep features. To deal with the image segmentation problem, U-Net-like [11] models have become popular due to their performance surpassing that of the conventional segmentation methods, e.g. [28,29]. Similar to other learning-based approaches, they try to learn a class label per pixel, which neglects the geometric characteristics of the centerline, resulting in a diffused centerline response. Recent research tends to focus on centerline refinement using geometric and topological properties. For instance, Mosinska et al. [18] propose iterative refinement to close small gaps in road segments. Topology loss is also introduced to supplement the usual pixel-wise loss to favor road-like structures. [30] measure the topological similarity between a prediction and the ground truth by comparing the number of their connected components and handles, and make this measurement differentiable for backpropagation through the network. In order to preserve overall geometric connectivity, [31] present a Feature Interactive Module to exchange features between semantic segmentation and boundary detection.

The output of the above methods consists of the probability that each pixel is part of a centerline, which cannot be directly used as a graph. Centerline and graph extraction are crucial steps of numerous applications in different domains [4,10,32,33]. For example, [32, 34] propose an anatomy structure analysis-based vein extraction algorithm to generate the vein graph. Then a vein indexing feature is extracted from the vein graph to assist the following probe-to-candidate matching in human identification. In Coronary Computed Tomography Angiography, to achieve quantitative measurements of coronary artery stenoses, [10,33] aim to extract the coronary centerline rather than its segmentation from image data. [33] proposes a multitask FCN to simultaneously generate centerline distance maps and detect branch endpoints. Then a minimal path algorithm is used to connect all branch endpoints. Wolterink et al. [10] propose to train a CNN to predict the direction and radius of the vessels. Starting from a seed point, the centerline is tracked by following the vessel centerline using the predictions of the CNN. The automated extraction of roads graphs has also been receiving increasing attention [9,35–38], since it benefits several related applications, such as vehicle navigation, urban planning, and updating geographic information systems. To obtain the graph from the score maps [9],apply standard Non-Maximum Suppression to keep the locations that correspond to a local maxima along radii. In [36], binary thresholding and morphological thinning are applied to produce single-pixel-width centerlines. Máttyus [35] further proposes a graph enhancement method by reasoning about missing connections. The method that we propose in this paper falls into this area. However, in contrast to the above methods, the centerline is extracted by *learning* non-maximum suppression with our novel DODN and is refined with a graph-cut algorithm in our proposed model.

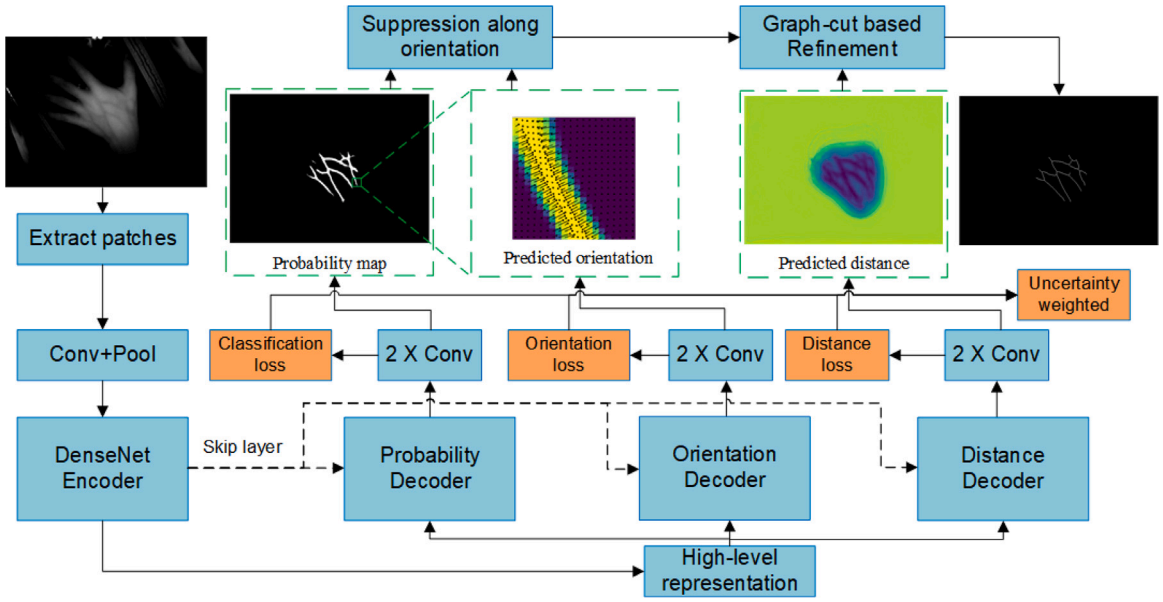


Fig. 1. Overview of the proposed approach. The proposed DODN takes an image as input and produces a pixel-wise classification, predicted orientation and predicted distance to the centerline. These tasks share the same encoder network, which uses a Densenet121 block [26] as shown in Fig. 2, and has separate decoder branches. Dashed arrows represent skip connections applied at the same scale between encoder and decoder layers, allowing feature maps from the downsampling path to be concatenated with corresponding feature maps in the upsampling path. Orange rectangles indicate uncertainty weighted multi-task loss for training our DODN. The centerline is obtained by suppressing pixels along the predicted orientation, and then smoothed by applying a morphological technique. Finally, graph-cut based refinement is proposed to improve centerline connectivity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Proposed method

In this section, the proposed approach for centerline detection is described. First, the method for constructing the oriented distance map is introduced. Then, a multi-task network for learning oriented distance transform and centerline pixel classification is proposed. Finally, a graph-cut based approach is proposed to refine the centerline mask.

3.1. Oriented distance transform for centerline

The goal of learning the oriented distance transform is two-fold. First, in order to obtain thinned centerlines or edges, typical approaches [2,9,20] compute the gradient direction of predicted score maps followed by non-maximum suppression, resulting in poor localization and connectivity. In contrast, we define the centerline points as sinks with flow from nearby points and model learning such representations as a sub-task of our multi-task network. By applying a multi-task learning strategy, the learning efficiency and prediction accuracy of the proposed model can be improved over training a separate model for each task. Second, the learned distance map is utilized to construct the graph-cut energy function to further improve the connectivity of previous centerline results.

Let I be a raw input image (e.g. image of roads, cracks or superficial veins) of size $L \times W$. Our goal is to predict its centerline mask $\hat{Y} = \left\{\hat{y}_{ij}, i = 1, \ldots, L, j = 1, \ldots, W\right\}$, where $\hat{y}_{ij} \in \{0, 1\}$ denotes the predicted label for each pixel at position (i, j), i.e., if the pixel at (i, j) is predicted as a centerline pixel, then $\hat{y}_{ij} = 1$, otherwise $\hat{y}_{ij} = 0$. Here, we also aim to predict the distance map $\hat{D} = \left\{\hat{d}_{ij}, i = 1, \ldots, L, j = 1, \ldots, W\right\}$ and its orientation map $\hat{\Theta} = \left\{\hat{\theta}_{ij}, i = 1, \ldots, L, j = 1, \ldots, W\right\}$. Given the ground-truth label map $Y = \left\{y_{ij}, i = 1, \ldots, L, j = 1, \ldots, W\right\}$ of the dorsal hand image I, we define the nearest centerline point of each pixel as

$$\left(i_{c'}, j_{c'}\right) = \underset{(i_c, j_c)}{\operatorname{argmin}} \sqrt{\left(i - i_c\right)^2 + \left(j - j_c\right)^2}, \tag{1}$$

where (i_c, j_c) is a point on the centerline. If multiple nearest centerline points are found, we will select the first one in clockwise order. Then, the distance map D and its orientation map Θ are computed in the polar representation

$$d_{ij} = \sqrt{\left(i - i_{c'}\right)^2 + \left(j - j_{c'}\right)^2}, \quad \theta_{ij} = tan^{-1}\left(\frac{j_{c'} - j}{i_{c'} - i + \epsilon}\right). \tag{2}$$

where ϵ is a small positive value to avoid the case: $i_{c'} - i = 0$.

3.2. Network for learning oriented distance transform and centerline pixel classification

Due to the similar appearance between centerline pixels and their neighboring pixels, deep learning based pixel-level classification algorithms suffer from low localization accuracy and poor connectivity for centerline detection. We address this problem by learning an oriented distance transform, i.e. predicting $\hat{D}$ and $\hat{\Theta}$, to help localize centerlines and improve their connectivity. In this section, we describe how to train a deep network for vessel centerline detection by focusing on $\hat{D}$, $\hat{\Theta}$ and $\hat{Y}$. The proposed architecture is based on the recently proposed and widely used U-net segmentation architecture, which is a deep encoder–decoder CNN. We use Densenet as our base feature encoder for producing a shared representation, followed by three task-specific convolutional decoders as shown in Fig. 1. For each task, we construct a 4-layer decoder, where each layer consists of one 2×2 up-convolution layer, one concatenation operation with related feature map by skip connections and two 3×3 convolution layers (as shown in Fig. 2). The first convolutional decoder performs per-pixel classification with a dice loss function $\mathcal{L}_{cls}$:

$$\mathcal{L}_{cls} = 1 - \frac{2\sum_{i,j} y_{ij}\psi\left(f_{ij}\left(W, w_{cls}\right)\right)}{\sum_{i,j}\left(\psi\left(f_{ij}\left(W, w_{cls}\right)\right)\right)^2 + \sum_{i,j} y_{ij}^2}, \tag{3}$$

where $\psi(\cdot)$ is a sigmoid function in the last layer, $f_{ij}\left(W, w_y\right)$ is the output of classification decoder at (i, j), W is the matrix of the weights of the shared encoder network and w_{cls} are the weights of this decoder branch.

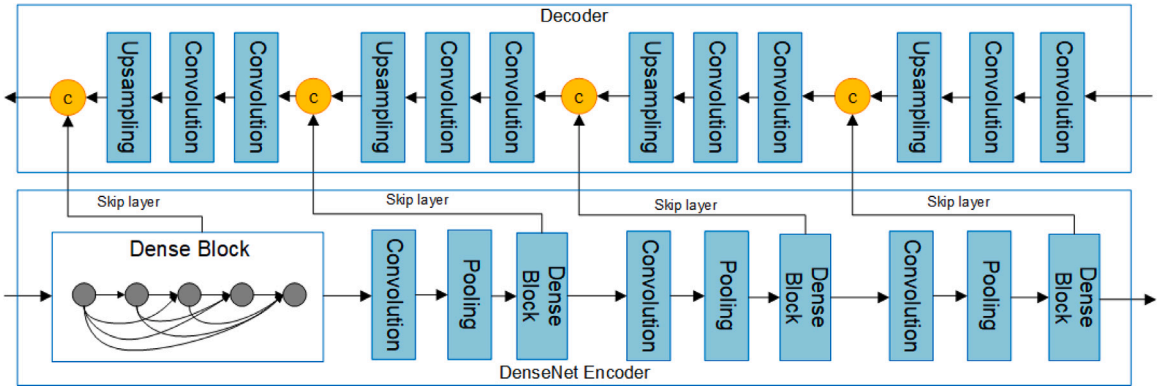


Fig. 2. Illustration of skip layers between the encoder and the decoder network. Yellow circles indicate the operation of concatenation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The second decoder branch is used to predict the orientation map, which is an essential step before performing non-maximum suppression. Similar to setting the direction angle in the Canny edge detector [39], the direction is discretized into $N = 8$ classes with the same angular interval, where the direction angle is rounded to one of eight values comprising vertical, horizontal and the four diagonals (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), $\theta_{ij}, \hat{\theta}_{ij} \in \{0, \dots, N\}$. To train this decoder branch, the orientation loss function is defined as:

$$\mathcal{L}_{ori} = \frac{1}{N}\sum_{n=1}^{N}\left(1 - \frac{2\sum_{i,j}\theta_{ij} f_{ij}\left(\theta_{ij} = n; W, w_{ori}\right)}{\sum_{i,j}\left(f_{ij}\left(\theta_{ij} = n; W, w_{ori}\right)\right)^2 + \sum_{i,j}\theta_{ij}^2}\right), \tag{4}$$

where $f_{ij}\left(\theta_{ij} = n; W, w_{ori}\right)$ is the probability that the output of the orientation decoder at (i, j) belongs to the direction class $n \in \{1, \dots, N\}$. This probability is obtained by squashing the decoder output through a softmax function.

The third decoder branch predicts the distance map D. To train this decoder branch, a combination of $L1$ and structural similarity is used:

$$\begin{aligned}\mathcal{L}_{dis} =& \alpha\frac{1 - SSIM\left(D, f\left(W, w_{dis}\right)\right)}{2} \\ &+ (1-\alpha)\left|f\left(W, w_{dis}\right) - D\right|\end{aligned} \tag{5}$$

with α commonly set to 0.85 [40]. $f\left(W, w_{dis}\right)$ is the output of the distance decoder branch and w_{dis} are the weights of this decoder branch. SSIM is defined as:

$$SSIM\left(D, \hat{D}\right) = \frac{2\mu_D\mu_{\hat{D}} + C_1}{\mu_D^2 + \mu_{\hat{D}}^2 + C_1} \cdot \frac{2\sigma_D\sigma_{\hat{D}} + C_2}{\sigma_D^2 + \sigma_{\hat{D}}^2 + C_2}, \tag{6}$$

where μ_D, σ_D and $\sigma_{D\hat{D}}$ are mean, variance and covariance, all of which are computed within a square window moving over the entire distance map. C_1 and C_2 are small positive constants to avoid division by zero.

With multiple regression loss ($\mathcal{L}_{dis}$) and classification loss ($\mathcal{L}_{ori}$ and $\mathcal{L}_{cls}$), the next goal is to train the multi-task deep network, which is inherently a multi-objective optimization problem. A common approach for this problem is to minimize a weighted linear combination of each individual task loss. However, such an approach would be heavily dependent on manual selection of the weights. Inspired by [41], we propose to control the multi-objective optimization by leaning an uncertainty weight of each individual task. We derive our multi-task loss function based on maximizing the log-likelihood:

$$\begin{aligned}&-\log\mathcal{L}\left(\hat{Y}, \hat{\Theta}, \hat{D} | W, w_{cls}, w_{ori}, w_{dis}, I\right) \\ &= -\log\mathcal{L}\left(\hat{Y} | W, w_{cls}, I\right) - \log\mathcal{L}\left(\hat{\Theta} | W, w_{ori}, I\right) \\ &\quad -\log\mathcal{L}\left(\hat{D} | W, w_{dis}, I\right).\end{aligned} \tag{7}$$

By modeling the predictive distribution of regression and classification by a Laplacian and a Boltzmann function respectively, our total optimization loss is obtained below:

$$\frac{\mathcal{L}_{cls}}{u_{cls}^2} + \frac{\mathcal{L}_{ori}}{u_{ori}^2} + \frac{\mathcal{L}_{dis}}{u_{dis}} + \log u_{cls} + \log u_{ori} + \log u_{dis}, \tag{8}$$

where u_{cls}, u_{ori} and u_{dis} are uncertainty weights that need to be automatically learned and are used to control the optimization.

3.3. Refinement with graph-cut

Many deep learning based methods [11,13–15] are based on pixel-wise classification, resulting in disjoint and poorer segmentation. To improve the connectivity of centerlines, a graph-cut based algorithm is proposed in this paper to connect the broken centerlines.

Let us define an undirected graph $(\mathcal{V}, \mathcal{E})$ whose nodes correspond to pixels. A standard graph cut algorithm aims to minimize the energy function below [44,45]:

$$E(l) = \sum_{p\in\mathcal{V}} E_p\left(l_p\right) + \sum_{(p,q)\in\mathcal{E}} E_{pq}\left(l_p, l_q\right)|l_p - l_q|. \tag{9}$$

Here $E_p\left(l_p\right)$ is a unary term based on the label $l \in \{0, 1\}$ of pixel p, where 0 and 1 correspond to the background and the foreground, respectively. E_{pq} is a boundary term which corresponds to a measure of similarity between pixels p and q. In order to reduce the size of the graph and accelerate computation, we suppose that the broken centerlines are located at the centerline endpoints. We firstly use a hit-miss transform [46] to detect all endpoints Ω from our previous output as shown in Fig. 9. Second, we find a set of pairs of endpoints (s, t) with the definition $\mathcal{ST} = \{(s, t) | dist(s, t) < \delta, s, t \in \Omega\}$, where $dist(s, t)$ denotes the distance between s and t. δ is a threshold for filtering out pairs (s, t) with large distance. Third, we determine a region of interest $ROI(s, t)$ where possible broken centerlines are detected. $ROI(s, t)$ is defined as a $m \times m$ square centered at the center point between s and t, where we set $m = 50$ empirically. Let us denote the total number of ROIs by $\mathcal{Z}$, we define an undirected subgraph $(\mathcal{V}_z, \mathcal{E}_z)$ for the z_{th} ROI. The energy function is defined below:

$$E(l) = \sum_{z=1}^{\mathcal{Z}}\sum_{p\in\mathcal{V}_z} E_p\left(l_p\right) + \sum_{(p,q)\in\mathcal{E}_z} E_{pq}\left(l_p, l_q\right)|l_p - l_q|. \tag{10}$$

In the unary term, we introduce a connectivity constraint in the form of $C_{l=1}(p)$. Specifically, the connectivity constraint is formulated as follows: there must exist a path $\mathcal{P}$ from s to t in each subgraph $(\mathcal{V}_z, \mathcal{E}_z)$

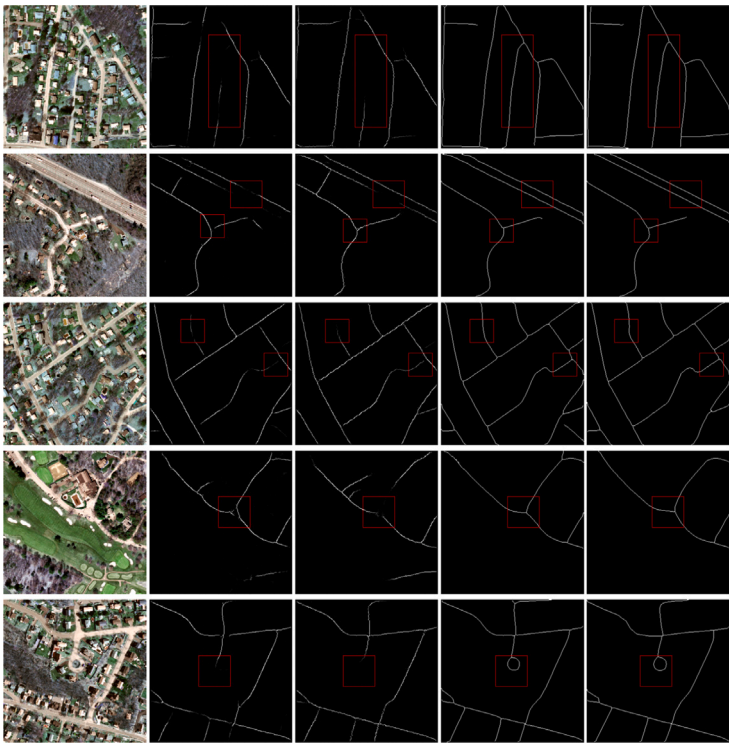


Fig. 3. Examples of detected road centerlines on the Massachusetts dataset. From left to right: Original image, Unet+NMS, Top-Aware [18], DODN-v2 and ground truth. The red rectangles highlight the superior performance of our method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

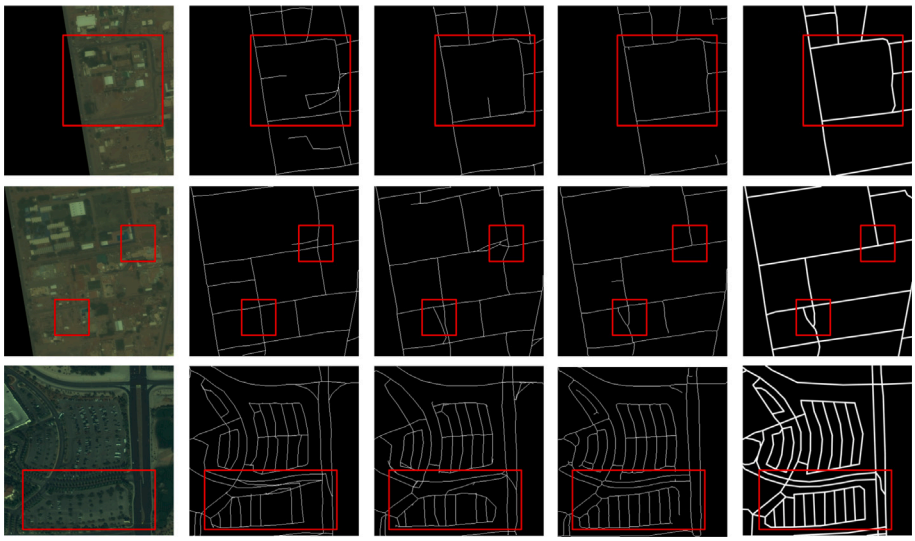


Fig. 4. Examples of detected road centerlines on the SpaceNet dataset. From left to right: Original image, RNGDet [42], RNGDet++ [43], DODN-v2 and ground truth. The red rectangles highlight the superior performance of our method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and all nodes p in the path $\mathcal{P}$ belong to the foreground. Then, the unary term is computed as follows:

$$
\begin{aligned}
&E_p\left(l_p\right) = C_{l=1}(p) + G_l(p) \\
&C_{l=1}(p) = \begin{cases} +\infty & \text{if} \quad p \in \Omega \quad or \quad p \in \mathcal{P} \\ 0 & \text{if} \quad otherwise \end{cases} \\
&G_{l=0}(p) = -log(\frac{f_p(W, w_d is)}{max(\hat{D})}) \\
&G_{1=1}(p) = -log(1 - \frac{f_p(W, w_d is)}{max(\hat{D})}),
\end{aligned}
\tag{11}
$$

where $G_l(p)$ is defined based on the distance map prediction. We define the boundary term E_{pq} in an 8-connected 2D grid graph as follow:

$$
E_{pq}\left(l_p, l_q\right) = \frac{1}{1 + \left\| f_p(W, w_d is) - f_q(W, w_d is) \right\|^2}. \tag{12}
$$

Algorithm 1 summarizes the steps of vessel centerline refinement which is inspired by the Dijkstra algorithm [47]. In this Algorithm, $e(p)$ indicates the cost of the feasible solution for the pair of nodes (s, p). $pre(\cdot)$ is a pointer to the next node, where the shortest path to node s can be obtained.

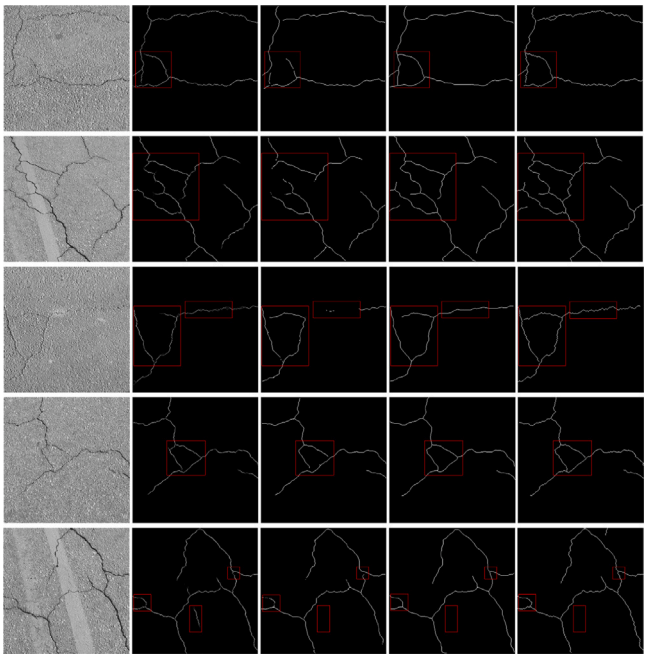


Fig. 5. Examples of detected cracks. From left to right: Original image, DeepCrack [2], Top-Aware [18], DODN-v2 and ground truth. The red rectangles highlight the superior performance of our method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Algorithm 1 Algorithm for centerline refinement with Graph-cut.

initialize: for all nodes $p \in \bigcup_{z=1:Z} \mathcal{V}_z - \{s_z\}$, $pre(p) = NULL$, $e(p) = +\infty$, for all s, $d(s) = min$ Eq. (10), for all z, $\mathcal{S}_z = \emptyset$

for each subgraph $(\mathcal{V}_z, \mathcal{E}_z)$ **do**
  while $t \notin \mathcal{S}$ and $\exists p \in \mathcal{V}_z - \mathcal{S}_z, e(p) < +\infty$ **do**
    find $p \in \mathcal{V}_z - \{s_z\}$ with min $e(p)$;
    add p to $\mathcal{S}$;
    for each neighbor node $q \in \mathcal{V}_z - \{s_z\}$ of p in 8-connected system **do**
      compute path $\mathcal{P}$ from s to q using $pre(q)$;
      compute a solution by minimizing $E(l)$ in Eq. (10) under the constraint $\mathcal{P}$;
      if $e(q) > E(l)$ **then**
        Update $e(q) = E(l)$;
        Update $pre(q) = p$;
      end if
    end for
  end while
end for

4. Experiments

4.1. Datasets

Roads Dataset: We conduct our experiments on two roads dataset: Massachusetts [48] and SpaceNet [49]. Massachusetts Roads dataset contains 1171 aerial color images taken of the state of Massachusetts. The task is to map the centerlines of the roads. Each image covers an area of 2.25 square kilometers with pixel resolution of 1500 × 1500. Following [48], the dataset is split into a training set of 1108 images and a test set of 49 images. To augment the training dataset, we split each image into patches of size 380 × 380 with an overlapping region of 160 pixels, thus providing ∼70K images. The test images are similarly split into patches but without overlapping regions yielding 784 images. SpaceNet contains 2551 RGB aerial images from four different cities with ground resolution of 1m/pixel and pixel resolution of 400 × 400. Following [43], the dataset is split into a training set of 2042 images, a validation set of 127 images and a test set of 382 images.

Cracks Dataset: This is the CrackTree200 dataset [50], which contains 206 images of cracks in roads with pixel resolution 600 × 800, for which we aim to trace the centers of the cracks. Following [50], the dataset is randomly split to 176 training images and 30 testing images. Each image is split into patches of the same size as the Massachusetts dataset, yielding ∼2k images for training and 120 images for testing. This is a challenging dataset since many images suffer from the problems of occlusions and shadows.

SuperID Dataset: This dataset was obtained from an earlier study carried out by *redacted for anonymous submission*. 112 participants were involved in this study. For each participant, the dorsal view of each hand was imaged separately in a flat pose using infrared imaging and color photography. The aim is to trace to centerlines of superficial veins. This yields a total of 224 color photographs (112 left hand, 112 right hand) of size 3504 × 2336 pixels and 224 infrared images (112 left hand, 112 right hand) of size 640 × 480 pixels. All images were reviewed by experts in forensic anthropology with experience of examining such images for casework in forensic identification. The visible vessel patterns were traced using Adobe Photoshop (Adobe Inc, California, USA) and recorded as vectors.

For experimentation, the images were split at the participant level so that the images of 100 participants' hands were used for algorithm training and the images from 12 participants' hands were used for evaluation, yielding 200 color (resp. infrared) images for training and 24 color (resp. infrared) images for evaluation. For efficiency, all images were resized to 640 × 480.

4.2. Implementation details

To make the comparison more clear, we name our full DODN model as **DODN-v2**, and its version without graph-cut refinement as **DODN-v1**. Our method was implemented using Keras with a TensorFlow backend. To reduce overfitting when training, data augmentation is conducted for the baseline approaches and the proposed method. For the Massachusetts Roads [48] and Cracks [50] datasets, we follow the approach in [18] to perform data augmentation by mirroring and rotating the training images by 90°, 180° and 270°. During testing, we employ the sliding window strategy, where the final predictions are obtained by aggregating the score maps predicted from the image patches and averaging values in overlapped regions. For SuperID dataset, given a training/validation hand dorsal image, we randomly sample image patches of size 320 × 320. The patches are randomly flipped in horizontal and vertical directions, and rotated within a $[-45, 45]$ degree interval. We also apply random brightness and contrast adjustment. All models are trained until convergence using the ADAM optimizer [51] with initial rate of 10^{-3}, weight decay of 0.1, mini-batch size of 8 and maximum training iterations of 100. All experiments were performed using a Dell Precision 5820 Workstation with a NVidia GeForce RTX 2080 Ti GPU, 32 GB RAM and an Intel(R) Xeon(R) W-2245 CPU. Our code is available here: https://github.com/ZhehengJiangLancaster/DODN.

4.3. Performance metrics

Following [18], we evaluate the results in terms of correctness, completeness, and quality [52]. Correctness (Corr) and completeness (Comp) correspond to relaxed precision and relaxed recall, where potential shifts of centerline positions are handled by relaxing the notion of a true positive with a distance threshold ρ. In the experiments, we follow [18] to set up a distance threshold of 2 pixels. Quality is a more general measure, which combines completeness and correctness into a single measure in the form:

$$quality = \frac{Comp * Corr}{Comp - Comp * Corr + Corr}. \tag{13}$$

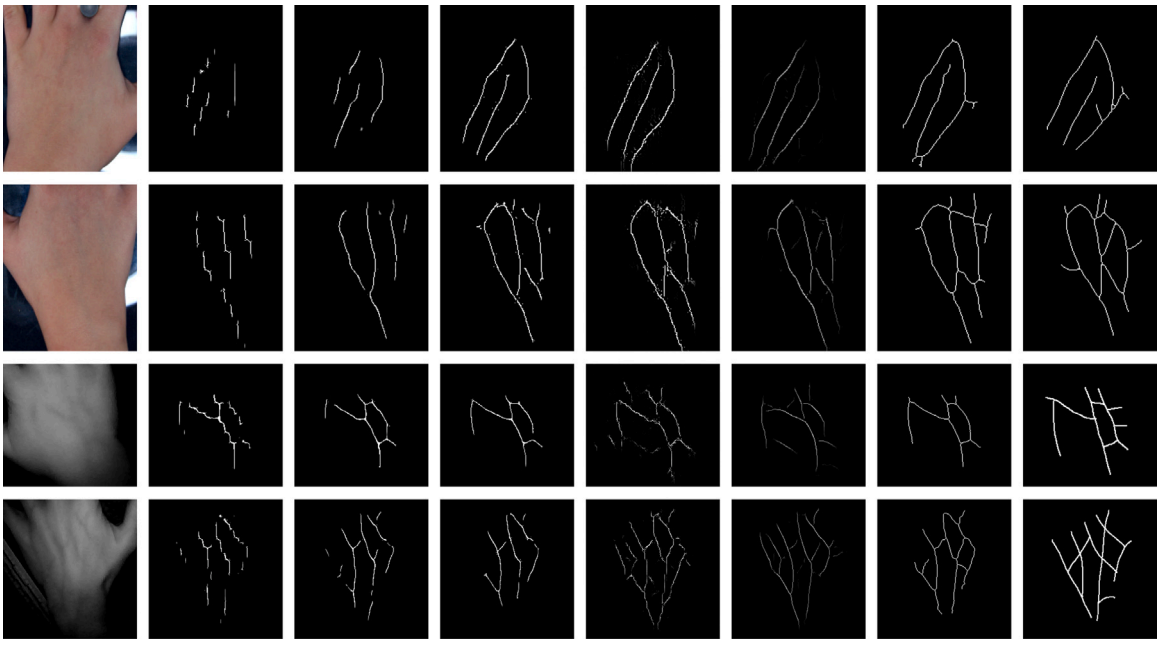


Fig. 6. Comparison with other deep learning based approaches on infrared images. From left to right: Original image, PSPNet+NMS, FPN+NMS, Unet+NMS, Topology-aware [18], DeepCrack [2], DODN-v2 and ground truth.

These metrics are designed specifically for linear structures. To obtain the precision–recall curves, we firstly assign different thresholds to the detected centerline response and obtain different binary maps. By matching binary maps and the ground truth centerline map, we can obtain a sequence of precision and recall pairs for the precision–recall curve. Then, the F-score ($2 * \frac{precision*recall}{precision+recall}$) can be computed as an overall metric of the precision–recall curve with a fixed threshold for every image. For experiments on SpaceNet dataset, we follow [42,43,53] to report TOPO metric [54], which calculates the average precision, recall, and F1-score based on randomly sampled seed locations across the entire region.

4.4. Performance evaluation

4.4.1. Comparison with state-of-the-art techniques

We compare our approach to previous state-of-the-art techniques [2,9,11,14,18,30,31,38,55]. Experimental results on the roads and cracks dataset are reported in Tables 1–3. Even without refinement (see), our DODN-v1 outperforms most of the state-of-the-art approaches while the refinement in DODN-v2 further improves the performance. For the experiments on the Massachusetts dataset, it is interesting to note that our final model achieves 11%, 4% and 11% improved results in terms of correctness, completeness, and quality, compared to competing methods. As shown in Tables 2 and 5, our approach consistently achieves the highest performance while maintaining faster inference speeds on the SpaceNet dataset. For the cracks dataset, our method also achieves state-of-the-art results. Figs. 3–5 show some examples of road centerline detection and crack detection respectively. It can be seen that the standard NMS results in poor localization as well as poor connectedness for centerline extraction. The proposed method models the geometric context of centerlines via an orientation representation, which allows one to associate centerline points with sinks. As highlighted in the red rectangle, we can see that our method has higher connectivity in various curvilinear structures, such as junctions and roundabouts in the road images. Our approach also performs well with crack detection in bright images and images with heavy shadows. We have also evaluated two state-of-the-art methods, Topological-aware [18] and DeepCrack [2] approaches, on the SuperID dataset. We used the authors' publicly available codes and followed their default experimental settings. Experimental results are reported in Table 4, demonstrating the superior performance of our approach on the SuperID dataset. Some experimental examples on RGB and infrared sets are displayed in Fig. 6. It is significant to note that the proposed approach trained on such a small-scale dataset can produce very similar vessel patterns to that of an expert human observer. Table 5 demonstrates the clear superiority of our approach in terms of inference speed compared to other state-of-the-art methods. Although DODN-v2 has a slightly longer inference time than DODN-v1, the inclusion of graph-cut refinement does not significantly impact the overall inference speed when compared to DODN-v1, especially considering the performance improvement achieved by the graph-cut refinement.

Table 1
Results on the Massachusetts dataset [48].

Method	Correctness	Completeness	Quality
Reg-AC [9]	0.254	0.348	0.172
PSPNet [14]	0.514	0.722	0.429
Unet [11]	0.623	0.751	0.515
TopoNet [30]	0.623	0.558	0.417
DRU [55]	0.606	0.570	0.416
JTFN [31]	0.687	0.634	0.492
PointScatter [38]	0.718	0.687	0.541
Top-Aware [18]	0.774	0.806	0.652
DODN-v1	0.864	0.827	0.732
DODN-v2	**0.887**	**0.843**	**0.761**

Table 2
Results on the SpaceNet dataset [49].

Method	Precision	Recall	F1	Correctness	Completeness	Quality
RNGDet [42]	0.909	0.733	0.811	0.723	0.728	0.570
RNGDet++ [43]	**0.913**	0.752	0.825	0.738	0.744	0.588
DODN-v1	0.897	0.761	0.815	0.784	0.746	0.619
DODN-v2	0.911	**0.771**	**0.828**	**0.802**	**0.753**	**0.635**

4.4.2. Effectiveness of DODN-v1 model

To evaluate the performance of our DODN-v1 model, several deep segmentation models with different backbones are used as baselines. Our baseline segmentation models include U-net [11], Linknet [13],

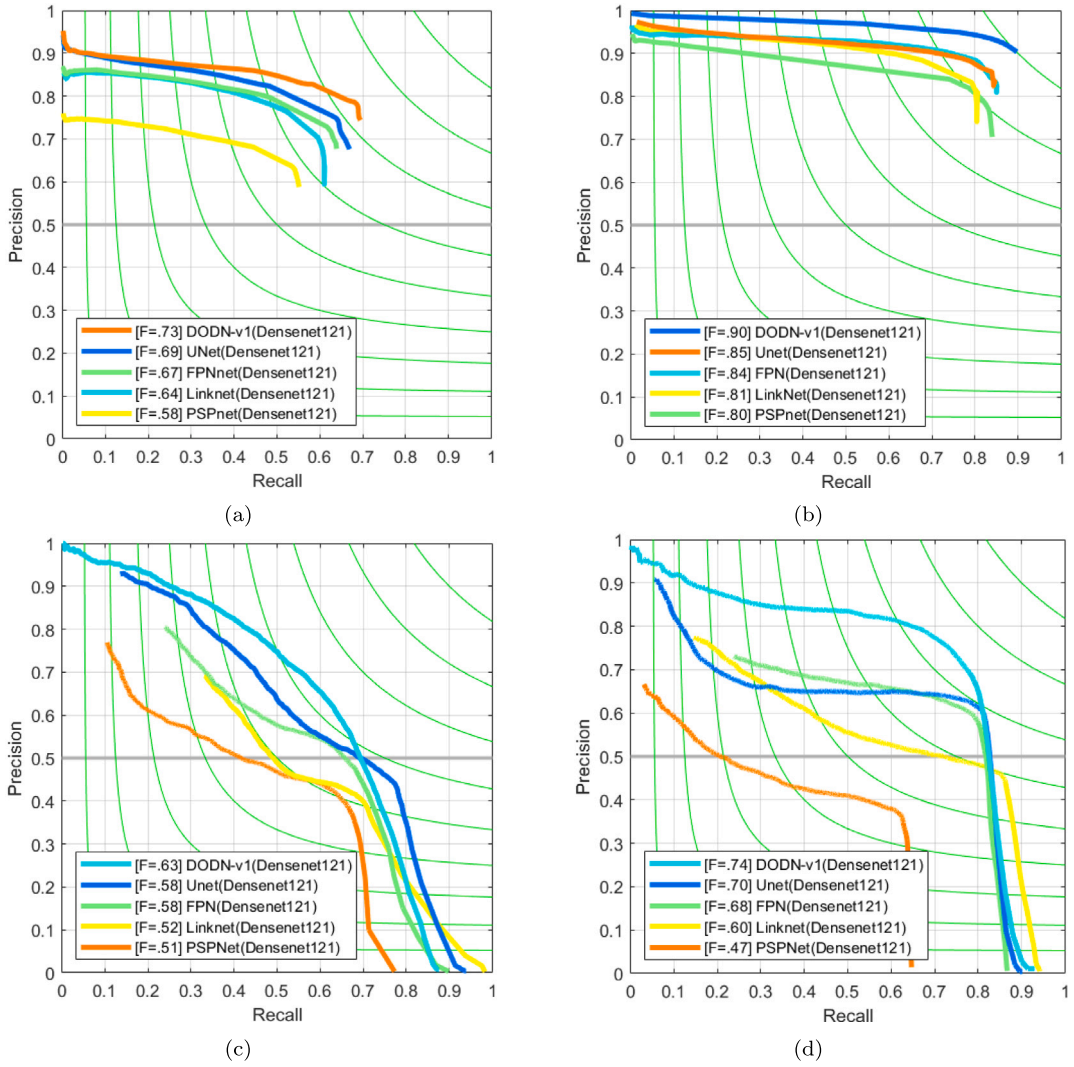


Fig. 7. Precision–Recall curves on datasets: (a) Roads, (b) Cracks, (c) RGB SuperID and (d) Infrared SuperID. The F-score of each method is shown in brackets, computed with a fixed threshold for every image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Results on the Cracks dataset [50].

Method	Correct.	Complet.	Quality
Unet [11]	0.840	0.881	0.755
TopoNet [30]	0.819	0.778	0.664
DRU [55]	0.848	0.775	0.680
JTFN [31]	0.883	0.874	0.783
Top-Aware [18]	0.766	0.904	0.708
DeepCrack [2]	0.900	**0.967**	0.873
DODN-v1	0.927	0.950	0.884
DODN-v2	**0.940**	0.951	**0.895**

Table 4
Results on the SuperID dataset.

Dataset	Method	Correct.	Complet.	Quality
Infrared	Top-Aware [18]	0.623	0.668	0.476
	DeepCrack [2]	0.688	0.622	0.485
	DODN-v1	0.821	0.717	0.621
	DODN-v2	**0.865**	**0.733**	**0.658**
Color	Top-Aware [18]	0.559	0.596	0.405
	DeepCrack [2]	0.629	0.573	0.428
	DODN-v1	0.636	**0.617**	0.456
	DODN-v2	**0.672**	0.616	**0.474**

Table 5
The inference time cost of all 382 SpaceNet testing images.

	Sat2Graph [53]	RNGDet [42]	RNGDet++ [43]	DODN-v1	DODN-v2
Time	1.15 h	1.22 h	1.88 h	0.32 h	0.35 h

PSPNet [14] and FPN [15]. For backbones, we choose Inceptionv3 [24], Resnet101 [25], Densenet121 [26], VGG16 and VGG19 [27]. Table 6 reports the F-scores of the different methods on the Cracks and Massachusetts dataset respectively. Table 7 reports the results on the RGB and infrared images of the SuperID dataset. We can observe that the proposed DODN model performs better than the baseline methods with the same Densenet121 backbone. The F-scores and Precision–Recall curves of our approach and these baseline methods are shown in Fig. 7. Compared to the Cracks and Massachusetts datasets, vessel detection from hand images is more challenging because of the low contrast between blood vessels and skin color, resulting from strong scattering and absorption by the tissue. From Table 7 and Figs. 7(c) and 7(d), we observe that the proposed method still achieves the best performance on both infrared and RGB images, showing the efficacy of the proposed approach for superficial vein pattern extraction. Fig. 8 compares our DODN-v1 model against the competing segmentation models on different backbones. The F-scores of standard U-net on most

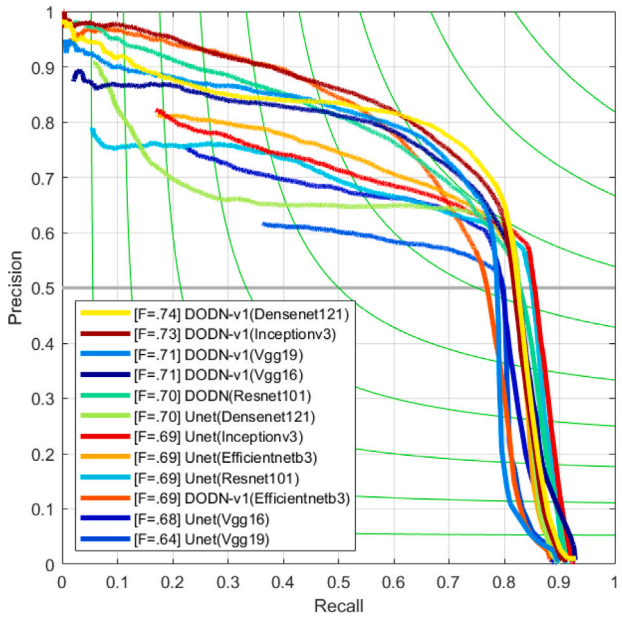


Fig. 8. Precision–Recall curves for showing the advantage of our DODN-v1 model on different backbones (using infrared images of SuperID dataset).

Table 6
F-scores for the proposed approach with the same backbone (Densenet) for the Cracks and Massachusetts datasets.

Setup	Roads	Cracks
PSPNet	0.58	0.80
LinkNet	0.64	0.81
FPN	0.67	0.84
UNet	0.69	0.85
DODN-v1	**0.73**	**0.90**
DODN-v2	**0.74**	**0.93**

Table 7
F-scores for the proposed approach with the same backbone (Densenet) for the RGB and infrared images in the SuperID dataset.

Setup	RGB	Infrared
PSPNet	0.51	0.47
LinkNet	0.52	0.60
FPN	0.58	0.68
UNet	0.58	0.70
DODN-v1	**0.63**	**0.76**
DODN-v2	**0.64**	**0.78**

backbones are significantly lower than our approach. In particular, for Densenet121, the DODN-v1 model outperforms standard Unet by 4% (F-score), demonstrating its effectiveness in detecting and localizing vessel centerlines. Note that, for standard Unet and other baselines, the vessel centerlines are detected from the score map by a non-maximum suppression operation [2,9,20].

4.4.3. Ablation analysis

As shown in Table 8, four ablation experiments were conducted to quantitatively evaluate the effectiveness of each component of the proposed approach. **Experiment D** is our final DODN-v2 model. **Experiment C** is our DODN-v1 model, which does not include the graph-cut based approach for centerline refinement. **Experiment B** removes uncertainty-weighted loss, weighting each task equally for training. **Experiment A** further removes geometric information during training, adopting only standard Dice loss for pixel classification, followed by NMS to thin the segmentation to centerlines. From Table 8, we observe that using geometric information in our base DODN model, improves

Table 8
Ablation study of the proposed approach on Roads, Cracks and infrared vessel dataset in terms of quality. GA = Geometric Awareness, UWL = Uncertainty Weighted Loss, GCR = Graph-Cut Refinement.

Experiment	GA	UWL	GCR	Dataset		
				Cracks	Roads	SuperID
A				0.854	0.681	0.564
B	✓			0.877	0.716	0.603
C (DODN-v1)	✓	✓		0.884	0.732	0.621
D (DODN-v2)	✓	✓	✓	**0.895**	**0.761**	**0.658**

the quality by 2.3%, 3.5% and 3.9% on Cracks, Roads and SuperID respectively. Experiment C (DODN-v1) outperforms B by 0.7%, 1.6% and 1.8%, showing the importance of using uncertainty to weight the losses. We can also observe from Experiment D (DODN-v2) that the overall performance is further boosted by 1.1%, 2.9% and 3.7% with the inclusion of the proposed Graph-cut based refinement. Fig. 9 shows that our algorithm has the ability to recover broken cracks, roads and vessels by considering all possible endpoint pairs. Our graph-cut based algorithm, whose hybrid energy function is constructed using the distance map representation, is able to determine endpoint pairs with a genuine connection and find the optimal path (path of least energy cost, shown in blue) between endpoints.

5. Conclusion

In this paper, a multitask Deep Orientated Distance-transform Network (DODN) is presented for accurate centerline detection, incorporating a Graph-cut based approach for improving the connectivity of the centerline. Experimental results show that DODN achieves excellent results, outperforming the current state-of-the-art for a diverse set of problems, including road mapping, crack detection and superficial vein centerline detection from both RGB and infrared images, demonstrating the generalizability of the proposed approach. Impressive results were achieved, improving over the state-of-the-art by 11%, 4% and 11% in terms of correctness, completeness, and quality on the Massachusetts road dataset. Further, this is the first approach to accurately detect vessel centerlines of hand dorsal images from RGB and infrared images, allowing us to directly build graph representations of vein patterns. Since our approach can produce continuous and thin centerlines of the object(s) of interest, it can benefit many real-world applications such as satellite navigation systems, engineering, forensic identification and medical investigation. In particular, our approach is able to facilitate downstream analysis and comparison on graph-structure data, which has strong applications in the analysis of medical and biomedical imaging of vascular structures and nerves, which is not currently feasible. Moreover, our method can serve as a strong baseline for researchers and engineers in the field.

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

The data that has been used is confidential.

Acknowledgments

The work in this publication is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 787768).

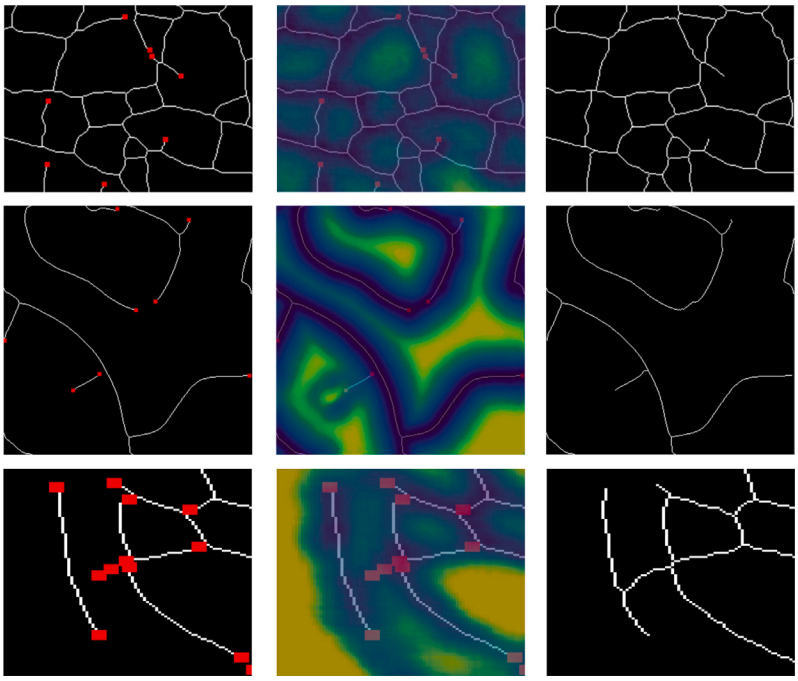


Fig. 9. Illustration of centerline refinement on Cracks (top row), Roads (middle row) and SuperID (bottom row) datasets. Red points indicate centerline endpoints. L-R: centerline, distance map representation and refinement. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

[1] G. Conte, P. Doherty, An integrated UAV navigation system based on aerial image matching, in: 2008 IEEE Aerospace Conference, 2008, pp. 1–10, http://dx.doi.org/10.1109/AERO.2008.4526556.

[2] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, S. Wang, Deepcrack: Learning hierarchical convolutional features for crack detection, IEEE Trans. Image Process. 28 (3) (2018) 1498–1512.

[3] B.M. Williams, D. Borroni, R. Liu, Y. Zhao, J. Zhang, J. Lim, B. Ma, V. Romano, H. Qi, M. Ferdousi, et al., An artificial intelligence-based deep learning algorithm for the diagnosis of diabetic neuropathy using corneal confocal microscopy: a development and validation study, Diabetologia 63 (2020) 419–430.

[4] Z. Jiang, H. Rahmani, P. Angelov, S. Black, B.M. Williams, Graph-context attention networks for size-varied deep graph matching, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2343–2352.

[5] A. Slot, Z.J. Geradts, The possibilities and limitations of forensic hand comparison, J. Forensic Sci. 59 (6) (2014) 1559–1567.

[6] A.H. Foruzan, R.A. Zoroofi, Y. Sato, M. Hori, A Hessian-based filter for vascular segmentation of noisy hepatic CT scans, Int. J. Comput. Assist. Radiol. Surg. 7 (2) (2012) 199–205.

[7] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, R. Kikinis, Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images, Med. Image Anal. 2 (2) (1998) 143–168.

[8] Y. Zheng, M. Loziczonek, B. Georgescu, S.K. Zhou, F. Vega-Higuera, D. Comaniciu, Machine learning based vesselness measurement for coronary artery segmentation in cardiac CT volumes, in: Medical Imaging 2011: Image Processing, Vol. 7962, International Society for Optics and Photonics, 2011, p. 79621K.

[9] A. Sironi, E. Türetken, V. Lepetit, P. Fua, Multiscale centerline detection, IEEE Trans. Pattern Anal. Mach. Intell. 38 (7) (2015) 1327–1341.

[10] J.M. Wolterink, R.W. van Hamersvelt, M.A. Viergever, T. Leiner, I. Išgum, Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier, Med. Image Anal. 51 (2019) 46–60.

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

[12] C. Becker, R. Rigamonti, V. Lepetit, P. Fua, Supervised feature learning for curvilinear structure segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 526–533.

[13] A. Chaurasia, E. Culurciello, Linknet: Exploiting encoder representations for efficient semantic segmentation, in: 2017 IEEE Visual Communications and Image Processing (VCIP), IEEE, 2017, pp. 1–4.

[14] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.

[15] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.

[16] X. Wang, X. Jiang, J. Ren, Blood vessel segmentation from fundus image by a cascade classification framework, Pattern Recognit. 88 (2019) 331–341.

[17] P. Bibiloni, M. González-Hidalgo, S. Massanet, A survey on curvilinear object segmentation in multiple applications, Pattern Recognit. 60 (2016) 949–970.

[18] A. Mosinska, P. Marquez-Neila, M. Koziński, P. Fua, Beyond the pixel-wise loss for topology-aware delineation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3136–3145.

[19] T. Zhou, Y. Yang, W. Wang, Differentiable multi-granularity human parsing, IEEE Trans. Pattern Anal. Mach. Intell. (2023) 1–14, http://dx.doi.org/10.1109/TPAMI.2023.3239194.

[20] P. Dollár, C.L. Zitnick, Fast edge detection using structured forests, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2014) 1558–1570.

[21] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 1998, pp. 130–137.

[22] M.W. Law, A.C. Chung, Three dimensional curvilinear structure detection using optimally oriented flux, in: European Conference on Computer Vision, Springer, 2008, pp. 368–382.

[23] M.W. Law, A.C. Chung, An oriented flux symmetry based active contour model for three dimensional vessel segmentation, in: European Conference on Computer Vision, Springer, 2010, pp. 720–734.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

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

[26] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[27] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[28] X.-F. Wang, D.-S. Huang, H. Xu, An efficient local Chan–Vese model for image segmentation, Pattern Recognit. 43 (3) (2010) 603–618.

[29] R.C. Dubes, A.K. Jain, S.G. Nadabar, C.-C. Chen, MRF model-based algorithms for image segmentation, in: [1990] Proceedings. 10th International Conference on Pattern Recognition, Vol. 1, IEEE, 1990, pp. 808–814.

[30] X. Hu, F. Li, D. Samaras, C. Chen, Topology-preserving deep image segmentation, Adv. Neural Inf. Process. Syst. 32 (2019).

[31] M. Cheng, K. Zhao, X. Guo, Y. Xu, J. Guo, Joint topology-preserving and feature-refinement network for curvilinear structure segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 7147–7156.

[32] L. Yang, G. Yang, X. Xi, K. Su, Q. Chen, Y. Yin, Finger vein code: From indexing to matching, IEEE Trans. Inf. Forensics Secur. 14 (5) (2018) 1210–1223.

[33] Z. Guo, J. Bai, Y. Lu, X. Wang, K. Cao, Q. Song, M. Sonka, Y. Yin, Deepcenterline: A multi-task fully convolutional network for centerline extraction, in: International Conference on Information Processing in Medical Imaging, Springer, 2019, pp. 441–453.

[34] L. Yang, G. Yang, Y. Yin, X. Xi, Finger vein recognition with anatomy structure analysis, IEEE Trans. Circuits Syst. Video Technol. 28 (8) (2017) 1892–1905.

[35] G. Máttyus, W. Luo, R. Urtasun, Deeproadmapper: Extracting road topology from aerial images, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3438–3446.

[36] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, C. Pan, Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network, IEEE Trans. Geosci. Remote Sens. 55 (6) (2017) 3322–3337.

[37] M. Yang, Y. Yuan, G. Liu, SDUNet: Road extraction via spatial enhanced and densely connected UNet, Pattern Recognit. 126 (2022) 108549.

[38] D. Wang, Z. Zhang, Z. Zhao, Y. Liu, Y. Chen, L. Wang, PointScatter: Point set representation for tubular structure extraction, in: European Conference on Computer Vision, Springer, 2022, pp. 366–383.

[39] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. (6) (1986) 679–698.

[40] C. Godard, O. Mac Aodha, M. Firman, G.J. Brostow, Digging into self-supervised monocular depth estimation, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 3828–3838.

[41] A. Kendall, Y. Gal, R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7482–7491.

[42] Z. Xu, Y. Liu, L. Gan, Y. Sun, X. Wu, M. Liu, L. Wang, Rngdet: Road network graph detection by transformer in aerial images, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–12.

[43] Z. Xu, Y. Liu, Y. Sun, M. Liu, L. Wang, RNGDet++: Road network graph detection by transformer with instance segmentation and multi-scale features enhancement, IEEE Robot. Autom. Lett. (2023).

[44] Y.Y. Boykov, M.-P. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in ND images, in: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, Vol. 1, IEEE, 2001, pp. 105–112.

[45] C. Rother, V. Kolmogorov, A. Blake, " GrabCut" interactive foreground extraction using iterated graph cuts, ACM Trans. Graph. (TOG) 23 (3) (2004) 309–314.

[46] M. Khosravi, R.W. Schafer, Template matching based on a grayscale hit-or-miss transform, IEEE Trans. Image Process. 5 (6) (1996) 1060–1066.

[47] G.R. Waissi, Network flows: Theory, algorithms, and applications, 1994.

[48] V. Mnih, Machine Learning for Aerial Image Labeling, University of Toronto, Canada, 2013.

[49] A. Van Etten, D. Lindenbaum, T.M. Bacastow, Spacenet: A remote sensing dataset and challenge series, 2018, arXiv preprint arXiv:1807.01232.

[50] Q. Zou, Y. Cao, Q. Li, Q. Mao, S. Wang, CrackTree: Automatic crack detection from pavement images, Pattern Recognit. Lett. 33 (3) (2012) 227–238.

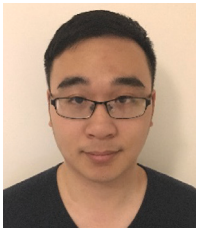
[51] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[52] C. Wiedemann, C. Heipke, H. Mayer, O. Jamet, Empirical evaluation of automatically extracted road axes, in: Empirical Evaluation Techniques in Computer Vision, Vol. 12, Citeseer, 1998, pp. 172–187.

[53] S. He, F. Bastani, S. Jagwani, M. Alizadeh, H. Balakrishnan, S. Chawla, M.M. Elshrif, S. Madden, M.A. Sadeghi, Sat2graph: Road graph extraction through graph-tensor encoding, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16, Springer, 2020, pp. 51–67.

[54] J. Biagioni, J. Eriksson, Inferring road maps from global positioning system traces: Survey and comparative evaluation, Transp. Res. Rec. 2291 (1) (2012) 61–71.

[55] W. Wang, K. Yu, J. Hugonot, P. Fua, M. Salzmann, Recurrent U-Net for resource-constrained segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 2142–2151.



Zheheng Jiang received the M.Sc. degree in Software Development from Queen's University of Belfast, Belfast, U.K and the Ph.D. degree in Computer Science from University of Leicester, Leicester, U.K. He is currently a Senior Research Associate in Lancaster University. His research interests include image processing, computer vision and machine learning.

Hossein Rahmani received the B.Sc. degree in computer software engineering from Isfahan University of Technology, Isfahan, Iran, in 2004, the M.Sc. degree in software engineering from Shahid Beheshti University, Tehran, Iran in 2010, and the Ph.D. degree from the University of Western Australia, in 2016. He has published several papers in top conferences and journals such as CVPR, ICCV, ECCV, and the IEEE Transactions on Pattern Analysis and Machine Intelligence. He is currently an associate professor (Senior Lecturer) in the School of Computing and Communications at Lancaster University. Before that he was a Research Fellow in the School of Computer Science and Software Engineering, University of Western Australia. He is Associate Editor of IET Computer Vision. His research interests include computer vision, video analysis, pose estimation and machine learning.

Plamen Angelov holds Ph.D. and DSc degrees and is Chair Professor of Intelligent Systems, Director of Research at the School of Computing and Communications and Director of the Lancaster Intelligent, Robotic and Autonomous systems (LIRA) Centre. Prof. Angelov is a Fellow of the IEEE, of the IET and of ELLIS and Governor of the International Neural Networks Society. He has over 380 publications cited over 13600 times with an h-index of 62 and active research portfolio in the area of interpretable deep learning and computational intelligence. He is recipient of numerous awards including the Dennis Gabor award (2020) for "outstanding contributions to engineering applications of neural networks", IEEE awards 'For outstanding Services' (2013 and 2017).

Dr. Ritesh Vyas is currently working as an Assistant Professor in Pandit Deendayal Energy University (PDEU), Gandhinagar, India. Earlier, he was a Senior Research Associate in Lancaster University, UK from 2020-2022. He received his Ph.D. degree from National Institute of Technology Delhi, India, in 2020. He completed his MTech and BTech degrees in ECE from YMCA University of Science & Technology (not known as JCBUST), India, and Kurukshetra University, India, in 2012 and 2009, respectively. He is a Senior Member of IEEE, and a life member of IUPRAI and ISTE. He is also an Associate Fellow of Higher Education Academy (AFHEA) of the United Kingdom. His areas of interest include biometrics, computer vision, image processing, pattern recognition, artificial intelligence, and machine learning.

Huiyu Zhou received a Bachelor of Engineering degree in Radio Technology from Huazhong University of Science and Technology of China, and a Master of Science degree in Biomedical Engineering from University of Dundee of United Kingdom, respectively. He was awarded a Doctor of Philosophy degree in Computer Vision from Heriot-Watt University, Edinburgh, United Kingdom. Dr. Zhou currently is a full Professor at School of Computing and Mathematical Sciences, University of Leicester, United Kingdom. He has published over 400 peer reviewed papers in the field. His research work has been or is being supported by UK EPSRC, ESRC, AHRC, MRC, EU, Royal Society, Leverhulme Trust, Puffin Trust, Invest NI and industry.

Professor Dame Sue Black is an anatomist and forensic anthropologist. She was Pro Vice Chancellor for Engagement at Lancaster University and is currently President of St. John's College, Oxford. She is a forensic expert witness.

Bryan M. Williams received the Ph.D. degree in mathematics from the University of Liverpool, Liverpool, U.K., in 2015. He is currently a Lecturer in biometrics and human identification with Lancaster University, Lancaster, U.K. Prior to that, he was a Research Associate with Universitat des Saarlandes, Saarbrücken, Germany. His research interests include developing computer vision research and