**ORIGINAL ARTICLE**

# Automatic Bone Surface Restoration for Markerless Computer-Assisted Orthopaedic Surgery

Xue Hu[*] and Ferdinando Rodriguez y Baena

## Abstract

An automatic markerless knee tracking and registration algorithm has been proposed in the literature to avoid the marker insertion required by conventional computer-assisted knee surgery, resulting in a shorter and less invasive surgical workflow. However, such an algorithm considers intact femur geometry only. The bone surface modification is inevitable due to intra-operative intervention. The mismatched correspondences will degrade the reliability of registered target pose. To solve this problem, this work proposed a supervised deep neural network to automatically restore the surface of processed bone. The network was trained on a synthetic dataset that consists of real depth captures of a model leg and simulated realistic femur cutting. According to the evaluation on both synthetic data and real-time captures, the registration quality can be effectively improved by surface reconstruction. The improvement in tracking accuracy is only evident over test data, indicating the need for future enhancement of the dataset and network.

**Keywords:** Bone surface reconstruction, Computer assisted orthopedic surgery, Markerless femur tracking

## 1 Introduction

Osteochondral defect (OCD) represents a common joint disease that causes great pain and discomfort in patients. 62% of the patients requiring arthroscopic intervention for knee pain report the OCD problem [1]. For severe defects that cannot be treated conservatively, synthetic implant replacement is an effective option. The potential of computer assistance in improving surgical outcome has been well recognised. Following the pre-operative plan displayed by the navigation system, the implant can be placed in high congruency with the surrounding anatomy [2]. It is vital to dynamically localise the target knee in a spatial coordinate, so that the computer-generated model initially registered onto the patient can be updated accordingly for surgeons' reference. Conventionally, optical markers are rigidly inserted into the target bones to infer the target movement from the marker tracking. Albeit the high tracking precision, the marker preparation and insertion lead to extra human-induced errors [3], longer surgical workflow [4, 5] and most importantly, higher invasiveness that may cause infection, nerve injury, and bone fracture to patients [6, 7].

Research effort has been recently allocated to automatic femur tracking for markerless knee surgery navigation [8–10]. These methods use a pre-trained convolutional neural network to segment the femur points from the RGBD captures by a commercial depth camera, and then register the segmented points to a 3D reference model created from pre-operative scanning to obtain the real-time spatial pose. However, these works only consider the intact target geometry: given the inevitable surgical interventions such as the bone drilling for lesion removal, the segmented surface points deviate from the pre-scanned intact surface. The increased spatial mismatch will degrade the registration quality (i.e., fitness and inlier matching error) and accuracy (i.e., registered

*Correspondence: xue.hu17@imperial.ac.uk
Mechatronics in Medicine Laboratory, Imperial College London, London, UK

spatial pose compared to the ground-truth pose), making the overall markerless tracking less reliable.

To compensate the spatial inconsistency caused by femur processing, we proposed a supervised surface restoration network based on the PointNet backbone. The network aims to improve the geometric consistency between the segmented surface and the intact model, while retaining the spatial pose of the input point cloud. Since modifying a large number of bone surfaces and capturing the pairwise images with the same target-in-camera pose is highly costly and tedious, we developed a synthetic dataset with simulated realistic surface deformation applied over the collected camera captures. The performance of our trained network was evaluated on both the test dataset and a model knee physically sampled by a depth camera in real-time.

## 2  Related Works

### 2.1  Femoral Surface Processing in OCD Treatment
With the degradation of articular cartilage caused by the overload or cyclical episodes on the knee joint, the subchondral bone may collapse, causing a lesion to form in the smooth joint surface [11]. OCD mainly affects the femoral condyle and/or patellofemoral articulation of the knee [12]. During OCD treatment, the affected area needs to be removed and processed in a regular shape (e.g., ellipse) to be compatible with an optimal implant chosen from the library built through the statistical modelling of real OCD dataset [2].

### 2.2  Learning-Based Surface Reconstruction
With the fast development of deep learning and the availability of large 3D datasets, many generative models have been proposed to reconstruct geometries with the implicitly learned structural features [13]. Most of the proposed networks are based on a vanilla architecture [14]: an encoder extract the latent features of the input data first. Then, the encoded features are depicted by a decoder to generate desired outputs [15–17]. The encoder usually consists of fully convolutional layers, followed by optional pooling, activation and/or fully connected layers. The decoder either contains deconvolutional layers or fully connected layers for upsampling.

The 3D geometry can be recorded in volumetric voxels, surface meshes, point cloud or multi-views. Among them, the point cloud-based learning offers high flexibility and efficiency in computation and memory. PointNet, based on the same encoder-decoder design, is regarded as the most popular backbone for point cloud-based learning [18]. It adopts three unique designs: a symmetry function to deal with unstructured input, a local and global information aggregation layer and a joint alignment network. Inspired by PointNet, many new

extensions and variants exist. An example is PointNetLK, a network trained for the point cloud registration [19]. By integrating a modified Lucas&Kanade (LK) algorithm in the PointNet framework, the spatial transformation can be depicted by a differentiable learning framework.

## 3  Materials and Methods
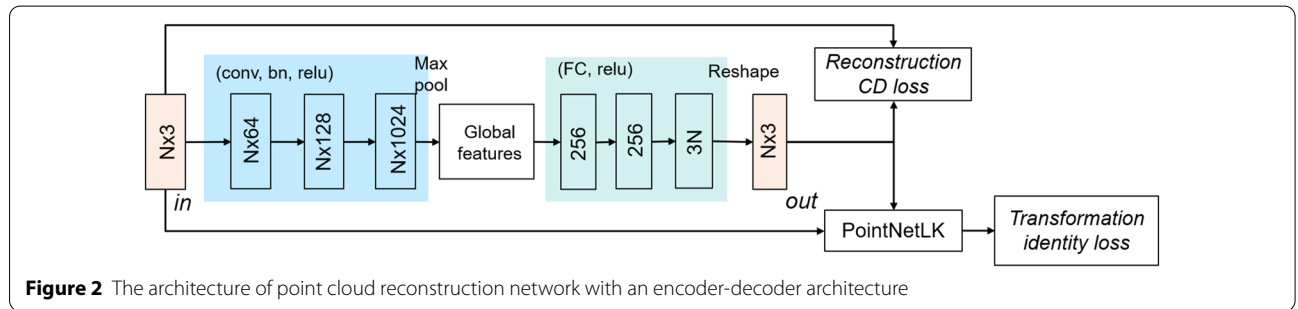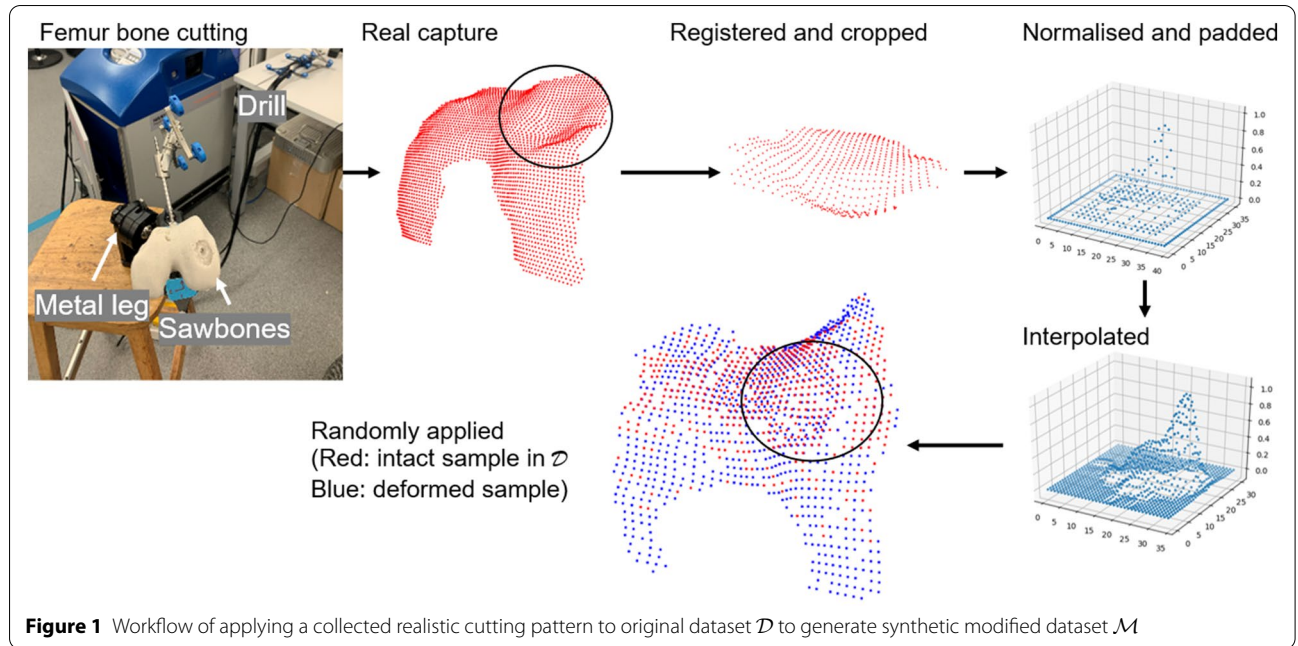
### 3.1  Synthetic Dataset with Simulated Bone Cutting
We collected a dataset $\mathcal{D}$ containing depth captures of a cadaver leg with an intact femur surface. A commercial depth camera, Realsense D415, was used to be consistent with the original work [8]. The camera can acquire depth data with less than 2% error at up to 90 frames per second, which is competent for real-time applications [20] The dataset includes 9334 depth captures of lab scenes. Each depth image is represented by a $160 \times 160 \times 4$ matrix. The first three channels in the last dimension are $(x, y, z)$ coordinates of a sampled pixel, and the last channel is the binary label that denotes whether the point belongs to the femur surface.

Sawbones were used to generate realistic femur cutting patterns. Figure 1 shows the overall workflow. The original femur surface $f_o$ was first captured and manually segmented from the depth frame. After cutting the femur around the condal following a conventional procedure for OCD removal, the modified femur surface $f_m$ was again captured. Due to the possible change in spatial pose, $f_o$ was registered to $f_m$, and the depth value $\tilde{z}$ of points fall in an annotated cutting $A$ area was interpolated. $\tilde{z}$ was subtracted by the depth value of $f_m$ in the cutting area (i.e., $\Delta z_A = z - \tilde{z}$) and normalised. To ensure the smooth connection between modified and unmodified surface, zeros were padded around the edge of $\Delta z_A$. The padded 3D variation was fitted by Clough Tocher 2D interpolation for a cutting pattern $f$. The same procedure was repeated 20 times to generate enough patterns with different cutting shapes and depth variations.

The intact femur points $p$ were segmented from $\mathcal{D}$ according to their binary labels. $K$ ($= 1\text{-}3$ in our trial) rectangular area were selected on the femur surface with the arbitrary size and location, to which the collected deformation patterns $f$ were mapped and scaled by the arbitrary maximum intrusion depth (i.e., 0–15 mm). The original and deformed point clouds were separately resampled in a $N \times 3$ point cloud and concatenated into a $N \times 6$ array. Since the time of later reconstruction is proportional to the number of points $N$, we chose $N = 2500$ as a compromise between speed and surface representation quality.

### 3.2  Network Architecture
The surface reconstruction network is shown in Figure 2. The encoder borrows the PointNet design of sequential

**Figure 1** Workflow of applying a collected realistic cutting pattern to original dataset $\mathcal{D}$ to generate synthetic modified dataset $\mathcal{M}$



**Figure 2** The architecture of point cloud reconstruction network with an encoder-decoder architecture

multilayer perceptrons (MLPs) and a max pooling layer to ensure the same loss expression regardless the order of input points. Note that for our need, no transformation (based on T-Net) is included for input and features, since it is designed to increase classification accuracy [19], and we want to keep the spatial consistency between input and output. After encoding the deformed points $p'$, the most critical point among $N$ points is selected for every latent dimension. The obtained latent feature vector is then passed to the decoder that consists of three fully convolutional layers to recover the $N \times 3$ points $\hat{p}$. The reconstruction loss is defined as the chamfer distance (*CD*) between the output $\hat{p}$ and the ground truth (GT) intact points $p$:

To ensure the spatial consistency, the reconstructed $\hat{p}$ and GT $p$ are fed to a pretrained PointNetLK to obtain a relative transformation $T$. We define a spatial regularisation loss $\mathcal{S}$ as the mean-squared error (MSE) between $T$ and an identity matrix. The overall cost function is defined as (where $\lambda$ is a weight factor):

$$\mathcal{C} = \mathcal{R} + \lambda\,\mathcal{S}. \tag{2}$$

## 4  Network Training and Evaluation
The networks were trained within the PyTorch framework. The full synthetic dataset $\mathcal{M}$ was randomly divided into training, validation and testing groups with a ratio

$$\mathcal{R} = CD(\hat{p}, p) = \frac{1}{|\hat{p}|} \sum_{x \in \hat{p}} \min_{y \in p} ||x - y||_2^2 + \frac{1}{|p|} \sum_{y \in p} \min_{x \in \hat{p}} ||x - y||_2^2. \tag{1}$$

of 6:2:2. The Adam optimiser with a descending learning rate starting from 0.0005 was used. The reconstruction network was first trained for 300 epochs to minimise reconstruction loss only. Then, the network was further trained with the regularisation by a PointNetLK registration network well-trained on ModelNet40 dataset. The trained registration network could achieve comparable performance with traditional iterative closest point (ICP) over our dataset (for registration between $p$ and $\hat{p}$, the mean square difference between network output and ICP output is 0.0008 in average). The weight $\lambda$ was chosen as 0.001. The training took 17 hours on an Nvidia Tesla P100-PCIe Graphic Processing Unit (GPU) before validation error further decreases.

## 5 Evaluation

### 5.1 Registration Test on Dataset

The performance of the proposed network was evaluated on the test dataset first. The points of reference geometry were scanned by a highly precise commercial scanner (HDI 3D scanner, LMI Technologies Inc.). After the RANSAC global alignment [21], the femur pose $T_{def}$ was computed by the standard ICP registration (Open3D library [22]) between reference points and the deformed points with or without reconstruction. The threshold corresponding distance was set to 5 mm. We evaluated the registration reliability in terms of fitness (i.e., the number of inlier correspondences divided by the total number of target points) and the root-mean-squared error (RMSE) between the registered inlier correspondences. The spatial tracking error induced by geometric changes was defined as the difference between registered poses $T_{def}$ and the GT poses $T_{undef}$ obtained between the undeformed frames and reference. As shown in Figure 3, by reconstruction, the registration fitness is improved

from 76.29±6.50% to 87.55±5.06%, and the RMSE is reduced from 2.26±0.17 mm to 1.96±0.11 mm. Both the 3D positional and rotational accuracy are significantly (p-values<0.05 by a Kruskal-Wallis one-way analysis of variance test).

We additionally ran a spatial consistency check by registering the reconstructed point clouds from multiple single-view captures to the same pre-scanned reference. As shown in Figure 4, the merged frames are consistent with the model, indicating that the 3D geometric features of the input were learned by the network and used for the surface restoration.

### 5.2 Test on Real-Time Captures

The combined markerless segmentation, reconstruction and registration workflow were then tested on a model knee (i.e., a drilled femur sawbones held by a metal leg). An optical marker was pinned into the leg and tracked by an optical tracker (Fusiontrack 500, Atracsys LLC) to provide the GT pose $T_{gt}$ for evaluation. The difference between the markerless femur pose ($T$) registered from the restored surface and the $T_{gt}$ obtained by marker-based tracking was defined as the tracking error:

$$T_{err} = T\,(T_{gt})^{-1}. \tag{3}$$

The pose registered from the raw segmented points without reconstruction was also evaluated as a reference.

The Python inference takes 0.05 s per frame for the real-time reconstruction. As shown in Figure 5, the reconstruction improves the registration fitness from 75.56±6.36% to 81.69±10.97%, and reduces the registration RMSE from 2.40±0.10 mm to 2.07±0.30 mm. However, the improvement in 3D spatial pose is not obvious: the Kruskal-Wallis test shows no significant difference
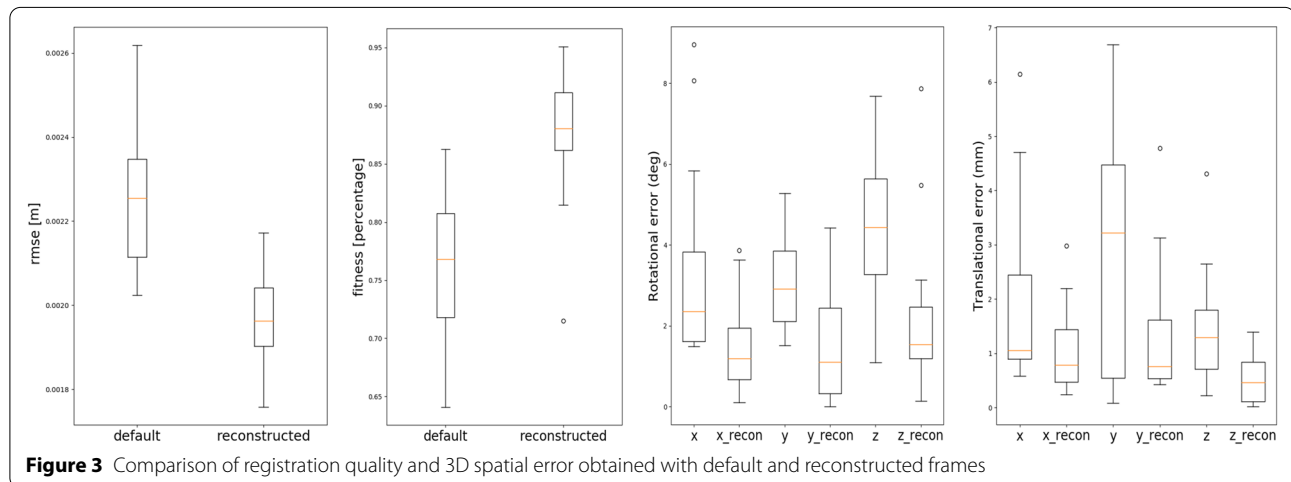


**Figure 3** Comparison of registration quality and 3D spatial error obtained with default and reconstructed frames

**Figure 4** Comparison between a reconstructed model fused from 20 frames (yellow) and the pre-scanned femur surface (blue)
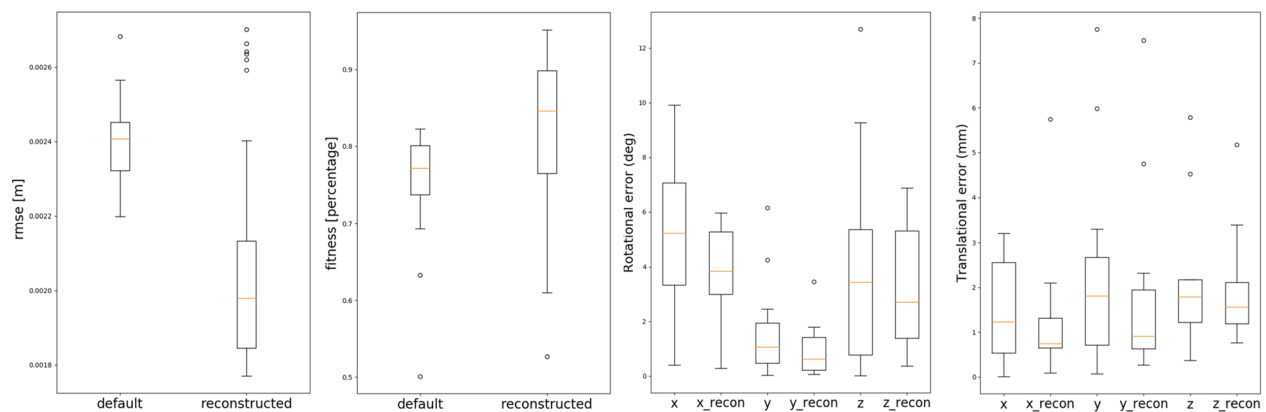


**Figure 5** Comparison of real-time femur tracking quality and errors for manipulated femur surface with (reconstructed) and without reconstruction (default)

between results obtained with and without reconstruction (p-value=0.83 for translational error and 0.28 for rotational error).

## 6 Discussion
While the camera can only provide maximum 2.5D partial views, the geometry modification should be consistent across viewpoints to match a common 3D geometry. Learning 3D features is essential for such depth frame-based modification. The network implicitly learns the structural features of the femur from a modified point cloud, and selects the ones relevant to the original frames through supervision.

The performance degradation from the test dataset to the actual capture is noticeable. We suspect that this may be due to the domain gap between the synthetic surface and the physically modified surface captured by a depth camera. In practice, the image sampling quality will be affected by the geometry variation, capturing angles and working distance. The effect of a shape modification on the actual capture cannot be directly superimposed. For example, a hole around a flat surface will look different from the same hole around the edge in the actual capture. Besides, the training dataset only contains one model geometry, which is slightly different from the shape of the actually tested sawbones. The decoder is prone to overfitting and not be fully generalised to a new geometry.

## 7 Conclusions
In this work, we explore the possibility of adopting a deep neural network for automatic femur surface restoration to improve the reliability of markerless tracking after surgical intervention. We train the network on a collect dataset with simulated realistic surface modification. While the registration quality is effectively increased, the improvement in tracking accuracy is obvious over test data but not under the physical setup. In the future, we will further improve the realism of synthetic deformation, involve more femur geometries (e.g., given by data-driven

statistical shape models), and tune the network architecture for better performance.

## Authors' Information
Xue Hu received the B.Eng. degree from Beihang University (BUAA), Beijing, China, in 2017, and the M.Sc. degree from Imperial College London, UK, in 2018. She is currently pursuing a PhD degree in the Department of Mechanical Engineering with the Mechatronics in Medicine Laboratory, Imperial College London, the UK. Her research interests include augmented reality, computer-assisted orthopaedic surgery, machine learning and 3D vision applications.

Ferdinando Rodriguez y Baena (Member, IEEE) received the M.Eng. degree from King's College London, U.K., in 2000, and the PhD degree in medical robotics from Imperial College London, in 2004. He is currently a Professor in medical robotics with the Department of Mechanical Engineering, Imperial College London, where he leads the Mechatronics in Medicine Laboratory. He is also co-director of the Hamlyn Centre, Imperial College London. His current research interests include the application of mechatronic systems to medicine, in the specific areas of clinical training, diagnostics, and surgical intervention.

## References
[1] D C Flanigan, J D Harris, T Q Trinh, et al. Prevalence of chondral defects in athletes' knees: a systematic review. *Medicine & Science in Sports & Exercise*, 2010, 42(10): 1795–1801.
[2] F Tatti, H Iqbal, B Jaramaz, et al. A novel computer-assisted workflow for treatment of osteochondral lesions in the knee. *The 20th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery,* 2020(4): 250–253.
[3] D K Bae, S J Song. Computer assisted navigation in knee arthroplasty. *Clinics in Orthopedic Surgery,* 2011, 3(4): 259–267.
[4] D C Beringer, J J Patel, K J Bozic. An overview of economic issues in computer-assisted total joint arthroplasty. Clinical Orthopaedics and Related Research®, 2007, 463: 26–30.
[5] A D Pearle, P F O'Loughlin, D O Kendoff. Robot-assisted unicompartmental knee arthroplasty. *The Journal of Arthroplasty,* 2010, 25(2): 230–237.
[6] R W Wysocki, M B Sheinkop, W W Virkus, et al. Femoral fracture through a previous pin site after computer-assisted total knee arthroplasty. *The Journal of Arthroplasty*, 2008, 23(3): 462–465.
[7] A P Schulz, K Seide, C Queitsch, et al. Results of total hip replacement using the robodoc surgical assistant system: clinical outcome and evaluation of complications for 97 procedures. *The International Journal of Medical Robotics and Computer Assisted Surgery,* 2007, 3(4): 301–306.
[8] H Liu, F R Y Baena. Automatic markerless registration and tracking of the bone for computer-assisted orthopaedic surgery. *IEEE Access,* 2020, 8: 42010–42020.
[9] P Rodrigues, M Antunes, C Raposo, et al. Deep segmentation leverages geometric pose estimation in computer-aided total knee arthroplasty. *Healthcare Technology Letters*, 2019, 6(6): 226–230.
[10] X Hu, A Nguyen, F R y Baena. Occlusion-robust visual markerless bone tracking for computer-assisted orthopaedic surgery. *IEEE Transactions on Instrumentation and Measurement*, 2021. https://doi.org/10.1109/TIM.2021.3134764.
[11] F M M da Cunha Cavalcanti, D Doca, M Cohen, et al. Updating on diagnosis and treatment of chondral lesion of the knee. *Revista Brasileira de Ortopedia*, 2012, 47(1): 12–20.
[12] C R Wheeless. Chondral and osteochondral injuries of the knee. In: Wheeless' Textbook of Orthopaedics, 2001.
[13] A Kurenkov, J Ji, A Garg, et al. Deformnet: free-form deformation network for 3D shape reconstruction from a single image. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018: 858–866.
[14] X Han, H Laga, M Bennamoun. Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(5): 1578–1604.
[15] M Tatarchenko, A Dosovitskiy, T Brox. Multi-view 3D models from single images with a convolutional network. *European Conference on Computer Vision*, 2016: 322–337.
[16] J Wang, B Sun, Y Lu. Mvpnet: multi-view point regression networks for 3D object reconstruction from a single image. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33: 8949–8956.
[17] D Eigen, C Puhrsch, R Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in Neural Information Processing Systems*, 2014: 2366–2374.
[18] C R Qi, H Su, K Mo, et al. Pointnet: deep learning on point sets for 3D classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 652–660.
[19] Y Aoki, H Goforth, R A Srivatsan, et al. Pointnetlk: robust & efficient point cloud registration using pointnet. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2019: 7163–7172.
[20] L Keselman, J I Woodfill, A Grunnet-Jepsen, et al. Intel(r) realsense(tm) stereoscopic depth cameras. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017: 1267–1276.
[21] S Choi, Q Y Zhou, V Koltun. Robust reconstruction of indoor scenes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 5556–5565.
[22] Q Y Zhou, J Park, V Koltun. Open3D: a modern library for 3D data processing. 2018, arXiv:1801.09847.