Motion Correction Image Reconstruction using NeuralCT Improves with Spatially Aware Object Segmentation

Zhennong Chen[1], Kunal Gupta[1], Francisco Contijoch[1,2]

***Abstract*— NeuralCT [1] has been recently proposed as an implicit neural representation-based image reconstruction that can produce time-resolved images from CT sinograms and reduce motion artifacts, even when undergoing complex motions. NeuralCT does not require the prior motion model or estimation of object motion. Instead, it utilizes a network to implicitly represent the time-varying object boundary by singed distance function and optimizes the network via differentiable rendering. In this work, we modify the NeuralCT framework to reconstruct scenes that have multiple moving objects with distinct attenuation levels. We show that the performance of NeuralCT reconstruction depends on the quality of the initialization of the network (in this case, object segmentation in motion corrupted FBP image). We show how spatially aware object segmentation can improve motion-corrected reconstruction in moving objects with multiple attenuation levels despite high angular motion and complex topological changes.**

***Index Terms*— Motion Correction, Implicit Neural Representation, Differentiable Rendering**

I. INTRODUCDTION

Cardiac computed tomography (CT) has emerged as a noninvasive method to evaluate the coronary artery disease and assess the cardiac function. However, image quality can be limited by motion of cardiac structures. For example, even slow coronary vessel motion (~15mm/s) can cause significant blurring of vessels [2]. Improved hardware such as faster gantry rotation or dual source designs can avoid/reduce motion artifacts but further improvement appears limited by physical constraints. Machine learning algorithms [3], [4] have been used to correct motion artifacts in reconstructed images. However, current approaches are limited by the need as a true motion vector field (for training) is unavailable in clinical data.

Recently, implicit neural representations (INR) [5] have been used to improve reconstruction of medical images [6], [7]. Gupta et al. [1] recently developed an INR-based framework to improve reconstruction of CT data corrupted by object motion. This framework, called "NeuralCT", takes CT sinograms as the input and produces time-resolved images and was shown to correct motion artifacts. A key benefit of NeuralCT is that it does not impose a motion model nor require estimates of the object motion. An overview is shown in **Fig 1**.

NeuralCT utilizes a neural network to implicitly represents (neural representation) the moving object boundary via the signed distance function (SDFs). Concretely, the INR maps the spatiotemporal domain of the moving object (a point at a particular position and time) to SDF value domain (the real-time relative position of this point with respect to the object boundary). In this work, the neural representation was initialized using intensity-based segmentation of the motion corrupted Filtered Backprojection (FBP) result. The representation was then optimized via differentiable rendering (DR) [5], a technique used to identify the shape of an object that best "explains" its acquired projection. Thus, NeuralCT aims to identify the optimal time-varying shape of moving object such that the resultant projection agrees with the CT sinogram (ground truth projections). We emphasize that NeuralCT is not a learning task that requires training and testing datasets as such approaches depend on data driven priors which have a tendency to introduce bias in the reconstruction. Instead NeuralCT builds on work where INR problems are solved via optimization. In this case, NeuralCT performs optimized reconstruction by forward rendering the moving object to acquire projection estimates, calculating the error between projection estimates and the true sinogram, and then updating the reconstruction by backpropagating the error via gradient descent.

In the initial description of NeuralCT, Gupta et al. showed *high-quality* motion-correction for a single foreground object with high angular motion (up to 200° displacement per gantry rotation) as well as complex topological deformation [1]. However, clinical CT scans are not composed of a single foreground class. Therefore, the core contribution to this study is to extend NeuralCT to successfully correct motion artifacts in scenes with multiple (i.e., different intensity) moving objects. In particular, we observed that imaging multiple moving objects with different attenuations can limit the accuracy of intensity-based segmentation and consequently decrease the reconstruction performance. As a result, we incorporate spatial information into the segmentation and compare our improved reconstruction result with the initial NeuralCT and FBP.

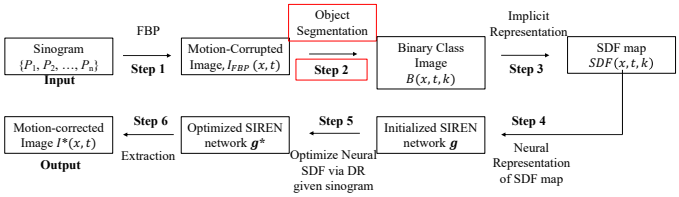

Fig. 1. NeuralCT framework. FBP = filtered backprojection, SDF = signed distance function, DR = Differentiable Rendering. In this study we proposed a new segmentation (red box) to extend NeuralCT to more complicated scenes.

This work is supported by NIH grant HL 143113.

[1]Department of Bioengineering, [1]Department of Radiology, University of California, San Diego, La Jolla, CA, 92093

II. METHODS

A. NeuralCT Framework

The NeuralCT framework is described in **Fig. 1** and the full description can be found in Ref [1]. The CT sinogram is the input and a time-resolved attenuation map $I^*(x,t)$ (motion-corrected image) is the output. The steps of the algorithm are:

Step 1: FBP images are created via backprojection of the sinogram P (comprised of a set of projections $\{P_1, P_2, \ldots, P_n\}$ for n gantry positions). This results in a series of motion-corrupted attenuation images $I_{FBP}(x,t)$.

Step 2: Segmentation *Seg* is used to identify different foreground objects from $I_{FBP}(x,t)$. The choice of *Seg* will be further discussed in Section II.B and II.C. Segmentation results in a binary time-varying images $B(x,t,k)$ where the k^{th} channel corresponds to the k^{th} foreground object.

Step 3: The time-varying scene of k binary images $B(x,t,k)$ is implicitly represented using the signed distance function (SDF). Specifically, $SDF(x,t,k)$ is generated to represent the position of the boundary as the signed distance of a point at location x in space at a particular time t to the boundary of the k^{th} object.

Step 4: For each location $x \in R^N$ where N is the number of spatial dimensions, the temporal evolution of an object's SDF was represented by Fourier Features (FF) using Fourier coefficients $\{A_0, A_1, \ldots, A_M, B_0, B_1, \ldots, B_M\}$:

$$SDF(x,t,k) \triangleq \frac{1}{M}\sum_{i=0}^{M} A_i(x,k)sin(2\pi\omega_i t) + B_i(x,k)cos(2\pi\omega_i t) \quad (1)$$

Here, ω_i are M randomly sampled frequencies. In our work, we approximated the SDF map $SDF(x,t,k)$ by a SIREN neural network [8] (an efficient framework to capture high frequency information). This neural network $\boldsymbol{g}(x,k;\boldsymbol{w})$, where $\boldsymbol{w}$ are weights in the network, was trained to output correct Fourier coefficients $\{A_i, B_i\}$ in Eqn. 1: $A_i(x,k;\boldsymbol{g})$ and $B_i(x,k;\boldsymbol{g})$.

The weights $\boldsymbol{w}$ were initialized randomly, and then updated by the standard gradient descent,

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \alpha\nabla\mathcal{L} \quad (2)$$

where $\mathcal{L} = \mathcal{L}_{SDF} + \lambda\mathcal{L}_E$. $\mathcal{L}$ is the total loss; $\mathcal{L}_{SDF}$ is the mean difference of the true SDF map (derived from FBP) versus $SDF(x,t,k;\boldsymbol{g})$ for all x,t,k; $\mathcal{L}_E$ is the Eikonal constraint computed as the mean value of absolute value of $\|\nabla_x SDF(x,t,k;\boldsymbol{g})\|_2$ -1 for all position x. λ is the regularization factor.

To conclude, after Steps 1-4 a SIREN neural network $\boldsymbol{g}$ is created that implicitly approximates the SDF map of the motion corrupted FBP images so $\boldsymbol{g}$ contains motion artifacts present after FBP.

Step 5: Differentiable Rendering (DR) is used to optimize $\boldsymbol{g}$ such that it represents a scene that is consistent with the acquired sinogram. Specifically, DR was used to identify the optimized shape $S^*$of an object that minimizes the projection loss $\mathcal{L}_P$ between the true projections (P_i) and the projections obtained via rendering of the estimated shape S:

$$\mathcal{L}_P = \sum_{i=0}^{n} |P_i - DR(S;\theta_i)| \quad (3)$$

Here, $DR(S;\theta_i)$ is the differentiable rendering operator; in CT, it represents the projection of an object shape S from "spatiotemporal attenuation space" $I(x,t)$ to the "projection space" P_i by the line integral of attenuation along the x-ray path u traversing through the scene at a gantry position θ_i:

$$DR(I(x,t);\theta_i) = \int_u I(x,t)\mathcal{R}_{\theta_i}(t)du \quad (4)$$

where $\mathcal{R}_\theta(t)$ is the time-varying rotation matrix describing the gantry rotation with angle θ_i.

Spatiotemporal attenuation maps $I(x,t)$ in Eqn. 4 were obtained from the SIREN SDF ($SDF(x,t,k;\boldsymbol{g})$) by first converting the SDF to an occupancy map $\mathcal{E}$ (where negative SDF value means the pixel is occupied) and then multiplying $\mathcal{E}$ with the object's attenuation $a(k)$ (Eqn. 5). $a(k)$ was approximated as the median attenuation of the k^{th} segmented object in the FBP image.

$$I(x,t) = \sum_k a(k) \times \mathcal{E}(SDF(x,t,k;\boldsymbol{g})) \quad (5)$$

Combining Eqn. 3-5, this approach enables the loss $\mathcal{L}_P$ to be defined as a differentiable function of $\boldsymbol{g}$**.** Additional loss terms – $\mathcal{L}_E$ (Eikonal constraint), $\mathcal{L}_{TVS}$ and $\mathcal{L}_{TVT}$ (total variances computed as the gradient of the SDF with respect to x and t) were added to constrain the result, leading to a total loss $\mathcal{L} = \mathcal{L}_P + \lambda_1\mathcal{L}_E + \lambda_2\mathcal{L}_{TVS} + \lambda_3\mathcal{L}_{TVT}$ where λ_1 to λ_3 serve as regularization weighting parameters.

Step 6: After optimization, the result $SDF(x,t,k;\boldsymbol{g}^*)$ was convert to the motion-corrected image $I^*(x,t)$ (i.e., the final product of NeuralCT reconstruction) via Eqn. 5.

B. NeuralCT with Intensity-based Segmentation

As outlined above, a key step in the NeuralCT framework is the initialization described in Step 4 where SIREN $\boldsymbol{g}$ aims to approximate the SDF map of the scene of interest. Gupta et al. [1] used a Gaussian Mixture Model (GMM) [9] that was solely based on the intensity histogram in $I_{FBP}(x,t)$. GMM fits a finite number of Gaussian distributions to the intensity histogram and assigns pixels with intensity from the same Gaussian distribution as the same class. After excluding the background, the top k classes with the most pixel were used to identify foreground objects. As shown in [1], this segmentation method, hereafter referred to as Seg_{GMM}, worked well in the scenes with a single foreground object – as it readily separates the object from the background, despite motion artifacts.

C. NeuralCT with Spatially Aware Segmentation

However, when Seg_{GMM} is applied to a scene with multiple moving objects, each with different attenuations, it becomes difficult to differentiate objects based solely on the intensity distribution. **Fig. 2** shows a failure of Seg_{GMM} when analyzing the FBP reconstruction of two moving dots with two different attenuations (top = 0.7, bottom = 0.2). Based on the histogram, GMM identifies the top two intensity values with the most

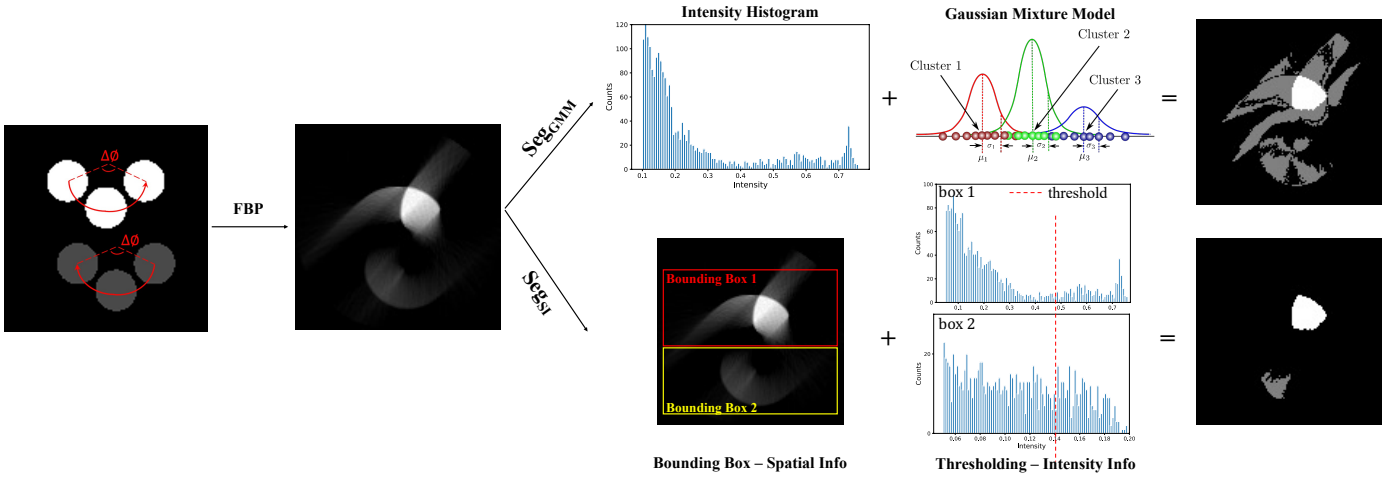

Fig.2. Two different object segmentation approaches used in NeuralCT. The first image shows the ground truth motion of two dots (top intensity = 0.7, moving from left to right, bottom intensity = 0.2, moving from right to left). $\Delta\emptyset$ is the angular displacement per gantry rotation. *Seg$_{GMM}$*: Gaussian mixture model incorrectly assigned the motion artifacts and the bottom dot as the same class. *Seg$_{SI}$*: Spatially aware segmentation utilized both spatial info (by setting bounding box in this example) and intensity info (thresholding) and led to correct detection of both top and bottom dots.

pixel counts from the distribution. However, this results in incorrect labeling of two dots as one bright foreground class and a second dimmer object spread throughout the image.

The core contribution of this study is to improve NeuralCT performance in the case of multiple intensity objects by resolving this segmentation error. We did so by applying a spatially-aware segmentation approach *Seg$_{SI}$* which incorporated both the Spatial (S) and Intensity (I) information of each object in the FBP image. *Seg$_{SI}$* aims to assign different classes to objects with different spatial positions and be aware of the different intensities between the real object and the motion artifacts. This can be achieved using various approaches such as Region-Of-Interest (ROI) definition plus thresholding or data-driven methods (e.g., deep learning segmentation). Here, we focus on demonstrating that this improvement in segmentation leads to improvements in NeuralCT performance. In **Fig. 2**, we show a simple approach to add spatial information. Specifically, bounding boxes were used to guide thresholding-based segmentation. Each bounding box was defined to only contain one moving dot such that we assigned one individual class to each box. In the box, we defined an intensity threshold = $\gamma \times I_{max}$ where I_{max} is the maximum intensity in the scene in each box to capture the real object. $\gamma = 0.7$ was set empirically.

Given that artifacts will always be present in the initial FBP images, we *highlight* here that the goal with this new segmentation is not to achieve a perfect segmentation but rather to provide a segmentation that is not so poor that it precludes improvement by the NeuralCT framework. We hypothesize that by improving the initial segmentation, we will avoid overt failures and improve image quality obtained with NeuralCT.

III. Experiments and Results

We performed two experiments to demonstrate the impact of the segmentation on the subsequent result and evaluate the improvement associated with our new segmentation approach.

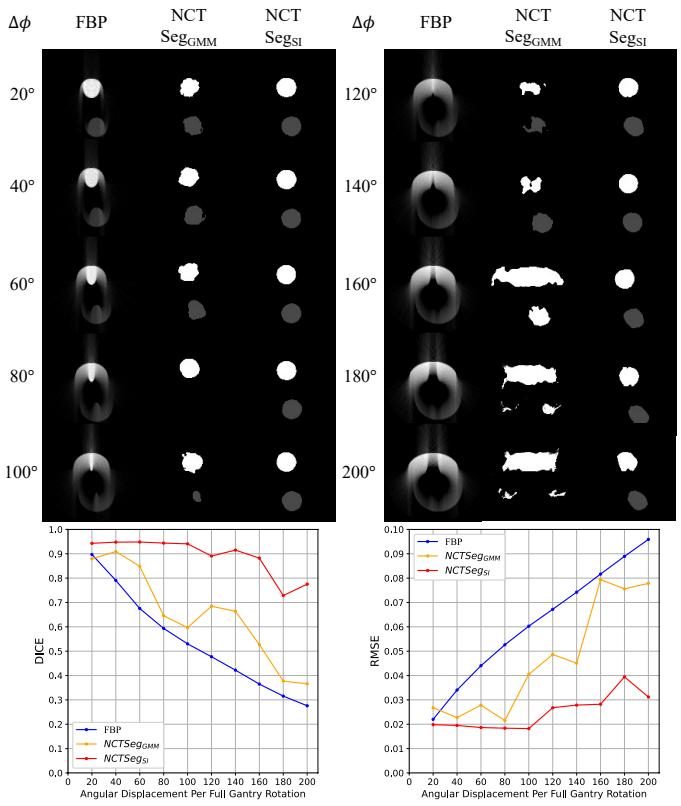

Fig. 3. NCT-*Seg$_{SI}$* accurately depicts the moving dots with two attenuations and high angular displacements. FBP suffers from motion artifacts for all $\Delta\emptyset$; NCT-*Seg$_{GMM}$* failed the reconstructions for high $\Delta\emptyset$ (>60); Only NCT-*Seg$_{SI}$* maintained high-quality motion-corrected reconstruction for all $\Delta\emptyset$ with higher DICE and lower RMSE when compared with FBP and NCT-*Seg$_{GMM}$*. $\Delta\emptyset$ = angular displacement per gantry rotation.

Experiment 1: Angular Displacement of Two Dots

As shown in **Fig. 2**, two circular dots which translate with angular displacement $\Delta\emptyset$ per full gantry rotation were imaged. The two dots had different attenuation levels (top = 0.7, bottom = 0.2), mimicking the difference between contrast-enhanced vessels and the myocardium in cardiac CT. Background = 0. The image resolution was set to 128×128 and a parallel beam

CT geometry was used with 720 gantry positions per rotation. Two NeuralCT frameworks were then evaluated – intensity-based segmentation (NCT-*Seg*$_{GMM}$) and spatially aware segmentation (NCT-*Seg*$_{SI}$) – across a range of ΔØ (from 20° to 200° per gantry rotation). Performance was evaluated using root-mean-square-error (RMSE) and DICE coefficients relative to the ground truth image.

As shown by the images and metrics in **Fig. 3**, FBP motion artifacts increased at higher ΔØ. Reconstruction with NCT-*Seg*$_{GMM}$ was limited when ΔØ > 60°. In contrast, NCT-*Seg*$_{SI}$ maintained high-quality motion-corrected reconstructions for all ΔØ and achieved low RSME (<0.028) and high (>0.89) DICE for ΔØ up to 160°.

Experiment 2: Complex Deformation of Letters

In experiment 2, we evaluated the ability of NCT-*Seg*$_{SI}$ to improve reconstruction of scenes with complex topological changes. As shown in **Fig. 4**, in this case, we simulated CT imaging during transformation of letters. The top letter transformed from "A" to "B" to "A" (attenuation = 0.7) while the bottom letter transformed from "B" to "A" to "B" (attenuation = 0.4). NCT-*Seg*$_{SI}$ (red line) significantly reduced the severity of artifacts observed with FBP (blue) and NCT-*Seg*$_{GMM}$ (orange), especially during transformation periods (2nd-3rd and 5th-6th columns). Quantitatively, median RMSE of NCT-*Seg*$_{SI}$ (median = 0.050 [0.042-0.061]) was significantly lower ($p<0.05$) than NCT-*Seg*$_{GMM}$ (0.090 [0.076-0.096]) and FBP (0.069 [0.047-0.085]). Median DICE for NCT-*Seg*$_{SI}$ (0.89 [0.86-0.93]) was significantly higher ($p<0.05$) than NCT-*Seg*$_{GMM}$ (0.72 [0.69-0.76]) and FBP (0.72 [0.64-0.87]). Lastly, NCT-*Seg*$_{SI}$ increased the percentage of the frames with RMSE< 0.05 (NCT-*Seg*$_{SI}$: 45.7%, NCT-*Seg*$_{GMM}$: 0%, FBP: 28.0%) as well as with DICE > 0.85 (NCT-*Seg*$_{SI}$: 89.6%, NCT-*Seg*$_{GMM}$: 0%, FBP: 27.6%).

IV. Summary

Reconstruction of moving scenes using a neural implicit representation-based framework (NeuralCT) can improve image quality the need for a prior motion model or estimation. Here, we show that when imaging scenes with multiple moving objects, performance of NeuralCT can be limited by poor segmentation of motion-corrupted FBP images. Using a spatially aware object segmentation method that incorporates both spatial and intensity information can result in an NeuralCT solution which maintains high reconstruction performance for moving objects with multiple attenuation levels despite high angular motion and complex topological changes.

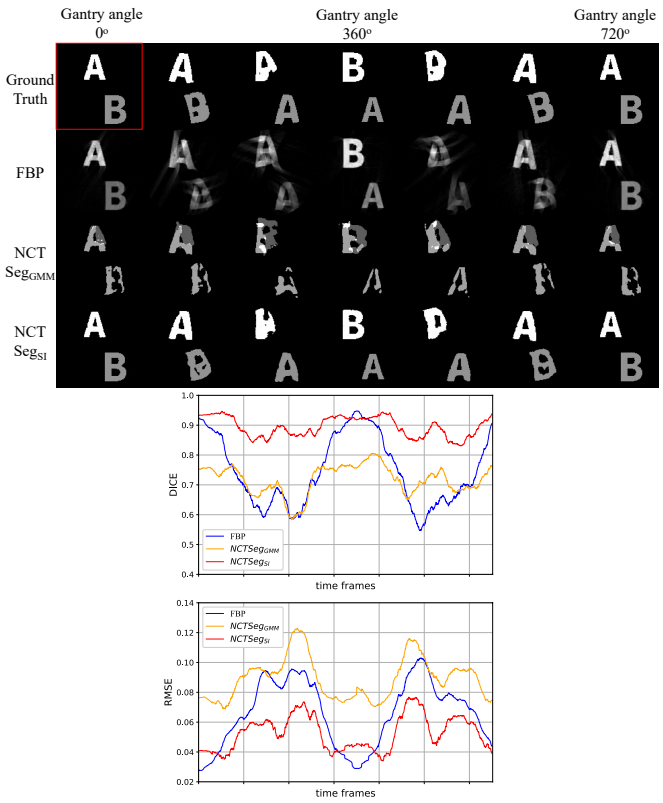

Fig. 4. NCT-*Seg*$_{SI}$ accurately depicts the complex topological change with multiple attenuations. The ground truth image (red box) contains two letters that transform over two gantry rotations. Seven frames including three stationary phases (column 1,4, 7) and four intermediate transformation phases (column 2-3, 5-6) are displayed. Both reconstructed images and the quantitative metrics indicates that NCT-*Seg*$_{SI}$ improved the imaging of a complex scene.

References

[1] K. Gupta, B. Colvert, and F. Contijoch, "Neural Computed Tomography,", Jan. 2022, Available: http://arxiv.org/abs/2201.06574

[2] Z. Chen *et al.*, "Precise measurement of coronary stenosis diameter with CCTA using CT number calibration," *Med. Phys.*, vol. 46, no. 12, pp. 5514–5527, Dec. 2019, doi: 10.1002/mp.13862.

[3] T. Lossau *et al.*, "Motion estimation and correction in cardiac CT angiography images using convolutional neural networks," *Comput. Med. Imaging Graph.*, vol. 76, p. 101640, Sep. 2019, doi: 10.1016/j.compmedimag.2019.06.001.

[4] Y. Ko, S. Moon, J. Baek, and H. Shim, "Rigid and non-rigid motion artifact reduction in X-ray CT using attention module," *Med. Image Anal.*, vol. 67, p. 101883, Jan. 2021, doi: 10.1016/j.media.2020.101883.

[5] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang, "NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction," Dec. 2021 Available: http://arxiv.org/abs/2106.10689

[6] L. Shen, J. Pauly, and L. Xing, "NeRP: Implicit Neural Representation Learning with Prior Embedding for Sparsely Sampled Image Reconstruction," Aug. 2021. Available: http://arxiv.org/abs/2108.10991

[7] Q. Wu *et al.*, "An Arbitrary Scale Super-Resolution Approach for 3-Dimensional Magnetic Resonance Image using Implicit Neural Representation," Oct. 2021. Available: http://arxiv.org/abs/2110.14476

[8] V. Sitzmann *et al.*, "Implicit Neural Representations with Periodic Activation Functions," Jun. 2020. Available: http://arxiv.org/abs/2006.09661

[9] D. Reynolds *et al.*, "Gaussian Mixture Models," *Encyclopedia of Biometrics*, 2009, pp. 659-664. doi: 10.1007/978-0-387-73003-5_196.