IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION

Received 2 April 2024, accepted 21 April 2024, date of publication 29 April 2024, date of current version 6 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3394895

APPLIED RESEARCH

Segmentation-Guided Coordinate Regression for Robust Landmark Detection on X-Rays: Application to Automated Assessment of Lower Limb Alignment

SEBASTIAN AMADOR SANCHEZ[1,2], PHILIPPE VAN OVERSCHELDE[3], AND JEF VANDEMEULEBROUCKE[1,2,4]

[1]Department of Electronics and Informatics, Vrije Universiteit Brussel, 1050 Brussels, Belgium
[2]imec, 3001 Leuven, Belgium
[3]moveUP, 9000 Ghent, Belgium
[4]Department of Radiology, Universitair Ziekenhuis Brussel, 1090 Brussels, Belgium

Corresponding author: Sebastian Amador Sanchez (sebastian.amador.sanchez@vub.be)

This work was supported by the Innoviris Grant of the Brussels-Capital Region, as part of the project AugmeNTed IntelligenCe In orthopaedics TrEatments (ANTICIPATE) on Augmented Intelligence in Orthopaedics Treatments, under Grant BHF/2020-RDIR-6a.

ABSTRACT In medical imaging, automated landmark detection estimates the position of anatomical points in images to derive measurements. Previous approaches commonly employ coordinate regression. Landmark segmentation, a technique in which masks centered at the target point are segmented, has recently shown promising results. Here, we present segmentation-guided coordinate regression, a methodology that fuses both approaches and balances accuracy and robustness. Our approach identifies masks centered at landmarks using a segmentation network. Then, a coordinate regression network estimates the coordinates by employing the input image and the segmentation output. We assessed the methodology's performance by detecting eight landmarks in full lower limb X-rays and investigated the impact of weight initialization, network backbone, and optimization of the loss function. The approach was contrasted with landmark segmentation and coordinate regression and applied to the analysis of lower limb malalignment. Results showed that deeper pretrained models with a weight of 0.2 at the segmentation loss detected landmarks more accurately. Segmentation-guided regression outperformed coordinate regression. Landmark segmentation was hampered by undetected landmarks and false positives. Due to its architecture, the proposed method did not suffer from failed detections, allowing lower limb malalignment to be reliably calculated. With respect to comparable literature, our approach leads to similar or improved results for landmark detection, translating to highly accurate and reliable lower limb malalignment analysis. In conclusion, we proposed a novel method for detecting landmarks in X-rays, which leads to a balance in accuracy and robustness and allows the measurement of lower limb malalignment.

INDEX TERMS Deep learning, landmark detection, X-rays, lower limb malalignment.

I. INTRODUCTION

Accurate identification of anatomical landmarks is a crucial process in medical image analysis. It is often the first step of a more extensive clinical analysis, supporting the measurement of clinical parameters, diagnosis or therapy planning [1]. It may also serve as initialization to other image analysis algorithms such as segmentation or registration [2]. While manual identification may seem trivial; such a process is considered cumbersome, time-consuming, and subject to

The associate editor coordinating the review of this manuscript and approving it for publication was Chulhong Kim.

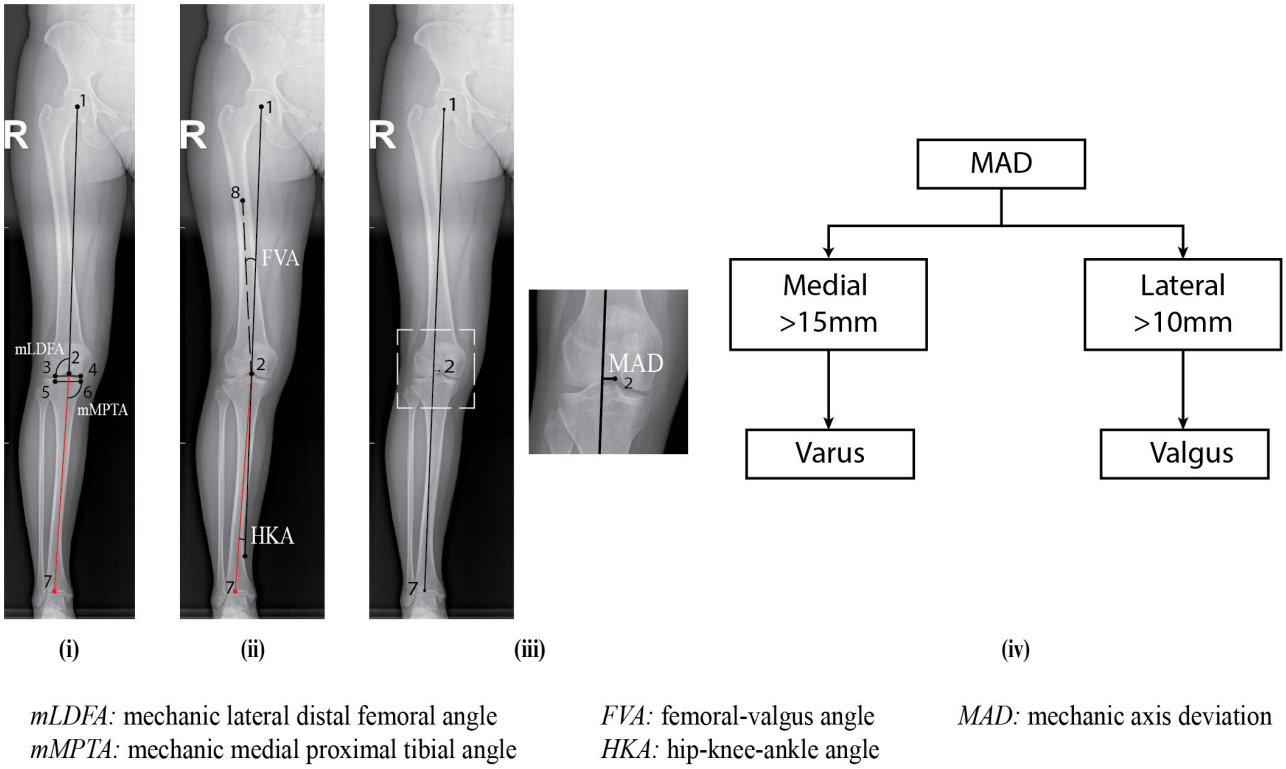

FIGURE 1. Angles and distances measured in a lower leg malalignment test. (i) ***mLDFA:*** Angle measured between the axis that runs from the head of the femur center to the center of the knee (1→2), and the line that runs tangent to the femoral condyles (3→4). ***mMPTA:*** Angle formed between the axis that connects the center of the knee and the center of the ankle (2→7), and the tangent line that joins the tibial plateaus (5→6). (ii) ***FVA:*** Angle between the axis that connects the head of the femur center and the center of the knee (1→2), and the axis that runs from the middle of the diaphysis to the center of the knee (8→2). ***HKA:*** Angle between the extended axis that joins the head of the femur center and the center of the knee (1→2), and the axis formed by connecting the center of the knee and the ankle(2→7). (iii) ***MAD:*** Distance from the center of the knee (2) to the axis that connects the head of the femur center and the center of the ankle (1→7). (iv) Evaluation of the deformity based on the MAD and the position of the axis (1→7) with respect to the center of the knee.

the physician's expertise. To alleviate this burden, several computer-aided diagnosis systems have been proposed, see Section I-A.

A common application of landmark detection in orthopedics is the assessment of lower limb alignment. This examination is done with CT or X-ray imaging, the latter being preferred due to the low radiation exposure and time needed to acquire an image [3]. To analyze the lower limb condition and to give a diagnosis or plan an orthopedic surgery, a lower limb malalignment (LLM) test is executed [4]. This test involves manually drawing axes over a full lower limb (FLL) X-ray with predefined landmarks of origin and end. After delineation of the axes, the angles formed between them are measured as is shown in Fig. 1. From these measured values, identification of the type of limb deformity can be achieved, see Fig. 1, or surgery planning can be decided.

Recently, deep learning algorithms have been explored to automate the previous process [5], [6], [7], [8]. These methods aim to automatically and accurately detect the landmarks required to draw the axes for an LLM test. To achieve such a goal, they employ standard landmark detection techniques: coordinate regression [6], [7], or segmentation [5], [8]. In this work, we propose a novel approach for landmark detection in FLL X-rays that combines both methodologies: coordinate regression and segmentation. Our approach was designed to balance accuracy and robustness, making it suitable for clinical applications such as the LLM test.

A. RELATED WORK

As Wu et al. [9] stated, multiple forms exist to execute landmark detection on images. This work on facial landmark detection highlights the three most common approaches: coordinate regression, heat map regression, and segmentation. Additionally, examples of how these can be combined to achieve better results are given. In each case, a two-step strategy is usually followed. First, a network to identify and isolate the region of interest (ROI) from the full image is employed. Next, a second network is used to locate the desired landmarks within the previously found ROI.

1) COORDINATE REGRESSION

Coordinate regression models aim to directly learn the mapping from an image to the landmark coordinates [9]. Usually, an encoder network is employed to extract features

from the image. Subsequently, a set of fully connected layers is used, where the final layer corresponds to the number of detected coordinates. A variation of this approach is known as *cascaded coordinate regression*, where two or more coordinate regression networks are stacked one after another. Thanks to this arrangement, each network refines the estimation of the previous one, achieving a more accurate landmark positioning [10].

Lee et al. [11] employed coordinate regression to estimate landmarks from cephalogram X-rays. First, a grid search was executed to get patch candidates for the ROIs of where the landmarks could be located. Then, a convolutional neural network (CNN) detected the patches that correspond to a possible landmark region. Afterwards, a set of 19 CNNs (1 network per landmark) performed coordinate regression on the correctly detected patches to estimate the coordinates. For the same task, Song et al. [12] also employed regression on patches extracted from the X-rays. However, image registration from labeled images was done to retrieve the ROI patches.

To determine the landmarks between the cartilage space of the knee, Tiulpin et al. [13] employed a regression network to detect the knee joint centers on bilateral knee AP X-rays. Two ROIs were cropped from these centers and input to a CNN that estimated a set of 16 landmarks via coordinate regression. Similarly, Nguyen et al. [6] used CNNs based on coordinate regression to get 10 ROIs centered on the positions of the landmarks from FLL X-rays. Subsequently, a second set of 10 CNNs refined the landmarks' position estimations. Contrary to them, Tack et al. [7] extracted the ROIs from FLL X-rays using a detection network (YOLO-v4). From the ROIs, landmarks were localized using CNNs trained following coordinate regression.

By implementing a global-to-local mindset, Noothout et al. [1] used a CNN to classify patches and regress the coordinates of the landmarks of multiple image modalities. Next, a second set of CNNs followed the same multi-task approach to refine the localization of the landmarks. Contrary to previous methods where two steps were employed to detect the landmarks, Watchareeruetai et al. [14] used a single model to regress facial landmarks coordinates. The novelty of their model relied on the employment of a transformer placed after an encoder backbone, used to give a sense of attention.

2) HEAT MAP REGRESSION

Models that follow a heat map strategy aim to localize the landmarks by indicating their position as a two-dimension heat map. Heat maps are pseudo-probability maps of a landmark representing its location at a specific pixel position [15]. Therefore, they are centered in the position of the landmarks and generally follow a two-dimensional Gaussian distribution. Often, a CNN backbone is used for feature extraction, and a subsequent de-convolutional part decodes the features to a set of heat maps corresponding to the number of landmarks [10].

Bier et al. [16] employed a two-stage pipeline that estimated heat maps to detect landmarks in hip X-rays. X-rays passed through a series of convolutional layers to yield a first set of heat maps. Then, in a second stage, the same X-rays were preprocessed through another series of convolutional layers and concatenated with the previously obtained heat maps. This new stack of features served as input for a final series of convolutional layers that generated a definitive estimation of the landmarks' coordinates. Using a similar logic, Payer et al. [15] estimated landmarks from different image modalities. Their model consisted of an encoder-decoder CNN that computed a first set of heat maps. These heat maps worked as input for a second CNN that generated a second set of heat map estimations. However, contrary to Bier et al. [16], both heat maps were multiplied to produce the final landmark estimations instead of going through more layers.

Tsai et al. [17] used direct heat map regression to localize landmarks in FLL X-rays. In this case, X-rays went through a single CNN model that yielded a unique set of heat maps. Likewise, Ye et al. [18] employed one encoder-decoder network to estimate a set of landmarks from lateral knee X-rays. Mahpod et al. [19] combined a cascade set of regression and heat map regression CNNs to estimate the position of facial landmarks. First, the heat map regression models were employed to generate a first-position estimation. Then, the regression networks refined the previous measure to achieve better localizations. Kim et al. [20] trained three HR-Nets using disentangled keypoint regression. First, the FLL X-rays were split into three regions (hip, knee, and ankle) using a rule-based partitioning system. Each cropped region was input to an HR-Net to detect a set of 19 landmarks.

3) SEGMENTATION

Few authors have employed segmentation models to perform landmark localization. This approach aims to classify an image's pixels in the foreground and background. What is commonly done is to train on small segmentation masks centered at the position of the desired landmarks. The objective of segmentation models is to correctly identify these pixel regions that correspond to the landmarks' localization. After the segmentation, the mask's centroid is calculated to retrieve the landmarks' coordinates.

Hsu et al. [10] implemented a segmentation approach to estimate facial landmarks. To achieve the latter, an encoder-decoder model was employed. The model was compared to regression and heat map regression models, outperforming both regression-based techniques. Pei et al. [5] employed segmentation to detect a set of 3 landmarks on FLL X-rays. For the detection of each landmark, a different CNN model was utilized. Likewise, Simon et al. [8] relied on segmentation to detect landmarks on FLL X-rays. However, in this case a first stage for ROI detection was used. After this step, landmarks were localized through segmentation within the ROIs.

Erne et al. [21] trained a model to first segment the femur, tibia, and fibula bones from FLL X-rays. Subsequently, four independent U-Nets segmented a total of 46 landmarks. Jo et al. [22] segmented four landmarks at the hip, knee, and ankle joints to obtain a first set of ROIs. These ROIs were input to a second set of U-Nets to segment 15 landmarks. Gai et al. [23] proposed a multi-task, multi-scale approach for segmenting bones and landmarks. One branch of their model segmented knee implants and the femur, tibia, and fibula bones. The second ramification segmented ten landmarks. To combine the features of both branches, they proposed a global-local attention module that relied on 1×1 convolution operations.

B. CONTRIBUTIONS OF THIS WORK

We propose a novel methodology for landmark detection that combines landmark segmentation with coordinate regression, termed segmentation-guided regression. Firstly, a segmentation network is employed to highlight the position of the landmarks over the image. Next, this prior positioning information is concatenated with the original image and fed to a coordinate regression branch to give the final estimate. We apply the method on FLL X-rays and demonstrate that the additional positioning information leads to more accurate estimations of the positions of the landmarks through coordinate regression. Simultaneously, the coupling with a regression model alleviates missed and erroneous detections of a direct landmark segmentation approach. The achieved balance between accuracy and robustness is deemed of great benefit for clinical use of landmark detection tools.

II. METHODOLOGY

A. GLOBAL WORKFLOW

We propose a workflow to automatically and accurately detect the necessary landmarks to measure malalignment on FLL X-rays. First, nine landmarks must be detected on each leg, as shown in Fig. 2. These landmarks allow the automatic definition of the axes and quantification of the following metrics employed in LLM assessment: MAD, mLDFA, mMPTA, FVA, and HKA. To automatically detect the nine landmarks, an approach consisting of two stages is proposed: a region of interest (ROI) identification stage and a landmark detection stage (Fig. 2).

The first stage consists of a Faster R-CNN network which detects and retrieves the hip, diaphysis, knee, and ankle regions in both legs of an FLL X-ray. Subsequently, these images are input into four independent deep-learning models (one model per ROI) to estimate the position of the required landmarks. Estimating the landmarks' location occurs according to the proposed segmentation-guided approach (detailed in Section II-D). An exception is the landmark identification of the diaphysis for the FVA delimitation. As this axis is not defined using a unique landmark, an alternative procedure was developed (see Section II-E). Once the complete set of landmarks is estimated, definition of the axes and quantification of LLM is done following the definitions mentioned in Fig. 1.

B. DATASET AND ANNOTATIONS

An anonymized private dataset of 919 FLL X-rays, including pre- and post-operative images, was employed. The X-rays were annotated using the software *V7 Darwin*.[1] First, ROIs were labeled by placing bounding boxes surrounding the desired joint. Next, nine landmarks were annotated on each leg side of an FLL X-ray, as shown in Fig. 3. As the center of the diaphysis does not correspond to a uniquely identifiable landmark, a segmentation that covers the diaphysis was used instead. From this segmentation mask, we took the midpoint at the most distal cross-section to identify the center of the diaphysis.

C. ROI DETECTION NETWORK

A Faster R-CNN [24] was trained to detect the hip, diaphysis, knee, and ankle ROIs from FLL X-rays. A pretrained Faster R-CNN with a ResNet-50 backbone was employed. However, adaptations were made such that it could identify the desired four regions. For its training, 346 FLL X-rays and their annotations were utilized, 80% for training and validation, and 20% for testing. The images were resampled from an original size of 2800×8100 to 1400×4200. Additionally, the intensity of the X-rays was rescaled between 0 and 1. Training took place during 50 epochs employing batches of 8 images. An Adam optimizer using a scheduler with an initial learning rate of 0.001 that decreased by a power of 10^{-1} after every 10 epochs was utilized. In addition, the same multi-task loss function, based on a classification and location component, was implemented as described by [24]. At inference, bounding boxes with a score less than 0.5 were pruned. Subsequently, the outputs were re-mapped to the original dimensions.

D. SEGMENTATION-GUIDED REGRESSION

The proposed segmentation-guided regression (SGR) for landmark detection uses a segmentation network that analyses image ROIs to highlight which regions of the image correspond to the landmark's location. The choice to incorporate this segmentation block is inspired by the work of Hsu et al. [10], who studied the detection of facial landmarks and found that a segmentation approach, called "pixel-wise classification," achieved better results in terms of accuracy than regression approaches. In our case, rather than directly using the resulting segmentation, the output probability map is combined with the original input and fed to a regression network to improve the stability and robustness of the approach. We theorize that by incorporating the output of the U-Net as an additional channel, the subsequent regression network focuses its attention on the immediate neighborhood of the points, which may translate to a more accurate regression of the landmark coordinates.

[1] https://www.v7labs.com/

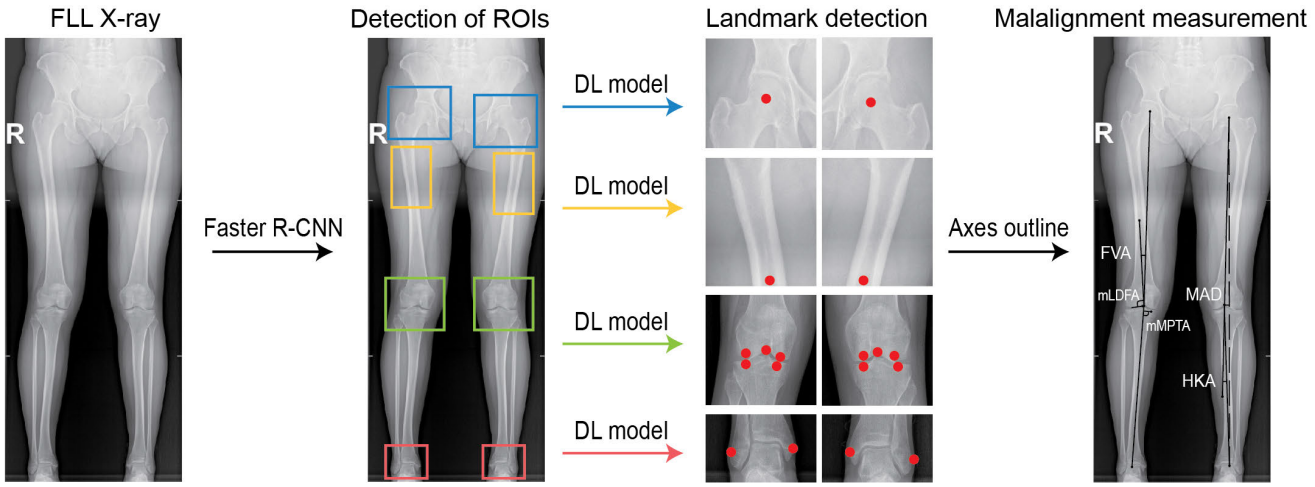

FIGURE 2. The proposed workflow consists of two main stages: detection of the regions of interest, followed by the detection and positioning of the required landmarks. ROI detection is done using a Faster R-CNN which segments the hip, diaphysis, knee, and ankle regions in both legs of a FLL X-ray. These regions are extracted and used as input to the next stage. Landmark identification is executed through four independent models (one per joint) trained via segmentation-guided regression. After obtaining the respective landmarks for each of the ROIs, delineation of the physiological axes is carried out. Finally, measurement of the malalignment metrics is executed in both legs.

The proposed model, therefore, consists of two blocks: a landmark segmentation and a coordinate regression, see Fig. 4. First, a U-Net segments small circular masks centered at the desired position. To generate these masks, we took the x-y coordinates of the targeted landmark and we placed a circular binary mask of radius equal to 15 pixels as shown in Fig. 3. This U-Net outputs probability maps with a number of channels equal to the number of landmarks present in the image. Subsequently, the output of the U-Net model is concatenated with the original X-ray image as an additional channel. Finally, this arrangement of X-rays and probability maps is input to a second CNN coupled with a fully connected (FC) layer at the end that regresses the landmarks' x-y coordinates. This last layer consists of two nodes for each landmark being estimated, where these nodes correspond to the targeted x-y coordinates.

1) ARCHITECTURE AND WEIGHT INITIALIZATION

To optimize the performance of our proposed approach, we investigated the impact of transfer learning and the depth of the networks. Shvets et al. [25] showed that segmentation could be enhanced if the convolutional layers of a VGG network pretrained on ImageNet are set as the encoder path of a U-Net. Hence, we compared the performance of networks with pretrained encoders to networks whose weights were randomly initialized. Additionally, two different backbone topologies with varying depths, VGG-11 and VGG-16 [26], were employed for comparison. As a result, four different model templates were evaluated. A U-Net with a VGG-11 topology encoder path coupled with a VGG-11 CNN to execute coordinate regression, where both have pretrained (SGR11-Pre) or random weight initialization (SGR11-Scr), and a U-Net with a VGG-16 topology encoder path coupled with a VGG-16 CNN to execute coordinate regression, where both have pretrained (SGR16-Pre) or random (SGR16-Scr) weight initialization.

2) LOSS FUNCTION

The employed loss function is composed of two terms, one for each optimized task. The first corresponds to the segmentation loss computed over the U-Net output and the ground truth landmark masks. The second term is the regression loss, calculated using the output of the VGG network and the ground truth landmark coordinates. In addition, we incorporated a hyperparameter α to modulate the impact of the segmentation loss over the regression loss. The decision to do this is because we envisioned the segmentation loss as an auxiliary component of the main task, which is estimating the coordinates of the landmarks.

a: SEGMENTATION BRANCH

Segmentation of small circular masks centered at an anatomical landmark position is an atypical segmentation task as it is highly imbalanced. For this reason, we chose to employ the Dice-Cross entropy (Dice-CE) loss. Such a loss involves coupling the Dice loss with the Cross-Entropy loss to leverage the flexibility of L_{Dsc} to class imbalance [27]. Though in literature there are multiple ways to combine these two losses, in our experiments, we used the one proposed by Isensee et al. [28]:

$$L_{Dsc-CE} = L_{Dsc} + L_{CE}\,, \tag{1}$$

$$L_{Dsc} = 1 - \frac{2TP}{2TP + FP + FN}\,, \tag{2}$$

$$L_{CE} = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y}))\,, \tag{3}$$

where TP, FP and FN correspond to true positive, false positive, and false negative classified pixels, respectively. For

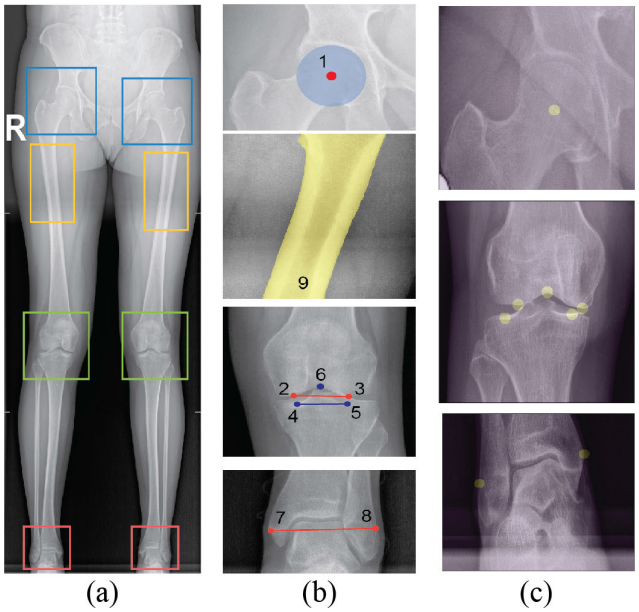

FIGURE 3. Labelling of the FLL X-rays (a) For the detection of the joints, bounding-boxes were placed on the femur, diaphysis, knee, and ankle ROIs. (b) For the landmark detection, a set of 9 landmarks were annotated on each leg. (1) A circle was fit on the head of the femur and its center was taken as the targeted landmark. (2)-(3) A line that connected the femoral condyles was delineated, the starting and ending points of it were taken as the needed landmarks. (4)-(5) A line that joint the tibial plateau was drawn and the corresponding landmarks were labelled. (6) A point was positioned on the femoral notch. (7)-(8) A line that went from the lateral to the medial malleolus parallel to the mortise's line was drawn, the starting and ending points were taken as the targeted landmarks. (9) A mask that covered the bone was drawn over the segmented X-ray, the center of the diaphysis at the most distal cross-section was taken as the landmark. (c) Circular shape masks employed to train the segmentation branch overlayed on their corresponding X-rays. For each type of joint, each image mask had a number of channels equal to the number of present landmarks.

L_{CE}, $\hat{y}$ and y represent the predicted value and the ground truth one, respectively.

b: REGRESSION BRANCH

In the regression branch of the proposed model we employed the mean squared error (MSE) loss. This one is expressed as follows:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \,, \tag{4}$$

where N is the number of samples, y_i the ground truth coordinate and $\hat{y}_i$ the estimated coordinate.

c: COMBINED LOSS

Training of previously described blocks is done in an end-to-end way. Our combined loss can be expressed as follows:

$$L = \alpha L_{Dsc-CE} + L_{MSE}. \tag{5}$$

The optimal value of α was explored on the validation set using a grid search that went from 0.1 to 0.9, on steps of 0.1.

3) TRAINING OF THE MODELS

Training and evaluation of the networks was done using a set of 919 paired FLL X-rays and annotations, 80% of these images were used for training and validation, and the remaining 20% for testing. Since the input images could vary in shape, resampling to a fixed size of 512 × 512 was done. Additionally, the intensity of the X-rays was rescaled between 0 and 1, and the number of channels was increased so that each network received batches of 8 images with dimensions of 3 × 512 × 512. During training, online data augmentation was executed. This consisted of an affine transformation involving rotations (±10°), translations (±0.1 factor) or scaling (±0.1 factor) of the images. The number of epochs was fixed to 100 and Adam optimizer with constant learning rate of 10^{-5} was set.

E. DIAPHYSIS SEGMENTATION

As described in Fig. 1, calculating the FVA angle requires the identification of three landmarks: the head of the femur, the center of the knee, and the center of the diaphysis. From these three, the automatic localization of the center of the diaphysis using landmark detection represents a challenge. The position of such a landmark is ill-defined with respect to surrounding features, i.e., many positions could be considered equivalent. Hence, the following alternative strategy was followed. First, we segmented the entire diaphysis bone using similar U-Nets as the ones described in Section II-D1. Next, we selected the midpoint of the caudal cross-section of the obtained mask as the desired landmark. Training and evaluation of the networks were done similarly to what was described above in Section II-D3, where the Dice-CE loss function was utilized. During inference, a threshold of 0.5 was employed to generate binary masks.

F. ABLATION STUDY

To evaluate the benefit of the proposed segmentation-guided regression approach, we performed an ablation study where we took each of the sub-components of the model and trained them independently to localize landmarks. Therefore, the study compares the proposed methodology to direct landmark segmentation and direct coordinate regression. For each case, the optimal architectures and weight initialization techniques described in Section II-D1 were likewise assessed. The same training-testing pipeline described in Section II-D3 was utilized to have concordant comparisons with the segmentation-guided method.

G. METRICS

1) MEAN AVERAGE PRECISION (mAP)

Mean average precision is the most widely employed metric to assess the performance of object detection algorithms. It relies on taking the mean of the average precision (AP) calculated on all the classes,

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \,. \tag{6}$$

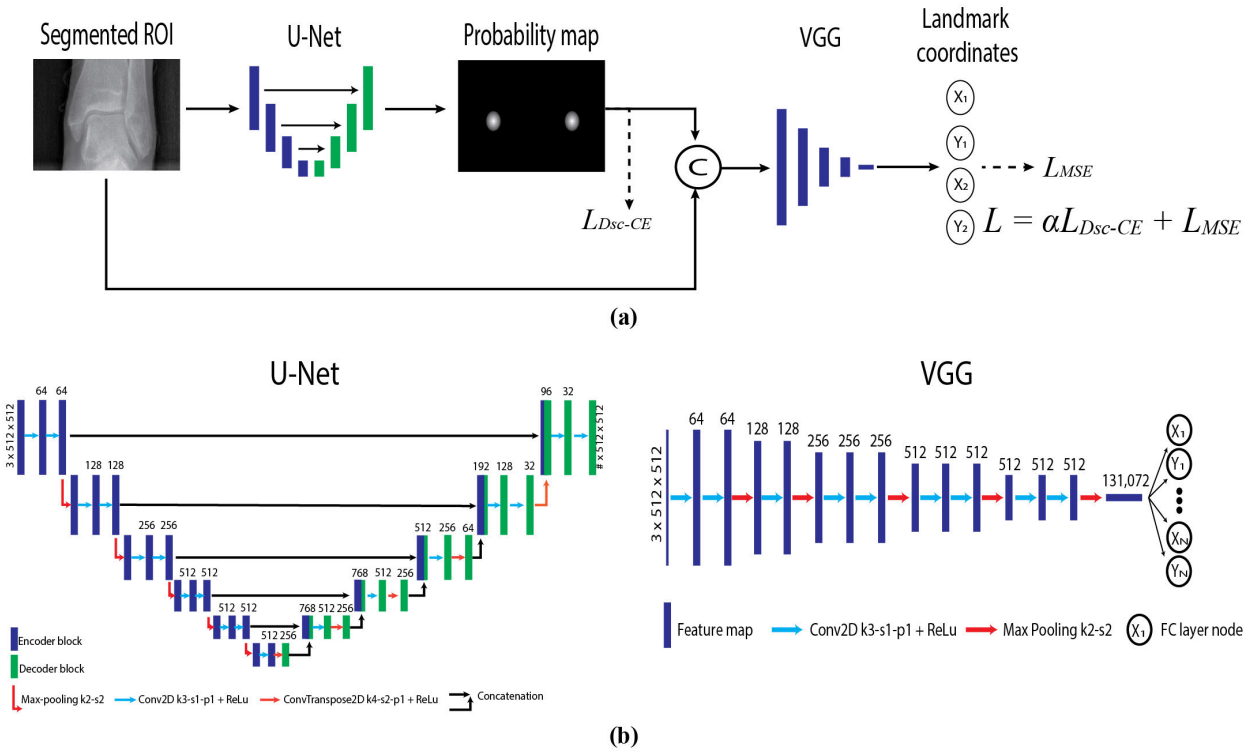

FIGURE 4. **(a) Segmentation-guided regression architecture that employs a U-Net to yield probability maps where the landmarks are located. Next, a VGG-like CNN uses the concatenated X-rays and probability maps to estimate the coordinates of the landmarks. (b) Architectures of the implemented networks. *Left:* U-Net topology utilized for the segmentation module. *Right:* VGG-16 backbone employed for coordinate regression.**

AP corresponds to the area under the curve obtained from a precision-recall plot. To construct this graph, the predicted bounding boxes are compared to their ground truths and their degree of overlapping is measured via the intersection over union (IoU) [29].

2) DICE SCORE

The Dice score is a metric utilized to address semantic segmentation tasks. It measures the overlap between an estimated segmentation and its corresponding ground truth. Therefore, it gives a sense of the quality of the segmentation. The closer to 1, the better the segmentation is. Mathematically,

$$Dice(I, \hat{I}) = \frac{2|I \cap \hat{I}|}{|I| + |\hat{I}|}, \tag{7}$$

where I corresponds to the ground truth and $\hat{I}$ to the estimated mask.

3) EUCLIDEAN DISTANCE

The Euclidean distance between the estimated landmark and the ground truth's position coordinates was measured to assess the detection accuracy. It is defined as

$$d(p, q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}, \tag{8}$$

where (p_x, p_y) and (q_x, q_y) correspond to the x-y coordinates of the ground truth and estimation, respectively.

4) MALALIGNMENT ANALYSIS

Finally, we evaluated the performance of the estimated landmark positions by conducting an LLM assessment. To this end, the error between the ground truth recorded values and the values obtained through the estimated positions of the landmarks was quantified. In each case, the mean absolute error (MAE) was employed,

$$MAE = \frac{\sum_{i=1}^{n} |\hat{M}_i - M_i|}{n}, \tag{9}$$

where $\hat{M}_i$ is the predicted metric value, M_i is the true metric value, and n is the number of employed test samples. Likewise, as proposed by Tack et al. [7], the percentage of images with HKA error $> 1.5°$ was measured.

H. IMPLEMENTATION

The algorithms were developed using Python 3.8.6. We employed Pytorch 1.7.1 and its sub-library Torchvision 0.8.2. These were complemented with MONAI 0.7.0, which includes the utilized Dice-CE loss function and data handling operations designed for medical images. The training of the neural networks was done using GPU computing through a Nvidia Tesla P100 GPU with 16GB of memory.

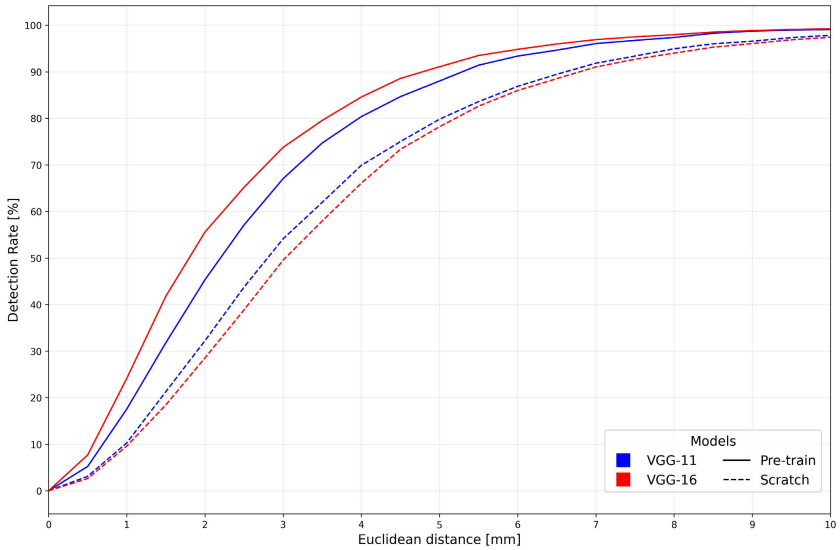

FIGURE 5. Comparison of the different possible SGR architecture combinations. The Euclidean distances of all the landmarks have been grouped and the percentage of landmarks below a specific Euclidean threshold is shown. (a) Evaluation of the different topologies and weight initialization techniques. (b) Evaluation of the pretrained VGG-16 SGR architecture, giving different weight to L_{Dsc-CE}.

TABLE 1. *mAP* values obtained for the proposed faster R-CNN at different IoU thresholds following the COCO protocol for object detection.

IoU	mAP
0.50	0.99
0.75	0.88
0.50:0.05:0.95	0.71

III. RESULTS

A. ROI DETECTION NETWORK

Table 1 summarizes the evaluation of the Faster R-CNN in terms of *mAP* at different IoU levels. First, *mAP* is calculated using an IoU of 0.50 and of 0.75. Subsequently, *mAP* is computed at IoU values ranging from 0.50 to 0.95 on intervals of 0.05; an average of these values is taken as the final score. From a test set of 101 FLL X-rays, the correct detection of the 8 ROIs was successful on 99 images. In the failed cases, the network detected only one of the two ankle regions. For a graphical representation of the results, see Supplementary Material, Fig. 1.

B. SEGMENTATION-GUIDED REGRESSION

Fig. 5 compares the four different setups where all the landmarks have been grouped and shows the percentage of landmarks below specific distance thresholds. From these results, it is observed how pretrained models outperformed their scratch counterparts. Higher successful detection rates at lower threshold values were achieved using pretrained models. For instance, at the 4 mm threshold, the pretrained SGR-11 configuration achieved a detection rate of 79.04%, whereas the SGR-16-Pre yielded 83.83%. In contrast, their randomly-scratch initialized counterparts reached 68.68% and 65.89%, respectively.

At 6 mm, detection rates above 90% are obtained on the pretrained models (SGR-11: 93.03%, SGR-16: 94.15), a condition that does not occur on the scratch configurations (SGR-11: 87.02%, SGR-16: 84.71%). Interestingly, the optimal depth of the VGG architecture depends on whether the weights were pretrained or not. In the pretrained case, the VGG-16-based topology detected the landmarks with lower error than the VGG-11 one. Yet, the scratch-initialized VGG-11 architecture reached lower Euclidean distance errors than the VGG-16 design. From the four models compared, a pretrained VGG-16-based architecture performed best and was adopted.

Since the SGR based on a pretrained VGG-16 configuration yielded the best results, we decided to use this setup to tune the value of α on Equation 5. Table 2 displays the average detection rates of landmarks at Euclidean thresholds that went from 0.5 mm to 10 mm in steps of 0.5 mm. The results show that tuning of α improves the overall detection rates. By adjusting α, average detection rates in the 77-78% boundary are achieved. Of the investigated values, the best metric is achieved at $\alpha = 0.2$ with an average detection rate of 78.08%.

C. DIAPHYSIS BONE SEGMENTATION

The performance of the networks trained to segment the diaphysis bone from the X-rays is displayed in Fig. 6. This figure shows how the pretrained models outmatch their random initialized counterparts. In addition, if both pretrained topologies are compared, there is statistically significant evidence that indicates that the U-Net with

TABLE 2. Average detection rates at different Euclidean distance thresholds [0.5:0.5:10 mm] for the pretrained VGG-16 SGR architecture, giving a different weight to L_{Dsc-CE}.

α	Avg. Detection rate [%]
0.1	77.57
0.2	**78.08**
0.3	77.88
0.4	77.92
0.5	77.25
0.6	77.22
0.7	77.98
0.8	77.82
0.9	77.39
1.0	75.44

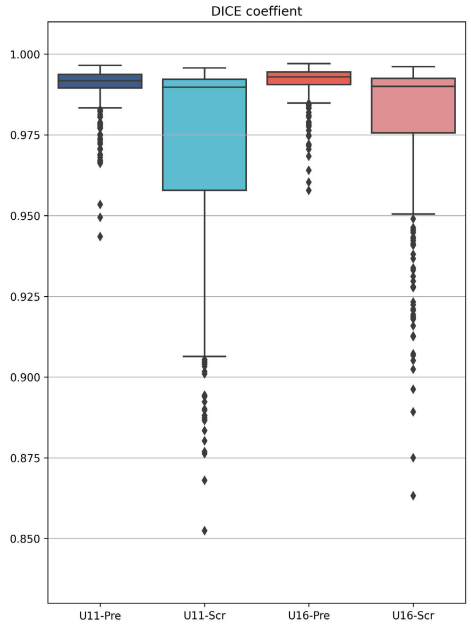

FIGURE 6. Diaphysis segmentation evaluation. **Notes:** ***U11-Pre***: U-Net with pretrained VGG-11 encoder path (Dice = 0.99 ± 0.007), ***U11-Scr***: U-Net with VGG-11 random initialized weights (Dice = 0.97 ± 0.031), ***U16-Pre***: U-Net with pretrained VGG-16 encoder path (Dice = 0.99 ± 0.005), and ***U16-Scr***: U-Net with VGG-16 random initialized weights (Dice = 0.98 ± 0.024).

the VGG-16 encoder path surpasses the VGG-11 topology (Wilcoxon test [$p < 0.05$]). For a visual comparison of the four networks, consult Fig. 2 of the Supplementary Material.

D. ABLATION STUDY

When considering the optimal depth for direct landmark segmentation and coordinate regression, we found that the combination of pretraining and VGG-16 yielded the best performance. Fig. 7 and Fig. 8 compare the segmentation-guided regression approach with $\alpha = 0.2$ against its independently trained sub-components: landmark segmentation and coordinate regression using pretrained VGG-16 encoders. For completeness, a comparison of the four possible configurations without α-tuning can be found in the Supplementary Material, Fig. 5 to 8.

When considering cases for which landmarks were successfully detected, segmentation-guided regression achieved

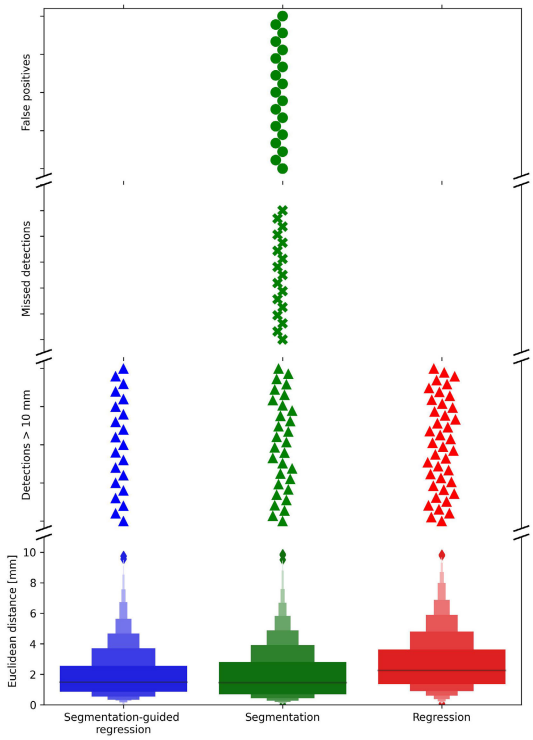

FIGURE 7. Comparison between the proposed approach and its sub-elements. On the Y-axis, from top to bottom: the number of false positives, the number of missed detections, the detections that were more than 10mm away from the ground truth position, and a box plot of all correctly estimated landmarks (femur, knee and ankle) below the 10mm distance threshold with respect to the ground truth position.

an averaged Euclidean error of 1.93 mm ± 1.53, whereas landmark segmentation reached 1.96 mm ± 1.64, and coordinate regression scored 2.67 mm ± 1.80. Hence, on average, segmentation-guided regression and landmark segmentation are more accurate than coordinate regression. The downside of landmark segmentation can be seen in Fig. 7. Direct segmentation suffered from 19 extra detected landmarks (false positives) and 17 missed detections, for a total of 2,944 ground truth landmarks. Of the 19 false positives, 11 occurred on the ankle region and 8 on the knee. The region with the highest number of missed detections was the ankle (12), then the knee (3), and finally the femur (2). Such failures impede further analysis and are problematic when considering clinical applications.

By design, regression approaches do not lead to missed detections or false positives, as each inference will lead to one set of coordinates. Coordinate regression did lead to considerably larger landmark identification errors. Segmentation-guided regression yielded a lower number of detections above the 10mm distance threshold, and lower average Euclidean error for the points within 10mm error.

E. MALALIGNMENT MEASUREMENT

To illustrate the impact of landmark detection on clinical applications, an evaluation of the measurement of lower limb malalignment was executed. A measurement was deemed

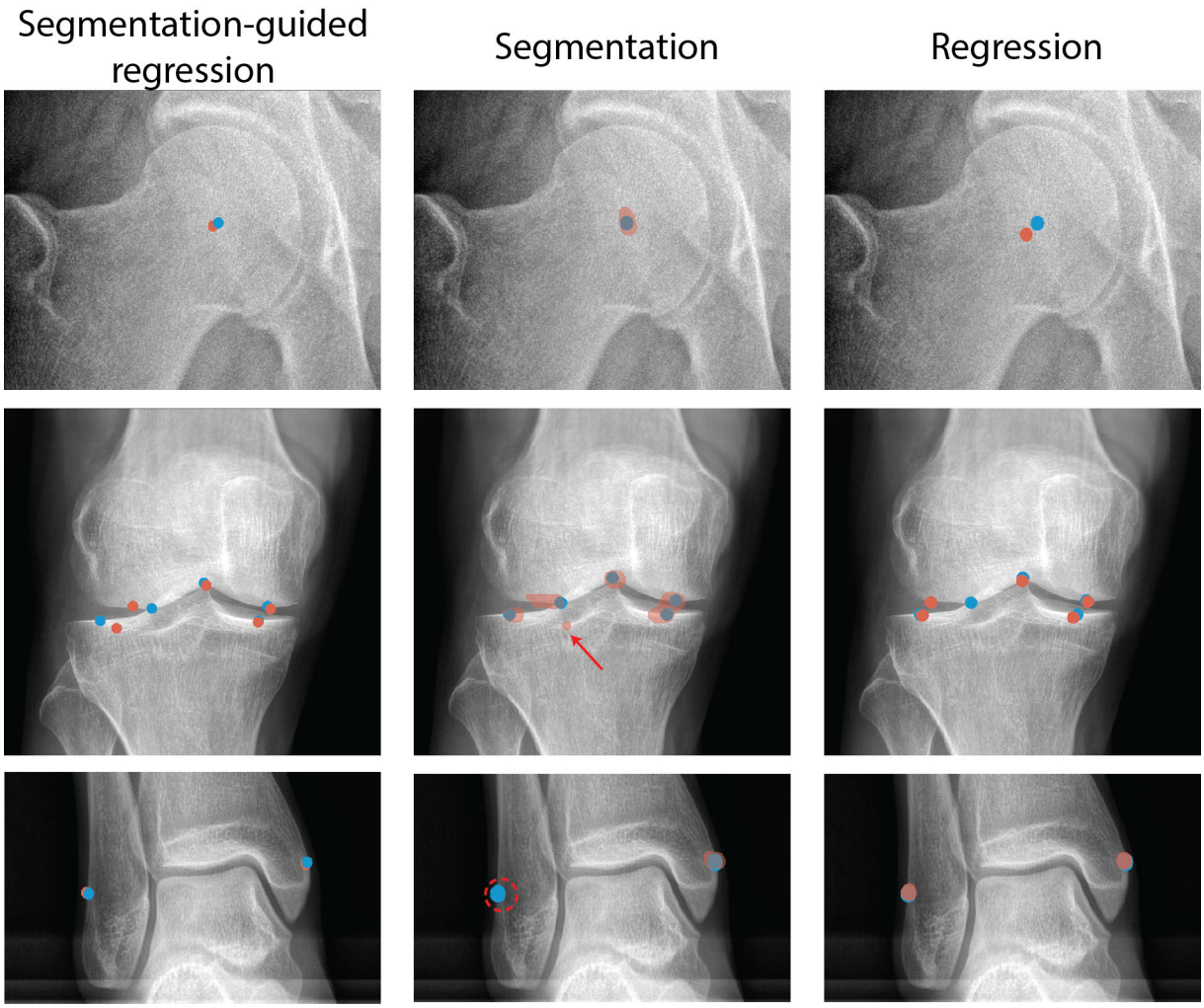

FIGURE 8. Graphic representation of the landmark detection approaches: ground truth landmark (blue) versus the estimation (red). The segmentation-guided positions the complete set of landmarks over the images in the desired location. Likewise, the segmentation centers the masks on the targeted position. Nonetheless, it is prone to false positives (red arrow) and missed detections (dashed red circle) that would impede the execution of the LLM test.

successful if all the necessary landmarks to draw the axes on both legs were detected. On the contrary, if a landmark was missing or an extra landmark was detected (false positive) on any of the ROIs, the image was considered a failure. For each malalignment metric, we computed the mean absolute error of the metric over the successful images and plotted it with respect to the number of failed images (Fig. 9). The results were compared to the segmentation and regression models tested in the ablation study.

From Fig. 9, it can be seen that the landmark segmentation approach achieves error values for successful images that are below or equal to those obtained through segmentation-guided regression and coordinate regression. However, this happens at the expense of a 3% to 11% rate of failed images for which the metrics could not be estimated. Execution of the complete LLM test (extraction of all metrics) was only possible on 86.41% of the images. Conversely, segmentation-guided regression and coordinate regression successfully analyze all X-rays. They do so at the expense of producing comparatively less accurate measurements on average. Segmentation-guided regression achieves the lowest errors of these two, consistently outperforming the regression approach for the five malalignment measurements.

IV. DISCUSSION

A. ROI DETECTION NETWORK

We found that the developed Faster R-CNN could accurately detect the hip, diaphysis, knee, and ankle regions from a full lower limb X-ray in 98% of the cases. Failing to detect the ankle regions occurred when these two were close to each other, and the network could only generate a single detection. Additionally, less accurate detections were observed when the ROIs were too close to the image's border. In such a scenario, the network predicted bounding boxes of lesser size compared to the ground truth. Nonetheless, the area where the landmark should be positioned was not affected, and accurate

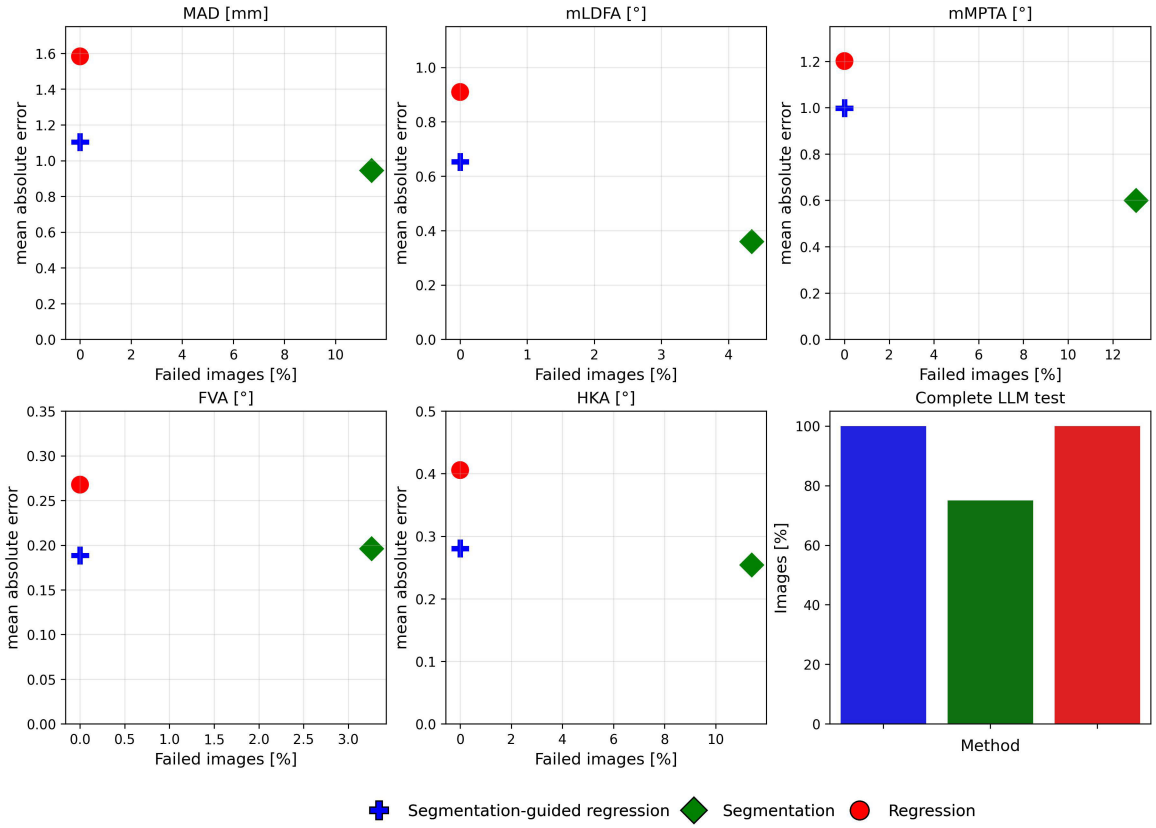

FIGURE 9. Robustness and accuracy evaluation of the proposed segmentation-guided method in contrast to the segmentation and regression ones. On the *y-axis*, MAE between the values obtained using the estimated landmark positions and the ground truth coordinates. On the *x-axis*, percentage of images where landmark detection failed. It is observed that the segmentation approach failed to detect the necessary landmarks on the five different malalignment metrics. Therefore, in only 86.41% of the images the LLM test was performed.

landmark detections could still be obtained, indicating the approach is not very sensitive to inaccurate ROI detections.

B. SEGMENTATION-GUIDED REGRESSION

A superiority for the employment of using transfer learning was identified. These results followed what Shvets et al. [25] and Iglovikov et al. [31] reported on the employment of pretrained encoders for enhanced image segmentation. Regarding the depth of the architectures, an obvious advantage was not perceived. Pretrained VGG-16 models were better than their VGG-11 counterparts. However, if weights were randomly initialized, VGG-11 topology turned out to be the optimal network architecture.

The previous observation can be explained given that the VGG-16 architecture incorporates extra trainable parameters. By having more trainable features and in the presence of relatively few data, the VGG-16 model could be overfitting. The scratch VGG-11 mitigates this since, by design, it incorporates fewer parameters to optimize. Thanks to transfer learning, a technique that has proven beneficial to avoid overfitting in the presence of few data, the VGG-16 with extra trainable parameters outperforms the other three configurations. On each landmark, lower errors were measured, which converted into higher detection rates at lower distance threshold values.

Balancing and optimizing multiple loss functions can be done in various ways [32]. In the presented job a hyper-parameter that limited the contribution of the auxiliary L_{Dsc-CE} to the total loss was implemented. This decision was due to its easy implementation and the evidence suggesting that a properly tuned parameter can yield better results [33]. Tuning α on Equation 5 translated to better landmark detections. Yet, a difference between the metrics achieved by tuning the weight of α was not observed. This situation indicates that the proposed loss function is moderately sensitive to the ratio between its elements, and extra research is required.

Regarding the ablation study between the segmentation-guided regression and its sub-components, the obtained results were inline with those reported by Hsu et al. [10]. Landmark segmentation more accurately localized the position of the target points compared to the regression approach. This condition translated to better detection rates and accurate malalignment metrics. However, for our experiments on FLL X-rays, landmark segmentation proved unreliable leading to false positives and missing

TABLE 3. Landmark positioning error comparison between the segmentation-guided regression and literature. Notes: *HoF*: head of the femur, *Right-FC*: right femur condyle, *Left-FC*: left femur condyle, *Right-TP*: right tibial plateau, *Left-TP*: left tibial plateau, *CoK*: center of knee, *LM*: lateral malleolus, *MM*: medial malleolus. *They employed the center of the ankle as landmark.

Model	Euclidean distance [mm]							
	HoF	Right-FC	Left-FC	Right-TP	Left-TP	CoK	LM	MM
Ours	1.43 ± 1.04	2.29 ± 2.25	2.11 ± 1.79	2.45 ± 2.53	2.49 ± 2.37	1.10 ± 1.14	2.30 ± 1.91	2.09 ± 1.48
Tack *et al.* [7]	1.72 ± 1.00	-	-	-	-	1.94 ± 1.33	$1.54 \pm 1.33^{*}$	
Tsai *et al.* [17]	3.7 ± 5.3	-	-	-	-	2.9 ± 6.3	$4.2 \pm 1.8^{*}$	

TABLE 4. Comparison of our approach with respect to what is published in the literature regarding LLM assessment. We report the mean absolute error (MAE) with the standard deviation (SD), the root mean squared error (RMSE), and the percentage of HKA errors above the 1.5° threshold. Notes: *Top corresponds to the right leg and bottom to the left leg measurements. **Two dataset were employed. ***Five values were reported, here we show the average of them.

Model	MAD [mm]		mLDFA [°]		mMPTA [°]		FVA [°]		HKA [°]		
	MAE (SD)	RMSE	MAE (SD)	RMSE	MAE (SD)	RMSE	MAE (SD)	RMSE	MAE (SD)	RMSE	>1.5°[%]
Ours	1.10 (1.30)	1.71	0.65 (0.63)	0.90	1.00 (0.98)	1.40	0.28 (0.34)	0.44	0.19 (0.16)	0.25	0.82
Pei *et al.* [5]	-	-	-	-	-	-	-	-	-	-	10.83
Nguyen *et al.* [6]*	-	-	0.90 (.80)	-	1.15 (0.94)	-	0.64 (0.50)	-	0.67 (.42)	-	17.7
	-	-	1.14 (0.89)	-	1.03 (.67)	-	1.30 (.57)	-	0.54 (.49)	-	
Tack *et al.* [7]**	-	-	-	-	-	-	-	-	-	-	3.38
	-	-	-	-	-	-	-	-	-	-	1.82
Kim *et al.* [20]	-	-	-	-	-	-	-	-	-	-	0.0
Erne *et al.* [21]***	-	-	-	1.89	-	1.42	-	-	-	-	-
Jo *et al.* [22]	-	-	0.52 (0.47)	-	0.46 (0.45)	-	-	-	0.22 (0.17)	-	0.0
Moon *et al.* [30]	0.80	1.30	0.53	0.63	0.63	0.87	-	-	-	-	-

landmarks. After a visual inspection of the failed cases on the landmark segmentation approach, we observed that missed detections occurred mainly due to the presence of external objects (bracelets, screws, and other orthopedic implants) or because the relevant anatomical region was not fully present on the X-ray.

Thanks to its architecture that incorporates a coordinate regression network, our methodology always led to the detection of the exact number of landmarks. Compared to coordinate regression, segmentation-guided regression detected landmarks more accurately and had fewer outliers (detections with error above 10mm). While equally accurate with respect to direct segmentation, we argue that the obtained balance in accuracy and robustness is of more value for clinical applications. In the following section, we will demonstrate that the achieved accuracy is competitive with respect to the state-of-the-art.

C. COMPARISON TO STATE-OF-THE-ART

We contrasted the performance of the proposed segmentation-guided regression to results previously reported in literature for comparable applications. While direct comparison of methodologies is not possible, due to the different datasets being used across the works, the comparison does allow us to evaluate the quality of our results with respect to the current state-of-the-art.

Table 3 displays the landmark positioning errors achieved by our model (with $\alpha = 0.2$) and compares them to accuracy values for the same landmark reported by others. From the table, one can note strongly different mean Euclidean distances for the different landmarks. For landmarks where the comparison with literature is possible, our segmentation-guided regression method achieves superior or similar results to what is currently reported in the literature for the head of the femur and the center of the knee. Tack et al. [7] achieved a better performance for the ankle. This could be explained given that their landmark was defined with respect to the talus bone, whereas ours was defined with respect to the lateral and medial malleolus.

A comparison regarding malalignment quantification is shown in Table 4. It is observed that our proposal achieves competitive results in the different assessed metrics. When comparing with Nguyen et al. [6], it can be seen that our approach obtained better results in all the malalignment metrics. A similar pattern is observed when contrasting against Pei et al. [5], Tack et al. [7], and Erne et al. [21]. Compared to our approach, Jo et al. [22] outperformed us on the measurement of mLDFA and mMPTA. This could be attributed to the data employed for training their models, where more than 10,000 X-rays were used, allowing the models to learn more features and better segment the landmarks, translating to improved metrics on LLM assessment. Moon et al. [30] achieved better metrics than us, a circumstance that could be explained by the fact that they defined the position of landmarks using hard-coded rules after segmenting the lower limb bones instead of detecting the landmarks.

Regarding the measurement of HKA, it is noticed that segmentation-guided regression achieves a percentage of images with HKA error $> 1.5°$ of 0.82%. Such a result makes our approach outperform what is reported by several authors. Yet, Kim et al. [20] and Jo et al. [22] yielded a 0% of measurements above the 1.5° threshold. Compared to us, they employed more than ten times the X-rays we used to train

our networks, allowing their models to learn more features and generalize better to new images. We employed 735 FLL X-rays; Kim et al. used 11,212, and Jo et al. 10,907.

Finally, as a limitation, we should note that the development of this investigation was entirely performed using a private dataset. Therefore, deployment and testing of the developed networks on datasets like the OAI or MOST should now be undertaken.

V. CONCLUSION

We proposed a novel method for the automated detection of landmarks in X-rays, termed segmentation-guided regression, based on two approaches for detecting landmarks: landmark segmentation and coordinate regression. By using a segmentation network and incorporating the output probability map into a regression network, we achieved a considerable increase in robustness with respect to direct segmentation and an improved accuracy with respect to direct regression. Compared to results reported in the literature, our approach led to similar or superior accurate landmark detections and highly reliable lower limb malalignment measurements.

REFERENCES

[1] J. M. H. Noothout, B. D. De Vos, J. M. Wolterink, E. M. Postma, P. A. M. Smeets, R. A. P. Takx, T. Leiner, M. A. Viergever, and I. Išgum, "Deep learning-based regression and classification for automatic landmark localization in medical images," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 4011–4022, Dec. 2020.

[2] M. Gillot, F. Miranda, B. Baquero, A. Ruellas, M. Gurgel, N. Al Turkestani, L. Anchling, N. Hutin, E. Biggs, and M. Yatabe, "Automatic landmark identification in cone-beam computed tomography," *Orthodontics Craniofacial Res.*, vol. 26, no. 4, pp. 560–567, 2023.

[3] A. Vashishtha and A. K. Acharya, "An overview of medical imaging techniques for knee osteoarthritis disease," *Biomed. Pharmacol. J.*, vol. 14, no. 2, pp. 903–919, Jun. 2021.

[4] N. Marques Luís and R. Varatojo, "Radiological assessment of lower limb alignment," *EFORT Open Rev.*, vol. 6, no. 6, pp. 487–494, Jun. 2021.

[5] Y. Pei, W. Yang, S. Wei, R. Cai, J. Li, S. Guo, Q. Li, J. Wang, and X. Li, "Automated measurement of hip–knee–ankle angle on the unilateral lower limb X-rays using deep learning," *Phys. Eng. Sci. Med.*, vol. 44, no. 1, pp. 53–62, Mar. 2021.

[6] T. P. Nguyen, D.-S. Chae, S.-J. Park, K.-Y. Kang, W.-S. Lee, and J. Yoon, "Intelligent analysis of coronal alignment in lower limbs based on radiographic image with convolutional neural network," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103732.

[7] A. Tack, B. Preim, and S. Zachow, "Fully automated assessment of knee alignment from full-leg X-Rays employing a 'YOLOv4 and resnet landmark regression Algorithm' (YARLA): Data from the osteoarthritis initiative," *Comput. Methods Programs Biomed.*, vol. 205, Jun. 2021, Art. no. 106080.

[8] S. Simon, G. M. Schwarz, A. Aichmair, B. J. H. Frank, A. Hummer, M. D. DiFranco, M. Dominkus, and J. G. Hofstaetter, "Fully automated deep learning for knee alignment assessment in lower extremity radiographs: A cross-sectional diagnostic study," *Skeletal Radiol.*, vol. 51, no. 6, pp. 1249–1259, Jun. 2022.

[9] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *Int. J. Comput. Vis.*, vol. 127, no. 2, pp. 115–142, Feb. 2019.

[10] C.-F. Hsu, C.-C. Lin, T.-Y. Hung, C.-L. Lei, and K.-T. Chen, "A detailed look at CNN-based approaches in facial landmark detection," 2020, *arXiv:2005.08649*.

[11] C. Lee, C. Tanikawa, J.-Y. Lim, and T. Yamashiro, "Deep learning based cephalometric landmark identification using landmark-dependent multi-scale patches," 2019, *arXiv:1906.02961*.

[12] Y. Song, X. Qiao, Y. Iwamoto, and Y.-W. Chen, "Automatic cephalometric landmark detection on X-ray images using a deep-learning method," *Appl. Sci.*, vol. 10, no. 7, p. 2547, Apr. 2020.

[13] A. Tiulpin, I. Melekhov, and S. Saarakkala, "KNEEL: Knee anatomical landmark localization using hourglass networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 352–361.

[14] U. Watchareeruetai, B. Sommana, S. Jain, P. Noinongyao, A. Ganguly, A. Samacoits, S. W. F. Earp, and N. Sritrakool, "LOTR: Face landmark localization using localization transformer," *IEEE Access*, vol. 10, pp. 16530–16543, 2022.

[15] C. Payer, D. Štern, H. Bischof, and M. Urschler, "Integrating spatial configuration into heatmap regression based CNNs for landmark localization," *Med. Image Anal.*, vol. 54, pp. 207–219, May 2019.

[16] B. Bier, F. Goldmann, J.-N. Zaech, J. Fotouhi, R. Hegeman, R. Grupp, M. Armand, G. Osgood, N. Navab, A. Maier, and M. Unberath, "Learning to detect anatomical landmarks of the pelvis in X-rays from arbitrary views," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 14, no. 9, pp. 1463–1473, Sep. 2019.

[17] A. Tsai, "Anatomical landmark localization via convolutional neural networks for limb-length discrepancy measurements," *Pediatric Radiol.*, vol. 51, no. 8, pp. 1431–1447, Jul. 2021.

[18] Q. Ye, Q. Shen, W. Yang, S. Huang, Z. Jiang, L. He, and X. Gong, "Development of automatic measurement for patellar height based on deep learning and knee radiographs," *Eur. Radiol.*, vol. 30, no. 9, pp. 4974–4984, Sep. 2020.

[19] S. Mahpod, R. Das, E. Maiorana, Y. Keller, and P. Campisi, "Facial landmarks localization using cascaded neural networks," *Comput. Vis. Image Understand.*, vol. 205, Apr. 2021, Art. no. 103171.

[20] S. E. Kim, J. W. Nam, J. I. Kim, J.-K. Kim, and D. H. Ro, "Enhanced deep learning model enables accurate alignment measurement across diverse institutional imaging protocols," *Knee Surgery Rel. Res.*, vol. 36, no. 1, p. 4, Jan. 2024.

[21] F. Erne, P. Grover, M. Dreischarf, M. K. Reumann, D. Saul, T. Histing, A. K. Nüssler, F. Springer, and C. Scholl, "Automated artificial intelligence-based assessment of lower limb alignment validated on weight-bearing pre- and postoperative full-leg radiographs," *Diagnostics*, vol. 12, no. 11, p. 2679, Nov. 2022.

[22] C. Jo, D. Hwang, S. Ko, M. H. Yang, M. C. Lee, H.-S. Han, and D. H. Ro, "Deep learning-based landmark recognition and angle measurement of full-leg plain radiographs can be adopted to assess lower extremity alignment," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 31, no. 4, pp. 1388–1397, Apr. 2023.

[23] L. Gai, Z. Qiao, L. Fan, X. Meng, S. Fang, P. Dong, and Z. Qian, "MMAN: Multi-task and multi-scale attention network for concurrently lower limbs segmentation and landmark detection," in *Proc. IEEE 20th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2023, pp. 1–5.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–11.

[25] A. A. Shvets, A. Rakhlin, A. A. Kalinin, and V. I. Iglovikov, "Automatic instrument segmentation in robot-assisted surgery using deep learning," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 624–628.

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

[27] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2020, pp. 1–7.

[28] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, "NNU-Net: Self-adapting framework for U-Net-based medical image segmentation," 2018, *arXiv:1809.10486*.

[29] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics*, vol. 10, no. 3, p. 279, Jan. 2021.

[30] K.-R. Moon, B.-D. Lee, and M. S. Lee, "A deep learning approach for fully automated measurements of lower extremity alignment in radiographic images," *Sci. Rep.*, vol. 13, no. 1, p. 14692, Sep. 2023.

[31] V. Iglovikov and A. Shvets, "TernausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation," 2018, *arXiv:1801.05746*.

[32] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, "Multi-task learning for dense prediction tasks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3614–3633, Jul. 2022.
[33] D. Xin, B. Ghorbani, J. Gilmer, A. Garg, and O. Firat, "Do current multi-task optimization methods in deep learning even help?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 13597–13609.

SEBASTIAN AMADOR SANCHEZ received the B.Eng. degree in biomedical engineering from the National Polytechnic Institute, Mexico City, Mexico, in 2015, and the M.S. degree in biomedical engineering from Vrije Universiteit Brussel, Brussels, Belgium, in 2019, where he is currently pursuing the Ph.D. degree with the Department of Electronics and Informatics.

During his time as the Ph.D. Candidate, he has been closely collaborating on projects involving developing computer-aided diagnosis systems to analyze Covid-19 CT images, the detection of landmarks in lower limb X-rays, and the suppression of bone structure in chest X-rays. His areas of interests include deep learning, computer vision, image segmentation, image registration, and object detection.

PHILIPPE VAN OVERSCHELDE received the Medical degree from Ghent University, Ghent, Belgium, in 1999. Subsequently, he specialized in orthopedic surgery and traumatology in Ghent and obtained his recognition in 2005. During his specialist training, he followed a two-year course in biomedical and clinical engineering techniques with Ghent University. After his general surgical orthopedic training, he left for a foreign fellowship in hip and knee surgery at the renowned center of Prof. J. N. Argenson in Marseille, France.

Since June 2006, he has been associated with Dr. Marc Goossens to further expand the service focused on hip and knee pathology–including arthroscopy, ligament surgery, primary prostheses, and revision surgery. During the Summer of 2007, he spent one month in Melbourne, Australia, to work with Dr. David Young. He further specialized in the surgical treatment of sports injuries of the hip and knee joint in top athletes and in the conservative surgical treatment of hip injuries, in particular hip arthroscopy. He is a consultant for various orthopedic companies and regularly gives lectures and surgical training abroad.

JEF VANDEMEULEBROUCKE received the master's degree in electronic engineering from the University of Ghent, Belgium. He specialized in artificial intelligence in Granada, Spain, for one year. He performed a postgraduate training in numerical optimization techniques with the Federal University of Santa Catarina (UFSC), Florianpolis, Brazil. His doctoral research was on lung motion estimation and modeling for image-guided radiation therapy, performed in collaboration with the Creatis Laboratory, University Lyon 1, France, and the Center for Machine Perception, Czech Technical University, Prague, Czech Republic.

He is an Associate Professor of medical image analysis with the Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Belgium. He is an Affiliated Researcher with imec, an international research and innovation hub in nanoelectronics and digital technologies. His current research interests include medical image analysis for applications in computer-aided diagnosis and image-guided interventions, with a particular focus on thoracic, whole-body and dynamic imaging for oncology, and musculoskeletal pathologies.