This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

SurfMark3D: Surface-Based Graph Convolution Framework for Distal Femoral Landmark Localisation in Total Knee Arthroplasty

Dharshan B.[1*], Pranav Hyagreev[1*], Vivek Maik[1†], Suhail Ansari T. A.[1], Manojkumar Lakshmanan[1], and Mohanasankar Sivaprakasam[2]

[1] Healthcare Technology Innovation Centre, IIT Madras, India
maik.vivek@htic.iitm.ac.in
[2] Indian Institute of Technology (IIT) Madras, India

Abstract. Accurate localisation of anatomical landmarks on distal femur is a critical prerequisite for surgical planning in image-based Total Knee Arthroplasty (TKA). Manual annotation is time-consuming and prone to variability, which can lead to inconsistent implant sizing. Existing classical and deep learning methods remain limited by latency, computational cost, and modality dependence. To address these issues, we operate on segmented 3D surface point clouds and formulate distal femoral landmark localisation as a heatmap regression task that predicts dense probability fields. Ground-truth heatmaps are constructed as Euclidean Gaussians on the femoral surface using learnable landmark-specific sigmas. We introduce a hybrid graph-convolution backbone that combines Low-Rank Adaptive Graph Convolution (LR-AGConv) for local feature extraction and Bi-level Routing Point Attention (BRPA) for efficient global semantic modelling. The network is trained using Adaptive Wing (AWing) loss and PointSTAR, a novel point-cloud adaptation of Self-adaptive Ambiguity Reduction (STAR) loss, with uncertainty-based multi-task weighting. We curate and release a clinically validated dataset of 550 distal femurs derived from the ShapeMed Knee study with expert landmark annotations. Experimental results demonstrate that the proposed method consistently outperforms state-of-the-art baselines, achieving a Mean Euclidean Error (MEE) of 1.24 ± 0.84 mm across all 11 anatomical landmarks. Code and dataset are available at https://github.com/IGRS-Imaging/SurfMark3D.

Keywords: Total Knee Arthroplasty · Landmark Localisation · Graph Convolution

1 Introduction

TKA success relies on accurate implant placement governed by morphometric measurements of distal femur and proximal tibia landmarks [6]. Manual anno-

*Equal contribution
†Corresponding author

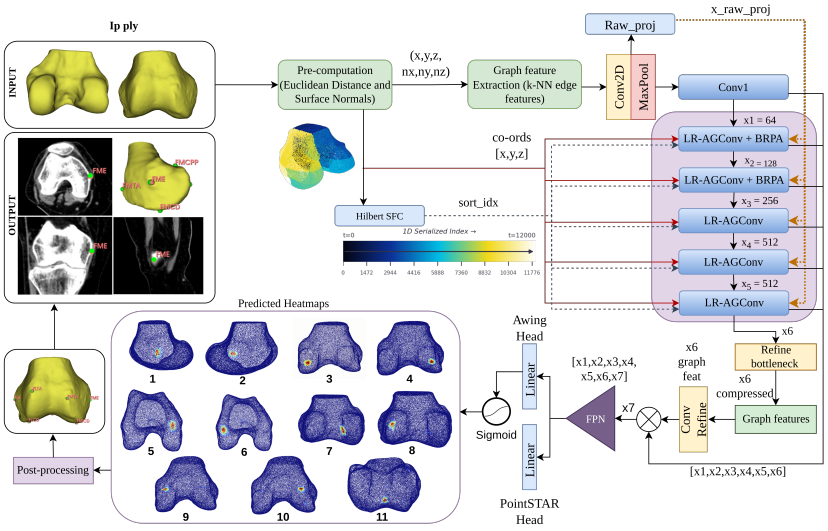


Fig. 1. Overall architecture of the proposed framework.

tation is labor-intensive and suffers from high inter- and intra-observer variability. Classical automated approaches (e.g., statistical shape models, atlas-based registration [1, 8]) report 2-6 mm errors but are sensitive to initialization and anatomical variance. Hybrid elastic template methods [5] improve accuracy at the cost of high latency. Recent volumetric [2] and multi-view 2D deep learning methods [15] reduce errors to 2-3 mm but remain computationally expensive and modality-dependent.

To be modality-agnostic, anatomical landmark localisation is increasingly performed directly on segmented 3D point clouds [10, 20]. Earlier methods relied on coordinate regression for landmark localisation. However, predicting a single coordinate is noisy and discards structural context. Probabilistic heatmap regression better models spatial context but typically relies on fixed isotropic Gaussian spreads that fail to capture the inherent distinctiveness of individual landmarks [20]. Furthermore, standard regression losses do not explicitly handle semantic ambiguity caused by irregular annotations and localisation uncertainty [26]. For instance, landmarks like FLTA (Table 1) form a highly anisotropic distribution along the trochlear ridge rather than a tight, isotropic peak.

Extracting local features with convolutional kernels involves a strict performance efficiency trade-off. Many kernels [17, 21, 24] either lack the adaptivity required for complex bone geometries or rely on complex parameter-heavy kernel constructions. Attention mechanisms are increasingly used to capture global context in point clouds. However, standard self-attention incurs prohibitive quadratic memory costs [18] and even efficient variants [22, 23] lack the selective regional filtering crucial for efficiently capturing long-range dependencies.

Table 1. Annotated distal femur landmarks. The prefix 'F' denotes 'Femoral'.

#	Abbr.	Name	#	Abbr.	Name
1	FME	Medial Epicondyle	7	FMTA	Medial Trochlea Anterior
2	FLE	Lateral Epicondyle	8	FLTA	Lateral Trochlea Anterior
3	FMCP	Medial Condyle Posterior	9	FMCPP	Medial Condyle Post. Prox.
4	FLCP	Lateral Condyle Posterior	10	FLCPP	Lateral Condyle Post. Prox.
5	FMCD	Medial Condyle Distal	11	Notch	Notch
6	FLCD	Lateral Condyle Distal			

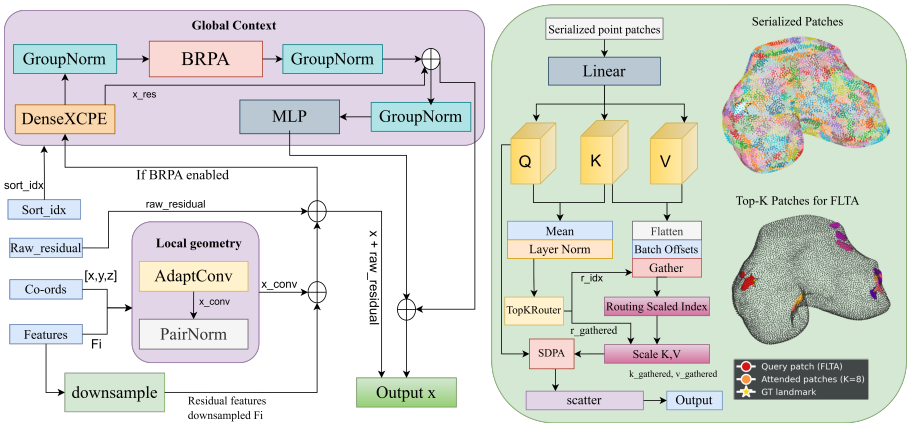
We formulate our task as probabilistic surface heatmap regression with dynamic, learnable landmark-specific Gaussian sigmas [16]. We propose a hybrid architecture featuring Low-Rank Adaptive Graph Convolution (LR-AGConv) [25] for efficient local expressivity, complemented by Bi-Level Routing Point Attention (BRPA), a novel adaptation of Bi-Level Routing Attention [27] for point clouds to selectively route queries to informative regions (Fig. 1). The network is optimized via Adaptive Wing loss [19] and PointSTAR, a novel 3D reformulation of STAR Loss [26] that resolves semantic ambiguity arising from surface anisotropy. Although open benchmarks (e.g., Teeth3DS+) have accelerated landmark localisation in other anatomies, orthopedics remains bottlenecked by a lack of public data. To advance research, we release a clinically validated benchmark dataset for distal femur landmark localisation.

In summary, our main contributions are: **(1)** A modality agnostic framework operating on 3D point clouds with dynamic heatmap spreads to model landmark-specific uncertainty; **(2)** A highly efficient hybrid backbone comprising LR-AGConv and BRPA; **(3)** A dynamic loss objective integrating AWing and PointSTAR loss; and **(4)** The release of a robust clinical benchmark of 550 expert-annotated distal femur meshes for landmark localisation.

2 Methods

2.1 Dataset and Annotation

We establish a novel clinical benchmark using distal femur meshes from ShapeMed Knee [4] (Baseline OAI 3D DESS MRIs), which reports expert-level segmentation quality (femoral DSC 0.99, ASSD 0.08-0.15 mm). Our dataset features 11 surgeon-validated manually annotated landmarks. Note that while the landmark framework itself is modality-agnostic, the pipeline requires an upstream segmentation step to generate meshes, which remains modality-dependent. We filter the cohort to non-osteophytic subjects (KL ≤ 1), consistent with [2,5,8], as degenerative morphological changes substantially increase landmarking errors [8], making reliable annotation infeasible on advanced OA surfaces. To ensure robust generalization, meshes undergo randomized Laplacian smoothing (15-25 iterations) and subjects are stratified across OAI demographic metadata (age, sex, race, ethnicity, BMI).



(a) Hybrid block (b) BRPA block

Fig. 2. Architectural components of the proposed framework.

2.2 Architecture

Input Representation The distal femur is processed as a point cloud ($N = 12000$ via FPS) with 6D features (spatial coordinates and normals). To preserve spatial locality efficiently for the attention mechanism, the cloud is serialized into a 1D sequence via a Hilbert space-filling curve, while maintaining the topology of the point cloud [22].

Hybrid Backbone Our 6-layer backbone maintains full spatial resolution N for vertex-level precision. Inspired by DeepGCN [11], we utilize dilated k-NN graphs, PairNorm and input residuals across all layers. The architecture (Fig. 1) operates in two phases: Stages 1-3 (Fig. 2a) combine LR-AGConv and BRPA (Fig. 2b) to jointly model local geometry and long-range semantics. Jointly leveraging global and local context has likewise proven effective in other medical imaging tasks [13]. Stages $4-6$ employ LR-AGConv only, isolating local surface details critical for precise localisation.

Low-Rank Adaptive Graph Convolution (LR-AGConv) To mitigate AGConv memory overhead, we propose a low-rank kernel decomposition. Input features of dimension C are compressed to $f'_i \in \mathbb{R}^{C_b}$ via a linear bottleneck, where C_b is the reduced channel size. Dynamic basis weights are synthesized from the feature graph. For point i and neighbor j, the convolution is:

$$A_{ij} = \mathrm{MLP}\big([f'_i,\ f'_j - f'_i]\big); \quad \tilde{f}_i = W_{\text{out}} \left(\max_{j \in \mathcal{N}(i)} \sum_{t=1}^{T} A_{ij}^{(t)}\ s_{ij}^{(t)} \right) \tag{1}$$

where, $A_{ij} \in \mathbb{R}^{R \times T}$ denotes low-rank adaptive kernels (rank R), $s_{ij} = [p_i, p_j - p_i] \in \mathbb{R}^T$ encodes spatial geometry, with $T = 6$ capturing absolute and rela-

tive 3D coordinates. Performing sequential aggregation directly in this low-rank space without generating large intermediate tensors reduces memory complexity to $\mathcal{O}(kC_bT)$. The projection W_{out} restores the output dimension, efficiently preserving geometric adaptivity.

Global Context with BRPA To capture long-range dependencies efficiently, we integrate BRPA. Inspired by [22], the N points are serialized via a Hilbert curve and partitioned into G non-overlapping regions of size P_s. Region-level queries and keys $(\bar{Q}_r, \bar{K}_r)$ are computed from token-level inputs (Q, K, V) via average-pooling and Layer Normalization (LN). A routing mechanism (TopkRouter) restricts fine-grained self-attention to top-k relevant regions:

$$\begin{aligned} \bar{Q}_r, \bar{K}_r = \text{LN}(\text{Avg}_{i\in r}\{Q_i, K_i\}), \quad W_r, I_r = \text{TopkRouter}\left(\frac{\bar{Q}_r \bar{K}^\top}{\sqrt{d}}\right) \\ \{K_r^g, V_r^g\} = \text{Gather}(K, V, I_r) \odot W_r, \quad \text{Attn}(Q_r) = \text{softmax}\left(\frac{Q_r (K_r^g)^\top}{\sqrt{d}}\right) V_r^g \end{aligned} \tag{2}$$

The gathered keys and values (K_r^g, V_r^g) are explicitly scaled by routing weights W_r before the token-level dot product (d is head dimension). Restricting computation to these top-k regions reduces quadratic complexity to $\mathcal{O}(N \cdot k \cdot P_s)$.

Multi-Scale Fusion and Prediction Multi-scale backbone features are projected to a uniform FPN dimension, concatenated, and enriched with globally pooled context. A shared MLP processes these features before branching into a primary AWing head and an auxiliary PointSTAR head (which acts as a training regularizer and is discarded during inference). During post-processing, the predicted landmark is computed as the weighted center-of-mass of the top-K points in the predicted heatmap response ($K = 10$).

2.3 Loss Functions

Adaptive Wing Loss We use AWing Loss to perform heatmap regression. For a ground truth y and prediction $\hat{y}$ at a given point, the loss is defined as:

$$\mathcal{L}_{AWing}(y, \hat{y}) = \begin{cases} \omega \ln\left(1 + \left|\frac{y-\hat{y}}{\epsilon}\right|^{\alpha-y}\right), & |y - \hat{y}| < \theta, \\ A|y - \hat{y}| - C, & \text{otherwise.} \end{cases} \tag{3}$$

where $(\alpha, \omega, \epsilon, \theta)$ are curvature hyperparameters, and A and C are the linear extension slope and intercept, respectively [19], ensuring continuity at the piecewise boundary. This base loss is subsequently integrated into our uncertainty formulation (Eq. 5).

PointSTAR Loss Unlike 2D STAR loss [26], PointSTAR respects the surface manifold by decoupling normal and tangential uncertainties. Given a predicted local covariance Σ and surface normal n, we project Σ onto the tangent plane via $P = (I - nn^\top)$, yielding $\Sigma_{\text{proj}} = P\Sigma P^\top$. Its eigen-decomposition produces eigenvalues ($\lambda_0 \leq \lambda_1 \leq \lambda_2$) and eigenvectors e_m. The smallest eigenvalue (λ_0) aligns with the normal, while λ_1, λ_2 span the tangent plane, capturing semantic ambiguity. Projecting the coordinate error $\Delta p = p - \hat{p}$ onto these axes, we apply normal (w_n) and tangential (w_t) weights to enforce strict surface-perpendicular localisation while tolerating manifold ambiguity:

$$\mathcal{L}_{\text{anis}} = w_n \frac{\mathcal{D}(\langle \Delta p, e_0 \rangle)}{\sqrt{\lambda_0}} + w_t \sum_{m=1}^{2} \frac{\mathcal{D}(\langle \Delta p, e_m \rangle)}{\sqrt{\lambda_m}} \tag{4}$$

where $\mathcal{D}(\cdot)$ is the Smooth L_1 distance. To prevent variance explosion and the consequent collapse of the distance penalty, we regularize $\mathcal{L}_{\text{var}} = \sum_{m=0}^{2} |\lambda_m|$. The final objective, $\mathcal{L}_{\text{PointSTAR}} = \mathcal{L}_{\text{anis}} + \beta \mathcal{L}_{\text{var}}$, projects the anisotropic penalty onto the manifold, ensuring strict convergence along the normal.

2.4 Hierarchical Uncertainty and Multi-Task Optimization

We propose a two-level homoscedastic uncertainty framework [3] to dynamically balance landmark-specific heatmap spreads and global multi-task objectives. At the landmark level, the spread σ_l of each landmark l is a learnable parameter. To stabilize training, $\mathcal{L}^{(l)}_{\text{AWing}}$ is scaled by a constant γ and weighted by σ_l. At the task level, we introduce homoscedastic uncertainties σ_{aw} and $\sigma_{\text{PointSTAR}}$ to automatically balance $\mathcal{L}^{(l)}_{\text{AWing}}$ and $\mathcal{L}_{\text{PointSTAR}}$ objectives. The joint optimization is formulated as:

$$\begin{aligned} \mathcal{L}^{(l)}_{\text{Reg}} &= \frac{\gamma}{2\sigma_l^2} \mathcal{L}^{(l)}_{\text{AWing}} + \log(\sigma_l), \\ \mathcal{L}_{\text{Total}} &= \frac{1}{2\sigma_{\text{aw}}^2} \left(\frac{1}{L} \sum_{l=1}^{L} \mathcal{L}^{(l)}_{\text{Reg}} \right) + \frac{\mathcal{L}_{\text{PointSTAR}}}{2\sigma^2_{\text{PointSTAR}}} + \log(\sigma_{\text{aw}} \sigma_{\text{PointSTAR}}) \end{aligned} \tag{5}$$

where L is the total number of landmarks. This hierarchical decoupling guarantees that landmark spreads modeled by σ_l do not interfere with the global balancing of multi-task gradients managed by σ_{aw} and $\sigma_{\text{PointSTAR}}$.

3 Experiments and Results

Implementation and Evaluation Details The dataset was partitioned into 440/55/55 samples (train/val/test), evaluated against a 50-case clinical baseline with inter- and intra-annotator variabilities of 1.24±1.05 mm and 1.18±0.90 mm. Models were trained for > 100 epochs using AdamW (LR=10^{-4}) on an NVIDIA A100. Key hyperparameters: LR-AGConv ($k = 60$, $C_b = 128$, $R = 32$); BRPA (8 heads, patch size 32, top-k = 8); and learnable sigmas initialized at 0.075

(clamped ≥ 0.03). We dynamically balanced Weighted AWing and PointSTAR loss ($k = 64$, ramped over epochs 10-20) via homoscedastic uncertainty. Results are averaged over three runs across MEE, SD (Standard Deviation), Median, and Success Detection Rate (SDR) at 1, 1.5, 2, and 4 mm thresholds. For clinical deployment, our full model requires only 1 GB VRAM with a median total inference latency of 1.7 s (0.3-0.5 s forward pass).

Comparative Analysis We evaluate architectural variants on the test set using our exact heatmap, loss formulations to isolate backbone impact (Table 2).

Backbone Comparison: PointNet++ [14], PointMLP [12] rely on hierarchical aggregation with global pooling and PtV3 uses an encoder-decoder architecture with serialized attentions. However, these are suboptimal (Table 2) for capturing both femoral global geometry and fine-grained local structure. Our hybrid architecture better balances local surface detail with global context.

GCN Operators: LR-AGConv generates dynamic kernels tailored to local anatomy, outperforming both DeltaConv and PAConv. Crucially, our low-rank formulation reduces peak training memory by 43.1% (40.4 GB $\rightarrow$ 23.0 GB) versus standard dense AGConv without degrading accuracy.

Attention Mechanisms: Standard MSA (Multi-Head Self Attention) is computationally prohibitive for training (65.5 GB VRAM, 1.18 s/scan). BRPA reduces this training memory to $\sim$27.5 GB. Although PTv3 is marginally faster, BRPA incurs negligible latency overhead while yielding better SDR (+4%).

Ablation Studies Table 2 also details our component-wise ablation. The base model (LR-AGConv + AWing) establishes a strong baseline. Integrating BRPA efficiently injects global semantic context; notably, it also stabilizes the convergence of the PointSTAR loss significantly smoother in V4 compared to V3 (Fig. 4). Adding PointSTAR loss explicitly addresses anisotropic ambiguity, notably reducing MEE for geometrically ambiguous landmarks like the FLTA (1.64 $\rightarrow$ 1.41 mm) and FMCD (1.86 $\rightarrow$ 1.60 mm). Furthermore, the addition of PointSTAR (V3, V4) accelerates overall training convergence to just 60 epochs with better performance, whereas models without it (V1, V2) require $>$ 100 epochs (Fig. 4). Finally, learnable sigmas outperform fixed variants by adapting to the distinctive learning dynamics of each landmark.

Clinical Generalizability and Cross-Modality Validation Our model demonstrates robust zero-shot transferability, achieving a 1.5$\pm$0.9 mm error on 56 CT-derived VSD point clouds despite the MRI-to-CT modality shift [9]. Adapting our architecture to geodesic heatmaps yielded a 1.8$\pm$1.35 mm error on a complex scapula dataset [7], proving broad anatomical generalizability.

Table 2. Performance evaluation on the test set. Top: Inter-architecture comparison against baselines. Bottom: Component-wise ablation study. MEE, SD, and Med. are in mm. SDR is in %. All models are trained with learnable sigmas, "Ours" refers to Base model trained with BRPA and PointSTAR.

Model	MEE±SD	Med.	SDR@t			
			1.0	1.5	2.0	4.0
Inter-architecture comparison						
PointNet++	1.39±0.93	1.16	40.6	63.6	78.4	98.1
PointMLP	1.38±1.13	1.12	42.1	68.3	82.3	97.2
PTv3	1.48±1.01	1.23	37.0	60.2	76.6	97.4
DeltaConv	1.54±1.04	1.30	36.0	58.1	74.2	96.3
PAConv	1.30±0.95	1.10	45.0	69.3	84.3	97.1
Self Attn	1.29±1.00	1.03	48.7	71.8	84.3	97.6
PTv2 Attn	1.43±1.32	1.10	44.2	68.5	81.5	95.3
PTv3 Attn	1.32±0.99	1.08	43.0	68.7	82.8	98.2
Intra-ablation study						
Base (LR-AGConv)	1.28±0.90	1.06	46.9	69.5	83.6	98.0
Base + BRPA	1.27±0.89	1.08	47.5	71.8	85.7	**98.7**
Base + PointSTAR	1.26±0.84	1.07	46.8	71.8	**86.3**	98.5
Ours w/o Learnable σ	1.28±0.92	1.08	46.8	70.3	84.0	97.9
Ours	**1.24±0.84**	**1.05**	**48.9**	**72.5**	85.6	**98.7**

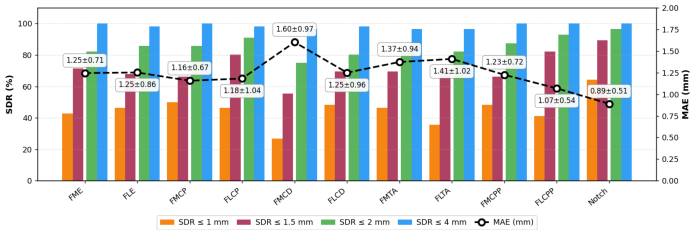


Fig. 3. Landmark-wise SDR and MAE ± SD for our model.

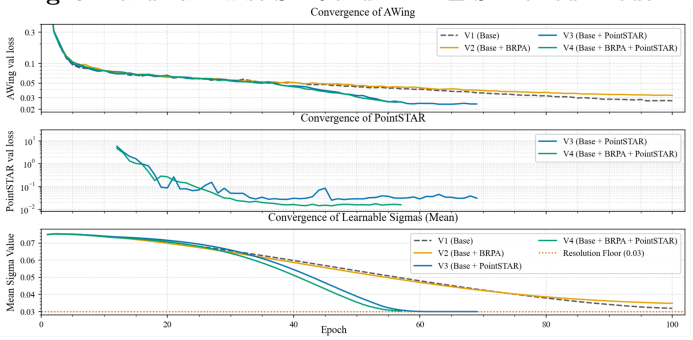


Fig. 4. Loss and sigma convergence.

4 Conclusion and Future Work

We present a modality-agnostic framework achieving state-of-the-art distal femur landmark localisation. Integrating LR-AGConv with BRPA captures rich local-to-global context while optimizing memory costs. Additionally, our novel PointSTAR loss explicitly resolves the geometric ambiguity of anisotropic surfaces to mitigate annotator variability. To advance research, we release an open, expert-annotated clinical benchmark dataset. Future work will extend this framework to explore downstream planning measurements from predicted landmarks, prediction robustness on osteophytic cases as well as tibial landmarks for comprehensive, fully automated TKA preoperative planning.

Disclosure of Interests. The authors have no competing interests.

References

1. Baek, S.Y., Wang, J.H., Song, I., Lee, K., Lee, J., Koo, S.: Automated bone landmarks prediction on the femur using anatomical deformation technique. Computer-Aided Design **45**(2), 505—510 (Feb 2013). https://doi.org/10.1016/j.cad.2012.10.033
2. Berger, L., Brößner, P., Ehreiser, S., Tokunaga, K., Okamoto, M., Radermacher, K.: Validation and comparison of three different methods for automated identification of distal femoral landmarks in 3d. Biomedical Engineering / Biomedizinische Technik **70**(5), 425—-431 (May 2025). https://doi.org/10.1515/bmt-2025-0026
3. Cipolla, R., Gal, Y., Kendall, A.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7482—-7491 (Jun 2018). https://doi.org/10.1109/cvpr.2018.00781
4. Gatti, A.A., Blankemeier, L., Van Veen, D., Hargreaves, B., Delp, S.L., Gold, G.E., Kogan, F., Chaudhari, A.S.: Shapemed-knee: A dataset and neural shape model benchmark for modeling 3d femurs (May 2024). https://doi.org/10.1101/2024.05.06.24306965
5. Grammens, J., Van Haver, A., Lumban-Gaol, I., Danckaers, F., Verdonk, P., Sijbers, J.: Automated landmark annotation for morphometric analysis of distal femur and proximal tibia. Journal of Imaging **10**(4), 90 (Apr 2024). https://doi.org/10.3390/jimaging10040090
6. Gromov, K., Korchi, M., Thomsen, M.G., Husted, H., Troelsen, A.: What is the optimal alignment of the tibial and femoral components in knee arthroplasty?: An overview of the literature. Acta Orthopaedica **85**(5), 480–487 (Jul 2014). https://doi.org/10.3109/17453674.2014.940573
7. Henninger, H.B.: 3d models of the human scapula and humerus with defined anatomic landmarks (2025). https://doi.org/10.5281/ZENODO.14590062
8. Horteur, C., Fougeron, N., Gaulin, B., Roch, Y., Bucki, M., Palluel, E., Payan, Y., Perrier, A.: Accuracy evaluation of an automatic landmarking method based on 3d segmented ct scans of full lower limbs with knee osteoarthritis. European Journal of Radiology **191**, 112292 (Oct 2025). https://doi.org/10.1016/j.ejrad.2025.112292

9. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The virtual skeleton database: an open access repository for biomedical research and collaboration. Journal of Medical Internet Research **15**(11), e245 (2013). https://doi.org/10.2196/jmir.2930
10. Kompella, G., Antico, M., Sasazawa, F., Jeevakala, S., Ram, K., Fontanarosa, D., Pandey, A.K., Sivaprakasam, M.: Segmentation of femoral cartilage from knee ultrasound images using mask r-cnn. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (Jul 2019). https://doi.org/10.1109/embc.2019.8857645
11. Li, G., Müller, M., Thabet, A., Ghanem, B.: Deepgcns: Can gcns go as deep as cnns? (2019). https://doi.org/10.48550/ARXIV.1904.03751
12. Ma, X., Qin, C., You, H., Ran, H., Fu, Y.: Rethinking network design and local geometry in point cloud: A simple residual mlp framework. In: Proceedings of the International Conference on Learning Representations (ICLR) (2022). https://doi.org/10.48550/arXiv.2202.07123
13. Murugesan, B., Vijaya Raghavan, S., Sarveswaran, K., Ram, K., Sivaprakasam, M.: Recon-GLGAN: A Global-Local Context Based Generative Adversarial Network for MRI Reconstruction, pp. 3—-15 (2019). https://doi.org/10.1007/978-3-030-33843-5_1
14. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 5099–5108 (2017). https://doi.org/10.48550/arXiv.1612.00593
15. Rostamian, R., Shariat Panahi, M., Karimpour, M., Kashani, H.G., Abi, A.: A deep learning-based multi-view approach to automatic 3d landmarking and deformity assessment of lower limb. Scientific Reports **15**(1) (Jan 2025). https://doi.org/10.1038/s41598-024-84387-z
16. Thaler, F., Payer, C., Urschler, M., Stern, D.: Modeling annotation uncertainty with gaussian heatmaps in landmark localization (2021). https://doi.org/10.48550/ARXIV.2109.09533
17. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds (2019). https://doi.org/10.48550/ARXIV.1904.08889
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need (2017). https://doi.org/10.48550/ARXIV.1706.03762
19. Wang, X., Bo, L., Fuxin, L.: Adaptive wing loss for robust face alignment via heatmap regression (2019). https://doi.org/10.48550/ARXIV.1904.07399
20. Wang, Y., Cao, M., Fan, Z., Peng, S.: Learning to detect 3d facial landmarks via heatmap regression with graph convolutional network. Proceedings of the AAAI Conference on Artificial Intelligence **36**(3), 2595—-2603 (Jun 2022). https://doi.org/10.1609/aaai.v36i3.20161
21. Wiersma, R., Nasikun, A., Eisemann, E., Hildebrandt, K.: Deltaconv: Anisotropic operators for geometric deep learning on point clouds (2021). https://doi.org/10.48550/ARXIV.2111.08799
22. Wu, X., Jiang, L., Wang, P.S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., Zhao, H.: Point transformer v3: Simpler, faster, stronger (2023). https://doi.org/10.48550/ARXIV.2312.10035
23. Wu, X., Lao, Y., Jiang, L., Liu, X., Zhao, H.: Point transformer v2: Grouped vector attention and partition-based pooling (2022). https://doi.org/10.48550/ARXIV.2210.05666

24. Xu, M., Ding, R., Zhao, H., Qi, X.: Paconv: Position adaptive convolution with dynamic kernel assembling on point clouds. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3172—-3181 (Jun 2021). https://doi.org/10.1109/cvpr46437.2021.00319
25. Zhou, H., Feng, Y., Fang, M., Wei, M., Qin, J., Lu, T.: Adaptive graph convolution for point cloud analysis (2021). https://doi.org/10.48550/ARXIV.2108.08035
26. Zhou, Z., Li, H., Liu, H., Wang, N., Yu, G., Ji, R.: Star loss: Reducing semantic ambiguity in facial landmark detection (2023). https://doi.org/10.48550/ARXIV.2306.02763
27. Zhu, L., Wang, X., Ke, Z., Zhang, W., Lau, R.: Biformer: Vision transformer with bi-level routing attention (2023). https://doi.org/10.48550/ARXIV.2303.08810