Medical Image Analysis 100 (2025) 103400

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Prediction of the upright articulated spine shape in the operating room using conditioned neural kernel fields

Sylvain Thibeault [b,*], Marjolaine Roy-Beaudry [a], Stefan Parent [a], Samuel Kadoury [a,b]

[a] Centre de Recherche du CHU Sainte-Justine, Montréal, QC, Canada
[b] Polytechnique Montréal, Montréal, QC, Canada

ARTICLE INFO

MSC:
41A05
41A10
65D05
65D17

Keywords:
Spine surgery
Neural kernel fields
Prediction model
Upright posture
Disentangled latent representation

ABSTRACT

Anterior vertebral tethering (AVT) is a non-invasive spine surgery technique, treating severe spine deformations and preserving lower back mobility. However, patient positioning and surgical strategies greatly influences postoperative results. Predicting the upright geometry from pediatric spines is needed to optimize patient positioning in the operating room (OR) and improve surgical outcomes, but remains a complex task due to immature bone properties. We propose a framework used in the OR predicting the upright spine geometry at the first visit following surgery in idiopathic scoliosis patients. The approach first creates a 3D model of the spine while the patient is on the operating table. For this, multiview Transformers that combine images from different viewpoints are used to generate the intraoperative pose. The postoperative upright shape is then predicted on-the-fly using implicit neural fields, which are trained from geometries at different time points and conditioned with surgical parameters. A Signed Distance Function for shape constellations is used to handle the variability in spine appearance, capturing a disentangled latent domain of the articulation vectors, with separate encoding vectors representing both articulation and shape parameters. A regularization criterion based on a pre-trained group-wise trajectory of spine transformations generates complete spine models. A training set of 652 patients with 3D models was used to train the model, tested on a distinct cohort of 83 surgical patients. The framework based on neural kernels predicted upright 3D geometries with a mean 3D error of 1.3 ± 0.5 mm in landmarks points, and IoU of 95.9% in vertebral shapes when compared to actual postop models, falling within the acceptable margins of error below 2 mm.

1. Introduction

To this day, surgery remains the preferred treatment option for severe cases of spine deformations, by restoring the spinal alignment along vertebral segments using forces exerted by metallic rods or tethers. Unfortunately, a reduction in spinal mobility can increase the risk of osteoarthritis which negatively impacts surgical outcomes (Cheng et al., 2015). Moreover, intraoperative positioning of the spine can significantly impact surgical results in the lateral alignment and global balance, especially for degenerative patients (Elysee et al., 2022). Studies have shown how crucial lumbar placement and natural sagittal alignment are post surgery (Karikari et al., 2018). Therefore, if online feedback of the posture is made available on the operating table, surgeons could attempt to reduce future pain and mobility complications. Diminished lumbar lordosis angulation has been shown to correlate with imbalance in the sagittal plane as well as with difficulties in gait when combined with specific fusion strategies (Koller et al., 2019; Wawrose et al., 2020). A predictive model generating the upright 3D spine of patients prior to or during surgery (Oren et al., 2019), whether for traditional fusion or tethering techniques, can help surgeons identify potential complications linked to lumbar degeneration (Yuan et al., 2020), thus improving post-surgical patient posture.

From a clinical objective perspective, it is likely that the combination of a given surgical prone position and a particular surgical plan affects the patient's upright pose. Previous works such as Nault et al. (2013) have shown that imaging and geometric features obtained from the preoperative reconstructed 3D spine, which are associated with posture changes, are correlated with adolescent idiopathic scoliosis-specific surgical outcomes. Further works have attempted to classify various 3D groups of deformity based on variational autoencoders from 3D shape models (Thong et al., 2016) or using deep networks (Mandel et al., 2020) with intervertebral disks and presurgical data. However, given the significant variations in pose during screw insertion and positioning of the pelvis and hips, this prediction task remains challenging. The goal of the current study is to tackle these

* Corresponding author.
E-mail addresses: sylvain.thibeault05@gmail.com (S. Thibeault), samuel.kadoury@polymtl.ca (S. Kadoury).

https://doi.org/10.1016/j.media.2024.103400
Received 30 October 2023; Received in revised form 7 November 2024; Accepted 19 November 2024
Available online 27 November 2024
1361-8415/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

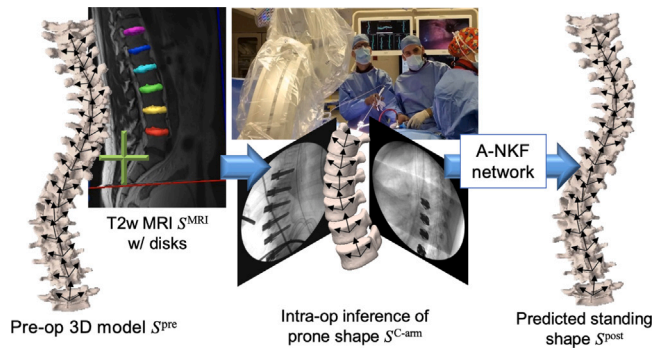

Fig. 1. Illustration of the proposed workflow in the OR, where prior to instrumentation, a 3D preoperative spine S^{pre} obtained in a upright pose along with the diagnostic MRI S^{MRI} and extracted intervertebral disks (IVD) are made available and used as priors (left). Once the patient is placed in a prone position, a 3D intraoperative mesh model $S^{C\text{-}arm}$ is inferred from calibrated biplanar C-arm images (middle). Finally, this intraoperative model is used as input with screw levels and tethered forces to the articulated neural kernel field (A-NKF) and predict the first upright spine pose S^{post} (right).

problems in the operating room (OR) via a shape generation framework, which instead of predicting outcomes, models the post-surgery upright 3D spine directly.

Neural Radiance Fields (NeRFs) have demonstrated tremendous potential in capturing implicit descriptions of shapes in different poses (Mildenhall et al., 2021). In medical imaging, NeRFs were used to generate synthetic views from C-arm projections (Xie et al., 2022). The network would thus provide a representation of the scene and object characteristics, namely density and radiance. Recent approaches extended the concept to 3D, producing volumes and anatomical shapes (Ge et al., 2019), point clouds (Ma et al., 2020), or mesh surfaces (Park et al., 2019). Contrary to explicit registration methods based on deep learning for aligning articulated preoperative data to intraoperative C-arm images (Esfandiari et al., 2019; Zhao et al., 2023), these kernel methods project data into a common latent domain which can then be used to train any generative model with this new feature space, instead of the original high-dimensional image inputs. They may also include a priori constraints (Williams et al., 2021), such as to complete partial shapes (Williams et al., 2022). Still, the complexity in adapting NeRFs for organ rendering applications, as well as the variations in spinal shape and posture, limit the overall adoption.

We propose here a data-driven surgical framework for spine interventions, allowing on-the-fly in the OR prediction of the spine's upright posture after corrective surgery by using both the patient's intraoperative prone pose and particular surgical strategy (Fig. 1).

The main contributions of the paper are as follows. First, we introduce a multi-view inference model based on view-divergent Transformers (Wang et al., 2021) to produce the 3D prone geometry from the 2D C-arm images on the operating table. Second, we introduce a prediction method based on a novel articulated neural kernel field (A-NKF) framework. These regressed kernels allow mapping between intra and post-surgical prone poses, conditioned by encoding the spatial parameters of a given surgical plan. Finally, we propose a pipeline that enables disentangling latent representations of the spine's posture across different poses and infer the first upright – or standing – spine shape with an articulation/shape network.

2. Related works

2.1. Intraoperative 2D/3D spine registration

Several methods were proposed for intraoperative 2D/3D registration to obtain a 3D model, but end-to-end approaches in spine interventions remain an open issue (Markelj et al., 2012; Unberath et al., 2021). Traditional registration techniques using intensity metrics are often solved iteratively via optimization processes. Similarity metrics such as locally weighted mutual information were proposed for local image guidance (Meng et al., 2019), yet remain sensitive to local minima in cases where the baseline registration is inaccurate. The spatial mapping between synthetic and actual X-ray images can also be predicted using convolutional neural networks (CNNs), but remains sensitive to initialization parameters (Miao et al., 2018). More recent supervised learning methods for landmark localization have increased the robustness of 2D/3D registration (Grupp et al., 2020), but still require manual annotations. This is time-consuming especially when dealing with challenging cases (Bier et al., 2018). To improve model robustness in the OR, patient-specific models were developed to refine the landmark localization scheme (Grimm et al., 2021). Regression neural networks were used to infer the pose of slices in relation to preoperative 3D volumes (Lee et al., 2022). Though the concept is simple, real-time performance is limited with respect to the initialization process. Recently, Transformer models with self-attention were shown to provide significant benefits over traditional 2D/3D registration methods, as they are able to infer global shapes without iterative processes or error accumulation (Chen et al., 2022).

2.2. Shape modeling with neural fields

Several types of shape topologies can be described with implicit functions that are both differentiable and continuous with efficient memory footprints (Park et al., 2019; Sitzmann et al., 2019). They attempt to capture 3D shapes as a low dimensional embedding via neural fields. Previous studies have indicated that these disentangled representations within the latent domain help capture 3D shape appearances and improve the synthesis process (Karras et al., 2019; Zhu et al., 2018). While auto-encoders are often used to represent body pose and shape (Tiwari et al., 2022), here we optimize latent codes obtained from a multi-layer perceptron (MLP), representing both mesh shape and global articulation vectors to infer articulated spine structures.

2.3. Prediction of articulated structures

In recent years, 3D shape modeling of articulated objects such as human poses and skeletal shapes has seen a surge in interest. Contrary to the modeling of single structures, priors can be integrated in relation to the geometry and labeling process. These priors can help to reduce the complexity linked to the pose, joint rotation and the dynamic nature of articulated structures (Mu et al., 2021). Predicting the motion of shape constellations seeks to estimate the future configuration of interconnected shapes with a series of 6D parameters (3 translation and 3 rotation) for each object (Weng et al., 2021), which was used namely for robotic control (Mittal et al., 2022). Recent work has demonstrated that implicit functions can be used to capture the temporal changes in unseen articulated structures, generated with their embedding code (Mu et al., 2021). In fact, the disentangled representation of these complex shape variations in a latent domain can help distinguishing individual factors of changes with meaningful global appearance as shown previously (Thibeault et al., 2023), but this work lacks the integration of surgical parameters.

3. Methods

The first step of the pipeline presented in Section 3.1 generates the intraoperative 3D mesh model $S^{C\text{-}arm}$ of the prone spine using a 3D view-divergent Transformer (3DV) network (Fig. 2), which integrates as a prior the diagnostic preoperative 3D spine reconstruction S^{pre}. This 3D mesh model $S^{C\text{-}arm}$ is an input to the second step — the predictive framework (Fig. 3) — which is based on an implicit representation from neural fields. It describes the articulation differences in a disentangled latent domain between prone (intraop) and upright (postop) shapes.

The kernel regression process (Section 3.2) is anchored on neural splines approximating a warping function. The training of this overall A-NKF model is presented in Section 3.3, which uses the regressed kernel within ShapeNet to produce mesh models conditioned on the pedicle screw strategy. Combined with an ArticulationNet model, it produces the regularized articulated upright 3D spine geometry $\mathcal{S}^{\text{post}}$ (Section 3.4). Finally, Section 3.5 describes the inference process to produce the upright spine at the first evaluation after surgery, as shown in Fig. 4.

Compared to previous works, the novelty of this work is three-fold:

1. We develop a 3D view-divergent (3DV) Transformer framework for modeling the 3D spine in prone position. As opposed to previous multi-view works (Wang et al., 2021), we propose a model for generating 3D spine geometries from multiple views incorporating prior knowledge from the preoperative pose in the training process to capture the deformation. Few supervised methods allow producing the prone 3D geometry; some require user inputs (Kadoury et al., 2016), increasing complexity and lengthening the procedure.
2. We propose a neural field based on Thibeault et al. (2023) to predict the postoperative upright spine conditioned on intraoperative (prone) spine from a C-arm scan, surgical parameters, and preoperative MRI by incorporating cross-attention layers prior to inference. Attention between the sets of tokens and the query is computed from the encoded surgical parameters (using a spatial adjacency graph with Laplacian encodings).
3. We demonstrate the clinical efficiency of the prediction tool on prospective surgery patients, with the goal to optimize prone positioning and surgical strategy. We evaluate the method's intraoperative performance on several deformation classes, and assess the clinical parameters used in routine practice.

3.1. 3D prone modeling of the spine from biplanar images

The initial phase of the proposed workflow infers a 3D spine model $\mathcal{S}^{\text{C-arm}}$ in the prone posture of the patient lying on the table before surgery: this is later used as an input to the A-NKF model. The 3D model obtained in the OR is computed from calibrated biplanar C-arm images $\mathcal{I} = \{\mathbf{I}^1, \mathbf{I}^2\}$ with approximately 90° in angulation between the image pair, and a preoperative EOS reconstruction $\mathcal{S}^{\text{pre}}$ of the spine generated automatically using a statistical approach (Humbert et al., 2009). A 3D inference technique anchored on view-divergent enhancing Transformers is used (Wang et al., 2021), as it was shown to be efficient for 3D mesh modeling using multi-views under noisy scenarios. The framework uses as input the multi-view images within a Transformer encoder, concatenating the extracted features and pairings between the different calibrated 2D C-arm images. This integrates generated view embeddings $\mathbf{Z}_0 = [\mathbf{z}^1, \mathbf{z}^2] \in \mathbb{R}^d$ obtained from a decoder based on Transformers processing the $\mathcal{I}$, as shown in Fig. 2.

First, a 2D-view encoder based on a view-divergent Transformer was used, which stacks vision attention layers from multiple heads thus enhancing the divergence between projections (`MH-DeAtt`). It also incorporates a forward propagation network integrating C-arm view embeddings $\mathbf{Z}$ which are position-dependent (`FeedFNet`). The encoder is defined as:

$$\overline{\mathbf{Z}}_l = \texttt{Normalize}(\texttt{MH}-\texttt{DeAtt}(\mathbf{Z}_{l-1}, \mathbf{Z}_0) + \mathbf{Z}_{l-1}) \tag{1}$$

$$\mathbf{Z}_l = \texttt{Normalize}(\texttt{FeedFNet}(\overline{\mathbf{Z}}_l) + \overline{\mathbf{Z}}_l) \tag{2}$$

which includes L normalization layers, with l indicating the basic block where the embedding is used for the encoder's 2D-projection output.

To mitigate the Transformer's inability to explore different associations at deeper levels of the network between projections from `ViewDiver`, the `MH-DeAtt` module slows down the divergence decay in the self-attention layers by maximizing the difference in low-dimensional features from the biplanar views. This Transformer integrates skip connections and combines learned features within the embedding of the input biplanar images, such that:

$$\texttt{MH}-\texttt{DeAtt}(\mathbf{Z}_{l-1}, \mathbf{Z}_0) = \texttt{ViewDiver}(\mathbf{A}, \mathbf{Z}_0)\mathbf{W}_i. \tag{3}$$

The parameter matrix $\mathbf{W}_i \in \mathbb{R}^{(Hd_k+d)\times d}$ includes the linear functions for view i, with d_k the feature dimension of each head, H is the total head count and d is the feature dimension of the view embedding from images $\mathcal{I}$. Here, $\mathbf{A} = (\mathbf{A}^1, \dots, \mathbf{A}^H)$ concatenates the different attention head operations $\mathbf{A}^h$ (defined by attention process on parameter matrix $\mathbf{W}^h$), with h being the `MH-DeAtt` layer's head number.

The proposed Transformer used to generate the prone 3D spine geometry concatenates features extracted from both the decoder of the biplanar views and the probabilistic outputs of each spatial query token. Here, the attention layers of the decoder links the 2D views and the 3D mesh models, namely from the nodes of the output 3D space and the 2D C-arm images used as input. On the other hand, attention layers within the 3D network computes the correlations between target 3D mesh points to capture the overall 3D representation.

To maintain the positional information of the spatial domain, position encodings $\mathbf{E}^{pos}$ are combined with the 3D mesh encodings from N different spines, such that $\mathbf{X}_0 = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N] + \mathbf{E}^{pos}$, with $\mathbf{x}$ being the 3D spine mesh models. Encodings provide information to each 3D vertebra mesh about the localization within the 3D spine. We use sine and cosine functions which transforms the positions into spherical coordinates, as it was previously shown to improve the spatial encoding (Carion et al., 2020). Then, the decoder block integrates the 3D attention, view-dependent attention layers and feed-forward networks:

$$\overline{\mathbf{X}}_l = \texttt{Normalize}(\texttt{MH}-\texttt{VolAttn}(\mathbf{X}_{l-1}) + \mathbf{X}_{l-1}), \tag{4}$$

$$\hat{\mathbf{X}}_l = \texttt{Normalize}(\texttt{MH}-\texttt{ViewVolAttn}(\overline{\mathbf{X}}_l, \mathbf{Z}_L) + \overline{\mathbf{X}}_l), \tag{5}$$

$$\mathbf{X}_l = \texttt{Normalize}(\texttt{FeedFNet}(\hat{\mathbf{X}}_l) + \hat{\mathbf{X}}_l), \tag{6}$$

with `MH-VolAttn` and `MH-ViewVolAttn` as the volume attention and projection-dependent layers of attention, respectively, each having multiple heads. The former includes layers learning global dependencies among different 3D mesh spines, while the latter includes pertinent feature information of all projections and 3D domains.

We introduce *a priori* knowledge about the overall spine to capture the patient-specific geometry, by encoding the upright pre-op 3D model $\mathcal{S}^{\text{pre}}$ reconstructed from biplanar X-rays and concatenating the codes at the input of the 3D model. The pre-surgical shape is used as a condition to the 3D shape embedding, which constraints the possible samples from the latent domain to be extracted only from this space. This helps to condition the inputs of the multi-head volume attention module to only focus on similar preoperative upright shapes. Using a binary cross-entropy loss between the mesh samples and known 3D sample point, the inferred 3D vertebral meshes are restructured to generate a spine $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_m\}$. $\mathbf{s}_m$ is defined as the mesh of a vertebra at a particular m level linked to a vector of articulations:

$$\mathbf{y}_i = [T_1; T_1 \circ T_2; \dots; T_1 \circ T_2 \circ \dots \circ T_m] \tag{7}$$

describing the M rigid registrations T_m with sequential bodies m and $m+1$, with every level capturing 6 degrees of freedom (rotation and translation), described via recursive compositions. Each $\mathbf{y}_i$ contains a series of M intervertebral transformations, describing rigid transformation based on the previous transformations for a given spine. For thoracolumbar spines, $M = 17$. During each epoch of the 3DV model training, the divergence decay — a similarity measure based on multi-view attention — is used to evaluate the divergence between the embedded features. This helps the 3DV model improve the 3D generative process by focusing on the complementary features between the views.

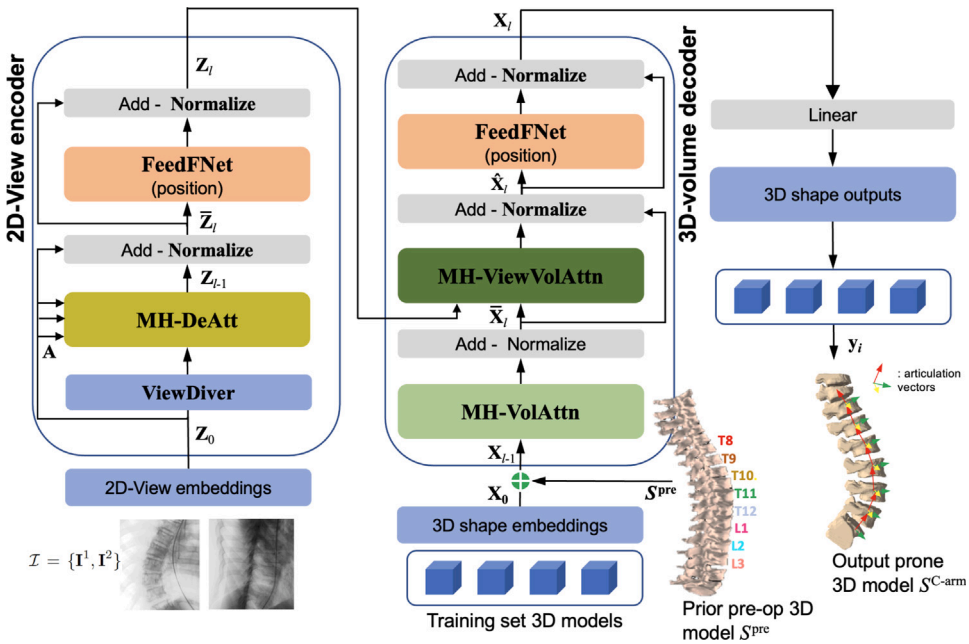

Fig. 2. Overview of the biplanar C-arm 3D view divergent (3DV) Transformer of the spine $S^{\text{C-arm}}$, which is obtained in the prone position with the patient lying on the operating table. The multi-view architecture integrates view-divergence process (ViewDiver) in order to augment 2D and 3D features within latent domain. The plus sign in green indicates element-wise addition.

3.2. Multi-pose shape deformation kernel

In this step, a shape deformation kernel is used to predict the post-op upright articulated spine S^{post} using as input the prone spine model $S^{\text{C-arm}}$ generated in Section 3.1.

To train the shape deformation kernel, every patient (N) had $K = 3$ scans of their spine: before S^{pre} (upright), during $S^{\text{C-arm}}$ (prone) and after S^{post} (upright) surgery. This resulted in a training dataset $\mathcal{S}$ of $P = N \times K$ cases. In other words, each spine sample $S = S_{n,k}$ belongs to a patient $n \in N$ with a particular pose $k \in K$. We express $s_i \in \mathbb{R}^3$ as a 3D vertex obtained with a shape S.

3.2.1. Data-dependent ShapeNet kernel for upright estimation

We use neural spline kernels to take a point cloud from a given spine in the prone posture to be mapped to the first upright position. These are trained from embeddings produced by auto-encoding neural networks. This has the benefit of capturing topological variations between poses while enabling the interpolation of new instances between shapes, as shown in Williams et al. (2022). Here, s_i are the input points linked to a code describing shape $\phi \in \mathbb{R}^D$ in latent domain, and D is the latent dimensionality, assigned to a feature vector $\rho(s_i|\mathcal{S}, \Omega, \tau)$, where ρ is the neural network (C-OcNet Peng et al., 2020, Fig. 3), which is conditioned with the parameters Ω based on the dataset $\mathcal{S}$. Feature vectors trained from the spine models in the upright S^{pre} and prone $S^{\text{C-arm}}$ positions are used with extracted MRI features τ. Intervertebral disks (IVD) are first segmented from the MRI S^{MRI} obtained supine, using an nn-Unet, followed by a feature extraction step in these regions of interest using a ResNet model, used to parameterize the output upright shapes based on the soft-tissue features. The data-dependent kernel is defined as:

$$K_{\mathcal{S},\Omega,\tau}(s_i, s_j) = K_{NS}([s_i : \rho(s_i|\mathcal{S}, \Omega, \tau)], [s_j : \rho(s_j|\mathcal{S}, \Omega, \tau)]) \tag{8}$$

with $[q : r]$ being the feature vectors concatenated together, describing the features from vertebral meshes i and j ($i \neq j$) obtained from pairs of prone and upright poses, while K_{NS} represents the function of the neural spline kernel. In order to produce features from points sampled in both prone and upright domains, we use a Convolutional Occupancy Network (Peng et al., 2020) mapping 3D points a to feature space and learning a 3D reconstruction function. A discretization process of the space around the shape is performed using a 3D grid space, with PointNet (Qi et al., 2017) applied at each grid cell containing input points. PointNet allows extracting features which can be structured in a hierarchy for feature learning. Extracted features are provided to the 3D fully convolutional network (FCN) component, resulting in transformed features with identical size, but in the upright pose. The implicit function of the neural network is defined with parameters α_j linked to every point s_j, with:

$$\alpha = [\alpha_j]_{j=1}^{2P} = (\mathcal{G}(\mathcal{S}, \Omega) + \lambda \mathbf{I})^{-1} \mathbf{y}_j \tag{9}$$

solving a linear system of $2P \times 2P$ to generate articulation vectors in the upright posture ($\mathbf{y}_j$), using as Gram matrix $\mathcal{G}(\mathcal{S}, \Omega)_{ij} = K_{\mathcal{S},\Omega,\tau}(s_i, s_j)$ taking as sample input points s_i and s_j from the training spine models. A regularization term λ is employed to ensure outputs produce realistic shapes. Here, λ is a user-defined parameter that when multiplied to the identity matrix $\mathbf{I}$, can be summed with the Gram matrix. This allows reducing irregular values in the Gram matrix, which comes from the neural spline kernel function. The function estimating unseen samples s is defined as:

$$f_s(s, \phi) = \sum_{s_j \in \mathcal{S}, \phi} \alpha_j K_{\mathcal{S},\Omega,\tau}(s, s_j) \gamma(s_i) \tag{10}$$

describing the ShapeNet network, projecting from the prone space embedding, feature points s onto the upright space embedding. So that feature vectors can be used as inputs to the ShapeNet model, data points s and shape codes ϕ are combined.

3.2.2. Kernel conditioning with surgical parameters

To condition the output upright shapes based on the planned surgical parameters $\gamma(s_i)$ (screw levels, screw tightening forces, screw trajectory and insertion point), the proposed framework parameterizes the neural field by incorporating cross-attention layers. The spatial information is encoded using spatial adjacency graphs that connect tokens, which include the series of surgical parameters for each surgical level. The motivation of graph-based positional encodings comes from the fact spatial adjacency graphs convey the relative position relations between tokens, thereby capturing the positional information. The positional information from the spatial adjacency graphs is encoded

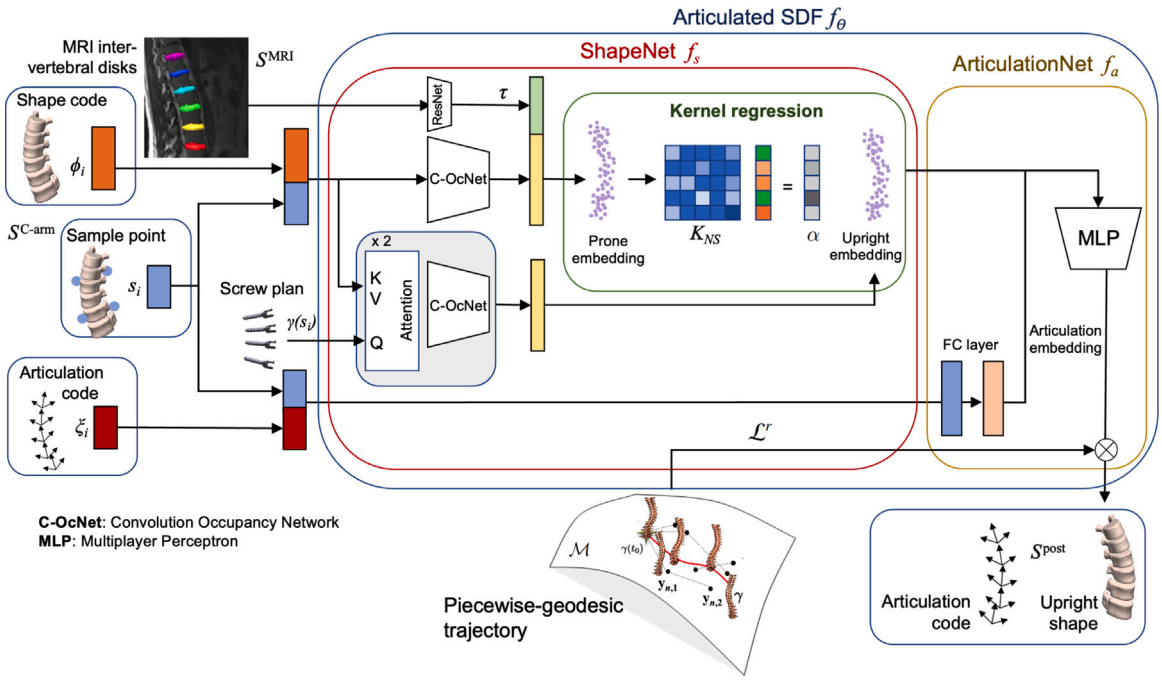

Fig. 3. Illustration of the prediction architecture for the upright spine shape at the first follow-up exam post-surgery. A selection of prone spine models (described by shape code and sample point vertices, in orange and blue boxes respectively) are used to train a neural spline kernel function (K_{NS}). These features (obtained with C-OcNet modules, in yellow) are parameterized with MRI features (green box) obtained from a ResNet, and with the upright spine pose as target, performs a regression of a neural spline kernel K_{NS}, which is conditioned on the screw plan strategy fed to an attention gate Q. The overall method for the articulated signed distance function (A-SDF) is composed of two models (ShapeNet + ArticulationNet) to produce upright spine shapes. First, a point cloud from the upright stance is produced using ShapeNet with the deformation kernel. Then, the point cloud is then provided to ArticulationNet using the articulation code (red box) as input with FC layers, to produce a model with series of connected vertebra. The first FC layers in blue indicate the projection of feature vectors within the embedding, with the latter layers in light orange, showing the classification module. The output of the MLP produces the final upright model.

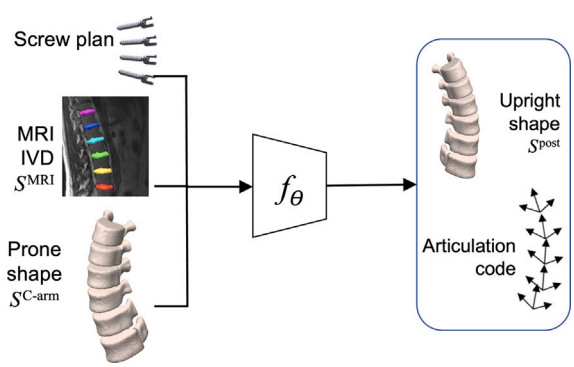

Fig. 4. Inference process of the proposed framework. As input, the model takes the prone 3D model $S^{\text{C-arm}}$, the intervertebral disks (IVD) from MRI model S^{MRI} and screw plan. The data is provided to the trained articulated SDF function (f_Θ), and generates the upright shape S^{post} and articulation code. During testing of the trained model, a joint inference process produces both shape and articulation codes for the post-operative upright poses.

with Laplacian positional encodings, using Laplacian eigenvectors as the positional encodings (Kreuzer et al., 2021). Specifically, a graph Laplacian $\boldsymbol{L}$ is defined from the k top eigenvalues selected from the graph edges. The Laplacian eigenvectors constitute a local coordinate system that retains the overall structure of the graph. Thus, the ith row of eigenvector matrix can be used as the positional encoding for each node of the graph. Attention between the sets of token and the query extracted from the encoded surgical parameters $\gamma(s_i)$ is achieved by means of concatenating the set of features (Rebain et al., 2022). This decomposes the latent code with tokens of constant width, offering a unique advantage by efficiently conditioning the high-dimensional shape generation with regards to the location and tension information of the screws. Encoded sets of latent tokens are generated with a shared network instead of performing individual predictions from each token.

3.2.3. Training of the ShapeNet kernel

Training of the kernel regression model is done with the set $\mathcal{S}$, where the input points estimate K_{NS}, with s_i^{3D} and $\mathbf{y}_i^{3D}$ as the input and output spine vertices within the dense 3D space, respectively. We note that the input points are only required in the supervised process for training. The output labels indicate the set of points:

$$\mathbf{y}_i^{3D} \begin{cases} 1 \text{ if } s_i^{3D} \text{ is inside the shape,} \\ 0 \text{ otherwise}. \end{cases} \tag{11}$$

The network's kernel is trained by estimating the implicit function, using the input spine shapes $\mathcal{S}$ and estimating the positions in the dense volume on the vertebra meshes $\mathbf{s}_m$:

$$L(f_s) = \sum_{i=1} CE(f(s_i^{3D}), \mathbf{y}_i^{3D}) + \lambda_{L1} \sum_{m=1} |f(\mathbf{s}_m)|. \tag{12}$$

The first loss term is the cross entropy (CE) estimating the occupancy within the target volume, while the second loss is the change in posture of the spine based on the alignment of sample vertices, weighted by λ_{L1}. The function $f(\mathbf{s}_m)$ is the mapping function that integrates the regression kernel K_{NS}, and is trained within the ShapeNet module. It is the target function that learns how to transform the input prone mesh models to the output post-surgical upright mesh model. Gradients are backpropagated using the global loss term which updates the network's weights, leading to a kernel which is data dependent.

3.3. Articulation SDF

The final phase of training is the A-NKF model, where the shape's articulation code is modified for each instance while the shape instance remains constant. Every shape instance $S_{n,k}$ is linked to $\xi \in \mathbb{R}^d$,

defining the embedded vector of articulations. For all n instances of different articulation vectors ξ_i (based on $\mathbf{y}_i$ defined in Eq.(1)), the shape embedding ϕ_n remains the same.

Inspired by the ArticulationNet model which proposed a general framework for inferring articulated shapes (Mu et al., 2021), we use the previous regressed kernel shape function f_s, based on the articulated signed distance function (A-SDF) obtained from sample vertices $s \in \mathbb{R}^3$ of the input shape. Here, ϕ and ξ represents, respectively, the code associated to shape and vertebral articulations obtained with an MLP. These articulation vectors $\mathbf{y}_i$ are used to create an embedding, with each individual sample ξ_i representing articulation codes. The shape code ϕ for each sample is obtained from the prior shape embedding, which is created from the collection of shapes S.

Auto-encoders (based on C-OcNet Peng et al., 2020 and MLPs, see Fig. 3) are used to implement the A-SDF for the spine shape generated with f_θ. The model incorporates f_s (the shape kernel function), in addition to f_a defined previously with ArticulationNet:

$$f_\theta(s, \phi, \xi) = f_a[f_s(s, \phi), s, \xi] = o. \tag{13}$$

Here, $o \in \mathbb{R}$ defines the generated mesh's SDF output value. Sampled points are determined to be inside the 3D spine shape, $\mathbf{s}_m$, if $f_\theta(.) > 0$.

The parameters of the A-SDF model and resulting shape codes are learned through articulation vectors ξ from ground-truth data (series of rigid transforms). Similarly, articulation codes ξ and sample points s are also concatenated. These are used to predict, for each sample point s, the SDF value vector from the ArticulationNet model, which takes as input in the first FC layers the feature vectors obtained from the latent domain describing the shape and articulation codes. Finally in the last hidden layer, a classification module (Softmax activation) is included to generate the vertebral series, by distinguishing the constellation of vertebral shapes $\mathbf{s}_m$, with $\mathbf{p}_m$ representing the level of the vertebra for points in $\mathbf{s}_m$. The $L1$ distance is used for the loss term, which performs a regression on the SDF values of V sample vertices, representing every spine model based on the f_θ function:

$$\mathcal{L}^s = \frac{1}{V}\sum_{v=1}^{V} \|f_\theta(s_v, \phi, \xi) - o_v\|_1 \tag{14}$$

with $s_v \in S$ describing a mesh vertex in the shape embedding, while o_v is the ground-truth SDF output for $v \in \{1, \dots, V\}$. The alignment of the series of vertebrae with the output rigid registrations represents the second loss term:

$$\mathcal{L}^p = \frac{1}{V}\sum_{v=1}^{V} [CE(f_\theta(s_v, \phi, \xi), p_v)]. \tag{15}$$

The loss term classifies the different constellations of vertebral shape $\mathbf{s}$, with p_v representing the level of the vertebral for point s_v. The proposed regressed neural spline kernel learns deformations across pairs of spine poses, regardless of the number of vertebrae in a tethered segment or field of view across the dataset, and is thus not limited to a specific segment size. Furthermore, because the kernel is parameterized with attention mechanisms on the surgical strategy, including the number of tethered segments, it allows including different vertebral levels in the training dataset.

3.4. Global shape regularization and overall loss

In order to produce spine shapes which are anatomically correct, we regularize outputs by ensuring predictions are mapped in the vicinity of a locally-linear geodesic path described in a spatio-temporal embedding. The loss term, denoted as $\mathcal{L}^r$, integrates a pre-trained geodesic trajectory, defining the longitudinal geometric spine changes after surgical treatment, initially proposed in Mandel et al. (2019). The embedded geodesic trajectory uses adjacent feature samples within the population of post-surgery cases, thus producing locally linear fields $\mathcal{N}(\mathbf{y}_i)$ capturing the pre- to postop variations of the shape constellation. The time component to the outcome prediction, depending on the gap between the time of surgery and the postoperative visit, is captured by the geodesic trajectory. The trajectory was pre-trained on longitudinal data of pre- and post-surgical cases (Mandel et al., 2020) and is parameterized by the time component. This helps constraining the outputs with regular shapes based on temporal variables. From data points $\mathbf{y}_i$ produced within the Riemannian field, i denoting a predicted case with constant time intervals, a regression of new data points is performed in $\mathbb{R}^D$. This follows the assumption that features represent the global shape changes in a surgical population, covered with the geodesic trajectory. We use a regression method to implement the piecewise-geodesic path, minimizing the geodesic distances between mapped data points and their location in the latent domain of longitudinal shapes (Boumal and Absil, 2011), in addition to a Euclidean distance that integrates velocity and acceleration components of the inferred path. These two measures produce smooth paths between the various target points and capture the series of rigid inter-vertebral transforms associated with the pose change with continuous transformations when the spine's global shape changes.

The global loss term is formulated as:

$$\mathcal{L}(S, \phi, \xi) = \mathcal{L}^s(S, \phi, \xi) + \beta_p \mathcal{L}^p(S, \phi, \xi) + \beta_\theta \mathcal{L}^r \tag{16}$$

where the importance of the constellation and smoothness terms are controlled by β_p and β_θ, respectively. The model's shape codes are randomly initialized using a normal distributions at training. Then, every shape code is optimized for all the articulated instances in the training set. Therefore, for all $N \times K$ instances, the overall loss function minimizes the following:

$$\arg\min_{\theta,\xi} \sum_{n=1}^{N}\sum_{k=1}^{K} \mathcal{L}(S_{n,k}, \phi_n, \xi_k) \tag{17}$$

where θ are the network parameters.

3.5. Upright shape inference

At inference, the 3D spine model $S^{\text{C-arm}}$ generated in the OR (obtained in 3.1) is provided as input with the surgical data $\gamma(s_i)$, and produces the upright shape model, in addition to the articulation vector, shown in Fig. 4. A random initialization of both codes (ϕ and ξ) representing shape as well articulation parameters is performed, and the parameters of the network remain constant. This is achieved by setting the objective function to:

$$\phi, \xi = \arg\min_{\phi,\xi} \mathcal{L}(S, \phi, \xi). \tag{18}$$

To avoid local minima of the term in Eq. (18) from the gradient-based optimization which can produce inaccurate results, codes related to articulations are first approximated and produced shapes are retained. This allows capturing the overall shape appearances. This is due to the fact that the articulation code typically converges to a good estimate first, while the obtained vertebral codes leads to noisy results. Then, based on the determined representation of the articulation vectors, we re-initialize the shape code ϕ, which is sent to the optimization process for the final shape inference.

4. Results

4.1. Dataset

We used a training set of 735 surgical patients for the predictive articulated neural kernel field (A-NKF) framework. The data consisted of thoracolumbar 3D models acquired before, during and after surgery. Patients had a preoperative mean Cobb angle of 48° (range of [35°,65°]), with postoperative visits 14 days after surgery. Pre- and first-upright 3D spine shapes were produced using the multi-view C-arm images based on a stereo-vision system (SterEOS+ software, Paris, France), with a learning-based technique to generate a complete

annotated spine model (Humbert et al., 2009). All scoliotic cases were treated with corrective spine surgery, and landmarks on the endplates and pedicles of vertebrae were validated by a trained radiologist, using the SterEOS+ software (Paris, France). To train the 3D View divergent Transformer (3DV) intraoperative inference model, ground-truth 3D reconstructions (sets of consecutive 3D vertebral meshes composing the spine) were obtained with a 3D stereo-triangulation method in the prone position, using intraoperative biplanar X-ray images which were annotated. These landmarks allowed defining a local coordinate system for each level, as well as warp high-resolution meshes obtained from cadaveric CT models and produced detailed geometric representations of the spine. These cadaveric CT models were obtained from a previous study that evaluated the morphological differences in vertebral anatomies, creating generic templates of each vertebral level (Kadoury et al., 2013). A warping process then mapped the template mesh to specific anatomical landmarks, located on the pedicle extremities and center of endplates. This was shown to be very accurate compared to ground-truth CT models (≤ 1 mm).

Once patients were positioned in a prone position on the OR table, coronal and sagittal X-rays were obtained using the C-arm before the start of surgery. The multi-view fluoroscopic images allowed producing the input 3D prone shape of the instrumented spine segment. Patients also had a T2 weighted turbo spin-echo 3D MRI of the lower-back (resolution of $2.5 \times 1.0 \times 1.0$ mm^3). Acquisitions were performed on a Siemens 1.5T MRI scanners 2 days before surgery. Then for each scan IVDs were automatically segmented in the lumbar region using a pre-trained nn-Unet approach (Isensee et al., 2021) trained on the in-house data.

4.2. Training procedure

For the 3DV Transformer inference model, a five-fold cross-validation approach was used for training and testing the model on a subset of 360 cases with patients from five deformation classes (1-hyper-kyphosis, 2-right-thoracic, 3-left-lumbar, 4-right-thoracic-left-lumbar and 5-left-thoracic.). We set the batch size at 64, and used resized 225×225 images for training. For the A-NKF framework, a dataset of 652 models was used for training and validation (80:20), with a separate hold-out set of 83 cases were used for testing. Here, a batch size of 32, a learning rate of 0.003, along with a dropout rate of 0.4, and $\beta_\theta = 0.5$, $\beta_p = 0.4$, $\lambda = 0.25$ and $\lambda_{L1} = 0.6$ were determined through an exhaustive grid search. For both models, the AdamW optimizer was adopted (Loshchilov and Hutter, 2017). The convergence criteria was when the mean square error fell below the value of 0.1 mm. Hyperparameters were determined through an exhaustive grid search, while using an exponential learning rate decay, with a decay rate of 0.9 and step size of 25. To make sure no overfitting was present when training the model, the validation folds were used to ensure the model fitting to out of distribution (OOD) samples.

The average run-time performance of the model was 0.4 s for the shape alignment step and 0.9 s for the inference of the predicted 3D model in the upright position. The training was performed on a workstation with a 3.50 GHz processor, 64 GB RAM and a GPU NVIDIA A100 GPU, taking around 10k iterations (roughly 20 h).

The source code of the proposed model is made publicly available at https://github.com/skadoury/spineA-SDF.

4.3. Multi-view 3D model inference

We first evaluated the intraoperative prone spine model generated from the 3DV Transformer-based model, using a five-fold cross-validation procedure, where four folds were used for the training, and the model tested on the fifth fold of 72 patients, repeating the process 5 times. The performance was compared to two other 2D/3D registration approaches which were adapted for this purpose: (1) a landmark-based neural network for 2D/3D registration (LNN) (Grupp et al., 2020) and (2) a pose estimation model based on a weighted regression network (WRN) (Grimm et al., 2021). For both approaches, once the preoperative CT was registered to fluoro images, the vertebral shapes were generated. This was used to evaluate the reliability of the proposed multi-view Transformer-based model. Manual landmark annotations identified by trained radiologists on biplanar C-arm images using the Slicer3D software served as ground-truth models for comparative purposes.

Table 1
Accuracy evaluation of the 3D View Divergent (3DV) Transformer using intraoperative C-arm images obtained after spine positioning, comparing the mean vertebral contour distances (in mm) of the proposed multi-view Transformer model with other approaches. The method was assessed on a dataset assembling between 5 deformation types, such as hyper-kyphosis (C1), right-thoracic (C2), left-lumbar (C3), right-thoracic-left-lumbar (C4) and left-thoracic (C5). The proposed approach (3DV) is compared with a landmark-based neural network (LNN) (Grupp et al., 2020) and a weighted regression network (WRN) (Grimm et al., 2021). All metrics are distance functions where lower is better.

	MOD (T_t) (mm)			OE (T_R) (^{o})			3D RMS (mm)		
	LNN	WRN	**3DV**	LNN	WRN	**3DV**	LNN	WRN	**3DV**
C1	0.76	0.79	**0.38**	**0.49**	0.51	0.65	0.72	0.79	**0.49**
C2	1.02	0.99	**0.70**	0.91	0.96	**0.90**	0.84	0.89	**0.66**
C3	1.43	1.54	**0.63**	1.08	1.50	**0.69**	1.12	1.39	**0.80**
C4	1.84	1.97	**0.77**	0.91	1.26	**0.89**	1.15	1.47	**0.85**
C5	1.35	1.11	**0.28**	**1.20**	1.42	1.21	0.96	1.01	**0.69**

Measurements such as the orientation error (OE) obtained in degrees (^{o}) the magnitude of differences (MOD), based on the maximal error (mm) and mean centroid distance (MCD) were calculated to assess the generation of the intraoperative 3D model $S^{\text{C-arm}}$ used as input for the A-NKF.

Results are shown in Table 1 for all 5 surgical groups of scoliosis, with an example of the 3D inference process shown in Fig. 5. The proposed model yields an overall 3D RMS error of 0.7 ± 0.3 mm, a MOD (based on overlap of GT and generated vertebral meshes) of 0.54 ± 0.2 mm and a difference of $0.9 \pm 0.3°$ in the orientation values. The levels of error are in the clinically acceptable ranges compared to other works, with a statistical improvement using Student-t tests. As the more frequent type of deformity for surgery is C2, results in this class show acceptable levels of error, indicating the model's ability to preserve the natural spine's articulated pose. These landmark errors fall within the tolerable range of errors, which are below 2 mm based on previous studies documenting the required level of accuracy for surgical guidance (Helm et al., 2015). Clinically applicable methods show errors between 1 and 2 mm in accuracy, which is needed to properly identify anatomical landmarks such as pedicles for screw insertion.

We also performed ablation experiments on the proposed 3DV model, to quantitatively evaluate the contribution of the attention mechanism, the multi-view-divergence layers, as well as the inclusion of the prior. Results in Fig. 6 shows that the enhancing view-divergence mechanisms plays an important part in improving the overall accuracy of the 3DV model. It also benefits the overall convergence when handling multiple views in the different layers of the enhancing function, which plays an essential role in improving the proposed enhancing volume Transformer performance.

4.4. Predictive model performance

In the next set of experiments, the accuracy of the predicted upright spine S^{post} was assessed based on a separate set of 83 AIS patients with various scoliotic profiles (18 hyper-kyphosis (C1), 24 with right-thoracic (C2), 19 with left-lumbar (C3), 15 with right-thoracic-left-lumbar (C4) and 7 left-thoracic (C5)) and varying bone maturity statuses, undergoing minimally invasive corrective spine surgery. Results are presented in Table 2, with Fig. 7 illustrating sample predictions in two cases. Errors in the annotated 3D landmarks were assessed

Table 2
3D RMS errors (mm), IoU (%) Chamfer distance, Cobb angle difference (°) and lordosis angle difference (°) for the proposed articulated neural kernel field (A-NKF) method, compared to a biomechanical registration model (Jobidon et al., 2019), ST-ResNet (Zhang et al., 2017), Convolutional Occupancy Network (Peng et al., 2020), DenseSDF (Park et al., 2019), NeuralPull (Ma et al., 2020) and ST-Manifold (Mandel et al., 2020). Upright 3D models acquired from multi-view X-ray images at first follow-up are used as ground-truth for comparative purposes. Ablation results are also shown in the second section.

	3D RMS (mm) ↓	IoU (%) ↑	Chamfer ↓	Cobb (°) ↓	Lordosis (°) ↓
Biomechanical (Jobidon et al., 2019)	5.7 ± 3.0	83.6 ± 3.1	7.6 ± 3.4	5.0 ± 3.1	4.8 ± 3.1
ST-ResNet (Zhang et al., 2017)	6.3 ± 3.8	78.5 ± 3.6	8.3 ± 4.2	5.6 ± 3.9	5.9 ± 4.2
C-OcNet (Peng et al., 2020)	4.6 ± 2.8	81.2 ± 3.4	7.0 ± 3.6	4.3 ± 2.9	4.8 ± 4.4
DeepSDF (Park et al., 2019)	3.8 ± 2.1	84.5 ± 2.8	6.1 ± 2.2	3.9 ± 2.4	4.2 ± 3.9
NeuralPull (Ma et al., 2020)	3.1 ± 1.4	86.9 ± 1.9	5.3 ± 1.8	3.7 ± 2.0	3.9 ± 3.6
ST-Manifold (Mandel et al., 2020)	3.0 ± 1.2	87.2 ± 1.5	5.5 ± 1.6	3.5 ± 1.8	3.8 ± 3.0
Standard NKF	3.7 ± 1.8	82.2 ± 3.1	5.8 ± 2.5	4.0 ± 3.7	4.2 ± 3.6
A-NKF	2.8 ± 1.0	86.6 ± 1.4	5.0 ± 1.8	3.3 ± 2.0	3.7 ± 3.0
A-NKF + MRI	2.4 ± 0.8	89.7 ± 1.1	4.5 ± 1.6	2.9 ± 1.7	3.2 ± 2.6
A-NKF + regular. + MRI	1.8 ± 0.4	93.2 ± 2.4	3.9 ± 0.4	2.2 ± 1.7	2.5 ± 2.1
A-NKF + kernel + MRI	1.5 ± 0.5	94.3 ± 2.0	3.8 ± 0.6	2.1 ± 1.5	2.3 ± 1.7
A-NKF + kernel + regular. + MRI	**1.3 ± 0.4**	**95.9 ± 1.8**	**3.1 ± 0.4**	**1.8 ± 1.0**	**2.0 ± 1.4**

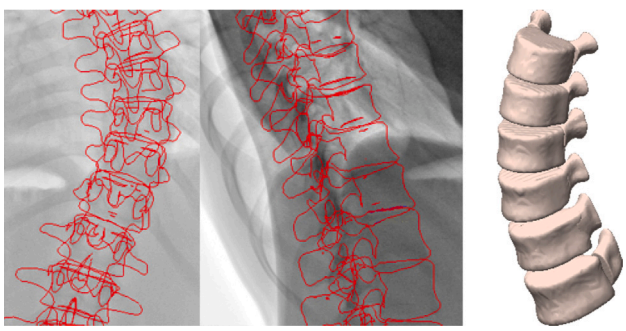

Fig. 5. Example of an output intraoperative 3D View divergent Transformer (3DV) model. The model uses biplanar C-arm images obtained prior to instrumentation with the spine in a lying position, and integrates the preoperative 3D model. Results show the 3D model $S^{\text{C-arm}}$ and projection of the full spine model on biplanar images.

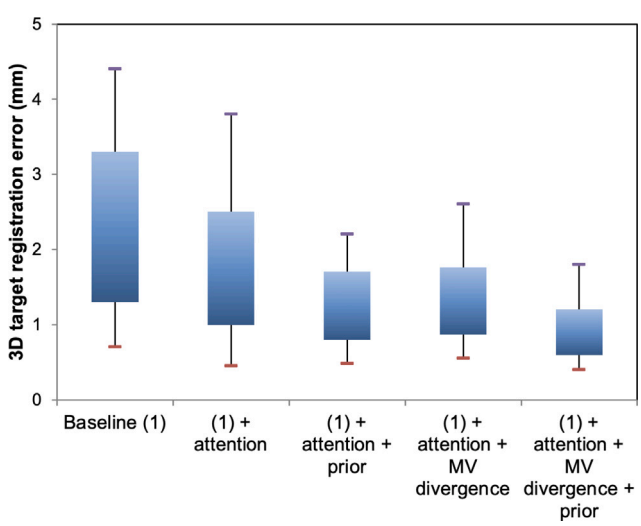

Fig. 6. Ablation experiment of the 3DV method, comparing the baseline (1) with the addition of a prior model, attention, and the view divergence mechanisms. The evaluation is based on 3D target registration errors from landmarks on the spine.

for every predicted upright spine shape, in addition to the Chamfer distance and Intersection over Union (IoU) measures. Errors in Cobb and lordosis angles were also determined for all predicted models in the test set, with comparative results to several state-of-the art spatio-temporal and neural fields models. We also conducted ablation studies, where statistically significant improvements ($p < 0.05$) can be observed in the overall accuracy when integrating the neural spline kernel and regularization term. Fig. 8 evaluates the model's performance by reporting the 3D RMS errors with respect to the number of sample points extracted from the input prone shape, as well as to number of vertebrae in tethered segments. From these results, we can notice the use of sparser point sets leads to higher mean errors (2.5 mm) with 250 points. However, using denser inputs (m=1000) provides an important gain in accuracy when using the conditioning neural fields.

In order to be applicable in a clinical setting, 3D digital models of the spine must yield an accuracy in the range of 1 to 2 mm, as identified by landmarks which are used for the insertion of pedicle screws (Helm et al., 2015) or for postoperative clinical assessment. These accuracy requirements are particularly crucial for minimally invasive surgical procedures of the spine. The overall higher accuracy of 1.3 mm demonstrated in this study compared to other state of the art methods may improve the adoption of predictive methods used intraoperatively. Indeed, leveraging data from preoperative models with surgical plans can potentially improve the positioning of the spine. Combined with image-guided interventions to place pedicle screws, optimal positioning and screw placement can lead to improved scoliosis surgery outcomes (Tjardes et al., 2010). Lately, data-driven methods offered similar levels of accuracy to biomechanical simulators replicating the dynamic behaviors of the spine (Kadoury et al., 2013), and integrating physiological considerations (Jobidon et al., 2019).

Finally, Fig. 9 illustrates the overall 3D vertebral landmark error distributions from the inferred upright shapes, in addition to the vertebral shape Dice scores compared to ground-truth models when using different sets of inputs. This includes the geometric shape features and articulated pose from the extracted IVD, integrating soft tissue and stiffness information into the regression problem.

4.5. Postoperative clinical assessment

As a final experiment, we compared the clinical parameters from the actual 3D models obtained at the postoperative evaluations, and the predicted 3D models which were obtained at the first-upright examination following surgery. This helped to assess the performance range of the method, and identify more challenging cases or scenarios for the model. Results are presented in Table 3 with the series of clinical parameters used in practice. We used Wilcoxon tests with paired analysis, and found no significant difference in all major angulation indices in the coronal plane (C^{PT}_{PA}, C^{MT}_{PA}, C^{L}_{PA}). The sagittal indices (C^{T4-T12}_{LAT}, C^{L1-L5}_{LAT}), reflecting kyphosis and lordosis angles, also showed no significant difference to the real first-upright model. The parameters obtained from the planes of maximal deformity (θ^{PT}_{PMC}, θ^{MT}_{PMC}, θ^{L}_{PMC}) yielded error levels which were slightly higher, however these remain insignificant based on the statistical analysis. Lower variability indicates the robustness across different scoliotic profiles and

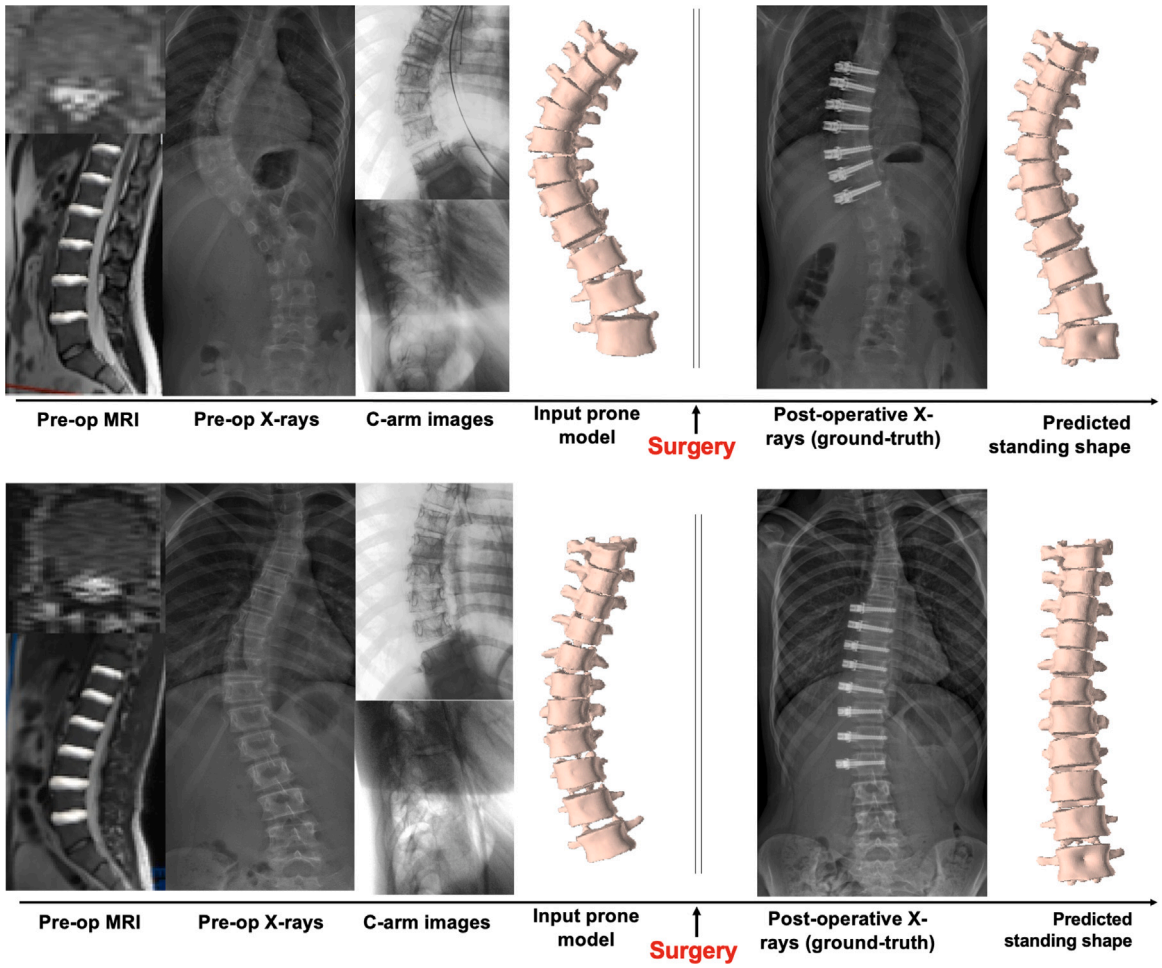

Fig. 7. Examples of predicted 3D spine shape models of the instrumented segment after minimally invasive spine surgery. Both samples illustrate the input 3D prone shape, C-arm images and MRI on the left, and the predicted spine model on the right, alongside the first follow-up EOS images.

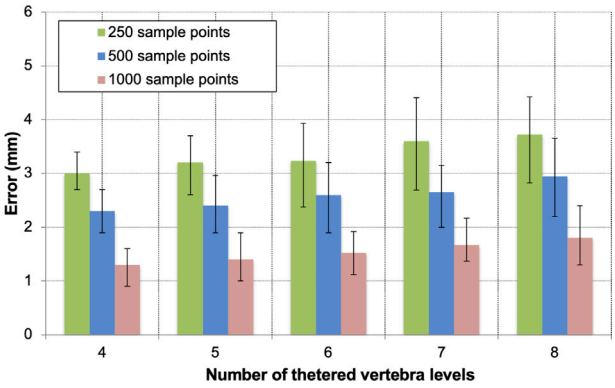

Fig. 8. Evaluation of predicted model accuracy based on the overall length of tethered segments and point sample size extracted from input data.

classes. Clinically acceptable levels of error remain below the 5 deg threshold, indicating that the predicted models can be used in a clinical setting, particularly for evaluating the immediate upright posture and preservation of lordosis. It should be noted that the lower variability in the computed Cobb angles can be explained by the fact measurements extracted from 3D models offer improved repeatability compared to manual measurements which offer lower reliability (Illés and Somoskeöy, 2013) (Tauchi et al., 2016). Finally, local parameters such as axial rotation show little differences to the ground-truth first follow-up examinations, indicating that local torsion and correction aspects of surgery are also captured with the predictive model, demonstrating no statistically significant difference.

We also evaluated the method's overall performance with respect to the deformity class (right thoracic, thoracolumbar and lumbar), bone maturity and overall balance. Results shown in Table 4 indicates the method has more difficulty with cases that have Risser grade above 3, indicating that highly mature patients have less predictable results due to complication with the tethering approach. Furthermore, we can observe cases with applied tethered forces over 300N produce less reliable results due to instabilities in segmental regions. Fig. 10 provides an example of such a scenario.

In order to evaluate the capability of the network to predict shape changes based on the intraoperative pose with different outcome stances, the network was retrained with the same patient dataset, using as target the lateral bending examinations ($N = 543$) which are also acquired at the first follow-up to estimate spine flexibility. Table 3 presents these experiments, while sample results are shown in Fig. 11. Both the architecture and strategy for re-training the network were identical to the configuration with an upright pose, however the weights were optimized for the bending shape targets using the ground-truth data. The shape inference accuracy was evaluated at 2.0 mm based on landmark localization errors and angular measurement errors were all

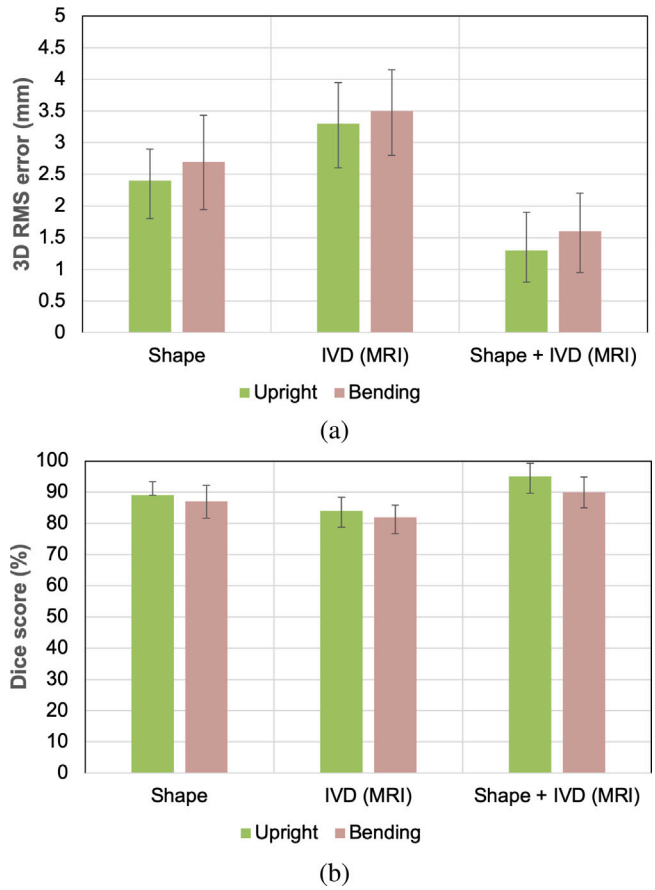

Fig. 9. Performance of the A-NKF model using combinations of shape and intervertebral disk (IVD) during training and testing. (a) 3D error plots for landmark localization. (b) Dice scores evaluation on target upright shapes.

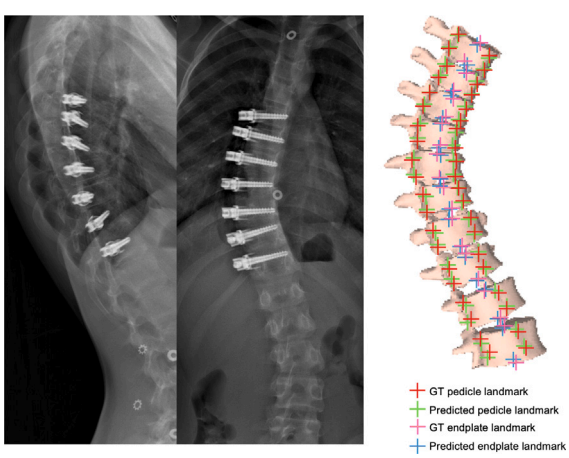

Fig. 10. Sample prediction case with tethered level over 300N and Risser grade of 4, yielding a geometric shape with angular difference of 10 deg and falling within in the 90th percentile of error distribution.

below 5deg, which remains acceptable for clinical assessment at follow-up. These experiments demonstrate the framework's ability to predict the dynamic behavior of the spine's shape, even though the errors are slightly higher compared to predicting the first-upright shape. This can therefore be used for multi-body structure flexibility prior to instrumentation.

Table 3
Differences (mean and standard deviation) in geometric measurements used for clinical assessment, with the geometrical models obtained at the first follow-up evaluation and the predicted shape. Statistical significance ($p < 0.05$) based on paired Wilcoxon tests (*).

Parameter	Symbol	Mean diff.(°) (Upright)	Mean diff.(°) (Side-Bending)
Cobb angle (PT)	C_{PA}^{PT}	0.8 ± 0.4	1.1 ± 0.6
Cobb angle (MT)	C_{PA}^{MT}	0.5 ± 0.3	0.7 ± 0.5
Cobb angle (L)	C_{PA}^{L}	1.1 ± 0.7	1.3 ± 0.7
Kyphosis	C_{LAT}^{T4-T12}	1.4 ± 0.8	1.8 ± 1.2
Lordosis	C_{LAT}^{L1-L5}	2.5 ± 1.1	2.8 ± 1.3
Max. deform. (PT)	θ_{PMC}^{PT}	2.4 ± 1.0	2.6 ± 1.0
Max. deform. (MT)	θ_{PMC}^{MT}	2.3 ± 1.0	2.5 ± 1.0
Max. deform. (L)	θ_{PMC}^{L}	2.9 ± 1.1	3.0 ± 1.4
Axial rotation	θ_{APEX}^{MT}	0.9 ± 0.7	1.2 ± 0.8
Frontal balance	y_{T1-L5}	3.1 ± 1.3*	NA
Sagittal balance	x_{T1-L5}	3.8 ± 2.0*	NA

Table 4
Clinical analysis of the method performance based on deformation types (hyper-kyphosis (C1), right-thoracic (C2), left-lumbar (C3), right-thoracic-left-lumbar (C4) and left-thoracic (C5)), bone maturity based on Risser grade and average screw tightening force applied on tethers.

	Risser grade			Screw tightening (N)	
	0–1	2–3	4–5	<300	> 300
C1	1.3 ± 0.3	1.5 ± 0.4	1.6 ± 0.4	1.4 ± 0.5	1.7 ± 0.6
C2	1.4 ± 0.4	1.6 ± 0.5	1.7 ± 0.4	1.5 ± 0.6	1.8 ± 0.6
C3	1.5 ± 0.5	1.7 ± 0.5	1.8 ± 0.4	1.5 ± 0.6	1.9 ± 0.7
C4	1.7 ± 0.5	1.9 ± 0.5	2.1 ± 0.6	1.8 ± 0.7	2.2 ± 0.8
C5	1.5 ± 0.4	1.6 ± 0.5	1.7 ± 0.4	1.5 ± 0.6	1.8 ± 0.6

5. Applications and extensions

We introduced a method to predict the upright spine posture after surgery using intraoperative data, namely from the inferred 3D prone spine model and extracted surgical parameters, such as vertebral screws and tightening forces on the tethers. We have demonstrated the competitiveness and robustness of the A-NKF framework on a cohort adolescents with idiopathic scoliosis treated with surgery.

Neural fields vs SOTA approaches. The proposed model offers the ability to (1) learn a disentangled latent domain of various spine poses within an embedded space; and to (2) produce a target shape pose using a sampling scheme of points which is conditional on external environment parameters, such as surgical strategy. Training the predictive regression kernel requires more memory and is computationally expensive. The method focuses on both upright and bending poses after surgery. As a consequence, when over 250 points are sampled from the input shape space in the prone posture, the proposed model outperforms other deep learning or biomechanical-based methods. This implies that a larger and more optimal inference network, relying solely on fully sampled geometric models, yields better performance than a smaller model provided with sparse representations.

This observation indicates that a trade-off must be made between training a bigger model and training a model with less sample points per model. However, when the input articulated model data is sparse, the model still preserves the overall shape with acceptable levels of errors (< 2 mm), even with a smaller network. Indeed, in this scenario, training the predictive network on target meshes becomes crucial as it has access to fewer data samples to train the kernel and the generated target prone model may be less reliable. Therefore we claim that the A-NKF has the potential to alleviate the landmark annotation process, which is a burden in intraoperative guidance workflows.

While the MRI was of sufficient resolution and quality, the extraction of intervertebral disks (IVD) allowed to parameterize the neural

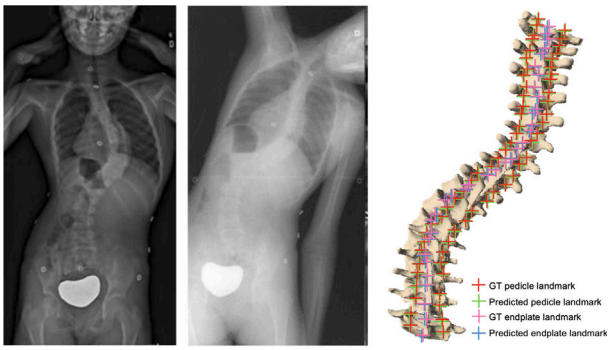

Fig. 11. Upright and sagittal side X-ray images used for the prediction of a 3D spine model in the bending position.

field of inter-vertebral articulations specifically. The model therefore requires the segmentation of the IVDs (in this case with an nn-Unet) from the diagnostic scan; however this step was shown to be accurate with Dice scores over 0.9 (Isensee et al., 2021). Furthermore the fact that the MRI is acquired in a supine position made it not possible for the MRI to be used as a prior in the 3D inference component. Standard geometric methods (non deep learning methods) have shown encouraging performance in recent studies (Jobidon et al., 2019). However these are highly dependent on explicit registration steps and are prone to misalignment. Thus the proposed approach offers an end-to-end alternative to the inference process.

Limitations. Surgical workflows and practices greatly influence patient outcomes. Due to the fact that the model was trained on data from a single institution with only 12 thoracic and 5 lumbar vertebrae (excluding transitional vertebrae), which follows similar practices in terms of patient positioning and AVT tethering, conditions were ideal: the training set is homogeneous with respect to strategies or surgical preferences. It was also acquired with the same imaging equipment. As the pathologies studied in this paper mostly showed deformities which are more easily categorized in one of five deformation classes, other scenarios such as with adults may be more complex. Bone maturity would no longer play a predominant factor, while degenerative cases such as osteoporosis and potential bone fracture should be studied. However, even though our experiments did not include these cases, we believe our results are representative of the actual behavior of the different models.

The A-NKF pipeline has demonstrated encouraging performance in the challenging task of upright shape prediction in the first follow-up exam. However, the multiple training shapes for the regression kernel which uses 1000 points for training different models (ShapeNet and ArticulationNet) leads to high training times and consequent computational resources. This is especially true for representing complex shapes or in longer segments where point densities will be higher.

MRI is increasingly used for surgery planning to visualize the soft tissues, in addition to the routine preoperative X-rays, to evaluate the global shape of the spine and obtain measurements. Recent studies have shown the increased use of MRI for identifying potential complications (Keshavarzi et al., 2023). On the other hand, biplanar EOS images for scoliosis has been a common practice for decades, as it is the preferred modality to the standard X-rays or CT due to the reduced dose (Hui et al., 2016). The practical limitations of transposing our method to the routine clinical practice are the computational resources required for the inference of the 3DV and A-NKF framework, which may prohibit generating the results in less than 3 seconds. Our method can be adapted to other types of surgeries, such as lateral approaches for spinal instrumentation. Models would need to be re-trained for these types of surgeries.

The prediction of outcomes also strongly depends on the positioning of the spine on the operating table, which is only known during surgery. Previous studies attempted to predict outcomes only from the preoperative 3D mesh model (Kadoury et al., 2017). While they showed promise, they nonetheless omitted crucial information about the intraoperative choices, such as the intraoperative posture and surgical parameters. Additional studies have demonstrated the important effect of the intraoperative positioning on the outcomes (Elysee et al., 2022). The issue with generating the spine from intraoperative images is the limited field of view, and therefore traditional methods would require long films to cover the entire spine, which increases surgery time and needs a stitching process between X-rays. Yet, excluding all intraoperative decision or parameters would limit the performance of the prediction.

Extensions. We have tested the A-NKF framework on an in-house dataset from a population of scoliotic patients which were followed longitudinally after surgery, offering an ideal environment to test the generative method for prediction applications, as well as for shape modeling purposes. However it remains limited to a single center which follows the same acquisition protocols, with regular time intervals between visits. We plan to explore other treatment options, such as brace wear which is notoriously more problematic to predict outcomes (Koutras et al., 2021). We consider evaluating the A-NKF method on other applications. Specifically, we consider that leveraging bone development in infants within the context of neural fields captured in a spatio-temporal domain can offer new insights on degenerative pathologies. We believe that the unsupervised nature of the A-NKF framework might provide benefits to mitigate limitation present in geometrical shape analysis.

6. Conclusion

This work presented an online surgical pipeline, predicting the first-upright spine shape based on the surgical approach and shape of the prone positioning on the OR table. The model learns with implicit neural fields and an articulation inference network, the anatomical variations in the vertebral constellation from the prone to the upright posture. During network training, geometric consistency is achieved by using a pre-trained locally-linear geodesic path, representing the correction evolution, leading to accurate outputs from the A-NKF. The framework produces results comparable to actual 3D spine models at the first follow-up exam taken in EOS biplanar systems (based on statistical significance tests), both in upright and side bending positions. The implicit A-NKF model describes the spine pose's physiological variations, potentially helping surgeons better plan the posture and surgical strategy in the OR and optimize outcomes. We plan in future work to evaluate the model within the context of a multi-centric prospective trial and assess the model's reliability and clinical integration for online surgical guidance.

CRediT authorship contribution statement

Sylvain Thibeault: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Marjolaine Roy-Beaudry:** Writing – review & editing, Resources, Project administration. **Stefan Parent:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Samuel Kadoury:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Academic Chair in Pediatric Spinal Deformities of CHU Ste-Justine and the Canada Research Chairs. We thank Compute Canada for providing the essential computational resources to complete this study.

Data availability

Data will be made available on request.

References

Bier, B., Unberath, M., Zaech, J.N., Fotouhi, J., Armand, M., Osgood, G., Navab, N., Maier, A., 2018. X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 55–63.

Boumal, N., Absil, P.A., 2011. A discrete regression method on manifolds and its application to data on SO (n). IFAC Proc. Vol. 44 (1), 2284–2289.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: Eur. Conf. on Computer Vision. pp. 213–229.

Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y., 2022. Transmorph: Transformer for unsupervised medical image registration. Med. Image Anal. 82, 102615.

Cheng, J.C., Castelein, R.M., Chu, W.C., Danielsson, A.J., Dobbs, M.B., Grivas, T.B., Gurnett, C.A., Luk, K.D., Moreau, A., Newton, P.O., et al., 2015. Adolescent idiopathic scoliosis. Nat. Rev. Dis. Primers 1 (1), 1–21.

Elysee, J.C., Lovecchio, F., Lafage, R., Ang, B., Huang, A., Bannwarth, M., Kim, H.J., Schwab, F., Lafage, V., 2022. Supine imaging is a superior predictor of long-term alignment following adult spinal deformity surgery. Glob. Spine J. 12 (4), 631–637.

Esfandiari, H., Anglin, C., Guy, P., Street, J., Weidert, S., Hodgson, A.J., 2019. A comparative analysis of intensity-based 2D–3D registration for intraoperative use in pedicle screw insertion surgeries. Int. J. Comput. Ass. Radiol. Surg. 14, 1725–1739.

Ge, L., Ren, Z., Li, Y., Xue, Z., Wang, Y., Cai, J., Yuan, J., 2019. 3D hand shape and pose estimation from a single rgb image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10833–10842.

Grimm, M., Esteban, J., Unberath, M., Navab, N., 2021. Pose-dependent weights and domain randomization for fully automatic X-ray to CT registration. IEEE Trans. Med. Imaging 40 (9), 2221–2232.

Grupp, R.B., Unberath, M., Gao, C., Hegeman, R.A., Murphy, R.J., Alexander, C.P., Otake, Y., McArthur, B.A., Armand, M., Taylor, R.H., 2020. Automatic annotation of hip anatomy in fluoroscopy for robust and efficient 2D/3D registration. Int. J. Comput. Assist. Radiol. Surg. 15, 759–769.

Helm, P.A., Teichman, R., Hartmann, S.L., Simon, D., 2015. Spinal navigation and imaging: history, trends, and future. IEEE Trans. Med. Imaging 34 (8), 1738–1746.

Hui, S.C., Pialasse, J.P., Wong, J.Y., Lam, T.-p., Ng, B.K., Cheng, J.C., Chu, W.C., 2016. Radiation dose of digital radiography (DR) versus micro-dose X-ray (EOS) on patients with adolescent idiopathic scoliosis: 2016 SOSORT-IRSSD "John Sevastic Award" Winner in Imaging Research. Scoliosis Spinal Disord. 11, 1–8.

Humbert, L., de Guise, J., Aubert, B., Godbout, B., Skalli, W., 2009. 3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences. Med. Eng. Phy. 31 (6), 681–687.

Illés, T., Somoskeöy, S., 2013. Comparison of scoliosis measurements based on three-dimensional vertebra vectors and conventional two-dimensional measurements: advantages in evaluation of prognosis and surgical results. Eur. Spine J. 22 (6), 1255–1263.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (2), 203–211.

Jobidon, H., Kadoury, S., Knez, D., Aubin, C.-É., 2019. Biomechanically driven intraoperative spine registration during navigated anterior vertebral body tethering. Phys. Med. Biol. 64 (11), 115008.

Kadoury, S., Labelle, H., Paragios, N., 2013. Spine segmentation in medical images using manifold embeddings and higher-order MRFs. IEEE Trans. Med. Imaging 32 (7), 1227–1238.

Kadoury, S., Labelle, H., Parent, S., 2016. Postoperative 3D spine reconstruction by navigating partitioning manifolds. Med. Phys. 43 (3), 1045–1056.

Kadoury, S., Mandel, W., Roy-Beaudry, M., Nault, M.L., Parent, S., 2017. 3-D morphology prediction of progressive spinal deformities from probabilistic modeling of discriminant manifolds. IEEE Trans. Med. Imaging 36 (5), 1194–1204.

Karikari, I.O., Lenke, L.G., Bridwell, K.H., Tauchi, R., Kelly, M.P., Sugrue, P.A., Bumpass, D.B., Elsamadicy, A.A., Adogwa, O., Lalezari, R., et al., 2018. Key role of preoperative recumbent films in the treatment of severe sagittal malalignment. Spine Deform. 6, 568–575.

Karras, T., Laine, S., Aila, T., 2019. A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4401–4410.

Keshavarzi, S., Ramchandran, S., Spardy, J., Dobyns, A., Errico, T., George, S., 2023. Utilization of pre-operative MRI to identify AIS patients at highest risk of intra-operative neuromonitoring alert. Spine Deform. 11 (6), 1419–1426.

Koller, H., Hitzl, W., Marks, M., Newton, P., 2019. Accurate prediction of spontaneous lumbar curve correction following posterior selective thoracic fusion in adolescent idiopathic scoliosis using logistic regression models and clinical rationale. Eur. Spine J. 28, 1987–1997.

Koutras, C., Pérez, J., Kardash, K., Otaduy, M.A., 2021. A study of the sensitivity of biomechanical models of the spine for scoliosis brace design. Comput. Methods Programs Biomed. 207, 106125.

Kreuzer, D., Beaini, D., Hamilton, W., Létourneau, V., Tossou, P., 2021. Rethinking graph transformers with spectral attention. Adv. Neural Inf. Process. Syst. 34, 21618–21629.

Lee, B.C., Sinha, A., Varble, N., Pritchard, W.F., Karanian, J.W., Wood, B.J., Bydlon, T., 2022. Breathing-compensated neural networks for real time C-arm pose estimation in lung CT-fluoroscopy registration. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, IEEE, pp. 1–5.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. In: ICLR 2019.

Ma, B., Han, Z., Liu, Y.S., Zwicker, M., 2020. Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. arXiv preprint arXiv:2011.13495.

Mandel, W., Oulbacha, R., Roy-Beaudry, M., Parent, S., Kadoury, S., 2020. Image-guided tethering spine surgery with outcome prediction using spatio-temporal dynamic networks. IEEE Trans. Med. Imaging 40 (2), 491–502.

Mandel, W., Turcot, O., Knez, D., Parent, S., Kadoury, S., 2019. Prediction outcomes for anterior vertebral body growth modulation surgery from discriminant spatiotemporal manifolds. IJCARS 14 (9), 1565–1575.

Markelj, P., Tomaževič, D., Likar, B., Pernuš, F., 2012. A review of 3D/2D registration methods for image-guided interventions. Med. Image Anal. 16 (3), 642–661.

Meng, C., Wang, Q., Guan, S., Sun, K., Liu, B., 2019. 2D-3D registration with weighted local mutual information in vascular interventions. IEEE Access 7, 162629–162638.

Miao, S., Piat, S., Fischer, P., Tuysuzoglu, A., Mewes, P., Mansi, T., Liao, R., 2018. Dilated FCN for multi-agent 2D/3D medical image registration. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, No. 1.

Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. Commun. ACM 65 (1), 99–106.

Mittal, M., Hoeller, D., Farshidian, F., Hutter, M., Garg, A., 2022. Articulated object interaction in unknown scenes with whole-body mobile manipulation. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, IEEE, pp. 1647–1654.

Mu, J., Qiu, W., Kortylewski, A., Yuille, A., Vasconcelos, N., Wang, X., 2021. A-sdf: Learning disentangled signed distance functions for articulated shape representation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13001–13011.

Nault, M.L., Mac-Thiong, J.M., Roy-Beaudry, M., Labelle, H., Parent, S., et al., 2013. Three-dimensional spine parameters can differentiate between progressive and nonprogressive patients with AIS at the initial visit: a retrospective analysis. J. Pediatr. Orthop. 33 (6), 618–623.

Oren, J.H., Tishelman, J.C., Day, L.M., Baker, J.F., Foster, N., Ramchandran, S., Jalai, C., Poorman, G., Cassilly, R., Buckland, A., et al., 2019. Measurement of spinopelvic angles on prone intraoperative long-cassette lateral radiographs predicts postoperative standing global alignment in adult spinal deformity surgery. Spine Deform. 7 (2), 325–330.

Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S., 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 165–174.

Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., Geiger, A., 2020. Convolutional occupancy networks. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. Springer, pp. 523–540.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. pp. 652–660.

Rebain, D., Matthews, M.J., Yi, K.M., Sharma, G., Lagun, D., Tagliasacchi, A., 2022. Attention beats concatenation for conditioning neural fields. arXiv preprint arXiv:2209.10684.

Sitzmann, V., Zollhöfer, M., Wetzstein, G., 2019. Scene representation networks: Continuous 3d-structure-aware neural scene representations. Adv. Neural Inf. Process. Syst. 32.

Tauchi, R., et al., 2016. Reliability analysis of cobb angle measurements of congenital scoliosis using X-ray and 3D-CT images. Eur. J. Orthop. Surg. Traumatol. 26 (1), 53–57.

Thibeault, S., Parent, S., Kadoury, S., 2023. Intra-operative forecasting of standing spine shape with articulated neural kernel fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 79–89.

Thong, W., Parent, S., Wu, J., Aubin, C.E., Labelle, H., Kadoury, S., 2016. Three-dimensional morphology study of surgical adolescent idiopathic scoliosis patient from encoded geometric models. Eur. Spine J. 1–10.

Tiwari, G., Antić, D., Lenssen, J.E., Sarafianos, N., Tung, T., Pons-Moll, G., 2022. Pose-ndf: Modeling human pose manifolds with neural distance fields. In: European Conference on Computer Vision. Springer, pp. 572–589.

Tjardes, T., Shafizadeh, S., Rixen, D., Paffrath, T., Bouillon, B., Steinhausen, E.S., 2010. Image-guided spine surgery: state of the art and future directions. Eur. Spine J. 19 (1), 25–45.

Unberath, M., Gao, C., Hu, Y., Judish, M., Taylor, R.H., Armand, M., Grupp, R., 2021. The impact of machine learning on 2D/3D registration for image-guided interventions: A systematic review and perspective. Front. Robotics AI 8, 716007.

Wang, D., Cui, X., Chen, X., Zou, Z., Shi, T., Salcudean, S., Wang, Z.J., Ward, R., 2021. Multi-view 3d reconstruction with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5722–5731.

Wawrose, R.A., LeVasseur, C.M., Byrapogu, V.K., Dombrowski, M.E., Donaldson, W.F., Shaw, J.D., Lee, J.Y., Anderst, W.J., Aiyangar, A.K., 2020. In vivo changes in adjacent segment kinematics after lumbar decompression and fusion. J. Biomech. 102, 109515.

Weng, Y., Wang, H., Zhou, Q., Qin, Y., Duan, Y., Fan, Q., Chen, B., Su, H., Guibas, L.J., 2021. Captra: Category-level pose tracking for rigid and articulated objects from point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13209–13218.

Williams, F., Gojcic, Z., Khamis, S., Zorin, D., Bruna, J., Fidler, S., Litany, O., 2022. Neural fields as learnable kernels for 3d reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18500–18510.

Williams, F., Trager, M., Bruna, J., Zorin, D., 2021. Neural splines: Fitting 3d surfaces with infinitely-wide neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9949–9958.

Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., Sridhar, S., 2022. Neural fields in visual computing and beyond. In: Computer Graphics Forum, Vol. 41, No. 2. Wiley Online Library, pp. 641–676.

Yuan, L., Zeng, Y., Chen, Z., Li, W., Zhang, X., Ni, J., 2020. Risk factors associated with failure to reach minimal clinically important difference after correction surgery in patients with degenerative lumbar scoliosis. Spine 45 (24), E1669–E1676.

Zhang, J., Zheng, Y., Qi, D., 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. pp. 1655–1661.

Zhao, L., Pang, S., Chen, Y., Zhu, X., Jiang, Z., Su, Z., Lu, H., Zhou, Y., Feng, Q., 2023. SpineRegNet: Spine registration network for volumetric MR and CT image by the joint estimation of an affine-elastic deformation field. Med. Image Anal. 86, 102786.

Zhu, J.Y., Zhang, Z., Zhang, C., Wu, J., Torralba, A., Tenenbaum, J., Freeman, B., 2018. Visual object networks: Image generation with disentangled 3D representations. Adv. Neural Inf. Process. Syst. 31.