NIMOSEF: Neural implicit motion and segmentation functions

Jaume Banus[1], Antoine Delaloye[2], Pedro M. Gordaliza[3,1], Costa Georgantas[1], Ruud B. van Heeswijk[1], and Jonas Richiardi[1,3]

[1] Department of Radiology, Lausanne University Hospital and University of Lausanne, Switzerland
[2] Ecole polytechnique fédérale de Lausanne (EPFL), Switzerland
[3] CIBM Center for Biomedical Imaging, Switzerland
jaume.banus-cobo@chuv.ch

Abstract. We present NIMOSEF, a novel unified framework that leverages neural implicit functions for joint segmentation, reconstruction, and displacement field estimation in cardiac magnetic resonance imaging (CMRI). By leveraging on a shared implicit representation for joint segmentation and motion estimation our approach improves spatio-temporal consistency with respect to conventional grid-based convolutional neural networks and implicit segmentation functions. NIMOSEF employs an auto-decoder architecture to learn subject-specific latent representations from unstructured point clouds derived from image intensities and reference segmentations. These latent codes, when combined with 4D space–time coordinates, enable the generation of high-resolution segmentation outputs and smooth, temporally coherent motion estimates. Experimental evaluation on a subset of 700 random patients from the UK Biobank demonstrates that our method achieves competitive segmentation accuracy—attaining Dice scores of up to 0.93 for the LV, 0.90 for the RV and 0.83 for the LV myocardium, with improved spatio-temporal consistency, predicting a smaller number of disconnected components. Simultaneously, it achieves an average registration error of the whole heart boundary of 3.08 ± 1.23mm measured by the Chamfer distance, and 8.57 ± 4.74mm according to the 95th percentile Hausdorff distance. Additionally, feature importance analysis reveals that the learnt implicit representation encodes physiologically relevant information. These results suggest that NIMOSEF offers a promising alternative for high-resolution, temporally consistent cardiac segmentation and motion estimation, with promising potential for advancing clinical assessment of cardiac function.

Keywords: Implicit functions · Segmentation · CMRI · Registration

1 Introduction

Image segmentation is key in the medical field for studying the morphology of objects of interest. In general, segmentation involves assigning a class label to every point in the input space. In cardiac magnetic resonance imaging (CMRI), accurate segmentation and motion estimation are essential for the clinical assessment of cardiac function, and is increasingly part of clinical routine.

Typically, since images are presented in a grid-like structure, Convolutional Neural Networks (CNNs) have become the de facto standard for medical image segmentation [3,1]. However, their reliance on grid-based (2D or 3D) representations limits the resolution of outputs and often requires additional post-processing to achieve high-quality results [15]. Furthermore, CNN training is computationally expensive [20], scaling exponentially with image resolution, and these models often generalize poorly to critical input transformations, such as scaling [27,10,18]. To address these challenges, several approaches have incorporated dataset-specific characteristics through augmentations and multi-scale feature extractors [11], achieving impressive scores on various challenges [1]. Other works have proposed lighter architectures [16] and mechanisms to capture long-range dependencies based on transformers [24] and sequence-state models [34]. Nevertheless, these methods require complex architectures and are computationally expensive with respect to memory. Recent studies explored the use of implicit neural representations (INRs) for segmentation [28,22], demonstrating that high-resolution outputs can be generated from sparse measurements. In these works, a neural network learns a mapping from a coordinate space to an arbitrary real-valued space (e.g., a segmentation label, a distance function, or the image intensity) [17]. However, due to the unstructured nature of the approach in which each voxel is treated independently, the obtained segmentation can result in non-connected components and temporal inconsistencies, and during inference, the selection of a proper validation criterion (e.g., determining when to stop training) is challenging. In parallel, implicit functions have also been used for modeling displacement fields in image registration [19,9]. However, to our knowledge, no joint model has tried to obtain an implicit representation valid for the joint estimation of motion and segmentation labels.

In this work, we introduce NIMOSEF, a novel unified framework that simultaneously estimates segmentation labels, reconstructs image intensity, and predicts displacement fields within an INR paradigm to promote spatio-temporal continuity. The core of the framework is an auto-decoder process that approximates a subject's latent representation (h) from pairs of coordinate-image intensity values (c, i). Given the latent representation, the image intensity, segmentation labels, and displacement field with respect to a reference frame, t_0, can be sampled at any arbitrary spatio-temporal coordinates. This joint formulation serves as an effective regularizer and improves the overall robustness of cardiac segmentation and motion estimation. We evaluate the proposed framework on a random subset of 700 subjects from the UK Biobank [29] using short-axis CMRI images. We report segmentation scores, registration metrics and investigate the latent space of the learnt implicit representation. The source code is available at https://github.com/jbanusco/nimosef-v1.

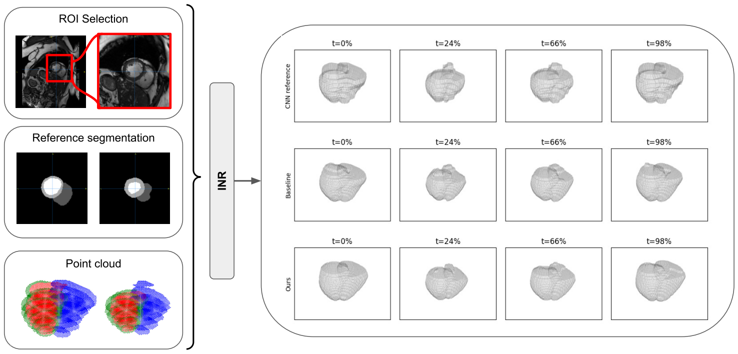


Fig. 1: Workflow schematic. We identify a region of interest (ROI) centered around the heart. Next, we segment the images using a pre-trained CNN [2]. The intensity and segmentation are sampled as unstructured point clouds and provided to the implicit neural network. Right panel shows the mesh obtained from the CNN segmentation and the outputs of the baseline and NIMOSEF.

2 Related work

2.1 Implicit neural representations

INRs are a paradigm shift from explicit representations by learning a continuous mapping from input coordinates to desired outputs. Early work in this area, such as occupancy networks [21,4] and signed distance functions [23], demonstrated that implicit representations can accurately capture complex 3D geometries. Due to their continuous nature, INRs can naturally interpolate sparse observations and enable evaluation at arbitrary resolutions without the need for post-processing, producing smoother representations. INRs have rapidly improved through novel activation functions and adjusting encoding strategies [8].

2.2 INR for Image Segmentation

Segmentation INRs usually adopt an auto-decoder framework that maps continuous coordinate spaces to segmentation probabilities and image intensities. This continuous formulation allows for sampling at any resolution during both training and inference, making the approach inherently robust to the anisotropic nature of medical imaging data. Recent studies have shown that INRs not only capture detailed anatomical shapes from sparse measurements but can also outperform conventional grid-based methods like U-Net [28,22] obtaining high-resolution versions of the objects from a sparse set of measurements.

2.3 INR for Image Registration

INR have also been applied in image registration [33]. For example, INRs were used to register the left ventricular myocardium in CT images by encoding signed distance fields and Hounsfield unit information [19]. In another approach, a diffeomorphic framework was introduced to estimate a continuous velocity field that is integrated via an ordinary differential equation (ODE), resulting in smooth, temporally consistent deformations with periodic constraints [9]. These advances highlight the potential of INRs to model complex, continuous deformations.

3 Methods

3.1 Shared Latent Space and Implicit Representation

Our network parametrizes a mapping between the image domain and a shared latent space $\mathcal{H}$ across all subjects. Each subject j is associated with a latent code $h_j \in \mathbb{R}^d$ that encodes subject-specific anatomical features and represents its location in the learned cardiac shape space. In our framework, h_j is concatenated with a 4D space-time coordinate $c \in \mathbb{R}^4$ to condition all downstream tasks: segmentation, intensity reconstruction, and displacement field estimation.

3.2 Model Architecture

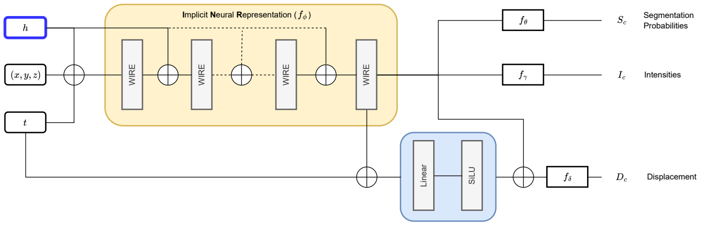


Fig. 2: Architecture overview. The orange box depicts the implicit neural representation, consisting of multiple residual layers with WIRE activations that take as input the subject-specific latent code (h) concatenated with the space-time coordinate (x, y, z, t). From this representation, the network predicts segmentation probabilities and image intensities. The blue box illustrates the displacement module, which applies a non-linear transformation to produce a “motion code", then combines it with the latent representation to yield the final displacement.

As shown in Figure 2, our neural network is implemented as an encoder-decoder architecture using a multilayer perceptron (MLP) with 8 residual layers, each comprising 128 units. We employ Wavelet Implicit Neural Representation (WIRE) activation functions [26] for space-frequency localization. The encoder receives as input the concatenation of the subject-specific latent code h_j and the space-time coordinate $c = (x, y, z, t)$. The encoded representation is used downstream for three prediction heads. The segmentation head, (f_θ), predicts a segmentation probability vector $s_c \in [0,1]^M$ for M classes, ensuring that $\sum_{i=1}^{M} s_c^i = 1$. The reconstruction head, (f_γ), predicts the image intensity $i_c \in [0,1]$. The displacement head (f_δ) predicts a displacement vector $d_c \in \mathbb{R}^3$. In the displacement branch, an additional linear layer followed by a sigmoid-weighted linear unit (SiLU) [7,12] is used to generate a "motion code", which is then concatenated with the encoder's representation to predict the displacement.

3.3 Training

At training, the latent codes (h_j) are randomly initialized and optimized jointly with the network parameters. The overall loss function is a weighted sum of several components. The segmentation loss is a combination of cross-entropy loss $\mathcal{L}_{CE}$ and Dice loss $\mathcal{L}_{Dice}$ between the predicted segmentation $f_\theta(c, h_j)$ and the ground truth s_c. The intensity reconstruction loss is defined as an L_1 loss, $\mathcal{L}_{L1}$, between the predicted intensity $f_\gamma(c, h_j)$ and the true intensity i_c. The displacement loss, $\mathcal{L}_{dsp}$, is based on the chamfer distance between the predicted boundary (after applying the predicted displacement) and the ground-truth segmentation boundary at time t. Since the reference frame is t_0, an $L2$ loss, $\mathcal{L}_{bd}$, penalises the predicted displacement at t_0, forcing it to vanish. In addition, a Jacobian regularization loss, $\mathcal{L}_J$, penalizes unrealistic deformations by discouraging negative determinants in the displacement field's Jacobian. Finally, an L_2 regularization term, $\mathcal{L}_{L_2}(h_j)$, is applied to the latent codes to prevent overfitting and promote a compact representation. The overall loss function is formulated as follows:

$$\begin{aligned}\mathcal{L}(\theta, \gamma, \delta, \phi, h_j) = {} & \lambda_{\text{seg}} \Big(\mathcal{L}_{CE}\big(f_\theta(c, h_j), s_c\big) + \mathcal{L}_{Dice}\big(f_\theta(c, h_j), s_c\big) \Big) \\ & + \lambda_{\text{rec}}\, \mathcal{L}_{L1}\big(f_\gamma(c, h_j), i_c\big) + \lambda_{\text{dsp}}\, \mathcal{L}_{dsp}\big(f_\delta(c, h_j), b_{t_0}, b_t\big) \\ & + \lambda_{reg} \mathcal{L}_{bd}\big(f_\delta(c, h_j), b_{t_0}, b_{t_0}\big) \\ & + \lambda_J\, \mathcal{L}_J(f_\delta(c, h_j), b_t) + \lambda_{\text{lat}}\, \mathcal{L}_{L_2}(h_j), \end{aligned} \tag{1}$$

where b_{t_0} and b_t denote the segmentation boundaries at the reference time, t_0, and target time, t, respectively. λ_{seg}, λ_{rec}, λ_{dsp}, λ_{reg}, λ_J, and λ_{lat} are the corresponding weighting coefficients. Displacement is computed from a fixed reference time t_0 (representing end-diastole, heart is fully relaxed) to any other time t in the cardiac cycle. The segmentation boundary is extracted at times t_0 and t. To identify boundary points, we compute the 5-nearest neighbors for each segmentation point and flag a point as a boundary if any of its neighbors has a value of 0. Subsequently, a 5-nearest neighbor graph is constructed, using a maximum distance threshold set to 1.5 times the average distance between boundary points. K-NN computations are performed using PyTorch3D [25], and the graph is built with Deep Graph Library (DGL) [32]. The graph is used to approximate the Jacobian of the displacement field along the edges as $J \approx \frac{\Delta u}{\Delta x}$, where Δu and Δx denote the changes in displacement and position, respectively.

4 Experiments

4.1 Data

We randomly sampled 700 subjects from the UK Biobank [29] cohort having short-axis CMRI. The images cover the base to the apex, with an in-slice resolution of $\sim 1.5\text{mm}^2$ and a slice thickness ranging from 5 to 10mm. The reference segmentations were obtained using a pre-trained CNN as described in [2]. A region of interest (ROI) covering the heart was defined based on the circular Hough transform [13,6]. After identifying an approximate LV center we extracted a 128×128 patch centered around it. The images were normalized as $I_{norm} = (I - I_2)/(I_2 - I_{98})$, where I corresponds to the image intensity and I_2 and I_{98} to the 2nd and 98th percentile respectively.

4.2 Experiment details

Each subject's latent vector was initialized as $h_j \sim \mathcal{N}(0, 10^{-2})$, promoting a compact representation that enables smooth interpolation. Both the network parameters and the subject-specific latent vectors (h_j) are optimized jointly. For each training batch, a reference time t_0 and an additional time frame t are sampled, and 90% of the corresponding 3D volume points are processed. The loss is computed from reconstruction, segmentation, and boundary mismatch errors after applying the predicted displacement. We trained each model for 500 epochs over 4 days on an NVIDIA RTX6000 using the Adam optimizer [14] with differential learning rates, 10^{-3} for the latent code, $5 \cdot 10^{-4}$ for the remaining parameters, and a weight decay of 10^{-4}. Loss weights were selected on a random subset of subjects and set to $\lambda_{\text{rec}} = 2$, $\lambda_{\text{seg}} = 1$, $\lambda_{\text{dsp}} = 0.5$, $\lambda_{\text{J}} = 1$, $\lambda_{reg} = 0.1$ and $\lambda_{\text{lat}} = 0.01$ for our model, while for the baseline (no motion estimation) λ_{dsp}, λ_{reg} and λ_{J} were set to zero, making it comparable to the NISF implementation [28]. To explore the latent space, we computed the mean shape code for each model and measured the $L2$ distance of each subject to this mean. A Random Forest regressor with 5-fold cross-validation was then used to assess the relationship between the distance and UK Biobank variables related to cardiovascular health, including cardiac function, clinical chemistry, demographic data and ECG recordings among others.

4.3 Results

Table 1 summarizes the segmentation performance. NIMOSEF predicts segmentations comparable to the CNN reference and the baseline while producing fewer disconnected components. Registration performance is reported in Table 2, NIMOSEF attains a Chamfer distance of 3.08 ± 1.23mm and a 95th percentile

Table 1: Segmentation performance: Dice score and average maximum number of connected components (CC) for each label.

	Dice Score (↑ better)			**Max CC** (↓ better)		
Method	LV	MYO	RV	LV	MYO	RV
Baseline	0.95 ± 0.01	0.87 ± 0.03	0.92 ± 0.02	1.08 ± 0.30	1.57 ± 0.75	1.21 ± 0.44
Ours	0.93 ± 0.02	0.83 ± 0.05	0.90 ± 0.03	1.02 ± 0.14	1.44 ± 0.69	1.16 ± 0.38
CNN Ref.	–	–	–	1.04 ± 0.20	1.35 ± 0.49	1.29 ± 0.47

Table 2: Chamfer distance and Hausdorff (95th percentile) distance between predicted boundary and target boundary of the CNN segmentation.

Method	**Chamfer Dist. [mm]** (↓)	**Hausdorff [mm] (95%)** (↓)
Ours	3.08 ± 1.23	8.57 ± 4.74

Hausdorff distance of 8.57 ± 4.74mm, comparable to CNN-based registration approaches [31]. Volume index analysis (Figure 4a) highlights that the temporal evolution of cardiac volumes estimated by NIMOSEF matches the reference segmentation, with smooth and temporally consistent profiles. The volume derivative curves (Figure 4b) of NIMOSEF are smoother compared to the other methods. Feature importance analysis (Figure 3), shows that the implicit representation learnt by NIMOSEF is better explained by a set of clinical, demographic, and laboratory features compared to the baseline (R^2 of 0.22 vs 0.18).

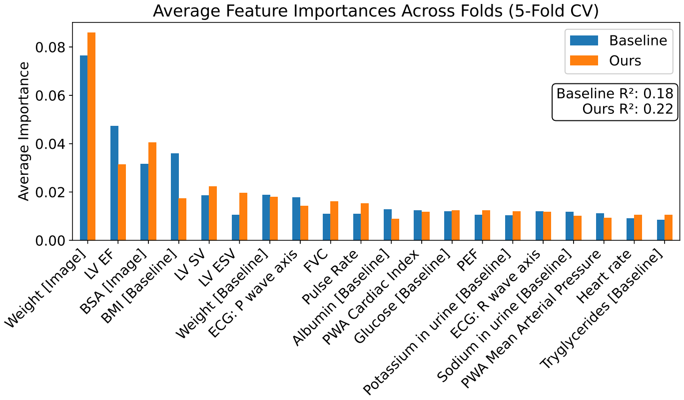
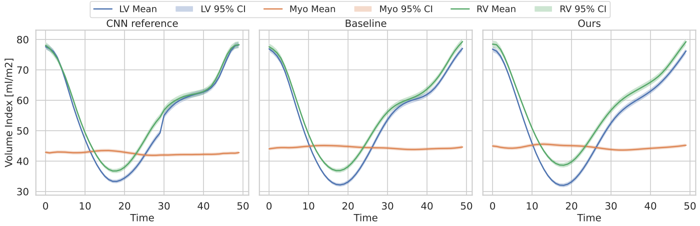
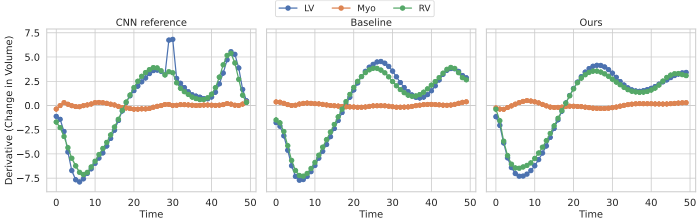


Fig. 3: Top-20 features based on average feature importance across 5-fold cross-validation. Demographic, clinical, and laboratory measurements were used to predict the distance to the mean shape code. Baseline: first visit date; Image: imaging visit date. BSA: body surface area; BMI: body mass index; FVC: forced vital capacity; PEF: peak expiratory flow; PWA: pulse wave analysis; ECG: electrocardiography; SV: stroke volume; ESV: end-systolic volume.



(a) Volume index.



(b) Volume derivatives.

Fig. 4: a) Average volume index curves for each segmentation approach. Shaded regions denote 95% confidence intervals. b) Volume derivative curves. LV: left ventricle; RV: right ventricle; Myo: Left ventricle myocardium.

5 Discussion

Experiments on the UK Biobank dataset demonstrate that NIMOSEF reconstructs image intensity, predicts segmentation labels with improved spatial coherence (evidenced by a reduced average maximum number of connected regions), and achieves motion estimation comparable to conventional approaches. Feature importance analysis indicates that our model captures clinically relevant information. Nonetheless, the auto-decoder framework is slow at inference due to latent code optimization. Recent studies suggest that conditioning on demographic variations [30] or clinical sub-populations [5] can enable a more informed latent initialization. Future work will explore whether conditioning improves performance without biasing predictions or limiting the interpretability of latent codes. Conditioning may boost accuracy but could also constrain the structure of learned representations. We also highlight that the reported segmentation scores are based on reference annotations derived from a pre-trained CNN [2]. These annotations contain spatial inconsistencies—particularly at challenging regions such as the apex or base. This limitation makes Dice scores less informative in those areas. In this context, NIMOSEF's ability to produce smoother and more connected structures could serve as a form of implicit label refinement. Moreover, since ground-truth boundaries are unavailable at inference, we plan to explore a warm-up period—guided by intensity reconstruction—followed by graph-based displacement metrics computed using the estimated boundaries from predicted segmentations. Finally, incorporating diffeomorphic or periodic constraints could further enhance motion consistency [9].

6 Conclusion

In this paper, we introduced NIMOSEF, a novel unified framework that learns an INR for joint segmentation, reconstruction, and displacement field estimation in CMRI. We demonstrated that a shared INR can be effectively used to obtain high-resolution segmentation and accurate displacement estimation. Future work will focus on enhancing computational efficiency and validating our results.

Acknowledgments. This research has been conducted using the UK Biobank Resource under Application Number 80108 and supported by the Swiss National Foundation Sinergia Grant CRSII5_202276/1.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Azad, R., Aghdam, E.K., Rauland, A., *et al.*: Medical Image Segmentation Review: The success of U-Net (2022)
2. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., *et al.*: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of Cardiovascular Magnetic Resonance **20**(1), 65 (Feb 2018)
3. Bernard, O., Lalande, A., Zotti, C., *et al.*: Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging **37**(11), 2514–2525 (2018)
4. Chen, Z., Zhang, H.: Learning Implicit Fields for Generative Shape Modeling (Sep 2019)
5. Dannecker, M., Kyriakopoulou, V., Cordero-Grande, L., *et al.*: CINA: Conditional Implicit Neural Atlas for Spatio-Temporal Representation of Fetal Brains (2024)
6. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Commun. ACM **15**(1), 11–15 (Jan 1972)
7. Elfwing, S., Uchibe, E., Doya, K.: Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Networks **107**, 3–11 (Nov 2018)
8. Essakine, A., Cheng, Y., Cheng, C.W., *et al.*: Where do we stand with implicit neural representations? a technical and performance survey (Fev 2025)
9. Garzia, S., Rygiel, P., Dummer, S., *et al.*: Neural Fields for Continuous Periodic Motion Estimation in 4D Cardiovascular Imaging. In: Statistical Atlases and Computational Models of the Heart. Springer Nature Switzerland (2024)
10. Hoffer, E., Fine, S., Soudry, D.: On the Blindspots of Convolutional Networks (Jul 2018)
11. Isensee, F., Jaeger, P.F., Kohl, S.A.A., *et al.*: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021)

12. Kazerouni, A., Azad, R., Hosseini, A., Merhof, D., Bagci, U.: INCODE: Implicit Neural Conditioning with Prior Knowledge Embeddings (Oct 2023). https://doi.org/10.48550/arXiv.2310.18846
13. Khened, M., Kollerathu, V.A., Krishnamurthi, G.: Fully convolutional multi-scale residual DenseNets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers. Medical Image Analysis **51**, 21–45 (Jan 2019)
14. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. In: ICLR. San Diego (2015)
15. Larrazabal, A.J., Martínez, C., Glocker, B., Ferrante, E.: Post-DAE: Anatomically Plausible Segmentation via Post-Processing With Denoising Autoencoders. IEEE Transactions on Medical Imaging **39**(12), 3813–3820 (Dec 2020), conference Name: IEEE Transactions on Medical Imaging
16. Liao, W., Zhu, Y., Wang, X., Pan, C., Wang, Y., Ma, L.: LightM-UNet: Mamba Assists in Lightweight UNet for Medical Image Segmentation (Mar 2024)
17. Lipman, Y.: Phase Transitions, Distance Functions, and Implicit Neural Representations. In: Proceedings of the 38th International Conference on Machine Learning. pp. 6702–6712. PMLR (Jul 2021), iSSN: 2640-3498
18. Liu, R., Lehman, J., Molino, P., *et al.*: An Intriguing Failing of Convolutional Neural Networks and the CoordConv Solution (Dec 2018)
19. Lowes, M.M., Pedersen, J.J., Hansen, B.S., *et al.*: Implicit neural representations for registration of left ventricle myocardium during a cardiac cycle. In: STACOM. Springer Nature Switzerland (2024)
20. Maji, P., Mullins, R.: On the Reduction of Computational Complexity of Deep Convolutional Neural Networks. Entropy **20**(4), 305 (Apr 2018)
21. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy Networks: Learning 3D Reconstruction in Function Space (Apr 2019)
22. Muffoletto, M., Xu, H., Xu, Y., *et al.*: Neural Implicit Functions for 3D Shape Reconstruction from Standard Cardiovascular Magnetic Resonance Views. In: STACOM 2023. pp. 130–139. Springer Nature Switzerland (2024)
23. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. In: 2019 IEEE/CVF CVPR. pp. 165–174. IEEE (Jun 2019)
24. Perera, S., Navard, P., Yilmaz, A.: SegFormer3D: an Efficient Transformer for 3D Medical Image Segmentation (Apr 2024)
25. Ravi, N., Reizenstein, J., Novotny, D., *et al.*: Accelerating 3d deep learning with pytorch3d (2020)
26. Saragadam, V., LeJeune, D., Tan, J., Balakrishnan, G., Veeraraghavan, A., Baraniuk, R.G.: Wire: Wavelet implicit neural representations. In: Conf. CVPR (2023)
27. Stabinger, S., Peer, D., Rodríguez-Sánchez, A.: Arguments for the Unsuitability of Convolutional Neural Networks for Non–Local Tasks (Feb 2021)
28. Stolt-Ansó, N., McGinnis, J., Pan, J., *et al.*: NISF: Neural Implicit Segmentation Functions. In: MICCAI 2023. pp. 734–744. Springer Nature Switzerland (2023)
29. Sudlow, C., Gallacher, J., Allen, N., *et al.*: UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Medicine **12**(3) (Mar 2015)
30. Sørensen, K., Diez, P., Margeta, J., *et al.*: Spatio-Temporal Neural Distance Fields for Conditional Generative Modeling of the Heart. In: MICCAI 2024. pp. 422–432. Springer Nature Switzerland (2024)
31. Upendra, R.R., Wentz, B.J., Shontz, S.M., Linte, C.A.: A Convolutional Neural Network-based Deformable Image Registration Method for Cardiac Motion Estimation from Cine Cardiac MR Images. Computing in cardiology **47** (Sep 2020)

32. Wang, M., Zheng, D., Ye, Z., *et al.*: Deep graph library: A graph-centric, highly-performant package for graph neural networks (2019)
33. Wolterink, J.M., Zwienenberg, J.C., Brune, C.: Implicit Neural Representations for Deformable Image Registration. In: Proceedings of The 5th International Conference on Medical Imaging with Deep Learning. pp. 1349–1359. PMLR (Dec 2022)
34. Xing, Z., Ye, T., Yang, Y., *et al.*: SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation (Sep 2024)