3DDX: Bone Surface Reconstruction from a Single Standard-Geometry Radiograph via Dual-Face Depth Estimation

Yi Gu[1], Yoshito Otake[1], Keisuke Uemura[2], Masaki Takao[3], Mazen Soufi[1], Seiji Okada[4], Nobuhiko Sugano[2], Hugues Talbot[5], and Yoshinobu Sato[1]

[1] Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, Japan
gu.yi.gu4@naist.ac.jp, {otake,yoshi}@is.naist.jp
[2] Department of Orthopeadic Medical Engineering, Osaka University Graduate School of Medicine, Japan
[3] Department of Bone and Joint Surgery, Ehime University Graduate School of Medicine, Japan
[4] Department of Orthopaedics, Osaka University Graduate School of Medicine, Japan
[5] CentraleSupélec, Université Paris-Saclay, France

Abstract. Radiography is widely used in orthopedics for its affordability and low radiation exposure. 3D reconstruction from a single radiograph, so-called 2D-3D reconstruction, offers the possibility of various clinical applications, but achieving clinically viable accuracy and computational efficiency is still an unsolved challenge. Unlike other areas in computer vision, X-ray imaging's unique properties, such as ray penetration and standard geometry, have not been fully exploited. We propose a novel approach that simultaneously learns multiple depth maps (front and back surfaces of multiple bones) derived from the X-ray image to computed tomography (CT) registration. The proposed method not only leverages the standard geometry characteristic of X-ray imaging but also enhances the precision of the reconstruction of the whole surface. Our study involved 600 CT and 2651 X-ray images (4 to 5 posed X-ray images per patient), demonstrating our method's superiority over traditional approaches with a surface reconstruction error reduction from 4.78 mm to 1.96 mm and further to 1.76 mm using higher resolution and pretraining. This significant accuracy improvement and enhanced computational efficiency suggest our approach's potential for clinical application.

Keywords: Monocular depth estimation · X-ray radiography · Deep learning · Inverse problem.

1 Introduction

Achieving monocular or 2D-3D reconstruction is a long-standing challenge in computer vision and medical engineering. Practical 3D reconstruction from radiographs has recently been a hot topic, considering the significance of clinical

applications. Usually, multiple radiographs are necessary to perform 3D reconstruction [2,40,3,9,1,8,23,12]. Only a few works have tried to achieve 3D reconstruction using single X-ray images [33,17,22,34]. However, existing works suffer from low reconstruction quality, accuracy, and resolution as well as high computational cost, which significantly limit clinical applications. On the other hand, monocular depth estimation from a single camera image [13], which offers impressive 3D reconstruction, has been extensively studied and widely applied, becoming an essential part of many vision models [26,38,35]. Nevertheless, the relation between a depth map and an X-ray image has barely been explored, especially for the topic of 2D-3D reconstruction.

In this paper, we shed light on a new path to the 3D reconstruction from a single X-ray image using depth estimation. Realizing the unique properties of penetrating rays in X-ray imaging, we propose simultaneous 3D dual-face (front and back) depth estimation from a single X-ray image (namely 3DDX) for 3D reconstruction. In the classic monocular depth estimation problem, the relative depth estimation (RDE) [32,39], which only cares about relative depth, and metric depth estimation (MDE), which estimates absolute physical-unit depth [13,24,4,5,6] are two major task categories. We focus on MDE for meaningful clinical application with physical units. However, conventional losses were designed to estimate a single depth map, whereas we try to estimate multiple depth maps from a single input. To tackle that, we propose generalizing the loss functions to multi-depth-map supervision. Furthermore, we take advantage of a standard imaging geometry, namely the relative position of the X-ray source with respect to the detector, by realizing that the diagnostic radiography is standardized [10]. To the best of our knowledge, we are the first to achieve 3D bone reconstruction from a single X-ray image acquired in a clinical setup using depth estimation.
Contribution: We propose a method (3DDX) for reconstructing 3D bone surfaces with absolute scaling and large field-of-view while retaining high-resolution details from single X-ray images acquired in a clinically standardized geometric setup. Our contribution is three-fold: 1) proposal of a dual-face depth estimation from a single X-ray image by exploiting information from the penetrating X-ray, 2) proposal of a new loss function in a depth map estimation network allowing the scale-specific training under a specific geometric constraint, 3) extensive evaluation using a large-scale hip X-ray image database (600 patients, 2651 X-ray images) paired with CT image through 2D-3D registration. Our code is available at https://github.com/Kayaba-Akihiko/3DDX.

2 Method

Fig. 1 shows an overview of the proposed method. We build a novel framework for estimating the complete 3D shape of the femur and pelvis (including unseen regions) from a single X-ray image. To this end, we propose to estimate front- and back-face depth maps for each target object (e.g., a hemipelvis) for the 3D reconstruction from a plain X-ray image. A depth maps estimation model G_d is trained to estimate all the depth maps. We propose a simple yet effective loss

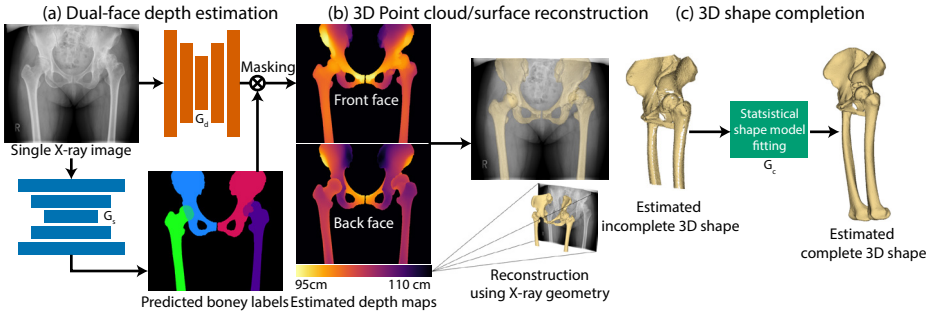

Fig. 1. Overview of the proposed method. (a) The dual-face depth estimation that uses a depth maps estimation model G_d and a bone segmentation model G_s to mask the invalid region. (b) 3D surface reconstruction from the estimated depth maps using X-ray geometry to produce initial 3D estimation. (c) 3D shape completion using bone statistical shape model fitting G_c.

function to improve the depth estimation performance by leveraging the standardized geometry information in X-ray imaging. We also train a segmentation model to generate the masks of target objects from an X-ray image, masking the invalid region (i.e., the non-target region) on the estimated depth maps for bone reconstruction. Using the given X-ray image geometry, the point cloud (PCD) of bone is constructed from the estimated depth maps. We perform 3D shape completion with the statistical shape model (SSM) fitting G_c to validate further the superiority of using dual-face depth.

2.1 Depth maps estimation

We revisited a popular MDE loss, scale-invariant (SI) loss [13,24,4] that preserves learning the global scale and shift for estimating a depth map by minimizing error variance defined as

$$\mathcal{L}_{si} = \alpha\sqrt{D(g)} = \alpha\sqrt{\frac{1}{|I|}\sum_{i\in I} g_i^2 - \frac{\lambda_{var}}{|I|^2}(\sum_{i\in I} g_i)^2}, \tag{1}$$

where $I \equiv \{i \in A : v_i = 1\}, A \equiv \{1, 2, ..., N\}$ are the indices of valid pixel in N pixels indicated by a label map v; $|I|$ represents the number of valid pixels; and $g_i = \log \hat{y}_i - \log y_i$ is the error logarithm between the predicted depth $\hat{y}_i$ and ground truth depth y_i at i-th pixel. α and λ_{var} are two hyper-parameters which we set to 10 and 0.85, respectively, following [24,4]. For the MDE tasks, the costly depth bin techniques [5,6,4] are often used. This work focuses on improving the SI loss for multiple-depth-maps supervision. Generalizing the SI loss to multiple depth maps leads to multiple functions, considering the inter-depth-map pixels relations since the loss considers pixel-to-pixel relations by minimizing the error variance. In the following subsections, we proposed the SI loss generalizations and improvement. We will discuss the performance difference in the Sec. 3.

Generalization to multiple depth maps. Eq. (2) and (3) are two straightforward ways to generalize the SI loss $\mathcal{L}_{si}$ into multiple depth maps. Assuming $J \equiv \{1, 2, ..., P\}$ are the indices of P depth maps. $I^j \equiv \{i \in A : v_i^j = 1\}$ are the indices of valid pixel in j-th depth map indicated by the j-th label map v^j. Thus, the g_i^j is the error logarithm at i-th pixel between the j-th ground truth and estimated depth maps. In particular, Eq. (2) is a simple averaging of the SI losses of all depth maps. In this way, the inter-depth-map pixels are independent of each other as the error variance is calculated separately. For considering inter-depth-map pixels relation, Eq. (3) is presented. The $\mathcal{L}_{si}$ Eq. (1) is a special case of $\mathcal{L}_{si}^{indep}$ and $\mathcal{L}_{si}^{dep}$ when only a single depth map ($|J| = 1$) in supervision.

$$\mathcal{L}_{si}^{indep} = \frac{\alpha}{|J|} \sum_{j \in J} \sqrt{D(g^j)} = \frac{\alpha}{|J|} \sum_{j \in J} \sqrt{\frac{1}{|I^j|} \sum_{i \in I^j} (g_i^j)^2 - \frac{\lambda_{var}}{|I^j|^2} (\sum_{i \in I^j} g_i^j)^2} \tag{2}$$

$$\mathcal{L}_{si}^{dep} = \alpha \sqrt{M(g)} = \alpha \sqrt{\frac{1}{\sum_{j \in J} |I^j|} \sum_{j \in J} \sum_{i \in I^j} (g_i^j)^2 - \frac{\lambda_{var}}{(\sum_{j \in J} |I^j|)^2} (\sum_{j \in J} \sum_{i \in I^j} g_i^j)^2} \tag{3}$$

Center-aligned scale-invariant loss. The vanilla SI error supervises both scale and shift, which is a general need but not in our case. To leverage the standard imaging geometry information, we propose the center-aligned SI loss (CASI), which supervises only the scale while allowing depth shifting by center alignment. A popular way to align the center is centralizing the prediction and ground truth to the depth origin. However, the scale-invariant log error only allows positive depth. Consequently, we propose to align the estimated depth center to the ground truth center using (4) to calculate the error logarithm, where the $t(\cdot)$ calculates the mean of given valid pixels. The ε is a numerical safeguard which we set to 1×10^{-6} to avoid division by zero. The proposed independent and dependent CASI losses were then defined as $\mathcal{L}_{casi}^{indep} = \frac{\alpha}{N} \sum_j \sqrt{D(h^j)}$ and $\mathcal{L}_{casi}^{dep} = \alpha \sqrt{M(h)}$, respectively. Thus, the proposed CASI loss does not introduce new tuning parameters, which lowers the hyperparameter search burden.

$$h_i^j = \log \left(\text{ReLU} \left(\hat{y}_i^j + t(y) - t(\hat{y}) \right) + \varepsilon \right) - \log \left(y_i^j + \varepsilon \right) \tag{4}$$

Segmentation of depth maps. We train a segmentation model G_s to generate the bone masks for removing the background region in the 3D bone surface reconstruction step. We use the Dice semimetric losses [36] and Cross-Entropy loss with label smoothing [30] for training. Segmentation in this task is considered a pixel-wise multi-class classification, allowing label overlay (e.g., in the hip joint region).

2.2 Surface reconstruction and 3D shape completion

We compare the 3D shape completion performance between the single-face-depth-map-reconstructed 3D shape (the conventional method) and the dual-face-depth-map-reconstructed 3D shape (our proposal). The object surfaces are

reconstructed from estimated depth maps with the predicted bone labels using standard imaging geometry. We perform SSM fitting [2,37] for 3D shape completion. We build an SSM for each object we target (i.e., two SSM models built from a training dataset in this paper). The GBCPD algorithm is used [20] for both rigid and non-rigid registration for constructing point-to-point correspondence. During the inference, the statistical shape models are fitted to the incomplete shape to estimate the complete shape. The cost function is defined as

$$\mathcal{L}_{ssm}(\theta) = \mathrm{dist}(\mathrm{clip}(\hat{s}(\theta), s), s) + \frac{\lambda_{l2}}{N_\theta} \sum_i \theta_i^2, \tag{5}$$

where θ is the N_θ-D vector for the optimization for fitting and the second term is a λ_{l2}-weighted $l2$ regularization. $\mathrm{clip}(\hat{s}(\theta), s)$ clips the estimated shape $\hat{s}(\theta)$ to as the same field-of-view as the fitting target shape s. The function $\mathrm{dist}(\cdot)$ measures the bi-directional shape distance if the fitting target is built from the proposed dual-face depth maps; otherwise, it measures the directional shape distance from the target to the shape model. The L-BFGS algorithm [25] was used to search the optimal θ. The λ_{l2} was set to 0.01.

3 Experiment and Result

We collected 2651 X-ray images (600 patients) paired with their respective CT images. CT bone segmentation [19] and X-ray 2D-3D registration [31] were performed to produce ground truth bone 3D shapes and depth maps. Ethical approval was obtained from the Institutional Review Boards at Osaka University and Nara Institute of Science and Technology (approval numbers 15056-3 and 2019-M-6, respectively). In this study, we aim to reconstruct the pelvis and femurs with left and right sides separated. Each object (hemi-bone) CT produced two depth maps (front and back faces) to train the depth model G_d and segmentation model G_s, i.e., eight depth maps (four objects) for a single X-ray, resulting in 10604 bone objects (8626 disease-affected, 1978 healthy, as graded by [29]). In Sec. 3.1, we evaluate the 3D shapes reconstructed from estimated single- and dual- face depth maps. Through 3D shape completion, we further show the performance difference between completion from single- and dual- face-depth-map-reconstructed 3D shapes, which we report in Sec. 3.2. We also compare the proposed CASI loss with conventional SI loss in depth map (2D space) in Sec. 3.3 and reconstructed shape (3D space) in Sec. 3.1. We started from training with a low 256×256 image resolution; however, we further explore performance improvement by image resolution scaling and incorporating pretraining with Masked Autoencoder [18]. A four-fold cross-validation policy was applied to all the experiments, including constructing the SSM models. We excluded 346 (3.26%) objects due to radiography-CT registration failure before gathering the fold results. The segmentation model G_s achieved a Dice coefficient of 0.988. For evaluating 3D shape, the average symmetric surface distance (ASSD), 95 percentile Hausdorff distance (HD95), earth mover's distance (EMD), and l2-chamfer distance (CD_{l2}) were used. The depth center of the estimated 3D shape

was aligned with that of the ground truth shape, i.e., a shift in the Z direction was added to the estimated 3D shape before calculating evaluation metrics. For evaluating the depth map, we used mean absolute error (MAE) and root mean square error (RMSE).

Implementation details The training policy was consistent across all experiments. The AdamW optimizer [28] with SGDR [27] with an initial learning rate of 2×10^{-4} was used, where the T_0 and T_i were set to 10 and 2, respectively. All the deep learning models were trained with 630 epochs, using RandAugment [11]. For the depth model G_d, we used the Norm-Free Network (F0 variant) [7] as the encoder for its training high efficiency and performance. The decoder in G_d followed [15,14]. We used a 2D nnU-Net [21] as the segmentation model G_s trained with 512 resolution. When converting an estimated depth map (after masking) to a point cloud, we created the rays based on the detector size and source-to-detector distance recorded in the DICOM header. We assumed a pinhole camera with a regular viewing frustum without the skew.

Table 1. Evaluation results of point cloud reconstruction with and without shape completion. For shape completion, the healthy and diseased bones are reported separately. *256*, *512*, and *1024* refer to the X-ray resolutions. ※ denotes 3D reconstruction using single-face depth maps. † denotes using pretraining. The mean(std.) of the metrics are reported. ASSD, HD95, EMD, are reported in mm unit; CD_{l2} is in mm^2 unit.

Point cloud evaluation ↓(↓)								
Method	Pelvis				Femur			
	ASSD	HD95	EMD	CD-l2	ASSD	HD95	EMD	CD_{l2}
256 $\mathcal{L}_{si}^{indep}$※	4.78(0.85)	18.0(2.13)	8.55(1.15)	115(30.8)	5.54(1.52)	21.1(2.63)	9.36(2.09)	152(68.9)
256 $\mathcal{L}_{si}^{indep}$	2.11(0.77)	5.82(2.08)	3.14(2.19)	21.2(21.5)	2.28(1.60)	5.75(4.10)	3.13(2.39)	25.6(61.9)
256 $\mathcal{L}_{casi}^{dep}$	1.96(0.77)	5.38(2.09)	2.93(1.18)	18.9(21.2)	2.20(1.66)	5.60(4.39)	3.03(2.49)	25.4(68.7)
256 $\mathcal{L}_{casi}^{indep}$	**1.95**(0.78)	**5.36**(2.10)	**2.92**(1.18)	**18.8**(21.3)	**2.15**(1.66)	**5.49**(4.38)	**2.97**(2.50)	**24.7**(68.5)
256 $\mathcal{L}_{casi}^{indep}$†	1.93(0.77)	5.30(2.11)	2.88(1.17)	18.5(21.2)	2.12(1.66)	5.42(4.42)	2.93(2.50)	24.4(77.3)
512 $\mathcal{L}_{casi}^{indep}$†	1.80(0.76)	4.93(2.07)	2.73(1.16)	16.9(20.7)	1.99(1.56)	5.14(4.17)	2.76(2.39)	21.8(60.0)
1024 $\mathcal{L}_{casi}^{indep}$†	**1.76**(0.75)	**4.82**(2.02)	**2.69**(1.13)	**16.3**(19.3)	**1.95**(1.55)	**5.07**(4.19)	**2.71**(2.34)	**21.2**(61.3)
3D completion evaluation ↓(↓)								
Fitting target	Healthy pelvis				Healthy femur			
	ASSD	HD95	EMD	CDl2	ASSD	HD95	EMD	CD_l2
256 $\mathcal{L}_{si}^{indep}$※	2.34(0.69)	5.95(2.14)	3.13(0.95)	19.7(17.6)	3.78(2.75)	9.21(6.87)	4.88(3.56)	72.3(148)
256 $\mathcal{L}_{casi}^{indep}$	1.95(0.61)	5.06(1.99)	2.73(0.86)	13.9(15.2)	2.19(1.16)	5.40(2.86)	2.93(1.64)	19.3(35.3)
1024 $\mathcal{L}_{casi}^{indep}$†	**1.91**(0.60)	**4.91**(1.98)	**2.66**(0.85)	**13.1**(15.1)	**2.11**(1.15)	**5.22**(2.83)	**2.85**(1.60)	**18.2**(33.7)
	Diseased pelvis				Affected femur			
256 $\mathcal{L}_{si}^{indep}$※	2.55(0.88)	6.84(2.85)	3.50(1.27)	24.8(23.7)	4.56(3.05)	11.5(7.71)	6.11(4.22)	101(151)
256 $\mathcal{L}_{casi}^{indep}$	2.15(0.80)	5.80(2.71)	3.03(1.18)	17.8(21.7)	2.62(1.68)	6.67(4.48)	3.54(2.45)	31.2(69.5)
1024 $\mathcal{L}_{casi}^{indep}$†	**2.05**(0.93)	**5.63**(2.80)	**2.95**(1.33)	**17.5**(37.8)	**2.51**(1.57)	**6.40**(4.19)	**3.40**(2.26)	**28.1**(61.2)

3.1 3D shape results without shape completion

Tab. 1 shows the evaluation results on the 3D shape reconstructed from estimated depth maps from the models trained with different settings. The first-row method (*256* $\mathcal{L}_{si}^{indep}$※) that produced single-face depth maps with $\mathcal{L}_{si}^{indep}$ loss is

regarded as the baseline. When predicting dual-face depth maps ($_{256}\ \mathcal{L}_{si}^{indep}$) with the same SI loss function significantly improved the 3D reconstruction performance, reducing the femur mean ASSD and HD95 from 5.54 and 21.1 mm to 2.28 and 5.75 mm, respectively. Using a higher resolution, the mean ASSD and HD95 for the femur were further improved to 1.95 and 5.07 mm, respectively. This suggests that this generalization of the SI loss is effective. The proposed CASI loss ($_{256}\ \mathcal{L}_{casi}^{indep}$) outperformed the conventional SI loss ($_{256}\ \mathcal{L}_{si}^{indep}$) on all the metrics. We observe that the generalization without inter-depth-map pixel dependency unexpectedly performed better. The behavior may be due to the size (thickness) difference in objects (femur and pelvis), which resulted in different error variance levels, influencing the training. In fact, we chose not to report the results by the SI loss with pixel dependency $\mathcal{L}_{si}^{dep}$, since the training is unstable and often fails. The proposed CASI losses $\mathcal{L}_{casi}^{dep}$ and $\mathcal{L}_{casi}^{indep}$ were stable during training. Fig. 2 shows the visual comparison between the methods on two representative samples, where the proposed methods improved the reconstruction quality significantly.

3.2 3D shape results with shape completion

We use 3D completion to demonstrate the effectiveness of estimating dual-face depth. Tab. 1 shows the evaluation results grouped by the disease. The completion on the proposed dual-face method was significantly better in the fact that much richer 3D information was accessible to fitting, as shown in Fig. 2 (a). The mean ASSDs for the healthy and diseased pelvis improved from 2.34 and 2.55 mm to 1.95 and 2.15 mm, respectively. The proposed dual-face-depth-reconstructed 3D also reduced the fitting outliers significantly indicated by CD_{l2}. The mean CD_{l2} values were reduced from 72.3 and 101 mm^2 to 19.3 and 31.2 mm^2 for the healthy and diseased femur, respectively. We also observed that the segmentation model failed in some regions of patients with unusual diseases, which resulted in degradation in 3D reconstruction performance. Despite that, our method showed robustness against bone deformation.

3.3 Depth map results

To better evaluate the proposed CASI loss, we also evaluated the estimated 2D depth maps, which involved pixel-to-pixel correspondence to the ground truth depth maps. For the femur, the conventional SI loss $\mathcal{L}_{si}^{indep}$ and the proposed CASI loss $\mathcal{L}_{casi}^{indep}$ achieved a mean RMSE of 3.7 mm and 3.52 mm, respectively. For the pelvis, CASI loss improved the mean RMSE from 4.8 to 4.5 mm. Further, bone and muscle volume estimation from an X-ray image had been studied previously by estimating 2D volume distribution [16]. Realizing that the 2D volume distribution is equivalent to thickness estimation at each pixel, our method with dual-face depth estimation is naturally capable of producing volume distribution by subtracting a front depth map from a back depth map to estimate bone volume. Using the proposed CASI loss, the Pearson correlation coefficient

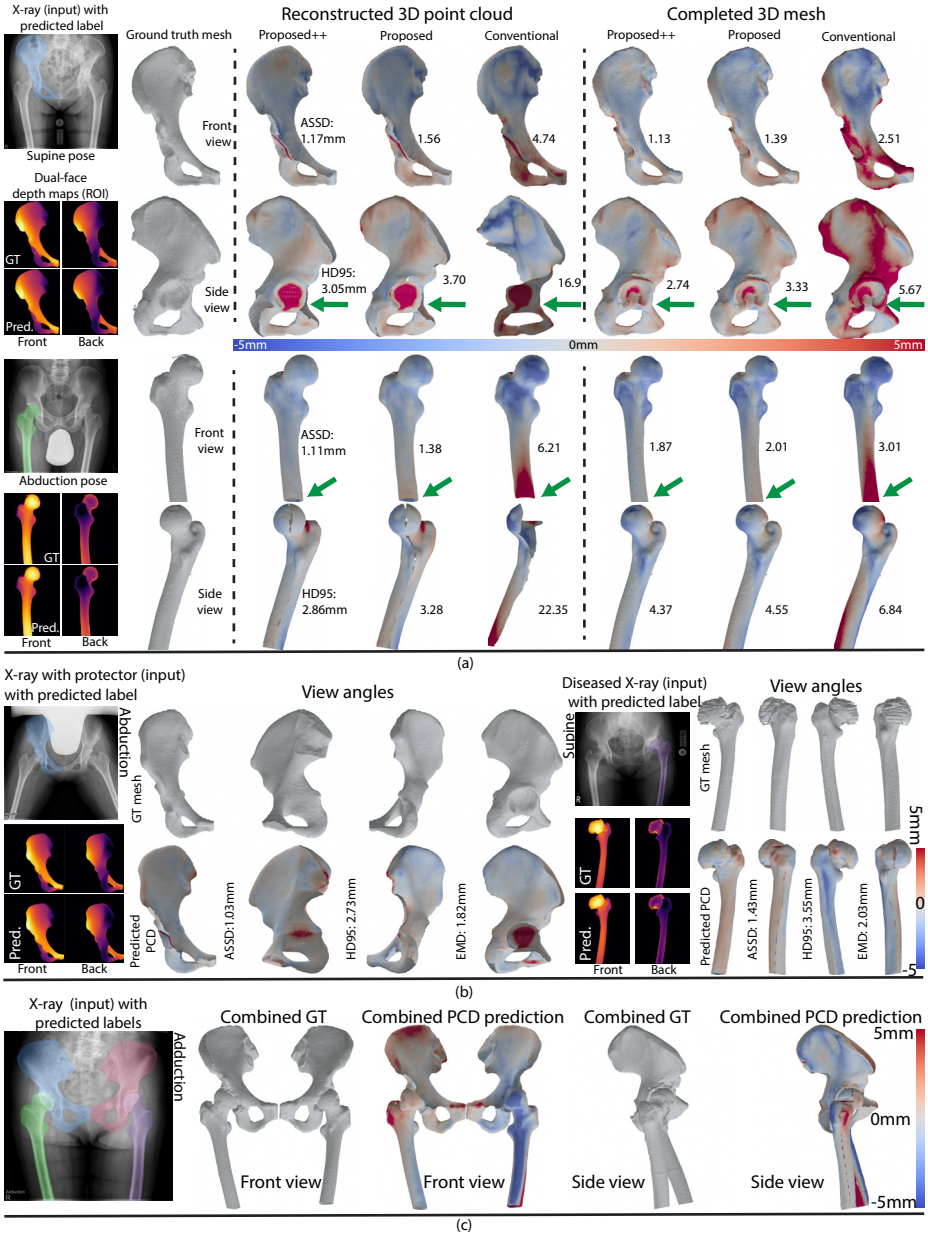

Fig. 2. Visualization of the reconstructed and completed 3D shapes. (a) Comparison between conventional ($_{256}\mathcal{L}_{si}^{indep}$), the proposed ($_{256}\mathcal{L}_{casi}^{indep}$), and the proposed++($_{1024}\mathcal{L}_{casi}^{indep}\dagger$) methods. (b) Visualization of two representative estimated samples (blocked region and diseased/deformed bone) using the proposed method. (c) A combined visualization of the target bones.

(PCC) between X-ray-derived and CT-derived pelvis volume was improved from 0.952 to 0.972 and further to 0.980 by pretraining and resolution scaling.

4 Conclusion and Summary

In this work, we propose a new approach to the fundamentally difficult problem of 2D-3D reconstruction from a single X-ray image, termed 3DDX, where we simultaneously estimate both the front and back faces of in-vivo bone structures of interest. Furthermore, we proposed the generalization of conventional loss to multi-depth-map supervision with improvement by utilizing known geometry information. Through rigorous experiments with a large-scale X-ray dataset on real patients, we demonstrate significant improvement in 3D reconstructions. This work offers potential for many novel and established clinical applications, such as posture estimation, low X-ray dose bone disease detection, diagnosis and follow-up on widely available equipment even outside of hospitals and specialized clinics, particularly in the developing world.

Acknowledgments. The research in this paper was funded by MEXT/JSPS KAKENHI (19H01176, 20H04550, 21K16655).

Disclosure of Interests. The authors have no competing interests to declare relevant to this article's content.

References

1. Almeida, D.F., et al.: 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). https://doi.org/10.1002/mp.14835
2. Baka, N., et al.: 2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models. Med Image Anal **15**(6), 840–850 (Dec 2011). https://doi.org/10.1016/j.media.2011.04.001
3. Balestra, S., et al.: Articulated Statistical Shape Model-Based 2D-3D Reconstruction of a Hip Joint. In: IPCAI. pp. 128–137 (2014)
4. Bhat, S.F., et al.: AdaBins: Depth Estimation using Adaptive Bins. In: CVPR. pp. 4008–4017 (Jun 2021). https://doi.org/10.1109/CVPR46437.2021.00400
5. Bhat, S.F., et al.: LocalBins: Improving Depth Estimation by Learning Local Distributions. In: ECCV. vol. 13661, pp. 480–496. Cham (2022)
6. Bhat, S.F., et al.: ZoeDepth: Zero-shot Transfer by Combining Relative and Metric Depth. arXiv (Feb 2023), https://arxiv.org/abs/2302.12288
7. Brock, A., De, S., Smith, S.L., Simonyan, K.: High-Performance Large-Scale Image Recognition Without Normalization. In: ICML. pp. 1059–1071 (Jul 2021)
8. Cafaro, A., et al.: X2Vision: 3D CT Reconstruction from Biplanar X-Rays with Deep Structure Prior. In: MICCAI. pp. 699–709 (2023)
9. Chênes, C., Schmid, J.: Revisiting Contour-Driven and Knowledge-Based Deformable Models: Application to 2D-3D Proximal Femur Reconstruction from X-ray Images. In: MICCAI. pp. 451–460 (2021)

10. Clohisy, J.C., et al.: A systematic approach to the plain radiographic evaluation of the young adult hip. J Bone Joint Surg Am **90 Suppl 4**(Suppl 4), 47–66 (Nov 2008). https://doi.org/10.2106/JBJS.H.00756
11. Cubuk, E.D., et al.: RandAugment: Practical Automated Data Augmentation with a Reduced Search Space. In: NeurIPS. vol. 33, pp. 18613–18624 (2020)
12. Dobbins III, J.T., McAdams, H.P.: Chest tomosynthesis: technical principles and clinical update. European journal of radiology **72**(2), 244–251 (2009)
13. Eigen, D., Puhrsch, C., Fergus, R.: Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. In: NeurIPS. vol. 27 (2014)
14. Gu, Y., et al.: BMD-GAN: Bone Mineral Density Estimation Using X-Ray Image Decomposition into Projections of Bone-Segmented Quantitative Computed Tomography Using Hierarchical Learning. In: MICCAI. pp. 644–654 (2022)
15. Gu, Y., et al.: Bone mineral density estimation from a plain X-ray image by learning decomposition into projections of bone-segmented computed tomography. Medical Image Analysis **90**, 102970 (Dec 2023)
16. Gu, Y., et al.: MSKdeX: Musculoskeletal (MSK) Decomposition from an X-Ray Image for Fine-Grained Estimation of Lean Muscle Mass and Muscle Volume. In: MICCAI. pp. 497–507 (2023)
17. Ha, H.G., et al.: 2D-3D Reconstruction of a Femur by Single X-Ray Image Based on Deep Transfer Learning Network. IRBM **45**(1), 100822 (Feb 2024)
18. He, K., et al.: Masked Autoencoders Are Scalable Vision Learners. In: CVPR. pp. 15979–15988 (Jun 2022)
19. Hiasa, Y., , et al.: Automated Muscle Segmentation from Clinical CT Using Bayesian U-Net for Personalized Musculoskeletal Modeling. IEEE Trans Med Imaging **39**(4), 1030–1040 (Apr 2020). https://doi.org/10.1109/TMI.2019.2940555
20. Hirose, O.: Geodesic-Based Bayesian Coherent Point Drift. IEEE TPAMI **45**(5), 5816–5832 (May 2023). https://doi.org/10.1109/TPAMI.2022.3214191
21. Isensee, F., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods **18**, 203–211 (Feb 2021)
22. Jiang, L., et al.: Reconstruction of 3D CT from A Single X-ray Projection View Using CVAE-GAN. In: 2021 IEEE International Conference on Medical Imaging Physics and Engineering (ICMIPE). pp. 1–6 (Nov 2021)
23. Kasten, Y., et al.: End-To-End Convolutional Neural Network for 3D Reconstruction of Knee Bones from Bi-planar X-Ray Images. In: MLMIR. pp. 123–133 (2020)
24. Lee, J.H., et al.: From Big to Small: Multi-Scale Local Planar Guidance for Monocular Depth Estimation (Sep 2021), arXiv:1907.10326 [cs]
25. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Mathematical Programming **45**(1-3), 503–528 (Aug 1989)
26. Liu, X., et al.: Multi-Modal Neural Radiance Field for Monocular Dense SLAM with a Light-Weight ToF Sensor. In: ICCV. pp. 1–11. IEEE (Oct 2023)
27. Loshchilov, I., Hutter, F.: SGDR: Stochastic Gradient Descent with Warm Restarts. In: ICLR (Nov 2016)
28. Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization. In: ICLR (Dec 2018)
29. Masuda, M., et al.: Automatic hip osteoarthritis grading with uncertainty estimation from computed tomography using digitally-reconstructed radiographs. IJCARS (in press) (2023), http://arxiv.org/abs/2401.00159
30. Müller, R., et al.: When does label smoothing help? In: NeurIPS. vol. 32 (2019)
31. Otake, Y., et al.: Intraoperative image-based multiview 2D/3D registration for image-guided orthopaedic surgery: incorporation of fiducial-based C-arm tracking and GPU-acceleration. IEEE Trans Med Imaging **31**(4), 948–962 (Apr 2012)

32. Ranftl, R., et al.: Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer. IEEE TPAMI **44**(03), 1623–1637 (Mar 2022). https://doi.org/10.1109/TPAMI.2020.3019967
33. Shiode, R., et al.: 2D–3D reconstruction of distal forearm bone from actual X-ray images of the wrist using convolutional neural networks. Sci Rep **11**(1), 15249 (Jul 2021). https://doi.org/10.1038/s41598-021-94634-2
34. Tan, Z., et al.: XctNet: Reconstruction network of volumetric images from a single X-ray image. Comput Med Imaging Graph **98**, 102067 (Jun 2022)
35. Wang, Q., et al.: Tracking Everything Everywhere All at Once. In: ICCV. pp. 19738–19749. Paris, France (Oct 2023)
36. Wang, Z., et al.: Dice Semimetric Losses: Optimizing the Dice Score with Soft Labels. In: MICCAI. pp. 475–485 (2023)
37. Whitmarsh, T., et al.: Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy X-ray absorptiometry. IEEE Trans Med Imaging **30**(12), 2101–2114 (Dec 2011)
38. Xiang, J., et al.: 3D-aware Image Generation using 2D Diffusion Models. In: ICCV. pp. 2383–2393. IEEE (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.00226
39. Yang, L., Kang, B., Huang, Z., Xu, X., Feng, J., Zhao, H.: Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. In: CVPR (Jan 2024)
40. Youn, K., et al.: Iterative approach for 3D reconstruction of the femur from uncalibrated 2D radiographic images. Medical Engineering & Physics **50**, 89–95 (Dec 2017). https://doi.org/10.1016/j.medengphy.2017.08.016