Learnable Skeleton-Based Medical Landmark Estimation with Graph Sparsity and Fiedler Regularizations

Yao Wang[1], Jiahao Chen[1], Wenjian Huang[2], Pei Dong[3], and Zhen Qian[1,*]

[1] United-Imaging Research Institute of Intelligent Imaging, Beijing, China,
[2] Southern University of Science and Technology, Shenzhen, China,
[3] United-Imaging Intelligent, Beijing, China,
{yao.wang; zhen.qian}@cri-united-imaging.com

Abstract. Recent development in heatmap regression-based models have been central to anatomical landmark detection, yet their efficiency is often limited due to the lack of skeletal structure constraints. Despite the notable use of graph convolution networks (GCNs) in human pose estimation and facial landmark detection, manual construction of skeletal structures remains prevalent, presenting challenges in medical contexts with numerous non-intuitive structure. This paper introduces an innovative skeleton construction model for GCNs, integrating graph sparsity and Fiedler regularization, diverging from traditional manual methods. We provide both theoretical validation and a practical implementation of our model, demonstrating its real-world efficacy. Additionally, we have developed two new medical datasets tailored for this research, along with testing on an open dataset. Our results consistently show our method's superior performance and versatility in anatomical landmark detection, establishing a new benchmark in the field, as evidenced by extensive testing across diverse datasets.

Keywords: Graph Convolution Networks · Fiedler Regularizations · Graph Sparsity · Landmark Detection.

1 Introduction

The precise and robust methods of anatomical landmarks localization in medical images are helpful for various diagnoses and treatment procedures [26, 12]. The connections between anatomical landmarks termed as the skeletal structure contain valuable anatomical and shape information. Graph Convolutional Networks (GCNs) are adept at capturing graph node attributes and relational structures through a sequence of graph-level convolutions [33].

In most previous research [2, 3], the construction of skeletal structures typically relies on predefined manual designs based on prior knowledge and assumptions about anatomy. However, in medical imaging, designing skeletal structures becomes more challenging due to the abundance of anatomical landmarks that frequently lack intuitive structure. This complexity makes it difficult to create a

graph that accurately represents the intricate relationships between landmarks. Additionally, adopting adaptive learning for graph connectivities, instead of predetermined structures, could markedly improve the model's generalization capabilities. Adaptive methods potentially allow the model to better accommodate unique and variable characteristics leading to more accurate and reliable analyses.

When considering the connectivity of the graph, it is essential to strike a balance between too few and excessive connected edges [10, 13]. Therefore, in this study, we aim to address a crucial question: Is it feasible to learn skeletal structures directly without the need for manual design across various tasks and attain outstanding performance through the network?

Algebraic connectivity, often referred to as the Fiedler value and denoted by λ_2, is a key concept in spectral graph theory [22, 23]. λ_2 denotes the second smallest eigenvalue of the Laplacian Matrix of a graph. This value reflects the connectivity and sparsity of the graph. Optimizing λ_2 reduces unnecessary and/or detrimental connections, leading to a more efficient network. It has received much attention for optimizing connections in various fields[15, 6, 9, 11], such as optical communication satellite networks, digital logistics networks and so on.

In this paper, we introduce a new method for learning skeletal structures, the Fiedler-regularized Graph Convolution Network (FRGCN), specifically designed to minimize the Fiedler value of a graph. The main contributions of this work can be summarized as:

- We present a novel skeleton reconstruction model using Fiedler regularization, which introduces graph-derived structural constraints to GCNs, representing a significant shift from traditional manual methods.
- We introduce the FRGCN, an effective model for landmark detection, which includes a Target-aware Encoder (TAE) and a Skeleton-aware Encoder (SAE). The TAE is crafted to capture information about landmarks and their interrelations, while the SAE is tailored to enforce constraints on skeletal structures.
- Extensive experiments show that FRGCN consistently outperforms SOTA methods on three medical image datasets.

2 Method

2.1 Related Work

Recent advancements in deep learning have showcased its efficacy in medical landmark detection [16, 34]. These methodologies can generally be categorized into three types: coordinate-based [24], heatmap-based [27], and graph-based approaches [35]. Graph-based methods, often building upon the other two methods, utilize GCNs to learn the interrelations of landmarks. Most GCNs are built on skeletal structures that are manually designed. To accommodate the changing relationships among human keypoints, dynamic graph convolution network

models [17] dynamically select preferred structures from manually pre-designed skeletons during training. Yet, these structural constraints are not applied during inference. While RSGNet [7] has improved upon this limitation, it still relies on manually designed priors, making the structural design for landmarks with non-intuitive relationships a persistent challenge.

Some approaches do not rely on manually designed structures, such as graph matching models [35] and deformable shape models [30]. However, their performance is significantly influenced by the initial configuration of the skeletal structures. Additionally, transformer-based methods enhance global relationships through attention mechanisms [19], eliminating the need for manually designed structures. However, these methods come with additional computational costs and may not perform as well on smaller datasets [36]. Consequently, there is an urgent need for a computationally inexpensive method for the automatic learning of skeletal structures.

2.2 Target-aware Encoder (TAE)

In this section, we introduce the three components of the FRGCN using lower limb landmark detection as an illustrative example. The lower limb landmarks in this study, based on DR images, are displayed in supplement Fig. 1. Here, a total of 20 landmarks are annotated to aid in the analysis of the mechanical axis of the lower limbs [18, 8].

Given an input image $I = \left\{ I_r \in R^{H_r \times W_r} \right\}_{r=1}^{N}$, where N represents the number of lower limb images and r is the image index, and H_r and W_r denote the input image height and width respectively. The positions of the K landmarks are represented by a set of coordinates $P \in R^{K \times 2}$.

First, we adopt HRNet [21] as a backbone to extract visual features f_r. Given the feature f_r, candidate landmarks' position $P' \in R^{K \times 2}$ can be generated. In order to further refine the coordinates for each landmarks, we calcuate the position vector $f_b \in [x, y, \Delta x, \Delta y]$, where (x, y) denotes the coordinates of the candidate landmark, and $(\Delta x, \Delta y)$ signifies the offset between the landmark and the center point of the image. The position encoder enhance the encoding of vector f_b.

To further aggregate the information between the visual vectors and position vectors, we also add spatial attention module [37]. The process is as follows:

$$E_c = \psi_c(w_c, concat(\psi_r(w_r, f_r), \psi_b(w_b, f_b))), \tag{1}$$

where ψ represents the encoding process. ψ_r,ψ_b are the visual and position encoder with parameters w_r and w_b. ψ_c is the spatial attention module of the weights w_c. E_c represents the encoded vector output of the TAE.

2.3 Graph Convolution Network using Fiedler Regularization

In a landmark graph $G = (V, E)$, the vertex set V comprises all the landmarks, and E is the edge set. The weights of E are either 0 or 1, indicating the presence or absence of connections between landmarks. The learning process for

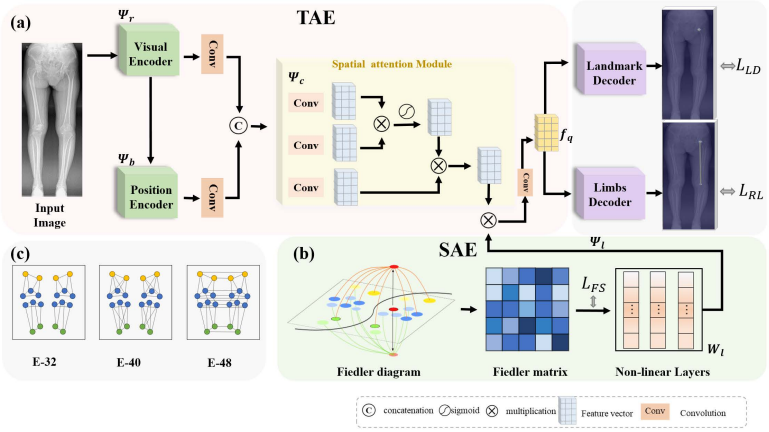

Fig. 1: The framework of our Graph Convolutional Network using Fiedler Regularization (FRGCN). (a) The Target-aware Encoder (TAE) comprises visual encoders ψ_r, positional encoders ψ_b and spatial attention module ψ_c. (b) The Skeleton-aware Encoder (SAE) utilizes Fiedler regularization to model Fiedler matrix and then updates using the non-linear layers. The SAE serves as the graph encoding module ψ_l. f_q is the feature vector aggregating the encoding outputs of TAE and SAE. The Fiedler diagram demonstrate the principle of minimizing the Fiedler value of the graph. We illustrate the FRGCN-based graph cut by selecting a representative landmark highlighted in red. This division into two subgraphs is represented by the yellow and green lines, positioned above and below, respectively. Nodes on the plane are categorized into lighter and brighter shades. (c) Illustrative diagrams depicting manually designed edges connecting lower limb points, featuring configurations with 32, 40, and 48 points.

skeletal structures becomes non-differentiable due to these dichotomized values, presenting a challenge for gradient-based optimization methods commonly used in neural networks.

In a general graph, the basic unit of graph connectivity consists of two nodes connected by an edge. We designate a new graph $G^{'} = (V^{'}, E^{'}, |F|)$ based on these units. Here, the vertex set $V^{'}$ includes all newly defined nodes, with each node representing a pair of connected nodes and the intervening edge in G, leading to a size of $(\frac{K(K-1)}{2}, 1)$, where K is the number of original nodes in G. The edge set $E^{'}$ corresponds to all possible connections within $V^{'}$, and we assume that all edges in $E^{'}$ to be present. The set F contains the weights of the edges in $E^{'}$, where each weight F_{ij} is a real number within the range $[-1, 1]$, and i and j indicate the respective rows and columns in F. The graph G' is expressed in matrix form. The degree matrix D' is derived from G', with each diagonal element D'_{ii} calculated as the sum of the absolute values of the weights connected to the i-th vertex, $\sum_{j=1}^{n} |F|_{ij}$. Subsequently, the Laplacian matrix of

the graph is defined as $L^{'} = D' - |F|$, which plays a critical role in analyzing the graph's properties.

Algorithm 1 The step of FRGCN

Input: Training data $\{I_r\}_{r=1}^{N}$
Hyperparameters: Learning rate η, batch size m, parameter γ_1,γ_2,γ_3, updating period T
Initialize parameters $W = \{w_c, w_r, w_b, W_l, F\}$ of the network
Compute the Laplacian $\mathbf{L}^{'}$ of the neural network
Compute the Fiedler vector $\mathbf{v}_2$ of the Laplacian $\mathbf{L}^{'}$
Initialize $c = 0$
while Stopping criterion not met **do**
  Sample minibatch $\{I_r\}_{r=1}^{m}$ from training set
  Set gradient $\delta = 0$
  for $i = 1$ to m **do**
    Compute gradient $\delta^{'} \leftarrow \delta$
    $+\nabla_{W_{c,q}} \gamma_1(\varphi((\mathcal{H}_K, A_{xy}^{K}), (\mathcal{H}^*)))$
    $+\nabla_{W_{c,q}} \gamma_2(\xi((\mathcal{H}_M, A_{xy}^{M}), (\mathcal{H}_M^*)))$
    $+\delta\nabla_F \gamma_3 \mathbf{v}_2^T \mathbf{L}_{|F|} \mathbf{v}_2$
  end for
  Apply gradient update $\mathbf{W} \leftarrow \mathbf{W} - \eta\delta$
  Update Laplacian matrix $\mathbf{L}^{'}$
  Update $c \leftarrow c + 1$
  if $c = T$ or $C \mod T = 0$ **then**
    Update second Laplacian eigenvector $\mathbf{v}_2$
  end if
end while

By introducing the new graph $G^{'}$, we shift the focus from directly learning about the original edge set E to minimizing the transformed edge set $E^{'}$ through adjustments in F. A higher value of $|F|$ signifies an increased likelihood of the existence of edges in $E^{'}$. Based on these likelihoods, the nodes are categorized into two separate subgraphs, aligning the distribution of edges with their probabilities. This shift allows us to transform the challenge of discerning connected edges into a classical network regularization problem, focusing on the decision of retaining or discarding each edge within $E^{'}$.

Classic regularization methods include dropout and L1 norm, among others. Edric [22] proposed leveraging spectral graph theory to improve the connectivity structure of the multi-layer nonlinear neural network through Fiedler regularization. Their method has demonstrated a significant boost in performance by minimizing hidden unit co-adaptation, presenting a more systematic approach compared to random dropouts during neural networks training.

The Fiedler value λ_2 is the second smallest eigenvalue of $G^{'}$'s Laplacian matrix $L^{'}$. A smaller λ_2 indicates a stronger connectivity in the subgraph [1]. However, during the training process, Fiedler value λ_2 cannot be optimized directly. Based on the theory of Cheeger's inequality and Rayleigh-Ritz variational characterization [20, 22], we are able to keep approaching the upper bound of Fiedler value, as shown in the following equation, by performing eigen-decomposition and further optimization of λ_2.

$$\lambda_2 \leq \boldsymbol{u}^T L^{'} \boldsymbol{u}, \tag{2}$$

where equality is achieved when unit vector $\boldsymbol{u} = v_2$, where v_2 is the eigenvector for λ_2. For the vertex set $V^{'}$, we denote $V^{'} = \{S \cup S', S \cap S' = \emptyset\}$, where S and S' are two subsets with dense connections within and sparsity between them.

During training, $G^{'}$ is constructed iteratively. The variable $\boldsymbol{u}$ is obtained through feature decomposition and serves as an upper bound to iteratively approximate the optimal value of λ_2, which is then utilized to update $L^{'}$, which in turn, aids in estimating $\boldsymbol{u}$ more accurately. The pseudo-code is as shown in Algorithm 1.

2.4 Skeleton-aware Encoder (SAE)

To incorporate the information of $G^{'}$, we include skeleton-aware encoder in our framework. From the aforementioned processes, we derive the F vector and E_c represents the encoded vector output of the TAE. We then apply a basic graph convolution network to model the relationships across landmarks. The operation of this graph convolution can be formulated as:

$$E_l = \psi_l(W_l E_c F), \tag{3}$$

where ψ_l is the skeleton-aware encoder and the W_l is the weights of non-linear layers, as shown in Fig. 1(b). E_l is the encoded vector output of skeleton-aware encoder.

The landmark detection task is reformulated to estimate K landmark heatmaps $\mathcal{H}_K \in R^{K \times H_h \times W_h}$ of size $H_h \times W_h$ and offset map $A_{xy}^K \in R^{2K \times H_h \times W_h}$. Offset map [32] is used to refine the landmark location. The limbs relation heatmaps is $\mathcal{H}_M \in R^{M \times H_h \times W_h}$ and the offset maps is $A_{xy}^M \in R^{2M \times H_h \times W_h}$. H_h and W_h denote the size of the feature map. It can also be derived from E_l.

$$\mathcal{H}_K, A_{xy}^K = D_{Ld}(f_q(W_q, E_l)), \mathcal{H}_M, A_{xy}^M = D_{Li}(f_q(W_q, E_l)), \tag{4}$$

f_q represents the feature vector of encoding output which aggregates information from both the target-aware encoder and the skeleton-aware encoder. D_{Ld} and D_{Li} represent the landmark decoder and the limbs decoder, respectively.

2.5 Loss Function

The overall loss to train FRGCN is a combination of three losses: 1) we calculate the Mean Squared Error (MSE) for the predicated landmark heatmap $\mathcal{H}_K$ and the relation heatmap $\mathcal{H}_M$. $\mathcal{H}_K^*$ and $\mathcal{H}_M^*$ represent the ground truth. 2) we calculate the $L1$ loss for the landmark offset maps A_{xy}^K and relation offset maps A_{xy}^M. $\mathcal{A}_{xy}^{K*}$ and $\mathcal{A}_{xy}^{M*}$ represent the ground truth. 3) we calculate the Fiedler score loss $\boldsymbol{u}^T L^{'} \boldsymbol{u}$ as the upper bound of Fiedler value λ_2

The learning objectives are defined as:

$$\begin{aligned} \boldsymbol{L} &= L_{LD} + L_{RL} + L_{FS} \\ &= \gamma_1 \varphi((\mathcal{H}_K, A^K_{xy}), (\mathcal{H}^*_K, \mathcal{A}^{K*}_{xy})) \\ &+ \gamma_2 \xi((\mathcal{H}_M, A^M_{xy}), (\mathcal{H}^*_M, \mathcal{A}^{M*}_{xy})) \\ &+ \gamma_3 \boldsymbol{u}^T L^{'} \boldsymbol{u}, \end{aligned} \tag{5}$$

where γ_1, γ_2, γ_3 are hyperparameters, and are respectively set to 1, 1 and 0.01 for balanced training. L_{LD} represent the loss of $\mathcal{H}_K$ and A^K_{xy}. L_{RL} represent the loss of $\mathcal{H}_M$ and A^M_{xy}. L_{FS} represent the loss of Fiedler regularization.

3 Experiment

In this section, we present the results obtained from two new datasets and one publicly datasets and conduct ablation studies to scrutinize the individual components.

3.1 Datasets and Evaluation

To evaluate the accuracy of landmark detection, we utilized the Mean Radial Error (MRE) and Successful Detection Rate (SDR) to evaluate the results in pixels.

Lower limb dataset The lower limb DR images are real-world data collected from collaborative hospitals. Two physicians manually annotated these images with 20 landmarks, as depicted in supplementary Figure 1.

When compared to VDNet[31],which integrates femur and tibia segmentation masks, VitPose [28], and RSGNet[7] as depicted in Table 1 (32, 40, and 48 manual edges, as well as full connectivity with 210 edges), our method demonstrates a significant decrease in MRE.

Pelvic dataset The pelvis DR images are the same as above. Each image is manually annotated with 22 landmarks. Our method achieves a remarkable reduction in MRE. Specifically, the reductions are 12.318, 1.49, 1.625, 1.239, and 2.389 pixels respectively as shown in supplementary Table1.

Cephalograms dataset The cephalograms dataset, publicly available and used in the IEEE 2015 ISBI Grand Challenge [25]. The train images and test images are delineated in prior research [5]. Our approach exhibits significant performance gains compared to previous works [4, 14, 29], as demonstrated in suppmentary Table2. Our model consistently achieves a reduction in MRE of at least 1.6 pixels.

Ablation Study The study explores the skeleton reconstruction using FRGCN. In order to prove that our proposed FRGCN is effective compared to backbone and other sparse methods, four ablation experiments are carried out: backbone, only non-linear layer, randomly dropping values in F, L1 norm of F. However, the dropout and L1 norm resulted in a decrease in performance, with increases

Table 1: Comparison of the SOTA methods on lower limb dataset with 20 landmarks.

Method	$L-hof$ $R-hof$	$L-gt$ $R-gt$	$L-lfc$ $R-lfc$	$L-mfc$ $R-mfc$	$L-fi$ $R-fi$	$L-ltc$ $R-ltc$	$L-mtc$ $R-mtc$	$L-ei$ $R-ei$	$L-lm$ $R-lm$	$L-mm$ $R-mm$	MRE ↓
VDNet	4.808 3.081	4.989 4.250	5.546 4.725	4.936 6.017	6.527 6.185	5.362 5.838	4.047 5.529	6.126 5.973	6.455 6.857	6.082 7.086	5.521
VDNet + mask	1.592 **1.363**	2.312 1.708	2.779 2.599	2.700 3.630	4.515 4.648	1.858 1.455	2.368 1.675	2.541 2.918	2.085 2.316	1.489 2.011	2.428
Vitpose-S	7.286 7.700	11.321 9.881	5.826 5.474	4.357 4.664	5.954 5.649	7.135 7.208	4.709 4.735	4.345 4.180	8.790 8.845	5.258 5.717	6.451
Vitpose-B	2.962 2.958	3.437 3.469	2.104 2.307	2.226 2.193	2.165 2.227	2.126 2.174	2.208 2.256	2.383 2.295	2.468 2.497	2.421 2.589	2.473
GCN-32	1.919 1.581	1.760 1.817	1.556 1.657	1.522 1.554	1.251 1.240	1.557 1.577	1.424 1.461	1.567 1.563	1.426 1.380	1.300 1.363	1.524
GCN-40	2.375 1.546	2.544 2.367	1.542 1.603	1.490 1.578	1.247 1.198	1.475 1.562	1.426 1.408	1.441 **1.494**	1.428 1.340	1.265 1.295	1.581
GCN-48	2.676 1.520	1.761 1.745	1.598 1.609	1.537 1.531	1.268 1.245	1.498 1.444	1.359 1.428	1.469 1.529	1.871 1.356	1.697 1.334	1.574
FRGCN	**1.504** 1.509	**1.640** **1.690**	**1.543** **1.581**	**1.439** **1.483**	**1.167** **1.201**	1.437 **1.461**	**1.334** **1.381**	**1.435** 1.527	**1.336** **1.272**	**1.200** **1.249**	**1.419**

Table 2: Ablation Study in lower limb dataset with 20 landmarks

Method	SAE		$L-hof$ $L-gt$ $L-lfc$ $L-mfc$ $L-fi$ $L-ltc$ $L-mtc$							MRE ↓
backbone[21]	non-linear layer	regularization term	$L-ei$ $L-lm$ $L-mm$ $R-hof$ $R-gt$ $R-lfc$ $R-mfc$							
			$R-fi$ $R-ltc$ $R-mtc$ $R-ei$ $R-lm$ $R-mm$ -							
✓	×	×	2.984	3.552	2.101	2.200	2.125	2.092	2.185	2.514
			2.353	2.659	2.621	2.989	3.496	2.276	2.254	
			2.121	2.153	2.274	2.291	2.731	2.823	-	
✓	✓	×	1.586	3.431	1.509	1.492	1.209	1.435	1.340	1.534
			1.483	1.358	1.246	1.482	1.712	1.617	1.491	
			1.195	1.469	1.373	1.560	1.361	1.328	-	
✓	✓	dropout	2.271	1.948	1.554	1.456	1.230	1.485	1.311	1.594
			1.405	1.345	1.253	1.527	3.721	1.620	1.548	
			1.218	1.446	1.383	1.448	1.352	1.354	-	
✓	✓	L1	1.514	1.644	1.490	1.430	1.195	1.454	1.399	1.486
			1.420	1.348	1.266	2.029	2.242	1.585	1.488	
			1.204	1.478	1.372	1.505	1.344	1.302	-	
✓	✓	FRGCN	1.458	1.628	1.537	1.446	1.166	1.452	1.342	**1.418**
			1.441	1.331	1.203	1.501	1.693	1.583	1.477	
			1.195	1.451	1.389	1.542	1.264	1.249	-	

in MRE of 0.176 and 0.068, respectively. These findings are detailed in Table 2, highlighting the impact of different manipulations of F on the effectiveness of the graph partitioning process.

The learned structures in different dataset are meaningful indicating strong connections.For example, the learned skeletal structure of the lower limb has a left-right approximate complementary connectivity map as shown in supplementary Figure2.

4 Conclusion

In this paper, we introduce an innovative model, FRGCN, for skeleton reconstruction in GCN-based landmark detection, signaling a notable shift from traditional manual methods by applying graph-derived structural constraints. FRGCN

is rooted in Fiedler regularization, a concept from spectral graph theory.The superiority of our model is not limited to theoretical aspects; we also illustrate its practicality and efficiency through an effective implementation approach, emphasizing its real-world applicability. Through extensive experiment, we have shown that our method surpasses existing state-of-the-art techniques in medical image analysis, across multiple performance metrics.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Biggs, N.: Algebraic graph theory: Vertex-partitions and the spectrum (1974)
2. Bin, Y., Chen, Z.M., Wei, X.S., Chen, X., Gao, C., Sang, N.: Structure-aware human pose estimation with graph convolutional networks. Pattern Recognition **106**, 107410 (2020)
3. Cai, Y., Ge, L., Liu, J., Cai, J., Cham, T.J., Yuan, J., Thalmann, N.M.: Exploiting spatial-temporal relationships for 3d pose estimation via graph convolutional networks. In: ICCV (October 2019)
4. Chen, R., Ma, Y., Chen, N., Lee, D., Wang, W.: Cephalometric landmark detection by attentive feature pyramid fusion and regression-voting. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22. pp. 873–881. Springer (2019)
5. Chen, R., Ma, Y., Chen, N., Lee, D., Wang, W.: Cephalometric landmark detection by attentivefeature pyramid fusion and regression-voting. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2019)
6. Cheung, K.F., Bell, M.G.: Improving connectivity of compromised digital networks via algebraic connectivity maximisation. European Journal of Operational Research **294**(1), 353–364 (2021)
7. Dai, Y., Wang, X., Gao, L., Song, J., Shen, H.T.: Rsgnet: Relation based skeleton graph network for crowded scenes pose estimation. AAAI **35**(2), 1193–1200 (May 2021)
8. Gieroba, T.J., Marasco, S., Babazadeh, S., Bella, C.D., Bavel, D.V.: Arithmetic hip knee angle measurement on long leg radiograph versus computed tomography—inter-observer and intra-observer reliability. Arthroplasty **5**(1) (2023)
9. He, Y., Gan, Q., Wipf, D., Reinert, G.D., Yan, J., Cucuringu, M.: Gnnrank: Learning global rankings from pairwise comparisons via directed graph neural networks. In: international conference on machine learning. pp. 8581–8612. PMLR (2022)
10. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors (2012)
11. Jiang, S.: Vision-Based Analysis of Human Face and Gesture: Dynamic Modeling, Synthesis and Recognition. Ph.D. thesis, Northeastern University (2022)
12. Lang, Y., Chen, X., Deng, H.H., Kuang, T., Barber, J.C., Gateno, J., Yap, P.T., Xia, J.J.: Dentalpointnet: Landmark localization onhigh-resolution 3d digital dental models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2022)

13. Larsen, J., Hansen, L.K., Svarer, C.: Regularization of neural networks using dropconnect (2001)
14. Lin, C., Zhu, B., Wang, Q., Liao, R., Qian, C., Lu, J., Zhou, J.: Structure-coherent deep feature learning for robust face alignment. IEEE Transactions on Image Processing **30**, 5313–5326 (2021)
15. Liu, X., Chen, X., Yang, L., Chen, Q., Guo, J., Wu, S.: Dynamic topology control in optical satellite networks based on algebraic connectivity. Acta Astronautica **165**, 287–297 (2019)
16. Nolte, D., Ko, S.T., Bull, A.M., Kedgley, A.E.: Reconstruction of the lower limb bones from digitised anatomical landmarks using statistical shape modelling. Gait & Posture **77**, 269–275 (2020)
17. Qiu, Z., Qiu, K., Fu, J., Fu, D.: Dgcn: Dynamic graph convolutional network for efficient multi-person pose estimation. AAAI **34**(07), 11924–11931 (Apr 2020)
18. Sled, E.A., Sheehy, L.M., Felson, D.T., Costigan, P.A., Lam, M., Cooke, T.D.V.: Reliability of lower limb alignment measures using an established landmark-based method with a customized computer software program. Rheumatology International **31**(1), 71–77 (2011)
19. Song, H., Liu, C., Li, S., Zhang, P.: Ts-gcn: A novel tumor segmentation method integrating transformer and gcn. Mathematical Biosciences and Engineering **20**(10), 18173–18190 (2023)
20. Spielman, D.: Spectral graph theory. Betascript Publishing (2010)
21. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: CVPR (June 2019)
22. Tam, E., Dunson, D.: Fiedler regularization: Learning neural networks with graph sparsity. In: III, H.D., Singh, A. (eds.) Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 9346–9355. PMLR (13–18 Jul 2020)
23. Tam, E., Dunson, D.: Spectral gap regularization of neural networks. arXiv preprint arXiv:2304.03096 (2023)
24. Valle, R., Buenaposada, J.M., Valdes, A., Baumela, L.: A deeply-initialized coarse-to-fine ensemble of regression trees for face alignment. In: Proceedings of the European Conference on Computer Vision (ECCV) (September 2018)
25. Wang, C.W., Huang, C.T., Hsieh, M.C., Li, C.H., Chang, S.W., Li, W.C., Vandaele, R., Marée, R., Jodogne, S., Geurts, P., Chen, C., Zheng, G., Chu, C., Mirzaalian, H., Hamarneh, G., Vrtovec, T., Ibragimov, B.: Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: A grand challenge. IEEE Transactions on Medical Imaging **34**(9), 1890–1900 (2015)
26. Wang, Z., Lv, J., Yang, Y., Lin, Y., Li, Q., Li, X., Yang, X.: Accurate scoliosis vertebral landmark localization on x-ray images via shape-constrained multi-stage cascaded cnns. Fundamental Research (2022)
27. Wei, S.E., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
28. Xu, Y., Zhang, J., ZHANG, Q., Tao, D.: Vitpose: Simple vision transformer baselines for human pose estimation. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems. vol. 35, pp. 38571–38584. Curran Associates, Inc. (2022)
29. Ye, Z., Yu, H., Li, B.: Uncertainty-aware u-net for medical landmark detection (2023)

30. Yu, X., Huang, J., Zhang, S., Metaxas, D.N.: Face landmark fitting via optimized part mixtures and cascaded deformable model. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(11), 2212–2226 (2016)
31. Zhang, L., Zhang, J., Shen, P., Zhu, G., Li, P., Lu, X., Zhang, H., Shah, S.A., Bennamoun, M.: Block level skip connections across cascaded v-net for multi-organ segmentation. IEEE Transactions on Medical Imaging **39**(9), 2782–2793 (2020)
32. Zhang, R., Zhu, Z., Li, P., Wu, R., Guo, C., Huang, G., Xia, H.: Exploiting offset-guided network for pose estimation and tracking (2019)
33. Zhang, S., Tong, H., Xu, J., Maciejewski, R.: Graph convolutional networks: a comprehensive review. Computational Social Networks **6**(1), 1–23 (2019)
34. Zhao, Q., Zhu, J., Zhu, J., Zhou, A., Shao, H.: Bone anatomical landmark localization with cascaded spatial configuration network. Measurement Science and Technology **33**(6), 065401 (mar 2022)
35. Zhou, F., Brandt, J., Lin, Z.: Exemplar-based graph matching for robust facial landmark localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (December 2013)
36. Zhu, H., Chen, B., Yang, C.: Understanding why vit trains badly on small datasets: An intuitive perspective. arXiv preprint arXiv:2302.03751 (2023)
37. Zhu, X., Cheng, D., Zhang, Z., Lin, S., Dai, J.: An empirical study of spatial attention mechanisms in deep networks. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6688–6697 (2019)