

X2BR: High-Fidelity 3D Bone Reconstruction from a Planar X-Ray Image with Hybrid Neural Implicit Methods

Gokce Guven¹, H. Fatih Ugurdag¹, Hasan F. Ates¹

¹Faculty of Engineering, Ozyegin University, Istanbul, Turkey
`{gokce.guven, fatih.ugurdag, hasan.ates}@ozu.edu.tr`

Abstract

Accurate 3D bone reconstruction from a single planar X-ray remains a challenge due to anatomical complexity and limited input data. We propose X2BR, a hybrid neural implicit framework that combines continuous volumetric reconstruction with template-guided non-rigid registration. The core network, X2B, employs a ConvNeXt-based encoder to extract spatial features from X-rays and predict high-fidelity 3D bone occupancy fields without relying on statistical shape models. To further refine anatomical accuracy, X2BR integrates a patient-specific template mesh, constructed using YOLOv9-based detection and the SKEL biomechanical skeleton model. The coarse reconstruction is aligned to the template using geodesic-based coherent point drift, enabling anatomically consistent 3D bone volumes. Experimental results on a clinical dataset show that X2B achieves the highest numerical accuracy, with an IoU of 0.952 and Chamfer-L1 distance of 0.005, outperforming recent baselines including X2V and D2IM-Net. Building on this, X2BR incorporates anatomical priors via YOLOv9-based bone detection and biomechanical template alignment, leading to reconstructions that, while slightly lower in IoU (0.875), offer superior anatomical realism, especially in rib curvature and vertebral alignment. This numerical accuracy vs. visual consistency trade-off between X2B and X2BR highlights the value of hybrid frameworks for clinically relevant 3D reconstructions.

Keywords: bone reconstruction, X-ray, ConvNeXt, non-rigid registration, implicit networks.

1 Introduction

Neural implicit representations have revolutionized 3D shape modeling by enabling high-fidelity reconstructions without explicit parameterization. These methods encode objects as continuous functions, allowing for precise 3D reconstruction, novel view synthesis, and shape interpolation. Their ability to model fine geometric details makes them essential for applications in virtual reality, robotics, and medical imaging.

Single-image 3D reconstruction, powered by deep learning, has significantly advanced spatial understanding from 2D inputs, benefiting fields such as augmented reality, autonomous navigation, and medical diagnostics. In medical imaging, neural implicit representations provide a powerful tool for high-resolution volumetric reconstruction, enabling detailed anatomical modeling for applications such as surgical planning, biomechanical analysis, and patient-specific treatment design. However, their continuous and unconstrained volumetric outputs may not inherently preserve anatomical structure or biomechanical validity. Therefore, integrating neural implicit methods with domain-specific priors—such as biomechanical templates or anatomical landmarks—is essential to ensure anatomically consistent and clinically reliable reconstructions.

Recent advances demonstrate the efficacy of neural implicit methods across various medical domains. X2Teeth [1] reconstructs individual teeth from panoramic radiographs, while Oral-3Dv2 [2] employs implicit functions to map 2D coordinates to 3D dental structures. ToothInpainter [3] fuses partial 3D models and X-rays for comprehensive dental reconstructions, including roots. In broader medical imaging, MedNeRF [4] generates high-resolution CT-like projections from sparse X-rays, ImplicitVol [5] reconstructs 3D ultrasound volumes without voxel grids, and SAX-NeRF [6] applies line-based transformers for improved sparse-view X-ray reconstructions. SNAF [7] further extends these capabilities by refining CBCT reconstructions with neural attenuation fields. These studies highlight the transformative impact of implicit representations in medical imaging.

This study introduces X2B and X2BR, two complementary neural implicit frameworks for 3D skeletal reconstruction from a single planar X-ray. X2B employs a ConvNeXt-based encoder to extract hierarchical spatial features and predict continuous occupancy fields, enabling 3D reconstruction of complex skeletal structures such as ribs and vertebrae. It effectively handles challenges such as overlapping Hounsfield Unit (HU) values and incomplete anatomical input without relying on voxel grids or predefined statistical templates. Building upon this, X2BR integrates a template-guided non-rigid registration step using a biomechanical skeleton model and geodesic-based coherent point drift (GBCPD++). This hybrid approach refines the initial reconstruction, ensuring anatomical consistency and improving alignment to patient-specific skeletal variations. Together, X2B and X2BR offer powerful solutions for accurate and personalized 3D bone modeling from sparse imaging data.

Experiments on clinical datasets demonstrate that X2B achieves state-of-the-art accuracy, significantly outperforming existing methods in volumetric IoU, Chamfer-L1 distance, and F-score. X2BR, while yielding slightly lower numerical accuracy, provides significantly improved anatomical consistency by incorporating a biomechanical template into the reconstruction process. This template serves as a prior anatomical structure, guiding the non-rigid registration of the occupancy-based output. By aligning the reconstructed volume to a deformable skeletal model, X2BR enforces biomechanical plausibility and better accommodates patient-specific anatomical variations. The framework’s ability to generate high-resolution 3D skeletal reconstructions from sparse imaging data—while leveraging a deformable anatomical template—makes it particularly well-suited for applications such as surgical planning, orthopedic assessment, and patient-specific biomechanical simulations.

In summary, the contributions of this study are as follows:

- Proposes a hybrid neural implicit framework for 3D skeletal reconstruction from a single planar X-ray, combining a template-free occupancy-based model (X2B) with a template-guided refinement module (X2BR) for enhanced anatomical consistency.
- Incorporates ConvNeXt-based encoder to enhance spatial feature extraction for precise skeletal reconstruction.
- Handles anatomical challenges, including overlapping HU values and missing anatomical regions.
- Enables high-resolution, continuous and anatomically consistent modeling of 3D bone structures.
- Introduces the largest real-patient dataset of 3D bone meshes and corresponding DRGs (digitally reconstructed radiograph).
- Achieves state-of-the-art performance, surpassing existing methods in IoU, Chamfer-L1, and F-score.

2 Related Work

2.1 Single-View Reconstruction with Implicit Surface Representations

Neural implicit methods commonly use MLPs to represent occupancy probabilities or signed distance functions (SDFs) for 3D reconstruction from single images [8, 9]. While previous neural implicit approaches effectively reconstruct general 3D shapes, they fail to address the anatomical complexity of skeletal structures from single planar X-rays; our proposed X2B and X2BR frameworks specifically bridge this gap, improving reconstruction accuracy for complex skeletal anatomies.

Building on these foundations, recent CNN-based methods such as DISN [9], MDISN [10], and Ray-ONet [11] improve reconstruction fidelity but continue to face limitations in accurately modeling complex geometries. More advanced frameworks like D2IM-Net [12], ED2IF2-Net [13], G2IFu [14], and LIST [15] further enhance topological accuracy and surface detail reconstruction. However, these approaches are primarily designed for general-purpose 3D reconstruction and remain insufficient when applied to anatomically intricate medical structures from sparse, single-view X-ray data.

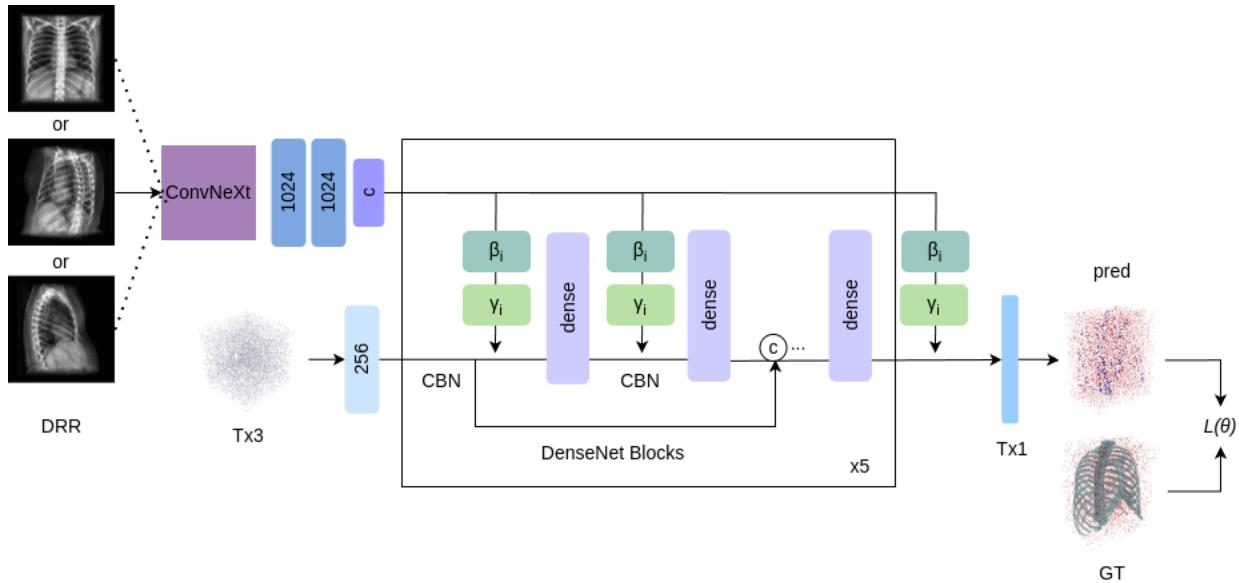


Figure 1: X2B network training pipeline. The figure illustrates the training process of the X2B network, which takes a DRR as input and uses a ConvNeXt backbone for feature extraction. The extracted features are passed through dense blocks with Conditional Batch Normalization (CBN) layers, parameterized by β_i and γ_i , to refine the latent representations.

2.2 X-ray to 3D Bone Reconstruction with Deformation Learning

Statistical Shape Models (SSMs) utilize atlases derived from healthy samples to model mean anatomical shapes and variations [16, 17]. Aubert et al. [18] combined SSMs with CNN-based landmark detection for automated 3D spinal reconstruction. Jiang et al. [19] proposed a 2D/3D registration method for spinal geometry reconstruction from frontal X-rays, while X23D [20] integrated multi-view stereo and X-ray calibration for intraoperative vertebrae modeling.

More recent deep learning approaches also leverage deformation modeling. BX2S-Net [21] employs bi-planar X-rays with encoder-decoder architectures and attention mechanisms for improved semantic alignment. Similarly, Yang et al. [22] adapted X2CT-GAN [23] to reconstruct spinal structures from bi-planar radiographs. For single-view reconstruction, the approach in [24] utilizes deep learning with deformation parameters to reconstruct accurate 3D femoral models, while FracReconNet [25] improves fracture reconstruction accuracy by augmenting training data.

Existing methods often depend on bi-planar inputs, explicit deformation models, or statistical shape templates, limiting their generalizability and patient specificity. In contrast, X2B reconstructs high-fidelity 3D skeletal structures directly from single planar X-rays without requiring such priors. X2BR further enhances anatomical consistency by integrating a patient-specific template and geodesic-based non-rigid alignment. Together, they combine the strengths of data-driven reconstruction and anatomy-aware refinement for robust single-view bone modeling.

2.3 Implicit Neural Representations in Medical Imaging

Recent advances in implicit neural representations have significantly improved medical imaging reconstruction. X2Teeth [1] employs three subnets—ExtNet, SegNet, and ReconNet—to extract features, enhance segmentation, and reconstruct individual teeth. Oral-3Dv2 [2] maps 2D coordinates to 3D voxel densities, leveraging dynamic sampling for improved detail. ToothInpaintor [3] reconstructs full dental models, including roots, using implicit representations. MedNeRF [4] employs neural radiance fields to generate CT projections from X-rays, while ImplicitVol [5] refines 3D volumes from 2D ultrasound without voxel grids. SAX-NeRF [6] and SNAF [7] enable 3D reconstruction from sparse X-ray views using specialized sampling and augmentation techniques. Our previous work X2V [26] utilizes a ViT-based implicit model to reconstruct 3D lung volumes from a single X-ray, without relying on mesh templates, and achieves state-of-the-art accuracy using occupancy networks.

Despite these advances, most existing methods are constrained to soft tissues, isolated anatomical regions, or require regular, high-contrast geometries. For instance, X2V is limited to organ volumes like lungs and relies on air-filled structures for accurate occupancy estimation. In contrast, our X2BR framework extends implicit neural representations to complex skeletal anatomy—including articulated structures such as vertebrae—using biomechanical template alignment and non-rigid registration. This enables anatomically consistent, high-fidelity 3D reconstructions from a single planar X-ray, even in the absence of dense multi-view inputs or statistical shape priors.

3 Proposed Method

Reconstructing accurate 3D bone structures from single planar X-rays is challenging due to anatomical complexity and overlapping tissue intensities. This study introduces two neural implicit models, X2B and X2BR, for precise and patient-specific 3D reconstructions.

3.1 X2B Network Architecture

The X2B network processes a 224×224 DRR image and T random 3D points to predict occupancy probabilities, indicating whether a point lies within the bone surface (see Figure 1). A ConvNeXt-based encoder extracts hierarchical spatial features, transforming the input into a 1024-dimensional latent representation. ConvNeXt optimizes computational efficiency using depthwise convolutions, GELU activations, and Layer Normalization (LN). The occupancy network combines ConvNeXt features with 3D points, processed through DenseNet blocks with Conditional Batch

Normalization (CBN) [27], which dynamically adjusts normalization based on contextual features. This architecture ensures robust feature extraction and smooth reconstruction, leveraging the occupancy function [8] to implicitly define the bone surface as a decision boundary. As explained below, DRRs captured from multiple angles are used during training of X2B to enhance the network’s ability to reconstruct 3D structures with greater precision.

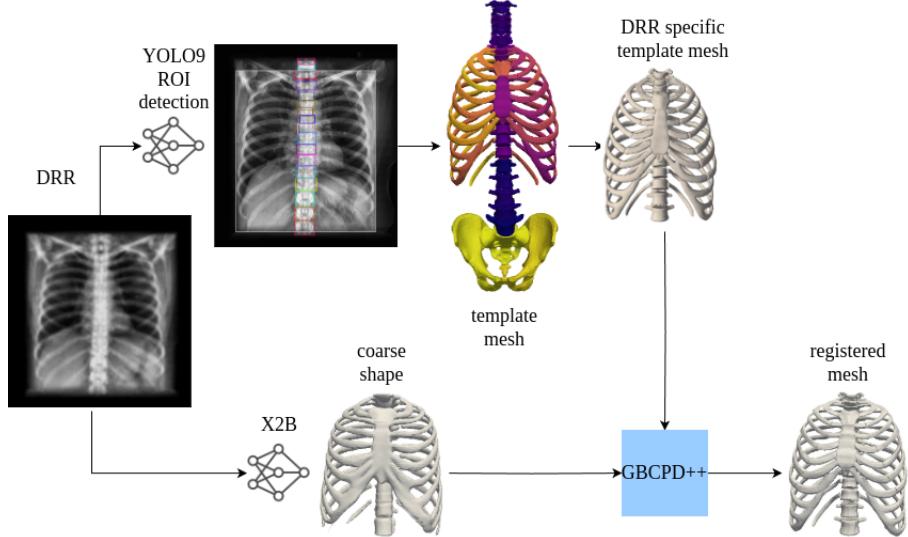


Figure 2: X2BR model architecture. Anterior-posterior DRR is used as input for both YOLOv9 and X2B for inference. The regions are detected via YOLOv9 network and DRR specific template mesh is extracted from the template mesh model using the detected regions. DRR specific template model is registered to the coarse shape, which is the output of the X2BR model.

3.2 X2BR Network Architecture

Building on X2B, X2BR integrates non-rigid registration to refine coarse reconstructions (see Figure 2). A fine-tuned YOLOv9 model identifies vertebrae (cervical, thoracic, lumbar) and ribs from DRR images. These detected regions are combined with the SKEL Biomechanical Skeleton Model (BSM) [28], developed using OpenSim [29], to construct a patient-specific template mesh. The Geodesic Bayesian Coherent Point Drift (GBCPD++) algorithm [27] aligns the coarse 3D shape from X2B with the patient-specific template, addressing complex deformations.

The combination of ConvNeXt encoding, CBN normalization, YOLOv9 object detection, and GBCPD++ alignment enables X2B and X2BR to address the limitations of traditional methods. These models deliver precise 3D reconstructions, making them suitable for applications in surgical planning, biomechanical analysis, and personalized treatment.

The SKEL Body Shape Model (BSM) provides watertight meshes of the thorax, pelvis, and spine, serving as the foundation for generating a DRR-specific template model. The X2BR model utilizes these meshes to construct subject-specific skeletal templates optimized for non-rigid registration. By leveraging the SKEL BSM, the X2BR model ensures high anatomical fidelity while efficiently adapting the template to align with input DRR data. This enables precise deformation and anatomically accurate reconstructions tailored to individual subjects.

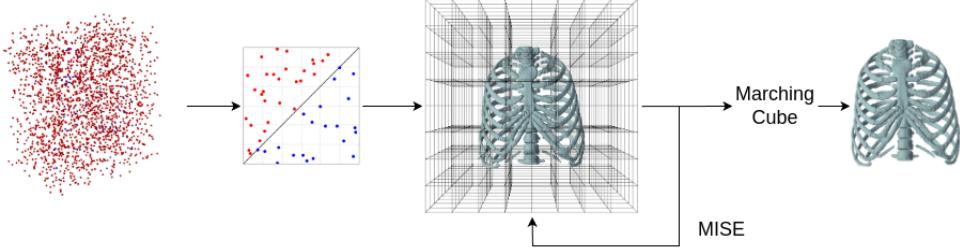


Figure 3: For 3D mesh inference with $X2B$ model, a modified Multiresolution IsoSurface Extraction (MISE) algorithm for high-resolution mesh extraction is integrated, starting with a base resolution and evaluating against the occupancy network. The occupancy threshold is set at $\tau = 0.2$ for balance in accuracy and completeness. The process involves subdividing voxels until the desired resolution is reached, using Marching Cubes for mesh generation, and refining the mesh with Fast-Quadric-Mesh-Simplification and gradient optimization. Our method achieves efficient and accurate mesh inference, optimized for an initial resolution of 32^3 , and is capable of extracting mesh normals effectively.

3.3 Training of X2B

The $X2B$ network is trained to reconstruct 3D skeletal structures from DRRs. As depicted in Figure 1, during training of the model a randomly selected input DRR image that is captured at different angles is passed through a ConvNeXt backbone, which extracts multi-dimensional feature representations. These features are then processed through a series of DenseNet blocks that include Conditional Batch Normalization (CBN) layers, parameterized by β_i and γ_i , to refine the latent representations. The network iteratively processes these features through multiple dense layers, producing a final prediction that reconstructs the 3D skeletal structure. The loss function $L(\theta)$ is computed by comparing the predicted 3D output to the ground truth, enabling the network to optimize its parameters for accurate reconstruction.

The training objective is to estimate the occupancy at every point $p \in \mathbb{R}^3$ to derive the coarse shape from the $X2B$ network’s output. The occupancy function

$$o : \mathbb{R}^3 \rightarrow \{0, 1\}$$

defines whether a 3D point is occupied. This function can be approximated by a neural network, which assigns each point p an occupancy probability between 0 and 1, using a binary classification network’s decision boundary to implicitly represent the object’s surface.

The network, represented as $f : \mathbb{R}^3 \times X \rightarrow \mathbb{R}$, maps the pair (p, x) to a real number indicating the probability of occupancy, where $x \in X$ is the observation conditioning the 3D reconstruction (3Dr) task.

Training the binary-classification network involves computing the mini-batch loss $L_B(\theta)$ at randomly sampled points within the object’s 3D bounding volume. This is expressed as:

$$L_B(\theta) = \frac{1}{|B|} \sum_{i=1}^{|B|} L(f_\theta(p_i, x_i), o_i) \quad (1)$$

where x_i is the i^{th} observation in batch B , $o_i \equiv o(p_i)$ represents the true occupancy at point p_i , and $L(\cdot, \cdot)$ is the binary cross-entropy classification loss. The method’s effectiveness depends on the sampling strategy for locations p_i during training, with optimal results achieved through uniform sampling within the object’s bounding box [8].

For all experiments, the Adam optimizer [30] was employed with a learning rate (η) of 10^{-4} . The default hyperparameters provided by PyTorch were used: $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.

3.4 Inference

Figure 3 demonstrates the inference of 3D meshes from occupancy predictions at the output of the X2B network. In the X2B implementation, a modified Multiresolution IsoSurface Extraction (MISE) algorithm [8] is employed to enable efficient high-resolution mesh extraction.

The process starts with the discretization of the volumetric space and the evaluation of occupancy values at this initial resolution. Voxelization involves thresholding to classify each voxel as inside or outside the object based on the model’s predictions. This critical step converts probabilistic outputs into a discrete 3D representation, facilitating accurate reconstruction from a single image viewpoint. Grid points exceeding an occupancy threshold, set to $\tau = 0.2$ for the X2B method, determine the surface thickness, balancing accuracy and completeness as per ONet recommendations [8].

Active voxels are identified and subdivided iteratively until the desired resolution is reached. The Marching Cubes algorithm then generates the initial mesh, which is refined using the Fast-Quadric-Mesh-Simplification algorithm. This algorithm applies iterative edge contraction and quadric error metrics, followed by gradient-based optimization, ensuring mesh quality enhancement and simplification.

3.5 Registration

The Geodesic-Based Bayesian Coherent Point Drift algorithm [27] introduces a non-rigid registration method addressing a significant drawback of the traditional Coherent Point Drift (CPD) algorithm. CPD, while prevalent for shape matching and deformation, often unnaturally deforms shapes with neighboring parts, such as human legs. This issue arises from the proximity-based deformation constraint known as motion coherence. The GBCPD++ method redefines motion coherence using geodesic distances, the shortest paths on a shape’s surface, to mitigate inappropriate deformations. Numerical experiments demonstrate the efficacy of the geodesic-based approach in avoiding CPD’s shortcomings and its scalability to handle shapes with millions of points. Key contributions include utilizing geodesic and Gaussian kernels for improved registration, providing a theoretical basis for converting indefinite geodesic kernels into positive-semidefinite ones, and employing the Nyström method and parallel computations to accelerate the process. The GBCPD++ method, integrated into the X2BR model, refines the reconstruction of 3D skeletal structures by enabling anatomically consistent registration of detailed vertebral features. By leveraging geodesic distances and scalable computations, it addresses complex deformations and enhances the accuracy of skeletal alignment.

4 Dataset

The dataset curated for this study is essential for advancing accurate 3D bone reconstruction from X-ray images. Due to the lack of publicly available paired datasets of X-rays and 3D bone meshes, a custom, high-quality dataset was created. Using advanced segmentation techniques, state-of-the-art DRR technology, and preprocessing pipelines, the dataset includes CT scans, DRR images, and 3D bone meshes systematically processed to ensure precision and suitability for training and validation. This section details the data acquisition, segmentation, DRR generation, and occupancy value computation, enabling efficient and accurate mesh reconstruction.

4.1 Segmentation with TotalSegmentator

The X2B network isolates organs from X-ray images, focusing on ribs, costal cartilages, and vertebrae due to their high HU contrast and the abundance of thorax data. Early experiments demonstrated the superior accuracy of TotalSegmentator [31] for bone segmentation, prompting its use in this study. Scapula bones were excluded due to frequent segmentation errors. Dataset generation steps are summarized as follows:

- Data collection from The Cancer Imaging Archive (TCIA).
- Bone segmentation from CT images into watertight manifold meshes using TotalSegmentator [31].
- Calculation of occupancy values for bone mesh points.
- Generation of corresponding DRR images from CT scans in multiple views.
- Enhancement of DRR images using CLAHE contrast adjustment.

CT scans from 964 subjects in the NLST dataset [32] were curated, excluding scans with incomplete thorax coverage or poor quality. DRR technology [33] was utilized to synthesize paired X-rays from real CT volumes. Scans were aligned to consistent anatomical planes, and low-quality segmentations were excluded.

A total of 3,640 CT scans from 964 subjects were processed. Segmentation achieved a Dice Similarity Coefficient (DSC) of 0.943 ± 0.04 [31]. A custom Python script [34–37] automated resampling to 1.5mm isotropic spacing, segmentation, mesh extraction, and merging of segmented regions into a unified watertight mesh.

3D-CT scans were resampled to 512x512x512 resolution with 1.0mm isotropic voxels. Using Siddon Ray Tracing [33, 38], 512x512 DRR images were generated and enhanced with CLAHE for improved contrast and detail.

4.2 Bone Localization and Detection

Anterior-Posterior (AP) DRR images at 512x512 resolution are annotated using Roboflow [39]. Initial manual annotations establish high-quality labels, followed by the use of Roboflow’s autolabeler for efficient dataset annotation. Automatically generated labels are reviewed and corrected to ensure precision. The final dataset is utilized to train the YOLOv9 model, achieving robust and accurate bone detection.

4.3 Occupancy Value Calculation

The trimesh Python library ensures watertight meshes for accurate occupancy value calculations [8]. All meshes are centered and normalized within a unit bounding box. Following Occupancy Networks (ONet) [8], 32^3 voxel grids are generated. Voxels intersecting the mesh surface are labeled as occupied, and 100,000 random points are analyzed for mesh inclusion using ray-intersection counting. During training of X2B, a subset of 2,048 points is randomly selected to calculate occupancy probabilities, enabling precise 3D reconstructions [40].

5 Experiments

This section presents the simulation results and comparative analysis of the X2B, and X2BR networks. It provides a comprehensive explanation of the evaluation metrics used and discusses the results in detail. The 3Dr performance of X2BR is compared against three recent single image 2D/3Dr models from the literature: X2V [26], D2IM-Net [12] and ED2IF2-Net [13]. The section includes an in-depth discussion of the evaluation metrics and a thorough analysis of the findings.

5.1 Comparative Studies

This section details the X2V, D2IM-Net and ED2IF2-Net models used for performance comparison.

5.1.1 X2V

X2V [26] reconstructs 3D organ volumes from a single X-ray using an implicit occupancy representation, eliminating the need for templates. Unlike D2IM-Net and ED2IF2-Net, it leverages a Vision Transformer (ViT) encoder for enhanced feature extraction. X2V [26] achieves superior performance for X-ray to 3D lung volume reconstruction.

5.1.2 D2IM-Net

D2IM-Net [12] is a single-view 3D reconstruction network that combines a coarse implicit field with displacement maps for front and back surfaces to recover topological structures and fine details. Using two decoders to extract global and local features, it achieves high reconstruction quality with metrics like Chamfer Distance (CD), IoU, and Edge Chamfer Distance (ECD). Its Laplacian loss improves surface detail representation [9], outperforming DISN.

5.1.3 ED2IF2-Net

ED2IF2-Net [13] leverages Pyramid ViT (PVT) for high-fidelity 3D reconstruction from single RGB images. It disentangles the implicit field into a deformed implicit field for topology and a displacement field for surface details. The architecture features three decoders: a coarse shape decoder, a deformation decoder using implicit field deformation blocks (IFDBs), and a surface detail decoder with hybrid attention modules (HAMS). A comprehensive loss function ensures robust learning of coarse and detailed features.

5.2 Evaluation Metrics

The proposed method and baselines are evaluated using Intersection over Union (IoU), Chamfer-L₁ (C-L₁) distance, F-score, mean absolute z -axis error (maze), and normal consistency (NC), following the metrics [26] established in X2V.

Chamfer-L₁ (C-L₁) measures the average nearest-point Euclidean distance between predicted and ground truth (GT) meshes. Voxelized IoU quantifies mesh overlap as the intersection-to-union ratio, using a voxelization (with voxel size = 0.5) for the generated meshes. NC evaluates surface normal alignment via the absolute dot product of neighboring points. F-score is the harmonic mean of precision and recall at distance threshold $t = 0.02$. Mean Absolute z -axis Error (maze) quantifies alignment deviations in ICP-aligned meshes, with maxe and maye extending this evaluation to x- and y-axes, respectively [17]. These metrics comprehensively assess reconstruction accuracy and precision [26].

6 Simulations and Results

6.1 X2B Performance Analysis

Figure 4 illustrates X2B reconstructions compared to GT meshes, showing DRR inputs, front and back views of GT and reconstructed meshes, and heatmaps highlighting reconstruction errors. The mean error is 2.192 mm, with a standard deviation of 1.886 mm, reflecting the model’s alignment accuracy and variability. Positive errors in X2B heatmaps result from over-smoothing in occupancy networks, limited training data resolution, and interpolation effects, which bias reconstructions toward overestimated boundaries.

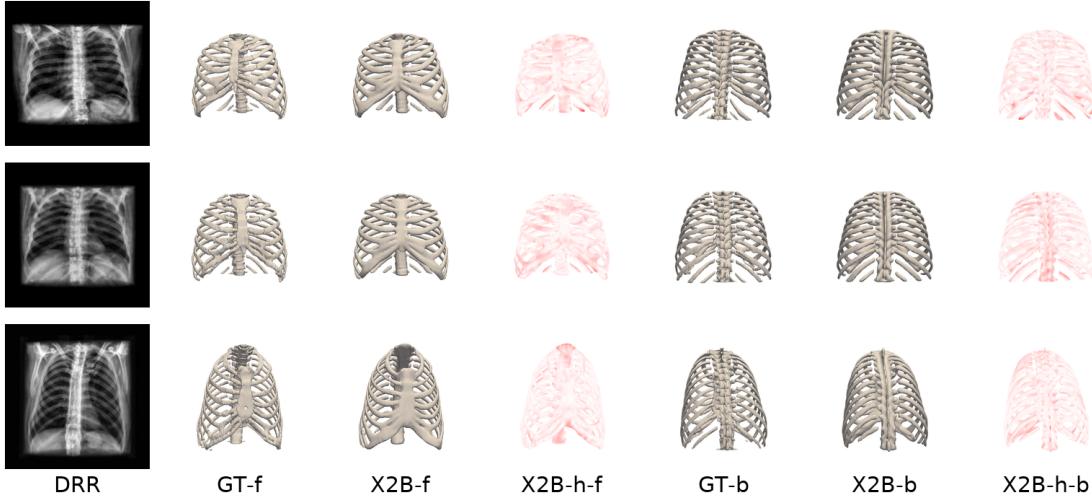


Figure 4: Comparison of X2B and GT. The first column displays the DRR inputs, while the second column presents the GT meshes. The subsequent columns show the X2B reconstructions (X2B-f), X2B heatmaps (X2B-h-f), GT-b, X2B-b, and X2B heatmaps in the back view (X2B-h-b).

6.2 X2B & X2BR Comparison with Existing Methods

Figure 5 compares 3Dr results from D2IM-Net, ED2IF2-Net, X2V [26], X2B, and X2BR, all trained on the same dataset using DRRs as inputs for a fair evaluation. X2B and X2BR demonstrate superior performance, with X2BR excelling in preserving fine anatomical features, such as rib curvature and spacing, making it particularly suitable for biomedical applications. In contrast, X2V can reconstruct overall organ topology but lacks fine structural details due to the limitations of its implicit representation. D2IM and ED2IF2 struggle with depth variations, overlapping structures, and complex backgrounds, leading to lower reconstruction accuracy.

Table 1 evaluates reconstruction accuracy based on IoU, Chamfer-L1 distance, F-score, and NC. X2B outperforms all methods, achieving the highest IoU, lowest Chamfer-L1 distance, and highest F-score. X2BR performs competitively, with improved NC.

Table 2 compares D2IM, ED2IF2, X2V [26], X2B, and X2BR using maxe, maye, and maze metrics. While X2B achieves the highest numerical accuracy, X2BR provides the most visually faithful reconstructions, especially in capturing fine anatomical details such as rib curvature, spacing, and vertebral alignment, as illustrated in Fig. 5. The slightly higher metric errors of X2BR are primarily due to localized deformations introduced during non-rigid registration, which enhance

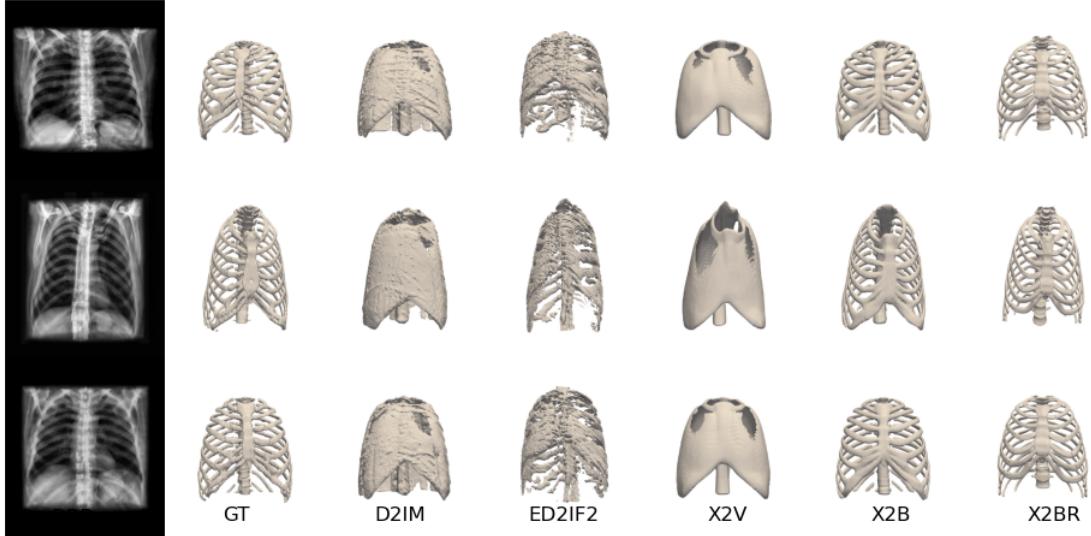


Figure 5: Comparison of reconstruction results across different methods. The figure presents DRR images (leftmost column) and their corresponding GT 3D meshes alongside reconstructed outputs from various methods: D2IM, ED2IF2, X2V, X2B, and X2BR. Each row corresponds to a different DRR input, while the columns illustrate the progression of reconstruction quality across the methods.

Table 1: Comparison of D2IM, ED2IF2, X2V, X2B, X2BR in terms of IoU, Chamfer-L1, F-score and NC metrics for mesh reconstruction accuracy.

Method	<i>IoU</i>	<i>Chamfer-L1</i>	<i>F-score</i>	<i>NC</i>
<i>D2IM</i>	0.903 ± 0.028	0.009 ± 0.001	0.912 ± 0.027	0.499 ± 0.006
<i>ED2IF2</i>	0.573 ± 0.041	0.019 ± 0.002	0.567 ± 0.048	0.495 ± 0.003
<i>X2V</i>	0.859 ± 0.057	0.009 ± 0.001	0.904 ± 0.050	0.496 ± 0.005
<i>X2B</i>	0.952 ± 0.038	0.005 ± 0.001	0.974 ± 0.038	0.505 ± 0.003
<i>X2BR</i>	0.875 ± 0.036	0.009 ± 0.001	0.913 ± 0.039	0.504 ± 0.003

anatomical realism but may slightly misalign with the ground truth mesh in a global coordinate system.

This trade-off reveals a crucial limitation of purely numerical metrics, which can underestimate perceptual or anatomical quality in cases involving fine-grained, patient-specific variations. The template-guided refinement in X2BR enables it to better model true anatomical structure, even if global alignment metrics slightly worsen. Without this qualitative improvement, X2BR’s contribution would be redundant; however, visual comparisons clearly demonstrate its added value in clinical interpretability and anatomical plausibility, which are critical for real-world applications such as surgical planning and biomechanical simulations.

6.3 Comparison of Registration Techniques

Table 3 and Figure 6 evaluate the effectiveness of different non-rigid registration techniques in aligning reconstructed thorax models. The objective is to assess how well each method preserves anatomical structures while handling global and local deformations.

GBCPD++ achieves the best metrics, including IoU, Chamfer-L1 distance, and F-score, demon-

Table 2: Comparison of D2IM, ED2IF2, X2V, X2B, and X2BR, in terms of maze, maxe and maye metrics in milimeter for mesh reconstruction accuracy.

Method	<i>maxe</i>	<i>maye</i>	<i>maze</i>
<i>D2IM</i>	3.468 ± 3.157	2.772 ± 2.474	3.304 ± 3.095
<i>ED2IF2</i>	11.871 ± 10.267	6.254 ± 6.655	7.114 ± 6.794
<i>X2V</i>	4.545 ± 5.996	4.706 ± 9.352	4.442 ± 6.698
<i>X2B</i>	2.821 ± 2.717	2.515 ± 2.320	2.590 ± 2.443
<i>X2BR</i>	3.562 ± 3.909	3.667 ± 4.203	3.003 ± 2.793

Table 3: Comparison of GBCPD++, NDP, CPD in terms of IoU, Chamfer-L1, F-score and NC metrics

Method	<i>IoU</i>	<i>Chamfer-L1</i>	<i>F-score</i>	<i>NC</i>
<i>NDP</i>	0.664 ± 0.111	0.008 ± 0.001	0.637 ± 0.124	0.498 ± 0.003
<i>CPD</i>	0.744 ± 0.049	0.014 ± 0.001	0.748 ± 0.051	0.498 ± 0.003
<i>GBCPD</i>	0.900 ± 0.019	0.010 ± 0.001	0.931 ± 0.019	0.500 ± 0.004

strating superior alignment and structural preservation. CPD shows moderate performance, while NDP exhibits significant misalignments, particularly in rib spacing and structural integrity. These results highlight GBCPD++ as the most effective method for accurate non-rigid registration.

6.4 Single vs. Double DRRs as Model Input

This section examines the impact of using single versus bi-planar DRRs on 3Dr accuracy. To address the limitations of a single-view approach, X2B2 extends X2B by integrating two orthogonal DRRs and employing a multi-head cross-attention mechanism for enhanced spatial feature fusion.

Table 4: Comparison of X2B and X2B2 results in terms of IoU, Chamfer-L1, F-score and NC metrics

Method	<i>IoU</i>	<i>Chamfer-L1</i>	<i>F-score(t=0.02)</i>	<i>NC</i>
<i>X2B</i>	0.952 ± 0.038	0.005 ± 0.001	0.974 ± 0.038	0.505 ± 0.003
<i>X2B2</i>	0.964 ± 0.020	0.004 ± 0.001	0.984 ± 0.017	0.505 ± 0.003

Tables 4 and 5 compare the performance of X2B and X2B2 in 3Dr tasks. Both models perform exceptionally well, but X2B2 demonstrates slight numerical advantages, achieving higher IoU and lower Chamfer-L1 distance , indicating improved geometric accuracy. X2B2 also attains higher F-score, showcasing its robustness in capturing fine structural details, while both models achieve identical NC scores.

Table 5 highlights X2B2’s improved reconstruction accuracy over X2B in maxe, maye, and maze, showing lower mean errors across all axes. These results reflect X2B2’s ability to reduce positional errors through bi-planar input processing and multi-view data fusion.

Figure 7 compares thorax reconstructions for the GT, X2B, and X2B2 models. The zoomed-in regions emphasize X2B2’s advantage in capturing finer anatomical details, illustrating its incremental improvement over X2B in reconstructing complex thorax structures.

X2B2 shows a slight numerical advantage over X2B, particularly in preserving fine anatomical details in 3Drs. By leveraging bi-planar DRR inputs and cross-attention, it enhances spatial feature

Table 5: Comparison of X2B and X2B2 results in terms of maze, maxe, and maye metrics in millimeters

Method	<i>maxe</i>	<i>maye</i>	<i>maze</i>
X2B	2.821 ± 2.717	2.515 ± 2.320	2.590 ± 2.443
X2B2	2.473 ± 2.143	2.431 ± 2.212	2.441 ± 2.247

fusion, improving reconstruction accuracy. These refinements make X2B2 well-suited for medical imaging and multi-perspective 3D reconstruction while maintaining X2B’s efficiency.

7 Conclusion

This study presents X2B and X2BR, two complementary neural implicit frameworks for high-fidelity 3D skeletal reconstruction from a single planar X-ray. X2B leverages a ConvNeXt-based encoder and Conditional Batch Normalization (CBN) layers to predict continuous occupancy fields, enabling template-free reconstruction of complex anatomical structures such as ribs and vertebrae. X2BR builds upon this by integrating a biomechanical template and applying non-rigid registration through GBCPD++, enhancing anatomical plausibility and improving alignment to patient-specific skeletal morphology.

Both frameworks capitalize on the strengths of neural implicit representations—modeling continuous volumetric structures without reliance on voxel grids—while addressing key challenges in X-ray-based reconstruction, including occlusion, overlapping intensities, and incomplete input data. Evaluations on clinical datasets demonstrate that X2B achieves state-of-the-art accuracy in metrics such as volumetric IoU, Chamfer-L1 distance, and F-score, while X2BR offers improved anatomical consistency through template-guided refinement.

In addition to methodological advances, this work introduces the largest known dataset of paired 3D bone meshes and corresponding digitally reconstructed radiographs (DRRs), contributing a valuable benchmark for future research. By addressing long-standing limitations in 3D reconstruction from sparse imaging data, X2B and X2BR offer practical tools for surgical planning, orthopedic assessment, and personalized biomechanical simulations.

References

- [1] Y. Liang, W. Song, J. Yang, L. Qiu, K. Wang, L. He, “X2Teeth: 3D teeth reconstruction from a single panoramic radiograph,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI International Conference Proceedings*, pp. 400–409. Springer International Publishing, Oct. 2020.
- [2] W. Song, H. Zheng, D. Tu, C. Liang, L. He, “Oral-3Dv2: 3D Oral Reconstruction from Panoramic X-Ray Imaging with Implicit Neural Representation,” arXiv preprint arXiv:2303.12123, Mar. 2023.
- [3] Y. Yang, Z. Cui, C. Li, W. Wang, “ToothInpaintor: Tooth Inpainting from Partial 3D Dental Model and 2D Panoramic Image,” arXiv preprint arXiv:2211.15502, 2022.
- [4] A. Corona-Figueroa, J. Frawley, S. Bond-Taylor, S. Bethapudi, H. P. Shum, C. G. Willcocks, “Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a

- single x-ray,” In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3843–3848, Jul. 2022.
- [5] P. H. Yeung, L. Hesse, M. Aliasi, M. Haak, W. Xie, A. I. Namburete, “Implicitvol: Sensorless 3d ultrasound reconstruction with deep implicit representation,” arXiv preprint arXiv:2109.12108, 2021.
 - [6] Y. Cai, J. Wang, A. Yuille, Z. Zho, A. Wang, “Structure-aware sparse-view x-ray 3d reconstruction,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11174–11183, 2024.
 - [7] Y. Fang *et al.*, “Snaf: Sparse-view cbct reconstruction with neural attenuation fields,” arXiv preprint arXiv:2211.17048, 2022.
 - [8] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, A. Geiger, “Occupancy networks: Learning 3d reconstruction in function space,” In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4460–4470, 2019.
 - [9] Q. Xu, W. Wang, D. Ceylan, R. Mech, U. Neumann, “Disn: Deep implicit surface network for high-quality single-view 3d reconstruction,” *Advances in neural information processing systems*, 2019.
 - [10] Y. Wang, Y. Zhuang, Y. Liu, B. Chenm, “MDISN: Learning multiscale deformed implicit fields from single images,” *Visual Informatics*, vol. 6, no. 2, pp. 41–49, 2022.
 - [11] W. Bian, Z. Wang, K. Li, V. A. Prisacariu, “Ray-ONet: efficient 3D reconstruction from a single RGB image,” arXiv preprint arXiv:2107.01899, Jul. 2021.
 - [12] M. Li, H. Zhang, “D2im-net: Learning detail disentangled implicit fields from single images,” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10246–10255, 2021.
 - [13] X. Zhu *et al.*, “ED²IF²-Net: Learning Disentangled Deformed Implicit Fields and Enhanced Displacement Fields from Single Images Using Pyramid Vision Transformer,” *Applied Sciences*, vol. 13, no. 13, 7577, 2023.
 - [14] R. Chen, Y. Yang, C. Tong, “G2IFu: Graph-based implicit function for single-view 3D reconstruction,” *Engineering Applications of Artificial Intelligence*, vol. 124, 106493, Sep. 2023.
 - [15] M. S. Arshad, W. J. Beksi, “LIST: Learning Implicitly from Spatial Transformers for Single-View 3D Reconstruction,” In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9321–9330, 2023.
 - [16] J. F. Hollenbeck, C. M. Cain, J. A. Fattor, P. J. Rullkoetter, P. J. Laz, “Statistical shape modeling characterizes three-dimensional shape and alignment variability in the lumbar spine,” *Journal of biomechanics*, vol. 69, pp. 146–155, 2018.
 - [17] P. R. Furrer, S. Caprara, F. Wanivenhaus, M. D. Burkhard, M. Senteler, M. Farshad, “Patient-specific statistical shape modeling for optimal spinal sagittal alignment in lumbar spinal fusion,” *European Spine Journal*, vol. 30, pp. 2333–2341, 2021.
 - [18] B. Aubert, C. Vazquez, T. Cresson, S. Parent, J. A. de Guise, “Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting,” *IEEE Trans. on Medical Imaging*, vol. 38, no. 12, pp. 2796–2806, 2019.

- [19] Y. Jiang *et al.*, “2D/3D Shape Model Registration with X-ray Images for Patient-Specific Spine Geometry Reconstruction,” *International Conference on Intelligent Robotics and Applications*, pp. 558–570, Jul. 2023.
- [20] S. Jecklin, C. Jancik, M. Farshad, P. Fürnstahl, H. Esfandiari, “X23D—Intraoperative 3D lumbar spine shape reconstruction based on sparse multi-view X-ray data,” *Journal of Imaging*, vol. 8, no. 10, pp. 271, 2022.
- [21] Z. Chen, L. Guo, R. Zhang, Z. Fang, X. He, J. Wang, “BX2S-Net: Learning to reconstruct 3D spinal structures from bi-planar X-ray images,” *Computers in Biology and Medicine*, vol. 154, pp. 106615, 2023.
- [22] C. J. Yang *et al.*, “Generative Adversarial Network (GAN) for Automatic Reconstruction of the 3D Spine Structure by Using Simulated Bi-Planar X-ray Images,” *Diagnostics*, vol. 12 no. 5, pp. 1121, 2022.
- [23] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, Y. Zheng, “X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks,” *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10619–10628, 2019.
- [24] H. G. Ha, J. Lee, G. H. Jung, J. Hong, H. Lee, “2D-3D Reconstruction of a Femur by Single X-Ray Image Based on Deep Transfer Learning Network,” *IRBM*, vol. 45, no. 1, pp. 100822, 2024.
- [25] D. Buttongkum *et al.*, “3D reconstruction of proximal femoral fracture from biplanar radiographs with fractural representative learning,” *Scientific Reports*, vol. 13, no. 1, pp. 455, 2023.
- [26] G. Guven, H. F. Ates and H. F. Ugurdag, “X2V: 3D Organ Volume Reconstruction From a Planar X-Ray Image With Neural Implicit Methods,” in *IEEE Access*, vol. 12, pp. 50898–50910, 2024.
- [27] O. Hirose, “Geodesic-Based Bayesian Coherent Point Drift,” in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5816–5832, 2022.
- [28] M. Keller *et al.*, “From skin to skeleton: Towards biomechanically accurate 3d digital humans,” *ACM Transactions on Graphics (TOG)*, vol. 42, no. 6, pp. 1–12, 2023.
- [29] S. L. Delp *et al.*, “OpenSim: open-source software to create and analyze dynamic simulations of movement,” in *IEEE transactions on biomedical engineering*, vol. 54, no. 11, pp. 1940–1950, 2007.
- [30] Z. Liu, H. Mao, C. Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie, “A convnet for the 2020s,” *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.
- [31] J. Wasserthal *et al.*, “TotalSegmentator: robust segmentation of 104 anatomic structures in CT images,” *Radiology: Artificial Intelligence*, vol. 5, no. 5, 2023.
- [32] N. L. S. T. R. Team, “Data from the national lung screening trial (nlst),” *The Cancer Imaging Archive*, vol. 10, 2013.
- [33] R. L. Siddon, “Fast calculation of the exact radiological path for a three-dimensional CT array,” *Medical physics*, vol. 12, no. 2, pp. 252–255, 1985.

- [34] R. Kikinis, S. D. Pieper, K. G. Vosburgh, “3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support,” in *Intraoperative imaging and image-guided therapy*, pp. 277–289, NY: Springer New York, 2013.
- [35] T. Kapur *et al.*, “Increasing the impact of medical image computing using community-based open-access hackathons: The NA-MIC and 3D Slicer experience,” *Medical Image Analysis*, vol. 33, pp. 176–180, 2016.
- [36] A. Fedorov *et al.*, “3D Slicer as an image computing platform for the Quantitative Imaging Network,” *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.
- [37] S. Pieper, B. Lorensen, W. Schroeder, R. Kikinis, “The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D slicer as an open platform for the medical image computing community,” In *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pp. 698–701, 2006.
- [38] M. De Greef, J. Crezee, J. C. Van Eijk, R. Pool, A. Bel, “Accelerated ray tracing for radiotherapy dose calculations on a GPU,” *Medical physics*, vol. 36, no. 9, pp. 4095–4102, Sep. 2009.
- [39] S. Alexandrova, Z. Tatlock, M. Cakmak, “RoboFlow: A flow-based visual programming language for mobile manipulation tasks,” In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5537–5544, May 2015.
- [40] C. B. Choy, D. Xu, J. Gwak, K. Chen, S. Savarese, “3d-r2n2: A unified approach for single and multi-view 3d object reconstruction,“. In *Computer Vision–ECCV: 14th European Conference, Proceedings*, pp. 628–644, Springer International Publishing, 2016.

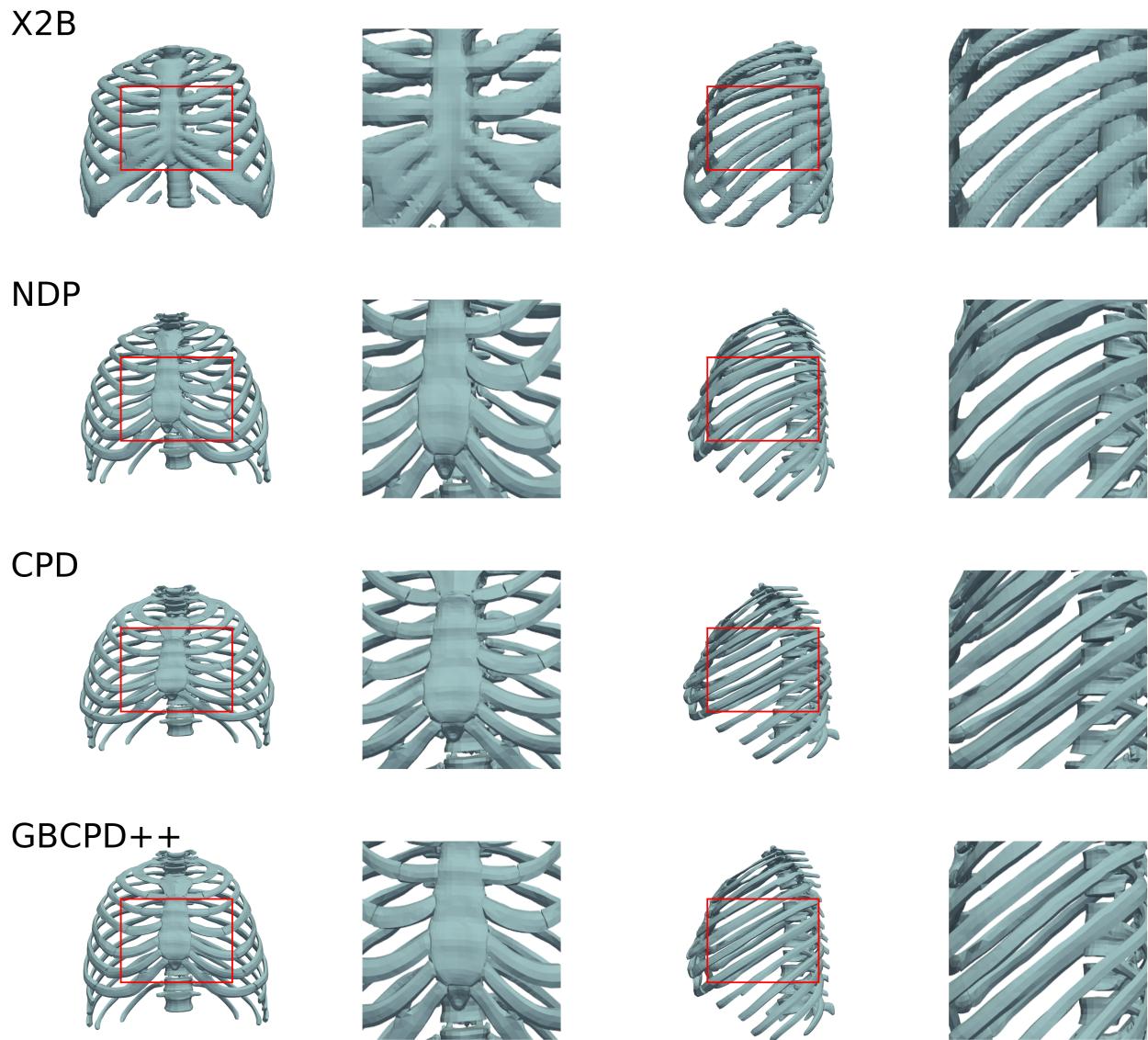


Figure 6: Thorax reconstructions using X2B, NPD, CPD, and GBCPD++. Each row shows full views (front and side) and corresponding zoom-ins highlighting fine anatomical details.

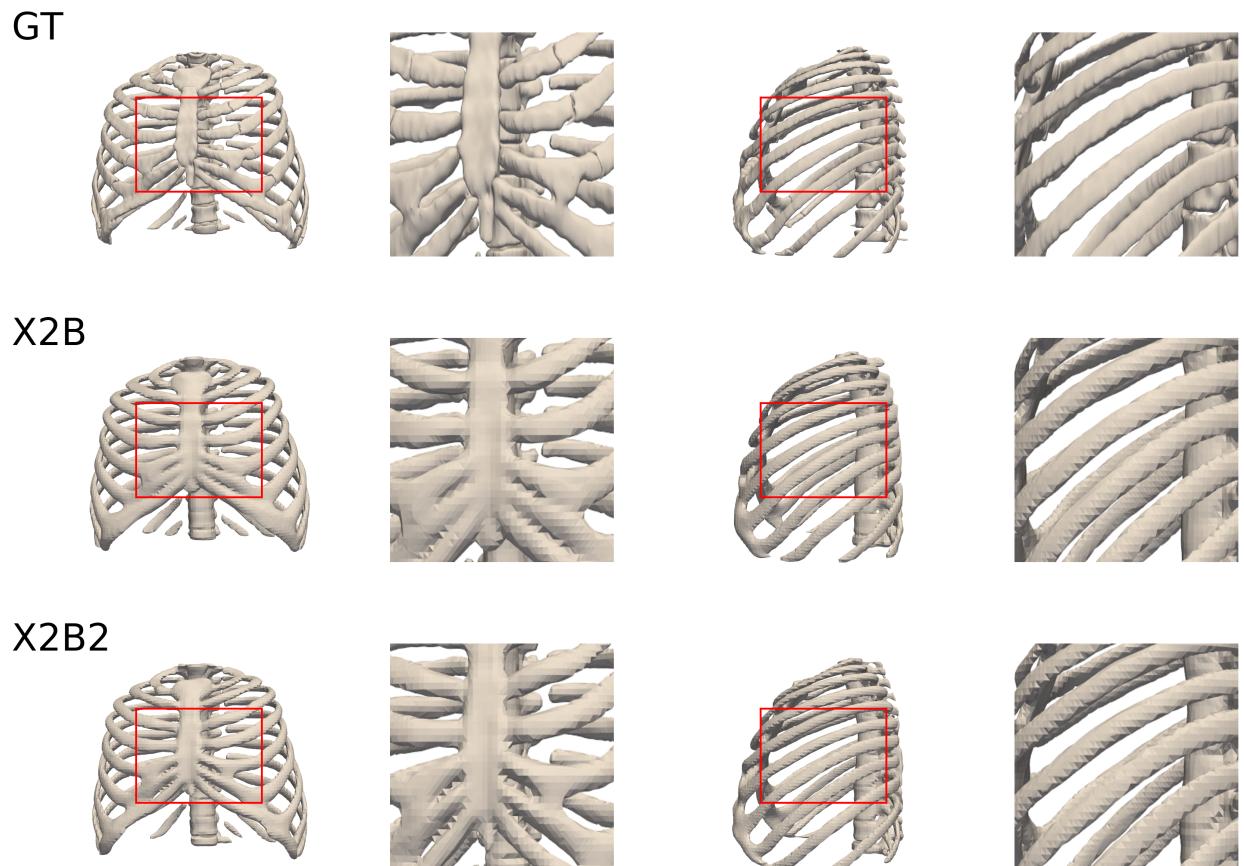


Figure 7: Thorax reconstruction results for GT, X2B, and X2B2. Each row shows full thorax views (front and side) and corresponding zoom-ins to highlight anatomical details.