

PPCNet: Projection-Conditioned Point Cloud Reconstruction of Spinal Vertebrae from Biplanar Radiographs

Rohith Reddy G K, Pandurang, Vivek Maik, Manojkumar Lakshmanan, and
Mohanasankar Sivaprakasam

Department of Electrical Engineering, Indian Institute of Technology (IIT) Madras,
India
`mohan@ee.iitm.ac.in`

Abstract. In this paper, we propose PPCNet (Projection-Conditioned Point Cloud Network), a query-refinement framework that reconstructs a dense 8,192-point cloud of spinal vertebrae from two orthogonal digitally reconstructed radiographs (DRRs) and their corresponding calibrated 3×4 projection matrices. PPCNet employs the calibrated projection geometry as the core 2D-to-3D lifting mechanism, combining occupancy-gated query initialisation with projection-conditioned refinement. Gap-preserving and curvature losses maintain inter-vertebral disc spaces and the spinal lordotic curve. Evaluated on the lumbar spine (L1–L5) across 105 held-out patients from a 1,037-patient dataset derived from VerSe and CTSpine1K, PPCNet achieves a mean Chamfer distance of 1.981 mm, F-score@5 mm of 0.973, and HD95 of 4.525 mm. Clinical validation across 525 vertebrae confirms a morphometric MAE of 2.80 mm with 97% of vertebrae at Vertebral Compression Level (VCL) Grade A. As a proof-of-concept, phantom-based navigation tracking yields a mean vertebra centroid error of 3.01 mm across L1–L5. The implementation and dataset are available at: PPCNet.

Keywords: Point Cloud Reconstruction · Biplanar Radiographs · DRR · Spinal Vertebrae · Projection-Conditioned Refinement · Minimally Invasive Spine Surgery

1 Introduction

In MISS, accurate intraoperative 3D visualization of spinal vertebrae is essential for safe instrument placement, where misjudging depth by a few millimetres can cause neural or vascular injury [1,2]. While intraoperative cone-beam CT can recover full 3D anatomy, its high radiation dose, cost, and limited deployment across surgical centres, particularly in resource-constrained settings [3] restrict routine use. Conventional C-arm fluoroscopy, though cost-friendly, captures only 2D projections without depth information. Due to these inherent limitations, there is a compelling need for methods that recover reliable 3D vertebral geometry directly from the biplanar views of C-arm fluoroscopy imaging already acquired during surgery, without additional radiation or equipment.

Classical statistical shape model (SSM) approaches fit parametric templates to biplanar radiographs [4,5,6], depending on pre-defined shape priors, limiting patient-specific fidelity. Deep learning methods have advanced rapidly: volumetric GAN-based synthesis [7,8], dimensionally-consistent encoders with attention-guided fusion [9], Swin Transformer encoders with cross-attention fusion [10], and graph convolutional decoders for single-vertebra reconstruction [11]. Despite this progress, these learning-based methods share two key limitations. First, volumetric approaches represent anatomy on fixed low resolution grids, where each voxel spans approximately 1.5 mm; fine structures such as the 3–5 mm inter-vertebral disc spaces occupy only 1–2 voxels at this resolution, and the network cannot distinguish disc from bone within a single voxel, causing these gaps to blur or vanish entirely. Second, standard C-arm setups produce calibrated 3×4 projection matrices that encode the exact 3D to 2D geometric mapping for each biplanar view, which can be utilized as an explicit reconstruction prior.

We address both limitations with PPCNet and our main contributions include:

- **Novel PPCNet Architecture:** By adopting a point cloud representation, PPCNet places points directly on spinal surfaces without discretising empty space, naturally handling disjoint structures such as separate vertebrae [12,13]. An occupancy-gated initialisation concentrates candidate points in bone-occupied regions, after which three refinement stages repeatedly project each point into both biplanar views via the calibrated matrices, sample local image and volumetric features at the projected locations, and predict bounded spatial offsets that progressively displace points toward the true surface. A gap-preserving penalty discourages points from collapsing into inter-vertebral disc spaces, and a curvature loss [14] preserves the sagittal spinal profile.
- **Clinical Validation through Anatomical Measurements and Navigation Tracking:** Beyond standard reconstruction metrics, we evaluate eight clinically relevant morphometric parameters and Vertebral Compression Level (VCL) grading [15] across 525 vertebrae (5 per patient, 105 test patients), and demonstrate clinical applicability through phantom-based surgical navigation tracking.
- **Open-Source Dataset for Biplanar Spine Reconstruction:** We release paired biplanar DRRs, projection matrices, and ground-truth point clouds for 1,037 patients derived from VerSe [16,17] and CTSpine1K [18] under a fixed 829/103/105 split.

2 Dataset and Methodology

2.1 Dataset

We curate a dataset from VerSe [16,17] and CTSpine1K [18], selecting 1,037 patients with complete L1–L5 lumbar segmentation labels. For each patient, two digitally reconstructed radiographs (DRRs) – one coronal (AP at 0°) and one

sagittal (lateral at 90°) – are generated using Plastimatch [19] with a source-to-axis distance (SAD) of 1,000 mm, source-to-image distance (SID) of 1,500 mm, a 500×500 mm detector, and 512×512 pixel resolution. Along with each DRR, a calibrated 3×4 projection matrix encoding the geometric relationship between the 3D volume and the 2D image plane is generated per view. Ground-truth point clouds are generated in four steps: (i) per-vertebra surfaces are extracted from the CT segmentation masks, (ii) the five vertebral meshes (L1–L5) are concatenated into a single lumbar spine surface, (iii) the concatenated mesh is aligned to the DRR coordinate space using Iterative Closest Point (ICP) registration with four axis-flip initialisations to resolve orientation ambiguity, and (iv) the aligned surface is uniformly subsampled to 5,120 points via farthest-point sampling.

2.2 Architecture

PPCNet (Fig. 1) takes two biplanar DRRs ($\mathbf{I}_{\text{AP}}$, $\mathbf{I}_{\text{LP}}$), their calibrated 3×4 projection matrices ($\mathbf{P}_{\text{AP}}$, $\mathbf{P}_{\text{LP}}$), and patient-specific normalisation parameters – center $\mathbf{c} \in \mathbb{R}^3$ and scale $\mathbf{s} \in \mathbb{R}^3$, computed from training-set statistics and provided per patient during inference – producing $N=8,192$ points through six stages.

Stage 1: Dual 2D Encoders. Two weight-independent ResNet-34 backbones [20] encode the AP and LP DRRs independently. Each backbone extracts features through its stem and residual blocks up to `layer3`, followed by a 1×1 convolution projecting to 192 channels (Eq. 1):

$$\mathbf{f}_{\text{AP}} = \mathcal{E}_{\text{AP}}(\mathbf{I}_{\text{AP}}), \quad \mathbf{f}_{\text{LP}} = \mathcal{E}_{\text{LP}}(\mathbf{I}_{\text{LP}}), \quad \mathbf{f} \in \mathbb{R}^{B \times 192 \times 32 \times 32} \quad (1)$$

where B is the batch size and $\mathcal{E}$ denotes each encoder; spatial resolution reduces from 512×512 to 32×32 through successive strided convolutions.

Stage 2: Feature Lift to 3D. Each 2D feature map $\mathbf{f} \in \{\mathbf{f}_{\text{AP}}, \mathbf{f}_{\text{LP}}\}$ is replicated along a new depth axis ($D=32$) to form a $32 \times 32 \times 32$ volume. A learnable depth embedding $\mathbf{e}_d \in \mathbb{R}^{192 \times 32}$ is added so that each depth slice carries a unique positional signature rather than identical replicated features, and two Conv3D–GroupNorm–GELU layers refine to 128 channels (Eq. 2):

$$\mathbf{V}_{\text{view}} = \text{Conv3D}(\mathbf{f} \oplus \mathbf{e}_d), \quad \mathbf{V}_{\text{view}} \in \mathbb{R}^{B \times 128 \times 32 \times 32 \times 32} \quad (2)$$

where $\oplus$ denotes addition along depth, producing one 3D feature volume per view: AP volume ($\mathbf{V}_{\text{AP}}$) and LP volume ($\mathbf{V}_{\text{LP}}$).

Stage 3: Biplanar Fusion. The $\mathbf{V}_{\text{AP}}$ is permuted so its depth axis aligns with the coronal direction, and the $\mathbf{V}_{\text{LP}}$ with the sagittal direction. The aligned volumes are concatenated along the channel dimension (128 channels from $\mathbf{V}_{\text{AP}}$ and 128 from $\mathbf{V}_{\text{LP}}$, yielding 256) and fused via 1×1 then 3×3 Conv3D back to 128 channels (Eq. 3):

$$\mathbf{V}_f = \text{Conv3D}_{3 \times 3}(\text{Conv3D}_{1 \times 1}([\mathbf{V}_{\text{AP}}^\pi; \mathbf{V}_{\text{LP}}^\pi])), \quad \mathbf{V}_f \in \mathbb{R}^{B \times 128 \times 32^3} \quad (3)$$

where π denotes view-specific axis permutation and $[\cdot; \cdot]$ is channel-wise concatenation.

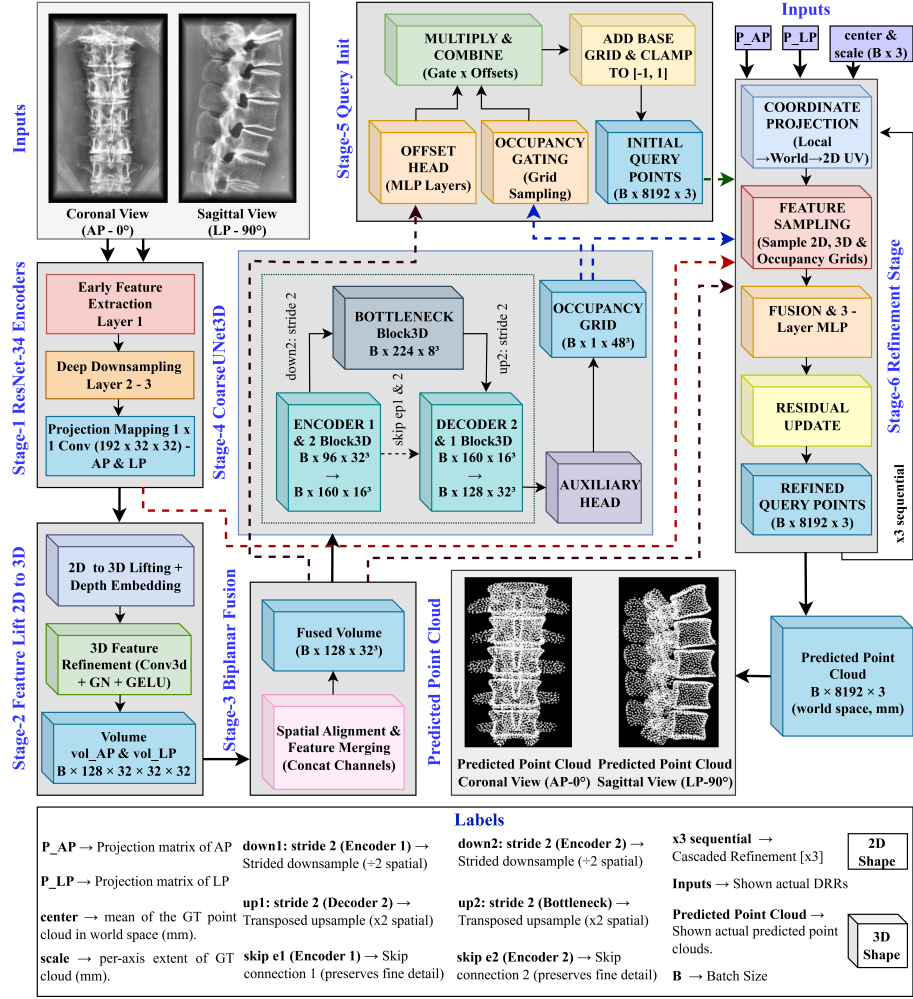


Fig. 1: PPCNet architecture: dual ResNet-34 encoders, 3D feature lifting and fusion, occupancy-gated query initialisation, and three projection-conditioned refinement stages.

Stage 4: Coarse 3D U-Net. A three-level U-Net [21] with progressively wider encoder levels (128→96→160→224 channels) and a decoder (160→128) refines V_f into an enriched volume v_f and simultaneously predicts a 48^3 occupancy volume o_{aux} via an auxiliary head supervised with Binary Cross-Entropy (BCE) + Dice loss. This occupancy volume provides a coarse binary estimate of bone-occupied regions, enabling the subsequent stage to concentrate query points in anatomically relevant space.

Stage 5: Query Initialisation. A regular $20 \times 20 \times 21$ grid subsampled to $N=8,192$ candidate points serves as the base set of initial query locations. A

global Multi-Layer Perceptron (MLP) predicts a spatial offset for each point, gated by the occupancy probability g_i sampled from the occupancy volume, so that points in empty regions receive near-zero shifts while points near bone are displaced toward occupied areas. The initial query position $\mathbf{q}_i^0$ (superscript 0 denoting stage zero, before refinement) is computed as (Eq. 4):

$$\mathbf{q}_i^0 = \text{clamp}\left(\mathbf{q}_i^{\text{base}} + g_i \cdot 0.25 \cdot \tanh(\mathbf{h})_i, -1, 1\right) \quad (4)$$

where $\mathbf{q}_i^{\text{base}}$ is the i -th grid point in normalised $[-1, 1]^3$ space, g_i is the sigmoid-activated occupancy probability sampled at the base grid position $\mathbf{q}_i^{\text{base}}$:

$$g_i = \sigma(\text{GridSample}(\mathbf{o}_{\text{aux}}, \mathbf{q}_i^{\text{base}}))$$

and $\mathbf{h} = \text{MLP}(\text{AvgPool}(\mathbf{v}_f))$ is a global offset vector derived from the average-pooled 3D volume. The $\tanh$ bounds each offset to ± 0.25 , and the final clamp constrains all queries within the normalised volume.

Stage 6: Projection-Conditioned Refinement. Three stages with independent weights iteratively refine queries, progressively improving both coarse shape and fine surface detail. Unlike methods [7,9,10] that fuse views implicitly, PPCNet uses calibrated matrices to determine *where* each 3D query projects onto each 2D feature map. At stage k , each query $\mathbf{q}_i^k \in \mathbb{R}^3$ is denormalised to world coordinates and projected onto both views (Eq. 5):

$$\mathbf{u}_i^v = \pi(\mathbf{P}_v[\mathbf{q}_i^k \odot \mathbf{s} + \mathbf{c}; 1]), \quad v \in \{\text{AP}, \text{LP}\} \quad (5)$$

where $\mathbf{q}_i^k \odot \mathbf{s} + \mathbf{c}$ maps normalised to world coordinates ($\odot$: element-wise multiply), $\mathbf{P}_v$ is the 3×4 projection matrix for view v , $[\cdot; 1]$ appends a homogeneous coordinate, and $\pi(\cdot)$ performs perspective division to yield the 2D pixel coordinate $\mathbf{u}_i^v$ of point i in view v . At each projected location, the network samples 2D image features (192-d, $\times 2$ views), 3D volumetric features (128-d), occupancy (1-d), 3D coordinates (3-d), and 2D pixel positions (2-d, $\times 2$ views), forming a 520-d vector. A three-layer MLP ($520 \rightarrow 384 \rightarrow 384 \rightarrow 3$, GELU, 5% dropout) predicts a bounded displacement (Eq. 6):

$$\Delta \mathbf{q}_i^k = 0.25 \cdot \tanh(\text{MLP}_k(\mathbf{x}_i)), \quad \mathbf{q}_i^{k+1} = \mathbf{q}_i^k + \Delta \mathbf{q}_i^k \quad (6)$$

where $\mathbf{x}_i$ is the 520-d concatenated feature vector and $\tanh$ bounds each stage- k displacement $\Delta \mathbf{q}_i^k$ to ± 0.25 (± 0.75 cumulative across three stages).

3 Results and Discussion

On 105 test data, PPCNet achieves a mean Chamfer distance of 1.981 ± 1.060 mm (median 1.770 mm), F@5 of 0.973 ± 0.074 , F@1 of 0.155 ± 0.070 , and HD95 of 4.525 ± 5.382 mm (median 3.423 mm). Per-patient sorted distributions (Fig. 2) confirm a long-tail behaviour driven by patients with severe deformities or transitional vertebrae. Since PPCNet outputs a single unlabelled point cloud, ICP registration assigns each point to its corresponding vertebra (L1–L5) for clinical evaluation.

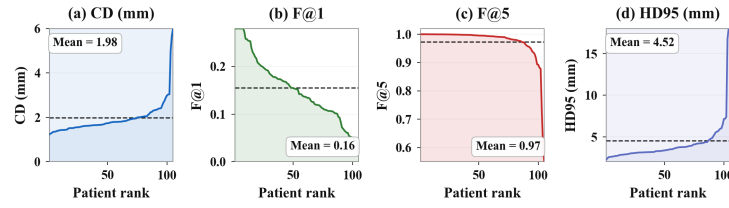


Fig. 2: Per-patient sorted test metrics (105 patients). (a) CD (mm), (b) F@1, (c) F@5, (d) HD95 (mm). Dashed lines = mean.

3.1 Clinical Validation

We evaluate eight clinically relevant parameters (Fig. 3) across 525 vertebrae (105 test patients) using automatic landmark detection [22]: VBH, aVBH, pVBH, VBW, VBD, DH, lumbar lordosis, and coronal Cobb angle. Grading thresholds: Excellent (<2 mm MAE), Good (2–4 mm), Acceptable (4–6 mm); VCL via Genant classification [15] (Grade A: <20% height deviation).

Per-metric MAE (Fig. 4(a)) shows DH at Excellent grade (1.26 mm), while VBH (3.07 mm), aVBH (2.77 mm), pVBH (2.86 mm), VBW (3.82 mm), and VBD (2.99 mm) fall within Good grade, yielding an overall morphometric MAE of 2.80 mm (Good). Lordosis (1.62°) and Cobb angle (0.83°) both achieve Excellent grade. For VCL grading (Fig. 4(b)), 97% of vertebrae achieve Grade A on VBH and 95% on aVBH. Bland-Altman analysis (Fig. 4(c)) confirms minimal bias for lordosis (bias = -0.88° , LoA = $[-5.6, 3.8]^\circ$) and Cobb angle (bias = -0.51° , LoA = $[-2.8, 1.8]^\circ$).

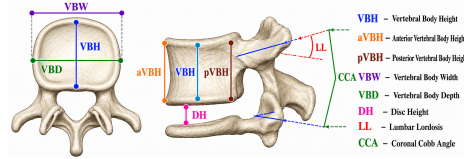


Fig. 3: Vertebral measurement landmarks for eight clinical parameters.

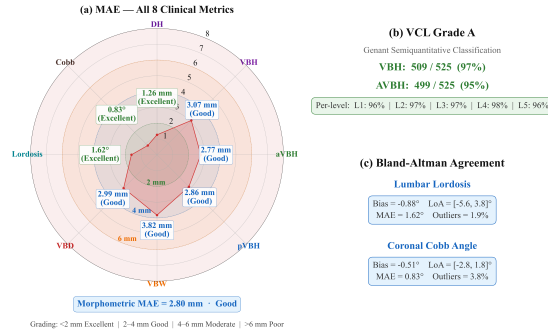


Fig. 4: Clinical validation: (a) MAE radar, (b) VCL Grade A, (c) Bland-Altman.

To validate clinical applicability, we perform phantom-based surgical navigation tracking using a 3D-printed lumbar spine phantom with an optical tracking system and fiducial-based CT-to-phantom registration (Fig. 5). The registration achieves a Fiducial Registration Error (FRE) of 0.41 mm, comparable to the 0.87 mm reported for clinical CT-navigated instrumentation [23]. Using the PPCNet-predicted point cloud as the intraoperative navigation model, the mean vertebra centroid localisation error across L1–L5 is 3.01 mm (range: 2.31–3.85 mm), approaching the clinically accepted navigation accuracy threshold of the Gertzbein-Robbins Grade B (<2 mm pedicle cortical breach) [24] and consistent with screw tip deviations of ~ 5 mm reported in CT-navigated spine surgery [1].

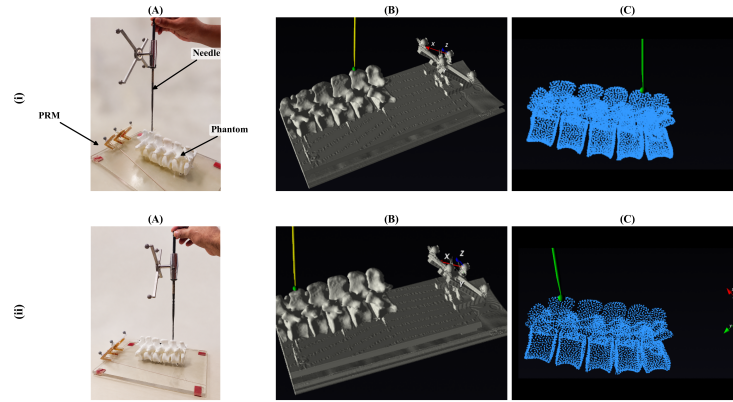


Fig. 5: Phantom-based navigation: (i) needle placed at L4, (ii) needle placed at L1. Each row shows (A) physical setup, (B) CT-derived STL navigation, (C) PPCNet point cloud navigation.

3.2 Ablation Study and Baseline Comparison

Table 1 traces PPCNet across six generations. PPCNet-v1 establishes the baseline with projection-conditioned refinement (2.672 mm). Versions v2–v4 introduce gap-occupancy gating, unified training, and Chamfer-ramp curriculum, collectively reducing CD by 0.222 mm on a shared 21.8M-parameter backbone. PPCNet-v5 adds phased loss scheduling with stable alternatives (1.999 mm), and the final PPCNet-v6 upgrades to ResNet-34, 8,192 points, and curvature/extent losses, reaching 1.981 mm. The progressive improvement is visible in Fig. 6.

Table 1: Six-generation ablation on the test set (105 patients). PPCNet-v1 to v5: 21.8M params, 5,120 points; PPCNet-v6: 38.6M, 8,192 points.

Model	Key addition	CD↓	F@1↑	F@2↑	F@5↑	HD95↓
PPCNet-v1	Baseline + proj. refinement	2.672	0.091	0.458	0.933	6.364
PPCNet-v2	+ gap-occ, gated queries	2.484	0.112	0.504	0.937	6.726
PPCNet-v3	+ unified pipeline	2.452	0.115	0.512	0.942	6.506
PPCNet-v4	+ Chamfer-ramp curriculum	2.450	0.111	0.514	0.944	6.051
PPCNet-v5	+ stable, phased losses	1.999	0.144	0.659	0.972	4.474
PPCNet-v6	+ curvature, extent	1.981	0.155	0.646	0.973	4.525

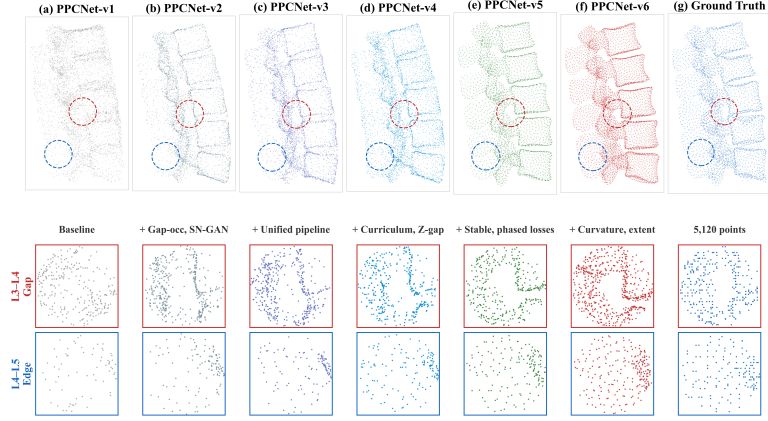


Fig. 6: Qualitative ablation: (a–f) v1 to v6, (g) ground truth. Red: L3–L4 gap sharpens; blue: L4–L5 endplate matches GT by v6.

We re-implement four methods under identical splits (Table 2). Volumetric methods [7,9,10] cluster around 2.47–2.56 mm CD on 64^3 grids (~ 1.5 mm/voxel). 3D-ReVert [11] (38.4M) achieves 2.084 mm. PPCNet yields 1.981 mm (4.9% improvement) via point clouds with explicit geometric projection.

Table 2: Quantitative comparison with re-implemented baselines (105 test patients, same dataset).

Method	Params	CD↓	F@1↑	F@2↑	F@5↑	HD95↓
X2CT-GAN [7]	344.1M	2.489	0.079	0.444	0.933	6.624
BX2S-Net [9]	23.2M	2.564	0.081	0.448	0.927	6.722
Swin-X2S [10]	67.3M	2.475	0.081	0.449	0.937	6.576
3D-ReVert [11]	38.4M	2.084	0.133	0.587	0.966	4.586
PPCNet (Ours)	38.6M	1.981	0.155	0.646	0.973	4.525

The comparison spans methods from 2019 to 2025, including the most recent Swin Transformer [10] and graph convolutional [11] approaches, confirming PPCNet’s consistent advantage across architectural paradigms.

4 Conclusion

We propose PPCNet, a projection-conditioned point cloud reconstruction framework that recovers lumbar vertebrae (L1–L5) as 8,192 points from biplanar

DRRs and calibrated 3×4 projection matrices. By making the projection geometry the core lifting mechanism rather than learning view correspondences implicitly from concatenated DRR features, PPCNet achieves a mean Chamfer distance of 1.981 mm on 105 held-out patients, with clinical validation confirming a morphometric MAE of 2.80 mm and 97% VCL Grade A across 525 vertebrae, and phantom-based navigation tracking yielding a mean centroid error of 3.01 mm. PPCNet demonstrates consistent improvements over recent volumetric and graph-based baselines spanning 2019–2025. Key limitations include reliance on synthetic DRRs and unevaluated robustness under imperfect calibration; bridging the sim-to-real gap remains an important direction for future work.

Acknowledgments. This work was carried out at the Healthcare Technology Innovation Centre (HTIC), IIT Madras.

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

References

1. Virk, S., Qureshi, S.: Navigation in minimally invasive spine surgery. *Journal of Spine Surgery* **5**(Suppl 1), S25–S30 (2019). <https://doi.org/10.21037/jss.2019.04.23>
2. Hoy, D., Brooks, P., Blyth, F., Buchbinder, R.: The epidemiology of low back pain. *Best Practice & Research Clinical Rheumatology* **24**(6), 769–781 (2010). <https://doi.org/10.1016/j.berh.2010.10.002>
3. Zheng, G., Nolte, L.P., Ferguson, S.J.: Scaled, patient-specific 3D vertebral model reconstruction based on 2D lateral fluoroscopy. *International Journal of Computer Assisted Radiology and Surgery* **6**(3), 351–366 (2011). <https://doi.org/10.1007/s11548-010-0515-7>
4. Aubert, B., Vázquez, C., Cresson, T., Parent, S., de Guise, J.A.: Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting. *IEEE Transactions on Medical Imaging* **38**(12), 2796–2806 (2019). <https://doi.org/10.1109/TMI.2019.2914400>
5. Fang, L., Wang, Z., Chen, Z., Jian, F., Li, S., He, H.: 3D shape reconstruction of lumbar vertebra from two X-ray images and a CT model. *IEEE/CAA Journal of Automatica Sinica* **7**(4), 1124–1133 (2020). <https://doi.org/10.1109/JAS.2019.1911528>
6. Humbert, L., de Guise, J.A., Aubert, B., Godbout, B., Skalli, W.: 3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences. *Medical Engineering & Physics* **31**(6), 681–687 (2009). <https://doi.org/10.1016/j.medengphy.2009.01.003>
7. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2CT-GAN: Reconstructing CT from biplanar X-rays with generative adversarial networks. In: *Proc. IEEE/CVF CVPR*. pp. 10619–10628 (2019). <https://doi.org/10.1109/CVPR.2019.01087>
8. Saravi, B., Guzel, H.E., Zink, A., et al.: Synthetic 3D spinal vertebrae reconstruction from biplanar X-rays utilizing generative adversarial networks. *Journal of Personalized Medicine* **13**(12), 1642 (2023). <https://doi.org/10.3390/jpm13121642>

9. Chen, Z., Guo, L., Zhang, R., Fang, Z., He, X., Wang, J.: BX2S-Net: Learning to reconstruct 3D spinal structures from bi-planar X-ray images. *Computers in Biology and Medicine* **154**, 106615 (2023). <https://doi.org/10.1016/j.compbimed.2023.106615>
10. Liu, K., et al.: Swin-X2S: Reconstructing 3D shape from 2D biplanar X-ray with Swin transformers. arXiv preprint arXiv:2501.05961 (2025). <https://doi.org/10.48550/arXiv.2501.05961>
11. Neeraja, R., Anusha, A., Maik, V., Purayath, A., Lakshmanan, M., Sivaprakasam, M.: 3D-ReVert: 3D reconstruction of vertebrae from a single radiograph for minimally invasive spine surgery. In: *Machine Learning in Medical Imaging (MLMI)*. pp. 30–40. LNCS, Springer (2025). https://doi.org/10.1007/978-3-032-09513-8_4
12. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3D object reconstruction from a single image. In: *Proc. IEEE/CVF CVPR*. pp. 2463–2471 (2017). <https://doi.org/10.1109/CVPR.2017.264>
13. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep learning on point sets for 3D classification and segmentation. In: *Proc. IEEE/CVF CVPR*. pp. 652–660 (2017). <https://doi.org/10.1109/CVPR.2017.16>
14. Vrtovec, T., Pernuš, F., Likar, B.: A review of methods for quantitative evaluation of spinal curvature. *European Spine Journal* **18**(5), 593–607 (2009). <https://doi.org/10.1007/s00586-009-0913-0>
15. Genant, H.K., Wu, C.Y., van Kuijk, C., Nevitt, M.C.: Vertebral fracture assessment using a semiquantitative technique. *Journal of Bone and Mineral Research* **8**(9), 1137–1148 (1993). <https://doi.org/10.1002/jbmr.5650080915>
16. Sekuboyina, A., Hussein, M.E., Bayat, A., et al.: VerSe: A vertebrae labelling and segmentation benchmark for multi-detector CT images. *Medical Image Analysis* **73**, 102166 (2021). <https://doi.org/10.1016/j.media.2021.102166>
17. Löffler, M.T., Sekuboyina, A., Jacob, A., et al.: A vertebral segmentation dataset with fracture grading. *Radiology: Artificial Intelligence* **2**(4), e190138 (2020). <https://doi.org/10.1148/ryai.2020190138>
18. Deng, Y., Wang, C., Hui, Y., et al.: CTSpine1K: A large-scale dataset for spinal vertebrae segmentation in computed tomography. arXiv preprint arXiv:2105.14711 (2021). <https://doi.org/10.48550/arXiv.2105.14711>
19. Milickovic, N., Baltas, D., Giannouli, S., Lahanas, M., Zamboglou, N.: CT imaging based digitally reconstructed radiographs and their application in brachytherapy. *Physics in Medicine and Biology* **45**(10), 2787–2800 (2000). <https://doi.org/10.1088/0031-9155/45/10/305>
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proc. IEEE/CVF CVPR*. pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
22. Taha, A.A., Hanbury, A.: Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging* **15**(29) (2015). <https://doi.org/10.1186/s12880-015-0068-x>
23. Gubian, A., Kausch, L., Neumann, J.O., Kiening, K., Ishak, B., Maier-Hein, K., Unterberg, A., Scherer, M.: CT-navigated spinal instrumentations – three-dimensional evaluation of screw placement accuracy in relation to a screw trajectory plan. *Medicina* **58**(9), 1200 (2022). <https://doi.org/10.3390/medicina58091200>

24. Gertzbein, S.D., Robbins, S.E.: Accuracy of pedicular screw placement in vivo. *Spine* **15**(1), 11–14 (1990). <https://doi.org/10.1097/00007632-199001000-00004>