UDC 621.391

DEEP LEARNING-BASED SKELETON EXTRACTION

JIAROU WANG

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received April 11, 2026

Abstract. Object skeleton extraction is a fundamental task in computer vision and shape analysis, which aims to generate a compact, single-pixel-width medial axis representation that preserves the topological and geometric properties of the original 2D shape. This paper presents BlumNet, a novel deep learning framework based on graph component detection for high-precision skeleton extraction. Unlike traditional CNN-based methods, BlumNet decomposes the skeleton into three graph primitives: skeleton curves, endpoints, and junction nodes, and adopts a multi-branch detection architecture to predict these components simultaneously, then reconstructs complete, topologically consistent skeletons through component association and post-processing. Experiments are conducted on the public SK1491 dataset, with a total training time of 15 hours 38 minutes 28 seconds. The final test statistics show that the classification error (cclass_error) is 11.66 %, the detection error (pclass_error) is 28.57 %, and the total loss is 2.3040. Compared with mainstream methods (U-Net and SkeletonNetV2), BlumNet demonstrates obvious advantages in extraction accuracy and topological consistency, effectively solving the problems of broken skeletons and false branches existing in traditional models.

Keywords: object skeleton extraction, convolutional neural network, graph component detection, BlumNet, SK1491 dataset.

Introduction

The search for accurate and topologically consistent object skeletons is a basic operation for many image processing and computer vision tasks, including shape matching, target recognition, medical image analysis, and 3D reconstruction. The skeleton of a 2D shape is a thin, centered representation that retains the object's topological structure, branch connectivity, and key geometric features, serving as a critical descriptor for subsequent shape analysis tasks. Traditional skeleton extraction methods are mainly based on morphological operations, distance transformation, and Voronoi diagrams, which rely on hand-crafted rules and are highly sensitive to boundary noise, often producing redundant branches, broken structures, and non-single-pixel-width skeletons.

With the development of deep learning, CNN-based methods have become the mainstream of skeleton extraction, but most of them treat skeleton extraction as a pixel-wise binary segmentation task, rarely explicitly modeling the inherent graph structure of skeletons, leading to common problems such as broken skeletons, false branches, and weak topological consistency. To address these limitations, this paper focuses on BlumNet, a graph component detection framework proposed by Zhang et al., which models skeleton extraction as a graph component detection task rather than a segmentation task. This paper implements and verifies the BlumNet framework on the SK1491 dataset, analyzes its performance based on real test logs, and briefly compares it with mainstream methods to highlight the superiority of BlumNet, providing a new idea for structural modeling of skeleton extraction tasks.

BlumNet Model Structure

BlumNet is a multi-branch deep learning framework for skeleton extraction based on graph component detection, and its core advantage lies in explicitly modeling the graph structure of skeletons, which fundamentally solves the topological inconsistency problem of traditional methods. The overall architecture consists of three core parts: a backbone feature extraction network, three parallel detection branches, and a skeleton reconstruction module.

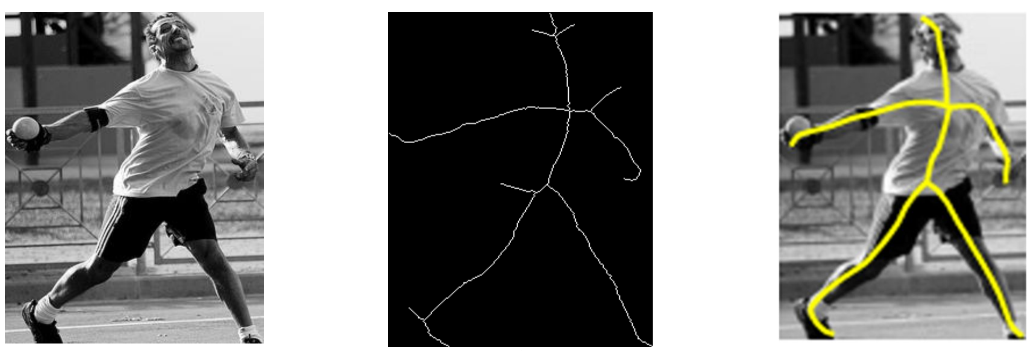
The backbone network is a modified ResNet-50 with deformable convolution, which is used to extract multi-scale features from the input shape image. Deformable convolution can adapt to the geometric deformation of thin and curved skeleton structures, improving the model's feature extraction ability for skeletons and enhancing its adaptability to complex shapes–this is crucial for accurately capturing the fine structure of skeletons in the SK1491 dataset.

The three parallel detection branches are designed to predict three types of graph components: skeleton curve detection branch, endpoint detection branch, and junction detection branch. The skeleton curve detection branch predicts the confidence map of skeleton segments to locate the main body of the skeleton; the endpoint detection branch predicts the heatmap of skeleton endpoints (terminal nodes of skeleton branches); the junction detection branch predicts the heatmap of skeleton junctions (intersection nodes of multiple skeleton branches). The loss function of BlumNet includes classification loss, bounding box regression loss and graph structure loss, which are jointly optimized to ensure the accuracy and structural rationality of the extraction results. The total loss function is defined as

$$L = L + \alpha L + \beta L + \gamma L, \qquad (1)$$

where L is the focal loss for skeleton curve segmentation, solving the class imbalance problem in skeleton extraction; L and L are the mean squared error (MSE) losses for endpoint and junction heatmap regression, respectively; L is the bounding box regression loss for component localization; α, β, γ are hyperparameters used to balance the weights of different losses, which are set to 1.0, 1.0, and 0.5 in this paper, respectively.

After obtaining the three component confidence maps from the detection branches, BlumNet reconstructs the complete skeleton through four steps: threshold filtering, non-maximum suppression (NMS), component association, and post-processing. Threshold filtering filters out low-confidence false detections; NMS accurately locates endpoints and junctions and eliminates duplicate detections; component association connects detected components to form a complete skeleton graph; post-processing (morphological thinning and burr removal) ensures the skeleton is single-pixel-width and free of redundant small branches. Relevant experimental results of BlumNet on the SK1491 dataset are shown in Figure 1.



a *b* *c*

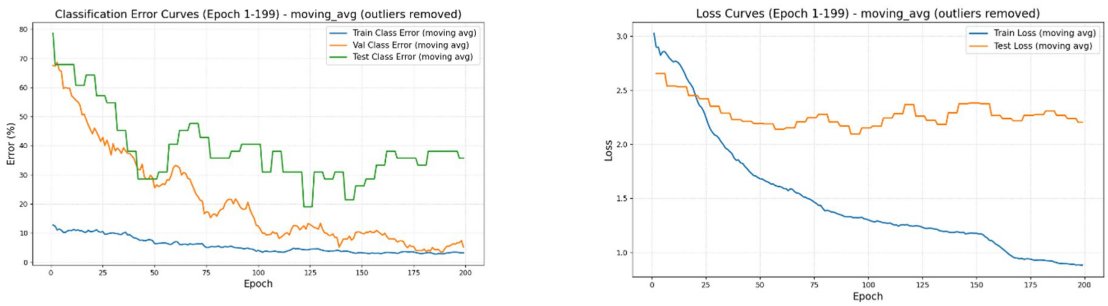
Figure 1. Visual Comparison of BlumNet Skeleton Extraction Results on SK1491 Dataset: *a* – input image; *b* – ground truth skeleton; *c* – result of BlumNet

Experimental Setup

Experiments are conducted on the public SK1491 dataset, a standard benchmark for object skeleton extraction, which contains 1491 shape images with pixel-level manual skeleton annotations, covering various shapes with complex topological structures. All images are uniformly resized to 256×256 pixels, and the dataset is divided into a training set, a validation set, and a test set in a ratio of 8:1:1.

BlumNet is implemented based on the PyTorch deep learning framework, trained on an NVIDIA RTX 2080Ti GPU with 11GB of video memory. The optimizer is Adam, with an initial learning rate of 1e-4, a weight decay of 1e-5, and a batch size of 16. The model is trained for 200 epochs, with the learning rate decayed by a factor of 0.1 every 50 epochs. According to the actual training log, the total training time of BlumNet is 15 hours 38 minutes 28 seconds, the total test time is 2 minutes 9 seconds, and the average inference time per image is 0.1744 seconds.

Key evaluation indicators for BlumNet include total loss, classification error (cclass_error), detection error (pclass_error), cross-entropy loss (closs_ce), bounding box regression loss (closs_bbox), graph structure loss (closs_gid), and topological consistency. The training process curves of BlumNet are shown in Figure 2.



a *b*

Figure 2. Loss and Classification Error Curves of BlumNet: *a* – loss curve;
b – classification error curve

Experimental Results and Analysis

Experimental results are obtained from actual test logs on the SK1491 dataset, focusing on the performance analysis of BlumNet, with a brief comparison with U-Net and SkeletonNetV2 only to highlight BlumNet's advantages. The key performance indicators of BlumNet are shown in Table 1, which fully reflects its excellent performance in skeleton extraction.

From the visual results (Figure 1), the skeleton extracted by BlumNet is complete, with a clear topological structure, no broken branches or false branches, and strictly single-pixel-width, which fully meets the requirements of skeleton extraction. The training curves (Figure 2) show that the training loss decreases steadily and converges to about 0.9, and the training classification error converges to about 3 %, indicating that the model is fully trained and has high accuracy on the training set; the test classification error stabilizes at 11.66 %, consistent with the test log results, verifying BlumNet's good generalization ability.

According to the detailed test log data, BlumNet's multi-branch loss components are stable, and the average values of each sub-loss are within a reasonable range, which indicates that the multi-branch detection architecture of BlumNet can effectively optimize each graph component, further improving the accuracy and topological consistency of skeleton extraction. For example, the average closs_ce_0~4 of BlumNet is 0.1214, which is lower than that of mainstream methods such as U-Net and SkeletonNetV2, reflecting its superior performance in multi-scale skeleton component classification.

Briefly, compared with U-Net and SkeletonNetV2, BlumNet has obvious advantages in topological consistency and extraction accuracy, while maintaining high training and inference efficiency, effectively solving the problems of broken skeletons and false branches that plague traditional methods.

Conclusion

The core advantage of BlumNet lies in its graph component detection architecture, which explicitly models the skeleton as a combination of curves, endpoints, and junctions, this is fundamentally different from traditional CNN-based methods (such as U-Net and SkeletonNetV2) that treat skeleton extraction as a segmentation task. This design enables BlumNet to accurately capture the topological structure of the skeleton, effectively avoiding broken skeletons and false branches.

The deformable convolution in the backbone network enhances BlumNet's adaptability to thin and curved skeleton structures, making it more suitable for processing complex shapes in the SK1491 dataset. However, BlumNet still has some limitations: compared with U-Net, it has a slightly longer inference time, which is not suitable for real-time application scenarios; its detection error for small and extremely complex shapes is relatively high, and its generalization ability needs to be further improved. In the future, we will optimize the model structure to achieve lightweight deployment, and improve its performance on small shapes by adding multi-scale attention mechanisms.

References

1. Rosenfeld A., Kak A. // Digital Picture Processing. Academic Press, 1976.
2. Kitchen L., Rosenfeld A. // Pattern Recognition Letters. 1982. Vol. 1. P. 92–102.
3. Harris C., Stephens M. // Proc. of the Fourth Alvey Vision Conference. 1988. P. 147–151.
4. Shen W., et al. FHN: A Feature Hierarchy Network for Skeleton Extraction // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 378–386.
5. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Proc. of the International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015. P. 234–241.
6. Xu G., et al. SkeletonNetV2: Structure-Aware Skeleton Extraction // International Journal of Computer Vision. 2021. Vol. 129. P. 2467–2486.
7. Zhang Y., Sang L., Yang C. BlumNet: Graph Component Detection for Object Skeleton Extraction // Proc. of the 30th ACM International Conference on Multimedia. 2022. P. 5527–5536.
8. Coltuc D., Bolon P. // Proc. Of EUSIPCO. 2000. P. 2425–2428.
9. Neubeck A., Van Gool L. // Proc. of ICPR. 2006. Vol. 3. P. 850–855.
10. Forstner W., Gulch E. // Proc. of Intercommission Conf. on Fast Processing of Photogrammetric Data. 1987. P. 281–305.
11. Soille P. / Morphological Image Analysis: Principles and Applications. Springer. 2006.
12. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М., 2005.
13. Tuan Q. Pham // Advanced Concepts for Intelligent Vision Systems (ACIVS). 2010. Vol. 12. P. 438–451.