

X-Ray to DRR Images Translation for Efficient Multiple Objects Similarity Measures in Deformable Model 3D/2D Registration

B. Aubert^{ID}, T. Cresson^{ID}, J. A. de Guise^{ID}, and C. Vazquez^{ID}, *Member, IEEE*

Abstract—The robustness and accuracy of the intensity-based 3D/2D registration of a 3D model on planar X-ray image(s) is related to the quality of the image correspondences between the digitally reconstructed radiographs (DRR) generated from the 3D models (varying image) and the X-ray images (fixed target). While much effort may be devoted to generating realistic DRR that are similar to real X-rays (using complex X-ray simulation, adding densities information in 3D models, etc.), significant differences still remain between DRR and real X-ray images. Differences such as the presence of adjacent or superimposed soft tissue and bony or foreign structures lead to image matching difficulties and decrease the 3D/2D registration performance. In the proposed method, the X-ray images were converted into DRR images using a GAN-based cross-modality image-to-images translation. With this added prior step of XRAY-to-DRR translation, standard similarity measures become efficient even when using simple and fast DRR projection. For both images to match, they must belong to the same image domain and essentially contain the same kind of information. The XRAY-to-DRR translation also addresses the well-known issue of registering an object in a scene composed of multiple objects by separating the superimposed or/and adjacent objects to avoid mismatching across similar structures. We applied the proposed method to the 3D/2D fine registration of vertebra deformable models to biplanar radiographs of the spine. We showed that the XRAY-to-DRR translation enhances the registration results, by increasing the capture range and decreasing dependence on the similarity measure choice since the multi-modal registration becomes mono-modal.

Index Terms—3D/2D deformable registration, biplanar X-rays, image similarity, image-to-image translation, 3D spine reconstruction.

Manuscript received 20 September 2022; revised 28 October 2022; accepted 28 October 2022. Date of publication 1 November 2022; date of current version 3 April 2023. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), in part by the Canada Research Chairs, in part by MEDTEQ—Mathematics of Information Technology and Complex Systems (MITACS) Program, and in part by EOS Image Inc. Company. (*Corresponding author: B. Aubert.*)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethic Committee C14.368 of Centre hospitalier de l'Université de Montréal (CHUM).

The authors are with the Laboratoire de recherche en imagerie et orthopédie (LIO), École de technologie supérieure, Centre de recherche du CHUM, Montreal, QC H2X 0A9, Canada (e-mail: benj.aubert@gmail.com; thierry.cresson@etsmtl.ca; jacques.deguese@etsmtl.ca; carlos.vazquez@etsmtl.ca).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TMI.2022.3218568>, provided by the authors.

Digital Object Identifier 10.1109/TMI.2022.3218568

I. INTRODUCTION

MULTI-MODAL 3D/2D registration is involved in biomedical imaging applications such as image guidance in computer assisted surgery, image fusion for surgical planning and three-dimensional (3D) reconstruction of bones from radiographs [1], [2], [3]. The registration process aims to align anatomy parts from either a 3D volume (MRI or CT-Scan) or a 3D model (surface mesh) on 2D planar radiograph(s). In the context of 3D bone reconstruction from biplanar radiographs (Fig. 1.A), a 3D deformable model is positioned and deformed using a 3D/2D registration method [4], [5], [6]. The correspondences between the 3D model projections and the 2D images can be based on geometric features (anatomical contours or landmarks) or intensity [1]. Automated registration using landmark detection on X-ray based on machine learning models is helpful for an automatic coarse initialization [3]; it enables a fully automated 3D/2D registration when combined with a fine 3D/2D registration [3].

Using anatomical points or contours as features for fine registration has limitations. The visibility of landmarks in X-rays varies considerably and is prone to unreliable localization in practice. Indeed, an unusual object appearance in projective views could occur for large deformities, such as the vertebral axial rotation that affect the pedicle visibility in frontal X-rays [7]. Maximizing the adequacy of 3D models' projected silhouettes to X-ray visible edges [4], [8] requires bone contours extraction in X-rays. However, this task is challenging and error-prone due to low contrast, air cavities and soft tissues superposition that often results in noisy and/or partial contours, especially for chest X-rays [9]. Since the contour extraction from X-rays is uncertain and depends on the object 3D orientation, most of the effort in contour-based 3D/2D registration must be done at the matching stage using, for instance, a robust matching method to build point pairs with an optimal sensitivity to noise and partial information [10]. Thus, the resulting accuracy depends on the accuracy of contour or point detection [1]. Finally, for articulated bone structures, all objects should be considered simultaneously to avoid mismatching – the structure overlays in projection, the extracted contours can belong to adjacent structure(s) [11].

Intensity-based methods, i.e., iconic registration, aims to find the optimal model's parameters that maximize the similarity between the X-ray images and the digitally recon-

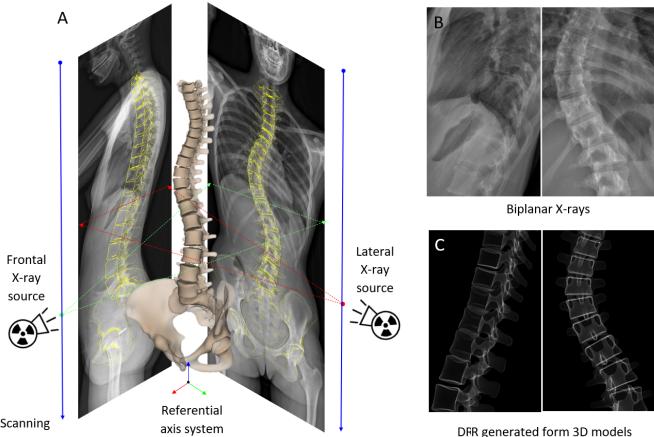


Fig. 1. Biplanar X-rays for 3D spine reconstruction (A). The 3D models are displayed in the 3D view and in both lateral and frontal projections (projected silhouettes in yellow). Differences between actual X-rays (B) and digitally reconstructed radiographs (DRR) generated from 3D models (C) showing the image domain gap and the multi-structure nature of the spine affecting the intensity-based 3D/2D registration process.

structured radiographs (DRR) generated from the 3D model [2], [11], [12]. They do not require extraction of point/contour features from images and are more robust and flexible for fine registration. However, the similarity measure, i.e., the criteria optimized in the registration process, must reliably reflect the degree of structures matching between both varying and target images. Some measures could provide accurate results but have low capture range due to their sensitivity to local minima in the objective function [3]. A main drawback is that the two images being matched differ in their modality. Even if the DRR generation tends to be realistic, domain gaps exist between both simulated and actual X-ray images (Fig. 1, B and C). Therefore, the similarity measure is often calculated on the gradient maps of the X-ray (instead of raw intensities) allowing object edges to align on both images even with this image domains gap. To enhance the gradient maps, images are generally pre-processed using local histogram normalization or advanced filters such as diffusion filter [11].

The 3D/2D registration process has already proven efficient for specific anatomical structures, specifically the pelvis and femur. However, the gradient maps are less exploitable for spine/chest x-ray images due to important superimpositions with other tissues and/or areas with visibility potentially altered [9]. To improve the images matching between real X-ray image (Fig. 1.B) and DRR image generated from a 3D model (Fig. 1.C), research focus is on improving the key components affecting the registration: the DRR X-ray simulation (involving an enriched model representation, such as intensity and shape statistical models) or developing advanced similarity metric and cost function [13].

In this paper, we changed the paradigm: instead of improving either DRR simulation or similarity metrics to increase the similarity level, the actual target X-ray image was converted into a set of DRR-like images separating bone structures using a cross-modality image-to-image translation model. We show

that using this prior image conversion step facilitates the 3D/2D registration since the image matching is more efficient in the DRR image domain where both images have similar style and characteristics. The method was applied to the 3D/2D registration of the multi-object and periodic structure of the spine from biplanar X-rays imaging modality. The similarity levels are increased and measured only on isolated structures of interest (i.e. the similarity is measured vertebra per vertebra). Even using both simple DRR generation method and standard similarity metrics, the proposed method enhances the vertebrae fine registration results. It increases the capture range and decreases the dependence on the similarity measure choice since the multi-modal registration becomes mono-modal. When combined with an automated initialization method [14], the fine-registration provides an accurate fully automated 3D spine reconstruction from biplanar X-rays.

II. RELATED WORKS

The intensity-based 3D/2D registration process is described in depth in survey papers [1] and [2]. The process is similar to a multi-modal image registration and relies on maximizing the similarity between a DRR generated from the 3D model and the X-ray image. In the context of a 3D/2D registration of a 3D model to planar radiographs, the ill-posed problem of finding the 3D model's parameters is often solved using [2]: (i) a transformation model, (ii) a 3D model projection algorithm to generate DRR image(s), (iii) a similarity metric, and (iv) an optimization algorithm to search for the optimal model parameters that reach the maximal image similarities. If the model is deformable, the algorithm solves for the 3D model's pose and shape using a regularized deformation such as modes of variations of a statistical shape model [8], [15].

The DRR generation process was initially developed to register a pre-operative 3D image (CT-scan volume) on interventional X-ray planar images [16], [17]. A ray-casting algorithm through the 3D volume provides the ray attenuation for each pixel on a DRR image. The 3D model can be represented by a floating volume incorporating density distribution knowledge. Statistical shape and intensity models (SSIM) integrate appearance and shape variations captured from a dataset of bone samples extracted from CT images [18]. However, the volume-to-volume matching and alignment is crucial to capture appropriate principal modes of variations because it strongly influences the statistical model behaviors [19]. Given that the model does not usually contain all the volumetric information with voxel densities [20], the DRR generation must be adapted for the registration of a 3D surface model. The choice of the 3D model representation defines the amount of information incorporated in the model that can be further used in DRR generation to simulate a realistic X-ray image. The densities can be integrated in the internal surface using tetrahedral meshing to generate virtual X-ray from the mesh [12], [20]. Much simpler methods generate the DRR using ray casting through a triangular meshed surface [11], [21], [22]. In [21], the DRR intensities are related to the length of X-rays intersecting the 3D meshes plus some attenuation factors, depending on the traversed tissues. In [22], the DRR

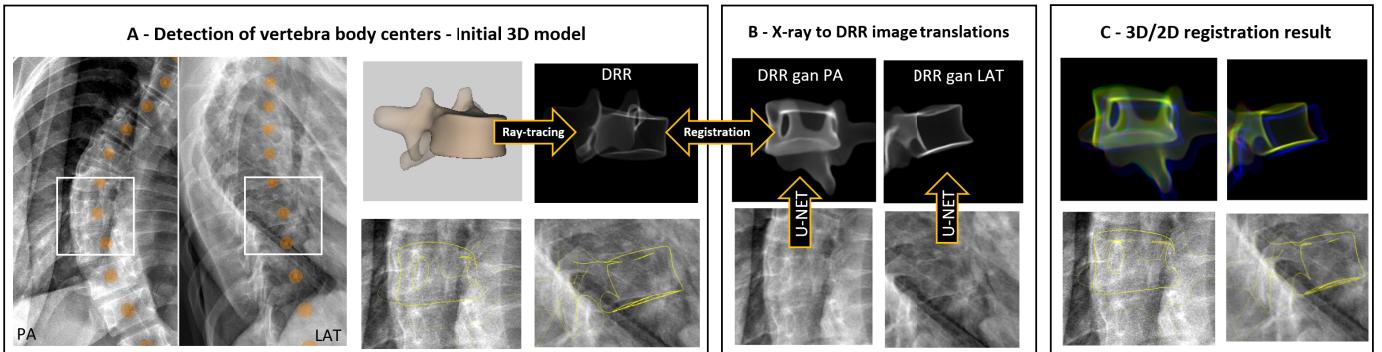


Fig. 2. Method overview. **(A)** The pre-personalized initial 3D models are provided by an automated method. **(B)** For the individual vertebra fine registration of T8, X-ray patches are converted in DRR-like images to apply a 3D/2D registration process. **(C)** After the registration, both DRR generated from the 3D model (red) and DRR converted (green) are well superimposed (initial model DRR showed in blue), aligning the 3D model's projected silhouettes (yellow contours) with X-ray image information.

generation takes into account two layers in the mesh to define the cortical and cancellous bone mediums. It requires some prior knowledge about the cortical thickness of the anatomical parts from at least one sample or from cadaveric studies.

While much effort may be devoted to generating DRR images that are similar to real X-rays, significant differences still remain between the DRR images and the real X-rays because of the limitations associated with the 3D model itself. These limitations include: the lack of structures such as soft tissues, skin, clothes, or even other bones; the absence of high-resolution density information; and the possible presence of other objects in the X-ray images.

One or more of these could contribute to making the two images very different and therefore the registration process sensitive and often unreliable. The X-rays and DRRs may be deemed to belong to two different image domains and finding correspondences between them is a cross-domain image matching problem. Advanced multi-modal similarity metrics analyzing image intensity statistical distributions such as normalized mutual information (NMI) are proven efficient in multi-modal 2D/3D registration [17], [20]. However, NMI increases the algorithmic complexity and the computational cost since it involves optimizing a non-convex cost function with no direct relation between the similarity score and the geometric space [3]. Reference[23] proposed to convert the images with an entropic filter (image structural representation) prior to the image registration in order to turn the multi-modal image registration into a mono-modal one. This image conversion enabled using the simpler sum of squared difference (SSD) metric giving similar performances as using mutual information [23]. Since simpler similarity metrics are less expensive to compute than mutual information, the optimization scheme is speeded up and can explore a higher number of potential solutions within the same timeframe. However, computing an entropic image from a DRR image remains not adapted for accurate local/fine registration because the entropic filter tends to blur the object's edges and entropic images computed from DRR and X-ray are still too different from each other.

With the advent of deep learning, recent studies have developed 3D/2D registration methods based on convolutional

neural networks (CNN) that directly regress the registration parameters [16], [24]. The CNNs replace the similarity measurement and the optimizer stages by capturing the link between both images and model parameters. However, it could be difficult to understand or trace the root cause of a registration failure. Since projection matrices are not considered in the problem formulation, it is generally assumed that the CNNs learn projection variations by themselves (appearance of objects in X-ray projections depends on their 3D positions and distance to X-ray source). This is also a limitation of recently proposed methods that directly infer a 3D model from 2D images [25], [26], even though some deep learning methods now consider varying the projection (camera parameters) in generating training data [27], [28]. Deep learning cross modality image-to-image translation gained important interest in medical imaging [29] for: segmentation and bone removal tasks [30], domain homogenization [31], or 3D/2D registration [16]. It allows one to learn a dense mapping to convert an image from its original image domain to an image domain where the downstream task is simplified and more efficient.

Even if the appropriate similarity metric is used for image correspondences, mismatching between articulated bone structures could occur because the structures are superimposed in projection (due to the line-integral nature of X-ray transmission imaging). Therefore, intensity-based object matching remains a challenge. To limit the mismatching in image correspondences, multi-object 3D models are often adopted to simultaneously match each projection part to the respective visible structures in X-rays [11], [32]. Indeed, trying to register a single 3D model in a scene containing multiple similar objects is prone to errors. However, multi-object models are more difficult to fit in images and raise the computational complexity. The optimization scheme solves for multiple poses, plus object shapes, and a trade-off could be necessary between pose and local shape deformation [11]. Registration of vertebrae in planar X-rays have one additional specificity related to the registration capture range - the spine is a periodic structure [33] and the presence of similar adjacent structures and overlaps in the environment can affect the registration performance (mismatches between vertebral endplates). This

is especially evident in non-rigid cases where the shape is also optimized because similarity measures can have local maxima for a registration solution that is degenerated.

III. MATERIALS AND METHODS

The proposed intensity-based 3D/2D registration method is applied to the fine registration of vertebrae on biplanar X-rays. An automated initialization method [14] provides the initial pose and shape of 3D models. Briefly, a spine statistical shape model with a detailed 3D surface representation is fitted on biplanar images using convolutional neural networks that detect the vertebral body centers. Then, a ray-casting through the two-layer 3D model (section III.B) produces a digitally reconstructed radiograph (DRR) image used in the registration (Fig. 2). Next, we introduced a prior image-to-image translation step (section III.C) which converts X-ray images into DRR images (Fig. 2) to: avoid mismatching between the DRR image and a radiograph having multiple superimposed objects, improve robustness, increase accuracy of the registration and reduce the computational complexity. This image conversion uses “pix2pix” GAN networks [34] and is beneficial to compute the image similarity (i.e. the criteria optimized in the registration) in the same image domain and on isolated objects (individual vertebra). The frontal and lateral X-ray patches centered on the vertebral center are therefore converted in DRR-like images using the U-Net generator trained by the GAN. These converted images become the target of an efficient 3D/2D registration process since measuring similarity on a single object from both images having the same DRR style improves the level of similarity (Fig. 2). The optimal pose and shape parameters maximizing the image similarities are found using a CMAE-ES optimizer. A statistical shape model of the vertebra is used for the shape regularization (section III.D) in the vertebra fine (non-rigid) registration. It provides an accurate subject-specific 3D model without user interventions.

A. Imaging Data

The dataset for the XRAY-to-DRRs converter training includes 463 adolescent idiopathic scoliosis (AIS) patients (average Cobb angle 43°, min./max. range [10 - 117]). Pre-operative or follow-up radiographs were retrospectively collected at the Sainte-Justine university hospital center (Montreal, Canada) after ethical approvals. Patients with over or sub numerary vertebra and post-op images were excluded. For each patient, biplanar X-ray acquisitions (frontal and lateral orthogonal views) and a 3D spine reconstruction (3D vertebra models) were available. The images were carried out using the fan-beam slot scanner (Fig. 1) EOS system (EOS Imaging, Paris, France) that provides full body bi-planar low dose standing X-rays [35]. This X-ray imaging device is geometrically calibrated and projection matrices are known [36]. An experienced user reconstructed the spine in 3D using the SterEOS software (EOS Imaging, Paris, France), which integrates a semi-automated 3D reconstruction method [37]. For the assessment of the proposed method, a separated clinical dataset of 40 AIS patients (mean age 14 ± 2 years,

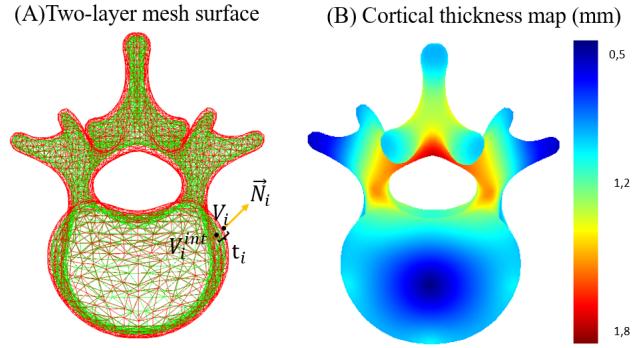


Fig. 3. Internal layer (green) creation from the external layer (red) using prior cortical thicknesses and normal vectors (A); Interpolated thickness map (mm) for the whole mesh of a L1 vertebra (B).

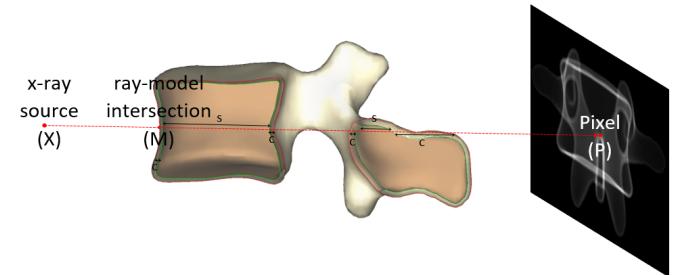


Fig. 4. Principle of the ray-casting through the two-layer mesh. The ray joining the X-ray source and a pixel on the DRR image alternatively traverses the cortical (c) and cancellous (s) mediums.

average main Cobb angle 56° and min./max. [32 - 102]) is used. Considering that there is no gold standard to compare the 3D model poses in standing position (in the EOS cabin referential axis system, cf. Fig. 1) and the 3D reconstructions are subject to intra- and inter-variability [37], three different experts carried out the 3D reconstruction of the 40 patients. Ground truth 3D models are built from the mean of three 3D reconstructions, defining a bronze standard for the test set [38]. For five patients, a localized lumbar CT-Scan is available. Their vertebrae are segmented with the 3D Slicer [39] to assess shape accuracy considering CT as the gold standard.

B. DRR Projection From Two-Layer 3D Model

The virtual X-ray images (DRR) are computed using a ray casting intersecting a 3D surface model composed of two layers delimiting the cortical and the cancellous bones. The two-layer surface is used to consider two different factors of attenuations corresponding to both material characteristics.

The shape of the 3D model is represented by a meshed surface $S = \{V, F\}$, defined by the set of 3D vertices $V \subseteq \mathbb{R}^3$, and vertex indexes defining the triangular faces $F \subseteq \mathbb{N}^3$. The two-layer mesh surface is created by adding an internal layer to the mesh. For each surface vertex V_i , the internal vertex is calculated using Eq. 1. where \vec{N}_i is the surface normal at the vertex V_i and t_i is the cortical thickness at the vertex V_i (Fig. 3). The normal of each vertex is computed as the sum

of the faces normal belonging to the vertex ring.

$$V_i^{int} = V_i t_i \frac{\vec{N}_i}{\|\vec{N}_i\|} \quad (1)$$

The average cortical thickness values of specific anatomical landmarks can be found in previous studies; for instance, for vertebral endplates [40] and pedicles [41]. The thickness for all the mesh vertices are then interpolated using a thin plate spline 3D interpolation [42] allowing computation of the cortical thickness map (Fig. 3).

The ray casting approximates the principle of the X-ray image formation [21] and computes the accumulation of traversed thicknesses through the two dual bone mediums (Fig. 4). The DRR image value at pixel P is related to the set of intersection points M between the model's faces and the corresponding ray $\overrightarrow{X(P)P}$, where $X(P) \in \mathbb{R}^3$ is the X-ray source point corresponding to pixel P and depends on the view projection matrix. By summing all distances between intersection points in M and the X-ray source, the traversed thickness along the ray is calculated (Eq. 2).

$$DRR(P) = \sum_{i=1}^{|M|} \gamma(M_i) \alpha(M_i) \|\overrightarrow{X(P)M_i}\| \quad (2)$$

$$\text{With } \gamma(M_i) = \begin{cases} -1, & \text{if the ray goes into the model} \\ 1, & \text{if the ray goes outside the model} \end{cases}$$

$$\text{and } \alpha(M_i) = \begin{cases} 1.0, & \text{if } M \text{ belongs to the cortical} \\ 0.9, & \text{if } M \text{ belongs to the cancellous} \end{cases}$$

γ is a sign that depends on the ray incidence (going into or outside the model) allowing to compute the thickness. Since the cortical structure accounts for the highest X-ray energy absorption and appears brighter in radiographs than the cancellous medium [22], a lower traversed thickness multiplier coefficient $\alpha = 0.9$ was empirically chosen to generate less attenuation in the cancellous medium of the bone and reinforce the cortical edges. The DRR image is thus a simplified X-ray image generated from the 3D model's external and internal meshing. To have a fast DRR image generation (the DRR image is updated for each view at each iteration during the registration process), the algorithm is implemented using GPU.

In brief, the ray lengths $\|\overrightarrow{X(P)V_i}\|$ and $-0.9 \|\overrightarrow{X(P)V_i^{int}}\|$ are computed for each V_i and V_i^{int} vertex in the vertex shader stage. A geometry shader generates the internal triangle faces at the rendering time. At the fragment shader stage, the ray lengths are interpolated for all pixel intensities inside the projected faces within the rasterization process. The sign γ is given by the front/back facing of the faces. Using an additive blending framebuffer, we directly obtain the DRR image in a float 2D texture.

C. XRAY to DRR Image Conversion

The XRAY-to-DRRs image converters are U-Net trained using generative adversarial networks (GAN). Following vertebra morphological changes along the spine, a per-view converter is trained for each spine segment: C7, T1 to T5,

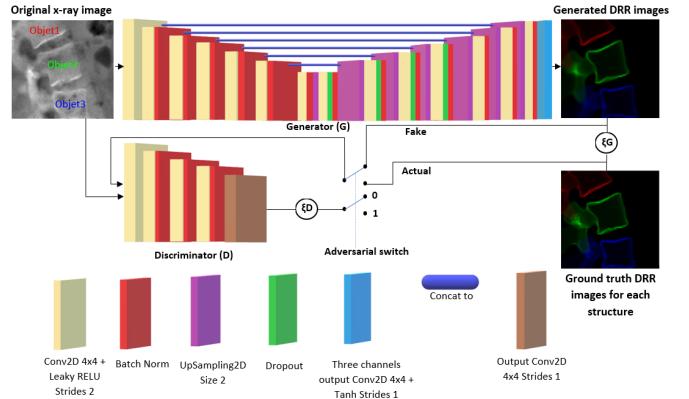


Fig. 5. XRAY-to-DRRs GAN network architecture and training. The U-Net generator is trained to convert the input XRAY patch into three DRR images (one channel per vertebra DRR represented by RGB image with object 1 in red, object 2 in green being the target and object 3 in blue).

T6 to T12 and L1 to L5 (including the S1 sacral endplate). The trained U-Net networks with weights W creates a dense pixel-wise mapping between paired input and output images, allowing the inference of DRR images from the original X-ray image (Eq. 3).

$$I_{DRR_GAN} = Infer(I_{XRAY}, W) \quad (3)$$

Deep abstract representations must be built to solve this problem [34]. The image-to-image translation using the U-Net transforms the image style from the X-ray image domain to the DRR image domain having a black uniform background. Moreover, even if there is a single channel for the input X-ray gray level image, three channels are used for the output DRR images - one channel is assigned to each of the different anatomical structures to separate them into the superior, middle, and inferior vertebra (Fig. 5). The process for extracting data is as follows. For each patient, the original X-ray images and the DRR images generated from the 3D reconstruction of the spine were available. The 3D center of the vertebra body center (VBC) and other vertebra dimensions (body width/depth) are extracted from the 3D model using the vertebral endplate corners. For a specific vertebral level, a crop is done around the VBC projected in 2D. The size of the crop is computed by dilating the vertebra dimension to include at least parts of the two adjacent vertebrae in the region of interest.

Then, an image resampling is performed to fit the cropped region in a 256×256 image size. The converters are not sensitive to the crop because there are a variety of patches with smaller/larger vertebrae and the three adjacent vertebrae are always visible.

A data augmentation is used to transform the images with: random translations of the VBC (± 5 mm), uniform scaling ($[0.9 - 1.1]$) and in-plane rotations (± 20 degrees) around the VBC, and deform the images with random vertebra corner movements to simulate changes in vertebra endplate orientations (± 5 degrees). Thanks to the data augmentation, the trained GAN can convert the X-ray image even if it is

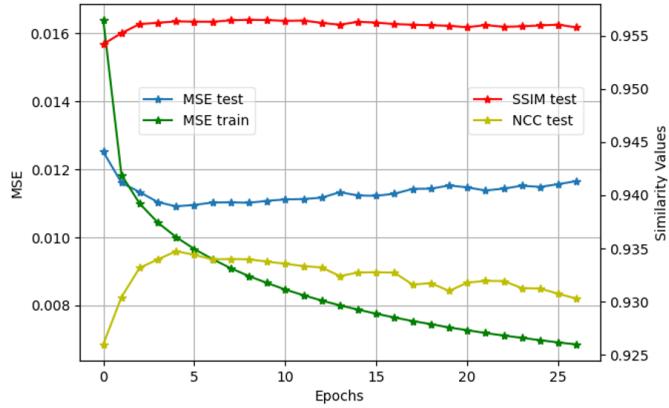


Fig. 6. Typical GAN training curves (MSE and similarity values).

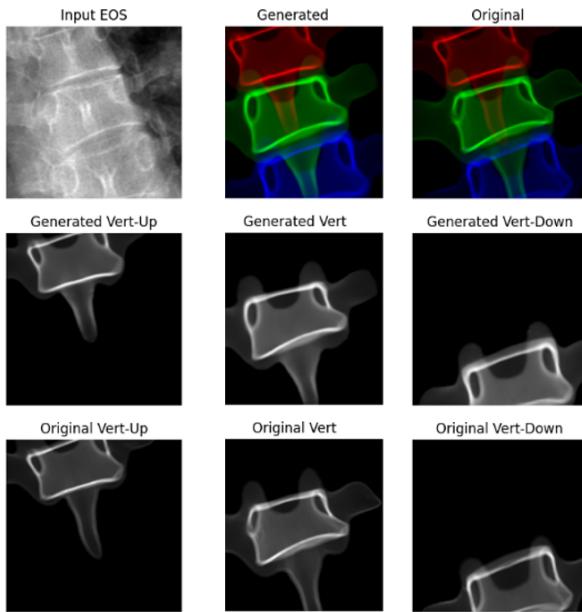


Fig. 7. Example of DRR images predicted by GAN for the thoracic level T6 in frontal view.

shifted in both X and Y directions (as well as rotated or scaled). However, it involves having one specific network per view trained separately because doing a frontal and lateral joint training would create discrepancies in biplanar X-rays correspondences (due to in-plane rotations).

A generative adversarial network (GAN) is composed of a U-Net generator (G) and a CNN discriminator (D) (Fig. 5). Therefore, two loss functions are defined. First, the residuals between the generated and actual DRR images from the training database (loss function ξG) are computed using the mean absolute error. The discriminator aims to classify a XRAY/DRR pair of images if the DRR is generated by the generator (fake class) or is the actual DRR image (real class). Therefore, the loss function ξD is defined by the binary cross-entropy loss and the discriminator output neuron has a sigmoid activation function. The architecture for the U-Net generator G is composed of 15 convolutional

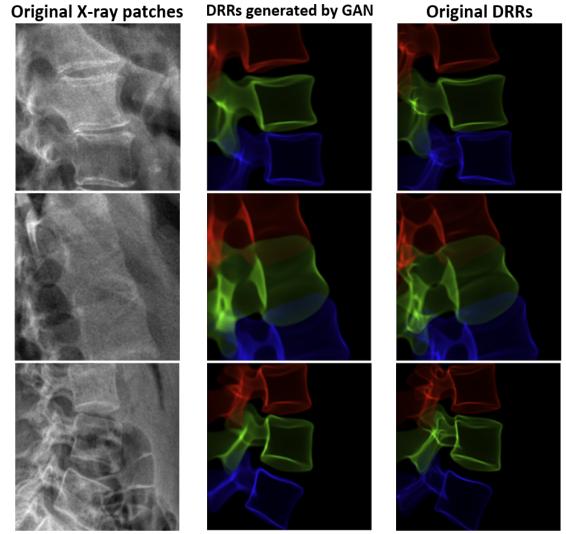


Fig. 8. Examples of L2-L3-L4 lumbar DRR predictions in lateral view.

layers for image features encoding-decoding [34]. Dropout layers ($p=0.5$) are used in the first three decoding layers to obtain a good generalization of the image-to-image translation model.

The training data are split into small minibatches of size 32. For each minibatch, the generator predicts the fake DRR images. Then, the discriminator is trained separately to predict the binary output with two images assigned to the generator output (Fig. 5) fake image (classifier output set to zero) or actual image (classifier output set to one). Finally, using the combined G and D networks and the total loss $L = \xi D + \xi G$, a retro-propagation of gradients is done to update the generator weight (W) (discriminators' weights are however frozen during this step) using the Adam optimizer. The training dataset is split into train (80%) and validation (20%) sets. An example of a typical training curve of the network is shown in Fig. 6. The training converged in a few epochs (5-15). At each epoch, the mean squared error (MSE) on predicted DRR (test set) is computed to select the best model over epochs. Similarity metrics of structural similarity index (SSIM) and normalized coefficient of correlation (NCC) are also calculated (Fig. 6). A qualitative result of XRAY-to-DRRs translation is shown in Fig. 7 and Fig. 8 for thoracic and lumbar levels respectively. The metrics computed on the validation dataset after training convergence are reported in Table I for each spine segment in both PA and LAT views. The similarity scores show that the shape of the original objects are well preserved in the generated DRRs (Table II).

D. Registration Process

The registration process aims to maximize a similarity measure between the DRR generated from the 3D model being registered and the target DRR image converted from the X-ray image using the GAN. The model transformation contains at least a rigid part (translations, rotations). Additional shape parameters could also be added to optimize the

TABLE I
QUANTITATIVE COMPARISONS (MEAN \pm SD) BETWEEN GROUND TRUTH DRR AND GAN-BASED GENERATED DRR FROM ACTUAL X-ray

Metrics / Views	Cervical (C7)	Thoracic Upper (T1-T5)	Thoracic Low (T6-T12)	Lumbar (L1-S1)
SSIM / PA	0.9398 ± 0.0096	0.929 ± 0.013	0.934 ± 0.013	0.931 ± 0.013
NCC / PA	0.9411 ± 0.0322	0.931 ± 0.032	0.932 ± 0.027	0.918 ± 0.024
MSE / PA	0.0141 ± 0.0062	0.0182 ± 0.008	0.0157 ± 0.0067	0.0143 ± 0.0054
SSIM / LAT	0.951 ± 0.0082	0.939 ± 0.014	0.944 ± 0.016	0.952 ± 0.013
NCC / LAT	0.9311 ± 0.0251	0.878 ± 0.043	0.889 ± 0.039	0.905 ± 0.03
MSE / LAT	0.0133 ± 0.0050	0.0232 ± 0.0082	0.0202 ± 0.0077	0.0122 ± 0.0045

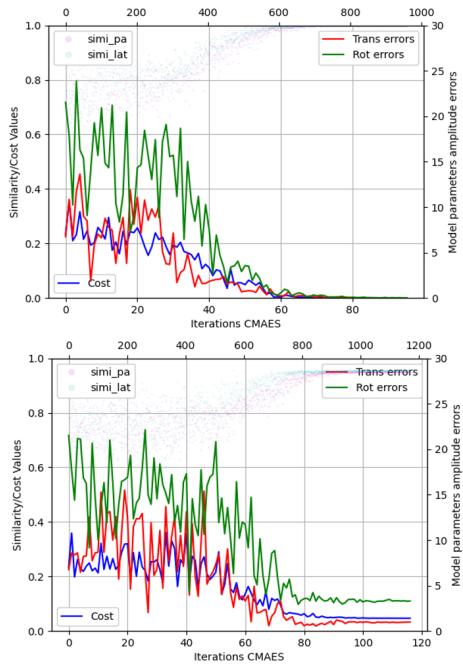


Fig. 9. Examples of CMA-ES optimization with a registration performed on the ground truth DRR that reaches perfect similarity values (top) and a registration performed on DRR predicted using GAN (bottom) (rigid case using the normalized cross correlation similarity metric).

3D model shape, for instance, using a principal components analysis (PCA) statistical shape model to restrict the shape to be statistically plausible. Formally, the cost function being maximized is defined by Eq. 4, with $S(I_1, I_2)$ being an image similarity metric comparing the images I_1 and I_2 :

$$Cost(\hat{v}) = \sum_{i \in \{PA, LAT\}} S(DRR^i(v), I_{DRR_GAN}^i) \quad (4)$$

$DRR^i(v)$ is a function to generate the DRR image on the view i from the 3D model parametrized by the vector $v = [t, r, s, u]$ with $t \in \mathbb{R}^3$ the translation vector, $r \in \mathbb{R}^3$ the Euler angles R_x, R_y, R_z around each axis, s an uniform scaling factor and u controls the 3D model shape. The dimensionality of u depends on the chosen number of modes of variations in

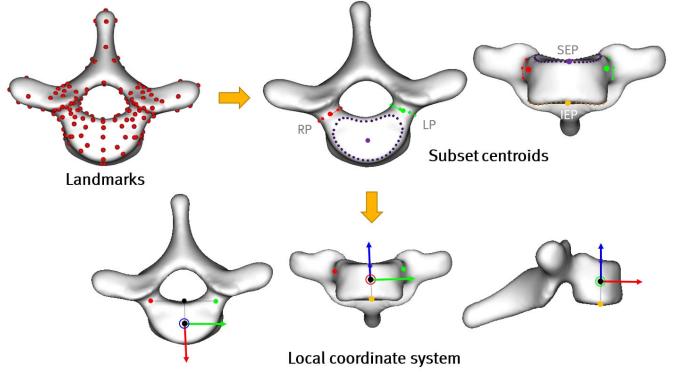


Fig. 10. Computation of the left pedicle (LP), right pedicle (RP), superior endplate center (SEP) and inferior endplate center (IEP) landmarks as centroids of subset of points. The four points define the vertebra local coordinate axis system (with X, Y and Z vectors shown in red, green and blue colors respectively).

the PCA model. The model vertices $V \in \mathbb{R}^{3,N}$ are calculated using the Eq. 5.

$$V(t, r, s, u) = s \cdot R(r) \cdot Shape(u) + t \quad (5)$$

With $R \in \mathbb{R}^{3,3}$ is the rotation matrix. For a rigid registration, the shape is frozen, otherwise $Shape() \in \mathbb{R}^{3,N}$ is a function returning the 3D model vertices using the statistical model.

We adopted a mesh moving least square (MLS) deformation technique combined to a PCA model [14]. The MLS deformation is controlled by a set of parametric geometrical handles described by parameters encoded in the PCA model. The prior definition of a simplified parametric model encompassing the object's shape allows a compact representation of the shape and directly provides subject-specific landmarks and geometrical primitives used for diagnosis from X-ray (such as the vertebral endplate orientations for patient's alignment assessment using Cobb, Kyphosis and Lordosis clinical angle measures). A statistical model is built for each individual vertebra registration using the same dataset of the global spine model proposed in [14].

In order to maximize the cost function (Eq. 4), a derivative-free exploration CMA-ES optimizer is used [43]. Each CMA-ES iteration evaluated the cost function 10 times to build the covariance matrix. The stopping criteria is reached once the parameters are stabilized. Upper and lower bounds are defined according to experiment to restrict the search space. Two examples of convergence curves are presented in Figure 9 (the negative cost function has been minimized): a registration against the DRR ground truth that quickly converges to the global minimum and a registration against a DRR generated using the GAN also showing the high level of similarity.

IV. REGISTRATION EXPERIMENTS

Three experiments have been done to demonstrate the advantages of the prior step image conversion and structure separation using GAN in vertebra model 3D/2D registration from calibrated biplanar X-rays. Five similarity metrics have been implemented [17], [33]: the sum of squared difference

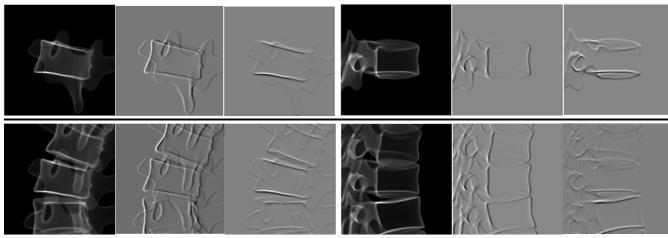


Fig. 11. PA and LAT DRR target image and XY gradients for isolated object (top row) or with adjacent objects (bottom row).

TABLE II

COMPARISONS OF THE VERTEBRA RIGID REGISTRATION ERRORS
(RMS 3D DISTANCES IN mm) WITHOUT AND WITH ISOLATED OBJECT

All objects superimposed					
Level	SSD	NCC	NGI	NCC Grad	NMI
C7	11.2	2.4	1.8	1.7	2.8
T8	7.4	1.8	1.2	1.9	1.5
T12	8.0	0.0	1.0	0.0	0.4
L4	8.1	7.1	0.1	0.7	0.5
Isolated vertebra object					
Level	SSD	NCC	NGI	NCC Grad	NMI
C7	0	0.1	0	0.5	0.1
T8	0.1	0	0.2	0	0.4
T12	0	0.4	0.3	0.35	0
L4	0.05	0.6	0	0.9	0

(SSD), the normalized cross correlation (NCC), the normalized gradient information (NGI), the normalized cross correlation on gradient images (NCC GRAD) and the normalized mutual information (NMI).

The ground truth 3D models have four landmarks defined as centroids of subsets of points extracted from a set of landmarks covering the surfaces (Fig. 10). A vertebra local coordinate axis system is defined as follows: the origin of the frame is defined as the vertebral body center $VBC = (SEP + IEP)/2$ (Fig. 10).

The center of both pedicles ($PC = (RP + LP)/2$) relatively to the VBC location give the orientation vector \vec{X} that express the vertebra axial orientation (Fig. 10). The definitions of the axis are given by equations 6, 7 and 8 (where \times represents the cross product). The vertebral frame and the landmarks are used to compute the registration errors in the experiments.

$$\vec{Z} = \overrightarrow{SEP - IEP} / \|\overrightarrow{SEP - IEP}\| \quad (6)$$

$$\vec{Y} = \vec{Z} \times (\overrightarrow{VBC - PC} / \|\overrightarrow{VBC - PC}\|) \quad (7)$$

$$\vec{X} = \vec{Y} \times \vec{Z} \quad (8)$$

A. Structure Separation Benefits

In this experiment, the ground truth 3D models were used to generate the target DRR images. The 3D models are randomly displaced in the 3D space (applying transformation

with translations in the interval $[-5, 5]$ mm and rotations around each X, Y, Z axis in the interval $[-10, 10]$ degrees). Then a rigid registration is applied. In theory, if the registration process is well implemented, the registration may result in finding exactly the inverse transformation that replaces the model at the ground truth pose (with maximal image similarity). However, in a multi-organ or multi-object 2D/3D registration, all volumetric information is projected in the image plane (Fig. 11). Consequently, objects are superimposed together, which reduce the similarity level and can introduce correspondence mismatch between structures. Table II reports the registration 3D errors in term of landmarks distances for four different vertebra levels and the five similarity functions in two conditions (Fig. 11). One condition is with isolated object (only the registered vertebra is present on DRR images for similarity computation) and the other condition is without structure separation (the registered vertebra has superimpositions with the two adjacent vertebrae).

The most sensitive image similarity in presence of superimpositions is the SSD metric (Table II). The C7 cervical and T8 thoracic areas have the highest degrees of superimpositions that perturbed the registration (being trapped in local minima). Gradient based metrics aligning object contours (NGI, NCC Grad) performed better than metrics using raw intensities (SSD, NCC), apart from the NMI metric. Note that the rigid registration considers a known shape, mismatch errors due to superimpositions should be significantly amplified if the shape is also optimized (non-rigid case). The registration against the isolated object showed zero or very small error. Some residuals were present revealing that some starting conditions caused non optimal solutions in the registration. This can be addressed by adapting the lower/upper bounds or other optimizer parameters.

B. GAN-Based Image Conversion Benefits

The same experimental setup described above was used in this second experiment. However, the target varied either by using the actual X-ray image or the DRR image converted from X-ray using GAN networks (Fig. 12). As previously demonstrated, because the isolated object is beneficial in the registration, only the middle channel was used from the GAN-based converted DRR (section III.C). Table III reports the registration errors. Using the actual X-ray as the registration target was inaccurate and gave high errors. It is not surprising that the registration failed with the most basic similarity metrics (SSD and NCC) since the X-ray image and the DRR generated from the 3D model are very different (Fig. 1). Even using the NMI metric (often chosen for robustness in multi-modal image registration), the registration did not converge in our case. When using gradient-based metrics (NGI and NCC Grad), lower errors are reached for the L2 lumbar vertebra which generally present a good visibility of structures and favorable environment with less superimpositions.

The 3D/2D registration against the GAN-based DRR allows reaching millimetric error range in pose recovery. It was the NCC and NGI metrics that gave the lowest RMS residual (mean 1.1 mm). In the rest of the experiments, the NCC

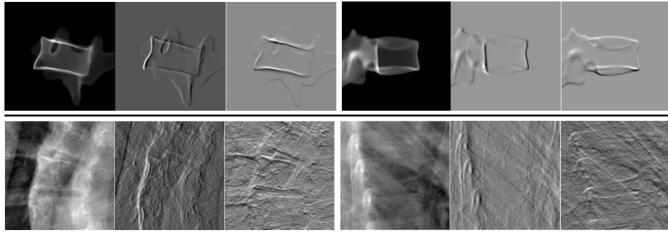


Fig. 12. PA and LAT (and XY gradients) target images for the converted DRR (top row) and the actual X-ray (bottom row).

TABLE III

COMPARISONS OF THE VERTEBRA RIGID REGISTRATION (RMS 3D ERRORS IN mm) USING X-ray OR DRR GAN AS REGISTRATION TARGET

Registration on X-ray image					
Level	SSD	NCC	NGI	NCC Grad	NMI
C7	10.0	10.5	8.5	6.3	12.3
T8	9.5	9.4	9.6	6.9	10.7
T12	10.4	9.9	6.8	4.2	12.5
L4	10.3	9.3	3.5	2.2	10.3
Registration on GAN DRR converted from X-ray					
Level	SSD	NCC	NGI	NCC Grad	NMI
C7	1.6	1.4	1.2	2.3	1.4
T8	1.4	1.35	1.3	1.8	1.4
T12	1.1	1	1.15	1.95	1
L4	1.0	0.8	0.9	0.8	0.8
Mean	1.3	1.1	1.1	1.7	1.2

similarity metric was chosen since it advantageously allows avoiding computing image gradients and thus saving computational time for the cost function evaluation. The boxplot errors for each translation and orientation components are presented in Figure 13. All translation median errors were inferior to 0.5 mm. The highest translation bias was observed for the C7 vertebra. The in-plane vertebra (frontal and lateral) pose orientations were accurate (median error < 1°). The out-of-plane axial vertebra rotation (AVR) was the most difficult orientation parameter to recover for C7 and T8/T12 thoracic vertebrae. However, it was very accurate for the lumbar vertebra L4 (Fig. 13). The variability was higher for the X translation for thoracic upper levels (T8), showing that the generated DRR is probably less accurate and could be shifted from the actual vertebra position.

C. Fully Automated 3D Spine Reconstruction

In this section, the performance of the fully automated 3D spine reconstruction is assessed. A global spine model fitted in images using CNNs provided the initial 3D models [14]. Only the initial step of fitting the vertebra body centers (VBC) is applied. Then, the non-rigid fine registration is applied, including the optimization of both pose and shape. The 3D/2D

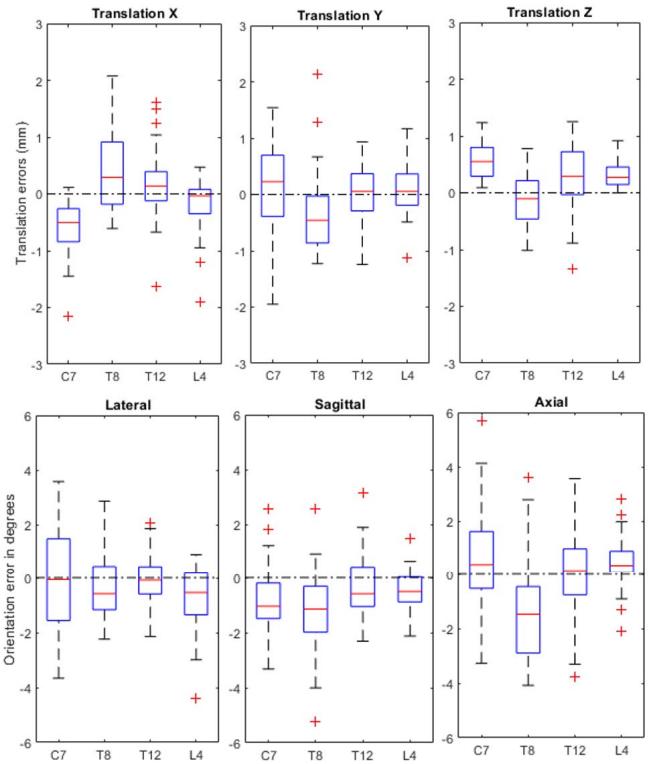


Fig. 13. Translations (top row) and orientations (bottom row) errors for C7-T8-T12-L4 vertebral levels (rigid registration using GAN DRR target and NCC similarity metric).

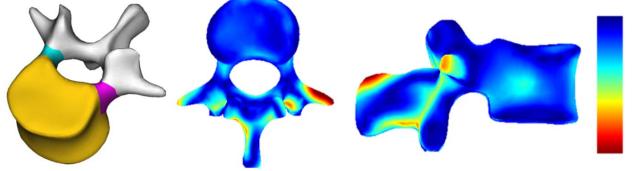


Fig. 14. Anatomical regions used for regional errors (left) and average errors map (L1) in top and lateral views (right). Color bar scale is 0 mm (blue) to ≥ 6 mm (red).

registration is launched individually for each vertebra and the spine model is fitted on resulting 3D vertebra landmarks for a final shape regularization of the whole spine [14].

1) Accuracy of Landmarks Locations: The 3D localization errors of the pedicle and endplate landmarks are reported in Table IV. These errors encompassed both shape and pose errors (absolute errors in 3D space). The fine registration improved all 3D vertebra models: the mean \pm SD error decreased from 2.6 ± 2.2 to 1.2 ± 0.9 mm, with 74.7% of landmarks having an error < 1.5 mm (Table IV). The C7 and L5 levels highlighted higher errors; they have the most challenging environment in X-ray (superimpositions, visibility...).

2) Clinical Measurements: The clinical parameters of the Cobb angle, the axial vertebra rotation (AVR) and Lordosis/Kyphosis angles for sagittal alignment are computed before and after the fine registration step. The improvements brought by the proposed method are reported in Table V. All parameters were improved, even those involving more difficult vertebra levels to detect such as T1 and L5. The accurate endplate and pedicle landmark locations provided by the

TABLE IV

ACCURACY OF 3D LANDMARKS LOCALIZATION BEFORE AND AFTER APPLYING THE 3D/2D FINE-REGISTRATION

Regions and levels	Initial solution (spine model fitted on VBC) [14]			After 3D/2D fine registration		
	Mean \pm SD	RMS	Error rate < 1.5 mm	Mean \pm SD	RMS	Error rate < 1.5 mm
Pedicle centers	2.9 \pm 2.2	3.7	16.9%	1.3 \pm 0.9	1.6	76.6%
C7	2.6 \pm 1.2	2.9	15.0%	1.6 \pm 0.7	1.8	52.5%
T1-T6	2.6 \pm 1.7	3.1	22%	1.3 \pm 0.9	1.6	70.8%
T7-T12	3.3 \pm 3.1	4.5	13.1%	1.3 \pm 1.0	1.6	72%
L1-L4	2.7 \pm 1.4	3	18.4%	1.2 \pm 0.9	1.5	77.8%
L5	3.5 \pm 1.3	3.8	5%	1.6 \pm 0.8	1.8	48.8%
Endplate centers	2.4 \pm 2.1	3.2	29.4%	1.2 \pm 0.8	1.5	74.7%
C7	1.5 \pm 0.9	1.8	62.5%	1.2 \pm 0.7	1.3	70.0%
T1-T6	2.5 \pm 1.7	3	20.6%	1.2 \pm 0.9	1.5	77.9%
T7-T12	2.7 \pm 3.1	4.1	26.2%	1.2 \pm 0.9	1.6	77.3%
L1-L4	1.9 \pm 1.0	2.2	38.4%	1.0 \pm 0.7	1.3	85.9%
L5	2.0 \pm 1.0	2.2	32.5%	1.1 \pm 0.5	1.3	72.5%
Global	2.6 \pm 2.2	3.4	23.2%	1.2 \pm 0.9	1.5	74.7%

TABLE V

CLINICAL MEASUREMENTS ABSOLUTE MEAN \pm SD ERRORS AND T-TESTS

Parameters	Initial solution [14]	After 3D/2D fine registration	p-value
Cobb angle ($^{\circ}$)	3.3 \pm 2.6	2.3 \pm 1.9	0.001
AVR ($^{\circ}$)	3.2 \pm 2.6	1.7 \pm 1.1	< 0.001
Kyphosis T1-T12	6.8 \pm 2.8	1.9 \pm 1.6	< 0.001
Kyphosis T4-T12	2.6 \pm 1.6	1.6 \pm 1.3	0.004
Lordosis L1-L5	4.5 \pm 2.6	2.1 \pm 1.4	< 0.001

TABLE VI

SHAPE ACCURACY RESULTS WITH POINT-TO-SURFACE ABSOLUTE MEAN (95% CI) ERRORS (mm)

Anatomical regions	L1	L2	L3	L4	L5
Vertebral body	0.8 (1.6)	0.8 (1.6)	0.9 (1.6)	0.9 (2.2)	1.0 (2.0)
Left pedicle	0.9 (1.6)	0.8 (1.6)	0.9 (2.0)	1.4 (3.4)	1.6 (3.6)
Right pedicle	0.8 (1.4)	0.9 (1.8)	1.0 (2.2)	1.3 (3.0)	1.4 (3.2)
Whole surface	1.3 (3.6)	1.3 (3.4)	1.3 (3.4)	1.4 (3.8)	1.6 (4.0)

fine-registration improved the AVR parameter represented by the out-of-plane rotation around the Z axis (Fig. 1 and 10).

3) *Shape Accuracy Versus CT*: The shape accuracy of the 3D vertebral model surface is independently assessed by aligning the 3D reconstructed models (standing position) and the 3D reference model generated from the CT segmentations (supine position). A coarse 3D/3D surface registration is done using the main inertial axis computed from both vertex point sets, then an iterative closest point (ICP) rigid alignment

is performed to minimize the distances between surfaces. The point-to-surface errors are reported in Table VI. The model is regionalized by anatomical regions allowing computation of errors for the most important parts used in clinical applications (the vertebral body and the pedicles) (Fig. 14). The mean error map for the lumbar L1 level is shown in Figure 14 and revealed larger errors for the posterior part of the vertebra.

Knowing the 3D rigid transformation between the reconstructed model and the CT model makes possible to transform back to the reconstructed model in the CT slices axis system to facilitate superimposing the reconstructed 3D models in the CT slices. Navigating into volume slices allows a qualitative analysis of the model's contours fit on image information against ground truth CT model (Figure 15).

V. DISCUSSIONS

In this paper, we proposed adding a preliminary XRAY-to-DRRs translation step to an intensity-based 3D/2D registration to improve its efficiency. The proposed approach is applied in the context of the 3D/2D registration of a 3D model to biplanar X-rays. We demonstrated that using XRAY-to-DRRs image converter has two main advantages that helped in improving the landmarks' localization results: (1) the image similarity calculation within the same image domain and (2) the structure separation that is beneficial for the multi-structure nature of the spine. Finally, the non-rigid fine registration of the vertebrae provided an automated subject-specific and accurate 3D modeling of the spine when combined with the automated initialization method proposed in [14].

As a first benefit, the image conversion improves the level of similarity between the two images by bringing the target image to the same domain as the varying DRR image. The conversion step thus reduces the dependency on the choice of the similarity metric over a set of metrics, and even enables using a common unimodal metric and simplified DRR generation from the 3D model. This is an interesting property since experiments are generally needed to select the best metric. However, this often results in a trade-off because some of the metrics produce good or bad results, depending on specific cases. As a second benefit, mismatches between adjacent and superimposed structures are avoided by selecting the separated structure of interest in the converter outputs. It enables a larger capture range of the registration process that is less sensitive to the initialization. Indeed, the structures were isolated in the XRAY-to-DRR translation process generating a regional DRR per object. This approach is equivalent to other research efforts that aim to decompose the X-ray signal in non-overlapped and clipped sub-volumes using a deep learning approach [44]. Each DRR is assigned to a separate layer in the multi-channel output image of the XRAY-to-DRR converter. We found that using three output layers with adjacent objects in the GAN training gave slightly better results in generated DRR than using only one channel with the middle vertebra. The reasoning behind this is that using three channels with adjacent levels helps the GAN training in cases where some levels have occluded parts. Moreover, it gives information on a larger image surface and in the overlay areas between

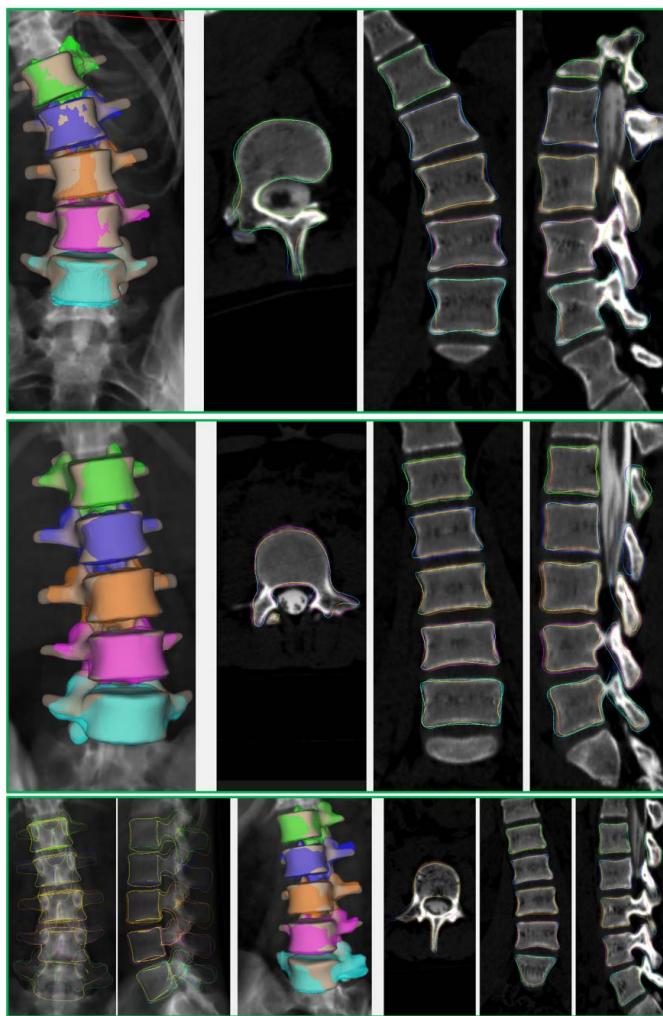


Fig. 15. Superimposition of 3D reconstructions from bi-planar X-ray images and 3D segmentations from CT-scans (L1-L5 ground truth colored in green, blue, orange, pink and cyan respectively) for three patients. The 3D surfaces are displayed in volume rendering and the projected silhouettes are displayed for the axial, coronal and lateral views. For the third patient in bottom row, the models are shown in the PA and LAT projection views of DRR generated from CT.

adjacent DRRs for gradient retro-propagation during U-Net training.

The automated fine-registration of vertebrae takes less than one minute of computation time for the whole spine (~ 3 sec. per vertebra with a Core i7 GPU with 32 MB RAM and a GTX 1080Ti Nvidia GPU) and achieved an accuracy of the 3D models comparable to state-of-the-art methods [37], [45]. The method [37] requires manual adjustments. The user must perform an elastic registration by manually moving some handles of deformation to obtain a good fit of projected silhouette on X-ray image. Among the total time needed to obtain accurate 3D models of vertebrae (more than 10 minutes) using this semi-automated 3D spine reconstruction, this adjustment step is the most operator dependent and time-consuming [45]. The mean error on the landmark 3D localization was 1.2 ± 0.9 mm, better than the 2.7 ± 1.7 mm found by a CNN-based landmark detection method for pedicles detection [46], and better than the 2.3 ± 1.2 provided by a non-linear spine model fit using a 3D-2D Markov random field [47].

Compared to a “quasi-” automated 3D reconstruction method [45] requiring user inputs and rigid manual adjustments of models on biplanar X-rays, our automated 3D/2D registration algorithm achieved better results for pose parameters on a population with more severe scoliosis. For instance, for the parameter of the axial vertebral rotation (AVR) used for scoliosis diagnosis, we found an absolute mean \pm SD error of $1.7 \pm 1.1^\circ$ after the fine-registration step (Table V) when [45] reported $3.5 \pm 3.1^\circ$ (for patients with Cobb curvature $> 30^\circ$). The fine-registration step avoids manual intervention and improves all clinical parameter accuracies.

This is a valuable result since these measured parameters feed predictive models developed for early scoliosis detection to enable better treatment [48]. However, the models are sensitive, and parameters extracted from biplanar X-rays should be accurate enough to guarantee the accuracy of the predictions. Similarly, Cobb measurements for small curvature angles (for those with mild scoliosis who require the most follow-up to observe a progression or not of the disease) are sensitive and we found a mean error of 2.3 ± 1.9 (Table V).

Compared to ground-truth lumbar CT segmentations, we found a millimetric mean shape error for the vertebral body and pedicles regions. For the whole vertebra, errors ranged from mean (95% CI) 1.3 (3.6) to 1.6 (4) mm for L5 (Table IV), versus 1 (2.7) mm in the reference semi-automated method [37]. Accuracy results were also similar to those of a method using a CT specimen as elastic mesh in an intensity-based 3D/2D registration, which reported a global error of 1.2 (4) mm for lumbar vertebrae [49]. However, manual segmentation of vertebrae in X-ray is needed to avoid matching on adjacent vertebrae in [49]. The posterior arch region in converted images was partially occluded in the LAT view due to the image square crop (Fig. 8). Therefore, the image-features were not used in this region where we observed the maximum errors (Fig. 14). Increasing the DRR image size could be a solution. Nevertheless, clinical diagnosis is less dependent on this region since clinical applications are generally focused on the pedicles and on the vertebral endplates.

A pre-requisite of the proposed approach is the requirement of paired X-ray/DRR images to train the GAN converters. In our study, we had an adequate dataset of DRR generated from 3D models reconstructed using the validated semi-automated method [37]. Therefore, the 3D models are already aligned on the calibrated biplanar radiographs. In the case of limited visibility in a given view area, the user can omit adjusting the model and the shape stays estimated by statistical inference in this region. Therefore, the information between the X-ray and the DRR derived from expert models could not match perfectly and, this potential mismatch could affect the training data. In practice, we observed no gap between the original X-ray and its conversion into DRR. Even for thoracic levels having the largest errors for DRR generation (Table I), the converted DRR from input X-ray patches which have a poor signal and/or occlusions remain plausible (Fig. 16). It still allows recovery of the 3D reconstructed model even in these challenging areas. When the input X-ray image has a very poor signal, the error on GAN prediction naturally rises (Fig. 16).

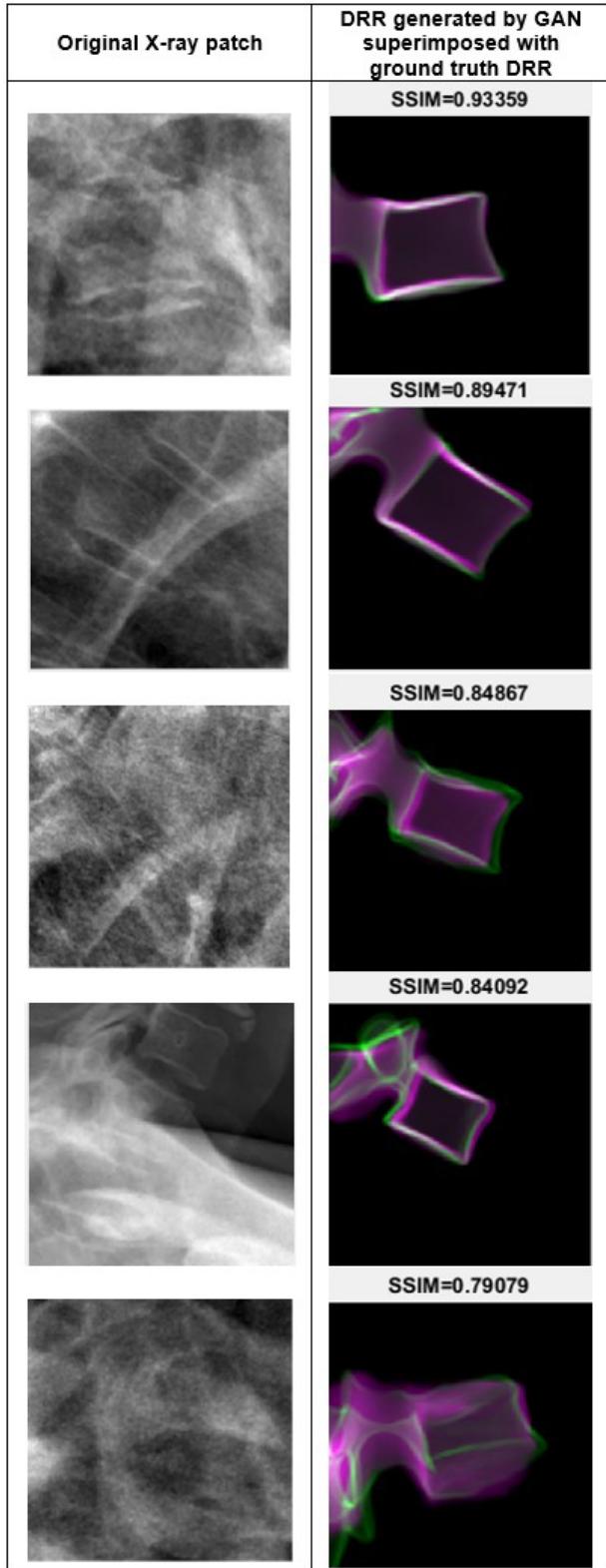


Fig. 16. DRR GAN predictions (violet in right column) from degraded input X-ray patch (left column) superimposed with ground truth DRR (green in right column) and SSIM scores computed between both DRR. All vertebra levels are thoracic vertebrae, except the fourth row that is a cervical C7.

In the future works, it could be useful to add a metric returning the uncertainty of DRR prediction to anticipate registration failure due to bad image quality. An uncertainty

map could be estimated in conjunction with the predicted DRR by calculating the entropy from multiple GAN prediction [50]. The map will give less importance to uncertain DRR image parts by using weighted similarity metric in the registration. In this case, some part of the spine in uncertain regions could be inferred from reliable regions using the whole statistical spine model. Other improvements may be dedicated to implement GAN training with other architectures such as using HR Net generator [51] to handle adult patients with degenerative deformities and/or the presence of surgical materials in X-ray that hide the image features. We already developed an initialization method that is robust to the presence of metal in post-operative images [53]. Moreover, we would like to investigate an unpaired training dataset using Cycle-GAN [52]. This should allow the use of unannotated X-ray images to massively extend the range of use of our framework, however, Cycle-GAN involves a second generator in the cycle with the inverse transformation and it could be very hard to infer the pixel intensities surrounding the bones from the simplified DRR to the actual X-ray.

VI. CONCLUSION

The proposed approach addresses the issue of low image similarity level when registration is done between simplified DRR, that has been generated from a 3D mesh, and X-ray image. The structure separation in the image conversion avoids mismatching, increases the capture range, and can be generalized to any multi-structure anatomy. In the context of the 3D spine reconstruction from biplanar X-rays, this brings a robust intensity-based registration process that replaces the time-consuming manual adjustments needed to obtain accurate 3D models semi-automatically. The proposed method should help the dissemination of 3D applications which require a robust 3D/2D registration process to be adopted in clinical routine and/or in interventional radiology.

REFERENCES

- [1] P. Markelj, D. D. Tomažević, B. Likar, and F. Pernuš, “A review of 3D/2D registration methods for image-guided interventions,” *Med. Image Anal.*, vol. 16, no. 3, pp. 642–661, Apr. 2012.
- [2] C. J. F. Reyneke, M. Lüthi, V. Burdin, T. S. Douglas, T. Vetter, and T. E. M. Mutsvangwa, “Review of 2D/3D reconstruction using statistical shape and intensity models and X-ray image synthesis: Towards a unified framework,” *IEEE Rev. Biomed. Eng.*, p. 1, 2018.
- [3] M. Unberath et al., “The impact of machine learning on 2D/3D registration for image-guided interventions: A systematic review and perspective,” *Front. Robot. AI*, vol. 8, p. 260, 2021.
- [4] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. De Guise, “3D/2D registration and segmentation of scoliotic vertebrae using statistical models,” *Comput. Med. Imaging Graph.*, vol. 27, pp. 321–337, 2003.
- [5] G. Zheng, L.-P. Nolte, and S. Ferguson, “Scaled, patient-specific 3D vertebral model reconstruction based on 2D lateral fluoroscopy,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 6, pp. 351–366, 2011.
- [6] H. Bennani, B. McCane, and J. Cornwall, “Three-dimensional reconstruction of in vivo human lumbar spine from biplanar radiographs,” *Comput. Med. Imag.*, vol. 96, 2022, Art. no. 102011.
- [7] S. Ebrahimi, L. Gajny, C. Vergari, E. D. Angelini, and W. Skalli, “Vertebral rotation estimation from frontal X-rays using a quasi-automated pedicle detection method,” *Eur. Spine J.*, vol. 28, pp. 3026–3034, 2019.
- [8] M. Fleure and S. Lavallée, “Nonrigid 3-D/2-D registration of images using statistical models,” in *Proc. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 1999, pp. 138–147.

- [9] T. Cresson, R. Chav, D. Branchaud, L. Humbert, B. Godbout, B. Aubert, W. Skalli, and J. A. De Guise, "Coupling 2D/3D registration method and statistical model to perform 3D reconstruction from partial X-rays images data," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2009, pp. 1008–1011.
- [10] N. Baka, C. T. Metz, C. J. Schultz, R.-J. van Geuns, W. J. Niessen, and T. van Walsum, "Oriented Gaussian mixture models for nonrigid 2D/3D coronary artery registration," *IEEE Trans. Med. Imag.*, vol. 33, no. 5, pp. 1023–1034, 2014.
- [11] X. Chen, J. Graham, C. Hutchinson, and L. Muir, "Automatic inference and measurement of 3D carpal bone kinematics from single view fluoroscopic sequences," *IEEE Trans. Med. Imag.*, 2013.
- [12] J. Yao and R. Taylor, "Tetrahedral mesh modeling of density data for anatomical atlases and intensity-based registration," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2000.
- [13] J. Wu and M. R. Mahfouz, "Reconstruction of knee anatomy from single-plane fluoroscopic X-ray based on a nonlinear statistical shape model," *J. Med. Imag.*, 2021.
- [14] B. Aubert, C. Vazquez, T. Cresson, S. Parent, and J. De Guise, "Towards automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting," *IEEE Trans. Med. Imag.*, vol. 38, no. 12, pp. 2796–2806, 2019.
- [15] J. Yao and R. Taylor, "Assessing accuracy factors in deformable 2D/3D medical image registration using a statistical pelvis model," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 2, 2003, pp. 1329–1334.
- [16] J. Zheng, S. Miao, Z. J. Wang, and R. Liao, "Pairwise domain adaptation module for CNN-based 2-D/3-D registration," *J. Med. Imag.*, vol. 5, no. 2, p. 21204, Apr. 2018.
- [17] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. Hill, and D. J. Hawkes, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *IEEE Trans. Med. Imag.*, vol. 17, pp. 586–595, 1998.
- [18] M. L. Picazo et al., "3D subject-specific shape and density estimation of the lumbar spine from a single anteroposterior DXA image including assessment of cortical and trabecular bone," *IEEE Trans. Med. Imag.*, vol. 17, no. 12, pp. 2651–2662, 2018.
- [19] A. Goparaju et al., "Benchmarking off-the-shelf statistical shape modeling tools in clinical applications," 2020, *arXiv:2009.02878*.
- [20] M. Ehlke, H. Ramm, H. Lamecker, H. C. Hege, and S. Zachow, "Fast generation of virtual X-ray images for reconstruction of 3D anatomy," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2673–2682, 2013.
- [21] F. P. Vidal and P.-F. Villard, "Development and validation of real-time simulation of X-ray imaging with respiratory motion," *Comput. Med. Imag. Graph.*, vol. 49, pp. 1–15, 2016.
- [22] A. Serrurier, S. Quijano, R. Nizard, and W. Skalli, "Robust femur condyle disambiguation on biplanar X-rays," *Med. Eng. Phys.*, vol. 34, pp. 1433–1440, 2012.
- [23] C. Wachinger and N. Navab, "Entropy and Laplacian images: Structural representations for multi-modal registration," *Med. Image Anal.*, vol. 16, no. 1, pp. 1–17, Jan. 2012.
- [24] S. Miao, Z. J. Wang, and R. Liao, "A CNN regression approach for real-time 2D/3D registration," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1352–1363, 2016.
- [25] P. Li, Y. Pei, Y. Guo, G. Ma, T. Xu, and H. Zha, "Non-rigid 2D-3D registration using convolutional autoencoders," in *Proc. Int. Symp. Biomed. Imag.*, 2020.
- [26] C.-J. Yang et al., "Generative adversarial network (GAN) for automatic reconstruction of the 3D spine structure by using simulated bi-planar X-ray images," *Diagnostics*, vol. 12, no. 5, 2022.
- [27] C. Neumann, K. D. Tönnies, and R. Pohle-Fröhlich, "Deep similarity learning using a siamese ResNet trained on similarity labels from disparity maps of cerebral MRA MIP pairs," 2020.
- [28] J. Van Houtte, X. Gao, J. Sijbers, and G. Zheng, "2D/3D registration with a statistical deformation model prior using deep learning," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform. (BHI)*, 2021, pp. 1–4.
- [29] S. Kaji and S. Kida, "Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging," *Radiol. Phys. Technol.*, vol. 12, no. 3, pp. 235–248, 2019.
- [30] M. Eslami, S. Tabarestani, S. Albarqouni, E. Adeli, N. Navab, and M. Adjouadi, "Image-to-images translation for multi-task organ segmentation and bone suppression in chest X-ray radiography," *IEEE Trans. Med. Imag.*, 2020.
- [31] R. Shiode, M. Kabashima, Y. Hiasa, K. Oka, T. Murase, Y. Sato, and Y. Otake, "2D-3D reconstruction of distal forearm bone from actual X-ray images of the wrist using convolutional neural networks," *Sci. Rep.*, vol. 11, no. 1, p. 15249, Jul. 2021.
- [32] T. J. T. Fotsin, C. Vázquez, T. Cresson, and J. De Guise, "Shape, pose and density statistical model for 3D reconstruction of articulated structures from X-ray images," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2019, pp. 2748–2751.
- [33] Y. Otake et al., "Robust 3D-2D image registration: Application to spine interventions and vertebral labeling in the presence of anatomical deformation," *Phys. Med. Biol.*, vol. 58, no. 23, pp. 8535–8553, Dec. 2013.
- [34] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017.
- [35] S. Deschênes et al., "Diagnostic imaging of spinal deformities: Reducing patients radiation dose with a new slot-scanning X-ray imager," *Spine*, vol. 35, pp. 989–994, 2010.
- [36] A. Allab, C. Vazquez, T. Cresson, and J. De Guise, "Calibration of stereo radiography system for radiostereometric analysis application," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, 2019.
- [37] L. Humbert, J. A. De Guise, B. Aubert, B. Godbout, and W. Skalli, "3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences," *Med. Eng. Phys.*, vol. 31, pp. 681–687, 2009.
- [38] P. Laurent, T. Cresson, C. Vázquez, N. Hagemeister, and J. de Guise, "A multi-criteria evaluation platform for segmentation algorithms," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2016, 2016, pp. 6441–6444.
- [39] A. Fedorov et al., "3D Slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imag.*, vol. 30, no. 9, pp. 1323–1341, 2012.
- [40] W. T. Edwards, Y. Zheng, L. A. Ferrara, and H. A. Yuan, "Structural features and thickness of the vertebral cortex in the thoracolumbar spine," *Spine*, vol. 26, no. 2, pp. 218–225, 2001.
- [41] R. Kothe, J. D. O'Holleran, W. Liu, and M. M. Panjabi, "Internal architecture of the thoracic pedicle. An anatomic study," *Spine*, vol. 21, no. 3, pp. 264–270, 1996.
- [42] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, 1989.
- [43] N. Hansen, "The CMA evolution strategy: A tutorial," 2016.
- [44] S. Albarqouni, J. Fotouhi, and N. Navab, "X-ray in-depth decomposition: Revealing the latent structures," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.
- [45] L. Gajny, S. Ebrahimi, C. Vergari, E. Angelini, and W. Skalli, "Quasi-automatic 3D reconstruction of the full spine from low-dose biplanar X-rays based on statistical inferences and image analysis," *Eur. Spine J.*, vol. 28, no. 4, pp. 658–664, Oct. 2018.
- [46] C. Bakhou, B. Aubert, C. Vazquez, T. Cresson, S. Parent, and J. De Guise, "Automatic pedicles detection using convolutional neural network in a 3D spine reconstruction from biplanar radiographs," *Proc. SPIE Med. Imag.*, vol. 10575, pp. 105750L–10575-9, 2018.
- [47] S. Kadoury, H. Labelle, and S. Parent, "Postoperative 3D spine reconstruction by navigating partitioning manifolds," *Med. Phys.*, vol. 43, no. 3, pp. 1045–1056, Mar. 2016.
- [48] W. Skalli, C. Vergari, E. Ebermeyer, I. Courtois, X. Drevelle, R. Kohler, K. Abelin-Genovois, and J. Dubousset, "Early detection of progressive adolescent idiopathic scoliosis: A severity index," *Spine*, vol. 42, no. 11, pp. 823–830, 2017.
- [49] L. Fang, Z. Wang, Z. Chen, F. Jian, S. Li, and H. He, "3D shape reconstruction of lumbar vertebra from two X-ray images and a CT model," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 4, pp. 1124–1133, 2020.
- [50] K. Hoebel, V. Andrearczyk, A. Beers, J. Patel, K. Chang, A. Depersinge, H. Müller, and J. Kalpathy-Cramer, "An exploration of uncertainty information for segmentation quality assessment," *Proc. SPIE*, vol. 11313, 2020, Art. no. 113131K.
- [51] Y. Gu, Y. Otake, K. Uemura, M. Soufi, M. Takao, N. Sugano, and Y. Sato, "BMD-GAN: Bone mineral density estimation using X-ray image decomposition into projections of bone-segmented quantitative computed tomography using hierarchical learning," 2022, *arXiv:2207.03210*.
- [52] B. Aubert, P. A. Vidal, S. Parent, T. Cresson, C. Vazquez, and J. De Guise, "Convolutional neural network and in-painting techniques for the automatic assessment of scoliotic spine surgery from biplanar radiographs," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2017, pp. 691–699.
- [53] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Task driven generative modeling for unsupervised domain adaptation: Application to X-ray image segmentation," in *Proc. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, 2018, pp. 599–607.