Point Cloud Learning for Automated Reconstruction and Quantification of Human Skeletal Anatomy

Suus van Hostaden
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
contactsuus@uc.edu

Abstract

Automated reconstruction and quantification of human skeletal anatomy from three-dimensional point clouds is complicated by partial coverage, thin cortical surfaces, sparse anatomical labels, and measurement requirements that differ from generic object segmentation. This manuscript develops an anatomy-constrained point-cloud learning framework that links semantic reconstruction, landmark localization, template-based shape completion, and uncertainty-aware quantification in one inferential sequence. The input is a calibrated but incomplete bone-surface point cloud; the outputs are anatomical labels, landmarks, a completed surface consistent with a statistical shape prior, and derived linear, angular, and curvature measures. A hierarchical point encoder extracts local geometry, a multi-task head combines dense segmentation with sparse landmark supervision, and a deformation energy regularizes completion and registration toward anatomically plausible shapes. Quantification is expressed as a measurement map defined on the reconstructed surface, with predictive uncertainty decomposed into epistemic and aleatoric components. An acceptance rule converts calibrated uncertainty and surface coverage into a review decision rather than a silent automated output. The analysis identifies the principal failure modes: landmark displacement near axes of rotation, template bias in rare morphology, hallucinated geometry in unobserved regions, and propagation of segmentation error into clinical measurements. The framework's contribution is methodological: it specifies how reconstruction, shape priors, and measurement uncertainty should be coupled so that quantified skeletal anatomy remains traceable to observed evidence. This framing supports reproducible evaluation and safer translation of point-cloud learning into surgical planning, orthopaedic assessment, and anatomical research.

Keywords

point cloud learning; skeletal anatomy; anatomical landmarks; shape completion; uncertainty quantification; statistical shape models

1. Introduction

The surface geometry of the human skeleton is a primary carrier of clinical and biomechanical information. Preoperative planning, deformity assessment, fracture reduction, implant sizing, forensic identification, and comparative anatomy all depend on the ability to recover bone surfaces and convert their geometry into stable measurements. Point clouds are an attractive representation for this task because they can be acquired from computed tomography surface extraction, intraoperative structured-light scanning, photogrammetry, laser scanning, and intraoral scanning without imposing a voxel grid. The same representation also accommodates irregular sampling, variable point density, and missing regions. These properties make point clouds well suited to skeletal surfaces, but they also make

reconstruction and quantification difficult to separate: a surface that looks plausible may still distort an anatomical axis, a landmark, or a derived angle.

Deep learning on unordered point sets has changed the available modeling options. Point-based architectures can learn permutation-invariant features directly from coordinates and auxiliary attributes [1], and hierarchical neighborhood aggregation enables local geometric detail to be captured across scales [2]. Graph-based and convolution-based formulations extend this capacity to anisotropic neighborhoods and thin structures [3,4]. Medical image analysis has followed a parallel trajectory, with deep networks now used for segmentation, detection, and registration across imaging modalities [5]. Skeletal point clouds nevertheless differ from the benchmarks that drove these methods. Bone surfaces are often partial, articulated, and thin; labels are sparse; and the target is not merely a class map but a set of measurements whose error must be understood.

Statistical shape models provide the complementary prior that point-cloud networks often lack. Active shape models established the idea that correspondences and modes of variation can constrain uncertain boundaries [7], and statistical shape modeling has since become a standard tool for medical segmentation and anatomical analysis [8]. Bone-specific shape models have demonstrated that population-level geometry can supply information when image evidence is incomplete [9]. Point-cloud completion networks can also fill missing regions [10,11], but unconstrained completion risks hallucinating anatomy that is not supported by the observation. A clinically meaningful framework therefore needs to combine learned geometric features with explicit anatomical constraints, while retaining a traceable account of which outputs are observed, which are inferred, and which are too uncertain to act upon.

Sparse supervision is a further constraint. Landmark annotations are expensive, and dense segmentation labels may exist only for a subset of scans. Segmentation-guided semi-supervised landmark detection on intraoral three-dimensional scans illustrates the useful principle that dense shape information can regularize sparse landmark supervision [6]. The present work adopts that principle within a broader reconstruction-and-quantification pipeline rather than treating landmark detection as an isolated task. The central contribution is an anatomy-constrained point-cloud learning framework that formulates reconstruction, landmarking, template-based completion, and measurement extraction as linked inference problems. The framework specifies where anatomical priors enter, how deformation should be regularized, how uncertainty should be decomposed, and how an automated output should be gated by calibrated confidence and coverage. Its value lies in making the assumptions and failure modes explicit before empirical claims are made.

2. Problem Formulation and Anatomical Requirements

A skeletal point cloud is a finite set of points, each carrying coordinates and optional attributes such as normal direction, intensity, scan confidence, or a learned descriptor. The basic observation is

where p_i is the i-th point, x_i, y_i, and z_i are its coordinates in a calibrated reference frame, f_i is an attribute vector, and N is the number of acquired points. The attributes may include surface normals, local curvature estimates, or scanner-specific confidence. The point cloud is unordered, so any inference function must be invariant to permutation of the indices. In practice, P is not a complete bone surface. It may contain only the visible portion of a femur, a hemipelvis, a mandible, or a set of intraoral surfaces. It may also contain soft-tissue points,

metal artifacts, or regions of highly variable sampling density. These conditions are part of the problem definition, not incidental noise.

The output of the framework has four components. The first is a semantic map that assigns each point or local region to an anatomical class. The second is a set of landmarks and axes expressed in the same coordinate frame as the input. The third is a completed surface S that separates observed geometry from inferred geometry. The fourth is a vector of quantitative measurements derived from S and from the landmarks. These outputs are hierarchical: segmentation supports landmark localization, landmarks and labels constrain completion, and completion supports measurement. The hierarchy matters because an error at one level can be hidden by plausible output at another. A completed femoral head may look smooth while its center is displaced, producing an incorrect neck-shaft angle.

The anatomical requirements differ by structure. Long bones require diaphyseal axes, cortical contours, and joint-surface landmarks. The pelvis requires bilateral symmetry and stable acetabular reference frames. The mandible and maxilla require dental and skeletal landmarks that remain meaningful under tooth loss and remodelling. The spine requires vertebral levels, endplate orientation, and curvature. In all cases, the measurement must be defined relative to a coordinate system that is reproducible. A distance between two points is not clinically equivalent to a distance between two anatomical axes, and a surface-distance metric does not capture angular error. The framework therefore treats measurement validity as a separate endpoint from reconstruction fidelity.

A boundary condition follows directly from the representation. A surface point cloud can describe the outer cortical surface but cannot, by itself, recover medullary geometry, bone mineral density, or cortical thickness unless an inner surface is also observed or inferred from volumetric imaging. Cortical thickness claims require paired surfaces or a registered volumetric prior. This boundary condition is important because point-cloud completion can otherwise produce an apparently detailed surface that overstates what the acquisition supports. The framework's quantification layer must therefore declare which measurements are observable, which are model-dependent, and which are outside the representable anatomy.

3. Methodological Framework

3.1. Hierarchical point encoding and geometric normalization

The first stage normalizes the point cloud and extracts local geometric features. Normalization includes rigid alignment to a reference frame when a pose is known, scale verification, removal of gross outliers, and density equalization through resampling. Because skeletal surfaces contain thin edges, ridges, and joint margins, a single neighborhood radius is inadequate. A hierarchical encoder aggregates information over several scales. A compact representation of one local aggregation step is

$$h_i = \max_j \in N(i)\varphi\left(p_i, p_j - p_i\right)$$

where h_i is the feature at point i, N(i) is the set of neighboring points, $p_j - p_i$ is the relative displacement from the central point to a neighbor, and φ is a learned transformation with shared parameters. The maximum operation supplies permutation invariance within the neighborhood. Multiple such layers with changing radii produce features that encode fine cortical detail and broader bone shape. Point-based set abstraction [2], dynamic graph construction [3], and deformable kernel convolution [4] are credible instantiations of this idea. The choice among them affects sensitivity to density and thin structures. Graph and kernel

methods can preserve local anisotropy better than isotropic ball queries, but they also impose greater computational cost.

The encoder should receive normals and curvature descriptors when available, because coordinates alone do not distinguish an outer cortical surface from an inner table or a fracture fragment. Normals must be oriented consistently before training; inconsistent orientation creates a systematic ambiguity that no amount of capacity can resolve. The normalized coordinates should be expressed relative to an anatomical or acquisition frame that is recorded with the scan. If the frame is estimated from the same partial surface that is being reconstructed, the estimation uncertainty must be propagated rather than discarded.

3.2. Segmentation, landmarking, and sparse supervision

The second stage combines dense anatomical segmentation with sparse landmark detection. Voxel-based medical segmentation networks provide a useful baseline, particularly when the point cloud is derived from volumetric imaging [12,13]. Point-native models are preferable when the input remains a surface sample, because voxelization can merge thin structures and introduce grid-dependent anisotropy. The multi-task objective couples the two predictions:

$$L = Lseg + \lambda Llmk + \mu Lcons$$

In this expression, Lseg is the segmentation loss, Llmk is the landmark or heatmap loss, Lcons penalizes inconsistency under augmentation or rigid perturbation, and λ and μ are nonnegative weights. Lseg may combine class-balanced cross-entropy with a region-overlap term. Llmk may use Gaussian heatmaps or direct coordinate regression with robust losses. Lcons enforces agreement between predictions on the original cloud and a transformed copy, which is useful when landmark labels are sparse and the network would otherwise overfit to local coordinate patterns.

Semi-supervision enters through unlabeled or weakly labeled scans. Dense segmentation predictions can supply shape context for landmark localization, while landmark locations can constrain the spatial arrangement of predicted classes. The segmentation-guided semi-supervised strategy used for intraoral scans [6] is relevant here because it shows how dense geometric labels can reduce reliance on exhaustive landmark annotation. The framework does not assume that this strategy transfers without adaptation. Intraoral surfaces include teeth, gingiva, and bone in close proximity, whereas long-bone scans may include only cortical surface. The consistency term must therefore be designed around the anatomy being modeled.

Landmark definitions must be fixed before training and documented in a coordinate-independent form. A landmark defined as the most prominent point on a curvature ridge is stable only if the ridge is observable; a landmark defined by a tangent intersection may remain stable under surface noise but can be sensitive to the fitted axes. The framework treats each landmark as having an observability class: directly visible, inferable from surrounding geometry, or dependent on a population prior. This classification determines whether the landmark can be used in a fully automated measurement or must be flagged for review.

3.3. Template-constrained completion and registration

Completion and registration are coupled because a template must be deformed and aligned to the observed anatomy. Let S be a template surface with vertices s_i, let d_i be a nonrigid displacement at vertex i, let R be a rotation, let t be a translation, and let q_i be the corresponding observed point. A regularized completion and registration energy is

$$E(D,R) = \sum_i \| R(s_i + d_i) + t - q_i \|^2 + \alpha \sum_i \| d_i \|^2$$

where D collects the displacements, α controls the trade-off between data fidelity and deformation magnitude, and the sum runs over established correspondences. The first term pulls the deformed template toward the observed surface. The second term suppresses implausible local deformation. A statistical shape prior can replace or augment the second term by penalizing displacement components outside the learned modes of variation [7,8]. Bone-specific models provide an anatomical basis for this regularization [9]. Completion networks that predict missing geometry from partial input [10,11] can initialize the template or propose residual displacements, but their output should not be accepted without an observation-aware constraint.

The rigid component can be estimated separately once correspondences are approximately known:

This is the classical orthogonal Procrustes problem, and its solution is used within iterative closest-point registration [14]. Generalized Procrustes analysis provides a related mechanism for aligning multiple shapes and estimating a mean shape [15]. In the proposed framework, registration is not a preprocessing step that is completed before learning. It is part of the inference loop: segmentation determines which observed points are eligible correspondences, landmark predictions constrain the rigid alignment, and the shape prior limits nonrigid deformation. This loop is necessary because partial skeletal surfaces often contain several locally plausible alignments. A femoral shaft alone may not disambiguate anteversion, and a hemipelvis may not determine the sagittal plane without sacral or pubic landmarks.

Completion should be represented as a distribution of plausible surfaces rather than a single deterministic output. The observed region supplies direct evidence; the unobserved region is filled by the prior and must carry a higher uncertainty. For this reason, the framework records an observation mask alongside the completed surface. Measurements that depend only on observed geometry can be reported with lower model dependence, whereas measurements that cross unobserved regions must be marked as prior-dependent. This distinction is more informative than a single global confidence score.

3.4. Quantification and uncertainty-aware acceptance

Quantification is a mapping from the reconstructed surface and landmark set to clinically meaningful measures. A general form is

where m_k is the k-th measurement, f_k is its defining rule, and θ_k contains any required axis, plane, or landmark definitions. Examples include a distance between two landmarks, an angle between anatomical axes, a curvature integral along a cortical ridge, a version angle relative to a reference plane, or a joint-space width when both articular surfaces are present. The measurement rules must be specified before evaluation so that the reconstruction cannot be tuned to a preferred measurement outcome. This separation also allows reconstruction accuracy and measurement validity to be assessed independently.

Uncertainty should be decomposed into epistemic and aleatoric parts. A practical predictive variance for an input x can be estimated from T stochastic forward passes:

$$V = (1/T)\sum_t y_t^2 - ((1/T)\sum_t y_t)^2 + \sigma^2$$

Here y_t is the predicted measurement or surface displacement on pass t, the first two terms estimate the variance of the predictive mean across passes, and σ^2 is the estimated observation noise. Monte Carlo dropout and related Bayesian approximations provide a computationally tractable route to the first component [16,17]. The decomposition distinguishes uncertainty that could be reduced by more data or a better model from uncertainty that reflects irreducible acquisition noise. For a clinical measurement, both matter, but they imply different responses: epistemic uncertainty suggests additional views or human review, whereas aleatoric uncertainty may require a tolerance-based decision.

The final acceptance rule converts uncertainty and coverage into a triage decision:

$$a(x) = 1 \; if \; V(x) \leq \tau \; and \; c(x) \geq c_0, \; else \; 0$$

where a(x) is the acceptance indicator, V(x) is the predictive variance, c(x) is a coverage or completeness score for the anatomical region required by the measurement, and τ and c_0 are tolerances fixed by the clinical or scientific use case. The rule is deliberately conservative. A low global surface error does not justify a measurement that crosses a large unobserved region, and a high coverage score does not compensate for a landmark with large predictive variance. The thresholds are not universal constants; they must be derived from measurement tolerances and metrological validation.

4. Analytical Assessment of Failure Modes and Design Trade-offs

The framework's main vulnerabilities can be analyzed without claiming empirical performance. The first is landmark displacement near axes of rotation. Small errors in landmark position can produce large angular errors when the landmarks define a short baseline. A neck-shaft angle, an acetabular version angle, or a mandibular plane angle may be highly sensitive to a few millimetres of displacement if the reference points are close together. The segmentation and landmark heads should therefore be evaluated with angular error, not only point error. The consistency term in L helps only if the augmentations preserve the anatomical definition; random rotations that move a landmark relative to a ridge can teach the network to ignore the very geometry that defines the landmark.

The second failure mode is template bias. A statistical shape prior improves completion when the observed anatomy is represented by the training population, but it can systematically distort rare morphology. The weight α in the completion energy controls this trade-off. Large α produces smooth, population-like surfaces that may be stable but biased. Small α follows noisy observations and may hallucinate local structure. A useful safeguard is to report the residual between the completed surface and the observed region, and to flag cases in which the nonrigid deformation lies outside the prior's plausible range. The framework should not use the prior to erase anatomy that is visible but unusual.

The third failure mode is segmentation error propagation. Segmentation errors that occur on a cortical ridge may displace a landmark; segmentation errors on a joint surface may bias a curvature or joint-space measure. Multi-task learning can reduce this propagation because landmark supervision shapes the segmentation features, but it can also create correlated errors if both heads rely on the same spurious cue. Independent evaluation of segmentation, landmarking, completion, and measurement is therefore necessary. Surface-overlap metrics are informative for segmentation but insufficient for measurement validity. Evaluation should include surface distance, landmark error, angular error, and agreement analysis for derived measures; established metric guidance for three-dimensional segmentation supports this separation [18]. Combined shape-model and convolutional approaches to bone segmentation

further illustrate why shape constraints and image evidence should be evaluated together rather than treated as interchangeable [19].

The fourth failure mode is distribution shift across scanners, anatomy, and populations. Point density, noise, and field of view can change the local feature statistics that the encoder learns. A model trained on reconstructed computed tomography surfaces may not transfer directly to intraoperative structured-light scans, where specularity and occlusion produce different missingness patterns. Normalization and multi-scale encoding reduce but do not remove this shift. The observation mask and uncertainty output are essential because they make shift visible at the level of individual cases. A system that returns a confident measurement for an out-of-distribution surface is more dangerous than one that returns a lower-confidence result with a coverage map.

The design choices also involve trade-offs between end-to-end and staged inference. An end-to-end network that predicts measurements directly from point clouds can reduce compounding errors, but it is difficult to audit and may learn dataset-specific shortcuts. A staged framework with explicit segmentation, landmarking, completion, and measurement keeps each step inspectable. The cost is that errors can accumulate across stages. The proposed framework mitigates this cost by sharing features across tasks and by propagating uncertainty, but it does not eliminate the need for stage-wise validation. Point-native models offer another trade-off: they preserve thin structures better than voxel grids but can be more sensitive to sampling density and neighborhood construction. The appropriate choice depends on the acquisition modality and on whether the clinical question concerns cortical detail or global alignment.

5. Technical Discussion and Implications

The central implication of this framework is that skeletal point-cloud learning should be evaluated as a measurement system, not only as a shape-prediction system. A reconstructed surface can be visually convincing while producing biased angles, and a segmentation can have high overlap while displacing a critical landmark. The proposed separation of observed, inferred, and quantified geometry makes these distinctions explicit. It also provides a basis for failure analysis: observed-region errors point to acquisition or encoding problems; completion-region errors point to prior misspecification; measurement errors that persist despite accurate surfaces point to landmark or axis definitions.

A second implication concerns the role of anatomical priors. Priors are not merely regularizers that improve smoothness. They encode population assumptions about shape, and those assumptions can be wrong for an individual. The framework therefore treats the prior as a source of information with a measurable influence. The deformation magnitude, the residual outside the observed region, and the uncertainty of prior-dependent measurements all expose that influence. This is preferable to a black-box completion that returns a full surface without indicating which parts were inferred.

A third implication concerns clinical deployment. Automated quantification should be paired with an acceptance rule that can withhold or escalate a case. The rule in this framework is intentionally simple, but its inputs are substantive: predictive variance, coverage, and measurement tolerance. In practice, the tolerances must be tied to the decision being supported. Implant sizing may require tighter angular and linear tolerances than an anatomical study of population variation. The same reconstruction can therefore be acceptable for one use

and unacceptable for another. The framework's explicit measurement map and uncertainty decomposition make this use-dependence visible.

The approach also has clear boundary conditions. It assumes a calibrated point cloud and a meaningful anatomical coordinate frame. It assumes that the observed surface contains enough evidence to constrain the pose and the major shape modes. It does not recover internal bone structure from an outer surface alone, and it does not replace volumetric imaging when medullary geometry or density is required. It depends on landmark definitions that are stable under the expected missingness patterns. These conditions are not defects of the framework; they define the domain in which the framework's outputs can be interpreted.

6. Conclusion

This manuscript has developed an anatomy-constrained point-cloud learning framework for automated reconstruction and quantification of human skeletal anatomy. The framework treats a skeletal point cloud as a partial, calibrated observation and links hierarchical geometric encoding, multi-task segmentation and landmark inference, template-constrained completion, registration, and measurement extraction in a single inferential sequence. Sparse supervision is handled through dense shape context and consistency constraints, while completion is regularized by statistical shape priors and by an explicit separation between observed and inferred geometry. Quantification is defined through measurement maps, and predictive uncertainty is decomposed into epistemic and aleatoric components that feed an acceptance rule based on variance and coverage. The analytical assessment identifies the main failure modes as landmark displacement near rotational axes, template bias in rare morphology, hallucinated completion in unobserved regions, and propagation of segmentation error into derived measurements. The framework's contribution is to make these mechanisms explicit and to specify how anatomical priors, learned geometry, and uncertainty should be coupled so that quantified skeletal anatomy remains traceable to observed evidence. This orientation supports reproducible evaluation and cautious translation of point-cloud learning into surgical planning, orthopaedic assessment, and anatomical research.

References

1. Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 652–660.

2. Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems, 30, 5099–5108.

3. Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics, 38(5), Article 146.

4. Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., & Guibas, L. J. (2019). KPConv: Flexible and deformable convolution for point clouds. Proceedings of the IEEE/CVF International Conference on Computer Vision, 6411–6420.

5. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

6. Liu, S., Jin, J., Li, W., & Yan, X. (2026). Segmentation-Guided Semi-Supervised Landmark Detection on Intraoral 3D Scans. Preprints. https://doi.org/10.20944/preprints202609.1722.v1

7. Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models—their training and application. Computer Vision and Image Understanding, 61(1), 38–59.

8. Heimann, T., & Meinzer, H.-P. (2009). Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis, 13(4), 543–563.

9. Sarkalkan, N., Weinans, H., & Zadpoor, A. A. (2014). Statistical shape and appearance models of bones. Bone, 60, 129–140.

10. Yuan, W., Khot, T., Held, D., Mertz, C., & Hebert, M. (2018). PCN: Point completion network. 2018 International Conference on 3D Vision, 728–737.

11. Xie, H., Yao, H., Zhou, S., Mao, J., Zhang, S., & Sun, W. (2020). GRNet: Gridding residual network for dense point cloud completion. Computer Vision – ECCV 2020, 12351, 365–381.

12. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 9351, 234–241.

13. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, 9901, 424–432.

14. Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2), 239–256.

15. Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society: Series B, 53(2), 285–339.

16. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? Advances in Neural Information Processing Systems, 30, 5574–5584.

17. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of the 33rd International Conference on Machine Learning, 48, 1050–1059.

18. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging, 15, Article 29.

19. Ambellan, F., Tack, A., Ehlke, M., & Zachow, S. (2019). Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative. Medical Image Analysis, 52, 109–118.