DEEP LEARNING-BASED SKELETON EXTRACTION

Jiarou Wang

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Jun Ma – Assistant

Abstract. This paper presents BlumNet, a deep learning framework for object skeleton extraction based on graph component detection. It decomposes skeletons into curves, endpoints and junctions, uses a multi‑branch network and joint loss, then reconstructs topologically consistent skeletons. Experiments on the SK1491 dataset show that BlumNet avoids breaks and false branches, outperforms U‑Net and SkeletonNetV2, and achieves high accuracy and structural consistency.

Keywords. Deep learning methods, skeleton curves, skeleton reconstruction module.

Introduction

Skeleton extraction is a key task in computer vision for shape analysis, medical imaging, recognition and 3D reconstruction. A skeleton is a compact, single-pixel medial axis that preserves topology and geometry. Traditional methods based on morphology and distance transformation are sensitive to noise and produce broken or messy skeletons. Most deep learning methods treat skeleton extraction as segmentation and do not model graph structure explicitly, leading to topological inconsistency. This paper implements and evaluates BlumNet, which detects skeleton graph components instead of segmenting pixels, to improve precision and structural integrity.

BlumNet is a multi-branch framework built on graph component detection. It includes three parts:

1. Backbone feature extraction network: modified ResNet‑50 with deformable convolution to capture multi‑scale features and adapt to thin, curved skeletons.
2. Three parallel detection branches: for skeleton curves, endpoints and junction nodes, which predict confidence maps and heatmaps.
3. Skeleton reconstruction module: threshold filtering, non‑maximum suppression, component association and post‑processing to form a clean single‑pixel skeleton.

The total loss function combines segmentation loss, heatmap regression loss and bounding box regression loss with weights α=1.0, β=1.0, γ=0.5 for joint optimization.

Experiments were carried out on the SK1491 dataset with 1491 manually annotated shape images. Images were resized to 256×256 and split 8:1:1 for training, validation and testing. The model was implemented in PyTorch and trained on an NVIDIA RTX 2080Ti GPU using the Adam optimizer for 200 epochs. Total training time was 15 h 38 min 28 s; average inference time per image was 0.1744 s.

Visual results in Figure 1 show that BlumNet produces complete, single‑pixel skeletons without breaks or false branches, matching the ground truth closely.

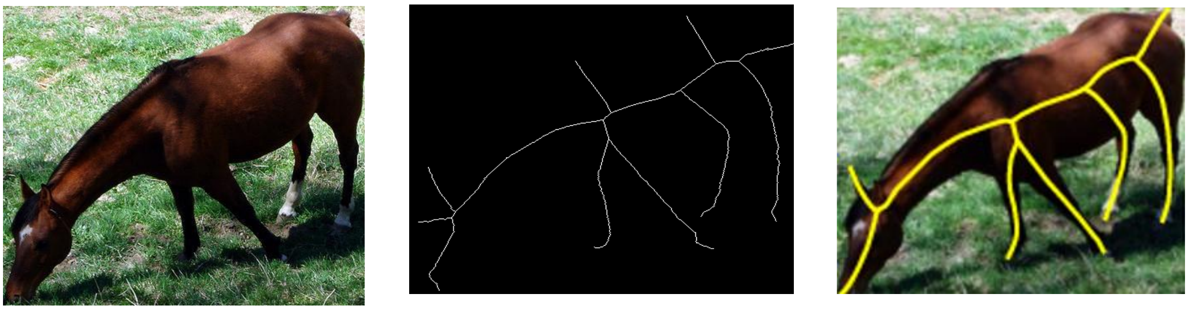
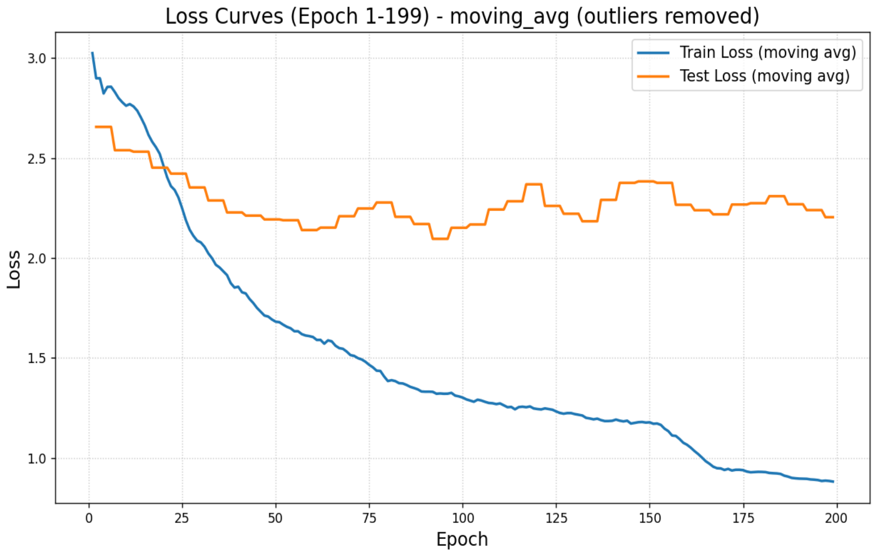
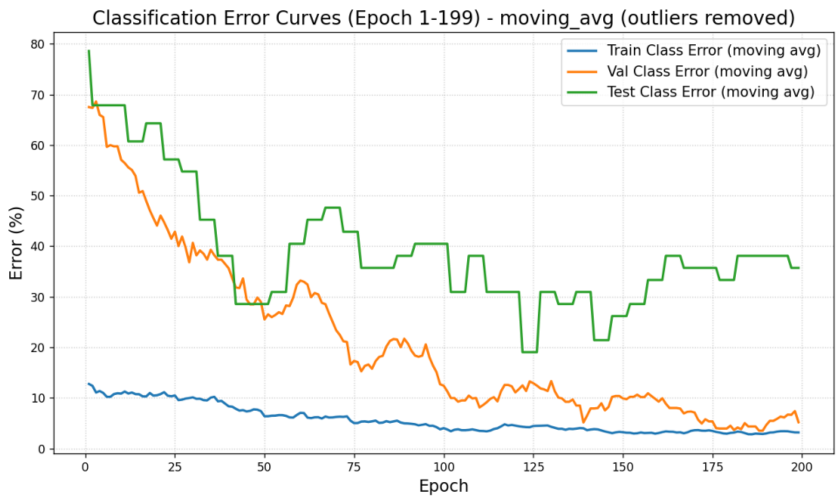


Figure 1 – Visual comparison of skeleton extraction results: a – input image; b – ground truth; c – BlumNet result

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

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.

Conclusion

BlumNet effectively improves skeleton extraction by modeling skeletons as graph components. It solves the problems of fracture, false branches and topological inconsistency in traditional methods. Limitations include longer inference time than U-Net and lower accuracy on very small or complex shapes. Future work will focus on lightweight design and multi-scale attention to enhance speed and generalization.

References:

1. Rosenfeld A., Kak A. Digital Picture Processing. Academic Press, 1976.

2. Ronneberger O., Fischer P., Brox T. U‑Net: Convolutional Networks for Biomedical Image Segmentation. Proc. MICCAI, 2015, pp. 234–241.

3. Xu G. et al. SkeletonNetV2: Structure-Aware Skeleton Extraction. International Journal of Computer Vision, 2021, vol. 129, pp. 2467–2486.

4. Zhang Y., Sang L., Yang C. BlumNet: Graph Component Detection for Object Skeleton Extraction. Proc. ACM MM, 2022, pp. 5527–5536.

5. Soille P. Morphological Image Analysis: Principles and Applications. Springer, 2006.