Article:

Alzaid, A., Lineham, B., Dogramadzi, S. et al. (3 more authors) (2023) Simultaneous Hip Implant Segmentation and Gruen Landmarks Detection. IEEE Journal of Biomedical and Health Informatics. ISSN 2168-2194

https://doi.org/10.1109/JBHI.2023.3323533

Simultaneous Hip Implant Segmentation and Gruen Landmarks Detection

Asma Alzaid , Beth Lineham, Sanja Dogramadzi, Hemant Pandit, Alejandro F. Frangi and Sheng Quan Xie

***Abstract*— The assessment of implant status and complications of Total Hip Replacement (THR) relies mainly on the clinical evaluation of the X-ray images to analyse the implant and the surrounding rigid structures. The current clinical practice depends on the manual identification of important landmarks to define the implant boundary and to analyse many features in arthroplasty X-ray images which is time-consuming and could be prone to human error. Semantic segmentation based on Convolutional Neural Network (CNN) has demonstrated successful results in many medical segmentation tasks. However, these networks cannot define explicit properties that lead to inaccurate segmentation, especially with the limited size of image datasets. Our work integrates clinical knowledge with CNN to segment the implant and detect the important features simultaneously. This is instrumental in arthroplasty complications diagnostics, particularly for implant loosening and implant-close bone fractures, where the fracture location in relation to the implant has to be accurately determined. In this work, we define the landmarks using Gruen zones that represent the implant's interface with the surrounding bone to build a Statistical Shape Model (SSM). We propose a multi-task CNN that combines regression of pose and shape parameters constructed from the SSM and semantic segmentation of the implant. This integrated approach has improved the estimation of the implant shape, from 74% to 80% dice score, making the segmentation realistic and allowing automatic detection of the Gruen zones. In order to train and evaluate our method, we generated a dataset of annotated hip arthroplasty X-ray images that will be made available.**

***Index Terms*— Arthroplasty, Image segmentation, Landmarks detection, Medical image analysis, Statistical Shape Model,**

I. Introduction

Total Hip Replacement (THR) follow-up radiographs are used in routine prosthetic joints evaluation and monitoring to identify any potential complications. These include loosening, infection, and other short and long-term problems related to the region surrounding the implant. For instance, Aseptic loosening, which is the most common cause for THR revision [1] is detected by visually assessing the radio-lucencies 'gaps' around the implant and determining the implant's positional variations in relation to the bone. The widely used clinical protocol for assessing the implant status is the Gruen zone system, which divides the interface between the bone and implant into seven zones (see Fig.1 (a)). In clinical practice, these landmarks and the surrounding boundary of the implant are defined by clinicians, often time-consuming, and prone to human error process that could lead to inconsistencies in outcomes between various clinical specialists. Automating the identification of these landmarks and segmenting the implant can minimize these problems and ultimately lead to more efficient and reliable diagnoses, better treatment planning, and, ultimately, improved patient outcomes.

In several research studies [2] [3] and medical imaging analysis and assisted tools in orthopaedics such as Ortho View and ELBRA, manual selection of anatomical landmarks or implant boundaries is used for subsequent analysis. To the best of our knowledge, there is currently no existing work on automated identification of the Gruen landmarks. On the other hand, several studies attempted to automate the segmentation of hip implant. The early work on implant segmentation considered the analysis of images based on hand-crafted features such as histogram thresholding [4] [5], Active Contour method initialized by using the Fast randomized circle detection method [3], and the region growing method initialized by applying the Hough transforms [2]. These methods are not generalized well towards THR radiographs and could provide good results only when the implant components are clearly presented in the X-ray images. Moving from traditional-based methods to Deep Learning (DL) based methods, Patel et al. [6] applied U-Net to segment hip implants as an initial step for classification of the type of implant. Even though the Convolution Neural Network (CNN) showed state-of-the-art results in many medical segmentation tasks, these networks map the global shape structure and can not define the local regional properties. In addition, these networks could produce unrealistic segmentations i.e. gaps or missing parts in the segmented implant, especially when the training dataset is limited, which is considered a major challenge in many medical imaging research. Similarly, the Gruen landmarks might not have simple distinguishable features to be learned by a CNN. It is defined based on the shape and geometry of the implant and

Asma Alzaid is with the School of Computing, University of Leeds, Leeds, UK (e-mail: scaalz@leeds.ac.uk).

Beth Lineham is with Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK.

Sanja Dogramadzi is with Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK.

Hemant Pandit is with Leeds Teaching Hospitals NHS Trust, Leeds, UK. and Leeds Orthopaedic Trauma Sciences, University of Leeds, Leeds, UK.

Alejandro Frangi is with Centre for Computational Imaging and Simulation Technologies in Biomedicine, School of Computing, University of Leeds, UK.

Sheng Quan Xie (Corresponding author) collaborates with Institute of Rehabilitation Engineering, Binzhou Medical University, Yantai, China. He is with School of Electrical and Electronic Engineering, University of Leeds, UK.

its surrounding bone. CNN exceeds at learning hierarchical representations of the visual features but may have difficulty capturing precise geometric features and shapes particularly if the training dataset is small.

Increasing the dataset size would improve the performance of CNN based methods, however, it is difficult and time-consuming to annotate a large number of THR X- ray images. In addition, the quality, complexity and variety of THR images may limit the effectiveness of synthesizing new data [7]. Therefore, we introduce a hybrid approach that leverages the shape knowledge of hip implants for simultaneous segmentation and detection of Gruen landmarks in the implant. Although several studies in the medical image analysis domain incorporate shape knowledge into DL such as segmentation of left ventricle [8], brain boundary [9] and skin lesions [10], this is the first work that uses such an approach for implant shape segmentation and landmark localization. This paper proposes a multi-task CNN to perform a binary segmentation map of the implant, detect the implant tip point and regress SSM parameters to compute the shape of the implant. We employed the Statistical Shape Model (SSM) to build a landmark-based shape model from a training dataset and fit this model to a new image using the shape coefficients and pose parameters. We combine the advantages of SSM for both imposing shape constraints and describing the important landmarks in the implant. In addition, we preserved the benefits of CNN to extract complex features from images.

Integrating segmentation and regression of the shape model parameters was utilised in other medical imaging domains, either using two parallel steps- one for predicting shape parameters and the other one for predicting the segmentation map such as in prostate segmentation in MRI image [11] or by combining the two steps in one pipeline [12] [13]. Regression of the SSM parameters and the distance map to segment the left ventricle was developed in [12], while [13] predicted shape coefficients and pose parameters to compute the coordinates of the landmark points that approximated the final segmentation. Compared to these methods, our approach improves the segmentation as well as landmarks identification by simultaneously predicting the SSM parameters, the implant tip point and perform the binary segmentation maps.

The novel contributions of this work are: (1) We propose integrated approach that allowed segmentation of implant and automatic detection of the landmarks of interest in the implant. (2) define Gruen zone landmarks and represent the shape of implant femoral component accordingly. (3) Annotated THR images dataset that defines implant landmarks. It will be publicly available to enhance the research in this field.

II. Related work

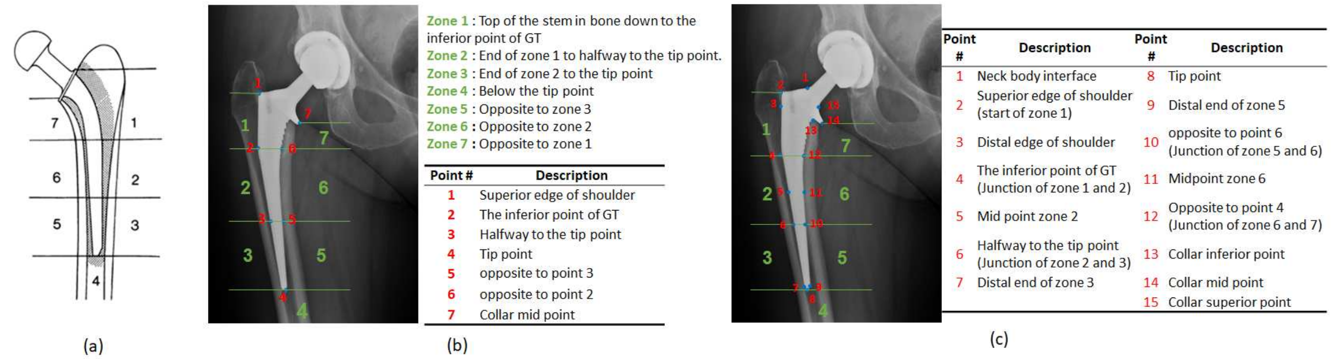
There are many currently adopted approches in the medical image domain that introduced the integration of shape knowledge with CNN. These approaches can be divided into five main categories: (1) post-processing by shape model, (2) prior knowledge, (3) multiple CNNs and shape models, (4) learning hidden representations of shape and (5) shape prior as regularization in the objective function.

follows this strategy for segmentation and tracking of the left ventricle. They swap out the CNN for a Faster-RCNN and use an improved ASM that allows to obtain matching points in greater ranges. ASM improves R-CNN segmentations for detection and tracking

The shape model is used as a post-processing step to refine the CNN segmentation. A method of using left ventricle segmentation that initialized the segmentation with a Faster R-CNN model for detection and tracking was reported in [15] followed by a selection-based sparse shape model and a local deformable model to perform the final segmentation. A modified Active Shape Model (ASM) to refine the segmentation of the left ventricle was implemented in [8]. Since the main limitation of ASM is high outliers as a result of searching for landmarks, the authors took advantage of CNN to maximize the quality of feature extraction from images. The Expectation-Maximization was selected to minimize the effect of outliers during the ASM optimization. Rather than using segmentation maps for initializing the shape model, Tabrizi et al. [16] predicted the bounding boxes as initializations, and the final segmentation using the weighted fuzzy ASM. A similar approach was introduced by Li et al. [17] for myocardial segmentation, where they applied random forest to build probability maps from the detected bounding box and utilized SSM for the final segmentation.

Prior shape knowledge is applied to generate the initial segmentation. Nguyen et al. [9], split images into groups with similar shapes and structures of brain boundaries. Then, prior ASM was used for each group to generate coarse segmentation, followed by a CNN and post-processing methods such as Conditional Random Field (CRF) and Gaussian processes to refine the segmented contours. Extended U-net architecture by incorporating multi-resolution input and integrating a shape prior as a template for cardiac MRI segmentation were reported by Zotti et al. [18]. Shape priors encoded the probability of a voxel being part of a specific class, which is used in segmentation and for predicting the central location of the object.

More accurate outcomes can be obtained by using multiple CNNs and shape models. A pipeline of multiple CNNs and SSMs to segment knee bone and cartilage from MRI images was proposed by Ambellan et al. [19]. The pipeline started with 2D U-Net to generate initial segmentation masks which are then regularized by SSMs. Then, 3D U-Net is employed to extract smaller MRI subvolumes. To further enhance the results, another SSM is used as a post-processing step. Finally, a third U-Net is used to segment the cartilage. Three steps pipeline of hippocampus segmentation from MRI was proposed by Brusini et al. [20], using U-Net, SSMs and a second U-Net. They utilized three orthogonal U-Nets and averaged their prediction to extract the final segmentation. A segmentation method for cardiac images that combined a multi-task DL approach with an atlas as a prior shape was reported by Duan et al. [21]. Their method trained a Fully Convolutional Network (FCN) for both segmentation and landmarks detection. The landmarks were used to initialize the atlas by selecting the most corresponding one, which is used to refine the segmentation.



Point #	Description
1	Superior edge of shoulder
2	The inferior point of GT
3	Halfway to the tip point
4	Tip point
5	opposite to point 3
6	opposite to point 2
7	Collar mid point

Point #	Description	Point #	Description
1	Neck body interface	8	Tip point
2	Superior edge of shoulder (start of zone 1)	9	Distal end of zone 5
3	Distal edge of shoulder	10	opposite to point 6 (Junction of zone 5 and 6)
4	The inferior point of GT (Junction of zone 1 and 2)	11	Midpoint zone 6
5	Mid point zone 2	12	Opposite to point 4 (Junction of zone 6 and 7)
6	Halfway to the tip point (Junction of zone 2 and 3)	13	Collar inferior point
7	Distal end of zone 3	14	Collar mid point
		15	Collar superior point

Fig. 1. (a) Femoral component zones according to Gruen et al. [14], (b) Modified definition of Gruen zones. (c) Shape landmarks description.

For learning the hidden representation of the anatomical shape and topological structures to impose the shape constraints on the initial segmentation, denoising autoencoders (DAE) were used for post-processing step for lung segmentation [22], constraint variational autoencoder (cVAE) for learning the latent representation of cardiac shapes [23], and a Shape-aware Multi-view AutoEncoder (Shape MAE) for learning the anatomical shape priors of cardiac anatomy [24].

The approaches that combine the shape priors as regularization terms in the loss function of the segmentation network are based on either using the landmarks distance [11] [25] or on the shape parameters [13] [12]. The normalized distant maps for the constructed contour from SSM parameters were combined in the segmentation as a parallel step to a network that generated the probability maps for prostate segmentation [11], or as the initialization step of the segmentation [25]. A stage-wise regression model is proposed in [13] that initially predicted the centre location of the prostate and subsequently incorporated shape parameters and rotation vector predictions. In contrast, [12] incorporated regression of shape and pose parameters along with the distance maps regression in one pipeline to segment the left ventricle. To enhance the segmentation of skin lesions, a shape prior was encoded as a new loss term in an FCN, with non-star shape segments being penalized in the prediction maps [10].

In this paper, we propose a hybrid approach that leverages the shape knowledge of hip implant for simultaneous segmentation and detection of important landmarks. A multi-task CNN is proposed that incorporates aspects of previous approaches to automatically extract an implant shape representation that can be utilized for several regional assessments. Compared with category 5 approaches to use the shape priors as regularization in the objective function, we regress the shape parameters of an SSM that helps us to identify the important landmarks in the implant, which enable further computation and extraction of implant surrounding regions. In contrast to other methods, we improve shape prediction by simultaneously detecting the implant tip point and performing semantic segmentation. In addition, a final alignment of the shape is calculated by applying the ICP algorithm. Our proposed architecture is designed as an encoder-decoder CNN where the features in the encoder part have shape-related information. These feature maps are shared by both branches- regression of pose and shape parameters and semantic segmentation. This automate the identification of the Gruen landmarks by constructing the implant shape from the predicted parameters.

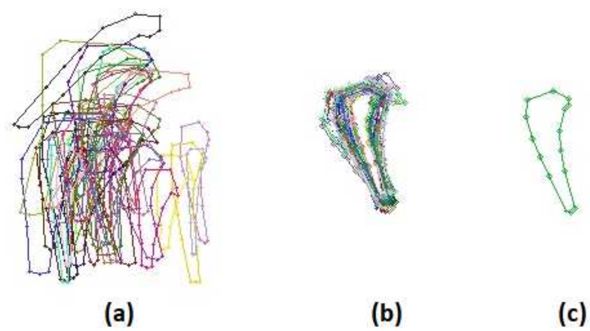


Fig. 2. GPA steps: (a) Samples of training shapes. (b) Aligned shapes. (c) mean shape

III. METHODS

A. Anatomical knowledge

The clinical assessment of THR postoperative radiographs includes examining the changes in the appearance of implant components and bone. Experienced clinicians depend greatly on their knowledge of the anatomical priors such as shape and position of the implant and bone for assessing radiograph images. We include this knowledge into our DL model to segment the implant and detect the important landmarks of the femoral component of the implant. The most widely used medical system for evaluating the status of the femoral stem is the Gruen system [14], which divides the femoral component into seven zones in Antero-Posterior (AP) radiograph (see Fig. 1 (a)). We introduce shape landmarks based on these zones (Fig. 1 (b) shows the definition of Gruen landmarks).

B. Shape Model

THR radiograph images significantly vary in appearance depending on the condition of the patient and the complication after THR surgery. A SSM, that describes the object shape and its variations [26], is generated from a training image set that is annotated by a human expert and built from the analysis of the

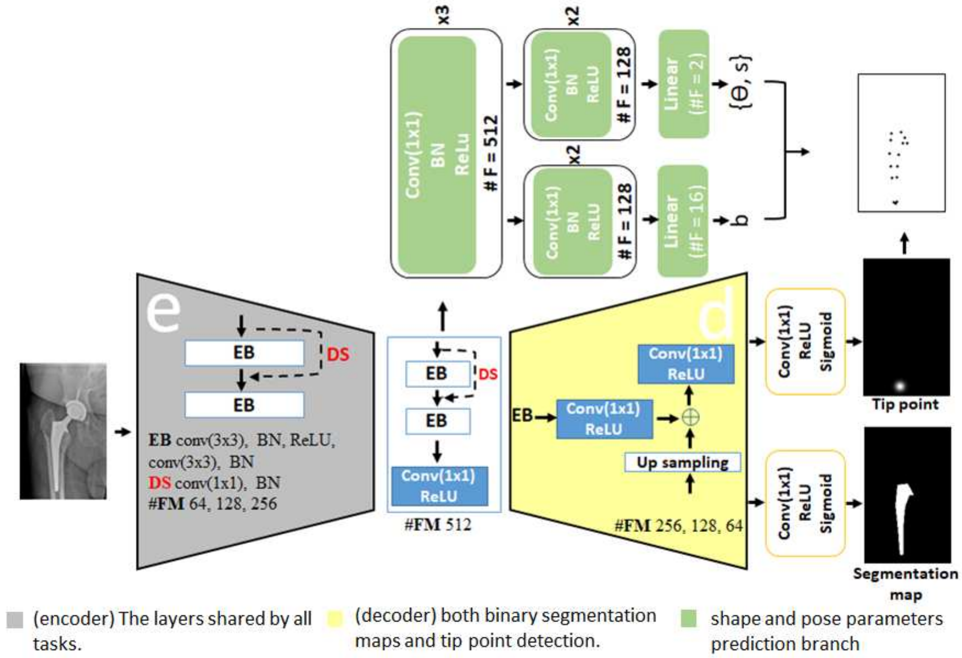


Fig. 3. The proposed GruenNet architecture. The encoder block (EB) consisted of 3×3 convolutional (conv) layer, batch normalization (BN), a parameterized rectified linear unit (ReLU) followed by another conv and BN. The Down sampling (SD) consisted of conv followed by BN. The number of feature maps (#FM) is presented for each block.

shape variations. The interpretation of a new image requires identifying the parameters that best match the model to the image. An accurate SSM requires correspondence mapping between shape landmarks. We define these landmarks using the Gruen zones.

The localization of these zones simplifies the analysis of the surrounding region of the implant which, consequently, approximates the shape of the implant and localizes the important landmarks.

Fig. 1 (c) shows a comprehensive description of the defined landmarks. Additional landmarks within each zone were added to accurately represent the shape of the implant.

After defining the shape landmarks, as a set of N connected landmarks $x = (x_1, ...x_N, y_1, ...y_N)$, the SSM can be constructed using the following steps. First, the mean shape is computed by aligning all the training S shapes together using Generalized Procrustes Analysis (GPA). GPA is an iterative method that starts by selecting a random shape from the training set as a mean shape. All shapes are aligned with reference to the mean shape, which is re-estimated and the alignment is repeated. The process ends when the estimated mean shape is equal to the previous one. The resulting aligned shape is defined as:

$$\begin{pmatrix} x^{'} \\ y^{'} \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + \begin{pmatrix} \cos\theta & -\sin\theta \\ s\sin\theta & s\cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{1}$$

Where (t_x, t_y), θ and s are the pose parameters (translation, rotation and scaling). The average shape can be estimated by:

$$\bar{x} = \frac{1}{S}\sum_{i=1}^{S} x_i^{'} \tag{2}$$

Where $x_i^{'}$ denotes the aligned shape vector and $i \in \{1, 2, ..S\}$. This process is presented in Fig. 2. The S samples of the training set are shown in Fig. 2(a) whereas the aligned shapes x_i are shown in Fig. 2 (B) and the mean shape $\bar{x}$ is presented in Fig. 2(C).

Then, the Principal Component Analysis (PCA) is applied to obtain shape variations. Given a set of shape vectors $\{x_i^{'}\}$, the mean shape is computed by using (2), and the covariance of the data is computed by:

$$C = \frac{1}{S-1}\sum_{i=1}^{S}(x_i - \bar{x})(x_i - \bar{x})^T \tag{3}$$

The eigenvectors $P = \{p_1, p_2, ...p_t\}$ and corresponding eigenvalues λ_t of C represent the directions of variation in the data about the mean. The first M largest eigenvalues are chosen such that:

$$\sum_{i=1}^{M} \lambda_i \geq f_v V_T \tag{4}$$

where f_v defines the proportion of the total variation V_T. Assuming that the shape follows a Gaussian probability distribution, the shape can be approximated using:

$$x \approx \bar{x} + Pb \tag{5}$$

where P contains the first m eigenvectors and b is a m dimensional vector given by:

$$b = P^T(x - \bar{x}) \tag{6}$$

C. Dataset Pre-processing

All images are resized to 224×224 px and are normalized by dividing by the largest pixel value (255). Pose parameters (θ and s) and b-coefficients are normalized by min-max feature scaling to values between 0 and 1. The tip point position is used to generate a heat-map image of size 224×224 by using a Gaussian kernel with $\sigma = 5$.

Online data augmentation is used to increase the dataset size. This is computed by first applying random transformations to the shape parameters as the following: the shape coefficients b are modified by adding a random uniform value $b_{aug} = b + a$ where $a \in [-2, 2]$, random shape rotation $\theta \in [-60, 60]$ and translation by a random value between [-10,10]. The images are transformed according to the computed augmented shape using the Thin Plate Spline Transformation method [27]. The masks and heat-maps are created respectively. In addition, brightness variation $[-0.2, 0.2]$ is applied for augmentation.

D. Gruen Net

The proposed Gruen network architecture for detecting Gruen landmarks and performing implant segmentation is presented in Fig. 3. The input to the network is the X-ray image and it has four outputs: (1) shape parameters b_m. (2) pose parameters θ and scale s. (3) implant tip point (c_x, c_y). (4) segmentation maps. The proposed architecture consisted of two branches; the green branch, which is responsible for learning b_m, θ and s, and the yellow branch which learns the binary segmentation map and the tip point heatmap. The grey layers are shared by all tasks. Semantic segmentation and tip point prediction share the same features that are conducted by an encoder-decoder to infer the probability label map. The encoder part includes three residual blocks that consist of two convolution layers with a kernel size of 3 × 3. Each convolution layer is followed by batch normalization and ReLu activation function. The encoder is followed by the bridge part which consists of one residual block. The decoder part uses both the features map from the bridge and the skip connections from different encoder blocks to learn the binary classification of each pixel for both segmentation and tip point localization tasks. Finally, the task-specific layers which consisted of convolution and ReLu layers followed by a sigmoid function are added to the network architecture. The regression of the SSM parameters branch starts from the bridge block. It shares three convolution layers and has specific two convolution layers and a Linear layer. Each convolution is followed by batch normalization and ReLu layers.

The network is trained using a weighted sum of multiple loss functions ($L_b, L_{\theta,s}, L_{sh}, L_{c_x,x_y}$ and L_{seg}). The shape parameters loss (L_b) is defined as the Mean Squared Error (MSE) between the ground truth shape parameters ($b_{i,true}$) and the predicted one ($b_{i,pred}$):

$$L_b = \frac{1}{N}\sum_{i=1}^{N}(b_{i,true} - b_{i,pred})^2 \tag{7}$$

The pose parameters loss ($L_{\theta,s}$) are defined as the sum of MSE loss between the ground truth ($\theta_{i,true}, s_{i,true}$) and predicted orientation and scale ($\theta_{i,pred}, s_{i,pred}$):

$$L_{\theta,s} = \frac{1}{N}(\sum_{i=1}^{N}(\theta_{i,true} - \theta_{i,pred})^2 + \sum_{i=1}^{N}(s_{i,true} - s_{i,pred})^2) \tag{8}$$

The heatmap regression is employed to detect the tip point. For each image, a heatmap image is formed using a Gaussian filter that is centred at the tip point location. The heatmap loss (L_{hm}) is defined using the Cross-Entropy (CE) loss function as:

$$L_{hm} = hm_{true} \cdot \log hm_{pred} + (1 - hm_{true}) \cdot \log(1 - hm_{pred}) \tag{9}$$

where hm_{true} is the ground truth label and hm_{pred} is the predicted probability of the point being tip point.

The implant shape is computed using the shape parameters (b_i, θ_i, s_i, and c_x, c_y) as described in section III-B. The shape loss (L_{sh}) is calculated by the MSE between the predicted shape ($sh_{i,pred}$)) and the ground truth shape ($sh_{i,true}$)):

$$L_{sh} = 1/N\sum_{i=1}^{N}(sh_{i,true} - sh_{i,pred})^2 \tag{10}$$

The binary segmentation loss is defined using the CE loss function:

$$L_{seg} = y_{true} \cdot \log y_{pred} + (1 - y_{true}) \cdot \log(1 - y_{pred}) \tag{11}$$

where y_{true} is the ground truth label and y_{pred} is the predicted probability.

IV. Experimental Settings

Multiple experiments have been carried out to validate the proposed method and the effect of each parameter. In addition, different loss functions and hyper-parameters have been explored to obtain the best results. For simplicity, BSM is referred to the segmentation resulting from the binary segmentation map and SP is referred to the segmentation constructed from the prediction of shape coefficients, pose parameters and tip points detection.

The purpose of the first experiment is to assess the performance of each task separately; (1) The prediction of a BSM of the image for semantic segmentation. (2) the prediction of the SP for segmentation and landmarks localization. The BSM was achieved by training the main branch of the proposed model (Fig. 3 the grey and yellow parts for segmentation task only), while the SP predictions were learned by training the grey and green part and yellow part for heatmap prediction. The effect of the data augmentation was also investigated on both tasks.

The second experiment is to evaluate the performance when combining both semantic segmentation and shape and pose parameters prediction in the learning process. In this experiment, we study also the effect of adding shape loss (L_{sh}) that is computed from the shape and pose parameters. The last experiment will study the impact of employing the ICP method to align the segmentation map with the predicted landmarks.

A. Dataset

To increase the variability of X-ray images, two different hip implant datasets were utilized to construct, train and validate the proposed method: Orthonet dataset [6] and in-house dataset [28]. Orthonet dataset is a publicly available dataset that was originally collected for the classification of the implant model type in knee and hip arthroplasty. It consisted of 1191 unilateral hip X-ray images with 8 different models of implant. Part of this dataset (198 images) was intended for implant segmentation. So, it includes the original x-ray images and the implant mask images. The images have various sizes and all the images represent the normal status of the implant. More details about this dataset and the generation of the implant masks can be found in [6]. The in-house dataset was generated for automated peri-prosthetic femur fracture diagnosis [28]. It consisted of X-ray images after the THR, which is considered normal cases, and X-ray images with various types of fractures. More details about this dataset can be found in [28].

Due to difficulties of manual annotation of the ground truth, this work has included part of both dataset. A total number of 330 images were used for training and validation of the proposed method. From Orthonet data, approximately, 30 images were randomly selected from each implant model. The remaining images were selected from the in-house data. The choice of in-house images was based on the fracture type. The fracture types B1, B2 and B3 occur within the implant region. Therefore, the images were randomly selected from these types (approximately 30 images per type). Table I demonstrates the distribution of the dataset.

Ground truth segmentations of implant femoral component and the SSM landmarks were annotated by a clinical expert using the Microsoft VOTT tool. The landmarks were annotated as described in Fig. 1 (c). Landmarks (2, 4, 6, 8, 10, 12, 14) are the Gruen zone landmarks, while the other points are added to define the implant boundary precisely. The implant masks were generated by filling the area of the defined shape.

TABLE I
DISTRIBUTION OF THE DATASET.

Dataset	Type	#images	Example
Orthonet	Depuy-Synthes Corail (with collar)	29	
	Depuy-Synthes Corail (no collar)	30	
	JRI Orthopaedics Furlong Evolution (with collar)	29	
	JRI Orthopaedics Furlong Evolution (no collar)	27	
	Smith & Nephew Anthology	30	
	Smith & Nephew Polarstem (no collar)	29	
	Stryker Accolade II	30	
	Stryker Exeter	30	
In-house	Fracture Type B1	30	
	Fracture Type B2	30	
	Fracture Type B3	29	

B. Implementation details

The femoral stem is represented by $N = 15$ landmarks and (θ, t_x, t_y, s) are computed as explained in Section III-B. The shape model has $M = 15$ modes of shape variation which explains 98% of shape variation. Fig. 4 shows examples of the shape variations related to the first 15 eigenmodes of the implant.

The dataset was divided into two parts: training and validation, with the ratio $75\% : 25\%$, respectively. Different augmentation methods have been applied as explained in section III-C to the dataset to minimize the effect of the small dataset size.

The network was trained on a Windows machine equipped with 8 GB RAM, Intel(R) Core(TM) CPU @ 3.00 GHz and GeForce RTX 2080 graphics card. It is trained over 200 epoch with AdamW optimizer, learning rate 1×10^{-4}, weight decay 5^{-4} and batch size = 8.

C. Evaluation settings

Multiple evaluation metrics were used to validate the proposed method. As explained earlier, we evaluated the accuracy of the femoral stem segmentation for both outcomes; BSM and SP. Dice coefficient and Hausdorff distance were used to evaluate the segmentation results.

Additionally, the performance of the predicted shape coefficients, pose parameters and tip point prediction were evaluated. For the pose parameters evaluation, the absolute error was utilized where the orientation error is defined as $\delta\theta = |\theta_{pred} - \theta_{true}|$ and the scale error is defined as $\delta s = |s_{pred} - s_{true}|$. The Euclidean distance was used to validate the tip point prediction. In addition, the impact of each parameter on the construction of the shape landmarks is analysed by taking into account the ground truth of all parameters except the studied one.

The shape landmarks were assessed using the Normalized Root Mean Square Error (NRMSE). NRMSE measures the average distance between the predicted and the ground truth

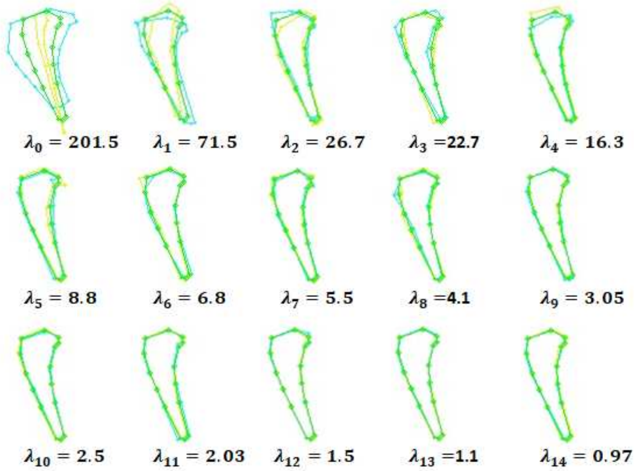


Fig. 4. 15 modes of shape variations. Green represents the mean model. Yellow represents the deformed shape by $-3\sqrt{\lambda_i}$ and blue represents the deformed shape by $3\sqrt{\lambda_i}$

TABLE II
DICE AND HD RESULTS FOR SEGMENTATION COMPUTED FROM BSM (UPPER ROW) AND SEGMENTATION COMPUTED FROM THE CONSTRUCTED IMPLANT SHAPE SP (BOTTOM ROW) IN THE ABLATION STUDIES. THE BEST RESULTS ARE HIGHLIGHTED.

Experiment	Dice (%)	HD (px)
U-Net	74 ± 13.3	16 ± 23.4
BSM	78 ± 23.3	20 ± 23.6
BSM + A	79 ± 24	12 ± 17.9
BSM + SP + A	77.7 ± 22.7	10 ± 14
SBSM + A + L$_{sh}$	**80 ± 22**	**8.8 ± 10.7**
SP$_{TXY}$	22.04 ± 24.11	34.7 ± 13
SP$_{HM}$	56.7 ± 16.8	26 ± 30.6
SP + A	57.4 ± 17.7	24.8 ± 23.4
SP + BSM + A	62 ± 15.7	20 ± 15.7
SP + BSM + A + ICP	66.7 ± 17.7	17.5 ± 17.6
SP + BSM + A + L$_{sh}$	62 ± 15.2	20.4 ± 9
SP + BSM + A + L$_{sh}$ + ICP	**69.7 ± 16.7**	**16.8 ± 9.8**

landmarks normalized by the distance between two adjacent ground truth landmarks (x_{i-1}, x_{i+1})

$$NRMSE = \frac{1}{N}\sum_{i=1}^{N}\frac{\sqrt{(x_i^p - x_i^t)^2 + (y_i^p - y_i^t)^2}}{\sqrt{(x_{i-1}^t - x_{i+1}^t)^2 + (y_{i-1}^t - y_{i+1}^t)^2}} \tag{12}$$

Where N is the number of the landmarks, (x_i^p, y_i^p) is the predicted landmark and (x_i^t, y_i^t) is the corresponding ground truth landmark. Furthermore, the cumulative error distribution (CED) was utilized to assess the detection of the landmarks. CED plots the cumulative NRMSE against the proportion of images with an NRMSE of less than or equal to a particular value.

The performance of using augmentation, adding shape loss and applying the ICP algorithm was validated by Dice coefficient, Hausdorff and NRMSE.

V. RESULTS

A. Ablation studies

In this paper, we integrated implant shape into a deep learning model to segment the implant and detect the Gruen landmarks. To demonstrate the effectiveness of our proposed method, we performed ablation experiments on the THR dataset. The results in Table II presented the validity of our proposed method. The upper rows in the table showed the segmentation result computed from the BSM component, while the bottom rows showed the segmentation result computed from the predicted shape and pose parameters SP. For simplicity, $\mathbf{A}$ represents the data augmentation, $\mathbf{L}_{sh}$ represents the shape loss. For the BSM task, the proposed model provided better segmentation results compared to U-net with a dice score of 78%. The performance was further improved when introducing the data augmentation with a dice score of 79% and HD of 12 px. The segmentation did not improve when joining the shape parameters prediction component in the training, however, introducing (L_{sh}) resulted in the best segmentation performance with a dice score of 80% and HD of 8.8 px. Fig. 5 illustrated examples of the binary segmentation results compared to the ground truth segmentation in different experimental settings. Also, the dice score is reported for each image. The predicted segmentation appeared disconnected when utilizing the BSM only, while the shape tends to be connected when joining the regression of the shape parameters, specifically when adding L_{sh}.

The bottom rows of Table II demonstrated the segmentation results computed from SP task. Two experiments were carried out to compute the implant shape. The first experiment regresses the translation, rotation, scale and shape parameters to compute the implant shape. For simplicity, we denoted this experiment as $\mathbf{SP}_{TXY}$. The second experiment differs from the first one by the computation of the translation parameter which is computed based on the position of the implant tip point. The tip point is predicted using the heatmap regression. We denoted this experiment as $\mathbf{SP}_{HM}$. The regression of shape and pose parameters only including the regression of the translation parameters (SP_{TXY}) produced poor segmentation results. The performance is enhanced significantly (by 34% dice score) when utilizing the tip point to calculate the translation parameter (SP_{HM}). The performance is further improved by adding data augmentation. When joining the BSM, the segmentation performance was improved in both metrics (Dice = 62% and HD = 20 px). On the other hand, the results have not changed when introducing the shape loss. Applying the ICP algorithm to align the predicted shape to the BSM results produced better shape segmentation with dice = 69.7% and HD = 16.8 px. Fig. 6 showed some examples in different experiments for the segmentation using the predicted shape. Additionally, a dice score is reported for each example. The shape results were improved with each change to the training method. Furthermore, it is illustrated in the images that when aligning the shape to the BSM the shape outcome is enhanced.

Table III listed a further validation to the predicted shape experiments by reporting the pose parameters (θ and s) errors, the implant tip point detection error and the constructed shape landmarks error. Regression of pose and shape parameters (SP_{TXY}) provided the best rotation and scale outcomes with $\Delta\theta = 4.48°$ and $\Delta s = 0.13$. The same scale error was produced when the BSM was combined with the training

TABLE III

MEAN AND STANDARD DEVIATION FOR ORIENTATION ERROR, SCALE ERROR, TIP POINT EUCLIDEAN DISTANCE, THE NRMSE FOR THE SHAPE LANDMARKS AND AFTER APPLYING ICP METHOD. THE BEST RESULTS ARE HIGHLIGHTED.

Experiment	$\theta(°)$	scale	Tip point (px)	Landmarks (px)	ICP (px)
SP_{TXY}	**4.48 ± 3.24**	**0.13 ± 0.09**	88.07 ± 19.73	1.54 ± 0.98	-
SP_{HM}	5.80 ± 4.45	0.16 ± 0.12	5.11 ± 31.31	0.80 ± 1.44	-
SP + A	6.09 ± 4.34	0.14 ± 0.10	3.46 ± 23.09	0.71 ± 1.13	-
SP + BSM + A	4.78 ± 4.32	0.14 ± 0.10	2.17 ± 8.28	0.57 ± 0.42	0.36 ± 0.27
SP + BSM + A + L_{Sh}	5.41 ± 4.23	**0.13 ± 0.10**	**1.29 ± 0.94**	**0.55 ± 0.30**	**0.33 ± 0.20**

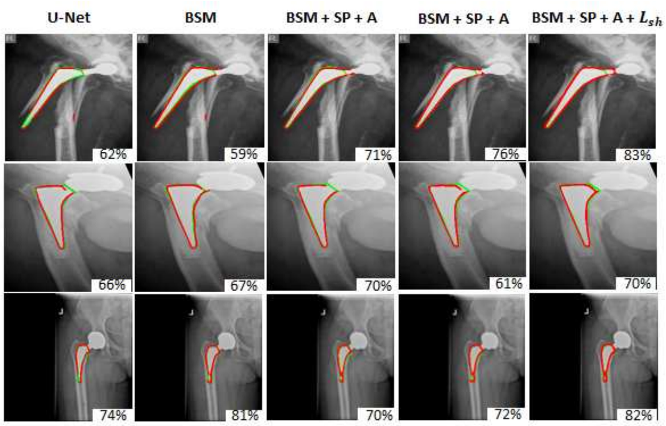


Fig. 5. Comparison of segmentation computed from BSM in ablation studies. The red is the predicted segmentation and the green is the ground truth. The dice score is presented in each image

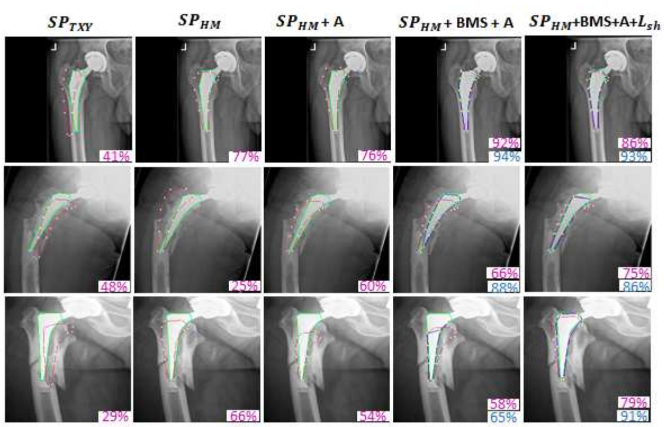


Fig. 6. Comparison of segmentation computed from SP in ablation studies. The green is the ground truth, the pink is the computed shape and the blue is the shape after applying the ICP algorithm. The dice score is presented in each image.

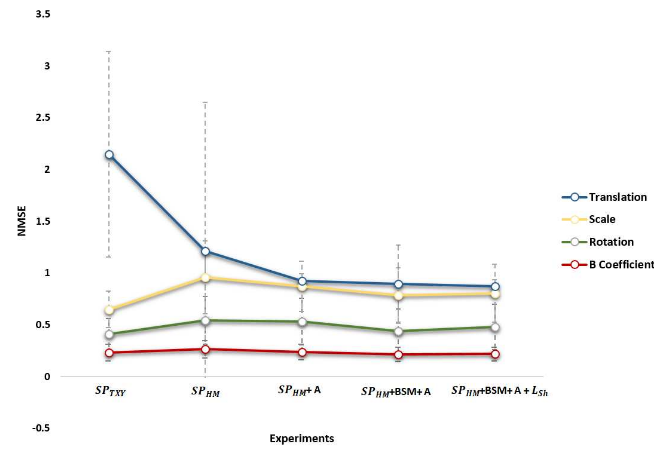


Fig. 7. The impact of error in translation, rotation, scale and B-coefficient on the computation of the implant shape landmarks. Each plot represents the mean NMSE for the shape computed by fixing all parameters as ground truth values except the studied parameter where the predicted value is used.

and the shape loss was added. However, we observed that the error difference among experiments for both orientation and scale parameters was slightly low, demonstrating that these parameters did not benefit from combined semantic segmentation to some extent.

We measured the translation error using the distance between the predicted implant tip point and the ground truth point. The results demonstrated that the translation parameter has improved significantly with each modification to the training method and provided the best result when the BSM joined the training and the L_{sh} is applied. The regression of translation parameters in the first experiment produced a large error. Introducing the tip point heatmap prediction to compute the translation parameters has improved the results from 88 px to 5.11 px. Similarly, the shape landmarks have improved in each alteration and the best outcome has resulted from the last experiment (SP + BSM + A + L_{Sh}). The shape landmarks have been considerably enhanced by aligning the constructed shape to the predicted segmentation, which has reduced the error by 0.22 px.

In addition, we studied the impact of the error in each shape component i.e translation, rotation, scale and B-coefficient to the final reconstruction of the shape landmarks. To study the impact, shape landmarks are constructed by fixing the values of all shape parameters to the ground-truth value except the parameter under investigation which involved the predicted value. Fig. 7 presented the NRMSE between the ground truth landmarks and the computed shape. The figure indicated that the landmarks error resulting from the error in the translation parameter has improved significantly in each modification. In addition, the B-coefficient error indicated a slight enhancement to the landmarks error. On the other hand, scale and translation errors have a major impact on the landmark error compared to the other parameters.

Furthermore, we summarised the performance of landmarks detection using the CED curves. It can be seen in Fig. 8 that in both experiments i.e. using the simultaneous training method with data augmentation and by adding shape loss the localization of the landmarks, 80% of the images are below

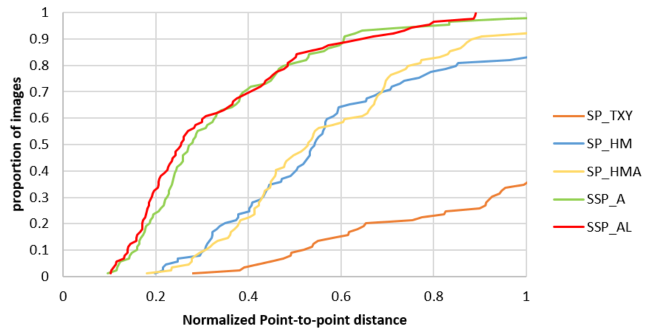


Fig. 8. cumulative error distributions. Comparing the performance of each experiment using point-to-point distance normalized by the two adjacent point distance.

TABLE IV
QUANTITATIVE RESULTS FOR IMPLANT SEGMENTATION ON OUR DATASET. BEST RESULTS ARE IN BOLD

Method	Dice (%)	HD (px)
UNet [29]	74.0 ± 13.3	16.0 ± 23.4
Res-Unet [30]	72.3 ± 12.0	17.5 ± 25.3
UNet ++ [31]	70.3 ± 13.1	33.0 ± 42.0
Attention UNet [32]	69.0 ± 16.0	44.2 ± 48.1
R2UNet [33]	48.2 ± 17.7	34.0 ± 25.9
CE-Net [34]	55.5 ± 5.50	135 ± 24.1
U2Net [35]	57.1 ± 9.32	124 ± 20.6
Our method	**80.0 ± 22.0**	**8.80 ± 10.7**

0.5 NRMSE. On the other hand, $\sim$ 40% of images are below 0.5 NRMSE using the prediction of shape parameters only (SP$_{HM}$+A and SP$_{HM}$+A $+L_{Sh}$). In addition, the maximum error produced by the simultaneous training method is lower than in other experiments.

B. Experimental Comparison on Test Dataset

To validate the advantages of the proposed method, state-of-the-art networks for both medical image segmentation and landmarks detection were considered as a comparison strategy. Seven state-of-the-art networks were utilized to compare the segmentation results: UNet [29], Res-Unet [30], UNet++, [31], Attention Unet [32] , R2Unet [33], CE-Net [34] , U2-net [35]. Table IV listed the results of the implant segmentation using different segmentation networks. We can observe that with a small-size dataset, the complex model might be more prone to overfitting, introduce complexity and cannot generalise from limited training samples. UNet tends to be a better solution because it has a relatively smaller number of parameters compared to other variants. However, employing shape priors has significantly improved the results to 80% dice score.

Regarding the Gruen landmarks detection, we compared our method with different CNN-based networks (UNet, ResNet50 [36], VGG16 [37], DenseNet121 [38] and SwinNet [39]) for predicting the landmarks as direct regression of the points or as heatmap regression. Table V listed the NRMSE of each tested model. The results indicated that our method improved landmarks detection significantly.

VI. DISCUSSIONS

.

TABLE V
QUANTITATIVE RESULTS FOR GRUEN LANDMARKS DETECTION ON OUR DATASET. THE BEST RESULTS ARE IN BOLD

Method	NMRSE (px)
UNet	3.21 ± 1.02
VGG16	3.06 ± 1.35
DenseNet121	2.78 ± 1.07
ResNet50	2.90 ± 1.14
SwinNet	3.00 ± 1.28
Our method	**0.55 ± 0.30**

Recent survey demonstrated that combining deep learning with medical knowledge has a huge impact on the outcomes of several medical image analysis tasks, including segmentation and diagnosis [40]. Therefore we adopt this strategy for implant joint images domain aiming to automate the segmentation of the implant and detecting the Gruen landmarks. Despite the challenges imposed by a limited dataset, incorporating implant shape knowledge into the CNN shows precise and valid implant segmentation and Gruen landmarks detection.

In this paper, we defined the implant shape using the Gruen landmarks definition and presented a deep learning method to predict the shape and pose parameters of the implant femoral component and perform its semantic segmentation. Compared to typical semantic segmentation where each pixel is binary classified, this approach predicts the shape and pose parameters which link to landmarks representation that can be used in many diagnostic tasks. Diagnosing implant complications depends mainly on the position in relation to the implant. Therefore, we defined the shape landmarks based on Gruen zones, to combine the advantage of both the segmentation and the detection of important landmarks. This is the first algorithm that can detect locations of the important landmarks and segment the femoral component. This has been successfully demonstrated through the comparison of the segmentation and landmarks detection results with the stste-of-the-art segmentation models and landmarks detection models. The landmarks localisation results could be considered state-of-the-art results. The dataset used in this work will be publicly available to enhance the research on this domain.

The results of the proposed approach indicated that the regression of the shape and pose parameters is a more challenging process compared to semantic segmentation. The shape and pose parameters regression is performed by training on and predicting a small number of uncorrelated values (19 values) per image using a limited-size dataset, whereas the semantic segmentation is predicted based on a large number of correlated values per image. Replacing the translation parameter regression with translation computed from the prediction of the tip point position has substantially enhanced the shape outcomes. The shape-based data augmentation is used to increase the size of the training dataset. Although orientation prediction did not benefit from the data augmentation, the prediction of the other parameters has improved which impacts positively on the computation of the position of the landmarks.

Combining the training of semantic segmentation with the shape and pose parameters regression has enhanced the outcomes of the segmentation using the constructed shape (see

TableII). The shared layers between the two tasks enable the learning of more relevant geometric features. We hypothesised that introducing the shape loss will impact the segmentation output of both tasks. It has enhanced the semantic segmentation performance which also makes indirect benefit to the segmentation based on the shape construction by aligning the resulting shape to the semantic segmentation outcome.

This paper focused on the segmentation and landmark detection of the implant femoral components, however, this method can be extended to other implant joints.

VII. Conclusion

In this paper, we proposed a new CNN approach for jointly segmenting the implant femoral component and regression of the pose and shape parameters. The implant landmarks' positions are computed from the predicted shape and pose parameters. Experiments demonstrated that combining semantic segmentation has enhanced the overall outcomes of the shape landmarks localisation. Results show that our method is accurate with an overall segmentation dice score of 80% and HD of 8.8 px. In addition, this work reported the state-of-the-art result of Gruen landmarks localisation with NRMSE of 0.33. Future work will consider extending this approach to other implant joints and utilise it as an initial stage for analysis of femur implant complications.

Acknowledgment

Professor Pandit is a National Institute for Health Research (NIHR) Senior Investigator. This study/research is funded by the NIHR Leeds Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

References

[1] United Kingdom National Joint Registry, "2020 17th Annual Report," tech. rep., 2020.

[2] M. Stark, *Automatic detection and segmentation of shoulder implants in x-ray images*. PhD thesis, Master's Thesis, San Francisco State University, San Francisco, CA, USA ..., 2018.

[3] N. Al-Zadjali, *Computer-aided diagnosis of complications of total hip replacement X-ray images*. PhD thesis, Loughborough University, 2017.

[4] A. Oprea and C. Vertan, "A quantitative evaluation of the hip prosthesis segmentation quality in x-ray images," in *2007 International Symposium on Signals, Circuits and Systems*, vol. 1, pp. 1–4, IEEE, 2007.

[5] L. Florea, C. Florea, C. Vertan, and A. Sultana, "Automatic tools for diagnosis support of total hip replacement follow-up," *Advances in Electrical and Computer Engineering*, vol. 11, no. 4, pp. 55–62, 2011.

[6] R. Patel, E. H. Thong, V. Batta, A. A. Bharath, D. Francis, and J. Howard, "Automated identification of orthopedic implants on radiographs using deep learning," *Radiology: Artificial Intelligence*, vol. 3, no. 4, p. e200183, 2021.

[7] Y. Skandarani, P.-M. Jodoin, and A. Lalande, "Gans for medical image synthesis: An empirical study," *Journal of Imaging*, vol. 9, no. 3, 2023.

[8] D. O. Medley, C. Santiago, and J. C. Nascimento, "Deep active shape model for robust object fitting," *IEEE Transactions on Image Processing*, vol. 29, pp. 2380–2394, 2019.

[9] D. H. Nguyen, D. M. Nguyen, T. T. Mai, T. Nguyen, K. T. Tran, A. T. Nguyen, B. T. Pham, and B. T. Nguyen, "Asmcnn: An efficient brain extraction using active shape model and convolutional neural networks," *Information Sciences*, vol. 591, pp. 25–48, 2022.

[10] Z. Mirikharaji and G. Hamarneh, "Star shape prior in fully convolutional networks for skin lesion segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds.), (Cham), pp. 737–745, Springer International Publishing, 2018.

[11] C. Qin, P. Tu, X. Chen, and J. Troccaz, "A novel registration-based algorithm for prostate segmentation via the combination of ssm and cnn," *Medical Physics*, 2022.

[12] S. Tilborghs, T. Dresselaers, P. Claus, J. Bogaert, and F. Maes, "Shape constrained cnn for cardiac mr segmentation with simultaneous prediction of shape and pose parameters," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 127–136, Springer, 2020.

[13] D. Karimi, G. Samei, C. Kesch, G. Nir, and S. E. Salcudean, "Prostate segmentation in mri using a convolutional neural network architecture and training strategy based on statistical shape models," *International journal of computer assisted radiology and surgery*, vol. 13, no. 8, pp. 1211–1219, 2018.

[14] T. A. M. GRUEN, G. M. MCNEICE, and H. C. M. AMSTUTZ, "'modes of failure' of cemented stem-type femoral components: A radiographic analysis of loosening.," *Clinical Orthopaedics and Related Research*, pp. 17–27, 1979.

[15] W.-Y. Hsu, "Automatic left ventricle recognition, segmentation and tracking in cardiac ultrasound image sequences," *IEEE Access*, vol. 7, pp. 140524–140533, 2019.

[16] P. R. Tabrizi, A. Mansoor, J. J. Cerrolaza, J. Jago, and M. G. Linguraru, "Automatic kidney segmentation in 3d pediatric ultrasound images using deep neural networks and weighted fuzzy active shape model," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1170–1173, IEEE, 2018.

[17] Y. Li, C. P. Ho, M. Toulemonde, N. Chahal, R. Senior, and M.-X. Tang, "Fully automatic myocardial segmentation of contrast echocardiography sequence using random forests guided by shape model," *IEEE transactions on medical imaging*, vol. 37, no. 5, pp. 1081–1091, 2017.

[18] C. Zotti, Z. Luo, A. Lalande, and P.-M. Jodoin, "Convolutional neural network with shape prior applied to cardiac mri segmentation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1119–1128, 2018.

[19] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative," *Medical image analysis*, vol. 52, pp. 109–118, 2019.

[20] I. Brusini, O. Lindberg, J.-S. Muehlboeck, Ö. Smedby, E. Westman, and C. Wang, "Shape information improves the cross-cohort performance of deep learning-based segmentation of the hippocampus," *Frontiers in neuroscience*, vol. 14, p. 15, 2020.

[21] J. Duan, G. Bello, J. Schlemper, W. Bai, T. J. Dawes, C. Biffi, A. de Marvao, G. Doumoud, D. P. O'Regan, and D. Rueckert, "Automatic 3d biventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach," *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2151–2164, 2019.

[22] A. J. Larrazabal, C. Martinez, and E. Ferrante, "Anatomical priors for image segmentation via post-processing with denoising autoencoders," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 585–593, Springer, 2019.

[23] N. Painchaud, Y. Skandarani, T. Judge, O. Bernard, A. Lalande, and P.-M. Jodoin, "Cardiac segmentation with strong anatomical guarantees," *IEEE transactions on medical imaging*, vol. 39, no. 11, pp. 3703–3713, 2020.

[24] C. Chen, C. Biffi, G. Tarroni, S. Petersen, W. Bai, and D. Rueckert, "Learning shape priors for robust cardiac mr segmentation from multi-view images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 523–531, Springer, 2019.

[25] J. Schock, M. Kopaczka, B. Agthe, J. Huang, P. Kruse, D. Truhn, S. Conrad, G. Antoch, C. Kuhl, S. Nebelung, *et al.*, "A method for semantic knee bone and cartilage segmentation with deep 3d shape fitting using data from the osteoarthritis initiative," in *International Workshop on Shape in Medical Imaging*, pp. 85–94, Springer, 2020.

[26] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[27] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[28] A. Alzaid, A. Wignall, S. Dogramadzi, H. Pandit, and S. Q. Xie, "Automatic detection and classification of peri-prosthetic femur fracture,"

International Journal of Computer Assisted Radiology and Surgery, vol. 17, no. 4, pp. 649–660, 2022.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241, Springer, 2015.

[30] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[31] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pp. 3–11, Springer, 2018.

[32] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[33] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation," *arXiv preprint arXiv:1802.06955*, 2018.

[34] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

[35] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-net: Going deeper with nested u-structure for salient object detection," *Pattern recognition*, vol. 106, p. 107404, 2020.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[38] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[39] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021.

[40] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Medical Image Analysis*, vol. 69, p. 101985, 2021.