Spatial-Division Augmented Occupancy Field for Bone Shape Reconstruction from Biplanar X-Rays

Jixiang Chen[1][0000−0001−9941−8324], Yiqun Lin[1][0000−0002−7697−0842], Haoran Sun[3][0009−0007−7390−2733], and Xiaomeng Li[1,2(✉)][0000−0003−1105−8083]

[1] The Hong Kong University of Science and Technology
eexmli@ust.hk
[2] HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen
[3] Koln 3D Technology (Medical) Limited Company, Science Park, Hong Kong

Abstract. Retrieving 3D bone anatomy from biplanar X-ray images is crucial since it can significantly reduce radiation exposure compared to traditional CT-based methods. Although various deep learning models have been proposed to address this complex task, they suffer from two limitations: 1) They employ voxel representation for bone shape and exploit 3D convolutional layers to capture anatomy prior, which are memory-intensive and limit the reconstruction resolution. 2) They overlook the prevalent occlusion effect within X-ray images and directly extract features using a simple loss, which struggles to fully exploit complex X-ray information. To tackle these concerns, we present Spatial-division Augmented Occupancy Field (SdAOF). SdAOF adopts the continuous occupancy field for shape representation, reformulating the reconstruction problem as a per-point occupancy value prediction task. Its implicit and continuous nature enables memory-efficient training and fine-scale surface reconstruction at different resolutions during the inference. Moreover, we propose a novel spatial-division augmented distillation strategy to provide feature-level guidance for capturing the occlusion relationship. Extensive experiments on the pelvis reconstruction dataset show that SdAOF outperforms state-of-the-art methods and reconstructs fine-scale bone surfaces. The code is available at https://github.com/xmed-lab/SdAOF.

Keywords: Bone Reconstruction · Biplanar X-Rays · Occupancy Field · Implicit Neural Representation

1 Introduction

3D reconstruction of patient-specific bone anatomy is a crucial task for various clinical applications, such as biomechanical load analysis, surgical planning, and customized implant design optimization [22,26,9]. The current gold standard

reconstruction of bones primarily involves segmentation from computed tomography (CT) images. However, CT acquisition involves relatively high ionizing radiation and is often unaffordable for many health centers [25]. In contrast, X-ray images are more accessible due to lower costs and reduced scanning time. Therefore, there is a significant need for the direct reconstruction of bone anatomy from a limited number of X-ray images. In this paper, we study the problem of bone mesh reconstruction from biplanar X-ray images.

Bone reconstruction from biplanar X-ray images tackles the challenging task of extracting 3D information from only two projection views, typically along the anterior-posterior (AP) and lateral view directions. Effectively modeling anatomical information is crucial for solving this ill-posed problem. Traditional approaches use statistical shape models (SSM) or statistical shape and intensity models (SSIM) to encode shape prior, but they suffer from cumbersome pre-modeling steps and sensitivity to model initialization [23,10,27,8]. With recent advances in deep learning techniques, several learning-based approaches are proposed [1,2,4,5,7,25,6,15]. These methods primarily utilize 3D convolutional layers to generate volumetric shape model, while adopting varied fusion strategy to address the dimensionality misalignment between 2D input and 3D output. However, the memory-intensive nature of voxel representation limits the reconstruction resolution and hinders the extraction of fine-scale bone meshes. An alternative approach for direct mesh reconstruction is to employ continuous implicit functions to represent the surface. This representation showcases superior efficiency for both model training and inference in different applications [24,16,18,17]. Inspired by this, we introduce the Occupancy Fields (OF) representation [13]. This function indicates whether a 3D point lies inside the surface and defines the surface as an iso-surface of the continuous function. Based on this representation, we reformulate the reconstruction problem as the occupancy value prediction task for any 3D point within the target reconstruction space. This field can be captured using an OF Network [24], which leverages pixel-aligned features from view-specific 2D feature planes and predicts per-point value through multi-layer perceptrons (MLP). This formulation facilitates fine-scale reconstruction in two respects: 1) The implicit representation does not require explicit storage, which is memory-efficient. 2) Its continuous nature allows us to scale into different resolutions during inference without retraining the model.

Despite leveraging the OF representation facilitates fine-scale mesh reconstruction, directly extracting features from input X-rays struggles to fully exploit information within complex X-ray images. The inherent complexity arises from the overlapped information in X-ray images, primarily due to common occlusion scenarios where the target bone is obscured by other tissues or its frontal parts. This occlusion stems from sequential attenuation along the view direction during the X-ray imaging process. While the occluded region of bones may vary due to different patient conditions, the anterior-posterior arrangement of organs follows a prior pattern when viewed from a fixed perspective. Therefore, we argue that it is possible to identify the occlusion pattern and further enhance the

reconstruction quality. To achieve this, we propose a spatial-division augmented distillation strategy. It divides the target reconstruction space along each view direction into several subspaces and uses the Digitally Reconstructed Radiography (DRR) technique to render an augmented X-ray image for each subspace. A teacher OF network is trained to extract subspace-specific features and capture occlusion relationships across subspaces from these X-ray images. These features are then employed to provide feature-level guidance, enabling the identification of occlusion patterns in the original input X-rays through knowledge distillation.

To this end, we present Spatial-division Augmented Occupancy Field (SdAOF), which integrates continuous implicit function representation and spatial-division augmented distillation strategy. In contrast to voxel-representation-based methods, SdAOF demonstrates memory efficiency during training and is scalable to different resolutions during inference, achieving state-of-the-art fine-scale bone shape reconstruction results. In summary, our main contributions include

1. To the best of our knowledge, we are the first to introduce the continuous occupancy field representation for biplanar bone shape reconstruction.
2. We argue that identifying the occlusion relationship within X-ray images is crucial for robust feature extraction. To this end, we propose a spatial-division augmented distillation strategy to capture this information.
3. We propose a novel framework SdAOF, which reconstructs fine-scale bone surfaces and is scalable to different resolutions without retraining.

2 Methodology

2.1 Occupancy Field Network

We use the continuous occupancy field [24,20] representation, where the occupancy value $o(p)$ of a 3D point $p \in \mathbb{R}^3$ indicates whether it lies inside the surface (1) or not (0). The complete set of continuous occupancy values across all points constitutes the OF. We can generate the mesh surface by densely sampling the field over the space and extracting the iso-surface of $o(p) = 0.5$ using the Marching Cube algorithm [19]. Therefore, the surface reconstruction is reformulated as per-point occupancy value prediction.

To this end, we introduce the OF network for bone reconstruction from multi-view X-rays, consisting of a feature extractor $\mathcal{F}$ and multi-layer perceptrons (MLP) for occupancy value prediction. Specifically, $\mathcal{F}$ extracts view-specific 2D features $\{F_i\}_{i=1}^{M} \subset \mathbb{R}^{C \times H \times W}$ for M input views (particularly $M = 2$ for biplanar reconstruction). Each 2D feature plane is regarded to be located at the X-ray detector plane of each respective view. For a 3D point p, we can obtain its pixel-aligned feature by first finding the projected location on each plane, and then querying the feature via interpolation. This is formulated as

$$f_i(p) = \text{Interp}(F_i, \pi(p)), \tag{1}$$

where $f_i(p) \in \mathbb{R}^C$ denotes the point-specific feature of the i-th view, $\text{Interp}(\cdot)$ denotes the bilinear interpolation operation, and $\pi(\cdot)$ refers to the projection
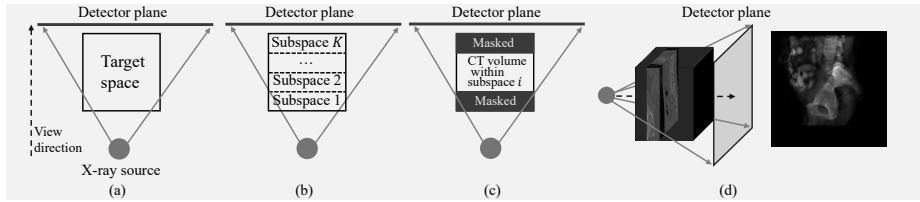
Fig. 1: Illustration of spatial-division augmentation for a single view. (a)-(c): Planform of X-ray cone-beam scanning. (b): The target reconstruction space is divided along the view-depth direction into distinct subspaces. (c)-(d): For each subspace, we mask the regions of the original CT volume that lie outside the subspace and use DRR to generate an augmented X-ray image.

operation. Multi-view per-point features are then concatenated and sent into MLP for occupancy value prediction, as formulated by

$$\hat{o}(p) = \text{MLP}\left(\text{Cat}(f_1(p), \ldots, f_M(p))\right). \tag{2}$$

The OF network uses mean-square-error (MSE) to calculate the per-point difference as the reconstruction loss

$$\mathcal{L}_{\text{recon}} = \frac{1}{N} \sum_{j=1}^{N} \left(\hat{o}(p_j) - o(p_j)\right)^2, \tag{3}$$

where N denotes the number of points used for training and $o(p_i)$ denotes the ground-truth occupancy value.

2.2   Spatial-division Augmentation

Identifying the view-specific occlusion pattern helps to extract robust features related to bone anatomy. Although the occluded region of bones may vary due to different patient conditions, the anterior-posterior arrangement of organs follows a prior pattern when viewed from a fixed perspective. Therefore, we propose spatial-division augmentation to separate this arrangement information along view-depth dimensions and render augmented X-ray images for each input view.

As depicted in Fig 1, we divide the target 3D space along the view direction into a set of K subspaces for each view. Subsequently, within each subspace, we utilize the corresponding CT image by masking information outside of it and use DRR to generate K augmented X-ray images, denoted by $\{X_k^{\text{aug}}\}_{k=1}^{K}$. These X-ray images serve as the inputs of our teacher OF network.

2.3   Teacher Occupancy Field Network

Since the OF network processes each input view identically except for the MLP prediction, in the subsequent illustrations, we explain the process for a single
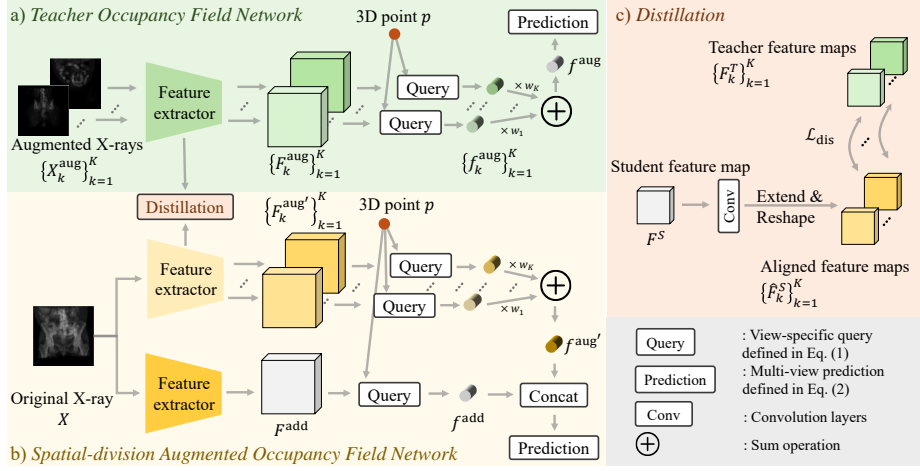
Fig. 2: Overall framework of SdAOF. Both input views undergo the same processing, and we demonstrate the procedure using one input view as an example.

view input and omit the view index for brevity. For each view, the teacher OF network extracts K 2D feature maps $\{F_k^{\mathrm{aug}}\}_{k=1}^K$ from $\{X_k^{\mathrm{aug}}\}_{k=1}^K$, where each feature map solely reflects information within the corresponding subspace. Following this, the view-specific feature of p consists of K folds $\{f_k^{\mathrm{aug}}(p)\}_{k=1}^K$. To capture the anterior-posterior arrangement along the view direction, we fuse these features based on the view-depth relationship. For this purpose, we consider each feature plane to be located in the middle of its subspace and perform a weighted sum to aggregate features, as formulated below:

$$f^{\mathrm{aug}}(p) = \sum_{k=1}^K w_k \cdot f_k^{\mathrm{aug}}(p), \text{ and } w_k = \frac{1/|d - d_k|}{\sum_{j=1}^K (1/|d - d_j|)}, \tag{4}$$

where d_k denotes the view-depth distance of the k-th feature plane and d refers to the view-depth distance of the projected point p. Intuitively, Eqn. (4) assigns a larger weight to features that are closer to the point along the view-depth direction. Therefore, a point should identify its occlusion relationship based on feature planes closer to itself rather than the ones further away. The multi-view $f^{\mathrm{aug}}(p)$ is then used as the input of MLP for reconstruction using $\mathcal{L}_{\mathrm{Recon}}$.

2.4 Spatial-division Augmented Occupancy Field Network

SdAOF aims to identify the occlusion patterns under the teacher OF network's guidance from original X-ray inputs. As depicted in Fig. 2, SdAOF employs two feature extractors, referred to as the student extractor $\mathcal{F}^S(\cdot)$ and the additional extractor $\mathcal{F}^{\mathrm{add}}(\cdot)$. They each extract a 2D feature map from the original X-ray images X, denoted by F^S and F^{add}, respectively. For each view, $\mathcal{F}^S(\cdot)$ performs

feature-level distillation from the teacher OF network. The subspace-specific teacher features $\{F_k^T\}_{k=1}^K$ consist of K feature maps, which are extracted from $\{X_k^{\text{aug}}\}_{k=1}^K$. In contrast, the student feature map F^S solely includes one fold. To facilitate subspace-aligned knowledge transfer, we employ additional convolution layers to increase the channel of F^S by K times $((c, w, h) \to (Kc, w, h))$, following by reshaping it for alignment $((Kc, w, h) \to (K, c, w, h))$. SdAOF then performs feature-level distillation as follows

$$\mathcal{L}_{\text{dis}} = \frac{1}{K} \sum_{k=1}^{K} \left\| F_k^T - \hat{F}_k^S \right\|^2, \tag{5}$$

where $\hat{F}^S$ refers to K extended student features and $\| \cdot \|$ indicates L_2 Norm for high-dimensional tensors. These features undergo the same fusion strategy as defined in Eqn. (4), generating $f^{\text{aug}'}(p)$. On the other hand, the extracted F^{add} explores additional information that complements the occlusion-aware ones. Its corresponding point feature $f^{\text{add}}(p)$ is obtained following Eqn. (1). SdAOF then concatenate $f^{\text{aug}'}(p)$ and $f^{\text{add}}(p)$ as the view-specific point feature for MLP inputs. The overall loss for training SdAOF includes the distillation objective and the reconstruction objective

$$\mathcal{L} = \mathcal{L}_{\text{Recon}} + \alpha \cdot \frac{1}{ML} \sum_{i=1}^{M} \sum_{j=1}^{L} \mathcal{L}_{\text{dis}}, \tag{6}$$

where α balances the contribution of two loss terms. In practice, the distillation is performed at L intermediate layers. During inference, only the SdAOF network is used, and the teacher OF network is disregarded.

3   Experiments

3.1   Experimental Settings

Dataset and Preprocessing. We conduct extensive experiments on the Pelvis dataset from the XrayTo3Dbenchmark [25] to validate the effectiveness of our proposed method in reconstructing bone shapes from biplanar X-rays. The original segmentation annotations are utilized to generate fine-scale meshes, serving as ground truth. We specifically exclude some noisy fragment annotations and omit the sacrum part due to insufficient quality in its annotation for fine-scale mesh derivation. The data split follows the original setting, with 321 samples used for training, 57 for validation, and 67 for testing. 2D X-ray images are generated by DRRs using the TIGRE [3] package under the cone-beam geometry. Please refer to supplementary material for implementation details.

Baseline Methods. We select 6 publicly available methods for comparison, including 5 state-of-the-art voxel-based methods: UNet [15], UNetr [12], Swin-UNetr [11], AttentionUNet [21], 3DReconNet [4] and one representative implicit-representation-based method in human-shape reconstruction, PIFu [24]. As PIFu

Table 1: Comparison results on XrayTo3Dbenchmark-Pelvis dataset. Mean and standard deviation results are reported. ∗: Evaluate at 336^3 output resolution.

| Method	Left pelvis		Right pelvis		
	CD (mm) ↓	EMD (mm) ↓	CD (mm) ↓	EMD (mm) ↓	Inference time (s)
Voxel-based methods					
UNet [15]	2.939 ± 0.778	7.203 ± 4.373	2.995 ± 0.623	6.779 ± 3.652	0.004
UNetr [12]	3.194 ± 0.716	8.651 ± 4.309	3.294 ± 0.684	8.934 ± 5.015	0.015
SwinUNetr [11]	2.817 ± 0.578	6.181 ± 2.076	2.887 ± 0.462	6.355 ± 2.212	0.018
3DReconNet [4]	2.628 ± 0.433	5.817 ± 1.856	2.675 ± 0.369	6.085 ± 1.810	0.016
AttentionUNet [21]	2.842 ± 0.562	6.426 ± 2.205	2.949 ± 0.708	6.579 ± 2.631	0.005
Implicit-representation-based methods (**Ours**)					
PIFu [24]	2.586 ± 0.418	5.899 ± 1.623	2.623 ± 0.291	$\mathbf{5.943 \pm 1.529}$	0.033
SdAOF	$\mathbf{2.465 \pm 0.367}$	$\mathbf{5.710 \pm 1.818}$	$\mathbf{2.571 \pm 0.290}$	6.160 ± 1.526	0.033
PIFu∗ [24]	2.494 ± 0.404	4.985 ± 1.601	2.407 ± 0.335	$\mathbf{4.502 \pm 1.385}$	8.822
SdAOF∗	$\mathbf{2.335 \pm 0.359}$	$\mathbf{4.556 \pm 1.692}$	$\mathbf{2.321 \pm 0.329}$	4.612 ± 1.480	11.852
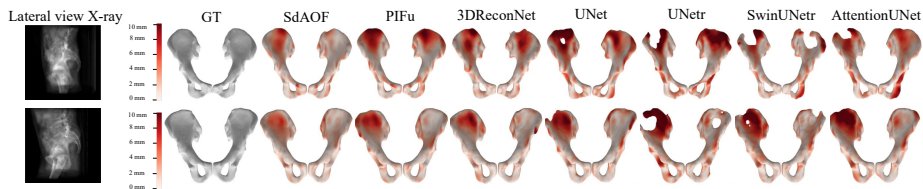
Fig. 3: Quantitative comparison results evaluated at 128^3 output resolution. The color indicates the distance between the reconstructed mesh and the GT mesh. More quantitative results are provided in supplementary material. The corresponding lateral view X-rays show the occlusion pattern.

has not been applied to this task previously, we implemented it ourselves using the original code base. We follow the configuration in [25] to use an input resolution of 128×128 and evaluate with a resolution of 128^3. For voxel-based methods, we derive the occupancy voxels and use the Marching Cube algorithm [19] to extract the mesh surface.

Evaluation Metrics. We follow prior works [24] to use the Chamfer Distance (CD) and Earth Mover's Distance (EMD) to assess reconstructed surface accuracy. Each method undergoes three runs to calculate the metrics. Lower CD/EMD values represent better higher reconstruction quality. Detailed metric explanations are provided in the supplementary material.
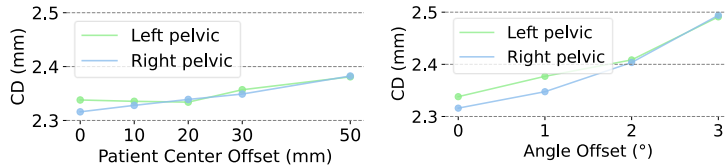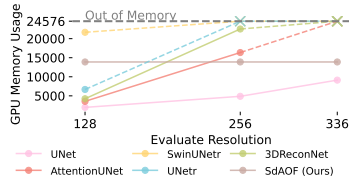
3.2 Results

Comparison with State-of-The-Arts. As indicated in Table 1, our method outperforms voxel-based approaches, achieving a substantial reduction in reconstruction error by 6% and 4% compared to the previously leading method (3DReconNet) for the left and right pelvis, respectively. The left pelvis, positioned distantly from the viewing source (from right to left in the lateral view), faces challenges of occlusion by the right pelvis. SdAOF leverages feature-level

Table 2: Ablation on spatial-division augmented strategy.

Settings			CD (mm)	
Index	Spatial division	Distillation	Left pelvis	Right pelvis
A)	✗	✗	2.586	2.623
B)	✓	✗	2.703	2.699
C)	✗	✓	2.547	2.612
D)	✓	✓	**2.465**	**2.571**

Fig. 4: GPU memory usage comparison when training with a batch size of 1.

Fig. 5: Results on different imaging parameters at 336^3 output resolution.

guidance to identify occlusion patterns, resulting in a notable improvement in left pelvis reconstruction, as illustrated in Fig. 3. Furthermore, methods based on implicit representation offer scalability to different output resolutions, facilitating fine-scale reconstruction. The results evaluated at 336^3 output resolution demonstrate SdAOF's further enhancement in reconstruction accuracy ($2.465 \rightarrow 2.335$ mm of CD on the left pelvis) while maintaining a reasonable reconstruction time (11 seconds per sample). Additionally, the training memory comparison results in Fig. 4 underscore the superiority of SdAOF. Unlike voxel-based methods that commonly struggle to train models at fine-scale resolutions, SdAOF's resolution scalability avoids the memory bottleneck associated with explicit representation.

Ablation Studies. The ablative analysis results of spatial division information and the distillation strategy are presented in Table 2, with the following settings: A): The vanilla PIFu baseline; B): Instead of guiding the model to capture spatial division information via distillation, we train an image-to-image translation network [14] that generates augmented X-ray images from original X-ray images. These predicted images are then used as additional inputs of SdAOF. The performance degrades as the generation network brings more noise than additional subspace-related information in the generated augmented X-rays. C): We train the teacher model by only inputting original X-ray images that are identical to the student's inputs. Then, we perform distillation using these features. The result highlights the crucial role of spatial-division information in improving reconstruction performance. Without such information, distillation degenerates into a vanilla ensemble of encoders, and the improvement is marginal. D): Our proposed SdAOF achieves the best result, validating the effectiveness of both the spatial division information and the distillation strategy.

Robustness to different imaging parameters. In real-world scenarios, the imaging direction may deviate from the perfect alignment with AP and lateral views, and the patient's center may vary. These deviations cause the magnification effect within X-ray images to differ from the training distributions. The results in Fig. 5 show that SdAOF is robust to minor variations in view angles and patient center locations.

4   Conclusion

In this study, we introduce SdAOF, a novel framework designed for the challenging task of reconstructing bone shapes from biplanar X-ray images. SdAOF achieves state-of-the-art fine-scale reconstruction results by integrating the continuous occupancy field representation and the spatial-division augmented distillation strategy. The proposed SdAOF is applicable to multi-view scenarios (whether single view or more than two views), offering adaptability across various settings to balance scanning cost, radiation exposure, and reconstruction quality. Our findings highlight the reconstruction benefits obtained from capturing occlusion patterns. Future research avenues could explore more generalized methods, such as incorporating adaptive subspace division based on 3D semantics. This promising direction remains a focus for our future work.

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

References

1. Almeida, D.F., Astudillo, P., Vandermeulen, D.: Three-dimensional image volumes from two-dimensional digitally reconstructed radiographs: A deep learning approach in lower limb ct scans. Medical Physics **48**(5), 2448–2457 (2021)
2. Bayat, A., Sekuboyina, A., Paetzold, J.C., Payer, C., Stern, D., Urschler, M., Kirschke, J.S., Menze, B.H.: Inferring the 3d standing spine posture from 2d radiographs. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. pp. 775–784. Springer (2020)
3. Biguri, A., Dosanjh, M., Hancock, S., Soleimani, M.: Tigre: a matlab-gpu toolbox for cbct image reconstruction. Biomedical Physics & Engineering Express **2**(5), 055010 (2016)
4. Buttongkum, D., Tangpornprasert, P., Virulsri, C., Numkarunarunrote, N., Amarase, C., Kobchaisawat, T., Chalidabhongse, T.: 3d reconstruction of proximal femoral fracture from biplanar radiographs with fractural representative learning. Scientific Reports **13**(1), 455 (2023)
5. Chen, C.C., Fang, Y.H.: Using bi-planar x-ray images to reconstruct the spine structure by the convolution neural network. In: Future Trends in Biomedical and Health Informatics and Cybersecurity in Medical Devices: Proceedings of the International Conference on Biomedical and Health Informatics, ICBHI 2019, 17-20 April 2019, Taipei, Taiwan. pp. 80–85. Springer (2020)
6. 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)
7. Chênes, C., Schmid, J.: Revisiting contour-driven and knowledge-based deformable models: application to 2d-3d proximal femur reconstruction from x-ray images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24. pp. 451–460. Springer (2021)
8. Clogenson, M., Duff, J.M., Luethi, M., Levivier, M., Meuli, R., Baur, C., Henein, S.: A statistical shape model of the human second cervical vertebra. International journal of computer assisted radiology and surgery **10**, 1097–1107 (2015)
9. Dreischarf, M., Shirazi-Adl, A., Arjmand, N., Rohlmann, A., Schmidt, H.: Estimation of loads on human lumbar spine: A review of in vivo and computational model studies. Journal of biomechanics **49**(6), 833–845 (2016)
10. Ehlke, M., Ramm, H., Lamecker, H., Hege, H.C., Zachow, S.: Fast generation of virtual x-ray images for reconstruction of 3d anatomy. IEEE transactions on visualization and computer graphics **19**(12), 2673–2682 (2013)
11. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2021)
12. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
13. He, T., Collomosse, J., Jin, H., Soatto, S.: Geo-pifu: Geometry and pixel aligned implicit functions for single-view human reconstruction. Advances in Neural Information Processing Systems **33**, 9276–9287 (2020)

14. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
15. Kasten, Y., Doktofsky, D., Kovler, I.: End-to-end convolutional neural network for 3d reconstruction of knee bones from bi-planar x-ray images. In: Machine Learning for Medical Image Reconstruction: Third International Workshop, MLMIR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 3. pp. 123–133. Springer (2020)
16. Lin, Y., Luo, Z., Zhao, W., Li, X.: Learning deep intensity field for extremely sparse-view cbct reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 13–23. Springer (2023)
17. Lin, Y., Wang, H., Chen, J., Li, X.: Learning 3d gaussians for extremely sparse-view cone-beam ct reconstruction. arXiv preprint arXiv:2407.01090 (2024)
18. Lin, Y., Yang, J., Wang, H., Ding, X., Zhao, W., Li, X.: C$\hat{}$ 2rv: Cross-regional and cross-view learning for sparse-view cbct reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11205–11214 (2024)
19. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: Seminal graphics: pioneering efforts that shaped the field, pp. 347–353 (1998)
20. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: 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)
21. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
22. Paxton, N.C., Nightingale, R.C., Woodruff, M.A.: Capturing patient anatomy for designing and manufacturing personalized prostheses. Current Opinion in Biotechnology **73**, 282–289 (2022)
23. Reyneke, C.J.F., Lüthi, M., Burdin, V., Douglas, T.S., Vetter, T., Mutsvangwa, T.E.: Review of 2-d/3-d reconstruction using statistical shape and intensity models and x-ray image synthesis: Toward a unified framework. IEEE reviews in biomedical engineering **12**, 269–286 (2018)
24. Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., Li, H.: Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 2304–2314 (2019)
25. Shakya, M., Khanal, B.: Benchmarking encoder-decoder architectures for biplanar x-ray to 3d bone shape reconstruction. Advances in Neural Information Processing Systems **36** (2024)
26. Yamazaki, A., Ito, T., Sugimoto, M., Yoshida, S., Honda, K., Kawashima, Y., Fujikawa, T., Fujii, Y., Tsutsumi, T.: Patient-specific virtual and mixed reality for immersive, experiential anatomy education and for surgical planning in temporal bone surgery. Auris Nasus Larynx **48**(6), 1081–1091 (2021)
27. Zhang, J., Lv, L., Shi, X., Wang, Y., Guo, F., Zhang, Y., Li, H.: 3-d reconstruction of the spine from biplanar radiographs based on contour matching using the hough transform. IEEE Transactions on Biomedical Engineering **60**(7), 1954–1964 (2013)