Adaptive Candidate Region Detection Method for Few-Shot Mesh Structures (Postprint)

Authors: Mou Lei, Chen Li

Date: 2018-10-11T00:00:00+00:00

Abstract

In image mesh structure detection tasks with limited labeled data, the detection performance of object detection models requiring large-scale training data degrades significantly. For region proposal-based object detection models, inference time increases proportionally with the number of targets. When such models generate a fixed number of candidate boxes while the quantity of mesh structure targets varies across images, this results in unnecessary computational overhead. To address this problem, we propose an adaptive candidate region detection method for few-shot mesh structures, based on density analysis of mesh structure targets in training samples and their feature distribution within images. This method obtains abundant training samples through a binary label map annotation approach and selects an appropriate number of candidate boxes via the adaptive candidate region mechanism. Compared with the baseline model, it accelerates detection speed with negligible accuracy loss, particularly demonstrating superior performance in data with sparse target densities.

Full Text

Preamble

Proposal Adaptive Detection Method for Small Sample Reticular Structures

Lei Moua,b, Li Chena,b

aSchool of Computer Science & Technology, bHubei Province Key Laboratory of Intelligent Information Processing & Real-time Industrial System, Wuhan University of Science & Technology, Wuhan 430065, China

Abstract: The detection performance of object detection models that require large amounts of training data degrades significantly when applied to reticular structure detection tasks with limited labeled data. For region proposal-based

detection models, detection time increases with the number of targets. If these models generate a fixed number of proposals regardless of the actual number of reticular structures in different images, it results in unnecessary time consumption. To address this issue, we propose a proposal adaptive detection method for small sample reticular structures by analyzing the density distribution of reticular targets in training samples and their feature distributions. This method generates extensive training samples through a binary-value labeled map marking approach and selects an appropriate number of proposals via an adaptive region proposal mechanism. Compared with unmodified models, our approach accelerates detection speed with negligible accuracy loss, with particularly pronounced advantages on data containing few objects.

Keywords: small sample; reticular structure; sample labeling; proposal adaptive

0 Introduction

Deep learning has achieved remarkable success in visual recognition tasks [1,2,3]. However, these models require massive amounts of labeled data and numerous training iterations to optimize their large parameter spaces. In practice, we often encounter scenarios with limited or no labeled data, where manual annotation is labor-intensive and costly. To address this data scarcity problem, various techniques have emerged, including data augmentation and few-shot learning [4].

Current deep learning-based object detection methods fall into two main categories: region proposal-based models and regression-based models [5]. Region proposal-based models represent a two-stage detection approach, while regression-based models constitute a single-stage approach. Compared with single-stage networks, two-stage networks achieve higher precision. Early methods such as R-CNN [7], Fast R-CNN [8], and the subsequent Faster R-CNN [9] have driven significant progress in object detection. In contrast, regression-based models employ a regression paradigm that requires pre-defined default boxes to establish relationships between predicted boxes, default boxes, and ground-truth object boxes for training [20,21], exemplified by YOLO [10] and SSD [11] among other excellent models. While the first category demonstrates superior detection performance, it suffers from slower detection speeds compared with the second category; conversely, the second category trades detection performance for significant speed improvements [1,2].

Region proposal-based models generate proposals through a Region Proposal Network (RPN) and then feed these proposals into fully connected layers for multi-class classification and regression. Due to the substantial number of parameters in fully connected layers, each neuron participates in computation during inference. As input to these layers increases, prediction time grows exponentially [12]. Different reticular structure samples contain varying numbers

of detection targets. Using a fixed output as input to fully connected layers for different data results in more proposals than actual targets, causing unnecessary computational overhead. Additionally, during annotation of reticular structures with interwoven, interconnected, and tree-like topological distributions, proposal labels must be continuously distributed to cover entire structures, making manual annotation extremely time-consuming and prone to imprecision.

To address these dual challenges, we propose a proposal adaptive detection method for small sample reticular structures. This approach employs a two-level preprocessing scheme that partitions reticular structures into linear collections and further subdivides them into block collections, ultimately generating numerous precisely annotated proposals to achieve data augmentation. Simultaneously, we design a proposal adaptive algorithm that selects an appropriate number of proposals as RPN output based on target density in training data, thereby reducing computational costs in fully connected layers. This achieves the dual objectives of accelerating prediction speed while maintaining model performance. Experimental results demonstrate that our method exhibits excellent performance for small sample reticular structure detection, providing significant improvements in detection time across different datasets.

1.2 Few-Shot Learning

In recent years, extensive research has addressed scenarios with limited labeled data. Early transfer learning techniques [13] pioneered few-shot learning by applying knowledge learned from source domains to target domain training. While transfer learning significantly improves deep model performance, substantial room for enhancement remains. Current deep learning methods extract target features through neural networks for classification and regression, yet most approaches focus on sample independence without considering similarity relationships between samples of the same class or between targets. This limitation has led to innovative architectures such as Siamese neural networks [14], matching networks [15], prototypical networks [16], and relation networks [17], which concentrate on learning inter-sample similarity relationships. These networks enable models to compare sample similarities, thereby achieving classification and recognition with minimal labeled data through similarity-based comparisons.

1.3 Region Proposal-Based Models

Region proposal-based models constitute a two-stage deep neural network architecture, in contrast to the single-stage nature of regression-based models. Two-stage networks achieve higher precision than their single-stage counterparts. Early methods like R-CNN [7] and Fast R-CNN [8] drove substantial progress

in object detection, while Faster R-CNN [9] further improved both speed and accuracy. The evolution from Selective Search [7] to Region Proposal Networks (RPN) [9] accelerated prediction speed, and incorporating fully connected layers after the RPN for classification and regression enhanced prediction precision.

1.4 Reticular Structures

Reticular structures are ubiquitous in real-world applications. Detecting and segmenting reticular structures such as human venous vessels and retinal blood vessels facilitates early disease detection and reduces health risks. Similarly, detecting satellite roads and ground cracks mitigates safety hazards associated with manual inspection [22]. Current segmentation methods for reticular structures primarily build upon U-Net [18], while detection methods remain relatively scarce. Due to their complex structures and high precision requirements, reticular structure detection mainly relies on region proposal-based models.

2.1 Few-Shot Reticular Structure Detection

Deep learning models require extensive labeled data for training. In practice, when only limited labeled data is available, model performance degrades significantly. Data augmentation techniques provide a solution for training models with scarce labeled data by increasing data volume without altering image categories, thereby improving model generalization. Geometric augmentation methods include flipping, translation, and cropping, while pixel-level transformations encompass color jittering and noise addition.

In two-dimensional reticular structure images, numerous interwoven and interconnected linear collections exist. For instance, retinal vessels in Figure 1: see original paper exhibit a tree-root distribution centered at the optic disc, while road networks in Figure 1: see original paper display elongated, noodle-like patterns. Reticular structure images feature extensive, interconnected, and widely distributed networks. Inspired by Kandinsky' s theory [19] that points, lines, and planes constitute fundamental elements of planar space, we conceptualize reticular structures as "planes" in two-dimensional images. Through hierarchical decomposition, we first partition reticular structures into linear structures ("lines"), then further divide linear structures into block structures ("points"). This layered processing approach, combined with a binary label map annotation method, generates training proposals. The entire hierarchical process is illustrated in [Figure 2: see original paper], where yellow borders in the left dashed region represent linear collections from the first decomposition, green borders represent block collections from the second decomposition, and subsequent steps calculate IoU to generate proposals for annotation in training samples.

The binary label map annotation method proceeds as follows: (a) Extract skeletons from foreground regions in the binary label map (Figure 3: see original paper) and use skeleton coordinates as proposal center coordinates, selecting the first proposal; (b) Locate the next proposal center coordinate on the skeleton at a distance equal to the current proposal' s side length; (c) Generate a new proposal using the current proposal' s side length and the newly generated center coordinate; (d) Calculate the ratio between the ground-truth area covered by the new proposal and the new proposal' s area (Figure 3: see original paper); (e) If the ratio falls within [S_min, 1], complete new proposal generation and repeat steps (b)-(e); (f) If the ratio is below S_min, maintain the center coordinate while scaling the proposal side length by factor m and regenerate; (g) If the ratio exceeds 1, replace the current side length with the new one in step (f) and continue; (h) Repeat steps (b)-(g) until completion.

The relationship between new proposal center coordinates and current proposal dimensions is given by:

$$\begin{cases} x_{i+1} = x_i + \min\{w_i, h_i\} \\ y_{i+1} = y_i + \min\{w_i, h_i\} \end{cases}$$

where (x_{i+1}, y_{i+1}) represents new proposal center coordinates, (x, y) denotes skeleton coordinates, (x_i, y_i) indicates current proposal center coordinates, and w_i, h_i are current proposal width and height. For a current proposal $P_i(x_i, y_i, w_i, h_i)$, the binary label map annotation method generates new proposals at each iteration as:

$$P_{i+1} = \mathrm{IoU}(L, P_i) \cdot C + \lambda \cdot L_{\mathrm{IoU}}(P_i, P_{i+1})$$

Equation (3) indicates that new proposal center coordinates depend on the current proposal, with size determined by new parameters. The final effect of the binary label map annotation method is shown in Figure 3: see original paper. When reticular structure samples are limited, this method annotates numerous proposals to augment training samples, providing a solution for data scarcity.

2.2 Adaptive Region Proposals

During proposal generation, N_0 anchors are randomly generated in an image, with each anchor undergoing classification and regression to output the top N proposals. In Faster R-CNN [15], the loss function simplifies to:

$$L(p_i, t_i) = L_{\mathrm{cls}}(p_i, p_i^*) + \lambda L_{\mathrm{reg}}(t_i, t_i^*)$$

where p_i represents the probability of anchor i being predicted as an object, p_i^* is the ground-truth label for anchor i (with $p_i^* = 1$ for positive samples and

$p_i^* = 0$ for negative samples), and t_i and t_i^* are coordinates related to anchor i. The optimization process requires classifying and regressing every anchor in each image, ultimately selecting the top N proposals.

Features from the proposal generation stage are fed into fully connected layers (FC) for classification and regression. FC layers act as classifiers in neural networks, mapping learned distributed feature representations to sample label spaces [20]. As shown in [Figure 4: see original paper], if the RPN generates n proposals of size $m \times m$, the FC layer input size becomes $n \times m \times m$, with output size $k \times n$. The FC layer computation is:

$$y = Wx + b$$

where W is the weight matrix and b is the bias. Computing all neurons requires $n \times m \times m \times k$ operations, yielding time consumption $T = n \times t_0$, which exhibits linear relationship with the number of proposals n.

To solve the problems of excessive prediction time due to fixed proposal counts and missed detections from insufficient proposals, we propose an adaptive region proposal algorithm that selects appropriate proposal numbers for different data.

In Section 2.1, each image generates numerous proposals through the binary label map method. We define block density ρ_i as the ratio between annotated proposals and image area:

$$\rho_i = \frac{N_{\text{proposal}}}{S_{\text{img}}}$$

representing the number of proposals per unit area. Higher block density indicates more targets, enabling target quantity estimation. Since training images exhibit varying block densities and the RPN accepts arbitrary image sizes while requiring complete target detection, we employ the maximum block density $\rho_{\max}$ across training samples:

$$\rho_{\max} = \max\{\rho_i | i \in [1, N]\}$$

During inference, given a test image' s dimensions (height H, width W), we calculate the appropriate N value for the proposal generation stage:

$$N = \rho_{\max} \times W \times H$$

This approach yields proposals that satisfy performance requirements while accelerating prediction. When block density is very small, the adaptively computed N may be insufficient for actual target counts. Moreover, models are sensitive to N values when target quantities are minimal. To prevent deviations, we set a minimum $N_{\min} = 50$. The final adaptive proposal count becomes:

$$N_{\text{adaptive}} = \max\{N_{\min}, \rho_{\max} \times W \times H\}$$

3 Experimental Results and Analysis

3.1 Experimental Data

Reticular structure detection models output numerous block collections, each containing positional information and category labels. Our method leverages block structures to reflect reticular distribution, making accurate block classification and localization paramount. We evaluate model performance using two datasets:

1. **STARE retinal image database** [22]: Training samples are 3504×2366 resolution images. Due to their large size, each image is divided into nine 1168×779 sub-images, yielding 90 training images and 36 test images.
2. **Satellite road dataset**: Training samples consist of 12 images at 1862×900 resolution, with 12 test images.

3.2 Mean Average Precision

In object detection, mean Average Precision (mAP) serves as the standard accuracy metric. For each category C, Precision-Recall curves are plotted, with the area under the curve representing Average Precision (AP). mAP is the mean of AP values across multiple categories, where higher mAP indicates more accurate detection and more complete reticular structure coverage. Detection results using our adaptive method are shown in [Figure 7: see original paper], achieving coverage rates of 74.8% (STARE) and 93.4% (satellite roads), effectively capturing fundamental distribution characteristics for subsequent retinal vessel segmentation and road detection.

3.3 Sample Augmentation

The binary label map annotation method generates extensive training samples, addressing the small sample challenge. When generating new proposal centers, the method locates points on skeleton coordinates at distance $m \times \min\{w_i, h_i\}$ from the current proposal. As shown in [Figure 6: see original paper], points A, B, and C represent three different proposal centers, with black dashed lines indicating inter-center distances. Given short side lengths l_A, l_B, l_C, the distance between A and B is $l_A \times m$ when $m = 1$, ensuring B remains within proposal A' s range. This prevents insufficient overlap that causes missed annotations (Figure 6: see original paper) and excessive overlap that leads to duplicate annotations (Figure 6: see original paper). Through sample augmentation, we obtain approximately 14,000 and 600 annotated proposals for the two datasets, respectively.

3.4 Adaptive Proposal Selection

We investigate the relationship between proposal count N (varied from 1 to 400) and model performance/prediction time on both datasets. [Figure 8: see original paper] shows performance curves, where model accuracy converges to maximum as N increases, plateauing at $N = 159$ for STARE and $N = 50$ for satellite roads. [Figure 9: see original paper] demonstrates that prediction time scales linearly with proposal count. The adaptive method calculates N values that correspond to the critical point where model performance is fully utilized while prediction speed is maximized. When N exceeds this critical value, increasing proposals no longer improves performance but reduces speed; when N falls below it, performance degrades substantially.

[Figure 5: see original paper] illustrates detection results on the same image with $N = 50$, $N = 100$, and $N = 300$, showing severe missed detections at $N = 50$ while $N = 100$ and $N = 300$ both detect 98 targets.

Comparative results between our adaptive method and conventional detectors are presented in . The adaptive approach reduces RPN output substantially, achieving approximately 43% higher accuracy than regression-based models (YOLO-v3, SSD) while reducing prediction time by ~57% compared with Faster R-CNN without accuracy loss. Despite being region proposal-based and incorporating FC layers, our method approaches the speed of YOLO and SSD while far surpassing their accuracy.

---

4 Conclusion

By analyzing proposal generation mechanisms and fully connected layer computations, we propose an adaptive region proposal method that selects optimal proposal counts, accelerating prediction without compromising performance. Additionally, based on reticular structures' distribution and structural characteristics, we introduce a binary label map annotation approach that generates extensive training proposals from limited labeled samples, achieving effective data augmentation. Combining these two innovations, our proposal adaptive detection method for small sample reticular structures demonstrates excellent performance on STARE and satellite road datasets, significantly improving prediction speed while maintaining accuracy in few-shot scenarios.

References

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]// Proc of International Conference on Neural Information Processing Systems. Curran Associates Inc., 2012.

[2] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

[3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014.

[4] Xian Yongqin, Lampert C H, Schiele B, et al. Zero-shot learning: a comprehensive evaluation of the good, the bad and the ugly [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018.

[5] 吴帅, 徐勇, 赵东宁. 基于深度卷积网络的目标检测综述 [J]. 模式识别与人工智能, 2018, 31(4): 335-346. (Wu Shuai, Xu Yong, Zhao Dongning. Survey of object detection based on deep convolutional network [J]. Pattern Recognition and Artificial Intelligence, 2018, 31(4): 335-346.)

[6] 李旭冬, 叶茂, 李涛. 基于卷积神经网络的目标检测研究综述 [J]. 计算机应用研究, 2017, 34(10): 2881-2886. (Li Xudong, Ye Mao, Li Tao. Review of object detection based on convolutional neural networks [J]. Application Research of Computers, 2017, 34(10): 2881-2886.)

[7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.

[8] Girshick R. Fast R-CNN [C]// Proc of IEEE International Conference on Computer Vision. 2015: 1440-1448.

[9] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]// Advances in Neural Information Processing Systems. 2015: 91-99.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.

[11] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot multibox detector [C]// Proc of European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[12] Chen Liangchieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS [J]. IEEE Trans on Pattern Analysis and Machine Intelligence 2018, 40(4).

[13] Pan J S, Yang Qiang. A survey on transfer learning [J]. IEEE Trans on knowledge and data engineering, 2010, 2(10): 1345-1359.

[14] Gregory K, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition [C]// Proc of ICML Deep Learning Workshop.

[15] Oriol V, Blundell C, Lillicrap T, et al. Matching networks for one shot learning [C]// Advances in Neural Information Processing Systems. 2016.

[16] Touretzky D S, Mozer M C, Hasselmo M E. Advances in neural information processing systems [C]// Cambridge: MIT Press, 1996.

[17] Sung F, Yang Yongxin, Zhang Li, et al. Learning to compare: relation network for few-shot learning [EB/OL]. (2018-03-27). https://arxiv.org/abs/1711.06025.

[18] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation [C]// Proc of International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.

[19] Kandinsky, Wassily, Hilla Rebay. Dover Publications. Point and line to plane [M]. Courier Corporation, 1979.

[20] Viola P. Robust real-time object detection [C]// Proc of International Workshop on Statistical and Computational Theories of Vision: Modeling, Learning, Computing, and Sampling. 2013: 87.

[21] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40(6): 1229-1251. (Zhou Feiyan, JIN Linpeng, Dong Jun. Review of Convolutional Neural Network [J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.)

[22] 王海, 蔡英凤, 贾允毅, 等. 基于深度卷积神经网络的场景自适应道路分割算法 [J]. 电子与信息学报, 2017, 39(2): 263-26. (Wang Hai, Cai Yingfeng, Jia Yunyi, et al. Scene Adaptive Road Segmentation Algorithm Based on Deep Convolutional Neural Network [J]. Journal of Electronics & Information Technology, 2017, 39(2): 263-26.)

Note: Figure translations are in progress. See original paper for figures.