Cephalometric Landmark Detection using Graph- and Image-aware Network with Landmark Contrastive Regularization

Source code

Presenting author:
Xuanang Xu (xux12@rpi.edu)
Corresponding author:
Pingkun Yan (yanp2@rpi.edu)

Xuanang Xu[1], Tianshu Kuang[2], Anwei Chen[2], Nathan Lampen[1], Jungwook Lee[1], Xi Fang[1], Daeseung Kim[2], Hannah H. Deng[2], Jaime Gateno[2], Pingkun Yan[1] ✉

1. Department of Biomedical Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
2. Department of Oral and Maxillofacial Surgery, Houston Methodist Research Institute, Houston, TX 77030, USA

ISBI 2025 — 2025 IEEE International Symposium on Biomedical Imaging, April 14-17, 2025 | Houston, TX, USA

Paper ID: #1571091698

Introduction

Cephalometric landmark detection (or digitization) is crucial for craniomaxillofacial (CMF) surgical planning, directly influencing surgical outcomes. Current methods rely either solely on cone-beam computed tomography (CBCT) images (**Fig. 1a**) or 3D bony mesh models (**Fig. 1b**), limiting their accuracy due to modality-specific constraints. Image-based methods lack sensitivity to subtle anatomical variations, whereas mesh-based methods often fail in locally ambiguous regions due to the absence of contextual image information. We propose an integrated approach utilizing both CBCT images and 3D bone models to improve landmark detection accuracy.

Our key contributions are:

1. Proposing a hybrid approach that leverages complementary information from CBCT images and 3D bone models.
2. Introducing a novel CNN-GCN hybrid network, Graph-/Image-aware Network (GrImNet), featuring an end-to-end differentiable mesh sampling layer.
3. Developing a landmark contrastive regularization strategy to enhance feature discriminability and detection accuracy.

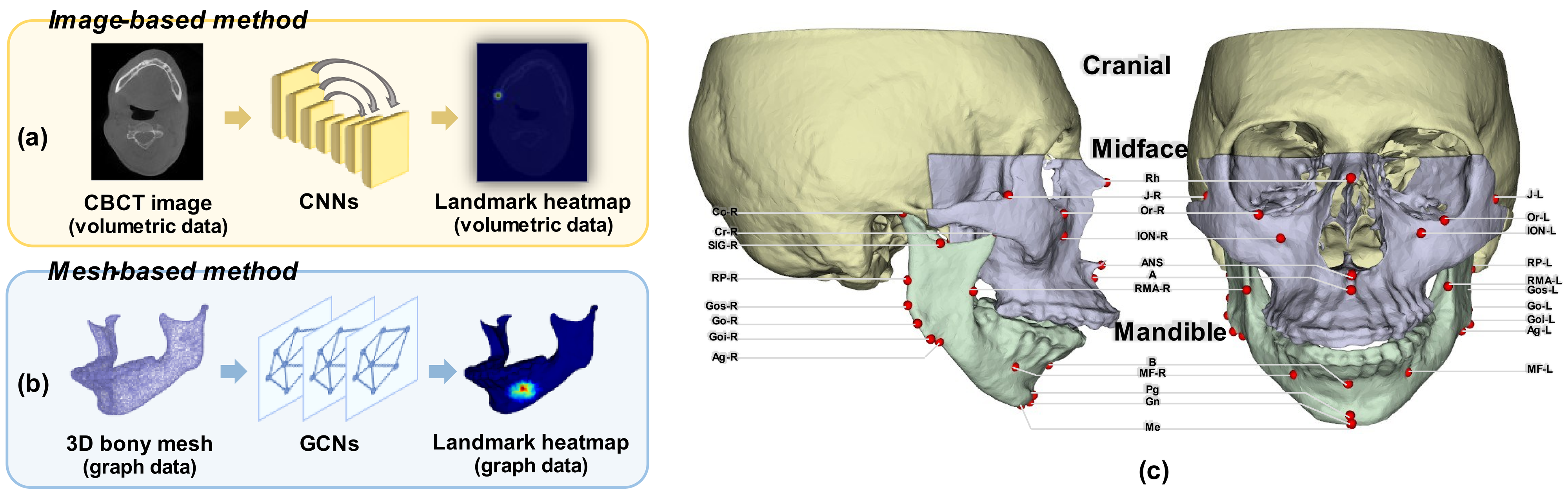

Fig. 1. Cephalometric landmark detection.

Method

A. Data preparation

CBCT images and corresponding bone segmentation masks are resampled uniformly to generate 3D triangle-mesh bone models. Each mesh vertex is associated with an 8-dimensional geometric feature vector describing its local geometry.

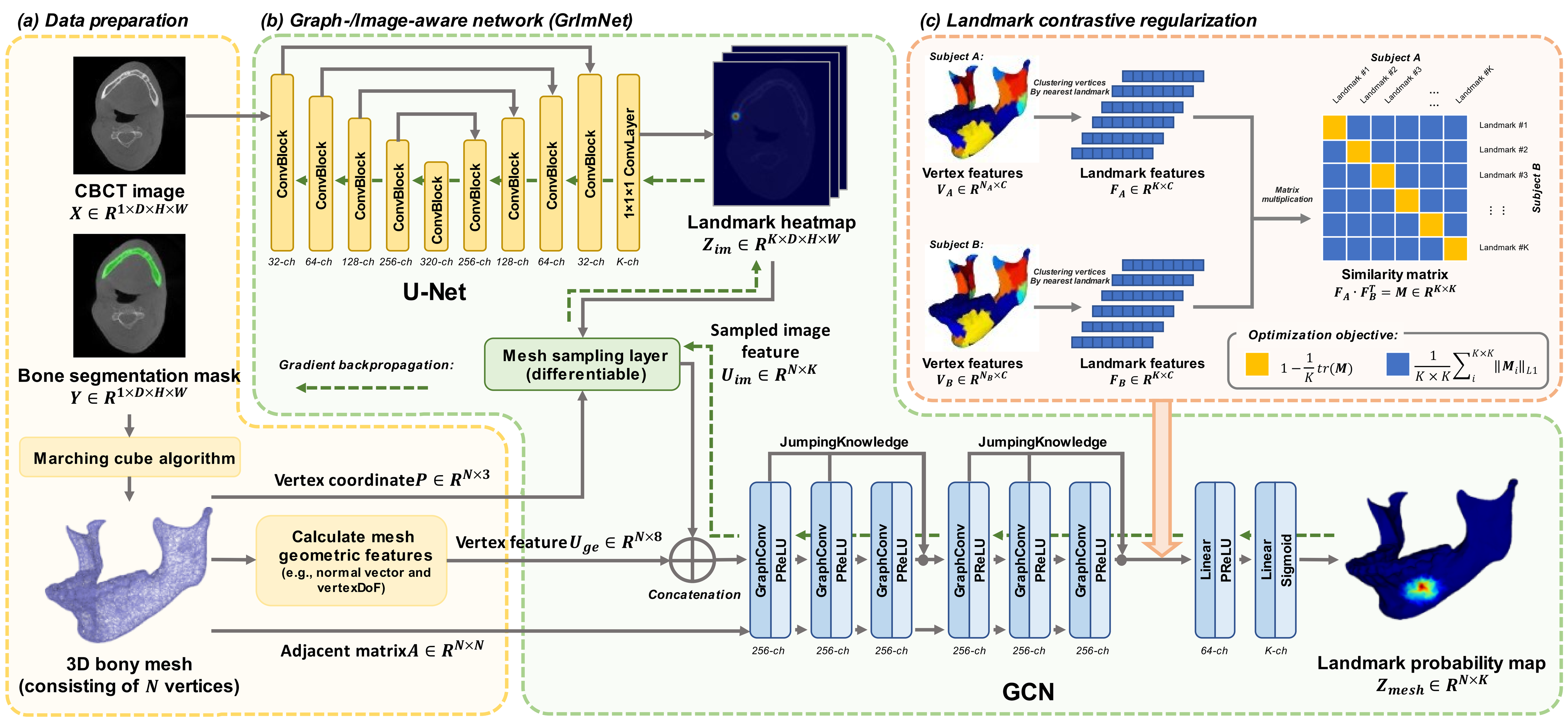

Fig. 2. Scheme of the proposed GrImNet method for cephalometric landmark detection.

B. Graph-/Image-aware Network (GrImNet)

The proposed GrImNet integrates image data via a U-Net[1] and geometric data via a Graph Convolutional Network (GCN)[2]. A differentiable mesh sampling layer connects these networks, enabling end-to-end training and aligning image and mesh features for precise landmark detection.

$$\boldsymbol{U}_{im}(\boldsymbol{p}_i) = \sum_{d,h,w}^{D,H,W} \boldsymbol{Z}_{im}(w,h,d)\kappa(x_i,w)\kappa(y_i,h)\kappa(z_i,d)$$

C. Landmark Contrastive Regularization

A novel contrastive regularization strategy enhances feature discriminability by maximizing similarity among identical landmarks and minimizing it across different landmarks, improving detection accuracy.

$$L_{con} = \left[1 - \frac{1}{K} tr(\mathbf{M})\right] + \frac{1}{K \times K} \sum_{i}^{K \times K} \|\mathbf{M}_i\|_{L_1}$$

Experimental Results

A. Dataset and Metric

We conducted experiments using a clinical dataset containing 56 subjects. Each subject includes a whole head CBCT image, a mandible segmentation mask with $K_{mandible} = 24$ mandible landmarks, and a midface segmentation mask with $K_{midface} = 12$ midface landmarks. We evenly divided the dataset into four parts and performed 4-fold cross-validation (three/one folds for training/testing, respectively) for evaluation. Distance error from the predicted landmark to the ground-truth landmark was used as the metric to quantitatively assess the model performance. Paired *t*-tests were conducted to investigate the statistical significance ("*": $p<0.05$) between the results yielded from different methods.

B. Comparison with Other Methods

Methods	Type	Distance error [Mean(SD) *mm*]		
		Midface	*Mandible*	*All*
U-Net [1]	Image-based	2.38(1.24)*	2.43(1.36)*	2.39(1.28)*
U-Net [1] (high resolution)	Image-based	2.24(1.22)*	2.29(1.36)*	2.25(1.26)*
Skull-engine [3]	Image-based	3.66(2.92)*	2.93(2.79)*	3.41(2.89)*
GCN in [4]	Mesh-based	1.89(2.20)*	2.41(3.53)*	2.06(2.73)*
U-Net + segmentation	Image&mesh	2.35(1.19)*	2.33(1.24)*	2.34(1.21)*
U-Net + mesh attaching	Image&mesh	2.14(1.32)*	2.08(1.52)*	2.12(1.39)*
GrImNet (proposed)	Image&mesh	**1.76(1.50)**	**1.94(1.64)**	**1.82(1.55)**

C. Ablation Study

Models	Distance error [Mean(SD) *mm*]		
	Midface	*Mandible*	*All*
U-Net	2.38(1.24)*	2.43(1.36)*	2.39(1.28)*
GCN	2.97(7.99)*	2.80(3.13)*	2.91(6.77)*
U-Net + GCN	1.84(1.35)*	2.01(2.32)	1.90(1.73)*
U-Net + GCN + L_{con} (w/o end-to-end training)	2.04(1.55)*	2.44(3.54)*	2.17(2.41)*
GrImNet (proposed)	**1.76(1.50)**	**1.94(1.64)**	**1.82(1.55)**

Conclusion

Our study demonstrates that integrating CBCT image information with geometric data from 3D bony mesh models significantly improves cephalometric landmark detection accuracy. The proposed GrImNet, featuring a differentiable mesh sampling mechanism and landmark contrastive regularization, provides a robust and precise solution suitable for clinical adoption, enhancing CMF surgical planning workflows.

References

1. Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in MICCAI, 2015.
2. Kipf and Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint, 2016.
3. Liu et al., "SkullEngine: a multistage CNN framework for collaborative CBCT image segmentation and landmark detection," in MLMI, 2021.
4. Elkhill et al., "Graph convolutional network with probabilistic spatial regression: Application to craniofacial landmark detection from 3d photogrammetry," in MICCAI, 2022.

Acknowledgment

This work was partially supported by NIH under award **R01DE021863**.