



Intra-operative Forecasting of Standing Spine Shape with Articulated Neural Kernel Fields

Sylvain Thibeault¹, Stefan Parent², and Samuel Kadoury^{1,2}(✉)

¹ MedICAL, Polytechnique Montreal, Montreal, QC, Canada
samuel.kadoury@polymtl.ca

² Sainte-Justine Hospital Research Center, Montreal, QC, Canada

Abstract. Minimally invasive spine surgery such as anterior vertebral tethering (AVT), enables the treatment of spinal deformities while seeking to preserve lower back mobility. However the intra-operative positioning and posture of the spine affects surgical outcomes. Forecasting the standing shape from adolescent patients with growing spines remains challenging with many factors influencing corrective spine surgery, but can allow spine surgeons to better prepare and position the spine prior to surgery. We propose a novel intra-operative framework anticipating the standing posture of the spine immediately after surgery from patients with idiopathic scoliosis. The method is based on implicit neural representations, which uses a backbone network to train kernels based on neural splines and estimate network parameters from intra-pose data, by regressing the standing shape on-the-fly using a simple positive definite linear system. To accommodate with the variance in spine appearance, we use a Signed Distance Function for articulated structures (A-SDF) to capture the articulation vectors in a disentangled latent space, using distinct encoding vectors to represent both shape and articulation parameters. The network's loss function incorporates a term regularizing outputs from a pre-trained population growth trajectory to ensure transformations are smooth with respect to the variations seen on first-erect exams. The model was trained on 735 3D spine models and tested on a separate set of 81 patients using pre- and intra-operative models used as inputs. The neural kernel field framework forecasted standing shape outcomes with a mean average error of 1.6 ± 0.6 mm in vertebral points, and generated shapes with IoU scores of 94.0 compared to follow-up models.

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

1 Introduction

Surgical treatment remains to this day the main approach to correct severe spinal deformities, which consists of rectifying the alignment of the spine over several

Supported by the Canada Research Chairs and NSERC Discovery Grant RGPIN-2020-06558.

vertebral bodies with intrinsic forces. However major limiting factors remain the loss in spinal mobility, as well as the increased risk of pain and osteoarthritis in years after surgery [1]. Furthermore, prone positioning during spinal surgery combined with hip extension can greatly affect post-operative outcomes of sagittal alignment and balance, particularly in degenerative surgical cases [2]. Recent literature has demonstrated the importance to preserve the lumbar positioning, as well as the normal sagittal alignment once surgery has been completed [5, 7]. Therefore, by optimizing the intra-op posture, surgeons can help to reduce long-term complications such as lower back pain and improve spinal mobility. A predictive model anticipating the upright 3D sagittal alignment of patients during spine surgery [15], such as anterior vertebral tethering (AVT), may not only have an impact on patients to preserve their post-operative balance, but also provide information about patients at risk of degenerative problems in the lumbar region of the spine following surgery [24].

To interpret the positioning effects during surgery, identifying features leading to reliable predictions of shape representations in postural changes is a deciding factor as shown in [14], which estimated correlations with geometrical parameters. However, with the substantial changes during surgery such as the prone positioning of the spine and global orientation of the pelvic area, this becomes very challenging. Previous classification methods were proposed in the literature to estimate 3D articulated shape based on variational auto-encoders [19] or with spatio-temporal neural networks [10], combining soft tissue information of the pre-operative models. The objective here is to address these important obstacles by presenting a forecasting framework used in the surgical workflow, describing the spine shape changes within an implicit neural representation, with a clinical target of estimating Cobb angles within 5° [14].

Recently, Neural Radiance Fields (NeRF) [12] have been exploited for synthetic view generation [23]. The implicit representation provided by the parameters of the neural network captures properties of the objects such as radiance field and density, but can be extended to 3D in order to generate volumetric views and shapes from images [4], meshes [16] or clouds of points [9]. As opposed to traditional non-rigid and articulated registration methods based on deep learning for aligning pre-op to intra-op 2D X-rays [3, 26], kernel methods allow to map input data into a different space, where subsequently simpler models can be trained on the new feature space, instead of the original space. They can also integrate prior knowledge [22], but have been mostly focused on partial shapes [21]. However, due to the difficulty to apply neural fields in organ rendering applications in medical imaging, and the important variations in the environment and spine posture/shape which can impact the accuracy of the synthesis process, their adoption has been limited.

This paper presents an intra-operative predictive framework for spine surgery, allowing to forecast on-the-fly the standing geometry of the spine following corrective AVT procedures. The proposed forecasting framework uses as input the intra-operative 3D spine model generated from a multi-view Transformer network which integrates a pre-op model (Sect. 2.1). The predictive model is based

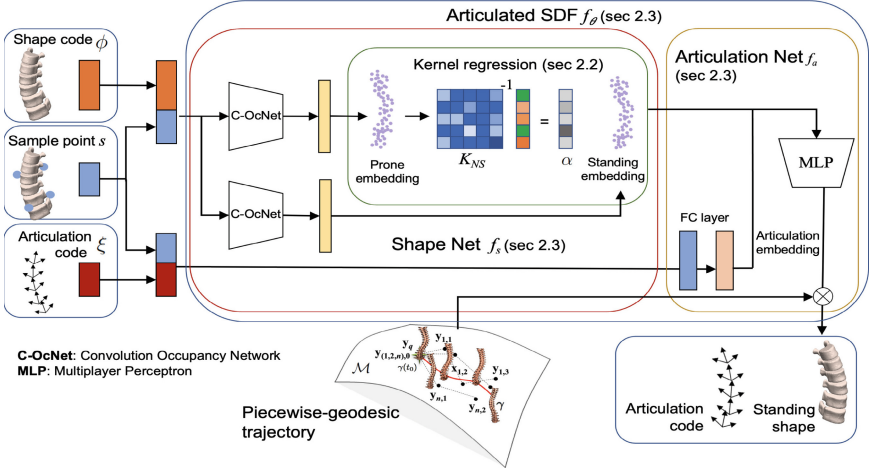


Fig. 1. The proposed forecasting architecture of the standing spine posture following surgery. During training, randomly selected spine shapes in the prone and first-erect standing posture are used to regress a neural spline kernel (K_{NS}). The generated shape in latent space (f_s) is then used as input to an articulated signed distance function (A-SDF) capturing the inter-vertebral shape constellation (f_a), transforming the signed distance values and infer the new instances from the trained kernel. Identical shape instances can yield multiple articulated instances, regardless of the shape code. At inference time, the shape and articulation code are inferred jointly by f_θ from back-propagation based on the input prone shape. A regularization term is added based on a pre-trained trajectory model, capturing the natural changes of spinal curvature.

on neural field representations, capturing the articulation variations in a disentangled latent space intra-operative positioning in the prone position. The method regresses a kernel function (Sect. 2.2) based on neural splines to estimate an implicit function warping the spine geometry between various poses. The inference process is regularized with a piecewise geodesic term, capturing the natural evolution of spine morphology (Sect. 2.3). As an output, the framework produces the geometry of the spine at the first-erect examination in a standing posture following surgery, as illustrated in Fig. 1.

2 Methods

2.1 Intra-operative 3D Model from Multi-view X-Rays

The first step of the pipeline consists of inferring a 3D model of the spine in the OR prior to instrumentation, using as input pair of orthogonal C-arm acquisitions $\mathcal{I} = \{\mathbf{I}^1, \mathbf{I}^2\}$ and previously generated pre-op 3D model [6], capturing the spine geometry in a lying prone position on the OR table. We use a multi-view 3D reconstruction approach based on Transformers [20]. The Transformer-based

framework integrates both a bi-planar Transformer-based encoder, which combines information and relationships between the multiple calibrated 2D views, and a 3D reconstruction Transformer decoder, as shown in Fig. 2.

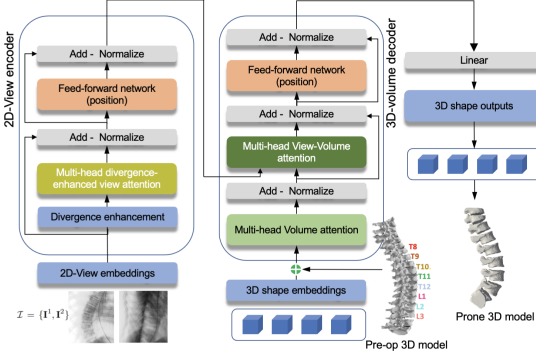


Fig. 2. Schematic illustration of the 3D shape inference of the spine from bi-planar X-rays in the prone position using multi-view Transformers. The model incorporates a view-divergence mechanisms to enhance features from the 2D view embeddings.

the 3D module to inject knowledge about the general geometry. By combining these modules (2D-2D, 2D-3D, 3D-3D) into a unified framework, feature correlations can be processed simultaneously using the attention layers in the encoder/decoder networks, generating as an output shape model $S = \{s_1, \dots, s_m\}$ with s_m as a vertebral mesh with a specific topology for the vertebral level m with an associated articulation vector:

$$\mathbf{y}_i = [T_1, ; T_1 \circ T_2, \dots, T_1 \circ T_2 \circ \dots \circ T_m] \quad (1)$$

representing the m inter-vertebral transformations T_i between consecutive vertebrae i and $i+1$, each consisting of 6 DoF with translation and rotation, expressed with recursive compositions. To accommodate with the Transformer’s inability to explore and synthesize multi-view associations at deeper levels, the divergence decay is slowed down within the self-attention layers by augmenting the discrepancy in multi-view embeddings.

2.2 Learnable Shape Deformation Kernel

We use a training set \mathcal{S} of N spine models from a surgical population, where each spine model is articulated into pre-operative, prone and first-erect poses ($K = 3$), leading to a training set of $L = N \times K$ spine models from a population

The 3D-reconstruction Transformer decodes and combines features from the biplanar views generated by the encoder, in order to produce a 3D articulated mesh model with probabilistic outputs for every spatial query token. The decoder’s attention layers captures the 2D-3D relationship between the resulting grid node of the mesh and input 2D X-rays. Attention layers in the 3D network on the other hand analyzes correlations between 3D mesh landmarks to learn a 3D representation. A conditional prior using the pre-op model is integrated at the input of

of scoliotic patients. We denote $S_{n,k}$ as the spine model n given a particular shape articulation k , where $n \in \{1, \dots, N\}$, and $k \in \{1, \dots, K\}$. We also define $s_i \in \mathbb{R}^3$ as a sample 3D point from a spine shape in \mathcal{S} .

We first train a kernel based on neural splines using the embeddings generated with a neural network [21], in order to map spine input points from the prone to the first-erect position. Input points $s_i \in \mathcal{S}$, associated to a shape code $\phi \in \mathbb{R}^D$ in latent space, with D indicating the dimensionality, are associated with a feature vector $\rho(s_i|\mathcal{S}, \Omega)$, where Ω conditions the parameters of the neural network ρ using the dataset \mathcal{S} . Based on the learned feature points of prone and upright spine models, the data-related kernel is given by:

$$K_{S,\Omega}(s_i, s_j) = K_{NS}([s_i : \rho(s_i|\mathcal{S}, \Omega)], [s_j : \rho(s_j|\mathcal{S}, \Omega)]) \quad (2)$$

with $[c : d]$ concatenating feature vectors c and d , representing the features of sample points i and j taken from a mesh model, respectively, and K_{NS} representing Neural Spline kernel function. A Convolutional Occupancy Network [17] is used to generate the features from sampled points in both spaces. Once the space around each input point is discretized into a volumetric grid space, PointNet [18] is used in every grid cell which contains points, to extract non-zero features from each 3D cell. These are then fed to a 3D FCN, yielding output features of same size.

The neural network’s implicit function is determined by finding coefficients α_j associated for each point s_j , such that:

$$\alpha = [\alpha_j]_{j=1}^{2L} = (\mathcal{G}(\mathcal{S}, \Omega) + \lambda \mathbf{I})^{-1} \mathbf{y}_j \quad (3)$$

which solves a $2L \times 2L$ linear system to obtain the standing articulations (\mathbf{y}_j), with $\mathcal{G}(\mathcal{S}, \Omega)$ as the gram matrix, such that $\mathcal{G}(\mathcal{S}, \Omega)_{ij} = K_{S,\Omega}(s_i, s_j)$, with s_i and s_j are the input sample points of the spine models, and regularized by the λ parameter using the identity matrix \mathbf{I} . For new feature points s , the regressed function is evaluated using α in the following term:

$$f_s(s, \phi) = \sum_{s_j \in \mathcal{S}, \phi} \alpha_j K_{S,\Omega}(s, s_j) \quad (4)$$

which represents the ShapeNet function and maps feature points s from the original prone embedding onto the standing embedding.

2.3 Articulation SDF

Each training shape $S_{n,k}$ is also associated with an articulation vector field, $\xi \in \mathbb{R}^d$. The shape embedding ϕ_n is common across all instances n with various articulation vectors. We then train ArticulationNet, where for each instance, the shape code is maintained and each individual code for shape is updated. The articulation vector \mathbf{y}_i defined in Eq.(1) is used to define ξ_i .

The shape function f_s trained from kernel regression is integrated into an articulation network [13] using the signed distance function (A-SDF) which is

determined by using $s \in \mathbb{R}^3$ as a shape’s sample point. We recall that ϕ and ξ are codes representing the shape and inter-vertebral articulation, respectively.

The A-SDF for the spine shape provided by f_θ is represented with an auto-decoder model, which includes the trained shape kernel function f_s , while f_a describes ArticulationNet:

$$f_\theta(s, \phi, \xi) = f_a[f_s(s, \phi), s, \xi] = v \quad (5)$$

with $v \in \mathbb{R}$ representing a value of the SDF on the output mesh, where the sign specifies if the sampled spine point falls within or outside the generated 3D surface model, described implicitly with the zero level-set $f_\theta(\cdot) = 0$.

At training, the ground-truth articulation vectors ξ are used to train the parameters of the model and the shape code. Sample points s are combined with the shape code ϕ to create a feature vector which is used as input to the shape embedding network. A similar process is used for the articulation code ξ , which are also concatenated with sample points s . This vector is used as input to the articulation network in order to predict for each 3D point s the value of the SDF. Hence, the network uses a fully connected layer at the beginning to map the vector in the latent space, while a classification module is added in the final hidden layer to output the series of vertebrae. We define the loss function using the L1 distance term, regressing the SDF values for the M number of 3D points describing each spine shape using the function f_θ :

$$\mathcal{L}^s = \frac{1}{M} \sum_{m=1}^M \|f_\theta(s_m, \phi, \xi) - v_m\|_1 \quad (6)$$

where $s_m \in \mathcal{S}$ is a sample point from the shape space, and v_m is the SDF value used as ground-truth for $m \in \{1, \dots, M\}$. The second loss term evaluates the vertebral constellation and alignment with the inter-vertebral transformations:

$$\mathcal{L}^p = \frac{1}{M} \sum_{m=1}^M [CE(f_\theta(s_m, \phi, \xi), p_m)] \quad (7)$$

with p_m identifying the vertebral level for sample s_m and CE represents the cross-entropy loss. This classification loss enables to separate the constellation of vertebral shapes \mathbf{s}_m .

To ensure the predicted shape is anatomically consistent, a regularization term \mathcal{L}^r is used to ensure generated models fall near a continuous piecewise-geodesic trajectory defined in a spatio-temporal domain. The regularization term integrates pre-generated embeddings capturing the spatiotemporal changes of the spine geometry following surgery [11]. The piecewise manifold is obtained from nearest points with analogue features in the training data of surgical spine patients in the low-dimensional embedding, producing local regions $\mathcal{N}(s_i)$ with samples lying within the shape change spectrum. A Riemannian domain with samples \mathbf{y}_i is generated, with i a patient model acquired at regular time intervals. Assuming the latent space describes the overall variations in a surgical

population and the geodesic trajectory covers the temporal aspects, new labeled points can be regressed within the domain \mathbb{R}^D , producing continuous curves in the shape/time domain. The overall loss function is estimated as:

$$\mathcal{L}(\mathcal{S}, \phi, \xi) = \mathcal{L}^s(\mathcal{S}, \phi, \xi) + \beta_p \mathcal{L}^p(\mathcal{S}, \phi, \xi) + \beta_\theta \mathcal{L}^r \quad (8)$$

with β_p and β_θ weighting the parts and regularization terms, respectively. During training, a random initialization of the shape codes based on Gaussian distributions is used. An optimization procedure is then applied during training for each shape code, which are used for all instances of articulations. Hence the global objective seeks to minimize for all $N \times K$ shapes, the following energy:

$$\arg \min_{\phi, \xi} \sum_{n=1}^N \sum_{k=1}^K \mathcal{L}(S_{n,k}, \phi_n, \xi_k). \quad (9)$$

At test time, the intra-operative spine model (obtained in 2.1) is given as input, generating the articulated standing pose, with shape and articulation vectors, using back-propagation. The shape and articulation codes ϕ and ξ are randomly initialized while the network parameters stay fixed. To overcome divergence problems and avoid a local minima for Eq. (9), a sequential process is applied where first articulated codes are estimated while shape outputs are ignored, allowing to capture the global appearance. Then, using the fixed articulated representation, the shape code is re-initialized and optimized only for ϕ .

3 Experiments

A mono-centric dataset of 735 spine models was used to train the articulated neural kernel field forecasting model. The dataset included pre-, intra- and post-operative thoraco/lumbar 3D models. Pre- and post-op models were generated from bi-planar X-rays using a stereo-reconstruction method (EOS system, Paris, France), integrating a semi-supervised approach generating vertebral landmarks [6]. Each patient in the cohort underwent corrective spine surgery, with a main angulation range of $[35^\circ - 65^\circ]$. Vertebral landmarks were validated by a surgeon.

During spine surgery, a pair of C-arm images at frontal and 90° angulations were acquired after prone positioning of the patient on the OR table and before starting the surgical instrumentation. The C-arm image pair was used to generate the input 3D model of the instrumented spine segment. For the Transformer model, batch size was set at 64 and image sizes for training were set at 225×225 . For the neural field model, batch size was set at 32, with a dropout rate of 0.4, a learning rate of 0.003, $\beta_p = 0.4$, $\beta_\theta = 0.5$, $\lambda = 0.25$. The AdamW optimizer was used for both models [8]. Inference time was of 1.2 s on a NVIDIA A100 GPU. We assessed the intra-operative modeling of the spine in a prone position using the Transformer based framework on a separate set of 20 operative patients. Ground-truth models with manually annotated landmarks on pairs of C-arm images were used as basis of comparison, yielding a 3D RMS error of 0.9 ± 0.4 mm, a Dice score (based on overlap of GT and generated vertebral meshes) of 0.94 ± 0.3 and a difference of $0.9^\circ \pm 0.3$ in the main spine angulation. These are in the clinical acceptable ranges to work in the field.

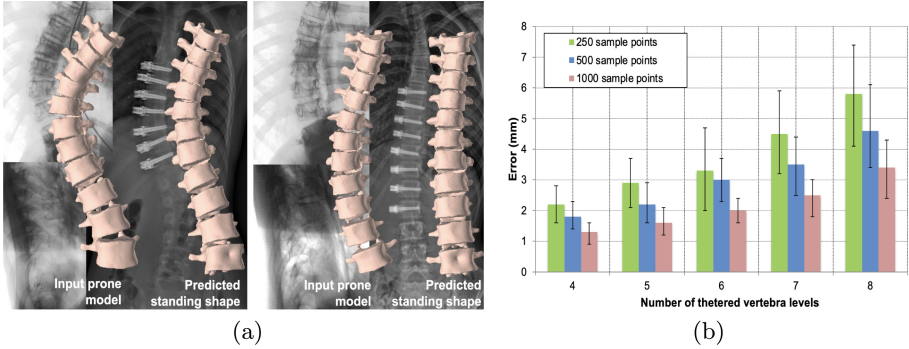


Fig. 3. (a) Sample predicted 3D geometric models of the thoraco/lumbar spine in standing posture following corrective procedures. Both examples show the input intra-operative model on the left, and the predicted geometric shape on the right, alongside the first-erect X-ray images at follow-up. (b) Effect of increasing number of tethered segments and number sampled points per model on the overall accuracy.

Finally, the predicted standing spine shape accuracy was evaluated on a hold-out set of 81 surgical patients who underwent minimally invasive corrective spine surgery. The cohort had an initial mean angulation of 48° , and immediate follow-up exam at 2 weeks. Results are presented in Table 1. For each predicted standing spine model, errors in 3D vertebral landmarks, IoU measures and Chamfer distance, as well as differences in Cobb and lordosis angles were computed for the inferred shapes, and were compared to state-of-the-art spatio-temporal and neural fields models. Ablation experiments were also conducted, demonstrating

Table 1. 3D RMS errors (mm), IoU (%) Chamfer distance, Cobb angle difference ($^\circ$) and lordosis angle difference ($^\circ$) for the proposed articulated neural kernel field (A-NKF) method, compared to ST-ResNet [25], Convolutional Occupancy Network [17], DenseSDF [16], NeuralPull [9] and ST-Manifold [10]. The ground-truth standing 3D spine models obtained at follow-up from bi-planar X-rays were used as basis for comparison. Bottom rows shows the ablation experiments.

	3D RMS ↓	IoU ↑	Chamfer ↓	Cobb ↓	Lordosis ↓
ST-ResNet [25]	6.3 ± 3.8	78.5 ± 3.6	8.3 ± 4.2	5.6 ± 3.9	5.9 ± 4.2
C-OccNet [17]	4.6 ± 2.8	81.2 ± 3.4	7.0 ± 3.6	4.3 ± 2.9	4.8 ± 4.4
DeepSDF [16]	3.8 ± 2.1	84.5 ± 2.8	6.1 ± 2.2	3.9 ± 2.4	4.2 ± 3.9
NeuralPull [9]	3.1 ± 1.4	86.9 ± 1.9	5.3 ± 1.8	3.7 ± 2.0	3.9 ± 3.6
ST-Manifold [10]	2.8 ± 1.2	88.2 ± 1.5	4.9 ± 1.6	3.1 ± 1.8	3.5 ± 3.0
A-NKF (no kernel + no regular.)	2.7 ± 1.0	87.4 ± 1.3	5.0 ± 1.8	3.3 ± 2.0	3.5 ± 2.9
A-NKF (no kernel)	2.1 ± 0.7	91.7 ± 2.5	4.3 ± 0.6	2.6 ± 1.8	2.9 ± 2.4
A-NKF (no regular.)	1.9 ± 0.7	92.2 ± 2.3	4.1 ± 0.7	2.5 ± 1.7	2.7 ± 2.0
Proposed A-NKF	1.6 ± 0.6	94.0 ± 2.0	3.7 ± 0.5	2.1 ± 1.2	2.3 ± 1.5

statistically significant improvements ($p < 0.05$) with the kernel and regularization modules to the overall accuracy. We finally evaluated the performance of the model by measuring the 3D RMS errors versus the input size of points sampled from each shape and tethered vertebral levels. We can observe limitations when using a sparser set of samples points at 250 points with a mean error of 2.8 mm, but a significant increase in performance of the kernel-based neural fields using denser inputs for each model ($n=1000$).

4 Conclusion

In this paper, we proposed an online forecasting model predicting the first-erect spine shape based on intra-operative positioning in the OR, capturing the articulated shape constellation changes between the prone and the standing posture with neural kernel fields and an articulation network. Geometric consistency is integrated with the network’s training with a pre-trained spine correction geodesic trajectory model used to regularize outputs of ArticulationNet. The model yielded results comparable to ground-truth first-erect 3D geometries in upright positions, based on statistical tests. The neural field network implicitly captures the physiological changes in pose, which can be helpful for planning the optimal posture during spine surgery. Future work will entail evaluating the model in a multi-center study to evaluate the predictive robustness and integrate the tool for real-time applications.

References

1. Cheng, J.C., et al.: Adolescent idiopathic scoliosis. *Nat. Rev. Dis. Primers* **1**(1), 1–21 (2015)
2. Elysee, J.C., et al.: Supine imaging is a superior predictor of long-term alignment following adult spinal deformity surgery. *Glob. Spine J.* **12**(4), 631–637 (2022)
3. Esfandiari, H., Anglin, C., Guy, P., Street, J., Weidert, S., Hodgson, A.J.: A comparative analysis of intensity-based 2D–3D registration for intraoperative use in pedicle screw insertion surgeries. *Int. J. Comput. Assist. Radiol. Surg.* **14**, 1725–1739 (2019)
4. Ge, L., et al.: 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 (2019)
5. Harimaya, K., Lenke, L.G., Mishihiro, T., Bridwell, K.H., Koester, L.A., Sides, B.A.: Increasing lumbar lordosis of adult spinal deformity patients via intraoperative prone positioning. *Spine* **34**(22), 2406–2412 (2009)
6. Humbert, L., de Guise, J., Aubert, B., Godbout, B., Skalli, W.: 3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences. *Med. Eng. Phys.* **31**(6), 681–87 (2009)
7. Karikari, I.O., et al.: Key role of preoperative recumbent films in the treatment of severe sagittal malalignment. *Spine Deform.* **6**, 568–575 (2018)
8. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *ICLR*, 2019 (2017)

9. Ma, B., Han, Z., Liu, Y.S., Zwicker, M.: Neural-pull: learning signed distance functions from point clouds by learning to pull space onto surfaces. arXiv preprint [arXiv:2011.13495](https://arxiv.org/abs/2011.13495) (2020)
10. Mandel, W., Oulbacha, R., Roy-Beaudry, M., Parent, S., Kadoury, S.: Image-guided tethering spine surgery with outcome prediction using Spatio-temporal dynamic networks. *IEEE Trans. Med. Imaging* **40**(2), 491–502 (2020)
11. Mandel, W., Turcot, O., Knez, D., Parent, S., Kadoury, S.: Prediction outcomes for anterior vertebral body growth modulation surgery from discriminant spatiotemporal manifolds. *IJCARS* **14**(9), 1565–1575 (2019)
12. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **65**(1), 99–106 (2021)
13. Mu, J., Qiu, W., Kortylewski, A., Yuille, A., Vasconcelos, N., Wang, X.: 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 (2021)
14. Nault, M.L., Mac-Thiong, J.M., Roy-Beaudry, M., Labelle, H., Parent, S., et al.: Three-dimensional spine parameters can differentiate between progressive and non-progressive patients with AIS at the initial visit: a retrospective analysis. *J. Pediatr. Orthop.* **33**(6), 618–623 (2013)
15. Oren, J.H., et al.: 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 (2019)
16. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: 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 (2019)
17. Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., Geiger, A.: Convolutional occupancy networks. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) *ECCV 2020. LNCS*, vol. 12348, pp. 523–540. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58580-8_31
18. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652–660 (2017)
19. Thong, W., Parent, S., Wu, J., Aubin, C.E., Labelle, H., Kadoury, S.: Three-dimensional morphology study of surgical adolescent idiopathic scoliosis patient from encoded geometric models. *Eur. Spine J.* **25**, 3104–3113 (2016)
20. Wang, D., et al.: Multi-view 3D reconstruction with transformers. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5722–5731 (2021)
21. Williams, F., et al.: Neural fields as learnable kernels for 3D reconstruction. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18500–18510 (2022)
22. Williams, F., Trager, M., Bruna, J., Zorin, D.: 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 (2021)
23. Xie, Y., et al.: Neural fields in visual computing and beyond. In: *Computer Graphics Forum*, vol. 41, pp. 641–676. Wiley Online Library (2022)
24. Yuan, L., Zeng, Y., Chen, Z., Li, W., Zhang, X., Ni, J.: 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 (2020)

25. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp. 1655–1661 (2017)
26. Zhao, L., et al.: 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 (2023)